BOURNEMOUTH UNIVERSITY

An Intelligent Destination Recommendation System for Tourists

Pree Thiengburanathum

Doctor of Philosophy

Department of Computing and Informatics, Faculty of Science and Technology

Bournemouth University, U.K.

March 2018

Acknowledgements

I would like to express my special appreciation and thanks to my main supervisor, Professor Hongnian Yu, for encouraging my research, for his support during the conduct of this research and for allowing me to grow as a research scientist. I owe my deepest gratitude to my second supervisor, Dr Shuang Cang. She has guided my steps through these years, and without her I would not have been able to achieve so much.

I would like to dedicate this thesis to my father and mother, as well as my brother and sister. During my PhD study, I would like to thank my girlfriend for her overwhelming care, encouragement and giving me a lot of support, even though we are many miles apart. Also, many thank to my colleagues in BU research groups and my roommates. I shared many impressive and wonderful moments with them.

Last but not least, I would like to express my gratitude to Bournemouth University, and my parents, for providing me with financial support to pursue my PhD study.

Abstract

Choosing a tourist destination from the information available is one of the most complex tasks for tourists when making travel plans, both before and during their travel. With the development of a recommendation system, tourists can select, compare and make decisions almost instantly. This involves the construction of decision models, the ability to predict user preferences, and interpretation of the results.

This research aims to develop a Destination Recommendation System (DRS) focusing on the study of machine-learning techniques to improve both technical and practical aspects in DRS. First, to design an effective DRS, an intensive literature review was carried out on published studies of recommendation systems in the tourism domain.

Second, the thesis proposes a model-based DRS, involving a two-step filtering feature selection method to remove irrelevant and redundant features and a Decision Tree (DT) classifier to offer interpretability, transparency and efficiency to tourists when they make decisions. To support high scalability, the system is evaluated with a huge body of real-world data collected from a case-study city. Destination choice models were developed and evaluated. Experimental results show that our proposed model-based DRS achieves good performance and can provide personalised recommendations with regard to tourist destinations that are satisfactory to intended users of the system.

Third, the thesis proposes an ensemble-based DRS using weight hybrid and cascade hybrid. Three classification algorithms, DT, Support Vector Machines (SVMs) and Multi-Layer Perceptrons (MLPs), were investigated. Experimental results show that the bagging ensemble of MLP classifiers achieved promising results, outperforming baseline learners and other combiners.

Lastly, the thesis also proposes an Adaptive, Responsive, Interactive Model-based User Interface (ARIM-UI) for DRS that allows tourists to interact with the recommended results easily. The proposed interface provides adaptive, informative and responsive information to tourists and improves the level of the user experience of the proposed system.

List of Publications

1. Thiengburanathum P., Cang S., Yu H., An Overview of Travel Recommendation System, The *IEEE* 22th International Conference on Automation and Computing 2016 (*published*)
2. Thiengburanathum P., Cang S., Yu H., A Decision Tree-based Destination Recommendation System for Tourist, The *IEEE* 21th International Conference on Automation and Computing, 2015 (*published*)
3. Thiengburanathum P., Cang S., Yu H., A Model-based Destination Recommendation System to Assist Travellers, Information and Management, 2017 (*Under review*)
4. Thiengburanathum P., Cang S., Yu H., Ensemble of Classifiers for a Destination Recommendation System, Decision Support System, 2017 (*Under review*)
5. Thiengburanathum P., Cang S., Yu H., Design and Implementation of a User Interface for a Model-based Destination Recommendation System, Travel Management, 2017 (*Preparing for submission*)

Table of Contents

List of Figures

List of Tables

Nomenclature

A	Recommended Attraction
AC	Accommodation
ACO	Ant Colony Optimization
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
AJAX	Asynchronous JavaScript and XML
ANN	Artificial Neural Network
API	Application Program Interface
AR	Augmented Reality
ARIM-UI	Adaptive, Responsive, Interactive Model-based User Interface
AT	Activities
AUC	Area Under the Curve
AUROC	Area Under Receiver Operatic Characteristic
AVG	Average
BHT	Baht
BN	Bayesian Network
CA	Correlation Analysis
CART	Classification and Regression Trees
CBF	Content-based Filtering
CBR	Case-based Reasoning
CF	Confidence Factor
CHAID	Chi-Squared-Automatic-Interaction Detection
CLF	Classification
CO	Cosine Similarity
COF	Collaborative Filtering

CSF	Constraint-based Filtering
CT	Clustering
CTF	Context-based Filtering
D	Recommended Destination
DF	Demographic Filtering
DM	Data Mining
DMT	Decision-Making Theory
DRG	Dynamic Route Guidance
DRS	Destination Recommendation System
DS	Descriptive Statistic
DSS	Decision Support System
DT	Decision Tree
EBM	Engel-Blackwell-Miniard
ERD	Entity Relationship Diagram
FA	Factor Analysis
FL	Fuzzy Logic
FN	False Negative
FP	False Positive
GA	Genetic Algorithm
GDP	Gross Domestic Product
GPS	Global Positioning System
GUI	Graphical User Interface
H	Heuristic
HCI	Human-Computer Interaction
HF	Hybrid Filtering
HTML	Hypertext Mark-up Language
IA	Intelligent Agent

ICT	Information Communication and Technology
IDRS	Intelligent Destination Recommendation System
ILS	Iterated Local Search
IoT	Internet of Things
ITAS	Intelligent Travel Attractions System
IUI	Intelligent User Interface
JFS	Java Server Faces
JSON	JavaScript Object Notation
KF	Knowledge-based Filtering
KKT	Karush-Kuhn-Tucker
KNN	K-Nearest Neighbour
KO	KnockOut java script library
MAR	Missing At Random
MAS	Multi-Agent System
MCDM	Multi-Criteria Decision-Making
MI	Mutual Information
ML	Machine Learning
MLP	Multi-Layer Perceptron
mRMR	Minimum-Redundancy, Maximum-Relevancy
mRMR MID	mRMR Mutual Information Differences
MVC	Model View Controller
MVVM	Model-View-View Model
NMIFS	Normalized Mutual Information Feature Selection
NoSQL	Not only SQL
ON	Ontology
OP	Orienteering Problem
OWL	Web Ontology Language

PHP	Personal Home Page
POI	Points of Interest
PTPS	Personalized Travel Planning System
RBF	Radial Basis Function
RDBMS	Relational Database Management System
RDF	Resource Description Framework
RE	Restaurants
RFID	Radio-Frequency Identification
RMSE	Root-Mean-Square Error
RO	Route recommendation
ROC	Receiver Operating Characteristic
RS	Recommendation System
SCG	Scaled Conjugate Gradient
SRDA	Survey Research Data Archive
SRM	Schedule Reasoning Method
STD	Standard Deviation
SVM	Support Vector Machine
T	Recommended Tour Package
TDIDT	Top-down Induction of DT
TIDP	Tourist Itinerary Design Problem
TM	Transport Mode
TN	True Negative
TOP	Team Orienteering Problem
TOPTW	Team Orienting Problem with Time Windows
TP	True Positive
TPL	Trip Planning
TRS	Tourist/Travel Recommendation System

TSP	Travelling Salesman Problem
TSPTW	Travelling Salesman Problem with Time Windows
TTDP	Tourist Trip Design Problem
UI	User Interface
UML	Unified Modeling Language
VSM	Vector Space Model
WPF	Windows Presentation Foundation
WTO	World Trade Organization
XML	eXtensiveble Mark-up Language

Chapter 1 Introduction

This chapter describes the background and motivation of this research. It explains the research questions, aims, and objectives. The contribution and innovative features of this research are presented. The structure of this thesis is presented at the end of this chapter.

Tourism is extremely important globally, contributing 10% to the world economy in 2015 and projected to grow to an estimated 10.3% average over the next decade (World Travel and Tourism Council, 2015). The number of tourists worldwide has increased rapidly. Over the same 10-year period, Southeast Asia is expected to be the fastest-growing region regarding travel and tourism's contribution to a country's or a region's Gross Domestic Product (GDP). Of particular note, Thailand, Indonesia, Singapore, and Myanmar were the countries identified as the most attractive tourist destinations in 2013 (Economic Impact of Travel & Tourism 2014 Annual Update: Summary, 2014).

Over the last decade, Thailand's tourism industry has boomed, with international tourist arrivals doubling over the past nine years (see Fig. 1.1). In 2013 alone, international arrivals increased by 18.8%, the second highest rate among the top-ten most visited destinations in the Asian and Pacific regions. Overall, Thailand was the 10th most visited destination worldwide, and attracting 26 million international tourists, and growing by 18.76% over the previous year (United Nations World Tourism Organization, (UNWTO)). Increasing both tourist numbers (international and domestic) and the benefits of tourism are a primary objective of the Thai government. In 2013, tourism generated 1.79 trillion BHT ($55.49 bn) in revenue for Thailand, an increase of 13% over the previous year (Thailand Annual Report, 2013).

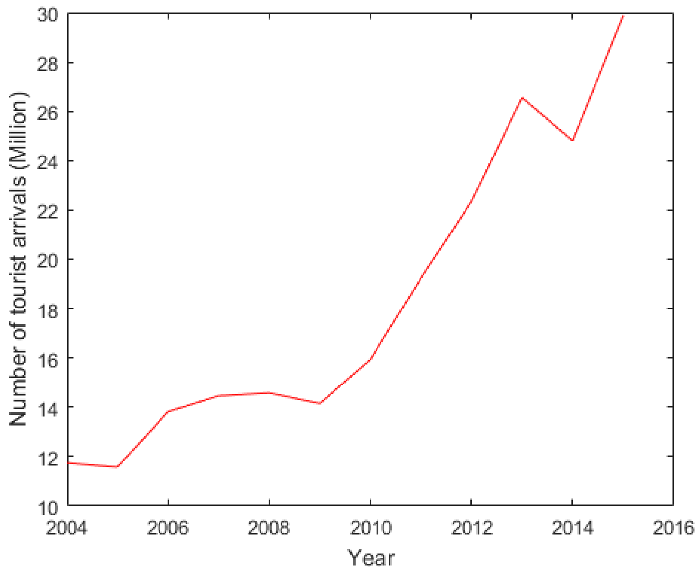


Figure 1.1 Number of international tourists arriving in Thailand in 2004-2015

Figure 1.2 shows tourist arrivals by country of origin for Thailand in 2013 (Thailand Annual Report, 2013). The top five countries' visits to Thailand are from Malaysia, China, Japan, Russia, and South Korea. It can be seen that of these five countries, the largest numbers are from Malaysia and China.

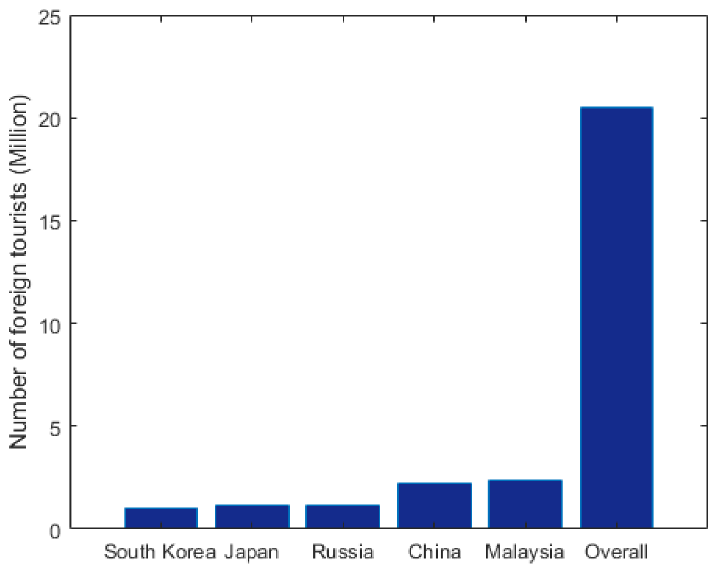


Figure 1.2 Numbers of international tourists from specific countries arriving in Thailand (millions) in 2013

With the boom in tourism over the last decade, information sources play an important role for tourists when making decisions and selecting destinations. The Internet is now considered to be the tourists' main information source for information on products and services (Pantano and Pietro, 2013). However, the sheer volume of data on the Internet has made it difficult for tourists to process information, whether in pre-trip planning or when making choices during travel. The travel-planning problem is highly complex, time-consuming, and dynamic as there are many factors involved in the decision-making process. Some of the factors involved in travel-planning include travel budget, number of nights one intends to stay at a given destination, food quality, the number of individuals travelling, transport mode, leisure activities, weather etc. (Pan and Fesenmaier, 2002).

1.1 Motivation

Recently, tourism has benefited substantially from Information and Communications Technology (ICT), and especially from Internet technology and its applications (Pitoska, 2013). Decision support tools, also known as Recommendation Systems (RSs), have been developed to address these concerns. In the tourism field, they are referred to as Tourism Recommendation Systems (TRSs). Tourists and tourism providers can search, select, compare and make decisions almost instantly, and more efficiently than ever.

Due to the enormous amount of heterogeneous information available on the Internet and through other information sources, TRSs can act as information filters. Selecting appropriate tourist services to match user preferences is one of the most complex tasks a tourist faces when planning a visit to an unfamiliar city. Even though search engines provide lists of tourism services, tourists are still overwhelmed with the information on offer. TRSs can be utilised extensively as a means of reducing information overload for tourists.

TRSs can help assist tourists to travel independently to an unfamiliar city, especially as regards searching, selecting and comparing tourism services. Not only can TRSs help travellers when planning their trip, but also during and after a trip, thanks to mobile and wireless communication. A well-developed TRS can suggest appropriate tourism services to tourists without interfering with their privacy and suggest other travel-related products to them.

Moreover, TRSs can help promote tourism in a city as well as market the tourist destinations. This will have a great impact on a city or country's tourism, especially tourism services, marketing and government marketing strategies. With regard to tourism-related companies, in order to be competitive and profitable and to make life easier for tourists, the tourism industry and travel agencies need to make use of TRSs to ensure they offer excellent services to tourists and thus improve their business.

To date, most TRSs have focused on estimates for choosing destinations, activities, attractions and tourism services (e.g. restaurants, hotels and transportation) based on users' preferences and interests. With regard to technical aspects, these TRSs only provide filtering, sorting and basic matching mechanisms between items and a user's hard and soft constraints. In order to assist tourists practically, a TRS needs to become 'intelligent' with regard to certain technical aspects, such as scalability, transparency, recommendation accuracy and validation methods; and certain practical aspects, such as user acceptance and usability – all of which should be taken into consideration when designing a system. Additionally, an effective TRS should strike a balance between practical and technical aspects. This research focuses on TRSs that recommend destinations to tourists, also known as Destination Recommendation Systems (DRSs).

To develop a successful DRS that effectively

addresses both practical and technical aspects, several challenges need to be overcome.

1. Enhance tourist decision-making

   One of the challenges for a DRS is to enhance the tourist's decision-making process. It is important for tourists to understand how recommendations that are generated by the system have been determined. To achieve this, it requires a deep understanding of tourists' decision-making and development of novel models for their information search process (Gretzel et al., 2012). Understanding the tourist decision-making process captures the attention of both researcher and practitioner.

2. Reduce users' efforts and preserve their privacy

   Uncertainties involved in the information search stage of a tourist's decision process need to be eliminated. Particularly, any user input that is insignificant to the search process should be excluded. Including more parameters in the system, may increase model complexity, decrease DRS recommendation performance, and decrease the level of user satisfaction with the system.

3. Increase recommendation performance

   Many existing TRSs only evaluate the system using an accuracy rate, and many of them do not have any evaluation method (Fouss and Saerens, 2008). This research focuses on improving recommendation performance using classification accuracy rate along with other aspects, such as confusion matrix, precision, recall, F-measure, ROC, and AUC after the data set has been pre-processed (i.e. after the processes of cleaning, transforming and feature selection). To increase recommendation performance during the model-building process is challenging, and there are many techniques for increasing the performance of the recommendation system. In this thesis, we focus on the investigation of classification algorithms, optimizing parameters, and combining classifiers. First, an investigation of multiple-classification algorithms needs to be conducted as some algorithms are better suited to our data sets than others. Different kinds of cross-validation methods can be applied to make sure that the model is not overly complex and that it is generalised enough for unseen data. Second, tuning hyper-parameters for classification algorithms is a crucial process for improving predictive accuracy. However, tuning hyper-parameters is considered an expensive and time-consuming process. These hyper-parameters play an important role in predictive results, and the goal is to find optimal ones. Third, the ensemble learning method has been proven to give better results, as the technique fuses the results of multiples of base classifiers (Saleh et al., 2017). The main challenge here is that it is not known which combination method will give better predictive results. Therefore, we need to construct a study to compare the results of two types of ensemble learning methods, including methods that combine multiple models of a similar type (e.g. bagging and boosting) and methods that combine multiple models of various types (e.g. vote classification).

4. Improve user satisfaction

   Another challenge in DRS development is related to improving user satisfaction with the system. For example, when a traveller uses a DRS on either from a mobile or desktop platform, they expect the user interface to be informative, responsive and interactive (Chu et al., 2001). Previous DRSs have improved the interaction between the user and the DRS. This expands the user experience and increases the level of satisfaction and enjoyment when searching for a destination (Buhalis and Law, 2008).

This thesis proposes an innovative DRS to respond to the afore-mentioned challenges. The proposed DRS is considered to be a model-based destination recommendation system. The supervised machine learning process, which runs offline, involves data acquisition, data pre-processing, data analysis and results interpretation.

1.2 Aims and objectives

The main aim of this research is to develop a model-based Destination Recommendation System (DRS) to assist tourists before they travel, or during their travel, to an unfamiliar city. The following objectives have been set in order to help achieve the mentioned main aim:

1. To review state-of-the-art Travel Recommendation Systems (TRSs) in the literature and identify research challenges and gaps (Chapter 2).
2. To design and develop a questionnaire for data collection for a case-study city (Chapter 3).
3. To identify features and data-processing techniques for the proposed system (Chapter 4).
4. To develop destination-choice models and evaluate them through the use of a variety of evaluation methods (Chapter 4).
5. To investigate and validate ensemble-learning techniques for destination classification (Chapter 5).
6. To develop an interactive and adaptive user interface for the proposed DRS (Chapter 6).

The vertical alignment of the machine-learning process flow follows the research objectives above, and its contribution and novelty are demonstrated in Figure 1.3.

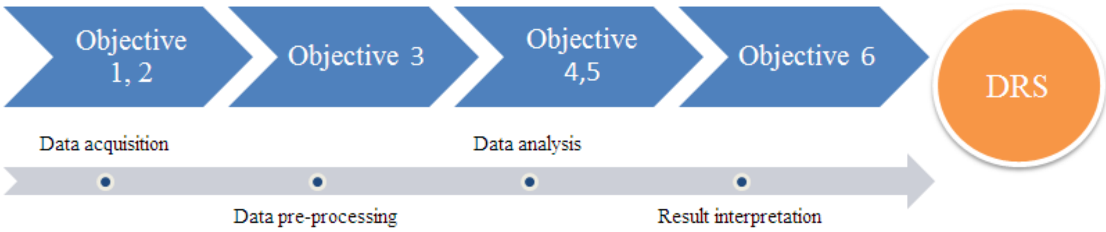


Figure 1.3 Flow-chart of research objectives

While previous DRSs have been developed, they have not taken technical issues, such as system accuracy, and practical issues such as usability and user acceptance (i.e. the system should be suited to real-world circumstances and usage), into consideration. To address the lack of technical and practical issues associated with DRSs, the following main research questions have been formulated:

RQ 1. How do you identify the preferred destination for a tourist using multiple human behaviour factors through a questionnaire?

RQ 2. Which set of factors plays an important role in making destination recommendations for tourists? Does using multiple factors help to improve recommendation accuracy? Do travel-motivation factors contribute to increasing the level of recommendation accuracy?

RQ 3. How can a tourist’s decision-making process be understood when selecting their preferred destination?

RQ 4. How can a user’s efforts be reduced, while still maintaining the same degree of recommendation performance and increasing the level of user satisfaction in the decision-making process when selecting a destination?

RQ 5. How can an optimal decision model be constructed when using multiple sets of factors with multiple tourist destinations?

RQ 6. How can the recommendation accuracy rate be improved using only relevant and non-redundant factors?

RQ 7. How can a tourist be helped to interpret and interact with the constructed decision model(s)?

1.3 Case study

In order to develop a successful and effective DRS (i.e. a DRS that has both technical and practical impact), a large-scale data set of human behaviour is needed to inform, e.g. a system design (Gretzel et al., 2012). In this study, five sets of factors that influence the tourist decision-making process, when selecting destinations, are investigated. We investigated trip characteristics, tourist expenditure behaviour, tourist behaviour, travel motivations and tourist socio-demographic information to understand how a tourist makes a decision when selecting a destination. There are no secondary data that can be used for this research. This study selected Chiang Mai, Thailand, as its location, due to its reputation as an internationally well-known tourist destination, and used the questionaire as the data collection method. Twenty popular tourist destinations in the city of Chiang Mai were used to evaluate the proposed DRS. The city has many religious sites, museums, art galleries and natural attractions, and it is host to many important festivals. According to TripAdvisor,[1] Chiang Mai was one of the top-25 best destinations in the world in 2014. Its cultural and historical sites were the main reason for Chiang Mai being added to a tentative list of world heritage sites by UNESCO in July 2015. Moreover, it has was ranked second in a list of the world's best cities in Travel + Leisure World's Best Awards 2016.

1.4 Summary of contributions

In this thesis the contributions to knowledge in the RS, TRS and DRS fields are as follows:

1. An extensive amount of literature based on various published studies of post-2008 TRSs has been produced from significant online databases and publications (Thiengburanathum et al. 2016.). This study conducted a comprehensive and systematic review of TRS techniques and their application to the tourism domain using the proposed review classification scheme. This was done through a study of the e-tourism services that TRSs currently provide, a review of the latest ICT concepts that have been applied to previous TRSs, the incorporation of possible research trends (e.g. group-based recommendations, integration of heterogeneous online information, tourist itinerary design problems, etc.), methodologies to improve the level of personalized services, and consideration of the trends in challenges that affect the direction and future development of TRSs. Research challenges and classification results will contribute significantly to knowledge in the TRS field.

2. The thesis proposes a novel model-based DRS framework that helps tourists to understand their decision-making processes using a machine-learning method. This involves a two-step feature-selection method based on Mutual Information (MI) in the data pre-processing phase, as well as a Decision Tree (DT) in the classification phase. Recommendation results were provided by a DT classifier. We selected C4.5 as the

classification algorithm, as that it offers several benefits including interpertablility, so that it provides meaningful decision rules that explain the importance of each feature and the relationship between them. Tourists can, therefore, understand how recommendations have been made. The DT-based recommendation system has never been applied to the TRS domain. An analysis of the C4.5 algorithm for different tourists' preferred destination choices was carried out. To reduce the complexity of the model and to inprove the classification accuracy rate, the data set was divided into eight destination categories using tourism-domain expert knowledge. Eight optimal destination choice models that offer explainability and transparency (i.e. a user can understand why an item is recommended to him/her) were found for each of the tourist destination categories. We believe this is the first study that has used a DT to represent tourists' destination choices.

3. Improve recommendation performance using hybrid recommendation: We propose a novel hybrid DRS that combines three filtering techniques: collaborative filtering, content-based and knowledge-based filtering. The purpose of the hybrid recommendation technique is to achieve the best performance and overcome the weaknesses/ disadvantages of one technique by complementing it with the strengths/ advantages of another technique. This technique combines two or more recommendation techniques to achieve better performance (Burke, 2002). A hybrid recommendation technique is more robust and efficient than a basic recommendation approach, such as a stand-alone user-based collaborative one (Badaro et al., 2013). The experimental results confirmed that our DRS performed well and was capable of providing personalised recommendations, with regard to tourist destinations, that are satisfactory to tourists.

4. A DRS that understands users and is scalable with real-world and large human behaviour data sets: Fully understanding the user is a crucial component of building any success recommendation system (Ricci et al., 2011). Large-scale human behavioural data are needed to inform such a design (Gretzel et al., 2012). In this research, 4,000 questionnaires related to human behaviour data were distributed and collected from participants in the top-20 favourite tourist destinations in Chiang Mai, Thailand. The factors that influenced tourists when selecting destinations were identified from

previous studies to construct destination-choice models. The sets of factors included tourist behaviours, tourist expenditure behaviours, social-demographic information, travel motivation, and trip characteristics. This data set is considered highly significant for the purposes of research on DRSs, TRSs and in the field of tourism generally. Destination-choice models were constructed from the data set. A model-based recommendation system can quickly generate a recommendation for a user and is more scalable than a user-based approach (Ghazanfar and Prugel-Bennett, 2010).

5. Reduce users' effort and preserve user's privacy: most existing TRSs require a lot of input to the system to generate acceptable recommendation accuracy. In an attempt to make a better recommendations, previous RSs have needed to elicit as much input as possible from users. Chu et al. (2012) found that feature selection helps in improving classification accuracy if using correct prior knowledge and methods. This study uses experts' tourism domain knowledge combined with a two-step feature selection method, based on Mutual Information (MI), to eliminate unnecessary input to the system while maintaining reasonable recommendation accuracy, which in turn improves the user experience. To the best of our knowledge, the feature-selection technique has not been used in the TRS field before. In this study we have applied a two-step filtering method to select the smallest number of variables that can map output classes. In the first filtering, irrelevant features are removed by applying the Max-Relevance feature selection algorithm based on MI. The second filtering method involves removing redundant features. Additionally, two well-known feature-selection algorithms based on MI are used, namely, Minimum-Redundancy Maximum-Relevancy (mRMR) and Normalized Mutual Information Feature Selection (NMIFS). Moreover, DT helps to reduce the search time, as that DT provides lists of recommendation items at its leaf nodes.

6. We conducted a comparative study of different classification algorithms for destination choice. To improve the level of recommendation performance of the system, three types of classifiers were investigated for this data set including DT, Support Vector Machines (SVMs) and Multi-Layer Perceptrons (MLPs). A recommendation performance comparison and an analysis of each of the classifiers in each of the data sets were carried out. The results demonstrate the importance of choosing optimal classfiers for a tourist's preferred destination-choice classification.

7. An ensemble learning method for a destination recommendation system: this is based on the use of combination rules and ensemble algorithms. Ensemble learning has been successfully applied in many applications, including face recognition, computer-aided medical diagnosis, text categorization etc. (Zhou, 2015). In this research we investigated the performance of seven combination rules to fuse individual classifiers and two popular ensemble learning methods: bootstrap aggregating (bagging) and boosting. The results indicated that all the ensemble classifiers achieve equal or higher classification accuracy than using only an individual classifier.

8. An adaptive, responsive and interative user interface: Chu et al., (2001) claim that a website needs to be interactive, responsive, informative and attractive to tourists. To increase the level of satisfaction with the system, along with letting tourists utilise the system efficiently on different devices, an Adaptive, Responsive, Interactive Model-based User Interface (ARIM-UI) for the DRS was proposed. The integrated interface has three main functionalities: adaptability, interactivity, and responsiveness. Emphasis was placed on the handling complexity of the DRS user interface, which is one of the most challenging tasks in Web semantics. We combined two language parsers (Rule2XML and XML2Jason), JQuery, Model-View-ViewModel (MVVM) design pattern framework and Bootstrap style for a responsive and interactive Web interface. Our proposed UI can automatically map the DT C4.5 model as an output from the proposed DRS. Based on user interaction with the interface our system can automatically generate dynamic new selection radio boxes, drop-down list menus and new information on the interactive and responsive Web user interface.

1.5 Thesis outline

Apart from the introduction, this thesis consists of six chapters and five appendices. The thesis is organised as follows:

Chapter 2 conducts a literature review and provides relevant background on the recommendation system in the tourism domain. Next, Information and Communication Technologies (ICT) used in recent TRSs and TRS applications are presented followed by a discussion of the current challenges and trends of TRSs. The research gap is specified at the end of the chapter.

Chapter 3 describes the research approach and the system architecture of the proposed DRS. The chapter covers the strategies used to collect the data sets and their related characteristics. Machine-learning methodology is presented, which involves data pre-processing focusing on feature selection and data analysis using a DT. This chapter outlines a two-step feature selection method based on MI measurements. Next, two popular feature selection algorithms, mRMR and NMIFS, are described in detail. At the end of the chapter, the evaluation techniques used to validate the performance of the algorithms and classifiers are discussed.

Chapter 4 presents the proposed DRS and this chapter is divided into two sections. The first section presents the implementation of two selected prototypes for a DRS, which includes a Personalised Travel Planning System (PTPS) (Chiang and Huang, 2015) and an Intelligent Travel Attractions System (ITAS) (Hsu et al., 2012). The results of a feasibility study of the two systems are presented, including identification of the problems with current DRS that need to be addressed. The results determine whether the problem are worth studying and can be processed within the proposed DRS. The second section describes the results of the feature and feature-selection studies using the Chiang Mai data set that we collected. This chapter also makes a performance comparison of the mRMR and NMIFS feature selection algorithms, as well as system performance, optimal models and extracted decision rules.

Chapter 5 conducts a comparative study of the different classifiers for the destination classification problem, including DT, SVM and MLP. The chapter discusses the use of ensemble learning methods, including different kinds of classifier-combination strategies including hard-voting methods such as majority vote, soft-voting methods such as combination rules, and the use of two popular ensemble-classifier algorithms involving bagging and boosting.

Chapter 6 focuses on two critical challenges in the design and implementation of a DRS user interface. The chapter discusses the proposed ARIM-UI framework, and the technology involved in the development of ARIM-UI for the DRS. The design and implementation of the user-interface system are also discussed in detail.

Chapter 7 concludes the thesis by revisiting the research objectives, summarising the contributions made and suggesting future research direction of this work.

Appendix A presents the questionnaire used in the process of data collection including English, Thai and Chinese versions. Appendix B lists the variable names and descriptions of the data sets that we collected. Appendix C shows the pilot form used during the interviews with the participants in the data-collection process. Appendix D presents the information sheet and the consent form used during the data collection. Appendix E presents the features and descriptions that were used in the study. Lastly, Appendix F lists classified post-2008 TRSs based on their system characteristics, focus stages, recommended items, methods and theories.

Chapter 2 State of the Art

Recently, ICT has been much applied to the tourism domain. This chapter reviews related work on post-2008 personalised TRSs. Its emphasis is on the use of ICT, its application, possible research trends and the challenges that arise in the development of a TRS. This chapter begins by providing the background to RSs and TRSs and discusses the post-2008 TRS overall framework. We present 33 different systems that were investigated and classified according to 11 dimensions. The chapter concludes with a discussion of the research gap identified in the literature.

2.1 Recommendation Systems

RS, a subset of Decision Support Systems (DSSs), is a tool that can recommend an item based on aggregating a user's preferences (Häubl and Trifts, 2000). It provides valuable information to help users make decisions based on priorities and concerns (Ricci et al., 2011). RSs usually apply their methodology from three fields. These are Information Retrieval (IR), Human-Computer Interaction (HCI) and Data Mining (DM) (Ricci et al., 2011). RSs play important roles in many popular e-commerce websites, such as Netflix, Spotify, Pandora, Amazon, and LinkedIn, along with others, by suggesting items to the user, including movies, music, news, articles, people, and URLs (Resnick and Varian, 1997). RSs have been applied in a wide range of domains and it would be impossible to cover them all. Therefore, this study focuses only on RSs in the tourism domain, referred to as TRSs.

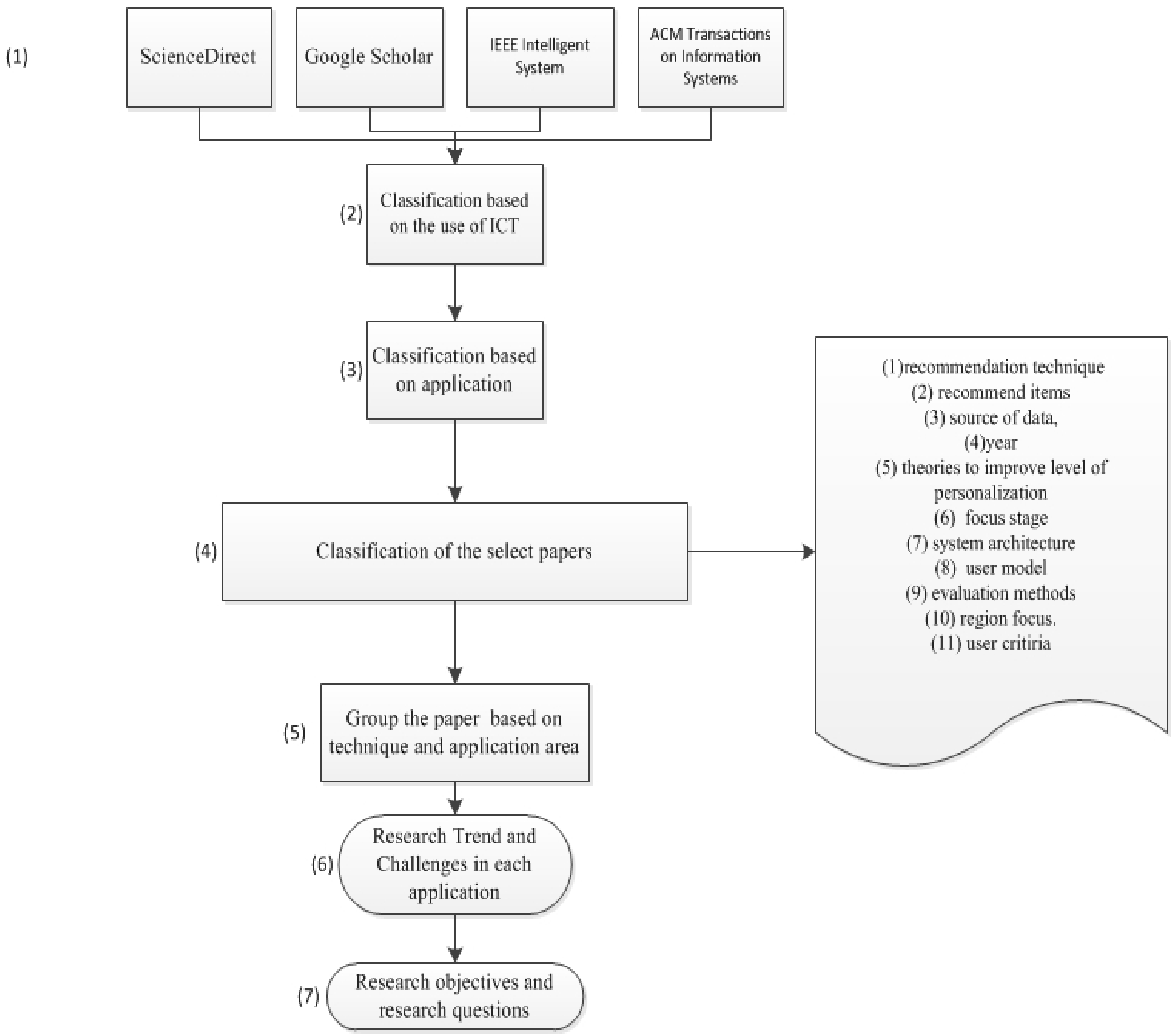


Figure 2.1 The proposed systematic literature review methodology

At this stage we aim to clarify the state of the art in ICT as it has emerged in TRS development. In addition, the TRS applications which have the greatest potential to contribute to the overall body of tourism knowledge, in terms of both academic and practical impacts, are identified. The literature review has been systematically updated, focusing on the use of ICT applications and theories/ methodology, to improve the level of personalised service and conduct an evaluation of TRSs. The main aim of this review is to analyse previous TRSs and identify research trends and challenges. This review can also serve as guidelines when designing a successful DRS. Figure 2.1, adapted from the review methodology framework by Mardani et al. (2016), presents the process followed in systematically reviewing the literature for this study.

According to Figure 2.1: (1) Papers were selected that relate to recommendations in tourism, using keywords and phrases such as: 'recommendation system in tourism'; 'tourism recommendation'; 'travel recommendation'; 'trip planning'; and 'travel recommender system'. Papers were selected from well-known online libraries: ScienceDirect[1], Google Scholar[2] and two major peer-reviewed journals: IEEE Intelligent Systems and ACM Transaction on Information Systems. From the selected papers, (2) and (3) are classified based on the use of ICT (e.g. Artificial Intelligent, Semantic Web, Multi-Agent System, etc.) and application of the TRS. The papers were classified based on 11 attributes (i.e. focus area, user criteria etc.) (4); these were categorised into two groups based on technique/ method and application (5); research trends and challenges were identified for each application (6); finally, research objectives and questions were devised based on the review findings (7).

2.2 Travel Recommendation Systems (TRSs)

Tourism is a leisure activity involving complex decision-making processes – for example, the process of selecting destinations, attractions, activities, hotels, restaurants, and services by the tourist or tourism agent. Thus, many academic and industry researchers are interested in TRSs. Over the past six years, most TSR studies have appeared in the *Expert Systems with Applications* journal. TSRs have been developed and deployed across many platforms, e.g. desktop, browser and mobile applications. Based on user input, TRSs may: 1) recommend results that are based on estimations of user interest; 2) recommend Points of Interests (POIs), tourism services, or routes; 3) rank suggested attractions/destinations in sequence; or 4) propose a holistic trip plan.

Although most current TRSs support individual tourists, systems also exist to support travel agents (Alptekin and Buyukozkan, 2011). They share similar frameworks but differ in their selection of technologies, theories to improve personalisation, data input, interaction style and recommendation techniques. Figure 2.2 shows the general framework of recent TRSs from the integration of information from various sources (e.g. sensors, GPS coordinates, surveys, reviews) to the repository (e.g. database schema, ontology).

[1] ScienceDirect www.sciencedirect.com
[2] Google Scholar https://scholar.google.com

The recommended engine can be composed of several subsystems, e.g. optimisation, statistical and intelligent subsystems. These subsystems are used to suggest, rank or predict items such as destinations, attractions, activities and services based on user requirements, preferences, hard and soft constraints such as user-demographic information, number of travel days, travel budgets and travel type.

Generally, before or during a trip, a tourist provides input (e.g. implicit, explicit, or both) to a TRS, which then creates a user profile and calculates recommended results based on the profile and various databases. A TRS may present results in many ways, such as using destination icons on a map interface with a point-to-point route, agenda, and itinerary. Most TRSs present results using spatial Web services and the Google Maps Application Program Interface (API).

Some TRSs are now able to adapt their results to the user by incorporating user-context information such as location or weather. Some TRSs let the user modify the results through the user feedback or user ratings; then, TRSs can update user profiles to make future recommendations (Anacleto et al., 2014; Sebastia et al., 2009).

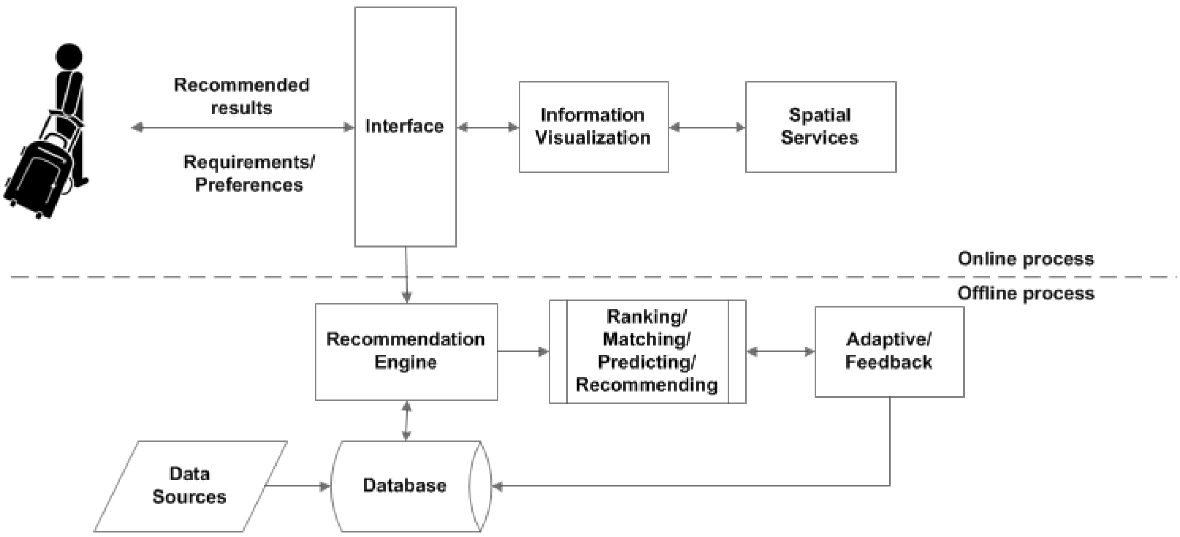


Figure 2.2 Conventional architecture of recent travel-recommendation systems

In this study, we aim to develop a TRS that recommends destinations to tourists. Our DRS has two main differences compared to previous systems found in the literature. This includes its contribution with regard to the recommendation engine and the system interface.

2.3 Recommendation engines and techniques

Schafer states that an RS can be classified by the degree of personalisation, including the usefulness and accuracy of the recommendations (Schafer et al., 2001). The degree of personalization can be defined from low to high, including non-personalization, ephemeral personalization (short-term), and persistent personalisation (long-term). The non-personalized RS is a relatively simple system that does not take user preferences into account when making recommendations. For instance, the RS only generates a list of the most popular items based on the number of reviews or number of purchases (i.e. editor's choices or top-sellers), in essence assuming that other generic users will probably like the recommended results. RS research has not focused on non-personalized RSs, due to their limited decision power (Ricci et al., 2011).

An ephemeral and personalised RS incorporating information about system users (e.g. user preferences, and socio-demographic information) is more advanced than a non-personalized RS. In other words, every user sees a different list of recommendations, depending on his/her preferences. For example, TripAdvisor (www.tripadvisor.com) recommends a destination based on the user's socio-demographic information. Previous studies have analysed many types of personalised RSs, and researchers have categorised them according to the information-filtering techniques employed (Burke, 2002; Jannach et al., 2010; Montaner et al., 2003; Ricci et al., 2011).

According to Jannach's findings, a recommendation engine (see Fig. 2.3) is composed of several recommendation techniques (Jannach et al., 2010).

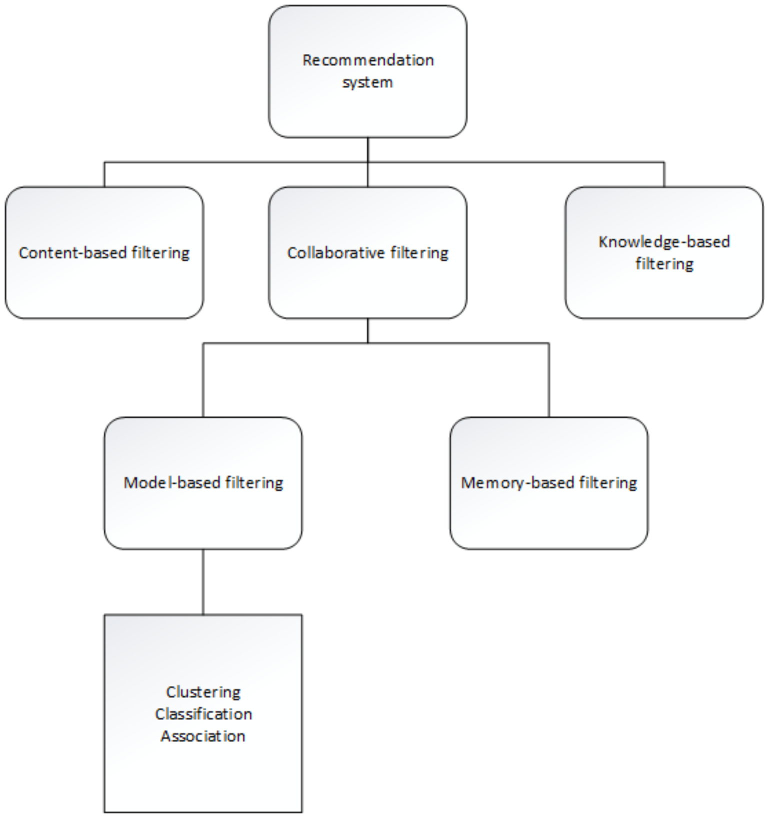


Figure 2.3 Inside a general recommendation engine

a) Collaborative filtering (also known as social filtering). This approach is the most widely implemented recommendation system. It recommends popular item(s) to the user based on the feedback of other users who share the same attributes. This approach suffers from a cold-start problem, whereby a new item or user needs to be rated before a recommendation can be made. The two most common approaches to this filtering technique are memory-based and model-based. The memory-based approach compares a user's historical records to other records in the database (Schiaffino and Amandi, 2009). The model-based approach uses statistical or learning methods, such as a Bayesian network (Huang and Bian, 2009), where a filtering technique classifies the user's historical records and builds a user model that is subsequently used in the recommendation process (Hsu et al., 2012). In fact, demographic filtering is a subset of collaborative filtering, as the system exploits demographic information (e.g. age, gender, and nationality) instead of user preferences (Jannach et al., 2010). Collaborative filtering has two main drawbacks: the cold-start problem and the data-sparsity problem. The cold-start problem occurs when the system does not have enough information regarding the item or user to make a prediction (Isinkaye et al., 2015). Therefore, the user needs to provide a significant amount of information before the system can generate a recommendation.

b) Content-based filtering: this recommendation technique suggests items to the user based on his/her previous searches or item queries. This approach suffers from the cold-start problem as the system needs to have a large historical data set in order to generate quality results (Burke, 2002). Another common problem is over-specialisation or content over-specialisation (Isinkaye et al., 2015) since the system is most likely to suggest the item that the user likes the most, with less diversity among the recommendations (Ricci et al., 2011).

c) Knowledge-based filtering: this technique recommends items to the user based on knowledge of the domain. In other words, the system has some knowledge of how a particular item relates to a particular user. This technique primarily uses case-based reasoning or ontological methods. This recommendation technique can be found in Alpekin and Buyukozkan (2011) and Santiago et al. (2012), where the system exploits travel agencies' and groups' expertise in past experiences. Constraint-based RSs (Chiang and Huang, 2015; Gavalas et al., 2012a; Vansteenwegen et al., 2011), in which the systems may not have the user's record and instead use knowledge about features in the domain to recommend specific items to the user, are also considered part of this category. For example, only certain attractions, such as stores, would be listed if a user's motivation is to shop. However, constraint-based RSs that simply generate a list of recommended items for a user based on constraints are less personalised. To avoid this, this kind of system should maintain a user's profile for future use.

d) Hybrid filtering: the afore-mentioned recommendation techniques have some strengths and weaknesses. The purpose of a hybrid recommendation technique is to achieve the best performance by mitigating the weaknesses/ disadvantages of one technique by complementing it with the strengths/ advantages of another. Many hybridisation methods for combining recommendation techniques exist, including weight, switching, mixing, feature combinations, cascades, feature augmentation and meta-levels (Burke, 2002).

One of the main tasks for a recommendation engine is to classify or cluster items (i.e. match the right item(s) to the right user(s)). Therefore, many measures of similarity methods (e.g. Euclidean distance, and correlation) have been applied in TRSs. The easiest and most common method is Euclidean distance. For example, one TRS approach uses the Euclidean distance between each pair of user and activity (Batet et al., 2012). Cosine similarity, or L2Norm, is another common method to determine the similarities between users (Schiaffino and Amandi, 2009). Another TRS approach uses Pearson correlation from statistics to find similarities between users/ items (i.e. linear relationships between two sets of data) (Hsu et al., 2012).

Previously, TRSs relied heavily on knowledge-based recommendation techniques (both case-based and constraint-based). More recently, TRSs have moved away from traditional recommendation approaches (collaborative, content-based and knowledge-based) towards context-based recommendations. The concept of context as used in RSs has appeared in the fields of Information Retrieval (IR), ubiquitous and mobile context-aware systems, marketing and management (Ricci et al., 2011). TRSs that use context-based approaches rely on a network of sensors to collect contextual information as they are mostly pull-based (i.e. requiring human intervention) (Lamsfus et al., 2012).

Tourism has substantially benefited from ICT, especially Internet technology (Pitoska, 2013). Today, tourists and tourism providers can research, select, compare and make decisions almost instantly. In 2013, 30 per cent of reservations were made online, a number that is expected to double in the next five years (Pitoska, 2013). For tourists, the Internet is the main information source with regard to tourism products and services (Pantano and Pietro, 2013). Given the huge volume of information on the Internet, the search for destinations, services and resources can overwhelm tourists and travel agencies. The tourism industry, travel agencies, and tourism companies require ICT to deliver quality services and remain competitive. Furthermore, online information searches in the future will respond to travellers' concerns when planning trips, booking reservations, and purchasing tickets (Jang, 2004). Decision support tools, also known as Recommendation Systems (RSs), have been developed to address these concerns. In tourism, they are referred to as Tourism Recommendation Systems (TRSs).

Kabassi (2010) has reviewed pre-2008 TRS studies, and Gavalas et al. (2014) has covered recent TRSs focusing on mobile applications. This study will review TRS studies (non-mobile applications) published between 2008 and 2014. It will focus on the latest developments in

TRS research, including ICT, methodological developments, research trends and challenges, features and system constraints, and combining recommendation techniques. Relevant academic journals were selected using Google Scholar, ScienceDirect and other academic websites (Gavalas et al., 2014; Kabassi, 2010).

2.4 The state of the art in technology

Judging by post-2008 TRSs, most of them rely heavily on hardware, software and communication technologies (see Fig. 2.4). In this section the ICT aspects that have been adopted in the TRS development process since 2008 are discussed. The main objective is to investigate whether there are new technologies, trends or challenges involved in TRSs.

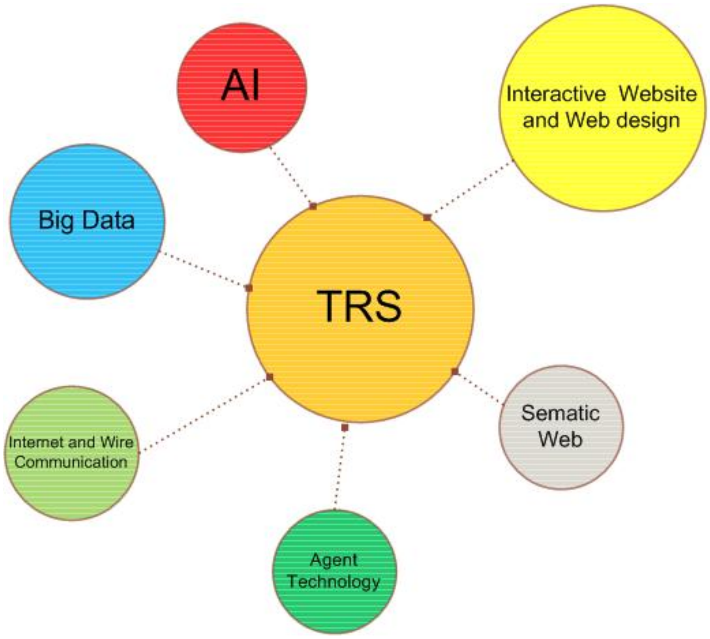


Figure 2.4 Emerging ICT

2.4.1 Wireless sensor networks

Recently, researchers have studied the effects of mobile and wireless technologies, including mobile telephones and wireless data communication, on TRSs. These technologies enhance the recommendation systems for tourists in terms of context-awareness, real-time recommendations, opportunities to re-design the route during the trip, and adapting to changed circumstances, as can be seen in García-Crespo et al. (2009), Lamsfus et al. (2012), Mochol et al. (2012) and Santiago et al. (2012). The Global Positioning System (GPS) and Geographic Information Systems (GIS) are used to retrieve user locations, provide user directions, detect nearby friends, calculate travel speed, and detect nearby POIs. GPS and GIS technologies help the user find the best POIs or routes, both before and during travel.

Many TRSs are not only deployed as stand-alone applications on desktops or browser platforms, but also supported on mobile devices due to the prevalence of smartphones embedded with GPS, compasses, accelerometers and other sensors. With mobile applications, parameters such as weather, noise level, and people nearby can be used for recommendations. Also, 3G, 4G, Wi-Fi, WiMAX and Bluetooth communication networks provide researchers with more opportunities and new state of the art resources.

Wireless technology has been used in recent TRSs. For instance, Tsai proposed a personalised recommendation system for theme parks to help tourists select a ride based on real-time information collected by radio-frequency identification (RFID) (Tsai and Chung, 2012). Gavalas et al. (2012b) implemented a Mobile Tourism Recommendation System (MTRS) that deploys a Wireless Sensor Networking (WSN) infrastructure to solve the problem of delivering a cost-effective means for remote content updates and to support proximity detection (Gavalas and Kenteris, 2011). There are two challenges regarding these innovations for a TRS. First, there is the use of context-aware ratings as a collaborative filtering approach in MTRSs where tourists can upload, review and make comments via their mobile devices. Second, there is the attempt to implement a Wireless Sensor Networking (WSN) infrastructure to solve the problem of providing a cost-effective means for remote content updates and to support proximity detection (rural positioning of POIs). Input data come from the user's website registration, where the input variables may include gender, marital status, age, education level, POI categories and favourite leisure activities as optional. WSN is an innovation which, due to the lack of developed network infrastructure and the high cost of mobile services in many countries, resultsin tourists mostly avoiding the use of 3G/ Edge connections (Gavalas et al., 2012b). However, this TRS still suffers from the implementation of unreliable networks.

The Internet of Things (IoT) is another concept that may play an important role in the tourism industry. According to Swan, IoT refers to the trend of merging the physical world with the world of information in a general Internet-like state of connectedness (Melanie Swan, 2012). For example, IoT connects many objects, stakeholders, agents and sub-systems in their business process. Therefore, tourists can now generate, send and receive data through communication devices, via a range of communication technologies, networking protocols, and data types, with little human intervention.

2.4.2 Artificial Intelligence (AI)

Artificial Intelligence (AI) is now being applied to tourism research. AI has many different definitions but, put simply, it is a technology that seeks to understand human thought processes and simulate human intelligence in machines (Turban et al., 2014).

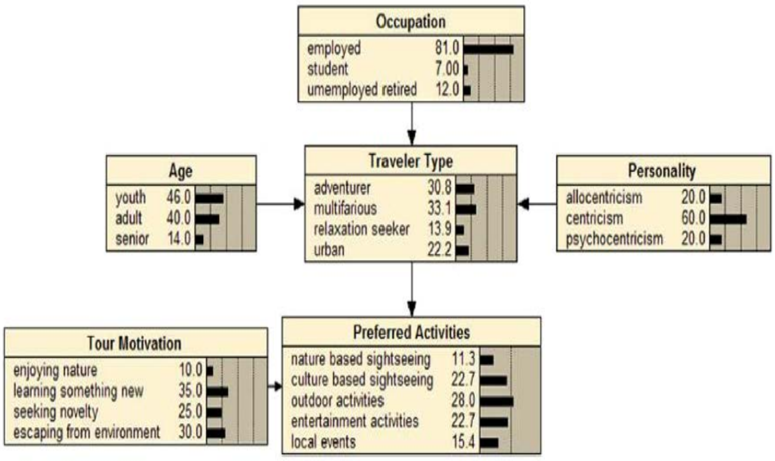


Figure 2.5 Bayesian Network model to predict a tourist's favourite attractions (Huang and Bian, 2009)

AI and machine learning have been heavily adopted in TRS to improve decision-making, optimisation, scheduling, clustering, knowledge representation and planning. Figure 2.5 shows that Bayesian Networks (BN), sometimes known as belief networks or probabilistic directed acyclic graphical models, are one of the most popular machine-learning techniques that TRS researchers use to estimate a user's favourite attractions based on user preferences. A BN combines Bayesian theories about knowledge. For example, given certain demographic tourist information, a BN estimates a tourist's preferred destination or activities (Hsu et al., 2012; Huang and Bian, 2009). A BN is a hybrid recommendation system that combines content-based filtering and collaborative filtering (Huang and Bian, 2009; Sparacino, 2003).

Fuzzy logic has also been adopted in previous studies, mostly for knowledge-based TRSs (Lucas et al., 2013). The fuzzy method has been used to deal with the uncertainties that surround linguistic assessments taken from sector experts and tourist feedback (Alptekin and Buyukozkan, 2011; Garcia-Crespo et al., 2011). It has also been used to understand uncertainty in driver behaviour in order to make the recommendation system more intelligent, e.g. by understanding the imprecise (fuzzy) way in which a driver picks a route (Pang and Takahashi, 1999).

Case-Based Reasoning (CBR), a machine-learning method, provides solutions to similar problems involving four processes: retrieve, reuse, revise and retain. Multiple-Criteria Decision-Making (MCDM), another problem-solving methodology, is a good method for evaluating and comparing criteria and then ranking alternatives. Alptekin and Buyukozkan (2011) proposed an intelligent tourism-destination planning system to help travel agencies reduce their workload. The system combines CBR and MCDM to increase system accuracy, where both methods share something in common in terms of decision-making. The challenges of this research study involved integrating of these two decision-making methods and having an understanding of how to increase the accuracy of the TRS. User requirements such as tour type (e.g. active, wandering, city), number of travellers, region, transport mode, tour length, season, accommodation type and rating (i.e. number of stars) are the parameters for the TRS. The output of this TRS is a travel plan with a quoted price. The advantages of the system are that the reliability of the results obtained and the framework can be adapted to suit other application domains. A major disadvantage of this system is the adaptation feature, which relies heavily on the experiences of travel agencies. For example, when a tourist creates a new case, it cannot be inserted directly into the database; rather, it has to be evaluated by the travel agency or accepted by the tourist first (i.e. the adaptation phase is done offline or manually). Another disadvantage is the cold-start problem (i.e. the system does not have sufficient information to make any inferences about users) because this TRS requires a long time to collect data and convey it to the database.

A Genetic Algorithm (GA) is a search heuristic that mimics the process of natural evolution. Ant Colony Optimization (ACO) is a metaheuristic method that mimics ant behaviour. Both have been used by personalised tourism-recommendation systems to learn about tourist personalities and context data in order to select a suitable route or POIs for them (Abbaspour and Samadzadegan, 2011; Liu et al., 2014; Mocholi et al., 2012).

There are many AI techniques that use recommendation engines beyond the field of TRSs. To name a few: Koren et al. (2009) proposed matrix-factorisation methods that are dedicated to the recommender and used in a collaborative filtering approach in movie recommendations, using the Netflix data set. Ge et al. (2011) developed a cost-aware recommendation system focused on making cost-aware tour recommendations. The system predicts travel package tours to the users based on travel costs and the tourist's interest. The system takes travel tour data collected from a travel company, using Gaussian processes to develop a model, and evaluates the system using an RMSE metric.

Scholz et al. (2015) proposed a utility-based recommendation system to predict consumer utility functions and their ability to pay. The system is designed from ordinal attributes input only and systems that use collaborative filtering methods could profit from their approach. De Bruyn et al. (2008) developed a RS that suggests optimal questions to be used on a website as the user's input. This paper also compares the performance of three algorithms: Bayesian treed regression, cluster classification and step-wise componential regression.

While these studies are of interest, their system goals are focused on prediction accuracy and not on the tourism-application challenge. For our proposed system we not only focused on the prediction accuracy but also concentrated on the transparency and interpretability of models. A DT is a hierarchic model, it provides decision rules which can make it easier to understand the decision-making process.

In the TRSs field, most of the developed models are considered to be black-box and do not provide this feature (white-box) as ours does. In addition, to the best of our knowledge, our approach has not yet been considered in any other TRS. The difference in our destination-recommendation system compared to other three systems is that we use a hybrid approach consisting of content-based, collaborative-based and knowledge-based filtering approaches.

2.4.3 Ontology and Semantic Web technology

The goal of the Semantic Web, also known as Web 3.0, is to efficiently share data and process information automatically and manually by promoting common exchange protocols and data formats. Many TRSs rely heavily on knowledge from the tourism domain. In order to represent knowledge in the tourism domain, a technology called ontology is commonly used. Ontology is a method used in Computer Science and the Information Sciences. It helps to represent knowledge in the domain, or at least part of it, as a set of concepts. It considers the relationships within the knowledge base and also plays a prominent role in the framework of the Semantic Web (Huang and Bian, 2009). Montejo-Ráez proposed a TRS which is called SAMAP (Castillo et al., 2008). This is an example of a TRS that has modelled and implemented its own ontology to represent tourists' interests (e.g. user, city, transport, place, personal preferences). Huang and Bian's work (Huang and Bian, 2009) is another example of a TRS that applies ontology. The goal was to model the attraction at Niagara Falls in New York State. In total, nine concepts were found by extracting information from many tourism websites. The concepts include attraction, opening times, admission fees, closed dates, minimum times and stay. Next, they applied the relationship between the concepts.

Semantic Web technology and ontology help researchers to integrate heterogeneous online information (Castillo et al., 2008; Horrocks, 2008; Huang and Bian, 2009; Mocholi et al., 2012; Petrevska and Koceski, 2012; Rodríguez et al., 2012; Santiago et al., 2012). The Resource Description Framework (RDF) and Web Ontology Language (OWL), the most commonly used languages (Horrocks, 2008), have been used to develop TRSs to represent the modelled tourist classes/concepts and their relationships.

2.4.4 Agent technology

Agent technology offers many benefits when modelling complex real-world problems (Kantamneni et al., 2015). Many personalised tourism recommendation systems have adopted this technology (Batet et al., 2012; Castillo et al., 2008; Lee et al., 2009). A Multi-Agent System (MAS) is composed of agents that interact with each other in the environment. Each agent has its own goal and tries to maximise resources, utilisation, and benefits (Siebers and Aickelin, 2008). There is no clear consensus on the definition of an agent (Siebers and Aickelin, 2008). MASs are promising tools for modelling problems of organisation or real-world problems, where people have to make decisions as a group (Payr et al., 2002). Some agents in the system are identified as Intelligent Agents (IAs), since they can make decisions, optimise, schedule, and solve complex problems.

Turist@ (Batet et al., 2012) is one example of a TRS that has been implemented with a MAS. It is a mobile-push and location-based TRS that has a high degree of dynamic adaptability, taking user locations from GPS into account (i.e. the system can adapt to changes in the trip schedule and incorporate new suggestions). The system also considers users' demographic information (e.g. age, education, nationality, language and disabilities), trip characteristics (e.g. travel-group type, trip duration) and user preferences. The system notifies the user when she/he is near an activity and suggests interesting activities. The TRS uses a hybrid filtering method (content-based filtering and collaborative filtering) to make a personalised recommendations. The TRS has a feature that can include dynamic management of a user's profile for use in the personalised recommendation process, such that the profile will be updated in both explicit and implicit ways.

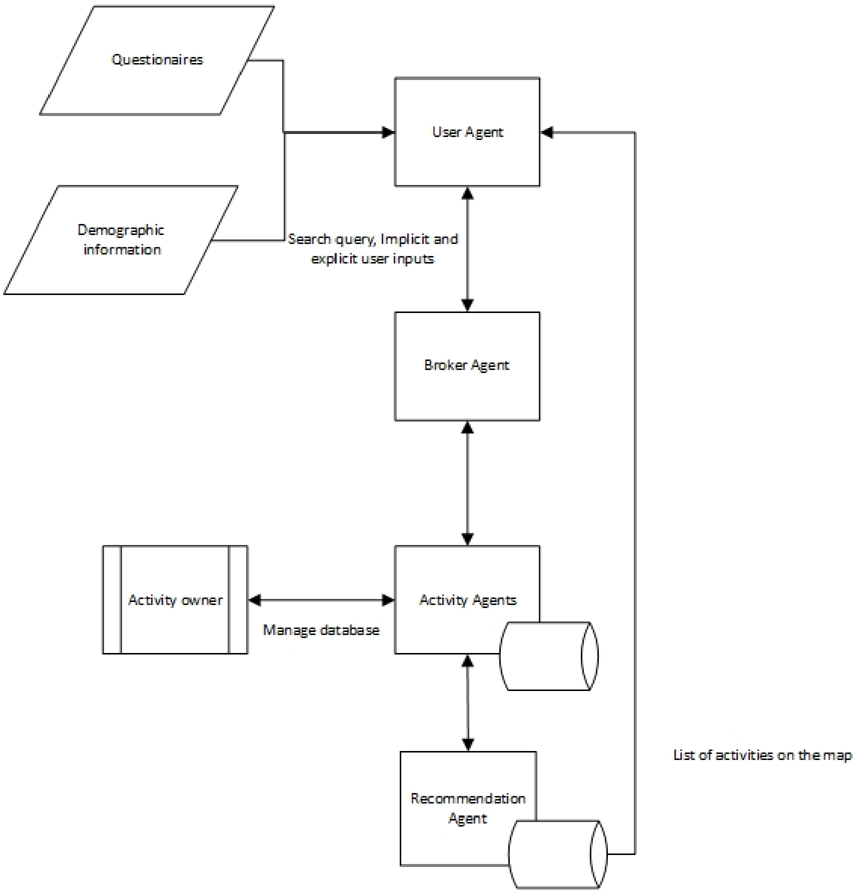


Figure 2.6 An overview of the Turist@ system architecture (Batet et al., 2012)

According to Figure 2.6, the use of a MAS has many advantages for a distributed system, in that there is an agent running on the mobile device, a broker agent running as a facilitator between the user agent and the activity agent to handle communication between them, and another agent responsible for maintenance of the databases so as to reduce server overload and so on. Moreover, the ability to adapt, adjust, add and remove agents seems to be a suitable concept for modularity design when modelling a distributed system and real-world problems. Also, there is a high degree of adaptive capability in the system, such that the system can adjust the plan based on a new location of the user at the time of execution. User feedback is based on both explicit (i.e. ratings approach) and implicit (i.e. monitoring his/her actions by analysing the time the user spends on a web page and the links the user followes etc.) factors.

2.4.5 Web design

When tourists browse travel websites they expect them to be interactive, responsive, informative and attractive (Chu, 2001). To meet this expectation, many personalised tourism recommendation systems use AJAX Web programming, which combines several technologies, such as HTML, JavaScript, XML and a document-object model, to create a sense of interaction between the user and the web application. Chiang and Huang (2015) and Montejo-Ráez et al. (2011) proposed a travel-planning system for recommending personalised travel schedules, it has an adjustable interface module that enhances travel-planning flexibility. Moreno et al. (2013) developed a Web-based TRS using Java Server Faces (JSF) and AJAX, a Web development technique combining XML and JavaScript, to create an asynchronous Web application for TRS. The ontology was also developed using the thesaurus of the WTO as a reference guide with OWL. Buhalis and Law (2008) indicate that Web design has been one of the most important technological innovations for the tourism industry. Moreover, accessibility features for disabled and elderly people should seen as a beneficial feature for an interactive website.

2.4.6 E-tourism services from TRS

Many recent TRSs have focused on recommending destinations, along with integrating certain tourism services, such as hotels and restaurants, into the content as shown in Figure 2.8. The output of most systems is itinerary-based. Lately, researchers have expanded their focus to include recommending routes and solving trip/ itinerary design problems. Many TRSs provide a holistic trip plan by mainly focusing on specific content. From the literature, TRSs can be categorised based on the e-tourism services they provide, including destination recommendations, tourist service recommendations, route recommendations, and trip planning /itinerary recommendations.

2.4.7 Destination and tourist-service recommendations

Even the simplest Destination TRSs and DRSs list destinations (e.g. POIs, attractions, activities, events) according to specific input constraints provided by users. Some of them take context information into account. DRSs are moving towards a point at which they will be able to rank the importance of destinations and predict destination suitability by the user (Yang and Hwang, 2013). Some DRSs have used decision-making theory to better understand how tourists select preferred destinations in order to improve prediction accuracy (Hsu et al., 2012; Huang and Bian, 2009).

Huang and Bian, (2009) proposed a tourists' personalised recommendation system (Web-based) combining BN and AHP as the recommendation engine, in order to produce a trip itinerary as the output. The system, over the Internet, suggests a set of tourist attractions in sequence at a given destination. Their recommendation system considers both the travel behaviour of the user and other tourists' behaviours, particularly using both content-based and collaborative filtering methods. The system has four main components: heterogeneous integration, personalised recommendations, adaptive capability and spatial-functions capability. The capability of integrating heterogeneous online tourism information (i.e. using ontology) and providing hybrid-personalised recommendations (i.e. collaborative filtering and content-based filtering) are the advantages of this work. The ability to predict a user's preferred activities using a machine-learning method, such as BN, is a novel approach. Moreover, the capability to rank attractions using decision-making theory (i.e. factors such as a user's preferred activities, cost and distance) is also considered interesting and challenging. Additionally, the system has adaptive capability, in that it provides an interactive Web interface so that the user can revise the recommendation results.

Regarding the disadvantages of the system, the ArcWeb service is now quite old, and the product line is retired. There are better GIS services that can provide spatial-function capabilities, such as the Google Map API. Moreover, more decision criteria could be added in order to rank attractions.

Hsu et al., (2012) developed a TRS, referred to as ITAS, that predicts preferred user tourism attractions based on decision theory, using machine-learning methods, such as the a Bayesian network to predict a user's preferred attractions based on a user's demographic information (e.g. age, gender, trip purpose, income, occupation, source of information, nationality). Statistical methods, such as factor analysis, are used to analyse the data set and remove non-required input from the user. Regarding the system output, the system generates a list of ranking attractions and provides interactive map interfaces and point-to-point route information to the tourist via Google Maps. The TRS achieves high recommendation accuracy from the results of the Receiver Operating Characteristic (ROC) curve. This is because the use of a combination of content-based filtering and collaborative filtering contributes to the quality of the data set obtained.

Yeh and Cheng (2015) proposed a knowledge-based TRS that recommends tourist attractions in Taiwan. The system uses knowledge from tourism domain experts to reconstruct target classes. It recommends attractions based on one of two user inputs – favourite travel category (e.g. natural, museums and galleries, heritage etc.), referred to as a construct-based recommendation, or a specific tourist attraction, referred to as an element-based recommendation. The challenge of this study was to increase the recommendation performance by reducing data sparsity using a novel method.

INTRIGUE (Ardissono et al., 2003) offers both Web-based and mobile (handheld devices) platforms for the city of Turin, Italy. The system recommends POIs (i.e. sightseeing destinations) and itineraries by taking the preferences of heterogeneous tourist groups (e.g. families with elderly members or children) into account, as this is one of the challenges in current TRS design. This TRS takes many user constraints into account as input, such as number of days, arrival/ departure time, start and end location, and preferred time of visit. The recommendation engine of this system relies heavily on the techniques of user-modelling and hypermedia. This system also supports tour scheduling both before and during travel, which is another challenge for TRS design.

PSiS (Anacleto et al., 2014) is a mobile TRS that makes POI recommendations focusing on user context (e.g. location, time, speed, direction, weather) and user preferences (i.e. through their previous work). The system has the capability to adapt dynamically to the recommended tour; for example, it can generate a new trip plan when the user is ahead of schedule. Another challenge of this TRS is the implantation of middleware that resides on the server. It synchronises data between the Web application and the mobile application. Another interesting feature is the architectonic tag, which can recommend POIs according to whether a destination is open or closed and is worth visiting. An additional feature is a tracking system, with the benefit of saving time.

SPETA (García-Crespo et al., 2009) takes advantage of Web 3.0 technologies by integrating social networks, the Semantic Web and context-awareness into a mobile TRS. The system aims to recommend tourism services, such as attractions or restaurants, to tourists who are new to the area. The TRS focuses on matching, searching and filtering items from the knowledge acquired via ontology (i.e. social and geo-location information). The system requires input – both explicit and implicit – from the user in order to make recommendations. The input includes user preferences (food and music types), user context information (weather, time, location), and derived variables such as speed and direction. The system also incorporates the opening and closing times and dates of attractions.

SigTur/E-Destination (Moreno et al., 2013) is a trip-planning, Web-based TRS that recommends leisure activities in Tarragona, Spain. The system takes many different kinds of input into account, both explicitly and implicitly. The user must explicitly input travel motivation, user demographic information (e.g. country of origin), travel budget, group composition, required destination, accommodation type, and travel dates (start and end dates) via a Web interface. When the user responds (i.e. adds or removes information) to the recommendation results, the RS takes this as implicit input to be factored into future recommendations. The advantages of this TRS are its hybrid recommendation approach and prediction method which analyse the massive data set.

Otium (Montejo-Ráez et al., 2011) is a personalised travel planning system that schedules leisure activities for tourists. Additionally, budget and current availability are factored into trip recommendations. This system relies on a Web-extraction methodology to retrieve information for its database. It uses an interactive Web-based interface so that the user can adjust the generated schedule according to his/her preferences. There are two input methods for the system. First, the tourist specifies a maximum budget and the travel area (city/province). Also, proximity, price, time, profile, and diversity are parameters that are needed to calculate a trip plan inside the recommender via a web interface. This wrapper method is an advantage when dealing with Web information sources. However, the wrapper can only parse an HTML file. It needs to keep up with the configuration file to be able to adapt to changes in the HTML file structure; also, it can only extract event attributes. This TRS lacks many important features, e.g. a transportation feature, whereby a user can search for a transport mode to select during the trip. Another drawback is the navigation system, in that the system could use the gathered geo-position to plot a route or location using Google Maps.

SAMAP (Castillo et al., 2008) is a TRS designed to assist tourists in planning a trip based on user history and other factors. It focuses on the Team Orienteering Problem with Time Windows (TOPTW) and recommending activities. SAMAP is based on a multi-agent system and is intended to work on mobile devices. System inputs include user preferences, personal information and user context. Transportation (e.g. bus, taxi, walking) and environmental information (e.g. traffic, street type) are also taken into account. The system recommends a trip plan, with a list of visitor activities, and a suggested route beginning at one POI and then taking the user to another.

e-Tourism (Sebastia et al., 2009) is a hybrid TRS that matches user demographics and preferences with a destination database to create a leisure plan with a list of recommended leisure activities in Spain. A taxonomy, a set of concepts, is used to describe tourist activities. The TRS uses AI planning to generate realistic activity plans incorporating opening hours, priorities, visiting duration, and utility as constraints. The system is adaptive, using a rating system upon user log in to obtain feedback in order to improve the user profile.

2.4.8 Route recommendations

Wireless Sensor Network technologies like GPS and RFID can retrieve context information, such as current location, as a parameter. A Route TRS can recommend route(s) through several destinations for a tourist. For example, it can learn user behaviour through context information to predict a route based on user or group preferences (Mocholi et al., 2012;Tsai and Chung, 2012).

Route TRSs make point-to-point recommendations with multi-model transportation services (Abbaspour and Samadzadegan, 2011; Castillo et al., 2008). Additionally, there is a TRS that provides real-time information to tourists to reduce congestion and avoid long queues at tourist hotspots (Liu et al., 2014).

Tsai and Chung (2012) proposed a route-recommendation system for theme-park tourists using a clustering technique. The authors used Radio-Frequency Identification (RFID) attached to a wristband provided to visitors to collect tourist profiles in real time, including the sequence of attractions visited and a corresponding timestamp. The system recommends itineraries based on tourist preferences and other travel behaviours. The advantage of this recommendation system is its idea of using RFID to collect and apply accurate and instant data to solve theme-park issues in real time. Regarding the system's disadvantages, first, the RFID system could be made more realistic by inputting the location of the information available at public booths into the system configuration. Second, the system parameters could be improved by using optimisation approaches to find better values for the system. Third, personal behaviours like spending habits and dietary favourites, could be used as input parameters in the system. Last, the problems could multiply if the park has multiple entrances and exits.

Lee et al. (2009) proposed a personalised tourism recommendation system for Tainan, Taiwan that acts as a travel agent for tourists by recommending POIs based heavily on the knowledge of domain experts. The system suggests a personalised tourist route in accordance with user requirements, such as the number of days, popularity, region, food types and classes of historical sites. The TSR combines Multi-Agent technology, ontology and ant-colony optimisation to present route plans with the aid of the Google Map API.

Pang and Takahashi, (1999) proposed a route-ranking recommendation system based on driver preferences (content-based approach) inside a vehicle's Dynamic Route Guidance (DRG) system. This requires a complex evaluation process, given that driver behaviour in terms of route choice is a complex problem. The proposed system models driver behaviour by using a fuzzy expert system; the system also has an adaptive mechanism function that responds to the driver's preferences and recent decisions. The inputs include the origin (obtained directly from the GPS) and the destination, along with any route attributes, such as travel distance, travel time, the degree of congestion, tolls, the degree of difficulty of travel, and scenery. From these, the system generates a recommended or optimum route, based on real-time traffic and road information, and displays it on the driver's console. The ability to learn from the uncertainty of the driver's behaviour makes the system more intelligent, and this is considered to be an advantage of this TRS.

Santiago proposes a knowledge-based system called GeOasis (Santiago et al., 2012). The system is integrated with GPS and acts as a tourist guide to suggest certain POIs, according to the tourist's location. The system is implemented for Jaen Province in Spain. The system behaves dynamically to adapt to user-context data, such as current location, time and space. This TRS has the capability to generate a trip plan in real time, with the use of a heuristic algorithm to improve the speed of computation time. The knowledge base is considered to be the greatest contribution to the system as knowledge is gathered from domain experts. Last, voice recognition seems to be a positive technological enhancement to user interaction with the system.

Mochili proposes a context-driven TRS called SACO (Mocholi et al., 2012), a system that is capable of learning the user's routes using context information involving GPS locations. The ability to deal with the problem of reducing the amount of information displayed on the map so that the user does not have to filter out non-interesting services by himself/herself manually is the main challenge of this TRS.

Regarding the disadvantages of this TRS, the communication between client and server is difficult to manage, since the client is constantly moving around. However, researchers have addressed this issue by implementing a buffer for the client, but this only seems to be a temporary solution to the problem.

2.4.9 Trip planning/itinerary recommendations

Trip planning is challenging; for example, tourists usually have specific requirements and needs, such as the number of travel days, number of travellers, budget, required destinations, the days that attractions are open, and starting locations. Trip planning/itinerary recommendation systems take these user preferences and context features into account when deriving the order of destinations on an itinerary. Moreover, these systems can create a new plan/itinerary for a traveller in response to changes occuring during the trip. For example, if the traveller is running out of time, the system may reschedule a destination.

While TRSs cover many different aspects of tourism services, few focus on the trip planning or scheduling problem, as this is a complex problem that requires the TRS to generate an automated optimal travel plan (i.e. the most realistic travel plan) for the user, based on many constraints.

According to Hagen et al. (2005), this problem has been termed the Tourist Itinerary Design Problem (TIDP) or the Tourist Trip Design Problem (TTDP) (Gavalas et al., 2012b, n.d.) (Gavalas et al., 2012a, n.d.). This problem resembles the classic Travelling Salesman Problem (TSP) in theoretical computer science and operations research. However, the TSP conundrum is concerned with minimizing travel time or travel distance; the simplest TIDP can be modelled as an Orienteering Problem (OP), where a set of vertices comprises given points of interest, each of them having a score (e.g. user satisfaction), and the goal is to create the best path to maximize the total score (time or budget) for each of the vertices. Golden et al., (1987) proved that OP is an NP-hard problem. TIDP can be modelled as a Team Orienteering Problem (TOP), where the problem is NP-complete (Vansteenwegen et al., 2009). The Team Orienteering Problem with Time Windows (TOPTW) (e.g. considering opening and closing times per day), which has appeared in recent studies (Gavalas et al., 2012a; Vansteenwegen et al., 2011), is an extension of TOP.

DailyTrip (Gavalas et al., 2012a) approaches TOPTW using a novel heuristic algorithm to find near-optimal itineraries that meet tourist requirements and other constraints. The system is a mobile Web-based application using exhaustive user constraints, e.g. user preferences, opening days of POIs, average visiting times of POIs, and weather conditions. The proposed heuristic algorithm, which outperformed the Iterated Local Search (ILS) algorithm is a significant contribution. However, an exhaustive number of explicit user inputs (i.e. from both context and user preferences) may be too intrusive for the user.

2.5 Research trends and challenges

The previous section has discussed the advances represented by each TRS, as well as the issues associated with each of them. For example, post-2008 TRSs have attempted to generate more realistic trip plans with different approaches (e.g. using more constraints, modified algorithms, etc.). In addition, ICT has been evolving, and TRSs have been building on each other over this period. This section is dedicated to the current state of TRSs which has been central to developing the proposed methodology.

2.5.1 User constraints and contextual information for a realistic trip plan

Recommending a near optimal or realistic trip itinerary is a major challenge, such that the following user constraints and context constraints can be added to the TRS to generate more realistic and effective recommended trip plans. This is done to satisfy user requirements and preferences (Gavalas et al., n.d.; Souffriau and Vansteenwegen, 2010; Vansteenwegen et al., 2011). The following user constraints and contextual information can be added to the TIDP model.

The City Trip Planner (Vansteenwegen et al., 2011) assists a tourist when planning routes for five cities in Belgium. The system addresses the TOPTW problem with the trip planning heuristic algorithm. In addition to incorporating traditional trip constraints, including number of days, start and end locations, start and end times, lunch breaks and opening and closing times, the system weights user preferences to estimate the level of interest in each POI.

PTPS (Chiang and Huang, 2015) is a Web-based TRS that schedules hotels, restaurants, and attractions based on user requirements (e.g. number of days, number of travellers, budgets, meal times, required POIs, and starting point). The proposed system recommends POIs based heavily on user needs/requirements in order to achieve maximum user satisfaction. The system also introduced an algorithm to solve TIDP. Moreover, with an adjustable interface feature embedded in the system, users can adjust their results to replace unsatisfactory items and to improve suggestions. The main contributions of this system are an introduction to the concept of a time framework and the planning of the algorithm, referred to as the Schedule Reasoning Method (SRM). However, the system has some major drawbacks, in that it needs to apply active learning methods to address any non-intrusive issues. Also, the travel matching module could be improved upon, and the SRM algorithm does not produce a realistic trip plan. In short, this TRS relies heavily on user requirements.

2.5.2 User constraints and contextual information for destination selection

TRSs provide options when selecting destinations and services by taking into consideration a user’s hard constraints including contextual information, requirements, preferences, interests, demographics and destination information. Future TRSs should provide the traveller with even more options (soft constraints) to force the system to collect information on the destination(s) that he/she wants to visit based on his/her needs. For instance, some tourists do not want to visit more than a specific number of destinations per day or destinations that he/she has already visited on a previous trip (Souffriau and Vansteenwegen, 2010). Since most users are budget-conscious, the travel budget should include limits for transportation fees, event entrance/admission fees and hotel/ restaurant bills. Also, lunch or dinner breaks, coffee breaks, and short breaks during the day should also be taken into system consideration. By giving the system the time frame for such breaks, the system would be able to locate other related destinations or services with opening hours to match the user’s specified available time.

Moreover, the number of travel days and accessibility issues(e.g. impaired vision or hearing, motor disabilities,) should be taken into account (Souffriau and Vansteenwegen, 2010). It can be seen that future TRSs that are concerned with a realistic trip plan need to explore intelligence mechanisms that can trigger itinerary updates when contextual information changes.

2.5.3 User constraints for tourist services selection

Soft constraints can be added to a TRS. For example, a TRS that recommends restaurants could be programmed to incorporate meal times, food type (Chinese, Thai or Japanese) and price range (low–high). With these soft inputs, the TRS could recommend restaurants with opening hours and a price range that match the user's selection criteria. For a TRS that recommends hotels, soft constraints can also be added, such as hotel type, price range, and amenities (Souffriau and Vansteenwegen, 2010). Transportation options should be based on a multi-option model (e.g. travellers can take a taxi so far, then walk to a POI) and some other aspects regarding transport services (e.g. transport fees) (Castillo et al., 2008; Gavalas et al., 2012a). Regarding contextual information; weather, traffic forecasting, and current date/ time to match the destination's operating dates/ times should be taken into consideration (Souffriau and Vansteenwegen, 2010).

There is room for more research on constraint-based and context-based recommendation systems, not only in the tourism domain but with regard to other applications, including map navigation, fleet management, weather information, roadside assistance, and personal location services (Lamsfus et al., 2012; Mocholi et al., 2012).

Recommending a near optimal or realistic trip itinerary that incorporates user and context constraints to satisfy user requirements and preferences is another challenge (Gavalas et al., n.d.; Souffriau and Vansteenwegen, 2010; Vansteenwegen et al., 2011).

2.5.4 Integration of heterogeneous online travel information

Integrating heterogeneous online travel information is a major challenge for TRSs (Huang and Bian, 2009). TRSs involve gathering large amounts of information from different information providers or tourism services (e.g. hotels, restaurants, POIs) with different, or even unique, types of categories or content in a variety of formats, including non-structural data. To address this challenge, information extraction techniques such as Web extraction/ crawlers (Montejo-

Ráez et al., 2011), semantic technologies, and Web 2.0 technologies, such as Mashup (a content aggregation technology) (Batet et al., 2012; Castillo et al., 2008; Huang and Bian, 2009), have been recently adopted by TRS researchers.

Wang et al. (2011) developed a TRS based on the work of Huang and Bian, (2009). The system is a Web-based personalised RS that has three main functions: integrating heterogeneous information on tourist attractions, estimating traveller preferences, and evaluating tourist attractions. This system enhances the ontology technologies of Huang and Bian, (2009). This study's main contribution is to have taken existing tourism ontology and re-modelled the approaches so as to define the outcome as travel and user-intelligent ontology (semantically integrated resources).

A traditional Relational Database Management System (RDBMS) would have difficulty managing the large amounts, and complex nature, of data used in TRSs, including geospatial data and continual and numerous user updates, given data availability and scalability issues. For TRSs, Not Only SQL (NoSQL) is a more promising technology for increasing system performance and reducing latency than RDBMSs. However, a trade-off of using NoSQL is that TRSs may lose database-wide or transaction consistency (Gavalas et al., 2014.)

2.5.5 Group-based recommendation

Group-based recommendation systems pose a challenge because, not only do groups of tourists have different individual preferences, but they must also be concerned with the preferences of other group members. Recommending an itinerary for a group that optimally satisfies differing individual interests is difficult. Given this difficulty, only one TRS study attempts to support both individual and group travellers, i.e. Garcia et al. (2011).

2.5.6 Interactive and responsive Web applications

Chu (2001) states that when tourists browse travel websites they expect them to be interactive, responsive, informative, and attractive. To meet these expectations, many personalised tourism recommendation systems have used AJAX Web programming that combines several technologies, such as HTML, JavaScript, XML, and document object models, to create a sense of interaction between the user and the Web application. Chiang and Huang (2015) and Montejo-Ráez et al. (2011) propose a planning system for recommending personalised travel schedules with an adjustable interface module that enhances travel-planning flexibility. Moreno et al., (2013) developed a Web-based TRS using Java Server Face (JSF) and AJAX; the ontology was developed using the thesaurus of the WTO as a reference guide with OWL. Accessibility features for disabled individuals and elderly people should be added to the TRS, to make it more responsive.

2.5.7 Tourists' decision-making and information processing through a human-centric approach.

Recently, a few TRSs have used quantitative research methods to understand tourists' search behaviours in assessing travel information and decision-making processes. According to Fesenmaier et al. (2006) and Gretzel et al. (2012), a successful DRS requires an understanding of tourists' decision-making and search processes. The factors identified below influence travel searches and travellers' decision-making processes for a truly human-centric DRS.

Personal characteristics of the traveller are of significant importance (e.g. socio-demographics, knowledge, personality, involvement, values, attitudes, cognitive style, decision-making style, vacation style) (Fesenmaier et al., 2006). Andereck et al., (1993) have stated that the major factors influencing consumer decisions when purchasing a product or service are information sources about that product or service. In addition, individual demographics may influence information-seeking behaviour.

Trip characteristics are also of significant value (e.g. travel purpose, trip length, travel distance, travel party, travel mobility) (Fesenmaier et al., 2006).

Mutinda and Mayaka (2012) have proposed two sets of factors, i.e. environmental factors and individual trait factors that impact on destination transformation or the destination funnelling process and the final choice. Environmental factors, such as sources of information, culture, family, lifestyle, and destination features are also of relevance. Individual trait factors (personal characteristics) include motivation, personality and past experience. Specific key factors that determine the choice of a domestic plan by tourists in Kenya include the following:

1. The need for knowledge and adventure
2. Economic concerns
3. Destination information
4. Travel arrangements

Regarding sources of information, the study found that personal experiences are the most significant factors in raising destination awareness.

Travel motivation has been addressed by Hsu et al. (2009) in their decision-making model based on the Analytic Hierarchy Process (AHP) for destination choice. The study not only evaluates the importance of motivational factors but also seeks an understanding of decision factors. The study identified the factors that influence a tourist's choice of destination and found the following to be the six most important motivational factors for inbound tourists visiting Taiwan:

1. Visiting friends/relatives
2. Personal safety
3. Escape
4. Rest and relaxation
5. Destination image (impressions that a person holds about a country in which they do not reside) (Hunt, 1971). A totality of impressions, beliefs, ideas, expectations, and feelings accumulated towards a place over time (Kim and Richardson, 2003)
6. Environmental safety and quality

When tourists are selecting their destinations, travel and tour motivation is one of the important factors found in the literature reviewed. This variable describes the reason why a tourist chooses to visit a particular destination (Leiper, 1990).

Crompton (1979) identified nine motives based on two kinds of motivation that influence the selection of a destination. Seven of them are classified as socio-psychological, the remaining two in the cultural category. Types of socio-psychological motivation are: escape from a perceived mundane environment, exploration and evaluation of oneself, relaxation, prestige, regression, improvement of kinship relationships and facilitation of social interaction. With regard to cultural motives, the main factors identified were novelty and education.

Figler et al. (1992) defined five factors that influence the selection of destinations: anomie/ authenticity-seeking, culture/ education, escape/ regression, wanderlust/ exploring the unknown, jet-setting / prestige-seeking.

2.5.8 Mobile recommendation systems in tourism

An increase in the use of mobiles and new developments in mobile computing and communication networks (i.e. GPS, Wi-Fi) offer state-of-the-art improvements to recommendation systems in the tourism domain. Context information from mobile device sensors such as that associated with a specific location, considers the speed used in the process of making recommendations. Mobile RS can provide tourists with a new experience when they are making decisions. For example, Balduini et al. (2012) proposed a mobile TRS application that is deployed on the Android operating system that using Augmented Reality (AR). The TRS assists tourists in the process of selecting restaurants in Insadongs, Seoul.

2.6 Identifying research gaps

Most previous TRSs have only supported individual tourists and have focused on estimates when choosing a destination, activities, attractions and tourism services (e.g. restaurants, hotels, transportation) based on the user's preferences and interests. With regard to technical aspects, these TRSs only provide filtering, sorting and basic matching mechanisms between items and the user's hard constraints.

It can be seen that the latest ICT provides new opportunities for researchers to design and implement a TRS that is more intelligent, interactive, adaptive, and automatable, one that supports a higher degree of user satisfaction than ever before.

In summary, future destination TRSs should be able to achieve the following:

1. *Enhanced tourist decision-making process*

   The travel decision-making process is complex. A deep understanding of how a traveller selects a destination is one of the biggest challenges when designing a TRS. A model-based approach TRS that aims to identify a tourist destination or other service selection process is necessary in order to develop a successful and useful DRS (Fesenmaier et al., 2006; Gretzel et al., 2012).

2. *Reduce user's effort*

   It can be seen that most current TRSs require massive input from users in order to generate a decent recommended result, but many user inputs may not be needed for the system (Chiang and Huang, 2015; Hsu et al., 2012). Current TRSs have begun to request more specific information from the user to generate an appropriate destination recommendation, in terms of route-planning, and trip-planning. However, having more parameters in the system could decrease TRS recommendation performance and the level of user satisfaction. Future TRSs should be able to understand relevant theories in order to improve accuracy, effectiveness, efficience, and satisfaction. Moreover, they should understand the factors that play an important role when tourists make decisions. They should be able to reduce the amount and types of information required to achieve system/ service satisfaction and still provide enjoyment in the process of searching for tourism information.

3. *Performance, speed, recommendation accuracy, and precision of DRS*

   Current TRS development needs to be concerned with recommendation performance and the selection of a proper scientific method to validate their systems. Future TRSs should combine recommendation techniques to find or modify recommendation algorithms and search for relevant factors. This could lead to an increase in system performance. Future DRS should provide proper scientific evaluation methods to validate the performance of the system.

4. *Intelligent user interface or website*

   Future TRSs should improve the interaction between the user and TRS to expand the user experience and increase satisfaction. Intelligent User Interface (IUI) technology comprises of both Artificial Intelligence (AI) and Human-Computer Interaction (HCI).

5. *Integration of heterogeneous information*

   Due to the heterogeneity of the information that is available on the Internet, future TRSs should provide a proper mechanism to automate the integration of information that is available from various travel information sources.

6. *Provide a holistic trip plan*

Future TRSs should provide a holistic trip plan and aim to create an even more realistic trip plan in real time. This can be done by taking massive amounts of data from several sensors, such as GPS and RFID data to generate real-time recommendations, or by having more of the user's soft constraints into the system.

7. *Support group recommendation*

   Most post-2008 TRSs only support a single user model. Future TRSs should support not only individual travellers but groups of travellers as well.

8. *Highly adaptive*

   Future TRSs should provide the ability to adapt to the user's contextual information features, enabling the user to modify the results by way of feedback mechanisms in order for it to be able to improve system accuracy and user satisfaction.

9. *Concerns about user privacy*

   Current TRSs are beginning to collect more information from the user, but the sharing of certain information can be considered a sensitive issue. For example, users may not be willing to share their age or gender information.

Due to the time constraints of this research study, this study only focuses on aspects 1, 2, 3 and 4.

2.7 Summary

This chapter has conducted a review of relevant literature regarding recommendation systems within the tourism domain. It can be seen that the majority of post-2008 TRSs focus on recommending destinations, routes, and realistic trip-planning/ itineraries. Also, we can see that the latest ICT provides a new opportunities for researchers to design and develop TRSa which are more intelligent, interactive, adaptive and automatable, as well able to offer a higher levels of user satisfaction and user experience than ever before. The literature review shows that current TRS developments are still at a stage that requires more variables than ever from the user in order to increase the predictive accuracy of destination recommendations, route plans or trip plans. However, this does not guarantee satisfaction in terms of the user's information search experience. This means that further TRSs should build on exisiting decision-making foundations in order to be more effective and less intrusive.

This research intends to contribute to the development of an improved DRS, as that previous DRSs are lacking in both technical methods, such as recommendation accuracy and evaluation, and practical aspects, such as user satisfaction. We propose a novel DRS that understands the tourist's destination choice by developing destination choice models using both quantitative and qualitative approaches, as well as increasing the level of user satisfaction by using machine learning and Web technology techniques. This is outlined in the next chapter.

Chapter 3 Research Approach, System Architecture and Pilot Study

The previous chapter reviewed studies of existing Travel Recommendation Systems (TRSs) and identified the key weaknesses of previous Destination Recommendation Systems (DRSs) for tourists. The aim now is to develop a DRS that overcomes current DRSs practicality issues in terms of understanding travellers' choices regarding the destinations they are planning to visit before or during a trip, as well as increasing levels of user satisfaction. Another aim relates to technical issues regarding improving the recommendation accuracy of the DRS. This chapter presents an overview of the research approach and system design and describes the proposed DRS framework, including the research methodology and system design included in the process of data collection. The design and development of the questionnaire used in the research and the survey sites are presented. At the end of this chapter the evaluation methods used to assess the system performance and system design of a practical DRS are presented.

3.1 Overview of the research approach

The research methodology used in this research consists of four main phases based on the KDD data-mining process flow by Fayyad et al. (1996), as illustrated in Figure 3.1: (1) First, the data sets of two existing DRSs were collected for as feasibility study (Chiang and Huang, 2015; Hsu et al., 2012). The first data set, referred to as the Chiang Mai POI data set, was collected from various travel websites; it contains information about POIs around the city of Chiang Mai. The second data set, referred to as the Annual Survey of Visitor Expenditure and Trends in Taiwan data set, was obtained from the Survey Research Data Archive. This data set contains five factors that influence the selection of tourists' favourite destinations in Taiwan. We used the first data set to develop a first DRS prototype and the second data set to develop a second prototype. Regarding the data collection for the proposed DRS, a pilot study was used in this phase to investigate user requirements and the design of the proposed DRS architecture. After that, a questionnaire was developed based on what we learned from implementation of the destination TRSs prototypes using the obtained data sets.

The designed questionnaire which contains six factors (five of which were used to predict tourists' preferred destinations in Chiang Mai, and one to increase levels of user satisfaction with the proposed DRS), was distributed and collected from 20 tourist destinations in Chiang Mai, Thailand. (2) After data had been collected, they needed to be pre-processed, using several data pre-processing techniques involving data cleaning, data transformation and feature-selection processes. (3) The third phase includes data analysis processes. A series of experiments was carried out to develop a DRS that required minimal input from the user but still achieved high recommendation accuracy. These experiments were conducted to identify suitable features and find optimal models from different classification of algorithms, as well as to evaluate the classification of combination methods. Once optimal models were obtained, they were validated with several validation methods, which are described in detail in Section 3.8. (4) The last phase involved interpretation of the results. The decision models were converted to set of decision rules for the development of an interactive, responsive and informative Web and mobile application in order for a tourist to interact with the proposed DRS.

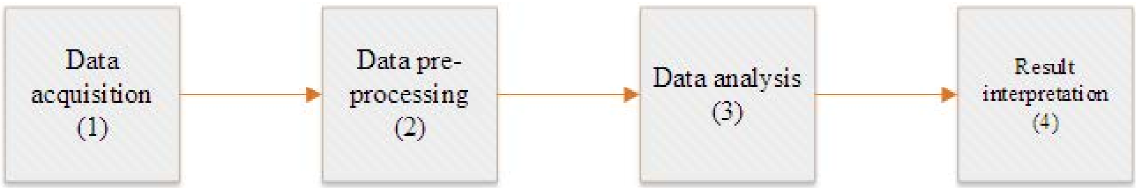


Figure 3.1 The proposed DRS framework using data mining process flow

3.2 Overview of the system design

Figure 3.2 presents the proposed system architecture for the DRS, which is a Web-based three-tier architecture model, more commonly known as client-server architecture. The architecture, which is composed of three layers, consists of presentation, application and data layers. The presentation layer is the user interface which was implemented with Web-browser technology. This layer receives inputs (e.g. demographics, user characteristics, user requirements) from tourists and displays the results to the users. The second layer is the application layer, and this acts as a middle layer. It is responsible for optimisation and logical decision-making, as well as data, evaluation and other calculations**.** The data layer takes and stores all the information from the upper layers. The information and data, such as geographical data and user and trip information, are stored in different layers using eXtensible Mark-up Language (XML) and JavaScript Object Notation (JSON). Moreover, the proposed system effectively supports mobile users.

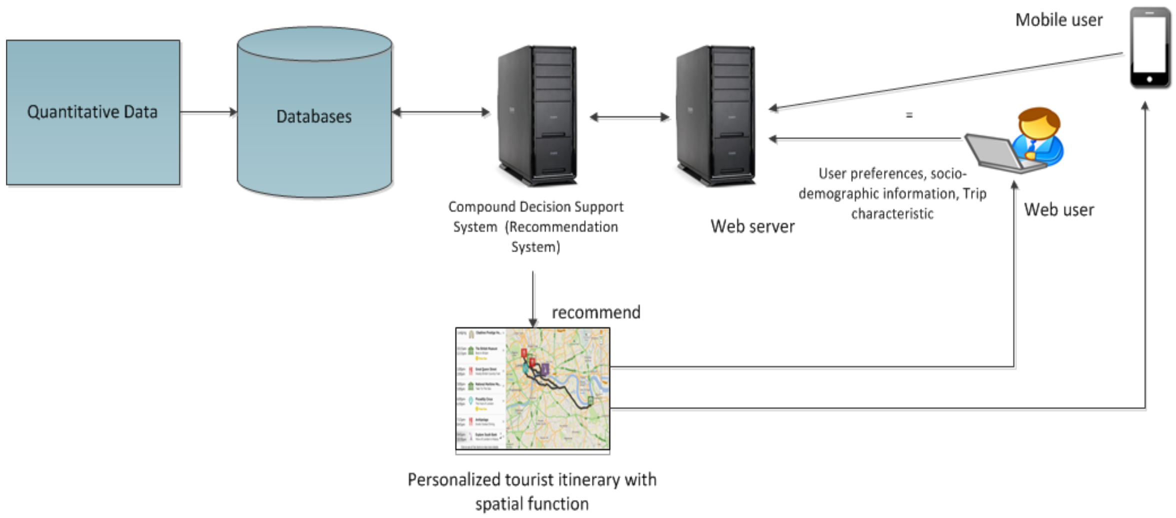


Figure 3.2 The proposed DRS system architecture

3.3 Data set acquisition

Three different data sets were collected for use in the development of the proposed DRS. The first two data sets were used to build DRS prototypes and were considered a benchmark for our data collection. The first data set, referred to as the Chiang Mai POIs data set, is a small data set containing all relevant information for trip planning, including destination names, geographical data including longitude and latitude, and attraction type. These data were collected from the Internet. The second data set was obtained from the Survey Research Data Archive (SRDA), available at https://srda.sinica.edu.tw, and is referred to as the Annual Survey of Visitors Expenditure and Trends in Taiwan data set. This data set was used as a benchmark to understand the factors that influence a tourist's preferred destination choice.

For the proposed DRS, five factors that influence tourists' preferred destinations were investigated, including travel characteristics, tourist expenditure behaviour, tourist behaviour and tourist demographic information. Additionally, user satisfaction factors were investigated and used in the results recommendation phase (i.e. the average satisfaction values of *n* users along with recommended destinations). The third data set, which is ours, was based on a questionnaire survey of 4,000 participants (both international and domestic) in Chiang Mai, Thailand. The questionnaire was designed to understand tourist-destination choices and levels of destination satisfaction in Chiang Mai by identifying the weaknesses of previous data sets, reducing irrelevant variables and adding more factors that are related to a tourist's preferred destination search in Chiang Mai. The following section describes each of the data sets in detail.

3.3.1 Chiang Mai POI data set

For this data set, information pertaining to 187 attractions, 48 hotels and 40 restaurants was manually collected from the Internet. Each POI's details include the name, description, address, longitude, latitude, type, opening time, closing time and opening days. This data set was used for the first prototype DRS (Section 4.1) to understand the current design of the DRS and recommendation methods, such as similarity measurement, trip planning and so on. Table 3.1 represents a description of the data including a sample of the attractions, hotels, and restaurants that were collected for this data set.

Table 3.1 Descriptions of attraction samples collected for the Chiang Mai POI data set.

POI name	Address	Description	Latitude	Longitude	Type
Patara Elephant Farm	135 Moo 10 Suthep Chiang Mai 50200 Thailand	This unique 14th-century temple is built into the side of Suthep mountain and is constructed of a series of tunnels.	18.78491	98.951175	Outdoors

Table 3.2 Descriptions of hotel samples collected for the Chiang Mai POI data set.

POI name	Address	Description	Latitude	Longitude	Price	Type	#stars
Ping Nakara Boutique	135/9 Charoenprathet Road |	The hotel's graceful gingerbread	18.7799	99.0047	7,900	Romantic	5

Hotel & Spa	Changklan Chiang Mai 50100 Thailand	architecture is accentuated by hand-carved fretwork and creates a relaxed and restful environment.	

Table 3.3 Descriptions of restaurant samples collected for the Chiang Mai POIs data set.

POI name	Address	Description	Latitude	Longitude
Anchan Vegetarian	Nimmanahaeminda Road opposite Soi 13 Opposite, Chiang Mai 50200, Thailand	We provide our clients with vegetarian meals so delicious you won't miss the meat.	18.79726	98.96536

Table 3.3 continued

Open-time	Close-time	Open day	Minimum price	Maximum price	Food type
11:00	17:00	MTWTHFS	10	30	Thai, vegetarian

3.3.2 Annual survey of visitor expenditure and trends in Taiwan data set

This data set is used as a benchmark for this research. The data set was obtained from the Survey Research Data Archive and was drawn from the "Annual Survey Report on Visitors Expenditure and Trends in Taiwan", https://srda.sinica.edu.tw. The data set contains information about the consumption behaviour of tourists during their stay in Taiwan and includes trip characteristics, trip plans, tourist behaviour and expenditure behaviour, along with demographic information. The time frame of the sample was from 1 January 2010 to 1 December 2012. It contains 270 variables and 12,024 cases. Hence, by using this data set as a benchmark, we saved time in the process of data collection and analysis. Our questionnaire contained fewer questions that were better related to the predicted variables. In this data set the factors that influenced tourists' favourite attractions included the four following factors:

1. Travel characteristics
2. Tourist expenditure behaviour
3. Tourist behaviour
4. Tourist demographic information

With regard to the second TRS prototype (see Chapter 4), we implemented the Annual Survey of Visitor Expenditure and Trends in Taiwan data set to provide practical aspects when recommending destinations to tourists. By using more factors than the Chiang Mai data set, including demographic information, tourist behaviour, spending behaviour and trip characteristics, the TRS provided a sense of the recommendations and a better level of performance. However, some variables from this data set were considered redundant and not related to tourists' preferred destination variables.

3.3.3 Chiang Mai Destination Data Set

This study used a questionnaire in the data-collection process as questionnaires are known to be effective mechanisms for collecting information from tourists (see Appendix A). A pilot study (see Appendix B) was also used as a pre-study in order to avoid overlooking errors.

3.3.3.1 Ethical issues

The study involves human interaction during the data collection process. Therefore, ethical issues were taken into consideration. Before distributing the questionnaire in the survey area, respondents were given a brief introduction to the study, and told the time that was needed to complete the questionnaire. Respondents were fully informed that the survey was completed anonymously and confidentially, and they would not be identified via any of their responses to the survey. They were also informed that they could withdraw at any time during the study if they wanted to. An ethical checklist (see Appendix D) was approved by the Faculty of Science and Technology, Bournemouth University, UK, before the data-collection process began.

3.3.3.2 Questionnaire design

The main aim of this questionnaire was to investigate the set of factors that influenced tourists' preferred destination choices as identified in the literature review and the set of factors from the Annual Survey of Visitor Expenditure and Trends in the Taiwan data set. In the study, motivation factors were added as a predictor of destination choice. The second aim of this questionnaire was to ascertain the level of tourist satisfaction with their preferred destinations using the set of factors found in the literature. In this research study, information regarding user satisfaction is used in the last phase of the research study to increase the level of user satisfaction with the proposed DRS.

Five sets of factors that influenced a tourist's preferred destinations were included in the questionnaire. These included a set of motivation factors, including self-actualisation, escape/ relaxation, novelty, adventure, learning experience, relationship, social status and shopping. At the end of the questionnaire, five satisfaction factors were inserted, namely, price, hospitality, food and beverages, facilities, and accessibility. The questionnaire was available in English, Thai and Chinese. The research team translated the feedback given in Thai and Chinese languages with assistance from instructors from relevant language departments. In summary, the questionnaire (45 questions in 7 sections) consisted of a set of six factors as follows:

1. *Travel characteristics* (purpose, travel party etc.)

   These variables are the most important ones when tourists select their destinations (Fesenmaier et al., 2006). They include trip length, travel purpose, trip composition, etc. Tourist characteristics include psychological, cognitive and socioeconomic status variables that influence a tourist's destination-choice process (Fesenmaier et al., 2006).

2. *Tourist expenditure behaviour*

   Trip expenditure has a significant influence on tourist destination selection (Guillet et al., 2011). These variables include the total expenditure that a tourist allots to trip and is divided into several parts (i.e. shopping, accommodation etc.)

3. *Tourist behaviour* (preferred activities etc.)

   These variables also include psychological, cognitive and socioeconomic status variables that influence a tourist's destination-choice process (Fesenmaier et al., 2006).

4. *Travel motivation* (escape, adventure etc.)

   Based on the literature, travel or tour motivation was found to be one of the most important factors for a tourist when selecting a destination. This variable describes the reasons why a tourist chooses to visit a particular destination (Leiper, 1990).

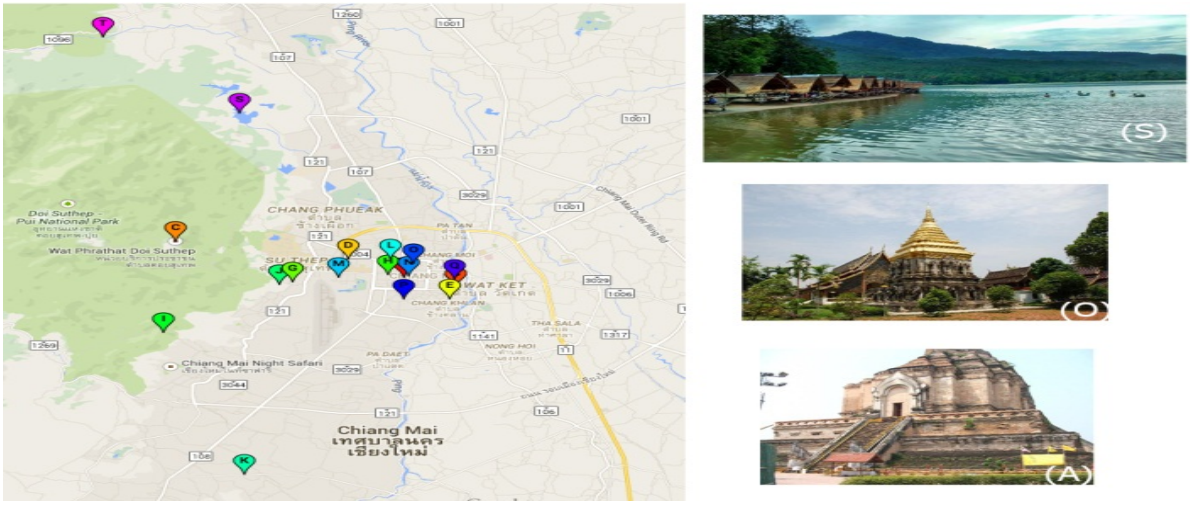
5. *Tourist satisfaction* (price, food etc.).

   These variables have a value range from 1 to 5. They were used in the results interpretation phase. For example, recommended destinations were presented to the user along with an average user-satisfaction value.

6. *Tourist demographic information* (age, gender, household income etc.)

   Individual demographics may influence information-seeking behaviour (Andereck and Caldwell, 1994).

Regarding the most popular tourist attractions in Chiang Mai, we obtained a list of attractions from the TripAdvisor website (www.tripadvisor.com) in the middle of August 2014. At that time, the website had 112 attractions in Chiang Mai tourist-ranked by registered users. We selected the top 20 tourist attractions and used knowledge acquired from a Chiang Mai tourism domain expert to validate the list we had obtained. Among the top 20 attractions, Wat Chedi Luang (see Fig. 3.3, indicated as A) was ranked number one, and Mae Sae Waterfall was ranked number 20 out of 112 attractions in Chiang Mai.



Label	Destination
A	Wat Chedi Luang
B	Chiang Mai Cabaret Show
C	Wat Phra That Doi Suthep
D	Museum of World Insects and Natural Wonders
E	Art in Paradise, Chiang Mai 3D Art Museum
F	Doi Inthananon
G	Wattana Art Gallery
H	Wat Phra Singh
I	Wat Phra That Doi Kham
J	Wat Umong
K	Wat Sri Suphan
L	Wat Lok Molee
M	Wat Suan Dok
N	Wat Pan Tao
O	Wat Chiang Man
P	Documentary Arts Asia
Q	Burklerk Gym- Muay Thai Training
R	Bua Thong Waterfalls
S	Huay Tung Tao Lake
T	Mae Sa Waterfall

Figure 3.3 Examples of top tourist-preferred destinations in Chiang Mai, Wat Chedi Laung (a) and Wat Chiang Man (b)

3.3.3.3 Survey sites

Four thousand questionnaires were distributed and collected at the top 20 most preferred tourist destinations in Chiang Mai, Thailand. The survey was distributed to both international (60%) and domestic tourists (40%) at 20 of the destinations. The participants took an average of 15–30 minutes to complete the questionnaire. To ensure that the questionnaire could be completed in an appropriate time frame and to check whether respondents would understand the terminology used in the questionnaire, a pilot test was first conducted with 350 questionnaires distributed at three tourist destinations. After that the survey was adjusted based on the pilot-study results. We then distributed 4,000 copies of the adjusted version at the 20 tourist-preferred destinations we had selected. Thirty-five samples were rejected as incompletely, 3,965 valid questionnaires, with 145 variables, were imported to the data pre-processing stage.

3.4 Pilot study

A pilot study was devised and distributed. The pilot study aimed to investigate users and the design of the proposed DRS approach. The objectives of the pilot study were to check the appropriateness of input parameters and the output of the proposed TRS in order to gather user requirements, check the research questions/ problems and identify any potential new ones.

The pilot study used a questionnaire with 20 open-ended questions and was administered over the duration of one hour. It was given to five selected participants. The pilot study was conducted as follow:

1. Participant introduction
2. Introduction to the personalised recommendation system
3. Open-ended questions

From the pilot study, we found that the Internet is users' primary source of information when planning a trip. It was also determined that having access to a personalised recommendation system would be a user's optimum objective.

Users felt that recommendation systems help individuals when facing difficult tasks and that they need to be extremely comprehensive, as in a holistic plan. Previous information that collected from the experiences of tourists has played a major role in developing a better system to assist users in making decisions. What also emerged from the pilot study is that the participants wanted software that has the most up-to-date information about points of interest. Regarding the system platform, a comprehensive platform is critical for the implementation of this service, as are efficiency of user interaction and software simplicity. Regarding appropriate input that a user is willing to feed into the system, users are more likely to provide input that does not include private or personal details, e.g. dates, budget etc. Individuals typically did not want to share specific details that are needed for establishing a demographic model, e.g. name, gender, race, home address, profession and date of birth.

Regarding the output of the system, the users would prefer it to be in the form of a summary of trip results with a combination of graphical visuals and a display of text. The presentation of results was very important and should be easy to understand. All the participants agreed that tourists would get the most benefit from the proposed system. Users would prefer to use the system before the trip began, but a system that lets the user adjust the plan during the trip was also considered significant. In addition, it has to be made available as a mobile application for the convenience of the user. Regarding the user feedback mechanism, a scaling and comment/ review function, or a combination of both seemed to be the most desirable.

In conclusion, user privacy, group recommendations, user interaction with the system, mobility, integration of heterogeneous information, and the desire for a holistic trip plan were found to be the most important common issues for the participants.

3.5 The proposed DRS framework

This section describes the proposed DRS framework (see Fig. 3.5). The proposed framework consists of five sub-systems based on a data-mining process flow: 1) data acquisition, 2) data pre-processing, 3) feature selection, 4) classification and model construction and 5) results interpretation. In terms of acquisition, the designed questionnaire was distributed among visitors to Chiang Mai, Thailand. The collected data were then pre-processed using a variety of data pre-processing methods: data cleaning, data transformation and feature selection methods. The process of data analysis involved several classification algorithms such as DT, SVM and MLP that serve as classifiers and used to develop optimal destination choice models, as well as decision rules. To improve recommendation performance, individual classifiers were combined using several combination methods. The proposed system was evaluated using several measurements, e.g. an accuracy matrix, a confusion matrix etc. Decision rules were passed on to the user interface engine to generate a Web user interface based on the given models.

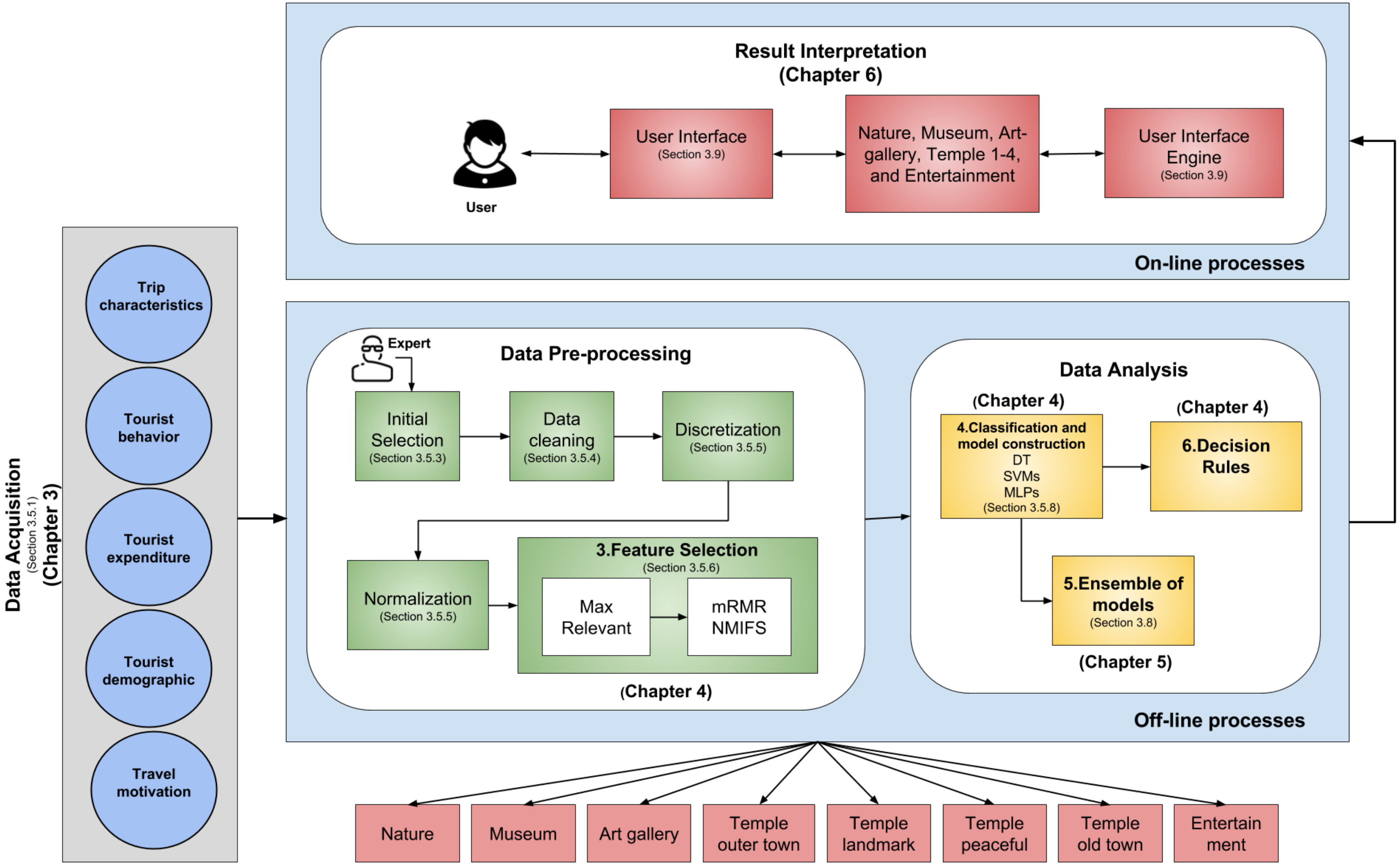


Figure 3.4 System framework of the proposed destination recommendation system for tourist.

3.5.1 Data acquisition

The proposed framework uses five factors as input variables, these were extracted from the questionnaire, as mentioned in Section 3.3.3. These were then employed as inputs to determine the classification of the tourist's preferred destinations. The potential inputs included travel characteristics, tourist behaviour, tourist expenditure behaviour, travel motivation and tourist demographic information. User satisfaction factors were used in the results presentation phase (Section 3.9).

3.5.2 Data pre-processing

Real-world data are incomplete, noisy, and inconsistent. For example, with surveys like ours, respondents may intentionally submit incorrect data because they do not want to submit personal information, or there may be data-entry errors. The best prediction results require good quality data. To achieve this, we pre-processed the survey data through data integration, cleaning, transformation, and reduction.

Data pre-processing – analysing missing values, identifying or removing outliers, discretising and resolving inconsistencies – is one of the most important components of data pre-processing. Data cleaning for this work consisted of six steps. The first step involved correcting inconsistencies in the data by selecting only relevant inputs and using tourism domain knowledge taken from the literature review. The aim of the second step was to remove cases and variables with many missing values. The third step aimed to smoothe noisy data by removing any extreme values. The next step involved reducting of a number of values of continuous features using a simple binning technique. Some features needed to be normalised, aggregated and generalised.

The last step aimed to reduce the dimensions of the data set by removing redundant and overlapping features that did not add to prediction power. For example, a user need only enter a few relevant inputs to obtain decent recommendation results from the system (i.e. the user only needs to enter three inputs instead of around 50 inputs to acquire the same recommended results. This can be achieved through this data pre-processing step).

3.5.2.1 Initial selection

The initial selection is the first step in the process of cleaning the data. In this phase, knowledge acquired from tourism domains is used to select variables that are not related to output classes. For example, satisfaction variables, survey location, survey date, comment, and survey ID were excluded from the data set.

3.5.2.2 Missing values

Missing values can significantly affect data analysis. Therefore, before proceeding to the next step, we considered simple remedies for deleting offending cases and variables with excessive levels of missing data. Based on Jr et al. (2009), we used the following rules to remove missing cases and variables:

1. Cases that involved missing data for dependent/predicted variables were deleted to avoid any artificial increases in their relationship with the independent variables
2. Variables missing at least 10 percent of data were candidates for deletion
3. Cases missing more than 15 percent of data were candidates for deletion.

For variables that are classified as Missing At Random (MAR), the imputation method was used to replace missing values. This stage was done to estimate missing values based on valid values of other variables or cases in the sample. One of the most popular methods used is mean or mode substitution. The advantages of using the mean/mode substitution method are that it is easy to implement and provides all cases with complete information. The mean and mode substitution method is best used when a variable has relatively low levels of missing data. The remedy which this study selected was mode substitution.

3.5.2.3 Outlier and extreme values

Outlier and extreme values usually appear in a data set. They neede to be identified and removed to reduce the variance of the models. For the Chiang Mai data set, they were acquired from the data entry process: 1,443 outliers were detected by combining an automated script (see Table 3.4) and human inspection. They were replaced manually by using original values from the corresponding questionnaire.

3.5.2.4 Data transformation using discretization and normalisation

The justification for using discretisation is that many algorithms do not perform well for continuous variables; therefore, they need to be converted into discrete variables. Continuous variables such as expenditure behaviour, contain many outliers and extreme values. We were not concerned with these values, we were more concerned with the range of values for each continuous variable that were significant for our purpose.

In this research study, two discretization methods were applied. The first discretisation method is referred to as simple binning. It divides the range into N intervals of equal size. Let A and B be the minimum and maximum values of a variable; then, the width (*W*) of the interval is defined as:

$$W = \frac{(B - A)}{N} \tag{3.1}$$

The second discretization method is applied to sort the data and partition them into equal sizes of bins; then each bin is smoothed using mean average sums. The third binning method involved the expert in the domain, setting the number of bins (i.e. categories) manually. The last binning method (Peng et al., 2005) is applied to handle continuous variables as described in the equation below, where the selection of a value for the variable *alpha* will have an effect on the process of feature selection, and this can be calculated as:

$$x = mean \pm alpha \times std \tag{3.2}$$

Table 3.4 Example of discretisation with regards to annual household income

Range	Description	Label
Less than $0	Very low income	1
$0.00–$49.99	Low income	2
$50.00–$99.99	Lower medium income	3
$100.00–$249.99	Medium income	4
$250.00–$499.99	Upper medium income	5
$500.00–$999.99	High income	6
$1000.00–$2000.00	Very high income	7

The main purpose of this process was to help improve the performance of the data mining algorithms. Three data-normalisation methods were applied: min-max normalisation, z-score normalisation and normalisation using the domain expert. However, the selected method depends on the chosen classifier. For example, min-max normalisation and z-score normalisation are particularly useful for the classification of algorithms involving support-vector machine neural-networks, such as nearest neighbour classification (Al Shalabi and Shaaban, 2006). However, they may not be very useful when using a DT as a classification model. It may help to increase the accuracy and simplicity of a tree model, but it may present difficulties with regard to data visualisation.

Min-max normalisation is done to perform a linear transformation of data to certain values, usually 0 and 1 or -1 and 1. Min-max normalisation is defined as:

$$Normalized(f_i) = \frac{f_i - F_{\min}}{F_{\max} - F_{\min}} \tag{3.3}$$

Z-score normalisation performs a linear transformation of data using mean and standard deviation. Z-score normalisation is defined as:

$$Normalized(f) = \frac{\left(f - \overline{f}\right)}{s} \tag{3.4}$$

Regarding the third method, data are scaled to a specific range based on the knowledge of the domain expert. For instance, a variable that describes ‘country of the user’ may contain 16 categories/countries. Hence, the data in the variable can be scaled as shown in Table 3.5.

Table 3.5 Data normalisation using expert knowledge

Country Type	Country name(s)	Label
Developed	Singapore, Korean, Japan, U.S.A, U.K., France, Germany, Sweden, Australia	1
Developing	China, Malaysia, India	2
Undeveloped	Laos	3
Domestic	Thailand	4

3.5.3 Feature selection

Feature selection is an important step in data pre-processing before moving on to the data-analysis process. It involves selecting a subset of relevant features for constructing classification models by removing irrelevant and redundant features. A feature-selection technique provides many benefits, e.g. improving the performance of a machine-learning algorithm, reducing the cost of data storage etc.

Feature selection has been used in many areas of research where data sets involve numerous variables, e.g. text processing and gene-expression array analysis (Guyon and Elisseeff, 2003). Feature selection was required in this study to better understand which variables/ features played important roles, to improve recommendation performance, to reduce the number of necessary user inputs, and to increase the performance of the classification model. An independent variable that is unrelated to the dependent variable is known as an irrelevant feature whereas an independent variable that is not useful is known as a redundant feature and needs to be removed before constructing a model (Hussein and Thomas G. Dietterich, 1991). There are three types of feature-selection techniques including filter, wrapper, and hybrid methods. In the filter method, variables are ranked and selected independently before being passed to a classification algorithm to be used. In the wrapper method, variables are selected by taking the classification algorithm into account. Last is the hybrid method in which variables are first selected using a filter method, followed by a wrapper method.

Mutual Information (MI) (Shannon, 2001) is a measure of the dependence on the amount of information one discrete random variable contains about another. MI was used to measure the similarities between set independent variables and dependent variables/ class variables. If they were found to be mutually independent, the MI value was zero. The greater the MI value, the more significant the dependent variable was. MI was used in our proposed TRS in the process of ranking features.

In this study, we carried out a two-step filtering method based on MI to rank features (first step) and remove irrelevant and redundant features (second step) from the data set. The Max-Relevance feature selection algorithm (Peng et al., 2005) was used in the first step, and the Minimum-Redundancy Maximum-Relevance (mRMR) (Peng et al., 2005) and Normalized

Mutual Information Feature Selection (NMIFS) (Estevez et al., 2009) algorithms were used in the second step. The feature selection method is described in Section 4.2.4.

3.5.4 Sampling strategy

Sampling is the primary technique used in data-mining or machine-learning to acquire a subset of a data set. In this research we used sampling for the purposes of creating training, validating and testing data sets for the model. The training data set was used to build the model, and the testing data set was used to evaluate the model (i.e. to make sure that the model performed well for any unseen data). For a real-world and imbalanced data set like ours there are many sampling strategies that have been developed by researchers to handle imbalanced data such as under-sampling, over-sampling and synthetic oversampling (SMOTE) (Chawla et al., 2002). In this study we used stratified sampling to reduce sampling errors and avoid any sampling biases that are usually generated by simple random-sampling methods. Stratified sampling is the most suitable method for the model selection process (Kohavi, 1995). In stratification sampling the divided data set contains the same proportions of the original classes.

1. *Hold-out method*

   The hold-out method is the simplest validation technique, sometimes known as a standard random sampling method. In this method, the data set is usually split into two partitions or sets – training and testing. However, the method has a few drawbacks. First, it wastes many samples from the original data set in dedicating them to the testing set. Second, over-specialisation may occur with the training set. In other words, the training set does not effectively represent the whole population of the data set. In this research, we used the hold-out method to split test data from the data set. This independent test data set was used to estimate the generalizability of the model.

2. *Repeat hold-out*

   To avoid over-specialisation of the hold-out method, we randomly re-sampled several times to generate the best representation of the population; we refer to this method as repeated hold-out cross-validation.

3.5.5 Classification and model construction

In this research study, we investigate three traditional classification algorithms – DT, SVM and MLP. In a set of given data, D = (x_i, y_i), $i = 1,\ldots,n$, x consists of the selected features from the previous stage and y is the destination associated with x, where $y \in \{c_1,\ldots c_n\}$ for n destinations. The input D is separated into two parts. One is called the training set, the other the testing set. The training set is used to train the model and the testing set is used to estimate the classification performance of the trained model. There are two main processes in the model construction: model selection and model assessment processes. In the process of model selection the training set is used to construct the model and the classifier's hyperparameters need to be tuned to obtain the optimised model, usually via cross-validation defined as follows:

$$CV(\theta) = \frac{1}{n}\sum_{k=1}^{K}\sum_{i \in F_k}\left(y_i - \hat{f}_{\theta}^{-k}(x_i)\right)^2 \tag{3.5}$$

Then we select the value of the tuning parameter that minimise CV error, defined as:

$$\hat{\theta} = \underset{\theta \in \{\theta_1,\ldots,\theta_n\}}{\arg\min}\, CV(\theta) \tag{3.6}$$

In the process of model assessment, cross-validation is used to estimate prediction accuracy value. In other words, cross-validation produces good estimates of the prediction accuracy of the model.

Each of the classification algorithms has its advantages and disadvantages, and the goal is to produce decision boundaries. For example, DT was chosen in the model-construction stage for the proposed DRS because it provides several benefits, such as simplicity, interpretability and efficiency. The relevant features of each tourist's preferred destination (e.g. nationality, household income etc.) are used to construct a model that describes the user's preferences. For the DRS, a dedicated DT can be built for each tourist's preferred destination choice. SVM is a theoretically well-founded classification algorithm and has been successfully applied in many real-world applications, e.g. face recognition, text recognition and so on. SVM is a supervised machine-learning algorithm that was originally designed for use in binary classification. The concept of SVM is based on the idea of finding an optimal hyperplane that can discriminate a data set into two classes. MLP is another supervised machine learning algorithm that extends the concept of single perceptron that has a problem with a non-linear separable. MLP, a feed-forward neural network, consists of one input layer, plus one output layer, and an arbitrary number of hidden layers located between the input and output layers. The data move from the input layer through hidden nodes to the output nodes. The MLP model is trained by a back-propagation algorithm. Lastly, an activate or transfer function is used in the network; it is a function that transforms a set of input signals into an output signal. There are several types of these activation functions such as *sigmoid* which maps input to a value ranging between 0 and 1, while *tanh* maps the input to a value ranging between -1 and 1. For multi-class classification problems the *softmax* function is used.

Comparing the performance of our DRS to other existing systems is challenging for several reasons, including the number of destinations, different cities and locations, performance criteria and the differences in evaluation methods.

1. *Number of user inputs and number of destinations*

   Existing TRSs aim to improve system accuracy and ignore practical aspects. Having a high recommendation accuracy, by eliciting a large number of inputs from a user, does not necessarily mean that the recommendation system is suffciently developed. This can easily be seen from two DRSs that applied very similar model-based approaches (Hsu et al., 2012; Huang and Bian, 2009). Both systems use similar input, but have different output. Hsu's system predicts the destination category while Huang and Bian's system predicts actual destinations. It can be seen that comparing our proposed system with others is difficult since the input and output of the system and the system goal are different.

2. *City and location*

   The city or location that the recommendation system applies plays a major role in its performance. Each city has it owns unique and complex nature. Using the same factors to associate with different destinations could produce different results. For example, tourist expenditure behaviour may not be correlated with the search process for destinations in some countries, but this factor may reveal a high correlation in other countries.

3. *Evaluation methods*

As can be seen from the literature review (Chapter 2), most existing TRSs do not provide any validation methods for their systems. The best way to evaluate a recommendation is to use an online-based method; here, one can see the direct impact of the recommendation system on the end user. A/B testing is one the methods used. However, this requires active user participation and is difficult to use as a benchmark in research. Based on the literature review, previous DRSs have used different methods to evaluate their systems, and these are mostly based off-line. For instance, Hsu et al. (2012) deployed ROC and AUC to evaluate the BN network while the Huang et al. (2009) system does not indicate how they evaluated their system. Chiang and Huang, (2015) were particularly concerned about user satisfaction and employed user studies to evaluate their system. Yeh and Cheng (2015) evaluated their system using only precision rates. To ensure there is no bias in the validation process, RSs and DRSs that apply collaborative filtering like our model-based one, need to ensure that all the ratings are evaluated using an out-of-sample approach. Methods such as hold-out and cross-validation are needed to make sure that the model is generalised enough for unseen data (Recommender Systems – The Textbook, Charu C. Aggarwal, Springer, 2016).

3.6 Ensemble of classifier methods for the proposed DRS

One promising way to solve complex problems in real life is to take votes from several experts, followed by a final decision obtained by combining their votes. This concept is also applied in machine-learning and is known as an ensemble of classifiers or ensemble learning. This method is a supervised learning algorithm that uses combination models, instead of an individual one, to obtain higher classification accuracy. Ensemble learning has been shown to potentially improve prediction performance and robustness, but this is not guaranteed (Dietterich, 2000).

3.7 Performance evaluation methods

In this research study, several performance criteria, sampling methods, and validation techniques were used to assess model performance and help in the model-selection process.

3.7.1 Measurement

In the TRS domain, especially in DRS, the most commonly accepted evaluation measures for TRS performance are accuracy, precision, recall, and f-score. In general, accuracy and error rates computed from a test data set are the main measurements used to evaluate a model's performance. Usually we want to have the model with the highest accuracy rate or the lowest error rate. However, accuracy or error rates alone do not guarantee that the test model performs well; several other measurements are also useful for comparing the performance of different models. In a multi-class classification problem, the model may obtain a decent accuracy rate but this may result in decreased performance for particular classes.

1. *Accuracy*

   Accuracy is a measurement of classifier performance. It represents the overall correctness of a model. It can be calculated as the sum of correct classifications divided by the total number of classifications, as shown in the following equation:

$$Accuracy = \frac{|TP| + |TN|}{|FN| + |FP| + |TN| + |TP|} \quad (3.6)$$

   Similarly, classifier performance can sometimes be expressed in terms of the misclassification error rate. The error rate can be calculated using the following formula:

$$Error\ rate = \frac{|FN| + |FP|}{|FN| + |FP| + |TN| + |TP|} \quad (3.7)$$

2. *Confusion matrix*

   A confusion matrix (Chawla et al., 2002), or table of confusion, contains information regarding the actual and predicted classifications generated by the classifier. Information consists of the True Positive (TP), True Negative (TN), False Positive

(FP), and False Negative (FN). The table below presents an example of a confusion matrix.

		Predict	
		Class = 1	Class = 0
Actual	Class 1	TP	FN
	Class 0	FP	TN

Figure 3.5 Confusion Matrix

3. *Precision and recall*

Using accuracy or error rate alone might be misleading in many cases, especially in real-world problems where the data set is usually imbalanced, as in our case. Imagine a binary classification problem in which there are 900 samples of class A and 100 of class B. If a classifier predicted everything to be class A, this would return a high classification accuracy rate of 90%. However, the classifier cannot detect class B. Precision and recall measures of relevance are used for evaluating classifier performance. Precision indicates how many selected items are relevant; recall indicates how many relevant items are selected. From the confusion matrix in Figure 3.5, precision and recall measurements are calculated using the following formulas (Buckland and Gey, 1994):

$$Precision = \frac{|TP|}{|FP|+|TP|} \tag{3.8}$$

$$Recall = \frac{|TP|}{|FN|+|TP|} \tag{3.9}$$

In the recommendation system domain, precision is more important than recall, as we want to achieve higher precision rather than recall (*An Introduction to Machine Learning, Miroslav Kubat*, *Springer*, 2015.)

4. *F-score*

The F-Score, sometimes known as the F-measure, represents a combination of two measurements: precision and recall (Buckland and Gey, 1994). The F-score can be thought of as an improvement in accuracy, as it takes class discrimination into account. The maximum value of F-score is 1, the lowest value is 0. The F-Score formula is presented below:

$$Fscore = 2 \times \left(\frac{precision \times recall}{precision + recall} \right) \tag{3.10}$$

5. *ROC curve and area under the curve*

The Receiver Operating Characteristics (ROC) curve (Swets, 1988), is a plot that represents the performance of a classifier by plotting TP against FP at several thresholds, as illustrated in Figure 3.7. The ROC curve has been used for comparing the performance of several machine-learning models and exhibits a number of desirable properties when compared to classification accuracy. The classifier which has a ROC curve close to the upper left is considered better than the others. On the other hand, the classifier which has a ROC curve below the diagonal line is considered worse than a random guess. According to Figure 3.7, classifier B is considered superior (i.e. better with respect to recommendation performance) to classifiers A and C.

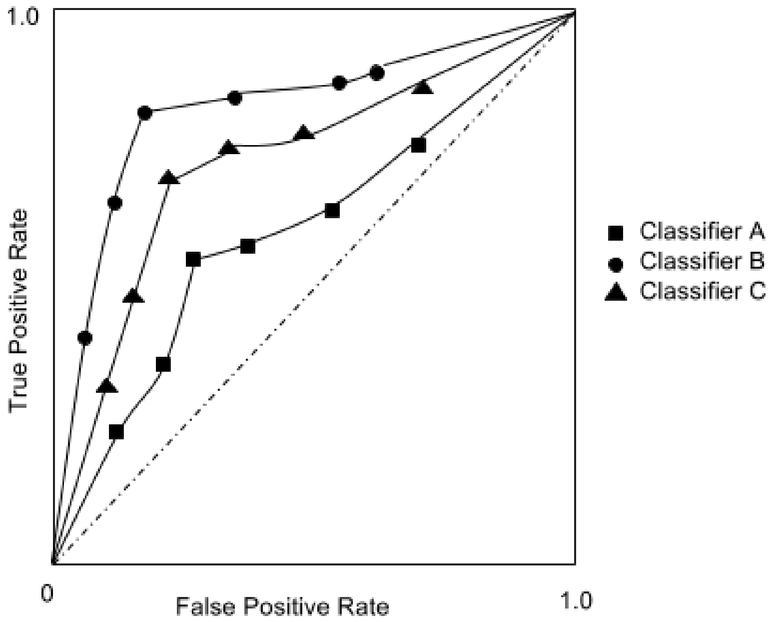


Figure 3.6 Comparison of classifiers' performance using ROC curves

The Area Under Receiver Operating Characteristics (AUROC) curve, also known as Area Under the Curve (AUC), is used as one of the metrics to evaluate the classification algorithm. AUC can be calculated by measuring the area under the AUC Curve (Bradley, 1997). AUC is used to tell how well the classification model can discriminate between two classes. The closer the value of AUC is to 1, the better the model is. A model that has an AUC value close to the baseline of 0.5 is considered useless and no better than a random guess.

3.7.2 Cross-validation

To select the optimal model, estimate the model's performance and protect against overfitting in a predictive model, cross-validation techniques were carried out in this study. We applied these techniques at the model regularisation and model assessment stages. Cross-validation, also known as a rotation estimate, is an extension of the hold-out method. This method tries to maximise training data. The simplest approach for cross-validation begins with two folds in which the data set is split into two partitions called training and testing. In the next iteration, the test data set is swapped with the training data set.

This method was generalised using *k*-fold cross-validation to split the data set into *k* partitions of approximately equal size. For each iteration, one-fold/partition was chosen to test the data set and the rest were chosen as a training data set; this process was repeated *k* times. The most common *k*-fold cross-validation involves 5-fold and 10-fold cross-validation. When choosing the number of the folds, the larger the *k* value, the less bias and high variance of the model. Leave-one-out extends *k*-fold cross-validation to another level, as the method sets *k*=N, where N is the number of samples in the data set. Leave-one-out is the most computationally extensive method. The accuracy rate of the model is estimated as the average of the accuracy of *k* models. In this research, *k* is set to 5 for all the experiments due to limited computation power.

3.7.3 Statistical tests

The purpose of using statistical tests in this study is to compare the overall performance of different classifiers and gauge the stability of the models. After the classification stage we applied two statistical tests. First, a Shapiro-Wilk normality test (Shapiro and Wilk, 1965) was used to test if the data were normally distributed. The Shapiro Wilk statistical test is defined as follows:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{3.11}$$

Where x_i is the smallest number in the sample, and $\bar{x}$ is the mean of the samples. The constant a_i can be calculated as follows:

$$(a_1,\ldots,a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}} \tag{3.12}$$

The Shapiro-Wilk method is used for samples sizes of less than 2,000. If the sample size is greater than 2,000, a Kolmogorov-Smirnov test is applied instead. Data are not considered normally distributed if the significance value is close to zero (>0.05). Next, if the data are normally distributed, a paired T-test with a 95% confidence level was conducted to determine whether the mean differences between paired samples differed by more than 0.5. Otherwise, a Wilcoxon signed rank test (Wilcoxon, 1945) was applied.

3.8 User interface development for the proposed DRS

The proposed system is based on a Web-based three-tier architecture model which is more commonly known as client-server architecture. The architecture, which is composed of three layers, consists of presentation, application and data layers. The presentation layer is the user interface, implemented via Web-browser technology, whereby it receives inputs such as demographics, user characteristics, and user requirements from tourists, and displays the results to users. The second layer is the application layer, which acts as a middle layer. It is responsible for optimisation and logical decision-making as well as data evaluation and other calculations. The data layer takes and stores all the information from the upper layers. Information and relevant data, such as geographical data and user trip information are stored in different layers by using the eXtensible Markup Language (XML) and JavaScript Object Notation (JSON) file formats. In this study, we aim to develop a UI for the proposed DRS that has adaptive, responsive, and interactive capabilities. The terms are clarified as follows:

According to Raymond, (2009), adaptation for a user interface needs to include some factors such as user performance, user goals, cognitive workload, user situation awareness, user knowledge, groups profiles, situation variables and task variables.

Decision Tree can be used as an adaptation algorithm and as one of the interface adaptation methods (Raymond, 2009). In the user interface, responsiveness refers to changes in the size of the browser window and how the content arranges itself.

Interactivity is one of the most promising aspects to consider in order to exploit the full potential of a DRS. Designing and implementing a real interactive website requires a lot of work involving collaborative attitudes of users, a clear process and standards for managing content, as well as designing research (Rubinelli et al., 2013). In this study we aim to increase the interactivity between the user and the system in order to display useful information (e.g. location of destination) to users through interactive maps. Moreover, advanced Web technologies, such as JQuery, CSS and HTML5, can be used to enhance the user experience and increase the response and interactivity of the system.

3.9 A practical Destination Recommendation System (DRS)

This section describes how the proposed DRS can be used to assist tourists during the pre-travel stage of their plan to visit an unfamiliar city. The proposed system is designed to be used by tourists and travel agents and consists of both online and offline phases. In the offline phase, the system performs a calculation of the optimal destination choice models to recommend destinations to tourists, saving the tourists additional hardware computation costs and time during the information search process. Raw data, such as survey records, are fed into the system via a data-management module. This module is responsible for integrating, cleansing, transforming, storing and maintaining survey data. Maintenance of the system simply requires feeding new data into the recommendation engine inside the data management module in this layer. For example, every year when new survey data are obtained, we can integrate it into the existing data set and new models will consequently be constructed and passed on to the Web server in the top layer. In the UI management module we can add, edit, delete or modify the models. The implementation of the administrator control panel is discussed in detail in Chapter 6.

In the Model Management module, DT classifiers and other machine learning classifiers are installed, including three well-known classification algorithms, DT, SVM and MLP and other ensemble learning models. These are used to discriminate between specific destinations in each data set. To make the complex model usable, and to interpret its results for the tourist, DT models are converted into decision rules and information is then passed to the UI management module. A brief description of the concepts and techniques of the classifiers used in this study is presented below:

In the online phase, the top layer can be considered the client layer, as it contains the user interface, where a tourist can interact with the system via different platforms such as mobile, desktop or Web browser. In the UI management module, decision rules are transformed into XML and JSON formats in order to generate a new user interface. Moreover, the system can connect to Google API to retrieve pertinent information that is related to maps and routes so that the system is able to display the results on the interface. Tourists can interact with the system via the user interface. To receive a recommended destination a tourist is required to submit a number of inputs, e.g. the trip's purpose and the user's income, as well as others, into the system by selecting from answers provided in lists. Subsequently, the recommended results will include the destination name and a travel route, which will be obtained by using the travel information obtained from the user's location and the selected destination. Geographical, spatial and route information are stored in this layer. The system connects to several Google APIs such as GMap and GLargeMap, to be able to load and control the maps.

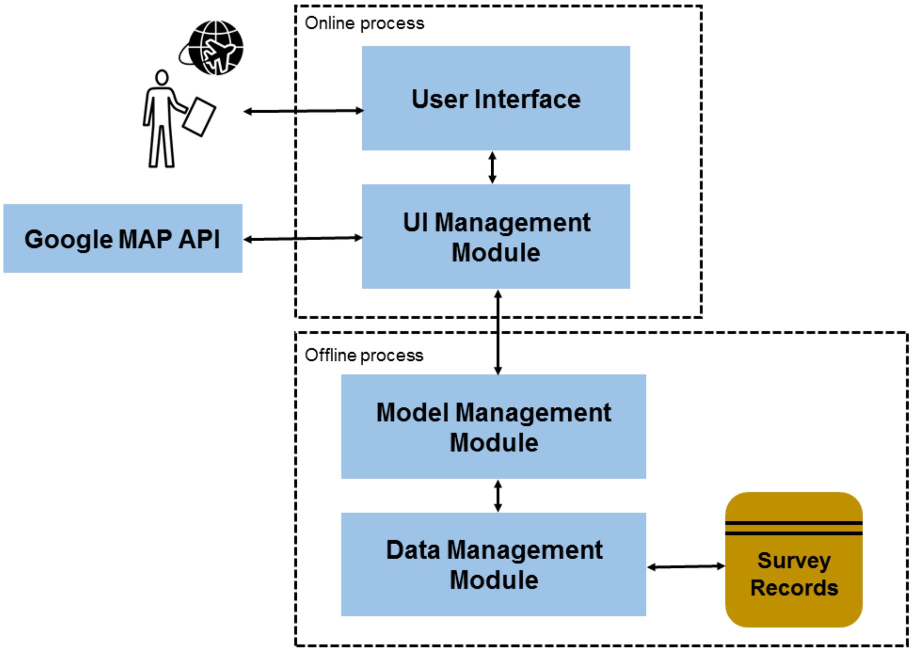


Figure 3.7 A practical recommendation system for tourists

3.10 Summary

In this chapter, the proposed research approach and system architecture have been presented. Details of the machine-learning techniques that will be used in the development of the DRS have been provided throughout this chapter. This chapter has also explained how data sets were collected, including the process designing questionnaire and the locations of survey sites. The data pre-processing techniques (e.g. initial selection, missing values, outlier detection etc.) were also discussed in detail, as well as the proposed two-step feature-selection methods based on MI to eliminate unnecessary inputs that are either irrelevant or redundant. In the TRS field, no studies has used any feature-selection methods to control input to the system. The proposed classification algorithms and technologies involved in the stage of results interpretation have also been presented.

The research approach presented in this chapter will be used for implementation of the Model-based and Ensemble-based DRS in Chapters 4 and 5, as well as the system interface development in Chapter 6.

Chapter 4 Model-Based Destination Recommendation System

This chapter consists of two parts and discusses the development of the DRS based on the proposed methodology described in the previous chapter. It begins by investigating the issues involved in developing the DRS by implementing two existing DRS prototypes. It then goes on to the process of development of the proposed model-based DRS including data pre-processing, construction of the classification mode, and system evaluation. The data set applied in this study was collected from Chiang Mai. This chapter addresses research questions 2, 3 and 4:

RQ 2. Which set of factors plays an important role in making destination recommendations for tourists? Does using multiple factors help improve recommendation accuracy? Do travel- motivation factors help to increasing the level of recommendation accuracy?

RQ 3. How can a tourist's decision-making process be understood when they select their preferred destination?

RQ 4. How can a user's efforts be reduced, while still maintaining the same degree of recommendation performance and increasing the level of user satisfaction in the decision-making process when selecting a destination?

4.1 Feasibility study of different DRSs

In seeking to investigate and analyse the results of different phases of the proposed DRS, two existing DRS prototypes were investigated, namely: Personalized Travel Planning System (PTPS) and Intelligent Tourist Attractions System (ITAS). The objectives of this feasibility study are explained below:

4.1.1 Objectives of the study

The first aim of this feasibility study was to identify existing issues in DRS development through the developed prototype and experimentation, and to determine if it was feasible to replace the BN model with our proposed DT inside the recommendation engine. The second aim was to compare existing similarity measurements from previous DRSs that shared similar types of data set, and to determine if it was feasible to use MI as the similarity measurement. The objectives below correspond to research questions 2:

1. To study the feasibility of using quantitative data for the DRS.
2. To investigate existing recommendation methods in DRSs, in both memory-based and model-based approaches.
3. To investigate and identify the factors that influence a tourist's preferred destination, acquired from data sets.

4.1.2 Personalised Travel Planning System study

We began by implementing our first prototype DRS – a user constraint-based DRS from Chiang and Huang's study, also known as the Personalized Travel Planning System (PTPS) (Chiang and Huang, 2015). Their system provides users with the novel concepts of travel planning and adjustable results by introducing a feedback mechanism, an adjustable interface, time framework and a schedule algorithm. However, we did not have time to implement them all, so our primary focus in this experiment involved the implementation of the basic matching mechanism, time framework, and their proposed Scheduling Reasoning algorithm.

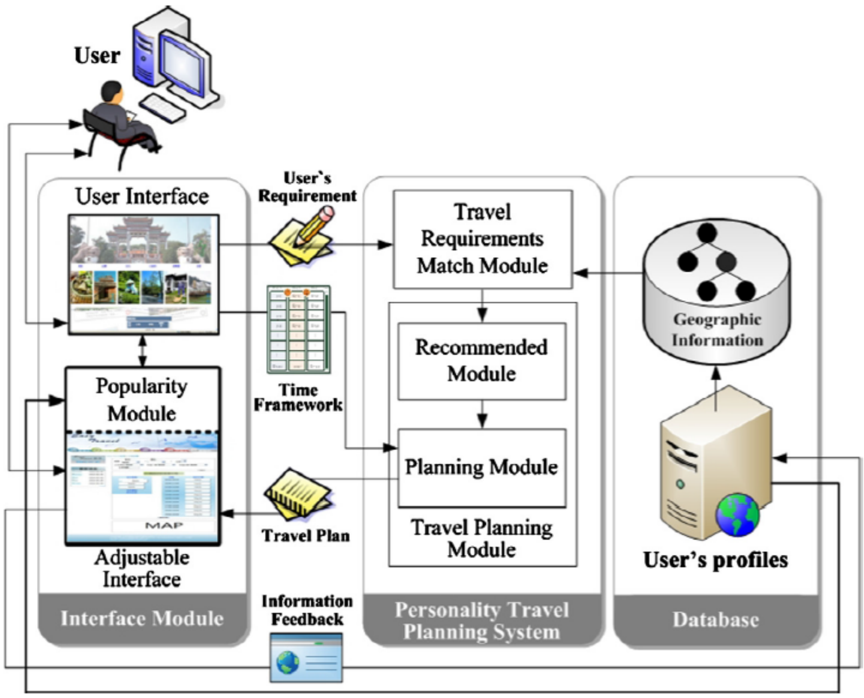


Figure 4.1 PTPS overall framework (Chiang and Huang, 2015)

As shown in Figure 4.1, the modules that this study focused on are the database module and the Personality Travel Planning System (PTPS) module, also the schedule reasoning algorithm that is used to produce a personalised travel schedule from a finite set of tourism services involving attraction locations, dining and restaurant locations, accommodation options, hotel locations, user requirements etc. The algorithm involves several steps for searching for a travel location or destination and calculations related to transportation and dwelling time. The feedback mechanism is a method applied to rank POIs (hotels, restaurants, accommodation), which is the cumulative value of user ratings of popularity.

The time framework is composed of d_s, v_z and *cl*. In this equation, d is the day number (e.g. day 1, day 3 etc.) is represented as $s = \{1,2,3..,n\}$, where v is the time block/hour, z ranges from 1 to 24, and *cl* is the category of POIs such as attractions, hotels or restaurants, respectively, represented as {A, R, H}.

4.1.2.1 Data collection and database management

Since we did not have access to the data set that was used in the afore-mentioned study (i.e. the database of Tai Chung, Taiwan, that contains all related information for the entire travel plan, such as points of interest, attractions, hotels, restaurants, time spent, geographical data and distance calculations), we had to input the data manually by collecting it from several websites and then loading it into a spreadsheet file. Specific examples of data could be: name, description, address, longitude, latitude, opening time, closing time, open days and so on.

The database system and structure of the software have been designed and are presented through an Entity Relationship Diagram (ERD) and Unified Modelling Language (UML) diagrams (see Figs 4.2 and 4.3). The UML diagram represents the overall implementation of the TRS through the following steps.

The Schedule Reasoning Method (SRM) was modified because we needed to search for location in the user requirement (A, H or R) tables first. If multiple locations were returned we picked the location that had the highest popularity value. If no locations were returned we searched for the most popular one in the A, H and R databases. The modified algorithm is presented in Table 4.2.

Entering the collected data into the database management system manually is a time-consuming task. Therefore, an Excel to SQL conversion tool implemented with JAVA language was created to handle the large amount of recorded data that needed to be inserted into the database. A front-end Web application management system was also developed for this experiment in order to manage the information in the database. The Web application was implemented with a PHP which was inter-connected with the created database. The user could then directly insert/ update/ edit records directly into the MySQL database.

4.1.2.2 Experiment setup

Table 4.1 presents the user requirements, such as choice of initial attractions, restaurants and hotels, travel type, points of departure, duration of travel, breakfast time, lunch time, dinner time, travel type, food type and the budget that the user is willing to spend.

Table 4.1 Example of user input of PTPS

Number of Days, Number of Travellers, Budget, Initial Point, Lunch Time, Dinner Time, Travel Type, Food Type, Region
3, 1, 500, Lanna Folk Life Museum, 13:00, 20:00, Literature Art, Thai Food, Chiang Mai
User Required Attraction/s
Chiang Mai Zoo, Big Game Fishing Adventure Tour
User Required Restaurant/s
NaN
User Required Hotel/s
NaN

4.1.2.3 Recommendation process

The Travel Requirement Match Module matches the user inputs (e.g. required attraction(s), hotel(s), restaurant(s) from the database). Then the recommended module executed the SRA (see Table 4.2). The following important stages were involved in three specific steps:

1. Travel location or destination searching.
2. Transportation and dwelling-time calculation (Note*:* the authors did not explain how they obtained dwellings time at the travel locations).
3. Addition of the selected travel location into the time framework.

Table 4.2 The modified SRM Algorithm

SRM algorithm:

if M(*locations*) = {A} **and** TF_{cur} (*cl*) = {A} **then**
  C_{n+1} = M(locations)
else
  C_{n+1} = Max{P(A)}
if M(*locations*) = {R} **and** TF_{cur} (*cl*) = {R} **then**
  C_{n+1} = M(locations)
else
  C_{n+1} = Max{P(R)}
if M(*locations*) = {H} **and** TF_{cur} (*cl*) = {H} **then**
  C_{n+1} = M(locations)
else
  C_{n+1} = Max{P(H)}

Figure 4.2 shows a structural diagram, known as a UML object diagram, that represents a snapshot of the system. The diagram describes the object names and their relationship in the implementation of our PTPS.

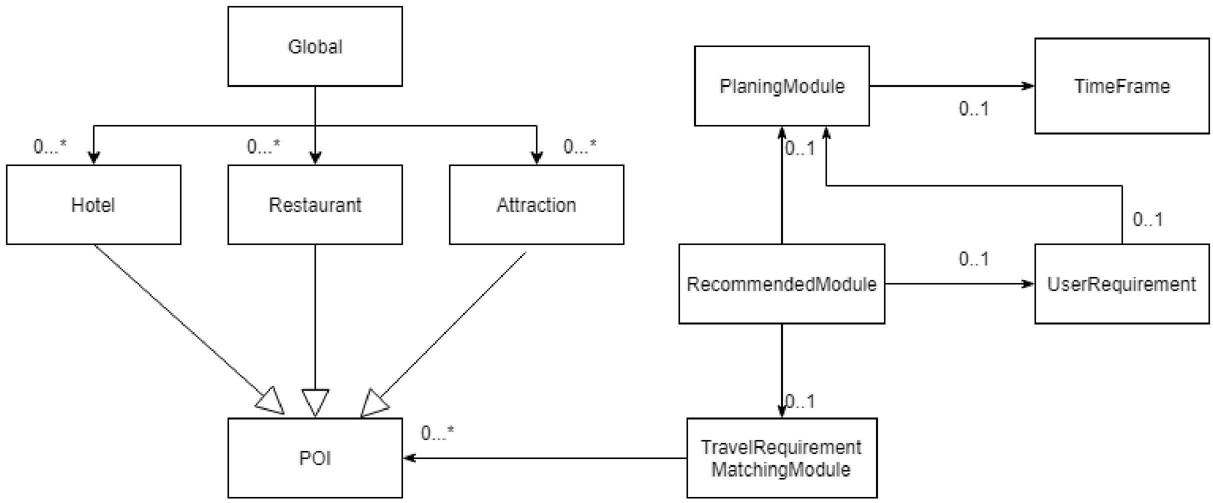
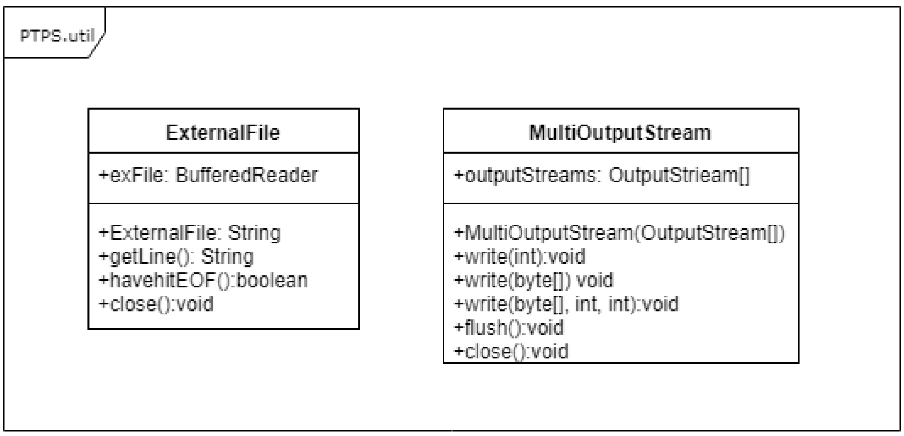
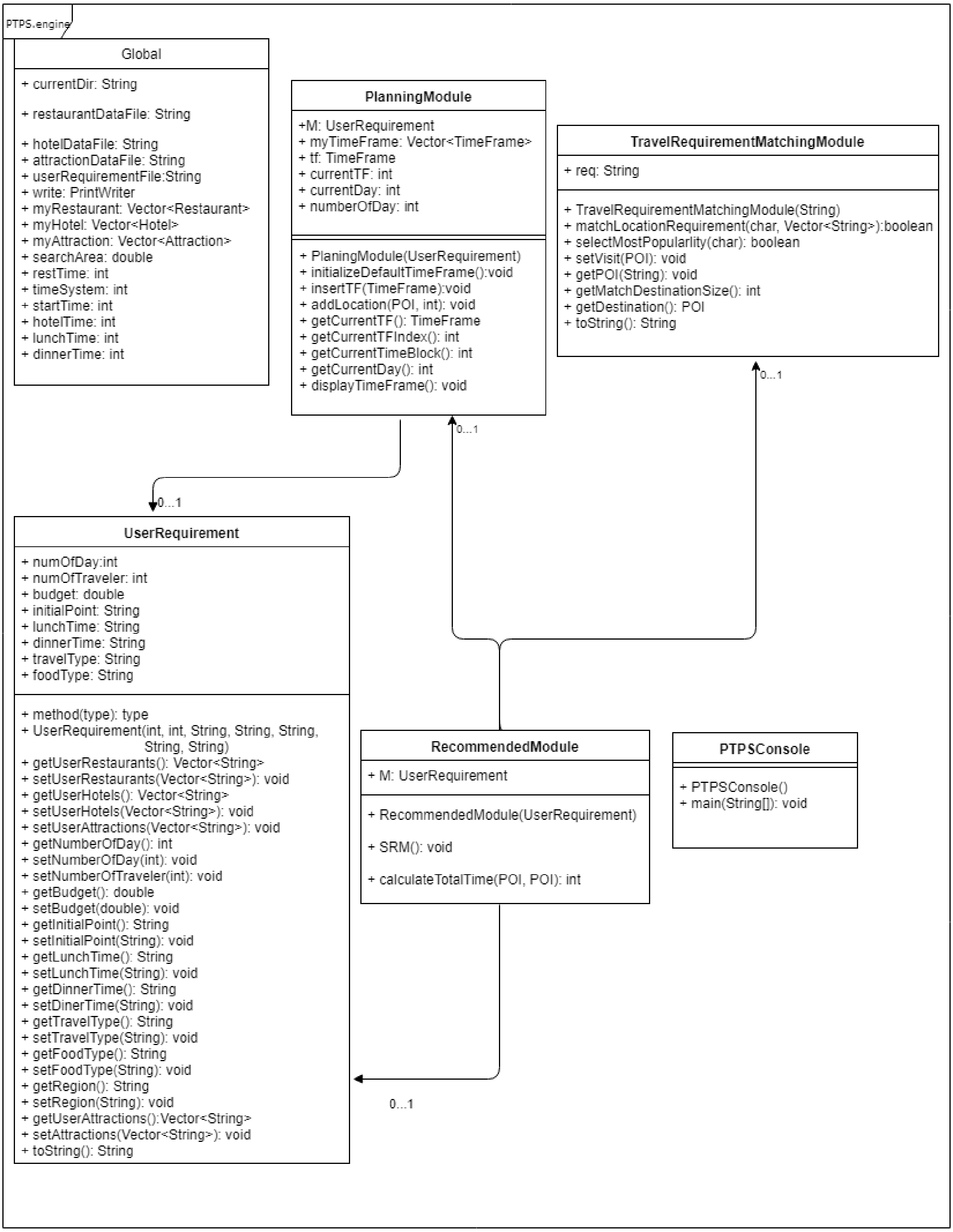


Figure 4.2 UML Object Diagram of PTPS

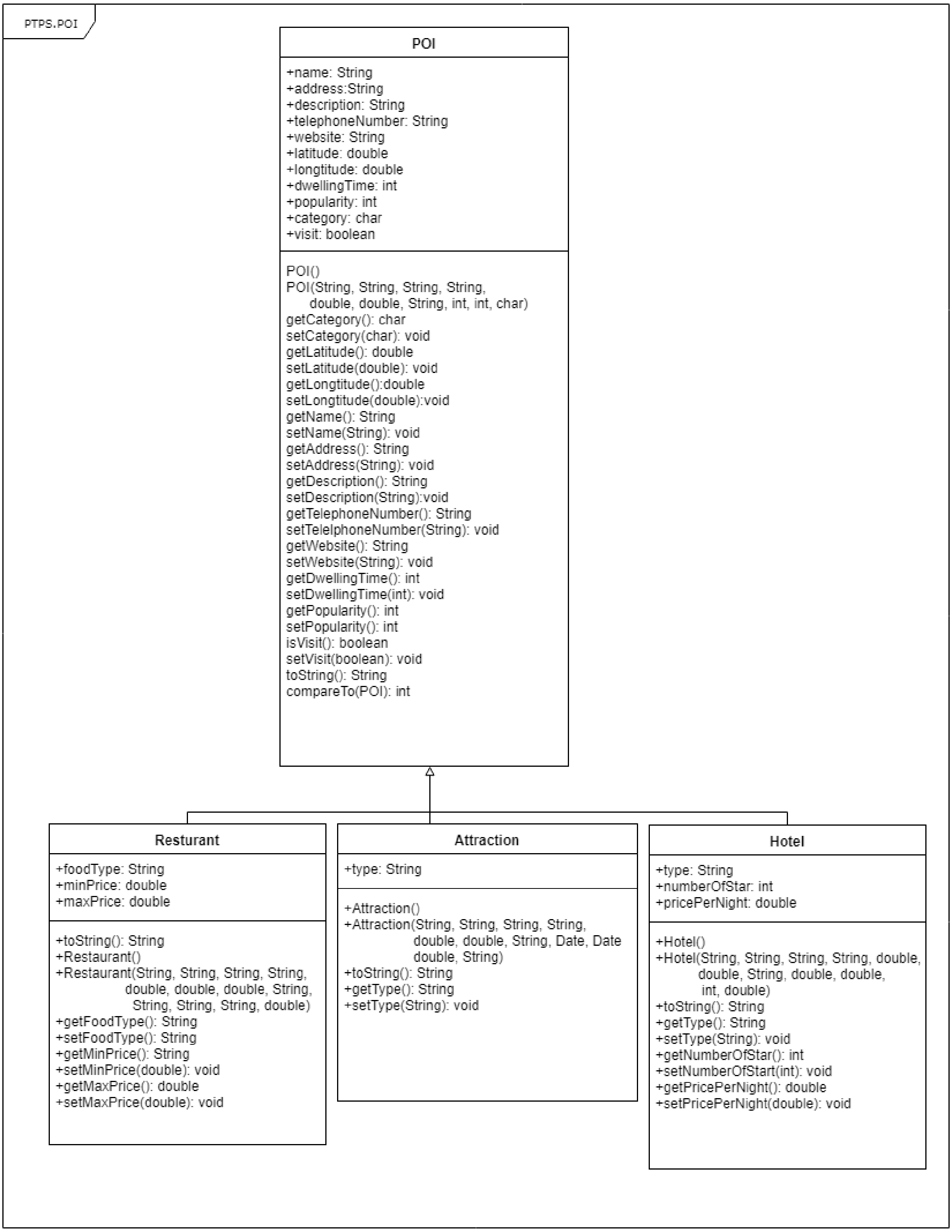
After the object diagram had been created, detailed UML class diagrams were created to illustrate the details, including the attributes and methods of each class, as well as how each class interacted with each other, along with capturing a picture of important entities in the PTPS. The class diagrams consisted of three main packages including utilities, reccommended engine and POI objects.



(a) UML class diagram of the utilities class used for reading and writing files in PTPS



(b) UML detailed class diagrams of the recommendation engine of PTPS



(c) UML class diagram demonstrates generalization between the superclass POI and three subclasses, i.e. Restaurant, Attraction and Hotel.

Figure 4.3 Class diagrams of the PTPS

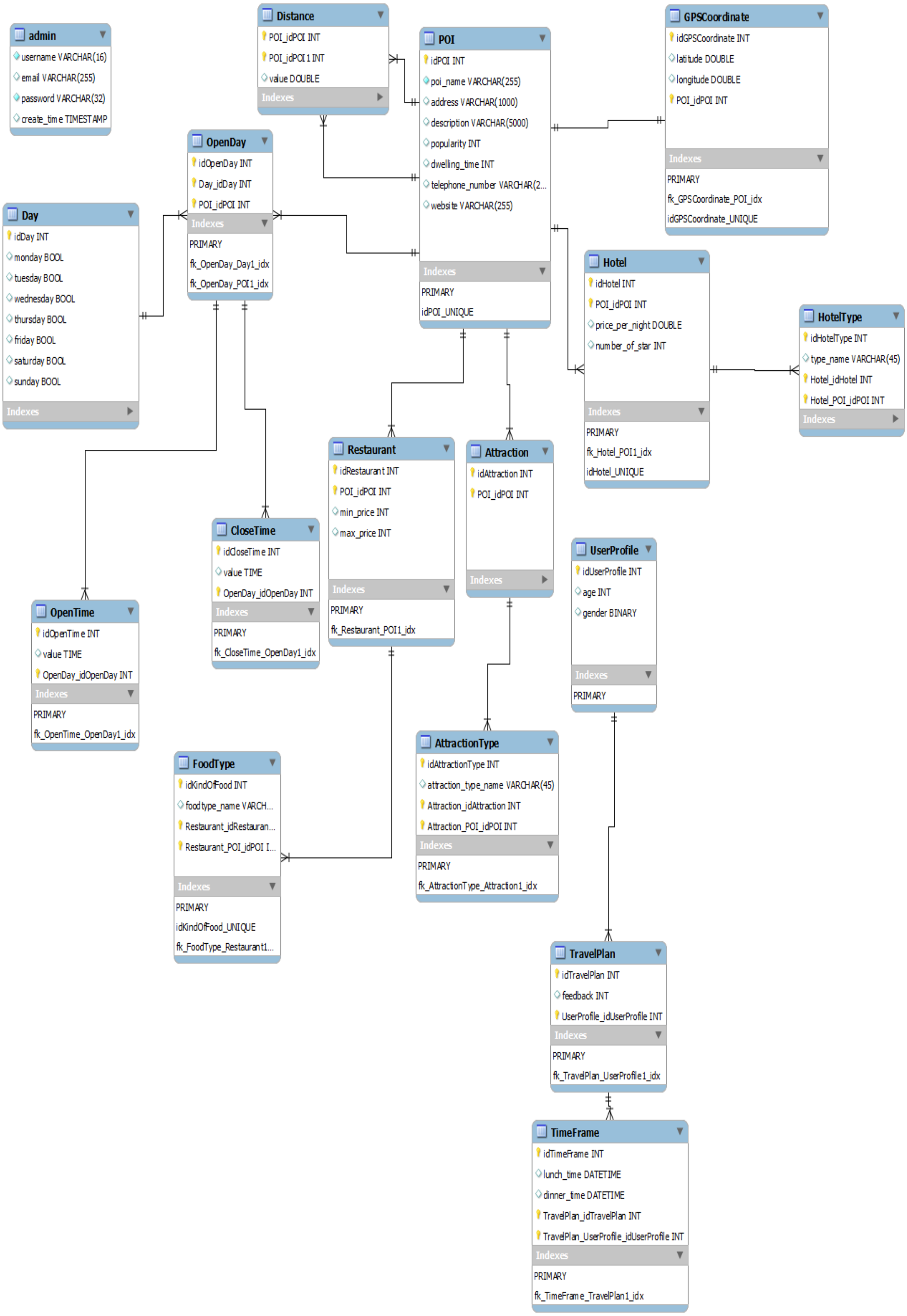


Figure 4.4 ER-diagram of the PTPS

4.1.2.4 Experimental results

From the experiment results we found that the proposed SRA suffered when a new location was added to the current schedule (i.e. when the time-frame of the new location overlapped with a lunch or dinner break). It could be seen that before an extra visit could be inserted into a tour plan, it had to be ascertained whether all the visits scheduled after the insertion place still satisfed their time windows. The total time, such as dwelling time and transportation time, for our experiment was set at one hour when traveling from one location to another. This was done because we did not have information related to the dwelling time at each location and had not implemented a program to retrieve transportation times via Google API, as this would have been a very time-consuming process (see Fig. 4.5). Moreover, the proposed algorithm does not generate a proper plan when dealing with a limited number of locations.

```
[0] TimeFrame [dayNumber=1, timeBlock=1, category=H, locationName=null]
[1] TimeFrame [dayNumber=1, timeBlock=2, category=H, locationName=null]
[2] TimeFrame [dayNumber=1, timeBlock=3, category=H, locationName=null]
[3] TimeFrame [dayNumber=1, timeBlock=4, category=H, locationName=null]
[4] TimeFrame [dayNumber=1, timeBlock=5, category=H, locationName=null]
[5] TimeFrame [dayNumber=1, timeBlock=6, category=H, locationName=null]
[6] TimeFrame [dayNumber=1, timeBlock=7, category=H, locationName=null]
[7] TimeFrame [dayNumber=1, timeBlock=8, category=A, locationName=Lanna Folklife Museum]
[8] TimeFrame [dayNumber=1, timeBlock=9, category=A, locationName=Big Game Fishing Adventure Tour]
[9] TimeFrame [dayNumber=1, timeBlock=10, category=A, locationName=Big Game Fishing Adventure Tour]
[10] TimeFrame [dayNumber=1, timeBlock=11, category=A, locationName=Chiang Mai Zoo]
[11] TimeFrame [dayNumber=1, timeBlock=12, category=A, locationName=null]
[12] TimeFrame [dayNumber=1, timeBlock=13, category=R, locationName=Himbannsoun restaurant]
[13] TimeFrame [dayNumber=1, timeBlock=14, category=A, locationName=Three Kings Monument Square]
[14] TimeFrame [dayNumber=1, timeBlock=15, category=A, locationName=Three Kings Monument Square]
[15] TimeFrame [dayNumber=1, timeBlock=16, category=A, locationName=Three Kings Monument Square]
[16] TimeFrame [dayNumber=1, timeBlock=17, category=A, locationName=Siam Insect-Zoo & Museum]
[17] TimeFrame [dayNumber=1, timeBlock=18, category=A, locationName=Siam Insect-Zoo & Museum]
[18] TimeFrame [dayNumber=1, timeBlock=19, category=A, locationName=Siam Insect-Zoo & Museum]
[19] TimeFrame [dayNumber=1, timeBlock=20, category=R, locationName=Aroon Rai Restaurant]
[20] TimeFrame [dayNumber=1, timeBlock=21, category=A, locationName=null]
```

[21] TimeFrame [dayNumber=1, timeBlock=22, category=H, locationName=Howie's HomeStay]

[22] TimeFrame [dayNumber=1, timeBlock=23, category=H, locationName=Howie's HomeStay]

[23] TimeFrame [dayNumber=1, timeBlock=24, category=H, locationName=Howie's HomeStay]

Figure 4.5 Example of output from a one-day travel plan for PTPS

Second, the similarity measurement of this TRS was not found to be appropriate, as the system uses string-matching between the POI name and user preferences. This is because the data set, in this set, does not contain enough relevant information regarding user preferences and POIs.

4.1.2.5 Discussion

The PTPS leaves a lot of room for improvement in the matching module and recommended module, both of which could make it a more intelligent and user-friendly system. By implementing the prototype DRS with the Chiang Mai data set (see Section 4.2), we learned that, when using only POI general information (i.e. POI name, type, locations, etc.), the DRS did not generate a satisfactory recommendation resul, e.g. when the user does not know where he/she wants to go or stay during or before his/her visit. As a result, the TRS characteristic are more like a planning system than a recommendation system.

4.1.3 Intelligent Tourist Attraction System study

For this study we implemented a model-based DRS, called the Intelligent Tourist Attractions System (ITAS), as utilised in the previous study (Hsu et al., 2012). This experiment aimed to understand the design and implementation of a system that involves a large data set and is model-driven. In the data-analysis phase, estimation of the user's prefered attractions were done through the use of BN. The experiment results that were obtained by using the 2012 inbound tourist data set were compared with other classifier methods (e.g. DT, Neural Networks (NN)).

The idea was to build a system based on the Engel-Blackwell-Miniard (EBM) decision model which involves many phases when a customer is about to make a decision. Another challenge in this study was calculating the probabilities of attractions for individual tourists, a Model-based CF approach, through statistical and machine learning using BN. The ROC curve is the only evaluation method that was used to evaluate the performance of the system. The ITAS methodology is summarised in the following four steps:

1. Extract measures from the EBM model for tourist attractions.
2. Collect data from the “2007 Annual Survey Report on Visitors Expenditure and Trends in Taiwan”. This information included demographic variables, such as gender, age, education, annual income, vacation, nationality, travelling motivation, information source and travel type.
3. Calculate the probability of an attraction’s appeal to a particular tourist by utilizing a BN. Descriptive statistical and factor analysis were applied to understand the factors that affect the overall satisfaction of inbound tourists to Taiwan. Correlation analysis was then applied to the selection variables to build the research model, and a ROC curve was used to evaluate the model’s performance.
4. Present recommended routes and tourist attractions through the system with Google Maps.

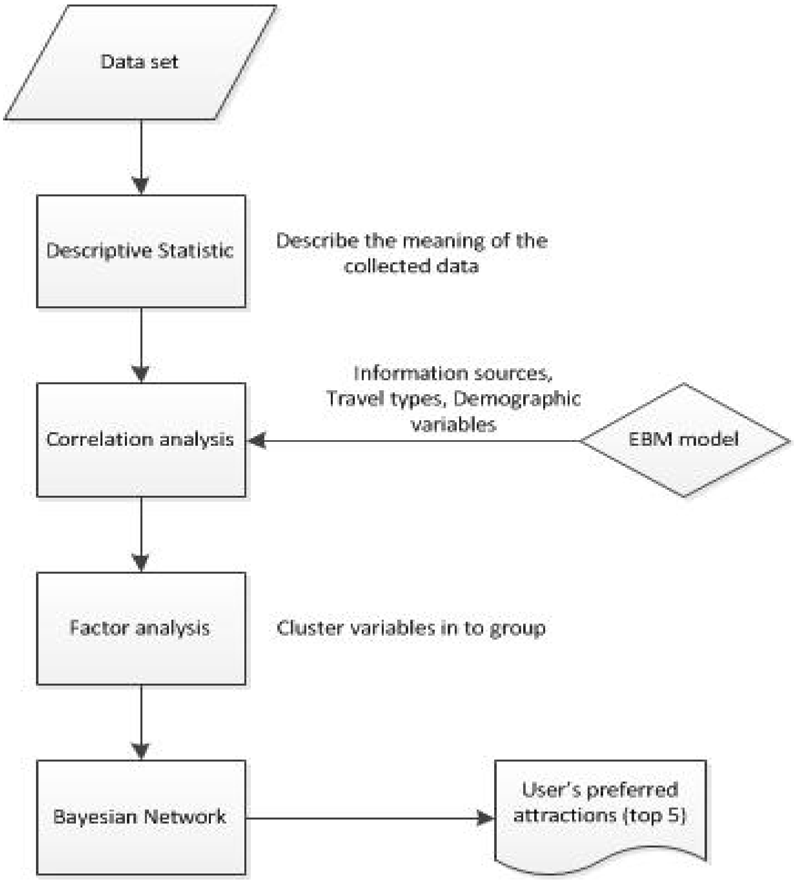


Figure 4.6 Overall diagram of the process flow of the TRS

Figure 4.6 represents the process flow, starting from the extraction of the meaning of the data set through to the construction of the network. The output of the experiment involves the user's preferred attractions, which are then ranked (e.g. top 5 based on user inputs).

4.1.3.1 Data set

The data set for the paper was obtained from the 2007 Annual Survey Report on Visitors Expenditure and Trends in Taiwan as Hsu et al. (2012), (previously mentioned in this paper), did. The sample size of the survey was 2,429. For the purpose of this project demonstration we used the same kind of survey but from a different year – we used a data set from 2012, which consisted of responses from 6,015 tourists.

As mentioned in Chapter 3, we applied for regular membership of the SRDA website https://srda.sinica.edu.tw in order to be able to use their survey data sets. Descriptive statistics are used in this phase to describe and summarise demographic information, travelling purpose, information source and travelling type. This is because, from the raw data we obtained, it was difficult to visualise what the data were showing and, therefore, difficult to present, describe and analyse the data of inbound tourists to Taiwan 2012 in meaningful ways.

The 2012 survey included 3,125 male travellers (52%) and 2,890 female travellers (48%). The survey was distributed to adult travellers, 91.7% of whom were aged between 20 and 60 years. In terms of education, most of the travellers (81.9%) had completed college, university, graduate school or higher. The main purpose of their trip was for sightseeing (66.2%) and 40.2% had come through group tours arranged through a travel agency (i.e. joined a tour group).

4.1.3.2 Experiment setup

In this experiment we mainly focused on the data pre-processing step, using the same methodology, and constructing the same experimental set-up, as described in the work of Hsu et al. (2012). We used descriptive statistics to analyse the data through percentage allocation. Also, a contingency coefficient was used to determine the correlation between independent variables and dependent variables. Netica software was used to build the BN recommendation engine. Most of the data pre-processing was done using SPSS software.

We began the experiment by obtaining the same data set (2007 Annual Survey Report on Visitors Expenditure and Trends in Taiwan) as that presented in the Hsu et al.'s (2012) work. We selected 22 tourist attractions and removed cases and variables that had excessive missing values. We then had approximately the same number of samples (around 3,000) as reported in Hsu et al. (2012).

Hsu et al. (2012) used factor analysis to find critical factors of inbound tourists’ satisfaction towards travel services and then extracted four factors (i.e. safety and friendliness, transportation convenience, entrance convenience, comfort and cleanliness). In our experiment we used the same approach with the 2012 data set to identify the important factors of inbound tourists’ satisfaction.

4.1.3.3 Similarity measurement

Correlation analysis was used to test the degree of association between the variables used in constructing the research model (i.e. to build the BN). In the case of this experiment, it was used to determine the correlation between tourists' choices of favourite attractions and other factors, including demographic variables, type of travel and purposes of travel.

Pearson's Correlation coefficient (commonly used) which determines the strength of the linear relationship between two variables, was applied. If a relationship exists between them, that relationship should be a linear one. When given the covariance of data points (*x*, *y*) and standard deviation σ , Pearson correlation is calculated as:

$$P(x, y) = \frac{\sum(x, y)}{\sigma_x \times \sigma_y} \tag{4.1}$$

To draw a conclusion about the relationship between two variables in the matrix we can look at the significance level and the correlation coefficient value. The correlation coefficient value will be between -1.0 and +1.0. If the coefficients are close to 0.0 they represent a weak relationship. Coefficients that are close to 1.0 or -1.0 represent a strong correlation.

Regarding the similarity measure, correlation analysis was used to test the degree of association between the variables to be used in constructing the research model (i.e. to build the BN). Here it was used to determine the correlation between tourists' choices of favourite attractions and other pertinent factors, including demographic variables, type of travel and purpose of travel. After relevant features were hand-selected by observing coefficient values, the process of model construction was carried out.

In the process of model construction, 20% of the data set was separated off and used for testing, while the remaining 80% was used to build the model. The C4.5 DT algorithm was applied to the data set.

4.1.3.4 Experimental results

Table 4.3 shows the correlation between the demographic variables and tourists' favourite attraction variable (i.e. predicted variable). The results show that the education variable was much closer to 0.0 and the significance value was 0.79, which represents a weak relationship.

Therefore, we can conclude that it is not necessary to use the education variable as a factor in building the model (this is similar to the findings presented in Hue et al. (2012).

Table 4.3 Correlation between demographic variables and the tourist's favourite attraction variable

Feature	Pearson correlation	*p*-value
Nationality	-0.003	0.836
Country of residence	0.001	0.934
Age	-0.053	0.000
Annual income (US$)	-0.006	0.660
Education	-0.004	0.794
Occupation	-0.007	0.605
Gender	0.012	0.388

We achieved the highest (30%) classification accuracy rate by using C4.5. Three major weaknesses of the system were found from the experiment. First, the system intentionally included 'undetermined' as one of the 22 target classes. Therefore, the system was likely to return a high predictive rate, and indeed the paper reported that AUC > 0.8. Second, the system did not provide proper validation criteria such as a basic evaluation of the system; for example, classification accuracy rate or a confusion matrix. Third, the model is not generalised enough to be applied elsewhere because the authors need to provide a proper sampling strategy (e.g. one part of the data set should be separated for testing purposes), and also the lack of a presentation of parameter learning of BN.

On the other hand, the BN, as a recommendation engine, provided both content-based filtering and collaborative filtering. Additionally, using Google Maps as an interactive geographical interface is a good feature of this system.

4.1.3.5 Discussion

This section discusses the investigation of existing DRSs. The weaknesses of each system are presented and the theories behind the recommendation engines of two DRSs are examined. Two prototypes were developed in order to demonstrate and identify the challenges of applying the proposed supervised machine-learning for the DRS. To construct an improved DRS, we proposed using a supervised machine-learning technique called Intelligent Destination Recommendation System (IDRS), comprised of model-based and ensemble-based approaches. IDRS is capable of generating a recommendation result for a user with better results regarding practical aspects. The proposed model-based DRS using feature selection and DT (based on the Chiang Mai data set) is discussed in the next section.

4.2 Feature extraction and model construction study

The first aim of this study is to investigate different features and feature-selection algorithms. The second aim is to build the optimal decision choice models. The proposed machine learning techniques were applied in this study to identify tourist destination choice processes that we do not understand yet. To make the model easier for a decision maker to intrepret, decision rules were generated from the models to describe the output classes. These rules will be used in the process of making recommendations as outlined in Chapter 6. The objectives which corresponded to the research questions 3 and 4 of this study are as follows:

4.2.1 Objectives of the study

1. To investigate and compare the performance of two well-established feature-selection algorithms.
2. To validate the proposed machine-learning techniques on the data set we collected.
3. To propose optimal destination-choice models using the proposed machine-learning techniques.
4. To evaluate the proposed models and estimate their generation errors on unseen data.
5. To generate decision rules from the models.

4.2.2 Representation of the Chiang Mai data set

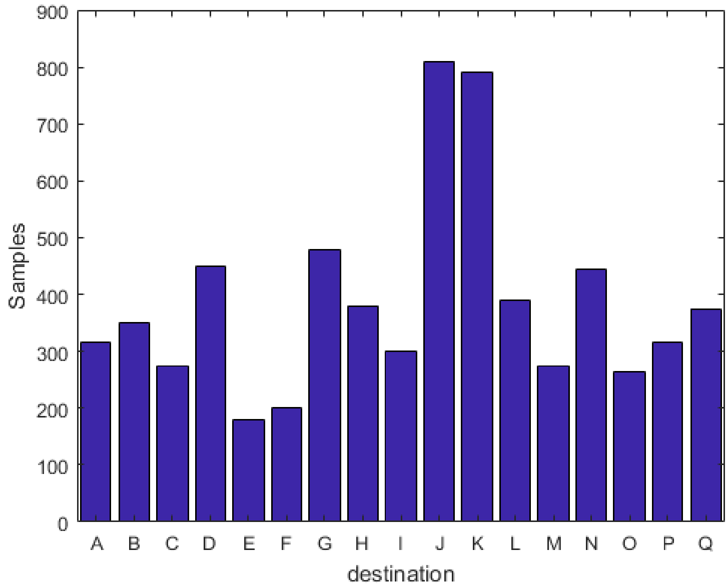


Figure 4.7 Class distribution for the Chiang Mai data set

Figure 4.7 represents the class distribution for the 20 destinations in Chiang Mai. It can be seen from the graph that it is an imbalanced data set in that the class distribution is not uniform among the classes. One of the challenges in this study was to develop a model that would be feasible for complicated real-world problems. The model that was constructed using all the 20 destinations achieved a very low classification accuracy rate of 17%, was complex and took a long time to construct. The model was too complex, as it had a large tree size and a large number of leaves. This made it difficult for the decision-maker to interpret. To solve this problem, we applied class decomposition in the pre-processing step. The goal was to identify groups of destinations with related patterns. Class decomposition offers us many advantages, including increased classification performance, scalability to a large database, increased comprehensibility, modularity and suitability for parallel computation.

Selecting an optimal decomposition method for a certain type of classification problem is difficult. There are many existing methods for class decomposition, such as clustering with k-mean, code matrix, concept aggregation etc. (Maimon and Rokach, 2005). Due to the fact that we considered the user experience and the meaning of the new cluster group/ destination category, the 20 multi-classes classification problem was decomposed explicitly into several sub-problems by investigating the types of tourists' preferred destinations (combining knowledge from the Chiang Mai tourism-domain experts and destination information from the Trip Advisor website). Machine-learning techniques may have led to better classification accuracy, but clustered group were meaningless to tourists. Hence ten destination categories were constructed and class distribution was applied (see Table 4.4). The models were constructed based on destination categories that featured in more than one class (i.e. a data set that represents the binary or multi-class classification problem). Regarding the characteristics of each data set, the Nature category consists of three classes (two of them represent waterfalls and one of them represents a lake); and the Museum and Art Gallery category consists of two classes (as there are both specialised museums and art galleries).

Consequently, ten tourist-preferred-destination categories were constructed (see Table 4.4). The models were configured based on categories that had more than one class. Regarding the characteristics of each category: (1) The Nature category consisted of three classes (two representing waterfalls and one representing a lake). Later, it was decided to exclude Bua Thong waterfall (A) from the category (as it overlaps with two official destination names containing Bua Thong waterfall) and Jed-See fountain (also known as Num-Poo-Jed-See) as this could confuse tourists. Also, during data collection, it became apparent that the fountain was difficult to find due to poor signage on the road. (2) The Museum and (3) Art Gallery categories are considered as two separate classes, as each of them is considered a specialised museum an art gallery.

Interestingly, most Chiang Mai tourist destinations are temples, as can be seen from the list of tourist destinations obtained from the Trip Advisor website (i.e. 11 out of 20 destinations we obtained involved temples). These temples and other attractions have already been categorised by the Trip Advisor website as religious sites and some of them are included in sub-categories such as heritage sites or landmarks. (4) The Temple-outer town category was constructed based on location. Destinations in this category were close to the university, restaurants and work places. (5) The Temple-landmark category was constructed based on the locations and reputations of temples as 'must-see temples' or landmarks. For instance, Wat Chedi Luang is a religious site and attractive to tourists as it is an impressive ruined temple. (6) The Temple-peaceful category consists of two classes. The temples in this category are not very well-known to tourists, and the structures share a similar style of architecture. They are located very close to each other in the central part of the city and surrounded by small pubs and bars. (7) The Temple-old town category contains two classes; the data set was constructed based on the fact that the locations of the temples are inside the town, and these temples are considered unique in their own way. Last, (8) the Entertainment category consists of two classes and both destination classes in this category provide tourists with a form of entertainment or a fun activity to do in Chiang Mai. The remaining destinations were categorized as Observation deck and National Park. After the destination categories had been constructed we repeated the experiment. The proposed data pre-process steps were applied to the eight constructed categories.

Table 4.4 Characteristics of the data set used in this study

Labels	Destination Name	Category name	# Sample	%	Trip Advisor' rank
A	Bua Thong Waterfall	Nature	230	2.50	18
B	Huay Tung Tao Lakw	Nature	313	3.40	19
C	Mae Sa Waterfall	Nature	360	3.91	20
D	Museum of World Insects and Natural Wonders	Museum	277	3.01	4
E	Art in Paradise, Chiang Mai 3D Art Museum	Museum	452	4.91	5
F	Wattana Art Gallery	Art gallery	186	2.02	7
G	Documentary Arts Asia	Art gallery	203	2.20	16
H	Wat Phra That Doi Kham	Temple-outer town	482	5.23	9
I	Wat Umong	Temple-outer town	385	4.18	10
J	Wat Suan Dok	Temple-outer town	311	3.38	13
K	Wat Chedi Luang	Temple-land mark	822	8.92	1
L	Wat Phra Singh	Temple-land mark	782	8.49	8
M	Wat Lok Molee	Temple-peaceful	391	4.24	12
N	Wat Pan Tao	Temple-peaceful	269	2.92	14
O	Wat Sri Suphan	Temple-old town	447	4.85	11
P	Wat Chiang Man	Temple-old town	278	3.02	15
Q	Chiang Mai Cabaret Show	Entertainment	314	3.41	2
R	Burklerk Gym- Muay Thai Training	Entertainment	376	4.08	17
S	Wat Phra That Doi Suthep	Observation deck	1538	16.70	3
T	Doi Inthananon	National Park	795	8.63	6

Figure 4.8 illustrates the class distribution of each destination choice category, in which each of them represents a separate data set and has a different number of samples. We can see that all the data sets are imbalanced as the classes are not represented equally. The proposed two-step filtering method was applied to each of them to remove irrelevant and redundant features.

Table 4.5 Participant characteristics of Chiang Mai tourists' preferred destinations data set.

Socio-Demographic variables		No.	%
Gender	Male	4525	49.1
	Female	4359	47.3
Age	18–25	2474	26.9
	26–35	3602	39.1
	36 and older	2967	32.2
Marital status	Single	4778	51.9
	Available	3489	37.9
Highest education	Less than high school	2423	26.3
	College or bachelor's degree	4700	51
	Higher than bachelor's degree	1827	19.8
	Other	261	2.8
Annual income	US$3,000–5,000	1746	19
	US$5,001–15,000	2632	28.6
	US$15,001–60,000	3206	34.8
	US$60,000 or more	1307	14.2
Employment	Employed	4336	47.1
	Self-employed	3001	32.6
	Un-employed	1304	14.2
	Other	570	6.2
Nationality	International	5315	57.7
	Local	3211	34.9
	Other	685	7.4

Regarding the summary of the data set, descriptive statistics were applied to analyse the background structure through percentage allocation (see Table 5.2). Of the inbound tourists, there were 4,525 male travellers (49.1%) and 4,359 females (47.3%). Regarding annual income, the largest group (34.8%) in terms of income included those earning more than US$15,000 per year or more than US$1,250 per month; 57.7% of the respondents were international tourists.

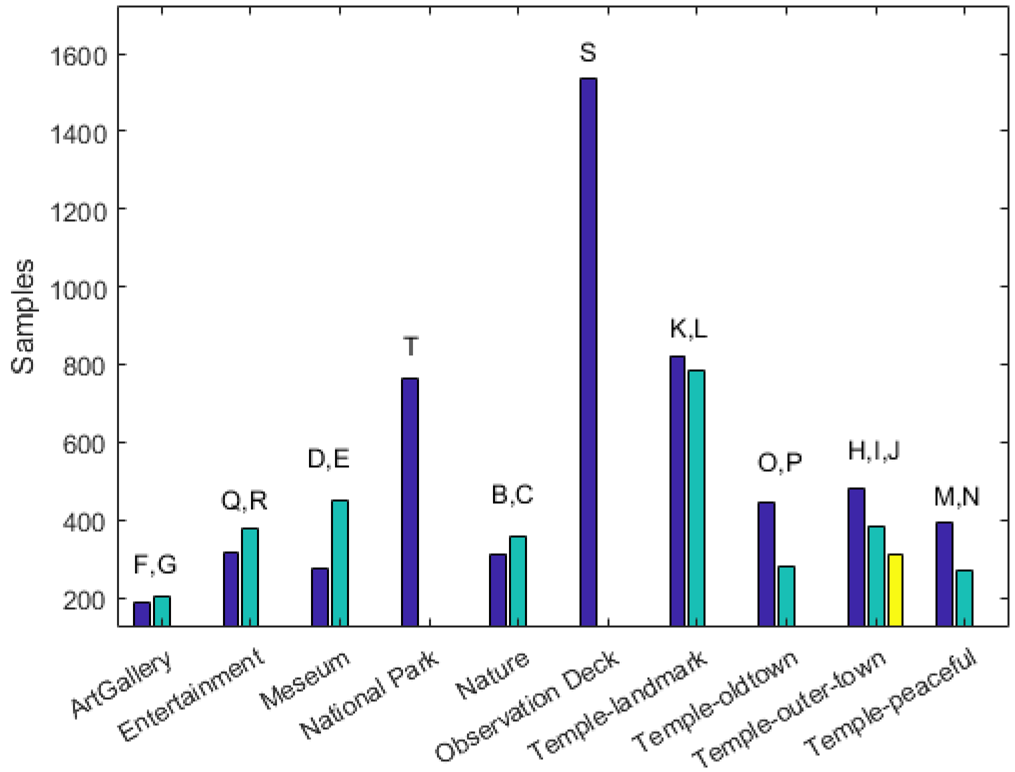


Figure 4.8 Class distribution of each destination choice category

4.2.3 Data pre-processing

After the initial input selection and MVA were applied, continuous variables were discretised using the binning method as outlined in Section 3.4.2; the number of bins was set to 10 for this study. Outliers were detected using the following proposed simple algorithm (see Table 4.6). Ordinal variables were scaled down from 5 to 3. Some of the variables were normalised using tourism-domain expert knowledge (i.e. g7 (nationality), g8 (country of residence), and g9 (origin)).

Table 4.6 Outlier detection algorithm

Algorithm 4.1: Outlier/Extreme value detection

```
Input: dataset
Output: number of detected values and survey_id,
id = [];      % list of survey id number
for i=1 to number of case
   for j=1 to number of variable
        if( isCategorialVariable )
             x = range of variable        % i.e. [1,5]
             if( isMemberOf (dataset(i, j), x) and dataset(i,  j) ~= missingvalue )
                n = n+1;
                id(end+1) = i;
        end
   end
end
return id, n
```

After the data set had been cleaned and transformed, the proposed two-step filtering method described in Section 3 was applied to the process of data reduction. This was done to remove irrelevant and redundant features from the data set.

4.2.4 Feature selection

MI is used as a similarity measurement in the feature-selection process to characterise both the relevance and redundancy of variables. In Equation (4.2), we are given a set of *X* and *Y*, $p(x)$ or $p(y)$ are the marginal probability distribution functions of *X* and *Y*, and $p(x,y)$ is the joint probability distribution function of *X* and Y:

$$MI(X,Y) = \iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy \qquad (4.2)$$

However, using continuous variables, the joint probability and marginal probability are difficult to estimate (Guyon and Elisseeff, 2003). In practice, continuous variables are often discretised to discrete variables and then MI can be calculated by using the following equation:

$$MI(X,Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \qquad (4.3)$$

$p(x, y)$ is the joint probability, which is the probability that two variables will occur simultaneously, where $p(x)$ or $p(y)$ is the marginal probability or the probability of occurrence of a single variable.

Marginal probability and joint probability can be calculated by constructing a joint probability mass function. For example, for $p(x_1)$ the marginal probability of x_1 is $(a+c)/n$; for $p(x_1, y_1)$, the joint probability of x_1 and y_1 is a/n; the marginal probability can then be calculated by the number of x occurrences in X divided by the total elements in the vector.

4.2.4.1 First filtering

The purpose of the first filtering step is to rank the variables and remove any independent variables that are unrelated to the dependent variable. We applied a Max-Relevance feature selection algorithm (Peng et al., 2005), in which we chose MI as the measurement to remove irrelevant features. We computed the MI score between each independent and dependent variable. Then we ranked them in descending order and used a threshold value (chosen manually) to remove features that contributed less or were not related to predictive power:

$$\max D(S, c), D = MI(\{x_i, i = 1, \dots, t\}; c) \tag{4.4}$$

Table 4.7 Max-Relevance Algorithm

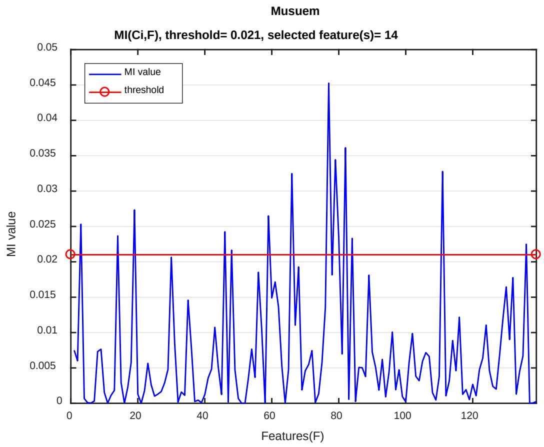
Algorithm 4.2: Max-Relevance
Input: Discretized data *d*, class *c* Output: feature set *F* 1: s = size(*d*); 2: **for** i = 1:s **do** 3: relevance(i) = mutual_info(d(:, i), *c*); 4: **end for** 5: **return** sort(relevance, ‘descend’);

Table 4.8 Description of the range of features regarding the factors influencing tourists' destination choices

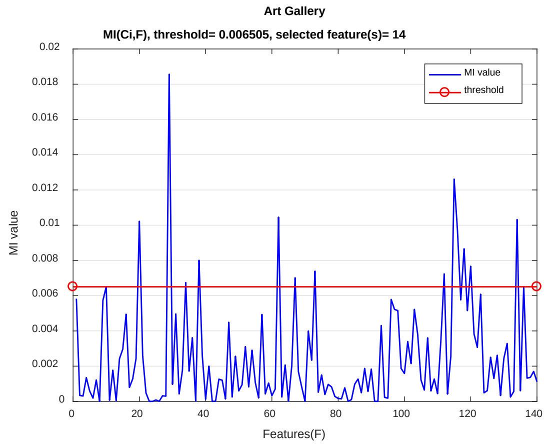
Set of factors	Feature numbers
Trip characteristics (TC)	1–25
Tourist expenditure behaviour (TEB)	26–38
Tourist behaviour (TB)	39–123
Travel motivations (TM)	124–136
Tourists' socio-demographic information (TSD)	136–145

In the feature-selection step, the first filtering method described in Section 3.2 was carried out. Different numbers of thresholds were used, based on each data set, to select 10% of the features. For example, the threshold was set to 0.0115 to select ten features from the Nature data set, while the threshold was set to 0.021 to select ten features from the Museum data set. Features that had an MI value less than the threshold line were removed from the data set. The experimental results show that the same feature is not significantly important for every item in the data set. For example, Tourist expenditure behaviour is an important factor for the Temple-old town data set but less significant for the Museum and Temple-peaceful data set.

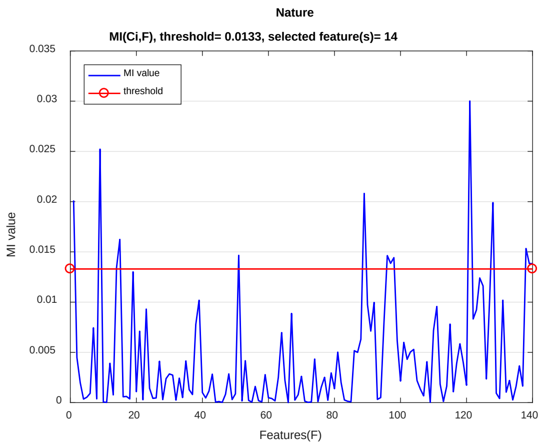
Table 4.8 describes the range of features regarding the factors that influence tourists' destination choices. The MI values for each feature in each data set are presented in Figure 4.9.



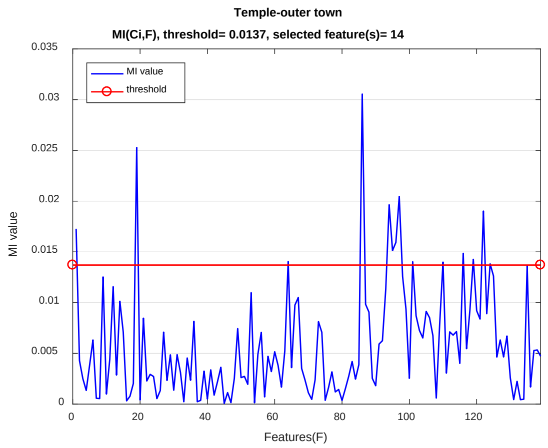
(a)



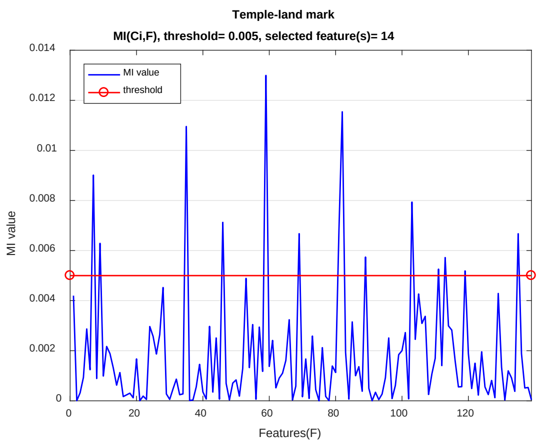
(b)



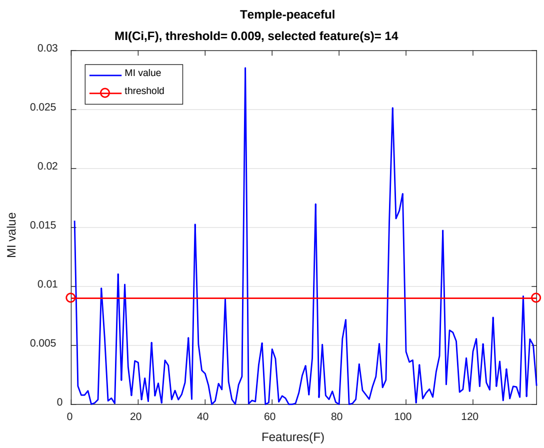
(c)



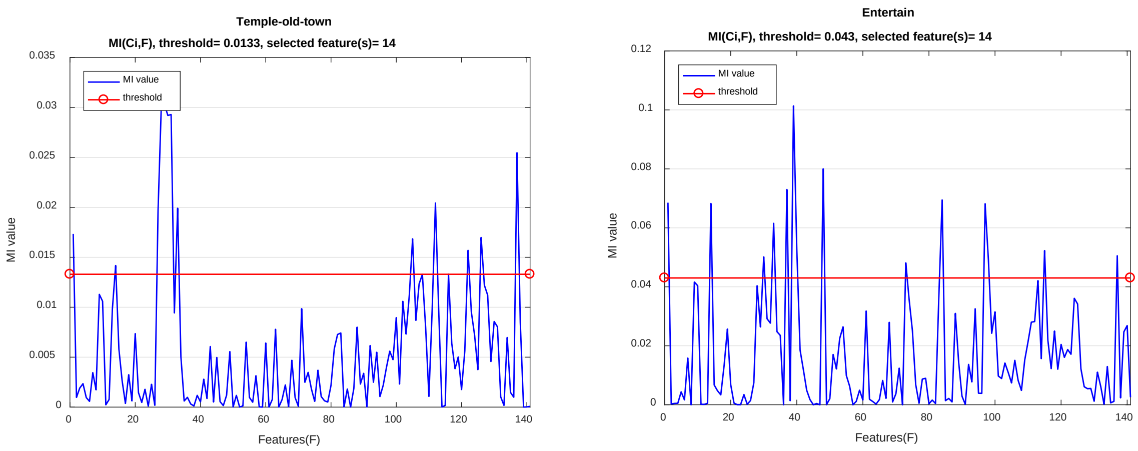
(d)



(e)



(f)



(g) (h)

Figure 4.9 MI value for each category

4.2.4.2 Second filtering

In the second filtering step we used two mutual information-based, feature-selection algorithms: Minimum Redundancy Maximum Relevance (mRMR) (Peng et al., 2005) and Normalized Mutual Information Feature Selection (NMIFS) (Estevez et al., 2009), to remove redundant variables from the data set.

mRMR algorithm

The idea of the mRMR algorithm (Peng et al., 2005) is that it uses MI value to rank features based on minimal redundancy and maximal relevant criteria. mRMR calculates the redundancy for every pair of features and the relevance between features and class. In this research we only considered MI for discrete variables and in the form of mRMR Mutual Information Differences (mRMR MID); it is formulated as equation (4.5). Table 4.9 shows the implementation of the algorithm.

$$MRMR = \max_{1\in\Omega_s}\left[I(i,h) - \frac{1}{|S|}\sum_{j\in s} MI(i,j) \right] \tag{4.5}$$

Table 4.9 Minimum-Redundancy Maximum-Relevance (mRMR) algorithm

Algorithm 4.3: Minimum Redundancy Maximum Relevance

Input: Discretized data *d*, class *c, max number of feature* **Output:** Selected feature set *F*.

```
s = size(d)
for i=1:s do
relevance(i) = mutual_info(d(:, i), c);
end for
idx = sort(relevance, 'descend');
F(1) = idx(1);
idx_left = idx(2: max number of feature)
for j=2:s do
n = length(idx_left);
last_fea = length(F);
    for k=1:n do
        mi(j) = mutual_info(F, c)
        redun(idx_left(j), least_fea) = mutual_info(F, c);
        redun_mi(i) = sum(redun(idxleft(i), :)) / last_fea;
    end for
[G, F(j)] = max( mi(1: n) - redun_mi(1: n) );
g_mi(j) = G;
tmp_idx = F(j);
F(j) = idx_left(tmpidx);
idx_left(tmp_idx) = [];
end for
```

NMIFS algorithm

NMIFS (Estevez et al., 2009) is a modification of the mRMR algorithm (see equation 4.8 and Table 4.10); it normalises the original MI value by the minimum entropy (H(i) and H(j)) of both features, as shown in equations (4.6) and (4.7).

$$H(X) = -\sum_{x} p(x) \log p(x) \tag{4.6}$$

Then, the modification of Mutual Information for the NMFIS algorithm can be written as:

$$MI2(i, j) = \frac{MI(i, j)}{\min\{H(i), H(j)\}} \tag{4.7}$$

Hence, NMIFS can be written as the equation below:

$$NMIFS = \max_{1\in\Omega s}\left[I(i,h) - \frac{1}{|S|}\sum_{j\in s} MI2(i,j)\right] \tag{4.8}$$

Table 4.10 Normalized Mutual Information Feature Selection (NMIFS) algorithm

Algorithm 4.4 Normalized Mutual Information Feature Selection

Input: Discretized data *d*, class *c, max number of feature***Output:** Selected feature set *F*.

```
s = size(d)
for i=1:s do
relevance(i) = mutual_info(d(:, i), c);
end for
idx = sort(relevance, 'descend');
F(1) = idx(1);
idx_left = idx(2: max number of feature)
for j=2:s do
n = length(idx_left);
last_fea = length(F);
    for k=1:n do
        mi(j) = mutual_info(F, c)
        redun(idx_left(j), least_fea) = mutual_info(F, c);
        tmp = sum(redun(idx_left(i), :)) / min( entropy (d(:, F(last_fea))), entropy( d(:, idx_left(i)) ))
        redun_mi2(i) = tmp/last_fea;
    end for
[G, F(j)] = max( mi(1: n) - redun_mi2(1: n) );
g_mi(j) = G;
tmp_idx = F(j);
F(j) = idx_left(tmpidx);
idx_left(tmp_idx) = [];
end for
```

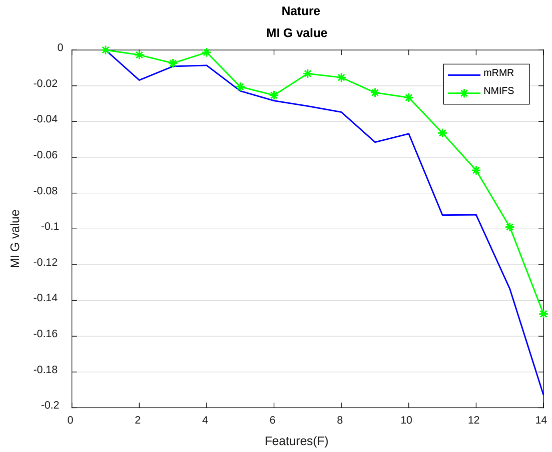
Table 4.11 presents the ten selected features by both of the feature-selection algorithms from each of the data sets. The bold variables indicate that the corresponding feature belongs to the optimal subset. Both mRMR and NMIFS selected the same features for every data set. However, they ranked them in a different order, except for the first few features which represent relevant features.

According to Table 4.11 we can see that the Nature category lacks relevant features to predict target classes. Only TM_1 (Number of times you have visited) was selected as an optimal feature by both feature-selection algorithms. For the Museum category, mRMR selected nine optimal features, and NMIFS selected eight. It can be seen that feature *TB$_3$* (Wildlife) was rated as the most important. This can be explained by the fact that one of the museums specialise in insects. For the Art Gallery category, six features were selected as optimal by both feature-selection algorithms, and *TEB$_1$* (Money for transportation) was found to be the most relevant feature. For the Entertainment category, the same characteristics as in the Museum category were shared in that there were many relevant and not redundant features used to estimate target classes. The results also show that all of the factors helped to contribute to all of the categories, especially the *TEB$_1$* (Money for transportation) factor, which was ranked as the most important factor in the Art Gallery and Temple-old town categories by two of the feature selection algorithms.

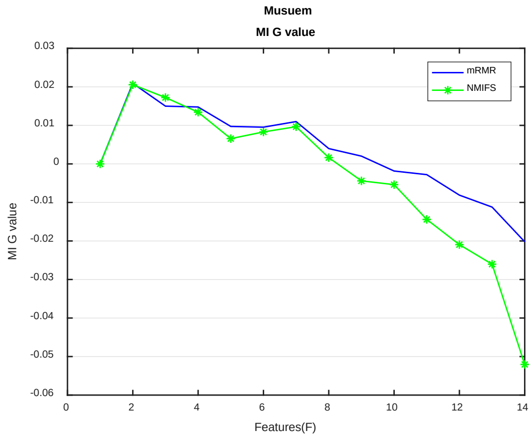
Table 4.11 Feature-ranking of each destination category where the subscript is the feature-identification number (see Appendix E)

Category	Threshold	Algorithm	Feature ranking									
Nature	0.013	mRMR	**TM$_1$**	TC$_1$	TB$_1$	TM$_2$	TB$_2$	TC$_2$	TC$_3$	TM$_3$	TM$_4$	TSD$_1$
		NMIFS	**TM$_1$**	TC$_1$	TM$_2$	TB$_1$	TB$_2$	TC$_2$	TM$_3$	TC$_3$	TC$_4$	TM$_4$
Museum	0.021	mRMR	**TB$_3$**	**TSD$_2$**	**TC$_5$**	**TC$_6$**	**TB$_4$**	**TB$_5$**	**TM$_5$**	**TB$_6$**	**TB$_7$**	TC$_3$
		NMIFS	**TB$_3$**	**TSD$_2$**	**TM$_5$**	**TC$_5$**	**TB$_4$**	**TB$_5$**	**TC$_6$**	**TC$_3$**	TB$_6$	TB$_7$
Art gallery	0.006	mRMR	**TEB$_1$**	**TM$_6$**	**TC$_7$**	**TB$_8$**	**TB$_9$**	**TB$_5$**	TB$_{10}$	TSD$_3$	TB$_{11}$	TM$_7$
		NMIFS	**TEB$_1$**	**TM$_6$**	**TB$_5$**	**TB$_{10}$**	**TB$_9$**	**TC$_7$**	TSD$_3$	TC$_8$	TB$_{11}$	TM$_8$
Temple-outer-town	0.013	mRMR	**TB$_{12}$**	**TB$_{13}$**	**TC$_5$**	**TM$_9$**	**TB$_{14}$**	**TM$_{10}$**	TM$_{11}$	TC$_1$	TM$_2$	TM$_7$
		NMIFS	**TB$_{12}$**	**TM$_9$**	**TB$_{13}$**	**TC$_5$**	**TB$_{14}$**	**TM$_{10}$**	TC$_1$	TM$_2$	TM$_{11}$	TM$_7$
Temple-landmark	0.005	mRMR	**TB$_6$**	**TC$_9$**	**TM$_{12}$**	**TB$_{15}$**	**TB$_{16}$**	**TB$_{17}$**	TSD$_4$	TM$_7$	TB$_{18}$	TM$_5$
		NMIFS	**TB$_6$**	**TC$_9$**	**TM$_{12}$**	**TB$_{16}$**	**TB$_{15}$**	**TB$_{17}$**	TSD$_4$	TM$_7$	TM$_5$	TB$_{18}$
Temple-peaceful	0.009	mRMR	**TB$_{19}$**	**TM$_{13}$**	**TC$_{10}$**	**TSD$_4$**	TM$_5$	TB$_{20}$	TM$_{14}$	TC$_3$	TC$_4$	TB$_8$
		NMIFS	**TB$_{19}$**	**TM$_{13}$**	**TC$_{10}$**	**TSD$_4$**	TM$_5$	TC$_3$	TC$_1$	TM$_{14}$	TB$_{20}$	TB$_8$
Temple-old town	0.013	mRMR	**TEB$_1$**	**TM$_{15}$**	**TEB$_2$**	TM$_6$	TSD$_4$	TM$_{12}$	TC$_3$	TM$_1$	TEB$_3$	TM$_9$
		NMIFS	**TEB$_1$**	**TM$_{15}$**	**TEB$_2$**	TM$_6$	TSD$_4$	TC$_3$	TM$_{12}$	TEB$_3$	TM$_9$	TM$_1$
Entertainment	0.04	mRMR	**TB$_8$**	**TM$_4$**	**TB$_7$**	**TEB$_4$**	**TM$_{16}$**	**TC$_3$**	**TB$_{21}$**	**TB$_{22}$**	**TB$_7$**	TB$_7$
		NMIFS	**TB$_8$**	**TM$_4$**	**TB$_7$**	**TEB$_4$**	**TC$_3$**	**B$_{21}$**	**TM$_{16}$**	**TB$_{22}$**	**TB$_7$**	TB$_7$

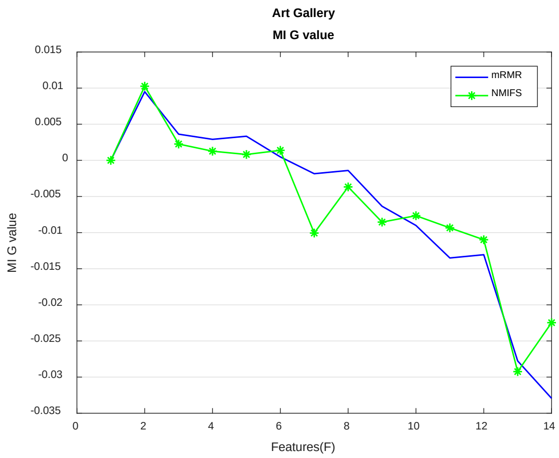
In this study we compared the performance of two feature-selection algorithms by observing MI G values. Note that MI G value is defined in algorithm 4.3, line 16, for mRMR, and in algorithm 4.4, line 18, for NMIFS as the maximum value that was chosen from the set of features *F*. From the second filtering step, by observing the performance graphs of both feature-selection algorithms (see Fig. 4.10), it can be seen that mRMR and NMIFS produced similar results (e.g. in terms of the selection of better sub-features). This is due to the fact that both feature-selection algorithms are based on MI for similarity measurements. However, mRMR selected marginally better sub-set features than NMIFS for the Art Gallery and Temple-land mark categories. For the Nature, Temple-outer town, Temple-old town and Entertainment categories, NMIFS performed better than mRMR (e.g. the NMIFS selected slightly better sub-features than the mRMR) (see Fig. 4.10 (b)).



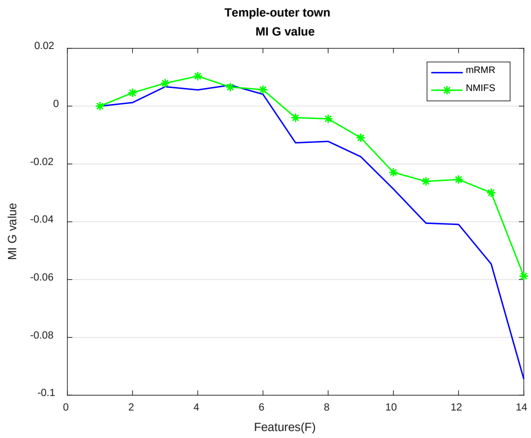
(a)



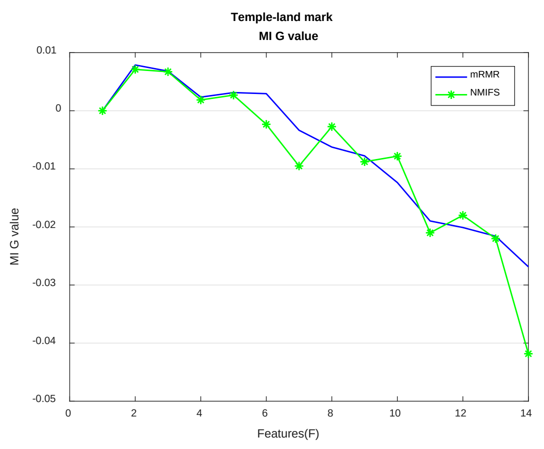
(b)



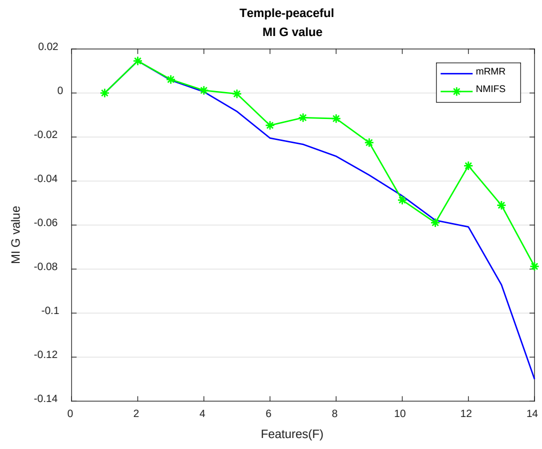
(c)



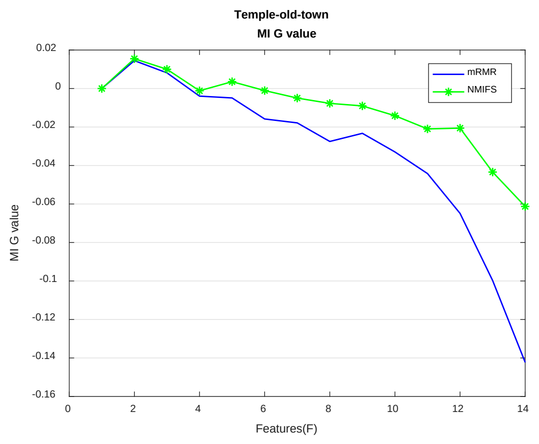
(d)



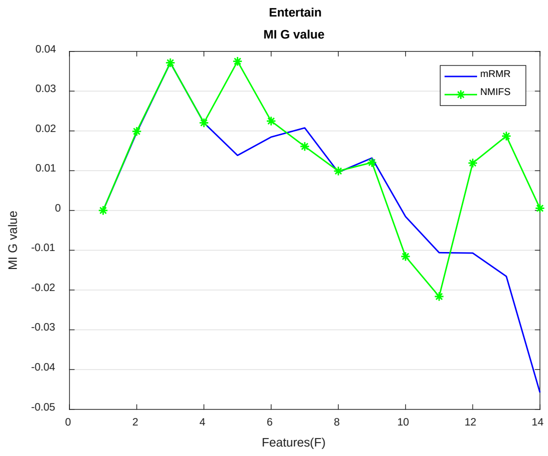
(e)



(f)



(g)



(h)

Figure 4.10 Performance comparison of mRMR and NMIFS for each data set.

4.2.5 Classification and model construction with a Decision Tree (DT)

After irrelevant and redundant features had been filtered out, and designated features had been selected, DT was chosen as the classifier to construct relevant models. Other classifiers could also have been used, including K-Nearest Neighbour (KNN), SVM or ANN. However, they are generally black boxes (i.e. from which we cannot acquire knowledge in a comprehensible way). The proposed feature-selection algorithm offers numerous benefits to DT models such as interpretability, accuracy, and simplicity. C4.5 was selected as the most appropriate classifier for this study as this algorithm is very simple to understand for decision-makers, and it is open source. Moreover, C4.5 can support both nominal and scale variables. In order to avoid the over-fitting problem and to minimise the complexity of the tree, C4.5 follows a post-pruning approach using either confidence-based or error-based pruning. Conversely, C4.5 supports both information gain and gain ratio approaches when measuring splitting. In this study, we used a gain-ratio based on the entropy concept. It is a modification of the information-gain approach from ID3 to reduce bias toward multi-valued attributes.

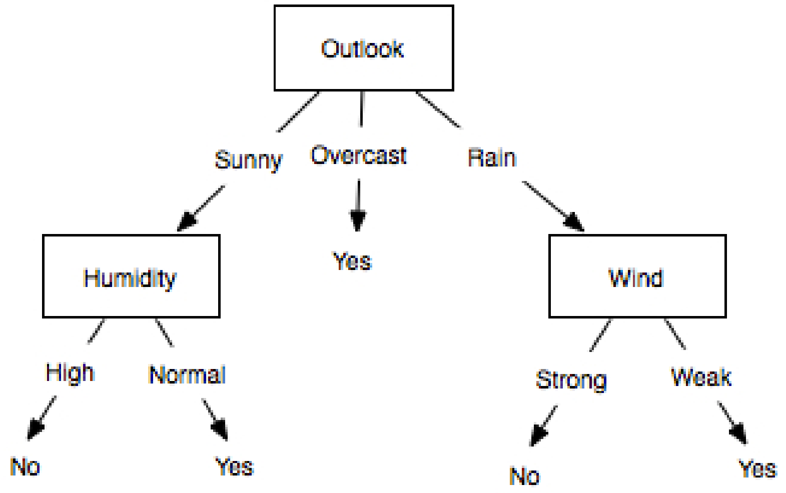


Figure 4.11 An example of a simple Decision Tree (DT)

A DT is a predictive hierarchical model that can be used to represent a trained classifier. It consists of nodes and leaves. The first node is called the root node, where instances from the test set start to navigate down to a leaf. Other nodes, referred to as internal nodes, involve testing a particular attribute; this is where the split – either binary or multiple – occurs. The leaf nodes represent class labels (i.e. output of classification) or the final decisions of instances from the test data (Witten and Frank, 2005). To the best of our knowledge, the DT algorithm has never been used in the TRS domain before.

Figure 4.11 presents a DT which is used to classify weather data, in which the problem is to learn how to classify new days as ‘to play’ or ‘not to play’. Starting from the top and going down through the leaf nodes, five rules were generated for this problem.

1. **If** outlook is sunny **and** humidity is high, **then** do not play
2. **If** outlook is sunny **and** humidity is normal, **then** play
3. **If** outlook is overcast, **then** play
4. **If** outlook is rain **and** wind is strong, **then** do not play

5. **If** outlook is rain **and** wind is weak, **then** play

One of the main advantages of a DT is its simplicity; decision-making is easily understood due to its flowchart-like nature. To recommend a destination to a tourist we must traverse the DT from root to leaf. Many DTs exist, such as Hunt's algorithm, Top-down Induction DT (TDIDT), ID3, CHAID, CART and C4.5. They differ in terms of splitting criteria, pruning, types of attributes, etc.

Chi-Squared-Automatic-Interaction Detection, known as CHAID (Kass, 1980), is a DT that uses a statistical method. CHAID uses *p*-value obtained from statistical tests in splitting criteria depending on the variable type. For example, Pearson's correlation coefficient and likelihood ratio are methods for determining node-splitting for nominal and ordinal variables. The DT was initially aimed at handling nominal variables, and it does not support tree pruning (Lior and Oded, 2008). The advantage of CHAID is that it is easy to interpret because the algorithm supports multiple ways of splitting and merging variables. Classification and Regression Trees (CART), developed by Breiman et al. (1984), only support binary splits and use the Gini index as the splitting criterion.

C4.5, an extension of ID3, was devised by Quinlan (1993). It was chosen for this study because C4.5 tried to solve the main drawbacks of ID3. ID3 (Quinlan, 1986a) is the most simple DT algorithm and has many drawbacks such as: an optimal solution is not guaranteed, overfitting problems when training the data set, supporting only nominal variables. C4.5 supports both nominal and scale variables. In order to avoid the over-fitting problem C4.5 supports tree-pruning (e.g. confidence-based and error-based pruning), it also allows attributes to be missed. On the other hand, C4.5 supports both information gain and gain ratio when measuring splitting, including two types of splitting criteria: information gained and the entropy-based criterion (see equations (4.9) and (4.10)). In this study, we used tagain ratio based on the entropy concept. This is a modification of the information gained from ID3 to reduce the bias toward multi-valued attributes. First, C4.5 calculated intrinsic or split information (*SI*) values as shown in equation (4.9). The gain ratio (*GR*), which represents a proportion of the information, is defined in equation (4.10).

$$SI(A) = -\sum_i \frac{n_i}{n} \log\left(\frac{n_i}{n}\right) \tag{4.9}$$

$$GR(A) = \frac{Gain(A)}{SI(A)} \tag{4.10}$$

The most recent version of this classifier is C5.0, the updated version of C4.5. It has more advantages than C4.5 in terms of memory, speed, and accuracy, and it generates a smaller DT than C4.5. Furthermore, C5.0 supports boosting, which is one of the ensemble techniques used to gain predictive accuracy.

Once the DT is constructed it can be converted to rules or rule-based classifiers. In order to build rule-based classifiers we can extract rules directly from constructed C4.5 models. The advantages of decision rules are that they are easier for decision-maker to understand and can classify new instances effectively. The simplest way is to have one rule for each class. An example of a decision rule is as follows:

Destination A, if (marriage status = single) or (income = USD 100-500)

C4.5 is known as J48 in Weka software. In this study we used J48, which was developed by the Weka project team (Witten and Frank, 2005). It is a DT model which involves the implementation of C4.5 algorithm, release version 8. J48, implemented in Java language. The interface between Matlab® and Weka software was developed in order to be able to use Weka's DT classifier (i.e. it was necessary to convert training, validating and testing data to .arff file format).

An investigation of C4.5 performance using two feature-selection algorithms was carried out. For each destination choice category, we ran the experiment 10 times with the same experimental setup. For each iteration, randomized permutation was applied to the data set and a stratified sampling method was applied to ensure that there was homogeneity within the strata and heterogeneity between them. A hold-out sampling method was used to split the data set into two partitions, where 85% of the data set was used as a training set and the remaining 15% was used for testing the true performance of the model. To find the optimal parameters and assess the model's performance, a stratified 5-fold cross-validation method was applied to the training set, wherein four folds were selected for training and one fold was used for validation. Different values of confidence factors for the error-based pruning algorithm were used. The confidence factors ranged from 0.01 to 1.0, with a step size of 0.01. The minimum number of instances per leaf was fixed at 2. The classification accuracies of the training and validating sets of the different iterations were averaged. The optimal model was found if it had the highest mean of validation classification accuracy, the smallest tree size and was not over-trained (i.e.

the mean accuracy of the validation set had to be less than or equal to the mean accuracy of the training set).

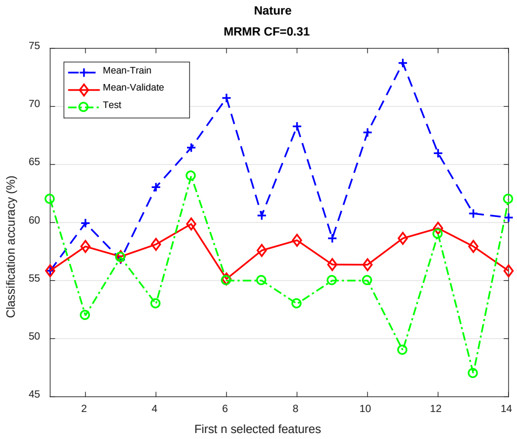
4.2.6 Experimental results

Table 4.12 represents the classification accuracy regarding the first n-selected features and the optimal models of the data set. It can be seen that optimal models were found when the confidence value was less than 0.59. The Entertainment data set reveals the highest classification accuracy rate of 78.64%, whereas Temple-outer town reveals the lowest rate.

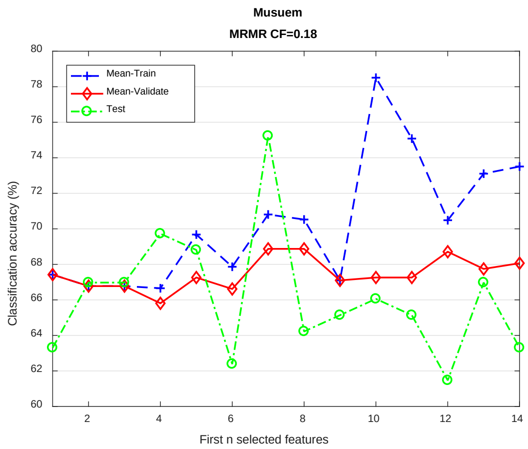
Table 4.12 Best classification accuracy rates results achieved by the C4.5 algorithm

Category	#Classes	#Features	Confidence factor	Mean-train rate (%)	Mean-validation rate (%)	SD	Test accuracy rate (%)	Most important factor
Nature	2	5	0.31	66.45	59.87	5.85	64	TM_1 (Visit friend)
Museum	2	7	0.18	70.80	68.87	1.34	75.23	TB_3 (Wild life)
Art Gallery	2	8	0.08	66.08	60.71	6.52	68.97	TC_7 (TV, radio is main information source)
Temple-outer town	3	3	0.59	46.36	44.71	2.49	51.13	TB_{12} (Overall food price)
Temple-landmark	2	4	0.1	58.99	58.87	1.86	62.08	TB_6 (Healthcare)
Temple-peaceful	2	10	0.21	70	63.28	3.62	68.68	TB_{19} (Entertainment)
Temple-old town	2	8	0.12	70.34	66.28	4.32	70.37	TEB_2 (Prepaid expense)
Entertainment	2	6	0.05	73.68	72.58	2.74	78.64	TB_8 (Attend festival)

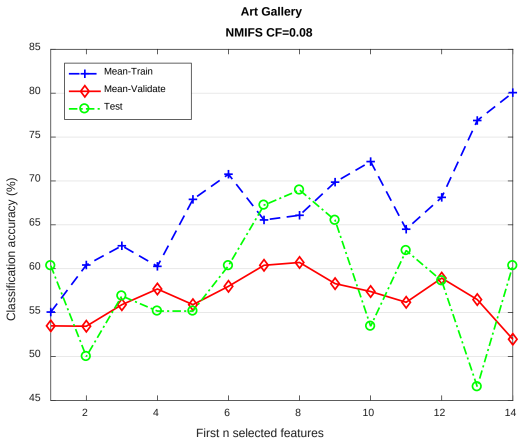
Furthermore, Figure 4.12 shows the classification accuracy for each data set including the mean classification accuracy of the training set, the mean classification accuracy of the validating set, and the classification accuracy of the test set with the most optimal Confidence Factor (CF). Note that CF is used to compute a pessimistic upper bound on the error rate at a tree node, and the smaller the value of CF, the heavier is the pruning. The results show that combining more features significantly improves the classification accuracy rate. For example, in the Museum category, in which 7 features were combined, we achieved the highest classification rate of 75.23%. However, if we continue adding features to the model and the features do not provide any significant relevance to the predicted class, the model will become more complex and very difficult to interpret for a decision-maker. Additionally, it can lead to an over-fitting problem where the model follows the training data set rigorously. Over-fitting can be easily seen in the results of the mean of training and the mean of validation accuracy rates acquired from the Museum category, as shown in Figure 4.12 (b), where there is no relevant feature to predict the target classes after using a combination of seven features. The results show that the best range for the CF value is between 0.1 and 0.6, and an increase in the CF value of more than 0.6 does not guarantee that a better classification result will be obtained.



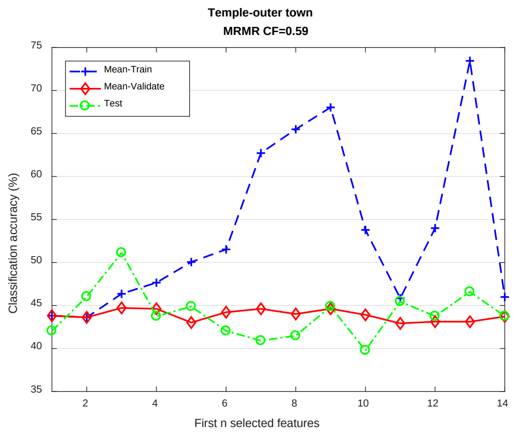
(a)



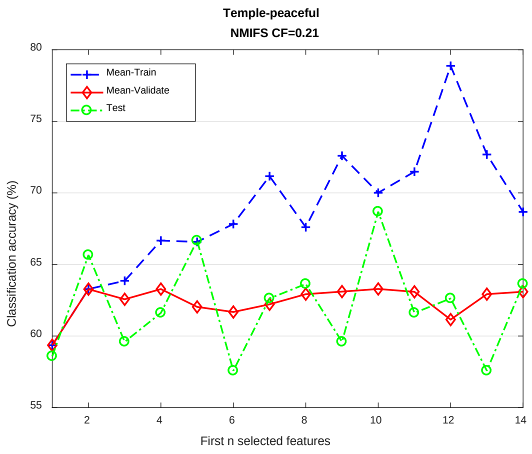
(b)



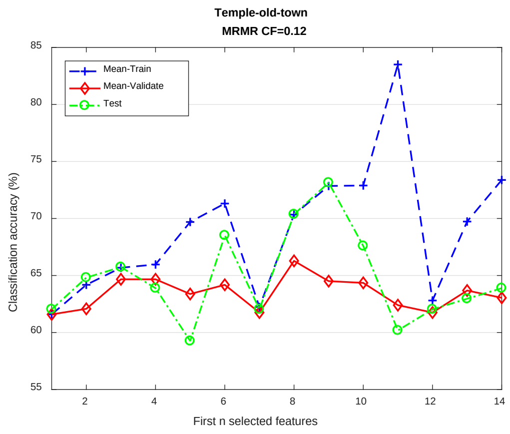
(c)



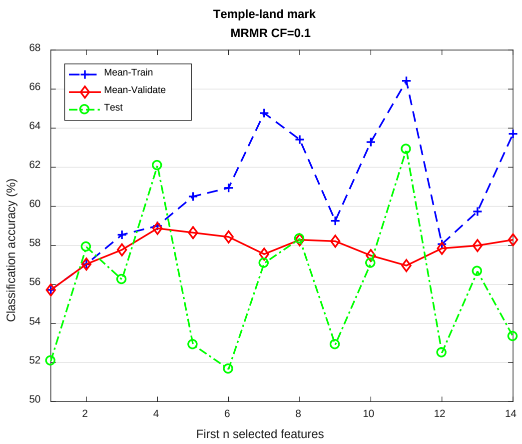
(d)



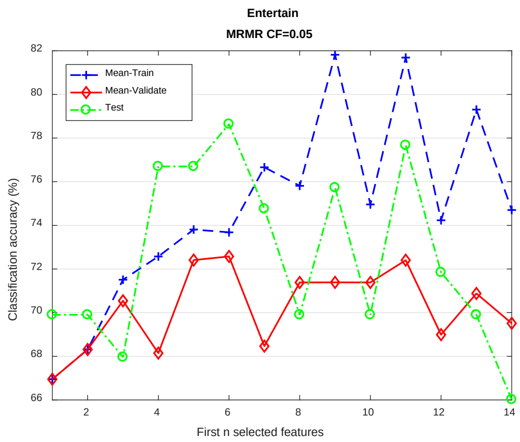
(e)



(f)



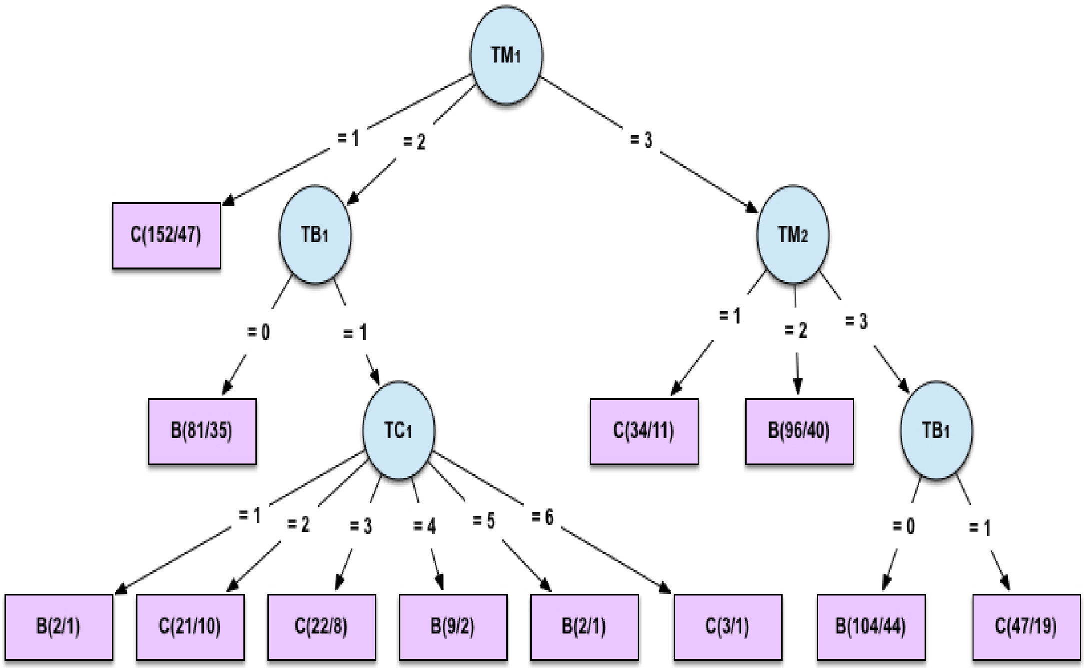
(g)



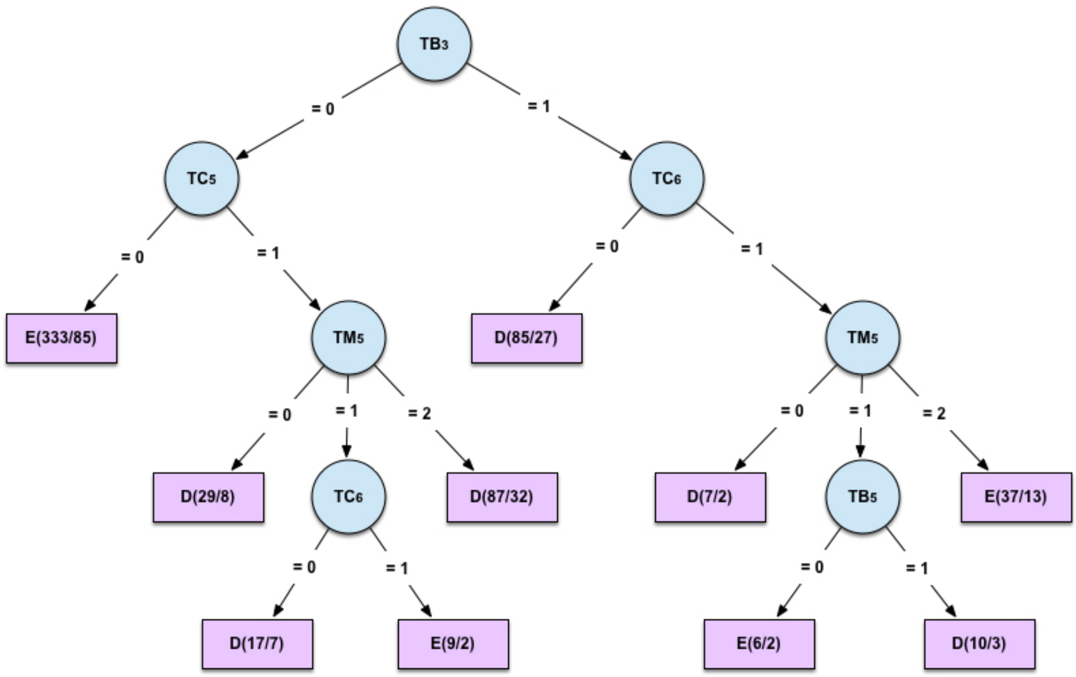
(h)

Figure 4.12 Mean of training Classification Accuracy Rate (CAR) (+), Mean of validation CAR (diamond), test CAR (circle)

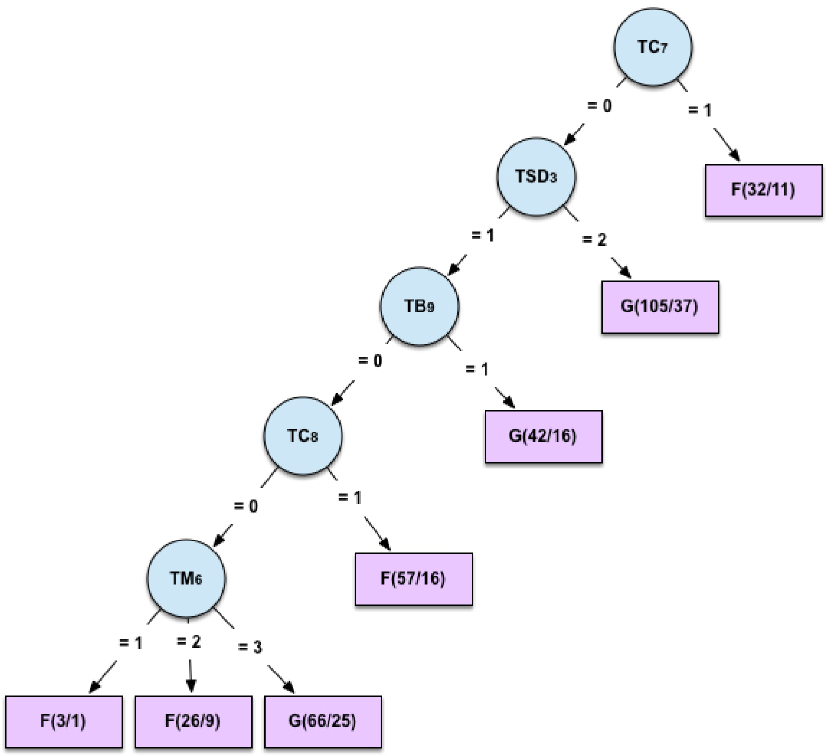
Eight optimal models were obtained, and decision rules were then extracted. However, it can be seen from the results that the model for the Temple-peaceful data is the most complex one (i.e. tree size = 33). The reason why the model is more complex than the others because it uses seven features in order to achieve the highest accuracy rate.



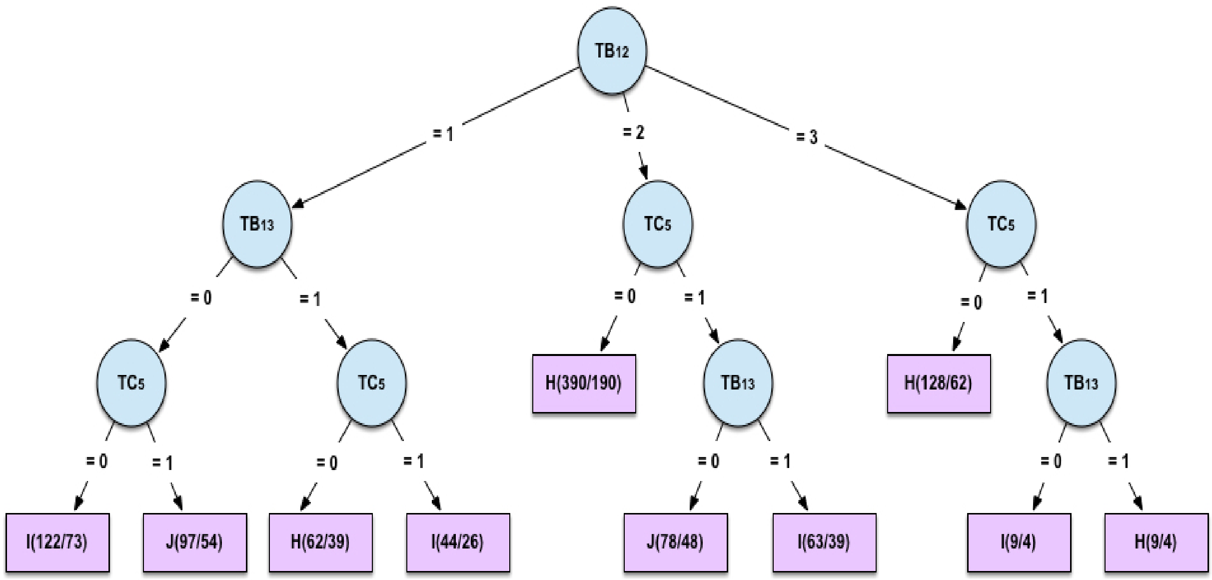
(a) Nature (Mae Sa Waterfall (B), Huay Tung Tao Lake (C))



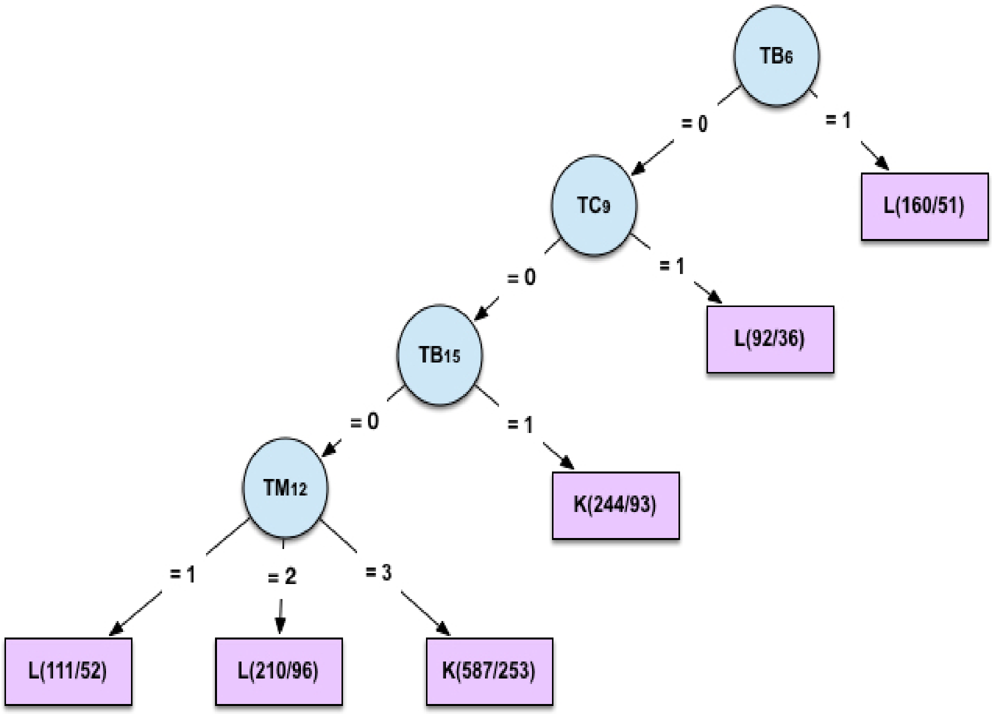
(b) Museum (Chiang Mai 3D Art Museum (D), Museum of World Insects and Natural Wonders (E))



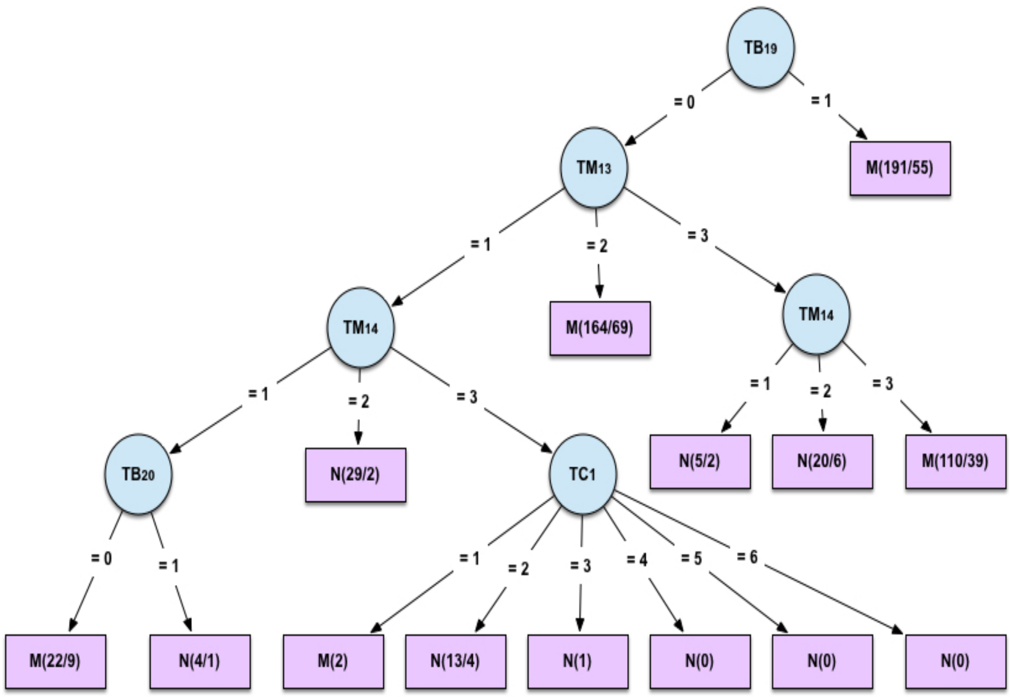
(c) Art Gallery (Wattana Art Gallery (F), Documentary Arts Asia (G))



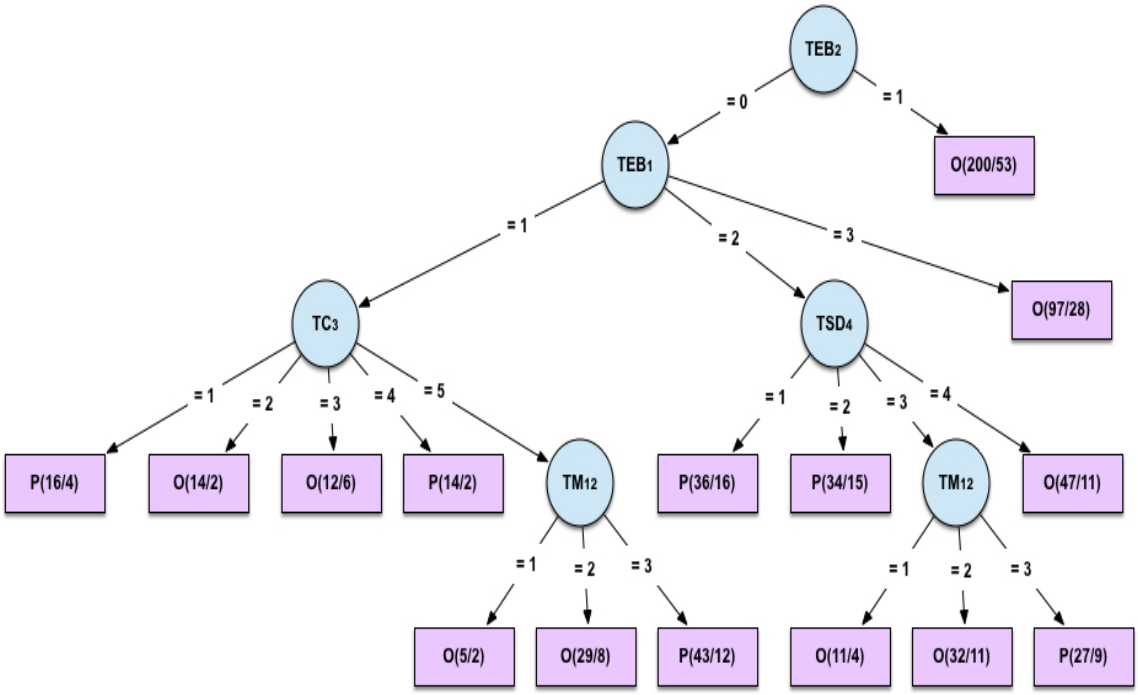
(d) Temple-outer town (Wat Phra That Doi Kham (H),
Wat Umong (I), Wat Suan Dok (J))



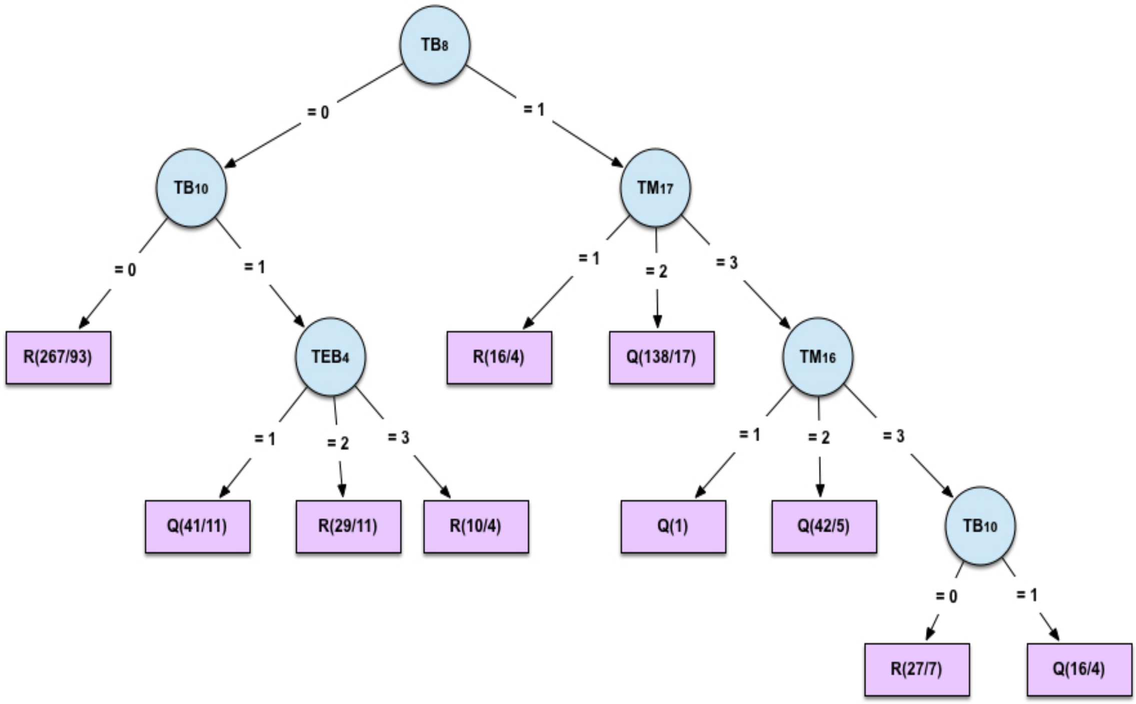
(e) Temple-landmark (Wat Chedi Luang (K), Wat Phra Singh (L))



(f) Temple-peaceful (Wat Lok Molee (M), Wat Pan Tao (N))



(g) Temple-old town (Wat Sri Suphan (O), Wat Chiang Man (P))



(h) Entertainment (Chiang Mai Cabaret Show (Q), Burklerk Gym- Muay Thai Training (R))

Figure 4.13 Decision Tree (DT) for each destination category

4.2.6.1 Decision rules

Eight optimal models for each tourist category were determined. In order to extract potentially useful information and make it simpler for decision-makers to understand the recommended results, decision rules were generated in the form of pseudo codes from the models using the depth-first search algorithm (Tarjan, 1972). For each model, decision rules are configured from the root node. Each feature that occurs in the model entails an 'IF' statement for the purposes of establishing a rule. The 'IF' statement ends in a leaf node with a 'THEN' statement. Table 4.13 presents the generated rules as they correspond to the number of leaves on the tree.

Temple-landmark has the fewest rules because its tree has the least number of leaves when compared to other trees. The rules from the Temple-old town model are more complex than other models because the tree has a depth level of five and there are two nodes with more than three leaves. From a DRS point of view, these constructed rules were parsed as eXtensible Mark-up Language for further development of the proposed DRS in the online phase (see Chapter 6).

Table 4.13 Decision rules for each data set

Model	Rule IF	THEN
M_{Nature}	$(TM_1=2\wedge TB_1=0)\vee(TM_1=2\wedge TB_1=1\wedge TC_1=1\vee TC_1=4\vee TC_1=5)\vee(TM_1=3\wedge TM_2=2)\vee(TM_1=3\wedge TM_2=3\wedge TB_1=0)$	B
	$(TM_1=1)\vee(TM_1=2\wedge TB_1=1\wedge TC_1=2\vee TC_1=3\vee TC_1=6)\vee(TM_1=3\wedge TM_2=1)\vee(TM_1=3\wedge TM_2=3\wedge TB_1=1)$	C
M_{Museum}	$(TB_3=0\wedge TC_5=1\wedge TM_5=0\vee TM_5=2)\vee(TB_3=0\wedge TC_5=1\wedge TM_5=1\wedge TM_6=0)\vee(TB_3=1\wedge TC_6=0)\vee(TB_3=1\wedge TC_6=1\wedge TM_5=0)\vee(TB_3=1\wedge TC_6=1\wedge TM_5=1\wedge TB_5=1)$	D
	$(TB_3=0\wedge TC_5=0)\vee(TB_3=0\wedge TC_5=1\wedge TM_5=1\wedge TC_6=1)\vee(TB_3=1\wedge TC_6=1\wedge TM_5=1\wedge TB_5=0)\vee(TB_3=1\wedge TC_6=1\wedge TM_5=2)$	E
$M_{Art\text{-}Gallery}$	$(TC_7=0\wedge TSD_3=1\wedge TB_9=0\wedge TC_8=0\wedge TM_6=1\wedge TM_6=2)\vee(TC_7=0\wedge TSD_3=1\wedge TB_9=0\wedge TC_8=1)\vee(TC_7=1)$	F
	$(TC_7=0\wedge TSD_3=1\wedge TB_9=0\wedge TC_8=0\wedge TM_6=3)\vee(TC_7=0\wedge TSD_3=1\wedge TB_9=1)\vee(TC_7=0\wedge TSD_3=2)$	G
$M_{Temple\text{-}outer\text{-}town}$	$(TB_{12}=1\wedge TB_{13}=1\wedge TC_5=0)\vee(TB_{12}=2\wedge TC_5=0)\vee(TB_{12}=3\wedge TC_5=0)$	H
	$(TB_{12}=1\wedge TB_{13}=0\wedge TC_5=0)\vee(TB_{12}=1\wedge TB_{13}=1\wedge TC_5=1)\vee(TB_{12}=2\wedge TC_5=1\wedge TB_{13}=1)\vee(TB_{12}=3\wedge TC_5=1\wedge TB_{13}=0)$	I
	$(TB_{12}=1\wedge TB_{13}=0\wedge TC_5=1)\vee(TB_{12}=2\wedge TC_5=1\wedge TB_{13}=0)$	J
$M_{Temple\text{-}landmark}$	$(TB_6=1)\vee(TB_6=0\wedge TC_9=1)\vee(TB_6=0\wedge TC_9=0\wedge TB_{15}=0\wedge TM_{12}=1\vee TM_{12}=2)$	K
	$(TB_6=0\wedge TC_9=0\wedge TB_{15}=1)\vee(TB_6=0\wedge TC_9=0\wedge TB_{15}=0\wedge TM_{12}=3)$	L
$M_{Temple\text{-}peaceful}$	$(TB_{19}=1)\vee(TB_{19}=0\wedge TM_{13}=2)\vee(TB_{19}=0\wedge TM_{13}=3\wedge TM_{14}=3)\vee(TB_{19}=0\wedge TM_{13}=1\wedge TM_{14}=1\wedge TB_{20}=0)\vee(TB_{19}=0\wedge TM_{13}=1\wedge TM_{14}=3\wedge TC_1=1)$	M

	$(TB_{19}=0\wedge TM_{13}=3\wedge TM_{14}=1\wedge TM_{14}=2)\vee$ $(TB_{19}=0\wedge TM_{13}=1\wedge TM_{14}=1\wedge TB_{20}=1)\vee$ $(TB_{19}=0\wedge TM_{13}=1\wedge TM_{14}=3\wedge TC_1=2\wedge TC_1=3\wedge TC_1=4\wedge TC_1=5\wedge TC_1=6)\vee$ $(TB_{19}=0\wedge TM_{13}=3\wedge TM_{14}=1\wedge TM_{14}=2)$	N
$M_{Temple\text{-}oldtown}$	$(TEB_2=0\wedge TEB_1=3)\vee(TEB_2=0\wedge TEB_1=2\wedge TC_3=2\wedge TC_3=3)\vee$ $(TEB_2=0\wedge TEB_1=1\wedge TC_3=5\wedge TM_{12}=1\wedge TM_{12}=2)\vee$ $(TEB_2=0\wedge TEB_1=2\wedge TSD_4=3\wedge TM_{12}=1\wedge TM_{12}=2)\vee$ $(TEB_2=0\wedge TEB_1=2\wedge TSD_4=4)\vee(TEB_2=1)$	O
	$(TEB_2=0\wedge TEB_1=1\wedge TC_3=1\wedge TC_3=4)\vee$ $(TEB_2=1\wedge TEB_1=1\wedge TC_3=5\wedge TM_{12}=3)\vee$ $(TEB_2=0\wedge TEB_1=2\wedge TSD_4=1\wedge TSD_4=2)\vee$ $(TEB_2=0\wedge TEB_1=2\wedge TSD_4=3\wedge TM_{12}=3)$	P
$M_{Entertainment}$	$(TB_8=0\wedge TB_{10}=0)\vee(TB_8=0\wedge TB_{10}=1\wedge TEB_4=2\wedge TEB_4=3)\vee$ $(TB_8=1\wedge TM_{17}=1)\vee(TB_8=1\wedge TM_{17}=3\wedge TM_{16}=3\wedge TB_{10}=0)$	Q
	$(TB_8=0\wedge TB_{10}=1\wedge TEB_4=1)\vee(TB_8=1\wedge TM_{17}=2)\vee$ $(TB_8=1\wedge TM_{17}=3\wedge TM_{16}=1\wedge TM_{16}=2)\vee$ $(TB_8=1\wedge TM_{17}=3\wedge TM_{16}=3\wedge TB_{10}=1)$	R

4.2.6.2 System evaluation

Besides the classification accuracy rate, a confusion matrix, presion, recall, and F-measure are also used to evaluate the model's performance. This study also provides ROC curves and calculates AUC values for better visualization and interpretation of the performance of the models.

Table 4.14 Confusion matrix for the test set (bold font indicates correctly classified instances)

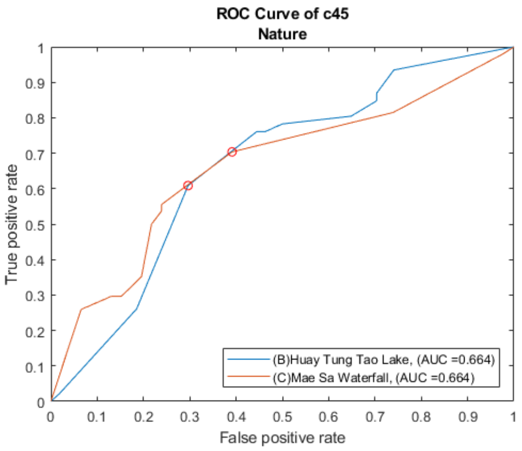
Actual	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
B	**35**	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	25	**29**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	**21**	21	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	6	**61**	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	**16**	11	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	7	**24**	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	**61**	8	3	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	29	**21**	7	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	25	14	**8**	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	**83**	40	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	51	**66**	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	**56**	3	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	28	**12**	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	**60**	7	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	25	**16**	**0**	**0**
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	**42**	**5**
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	**17**	**39**

Both the confusion matrix and the F-measure (see Tables 4.14 and 4.15) revealed that it was very difficult to classify three destinations: Wat Umong (I) in the Temple-outer town category and Wat Pan Tao (N) in the Temple-peaceful category. This is because the categories have a high ratio in the imbalance class. Additionally, these models do not have any related significant features to classify the similarity of the destinations.

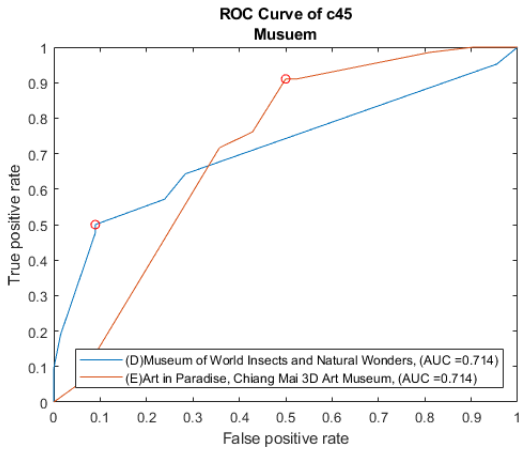
Table 4.15 The precision, recall and F-measure of each destination.

Destination	Precision	Recall	F-score
B	0.583	0.761	0.660
C	0.725	0.537	0.617
D	0.778	0.500	0.609
E	0.744	0.910	0.819
F	0.696	0.593	0.640
G	0.686	0.774	0.727
H	0.530	0.847	0.652
I	0.488	0.368	0.420
J	0.444	0.170	0.246
K	0.619	0.675	0.646
L	0.623	0.564	0.592
M	0.667	0.949	0.783
N	0.800	0.300	0.436
O	0.706	0.896	0.789
P	0.696	0.390	0.500
Q	0.712	0.894	0.792
R	0.886	0.696	0.780

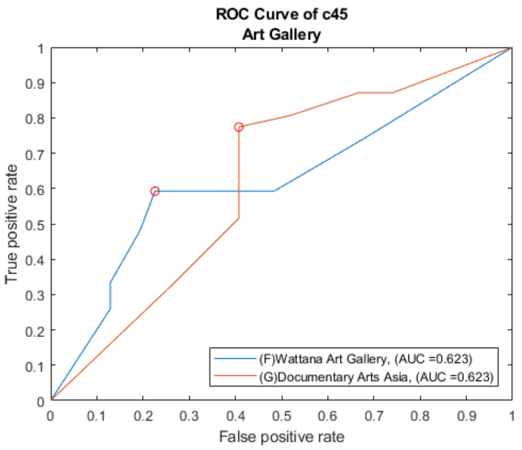
The ROC curve plots reveal the true positive rate (Sensitivity) against the false positive rate (Specificity) for each data set. Also, the plots present the area under the curve (AUC). We can see that the classifier cannot discriminate (I) Wat-Umong from other destinations. Wat-Umong reveals an AUC of 0.58, which is 0.8 better than random guessing (see Fig. 4.14 (d)). The Entertainment data set reveals the highest AUC value of 0.77. In this study we prefer precision over recall. Therefore, we consider classes that archives the high true positive rates while still having low false positive rates, such as the destinations C, E, G, H, K. M, O and Q.


ROC Curve of c45
Nature
True positive rate
False positive rate
(B)Huay Tung Tao Lake, (AUC =0.664)
(C)Mae Sa Waterfall, (AUC =0.664)

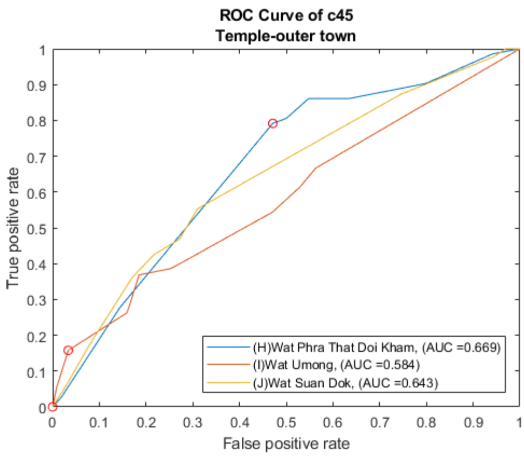
(a)


ROC Curve of c45
Musuem
True positive rate
False positive rate
(D)Museum of World Insects and Natural Wonders, (AUC =0.714)
(E)Art in Paradise, Chiang Mai 3D Art Museum, (AUC =0.714)

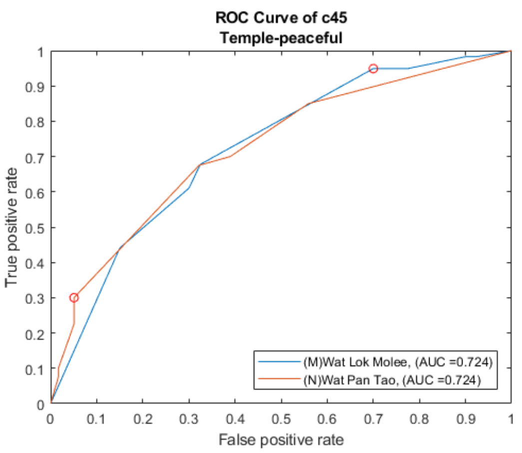
(b)


ROC Curve of c45
Art Gallery
True positive rate
False positive rate
(F)Wattana Art Gallery, (AUC =0.623)
(G)Documentary Arts Asia, (AUC =0.623)

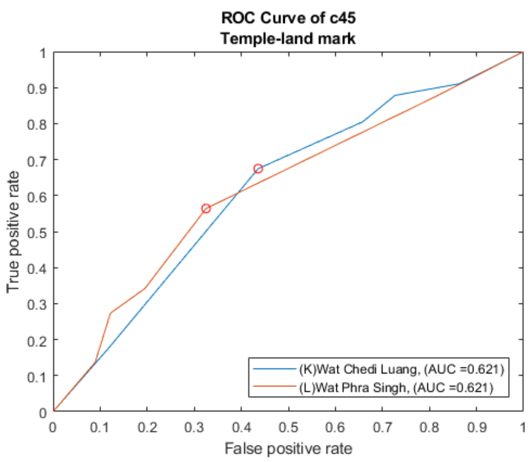
(c)


ROC Curve of c45
Temple-outer town
True positive rate
False positive rate
(H)Wat Phra That Doi Kham, (AUC =0.669)
(I)Wat Umong, (AUC =0.584)
(J)Wat Suan Dok, (AUC =0.643)

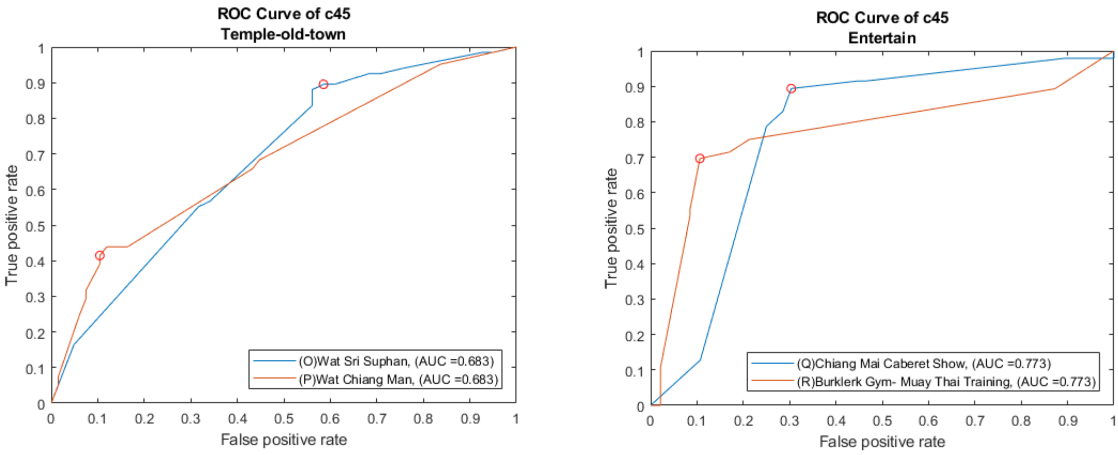
(d)


ROC Curve of c45
Temple-peaceful
True positive rate
False positive rate
(M)Wat Lok Molee, (AUC =0.724)
(N)Wat Pan Tao, (AUC =0.724)

(e)


ROC Curve of c45
Temple-land mark
True positive rate
False positive rate
(K)Wat Chedi Luang, (AUC =0.621)
(L)Wat Phra Singh, (AUC =0.621)

(f)



(g) (h)

Figure 4.14 ROC curve summarizes the C4.5 performance of the data sets

4.2.7 Discussion

The results of test-classification accuracy, using different numbers of features, confirmed that using more features does improve classification accuracy. It can be seen from the results that there are no common 'most important factors' to estimate destinations for all the data sets. For example, Trip purpose is the most important factor when classifying the Nature data set, while Wild life is the most important factor in classifying the Museum data set. The experimental results also show that, by combining sets of motivation factors, classification accuracy is increased for all data sets.

In this study we developed a novel model-based DRS that recommends 20 destinations to tourists using a set of human factors. The proposed DRS focused on pre-travel considerations before a tourist planned to visit, or during their visit, to the city of Chiang Mai. The aim of this study was to solve the current challenges of the destination TRSs in terms of practical issues, such as a non-intrusive system, and technical issues, such as recommendation accuracy and recommendation transparency. With regard to a non-intrusive system, we achieved this aim by reducing the user's efforts while maintaining a reasonable system accuracy rate using the proposed feature-selection method. For recommendation accuracy, the data set was decomposed into seven sub-data sets using relevant tourism-domain knowledge; this was done to increase the classification accuracy rate and reduce the complexity of the DT. We achieved a classification accuracy rate of 78.65% for the Entertainment category, 75.23% for the Museum category, 70.37% for the Temple-old town category, 68.97% for the Art Gallery category, 68.68% for the Temple-peaceful category, 64% for the Nature category, 62.91% for the Temple-landmark category and 51.13% for the Temple-outer-town category.

Table 4.16 Description of the eight optimal models for tourist destinations

Model	#Feature used	Tree size	Depth	#Rules	Features
M_{Nature}	4	17	4	12	$\mathbf{TM_1}$, TB_1, TM_2, TC_1
M_{Museum}	5	17	4	10	$\mathbf{TB_3}$, TC_6, TM_5, TB_5, TC_5
$M_{Art\text{-}Gallery}$	5	12	5	7	$\mathbf{TC_7}$, TSD_3, TB_9, TC_8, TM_6
$M_{Temple\text{-}outer\text{-}town}$	3	18	3	10	$\mathbf{TB_{12}}$, TB_{13}, TC_5
$M_{Temple\text{-}landmark}$	4	10	4	6	$\mathbf{TB_6}$, TC_9, TM_{12}, TB_{15}
$M_{Temple\text{-}peaceful}$	5	20	4	14	$\mathbf{TB_{19}}$, TM_{13}, TM_{14}, TC_1, TB_{20}
$M_{Temple\text{-}oldtown}$	5	21	5	15	$\mathbf{TEB_2}$, TEB_1, TC_3, TM_{12}, TSD_4
$M_{Entertainment}$	4	16	4	10	$\mathbf{TB_8}$, TB_{10}, TM_{16}, TEB_4

Table 4.16 summarises information from the eight optimal models, including a number of features that the model used after pruning, tree-size consideration and a number of generated rules. The common features that were used for each data set are also presented in Table 5.6.

This study also investigated five sets of factors that influenced tourists' preferred destinations, including trip characteristics, tourist characteristics, tourist expenditure behaviour, travel motivation and tourists' socio-demographic information based on qualitative research. The bold variables indicate the most important features of the model (e.g. tourist behaviour is the most important factor used to classify the Museum, Temple-outer town, Temple-landmark, Temple-peaceful and Entertainment categories. Trip characteristic is the most important factor in classifying the Nature and Art Gallery categories. Tourist expenditure behaviour is the most important factor when classifying the Temple-old town category). Thirty-five features were detected as having the largest influence on the proposed DRS.

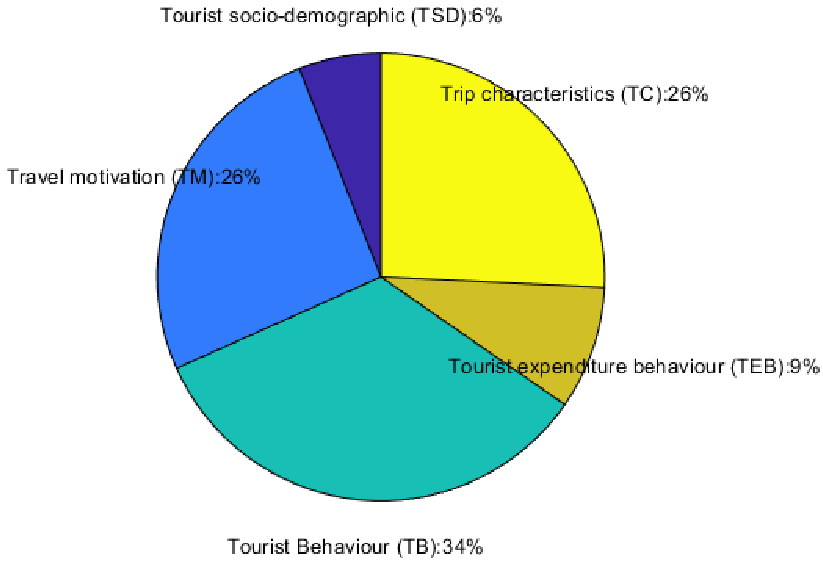


Figure 4.15 Summary of the factors that were used in the destination choice models.

Figure 4.15 illustrates the contribution of the factors that play an important role in the destination-choice models. It can be seen that the tourist behaviour factor was the one most commonly used (34%), followed by the travel characteristic (TC) (26%) and travel motivation (TM) (26%) factors. The tourist social demographic (TSD) factor makes the least significant contribution to the system (6%) and is only used in the Art Gallery and Temple-old town categories. The experimental results also support findings from the literature (Leiper, 1990) that indicate that combining tourist-motivation factors helps to increase classification accuracy, especially for the Temple-peaceful category, as this factor was identified as having the greatest influence and was used in the model as the most relevant feature.

In terms of practical aspects, the proposed DRS used a small number of relevant and non-redundant inputs from 3–5 features to achieve the best recommendation results. This means that the proposed system is considered non-intrusive and likely to be accepted by users. The constructed models can assist decision-makers with an overview of the multiple stages that will follow each possible decision when selecting a destination in Chiang Mai. Additionally, decision rules from the optimal models were extracted for decision-makers' ease in understanding the results, which show that Temple-landmark and Temple-peaceful had the fewest rules. These rules will be used when we integrate the online phase into the system.

Table 4.17 Optimum feature selection on each data set

Data set	Optimum feature selection algorithm
Nature	mRMR
Museum	mRMR
Art Gallery	NMIFS
Temple-outer-town	mRMR
Temple-landmark	mRMR
Temple-peaceful	NMIFS
Temple-oldtown	mRMR
Entertainment	mRMR

The performance of both modern feature-selection algorithms was investigated. Based on experimental results using eight data sets, the classification accuracy results (see Table 4.18) show that, in general, mRMR is the optimum feature-selection algorithm. The mRMR algorithm outperforms the NMIFS algorithm for the Nature, Museum, Temple-outer town, Temple-land mark, Temple-old town and Entertainment categories, while NMIFS outperforms mRMR for the Art Gallery and Temple-peaceful categories. Based on the experimental results, mRMR is best suited for the categorical data set. However, by observing the performance graphs of mRMR and NMIFS we can see that there are still some features that should be pre-selected.

4.2.8 Concluding remarks

This study demonstrates that human factors can be used to suggest tourist destinations to a user. A DT can provide transparency to the proposed system. However, recommendation performance is still a challenge; it can be improved by modifying the feature-selection algorithms or using other better feature selection algorithms that can measure the redundant and irrelevant features more effectively than the mRMR and NMIFS ones. In the next chapter we discuss the ensemble learning methods used to improve destination recommendation performance.

Chapter 5 Ensemble-Based Destination Recommendation System (DRS)

In the previous chapter, a model-based DRS, using a hybrid approach, was discussed. In this chapter we propose an ensemble-based hybrid approach to improve the effectiveness of our model-based DRS in terms of classification performance. Classification results such as prediction label, probability score, and ranking from classification algorithms are combined in order to produce a single and more robust final output. This chapter focuses on the weighted and cascade hybrid methods involving seven combination rules and bagging and boosting algorithms. This chapter addresses the following research objective:

RQ 6. How can the recommendation accuracy rate be improved using only relevant and non-redundant factors?

5.1 Destination classification algorithms study

5.1.1 Objectives of the study

The aim of this study is to improve the classification performance of the proposed DRS by investigating other traditional classification algorithms including Decision Tree (DT), Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) for the DRS. The performance of the classifiers is evaluated using the eight Chiang Mai destination choice data sets that we constructed in the previous study. The objectives are to evaluate and compare different classification algorithms with C4.5 as the baseline classifier.

5.1.2 Experimental design and data set

In this study, six classifiers were generated from SVM and MLP to compare with C4.5 that was investigated in the previous study. Experiments were conducted to compare SVM and MLP with C4.5 using several performance metrics including classification accuracy, confusion matrix, and f-score as the evaluation criteria. ROC curves for each test and AUC plots were also applied. We used the same Chiang Mai destination data set that has only the relevant and non-redundant features (e.g. selected features from the feature-selection process). The data set was decomposed into eight classification problems to solve the original problem (see Table 5.1). The eight distinct data sets were then constructed: ‘Nature’, ‘Museum’, ‘Art Gallery’, ‘Temple-outer town’, ‘Temple-landmark’, ‘Temple-peaceful’, ‘Temple-old town’ and ‘Entertainment’. The data sets were cleaned with regards to missing values, removal of noise and outliers, and normalised. Table 5.1 presents the data sets and variable descriptions.

Table 5.1 Features selected by the two-step feature selection method

Data set	**Destinations**	**Label**	**Features**
Nature	Huay Tung Tao Lake Mae Sa Waterfall	B C	TM_1: To visit relative(s)/friend(s) TC_1: Number of nights you plan to stay TB_1: Visit markets, walking streets TM_2: To work on my personal/spiritual values TB_2: The transport mode that you plan to use during this visit is walking
Museum	Museum of World Insects and Natural Wonders Art in Paradise, Chiang Mai 3D Art Museum	D E	TB_3: Wildlife has made the deepest impression upon you TSD_1: Your country of residence TC_2: Books, guides are the information sources that have influenced your decision to visit TC_3: People whom you are accompanied by are friends TB_4: Museums have made the deepest impression upon you TB_5: Outdoor areas are of interest to you and you plan to visit them during this visit TM_3: To visit places I have never been before
Art Gallery	Wattana Art Gallery Documentary Arts Asia	F G	TEB_1: The amount of money you plan to spend per person on transportation during this visit TM_4: To develop new abilities TB_5: Outdoor areas are of interest to you and you plan to visit them during this visit TB_6: Thai food has made the deepest impression upon you TB_7: Observing wildlife is the activity you plan to participate in during this visit TC_4: TV, radio are the information sources that have influenced your decision to visit TSD_2: Marital status TC_5: Adventurer is defined as your travel style
Temple-outer-town	Wat Phra That Doi Kham Wat Umong Wat Suan Dok	H I J	TB_8: Overall cost of meals/food TB_9: Transport mode you plan to use is private car/motorcycle, van, coach for this visit TC_2: Books, guides are the information sources that have influenced your decision to visit

Temple-landmark	Wat Chedi Luang Wat Phra Singh	K L	TB_{10}: Heath care is the primary focus of this visit TC_6: People whom you are accompanied by are children TM_5: To not worry about time and work TB_{11}: The transport mode you plan to use during this visit is a bicycle
Temple-peaceful	Wat Lok Molee Wat Pan Tao	M N	TB_{12}: Entertainments activities are planned during this stay TM_6: To gain a new perspective on life TC_7: Friends/relatives have influenced your decision to visit TSD_3: Household annual income TM_3: To visit places I have never been before TC_8: The arrangements pertaining to this visit TC_1: Number of nights you plan to stay TM_7: To experience solitude and calm TB_{13}: Nightlife has made the deepest impression upon you TB_{14}: Attending festivals is the activity you plan to participate in during this visit
Temple-oldtown	Wat Sri Suphan Wat Chiang Man	O P	TEB_1: The amount of money you plan to spend per person on transportation during this visit TM_8: To improve my romantic life TEB_2: Miscellaneous expenses you have pre-paid before this visit TM_4: To develop new abilities TSD_3: Household annual income TM_5: To not worry about time and work TC_8: The arrangements pertaining to this visit TM_1: To visit relative(s)/friend(s)
Entertainment	Chiang Mai Cabaret Show Burklerk Gym- Muay Thai Training	Q R	TB_{14}: Attending festivals is the activity you plan to participate in during this visit TM_8: To understand more about myself TB_{15}: Thai boxing has made the deepest impression upon you TEB_3: The amount of money you plan to spend per person on shopping during this visit TM_9: To see famous cultural and historical sites TC_8: The arrangements pertaining to this visit

Table 5.2 represents the eight Chiang Mai destination choice data sets that we constructed in the previous study. Each data set used different kinds of features and different numbers. For example, we can see that each destination category is composed of both nominal and ordinal variables ranging from 3–10 variables. The Temple-peaceful data set used more features than other data sets (total of 10) to acheive its highest classification accuracy rate. On the other hand, Temple-outer town only used three features. Regarding feature type, Tourist behaviour (TB) was used the most (35%), while Tourist expenditure behaviour (TEB) appeared in only 4% of data sets. In this study, the same features that were built into the DT models were used for the construction of the SVM and MLP classifiers, as well as the same proportion of training and testing sets (85% and 15%).

In this study we constructed two experiments using SVM and MLP (as mentioned in Section 3) to stack up against with our baseline classifier C4.5 from the previous experiment. In total, six different classifiers were considered for the eight Chiang Mai destination choice data sets. We repeated all the experiments in this study 10 times using a stratified 5-fold cross-validation (i.e. four folds were selected for training, the rest were used for validation) procedure for each data set. For each classification algorithm we chose the classifier that achieved the highest classification accuracy rate. Table 5.2 presents the sizes of the training and testing data sets.

Table 5.2 Description of the Chiang Mai data sets for classification performance comparison.

Data set	#training	#testing	#classes	#features	#nominal	#ordinal
Nature	573	100	2	5	3	2
Museum	620	109	2	7	6	1
Art Gallery	331	58	2	8	7	1
Temple-outer town	1002	176	3	3	2	1
Temple-landmark	1364	240	2	4	3	1
Temple-peaceful	561	99	2	10	7	3
Temple-old town	617	108	2	9	5	4
Entertainment	587	103	5	6	4	2

Data pre-processing

All the variables in the data sets are categorical variables (e.g. ordinal, nominal) and it has been observed that these types of variables can cause a discontinuous relationship between an independent variable and a dependent variable (Brouwer, 2002). In order to prepare data for the SVM and MLP classifiers, nominal and ordinal variables for both inputs and outputs (only for the MLP classifier) need to be transformed into numerical variables (see Section 3.4.2), otherwise they may lead to an incorrect model. To ensure the generalisation capability of the models we have proposed two encoding schemes. First, One-of-N encoding scheme was used to represent each category as an integer (e.g. cat = (1 0 0), dog (0 1 0), mouse (0 0 1)). Second, the scheme employed a Thermometer encoding approach which is meaningful for ordinal variables. For instance, the variable could be coded using binary inputs such as (0 0), (0 1) and (1 1). Hence, all the inputs are scaled to the [0, 1] range. Since categorical variables lack a natural order in MLP, the data pre-processing set for the dependent variables and independent variables was encoded with the One-of-N encoding scheme.

Classification algorithms

Three classifiers, DT, SVM and MLP, are used in this study. An investigation into the classification performance of SVM and MLP was carried out. Details of the classification algorithms mentioned above are discussed as follows:

1. Decision Tree (DT)

   C4.5 (Quinlan, 1986b) is also used in this study, DT C4.5 is discussed in detail in Chapter 4. This study chose a post-pruning algorithm using 'subtree raising with confidence' to prevent over-fitting. Regarding hyper-parameter tuning, the confidence-factor ranged from 0.1 to 1 and the step size was set to 0.01. In this study, we deployed C4.5 as our baseline learner and benchmark model for the data sets.

2. Support Vector Machine

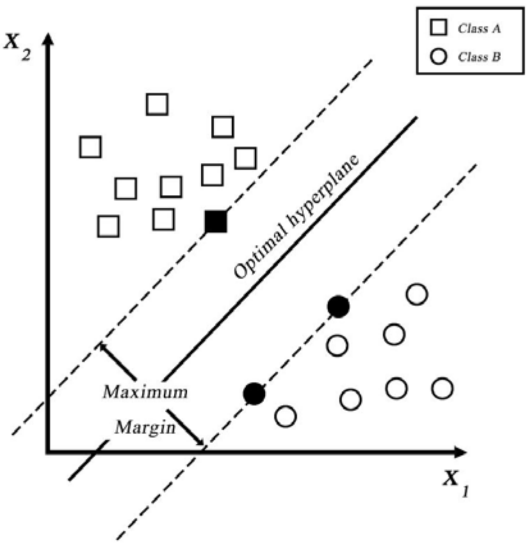


Figure 5.1 An example of a Hyper-plane in 2D space for a binary classification problem

A Support Vector Machine (SVM) (Chang and Lin, 2011), also known as a Support Vector Network, is typically used to address classification and regression problems. SVM has been successfully applied in many domains to address classification tasks, such as handwriting digital-character recognition, face detection and so on. This approach projects input into higher dimensional spaces so that non-linear data can be separated. The goal is to optimise the hyperplane, which can be separated into two classes of objects indicated by squares and circles, while maximising the distance of each point to the hyperplane as shown in Figure 5.1. SVM consists of two main phases. First, the kernel function is used to map the data to a higher dimension (i.e. linear, polynomial, radius bias function (RBF)). At this point the hyperplane can be used to separate the two classes. For a data set that cannot be perfectly separated linearly, the goal of the process is to find a set of weights that specify two hyperplanes, as defined below:

$$\begin{aligned} \vec{w} \cdot \vec{x} + b \geq +1 \\ \vec{w} \cdot \vec{x} + b \leq -1 \end{aligned} \tag{5.1}$$

In the case of non-linearly separable data, SVMs can handle non-separable points by introducing slack variables, as shown below:

$$y_i(\mathrm{w}^T x_i + b) \geq 1 - \xi_i \tag{5.2}$$

Hence, for a non-separable data set, the goal of SVM is to find the hyperplane with a maximum margin that also minimises slack terms. Many kernels have been proposed by researchers including linear, polynomial and sigmoid kernels. In this study the Gaussian RBF kernel was selected as the most suitable kernel function because our data set consists of a small number of features (i.e. 3–14) and RBF uses fewer hyper-parameters than the polynomial kernel. The Gaussian RBF, as defined in equation (5.3), was selected for this study.

$$f(x_i) = \exp(-\frac{1}{(2\sigma^2)}\left\| x_i - x_j \right\|^2) \tag{5.3}$$

The term $\frac{1}{2\sigma^2}$ can be replaced by γ, where $\gamma > 0$, and $\left\| x_i - x_j \right\|^2$ is the distance between the two feature support vectors.

Regarding the advantages of SVM, this classifier is capable of finding a global minimum and its simple geometric interpretation provides fertile ground for future investigations. The most advantageous characteristic of the nonlinear SVM classifier is convexity. However, SVM also has a few drawbacks: it is very sensitive to kernel parameters, and choice of kernel; therefore, selecting a slightly out-of-margin parameter may result in low classification performance.

Tuning these parameters is usually necessary for good performance. For example, choosing a cost parameter is critical. Using a larger cost value may lead to over-fitting of the model. Also, developing a model with SVM requires a laborious trial-and-error approach and is quite time-consuming, especially for a large volume of data. Table 5.3 illustrates the algorithm that was used to optimize the hyper-parameters in SVM in this study.

Table 5.3 Grid search cross validation algorithm

Algorithm 5.1: Grid search cross-validation

1: **Input:** D_{train}, $\log_2$c_vector, $\log_2$g_vector
2: **Output:** $w^*(c,\gamma)$ % Large scale search
3: stepsize = 1;
4: **for** i=1:numl($\log_2$c_vector) % loop through every element in the list.
5: **for** j=1:numel($\log_2$g_vector)
6: $w^*(c,\gamma) = \arg\max_w CV(w,c,\lambda,D_{train})$
7: **if** w*>best w
8: c*=c,g*=g;
9: **end**
10: **end**
11: **end**
12: stepsize = prev_stepsize ÷ 2; % Adjust the medium-scale and small-scale search
13: $\log_2$c_vector = c*-prev_stepSize:stepsize:g*+prev_stepsize;

SVM can only separate binary-class problems. So, to handle a multi-class problem, *N* different binary classifiers were created, and the one with the highest classification accuracy rate was selected. This technique is commonly known as the one-vs-all method or one-vs-rest method. For instance, we need to build *K* classifiers for the three-classification problem, and each classifier is dedicated to one class. The condition can be justified as:

$$y_i = \begin{cases} +1, x_i \in C_k \\ -1, x_i \notin C_k \end{cases} \tag{5.4}$$

Another method to deal with a multi-class classification problem is to train *K*(*K*-1)/2, also known as a One-vs-One or pairwise SVM method, in which a voting scheme is applied. In short, the procedure of the SVM model's construction is as follows:

1. Conduct scaling on the data sets

2. Try different kernels (linear, polynomial, RBF)
3. Use grid-search cross-validation to find the optimal parameters
4. Train the SVM model using the training set with the obtained optimal parameters
5. Test using a test data set and evaluate using performance metrics

3. Multi-layer perceptron

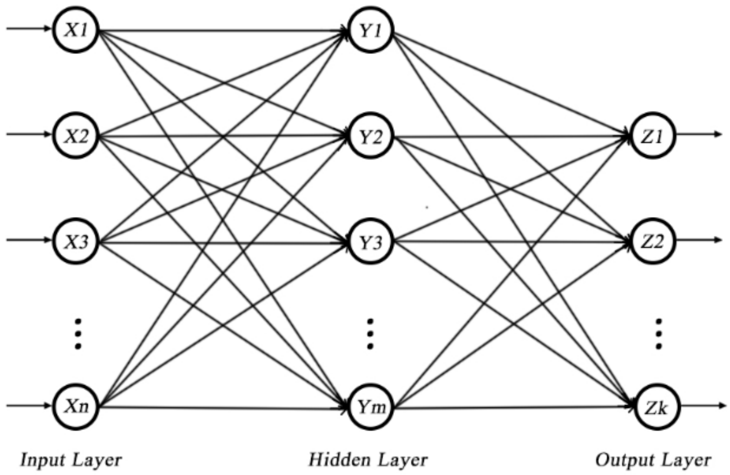


Figure 5.2 An architecture of the multi-layer perceptron with one hidden layer

A multi-layer perceptron (MLP) (Bishop, 1996) is considered to be a feed-forward network, a universal approximator inspired by the biological neural networks in the human brain. It is the most commonly applied method in the area of artificial neural networks (ANN) for handling classification tasks. A neural network can be trained to predict a class variable. There are many types of ANNs used for classification including MLP, radial basic function and probabilistic neural networks. In this study, MLP was selected as the network type; its architecture consists of one or more hidden layers between the input and output nodes and each of the nodes in the network is connected and has a certain weight. Figure 5.2 illustrates the overall network architecture of the MLP. MLP maps the data from feature space to classification output space and prediction can be selected as the encoding vector that is closest to the output (i.e. the output that displays the highest value is the winning class). The effective back propagation algorithm was used to train the network at the error-correction stage. An MLP model was designed using the following criteria:

1. The network architecture consists of one input layer, one hidden layer and one output layer. The input layer contains input vectors and there is no computation performed here. In the hidden layer, we choose one hidden layer which receives input from the input layer, most complex problems can be solved using one hidden layer (Heaton, 2008). The output layer contains the output vector where the activate function is applied.
2. Selection of the number of hidden nodes. As far as we know, no conclusion has been reached regarding the number of hidden neurones that should be used in the hidden layer; therefore, the number of optimal hidden nodes is based on a process of trial and error. Deciding on the number of hidden neurones in the hidden layer is critical as it may lead to over-fitting and longer computation times if we use too many hidden neurones or under-fitting where there are too few neurones in the hidden layer. In this study we consider two approaches. For the first approach, the number of hidden neurones was adjusted and ranged from 1 to 100 nodes, which were trained, validated, and tested with a step size of 1. The second approach involved the selection of a number of neurones based on a rule of thumb defined as follows:
   - The number of hidden neurones is 2/3 of the size of the input layer (Boger and Guterman, 1997). The formula can be defined as 2/3(N_i), where N_i is the number of input neurones in the network.
   - The hidden output-connection weight becomes as small as the number of hidden neurones N_h becomes large (Shibata and Ikeda, 2009). The formula can be written as $N_h = sqrt(N_i * N_o)$, where N_i represents the number of input neurones and N_o represents the number of output neurones acquired from the network.
   - Trenn (2008) defined the number of hidden neurones as $N_h = n+n_0 - (1/2)$, where n represents the number of inputs and n_0 represents the number of outputs.
3. The *softmax* function was used as the activation function for all the layers and both binary and multi-class classification problems. The function guarantees that the sum of all class probabilities is equal to 1. Considering that we have a vector *x* of *K* outcomes, the function can be calculated as:

$$f(x_i) = \frac{\exp(x_i)}{\sum_{j=0}^{K} \exp(x_j)}, i = 0..K \qquad (5.5)$$

4. The Scaled Conjugate Gradient (SCG) back-propagation algorithm (Møller, 1993) was used in this study when training the network. SCG is thought to be better than the standard back-propagation algorithm as it eliminates certain important disadvantages, such as poor convergence rate and user-parameter dependency. The network was trained and validated 10 times due to the disadvantages of ANN, in which suffers from multiple local minima. The network that displayed the highest accuracy rate was selected. In short, the procedure of the MLP model selection and assessment was as follows:

    1. Conduct scaling of the input and output
    2. Use a cross-validation search for the optimal number of hidden neurons from 1:1:100
    3. Use a cross-validation search for the optimal number of hidden neurons using a rule of thumb
    4. Train the network with the obtained optimal number of hidden neurons
    5. Test with test data and evaluate using performance metrics.

5.1.3 Experimental results

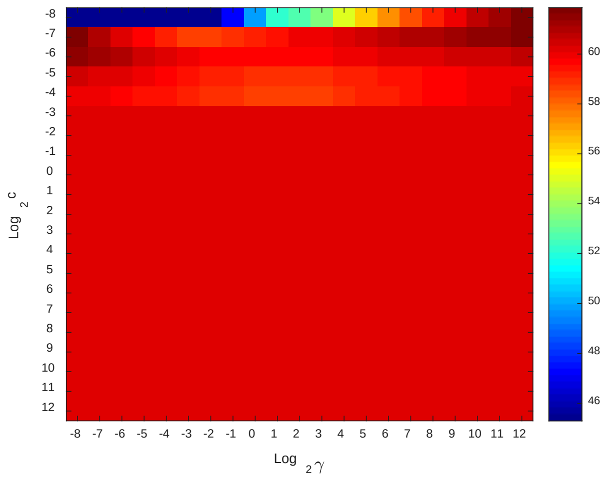
In this section, the experiments performed on destination classification of the eight data sets are described and the results compared and discussed.

The SVM results were obtained by using *LibSVM* library (Chang and Lin, 2011), an open source library for constructing the SVM model. Moreover, two other SVM libraries from the Matlab®, Statistics and Machine Learning toolbox, i.e. *svmtrain*, *svmclassify* from the earlier Matlab version and *fitcsvm*, were investigated in this study. The two implementations of SVM from Matlab have different parameters to configure. For instance, a number of iterations were required for *svmtrain* which do not appear in *fitcsvm*. For all the SVM classifiers, the random seed was set to 1 in order to be able to reproduce the results.

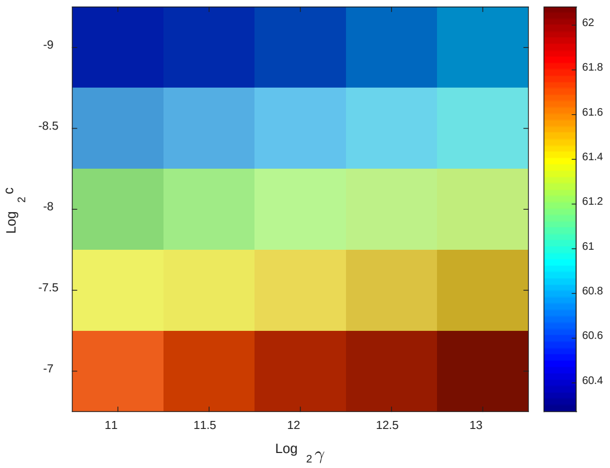
To use *LibSVM* we first transformed the data into a relevant format in the SVM package. Training data sets did not require to be shuffled as SVM will always converge to the same solution for a given data set (Veropoulos et al., 1999). After that, both training and testing data sets that were used to construct C4.5 baseline learners were transformed using the One-of-N encoding scheme for nominal variables, and the thermometer encoding scheme for ordinal variables as described in Section 5.3.2.

A process of model development and model selection was carried out, the goal being to identify optimal hyperparameters (C, γ). The parameters C (penalty for misclassification) and gamma (a function of the deviation of the Gaussian kernel) were determined by using stratified 5-fold cross-validation (i.e. four folds were selected for training, and the rest were used for validation).

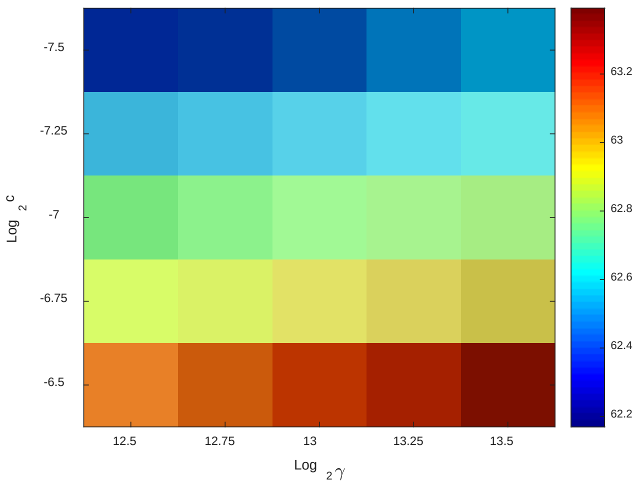
A grid-search technique (including large-, medium-, and small-scale parameters) and stratified 5-fold cross-validation were applied to the training sets for the process of model regularisation. A large-scale search (see Fig. 5.3(a)) was first conducted to identify a better region in the grid, so that finer grid searches (see Figs 5.3(b) and (c)) could be conducted in the neighbourhood later. The three SVM classifiers were experimented in different ranges of hyper-parameters. For each data set we estimated the generalised rate of accuracy using all combinations of kernel parameters C and parameters γ, as shown in Table 5.4. For instance, in SVM*libsvm*, the ranges of C and γ values are 2^{-10} to 2^{30}. After the best C and γ values were found, based on the highest cross-validation accuracy rate, the entire training set was trained again using the obtained (C, γ) and tested with the testing set (unseen data). To handle multi-class problems, such as the Temple-outer town data set, both One-Vs-One and One-vs-All methods were used for all the SVM classifiers.



(a) Large-scale grid search



(b) Medium-scale grid search



(c) Small-scale grid search

Figure 5.3 Heat maps for the Museum data set generated by SVM using the *svmtrain* function

Next, we investigated the classification performance of all eight data sets. By observing the cross-validation accuracy rates from the heat maps generated by all the SVM classifiers we noticed difference in classification-accuracy results from three of the SVM toolboxes with respect to the range of C and γ values. First, increases in the C and γ values for *libsvm* and *svmtrain* from 15 did not increase the classification accuracy. Additionally, *svmtrain* took longer to compute when the C and γ values were higher, especially after a value of 15. On the other hand, *fitcsvm* is very sensitive to these values, therefore we increased the value of the kernel parameters ranging from [-5, 15] to [-5, 30]. The SVM experimental setup is described in detail in Table 5.4.

Table 5.4 Experiment designs for SVM classifiers with details of parameters

Objective function	$\frac{1}{2}\|w\|\frac{2}{2}+C\sum_{i=0}^{N}\max(0,1-o_i y_i)$
Kernel function	$f(x_i)=\exp(-\frac{1}{(2\gamma^2)}\|x_i - x_j\|^2)$
Cost (denoted as *C*)	2^{-20}, 2^{-19},…, 2^{30}(*LIBSVM*) 2^{-8}, 2^{-7},…, 2^{10}(*svmtrain*) 2^{-5}, 2^{-4},…, 2^{30} (*fitcsvm*)
Gamma (denoted as γ)	2^{-20}, 2^{-19},…, 2^{30}(*LIBSVM*) 2^{-8}, 2^{-7},…, 2^{10}(*svmtrain*) 2^{-5}, 2^{-4},…, 2^{30} (*fitcsvm*)

With respect to the results for SVM_{libsvm}, the highest classification accuracy rate obtained for the Nature data set was 65%, using C = -0.25, γ = -1.25. The highest classification accuracy rate for the Museum data set was 70.64%, using C = 30, γ = -13.5. For the Art Gallery data set, the classifier achieved a highest classification accuracy rate of 58.62%, using C = 28.75, γ = -13. For the Temple-outer town data set the highest classification accuracy rate obtained was 47.16%, using C = 5, γ = -3.5. The highest classification accuracy rate obtained for the Temple-landmark data set was 62.08%, using C = -2.25, γ = 31.5. For the Temple-peaceful data set the highest classification accuracy rate obtained was 60.61%, using C = 1, γ =-5.25. For the Temple-old town data set the highest classification accuracy rate of 63.89%, C = 1, γ = -3 was used. Last, the highest classification accuracy for the Entertainment data set was 74.75% and the value of cost and gamma that were used were (19.5, -15.5). The confusion matrix, precision, recall and F-score for SVM_{libsvm} are presented in Tables 5.5 and 5.8.

Surprisingly, the SVM_{m1} results for the Museum data set and Temple-outer town data set were quite acceptable. As regards the total training time, we found that SVM_{m1} took relatively longer to converge than SVM_{libsvm} and SVM_{m2} for all data sets. Out of all the SVM classifiers, the overall training time for the Temple-outer town data set was longer than for the other data sets. This is because this data set has more classes than the others. Concerning the speed of convergence of *svmtrain*, we lowered the cost values and increased the value of the parameter 'tolkkt', which specifies the tolerance with which the Karush-Kuhn-Tucker (KKT) conditions are checked for the Sequential Minimal Optimazation (SMO) train method.

Regarding the results for SVM_{m1}, the highest classification accuracy rate obtained for the Nature data set was 58%, using C = 16.5, γ = 16. The highest classification accuracy rate obtained for the Museum data set was 80.73%, using C = -5, γ = 2. The highest classification accuracy rate obtained for the Art Gallery data set was 70.69%, using C = 15, γ = 6. The highest classification accuracy rate obtained for the Temple-outer-town data set was 62.5%, using C = 2, γ = -2. The highest classification accuracy rate obtained for the Temple-landmark data set was 55%, using C = 11.5, γ = 11. SVM_{m1} performed poorly in this data set and, when observing classification-accuracy rates, it appears that the model is over-fitted because the cross-validation accuracy of this model reached 90.98% but the test rate only reveals a classification rate of 55%. For the Temple-peaceful data set the classifier obtained a highest classification accuracy rate of 65.66%, using C = 0.5, γ = 1.5. The highest classification accuracy rate obtained for the Temple-old town data set was 69.44%, using C = 3, γ = -7. Finally, the highest classification accuracy rate obtained for the Entertainment dataset was 75.73%, and the values of cost and gamma that were used were (6, 13.5). The confusion matrix, precision, recall and F-score for SVM_{m1} are presented in Tables 5.6 and 5.9.

For SVM_{m2} we can see that the best cross-validation classification accuracy rate was found with higher values of C and γ than for the other SVMs. Regarding the results, the highest classification accuracy rate obtained for the Nature data set was 58%, using C = 27, γ = 1. The highest classification accuracy rate obtained for the Museum data set was 74.31%, using C = 27, γ = 17. The highest classification accuracy rate obtained for the Art Gallery data set was 68.97%, using C = 5, γ = 2.5. The highest classification accuracy rate obtained for the Temple-outer town data set was 50%, using C = -1, γ = -1. The highest classification accuracy rate obtained for the Temple-landmark data set was 62.92%, using C = 27, γ = 17. The highest classification accuracy rate obtained for the Temple-peaceful data set was 62.63%, using C = 15, γ = 6. The highest classification accuracy rate obtained for the Temple-old town dataset was 69.44%, using C = 27, γ = 14. Finally, the Entertainment data set achieved a highest classification accuracy rate of 71.84%, and the value of cost and gamma that were used were (18, 27). The confusion matrix, precision, recall and F-score for SVM_{m2} are presented in Tables 5.7 and 5.10. The bold font in the confusion matrix indicates correctly classified instances.

Table 5.5 Confusion matrix for SVM$_{libsvm}$ for all data sets

	Predict																
Actual	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
B	**25**	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	14	**40**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	**19**	23	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	9	**58**	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	**13**	14	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	10	**21**	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	**63**	6	3	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	38	**12**	7	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	27	12	**8**	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	**79**	44	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	47	**70**	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	**51**	8	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	31	**9**	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	**58**	9	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	30	**11**	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	**30**	17
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	9	**47**

Table 5.6 Confusion matrix for SVM$_{m1}$ for all data sets

	Predict																
Actual	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
B	**37**	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	33	**21**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	**28**	14	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	7	**60**	11	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	**17**	10	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	7	**24**	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	**58**	12	2	21	0	0	0	0	0	0	0
I	0	0	0	0	0	0	31	**16**	10	10	0	0	0	0	0	0	0
J	0	0	0	0	0	0	21	18	**8**	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	**119**	4	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	104	**13**	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	**38**	21	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	13	**27**	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	**57**	10	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	23	**18**	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	**40**	7
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	18	**38**

Table 5.7 Confusion matrix for SVM$_{m2}$ for all data sets

	Predict																
Actual	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**
B	***35***	*11*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
C	*31*	***23***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
D	*0*	*0*	***21***	*21*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
E	*0*	*0*	*7*	***50***	*11*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
F	*0*	*0*	*0*	*0*	***18***	*8*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
G	*0*	*0*	*0*	*0*	*9*	***22***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
H	*0*	*0*	*0*	*0*	*0*	*0*	***62***	*7*	*3*	*21*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
I	*0*	*0*	*0*	*0*	*0*	*0*	*32*	***18***	*7*	*10*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
J	*0*	*0*	*0*	*0*	*0*	*0*	*25*	*14*	***8***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
K	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*77*	*46*	*0*	*0*	*0*	*0*	*0*	*0*
L	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*43*	***74***	*0*	*0*	*0*	*0*	*0*	*0*
M	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***53***	*6*	*0*	*0*	*0*	*0*
N	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*31*	***9***	*0*	*0*	*0*	*0*
O	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***61***	*6*	0	0
P	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*27*	***14***	0	0
Q	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***0***	***46***	*1*
R	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***0***	*28*	***28***

Table 5.8 Precision, recall and F-score for SVM$_{libsvm}$

Destination	Precision	Recall	F-score
B	0.6410	0.5435	0.5882
C	0.6557	0.7407	0.6957
D	0.6786	0.4524	0.5429
E	0.7160	0.8657	0.7838
F	0.5652	0.4815	0.5200
G	0.6000	0.6774	0.6364
H	0.4922	0.8750	0.6300
I	0.4000	0.2105	0.2759
J	0.4444	0.1702	0.2462
K	0.6270	0.6423	0.6345
L	0.6140	0.5983	0.6061
M	0.6220	0.8644	0.7234
N	0.5294	0.2250	0.3158
O	0.6591	0.8657	0.7484
P	0.5500	0.2683	0.3607
Q	0.7692	0.6383	0.6977
R	0.7344	0.8393	0.7833

Table 5.9 Precision, recall and F-score for SVM_{m1}

Destination	Precision	Recall	F-score
B	0.5286	0.8043	0.6379
C	0.7000	0.3889	0.5000
D	0.8000	0.6667	0.7273
E	0.8180	0.8955	0.8511
F	0.7083	0.6296	0.6667
G	0.7059	0.7742	0.7385
H	0.5273	0.8056	0.6374
I	0.3478	0.2807	0.3107
J	0.4000	0.1702	0.2388
K	0.5336	0.9675	0.6879
L	0.7647	0.1111	0.1940
M	0.7451	0.6441	0.6909
N	0.5625	0.6750	0.6136
O	0.7125	0.8507	0.7755
P	0.6429	0.4390	0.5217
Q	0.6897	0.8511	0.7619
R	0.8444	0.6786	0.7525

Table 5.10 Precision, recall and F-score for SVM_{m2}

Destination	Precision	Recall	F-score
B	0.5303	0.7609	0.6250
C	0.6765	0.4259	0.5227
D	0.7500	0.5000	0.6000
E	0.7407	0.8955	0.8108
F	0.6667	0.6667	0.6667
G	0.7097	0.7097	0.7097
H	0.5210	0.8611	0.6492
I	0.4615	0.3458	0.3750
J	0.4444	0.1702	0.2462
K	0.6417	0.6260	0.6337
L	0.6167	0.6325	0.6245
M	0.6310	0.8983	0.7413
N	0.6000	0.2250	0.3273
O	0.6932	0.9104	0.7871
P	0.7000	0.3415	0.4590
Q	0.6216	0.9787	0.7603
R	0.9655	0.5000	0.6588

The experimental results show that SVM is very sensitive to the hyperparameter. A general observation for SVM was acknowledged when increasing the values of the cost and grammar parameters, especially for the *fitcsvm* function. The validation classification accuracy values for the Nature, Art Gallery and Temple-peaceful categories were found to have increased. However, this led to longer training times for the models.

With respect to MLP, two different MLP libraries were used. One is referred to as MLP_m and is derived from the Matlab® Neural Network Toolbox, and the other one is referred to as MLP_n, and is derived from the Netlab Neural Network Toolbox (*NETLAB – Algorithms for Pattern Recognition, Ian T. Nabney, Springer*, 2002). Different search strategies, according to the number of hidden neurons were used, as described above. First, different numbers of hidden neurons ranging from 1 to 100 with an increment of 1 were validated in order to find the optimal model. After that, the four rules of thumb to find the optimal number of hidden neurons, referred to earlier, were utilised. Due to the instability of ANN (it usually suffers from multiple local minima), the feed-forward network was trained and validated 50 times. An optimal number of hidden neurons was selected based on the highest cross-validation accuracy rate. The test data set was then applied to the network to obtain the true performance of the model.

Regarding MLP, a random selection of the number of hidden neurons usually causes over-fitting and under-fitting problems. For example, excessive hidden neurons will cause overfitting because the network has overestimated the complexity of the problem. In this report, we applied four methods to fix the number of neurons in the hidden layer. This includes three rules of thumb found in previous studies and a sequential-search approach ranging from 1 to 100 with a step size of 1. Figure 5.14 shows the impact of accuracy on the number of hidden neurons. It can be seen that all the MLP models used fewer number of hidden neurons to achieve the highest cross-validation accuracy rates. For both MLP classifiers we can see that increasing the number of hidden neurons does not guarantee that a better classification accuracy rate will be achieved for all data sets.

For MLP_n the highest cross-validation accuracy rate obtained was 56.54% ±4.41 for the Nature data set, 69.03%±3.85 for the Museum data set, 61.29%±5.92 for the Art Gallery data set, 46%±2.58 for the Temple-outer town data set, 59.09%±2.36 for the Temple-landmark data set, 63.46%±4.57 for the Temple-peaceful data set, 65.82%±5.51 for the Temple-old town data set and 71.04%±1.96 for the Entertainment data set.

For MLP_m we achieved a higher cross-validation accuracy than MLP_n, except in the Museum data set, where MLP_n obtained a slightly higher rate. The maximum cross-validation accuracy rate obtained was 60.21% ±3.51 for the Nature data set, 68.87%±6.39 for the Museum data set, 63.15%±3.95 for the Art Gallery data set, 44.41%±2.97 for the Temple-outer town data set, 59.23%±1.60 for the Temple-landmark data set, 63.63%±3.36 for the Temple-peaceful data set, 67.59%±3.75 for the Temple-old town data set and 71.72%±4.98 for the Entertainment data set.

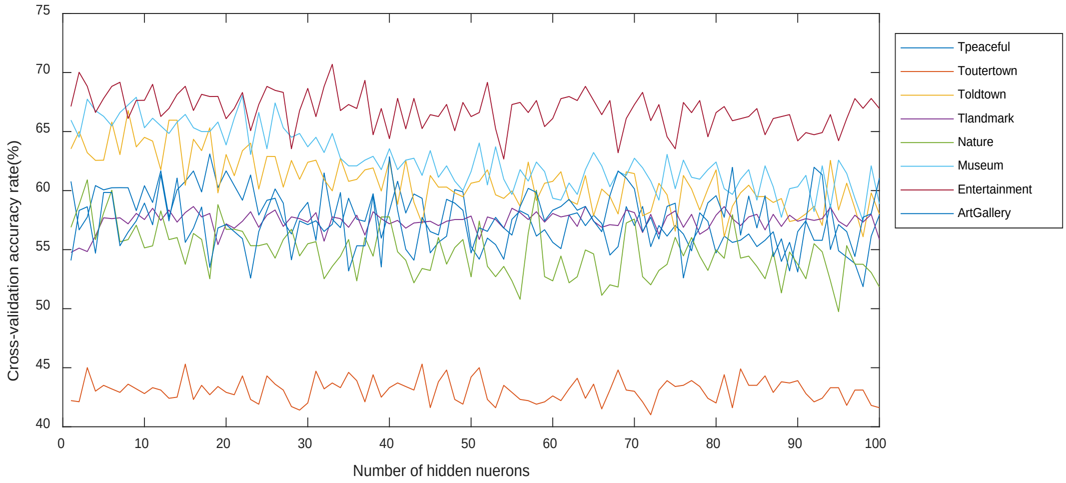


Figure 5.4 Cross-validation accuracy of MLP$_m$ on data sets and the number of hidden neurons

Table 5.11 shows that the Temple-outer town data set used the largest number of hidden neurons: 16 for the MLP$_m$ and 23 for the MLP$_n$ classifiers. The table also reveals that MLP$_\mathbf{m}$ used a higher number of hidden neurons than did MLP$_n$.

Table 5.11 Optimal numbers of hidden neurons in the data sets

Data set	MLP*n*	MLP*m*	d
Nature	14	10	4
Museum	1	6	-5
Art Gallery	1	15	-14
Temple-outer-town	23	16	7
Temple-landmark	2	10	-8
Temple-peaceful	1	20	-19
Temple-oldtown	2	9	-7
Entertainment	14	5	9
AVG.	7.25	11.38	-4.13

Table 5.12 Precision, recall and F-score for MLP_n

Destination	Precision	Recall	F-score
B	0.6067	0.5870	0.6067
C	0.6847	0.7037	0.6847
D	0.7576	0.5952	0.6667
E	0.7763	0.8806	0.8252
F	0.7500	0.5556	0.6383
G	0.6842	0.8387	0.7536
H	0.5304	0.8472	0.6524
I	0.4884	0.3684	0.4200
J	0.4444	0.1702	0.2462
K	0.6357	0.6667	0.6508
L	0.6306	0.5983	0.6140
M	0.7077	0.7797	0.7419
N	0.6176	0.5250	0.5676
O	0.7432	0.8209	0.7801
P	0.6471	0.5366	0.5867
Q	0.7500	0.8298	0.7879
R	0.8431	0.7679	0.8037

Table 5.13 Precision, recall and F-score for MLP_m

Destination	Precision	Recall	F-score
B	0.6667	0.6400	0.6667
C	0.6923	0.7200	0.6923
D	0.8571	0.5714	0.6857
E	0.7778	0.9403	0.8514
F	0.7391	0.6296	0.6800
G	0.7143	0.8065	0.7576
H	0.5304	0.8472	0.6524
I	0.4884	0.3684	0.4200
J	0.4444	0.1702	0.2462
K	0.6277	0.6992	0.6615
L	0.6408	0.5641	0.6000
M	0.7333	0.7458	0.7395
N	0.6154	0.6000	0.6076
O	0.7250	0.8657	0.7891
P	0.6786	0.4634	0.5507
Q	0.7600	0.7600	0.7835
R	0.8302	0.8302	0.8073

The experimental results show that our approach achieved the highest classification accuracy rate. By using a statistical test, we found that MLP_m was a significant improvement over other models (p-value < 0.05).

Table 5.14 Classification accuracy rates for C4.5, SVMs and MLPs

Individual models												
Data set	**C4.5**		**SVM*libsvm***		**SVM*m1***		**SVM*m2***		**MLP*n***		**MLP*m***	
	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test
Nature	59.87	64	62.47	65	57.57	58	59.35	58	56.54	65	60.21	**68**
Museum	68.87	75.23	67.42	70.64	69.84	**80.73**	66.04	74.31	69.03	77.06	68.87	79.82
Art-gallery	60.71	68.97	61.63	58.62	62.52	70.69	70.91	68.97	61.29	70.69	63.15	**72.41**
T-outer town	44.71	51.14	45.21	47.16	60.08	**62.50**	45.35	50	46.01	51.14	44.41	51.14
T-land mark	58.87	62.08	59.01	62.08	90.98	55.00	54.37	62.92	59.09	**63.33**	59.05	**63.33**
T-peaceful	63.29	**68.69**	64.71	60.61	60.07	65.66	68.82	62.63	63.46	67.68	63.63	**68.69**
T-old town	66.28	70.37	68.23	63.89	65.97	69.44	69.39	69.44	65.81	**71.30**	67.59	**71.30**
Entertainment	72.57	78.64	72.91	74.75	70.87	75.73	65.19	71.84	71.04	**79.61**	71.72	**79.61**
AVG.	61.89	67.39	62.7	62.84	67.23	67.22	62.43	64.76	61.53	68.23	62.33	**69.29**

SVM$_{m1}$ indicates SVM using Matlab *svmtrain* and *svmclassify* functions SVM$_{m2}$ indicates SVM using Matlab *fitcsvm* function; *CV indicates cross-validation accuracy rate.*

Table 5.14 shows the results for cross-validation accuracy and a test set comparing six classifiers. Regarding the overall comparison, the global best for each data set is denoted using a bold font. By observing the averaged classification results we can conclude that the three classifier algorithms produce similar averaged accuracy performance for most of the data sets. We can see that the averaged classification accuracy for MLPs is slightly higher than that of the other algorithms. When using a Shapiro-Wilk statistical test, all the data sets show a normal distribution (*p-value* > 0.05). Next, a paired T-test was used and the difference between each model is statistically significant in terms of an improvment in accuracy rate: MLP$_m$ >* MLP$_n$ >* SVM$_{m2}$ > SVM$_{m1}$ > SVM$_{libsvm}$ > C4.5, where >* indicates 'significantly better at a 95% confidence interval' and > indicates 'no significant difference'. The experimental results also show that MLP$_m$ reaches a higher classification accuracy rate for all data sets than other classifiers, except the Museum, and Temple-outer town data sets. Interestingly, SVM$_{m1}$ achieved the highest classification accuracy rate for the Museum and Temple-outer town data sets but the model is not significantly better than the baseline (*p-value* = 0.94). It can be stated that none of the SVM models performed very well, especially SVM$_{libsvm}$ and SVM$_{m2}$ which are significantly worse than the other classification algorithms and the baseline learner.

When observing the classification accuracy rate for each data set we can see that MLP$_m$ and MLP$_n$ are superior to the other classifiers, while SVM$_{libsvm}$ delivers the worst performance. As is evident from Figure 5.11 and Table 5.13, MLP$_m$ provides a higher classification accuracy rate than the other classifiers for all data sets except the Museum data set, for which C4.5 reveals the highest classification accuracy rate. It is plain to see that MLP$_m$ performed better than the rest of the classifiers for all the data sets. It can stated that SVM and C4.5 did not perform very well. Interestingly, SMV$_{m1}$ achieved the highest classification accuracy value of 79.82%, the same as MLP$_m$ for the Art Gallery data set.

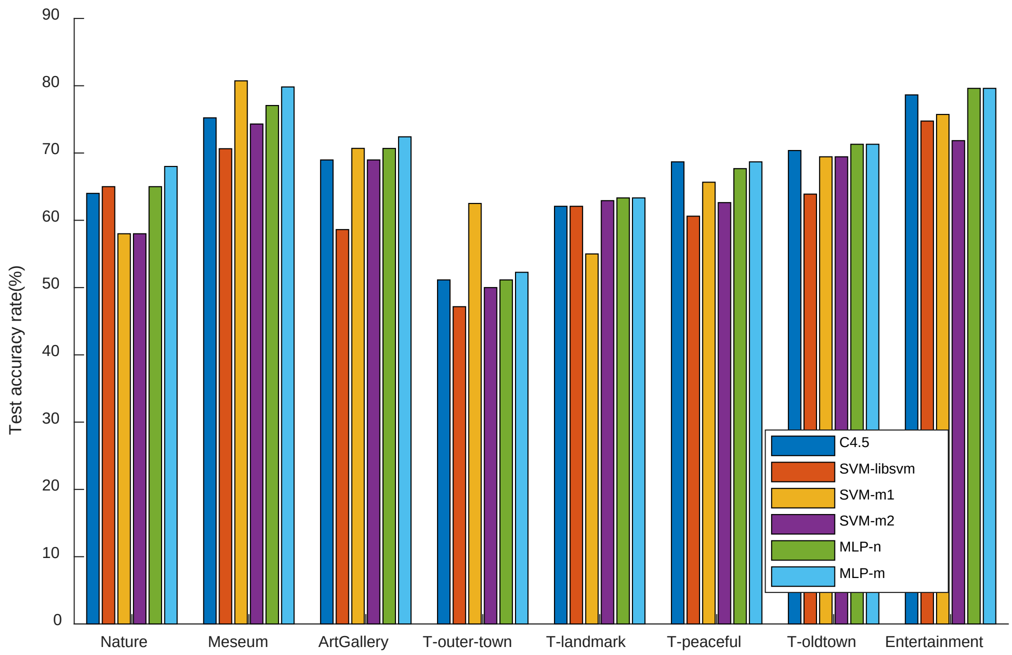


Figure 5.5 Test classification accuracy-rate comparison of individual classifiers

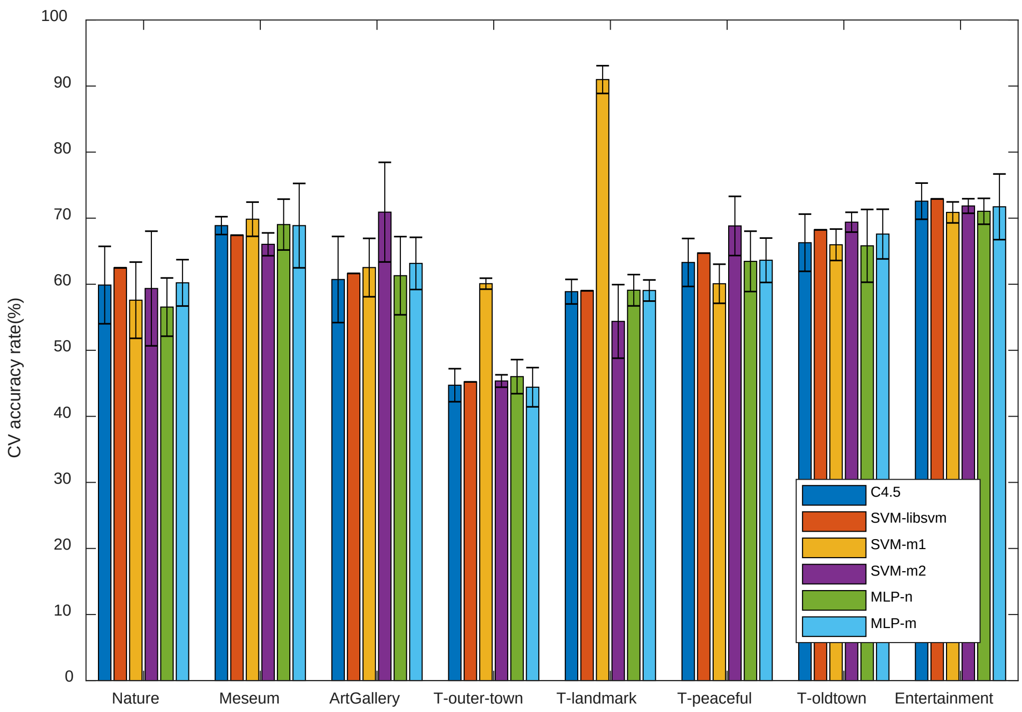


Figure 5.6 Cross-validation accuracy-rate comparison of individual classifiers

We achieved a highest validation classification accuracy value of 68% for the Nature category, 79.82% for the Museum category, 72.41% for the Art Gallery category, 51.14% for the Temple-outer town category, 63.33 for the Temple-landmark category, 68.69% for the Temple-peaceful category, 71.30% for the Temple-old town category and 79.61% for the Entertainment category.

Table 5.15 F-score comparison of classifiers for each data set.

Model	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
C4.5	0.660	0.617	0.609	0.819	0.640	0.727	0.652	0.420
SVM$_{lib}$	0.588	0.696	0.543	0.784	0.520	0.636	0.630	0.276
SVM$_{m1}$	0.638	0.500	0.727	0.851	0.667	0.739	0.637	0.637
SVM*m2*	0.625	0.523	0.600	0.811	0.667	0.710	0.649	0.375
MLP$_{n}$	0.607	0.685	0.667	0.825	0.638	0.754	0.652	0.420
MLP$_{m}$	0.667	0.692	0.686	0.851	0.680	0.758	0.652	0.420

Model	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**
C4.5	0.246	0.646	0.592	0.783	0.436	0.789	0.500	0.792	0.780
SVM$_{libsvm}$	0.246	0.635	0.606	0.723	0.316	0.748	0.361	0.698	0.783
SVM$_{m1}$	0.239	0.688	0.194	0.691	0.614	0.776	0.522	0.762	0.753
SVM$_{m2}$	0.246	0.634	0.625	0.741	0.327	0.787	0.459	0.760	0.659
MLP$_{n}$	0.246	0.651	0.614	0.741	0.568	0.780	0.587	0.788	0.804
MLP$_{m}$	0.246	0.662	0.600	0.741	0.608	0.789	0.551	0.784	0.807

As can be seen, the SVMs classifiers did not perform well for any data set. This is due to an imbalance among them. A general observation of SVM was acknowledged when increasing the values of the cost and grammar parameters, in particular for the *fitcsvm* function. The validation classification accuracy value for the Nature, Art Gallery and Temple-peaceful categories were found to have increased. However, this leads to longer training times for the models.

Table 5.16 Confusion matrix of MLP$_m$ for all data sets

	Predict																
Actual	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**
B	**32**	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	14	**36**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	**24**	4	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	18	**63**	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	**17**	10	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	6	**25**	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	**61**	8	3	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	29	**21**	7	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	25	14	**8**	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	**86**	51	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	37	**66**	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	**44**	16	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	15	**24**	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	**59**	22	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	9	**19**	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	**38**	12
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	9	**44**

Table 5.17 Confusion matrix of MLP$_n$ for all the data sets

	Predict																
Actual	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**
B	***27***	*19*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
C	*16*	***38***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
D	*0*	*0*	***25***	*17*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
E	*0*	*0*	*8*	***59***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
F	*0*	*0*	*0*	*0*	***15***	*12*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
G	*0*	*0*	*0*	*0*	*5*	***26***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
H	*0*	*0*	*0*	*0*	*0*	*0*	***61***	*8*	*3*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
I	*0*	*0*	*0*	*0*	*0*	*0*	*29*	***21***	*7*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
J	*0*	*0*	*0*	*0*	*0*	*0*	*25*	*14*	***8***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
K	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***86***	*41*	*0*	*0*	*0*	*0*	*0*	*0*
L	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*47*	***70***	*0*	*0*	*0*	*0*	*0*	*0*
M	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***46***	*13*	*0*	*0*	*0*	*0*
N	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*19*	***21***	*0*	*0*	*0*	*0*
O	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***55***	*12*	0	0
P	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*19*	***22***	0	0
Q	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***0***	***39***	*8*
R	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***0***	*13*	***43***

Tables 5.16 and 5.17 show the confusion matrices for MLP classifiers for all eight data sets. They show that MLP$_m$ is better at discriminating classes in the Temple-outer town data set than is MLP$_n$. It can be seen that the classifier MLP$_n$ is confused between destination I (Wat Umong) and destination J (Wat Suan Dok). On the other hand, classifier MLP$_n$ is confused between destination H (Wat Phra That Doi Kham) and I (Wat Umong).

5.1.1 Experiment 1*:* Discussion

In this study, different classification algorithms were compared, including C4.5, SVM and MLP. We investigated the performance of six classifiers. The experimental results indicate that MLP is the most robust classification algorithm for the Chiang Mai data sets. We achieved the highest classification accuracy rate of 79.82% for the Museum data set and 69.3% for the mean of all data sets. Figures 5.6–5.13 show the ROC curves and AUC plots for the data sets. When observing ROC curves and AUC values, it can be seen that SVM$_{m1}$ and SVM$_{m2}$ are better at discriminating between classes than other classifiers for the Nature, Temple-old town and Entertainment data sets. All the classifiers show the same ability in the tests to correctly classify two destinations in the Temple-landmark and Temple-peaceful data sets (see Figs 5.10 and

5.11). All the classifiers found it easy to discriminate between two destinations in the Entertainment data set; this is because destination Q is related to night life and destination R is related to outdoor entertainment, so they are easy to classify.

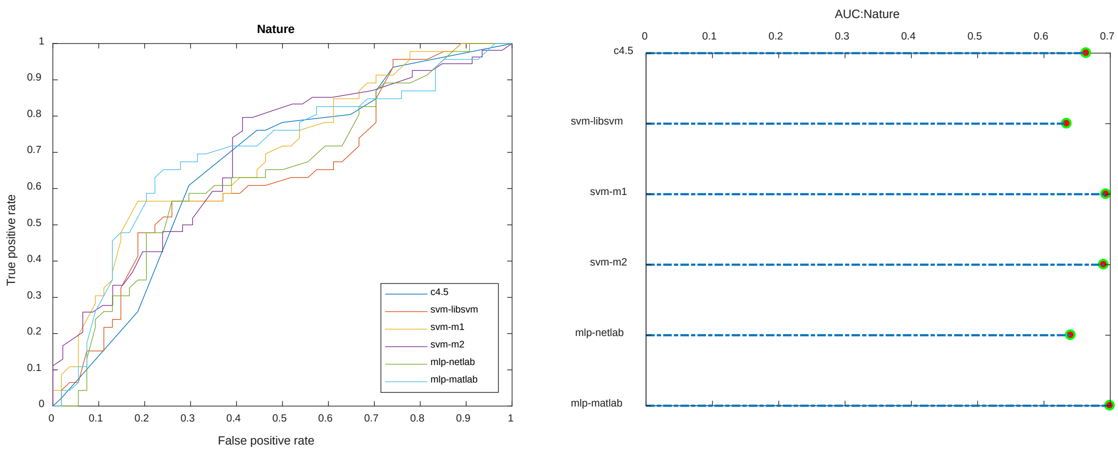


Figure 5.6 Comparative evaluation using ROC curves and AUC plots of DT, SVMs and MLPs for the Nature data set

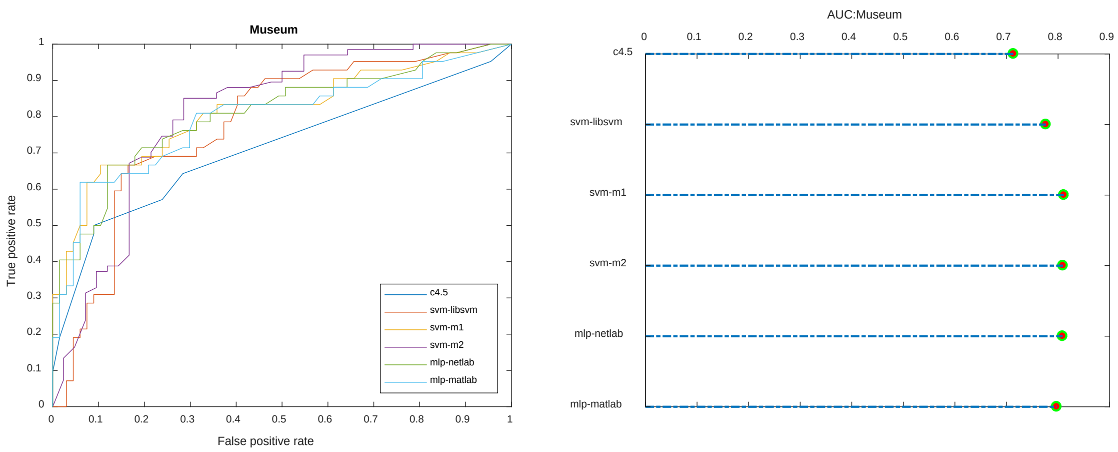


Figure 5.7 Comparative evaluation using ROC curves and AUC plots of DT, SVMs and MLPs for the Museum data set

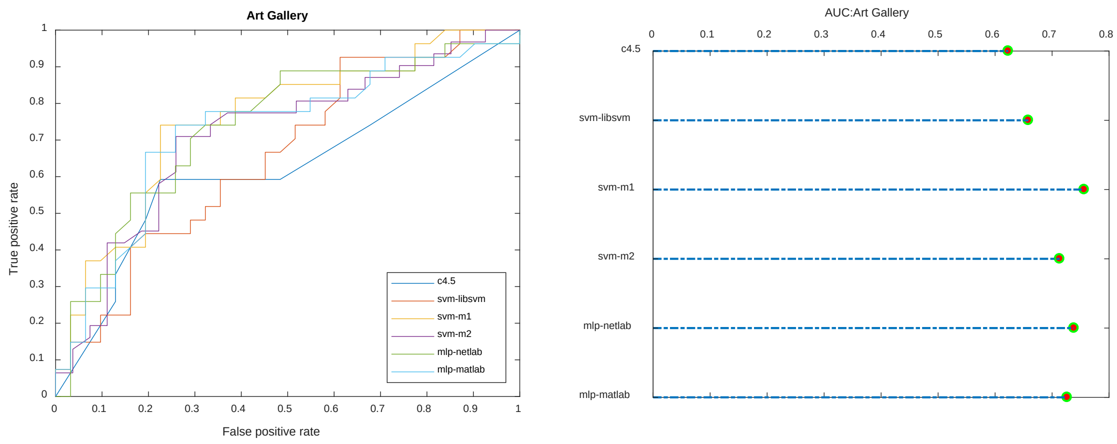


Figure 5.8 Comparative evaluation using a ROC curves and AUC plots of DT, SVMs and MLPs for the Art Gallery data set

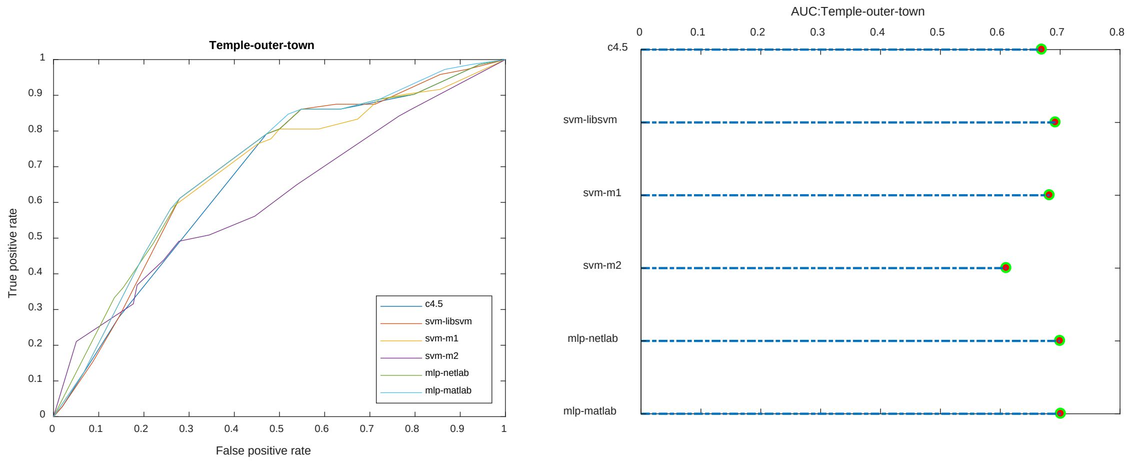


Figure 5.9 Comparative evaluation using a ROC curve and AUC plots of DT, SVMs and MLPs for the Temple-outer town data set

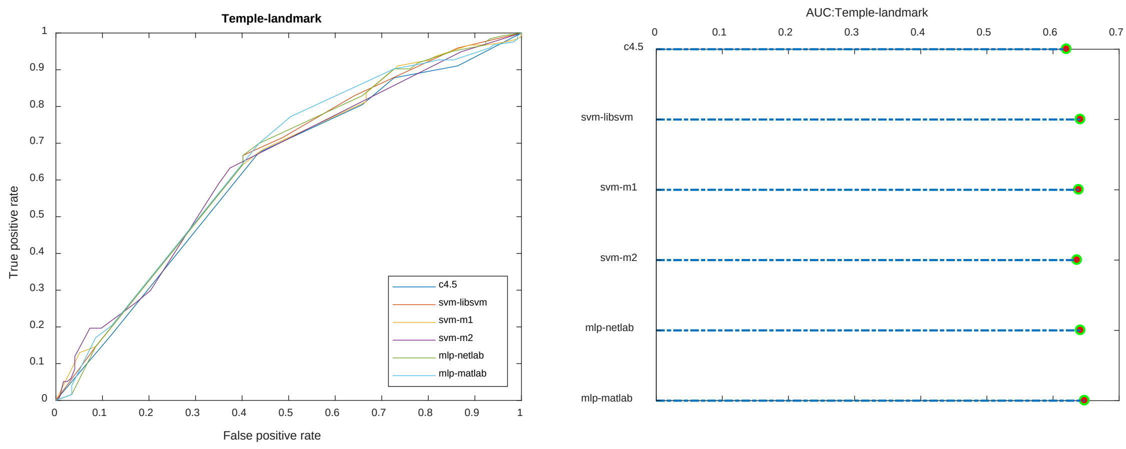


Figure 5.10 Comparative evaluation using a ROC curve and AUC plots of DT, SVMs and MLPs for the Temple-landmark data set

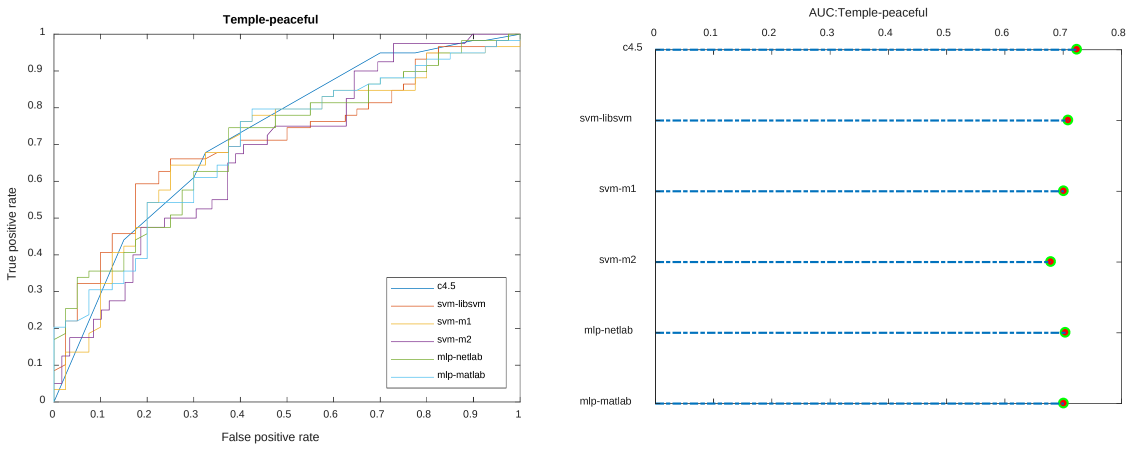


Figure 5.11 Comparative evaluation using a ROC curve and AUC plots of DT, SVMs and MLPs for the Temple-peaceful data set

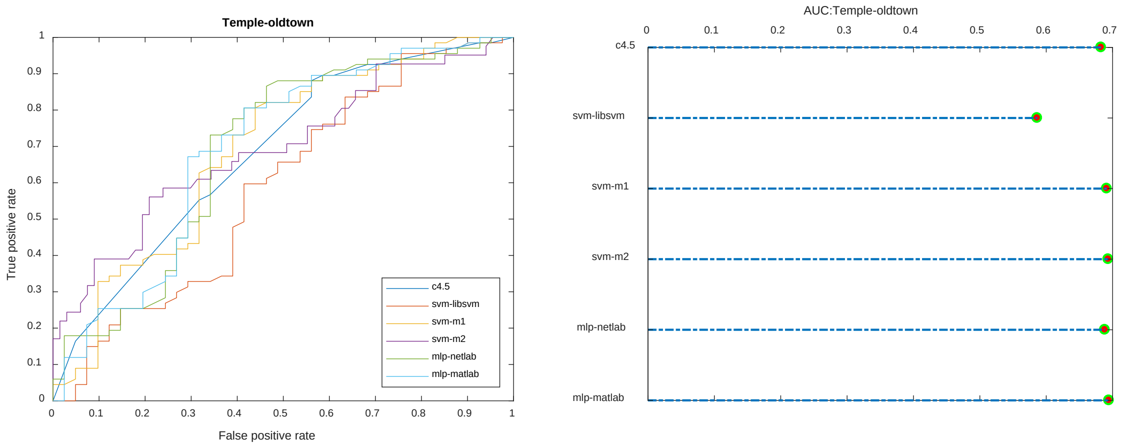


Figure 5.12 Comparative evaluation using a ROC curve and AUC plots of DT, SVMs and MLPs for the Temple-old town data set

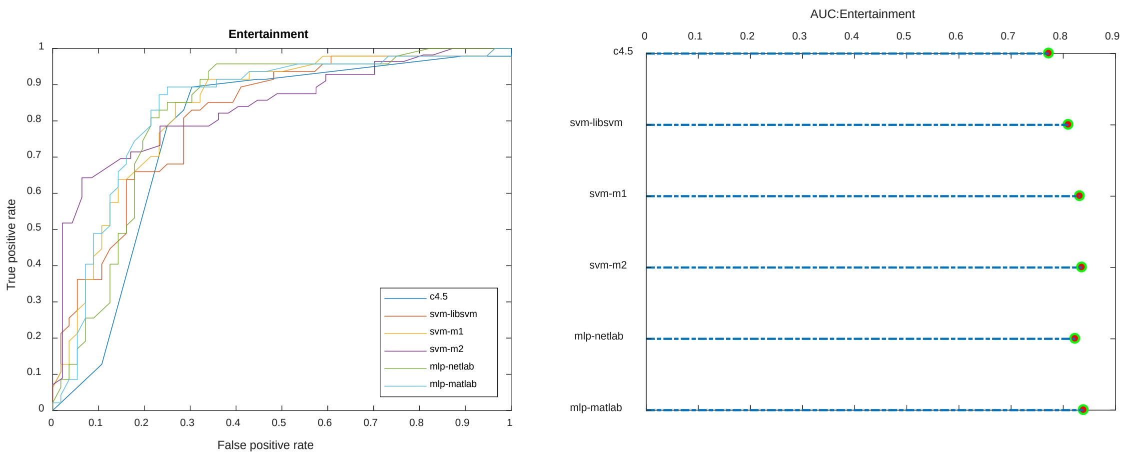


Figure 5.13 Comparative evaluation using a ROC curve and AUC plots of DT, SVMs and MLPs for the Entertainment data set

5.1.2 Concluding remarks

The experimental results show that MLP achieved the highest classification accuracy rate, followed by DT and SVM. According to the evaluation metrics, it can be seen that different classification algorithms are better at classifying different destination-choice data sets. Classification performance can be improved by using a combination of these classifiers and by ensemble of classifiers methods (Catal et al., 2015).

5.2 Ensemble of Classifiers using combination rules

5.1.3 Objectives of the study

In the previous study ordinary learning approaches were experiments in which several individual learners were used to classify data sets. In order to increase the prediction rate of the models an ensemble method can be used. The purpose of this study is to investigate and analyse the performance of several classification combination rules and to investigate the available ensemble learning methods, including combination rules and ensembles of classifier algorithms.

5.2.1 Experimental design and data set

In this study our experiment setup consisted of three steps involving selecting the most valuable individual classifier, choosing appropriate combination rules, and evaluating the classifier. We used six classifiers generated by DT, SVM, and MLP.

Since the functions of SVM_{m1} in original implementation of the classifier did not support the calculation of scores (i.e. distance from the hyper-plane) or posterior probabilities, we had to deploy a function to calculate the posterior probability for this SVM function by finding decision values using Platt’s scaling (Platt, 1999) and fit a score vector to a sigmoid function in order to find the probability distribution. To be able to transform classifier scores into accurate multiclass probability estimates, the one-vs-all method was used. Then, the least-squares method was used to normalize the probabilities to 1. The output of these classifiers can be combined by using three types of output: an abstract output (i.e. predicted labels), measurement-level output (i.e. posterior probability) and ranked-classes output. This study used the Chiang Mai destination data sets.

Combining rules

In this study we investigated seven combinations, including majority vote, maximum, minimum, average, summation, product and ranking. Given that $pred_i$ is the prediction label of vector x_i from classifier *j*, $\hat{P}_{ik}^{(j)}$ is the posterior probability that x_i belongs to class *k* and w_j is the weight of the classifier *j*. Therefore, the seven combination rules can be computed as:

1. Majority Vote (MV)

$$pred_i = Mode_J\{pred_i^{(j)}\} \tag{5.6}$$

The first rule is considered as hard voting, using the predicted label output from each classifier. Majority voting is associated with binary-class problems. In the case of multi-class problems, it is referred to as plurality voting. In this method we treat each of the classifiers as an expert. Regarding the simplest cast of majority voting, the decision of the final predicted label is based on the following rule:

The second to the sixth rules are referred to as soft voting that includes weight in the calculation. By default, weight w_j is set to 1, and posterior probability output is used to determine $pred_i$.

2. Maximum (MAX)

$$pred_i = \max_K \left\{ \frac{\max_J (\hat{P}_{ik}^{(j)}) w_j}{\sum_{k=1}^{K} \max(\hat{P}_{ik}^{(j)}) w_j} \right\} \tag{5.7}$$

3. Minimum (MIN)

$$pred_i = \max_K \left\{ \frac{\min_J (\hat{P}_{ik}^{(j)}) w_j}{\sum_{k=1}^{K} \min(\hat{P}_{ik}^{(j)}) w_j} \right\} \tag{5.8}$$

4. Average (AVG)

$$pred_i = \max_K \left\{ \frac{mean_J (\hat{P}_{ik}^{(j)}) w_j}{\sum_{k=1}^{K} mean(\hat{P}_{ik}^{(j)}) w_j} \right\} \tag{5.9}$$

5. Summation (SUM)

$$pred_i = \max \left\{ \frac{1}{J} \sum_{j=1}^{J} (\hat{P}_{ik}^{(j)}) w_j \right\} \tag{5.10}$$

6. Product (PRO)

$$pred_i = \max_K \left\{ \frac{1}{p(C_k)^{J-1}} \prod_{J=1}^{J} (\hat{P}_{ik}^{(j)}) w_j \right\} \tag{5.11}$$

The ranking combination rule uses ranked-class output. To determine the rank of the classes the posterior probabilities of the predicted classes were sorted in descending order. Hence, $pred_i$ can be computed from the sorted probability values.

7. Ranking (RANK)

$$pred_i = \max_K \sum_{j=1}^{J} \hat{P}_{ik}^{(j)} \tag{5.12}$$

5.1.4 Experimental results

It can be seen from Table 5.18 that the majority vote rule achieves the highest average accuracy. Both sum and average rules reveal the same performance of 67.28%. The ranking rule reveals the worst average result of 52.7%. When comparing the classification-accuracy averages of the data sets, only Majority vote outperformed the baseline classifier, by 1.76%.

Regarding statistical tests, a normality test using Shapiro-Wilk was applied first before using paired t-tests. Statistical results show that all the classifier combination rules are normally distributed (*p-value* > 0.05, so the null hypothesis is retained at a 95% level of significance). Next is a parametric test where a paired t-test is applied to compare with the baseline learner. The statistical results show that there was no significant improvement between the baseline learner (C4.5) from the previous experiment and the combination rules: *Majority vote* > *Product* > *Summation* > *Average* > *Minimum* >* *Maximum* >* *Ranking*, where >* indicates 'significantly better at a 95% confidence interval'. The results show that the average and ranking rules are significantly lower than the baseline learner. However, combination rules reveal a higher classification accuracy rate than the baseline learner for the Museum, Art Gallery, Temple-outer town, Temple-landmark, Temple-old town, and Entertainment data sets, especially for the Temple-outer town data set where combination rules show a 10.97% improvement in classification accuracy rate.

We can see from Table 5.20, when observing precision and recall scores, that MV achieved a slightly better score than other combination rules, except for the Museum, Art Gallery and Temple-landmark data sets where MAX, SUM, PRO, and RANK obtained better scores. When compared with the single best learner (MLP_m), from the previous experiment, MV has a better precision score than MLP_m when detecting classes D (Museum of World Insects), P (Wat Chiang Man) and Q (Chiang Mai Cabaret show). MV achieved slightly better recall than MLP_m for classes B (Huay Tung Tao Lake), F (Wattana Art Gallery), H (Wat Phra That Doi Kham), M (Wat Lok Malee) and R (Burklerk Gym-Muay Thai Training). With regard to f-score, MV obtained a higher score than other combination rules but was slightly lower than MLP_m in all classes except classes F (Wattana Art Gallery), O (Wat Sri Suphan), and Q (Chiang Mai Cabaret Show) where MV achieved better scores than MLP_m.

Table 5.18 Test classification accuracy rates for combination rules for each data set

Data set	MV	MAX	MIN	SUM	AVG	PRO	RANK	Baseline
Nature	**62.0**	57.0	57.0	60.0	60.0	57.0	54.0	64.0
Museum	76.15	78.90	78.90	**77.06**	**77.06**	**77.98**	61.47	75.23
Art-Gallery	72.41	**75.86**	75.86	72.41	72.41	72.41	53.45	68.97
T-outer-town	**61.93**	46.59	46.59	46.59	46.59	46.59	59.09	51.14
T-landmark	**63.33**	55.00	55.00	57.08	57.08	57.08	47.92	62.08
T-peaceful	63.64	64.65	64.65	**66.67**	**66.67**	**66.67**	40.40	68.69
T-oldtown	**70.37**	69.44	69.44	71.30	71.30	71.30	37.96	70.37
Entertainment	**81.55**	75.73	75.73	77.67	77.67	76.67	76.70	78.64
AVG.	**68.92**	65.39	65.39	66.09	66.09	65.71	53.87	67.39

Note: Bold font indicate the highest accuracy among the rules.

Table 5.19 Confusion matrix for Majority vote

	Predict																
Actual	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
B	***33***	*13*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
C	*25*	***29***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
D	*0*	*0*	***23***	*19*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
E	*0*	*0*	*7*	***60***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
F	*0*	*0*	*0*	*0*	***19***	*8*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
G	*0*	*0*	*0*	*0*	*8*	***23***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
H	*0*	*0*	*0*	*0*	*0*	*0*	***62***	*7*	*3*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
I	*0*	*0*	*0*	*0*	*0*	*0*	*32*	***18***	*7*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
J	*0*	*0*	*0*	*0*	*0*	*0*	*25*	*14*	***8***	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
K	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***86***	*37*	*0*	*0*	*0*	*0*	*0*	*0*
L	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*51*	***66***	*0*	*0*	*0*	*0*	*0*	*0*
M	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***51***	*8*	*0*	*0*	*0*	*0*
N	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*28*	***12***	*0*	*0*	*0*	*0*
O	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***62***	*5*	0	0
P	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*27*	***14***	0	0
Q	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***0***	***44***	*3*
R	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	***0***	*16*	***40***

Table 5.20 Precision, recall and F-score for Majority vote

Destination	Precision	Recall	F-score
B	0.5690	0.7174	0.6346
C	0.6905	0.5370	0.6042
D	0.7667	0.5476	0.6389
E	0.7595	0.8955	0.8219
F	0.7037	0.7037	0.7037
G	0.7419	0.7419	0.7419
H	0.5210	0.8611	0.6492
I	0.4615	0.3458	0.3750
J	0.4444	0.1702	0.2462
K	0.6277	0.6992	0.6615
L	0.6408	0.5641	0.6000
M	0.6456	0.8644	0.7391
N	0.6000	0.3000	0.4000
O	0.6966	0.9254	0.7949
P	0.7368	0.3415	0.4667
Q	0.7333	0.9362	0.9362
R	0.9302	0.7143	0.7143

5.2.2 Experiment 2*:* Discussion

In this study, experiments on seven classifier combination rules, majority vote, maximum, minimum, summation, average, product and ranking, were performed. The results show that majority vote (hard voting) is the most effective rule but not significantly different in terms of improving from the base line classifier (*p-value* = 0.366). The experimental results also show that the ranking rule is the worse combination rule. This is because converting posterior probability to ranked classes loses some information.

5.1.5 Concluding remarks

The performance of an ensemble of classifiers using combination rules was investigated. The combiners were compared with the baseline learner. The experiment results show that there is no statistical significance in terms of improvement in classification accuracy rates. However, Majority vote has a higher mean for the classification accuracy of data sets than the baseline learner. The reason why there is no significant improvement is due to the fact that each classifier may be superior to the others, as can be seen from the Temple-outer town data set, where the combiners in this study performed 10.97% better than the baseline learner. By adjusting the weight parameter in the soft voting rules, using various weight functions, the classification accuracy rate could be improved.

5.3 Ensemble of classifiers by weight and cascade

This study investigated bagging and boosting algorithms, specifically the Adaptive Boosting algorithm (AdaBoost) in the destination classification.

1. Bootstrap aggregation

   Bootstrap aggregation is also known as bagging. In the bagging method, diverse classifiers are generated only if the base learning algorithm is unstable, such as a DT algorithm (Breiman, 1996). Bagging uses *random sampling with replacement* (cases can be selected more than once for the sample, and they are not removed from the data set once selected) and they are used to sample the population for training. The rest of the samples that were not selected were allocated to the test set. To find the final answer to the classification problem, Majority voting or plurality voting algorithms were applied. The bagging algorithm below, was applied in this study as follows:

Algorithm 5.2: Bagging

1: **Input:** Dataset D= $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$;
2: Base learning algorithm €;3
3: Number of bags T.
4**: Process:**
5**:** **for** t = 1,..,T
6: ht = €(D, D_{bs}) % D_{bs} is the bootstrap distribution
7: **end**
8**: Output:** $H(x) = \frac{\arg\max}{y \in Y} = \sum_{t=1}^{T} \coprod (h_t(x) = y)$

2. Boosting

   The concept of the boosting method is to construct a strong learner from a set of weak ones. Boosting works by training a set of learners sequentially and then combining them for prediction. The later learners become stronger and focus more on the mistakes of the earlier ones. In the training stage, the initial weight of each training sample is assigned equally. For each boosting round, the model is trained using the training set, and the error is calculated. Then, the weight is updated using the alpha value. This process continues until the last classifier has been trained. The final model is calculated by using the weighted sum of the M classifiers. The weight of the incorrectly classified sample is increased. In this study we applied AdaBoost (Freund and Schapire, 1999), a well-known boosting algorithm. AdaBoostM1 supports multi-class problems by choosing the class that has the highest total vote. The algorithm that was applied to the Chiang Mai data set is described below:

Algorithm 5.3: ADABoost-m1

Input: Dataset D = $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$;

Base learning algorithm €;

A number of learning rounds T.

Process:

$D_1(x) = 1/m$ % Initialize the weight distribution

for t=1,..,T:

ht = €(D, D_t) % Train a classifier ht from D under distribution D_t

$$error_t = \sum_{i=1}^{n} w_i I(c_i \neq ht(x)) / \sum_{i=1}^{n} w_i$$

$$\text{Set} \quad \partial_t = \log\left(\frac{(1-error_t)}{error_t}\right)$$

$w_n = w_n \exp(\alpha_m I)$ %Updated weight distribution

end

Output: $H(x) = \arg\max_{y \in Y} \sum_{h(x)=y} \log \frac{1}{\partial_t}$

Table 5.21 Summary of cross-validation accuracy and test accuracy rates for the data sets (best values in bold)

Data set	Ensemble models											
	Bagged-C4.5		Bagged-SVM$_{libsvm}$		Bagged-SVM$_{m1}$		Bagged-SVM$_{m2}$		Bagged-MLP$_n$		Bagged-MLP$_m$	
	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test
Nature	66.32	67	59.87	62	60.38	59	59.35	65	59.16	67	60.39	**68(9)**
Museum	69.19	76.15	69.03	75.22	70.32	**79.82**	69.19	75.23	67.58	**79.82**	69.03	**79.82(1)**
Art-gallery	66.77	**74.14**	61.34	65.52	62.56	67.24	62.96	58.62	62.55	68.97	63.13	**74.14(14)**
T-outer town	45.21	52.84	45.81	48.30	62.48	**64.77**	59.58	48.30	46.01	51.14	45.31	52.27(8)

T-land mark	59.09	63.33	59.02	62.08	59.09	62.08	59.10	62.08	62.38	**70.71**	59.24	63.33(7)
T-peaceful	66.84	**70.71**	64.88	61.62	62.20	67.68	63.10	62.63	62.03	73.74	61.33	69.70(13)
T-old town	72.61	70.37	67.90	68.52	68.07	70.37	67.10	69.44	68.07	70.37	68.23	**73.15(14)**
Entertain	74.11	**80.58**	73.09	75.73	72.06	75.73	73.26	75.73	72.75	78.64	72.91	**80.58(71)**
AVG.	65.02	69.39	62.62	64.87	64.65	68.34	64.21	64.63	63.47	68.80	62.41	**70.12**

5.1.6 Experimental results

Regarding the bagging method, we applied the same experimental setup as that applied to individual learners. When observing the paired-sample test differences between the individual classifier and the ensemble classifier, we found that bagged-MLP$_m$ showed the statistical difference and outperformed the rest of the ensemble classifiers in all data sets. This was statistically significant as: bagged-MLP$_m$ >* bagged-C4.5 > bagged-MLP$_n$ > bagged-SVM$_{m1}$ > SVM$_{libsvm}$ > bagged-SVM$_{m2}$ (p-value < 0.05). According to Table 5.21, the gain in the classification accuracy rate reached up to 19.8% in the Temple-outer town data set and 6.89% in the Art Gallery data set. Interestingly, by observing the classification alone, bagged-SVM$_{m1}$ outperformed its single model, especially in the Temple-outer town data set, which involved the multi-class problem. Regarding improvements over the base learner (C45), all the bagging models outperformed the base learner and the statistical results showed that bagged-C45 outperformed the rest of the bagging classifiers. This was statistically significant as: bagged-C4.5 >* bagged-MLP$_m$ >* bagged-SVM$_{libsvm}$ >* bagged-MLP$_n$ >* bagged-SVM$_{m2}$ > bagged-SVM$_{m1}$ (p-value < 0.05). The bagged-SVM$_{m1}$ was the only classifier that did not improve in a statistically significant way compared to the base learner.

In determining whether or not there were any statistically significant differences between the boosting method and the individual models, the results showed no significance differences between them (p-value <0.05). Regarding the performance over the base learner, Boost-MLP$_m$ was the only classifier whose performance was significantly better than the baseline learner, while the rest of the classifiers under performed.

Table 5.22 Summary of cross-validation and test accuracy rates for the data sets (best values in bold)

Ensemble models												
Data set	Boost-C4.5		Boost-SVM$_{libsvm}$		Boost-SVM$_{m1}$		Boost-SVM$_{m2}$		Boost-MLP$_n$		Boost-MLP$_m$	
	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test
Nature	59.87	64	60.21	63	59.35	61	60.05	65	59.00	68	59.34	**68(5)**
Museum	68.87	76.15	69.19	74.32	70.48	78.89	68.87	70.64	65.97	77.06	68.23	77.98(5)
Art-gallery	65.26	63.79	61.65	72.41	61.66	67.24	61.95	60.34	63.44	67.24	62.24	74.14(8)
T-outer town	46.11	50	45.11	47.16	45.31	47.72	45.71	50.57	45.40	51.14	45.01	**51.14(4)**
T-land mark	58.94	62.08	59.09	62.08	58.87	62.08	59.02	62.08	59.17	63.33	59.09	63.33(3)
T-peaceful	65.42	66.67	63.45	60.61	62.38	65.65	62.74	63.64	62.38	70.70	61.32	**68.69(5)**
T-old town	69.53	67.59	66.61	71.29	66.60	68.59	66.44	70.37	67.09	69.44	66.29	**70.37(6)**
Entertain	73.42	78.64	72.75	73.79	72.23	75.73	73.43	74.76	71.38	80.58	72.76	**79.61(14)**
AVG.	63.43	66.12	62.26	65.58	62.11	65.86	62.28	64.68	61.78	68.20	61.79	**69.15**

5.3.1 Experiment 3: Discussion

In this study, bagging and boosting algorithms were compared. The main conclusion of this study is that bagging performed significantly better than boosting and the base learner. In determining whether or not there were any statistically significant differences between the boosting method and the individual models, the results showed that there were no significance differences between the boosting and individual models (p-value < 0.05). Regarding the performance over the base learner, Boost-MLP$_m$ was the only classifier that showed a statistically significant improvement over the base learner, while the rest of the classifiers under performed. The experimental results revealed that the bagging method achieved a better classification accuracy rate than the other ensemble learner methods, including Majority vote (see Section 5.2) and boosting. The reason why the bagging method showed a significant improvement in C4.5 and both the MLP classifiers was because that the DT and neuron network are known to be sensitive to perturbation on the training samples, and they are also known to be unstable learners. C4.5 DT is known for the ease with which randomness can be injected.

Table 5.23 Precision, recall and F-score of Bagged-C4.5

Destination	Precision	Recall	F-score
B	0.644	0.630	0.637
C	0.691	0.704	0.704
D	0.786	0.524	0.629
E	0.753	0.910	0.824
F	0.773	0.630	0.694
G	0.722	0.839	0.776
H	0.530	0.847	0.652
I	0.682	0.263	0.380
J	0.436	0.362	0.395
K	0.628	0.699	0.662
L	0.641	0.561	0.600
M	0.688	0.932	0.791
N	0.789	0.375	0.508
O	0.692	0.940	0.797
P	0.765	0.317	0.448
Q	0.737	0.894	0.808
R	0.891	0.732	0.804

5.3.2 Concluding remarks

This chapter demonstrates the power of ensemble learning in predicting preferred tourist destinations to the traveller, which most researchers still consider to be an art form and, therefore, cannot be forecast to reveal an acceptable result. Selecting the right classifier for DRS is not an easy task and is data-dependent. Therefore, we have investigated ensemble learning approaches, starting with the simplest one. The efficacy of three classification algorithms, DT C4.5, SVMs and MLPs has been investigated and compared to each other with respect to the differences revealed in the data sets.

The classification algorithms were evaluated with proper scientific methods including classification accuracy rate, confusion matrix, precision, recall and F-measure score. This study applied three ensemble methods to construct predictive models, including combination rules, bagging, and boosting. For future enhancements of the system in terms of improving recommendation accuracy and reducing redundant features, we suggest employing ensemble learning methods such as stacking, random forest, random sub-spaces or pasting.

Other traditional classifier algorithms, such as RTree and REPTree, could be used as base learners, along with the intention to construct deep-learning neural networks for each of the destination-choice models.

Chapter 6 Model-Based User Interface for DRS

In the field of TRS, information presentation plays a major role in service and is an attractive application for the end-user. To efficiently design and develop a user interface for the proposed DRS, we proposed an Adaptive, Responsive, Interactive Model-based User Interface (ARIM-UI) framework for the DRS as it handles decision model-to-user interface complexity, which is one of the greatest challenges when designing a semantic web. Our ARIM-UI can automatically convert constructed decision models from the C4.5 algorithm into a user interface, as well as support ease of usage through heterogeneous interfaces. By combining JavaScript library based on a MVVM design pattern, two popular web frameworks, Google Maps API and two language parsers, the proposed ARIM-UI can provide three main functionalities: rich responsive display, interaction and adaption. Moreover, ARIM-UI supports back-end login, which lets experts directly modify knowledge. This chapter addresses the last research objective:

RQ 7. How can a tourist be helped tointerpret and interact with the constructed decision model(s)?

Buhalis and Law (2008) claim that Web design is one of the most significant technological innovations for the tourism industry; and besides user interaction, accessibility features for disabled and elderly people should be more responsive.

Previous DRSs have improved the interaction between user and system. DRSs started with static and unfriendly user interfaces on their websites and have improved to provide more dynamically advanced, informative ones. PHP, MySQL and AJAX technologies combine several technologies such as HTML, JavaScript and XML and have been applied to create a sense of interaction between the user and the web application interface (Chiang and Huang, 2015). This has helped to improve the user experience and increase the level of satisfaction and enjoyment during searches for tourism services. For example, Web technologies such as jQuery and JavaScript were used in the user-interface development in order to provide a dynamic-drag interface design (Chiang and Huang, 2015). Hsu et al. (2012) produced an interactive GUI using Google map API to allow the user to adjust geographic data according to personal needs. SigTur/E-destination (Moreno et al., 2013) applied several open-source Web technologies comprising Java Server Faces (JSF), AJAX and integrated Google Map API, to generate a sense of interaction between tourist and system.

One of the biggest challenges when designing and developing a successful Web user interface is to make complex functionality available to the user in an easy way (Khalili and Auer, 2013). From 2011 to 2018, global mobile data traffic increased 11-fold (Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper – Cisco, 2017). This led to another challenge, namely, how to make the user interface of DRS more responsive and interactive, while supporting different devices. In this chapter we focus on afore-mentioned challenges. The objectives below correspond to the Research Question 7.

6.1 Objectives of the study

1. To provide adaptive capability such that when the decision model is changed, the interface and related information will automatically change
2. To provide a simple and proper connection between the UI and data models
3. To provide interactive and responsive capabilities
4. To provide geographic capability.

6.2 Methodology and User Interface System framework

In this chapter, we propose an adaptive model-based user interface that also provides a sense of interaction and response as a front-end to support the proposed DRS. By integrating the MVVM design pattern, Bootstrap style web framework, KnockoutJS framework, Google Maps API, and two languages parsers, our proposed ARIM-UI will have the following five features:

- Automatically update the correct parts of the UI (i.e. drop-down menu, radio-check boxes, and dynamic information, such as question and answer choices), whenever the data model changes or when the user selects or interacts with the UI, by using KnockoutJS (KO) Javascript libraries with the MVVM design pattern. The MVVM pattern provides a clear separation of concerns between the user-interface controls and their logic.
- Provide a responsive design front-end to the user where the layout of web-pages adjusts dynamically, by detecting the user's screen size and orientation, and changes the layout accordingly. This includes desktop, tablet, and mobile platforms. It supports all mainstream browsers, such as Internet Explorer, Firefox, Google Chrome, Opera and Safari. This is done by using Bootstrap, the world's most popular HTML, CSS and JavaScript framework. Users can choose the start point and the system will automatically arrange the route to the destination.
- Provide an intuitive and interactive user interface using Google Maps API, Google Directions API and Google Geocoding API. The proposed UI is connected to three different APIs, thereby allowing it to show points of interest, select modes of transport, provide a route from the current location to the destination, and predict travel time and current and future traffic. By using Google Map Geocoding, the user can type the address; then, the geocoding will return the latitude and longitude, and this will be used to place markers on the map.
- Provide a simple connection between the UI and the data model by using two language parsers: DecisionTree2XML and XML2Jason.
- Provide a model management system via a back-end for an administrator or other power user. They will be required to log in with a username and password, and can then add, edit, delete and upload new model files.

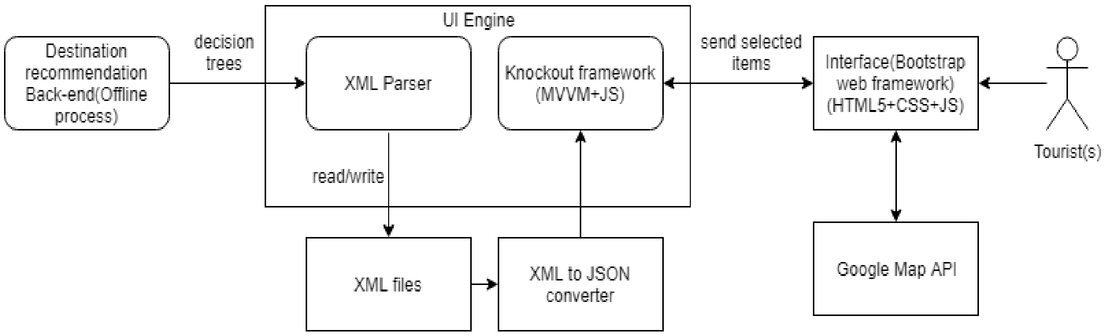


Figure 6.1 Our proposed UI framework to support the DRS

Figure 6.1 illustrates the proposed adaptive user interface. First, we provided automatic updates for the correct parts of the user interface (e.g. drop-down menu, radio-check boxes and dynamic information, such as question-and-answer choices), whenever the data model changes or the user selects or interacts with the UI. This was achieved by using KO Javascript libraries with the MVVM design pattern. Second, we implemented a straightforward and proper connection between the UI and the data model by using two language parsers: DecisionTree2XML and XML2JSON. Third, we provided an interactive and responsive front-end to the user for desktop, tablet and mobile platforms with the navigating system on Google maps, showing, for instance, points of interest and a route from the current location to the next destination. This was done be using Bootstrap, the world's most popular HTML, CSS and JavaScript framework. It supports all mainstream browsers, such as Chrome, Firefox, Internet Explorer and Safari. Finally, our UI provides spatial functions capability by integrating with Google Map service API; the system can plot the current location and the route to a destination, as well as information regarding how to get there.

Both SVM and MLP are black-box methods, which make them difficult to interpret. Therefore, to be able to develop an adaptive user interface, a way to rank input features is needed. Our proposed DRS generated several decision models from the C4.5 algorithm and these models were converted to decision rules, as shown in Figure 5(a). First, these decision rules needed to be converted to a specific format, such as XML, by using the XML parser program, for use across the Internet in an understandable form of data structure, work in conjunction with Web services and connect with the API. Second, XML files were converted to Jason objects using JQuery language in order to connect with an observable variable which was constructed from the KnockoutJS framework (see Fig. 5(a)). Third, the KO used observable variables to connect with the Bootstrap-style view model. Last, when a user answers a question or a new question is created, the KnockoutJS will automate a new HTML page and re-process by selecting the correct data from JSON objects.

6.3 Technologies involved

To develop the proposed UI, the best Web programming languages and most advanced Web technologies framework were selected, i.e. XML, CSS, JSON, KO and Bootstrap. We also implemented two language parsers/ converters, DecisionTree2XML and XML2JSON, in order to make the data/information flow properly within the proposed system.

1. eXtensible Markup Language

   eXtensible Markup Language (XML) is a markup language developed by W3C for organising and tagging the elements of a document so that the document can be transmitted and interpreted by applications and organisations in the same protocol. It is designed to be both human- and machine-readable.

2. Cascading Style Sheets (CSS)

   Cascading Style Sheets (CCS) is a mark-up language maintained by W3C and originally designed to separate the content and presentation of HTML documents. A Web developer can easily add to, edit and delete styles from Web documents without having to go through each document. CSS provides several advantages in terms of bandwidth reduction, consistency and browser compatibility. This enables the website to look better and load faster. Figure 6.2 shows an example of CSS implemented in the proposed DRS user interface.

```
/* Global
------------------------------------
html, body {
    height: 100%;
    padding: 0;
    margin: 0;
}

body {
  display: table;
  width: 100%;
  height: 100%;
  min-height: 100%;
}

.header, .footer, .main-content {
    display: table-row;
}

.header{
    height:520px;
}

```

Figure 6.2 Example of CSS for the proposed DRS UI

3. Decision tree to XML parser

We used the C4.5 algorithm from Weka software to generate decision models; the output of Weka software is either in a text-tree format (see Fig. 6.3) or as a graphical model. Therefore, the goal of the XML parser is to convert the output from Weka DT J48 or similar, such as C4.5 algorithm text syntax (Quinlan, 1993), to XML format as shown in Figure 6.4. This provides two benefits, namely, storing model data in a proper data structure schema and the possibility to create other new data types. Second, XML can be used with other Web services. Figure 6.4 presents an XML file from the Nature data set. The C4.5 model in XML defines all the tree nodes and features used in the model.

```
e1_6_4 = 1: 20 (144.0/44.0)
e1_6_4 = 2
|   b3_4 = 0: 19 (80.0/32.0)
|   b3_4 = 1: 20 (63.0/26.0)
e1_6_4 = 3
|   e1_1_3 = 1: 20 (29.0/9.0)
|   e1_1_3 = 2: 19 (97.0/41.0)
|   e1_1_3 = 3
|   |   d10_1 = 0: 19 (108.0/44.0)
|   |   d10_1 = 1: 20 (52.0/20.0)
```

Figure 6.3 Example of decision-tree output from C4.5 algorithm

```xml
<?xml version="1.0" encoding="UTF-8"?>
<DecisionTree type="nature">
    <Test attribute="e1_6_4" operator="=" value="1">
        <Output decision="20" info="(152.0/47.0)"/>
    </Test>
    <Test attribute="e1_6_4" operator="=" value="2">
        <Test attribute="d10_1" operator="=" value="0">
            <Output decision="19" info="(81.0/35.0)"/>
        </Test>
        <Test attribute="d10_1" operator="=" value="1">
            <Test attribute="a1" operator="=" value="1">
                <Output decision="19" info="(2.0/1.0)"/>
            </Test>
            <Test attribute="a1" operator="=" value="2">
                <Output decision="20" info="(21.0/10.0)"/>
            </Test>
            <Test attribute="a1" operator="=" value="3">
                <Output decision="20" info="(22.0/8.0)"/>
            </Test>
            <Test attribute="a1" operator="=" value="4">
                <Output decision="19" info="(9.0/2.0)"/>
            </Test>
            <Test attribute="a1" operator="=" value="5">
                <Output decision="19" info="(2.0/1.0)"/>
            </Test>
            <Test attribute="a1" operator="=" value="6">
                <Output decision="20" info="(3.0/1.0)"/>
            </Test>
        </Test>
    </Test>
    <Test attribute="e1_6_4" operator="=" value="3">
        <Test attribute="e1_1_3" operator="=" value="1">
            <Output decision="20" info="(34.0/11.0)"/>
        </Test>
        <Test attribute="e1_1_3" operator="=" value="2">
            <Output decision="19" info="(96.0/40.0)"/>
        </Test>
        <Test attribute="e1_1_3" operator="=" value="3">
            <Test attribute="d10_1" operator="=" value="0">
                <Output decision="19" info="(104.0/44.0)"/>
            </Test>
            <Test attribute="d10_1" operator="=" value="1">
                <Output decision="20" info="(47.0/19.0)"/>
            </Test>
        </Test>
    </Test>
</DecisionTree>
```

Figure 6.4 The Nature model in XML format

4. JavaScript Object Notation (JSON)

JSON is the most modern, lightweight, and simple syntax and data-exchange format and replaces XML (which is commonly used by AJAX technology). The goal of this data format is to be able to transfer between a Web browser and a Web server. For example, the Bootstrap framework offers JSON API (i.e. it needs JSON as an input). JSON can easily be converted back to the original XML (converting Between XML and JSON, 2006.). Figure 6.5 shows the output of the Nature data set after it being converted from XML to JSON format.

.


```
{"DecisionTree":{"@type":"nature",
"Test":[{"@attribute":"e1_6_4","@operator":"=","@value":"1","Output":{"@decision":"20","@info":"(144.0/44.0)"}},{"@attribute":
"e1_6_4","@operator":"=","@value":"2",
"Test":[{"@attribute":"b3_4","@operator":"=","@value":"0","Output":{"@decision":"19","@info":"(80.0/32.0)"}},
{"@attribute":"b3_4","@operator":"=","@value":"1","Output":{"@decision":"20","@info":"(63.0/26.0)"}}]},
{"@attribute":"e1_6_4","@operator":"=","@value":"3",
"Test":[{"@attribute":"e1_1_3","@operator":"=","@value":"1","Output":{"@decision":"20","@info":"(29.0/9.0)"}},
{"@attribute":"e1_1_3","@operator":"=","@value":"2","Output":{"@decision":"19","@info":"(97.0/41.0)"}},
{"@attribute":"e1_1_3","@operator":"=","@value":"3",
"Test":[{"@attribute":"d10_1","@operator":"=","@value":"0","Output":{"@decision":"19","@info":"(108.0/44.0)"}},
{"@attribute":"d10_1","@operator":"=","@value":"1","Output":{"@decision":"20","@info":"(52.0/20.0)"}}]}]}]}}
```


Figure 6.5 JSON format of the Nature data set

5. KnockoutJS

After the JSON files were created they were passed to the Knockout (KO) framework (http://knockoutjs.com). This open-source Web framework helps to create rich and responsive UI interaction. One of the key concepts of this framework is that it provides a response to any data source change, e.g. automatic user-interface refresh by using JavaScript based on the MVVM (Gamma, 1995) design pattern, as shown in Figure 6.6. In MVVM, the data from HTML are connected with the ViewModel module, known as declarative biddings, so the web page can be generated in a dynamic way depending on the actions of the user. Two more advantages of KnockoutJS are dependency tracking and templating. KnockoutJS version 3.4.2 (*knockout*, 2017) was used in the study.

6. MVVM design pattern

In this study we used a software design pattern which offers existing solution to a common problem. The MVVM design pattern (Anderson, 2012), a modern variant of MVC, was selected for this study as it provides a clean separation of concerns between user-interface controls and their logic. It was designed to make use of the data-binding functions in Windows Presentation Foundation (WPF).

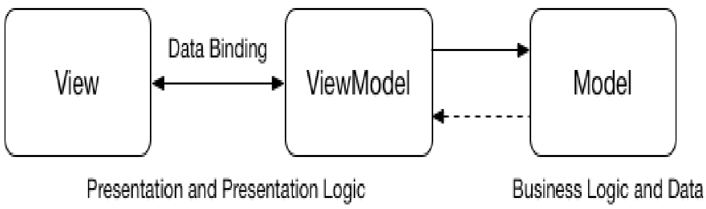


Figure 6.6 MVVM design pattern framework

7. Bootstrap Web framework

   To provide a responsive front-end for the proposed DRS, we selected the Bootstrap framework (http://www.getbootstrap.com), the most popular responsive Web framework. Its open-source code consists of three main technologies comprising of HTML, CSS, and JavaScript front-end framework. This framework is very effective on web browsers, tablets, and mobile phones. Bootstrap provides a responsive Web interface.

8. Spatial Web service

   Regarding interactive geography, our UI provides a spatial functions capability by integrating the system with Google Maps service API. This system can plot the current location and the route to destination, as well as provide essential information on how to get there. In this project we connect to several Google APIs, such as GMap and GLargeMap, to be able to load and control the maps. Additionally, the system uses Google directions API and geocoding in order to retrieve multi-part directions for a series of waypoints including transport mode, travel time, and current and future traffic status.

6.4 Internal work flow and UML diagrams

Figure 6.7 illustrates the workflow of the UI system from beginning to end. The first process begins with the input decision rule(s) from the Weka C4.5 algorithm in text syntax. This then needs to be converted to XML to be stored as the database in the server. To be able to connect to KnockoutJS JavaScript library we needed to parse our XML to JSON object data format, so we wrote a script to convert XML files to JSON. Next, the JSON file is loaded to an array data structure and bound to UI.

The purpose of the sequence diagram is to demonstrate the interaction between the objects (GUI Interface, KO object, and XML2Json) in a sequential order. In other words, Figure 6.7 below shows how our system would behave during the design phase.

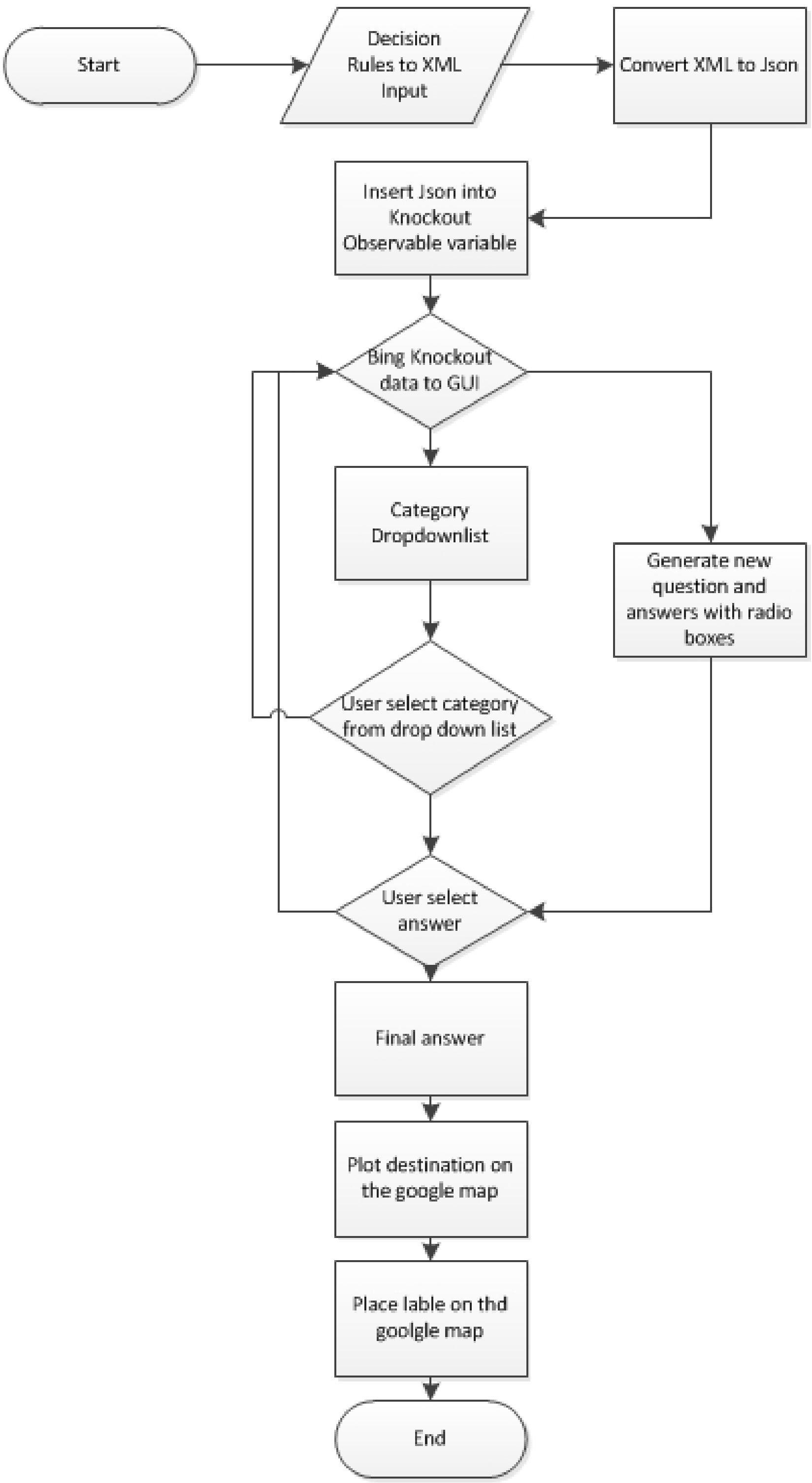


Figure 6.7 Data flow diagram for implementation of the UI

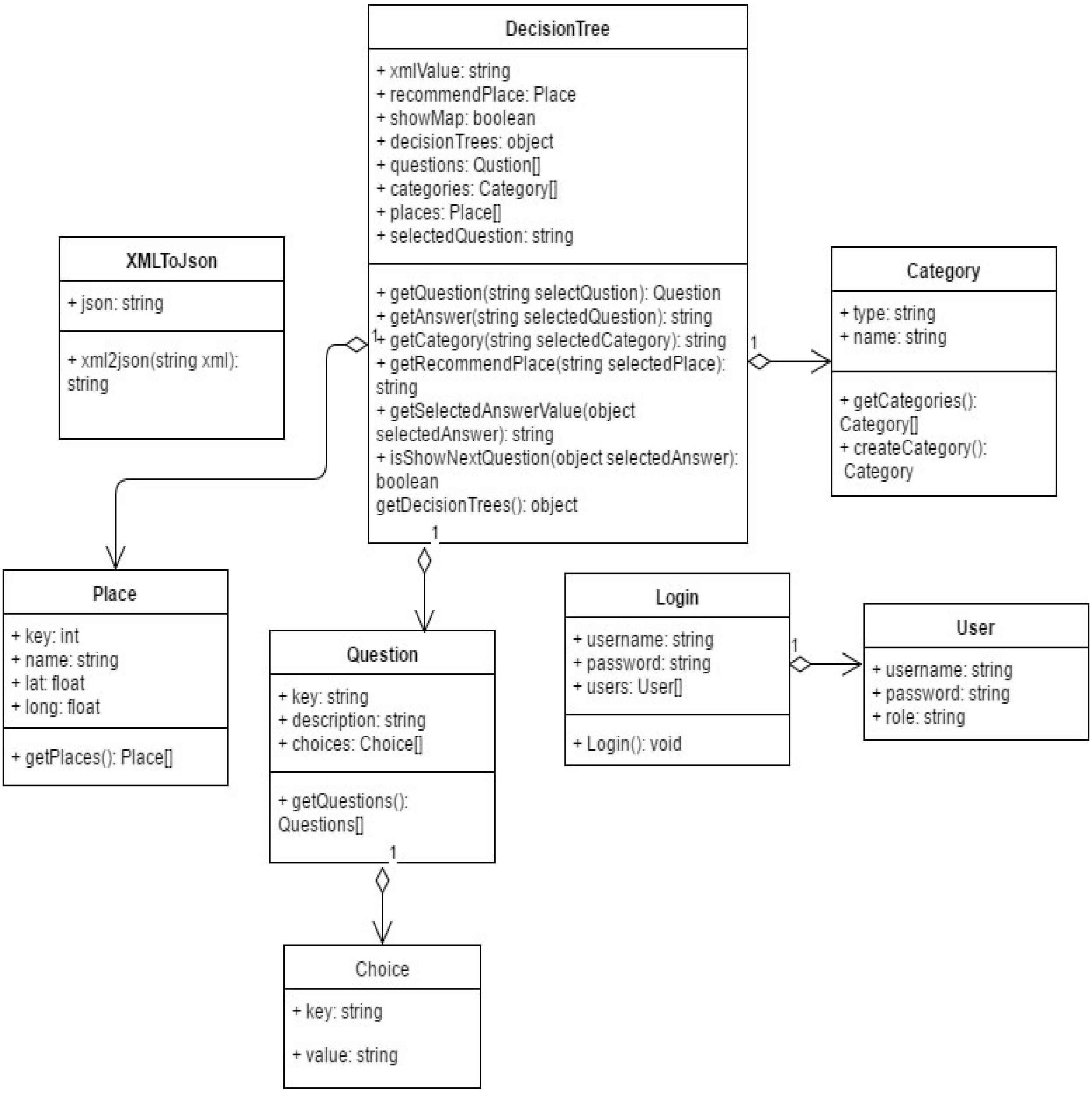


Figure 6.8 Class diagram for the back-end of the UI engine

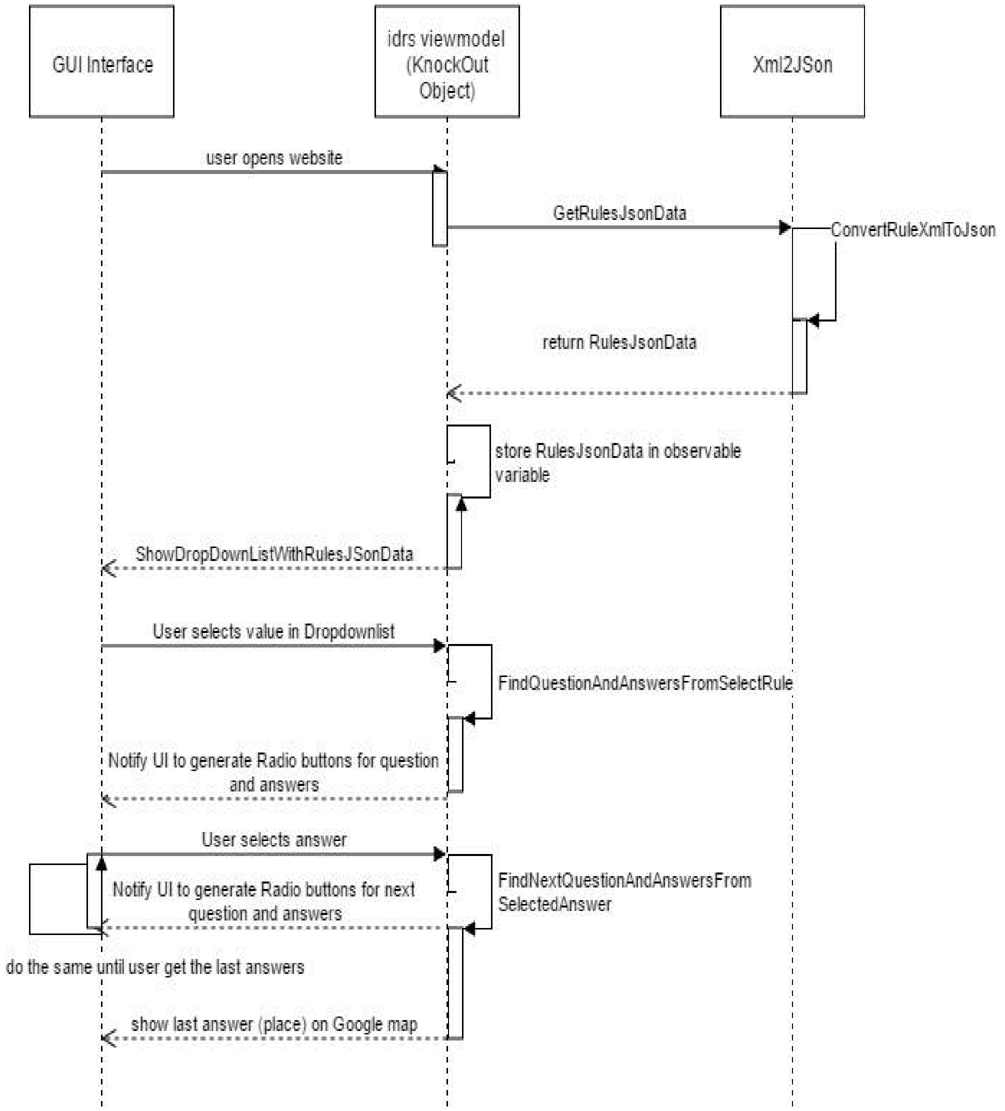


Figure 6.9 ARIM-UI sequence diagram

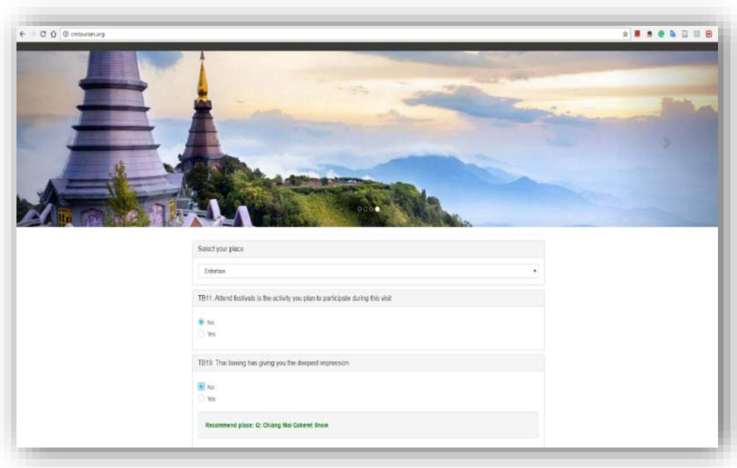
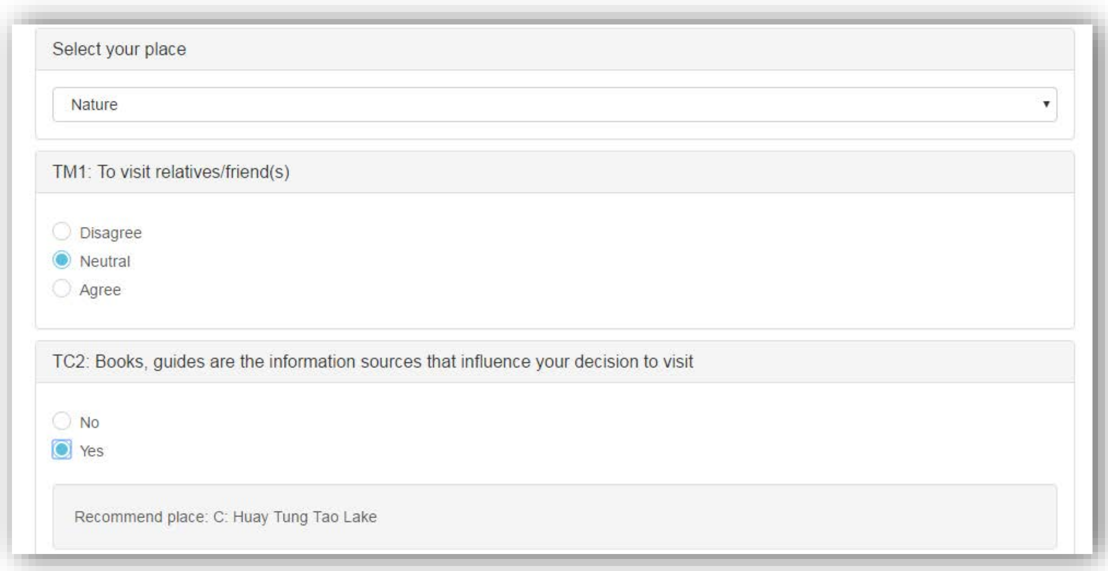
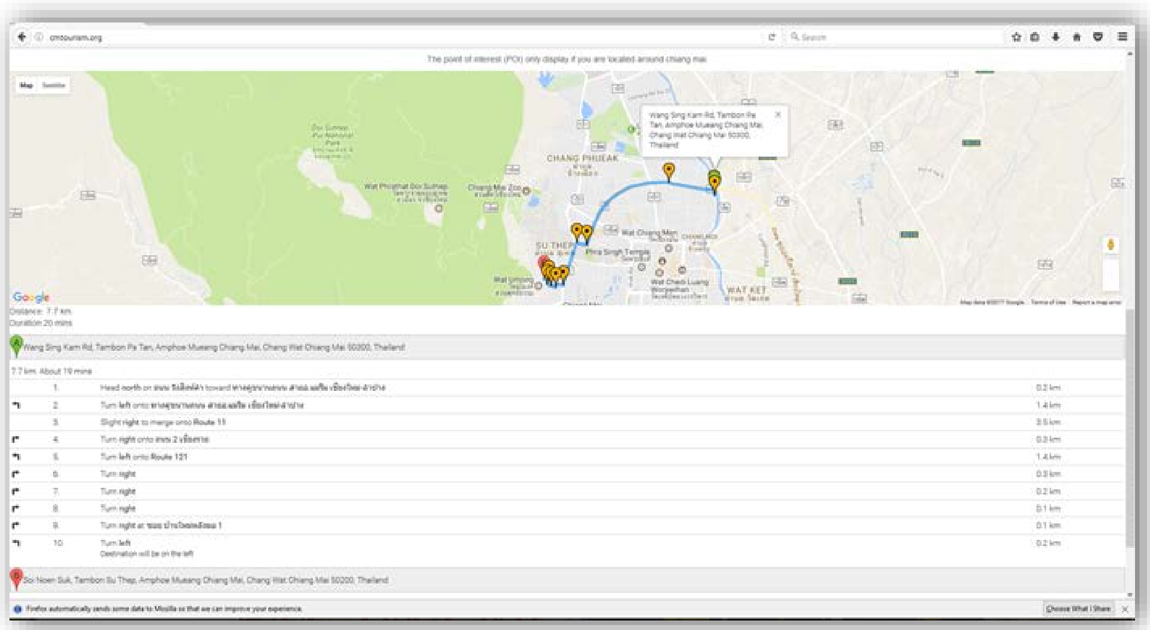


Figure 6.10 SAUI-DRS



(a)



(b)

Figure 6.11 UI for the DRS (a) Route from current user location to recommended destination (b) with detailed travel information

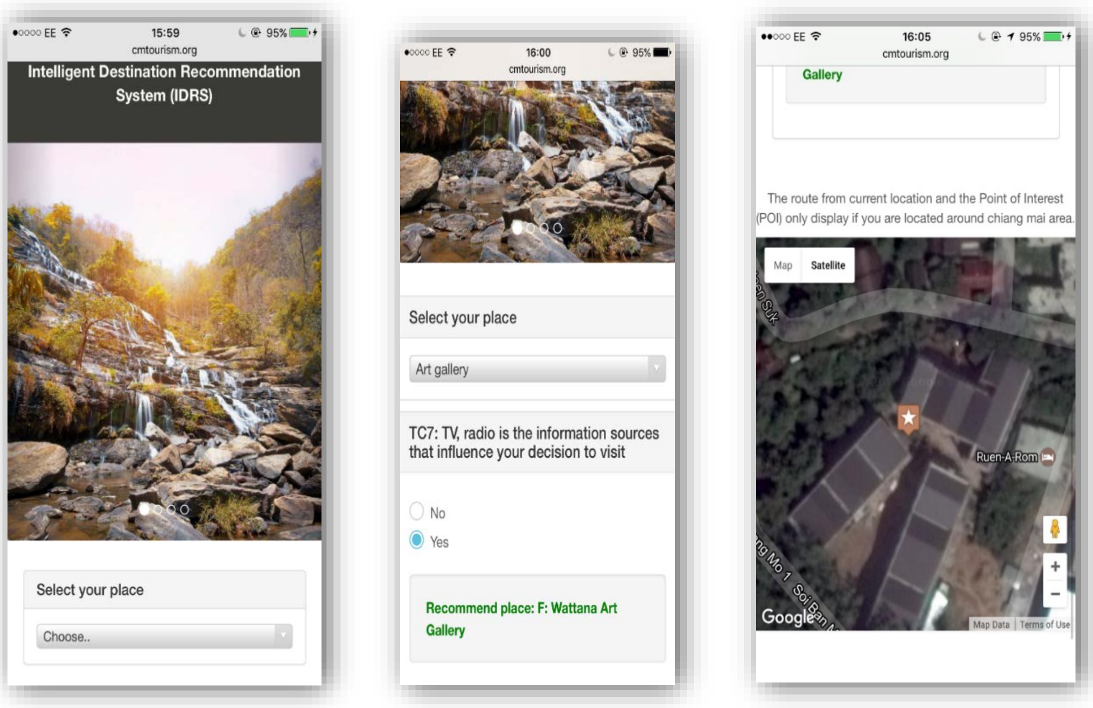
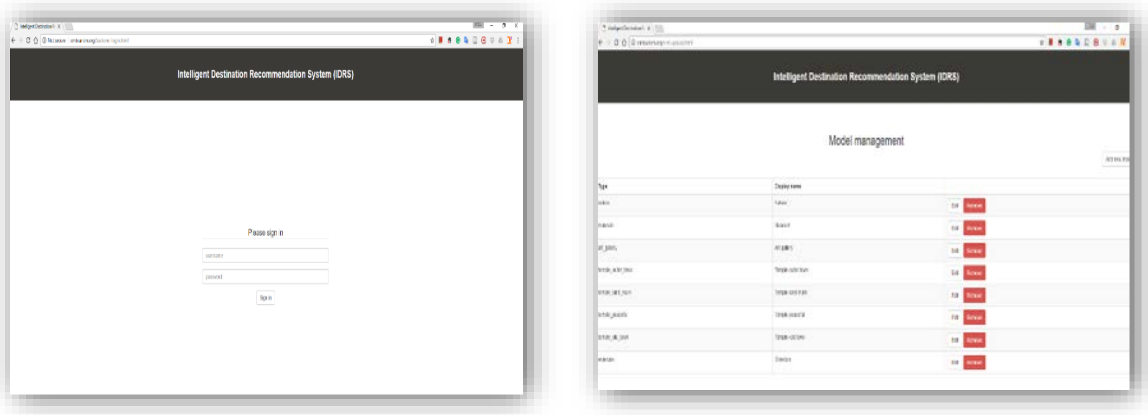


Figure 6.12 UI displays on a mobile device

6.5 Management System of the ARIM-UI

ARIM-UI supports back-end login (see Figure 6.13(a)), which acts as an administration control panel for superusers. A superuser is required to enter their username and password in order to add information to, edit or delete the exisiting model or change other information related to the website and destination choices (see Figs 6.13 and 6.14). Global.js is responsible for storing all the static information such as login information; POI information comprises ID, name, longitude and latitude, and question and answer choices.



(a) (b)

Figure 6.13 Administration control panel login (a) and model management panel (b)

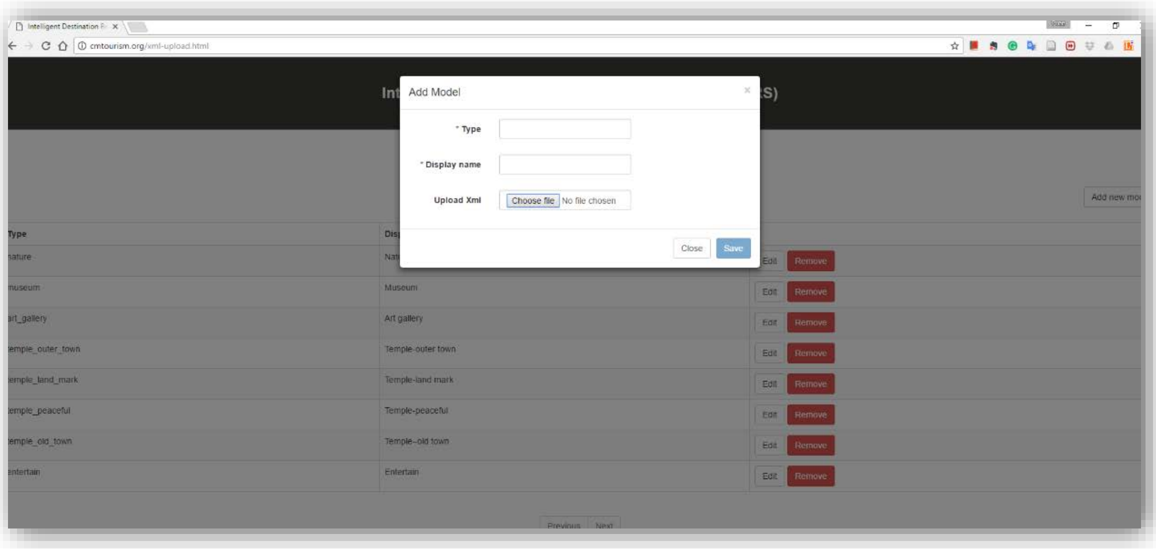


Figure 6.14 XML upload panel

6.6 Discussion

The user interface evaluation such as usability testing and congnitive walkthrough involves time-consuming and expensive processes (Jeffries et al., 1991). Due to the limited time of this project therefore, the evaluation of the interface has not been evaluated. Our web application can be accessed from various computing platforms (i.e. web browsers, tablets, mobile phones).

The SAUI displays beautifully and offers adaptive, interactive and responsive functions to the user. The user can begin by navigating from the drop-down menu (i.e. selecting from eight destination types). Next, the user will need to answer questions based on nodes and leaves from the DT C4.5 model. The last question (last node) will provide the user with the recommended destination name. Moreover, the destination will be plotted on Google Map with a route from the current location to the recommended destination (see Fig. 7(b)). The proposed UI was developed and deployed on a Linux operating system running Intel® Xeon® CPU E5-2630 v3@2.4GHz.

6.7 Concluding Remarks

In this chapter, we have discussed the design and implementation of ARIM-UI to support the proposed DRS. The proposed UI provides three main functionalities, being: responsive, interactive and adaptive. First, this study proposes an adaptive and responsive UI by using an MVVM design pattern, enabling it to create a rich, responsive user interface with a clean underlying data model. For instance, every time sections of the UI change, either from the user's actions or from new data or source changes, our UI automatically updates the correct parts. Second, to make a website that was even more responsive, and one that can dynamically adjust to proper screen resolution on any device, a front-end Web framework comprising HTML, CSS and JavaScript was applied. Third, a Google Maps API was embedded into the website, which makes the interface more interactive for the user. Fourth, a proper conversion from decision models and UI was done by using two language parsers, involving converting from the model to XML and from XML to JSON. Last, an administration control panel was implemented to let superusers modify and maintain data and models on the fly.

Chapter 7 Conclusion and Future Works

As a result of the rapid growth in the numbers of tourists who are travelling, the Internet is becoming increasingly populated with travel information. When selecting their preferred destinations before or during their travel to an unfamiliar city, tourists can therefore easily be overwhelmed. Destination recommendation systems (DRSs) are recognised as a valuable decision-support tool for online travel as well as for tourism marketing. A model-based DRS and an ensemble-based DRS with an adaptive, responsive and interactive user interface has been successfully developed and implemented. The DRS aims to assist tourists plan before or during their visit to an unfamiliar city. Both technical and practical aspects were considered, including data sparsity, scalability, transparency, system accuracy, usability and user acceptance.

7.1 Objectives revisited

1. *To review Travel Recommendation Systems (TRSs) from the available literature and identify research challenges and gaps*

   An extensive literature review was carried out with regard to travel-recommendation systems for the purposes of this research study. The review began with an overview of RSs and their engines. TRS developments in the period between 2008 and 2015 were then reviewed. Published studies on TRSs were selected from well-known online libraries and classified according to different criteria, including the technologies involved in TRS development, e-tourism services that TRSs currently provide, theories to improve the level of personalisation, methodologies and system evaluation. Based on the proposed semantic review method, the general system framework of a TRS was presented. Based on the literature review (Chapter 2), challenges and research gaps in TRS development were identified.

2. *To design and develop a questionnaire for data collection from a case-study city*

Three data sets were used in the process of developing the proposed DRS. Chapter 3 describes the data acquisition of the system. First, a Chiang Mai POI data set was collected for a first DRS prototype implementation. Second, a Taiwanese data set was collected for a second DRS prototype. Finally, a questionnaire was designed based on an empirical study and the Taiwanese data set. At the beginning of Chapter 4, two destination TRS prototypes were implemented and investigated to determine the weaknesses of current systems' characteristics.

3. *To identify features and data-processing techniques for the proposed system*

The proposed DRS was implemented based on a DM approach, using data collected from Chiang Mai through the designed questionnaire. The data set obtained was decomposed into eight sub-data sets using relevant tourism-domain knowledge. This was done to increase the system performance and reduce the complexity of the DT model.

4. *To investigate techniques for the classification of tourists' preferred destinations and evaluate classification results that generated through the use of a variety of techniques*

Eight optimal C4.5 DTs were built as our baseline classifiers. Two classifications of algorithm performance, SVM and MLP, were compared and investigated. This included different results from three SVM toolboxes and two MLP toolboxes. The experiment results indicated that MLP outperformed DT and SVM.

In this study we developed a novel model-based DRS that recommends 20 destinations to tourists before or during their visit to the city of Chiang Mai. The aim of this study was to solve the current practical and technical issues that beset destination TRSs. We achieved this by reducing users' efforts while maintaining a decent system-accuracy rate. This study also investigated five sets of factors that influenced tourists' preferred destinations, including trip characteristics, tourist characteristics, tourist expenditure behaviour, travel motivation and tourist socio-demographic information. The data set was decomposed into seven sub-data sets using relevant tourism-domain knowledge; this was done to increase the classification-accuracy rate and reduce the complexity of the DT. Seven DTs were obtained along with the highest classification-accuracy rate for each data set.

Three scientific evaluation methods were used to assess the performance of predictive models: accuracy rate, confusion matrix and f-measure. Regarding system performance, we achieve a 80% classification-accuracy rate for the Museum data set, 71.9% for the Temple-peaceful data set, 71.72% for the Temple-old town data set, 64.1% for the Art Gallery data set, 61.25% for the Temple-landmark data set, 52.76% for the Temple-outer town data set and 49.72% for the Nature data set. Regarding the performance of the two feature-selection algorithms, the NMIFS algorithm is considered superior to the mRMR algorithm, except in the case of the Temple-outer town data set, where mRMR performs better. It can be seen that NMIFS is the optimum method because it uses fewer features than mRMR for both data sets. Optimal DTs, with the highest accuracy rate and simplicity (i.e. fewer leaves and smaller size), were constructed for each data set. Decision rules were extracted from the DTs. Finally, the experimental results confirmed the applicability of the proposed DRS. The proposed DRS satisfied the requirements of tourists who planned to visit the city of Chiang Mai or proved satisfying to those tourists during their visit to that city.

5. *To develop an interactive and adaptive user interface for the proposed DRS*

We have proposed a front-end adaptive, responsive and interactive model-based user interface (ARIM-UI). Several Web technologies including JavaScript, MVVM pattern, HTML, XML and CSS were chosen in order to develop the proposed UI for the DRS. Our proposed user interface provides three main functionalities: responsiveness, interactivity and adaptability. Additionally, we demonstrated the design and implementation of the UI system by providing three important diagrams: a class diagram, a workflow diagram and a sequence diagram.

7.2 Empirical findings

This section summarises the findings regarding the research questions.

1. *How to detect a tourist's preferred destination?*

   Destination preference plays a major role in chhosing tourist destinations to visit. Detecting a tourists' preferred destinations is extremely challenging as they are often hidden and not explicitly known by people at the start of or during travel (LOH et al., 2003). There are two approaches to detecting a preferred destination. The first is based on content-based filtering techniques, such as a tourist's past travel behaviour; and the second one is based on collaborative filtering techniques, such as those based on other travellers. We estimate a tourist's preferred destination by combining the two approaches as a hybrid filtering technique using a DT. The destination-search process needs to be understood. Therefore, we used a questionnaire as the data-collection method to investigate five sets of factors that influence tourists' preferred destinations, including trip characteristics, tourist characteristics, tourist expenditure behaviour, travel motivation and tourists' socio-demographic information based on qualitative research. There are no secondary data that can be used for this research. The primary data used were both qualitative and quantitative, using a mixture of qualitative and quantitative methods. In addition to physical/ sensor data for the destination-recommendation system itself, a quantitative method is the best approach in practice. In practical applications, questionnaires can be collected every year to acquire input that can be updated within the system's back-end.

2. *Which set of factors plays an important role in making destination recommendations to tourists? Does using multiple factors help increase recommendation accuracy? Do travel-motivation factors contribute to increasing the level of recommendation accuracy?*

   Based on the experimental results presented in Chapter 4, tourist behaviour was the most commonly used (28.5%) followed by travel characteristics (25.7%). It can be seen from the results that there are no common 'most important factors' to estimate destinations for all the data sets. The results confirmed that using different features from multiple factors does improve recommendation accuracy. The experimental results from Chapter 4 also indicate that combining tourist-motivation factors does improve recommendation accuracy.

3. *How can users' efforts be reduced, while still maintaining the same degree of recommendation performance and increasing levels of user satisfaction in the decision-making process when selecting destinations?*

   Unnecessary inputs that are either irrelevant or redundant were eliminated using our proposed two-feature selection method. The experimental results presented in Chapter 4 confirmed that the proposed DRS used a small number of relevant and non-redundant inputs from 3–7 features to achieve the best recommendation results. This means that the proposed system is considered non-intrusive and more likely to be accepted by users.

4. *How can an optimal decision model be constructed when using multiple sets of factors for multiple tourist destinations?*

   The process of constructing a destination-choice model was divided into two phases. The first phase involved decomposing the classes into a group of clusters. The second process involved pre-processing data and applying several supervised machine-learning algorithms to build decision models. The models were evaluated using appropriate scientific methods.

5. *How can the recommendation accuracy rate be improved using only the relevant and non-redundant factors?*

   By combining the results generated by different classifiers, and using different voting strategies, recommendation performance was improved, as presented in Section 5.3.

6. *How can tourists be encouraged to interpret and interact with the constructed decision model(s)?*

   For each destination-choice model, input variables were extracted from a C4.5 algorithm and converted to XML format. Each XML file represents one destination-choice model. Next, the XML files were uploaded to the proposed user interface (ARIM-UI), which supports three main functionalities: adaptability, responsiveness and interactivity. Details of the design and implementation of the proposed UI are presented in Chapter 6.

7.3 Research limitations

1. *Deploying the system in a new city*

   This recommendation system has only been applied to Chiang Mai. To use the system with another city or other destinations, a new data set would need to be collected, and the factors that influence tourists' destination choices may be different (e.g. people's behaviours are different, destinations are different etc.), and these are automatically identified through the DRS. The system can be maintained by updating it with a new data set.

2. *A limit in the number of training samples*

   Since no secondary data can be used for this research, acquiring a data set for this project was expensive and time-consuming. At the beginning of the project, 4,000 samples were collected, which is the optimal number of samples when using the machine-learning approach. However, due to the complexity of the problem, the data set needed to be broken down into several sub-data sets. This led to a lower number of training samples for each model. Therefore, this would affect the classification accuracy, as well as the performance of the recommendation system.

7.4 Future work

More research on DRS can be conducted based on the research limitations described above. Furthermore, future research on the proposed DRS regarding improvements in recommendation accuracy can be carried out within the process of machine-learning. The future research directions proposed are grouped into the following four aspects:

7.4.1 Soft-constraint aspect

Multiple types of user input through combining tourist behaviour factors and factors from users' mobile contexts

Due to the increase in mobile tourism, and improvements in technology such as wireless communication and sensors, temporary factor,s such as location, time, opening times, closing times and weather, can be integrated into a system account.

7.4.2 Data pre-processing aspect.

Dealing with class imblances

Real-world data sets are usually characterised by class imbalance, in other words, classes are usually not equally represented, such as in the data set we collected for this research. If collecting more data is not an option, then to deal with imbalances, the recommendation performance of the proposed DRS can be improved in the data pre-processing stage. For instance, we need to handle an imbalanced data set before passing it to the process of model construction. In a future research direction, resampling techniques such as undersampling the majority class or oversampling the minority class could be used to deal with imbalanced data. Also, synthetic sampling algorithms, such as SMOTE (Chawla et al., 2002), could be applied with respect to imbalanced data. Despite the fact that these methods can improve the predictive performance of the model, they could also cause bias in the data set. Therefore, it is critical to understand how bias affects the outcomes of models.

7.4.3 Class-decomposition aspect

Class decomposition is a crucial step in data-mining and machine-learning, where the goal is to separate each class into a group of clusters before constructing a predictive model. Many class-decomposition techniques have been proposed, such as decomposition using K-mean and Hierarchical Clustering (Banitaan et al., 2015), Error Correcting Output Coding (ECOC) (Zhou et al., 2008), one-vs-one (Garcia et al., 2015) etc. However, in real-world problems, these techniques usually generate an inappropriate model, even though the techniques may return a better classification performance. An example is the real-world problem presented in this thesis, where there are 20 classes involved. By applying simple decomposition techniques such as K-mean, the results might have returned 6–7 different destination-choice models, and those models could have returned better performances than the models proposed in this thesis. However, the models generated by these techniques are only concerned with accuracy and are meaningless in practice, where we also have to consider user satisfaction with the system. Therefore, when handling class decomposition, the best approach is to strike a balance between technique and practical aspects.

7.4.4 Classification algorithm aspect

1. *Rule-based classification approach*

   A promising approach to increase classification accuracy is to use rule-based classifiers because we can benefit from the rules derived from models. Rules can be pruned by using a tourist-domain expert to generate higher predictive accuracy. Moreover, irrelevant or redundant features can also be eliminated during the process of converting from the DT to rules by integrating an existing algorithm or modifying the C4.5 algorithm.

2. *More diverse combination rules*

   Further studies should attempt to construct more combination rules such as Bayes, Decision Template, Dempster-Shafer (DS) or Behaviour Knowledge Space (BKS).

3. *More traditional classification algorithms*

Other traditional classification algorithms, such as KNN, RTree and RBF, could be used as base learners and could benefit from the utilisation of both boosting and bagging methods. In addition, ensemble learning methods, such as stacking, random forest, random sub-spaces or pasting, could be employed.

4. *Deep learning approach*

Another emerging paradigm in machine-learning society is deep learning. Deep learning has been applied, and been successful, in computer-vision applications, such as image recognition. It would be interesting to see how this machine-learning method could be used in categorical data sets like ours, what data pre-processing steps would be necessary before training the model, and what the selection of the network architecture for the destination classification problem would be. In the previous experiment we constructed a three-layered feed-forward neural network consisting of input, hidden and output layers, also known as MLP, to classify tourist-destination problems. The data move from the input layer through hidden nodes to the output nodes. The experimental results show that the MLP classifiers outperformed other classical classifiers such as C4.5 and SVM. In the next study, we can apply deep neural networks, i.e. ones which have multiple hidden layers. The term ‘deep’ refers to the nesting of non-linear functions (Bengio, 2009). The concept of having many hidden layers will allow us to compute much more complex features of given input.

7.4.5 User Interface aspect

There are three research directions for the proposed user interface:

1. *Towards semantic websites*

   The first involves bridging the gap between a generated model file, such as XML and JSON, and semantic Web rule language.

2. *Feedback mechanism*

   The second direction is to implement a feedback mechanism, such that the tourist can rate and review destinations. Integrating user reviews and ratings, this could enhance our DRS and bridge the gap between model-based and review-based RS. User reviews and ratings could be exploited using text analysing to design a more effective UI.

3. *User-interface evaluation*

   Further development of the user interface for the proposed DRS should focus on the evaluation method. Methods involving heuristic evaluation, usability testing, guidelines, and cognitive walkthrough should be reviewed carefully, because each evaluation method has its own advantages and disadvantages. According to Jeffries (1991), heuristic evaluation can identify a severe problem in the UI, but the method requires UI expertise to apply heuristic critique to an interface effectively. A guidelines method is considered the best one to find general and recurring problems. However, this method has a problem when identifying severe problems. A usability method is capable of finding general and recurring problems, as well as severe problems, but it is not as good as a heuristic method; however, the cost of using this method is high (Jeffries et al., 1991). A cognitive walkthrough is very good at identifying users' goals and assumptions, but the method is time-consuming and less effective in terms of finding general, recurring and severe problems. The effectiveness and success of DRS depend on system usability; therefore, selecting the most effective evaluation method is an important aspect and is a crucial step towards developing a successful DRS.

In this thesis, we have proposed an intelligent DRS using model-based and ensemble-based approaches based on machine-learning techniques. We have compared and studied several well-known classification algorithms, and we found that MLP was superior to the others for the data sets. We have shown, in an experimental study, how ensemble learning methods could be exploited to improve the classification-accuracy rate of the DRS. Moreover, the development of a model-based user interface that has adaptive, responsive and interactive capability was carried out at the end of this thesis in order to increase the level of user satisfaction with the system.

Appendix A

The questionnaire that was used for data collection:

Survey ID_______ Survey location______________ Day___Month___2014

<u>Tourist Destination Choice and Satisfaction Survey</u>

This survey is being conducted for PhD. research purpose, as a part of the e-tourism project and on behalf of Chiang Mai University and Bournemouth University. The survey results will help us to understand tourist destination choice in the city of Chiang Mai. We greatly appreciate your time for 10-15 minutes to answer this questionnaire. Your participation in this study is entirely voluntary and you will not be identified with any of your responses to the survey. Your viewpoint is very important to us. Thank you for taking the time to complete this survey. All the questions below ask you for this trip.

A. Trip characteristic information

A1. How many night(s) do you plan to stay in Chiang Mai?
☐(1) 0-1 ☐(2) 2-3 ☐(3) 4-7
☐(4) 8-14 ☐(5) 15-30 ☐(6) 31 or more

A2. Is anyone accompanying you on this trip?
☐(1) Yes **(Please go to A3.)**
☐(2) Yes **(Please go to A3.)**

A3. The people who are accompanying you are: **(Please tick all that apply)**
☐(1) Friends ☐(2) Parents ☐(3) Spouse
☐(4) Relatives ☐(5) With children ☐(6) Colleagues

A4. How many times have you visited Chiang Mai **(Including this trip)** in the last five years?
☐(1) First time ☐ (2) 2-3 times
☑(3) 4-7 times ☐(4) 8-20 times ☐(5) 21 times or more

A5. Which of the following define your travel style? **(Please tick all that apply)**
☐(1) Adventurer, prefers outdoor adventure and general sightseeing
☐(2) Multiple interests, prefers diverse activities
☐(3) Relaxation seeker, prefers amusement, relaxation, and general sightseeing activities
☐(4) Cultural, prefers performing arts and local events

B. Plan for your trip

B1. How did you arrange this trip to Chiang Mai?
☐(1) Group tour arranged through a travel agency. (Join a tour group)
☐(2) Independent tour arranged through a travel agency. (e.g., family and friends tour group, school holiday trip, etc.)
☐(3) Flight and accommodation tour booked by agency.
☐(4) Personally planned tour arranged by Chiang Mai travel agency.
☐(5) No arrangements by a Chiang Mai travel agency after arrival.

B2. What is the purpose of taking this trip to Chiang Mai?
☐(1) Business reasons
☐(2) Holiday or Vacation
☐(3) Visit relatives and friends
☐(4) Religious reasons
☐(5) Conference, meeting, seminar, exhibition
☐(6) Heath related
☐(7) Shopping
☐(8) Other______________________

B3. Which of the following information sources influences your decision to visit Chiang Mai? **(Please tick all that apply)**
☐(1) Friend or relative recommendation
☐(2) The Internet ☐(3) Travel agency
☐(4) Books and guides ☐(5) TV, radio
☐(6) Personal Experience
☐(7) Other______________________

C. Expenditure behavior

C1. Which of the following expenses have you prepaid before arrival in Chiang Mai? **(Please tick all that apply)**
☐ (1) Hotel ☐ (2) Meals outside hotel
☐ (3) Local transportation ☐ (4) Entertainment
☐ (5) Miscellaneous expenses

C2. How much money do you plan to spend on this trip (in US dollars)? (*1 dollar equals 32 Thai baht)*
☐(1) Less than $100 ☐(2) $101-$500
☐(3) $501-$1000 ☐(4) $1001- $5000
☐(5) $5001-$10,000 ☐ (6) $10,000 and over

C3. How much money do you plan to spend <u>per person</u> on the following items during your stay in Chiang Mai (in US dollars)? (*1 dollar equals 30 Thai baht)*

(1) Transportation	Approx.__________US$
(2) Accommodation	Approx.__________US$
(3) Restaurants, cafes	Approx.__________US$
(4) Souvenirs	Approx.__________US$
(5) Entertainment	Approx.__________US$
(6) Shopping	Approx.__________US$
(7) Other expenses	Approx.__________US$

D. Tourist behavior

D1. Please tick all the attractions which you plan to visit during your stay in Chiang Mai. **(Please tick all that apply)**
☐(1) Wat Chedi Luang
☐(2) Chiang Mai Caberet Show
☐(3) Wat Phra That Doi Suthep
☐(4) Museum of World Insects and Natural Wonders
☐(5) Art in Paradise, Chiang Mai 3D Art Museum
☐(6) Doi Inthananon
☐(7) Wattana Art Gallery
☐(8) Wat Phra Singh
☐(9) Wat Phra That Doi Kham
☐(10) Wat Umong
☐(11) Wat Sri Suphan
☐(12) Wat Lok Molee
☐(13) Wat Suan Dok
☐(14) Wat Pan Tao
☐(15) Wat Chiang Man
☐(16) Documentary Arts Asia
☐(17) Burklerk Gym- Muay Thai Training
☐(18) Bua Thong Waterfalls (Nam Phu Chet Si)
☐(19) Huay Tung Tao Lake
☐(20) Mae Sa Waterfall
☐(21) Other______________________

Survey ID_______ Survey location______________ Day___Month___2014

D2. Which one of the followings is your most favorite attraction which you plan to visit during your stay in Chiang Mai?
☐(1) Wat Chedi Luang
☐(2) Chiang Mai Caberet Show
☐(3) Wat Phra That Doi Suthep
☐(4) Museum of World Insects and Natural Wonders
☐(5) Art in Paradise, Chiang Mai 3D Art Museum
☐(6) Doi Inthananon
☐(7) Wattana Art Gallery
☐(8) Wat Phra Singh
☐(9) Wat Phra That Doi Kham
☐(10) Wat Umong
☐(11) Wat Sri Suphan
☐(12) Wat Lok Molee
☐(13) Wat Suan Dok
☐(14) Wat Pan Tao
☐(15) Wat Chiang Man
☐(16) Documentary Arts Asia
☐(17) Burklerk Gym- Muay Thai Training
☐(18) Bua Thong Waterfalls (Nam Phu Chet Si)
☐(19) Huay Tung Tao Lake
☐(20) Mae Sa Waterfall
☐(21) Other________________________

D3. Which one of the followings is your favorite for this trip? **(Please tick all that apply)**
☐ (1) Cultural and historical
☐ (2) Performances
☐ (3) Natural scenery and landscape
☐ (4) Educational and sport site
☐ (5) Museums and art galleries

D4. Please tick all the activities which you plan to participate in during your stay in Chiang Mai. **(Please tick all that apply)**
☐(1) Attend festivals (2) Attend performances
☐(3) Attend cultural events
☐(4) Attend exhibitions ☐(5) Outdoor recreation
☐(6) Biking ☐(7) Rafting ☐(8) Golfing
☐(9) Hot springs ☐(10) Thai Boxing
☐(11) Shopping ☐(12) Hiking/Climbing
☐(13) Visit historical places
☐(14) Visit markets, walking streets
☐(15) Visit health spas, massage, sauna
☐(16) Visit entertainment places, nightclubs, bars
☐(17) Visit art galleries ☐(18) Visit mountains
☐(19) Visit national park/forests
☐(20) Visit museums ☐(21) Restaurants/dinning out
☐(22) Sightseeing in cities
☐(23) Health care ☐(24) Thai cooking
☐(25) Observing wildlife
☐(26) Sampling local food
☐(27) Other_______________

D5. What makes you interested in plan to participate in the activities? **(Please tick all that apply)**
☐ (1) Entertainment activities
☐ (2) Culture based sightseeing
☐ (3) Outdoor activities
☐ (4) Thai cuisine
☐ (5) Thai spa and traditional message activities
☐ (6) Local activities
☐ (7) Nature based activities

D6. Which part of Chiang Mai has given you the deepest impression? **(Please tick all that apply)**
☐ (1) Temple ☐ (2) Thai food ☐ (3) Night life
☐ (4) Art gallery ☐ (5) Night market, walking street
☐ (6) Heath Massage, spa ☐ (7) Wildlife
☐ (8) Biking ☐ (9) Golfing ☐ (10) Hot spring
☐ (11) Shopping ☐ (12) Thai Boxing
☐ (13) Nature ☐ (14) Museum

D7. What type of accommodation do you plan to use in Chiang Mai?
☐(1) Hotel ☐(2) Guest house ☐(3) Youth hostels
☐(4) Home of relatives or friends
☐(5) Dormitory ☐(6) Rental apartments ☐(7) Temple
☐(8) Other________________________

D8. Please rate the overall price that you plan to spend on your meal/food in Chiang Mai.
(1-Least expensive, 5-Very expensive)
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

D9. Please rate the overall price that you plan to spend on accommodation in Chiang Mai.
(1-Least expensive, 5-Very expensive)
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

D10. Please tick the transport modes that you plan to use during this trip in Chiang Mai. **(Please tick all that apply)**
☐(1) Walk ☐(2) Bicycle ☐(3) Rental car
☐(4) Shared Taxi (Rod dang/Red cab) ☐(5) Taxi
☐(6) Bus ☐(7) Private car/motorcycle/van/coach

E. Travel Motivation
E1. How would you rate the following motives for this trip in Chiang Mai?
(1-Strongly disagree, 5-Strongly agree)

E1.1 Self-actualize:
(1) To understand more about myself
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To gain a new perspective on life
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3) To work on my personal/spiritual values
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(4) To seek a better existence
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

E1.2 Escape/Relaxation:
(1) To experience solitude and calm
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To experience inner harmony and peace
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3) To refresh mentally and physically
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(4) To rejuvenate myself
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(5) To not worry about time and work
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Survey ID_______ Survey location______________ Day___Month___2014

E1.3 Novelty:
(1) To have fun
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To experience something different
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3) To feel the special atmosphere of the destination
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(4) To visit places related to my personal interests
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

E1.4 Adventure:
(1) To find excitement
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To experience the risk involved
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3) To experience danger and thrills
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(4) To visit places I have never been before
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

E1.5 Learning experience:
(1) To discover new people, places and things
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To see famous cultural and historical sites
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3) To develop new abilities
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(4) To learn about Thai cuisine
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(5) To learn about nature
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

E1.6 Relationship:
(1) To do things with family and friend(s)
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To do something with my companion(s)
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3) To enhance relationships with friend(s)/family
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(4) To visit relatives/friend(s)
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

E1.7 Social status:
(1) To visit a destination that would impress my friends or family
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To share what I have learned with others
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3) To reveal my thoughts, feelings, or physical skills to others
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

E1.8 Romance:
(1) To improve my romantic life
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To experience fantasy of travel
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3) To reflect on past memories
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(4) To be with people of the opposite sex
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

E1.9 Shopping:
(1) To go shopping
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(2) To buy local Thai product
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
(3)To buy world famous brand-name products
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

F. Tourist Satisfaction
F1. How would you rate the level of importance and satisfaction for the attraction you are currently visiting?
(1-Not very important, 5-Very important)
(1-Very Unsatisfied, 5-Very Satisfied)

F1.1 Price:
(1) Price of general goods and services
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(2) Fairness of services and goods relative to price
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐

F1.2 Hospitality:
(1) General friendliness of the people in the area
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(2) General friendliness of the employees of facilities
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(3) Willingness of the employees of facilities to aid tourists
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐

F1.3 Food and beverage
(1) Variety of food and beverage
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(2) Seating space
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(3) Quality of services
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(4) Quality of food and beverage
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐

F1.4 Facility:
(1) Cleanliness
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(2) Personal safety and security
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(3) Climate condition
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐

Survey ID_______ Survey location______________ Day___Month___2014

Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(4) Beauty of the scenery
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(5) Variety of activity
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(6) Cultural and historical content
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐

F1.5 Accessibility:
(1) Availability of travel information
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(2) Availability of local parking
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(3) Convenience of public transport
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
(4) Efficiency and safety of public transport
Importance 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐
Satisfaction 1☐ 2☐ 3 ☐ 4 ☐ 5 ☐

G. Demographic information
G1. Gender: ☐ (1) Male ☐(2) Female

G2. Age (years old): ☐(1)18-25 ☐(2) 26-35
☐(3) 36-45 ☐(4) 46-55 ☐(5) 56-65
☐(6) 66 and over

G3. Marital status: ☐(1) Single ☐(2) Married
☐(3) Separated ☐(4) Widowed ☐(5) Divorced

G4. Highest Education
☐(1) Middle school
☐(2) High school
☐(3) College graduate
☐(4) Bachelor's degree
☐(5) Master's degree
☐(6) Doctorate degree
☐(7) Other________________________

G5. What is your current household annual income in U.S. dollars ($)?
☐(1) Under $3,000 ☐(2) $3,000-5,000
☐(3) $5,001 - $10,000 ☐(4) $10,000 - $15,000
☐(5) $15,001 - $30,000 ☐(6) $30,001 - $60,000
☐(7) $60,001 - 90,000 ☐(8) $90,000 or more

G6. Which of the following categories best describes your primary area of employment (regardless of your actual position)?
☐(1) Company employee
☐(2) Business owner
☐(3) Self-employed
☐(4) Government sector
☐(5) Professionals/Scientist
☐(6) Housewife/Househusband
☐(7) Student
☐(8) Retired
☐(9) Unemployed
☐(10) Other________________________

G7. What is your nationality?
☐(1) China ☐(2) Laos ☐(3) Malaysia
☐(4) Singapore ☐(5) Korea ☐(6) Japan
☐(7) U.S.A ☐(8) U.K. ☐(9) France
☐(10) Germany ☐(11) Russia ☐(12) Sweden
☐(13) India ☐(14) Australia or New Zealand
☐(15) Thailand
☐(16) Other________________________

G8. What is the country of your residence?
☐(1) China ☐(2) Laos ☐(3) Malaysia
☐(4) Singapore ☐(5) Korea ☐(6) Japan
☐(7) U.S.A ☐(8) U.K. ☐(9) France
☐(10) Germany ☐(11) Russia ☐(12) Sweden
☐(13) India ☐(14) Australia or New Zealand
☐(15) Thailand
☐(16) Other________________________

G9. Where is your origin?
Overseas region (for international tourist):
☐(1) East Asia
☐(2) Europe
☐(3) The Americas
☐(4) South Asia
☐(5) Oceania
☐(6) Middle East
☐(7) Africa
Domestic region (for local/Thai tourist):
☐(8) Central Thai
☐(9) Northeastern Thai
☐(10) Northern Thai
☐(11) Southern Thai

H. Other suggestion:
____
____
____
____
____
____

Thank you for completing this questionnaire. Your responses will be a valuable contribution to what is presently know about the importance of tourist destination choice and satisfaction in Chiang Mai.

If you have any questions regarding the survey, please contact Mr. Pree Thiengburanthum (pree.t@cmu.ac.th, 09-4603-8726) Have a great trip in Chiang Mai!

แบบสำรวจเลขที่______ สถานที่สำรวจ______________ วันที่___เดือน____พ.ศ.2557

แบบสอบถามสำรวจตัวเลือกปลายทาง และ ความพึงพอใจของนักท่องเที่ยว

แบบสอบชุดนี้มีวัตถุประสงค์เพื่องานวิจัยระดับปริญญาเอก โดยงานวิจัยนี้เป็น ส่วนหนึ่งของโครงการ e-tourism ซึ่งเป็นความร่วมมือกันระหว่างมหาวิทยาลัยเชียงใหม่ และมหาวิทยาลัย Bournemouth University ประเทศอังกฤษ ทั้งนี้ผลของการสำรวจแบบสอบถามจะช่วยให้ทีมงานวิจัยเข้าใจถึงตัวเลือกของสถานที่ท่องเที่ยวในเมืองเชียงใหม่ของนักท่องเที่ยว ทางทีมงานวิจัยขอขอบพระคุณท่านเป็นอย่างยิ่งในการสละเวลาประมาณ 10-15 นาที สำหรับการตอบแบบสอบถามครั้งนี้ และท่านจะไม่มีส่วนกระทบ และรับผิดชอบใดๆ ในการตอบแบบสอบถามชุดนี้ การตอบแบบสอบถามของท่านครั้งนี้เป็นความสมัครใจและขอบพระคุณมากสำหรับตอบแบบสอบถามชุดนี้ คำถามทุกข้อในแบบสอบถามชุดนี้ถามเกี่ยวกับการเดินทางครั้งนี้ของท่านเท่านั้น

A. ลักษณะการเดินทางท่องเที่ยว

A1. ท่านมีแผนค้างคืนในจังหวัดเชียงใหม่ เป็นเวลาจำนวนกี่คืน

☐(1) 0-1คืน ☐(2) 2-3 คืน ☐(3) 4-7 คืน

☐(4) 8-14 คืน ☐(5) 15-30 คืน ☐(6) 31 คืนขึ้นไป

A2. ท่านมีคนเดินทางมาด้วย สำหรับการเดินทางครั้งนี้ใช่หรือไม่

☐(1) ใช่ (โปรดทำแบบสอบถามข้อ A3.)

☐(2) ไม่ใช่ (โปรดข้ามไปทำแบบสอบถามข้อ A4.)

A3. ท่านเดินทางมากับใครที่จังหวัดเชียงใหม่ในครั้งนี้ (ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)

☐(1) เพื่อน ☐(2) ผู้ปกครอง ☐(3) คู่สมรส

☐(4) ญาติ ☐(5) ลูก ☐(6) เพื่อนร่วมงาน

A4. ภายในเวลา 5 ปีที่ผ่านมา ท่านเคยเดินทางมาที่จังหวัดเชียงใหม่แล้วเป็นจำนวนกี่ครั้ง (นับรวมการเดินทางครั้งนี้ด้วย)

☐(1) ครั้งแรก ☐(2) 2-3 ครั้ง

☐(3) 4-7 ครั้ง ☐(4) 8-20 ครั้ง ☐(5) 21 ครั้งขึ้นไป

A5. ท่านมีการเดินทางในลักษณะแบบไหนในข้อต่อไปนี้ (ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)

☐(1) การเดินทางแบบผจญภัย เช่น ชอบการผจญภัยกลางแจ้ง และเที่ยวชมสถานที่ท่องเที่ยวทั่วไป

☐(2) การเดินทางแบบหลากหลาย เช่นชอบการเดินทางที่รวมหลากหลายกิจกรรม

☐(3) การเดินทางแบบผ่อนคลาย เช่น เน้นสถานที่บันเทิง มีกิจกรรมที่ผ่อนคลาย และเที่ยวชมสถานที่ทั่วไป

☐(4) การเดินทางที่เน้นวัฒนธรรม เช่นชอบชมศิลปะและ กิจกรรมในท้องถิ่น

B. แผนการเดินทางของท่าน

B1. ท่านวางแผนการเดินทางมาท่องเที่ยวที่จังหวัดเชียงใหม่อย่างไร

☐(1) ท่องเที่ยวแบบกลุ่ม โดยเข้าร่วมท่องเที่ยวกับพร้อมกับกลุ่มนักท่องเที่ยวอื่นๆ โดยใช้บริการของบริษัทรับจัดการท่องเที่ยว

☐(2) ท่องเที่ยวแบบกลุ่มส่วนตัว (เช่น ท่องเที่ยวมากับครอบครัว หรือกับกลุ่มเพื่อนสนิท) โดยใช้บริการของบริษัทรับจัดการท่องเที่ยว

☐(3) ใช้เพียงบริการจองตั๋วเครื่องบิน และที่พัก ผ่านบริษัทจัดการท่องเที่ยว

☐(4) ใช้บริการการจัดการท่องเที่ยวแบบส่วนตัว ผ่านบริษัทการท่องเที่ยวที่เชียงใหม่

☐(5) ไม่ใช้บริการใดๆ จากบริษัทจัดการท่องเที่ยวที่เชียงใหม่

B2. วัตถุประสงค์ของการเดินทางมาเชียงใหม่ของท่านคือ

☐(1) การทำงาน หรือ ธุรกิจ

☐(2) วันหยุด หรือ พักร้อน

☐(3) เยี่ยมญาติ หรือ เพื่อน

☐(4) กิจกรรมทางศาสนา

☐(5) ประชุม, สัมมนา และนิทรรศการ

☐(6) เกี่ยวกับ สุขภาพ อนามัย

☐(7) ช้อปปิ้ง

☐(8) อื่นๆ________________

B3. แหล่งข้อมูลใดในข้อต่อไปนี้มีอิทธิพลในการตัดสินใจในการเดินทางมาจังหวัดเชียงใหม่ของท่าน (ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)

☐(1) คำแนะนำจากเพื่อน หรือญาติ

☐(2) อินเตอร์เน็ต ☐(3) บริษัทการท่องเที่ยว

☐(4) หนังสือและคู่มือแนะนำสถานที่ท่องเที่ยว ☐(5) โทรทัศน์ และ วิทยุ

☐(6) ประสบการณ์ส่วนตัว

☐(7) อื่นๆ________________

C. พฤติกรรมการใช้จ่าย

C1. ค่าใช้จ่ายใดที่ท่านได้ชำระล่วงหน้าก่อนที่จะเดินทางมาถึงจังหวัดเชียงใหม่(ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)

☐ (1) ค่าโรงแรม ☐ (2) ค่าอาหารนอกโรงแรม

☐ (3) ค่าเดินทางขนส่งท้องถิ่น ☐ (4) ค่าบันเทิง ☐ (5) ค่าใช้จ่ายเบ็ดเตล็ด

C2. จำนวนเงินค่าใช้จ่ายทั้งหมดโดยประมาณที่ท่านวางแผนจะใช้ในการเดินทางครั้งนี้

☐(1) น้อยกว่า 3,000 บาท ☐(2) 3,001 – 15,000 บาท

☐(3) 15,001 – 30,000 บาท ☐(4) 30,001 – 150,001 บาท

☐(5) 150,001 – 300,000 บาท ☐ (6) มากกว่า 300,000 บาท

C3. จำนวนเงินค่าใช้จ่ายตามรายการข้างล่าง<u>ต่อคน</u>ที่ท่านวางแผนจะใช้ในการท่องเที่ยวในจังหวัดเชียงใหม่

(1) ค่าเดินทาง	ประมาณ__________บาท
(2) ค่าที่พัก	ประมาณ__________บาท
(3) ค่าอาหาร และ เครื่องดื่ม	ประมาณ__________บาท
(4) ค่าของที่ระลึก	ประมาณ__________บาท
(5) ค่าความบันเทิง	ประมาณ__________บาท
(6) ค่าช้อปปิ้ง	ประมาณ__________บาท
(7) ค่าอื่นๆ	ประมาณ__________บาท

D. พฤติกรรมของนักท่องเที่ยว

D1. โปรดเลือกสถานที่ท่องเที่ยวในจังหวัดเชียงใหม่ ที่ท่านวางแผนที่จะไปเยี่ยมชมในระหว่างท่องเที่ยวในจังหวัดเชียงใหม่ (ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)

☐(1) วัดเจดีย์หลวง

☐(2) เชียงใหม่ คาบาเร่ต์ โชว์

แบบสำรวจเลขที่______ สถานที่สำรวจ________________ วันที่___เดือน____พ.ศ.2557

☐(3) วัดพระธาตุดอยสุเทพ
☐(4) พิพิธภัณฑ์แมลงโลกและสิ่งมหัศจรรย์ธรรมชาติ
☐(5) พิพิธภัณฑ์ภาพวาด 3 มิติ
☐(6) ดอยอินทนนท์
☐(7) หอศิลป์ วัฒนะ
☐(8) วัดพระสิงห์
☐(9) วัดพระธาตุดอยคำ
☐(10) วัดอุโมงค์
☐(11) วัดศรีสุพรรณ
☐(12) วัดโลกโมฬี
☐(13) วัดสวนดอก
☐(14) วัดพันเตา
☐(15) วัดเชียงมั่น
☐(16) เชียงใหม่ ภาพถ่ายและสารคดีแห่งเอเชีย
☐(17) ค่ายมวยไทย เม็กฤทธิ์ยิม
☐(18) น้ำตกบัวตอง
☐(19) ห้วยตึงเฒ่า
☐(20) น้ำตกแม่สา
☐(21) อื่นๆ__________________

D2. โปรดเลือกสถานที่ท่องเที่ยวที่ท่านชื่นชอบมากที่สุดที่ท่านวางแผนจะไปเยี่ยมชมในระหว่างท่องเที่ยวในจังหวัดเชียงใหม่
☐(1) วัดเจดีย์หลวง
☐(2) เชียงใหม่ คาบาเร่ต์ โชว์
☐(3) วัดพระธาตุดอยสุเทพ
☐(4) พิพิธภัณฑ์แมลงโลกและสิ่งมหัศจรรย์ธรรมชาติ
☐(5) พิพิธภัณฑ์ภาพวาด 3 มิติ
☐(6) ดอยอินทนนท์
☐(7) หอศิลป์ วัฒนะ
☐(8) วัดพระสิงห์
☐(9) วัดพระธาตุดอยคำ
☐(10) วัดอุโมงค์
☐(11) วัดศรีสุพรรณ
☐(12) วัดโลกโมฬี
☐(13) วัดสวนดอก
☐(14) วัดพันเตา
☐(15) วัดเชียงมั่น
☐(16) เชียงใหม่ ภาพถ่ายและสารคดีแห่งเอเชีย
☐(17) ค่ายมวยไทย เม็กฤทธิ์ยิม
☐(18) น้ำตกบัวตอง
☐(19) ห้วยตึงเฒ่า
☐(20) น้ำตกแม่สา
☐(21) อื่นๆ__________________

D3. สิ่งใดต่อไปนี้ ที่ท่านชื่นชอบสำหรับการท่องเที่ยวครั้งนี้(ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)
☐ (1) วัฒนธรรมและประวัติศาสตร์
☐ (2) งานแสดงสด
☐ (3) ธรรมชาติและภูมิทัศน์
☐ (4) การศึกษาและเกี่ยวกับกีฬา
☐ (5) พิพิธภัณฑ์และหอศิลป์

D4. โปรดเลือกกิจกรรมที่ท่านวางแผนที่จะเข้าร่วมระหว่างอยู่ในจังหวัดเชียงใหม่ (ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)
☐ (1) งานเทศกาลต่างๆ ☐(2) งานแสดงสด
☐ (3) กิจกรรมทางวัฒนธรรม
☐ (4) งานนิทรรศการ ☐(5) นันทนาการกลางแจ้ง
☐(6) ปั่นจักรยาน ☐(7) ล่องแพ ☐ (8) เล่นกอล์ฟ
☐(9) น้ำพุร้อน ☐(10) มวยไทย ☐(11) ช้อปปิ้ง
☐(12) เดินเขา ปีนเขา ☐(13) เยี่ยมชมสถานที่ทางประวัติศาสตร์
☐(14) ใช้บริการสถานที่นวดสุขภาพ สปาห้องอบเซาน่า
☐(15) เยี่ยมชมตลาด ถนนคนเดิน
☐(16) เยี่ยมชมสถานที่บันเทิง ไนท์คลับ บาร์
☐(17) เยี่ยมชมหอศิลป์ ☐(18) เยี่ยมชมภูเขา
☐(19) เยี่ยมชมอุทยานแห่งชาติ / ป่า
☐(20) เยี่ยมชมพิพิธภัณฑ์ ☐(21) ร้านอาหาร /รับประทานอาหารนอกบ้าน
☐(22) เที่ยวชมในตัวเมือง
☐(23) การดูแลสุขภาพ ☐(24) การทำอาหารไทย
☐(25) การสังเกตสัตว์ป่า
☐(26) การชิมอาหารท้องถิ่น
☐(27) อื่นๆ____________

D5. อะไรทำให้ท่านสนใจจะมาเข้าร่วมกิจกรรม(ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)
☐ (1) กิจกรรมเพื่อความบันเทิง
☐ (2) เยี่ยมชมประเพณีและวัฒนธรรม
☐ (3) กิจกรรมกลางแจ้ง
☐ (4) การทำอาหารไทย
☐ (5) สปา และนวดแผนโบราณ
☐ (6) กิจกรรมท้องถิ่น
☐ (7) กิจกรรมเกี่ยวกับธรรมชาติ

D6. สิ่งไหนที่ท่านประทับใจมากที่สุดในจังหวัดเชียงใหม่ (ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)
☐ (1) วัด ☐ (2) อาหารไทย ☐ (3) ชีวิตกลางคืน
☐ (4) หอศิลป์ งานแสดงศิลปะ ☐ (5) ตลาดตอนกลางคืน ถนนคนเดิน
☐ (6) นวดแผนโบราณ สปา ☐ (7) ชมสัตว์ป่า
☐ (8) ปั่นจักรยาน ☐ (9) เล่นกอล์ฟ ☐ (10) น้ำพุร้อน
☐ (11) ช้อปปิ้ง ☐ (12) มวยไทย
☐ (13) ธรรมชาติ ☐ (14) พิพิธภัณฑ์

D7. สถานที่พักแบบไหนที่ท่านวางแผนไว้ที่จะพักอาศัยอยู่ในจังหวัดเชียงใหม่
☐(1) โรงแรม ☐(2) เกสต์เฮาส์ ☐(3) หอพักเยาวชน
☐(4) บ้านญาติ หรือบ้านเพื่อน
☐(5) หอพัก ☐(6) อพาร์ตเมนต์ให้เช่า☐ (7) วัด
☐(8) อื่นๆ ____________________

D8. โปรดประเมินราคาค่าอาหารโดยรวมของร้านอาหารที่ท่านวางแผนจะไปใช้บริการระหว่างอยู่ท่องเที่ยวที่ในจังหวัดเชียงใหม่
(1-ราคาแพงน้อยที่สุด, 5- ราคาแพงมากที่สุด)

แบบสำรวจเลขที่______ สถานที่สำรวจ______________ วันที่___เดือน____พ.ศ.2557

1☐ 2☐ 3☐ 4☐ 5☐

D9. โปรดประเมินราคาค่าห้องพักโดยรวม ของที่พักที่ท่านวางแผนจะไปใช้บริการระหว่างท่องเที่ยวอยู่ในจังหวัดเชียงใหม่

(1=ราคาแพงน้อยที่สุด, 5=ราคาแพงมากที่สุด)

1☐ 2☐ 3☐ 4☐ 5☐

D10. ท่านวางแผนจะใช้ประเภทการเดินทางแบบใด ระหว่างที่ท่านเที่ยวในจังหวัดเชียงใหม่(ท่านสามารถเลือกได้มากกว่าหนึ่งข้อ)

☐(1) เดิน ☐(2) จักรยาน ☐(3) รถเช่า

☐(4) รถสี่ล้อแดง/รถสองแถว ☐(5) แท็กซี่ ☐(6) รถเมล์

☐(7) ยานพาหนะส่วนตัว (รถยนต์ รถตู้ รถทัวร์ หรือ รถมอเตอร์ไซค์)

E. ปัจจัยสำคัญเกี่ยวกับแรงจูงใจของนักท่องเที่ยว

E1. โปรดประเมินสิ่งจูงใจในการมาท่องเที่ยวที่จังหวัดเชียงใหม่

(1=เห็นด้วยน้อยที่สุด, 5=เห็นด้วยมากที่สุด)

E1.1 ด้านการค้นหาตัวเอง:

(1) เพื่อต้องการเข้าใจตัวเองให้มากขึ้น

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อเพิ่มมุมมองใหม่ให้กับชีวิต

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อพัฒนาคุณค่าทางจิตใจให้กับตัวเอง

1☐ 2☐ 3☐ 4☐ 5☐

(4) เพื่อแสวงหาชีวิตที่ดีกว่า

1☐ 2☐ 3☐ 4☐ 5☐

E1.2 ด้านการพักผ่อน / การหลีกเลี่ยง

(1) เพื่อความสงบภายในจิตใจ

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อเผชิญกับความสันโดษและความสงบ

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อทำให้จิตและกายสดชื่น ทำให้จิตและกายมีพลังอีก

1☐ 2☐ 3☐ 4☐ 5☐

(4) เพื่อทำให้ตัวเองมีชีวิตชีวาอีก

1☐ 2☐ 3☐ 4☐ 5☐

(5) เพื่อไม่ต้องกังวลเกี่ยวกับเวลา และ งาน

1☐ 2☐ 3☐ 4☐ 5☐

E1.3 ด้านความแปลกใหม่:

(1) เพื่อความสนุกสนาน

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อประสบการณ์ที่แปลกใหม่กว่าเดิม

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อสัมผัสและรู้สึกบรรยากาศที่พิเศษของสถานที่

1☐ 2☐ 3☐ 4☐ 5☐

(4) เพื่อไปสถานที่ๆชื่นชอบ

1☐ 2☐ 3☐ 4☐ 5☐

E1.4 ด้านการผจญภัย:

(1) เพื่อความตื่นเต้น เร้าใจ

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อหาประสบการณ์เสี่ยงภัย

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อหาประสบการณ์อันตรายและหวาดเสียว

1☐ 2☐ 3☐ 4☐ 5☐

(4) เพื่อเยี่ยมชมสถานที่ๆไม่เคยมา

1☐ 2☐ 3☐ 4☐ 5☐

E1.5 ด้านการเรียนรู้:

(1) เพื่อเจอผู้คนใหม่ ๆสถานที่ใหม่ๆ และ วิถีใหม่ๆ

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อเรียนรู้เกี่ยวกับวัฒนธรรมและประวัติศาสตร์

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อเรียนทักษะใหม่ๆให้กับตัวเอง

1☐ 2☐ 3☐ 4☐ 5☐

(4) เพื่อเรียนรู้เกี่ยวกับอาหารไทย

1☐ 2☐ 3☐ 4☐ 5☐

(5) เพื่อเรียนรู้เกี่ยวกับธรรมชาติ

1☐ 2☐ 3☐ 4☐ 5☐

E1.6 ด้านความสัมพันธ์:

(1) เพื่อหาอะไรทำกับเพื่อนกับหรือครอบครัว

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อหาอะไรทำกับเพื่อนร่วมงาน

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อเพิ่มความสัมพันธ์ที่ดีกับครอบครัว หรือ เพื่อน

1☐ 2☐ 3☐ 4☐ 5☐

(4) เพื่อเยี่ยมญาติและเพื่อน

1☐ 2☐ 3☐ 4☐ 5☐

E1.7 ด้านสถานะทางสังคม:

(1) เพื่อไปสถานที่ทำให้เพื่อนให้เพื่อนหรือครอบครัวประทับใจ

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อแบ่งปันสิ่งที่ได้รู้ได้เห็นกับผู้อื่น

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อแสดงให้เป็นความรู้สึก ความคิด หรือ ความสามารถให้ผู้อื่นเห็น

1☐ 2☐ 3☐ 4☐ 5☐

E1.8 ด้านความรัก:

(1) เพื่อทำให้ชีวิตรักดีขึ้น

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อได้สัมผัสกับจินตนาการของการเดินทาง

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อทบทวนความทรงจำเก่าๆ

1☐ 2☐ 3☐ 4☐ 5☐

(4) เพื่ออยู่กับเพศตรงกันข้าม

1☐ 2☐ 3☐ 4☐ 5☐

แบบสำรวจเลขที่________ สถานที่สำรวจ________________ วันที่___เดือน____พ.ศ.2557

E1.9 ด้านการจับจ่ายใช้สอย

(1) เพื่อช้อปปิ้ง

1☐ 2☐ 3☐ 4☐ 5☐

(2) เพื่อซื้อสินค้าท้องถิ่น

1☐ 2☐ 3☐ 4☐ 5☐

(3) เพื่อซื้อสินค้าแบรนด์เนม

1☐ 2☐ 3☐ 4☐ 5☐

F. ความพึงพอใจของนักท่องเที่ยว

F1. โปรดให้คะแนนระดับความสำคัญและความพึงพอใจสำหรับสถานที่ที่ท่านกำลังเยี่ยมชมอยู่ขณะนี้ในจังหวัดเชียงใหม่

(1-สำคัญน้อยที่สุด, 5-สำคัญมากที่สุด)

(1-พึงพอใจน้อยที่สุด, 5-พึงพอใจมากที่สุด)

F1.1 ด้านราคา:

(1) ราคาของสินค้าที่ขายและบริการทั่วไปของสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(2) สินค้าบริการมาตรฐาน ราคายุติธรรมของสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

F1.2 ด้านการต้อนรับ:

(1) อัธยาศัยของผู้คนทั่วไป

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(2) อัธยาศัยของพนักงานหรือเจ้าหน้าที่ในสถานที่ท่องเที่ยวนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(3) ความตั้งใจของพนักงานของเพื่ออำนวยความสะดวกและให้ความช่วยเหลือนักท่องเที่ยว

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

F1.3 ด้านอาหารและเครื่องดื่ม

(1) ความหลากหลายของอาหารและเครื่องดื่มจากสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(2) จำนวนที่นั่งรับประทานและดื่มของสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(3) คุณภาพการให้บริการของสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(4) คุณภาพอาหารและเครื่องดื่มของสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

F1.4 ด้านสถานที่

(1) ความสะอาดของสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(2) ความปลอดภัยโดยรวม

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(3) สภาพภูมิอากาศของสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(4) ความสวยงามของทิวทัศน์จากสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(5) ความหลากหลายของกิจกรรมจากสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(6) การได้รับวัฒนธรรมและประวัติศาสตร์จากสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

F1.5 ด้านการเข้าถึง:

(1) ข้อมูลการท่องเที่ยวในสถานที่นี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(2) การอำนวยความสะดวกสำหรับที่จอดรถของสถานที่แห่งนี้

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(3) ความสะดวกสบายในการใช้ระบบขนส่งสาธารณะ

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

(4) ความปลอดภัยและประสิทธิภาพของระบบขนส่งสาธารณะ

ระดับความสำคัญ 1☐ 2☐ 3☐ 4☐ 5☐

ระดับความพึงพอใจ 1☐ 2☐ 3☐ 4☐ 5☐

G. ข้อมูลทั่วไปส่วนบุคคล

G1. เพศ: ☐(1) ชาย ☐(2) หญิง

G2. อายุ(ปี): ☐(1)18-25 ☐(2) 26-35
☐(3) 36-45 ☐(4) 46-55 ☐(5) 56-65
☐(6) 66 หรือมากกว่า

G3. สถานภาพสมรส ☐(1) โสด ☐(2) แต่งงาน
☐(3) แยกกันอยู่ ☐(4) แม่ม่าย ☐(5) หย่า

G4. การศึกษาสูงสุดของท่าน
☐(1) ระดับประถมศึกษา
☐(2) ระดับมัธยมศึกษา
☐(3) ระดับ ปวช ปวส หรือ วิทยาลัย

แบบสำรวจเลขที่______ สถานที่สำรวจ________________ วันที่___เดือน____พ.ศ.2557

☐(4) ระดับปริญญาตรี
☐(5) ระดับปริญญาโท
☐(6) ระดับปริญญาเอก
☐(7) อื่นๆ____________________

G5. รายได้รวมปัจจุบันต่อปีของครอบครัวท่าน
☐(1) น้อยกว่า 90,000 บาท ☐(2) 90,001 บาท ถึง 150,000 บาท
☐(3) 150,001 บาท ถึง 300,000 บาท
☐(4) 300,001 บาท ถึง 450,000 บาท
☐(5) 450,001 บาท ถึง 900,000 บาท
☐(6) 900,001 บาท ถึง 1,800,000 บาท
☐(7) 1,800,001 บาท ถึง 2,700,000 บาท ☐(8) มากกว่า 2,700,000 บาท

G6. อาชีพใดในข้อต่อไปนี้บ่งบอกถึงลักษณะงานของท่านได้ดีที่สุด
☐(1) พนักงานบริษัท
☐(2) ประกอบธุรกิจส่วนตัว
☐(3) ทำงานด้วยตนเอง
☐(4) ข้าราชการ
☐(5) ผู้เชี่ยวชาญในด้านต่างๆ / นักวิจัย
☐(6) พ่อบ้าน แม่บ้าน
☐(7) นักเรียน
☐(8) เกษียณ
☐(9) ว่างงาน
☐(10) อื่นๆ____________________

G7. สัญชาติของท่านคืออะไร
☐(1) จีน ☐(2) ลาว ☐(3) มาเลเซีย
☐(4) สิงคโปร์ ☐(5) เกาหลี ☐(6) ญี่ปุ่น
☐(7) สหรัฐอเมริกา ☐(8) สหราชอาณาจักร
☐(9) ฝรั่งเศส ☐(10) เยอรมัน ☐(11) รัสเซีย
☐(12) สวีเดน ☐(13) อินเดีย ☐(14) ออสเตรเลีย หรือ นิวซีแลนด์
☐(15) ไทย ☐(16) อื่น____________________

G8. ประเทศที่พำนักของท่านคือที่ใด
☐(1) จีน ☐(2) ลาว ☐(3) มาเลเซีย
☐(4) สิงคโปร์ ☐(5) เกาหลี ☐(6) ญี่ปุ่น
☐(7) สหรัฐอเมริกา ☐(8) สหราชอาณาจักร
☐(9) ฝรั่งเศส ☐(10) เยอรมัน ☐(11) รัสเซีย
☐(12) สวีเดน ☐(13) อินเดีย ☐(14) ออสเตรเลีย หรือ นิวซีแลนด์
☐(15) ไทย ☐(16) อื่น____________________

G9. ภูมิลำเนาของท่านคือที่ใด
(สำหรับนักท่องเที่ยวต่างชาติ):
☐(1) เอเชียตะวันออก
☐(2) ยุโรป
☐(3) อเมริกา
☐(4) เอเชียตะวันออกเฉียงใต้
☐(5) หมู่เกาะทางแปซิฟิก
☐(6) ตะวันออกกลาง
☐(7) แอฟริกา

(สำหรับนักท่องเที่ยวไทย หรือ ท้องถิ่น):
☐(8) *ภาคกลาง*
☐(9) *ภาคอีสาน*
☐(10) *ภาคเหนือ*
☐(11) *ภาคใต้*

H. ความเห็นเพิ่มเติม หรือ คำแนะนำ:

____
____
____
____
____
____
____

ขอบพระคุณท่านที่ให้ความร่วมมือในการตอบแบบสอบถาม คำตอบของท่านจะมีคุณค่ากับการเลือกสถานที่ท่องเที่ยวในจังหวัดเชียงใหม่ ถ้าท่านมีคำถามและข้อสงสัยประการใด โปรดติดต่อ คุณปรีดิ์ เพียรบูรณธรรม ได้ที่ prcc.t@cmu.ac.th หรือ โทร. 09-46038726 ขอให้ท่านมีความสุขกับการท่องเที่ยวในจังหวัดเชียงใหม่

แบบสำรวจเลขที่________ สถานที่________ วันที่__เดือน______พ.ศ.2557

游客对旅游目的地满意程度问卷调查表

此份调查表是为了清迈大学和英国伯恩茅斯大学关于电子旅游(e-tourism)的合作项目的博士级论文收集信息而制定.

问卷调查的结果将会为本团队对游客选择清迈作为旅游地的原因的研究起到重要作用,我们团队对各位愿意花 10-15 分钟为我们填写这份表格表示衷心的感谢, 此次调查本着自愿的原则对游客的本次旅游情况做一些相关调查,调查者不用担心对您的旅游有任何影响也不用对你的任何答案负责.

A.旅游性质

A1.您计划在清迈停留几晚
□(1) 0-1 晚 □(2) 2-3 晚□(3) 4-7 晚
□(4) 8-14 晚□(5) 15-30 晚□(6) 30 晚以上

A2. 此次旅游有人随行吗
□(1) 有 (继续做 A3.题)
□(2) 没有 (请直接做 A4.题)

A3. 此次清迈之旅您与谁同行(**多项选择**)
□(1) 朋友 □(2) 家长 □(3) 情侣
□(4) 亲戚 □(5) 儿女 □(6) 同事

A4. 在过去的 5 年内,您来过清迈几次(**包含本次**)
□(1) 第 1 次 □(2) 2-3 次
□(3) 4-7 次 □(4) 8-20 次 □(5) 21 次以上

A5. 你本次旅游的性质有 (**多项选择**)
□(1) 探险式旅游, 户外探险及景点游览
□(2) 活动式旅游, 参加各种各样的旅游活动
□(3) 放松式旅游 如:娱乐场所 参加休闲活动和景点游览
□(4) 文青式旅游 对艺术 文化及当地风情的旅游

B. 旅游计划

B1. 您是如何安排本次清迈旅游
□(1) 团体旅游 通过旅行社的安排和其他游客一起旅游
□(2) 自由团体(家人或朋友), 同样通过旅行社安排旅游
□(3) 只是通过旅行社订购机票和预定房间
□(4) 借助清迈本地旅行社自行安排旅游
□(5) 完全自行,不借助任何旅行社

B2. 本次到清迈的目的是
□(1) 公务(业务访问)
□(2) 假日
□(3) 探亲访友
□(4) 宗教活动
□(5)会议,研讨会或展览
□(6) 卫生健康方面
□(7) 购物
□(8) 其他________________

B3.何原因使您对来清迈旅游感兴趣 (**多项选择**)
□(1) 亲朋好友介绍□(2) 互联网 □(3) 旅游公司
□(4) 旅游杂志 □(5) 电视 或 电台
□(5) 自身经验
□(6) 其他________________

C. 支付方式

C1. 到清迈前已经支付的项目
□ (1) 酒店费用 □ (2) 酒店订餐
□ (3) 区域间交通费 □ (4) 娱乐项目费 □ (5) 杂项开支

C2. 本次旅行您的预计开支
□(1) 5,000 铢以下 □(2) 5,001 -15,000 铢
□(3) 15,001 -50,000 铢 □(4) 50,001-150,000 铢
□(5) 150,001 – 500,000 铢□ (6) 500,000 铢

C3. 您本次清迈旅游中在以下的项目的预计花费
(1)交通费 大约__________铢
(2) 住宿费 大约__________铢
(3) 饮食费 大约__________铢
(4) 纪念方面 大约__________铢
(5) 娱乐方面 大约__________铢
(6) 购物方面 大约__________铢
(7) 其他 大约__________铢

D. 游客方面

D1. 请选择本次清迈之行您计划去的景点(多项选择)
□(1) 柴迪隆寺
□(2)清迈人妖歌舞表演
□(3) 素贴山双龙寺
□(4) 昆虫和自然景观博物馆
□(5) 3D 画像馆
□(6) 因它暖山
□(7) 艺术文化展
□(8) 帕邢寺
□(9) 对康寺
□(10) 乌孟寺
□(11) 希苏娑寺
□(12) 洛莫利寺

ผู้เก็บ________

แบบสำรวจเลขที่________ สถานที่________ วันที่__เดือน______พ.ศ.2557

☐(13) 松达寺
☐(14) 潘岛寺
☐(15) 清曼寺
☐(16)清迈亚洲图片和摄影展
☐(17) 搏击泰拳馆
☐(18) 播东瀑布
☐(19) 惠登湾公园
☐(20) 眉沙瀑布
☐(21) 其他________________

D2. 本次出行您最喜欢的是(**多项选择**)
☐(1) 历史、 文化氛围
☐(2) 各类表演
☐(3) 自然风情
☐(4) 教育、体育
☐(5) 博物馆、艺术中心

D3. 本次清迈行您打算参加的项目有(**多项选择**)
☐(1) 各类节日 ☐(2) 表演
☐(3)文化活动
☐(4) 展览 ☐(5) 户外娱乐
☐(6) 骑自行车☐(7) 漂流 ☐(8)高尔夫
☐(9) 温泉 ☐(10)泰拳☐(11) 购物
☐(12) 爬山 ☐(13) 游览名胜古迹
☐(14) 泰式按摩 桑拿 ☐(15) 娱乐场所
☐(16) 参观美术展☐(17) 游山外水
☐(18) 参观国家公园/森林☐(19) 参观博物馆
☐(20) 餐馆各类美食 ☐(21) 古城游览
☐(22) 养身 ☐(23) 学做泰国菜
☐(24) 观察野生动物
☐(25) 品尝地方美食
☐(26)其他____________

D4. 你感兴趣的活动有(**多项选择**)
☐(1) 娱乐类的活动
☐(2) 参观当地文化风俗的活动
☐(3) 户外活动
☐(4) 做泰国菜
☐(5) SPA 泰式按摩
☐(6) 地方性活动
☐(7) 亲近自然

D5. 您对清迈最满意的是(**多项选择**)
☐(1) 寺庙 ☐(2) 泰国菜 ☐(3) 夜生活
☐(4) 画展 艺术展 ☐(5) 夜市
☐(6) 泰式按摩 SPA ☐(7) 野生动物
☐(8)骑自行车 ☐(9) 打高尔夫 ☐(10) 温泉
☐(11) 购物☐(12) 泰拳
☐(13) 自然风景 ☐(14) 博物馆

D6. 本次清迈行你打算住哪类居所
☐(1) 酒店☐(2)青年旅社☐(3) 民居
☐(4) 亲朋好友家
☐(5) 宿舍 ☐(6) 出租公寓 ☐(7) 寺庙
☐(8) 其他 ________________

D7. 请对清迈的饮食价格打分
(1-最便宜, 5- 最贵)
1☐ 2☐ 3☐ 4☐ 5☐

D8. 请对清迈的居住情况打分
(1-最便宜, 5-最贵)
1☐ 2☐ 3☐ 4☐ 5☐

D9. 在清迈旅行期间您选择哪种交通出行(**多项选择**)
☐(1) 步行 ☐(2) 自行车☐(3) 租车
☐(4) 红车/双条车☐(5) 出租车☐(6) 公车
☐(7) 自驾(汽车、面包车、旅行大巴、摩托车)

E. 吸引游客赴清迈旅游的重要因素
E1. 请对吸引游客来清迈旅游的各项因素打分
(1-最不赞同 5-十分赞同)

E1.1 对自身方面:
(1) 为了更清楚的认识自己
1☐ 2☐ 3☐ 4☐ 5☐
(2)为了发现生活新视角
1☐ 2☐ 3☐ 4☐ 5☐
(3) 提高自我身心修养
1☐ 2☐ 3☐ 4☐ 5☐
(4) 为了寻找更好的生活
1☐ 2☐ 3☐ 4☐ 5☐

E1.2 休息休闲
(1) 为了心灵的平静
1☐ 2☐ 3☐ 4☐ 5☐
(2) 为了享受自由和宁静
1☐ 2☐ 3☐ 4☐ 5☐
(3) 为了身心健康舒适
1☐ 2☐ 3☐ 4☐ 5☐
(4) 为了使自己充满活力
1☐ 2☐ 3☐ 4☐ 5☐
(5) 为了从工作和时间的压力中释放
1☐ 2☐ 3☐ 4☐ 5☐

E1.3 新鲜事物方面

ผู้เก็บ________

แบบสำรวจเลขที่________ สถานที่________ วันที่__เดือน______พ.ศ.2557

(1) 为了寻求快乐
1☐ 2☐ 3☐ 4☐ 5☐
(2) 为了体验新鲜事物
1☐ 2☐ 3☐ 4☐ 5☐
(3) 为了体验和认识景区的新鲜环境
1☐ 2☐ 3☐ 4☐ 5☐
(4) 为了去喜欢的景点
1☐ 2☐ 3☐ 4☐ 5☐

E1.4 冒险方面
(1)为了体验兴奋和刺激
1☐ 2☐ 3☐ 4☐ 5☐
(2)为了得到冒险的经历
1☐ 2☐ 3☐ 4☐ 5☐
(3) 为了得到刺激冒险感受的经验
1☐ 2☐ 3☐ 4☐ 5☐

(4) 为了参观没去过的地方
1☐ 2☐ 3☐ 4☐ 5☐

E1.5 学习方面
(1) 为了遇到不同的人 不同的景点 和不同的事物
1☐ 2☐ 3☐ 4☐ 5☐
(2) 为了得到历史文化方面的知识
1☐ 2☐ 3☐ 4☐ 5☐
(3) 为了学会一些新的技能
1☐ 2☐ 3☐ 4☐ 5☐
(4) 为了学习泰国菜
1☐ 2☐ 3☐ 4☐ 5☐
(5) 为了学习自然知识
1☐ 2☐ 3☐ 4☐ 5☐

E1.6 关系方面
(1) 为了和朋友或家人有更多接触
1☐ 2☐ 3☐ 4☐ 5☐
(2)为了和同事有跟多接触
1☐ 2☐ 3☐ 4☐ 5☐
(3) 为了和家人或朋友增进感情
1☐ 2☐ 3☐ 4☐ 5☐
(4) 为了探访亲友
1☐ 2☐ 3☐ 4☐ 5☐

E1.7 社会角色方面
(1) 为了去一个地方使家人和朋友都满意
1☐ 2☐ 3☐ 4☐ 5☐
(2) 为了和朋友分享所见所闻
1☐ 2☐ 3☐ 4☐ 5☐
(3) 为了向别人展示自己的体验 想法和能力
1☐ 2☐ 3☐ 4☐ 5☐

E1.8 爱情方面
(1) 为了让生活充满爱
1☐ 2☐ 3☐ 4☐ 5☐
(2)1 为了体验旅游的无限想象
1☐ 2☐ 3☐ 4☐ 5☐
(3) 为了召回曾经的回忆
1☐ 2☐ 3☐ 4☐ 5☐

(4) 为了和异性在一起
1☐ 2☐ 3☐ 4☐ 5☐

E1.9 支付方面
(1) 为了购物
1☐ 2☐ 3☐ 4☐ 5☐

(2) 为了购买当地特产
1☐ 2☐ 3☐ 4☐ 5☐

(3)为了购买名牌
1☐ 2☐ 3☐ 4☐ 5☐

F. 游客的满意程度
F1. 请对清迈旅游各方面的重要程度和您的满意程度打分
(1-最不重要 5-最重要)
(1-最不满意 5-最满意)

F1.1 价钱:
(1) 清迈物价和各类服务费
重要程度 1☐ 2☐ 3☐ 4☐ 5☐
满意程度 1☐ 2☐ 3☐ 4☐ 5☐
(2) 景区的服务水准和是否明码标价
重要程度 1☐ 2☐ 3☐ 4☐ 5☐
满意程度 1☐ 2☐ 3☐ 4☐ 5☐

F1.2 接待方面:
(1) 人民的热情程度
重要程度 1☐ 2☐ 3☐ 4☐ 5☐
满意程度 1☐ 2☐ 3☐ 4☐ 5☐
(2) 清迈景点的员工的热情程度
重要程度 1☐ 2☐ 3☐ 4☐ 5☐
满意程度 1☐ 2☐ 3☐ 4☐ 5☐
(3) 景区员工是否为游客提供方便和帮助
重要程度 1☐ 2☐ 3☐ 4☐ 5☐
满意程度 1☐ 2☐ 3☐ 4☐ 5☐

F1.3 饮食方面
(1) 景区饮食的多样性
重要程度 1☐ 2☐ 3☐ 4☐ 5☐
满意程度 1☐ 2☐ 3☐ 4☐ 5☐

ผู้เก็บ________

แบบสำรวจเลขที่________ สถานที่________ วันที่__เดือน______พ.ศ.2557

(2) 景区座位是否够用
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(3) 景区服务质量
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(4) 景区饮食的质量
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐

F1.4 景区方面
(1) 景区的清洁程度
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(2) 景区的安全程度
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(3) 景区的气候
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(4) 景区的风景
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(5) 景区活动的多样性
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐

(6) 景区的文化和历史氛围
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐

F1.5 景区交通方面
(1)旅游信息方面
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(2) 景区停车是否方便
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(3) 公共交通运输的方便快捷方面
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐
(4) 交通安全和效率方面
重要程度 1☐2☐3☐4☐5☐
满意程度 1☐2☐3☐4☐5☐

G. 游客信息

G1. 性别: ☐(1) 男 ☐(2)女

G2. 年龄: ☐(1)18-25 ☐(2) 26-35
☐(3) 36-45 ☐(4) 46-55 ☐(5) 56-65
☐(6) 65 以上

G3. 婚姻状况 ☐(1) 单身 ☐(2) 已婚
☐(3) 分居 ☐(4) 丧偶 ☐(5) 离异

G4. 最高学历
☐(1) 小学
☐(2) 中学
☐(3) 专科
☐(4) 本科
☐(5) 硕士
☐(6) 博士
☐(7)其他________________

G5. 目前年收入
☐(1) 90,000 铢 ☐(2) 90,001 铢到 150,000 铢
☐(3) 150,001 铢到 300,000 铢
☐(4) 300,001 铢到 450,000 铢
☐(5) 450,001 铢 到 900,000 铢
☐(6) 900,001 铢 到 1,500,000 铢
☐(7) 1,500,001 铢 到 2,700,000 铢 ☐(8) 2,700,000 铢以上

G6. 您的职业范围
☐(1) 公司职员
☐(2) 私人企业
☐(3) 自由职业
☐(4) 公务员
☐(5) 学者 专家
☐(6) 家庭主妇（夫）
☐(7) 学生
☐(8)退休
☐(9) 其他________________

G7. 您的国籍
☐(1) 中国 ☐(2) 老挝 ☐(3) 马来西亚
☐(4) 新加坡 ☐(5) 韩国 ☐(6) 日本
☐(7) 美国 ☐(8) 英国
☐(9) 法国 ☐(10) 德国 ☐(11) 俄罗斯
☐(12) 瑞士 ☐(13)印度 ☐(14) 澳大利亚 新西兰
☐(15) 其他________________

G8. 您现居的国家是
☐(1)中国 ☐(2) 老挝 ☐(3) 马来西亚
☐(4)新加坡 ☐(5)韩国 ☐(6) 日本
☐(7) 美国 ☐(8) 英国
☐(9) 法国 ☐(10) 德国 ☐(11) 俄罗斯
☐(12)瑞士 ☐(13) 印度 ☐(14) 澳大利亚 新西兰
☐(15) 其他________________

ผู้เก็บ________

แบบสำรวจเลขที่________ สถานที่________ วันที่__เดือน______พ.ศ.2557

๑๙. 你所处的版块

(针对外国游客):

☐(1)东亚

☐(2) 欧洲

☐(3) 美洲

☐(4) 东南亚

☐(5) 太平洋群岛

☐(6) 中东

☐(7) 非洲

(针对泰国游客):

☐(8) 中部

☐(9) 东北部

☐(10) 北部

☐(11) 南部

๒๐. 其他意见和建议

____

____

____

____

____

____

十分感谢您花时间对这份清迈旅游问卷做出详细的填写，您的答案对我们理解游客为什么选择到清迈旅游十分宝贵，如果您有什么问题和疑问，欢迎致电 09-46056726 或电邮：prcc.t@cmu.ac.th [illegible] 女士，祝您在清迈旅途愉快！ 谢谢！

ผู้เก็บ________

Appendix B

List of variable names and labels for the data sets that we collected:

Variable Information

Variable	Label
id	Survey identification number
a1	How many night(s) do you plan to stay in Chiang Mai?
a2	Is anyone accompanying you on this trip?
a3_1	The people who are accompanying you are: (Please tick all that apply) (1) Friends
a3_2	The people who are accompanying you are: (Please tick all that apply) (2) Parents
a3_3	The people who are accompanying you are: (Please tick all that apply) (3) Spouse
a3_4	The people who are accompanying you are: (Please tick all that apply) (4) Relatives
a3_5	The people who are accompanying you are: (Please tick all that apply) (5) With children
a3_6	The people who are accompanying you are: (Please tick all that apply) (6) Colleagues
a4	How many times have you visited Chiang Mai (Including this trip) in the last five years?
a5_1	Which of the following define your travel style? (Please tick all that apply) (1) Adventurer prefers outdoor adventure and general sightseeing
a5_2	Which of the following define your travel style? (Please tick all that apply) (2) Multiple interests prefers diverse activities
a5_3	Which of the following define your travel style? (Please tick all that apply) (3) Relaxation seeker prefers amusement relaxation and general sightseeing activities
a5_4	Which of the following define your travel style? (Please tick all that apply) (4) Cultural prefers performing arts and local events
	(1) Group tour arranged through a travel agency. (Join a tour group)
b1	How did you arrange this trip to Chiang Mai?
b2	What is the purpose of taking this trip to Chiang Mai?
b3_1	Which of the following information sources influences your decision to visit Chiang Mai? (Please tick all that apply) (1) Friend or relative recommendation
b3_2	Which of the following information sources influences your decision to visit Chiang Mai? (Please tick all that apply) (2) The Internet
b3_3	Which of the following information sources influences your decision to visit Chiang Mai? (Please tick all that apply) (3) Travel agency
b3_4	Which of the following information sources influences your decision to visit Chiang Mai? (Please tick all that apply) (4) Books and guides
b3_5	Which of the following information sources influences your decision to visit Chiang Mai? (Please tick all that apply) (5) TV radio
b3_6	Which of the following information sources influences your decision to visit Chiang Mai? (Please tick all that apply) (6) Personal Experience
b3_7	Which of the following information sources influences your decision to visit Chiang Mai? (Please tick all that apply) (7) Other
c1_1	Which of the following expenses have you prepaid before arrival in Chiang Mai? (Please tick all that apply) (1) Hotel
c1_2	Which of the following expenses have you prepaid before arrival in Chiang Mai? (Please tick all that apply) (2) Meals outside hotel
c1_3	Which of the following expenses have you prepaid before arrival in Chiang Mai? (Please tick all that apply) (3) Local transportation
c1_4	Which of the following expenses have you prepaid before arrival in Chiang Mai? (Please tick all that apply) (4) Entertainment
c1_5	Which of the following expenses have you prepaid before arrival in Chiang Mai? (Please tick all that apply) (5) Miscellaneous expenses
c2	How much money do you plan to spend on this trip (in US dollars)? (1 dollar equals 32 Thai baht)

c3_1	How much money do you plan to spend per person on the following items during your stay in Chiang Mai (in US dollars)? (1 dollar equals 30 Thai baht) (1) Transportation
c3_2	How much money do you plan to spend per person on the following items during your stay in Chiang Mai (in US dollars)? (1 dollar equals 30 Thai baht) (2) Accommodation
c3_3	How much money do you plan to spend per person on the following items during your stay in Chiang Mai (in US dollars)? (1 dollar equals 30 Thai baht) (3) Restaurants cafes
c3_4	How much money do you plan to spend per person on the following items during your stay in Chiang Mai (in US dollars)? (1 dollar equals 30 Thai baht) (4) Souvenirs
c3_5	How much money do you plan to spend per person on the following items during your stay in Chiang Mai (in US dollars)? (1 dollar equals 30 Thai baht) (5) Entertainment
c3_6	How much money do you plan to spend per person on the following items during your stay in Chiang Mai (in US dollars)? (1 dollar equals 30 Thai baht) (6) Shopping
c3_7	How much money do you plan to spend per person on the following items during your stay in Chiang Mai (in US dollars)? (1 dollar equals 30 Thai baht) (7) Other expenses
d2	Which one of the followings is your most favorite attraction which you plan to visit during your stay in Chiang Mai?
d3_1	Which one of the followings is your favorite for this trip? (Please tick all that apply) (1) Cultural and historical
d3_2	Which one of the followings is your favorite for this trip? (Please tick all that apply) (2) Performances
d3_3	Which one of the followings is your favorite for this trip? (Please tick all that apply) (3) Natural scenery and landscape
d3_4	Which one of the followings is your favorite for this trip? (Please tick all that apply) (4) Educational and sport site
d3_5	Which one of the followings is your favorite for this trip? (Please tick all that apply) (5) Museums and art galleries
d4_1	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (1) Attend festivals
d4_2	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (2) Attend performances
d4_3	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (3) Attend cultural events
d4_4	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (4) Attend exhibitions
d4_5	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (5) Outdoor recreation
d4_6	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (6) Biking
d4_7	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (7) Rafting
d4_8	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (8) Golfing
d4_9	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (9) Hot springs
d4_10	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (10) Thai Boxing
d4_11	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (11) Shopping
d4_12	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (12) Hiking/Climbing
d4_13	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (13) Visit historical places
d4_14	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (14) Visit markets walking streets
d4_15	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (15) Visit health spas massage sauna
d4_16	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (16) Visit entertainment places nightclubs bars
d4_17	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (17) Visit art galleries
d4_18	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (18) Visit mountains
d4_19	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (19) Visit national park/forests
d4_20	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (20) Visit museums
d4_21	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (21) Restaurants/dinning out
d4_22	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (22) Sightseeing in cities
d4_23	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (23) Health care
d4_24	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (24) Thai cooking
d4_25	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (25) Observing wildlife
d4_26	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (26) Sampling local food
d4_27	Please tick all the activities which you plan to participate in during your stay in Chiang Mai. (Please tick all that apply) (27) Other

d5_1	What makes you interested in plan to participate in the activities? (Please tick all that apply) (1) Entertainment activities
d5_2	What makes you interested in plan to participate in the activities? (Please tick all that apply) (2) Culture based sightseeing
d5_3	What makes you interested in plan to participate in the activities? (Please tick all that apply) (3) Outdoor activities
d5_4	What makes you interested in plan to participate in the activities? (Please tick all that apply) (4) Thai cuisine
d5_5	What makes you interested in plan to participate in the activities? (Please tick all that apply) (5) Thai spa and traditional message activities
d5_6	What makes you interested in plan to participate in the activities? (Please tick all that apply) (6) Local activities
d5_7	What makes you interested in plan to participate in the activities? (Please tick all that apply) (7) Nature based activities
d6_1	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (1) Temple
d6_2	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (2) Thai food
d6_3	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (3) Night life
d6_4	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (4) Art gallery
d6_5	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (5) Night market walking street
d6_6	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (6) Heath Massage spa
d6_7	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (7) Wildlife
d6_8	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (8) Biking
d6_9	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (9) Golfing
d6_10	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (10) Hot spring
d6_11	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (11) Shopping
d6_12	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (12) Thai Boxing
d6_13	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (13) Nature
d6_14	Which part of Chiang Mai has given you the deepest impression? (Please tick all that apply) (14) Museum
d7	What type of accommodation do you plan to use in Chiang Mai?
d8	Please rate the overall price that you plan to spend on your meal/food in Chiang Mai.
d9	Please rate the overall price that you plan to spend on accommodation in Chiang Mai.
d10_1	Please tick the transport modes that you plan to use during this trip in Chiang Mai. (Please tick all that apply) (1) Walk
d10_2	Please tick the transport modes that you plan to use during this trip in Chiang Mai. (Please tick all that apply) (2) Bicycle
d10_3	Please tick the transport modes that you plan to use during this trip in Chiang Mai. (Please tick all that apply) (3) Rental car
d10_4	Please tick the transport modes that you plan to use during this trip in Chiang Mai. (Please tick all that apply) (4) Shared Taxi (Rod dang/Red cab)
d10_5	Please tick the transport modes that you plan to use during this trip in Chiang Mai. (Please tick all that apply) (5) Taxi
d10_6	Please tick the transport modes that you plan to use during this trip in Chiang Mai. (Please tick all that apply) (6) Bus
d10_7	Please tick the transport modes that you plan to use during this trip in Chiang Mai. (Please tick all that apply) (7) Private car/motorcycle/van/coach
e1_1_1	How would you rate the following motives for this trip in Chiang Mai? E1.1 Self-actualize: (1) To understand more about myself
e1_1_2	How would you rate the following motives for this trip in Chiang Mai? E1.1 Self-actualize: (2) To gain a new perspective on life
e1_1_3	How would you rate the following motives for this trip in Chiang Mai? E1.1 Self-actualize: (3) To work on my personal/spiritual values
e1_1_4	How would you rate the following motives for this trip in Chiang Mai? E1.1 Self-actualize: (4) To seek a better existence
e1_2_1	How would you rate the following motives for this trip in Chiang Mai? E1.2 Escape/Relaxation: (1) To experience solitude and calm
e1_2_2	How would you rate the following motives for this trip in Chiang Mai? E1.2 Escape/Relaxation: (2) To experience inner harmony and peace
e1_2_3	How would you rate the following motives for this trip in Chiang Mai? E1.2 Escape/Relaxation: (3) To refresh mentally and physically
e1_2_4	How would you rate the following motives for this trip in Chiang Mai? E1.2 Escape/Relaxation: (4) To rejuvenate myself
e1_2_5	How would you rate the following motives for this trip in Chiang Mai? E1.2 Escape/Relaxation: (5) To not worry about time and work

e1_3_1	How would you rate the following motives for this trip in Chiang Mai? E1.3 Novelty: (1) To have fun
e1_3_2	How would you rate the following motives for this trip in Chiang Mai? E1.3 Novelty: (2) To experience something different
e1_3_3	How would you rate the following motives for this trip in Chiang Mai? E1.3 Novelty: (3) To feel the special atmosphere of the destination
e1_3_4	How would you rate the following motives for this trip in Chiang Mai? E1.3 Novelty: (4) To visit places related to my personal interests
e1_4_1	How would you rate the following motives for this trip in Chiang Mai? E1.4 Adventure: (1) To find excitement
e1_4_2	How would you rate the following motives for this trip in Chiang Mai? E1.4 Adventure: (2) To experience the risk involved
e1_4_3	How would you rate the following motives for this trip in Chiang Mai? E1.4 Adventure: (3) To experience danger and thrills
e1_4_4	How would you rate the following motives for this trip in Chiang Mai? E1.4 Adventure: (4) To visit places I have never been before
e1_5_1	How would you rate the following motives for this trip in Chiang Mai? E1.5 Learning experience: (1) To discover new people places and things
e1_5_2	How would you rate the following motives for this trip in Chiang Mai? E1.5 Learning experience: (2) To see famous cultural and historical sites
e1_5_3	How would you rate the following motives for this trip in Chiang Mai? E1.5 Learning experience: (3) To develop new abilities
e1_5_4	How would you rate the following motives for this trip in Chiang Mai? E1.5 Learning experience: (4) To learn about Thai cuisine
e1_5_5	How would you rate the following motives for this trip in Chiang Mai? E1.5 Learning experience: (5) To learn about nature
e1_6_1	How would you rate the following motives for this trip in Chiang Mai? E1.6 Relationship: (1) To do things with family and friend(s)
e1_6_2	How would you rate the following motives for this trip in Chiang Mai? E1.6 Relationship: (2) To do something with my companion(s)
e1_6_3	How would you rate the following motives for this trip in Chiang Mai? E1.6 Relationship: (3) To enhance relationships with friend(s)/family
e1_6_4	How would you rate the following motives for this trip in Chiang Mai? E1.6 Relationship: (4) To visit relatives/friend(s)
e1_7_1	How would you rate the following motives for this trip in Chiang Mai? E1.7 Social status: (1) To visit a destination that would impress my friends or family
e1_7_2	How would you rate the following motives for this trip in Chiang Mai? E1.7 Social status: (2) To share what I have learned with others
e1_7_3	E1. How would you rate the following motives for this trip in Chiang Mai? E1.7 Social status: (3) To reveal my thoughts feelings or physical skills to others
e1_8_1	E1. How would you rate the following motives for this trip in Chiang Mai? E1.8 Romance: (1) To improve my romantic life
e1_8_2	E1. How would you rate the following motives for this trip in Chiang Mai? E1.8 Romance: (2) To experience fantasy of travel
e1_8_3	E1. How would you rate the following motives for this trip in Chiang Mai? E1.8 Romance: (3) To reflect on past memories
e1_8_4	E1. How would you rate the following motives for this trip in Chiang Mai? E1.8 Romance: (4) To be with people of the opposite sex
e1_9_1	E1. How would you rate the following motives for this trip in Chiang Mai? E1.9 Shopping: (1) To go shopping
e1_9_2	E1. How would you rate the following motives for this trip in Chiang Mai? E1.9 Shopping: (2) To buy local Thai product
e1_9_3	E1. How would you rate the following motives for this trip in Chiang Mai? E1.9 Shopping: (3)To buy world famous brand-name products
g1	G1. Gender:
g2	G2. Age (years old):
g3	G3. Marital status:
g4	G4. Highest Education
g5	G5. What is your current household annual income in U.S. dollars ($)?
g6	G6. Which of the following categories best describes your primary area of employment (regardless of your actual position)?
g7	G7. What is your nationality?
g8	G8. What is the country of your residence?
g9	G9. Where is your origin?

Appendix C

The answer sheet used in the pilot study:

Pilot study for proposed personalizing recommendation system for tourists.
Answer sheet

Please write something about yourself.

Name:_____________________
Race:_____________________
Nationality/Region:_____________________
Gender:_____________________
Expertise:_____________________

Question 1:

____

____

Question 2:

____

____

Question 3:

____

____

Question 4:

Desktop, *deploy an application for a desktop computer that run on Window, Mac or Linux.*

__1(not useful) __2 __3 __4 __5(very useful)

Mobile, *deploy an application on smart phone or tablet.*

__1(not useful) __2 __3 __4 __5(very useful)

Browser, *deploy an application that can run on web browsers e.g. Firefox, IE, chrome and so on.*

__1(not useful) __2 __3 __4 __5(very useful)

Question 5:

____

____

Question 6:

____

____

Question 7:

__1(not important) __2 __3 __4 __5(very important)

Question 8:

Suggest an attraction

__1(not important) __2 __3 __4 __5(very important)

Suggest an restaurant/café shop

__1(not important) __2 __3 __4 __5(very important)

Suggest a hotel

__1(not important) __2 __3 __4 __5(very important)

Suggest a flight

__1(not important) __2 __3 __4 __5(very important)

Suggest general information

__1(not important) __2 __3 __4 __5(very important)

Suggest a route (Map guidance) (A ->B->C)

__1(not important) __2 __3 __4 __5(very important)

Suggest a route with visiting sequence (Map guidance) (A->C->B)

__1(not important) __2 __3 __4 __5(very important)

Suggest a whole/holistic travel package

__1(not important) __2 __3 __4 __5(very important)

Other (please comment)____

Question 9:

____

____

Question 10:

Tourists __1(not important) __2 __3 __4 __5(very important)

Travel agencies __1(not important) __2 __3 __4 __5(very important)

Tourism provider __1(not important) __2 __3 __4 __5(very important)

Others____

Question 11:

Before trip __1(not important) __2 __3 __4 __5(very important)

During trip __1(not important) __2 __3 __4 __5(very important)

After trip __1(not important) __2 __3 __4 __5(very important)

Others____

Question 12:

Like/dislike

__1(not important) __2 __3 __4 __5(very important)

Scaling

__1(not important) __2 __3 __4 __5(very important)

Comment/Review

__1(not important) __2 __3 __4 __5(very important)

Others____

Question 13:

__1(not important) __2 __3 __4 __5(very important)

Question 14:

Budget

__1(not important) __2 __3 __4 __5(very important)

Time/date (trip duration)

__1(not important) __2 __3 __4 __5(very important)

Point of interest

__1(not important) __2 __3 __4 __5(very important)

Events

__1(not important) __2 __3 __4 __5(very important)

Travel theme (romance, historical and etc.)

Weather

__1(not important) __2 __3 __4 __5(very important)

Season

__1(not important) __2 __3 __4 __5(very important)

Others____

Question 15:

__1(not important) __2 __3 __4 __5(very important)

Question 16:

Based on your travel preference

__1(not important) __2 __3 __4 __5(very important)

Based on other tourists preference

__1(not important) __2 __3 __4 __5(very important)

Based on travel agencies (knowledge expertise)
__1(not important) __2 __3 __4 __5(very important)

Based on your social network
__1(not important) __2 __3 __4 __5(very important)

Based on you and your group demographic
__1(not important) __2 __3 __4 __5(very important)

Others____

Question 17:

__1(not important) __2 __3 __4 __5(very important)

Question 18:

__1(not important) __2 __3 __4 __5(very important)

Question 19:

Group

__1(not useful) __2 __3 __4 __5(very useful)

Individual

__1(not useful) __2 __3 __4 __5(very useful)

Question 20:

__1(not likely) __2 __3 __4 __5(very much like)

Comments/ideas/brain storming

____

____

____

____

Thanks for your participation!

Appendix D

An example of an information sheet and consent form used in the data collection:

Bournemouth University

Research Ethics Checklist

Reference Id	4793
Status	Submitted

Researcher Details

Name	Pree Thiengburanathum
School	School of Tourism
Status	Postgraduate Research (PhD, MPhil, DProf, DEng)
Course	Postgraduate Research
Have you received external funding to support this research project?	No
Please list any persons or institutions that you will be conducting joint research with, both internal to BU as well as external collaborators.	Chiang Mai University (CMU), Thailand

Project Details

Title	A Route Recommendation System for Tourists in IoT Enviroment
Proposed Start Date	16/09/2013
Proposed End Date	15/09/2014

Summary (including detail on background methodology, sample, outcomes, etc.)

Page 1 of 4 Printed On 04/08/2014 21:58:45

Nowadays, huge volumes of information that is generated from the Internet and other sources, such as communication devices, sensors, guide books and maps have made it a difficult task for tourists to make decisions in terms of their preferences in traveling. This is true before the trip and during their trips and involves, selecting destinations, organizing trip plans, and making other decisions related to travel. It is considered to be a complex problem due to several factors, such as, number of days, number of travellers, budgets, user requirements, user profile, etc. Tourism industries/city/travel agency needs to address this problem to provide a quality of service to tourist, increase satisfaction and promote loyalty.By bringing a latest concept in Information and communications technology (ICT) such as Artificial Intelligent (AI) to the tourism domain, this could help reduce the complex problem when tourists planning their trip. The aims of this project are to study the impact (in tourism) of Compound Decision Support Systems (DSS), which is going to be an integration of Intelligent System (IS), Data mining, Database Management Systems, and Web based application. Another aim is to examine the needs and preference of tourists when they visit the destination. The proposed system would build based on Internet of Thing (IoT) framework system which involves the process of data acquisition, information creation, meaning-making and action taking. It should shorten the duration of the research and make research more efficient when dealing with a large volume of data from various sources, for examples, primary data from questionnaires and large data from GPS and RFID.Regarding the outcome of the project, the proposed system should be able to rank priority of point of interest, includes top destination/attractions, restaurants and hotels of the city. Also, it could generate the holistic trip plan based on the hard and soft constraints (user requirements and their demographic/characteristic profile) of the user. Machine learning (Decision Tree, Artificial Neural Network) and feature selection algorithms will be implemented for predict the right destination to the user, and optimize the trip plan.We have selected the city of Chiang Mai (One of the top cities for tourist destination in Thailand) to be our case study where we will obtain data for analyse the significant factors, test the classifier models and validate/modify the new algorithms.

External Ethics Review

Does your research require external review through the NHS National Research Ethics Service (NRES) or through another external Ethics Committee?	No

Research Literature

Is your research solely literature based?	No

Human Participants

Will your research project involve interaction with human participants as primary sources of data (e.g. interview, observation, original survey)?	Yes
Does your research specifically involve participants who are considered vulnerable (i.e. children, those with cognitive impairment, those in unequal relationships—such as your own students, prison inmates, etc.)?	No
Does the study involve participants age 16 or over who are unable to give informed consent (i.e. people with learning disabilities)? NOTE: All research that falls under the auspices of the Mental Capacity Act 2005 must be reviewed by NHS NRES.	No

Will the study require the co-operation of a gatekeeper for initial access to the groups or individuals to be recruited? (i.e. students at school, members of self-help group, residents of Nursing home?)	No
Will it be necessary for participants to take part in your study without their knowledge and consent at the time (i.e. covert observation of people in non-public places)?	No
Will the study involve discussion of sensitive topics (i.e. sexual activity, drug use, criminal activity)?	No

Are drugs, placebos or other substances (i.e. food substances, vitamins) to be administered to the study participants or will the study involve invasive, intrusive or potentially harmful procedures of any kind?	No

Will tissue samples (including blood) be obtained from participants? Note: If the answer to this question is 'yes' you will need to be aware of obligations under the Human Tissue Act 2004.	No

Could your research induce psychological stress or anxiety, cause harm or have negative consequences for the participant or researcher (beyond the risks encountered in normal life)?	No
Will your research involve prolonged or repetitive testing?	No
Will the research involve the collection of audio materials?	No
Will your research involve the collection of photographic or video materials?	No
Will financial or other inducements (other than reasonable expenses and compensation for time) be offered to participants?	No

Please give a summary of the ethical issues and any action that will be taken to address these. Explain how you will obtain informed consent (and from whom) and how you will inform the participant about the research project (i.e. participant information sheet).
At the survey area, the participant (i.e. local tourist, international tourist) will be asked to complete the survey to understand his/her destination choice in the city of Chiang Mai. The survey would take approximate 10 minutes to complete, and the participant will not be identified with any of his/her response to the survey.

Final Review

Will you have access to personal data that allows you to identify individuals OR access to confidential corporate or company data (that is not covered by confidentiality terms within an agreement or by a separate confidentiality agreement)?	No
Will your research involve experimentation on any of the following: animals, animal tissue, genetically modified organisms?	No

Will your research take place outside the UK (including any and all stages of research: collection, storage, analysis, etc.)?	Yes
Does the country in which you are conducting research require that you obtain internal ethical approval (i.e. beyond that required by Bournemouth University)?	No

Please use the below text box to highlight any other ethical concerns or risks that may arise during your research that have not been covered in this form.

Researcher Statement

JOURNALISM / BROADCAST RESEARCHERS: I confirm that I have consulted and understand the Research Ethics Supplementary Guide: For Reference by Researchers Undertaking Journalism and Media Production Projects (available on the Research Ethics page)	Yes

Appendix E

Features used to determine preferred destinations for tourists visiting Chiang Mai:

Factor	Feature	Description
Trip Characteristic (TC)	TC_1	Number of times you have visited
	TC_2	The purpose of this visit
	TC_3	The arrangements pertaining to this visit
	TC_4	Number of nights you plan to stay
	TC_5	Books, guides are the information sources that have influenced your decision to visit
	TC_6	People whom you are accompanied by are friends
	TC_7	TV, radio is the information source that has influenced your decision to visit
	TC_8	Adventurer is defined as your travel style
	TC_9	People whom you are accompanied by are children
	TC_{10}	Friends/relatives have influenced your decision to visit
Tourist Expenditure Behavior (TEB)	TEB_1	The amount of money you plan to spend per person on transportation during this visit
	TEB_2	Miscellaneous expenses you have pre-paid before this visit
	TEB_3	The amount of money you plan to spend on this visit
	TEB_4	The amount of money you plan to spend per person on shopping during this visit
Tourist Behavior (TB)	TB_1	Visiting markets and the walking streets is the activity you plan to participate in during this visit
	TB_2	The transport mode that you plan to use during this visit is walking
	TB_3	Wildlife has made the deepest impression upon you
	TB_4	Museums have made the deepest impression upon you
	TB_5	Outdoor is the activity you plan to participate in during this visit
	TB_6	Heath care is the activity that you plan to participate in during this visit
	TB_7	Thai boxing is the activity that you plan to participate in during this visit
	TB_8	Thai boxing has made the deepest impression upon you
	TB_9	Golfing has made the deepest impression upon you
	TB_{10}	Attending festivals is the activity you plan to participate in during this visit
	TB_{11}	Observing wildlife is the activity you plan to participate in during this visit
	TB_{12}	Thai food has made the deepest impression upon you
	TB_{13}	Performances are the primary focus on this visit
	TB_{14}	Overall cost of meals/food
	TB_{15}	Transport mode you plan to use is private car/motorcycle, van, coach for this visit
	TB_{16}	The transport mode you plan to use during this visit is the bicycle
	TB_{17}	Local activities are planned during your stay
	TB_{18}	Shopping has made the deepest impression upon you
	TB_{19}	Visiting entertainment places, nightclubs, bars is the activity that you plan to participate in during this visit
	TB_{20}	Nightlife has made the deepest impression upon you
	TB_{21}	Educational and sport sites are your favorite sites on this visit
	TB_{22}	Temple has made the deepest impression upon you
	TB_{23}	Attending performances is the activity you plan to participate in during this visit
Travel Motivation (TM)	TM_1	To work on my personal/spiritual values
	TM_2	To reflect on past memories
	TM_3	To reveal my thoughts, feelings, or physical skills to others
	TM_4	To visit relatives/friend(s)
	TM_5	To seek a better existence
	TM_6	To develop new abilities
	TM_7	To enhance relationships with friend(s)/family
	TM_8	To do things with family and friend(s)
	TM_9	To experience danger and trills
	TM_{10}	To share what I have learned with others
	TM_{11}	To learn about nature
	TM_{12}	To not worry about time and work

	TM_{13}	To visit places I have never been before
	TM_{14}	To gain a new perspective on life
	TM_{15}	To experience solitude and calm
	TM_{16}	To improve my romantic life
	TM_{17}	To understand more about myself
	TM_{18}	To see famous cultural and historical sites
Tourist Socio-Demographic information (TSD)	TSD_1	Primary area of employment
	TSD_2	Marital status
	TSD_3	Household income

Appendix F

System	Recommended item	RS Focus stage	System constraints	Recommended technique	Theories/Methods	Other features/service	Regional focus	System architecture	Adaptive capability	Spatial service	Ontology
Huang, Bian (2009) (Huang and Bian, 2009)	A	A, TPL	Age, tour motivation, occupation, travel type, personality, preferred activity, cost, distance	Critique-based, hybrid filtering (content-based filtering and collaborative filtering)	BN, AHP, DMT	Prediction of user preferred activity, ranking attractions, integration of heterogeneous online travel information	New York, USA	W	Yes	Yes	Yes
PSiS(Anacleto et al., 2014)	A, AC, RO	TPL	Location, time, speed, direction, weather and user preferences	Context-based	Algorithm (ranking POIs)	Architectonic tag(recommend the POI beyond the regular schedule), dynamic tour adaption, device-aware	Porto, Portugal	M, W	Yes	Yes	No
PTPS(Chiang and Huang, 2015)	A, AC, RO	TPL, TIDP	Number of days, budget, lunch time, dinner time,	User-constraint based	Algorithm (Matching, ranking, and planning)	Rank attraction by user feedback, time arrangement	Taiwan	W	Yes	Yes	No

			must see POIs, start point, travel type, food type, dwelling time, transport time			mechanism, Solving trip design problem (TSPTW)					
Otium(Montejo-Ráez et al., 2011)(Montejo-Ráez et al., 2011)	AT (e.g. theatre event)	TPL	budget, start and end date	User-constraint based	Algorithm, VSM, CO	Web extraction for heterogeneous online travel information	Spain	W	Yes	No	No
ITAS (Hsu et al., 2012)	A	Recommended attractions in sequences	User demographic information (nation, gender, age, income, occupation) purpose of travel, source of information, travel type	User-constraint based, hybrid filtering (content-based filtering and collaborative filtering)	BN, DMT, CLF, CA, DS	Prediction of user preferred attractions	Taiwan	W	No	Yes	No
DailyTrip (Gavalas et al., 2012a)	POI (museum, archaeological site, monument, etc.)	TPL, TIDP	User demographic information (age, educational level),	User-constraint based, context-based	Algorithm, H	solving trip design problem (TOPTW)	Not specified	WM	Yes	Yes	No

			disability, budget, time, transport mode, time available for sightseeing, open days of sites, average visiting time for the sites.							
(P. Vansteenwegen en et al., 2011)(Vansteenwegen et al., 2011)	POI	TPL, TIDP	Number of days, start and end location, start and end time, lunch break, multiple opening and closing times per day, and user interest	User-constraint based	Algorithm, H	Solving trip design problem (TOPTW)	Belgium	W	Yes	Yes	No
(Lee et al., 2009)(Lee et al., 2009)	A (historical sites), R	RE (historical sites and restaurant)	Number of days, popularity, region, food type, classes of historical sites.	User-constraint based	ACO , Planning Algorithm, FL	POIs location transfer mechanism, solving TSP problem	Taiwan	W	No	Yes	Yes

(Montejo-Ráez et al., 2011)SAMAP(Castillo et al., 2008)	POI, RO, AT	TPL, R (point to point)	Demographic information, interest, the number of days, current time, transport price, transport duration, required activities, budget, food type, meal time, the open time of the place.	User-constraint based, hybrid filtering (context-based, collaborative filtering)	KNN, CBR, AI planners	Ranking attractions, support user the point to point route and the transport mode (bus, taxi, walking, etc.) to take, solving trip design problem (TOTPW)	Not specified	SM	No	Yes	Yes
(Wang et al., 2011) (Wang et al., 2011)	AT	TPL	Age, tour motivation, occupation, travel type, personality, preferred activity.	Hybrid filtering (content-based filtering and collaborative filtering)		Prediction of user preferred activity, integration of heterogeneous online travel information by using Mashup	Beijing and Shanghai China	W	No	Yes	Yes
Sig Tur(Moreno, Valls, Isern, Marin, & Borràs, 2013)	POI, AT	TPL	Demographic information (country of origin), Tour characteristic	Hybrid filtering (collaborative filtering,	Aggregation operators, CLF using k-means clustering algorithm	Ranking activities, feedback	Tarragona, Spain	W	Yes	Yes	Yes

			(travel budget, group composition, required destination, accommodation type, budget, travel date (starting and ending) date) motivations	content-based filtering)							
Turist@(Batet et al., 2012)	AT	TPL	Demographic information (birth date, nationality, education, language, interest, disability. Travel group type, start and end date of the trip, discounts, price, free entrances	Hybrid filtering (content-based filtering and collaborative filtering)	VSM, normalised Euclidean distance, CFT,CLT	Explicit and implicit feedback.	Tarragona, Spain	M	Yes	No	No
SPETA(García-Crespo et al., 2009)	A	T	location, weather, speed, direction,	Hybrid filtering (context-based,	(feature-based similarity algorithms, VSM, SVM	Filter the attraction using and open/close time, date, and	Not specified	SM		No	Yes

			time, user preferences (food type), social network, user's history	knowledge filtering, collaborative filtering)		user context information.					
(Alptekin and Buyukozkan, 2011)	A, RO , T	T	Number of travellers, trip length, Region, duration, trip type, hotel type, season	Knowledge-based filtering	AHP, CBR, distance calculation	Price with the trip plan, develop for travel agency use.	Not specified	W	No	No	No
Traveller(Schiaffino and Amandi, 2009)	D, AC, T(Holiday tour package)	D	User preferences	Hybrid-filtering (content-based filtering, collaborative filtering, demographic filtering)	Association rule, Cosine-Similarity			W		No	No
Yeh, Cheng(2014)	A	A	User preferences	Knowledge filtering, Delphi panel and Repertory grid	Cosine Similarity, FOCUS analysis	Predicting attraction using constructed and elemenet0-based recommendation	Taiwan	W	No	No	No

GeOasis(Santiago et al., 2012)	POI	TPL	User preferences, current location, time, and space	Knowledge filtering, context-based filtering	Planning algorithm, H	Voice-based interface to improve user interactive, real-time recommendation	Jaen, Spain	S	Yes	Yes	Yes
SACO (Mocholi et al., 2012)	RO, AT	RE	User preferences, context information	Context-based filtering	ACO, Sematic searching	The feature that let user define his/her ontology	Not specified	S	Yes	Yes	Yes
BOTTARI (Balduini et al., 2012)	A	RE	Context information	Location-based	Inductive and deductive stream reasoner	Use augmented reality	Insadong, Seoul	M	Yes	Yes	Yes

Table 2 (Continued)

System	User model	User input	Relevance feedback	System Evaluation
Huang, Bian (2009) (Huang and Bian, 2009)	I	E	Yes	No evaluation
PSiS(Anacleto et al., 2014)	I	I, E	Yes	Survey
PTPS(Chiang and Huang, 2015)	I	I, E	Yes	Satisfaction, Questionnaires PRE, F, RE

Otium (Montejo-Ráez et al., 2011)	I	E	No feedback	No evaluation
ITAS (Hsu et al., 2012)	I	E	No feedback	ACC, ROC
DailyTrip (Gavalas et al., 2012a)	I	E	No feedback	Algorithm performance
P. Vansteenwegen en et al., 2011)(Vansteenwegen et al., 2011)	I	E	No feedback	Satisfaction, Questionnaires, usage statistics
(Lee et al., 2009**)**(Lee et al., 2009)	I	E	No feedback	No evaluation
(SAMAP(Castillo et al., 2008)	I	E	No feedback	No evaluation
(Wang et al., 2011) (Wang et al., 2011)	I	E	Yes	No evaluation
Sigtur/E-destination (Moreno, Valls, Isern, Marin, & Borràs, 2013)				
Turist@(Batet et al., 2012)	I	I, E	Yes	No evaluation
SPETA(García-Crespo et al., 2009)	I	I, E	Yes	No evaluation
(Alptekin and Buyukozkan, 2011)	I	E	No feedback	No evaluation
Traveller (Schiaffino and Amandi, 2009)	I	E	No feedback	Comparing prediction and precision values

Yeh, Cheng (2014)	I	I	Yes	ACC
GeOasis (Santiago et al., 2012)	I	I, E	Yes	No evaluation
SACO (Mocholi et al., 2012)	I	I, E	No feedback	No evaluation
BOTTARI (Balduini et al., 2012)	I	I,E	No feedback	ACC

References

Abbaspour, R.A., Samadzadegan, F., 2011. Time-dependent personal tour planning and scheduling in metropolises. Expert Systems with Applications 38, 12439–12452. https://doi.org/10.1016/j.eswa.2011.04.025

Al Shalabi, L., Shaaban, Z., 2006. Normalization as a Preprocessing Engine for Data Mining and the Approach of Preference Matrix, in: International Conference on Dependability of Computer Systems, 2006. DepCos-RELCOMEX '06. Presented at the International Conference on Dependability of Computer Systems, 2006. DepCos-RELCOMEX '06, pp. 207–214. https://doi.org/10.1109/DEPCOS-RELCOMEX.2006.38

Alptekin, G.I., Buyukozkan, G., 2011. An integrated case-based reasoning and MCDM system for Web based tourism destination planning. EXPERT SYSTEMS WITH APPLICATIONS 38, 2125–2132.

An Introduction to Machine Learning | Miroslav Kubat | Springer, n.d.

Anacleto, R., Figueiredo, L., Almeida, A., Novais, P., 2014. Mobile application to provide personalized sightseeing tours. Journal of Network and Computer Applications 41, 56–64. https://doi.org/10.1016/j.jnca.2013.10.005

Andereck, K.L., Caldwell, L.L., 1994. The Influence of Tourists' Characteristics on Ratings of Information Sources for an Attraction. Journal of Travel & Tourism Marketing 2, 171–190. https://doi.org/10.1300/J073v02n02_11

Andereck, K.L., Caldwell, L.L., Uysal, M., 1993. The influence of tourists' characteristics on ratings of information sources for an attraction. Journal of Travel & Tourism Marketing 2, 171–189.

Anderson, C., 2012. The Model-View-ViewModel (MVVM) Design Pattern, in: Pro Business Applications with Silverlight 5. Apress, pp. 461–499. https://doi.org/10.1007/978-1-4302-3501-9_13

Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P., 2003. Intrigue: Personalized Recommendation of Tourist Attractions for Desktop and Hand Held Devices. Applied Artificial Intelligence 17, 687.

Badaro, G., Hajj, H., El-Hajj, W., Nachman, L., 2013. A hybrid approach with collaborative filtering for recommender systems, in: 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC). Presented at the 2013

9th International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 349–354. https://doi.org/10.1109/IWCMC.2013.6583584

Balduini, M., Celino, I., Dell'Aglio, D., Della Valle, E., Huang, Y., Lee, T., Kim, S.-H., Tresp, V., 2012. BOTTARI: An augmented reality mobile application to deliver personalized and location-based recommendations by continuous analysis of social media streams. Web Semantics: Science, Services and Agents on the World Wide Web, The Semantic Web Challenge 2011 16, 33–41. https://doi.org/10.1016/j.websem.2012.06.004

Batet, M., Moreno, A., Sanchez, D., Isern, D., Valls, A., 2012. Turist@: Agent-based personalised recommendation of tourist activities. EXPERT SYSTEMS WITH APPLICATIONS 39, 7319–7329.

Bengio, Y., 2009. Learning Deep Architectures for AI. Foundations and Trends® in Machine Learning 2, 1–127. https://doi.org/10.1561/2200000006

Bishop, C.M., 1996. Neural Networks for Pattern Recognition. Oxford University Press, USA, Oxford : New York.

Boger, Z., Guterman, H., 1997. Knowledge extraction from artificial neural network models, in: , 1997 IEEE International Conference on Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation. Presented at the , 1997 IEEE International Conference on Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation, pp. 3030–3035 vol.4. https://doi.org/10.1109/ICSMC.1997.633051

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 30, 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Breiman, L., 1996. Bagging predictors. Mach Learn 24, 123–140. https://doi.org/10.1007/BF00058655

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees, New Ed edition. ed. Chapman and Hall/CRC, New York, N.Y.

Brouwer, R.K., 2002. A feed-forward network for input that is both categorical and quantitative. Neural Networks 15, 881–890. https://doi.org/10.1016/S0893-6080(02)00090-4

Buckland, M., Gey, F., 1994. The relationship between Recall and Precision. J. Am. Soc. Inf. Sci. 45, 12–19. https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L

Buhalis, D., Law, R., 2008. Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tourism Management 29, 609–623. https://doi.org/10.1016/j.tourman.2008.01.005

Burke, R., 2002. Hybrid Recommender Systems: Survey and Experiments. User Model User-Adap Inter 12, 331–370. https://doi.org/10.1023/A:1021240730564

Castillo, L., Armengol, E., Onaindia, E., Sebastia, L., Gonzalez-Boticario, J., Rodriguez, A., Fernandez, S., Arias, J.D., Borrajo, D., 2008. SAMAP: An user-oriented adaptive system for planning tourist visits. EXPERT SYSTEMS WITH APPLICATIONS 34, 1318–1332.

Catal, C., Tufekci, S., Pirmit, E., Kocabag, G., 2015. On the use of ensemble of classifiers for accelerometer-based activity recognition. Applied Soft Computing 37, 1018–1022. https://doi.org/10.1016/j.asoc.2015.01.025

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A Library for Support Vector Machines. ACM Trans. Intell. Syst. Technol. 2, 27:1–27:27. https://doi.org/10.1145/1961189.1961199

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Int. Res. 16, 321–357.

Chiang, H.-S., Huang, T.-C., 2015. User-adapted travel planning system for personalized schedule recommendation. Information Fusion. https://doi.org/10.1016/j.inffus.2013.05.011

Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C., 2012. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. NeuroImage 60, 59–70. https://doi.org/10.1016/j.neuroimage.2011.11.066

Chu, R., 2001. What online Hong Kong travelers look for on airline/travel websites? International Journal of Hospitality Management 20, 95–100. https://doi.org/10.1016/S0278-4319(00)00046-3

Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper - Cisco [WWW Document], n.d. URL http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html (accessed 6.27.17).

Converting Between XML and JSON [WWW Document], n.d. URL http://www.xml.com/pub/a/2006/05/31/converting-between-xml-and-json.html (accessed 6.27.17).

Crompton, J.L., 1979. Motivations for pleasure vacation. Annals of Tourism Research 6, 408–424. https://doi.org/10.1016/0160-7383(79)90004-5

De Bruyn, A., Liechty, J.C., Huizingh, E.K.R.E., Lilien, G.L., 2008. Offering Online Recommendations with Minimum Customer Input Through Conjoint-Based Decision Aids. Marketing Science 27, 443–460. https://doi.org/10.1287/mksc.1070.0306

Dietterich, T.G., 2000. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. Machine Learning 40, 139–157. https://doi.org/10.1023/A:1007607513941

Economic Impact of Travel & Tourism 2014 Annual Update: Summary, 2014.

Estevez, P.A., Tesmer, M., Perez, C.A., Zurada, J.M., 2009. Normalized Mutual Information Feature Selection. IEEE Transactions on Neural Networks 20, 189–201. https://doi.org/10.1109/TNN.2008.2005601

Fayyad, U., Piatetsky-shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases. AI Magazine 17, 37–54.

Fesenmaier, D.R., Wöber, K.W., Werthner, H., 2006. Destination recommendation systems [electronic resource] : behavioural foundations and applications / edited by Daniel R. Fesenmaier, Karl W. Wöber, Hannes Werthner. Wallingford, UK ; Cambridge, MA : CABI Pub., c2006.

Figler, M.H., Weinstein, A.R., Iii, J.J.S., Devan, B.D., 1992. Pleasure travel (tourist) motivation: A factor analytic approach. Bull. Psychon. Soc. 30, 113–116. https://doi.org/10.3758/BF03330412

Fouss, F., Saerens, M., 2008. Evaluating Performance of Recommender Systems: An Experimental Comparison, in: 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Presented at the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 735–738. https://doi.org/10.1109/WIIAT.2008.252

Freund, Y., Schapire, R.E., 1999. A Short Introduction to Boosting.

Gamma, E., 1995. Design patterns: elements of reusable object-oriented software. Addison-Wesley, Reading, Mass.

Garcia, I., Sebastia, L., Onaindia, E., 2011. On the design of individual and group recommender systems for tourism. Expert Systems with Applications 38, 7683–7692. https://doi.org/10.1016/j.eswa.2010.12.143

García-Crespo, A., Chamizo, J., Rivera, I., Mencke, M., Colomo-Palacios, R., Gómez-Berbís, J.M., 2009. SPETA: Social pervasive e-Tourism advisor. Telematics and Informatics 26, 306–315. https://doi.org/10.1016/j.tele.2008.11.008

Garcia-Crespo, A., Lopez-Cuadrado, J.L., Colomo-Palacios, R., Gonzalez-Carrasco, I., Ruiz-Mezcua, B., 2011. Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain. EXPERT SYSTEMS WITH APPLICATIONS 38, 13310–13319.

Gavalas, D., Kenteris, M., 2011. A web-based pervasive recommendation system for mobile tourist guides. PERSONAL AND UBIQUITOUS COMPUTING 15, 759–770.

Gavalas, D., Kenteris, M., Konstantopoulos, C., Pantziou, G., 2012a. Web application for recommending personalised mobile tourist routes. IET Software 6, 313–322. https://doi.org/10.1049/iet-sen.2011.0156

Gavalas, D., Kenteris, M., Konstantopoulos, C., Pantziou, G., 2012b. Web application for recommending personalised mobile tourist routes. IET Software 6, 313–322. https://doi.org/10.1049/iet-sen.2011.0156

Gavalas, D., Konstantopoulos, C., Mastakas, K., Pantziou, G., n.d. Mobile recommender systems in tourism. Journal of Network and Computer Applications. https://doi.org/10.1016/j.jnca.2013.04.006

Gavalas, D., Konstantopoulos, C., Mastakas, K., Pantziou, G., n.d. A survey on algorithmic approaches for solving tourist trip design problems. J Heuristics 1–38. https://doi.org/10.1007/s10732-014-9242-5

Ge, Y., Liu, Q., Xiong, H., Tuzhilin, A., Chen, J., 2011. Cost-aware Travel Tour Recommendation, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’11. ACM, New York, NY, USA, pp. 983–991. https://doi.org/10.1145/2020408.2020568

Ghazanfar, M.A., Prugel-Bennett, A., 2010. A Scalable, Accurate Hybrid Recommender System, in: 2010 Third International Conference on Knowledge Discovery and Data Mining. Presented at the 2010 Third International Conference on Knowledge Discovery and Data Mining, pp. 94–98. https://doi.org/10.1109/WKDD.2010.117

Golden, B.L., Levy, L., Vohra, R., 1987. The orienteering problem. Naval Research Logistics 34, 307.

Gretzel, U.(1), Hwang, Y.-H.(2), Fesenmaier, D. r. (3), 2012. Informing destination recommender systems design and evaluation through quantitative research.

International Journal of Culture, Tourism, and Hospitality Research 6, 297–315. https://doi.org/10.1108/17506181211265040

Guillet, B.D., Lee, A., Law, R., Leung, R., 2011. Factors Affecting Outbound Tourists' Destination Choice: The Case of Hong Kong. Journal of Travel & Tourism Marketing 28, 556–566. https://doi.org/10.1080/10548408.2011.588120

Guyon, I., Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. J. Mach. Learn. Res. 3, 1157–1182.

Hagen, K. ten, Kramer, R., Hermkes, M., Schumann, B., Mueller, P., 2005. Semantic Matching and Heuristic Search for a Dynamic Tour Guide, in: Frew, D.A.J. (Ed.), Information and Communication Technologies in Tourism 2005. Springer Vienna, pp. 149–159.

Häubl, G., Trifts, V., 2000. Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. Marketing Science 19, 4.

Heaton, J., 2008. Introduction to Neural Networks with Java, 2nd Edition, ed. Heaton Research, Incorporated, St. Louis, Mo.

Horrocks, I., 2008. Ontologies and the Semantic Web. Commun. ACM 51, 58–67. https://doi.org/10.1145/1409360.1409377

Hsu, F.M., Lin, Y.T., Ho, T.K., 2012. Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps. EXPERT SYSTEMS WITH APPLICATIONS 39, 3257–3264.

Hsu, T.-K., Tsai, Y.-F., Wu, H.-H., 2009. The preference analysis for tourist choice of destination: A case study of Taiwan. Tourism Management 30, 288–297. https://doi.org/10.1016/j.tourman.2008.07.011

Huang, Y.X., Bian, L., 2009. A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet. EXPERT SYSTEMS WITH APPLICATIONS 36, 933–943.

Hussein, A., Thomas G. Dietterich, 1991. Learning with Many Irrelevant Features. MIT Press, Cambridge, MA 547–553.

Isinkaye, F.O., Folajimi, Y.O., Ojokoh, B.A., 2015. Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal 16, 261–273. https://doi.org/10.1016/j.eij.2015.06.005

Jang, S. (Shawn), 2004. The Past, Present, and Future Research of Online Information Search. Journal of Travel & Tourism Marketing 17, 41–47. https://doi.org/10.1300/J073v17n02_04

Jannach, D., Zanker, M., Felfernig, A., Friedrich, G., 2010. Recommender Systems: An Introduction. Cambridge University Press, New York.

Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K., 1991. User Interface Evaluation in the Real World: A Comparison of Four Techniques, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’91. ACM, New York, NY, USA, pp. 119–124. https://doi.org/10.1145/108844.108862

Jr, J.F.H., Black, W.C., Babin, B.J., Anderson, R.E., 2009. Multivariate Data Analysis, 7 edition. ed. Prentice Hall, Upper Saddle River, NJ.

Kabassi, K., 2010. Personalizing recommendations for tourists. Telematics & Informatics 27, 51–66. https://doi.org/10.1016/j.tele.2009.05.003

Kantamneni, A., Brown, L.E., Parker, G., Weaver, W.W., 2015. Survey of multi-agent systems for microgrid control. Engineering Applications of Artificial Intelligence 45, 192–203. https://doi.org/10.1016/j.engappai.2015.07.005

Kass, G.V., 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. Journal of the Royal Statistical Society. Series C (Applied Statistics) 29, 119–127. https://doi.org/10.2307/2986296

Khalili, A., Auer, S., 2013. User interfaces for semantic authoring of textual content: A systematic literature review. JOURNAL OF WEB SEMANTICS 22, 1–18.

knockout: Knockout makes it easier to create rich, responsive UIs with JavaScript, 2017. . Knockout.js.

Kohavi, R., 1995. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI’95. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1137–1143.

Koren, Y., Bell, R., Volinsky, C., 2009. Matrix Factorization Techniques for Recommender Systems. Computer 42, 30–37. https://doi.org/10.1109/MC.2009.263

Lamsfus, C., Martin, D., Alzua-Sorzabal, A., López-de-Ipiña, D., Torres-Manzanera, E., 2012. Context-based tourism information filtering with a semantic rule engine. Sensors (Basel, Switzerland) 12, 5273–5289. https://doi.org/10.3390/s120505273

Lee, C.-S., Chang, Y.-C., Wang, M.-H., 2009. Ontological recommendation multi-agent for Tainan City travel. Expert Systems with Applications 36, 6740–6753. https://doi.org/10.1016/j.eswa.2008.08.016

Leiper, N., 1990. Tourist attraction systems. Annals of Tourism Research 17, 367–384. https://doi.org/10.1016/0160-7383(90)90004-B

Lior, R., Oded, M., 2008. Data Mining with Decision Trees- Theroy and Applications.

Liu, L., Xu, J., Liao, S.S., Chen, H., 2014. A real-time personalized route recommendation system for self-drive tourists based on vehicle to vehicle communication. Expert Systems with Applications 41, 3409–3417. https://doi.org/10.1016/j.eswa.2013.11.035

LOH, S., Lorenzi, F., Salana, r., Licthnow, D., 2003. A Tourism Recommender Syste based on Collaboration and Text Analysis. Information Technology & Tourism 6, 157–165. https://doi.org/10.3727/1098305031436980

Lucas, J.P., Luz, N., Moreno, M.N., Anacleto, R., Almeida Figueiredo, A., Martins, C., 2013. A hybrid recommendation approach for a tourism system. Expert Systems with Applications 40, 3532–3550. https://doi.org/10.1016/j.eswa.2012.12.061

Maimon, O., Rokach, L., 2005. Decomposition Methodology for Knowledge Discovery and Data Mining, in: Maimon, O., Rokach, L. (Eds.), Data Mining and Knowledge Discovery Handbook. Springer US, pp. 981–1003. https://doi.org/10.1007/0-387-25465-X_46

Mardani, A., Jusoh, A., Zavadskas, E., Kazemilari, M., Ahmad, U., Khalifah, Z., 2016. Application of Multiple Criteria Decision Making Techniques in Tourism and Hospitality Industry: a Systematic Review. Transformations in Business and Economics 15.

Melanie Swan, 2012. Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. Journal of Sensor and Actuator Networks 217.

Mocholi, J.A., Jaen, J., Krynicki, K., Catala, A., Picón, A., Cadenas, A., 2012. Learning semantically-annotated routes for context-aware recommendations on map navigation systems. Applied Soft Computing 12, 3088–3098. https://doi.org/10.1016/j.asoc.2012.05.010

Møller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6, 525–533. https://doi.org/10.1016/S0893-6080(05)80056-5

Montaner, M., Lopez, B., de la Rosa, J.L., 2003. A taxonomy of recommender agents on the Internet. ARTIFICIAL INTELLIGENCE REVIEW 19, 285–330.

Montejo-Ráez, A., Perea-Ortega, J.M., García-Cumbreras, M.Á., Martínez-Santiago, F., 2011. Otiŭm: A web based planner for tourism and leisure. Expert Systems with Applications 38, 10085–10093. https://doi.org/10.1016/j.eswa.2011.02.005

Moreno, A., Valls, A., Isern, D., Marin, L., Borràs, J., 2013. SigTur/E-Destination: Ontology-based personalized recommendation of Tourism and Leisure Activities. Engineering

Applications of Artificial Intelligence 26, 633–651. https://doi.org/10.1016/j.engappai.2012.02.014

Moreno, A., Valls, A., Isern, D., Marin, L., Borràs,. SigTur/E-Destination: Ontology-based personalized recommendation of Tourism and Leisure Activities. Engineering Applications of Artificial Intelligence 26, 633–651. https://doi.org/10.1016/j.engappai.2012.02.014

Mutinda, R., Mayaka, M., 2012. Application of destination choice model: Factors influencing domestic tourists destination choice among residents of Nairobi, Kenya. Tourism Management 33, 1593–1597. https://doi.org/10.1016/j.tourman.2011.12.008

NETLAB, Algorithms for Pattern Recognition, Ian T. Nabney, Springer, n.d.

Pang, G.K.H., Takahashi, K., 1999. Adaptive Route Selection for Dynamic Route Guidance System based on Fuzzy-Neural Approaches. IEEE Transactions on Vehicular Technology 48, 2028.

Pantano, E., Pietro, L.D., 2013. From e-tourism to f-tourism: emerging issues from negative tourists, online reviews. Journal of Hospitality and Tourism Technology 4, 211–227. https://doi.org/10.1108/JHTT-02-2013-0005

Payr, S., Petta, P., Trappl, R., 2002. Emotions in Humans and Artifacts. MIT Press, Cambridge, Mass.

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1226–1238. https://doi.org/10.1109/TPAMI.2005.159

Petrevska, B., Koceski, S., 2012. Tourism Recommendation System: Empirical Investigation. Revista de Turism - Studii si Cercetari in Turism 11–18.

Pitoska, E., 2013. E-Tourism: The Use of Internet and Information and Communication Technologies in Tourism: The Case of Hotel Units in Peripheral Areas. Tourism in South East Europe 2, 335–344.

Platt, J.C., 1999. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods, in: Advances in Large Margin Classifiers. MIT Press, pp. 61–74.

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Quinlan, J.R., 1986a. Induction of Decision Trees. MACH. LEARN 1, 81–106.

Quinlan, J.R., 1986b. Induction of Decision Trees. Mach Learn 1, 81–106. https://doi.org/10.1023/A:1022643204877

Raymond, C., 2009. Intelligent Systems for Automated Learning and Adaptation: Emerging Trends and Applications: Emerging Trends and Applications. IGI Global.

Recommender Systems - The Textbook Charu C. Aggarwal Springer, n.d.

Resnick, P., Varian, H.R., 1997. Recommender Systems. Commun. ACM 40, 56–58. https://doi.org/10.1145/245108.245121

Ricci, F., Rokach, L., Shapira, B., 2011. Introduction to Recommender Systems Handbook, in: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (Eds.), Recommender Systems Handbook. Springer US, pp. 1–35.

Rodríguez, B., Molina, J., Pérez, F., Caballero, R., 2012. Interactive design of personalised tourism routes. Tourism Management 33, 926–940. https://doi.org/10.1016/j.tourman.2011.09.014

Rubinelli, S., Collm, A., Glässel, A., Diesner, F., Kinast, J., Stucki, G., Brach, M., 2013. Designing interactivity on consumer health websites: PARAFORUM for spinal cord injury. Patient Education and Counseling 93, 459–463. https://doi.org/10.1016/j.pec.2013.09.015

Saleh, E., Błaszczyński, J., Moreno, A., Valls, A., Romero-Aroca, P., de la Riva-Fernández, S., Słowiński, R., 2017. Learning ensemble classifiers for diabetic retinopathy assessment. Artificial Intelligence in Medicine. https://doi.org/10.1016/j.artmed.2017.09.006

Santiago, F.M., López, F.A., Montejo-Ráez, A., López, A.U., 2012. GeOasis: A knowledge-based geo-referenced tourist assistant. Expert Systems with Applications 39, 11737–11745. https://doi.org/10.1016/j.eswa.2012.04.080

Schafer, J.B., Konstan, J.A., Riedl, J., 2001. E-Commerce Recommendation Applications, in: Kohavi, R., Provost, F. (Eds.), Applications of Data Mining to Electronic Commerce. Springer US, pp. 115–153.

Schiaffino, S., Amandi, A., 2009. Building an expert travel agent as a software agent. Expert Systems with Applications 36, 1291–1299. https://doi.org/10.1016/j.eswa.2007.11.032

Scholz, M., Dorner, V., Franz, M., Hinz, O., 2015. Measuring consumers’ willingness to pay with utility-based recommendation systems. Decision Support Systems 72, 60–71. https://doi.org/10.1016/j.dss.2015.02.006

Sebastia, L., Garcia, I., Onaindia, E., Guzman, C., 2009. e-TOURISM:: A TOURIST RECOMMENDATION AND PLANNING APPLICATION. International Journal on Artificial Intelligence Tools 18, 717–738.

Shannon, C.E., 2001. A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review 5, 3–55.

Shapiro, S.S., Wilk, M.B., 1965. An Analysis of Variance Test for Normality (Complete Samples). Biometrika 52, 591–611. https://doi.org/10.2307/2333709

Shibata, K., Ikeda, Y., 2009. Effect of number of hidden neurons on learning in large-scale layered neural networks, in: ICCAS-SICE, 2009. Presented at the ICCAS-SICE, 2009, pp. 5008–5013.

Siebers, P.-O., Aickelin, U., 2008. Introduction to Multi-Agent Simulation.

Souffriau, W., Vansteenwegen, P., 2010. Tourist Trip Planning Functionalities: State of the Art and Future, in: Daniel, F., Facca, F.M. (Eds.), Current Trends in Web Engineering, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 474–485.

Sparacino, F., 2003. Sto(ry)chastics: A Bayesian Network Architecture for User Modeling and Computational Storytelling for Interactive Spaces, in: Dey, A., Schmidt, A., McCarthy, J. (Eds.), UbiComp 2003: Ubiquitous Computing, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 54–72.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

Tarjan, R., 1972. Depth-First Search and Linear Graph Algorithms. SIAM J. Comput. 1, 146–160. https://doi.org/10.1137/0201010

Thailand Tourism Annual Report 2013, n.d.

Thiengburanathum, P., Shuang, C., Hongnian, Y., 2016. Overview of Personalized Travel Recommendation Systems, in: Proceedings of the 20th International Conference on Automation & Computing. Presented at the ICAC, Cranfield University, Bedforshire, UK.

Trenn, S., 2008. Multilayer Perceptrons: Approximation Order and Necessary Number of Hidden Units. IEEE Transactions on Neural Networks 19, 836–844. https://doi.org/10.1109/TNN.2007.912306

Tsai, C.-Y., Chung, S.-H., 2012. A personalized route recommendation service for theme parks using RFID information and tourist behaviour. Decision Support Systems 52, 514–527. https://doi.org/10.1016/j.dss.2011.10.013

Turban, E., Sharda, R., Delen, D., 2014. Business Intelligence and Analytics: Systems for Decision Support, 10 edition. ed. Pearson, Harlow.

Vansteenwegen, P., Souffriau, W., Van Oudheusden, D., Vanden Berghe, G., 2009. Iterated local search for the team orienteering problem with time windows. Computers and Operations Research 36, 3281–3290. https://doi.org/10.1016/j.cor.2009.03.008

Vansteenwegen, P., Souffriau, W., Berghe, G.V., Oudheusden, D.V., 2011. The City Trip Planner: An expert system for tourists. Expert Systems with Applications 38, 6540–6546. https://doi.org/10.1016/j.eswa.2010.11.085

Veropoulos, K., Campbell, C., Cristianini, N., 1999. Controlling the Sensitivity of Support Vector Machines, in: Proceedings of the International Joint Conference on AI. pp. 55–60.

Wang, W., Zeng, G., Tang, D., 2011. Bayesian intelligent semantic mashup for tourism. Concurrency and Computation: Practice and Experience 23, 850–862. https://doi.org/10.1002/cpe.1676

Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. Biometrics Bulletin 1, 80–83. https://doi.org/10.2307/3001968

Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd Revised edition. ed. Morgan Kaufmann Publishers in, Amsterdam, Boston, MA.

World travel and tourism council, n.d. Travel and Tourism enconomic impact 2015 Thailand.

Yang, W.-S., Hwang, S.-Y., 2013. iTravel: A recommender system in mobile peer-to-peer environment. Journal of Systems and Software 86, 12–20. https://doi.org/10.1016/j.jss.2012.06.041

Yeh, D.-Y., Cheng, C.-H., 2015. Recommendation system for popular tourist attractions in Taiwan using Delphi panel and repertory grid techniques. Tourism Management 46, 164–176. https://doi.org/10.1016/j.tourman.2014.07.002

Zhou, Z.-H., 2015. Ensemble Learning, in: Li, S.Z., Jain, A.K. (Eds.), Encyclopedia of Biometrics. Springer US, Boston, MA, pp. 411–416. https://doi.org/10.1007/978-1-4899-7488-4_293