PERPUSTAKAAN UMP
0000113751

CUSTOMER CHURN CLASSIFICATION IN TELECOMMUNICATION COMPANY USING ROUGH SET THEORY

NUR SYAFIQAH MOHD NAFIS

Thesis Submitted in Fulfilment of Requirement for the Degree of Master of Science in the Faculty of Informatics and Computing Universiti Sultan Zainal Abidin
2016

PENGKELASAN PERPINDAHAN MENGGUNAKAN TEORI SET KASAR PELANGGAN DI SYARIKAT TELEKOMUNIKASI

ABSTRAK

Perpindahan boleh ditakrifkan sebagai tingkah laku pelanggan untuk meninggalkan atau menamatkan perkhidmatan. Perilaku ini mengakibatkan kehilangan keuntungan yang dijana oleh syarikat kerana bagi mendapatkan pelanggan baharu, ia memerlukan pelaburan yang lebih tinggi untuk iklan dan promosi berbanding mengekalkan pelanggan sedia ada. Oleh itu, model ramalan yang cekap perlu diambil kira bagi mengurangkan kadar perpindahan. Dalam pendekatan tradisional pemodelan ramalan, ia tidak menghasilkan tafsiran keputusan yang mudah. Oleh itu, mengenal pasti model ramalan terbaik untuk mengurangkan kadar perpindahan sememangnya merupakan satu tugas yang mencabar. Objektif utama tesis ini ialah mencadangkan algoritma pengelasan baharu untuk meramalkan perpindahan pelanggan menggunakan Teori Set Kasar. Penyelidikan ini menggunakan proses Penemuan Pengetahuan dalam Pangkalan Data (KDD) yang melibatkan data sebelum pemprosesan, pendiskretan data, pengurangan atribut, penghasilan peraturan, proses pengelasan, serta analisis data, dengan menggunakan kit panduan set kasar. Elemen Teori Set Kasar terdiri daripada hubungan yang tidak dapat dilihat dengan jelas, anggaran di aras bawah dan atas, serta set pengurangan. Elemen-elemen tersebut digunakan untuk mengelaskan perpindahan pelanggan dari set data yang tidak menentu dan tidak tepat. Keputusan model yang dicadangkan dibandingkan dengan beberapa pendekatan sedia ada yang telah diiktiraf. Keputusan kajian menunjukkan bahawa model pengelasan yang dicadangkan mengatasi model yang sedia ada dan menyumbang kepada peningkatan yang ketara berbanding dengan pengelas sedia ada. Ketepatan model ini diuji menggunakan set data daripada syarikat telekomunikasi tempatan yang mencapai 90.32%. Kesimpulannya, keputusan tersebut membuktikan bahawa model pengelasan berasaskan Teori Set Kasar mampu mengelaskan perpindahan pelanggan berbanding pengelas sedia ada.

CUSTOMER CHURN CLASSIFICATION IN TELECOMMUNICATION COMPANY USING ROUGH SET THEORY

ABSTRACT

Churn is perceived as the behaviour of a customer to leave or to terminate a service. This behaviour causes the loss of profit to companies because acquiring new customer incurred high investment for advertisements and promotions compared to retaining existing ones. Thus, it is necessary to consider an efficient classification model to reduce the rate of churn. In the traditional approach of classification modelling, it do not produce straightforward result interpretation. Therefore, identifying the best classification model to reduce the rate of churn is indeed a challenging task. The main objective of this thesis is to propose a new classification model based on the Rough Set Theory to classify customer churn. This research utilized the Knowledge Discovery in Database (KDD) process involving data pre-processing, data discretization, attribute reduction, rule generation, classification process, as well as data analysis, using the Rough Set toolkit. The Rough Set theory elements consist of indiscernibility relation, lower and upper approximations, as well as reduction set. Those elements are applied to classify customer churn from uncertain and imprecise dataset. The results of the proposed model are compared with a few established existing approaches. The results of the study show that the proposed classification model outperformed the existing models and contributes to significant accuracy improvement. The model is tested using dataset form local telecommunication company which achieves 90.32%. In conclusion, the results proved that the classification model based on Rough Set Theory had been capable to classify customer churn compared to the existing model.

ACKNOWLEDGEMENTS

Alhamdulillah, all praises to the almighty Allah S.W.T for giving me good health and strength, and with His blessing, my thesis had finally been successfully completed with my best efforts. This project would have taken much longer to create, if I did not have the help and criticism of the following people. Firstly, I wish to acknowledge my thanks to Universiti Sultan Zainal Abidin (UniSZA), the Malaysian Ministry of Higher Education, Universiti Malaysia Pahang and Majlis Amanah Rakyat (MARA) for giving me the chance to pursue my studies as well as the financial aid rendered all these years.

I am pleased to express my sincerest and deepest gratitude to my supervisor, Dr. Mokhairi Makhtar and Prof. Dr. Mohd Nordin Abdul Rahman for their suggestions, comments, assistance and generous guidance for improvements during the preparation of this thesis. I am very grateful to them for reviewing the thesis and making valuable suggestions. I would also like to address my special appreciation to the many lecturers in the Faculty of Informatics and Computing (FIK) for their cooperation, help and strong support. I am also thankful to all the staff members of the Graduate School for their full cooperation and help. This acknowledgement would be incomplete if I did not mention the emotional support and blessings from my friends. I had a pleasant, enjoyable and fruitful company with them.

Last but not least, I would like to express my greatest thanks to my beloved family, especially my husband and mother for her encouragement, patience, support, financial support and all the sacrifice she had given me during the course of this thesis.

APPROVAL

I certify that an Examination Committee has met on 23rd February 2015 to conduct the final examination of Nur Syafiqah Mohd Nafis, on her Master of Science thesis entitled 'Customer Churn Classification in Telecommunication Company using Rough Set Theory' in accordance with the regulations approved by the Senate of Universiti Sultan Zainal Abidin. The Committee recommends that the candidate be awarded the relevant degree, and it has been accepted by the Senate of Universiti Sultan ZainalAbidin as fulfilment for the Master of Science. The members of the Examination Committee are as follows:

Saiful Bahari Mohamed, PhD.
Associate Professor
Faculty of Innovative Design and Technology
Universiti Sultan Zainal Abidin
(Chairperson)

Mumtazimah Mohamad, PhD.
Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin
(Internal Examiner)

Rozaida Ghazali, PhD.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia
(External Examiner)

PROF. DR. MUSTAFA BIN MAMAT, PhD.
Dean of Graduate School
Universiti Sultan Zainal Abidin

Date: 8/6/2016

DECLARATION BY CANDIDATE

I hereby declare that this thesis is based on my original work except for quotation and citation, which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Sultan Zainal Abidin or other institutions.

NUR SYAFIQAH MOHD NAFIS
Date: 7/6/2016

DECLARARTION BY THE SUPERVISORS

This is to confirm that:
The research conducted and the writing of this thesis was under our supervision.

Signature: ____________________
Name of Main Supervisor: Dr. Mokhairi bin Makhtar
Faculty: Informatics and Computing

Signature: ____________________
Name of Co-Supervisor : Prof. Dr. Mohd Nordin Bin Abdul Rahman
Faculty: Informatics and Computing

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

UMP

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BP	Back-propagation
CRM	Customer Relationship Management
CSV	Comma Separate Values
EFB	Equal Frequency Binning
KDD	Knowledge Discovery in Database
LEM2	Learning from Example Module 2
LHS	Left-Hand-Side
MLP	Multi-layer Perceptron
PHP	Hypertext Pre-processor
RHS	Right-Hand-Side
ROSETTA	Rough Set Toolkit for Analysis Data
RST	Rough Set Theory
RSES	Rough Set Exploration System
UCI	University of California, Irvine
WEKA	Waikato Environment for Knowledge Analysis

UMP

INTRODUCTION

1.1 Research Background

Due to rapidly growing industries, the volume of data stored by telecommunication companies has increased. These data are valuable for business strategies. Customer records, which consist of hundreds or thousands of attributes, are useful for data mining purposes. Data mining is the process of extracting and analysing patterns, relationships and useful information from massive databases (Oseman, Mohd Shukor, Abu Haris, & Abu Bakar, 2010). Modern massive databases are large and consist of noise, as well as, mixed data of numerical and categorical predicators (Tuv, Borisov, Runger, & Torkkola, 2009) which are useful for marketing strategies.

In the recent past, data mining techniques have been extensively used in many fields; for example, mining student's academic performance (Abdul Aziz, Ismail, & Ahmad, 2013) and developing innovative applications in agriculture (Cunningham & Holmes, 1999). Classification is one of the techniques available in data mining. Classification refers to a process of pattern extraction and categorisation. Rapid research developments in the classification field allows researchers to develop classification models for customer churn using various data mining techniques such as soft computing and classical statistic approach. However, classical statistic approach cannot answer broad questions derived from large sample datasets (Yahia & El-Taher, 2010). On the other hand, soft computing such as Rough Set Theory (RST) is very popular in the classification problem.

This research discussed how to construct and improve the classification model for a local telecommunication company using RST. RST was selected to model the classification task since it is widely used in many fields such as email classification (Zhoa & Zhang, 2005), multimedia data management (Abdul Rahman, Mohd Lazim, & Mohamed, 2011) and disease diagnosis (Anouncia, Clara Madonna, P, & Nandhini, 2013) . Three sets of data were used in this research. The local telecommunication company dataset was chosen as a case study for the research. Moreover, data retrieved from the University of California, Irvine (UCI) and IBM Watson Analytics repository were utilised for validation and verification of the proposed classification framework.

1.2 Research Motivation

Telecommunication is a very important need, because without it, businesses will shut down and people will have a hard time performing daily activities. Speedy developments in the telecommunication industry have caused a vigorous competition between local telecommunication companies. In addition, customers have the right to switch to their desired service provider. In the process of retaining existing subscribers, telecommunication companies require a precise and accurate classification system. Misclassification will increase expenses in retention programs.

For that reason, classification using RST was applied to classify customer churn. RST-based classifier assisted in handling uncertain and imprecise customer churn data. With the combination of feature selection, feature reduction and rule filtering, it produced classification accuracy with great improvement. In addition, RST also offers a human-readable if-else rule which assisted in developing a customer churn classification system.

1.3 Problem Statement

Due to higher demands from telecommunication customers in Malaysia, this industry is rapidly and fiercely growing (Ong, Tan & Andrews, 2012). The rapid evolution from the digital telephone to the smart phone, as well as, the computer made telecommunication's style and control more complex. In addition, the telecommunication industry covers a wide range of innovative technologies from the daily call plan to the internet service provider. Currently, there are more than five telecommunication service providers in Malaysia. Thus, customers have many options to choose, from one company to another, which satisfies their needs. It is vital to identify customer churn in the near future since acquiring a new customer is more expensive compared to retaining existing subscribers (Kirui, Hong, Cheruiyot & Kirui, 2013).

Company revenues are affected by churn. A high churn rate will increase costs or decrease service quality (Godfrey, Shenker & Stoica, 2006). The fact is, the telecommunication industry experiences an average of 30 to 35% annual churn rate and it costs 5 to 10 times more to recruit a new customer than to retain an existing one (Lu, 2002). Therefore, keeping existing customers should be the focus. If the telecommunication companies are unable to keep their existing customers to continuously subscribe to the service, they will automatically increase the churn rate. Hence, predicting customers is not an easy task. In addition, various and complex customer behaviours make it even difficult. A churn classification model should be precise and accurate to classify churners and non-churners because failure to predict customer churn will lead to an increase in the customer churn rate.

However, researchers had proposed various methods that can help a telecommunication company to classify customer churn such as soft computing, neural network, and decision tree and regression analysis. As classifying customer churn is an extremely difficult undertaking, this research proposed a classification model using the RST classification. Classification is the most popular data mining technique to classify customer churn in the telecommunications industry. The RST was chosen as a technique for classification because it offers efficient algorithms for finding hidden patterns in data (Pawlak, 1996). In addition, it produces a minimal human readable set of decision rules with straightforward result interpretation. As a result, a new classification framework and model, which yields the highest classification accuracy, was considered in this research.

1.4 Research Objectives

This research target achieves the following objectives:

i. To propose a new classification model to classify customer churn using the RST.
ii. To implement the proposed model with feature selection, attribute reduction method and rule filtering.
iii. To compare and evaluate the RST classification model with non-RST classifiers (Neural Network, Decision Tree and Regression Analysis).

1.5 Research Scope

In order to successfully achieve these objectives, it is crucial to identify the research scope clearly as follows:

i. This research will focus on deliberate voluntary customer churn classification using the RST.
ii. This research will construct the model using two main components which is the customer churn dataset and the RST.
iii. This research will utilize the Rough Set Toolkit for Analysis Data (ROSETTA) (built-in feature reduction and rule filtering) and Waikato Environment for Knowledge Analysis (WEKA) (Feature Selection) to analyse the capability and performance of the newly generated model for customer churn classification.

1.6 Thesis Organization

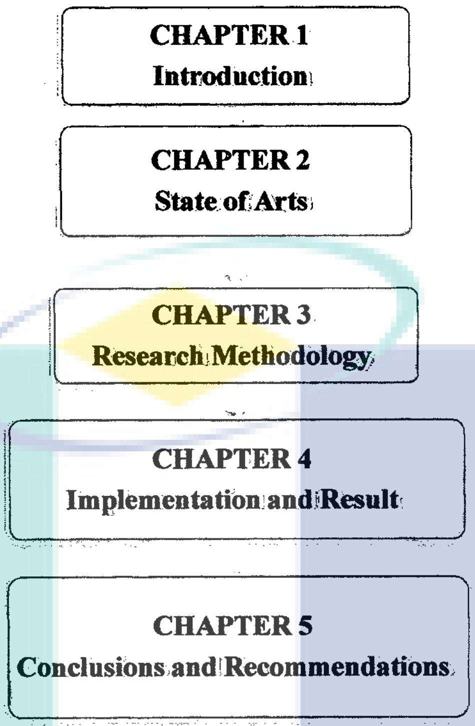


Figure 2.1: Thesis Structure

This thesis consists of five (5) chapters, as shown in Figure 1.1. Chapter 2 discussed the related past researches for customer churn management, customer churn classification modelling, the classification approach taken and RST concept. The main focus of Chapter 2 was to find a research gap between the existing approach and the approach taken in this research. The research gap was provided at the end of Chapter 2. Next, in Chapter 3, the three research methodology framework was explained in detail. In Chapter 4, the overall research implementations and findings on the proposed classification technique was discussed. Finally, in Chapter 5, the summary on the research findings, research contributions, research limitations and suggestion for future work were discussed in detail.

1.7 Concluding Remarks

In this introductory chapter, the research background and motivation is outlined. The research problem statements, objectives, scopes and contributions of the research also will be provided in order to fulfil this research aims to classify customer churn. Finally the thesis organization is presented at the end of this chapter.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Identifying and classifying customer churn are essential for company profit. Since competition in the telecommunications industry is very fierce, many carriers consider reducing churners as it is an important business venture to maintain profitability (Au, Chan, & Yao, 2003). Jadhav and Pawar (2011) stated that the biggest revenue leakages in the telecommunication industry are the increase of customers' churn behaviour. In order to counter this problem, firms must recognize churners before they churn, so that in the near future, firms or service providers can provide a special personalized offer and service to those customers who are classified with higher likelihood to churn. Hence, the key to survive and win the competition in industries is to encounter churning problems.

So, in this chapter, it recalls on the approaches and methodologies used in previous works in solving customer churn classification in multiple fields especially in telecommunication area. The importance of literature review is to give a deeper understanding towards the selected technique in order to complete the comparative study and get the most effective and accurate results. At the very beginning of this chapter, it will explain on customer churn concepts in section 2.2. Next in Section 2.3, it will provide reviews on previous work on data mining approach in customer churn. Meanwhile, Section 2.4 will discuss on the customer churn classification approaches and RST concepts are explained in Section 2.5. Identified research gaps are provided in Section 2.6. Finally, concluding remarks on Section 2.7. Figure 2.1 shows the literature review structure for this thesis.

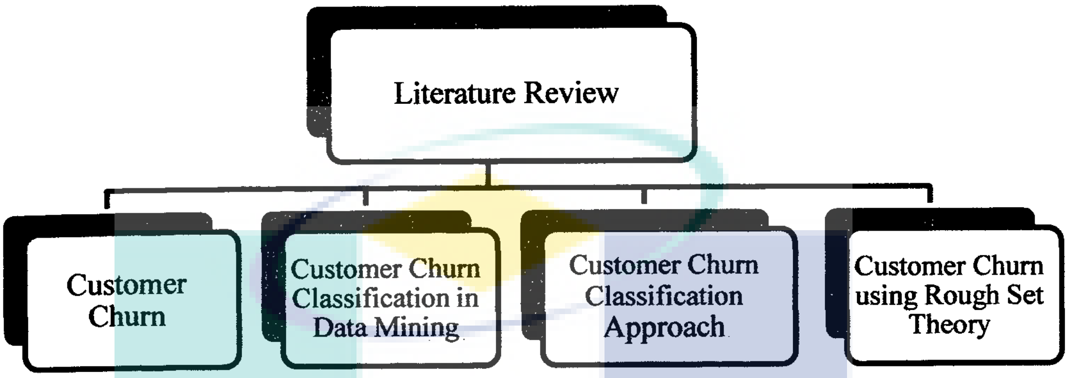


Figure 2.1: Literature Review Structure

2.2 Introduction to Customer Churn

Literally, churn is defined as the turnover action by the customer of a business or services to the competitor. Recent phenomenon has shown that telecommunication is becoming the number one need for the young or the old, individuals or corporate members worldwide. This phenomenon affects many telecommunication companies that need to provide excellent services in order to win over the competition. Thus, customers churn should be the number one concern of the firms (Khan, Usman, Usman, Rehman, & Ateeq, 2013; Oseman et al., 2010) and it is crucial to classify which customer will shift to the competitor since acquiring a new customer is more costly compared to retaining existing subscribers had been clearly mentioned in previous research (Huang et al., 2015). Furthermore, there are also common clashes where customers want good service with cheaper price while the firms are strictly focusing on their business goals, which may increase the company's churn rate. Hence, customer churn prevention should be taken into action in order to minimize churn.

Churn can be distinguished into two types; voluntary churn and involuntary churn (Hadden, 2008). Involuntary churn occurs when customers are disconnected by the service provider for fraud or non-payment. Voluntary churn can be varied and more complex. However, it can be differentiated between incidental churn and deliberate churn. Incidental churn includes the customers' financial problems and the customers relocating to a new geographical location where the company's service is not provided. For deliberate churn, the customer decides to terminate the service due to poor services provided and choose to subscribe to a new service with the competitor. Companies should put more effort on preventing deliberate customer churn since there are many reasons for deliberate churn to occur such as a better offer from competitor, bad network service and technology matters. The focus in this research is to overcome deliberate voluntary churn.

UMP

Overall, good customer management must be able to capture feedback from customers (Ismail, 2015). Feedbacks gathered from customers should be able to convert into superb business strategies to offer great services to the subscribers. Thus, churn will be avoided.

2.2.1 Customer Churn Management

Customer churn management can be defined as the service provider's strategies to prevent churn from occurring. Good customer churn management will ensure the service provided meet the customer's satisfactions to avoid churn. Referring to Kojo (2011), a well-functioning churn management will retain high value subscribers resulting in an increase in Average Revenue Per User. Customer churn management should be the main focus for many telecommunication companies because it can increase profitability and decrease churn rate. If the churn prevention program is effective, firms can look forward to reach benefits from its efforts.

Econsultancy is the leading source of independent advice organisation and insight on digital marketing and ecommerce founded in 1999. In order to measure the awareness on churn management in a company, *Econsultancy* led a survey. *Econsultancy* had published Cross-channel Marketing Report 2013 which denied some compelling statistics for those who believe that customer retention is undervalued. It is illustrated in Figure 2.2 on how a company concerns on relationship marketing or churn management. It shows that only 30% companies committed in customer relationship management, while 48% committed on certain cases of churn and about 22% really do not care about relationship marketing which reflect that awareness on customer relationship is still the least.

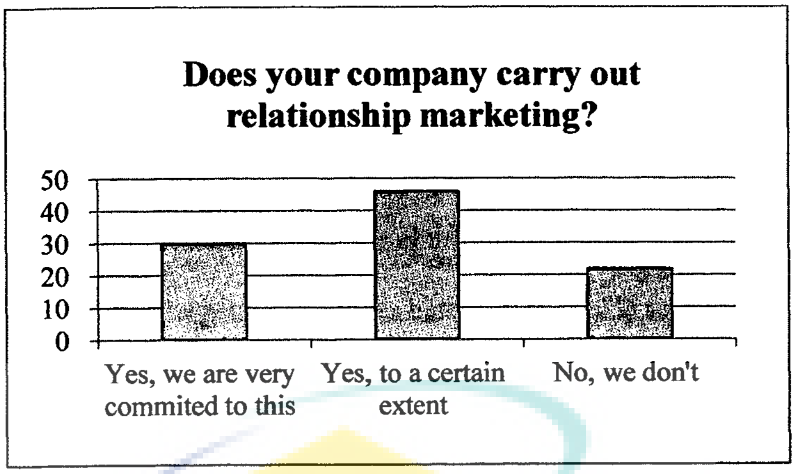


Figure 2.2: Customer Relationship Awareness

According to Khan, Jamwal, and Sepehri (2010) , there are two basic approaches to fight customer churn which are untargeted and targeted approach. Relying on superior products and mass advertisement to increase brand loyalty and retain customer are categorized as untargeted approach while in target approach, firms will identify customers who are likely to churn and offer direct incentives to avoid churn. Targeted approach is divided into two categories; 1) reactive approach and 2) proactive approach. For reactive approach to occur, the company will wait until customers contact it to terminate their subscription. Then, the company offers incentives on the spot, for example a rebate or cheaper packages for a customer to stay.

Meanwhile, in proactive approach, the company tries to identify subscribers who are likely to churn in the near future, then targets these customer with special packages or incentives to keep them from churning. Proactive approach has an advantage in lowering incentive cost. However, this approach seems a waste if the customer is not accurately classified because firms will waste on incentive money for customers who would stay away. For that reason, customer churn classification should be as accurate as possible. In conclusion, this research is one of the proactive approaches to classify deliberate customer churn. As a conclusion, a proactive action should be applied on the targeted customers before they decide to move to competitors (Rashid, 2008).

Therefore, investment in Customer Relationship Management (CRM) technology is important. The prevention of customer churn through customer retention is a core issue of CRM (Prasad & Madhavi, 2012). Besides, many highly competitive organizations understood that retaining existing and valuable customer is their core managerial strategy to survive in the industry (Tsai & Chen, 2010). Consequently, analysing and controlling churn are critical for any telecommunication company to improve its revenues.

2.2.2 Customer Churn Effects to Telecommunication Company

Customer churn has a negative impact on a telecommunication company's revenue. The customer movement is closely related to the customer retention rate and loyalty (Kojo, 2011). If customer retention is effective, customer churn prevention program could be avoided because offering prevention program to the customer who does not have intention to leave will effect on the cost expenses. Thus, customer churn prevention program cost can be cut off. In addition, churn rate has a strong impact to the long term value because it affects the length of service and the future revenue.

Understanding the true value of possible churners will help the company in customer relationship management because a small change in retention rate will impact the business significantly (Van den Poel & Larivière, 2004).

Besides, an increase in customer churn rate will have an impact on the company's reputation as well. Dissatisfied customers may tell others about their experience. Customers who are still subscribing the service may be in doubt to continue using the services and also decide to churn. However, high churn rates and substantial revenue loss due to churning have turned correct future churn prediction and prevention to a vital business process (Lazarov & Capota, 2007).

2.3 Data Mining in Customer Churn Classification

Data Mining is one important step in the Knowledge Discovery in Database (KDD). It is a process of extracting the hidden information from data which is yet unknown and potentially beneficial (Bal, 2013). Knowledge Discovery in Database (KDD) is the most popular and established method used in many fields such as educational, financial, medical and engineering for data extraction (Ozer, 2008). Figure 2.3 summarizes the KDD. KDD phases are briefly explained as follows (Fayyad, Piatetsky-Shapiro, & Smyth, 1996):

i. *Selection:* Selecting the correct dataset to create *target data* to represent the tabular dataset. Dataset must be flat and two-dimensional data tables. The selection of the most significant features was also involved in this selection phase.

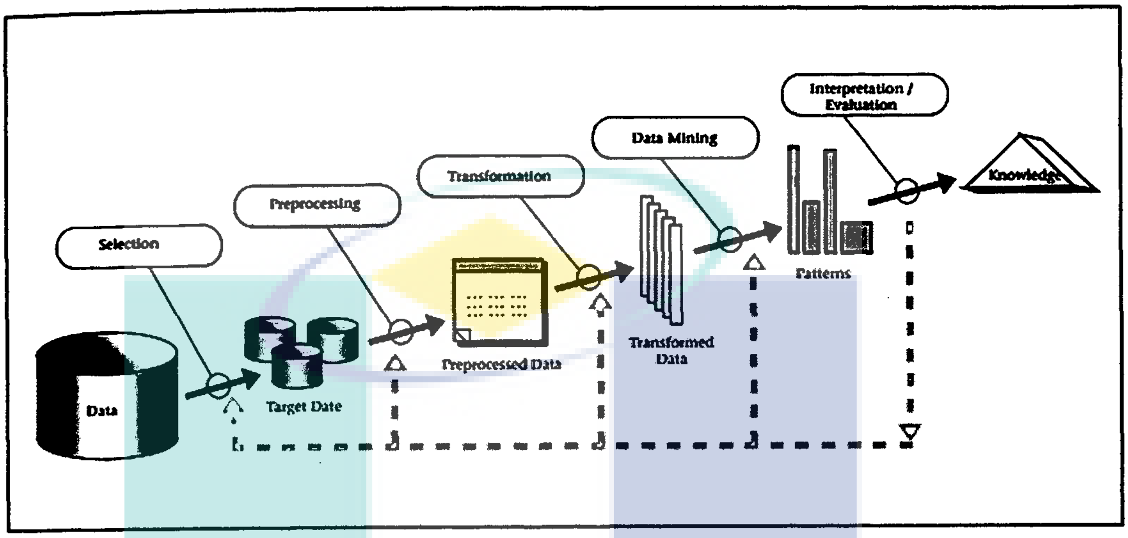


Figure 2.3: Overview of KDD Process

ii. *Pre-processing:* Raw data from database may be incomplete and contain noise. Data must be processed to yield a clean tabular data.

iii. *Transformation:* Transformation process included pre-processed data normalization and discretization. It involved the searching for consistent partitions of attribute domains into intervals and unifying the values over each interval on numerical attributes to produce a transformed data.

iv. *Data Mining:* RST if-then rules were mined in a two-step process. At this step, minimal attribute subset was produced, followed by rules generation from these subsets using transformed data to build a classification model which contained rules or patterns.

v. *Interpretation/Evaluation:* Patterns and rules can be measured and manually inspected. Patterns or rules can be used to classify objects and classification performance can be measured using confusion matrix.

In practise, *'predictive'* and *'descriptive'* are the two high-level goals in data mining. Predictive data mining aims to predict the performance of various variables while descriptive data mining generates new data and performance while describing the behavioural patterns of variables based on the available data set. This research will focus on the predictive goal of data mining. Classification, regression and clustering are the most common task in data mining

Classification can be simply referred as process to categorize objects according to the characteristics of the objects. In data mining, classification is defined as an analysing task for a set of pre-classified data objects to study a model or function that can be used or applied to unseen data objects before being placed into one of several predefined classes (An, 2005). A robust study on classification had caused for the emergence of many classification techniques, for example the artificial-based model, Bayesian-based model, regression model, rule-based model and so on (F. Li, Lei, Tian, Punyapatthanakul, & Wang, 2011). However, a thorough search of classification methodologies used in literature revealed that no one methodology can be served as a recipe for a comparison of all computational classification methods (Al-radaideh, 2008). From the broad and deep search on previous research, classification has been widely used in many domains such as the application of images classification (Kamarudin et al., 2015), text classification (Zhang, Li, Sun, & Nadee, 2013), classification in multimedia data management (Abdul Rahman et al., 2011) and classification of multispectral image (Perumal & Bhaskaran, 2010).

2.4 Customer Churn Classification Approaches

In the classification area of study, there are various techniques that exist. The most common techniques are artificial neural network, decision tree and regression analysis. In order to evaluate the RST classification, the concepts of these three classifiers (artificial neural network, decision tree and regression analysis) were also investigated.

2.4.1 Artificial Neural Network (ANN)

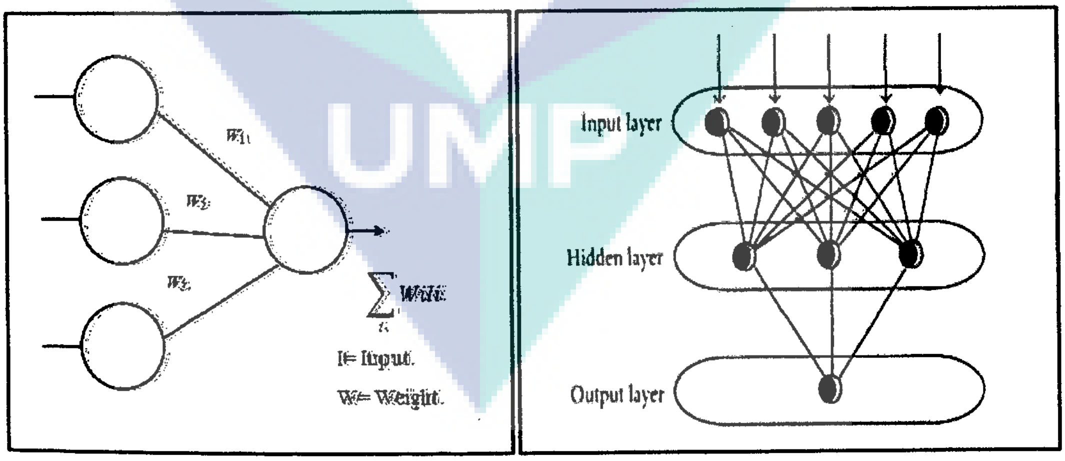


Figure 2.4: Single Layer Perceptron Neural Network and Multi-layer Perceptron Neural Network

Artificial neural network (ANN) can be defined as a model of thought which mimic human brains. In ANN, nodes and links connected between them. ANN consists of three main elements which are weight, bias, and activation function. ANN can be categorized into single layer or multi-layer perceptron (MLP). Figure 2.4 illustrates three elements in ANN for single layer and multi-layer perceptron (Mohammadi, Tavakkoli-Moghaddam, & Mohammadi, 2013).Through deep research in churn management, it was found that many researches had been done implementing ANN. Rashid (2008) implemented dynamic training to classify non-churners and churners. Dynamic training is based on two ANN configurations to reduce the total erroneous network. Dynamic training is selected to improve the prediction performance of the standard ANN. This dynamic training approach performed better than artificial neural network alone in customer churn classification which proved that the new dynamic training of ANN can solve customer churn classification problems in standard ANN approach.

Meanwhile, in a study conducted by Ismail, Awang, Abdul Rahman, and Makhtar (2015), they proposed a multilayer–perceptron (MLP) ANN algorithms to predict customer churn in the local telecommunication companies in Malaysia. The study chose nine (9) algorithms from the MLP neural network to train and test the dataset. Among these nine algorithms, Levenberg-Marquardt gave the best performance accuracy. In order to evaluate the MLP neural network algorithms, two classifiers were selected; Multiple Regression Analysis and Logistic Regression. Result which showed MLP neural network is better than Multiple Regression Analysis and Logistic Regression Analysis techniques in classifying churners and non-churners.

Overall, this study proved that ANN classifier offers a feasible alternative in classifying customer churn compared to traditional predictive approach.

2.4.2 Decision Tree

Another popular classifier applied in customer churn risk analysis is the decision tree. Decision tree is a nonparametric approach for building classification models (Bin, Peiji & Juan, 2007). In the decision tree, it defines *'nodes'* to classify objects. There are three types of nodes: root node, internal node and leaf or terminal node. Leaf or terminal node is assigned to a class label, while non-terminal node (internal and root node) contains attribute test conditions to separate records that have different characteristics. Figure 2.5 showed the example of mammal classification problem using the decision tree (Tan, Steinbach, & Kumar, 2006).

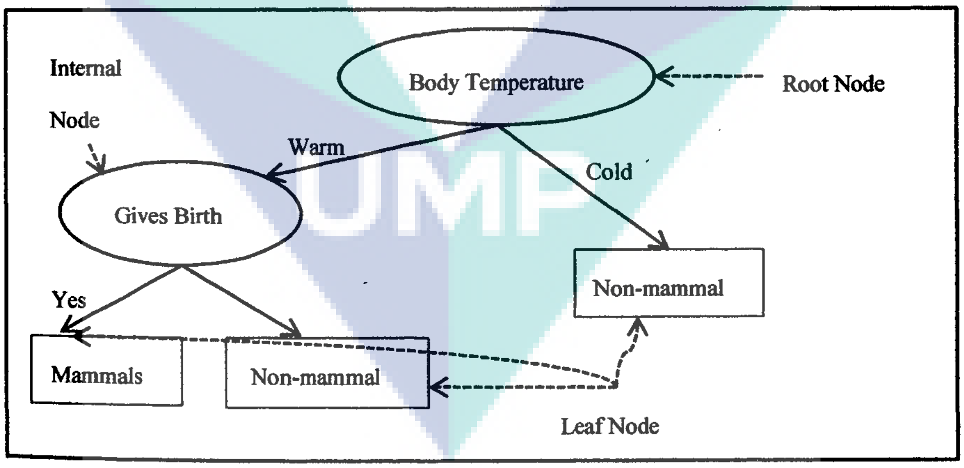


Figure 2.5: Example of decision tree classification

According to Hongxia and Min (2009), the decision tree is feasible and effective enough to classify customer churn. In their research, through the construction of the decision tree and data classification, business factors and high possibility churners were found. Moreover, some rules were also defined. However, this research analysed the characteristics and factors that lead to customer churn instead of classifying churners and non-churners.

2.4.3 Regression Analysis

Regression analysis is a statistical classifier approach to investigate relationships between variables. It was stated by Lazarov and Capota (2007) that regression analysis is a good technique for identifying and predicting customer satisfaction. Equation 2.1 displayed a simple linear regression:

$$y = a + bx + e_i \tag{2.1}$$

Where,

y = dependent variables or predicted values,

α = constant population value when the value of x is zero,

x = value of independent variables,

b = constant of independent variables (slope for the population), and

e_i = error values or noise or disturbance.

Gürsoy and Tuğba (2010) conducted a research to determine reasons for churning in a major telecommunication company in Turkey using logistic regression analysis and the decision tree. Results showed that the accuracy of correctly predicting non-churners was 74.3% and the accuracy of correctly predicting churners was only 66%. Even though the prediction accuracy was low, the logistic regression results were significant according to the significance level which was less than 5%.

The innovation and implementation of customer churn management research can be varied. Hadden, Tiwari, Roy and Ruta (2006) utilised customer's complaint data to classify churn. This research also provided a comparative analysis of standard neural network, Bayesian neural network, regression tree and regression in order to evaluate the classifiers' capabilities in classifying churners. Overall, the decision tree was the best among the tested classifiers (82% accuracy) while the standard neural network performed the worst (72% accuracy). However, in terms of classifying churners, the Bayesian neural network was the most accurate with 70% accuracy. Table 2.1 briefly explained the various methods applied in customer churn classification in different domains using different methods.

Table 2.1: Various Methods in Customer Churn Classification

Case Study	**Method**
Wireless Industry (Nath & Behara, 2003)	Naïve Bayes
Telecommunications Industry (Oseman et al., 2010)	Decision Tree Analysis
Telecommunication Company (Jadhav & Pawar, 2011)	Neural Network (Back Propagation)
Insurance Customer (Soeini & Rodpysh, 2012)	Churn Index and Decision Tree CART.
Mobile Telephony Industry (Kirui et al., 2013)	Naïve Bayes and Bayesian Network
Australia Nursing Industry (Dawson, Stasa, Roche, Homer, & Duffield, 2014)	Qualitative Design

2.4.4 Rough Set Theory

Rough Set Theory had been introduced by Pawlak in 1982 had attracted many researchers' and practitioners' attention because these new mathematical approach tackle imperfect and uncertain knowledge.

From a broad and deep research in solving customer churn, RST is one of the common approaches used. For example, researchers had utilized RST and back-propagation (BP) neural network to solve customer churn problem (Xu, Liangshan, Xuedong, & Baofeng, 2006). They used RST attribute discretization and attribute reduction algorithm to refine initial data. After being refined, data is said to be more precise in structure and more convenient to be applied in BP neural network classifier. Classification accuracy using RST refined data is higher compared to initial classification without data refining. At this point, it proved that RST-based attribute discretization and reduction is powerful to assist in improving classification accuracy.

Hu (2008) also had employed RST in classifying customer churn based on historical data by proposing RST-based feature reduction algorithm based. This study aimed to implement RST to mine historical customers' data to infer useful rules for the customers that will probably churn. In addition, this study also produced the causal relationships model between the customer churn and the evidences of observations customers' behaviour. The results have shown the practical viability of the RST-based data mining approach for predication of customers churn and extracting rules. Extracted rules can be used to build a CRM decision support system for future work.

Lin, Tzeng, and Chin (2011) combined RST and flow network graph to predict customer churn in credit card in Taiwan. This study utilized RST as a rule-based decision-making technique to extract rules related to customer churn and then uses a flow network graph (path-dependent approach) to infer decision rules and variables. Finally, they present the relationships between rules and different kinds of churn. As a result, this hybrid model can fully predict customer churn and provide useful information for decision-makers in devising marketing strategy. Furthermore, empirical study shows that combining RST and flow network graph is very effective to support CRM decision making as RST expressed decision rules in natural language which can be easily deduced and understood for decision-makers to make accurate prediction and act accordingly.

Amin, Shehzad, Khan, Ali and Anwar (2015) also utilised RST in solving the customer churn problem. They explored four RST-based reduction algorithms which were Exhaustive, Genetic, Covering and Learning from Example Module (LEM2) to identify the most appropriate algorithm for generating rule sets for rough set classification approach in the telecommunication sector. The obtained results demonstrated that genetic algorithm rules generation yielded the most suitable performance out of the four rules generation algorithms. This may assist decision-makers in strategic decision making and planning of the process of customer churn. However, rules obtained from these experiments were not filtered. Rules' filtering is necessary in order to remove any unused rules which do not assist in improving the classification accuracy. Therefore, this research also aimed to identify which features are more indicative for churn prediction in the telecommunication sector.

Overall, RST promised an efficient and feasible algorithm to find hidden patterns and rules in data mining. These hidden patterns and rules can be found through data reduction to make up a minimal set of data. According to Bal (2013), RST has advantages in locating minimal datasets. Patterns and rules are in human-readable format which can be easily understood. Furthermore, results obtained can be clearly interpreted and are suitable for parallel processing. In addition, it does not require any preliminary or additional information about data, just like probability in statistics and grade of membership in the fuzzy set theory. Therefore, implementing the Rough Set Theory for customer churn classification modelling is currently relevant.

2.5 Customer Churn Classification using Rough Set Theory

Rough Set Theory is one of the new generation techniques available for classification. It is an extend from the traditional set approach (Pawlak, 1996). Since then, RST has attracted many researchers and practitioners in various fields of science and technology. In classification area, RST has been applied in many real-life applications such as in image segmentation (Senthilkumaran & Rajesh, 2009), marketing evaluation (Jiang & Ruan, 2010), medical diagnosis (Tripathy, Acharjya, & Cynthya, 2011), stock prices (Khoza & Marwala, 2011) and multimedia data management (Abdul Rahman et al., 2011). The advantages of utilizing RST over other techniques are it does not need any preliminary or additional information about the data and it offers straightforward interpretation from the obtained results (Suraj, 2004). Moreover, RST is employed to represent imprecise and uncertain information.

Rough set deals with data analysis in a tabular format called the decision table which make up an information system. Figure 2.6 illustrates a decision table. Each row in the table represents an object, for instance a case or event. Meanwhile, each column in the table represents an attribute or feature of the object such as a property or a variable. There are two types of attributes which are condition attribute and decision attribute and each object is assigned with some attribute values. The Rough Set theory elements consist of indiscernibility relation, lower and upper approximations, as well as attribute reduction.

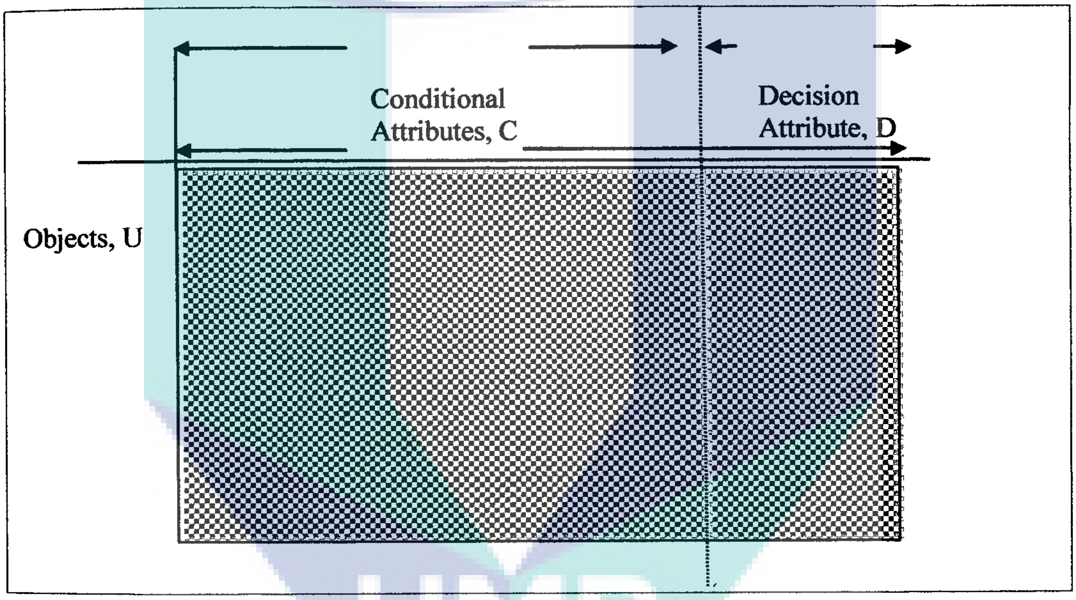


Figure 2.6: Decision Table Illustration

The relation between two objects or more is called indiscernibility where all the values are identical in relation to a subset of a considered attribute. Given a subset of attributes, $\alpha \in A$ *and* $B \subseteq A$, each such subset defines an equivalence relation $IND_A(B)$ called an indiscernibility relation. This indiscernibility relationship can be defined as follows:

$$(B) = \{(x, x) \in U^2 \mid \forall \alpha \in B, \alpha(x) = \alpha(x)\} \quad 2.2$$

From equation 2.2, the subset of attributes will be defined as a classification process of the universe into sets such that each object in a set cannot be distinguished from other objects in the set using only the attributes in *B*. The sets which the objects are divided into are known as equivalence class.

The RST concept of border can be expressed by lower and upper approximation. The concept of approximation is required to class the objects based on the equivalence class. Now, to define two approximations are called the P-lower and the P-upper approximation of *X* respectively where,

$$\underline{P(X)} = \{x \in U : [x]P \subseteq X\} \quad 2.3$$

$$\overline{P(X)} = \{x \in U : [x]P \cap X \neq \emptyset\} \quad 2.4$$

Figure 2.7 illustrates RST approximation concept. The lower approximation is the set that contains all objects for which the equivalence class corresponds to the object which is the subset of the set. This set contains all objects which certainly belong to the set *X*. Meanwhile, upper approximation is the set containing the objects for which the intersection of the object equivalence class and the set is not the empty set. This set contains all objects which possibly belong to set *X*. The boundary of *X* for the given $B \subseteq A$ and $X \subseteq U$ in IS can be defined as:

$$PN_P = \overline{P(X)} - \underline{P(X)} \quad 2.5$$

PN_P consists of objects that certainly do not belong to X on the basis of A.

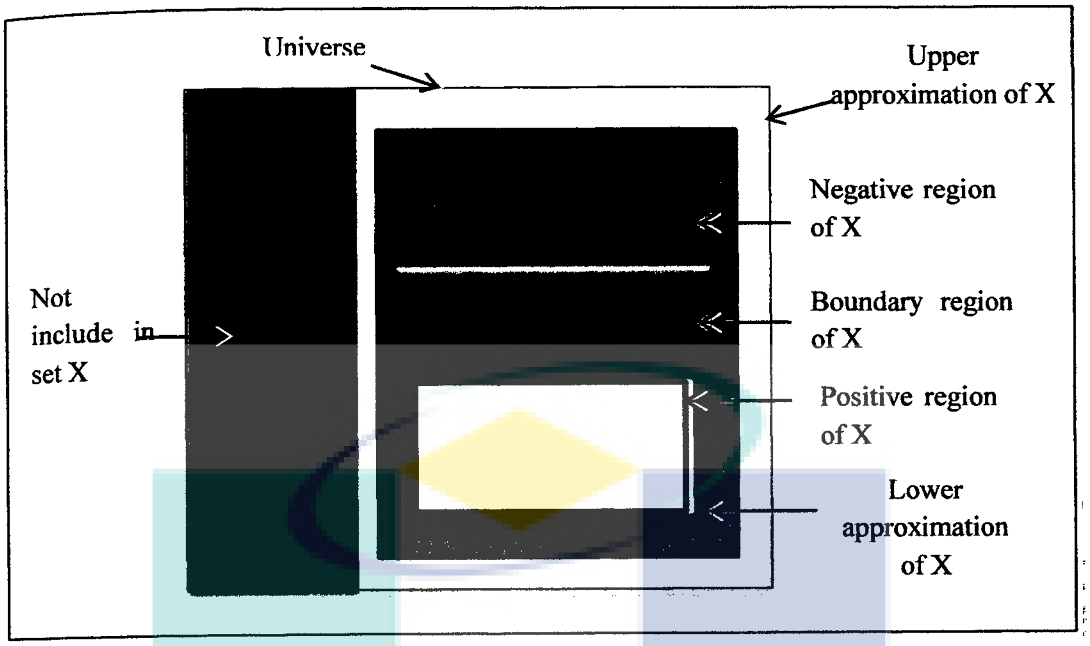


Figure 2.7: RST Approximation Concepts

Table 2.2: Example of Information Table

Object	Conditional Attribute				
	Gender	Packages	Monthly commitment	Call Plan	Service Rental
O_1	1	1	1	1	1
O_2	1	2	2	1	4
O_3	1	3	3	2	2
O_4	1	4	4	2	2
O_5	0	4	4	2	2
O_6	0	5	5	2	2

In Table 2.2, it shows the example of an information table. Then, let $IS = (U, A, C, D)$ be the information system in which U is a non-empty finite set called universe, and A is an attribute set; A consists of condition attribute C and decision attribute D such that $A = C \cup D$ and $C \cap D = \emptyset$ .D is not necessarily constant on the equivalence class. Therefore, two objects may belong to the same equivalence class but the decision attribute may be varied. For example, if D is inserted into IS in Table 2.2, it will produce an IS as in Table 2.3.

Table 2.3: Example of Decision Table

Object	**Condition Attribute**					**Decision Attribute**
	Gender	Packages	Monthly commitment	Call Plan	Service Rental	Churn?
O_1	1	1	1	1	1	1
O_2	1	2	2	1	4	1
O_3	1	3	3	2	2	0
O_4	0	4	4	2	2	0
O_5	0	4	4	2	2	1
O_6	0	5	5	2	2	1

From Table 2.3, O_4 and O_5 are having the same equivalence class respected to Packages, Monthly Commitment and Call Plan attributes but they are classified differently. This information table can be referred to as inconsistent. As a solution, a condition application of the object is required. Objects O_4 and O_5 in Table 2.3 also can be classified as indiscernibility relation. The example of indiscernibility relations are:

i. *IND (Gender)* = $\{\{O_1, O_2, O_3\}, \{O_4, O_5, O_6\}\}$

ii. *IND (Call Plan, Service Rental)* = $\{\{O_1\}, \{O_2\}, \{O_3, O_4, O_5, O_6\}\}$

iii. *IND (Gender, Packages, Monthly Commitment)* = $\{\{O_1\}, \{O_2\}, \{O_3\}, \{O_4, O_5\}, \{O_6\}\}$

Based on Equation 2.3 and 2.4, the lower approximation and upper approximation of X from Table 2.3 be classified as follows:

$$\underline{P(X)}_{\square} = \{O_1, O_2, O_3, O_6\}$$

$$\underline{P(X)}_{\square} = \{O_1, O_2, O_3, O_4, O_5, O_6\}$$

And boundary has been explained in Equation 2.5 can be defined as, $PN_P = \{O_4, O_5\}$.

A reduction set or the so-called 'reduct' is a minimal set of attribute after removing redundant and insignificant attributes but preserve the original classification. In some cases, not all attributes are required to classify an object. A reduct of A is defined as minimal set of attributes in original classification defined by $B \subseteq$ A such that $IND_A(B) = IND_A(A)$. In this designed model example, to discern between the different equivalence classes, only attributes Packages and Call Plan are necessary and the example of reduct is:

$$IND_A(\{Packages, Call\ Plan\}) = IND_{(A)}$$

Table 2.4 shows example of the decision table after the reduction process which attributes Gender, Monthly Commitment and service Rental are dropped. As a result, decision rules in row 4th and 5th in Table 2.4 have the same conditional attributes but different decisions. Thus, rules as known as inconsistent. Meanwhile rules in the 1st and 2nd rows are consistent. In order to handle inconsistency in such decision tables, the RST approximation concept is required has been explained previously.

Table 2.4: Example of Decision Table after Reduction

Object	Decision Attribute		Condition Attribute
	Packages	Call Plan	Churn?
O_1	1	1	1
O_2	2	1	1
O_3	3	2	0
O_4	4	2	0
O_5	4	2	1
O_6	5	2	1

The approximation of the decision, D can be defined by constructing the decision rules from Table 2.4. From this generated approximate decision rules, objects can be classified. From the Table 2.4, rules cannot be exactly classified by approximately only. Basically, rules are presented as the implication *"if....then..."* rules. The rules are constructed as follows:

- Rule 1, if (packages,1) and (call plan,1) then (churn,1)
- Rule 2, if (packages,2) and (call plan,1) and then (churn,1)
- Rule 3, if (packages,5) and (call plan,2) and then (churn,1)
- Rule 4, if (packages,3) and (call plan,2) and then (churn,0)
- Rule 5, if (packages,5) and (call plan,2) and then (churn,1)
- Rule 6, if (packages,5) and (call plan,2) and then (churn,0)

In conclusion, Rule 1, Rule 2 and Rule 3 can be certainly classified as causing churn while Rule 4 can be certainly classified as not causing churn. Lastly, Rule 5 and Rule 6 can be possibly classified as causing churn and not causing churn.

The study of previous research indicated that RST had also been widely used in various areas. For example, RST was successfully implemented to reduce the error rate that discriminated a non-spam to spam in classifying emails (Zhoa & Zhang, 2005). Meanwhile, in the medical research area, RST had been utilised in image segmentation that was carried out by Senthilkumaran and Rajesh (2009). This research successfully identified the segments of medical images according to the images' characteristics; for example, image colour, objects, shapes and positions. The poor medical image contrast and artefacts can affect missing or diffused organ/tissues. Lastly, Tian, Yang and Zhang (2010) used the RST concept to develop road traffic cause analysis. The research proposed the analysis method of road traffic accidents cause based on data mining to greatly reduce personal casualty and property loss due to road traffic accidents. This study was able to provide a powerful data analysis support for the road traffic safety related research by applying RST.

RST also offers a powerful attribute reduction method. Attribute reduction is an important topic in machine learning and a crucial phase in data mining. It becomes the focus of much research in the areas of application for which datasets with tens or hundreds of thousands of variables are available (Guyon & Elisseeff, 2003). The objective of feature reduction is to select the best subset containing the least number of dimensions that most contribute to accuracy; to discard the remaining, unimportant dimensions (Ladha & Deepa, 2011). Besides, it removes unnecessary attributes to reduce processing time. Performing data reduction can significantly reduce the amount of information required in order to discern between the decision rules (Sarkar, 2014). In short, it is a process of finding the optimal subset of features that satisfy certain criteria. Dynamic reduction is one of the approaches used in attribute reduction.

2.5.1 Dynamic Attribute Reduction

Based on Bazan (1998), dynamic reduction was the most stable reduction set of a given decision table because the reduction set was the most frequent reduction set appearing in sub tables created by random samples of a given decision table. Hence, it gave the most stable reduction set of a given decision table. Three dynamic reduction methods were used in these experiments; Dynamic Exhaustive Calculation, Dynamic Johnson's Algorithms and Dynamic Genetic Algorithms. This dynamic reduction is available in Rough Set Exploration System (RSES) built-in library in ROSETTA.

2.5.1.1 Exhaustive Calculation

In exhaustive calculation, searching is done through the subsets of features to find the best feature among competing 2^N candidate subsets according to some evaluation criterion (Dash & Liu, 1997). However, the exhaustive search could only be implemented in a domain where N is relatively small. Large N will make the search intractable in many real world applications (M. Zhang & Yao, 2004). The exhaustive search is not feasible due to its high time complexity since the searching method evaluates all possibilities of the feature subset. This method requires a stopping criterion to prevent feature selection process from running exhaustively or forever through the space of subsets.

2.5.1.2 Johnson's Algorithm

Johnson's reduction algorithm is one of the heuristic algorithms. The Johnson reducer uses a simple greedy algorithm to compute single reduction set only (J. Li, 2007). Hence, it provides a reduction method which is fast and efficient with the minimum overhead (Akbar, 2003). The Johnson's algorithm was outlined as follows (Ohrn, 2000b):

Let B be the reduction set, where S denotes the set of sets corresponding to the discernibility function, and $w(S)$ denotes a weight for set S in S that automatically gets computed from the data.

i. Let $B = \emptyset$.
ii. Let a denote the attribute that maximises $\sum w(S)$, where the sum is taken over all sets S in S that contain a. Ties are currently resolved arbitrarily.
iii. Add a to B.
iv. Remove all sets S from S that contains a.
v. If $S = \emptyset$ return to B. Otherwise, go to step 2.

The support for computing approximate solutions are provided by aborting the loop when "enough" sets have been removed from *S*, instead of requiring that *S* has to be fully emptied. The support count associated with the computed reduction set equals to the reduction set's hitting fraction multiplied by 100; i.e., the percentage of sets in *S* that *B* has a non-empty intersection with.

2.5.1.3 Genetic Algorithm

Genetic algorithm is an evolutionary computing technique used to solve optimisation problems which are inspired by biological evolution. Unlike Johnson's algorithm, genetic algorithms produce multiple numbers of reduction sets. Thus, it is used to reduce the computational cost in large and complex decision tables. Genetic algorithms scan the training set object-by-object and generate rule sets by matching the objects and attributes with the reduction set. The genetic algorithm was explained as follows (Ohrn, 2000b):

The Genetic algorithm has support for both cost information and approximate solutions. The algorithm's fitness function *f* was defined below,

$$f\,(B) = (1-\alpha) \times \frac{cost\,(A) - cost\,(B)}{cost\,(A)} + \alpha \times \min\{\varepsilon, \frac{|[S\ in\ S|S \cap B \neq \emptyset]|}{|S|}\} \qquad 2.6$$

Where,

S = set of sets corresponding to the discernibility function

α = weighting between subset cost and hitting fraction, while

ε = is relevant in the case of approximate solutions.

The subsets B of A that are found through the evolutionary search driven by the fitness function and are "good enough" hitting sets (for example, they have a hitting fraction of at least ε), are then collected in a "keep list". The size of the keep list can be specified. The function cost specifies the cost of an attribute subset. If no cost information is used, then a default unit cost is assumed, effectively defining cost (B) = |B|.

Meanwhile, approximate solutions are controlled through two parameters, ε and k. The parameter ε signifies a minimal value for the hitting fraction, while k denotes the number of extra keep lists in use by the algorithm. If $k = 0$, then only minimal hitting sets with a hitting fraction of approximately ε are returned. If $k > 0$, then k+1 groups of minimal hitting sets are returned, each group having an approximate (but not smaller) hitting fraction evenly spaced between ε and 1. Note that $\varepsilon = 1$ implies proper minimal hitting sets. Table 2.5 summarised the three reduction methods used.

Table 2.5: Summary of Reduction Methods

Name of the Reduction Method	Description
Dynamic Exhaustive Calculation(RSES)	Computes all reduction set by brute force. It is only suitable for moderate size tables because this algorithm do not scale up well. Apply boundary region thinning or approximate solutions.
Dynamic Genetic Algorithm(RSES)	Utilized a genetic algorithm to search for reduction set, either until the search space is exhausted or until a given maximum number of reduction set have been found.
Dynamic Johnson's Algorithm(RSES)	Invokes the RSES implementation of the greedy algorithm of Johnson.

2.5.2 Voting for Classification using Rough Set Theory

Voting is an ad-hoc technique for rule-based classification that works reasonably well in practice which can be implemented in RST classification. The algorithm that utilises decision rules for classification usually employs voting to obtain decisions made by the rules. When a collection of decision rules is pointing at different decisions, voting can be applied to make up a simple rule-based decision support system. The examples of voting approaches include simple voting, majority voting and tune voting. Voting can be implemented in various rule-based classification techniques such as decision tree classifier and RST-based classifier.

2.5.3 Rules Filtering

The feature reduction process yielded a set of minimal subsets. Hundreds or even thousands of minimal subsets will be generated as rules for classification step in the next stage. The number of rules generated depended on the number of attributes and number of objects in the dataset. However, rules can be removed without lowering the performance significantly (Agotnes, 1999). These generated rules were known as the classification model.

Rule filtering attempted to find good sub models by removing whole rules and retaining the classification performances. In ROSETTA it offered two types of filtering, basic rules filtering and quality filtering. Basically, there will be few aspects to be considered in basic rule filtering, for example support, accuracy, coverage and stability. Support was the count of objects that satisfied a conjunctive formula or a decision rule while accuracy can be interpreted as the proportion of making a correct classification for an object that satisfied the rules. Accuracy can be used as performance measures for rule filtering. In order to measure the completeness of a rules, coverage of a rule was an estimate of the fraction of all the objects in the information system belonging to the indicated equivalence class that the rule can classify into this class. Lastly, stability was for the fraction of the sub model in which it was a proper reduction set.

Rule filtering aspects selected is displayed in Figure 2.8. Rule filtering was a process taken to avoid complex rules with low coverage on those containing irrelevant literals that had only been added to exclude noisy examples. The rule filtering problem was described as an optimization problem because the performance of the pruned models could be assessed.

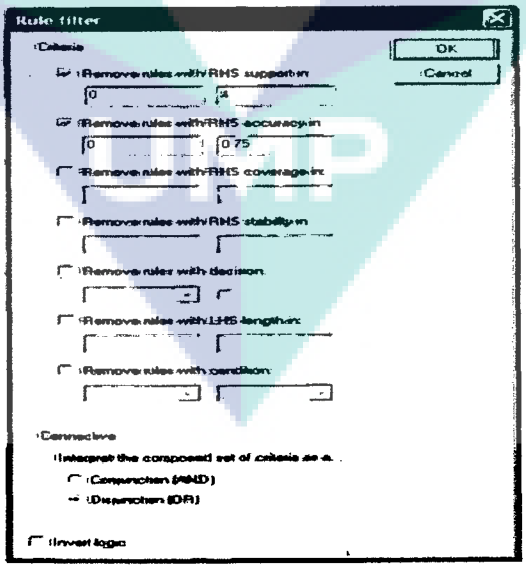


Figure 2.8: Rule Filtering Options

2.5.4 Rough Set Data Mining Tools

ROSETTA is a toolkit application specially built to analyse tabular data by implementing RST. According to Abdul Rahman, Mohd Lazim, Mohamed, Saany, and Mohd Yusof (2013), ROSETTA is a user-friendly graphical interface and easy to use software. In addition, ROSETTA is freely distributed and available for both Windows and Linux operating systems (J. Li, 2007). Figure 2.9 illustrates the example of ROSETTA data mining processes.

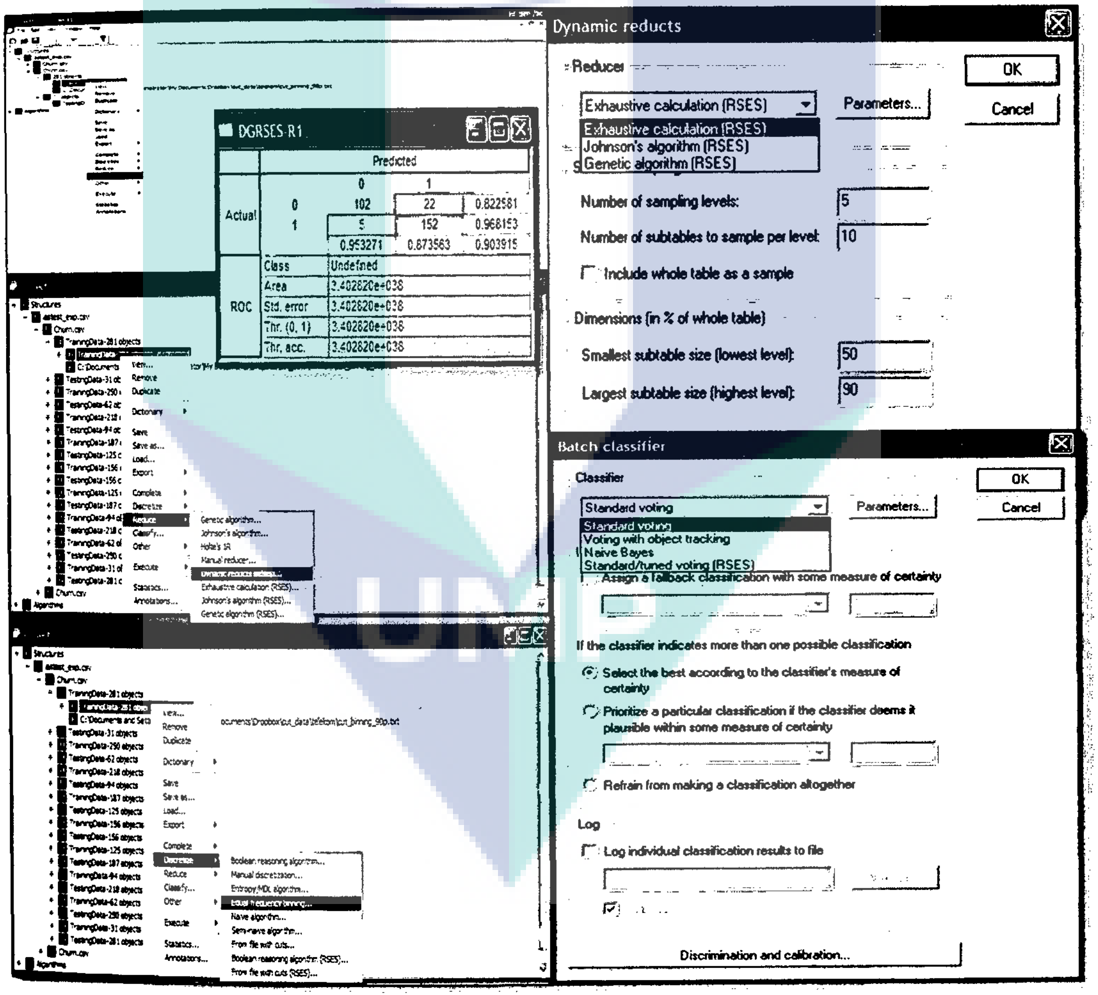


Figure 2.9: Example of ROSETTA Data Mining Processes

Rough Set Exploration System (RSES) is an embedded library in ROSETTA. It is a collection of algorithms and data structures for RST computations. It was developed by Group of Logic, Inst. of Mathematics, University of Warsaw, Poland. ROSETTA was designed so as to be able to make use of this legacy code, and suitable wrappers have been written so that the RSES library can be linked into the ROSETTA kernel. The ROSETTA kernel is also fully functional without RSES present. RSES library allows many types of data exploration such as calculation of reduction set, generation of decision rules with use of reduction set, discretization of numerical attributes and data manipulation and edition. However, RSES is not suitable for a very large data analysis.

2.5.5 Customer Churn Datasets

Data privacy is one of the obstacles to retrieve genuine data set for better end results. It is a common practise for a company to protect customer data for any privacy and safety purposes. However, records containing customers' private data such as address, telephone number and email address were not suitable for classification. Two different datasets were analysed; a genuine dataset from a local telecommunication company and established dataset retrieved from UCI repository. The local telecommunication company dataset were used as the main test case while dataset form UCI repository and IBM Watson Analytic repository were used as a validation for the classification framework. For the local telecommunication company dataset, it consisted of 312 records and 21 attributes including one decision attributes. The instances were briefly explained as follows in Table 2.6.

Table 2.6: Local Telecommunication Company Customer Churn Dataset Descriptions

Attribute Name	Description
Age	Integer-valued variable determine age of customer
Package	Categorical variable for customer subscribe package
Gender	Categorical variable to determine gender of customer
Monthly Commitment	Integer valued variable determine monthly commitment agreed by customer
Call Plan	Integer valued variable for customer call plan package
Service Rental	Integer valued variable to determine service rental selected by customer
Method of Payment	Categorical variable identifying method of payment choose by customer
Total Usage (RM)	Continuous-valued variable which determine sum of total monthly usage in Ringgit Malaysia
Total Local Call	Integer valued variable to count number of local call made by customer
Total Bill Local Call (RM)	Continuous valued variable which determine total bill for local call in Ringgit Malaysia
Total National Call Same Operator	Integer valued variable to count number of call made by customer to same operator user
Total Duration National Call Same Operator (Minutes)	Continuous valued variable which determine total duration for national local call same operator made by customer in minutes count
Total Bill National Call Same Operator (RM)	Continuous valued variable which determine total bill for national local call same operator made by customer in Ringgit Malaysia
Total National Unlimited Call For Same Operator	Integer valued variable to count number of national unlimited call made by customer to same operator user
Total National Call Different Operator	Integer valued variable to count number of call made by customer to different operator user
Total Duration National Call Different Operator (Minutes)	Continuous valued variable which determine total duration for national local call different operator made by customer in minutes count
Total Bill National Call Different Operator (RM)	Continuous valued variable which determine total bill for national local call different operator made by customer in

Attribute Name	Description
	Ringgit Malaysia
Total Free National Call Different Operator	Integer valued variable to count free national call to different operator received by customer
Total Payment (RM)	Continuous valued variable to determine total payment should be made by customer
Account Receivable (AR) Days	Continuous valued variable to determine the length of time it takes to clear all Accounts Receivable in Days count

The second selected data were retrieved from UCI repository consisted of 1667 objects belonging to 21 instances including one decision attribute used to discover churn pattern. The attribute *'phone'* in this dataset was removed since *'phone'* was used as identification only in a study done by Sharma and Panigrahi (2011). This established dataset was used for the framework validation. The instances were briefly explained in Table 2.7.

Table 2.7: UCI Repository Customer Churn Dataset Descriptions

Attribute Name	Description
State	Categorical variable, for the 50 states and the district of Columbia
Account length	Integer-valued variable for how long account has been active
Area code	Categorical variable to determine state of the customer
Phone number	Essentially a surrogate key for customer identification
International Plan	Dichotomous categorical having yes or no value
Voice Mail Plan	Dichotomous categorical variable yes or no value
Number of voice mail messages	Integer-valued variable
Total day minutes	Continuous variable for number of minutes customer has used the service during the day

UMP

Attribute Name	Description
Total day calls	Integer-valued variable
Total day charge	Continuous variable based on foregoing two variables
Total evening minutes	Continuous variable for minutes customer has used the service during the evening
Total evening calls	Integer-valued variable
Total evening charge	Continuous variable based on previous two variables
Total night minutes	Continuous variable for storing minutes the customer has used the service during the night
Total night calls	Integer-valued variable
Total night charge	Continuous variable based on foregoing two variables
Total international minutes	Continuous variable for minutes customer has used service to make international calls
Total international calls	Integer-valued variable
Total international charge	Continuous variable based on foregoing two variables
Number of calls to customer service	Integer-valued variable

Lastly, a dataset from IBM Watson Analytic Repository is taken to validate customer churn classification. *'Customer ID'* is permanently remove form this dataset because it is used for customer identification only. Table 2.8 summarizes the datasets selected.

Table 2.8: IBM Watson Analytic Repository Dataset Descriptions

Attribute Name	Description
Customer ID	Essentially a surrogate key for customer identification
Gender	Categorical variable to determine gender of customer
Senior Citizen	Categorical variable to determine customer is a senior citizen or not

Attribute Name	Description
Partner	Categorical variable to determine partnership
Dependents	Categorical variable to determine customer is having dependents or not
Tenure	Integer-valued variable for how long account has been tenured
Phone Service	Categorical variable to determine customer is subscribing phone service or not
Multiple Lines	Categorical variable to determine customer is having multiple lines or not
Internet Service	Categorical variable to determine customer is subscribing internet service or not
Online Security	Categorical variable to determine customer is subscribing online security service or not
Online Backup	Categorical variable to determine customer is subscribing online backup service or not
Device Protection	Categorical variable to determine customer is on device protection or not
Technical Support	Categorical variable to determine customer is having technical support or not
Streaming TV	Categorical variable to determine customer is subscribing streaming TV or not
Streaming Movies	Categorical variable to determine customer is subscribing streaming movies or not
Contract	Categorical variable identifying type of customer contract
Paperless Billing	Categorical variable identifying whether customer subscribing paperless billing or not
Payment Method	Categorical variable identifying method of payment choose by customer
Monthly Charges	Continuous variable based average monthly charges
Total Charges	Continuous variable based on total charges for a period

Table 2.9 summarized the selected dataset which will be used for customer churn classification using RST. Since this UCI and IBM Watson Analytic repository dataset are the established dataset, minimal data cleaning process was needed.

Table 2.9: Summary of Selected Dataset

Dataset	**Total Number of Conditional Attributes**	**Total Number of Decision Attribute**	**Total Number of Attribute**	**Number of Objects**
Local Telecommunication Company	20	1	21	312
UCI Machine Learning Repository	20	1	21	1667
IBM Watson Analytic	19	1	20	4900

2.6 Research Gaps

The research gaps identified from the literature review were as follows:

i. From a deep research in the telecommunication industries in Malaysia, no research had been found that focused on customer churn classification using RST. Based on (Malaysian Communications and Multimedia Commission, 2013) Annual Report, about 70% of the 11,395 complaints received in 2013 were related to the provision of service provider and service performance. These may cause customer churning.

ii. The next gap in previous researches was a lower classification accuracy which may lead to misclassification of churners and non-churners.

iii. The third gap was most of the previous researches using ROSETTA did not employ rule filtering to optimize classification rules accuracy.

iv. Lastly, using the filtered decision rules obtained, a development of customer churn classification system will be considered.

2.7 Concluding Remarks

In this chapter, literature reviews according to past works which applied similar or related approaches, frameworks, methods or techniques with customer churn, classification and RST were discussed. This chapter also briefly explained the definition for every subtopic mentioned and toolkits used as well as selected datasets.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Research Methodology is a scientific and systematic process guideline to a research problem. The methodology acts as a guideline so that at the end of every phase, an operation or process implies precise deliveries. Section 3.2 will discuss on research methodology activity which comprise of three phases. Later in Section 3.3, proposed research methodology framework. Section 3.4 provided summarized key aspects of the parser presented in this chapter.

Figure 3.1 illustrates integrated research methodology activity, RST framework and Knowledge Discovery in Database process. KDD processes include five phases and RST framework begins with creating decision table until result analysis. Detailed of this framework will be provided in later section.

3.2 Research Framework

Research framework acts as a guideline for the researchers in order to ensure research objectives will be satisfied at the end of the research. Figure 3.1 provided the framework employed in this research which is the integration of KDD processes and RST classification.

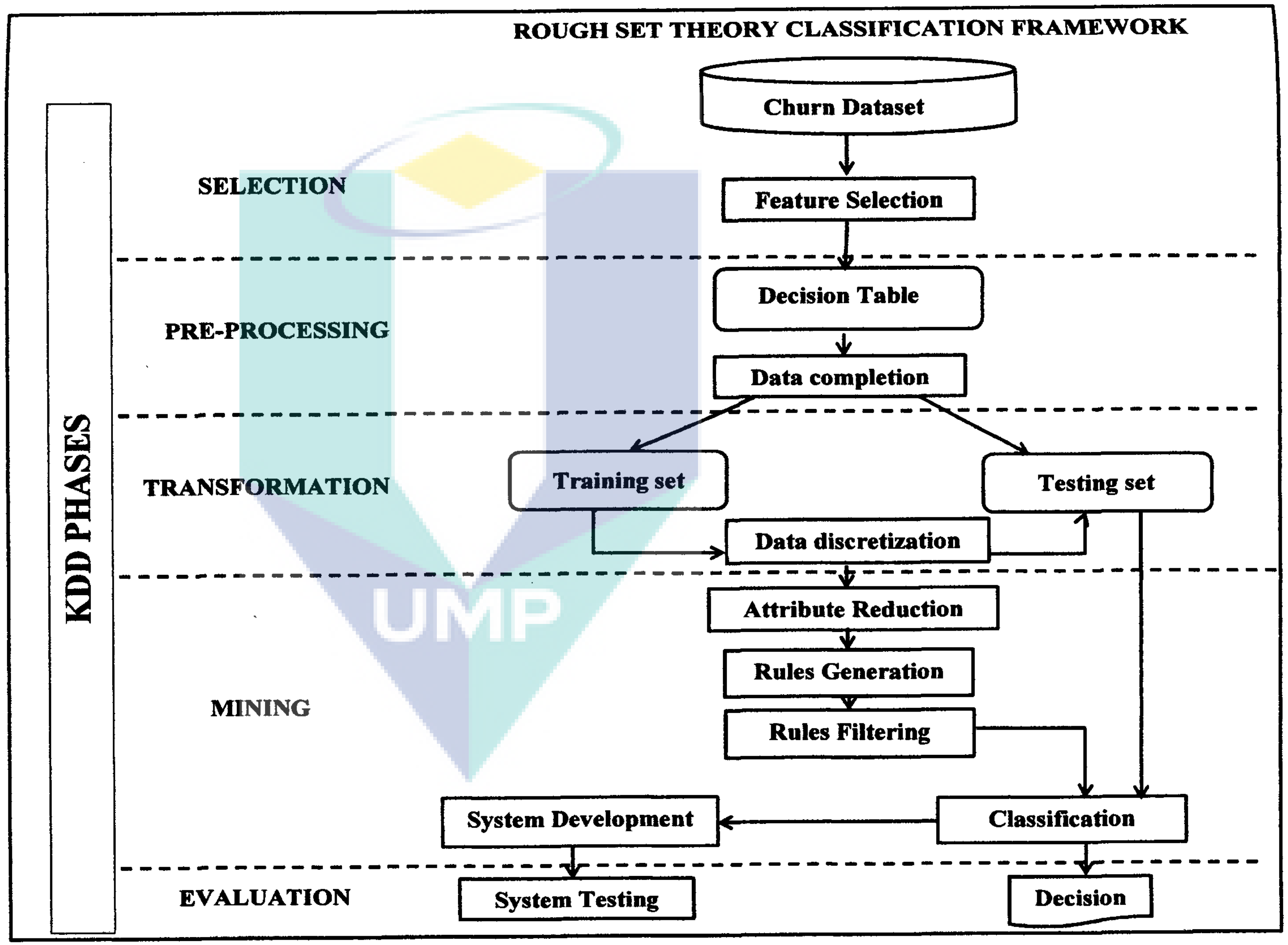


Figure 3.1: RST and KDD Integrated Framework

3.2.1 Phase 1: Selection

The very first step of the KDD process was data selection that was to be used throughout the study. In this research, customer churn dataset was retrieved from the local telecommunication company, UCI repository and IBM Watson Analytic repository. After that, the data will undergo some data modifications using Microsoft Excel to reformat the data into ROSETTA readable format and will be saved as Comma Separate Values file (CSV). CSV data format is commonly compatible for WEKA and ROSETTA toolkits. By using WEKA data mining tool, dataset went through an attribute selection step. This attribute selection mainly aimed to select only relevant and significant attributes from conditional attributes for the next process. Figure 4.1 shows the attribute selection process in WEKA using Correlation-based Feature Selection Subset Evaluator (CfsSubsetEval) attribute evaluator and exhaustive search method. CfsSubsetEval was selected because it evaluated the worth of a subset of attributes by considering the individual predictive ability of each feature, along with the degree of redundancy between them (Kamarudin, 2015) while exhaustive search method used to find the possible optimal subset when the number of attributes, *N* was not too large.

3.2.2 Phase 2: Pre-processing

After the data were ready in the decision table format, the pre-processing of the data needed to be performed. In this step, the data will undergo data completion or correction, cleaning and investigating the quality. The decision table underwent data completion checks to make sure no incomplete data existed. In this research, built-in *Mean/Mode Completer* in ROSETTA were used. Then, the dataset underwent some data modification called data normalization. Normalization is a process to transform the data into a predefined group so that data is readable and compatible for the toolkit.

3.2.3 Phase 3: Transformation

In data transformation, the experiment started with importing customer churn dataset into the ROSETTA toolkit. The customer churn dataset now consisted of eight conditional attributes and one decision attributes. About 312 objects were to be imported into the system. Transformation phase will transformed the data so that there are limited numbers of possible states.

The data transformation phase began with splitting the data into two parts; training set and testing set which were the most common approaches applied in supervised learning. The separation was based on the splitting factor. Training dataset was used to implement learning process which will produce patterns or rules of the data while testing data was employed for testing steps. Training dataset will be used in the attribute or feature reduction step to find the minimal subset which had similar ability to discern groups when the full set of feature attributes were used (Venkateswara Reddy, Suresh, Reddy, & Shaik, 2011). This minimal subset was known as ***reduction set***.

The next process was data discretization. The purpose of discretization was to search for a consistent partition of attribute domains into intervals and unifying the values over each interval and the quality of classification depended on discretization process (Nguyen, 1998) to group or categorize the data which had continuous values for certain attributes. From the discretization process, a cut data was produced to be applied to the testing dataset. There were a few examples of discretization algorithm such as Naive Algorithm, Semi-Naive Algorithm, Equal Frequency Binning (EFB) and Boolean Reasoning. For this experiment, Equal Frequency Binning was employed to discretize data. Equal Frequency Binning was utilized because this discretization method divided the data into a specified number of intervals, each containing an equal number of objects.

3.2.4 Phase 4: Mining

The data mining phase is the most vital part in the KDD process. In this phase, the classification process takes place. Prior to classification, the decision table must undergo the attribute reduction process. The minimal subset of attributes for the decision table will be used for rules generation. The more numbers of reduction sets, the more of an increase in the number of rules. In this phase, the classification process takes place. Majority voting is chosen as the RST-based classifier (Ohrn, 2000a) explained voting as follows:

Let R be denoted as an unordered set of decision rules.

i. The set R is scanned for rules that fire; i.e., rules that have an antecedent that matches x. Let $R(x) \subseteq R$ denote the set of rules that fire for object x.

ii. If $R = \emptyset$, then no classification can be made. Typically, a predetermined fall back classification is then invoked. Alternatives include reverting to a nearest-neighbour method, and instead, considering a collection of rules "close" to object x in some sense.

iii. An election process among the rules in *R(x)* is performed in order to resolve conflicts and rank the decisions. The election process is performed as follows:

(a) Let each rule $r \in R(X)$ cast a number of *votes (r)* in favour of the decision class the rule indicates. The number of votes a rule gets to cast may vary. Typically, *votes (r)* is based on the support of the rule, but more complex quality-based measures are also possible as in Equation 3.1.

$$RUL_\beta = \{r \in R(X) | r\ predict\ \beta\} \qquad 3.1$$

(b) Compute a normalisation factor for *norm (x)*. The normalisation factor can be computed in different ways. In Equation 3.2, *norm (x)* is simply the sum of all cast votes, and only serves as a scaling factor.

$$votes\ (\beta) = \sum_{r \in RUL(\beta)} votes\ (r) \qquad 3.2$$

(c) Divide the accumulated number of votes for each possible decision class β by the normalisation factor *norm (x)* in order to arrive at a certainty coefficient *certainty (x, β)* for each decision class, as in Equation 3.3.

$$certainty(x, \beta) = {votes(\beta)}/{norm(x)} \qquad 3.3$$

Majority voting is one of the approaches applied in voting. Majority voting is based on support voting but with no tolerance for missing values (Ohrn, 2000). In majority voting, votes are summed up and the decision that got the majority of votes is selected.

3.2.5 Phase 5: Interpretation/ Evaluation

Interpretation or evaluation is the last step in KDD process. In this phase, the results obtained in the classification process were interpreted into the knowledge that can be understood by human for example confusion matrix. From the confusion matrix, it showed three aspects to be measured in a classification such as accuracy, specificity and sensitivity. Examples of interpretation for confusion matrix were explained in next subsection of Performance Measure.

3.2.5.1 Performance Measure

Generally, four terms, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were used to determine the accuracy of the measurement in the confusion matrix. This term was very useful to interpret the results obtained as well as the performance measure to the classification process. The terms were summarized in Table 3.1. Table 3.2 illustrates how these four terms were posted in the Confusion Matrix. Meanwhile, Table 3.3 summarizes the equations for each term.

Table 3.1: Confusion Matrix Terms Description

Term Name	Description
True Positive (TP)	Churn customer correctly classified as churn
False Positive (FP)	Non-churn customer incorrectly classified as churn
True Negative (TN)	Non-churn customer correctly classified as non-churn
False Negative (FN)	Churn customer incorrectly classified as non-churn

Table 3.2: Confusion Matrix Terms Allocation

		Classified	
		Non-churn	Churn
Actual	Non-churn	TN(a)	FP(b)
	Churn	FN(c)	TP(d)

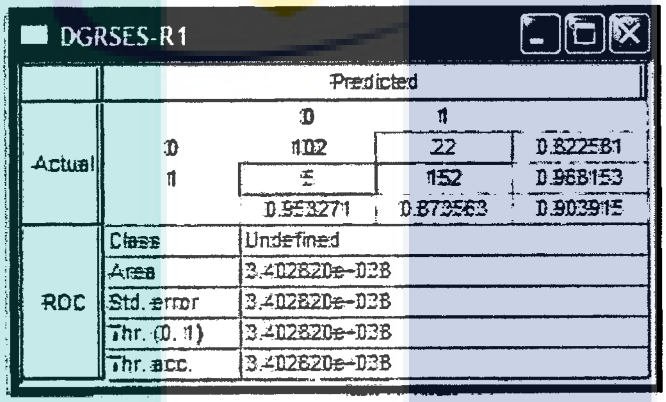
Table 3.3: Confusion Matrix Terms Equation

		Classified	
		Non-churn	Churn
Actual	Non-churn	$TN = \frac{a}{a+b}$	$FP = \frac{b}{a+b}$
	Churn	$FN = \frac{c}{c+d}$	$TP = \frac{d}{c+d}$

The performance measure indicated how good the classifier performed. Besides, performance measure will tell if the research objectives fulfilled and any improvement needs or not. In the business line, performance measure presented how satisfied the customers were. RST classification used confusion matrix as the evaluation of the performance. From the confusion matrix, sensitivity (TP), specificity (TN) and accuracy (AC) can be easily calculated. Equation 3.4 shows how accuracy is calculated:

$$AC = \frac{a + d}{a + b + c + d} \quad (3.4)$$

Figure 3.2 displays the example of confusion matrix obtained from classification.


DGRSES-R1

		Predicted		
		0	1	
Actual	0	102	22	0.822581
	1	5	152	0.968153
		0.953271	0.873563	0.903915
ROC	Class	Undefined		
	Area	3.402820e-038		
	Std. error	3.402820e-038		
	Thr. (0, 1)	3.402820e-038		
	Thr. acc.	3.402820e-038		

Figure 3.2: ROSETTA Confusion Matrix

3.3 Concluding Remarks

This chapter was all about the research methodology conducted along these researches. It was important to have a research guideline in order to complete the research on time. The research methodology is explained through the research framework.

CHAPTER 4

IMPLEMENTATION AND RESULTS

4.1 Introduction

This chapter provided the implementation of classification on customer churn datasets, along with a discussion of the results obtained. The results obtained were presented in tables and charts, based on accuracy achieved from different experiments. At the end of this chapter, the development of customer churn classification system was explained.

4.2 The Data Manipulation Process on Customer Churn Dataset for Local Telecommunication Company

The data manipulation processes were conducted to evaluate how the data was manipulated prior to the classification process. The data manipulation was highlighted in Figure 4.1. The purpose of data manipulation is to provide the most significant dataset format in the classification. These processes include the feature selection, data normalisation, and data transformation, attribute reduction and rules generation.

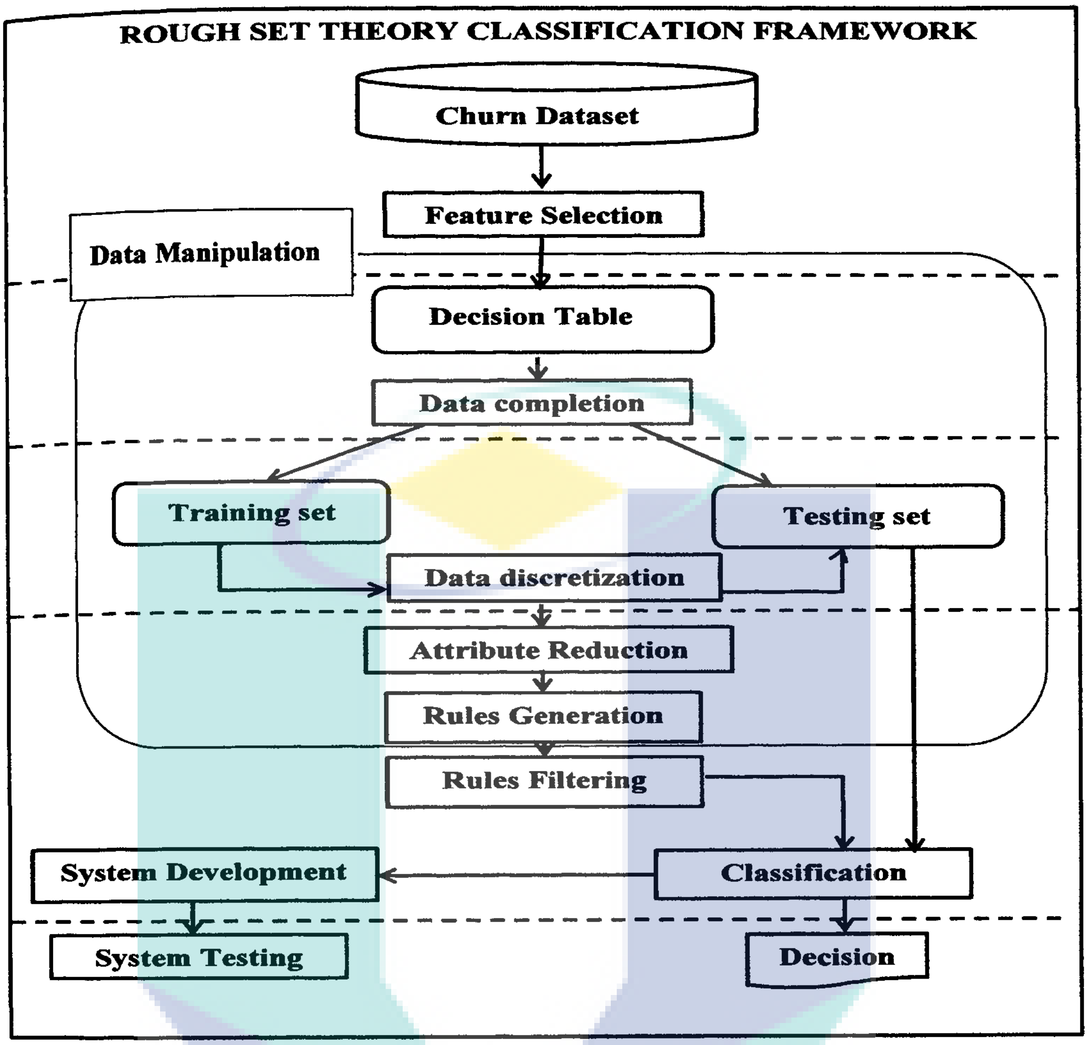


Figure 4.1: Data Manipulation in Research Framework

4.2.1 Feature Selection

In order to select the most significant features for classification, with the aid of WEKA tool, features "*Age, Gender, Packages, Call Plan, Method of Payment, Total duration National Call Same Operator, Total duration National Call Different Operator and AR Days*" were selected. In Figure 4.2 it displays feature selection process took place in WEKA. After feature selection process, the decision table has eight conditional attributes and one decision attribute.

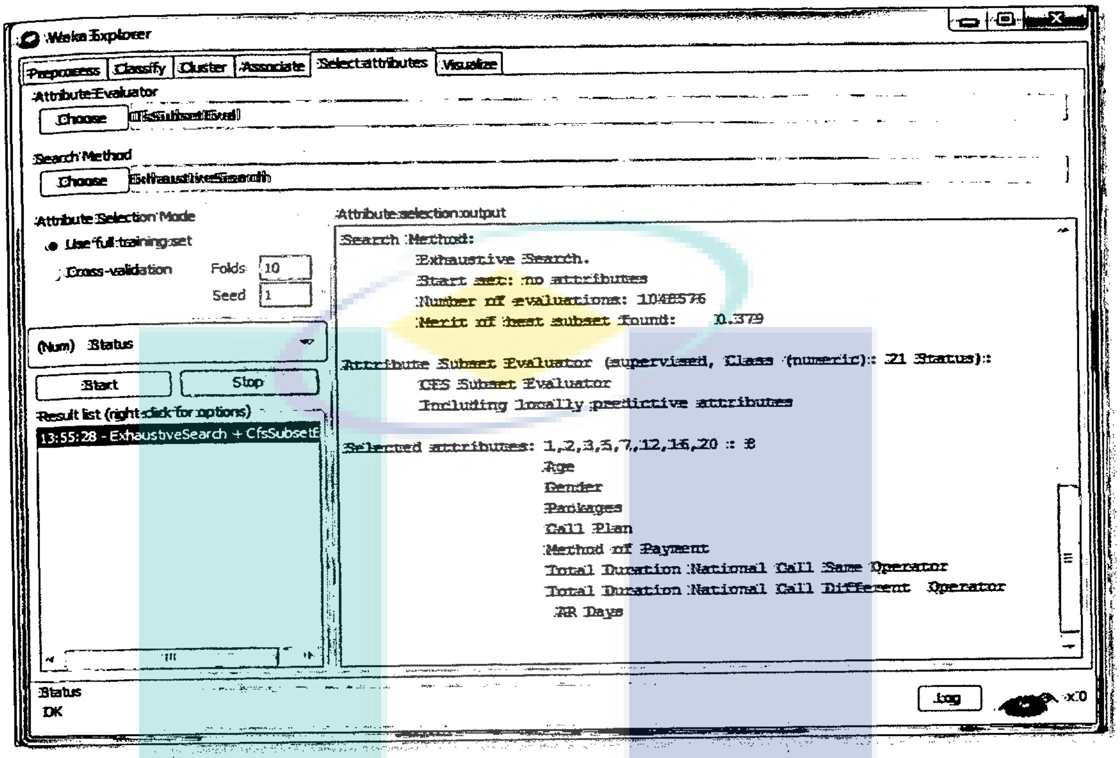


Figure 4.2: Feature Selection in WEKA

4.2.2 Data Normalization

After the dataset undergoes data completion, it proceeds with data normalisation to standardise the values of all parameters. Table 4.1 displayed an example of the normalised customer churn dataset where non-continuous-valued variable was assigned to certain values, such as 1 for male and 0 for female.

Table 4.1: Attribute Normalization Summary

	Name of Attribute	Values
A_1	Age	Integer-Valued Variable
A_2	Gender	0 = Female 1 = Male
A_3	Packages	1 = 4000 2 = 384 3 = 2000 4 = 1000 5 = 512
A_4	Call Plan	1 = 10 2 = 0 3 = 28 4 = 38
A_5	Method Of Payment	1 = Manual 0 = Auto
A_6	Total duration National Call Same Operator	Continuous-valued variable
A_7	Total duration National Call Different Operator	Continuous-valued variable
A_8	AR Days	Continuous-valued variable
D	Status	0 = Not Churn 1 = Churn

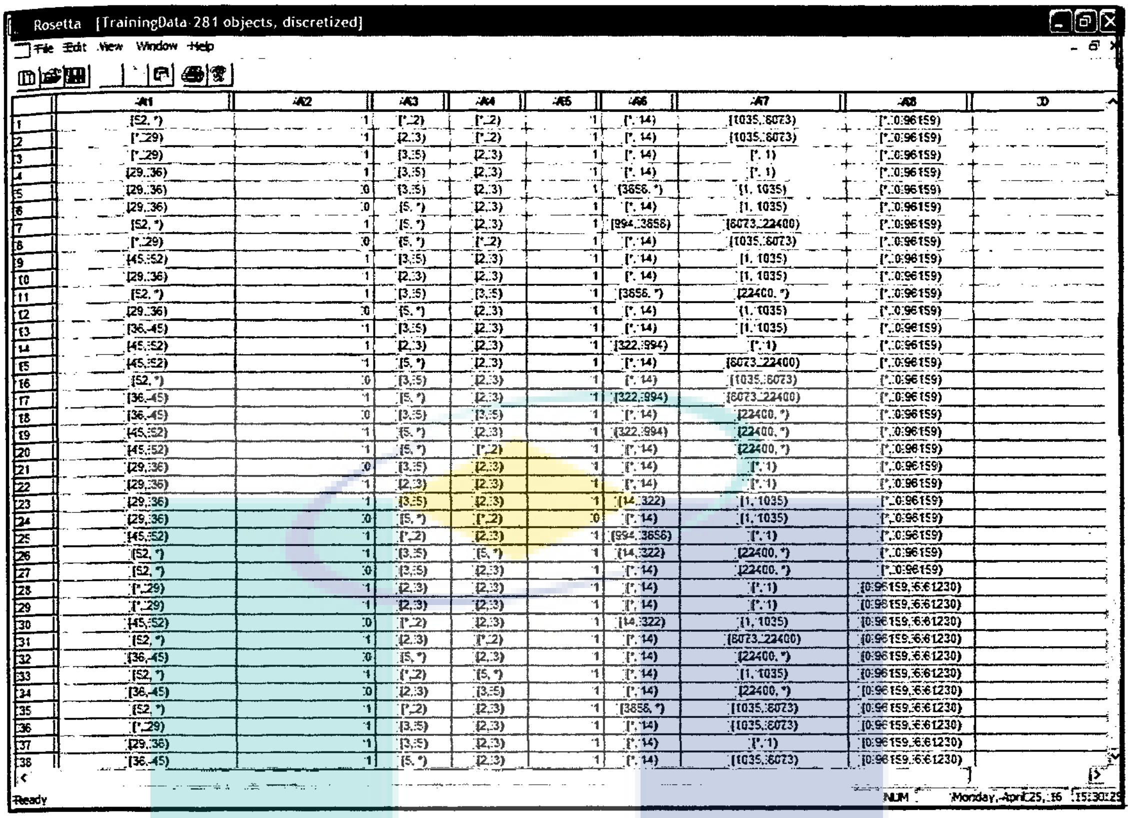
4.2.3 Data Transformation

In data transformation, data is discretised. In order to discretise data, the data needs to split. The split factor is used to divide datasets into a training set and a testing set. As an example of data splitting, if the splitting factor is 0.5, it would indicate that 50% was allocated to the training set and the other 50% for the testing set. The purpose of using different splitting factors is to identify which splitting factors contribute to optimising the accuracy. The splitting factor uses the range of 0.1 to 0.9. Table 4.2 showed the percentage and amount of training and testing of datasets based on the splitting factor.

Table 4.2: Data Splitting Factors

Split Factor	Percentage of Training Data (%)	Percentage of Testing Data (%)	Total Number of Training Data	Total Number of Testing Data
0.9	90	10	281	31
0.8	80	20	250	62
0.7	70	30	218	94
0.6	60	40	187	125
0.5	50	50	156	156
0.4	40	60	125	187
0.3	30	70	94	218
0.2	20	80	62	250
0.1	10	90	31	281

Data is discretised using Equal Frequency Binning (EFB). Figure 4.3 provided a sample of discretised decision table after the discretisation process. For non-binary valued variables, discretised algorithms assign a range of value. For example, the value of 48 at the attribute A1 will be [48, *) after the discretisation process. This discretised decision table was used for the entire research.



	A1	A2	A3	A4	A5	A6	A7	A8	D
1	[52, *)	1	[*, 2)	[*, 2)	1	[*, 14)	[1035, 6073)	[*, 0.96159)	
2	[*, 29)	1	[2, 3)	[*, 2)	1	[*, 14)	[1035, 6073)	[*, 0.96159)	
3	[*, 29)	1	[3, 5)	[2, 3)	1	[*, 14)	[*, 1)	[*, 0.96159)	
4	[29, 36)	1	[3, 5)	[2, 3)	1	[*, 14)	[*, 1)	[*, 0.96159)	
5	[29, 36)	0	[3, 5)	[2, 3)	1	[3656, *)	[1, 1035)	[*, 0.96159)	
6	[29, 36)	0	[5, *)	[2, 3)	1	[*, 14)	[1, 1035)	[*, 0.96159)	
7	[52, *)	1	[5, *)	[2, 3)	1	[994, 3656)	[6073, 22400)	[*, 0.96159)	
8	[*, 29)	0	[5, *)	[*, 2)	1	[*, 14)	[1035, 6073)	[*, 0.96159)	
9	[45, 52)	1	[3, 5)	[2, 3)	1	[*, 14)	[1, 1035)	[*, 0.96159)	
10	[29, 36)	1	[2, 3)	[2, 3)	1	[*, 14)	[1, 1035)	[*, 0.96159)	
11	[52, *)	1	[3, 5)	[3, 5)	1	[3656, *)	[22400, *)	[*, 0.96159)	
12	[29, 36)	0	[5, *)	[2, 3)	1	[*, 14)	[1, 1035)	[*, 0.96159)	
13	[36, 45)	1	[3, 5)	[2, 3)	1	[*, 14)	[1, 1035)	[*, 0.96159)	
14	[45, 52)	1	[2, 3)	[2, 3)	1	[322, 994)	[*, 1)	[*, 0.96159)	
15	[45, 52)	1	[5, *)	[2, 3)	1	[*, 14)	[6073, 22400)	[*, 0.96159)	
16	[52, *)	0	[3, 5)	[2, 3)	1	[*, 14)	[1035, 6073)	[*, 0.96159)	
17	[36, 45)	1	[5, *)	[2, 3)	1	[322, 994)	[6073, 22400)	[*, 0.96159)	
18	[36, 45)	0	[3, 5)	[3, 5)	1	[*, 14)	[22400, *)	[*, 0.96159)	
19	[45, 52)	1	[5, *)	[2, 3)	1	[322, 994)	[22400, *)	[*, 0.96159)	
20	[45, 52)	1	[5, *)	[*, 2)	1	[*, 14)	[22400, *)	[*, 0.96159)	
21	[29, 36)	0	[3, 5)	[2, 3)	1	[*, 14)	[*, 1)	[*, 0.96159)	
22	[29, 36)	1	[2, 3)	[2, 3)	1	[*, 14)	[*, 1)	[*, 0.96159)	
23	[29, 36)	1	[3, 5)	[2, 3)	1	[14, 322)	[1, 1035)	[*, 0.96159)	
24	[29, 36)	0	[5, *)	[*, 2)	0	[*, 14)	[1, 1035)	[*, 0.96159)	
25	[45, 52)	1	[*, 2)	[2, 3)	1	[994, 3656)	[*, 1)	[*, 0.96159)	
26	[52, *)	1	[3, 5)	[5, *)	1	[14, 322)	[22400, *)	[*, 0.96159)	
27	[52, *)	0	[3, 5)	[2, 3)	1	[*, 14)	[22400, *)	[*, 0.96159)	
28	[*, 29)	1	[2, 3)	[2, 3)	1	[*, 14)	[*, 1)	[0.96159, 6.61230)	
29	[*, 29)	1	[2, 3)	[2, 3)	1	[*, 14)	[*, 1)	[0.96159, 6.61230)	
30	[45, 52)	0	[*, 2)	[2, 3)	1	[14, 322)	[1, 1035)	[0.96159, 6.61230)	
31	[52, *)	1	[2, 3)	[*, 2)	1	[*, 14)	[6073, 22400)	[0.96159, 6.61230)	
32	[36, 45)	0	[5, *)	[2, 3)	1	[*, 14)	[22400, *)	[0.96159, 6.61230)	
33	[52, *)	1	[*, 2)	[5, *)	1	[*, 14)	[1, 1035)	[0.96159, 6.61230)	
34	[36, 45)	0	[2, 3)	[3, 5)	1	[*, 14)	[22400, *)	[0.96159, 6.61230)	
35	[52, *)	1	[*, 2)	[2, 3)	1	[3656, *)	[1035, 6073)	[0.96159, 6.61230)	
36	[*, 29)	1	[3, 5)	[2, 3)	1	[*, 14)	[1035, 6073)	[0.96159, 6.61230)	
37	[29, 36)	1	[3, 5)	[2, 3)	1	[*, 14)	[*, 1)	[0.96159, 6.61230)	
38	[36, 45)	1	[5, *)	[2, 3)	1	[*, 14)	[1035, 6073)	[0.96159, 6.61230)	

Figure 4.3: Discretized Decision Table

4.2.4 Attribute Reduction and Rules Generation

There are three reduction algorithms used to develop the best classification model such as Exhaustive Calculation, Johnson's Algorithm and Genetic Algorithm. The best reduction algorithm will provide great improvement on the classification accuracy. Figure 4.4 illustrates the reduction set sample produced using Dynamic Exhaustive Calculation Algorithm. The length for every reduction set may vary depending on functional dependencies of the object, and the *'support'* defines the count of objects that satisfy a conjunctive formula or a decision. For example, 46 *'support'* means 46 objects satisfied this reduction set (A2, A4 and A7). This reduction set will generate a group of readable if-else rules for the classification phase.

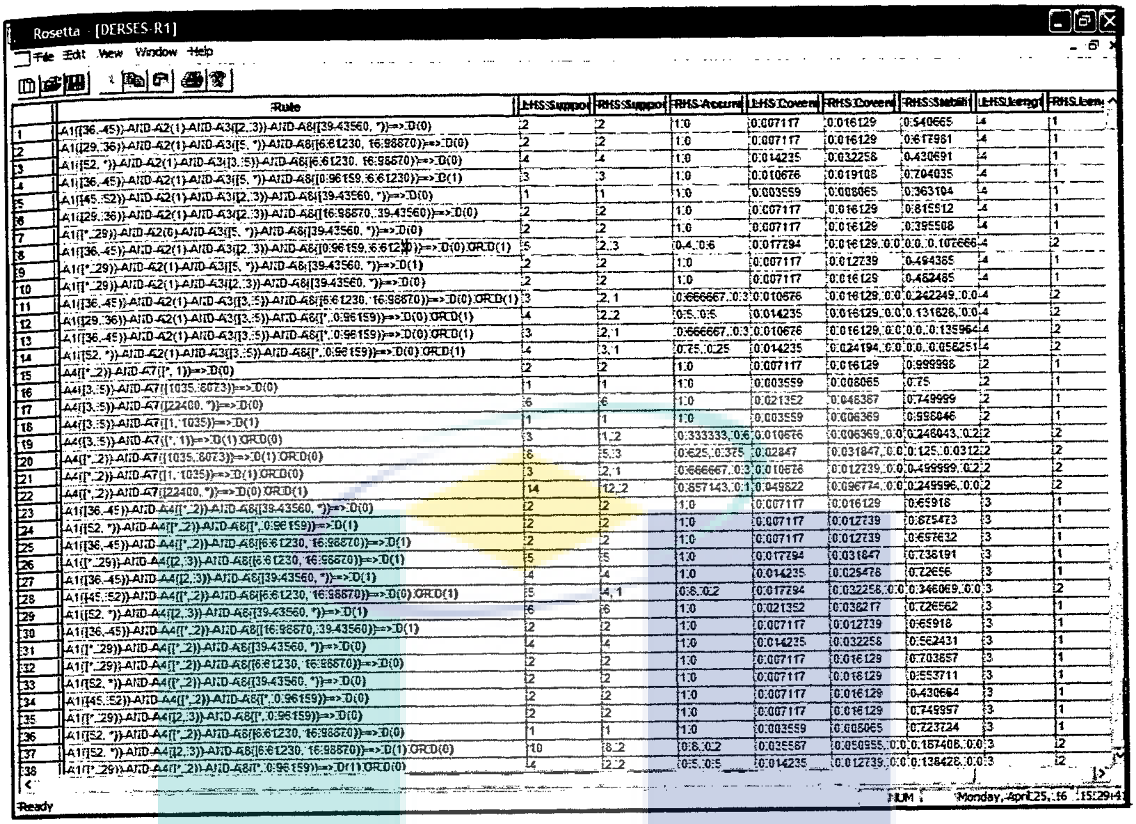
DERSES-Reduct

	Reduct	Support	Length
1	{A2, A4, A7}	46	3
2	{A3, A4, A7}	52	3
3	{A4, A5, A7}	44	3
4	{A4, A7, A8}	59	3
5	{A1, A2, A3, A7}	60	4
6	{A1, A2, A6, A7}	60	4
7	{A1, A2, A8}	60	3
8	{A2, A3, A4}	60	3
9	{A2, A3, A6, A7}	60	4
10	{A2, A3, A6, A8}	60	4
11	{A2, A4, A7, A8}	60	4
12	{A2, A6, A7, A8}	60	4
13	{A1, A4, A7}	60	3
14	{A1, A7, A8}	60	3
15	{A3, A4, A7, A8}	60	4
16	{A3, A6, A7, A8}	60	4
17	{A1, A2, A3, A5, A8}	38	5
18	{A1, A2, A3, A7, A8}	57	5
19	{A1, A4}	53	2
20	{A1, A6, A8}	60	3
21	{A3, A4}	60	2
22	{A4, A6}	50	2
23	{A4, A7}	54	2
24	{A4, A8}	52	2

Figure 4.4: Reduction Set Sample

UMP

Figure 4.5 displayed the sample of rules generated. The length of the rule depends on the length of the reduction set. The reduction set of five will produce a rule with five in length. Left-Hand-Side (LHS) length defines the length of the rule and Right-Hand-Side (RHS) length means the length of the decision. A rule may have two choices of decisions which are either 0 or 1. When classification hits this rule, RST approximation concept will be applied, along with the voting technique.



Rosetta - [DERSES-R1]

File Edit View Window Help

	Rule	LHS Suppo	RHS Suppo	RHS Accura	LHS Covera	RHS Covera	RHS Stabili	LHS Lengt	RHS Len
1	A1([36, 45)) AND A2(1) AND A3([2, 3)) AND A8([39.43560, *)) => D(0)	2	2	1.0	0.007117	0.016129	0.540665	4	1
2	A1([29, 36)) AND A2(1) AND A3([5, *)) AND A8([6.61230, 16.98870)) => D(0)	2	2	1.0	0.007117	0.016129	0.617981	4	1
3	A1([52, *)) AND A2(1) AND A3([3, 5)) AND A8([6.61230, 16.98870)) => D(0)	4	4	1.0	0.014235	0.032258	0.430691	4	1
4	A1([36, 45)) AND A2(1) AND A3([5, *)) AND A8([0.96159, 6.61230)) => D(1)	3	3	1.0	0.010676	0.019108	0.704035	4	1
5	A1([45, 52)) AND A2(1) AND A3([2, 3)) AND A8([39.43560, *)) => D(0)	1	1	1.0	0.003559	0.008065	0.363104	4	1
6	A1([29, 36)) AND A2(1) AND A3([2, 3)) AND A8([16.98870, 39.43560)) => D(0)	2	2	1.0	0.007117	0.016129	0.815512	4	1
7	A1([*, 29)) AND A2(0) AND A3([5, *)) AND A8([39.43560, *)) => D(0)	2	2	1.0	0.007117	0.016129	0.395508	4	1
8	A1([36, 45)) AND A2(1) AND A3([2, 3)) AND A8([0.96159, 6.61230)) => D(0) OR D(1)	5	2, 3	0.4, 0.6	0.017794	0.016129, 0.0	0.0, 0.107666	4	2
9	A1([*, 29)) AND A2(1) AND A3([5, *)) AND A8([39.43560, *)) => D(1)	2	2	1.0	0.007117	0.012739	0.494365	4	1
10	A1([*, 29)) AND A2(1) AND A3([2, 3)) AND A8([39.43560, *)) => D(0)	2	2	1.0	0.007117	0.016129	0.482485	4	1
11	A1([36, 45)) AND A2(1) AND A3([3, 5)) AND A8([6.61230, 16.98870)) => D(0) OR D(1)	3	2, 1	0.666667, 0.3	0.010676	0.016129, 0.0	0.242249, 0.0	4	2
12	A1([29, 36)) AND A2(1) AND A3([3, 5)) AND A8([*, 0.96159)) => D(0) OR D(1)	4	2, 2	0.5, 0.5	0.014235	0.016129, 0.0	0.131626, 0.0	4	2
13	A1([36, 45)) AND A2(1) AND A3([3, 5)) AND A8([*, 0.96159)) => D(0) OR D(1)	3	2, 1	0.666667, 0.3	0.010676	0.016129, 0.0	0.0, 0.135964	4	2
14	A1([52, *)) AND A2(1) AND A3([3, 5)) AND A8([*, 0.96159)) => D(0) OR D(1)	4	3, 1	0.75, 0.25	0.014235	0.024194, 0.0	0.0, 0.056251	4	2
15	A4([*, 2)) AND A7([*, 1)) => D(0)	2	2	1.0	0.007117	0.016129	0.999998	2	1
16	A4([3, 5)) AND A7([1035, 8073)) => D(0)	1	1	1.0	0.003559	0.008065	0.75	2	1
17	A4([3, 5)) AND A7([22400, *)) => D(0)	6	6	1.0	0.021352	0.048387	0.749999	2	1
18	A4([3, 5)) AND A7([1, 1035)) => D(1)	1	1	1.0	0.003559	0.006369	0.998046	2	1
19	A4([3, 5)) AND A7([*, 1)) => D(1) OR D(0)	3	1, 2	0.333333, 0.6	0.010676	0.006369, 0.0	0.248043, 0.2	2	2
20	A4([*, 2)) AND A7([1035, 8073)) => D(1) OR D(0)	8	5, 3	0.625, 0.375	0.02847	0.031847, 0.0	0.125, 0.0312	2	2
21	A4([*, 2)) AND A7([1, 1035)) => D(1) OR D(0)	3	2, 1	0.666667, 0.3	0.010676	0.012739, 0.0	0.499999, 0.2	2	2
22	A4([*, 2)) AND A7([22400, *)) => D(0) OR D(1)	14	12, 2	0.857143, 0.1	0.049822	0.096774, 0.0	0.249996, 0.0	2	2
23	A1([36, 45)) AND A4([*, 2)) AND A8([39.43560, *)) => D(0)	2	2	1.0	0.007117	0.016129	0.65918	3	1
24	A1([52, *)) AND A4([*, 2)) AND A8([*, 0.96159)) => D(1)	2	2	1.0	0.007117	0.012739	0.675473	3	1
25	A1([36, 45)) AND A4([*, 2)) AND A8([6.61230, 16.98870)) => D(1)	2	2	1.0	0.007117	0.012739	0.697632	3	1
26	A1([*, 29)) AND A4([2, 3)) AND A8([6.61230, 16.98870)) => D(1)	5	5	1.0	0.017794	0.031847	0.736191	3	1
27	A1([36, 45)) AND A4([2, 3)) AND A8([39.43560, *)) => D(1)	4	4	1.0	0.014235	0.025478	0.72656	3	1
28	A1([45, 52)) AND A4([*, 2)) AND A8([6.61230, 16.98870)) => D(0) OR D(1)	5	4, 1	0.8, 0.2	0.017794	0.032258, 0.0	0.346069, 0.0	3	2
29	A1([52, *)) AND A4([2, 3)) AND A8([39.43560, *)) => D(1)	6	6	1.0	0.021352	0.038217	0.726562	3	1
30	A1([36, 45)) AND A4([*, 2)) AND A8([16.98870, 39.43560)) => D(1)	2	2	1.0	0.007117	0.012739	0.65918	3	1
31	A1([*, 29)) AND A4([*, 2)) AND A8([39.43560, *)) => D(0)	4	4	1.0	0.014235	0.032258	0.562431	3	1
32	A1([*, 29)) AND A4([*, 2)) AND A8([6.61230, 16.98870)) => D(0)	2	2	1.0	0.007117	0.016129	0.703657	3	1
33	A1([52, *)) AND A4([*, 2)) AND A8([39.43560, *)) => D(0)	2	2	1.0	0.007117	0.016129	0.553711	3	1
34	A1([45, 52)) AND A4([*, 2)) AND A8([*, 0.96159)) => D(0)	2	2	1.0	0.007117	0.016129	0.430664	3	1
35	A1([*, 29)) AND A4([2, 3)) AND A8([*, 0.96159)) => D(0)	2	2	1.0	0.007117	0.016129	0.749997	3	1
36	A1([52, *)) AND A4([*, 2)) AND A8([6.61230, 16.98870)) => D(0)	1	1	1.0	0.003559	0.008065	0.723724	3	1
37	A1([52, *)) AND A4([2, 3)) AND A8([6.61230, 16.98870)) => D(1) OR D(0)	10	8, 2	0.8, 0.2	0.035587	0.050955, 0.0	0.187408, 0.0	3	2
38	A1([*, 29)) AND A4([*, 2)) AND A8([*, 0.96159)) => D(1) OR D(0)	4	2, 2	0.5, 0.5	0.014235	0.012739, 0.0	0.138428, 0.0	3	2

Ready NUM Monday, April 25, 16 15:29:41

Figure 4.5: Rules Generated (Model) Sample

RHS accuracy defines the proportion of making a correct classification for an object that satisfies the rules. For example, accuracy of 0.5 refers to 50% correct classifications by a rule. LHS Coverage means an estimation of the fraction of all the objects in the information system belong to the indicated equivalence class that the rule can classify into this class. Overall rules generated were attached in the Appendices section.

4.3 The Implementation of Classification Technique

The classification process consists of two main parts: the training and testing parts. Training is also known as the learning part for the cognitive operation of defining criteria, by which the characteristics are distinguished in order to build up the classification model. In this process, the classifier learns its own classification rules from a training set. In the training process, the data collected and stored in the database will be converted into a decision table. Then, it will be continued with the process of data discretisation, feature reduction, rules generation and rules filtering. While in the testing stage, rules generated from the training step will be utilised to measure the performance of the generated model.

4.3.1 Applying Rough Set Theory Classification on Local Telecommunication Company Dataset

This section presented the experiments performed in this research to classify objects. Firstly, the effect of feature selection was evaluated. Rules obtained in data mining were also tested towards rules filtering to evaluate how it affected the classification accuracy. The comparison between classifications using RST-based classifier and non-RST classifiers were also provided.

4.3.1.1 Customer Churn Classification without Feature Selection

Feature selection is highly important and beneficial when the amount of data is larger. However, in this section, the selected dataset was experimented without the involvement of feature selection prior to the classification. The purpose of this experiment was to evaluate the effects of feature selection in a later section. The data were tested using different split factors and reduction methods.

Table 4.3: Classification Accuracies without Feature Selection

Split Factor	Reduction Method	Training Accuracy (%)	Testing Accuracy (%)	Overall Accuracy (%)
0.9	Exhaustive Calculation	90.04	83.87	86.95
	Johnson's Algorithm	86.83	83.87	85.35
	Genetic Algorithm	90.04	83.87	86.95
0.8	Exhaustive Calculation	89.20	79.03	84.12
	Johnson's Algorithm	87.2	82.26	84.73
	Genetic Algorithm	89.60	77.42	83.51
0.7	Exhaustive Calculation	84.40	68.09	76.24
	Johnson's Algorithm	89.45	77.66	83.55
	Genetic Algorithm	86.70	76.60	81.65
0.6	Exhaustive Calculation	91.44	72.80	82.12
	Johnson's Algorithm	89.31	74.40	81.85
	Genetic Algorithm	88.77	76.00	82.39
0.5	Exhaustive Calculation	92.95	75.64	84.29
	Johnson's Algorithm	91.67	74.36	83.01
	Genetic Algorithm	85.90	78.21	82.05
0.4	Exhaustive Calculation	88.80	72.73	80.76
	Johnson's Algorithm	88.00	74.87	81.43
	Genetic Algorithm	88.80	74.33	81.57
0.3	Exhaustive Calculation	84.04	72.02	78.03
	Johnson's Algorithm	82.98	70.64	76.81
	Genetic Algorithm	89.36	72.94	81.15
0.2	Exhaustive Calculation	85.48	69.60	77.54
	Johnson's Algorithm	83.87	67.60	75.74
	Genetic Algorithm	82.26	70.40	76.33
0.1	Exhaustive Calculation	83.87	60.85	72.36
	Johnson's Algorithm	83.87	58.72	71.29
	Genetic Algorithm	83.87	60.14	72.01

Table 4.3 summarised the classification results when prior feature selection was not involved. Exhaustive Calculation (RSES) reduction method yielded decision rules for the highest classification accuracy at 0.5 splitting factor (92.95% accuracy) for the training set, and the lowest accuracy at 82.26% for 0.2 splitting factor. For testing set accuracy, all methods hit the highest accuracy at 0.9 splitting factor which was 83.87%. Meanwhile, at the splitting factor of 0.1, these three reduction methods produced rules which cannot improve the classification accuracies for the testing set. At this point, it can be concluded that testing data had an under-fitting problem. As a result, a later section of feature selection was implemented prior to the data mining process to improve the classification accuracy.

4.3.1.2 Customer Churn Classification using Rough Set Theory

The main objective of this experiment was to evaluate the RST-based classifier and non-RST based classifier, as well as, to access the classification accuracy with feature selection. Table 4.4 showed the results for RST-based classifier with different splitting factor ranging from 0.9 to 0.1. For the overall dataset, Genetic Algorithm achieved the highest accuracy for training dataset which was 94.40% accuracy. Then, it can be concluded that 40% of randomly picked training datasets are able to produce rules which fit to classify another 60% of the testing datasets. Exhaustive Calculation hit the lowest classification accuracy at 0.1 splitting factor which was 80.65%. This may be caused by the under-fitting problem. Meanwhile, in the testing set, Exhaustive Calculation achieved highest accuracy of 90.32% with 0.9 splitting factor and 90.39% for the training set. At this point, it can be declared that the learning rules obtained were capable of classifying the testing set even though it may face over-fitting problems.

Table 4.4: Customer Churn Classification Accuracy for local Telecommunication Company Dataset

Split Factor	Reduction Method	Training Set Accuracy (%)	Testing Set Accuracy (%)	Overall Accuracy (%)
0.9	Exhaustive Calculation	90.39	90.32	90.36
	Johnson's Algorithm	89.68	90.32	90.00
	Genetic Algorithm	90.39	77.42	83.91
0.8	Exhaustive Calculation	91.60	75.81	83.7
	Johnson's Algorithm	89.60	80.64	85.12
	Genetic Algorithm	90.00	79.03	84.52
0.7	Exhaustive Calculation	86.24	75.53	80.89
	Johnson's Algorithm	85.32	73.40	79.36
	Genetic Algorithm	84.86	70.21	77.54
0.6	Exhaustive Calculation	87.17	76.80	81.98
	Johnson's Algorithm	84.49	76.00	80.25
	Genetic Algorithm	88.77	78.40	83.59
0.5	Exhaustive Calculation	87.82	76.92	82.37
	Johnson's Algorithm	88.46	73.72	81.09
	Genetic Algorithm	87.82	75.64	81.73
0.4	Exhaustive Calculation	88.80	69.52	79.16
	Johnson's Algorithm	87.20	65.78	76.49
	Genetic Algorithm	94.40	73.80	84.10
0.3	Exhaustive Calculation	89.36	71.56	80.46
	Johnson's Algorithm	84.04	67.89	75.97
	Genetic Algorithm	91.49	72.48	81.98
0.2	Exhaustive Calculation	87.10	72.40	79.75
	Johnson's Algorithm	82.26	70.80	76.53
	Genetic Algorithm	83.87	71.60	77.74
0.1	Exhaustive Calculation	80.65	60.85	70.75
	Johnson's Algorithm	87.10	60.85	73.98
	Genetic Algorithm	83.87	61.21	72.54

Lastly, at 0.1 split factor, Exhaustive Calculation reached the lowest accuracy at 80.65% for the training set and 60.85% for the testing set. This may be caused by an imbalanced number of data between the training set and the testing set which led to under-fitting. Splitting factors ranging from 0.6 to 0.8 gave the most optimum classification accuracy since there was no significant difference between the training and testing sets.

Table 4.5: Classification Accuracy without Feature Selection and With Feature Selection

Split Factor	Reduction Method	Without Feature Selection (%)	With Feature Selection (%)
0.9	Exhaustive Calculation	86.95	90.36
	Johnson's Algorithm	85.35	90.00
	Genetic Algorithm	86.95	83.91
0.8	Exhaustive Calculation	84.12	83.70
	Johnson's Algorithm	84.73	85.12
	Genetic Algorithm	83.51	84.52
0.7	Exhaustive Calculation	76.24	80.89
	Johnson's Algorithm	83.55	79.36
	Genetic Algorithm	81.65	77.54
0.6	Exhaustive Calculation	82.12	81.98
	Johnson's Algorithm	81.85	80.25
	Genetic Algorithm	82.39	83.59
0.5	Exhaustive Calculation	84.29	82.37
	Johnson's Algorithm	83.01	81.09
	Genetic Algorithm	82.05	81.73
0.4	Exhaustive Calculation	80.76	79.16
	Johnson's Algorithm	81.43	76.49
	Genetic Algorithm	81.57	84.10
0.3	Exhaustive Calculation	78.03	80.46
	Johnson's Algorithm	76.81	75.97
	Genetic Algorithm	81.15	81.98
0.2	Exhaustive Calculation	77.54	79.75
	Johnson's Algorithm	75.74	76.53
	Genetic Algorithm	76.33	77.74
0.1	Exhaustive Calculation	72.36	70.75
	Johnson's Algorithm	71.29	73.98
	Genetic Algorithm	72.01	72.54

Table 4.5 presented the classification accuracy without feature selection and with feature selection prior to classification. Feature selection did not improve classification accuracy at all splitting factors. However, at split factor 0.9, it gave better classification accuracy for Exhaustive Calculation and Johnson's Algorithm

which ranged from 86.95 % to 90.36% and 85.35 to 90.00%, respectively. Meanwhile, at the lower split factors which ranged from 0.1 to 0.3, feature selection also gave a significant improvement for the classification accuracy. Then, it can be concluded that feature selection does provide a good effect on the classification accuracy. Thus, research objectives number two (to implement the proposed model with feature reduction method and rule filtering) was successfully achieved.

4.3.1.3 Effect of Rule Filtering on Classification Accuracy

Rule filtering is a process of removing insignificant rules from the rule set which cannot improve the classification performance (Simon, Kumar & Li, 2011). In ROSETTA, it offered two types of filtering: basic rules filtering and quality filtering. An example of rule filtering would be that if there were 150 rules generated, after rules filtering, the number of rules should be lesser than 150 with the accuracies maintained or higher. In the case of maintaining the accuracy, it can be concluded that rule filtering successfully removed unused rules; while in the case of hitting higher accuracy, after filtering the rules, it proved that filtering removes insignificant rules.

Two basic aspects, RHS support and RHS accuracy, were selected. As rule filtering is an optimisation problem, any combination of aspects which leads to optimum accuracy will be accepted. By using the existing rule in Section 4.2, rules were filtered. Table 4.6 summarised the number of rules obtained before and after the rule filtering process for each reduction method. The new decision rules set were applied to the training and testing sets to investigate the effect of rules filtering towards the classification accuracy.

Table 4.6: Rules Obtained Before and after Rule Filtering

Reduction Method	Rules Generated	Rules after Filtered
Exhaustive Calculation	930	114
Johnson's Algorithm	284	134
Genetic Algorithm	932	258

Table 4.7: Rule Filtering 1

	Split factor	Reduction Method	Accuracy Without Rule Filtering (%)	Accuracy With Rule Filtering (%)
Training Set	0.5	Exhaustive Calculation	87.82	84.62
Testing Set			76.92	75.64

Referring to Table 4.7, rules induced from Exhaustive Calculation reduction method were filtered. Prior to the rules filtering process, 930 rules were available. When basic rule filtering was applied to these rules, only 114 rules were left. When these new rule sets were applied to the training and testing datasets, it reduced the classification accuracies; as in Table 4.7. In conclusion, any significant and important rules were removed. Thus, it reduced the classification accuracy. So, these new decision rule sets are not suitable for customer churn classification.

Table 4.8: Rule Filtering 2

	Split factor	Reduction Method	Accuracy Without Rule Filtering (%)	Accuracy With Rule Filtering (%)
Training Set	0.5	Johnson's Algorithm	88.46	90.38
Testing Set			76.92	76.92

Referring to Table 4.8, rules induced from Johnson's Algorithm reduction method were filtered. Prior to the rules filtering process, 284 rules were available. When basic rule filtering was applied to these rules, only 134 rules were left without reducing the testing set accuracy. Then, it can be concluded that insignificant rules were removed by improving the classification accuracy to 90.38% at the training set and maintaining classification accuracy for the testing set. So, these new decision rules sets can be applied on customer churn classification.

Table 4.9: Rule Filtering 3

	Split factor	Reduction Method	Accuracy Without Rule Filtering (%)	Accuracy With Rule Filtering (%)
Training Set	0.5	Genetic Algorithm	87.82	87.18
Testing Set			75.64	76.28

Lastly, the experiment employed Genetic Algorithm induced rules. According to Table 4.6, before basic rules filtering were applied to these rule sets, 932 rules were available but only 258 rules were left after filtering and referring to Table 4.9, these 258 rules improved classification accuracy from 75.64% to 76.28%. Hence, it can be concluded that rule filtering removed insignificant and irrelevant rules. Moreover, it fastens the speed of the classification process, as when classifying objects, it will go through less rules-checking. However, Suraj, Gayar and Delimata (2006) stated that in order to classify data properly, a large set of decision rules is required. In this case, 258 rules were enough to classify data since it helped improve classification accuracies. Thus, research objective number two was successfully achieved.

4.3.1.4 Comparison of Classifier Accuracy

In order to evaluate the performance of non-RST based and RST-based classifiers, Regression Analysis, J48 (Decision Tree) classifier and Voted Perceptron (Neural Network) classifier were applied to the local telecommunication company dataset using the WEKA software. The same supervised learning approach was utilised in these experiments. For the RST-based classifier, Standard Voting/Tuned (RSES) with Genetic Algorithm reduction method results from a previous section were chosen. Table 4.10 depicted the classification accuracy for different non-RST based and RST-based classifiers.

In Figure 4.6, the performance of RST-based classifier and Non-RST classifier were shown. At all splitting factors, RST-based classifier achieved great improvements in classification accuracies. It can be concluded that RST was capable and efficient in classifying churners and non-churners; thus, the third research objective was achieved.

Table 4.10: Classification Accuracies for Non-RST Classifier and RST Classifier

		Classifiers with Classification Accuracy (%)			
		Linear Regression	**J48**	**Voted Perceptron**	**Rough Set Based Classifier**
Splitting Factor	0.9	19.45	54.80	54.80	90.32
	0.8	38.89	64.50	69.40	75.81
	0.7	43.85	55.30	73.40	75.53
	0.6	44.38	55.30	70.40	76.80
	0.5	43.89	55.10	66.00	76.92
	0.4	39.13	57.80	68.40	69.52
	0.3	31.32	57.80	64.70	71.56
	0.2	27.91	56.00	60.40	72.40
	0.1	37.12	59.89	54.80	60.85

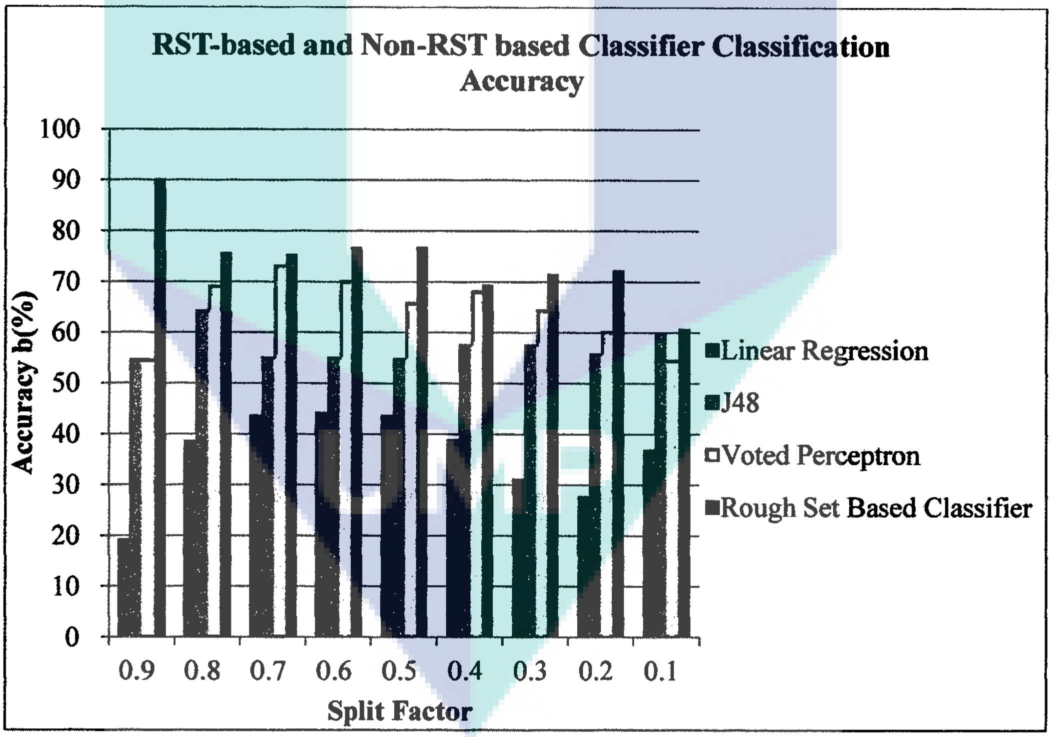


Figure 4.6: RST-based and Non-RST based Classifier Classification Accuracy

4.3.2 Applying Rough Set Theory Classification on Customer Churn Datasets

The main objective of this experiment was to prove that the chosen classification technique for this research was the best technique to classify customer churn. In addition, this experiment validated the proposed framework. UCI Repository and IBM Watson Analytic customer churn datasets were tested only in the classification process, without rules filtering and system development.

4.3.2.1 UCI Machine Learning Repository Customer Churn Dataset

UCI Machine Learning Repository customer churn dataset consisted of 1667 objects. From 21 features, only 5 were obtained from the feature selection process. The classification results were presented in Table 4.11. At all splitting factors, classification accuracies performed well, which were above 80% accuracy. The highest performance model was developed by Genetic Algorithm at 0.9 splitting factor, while the lowest performance model developed by Exhaustive Calculation and Genetic Algorithm reduction methods were at split factor 0.3 which was 86.52%.

Table 4.11: UCI Machine Learning Repository Customer Churn Dataset Classification Accuracy

Split Factor	Reduction Method	Training Set Accuracy (%)	Testing Set Accuracy (%)	Overall Accuracy (%)
0.9	Exhaustive Calculation	86.04%	90.59	88.32
	Johnson's Algorithm	86.18%	90.59	88.38
	Genetic Algorithm	86.24%	90.59	88.42
0.8	Exhaustive Calculation	85.77%	89.48	87.62
	Johnson's Algorithm	85.92%	89.48	87.70
	Genetic Algorithm	85.77%	89.48	87.62
0.7	Exhaustive Calculation	86.18%	87.66	86.92

Split Factor	Reduction Method	Training Set Accuracy (%)	Testing Set Accuracy (%)	Overall Accuracy (%)
	Johnson's Algorithm	86.09%	87.86	86.98
	Genetic Algorithm	86.18%	87.66	86.92
0.6	Exhaustive Calculation	86.26%	87.64	86.95
	Johnson's Algorithm	86.16%	87.64	86.90
	Genetic Algorithm	86.45%	87.64	87.04
0.5	Exhaustive Calculation	86.81%	86.44	86.62
	Johnson's Algorithm	86.93%	86.44	86.68
	Genetic Algorithm	87.04%	86.44	86.74
0.4	Exhaustive Calculation	87.17%	86.36	86.76
	Johnson's Algorithm	87.17%	86.36	86.76
	Genetic Algorithm	86.88%	86.46	86.67
0.3	Exhaustive Calculation	87.63%	85.41	86.52
	Johnson's Algorithm	87.83%	85.50	86.66
	Genetic Algorithm	87.63%	85.41	86.52
0.2	Exhaustive Calculation	87.93%	86.15	87.04
	Johnson's Algorithm	87.93%	86.15	87.04
	Genetic Algorithm	87.93%	86.15	87.04
0.1	Exhaustive Calculation	88.82%	85.64	87.23
	Johnson's Algorithm	88.24%	86.31	87.27
	Genetic Algorithm	90.59%	84.50	87.54

4.3.2.2 IBM Watson Analytic Customer Churn Dataset

This dataset contained 4900 instances with 20 features. Through feature selection, only eight features were selected. Table 4.12 presented the accuracy value extracted from the classification process. The highest accuracy value of 80.74% was extracted from Johnson's algorithm for 0.7 split factor, while the lowest accuracy hit at split factor 0.3 for the Exhaustive Calculation.

Table 4.12: IBM Watson Analytic Customer Churn Dataset Classification Accuracy

Split Factor	Reduction Method	Training Set Accuracy (%)	Testing Set Accuracy (%)	Overall Accuracy (%)
0.9	Exhaustive Calculation	82.04	78.78	80.41
	Johnson's Algorithm	82.20	79.18	80.69
	Genetic Algorithm	82.00	79.18	80.5
0.8	Exhaustive Calculation	82.40	78.27	80.33
	Johnson's Algorithm	82.78	78.37	80.57
	Genetic Algorithm	82.60	78.47	80.54
0.7	Exhaustive Calculation	82.42	78.23	80.33
	Johnson's Algorithm	82.71	78.78	80.74
	Genetic Algorithm	82.57	78.30	80.43
0.6	Exhaustive Calculation	82.01	78.88	80.44
	Johnson's Algorithm	82.14	78.93	80.54
	Genetic Algorithm	82.14	79.13	80.64
0.5	Exhaustive Calculation	80.98	78.82	79.90
	Johnson's Algorithm	81.39	78.69	80.04
	Genetic Algorithm	81.63	79.14	80.39
0.4	Exhaustive Calculation	82.14	77.86	80.00
	Johnson's Algorithm	81.79	77.86	79.82
	Genetic Algorithm	82.35	77.89	80.12
0.3	Exhaustive Calculation	81.56	76.97	79.27
	Johnson's Algorithm	81.97	77.49	79.73
	Genetic Algorithm	82.31	77.08	79.70
0.2	Exhaustive Calculation	83.06	76.76	79.91
	Johnson's Algorithm	82.96	76.61	79.78
	Genetic Algorithm	82.76	76.81	79.78
0.1	Exhaustive Calculation	86.12	76.53	81.33
	Johnson's Algorithm	84.08	75.83	79.95
	Genetic Algorithm	83.24	76.15	79.69

4.3.3 Customer Churn Classification System

The best rule set from previous experiments had been chosen to develop a simple customer churn classification system. Referring to rules filtering step, it can be concluded that Genetic Algorithm induced rules was better for selection for the development of classification system. In Figure 4.7, it shows the system flowchart.

Hypertext Pre-processor (PHP) was employed to develop the online based system and MySQL for database.

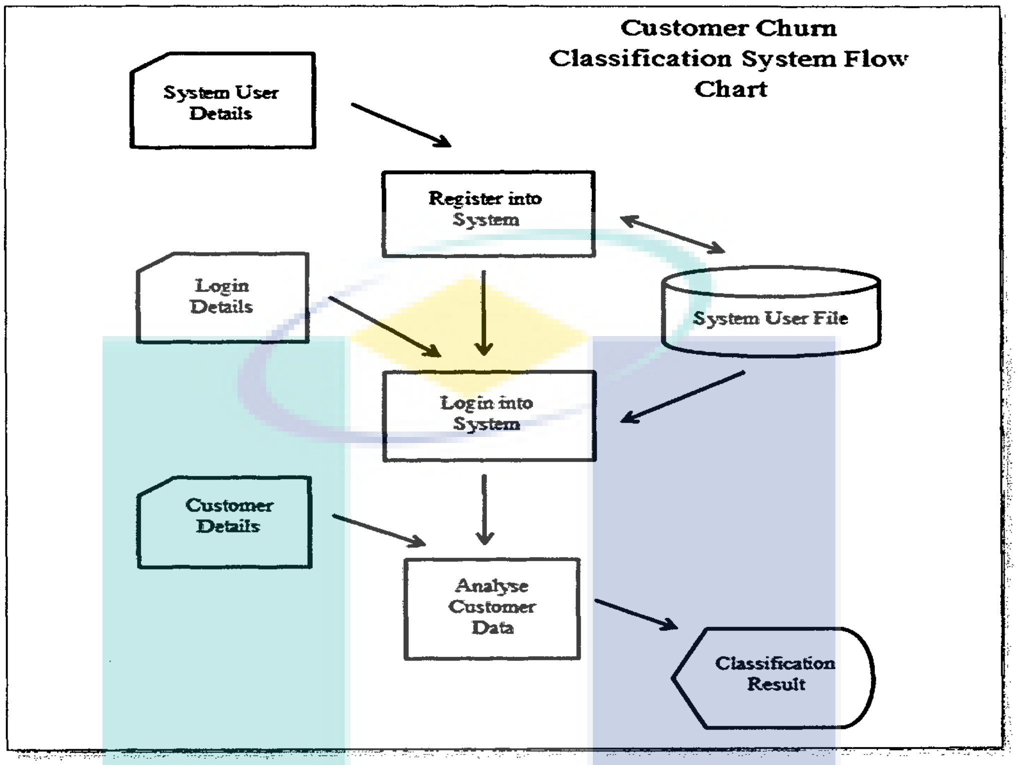


Figure 4.7: Customer Churn Classification System Flow Chart

Figure 4.8 illustrates the Login Page of Customer Churn Classification System. This system required the user to register into the system and login as an authorize user and Figure 4.9 displays the user registration page.

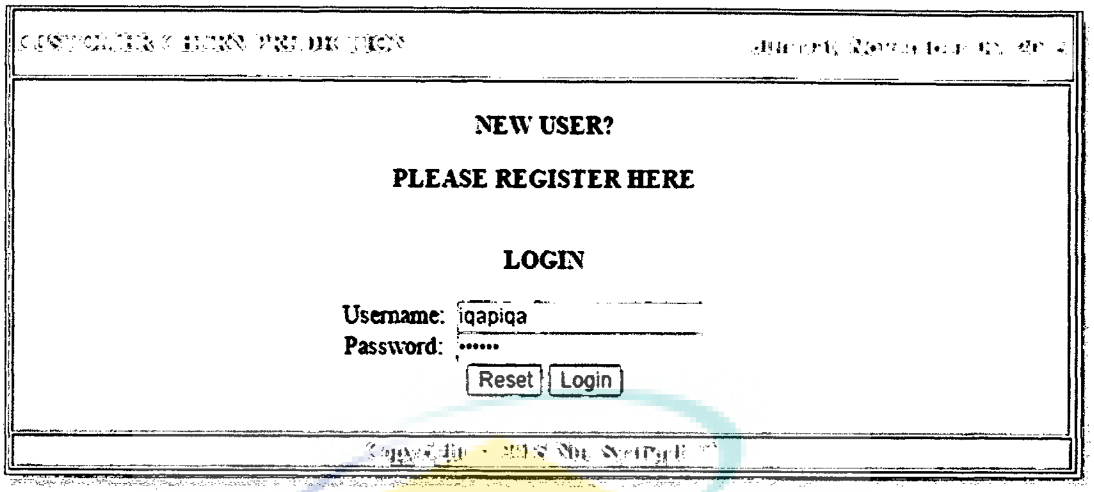


Figure 4.8: Login Page of Customer Churn Prediction System

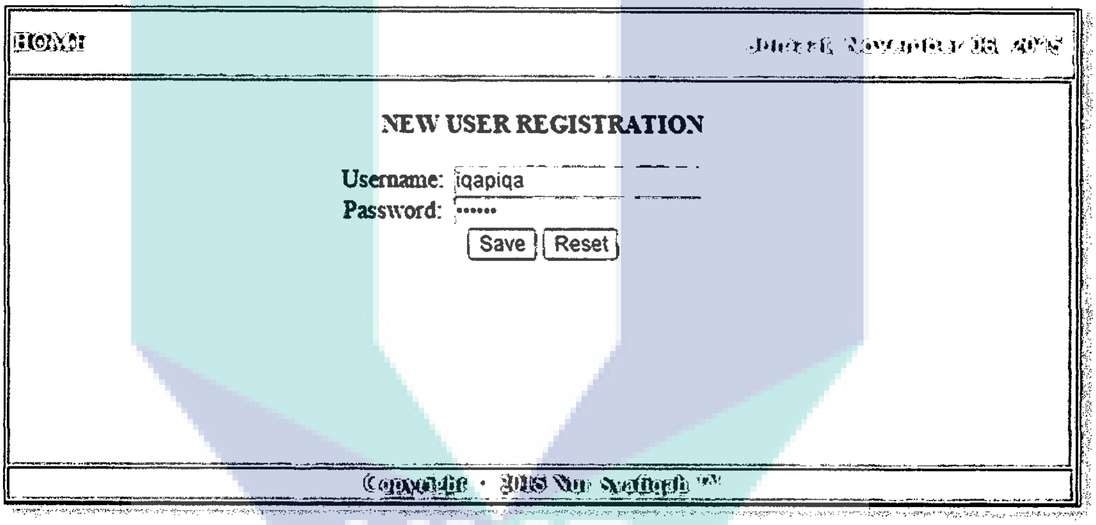


Figure 4.9: System Registration Page of Customer Churn Prediction System

The main page where user input the customer records was displayed in Figure 4.10. This system consisted of two-panels; left-side panel was the input panel and the right-side panel was the output panel. In input panel, user needs to input seven inputs. These inputs are the seven attributes retrieved from attribute selection as had been discussed in Section 4.3.1.3. The example of output after record had been submitted into the system is shown in Figure 4.11 and Figure 4.12.

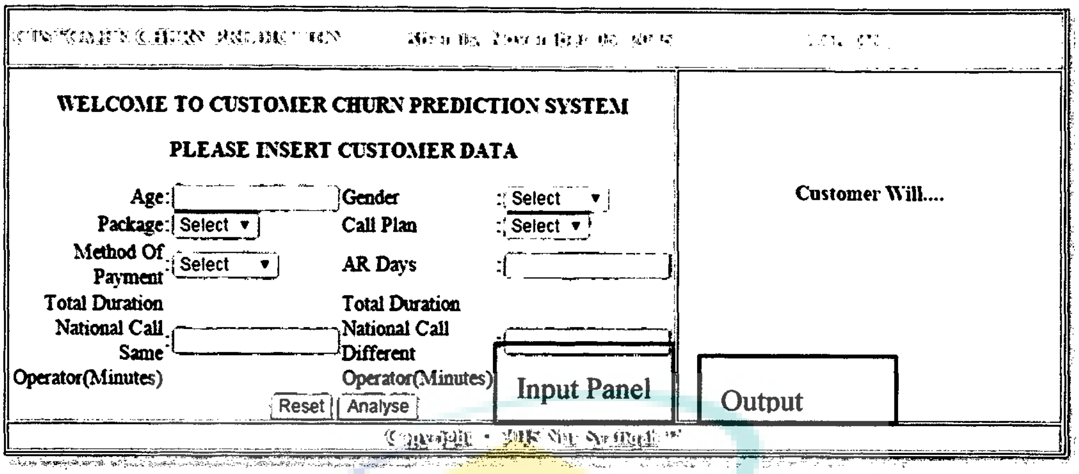


Figure 4.10: Customer Churn Classification System Interface

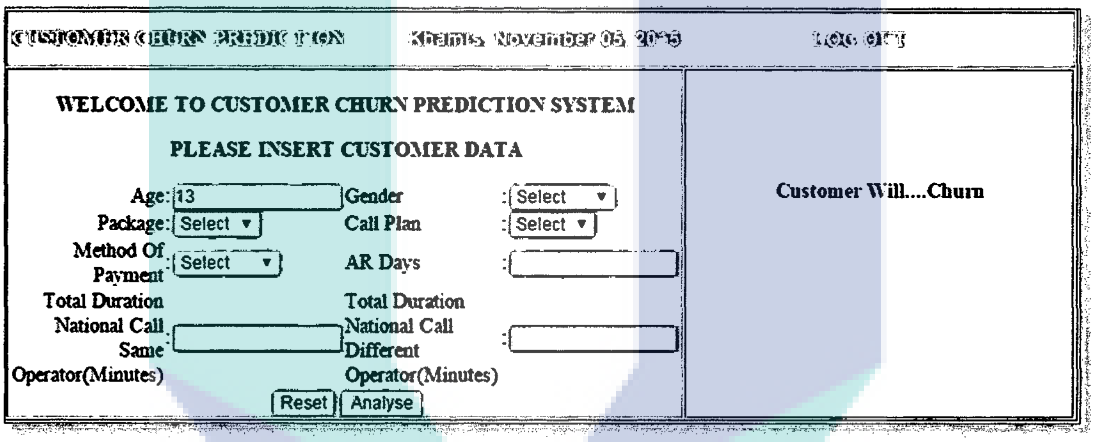


Figure 4.11: Output if Customer is classified as Churner

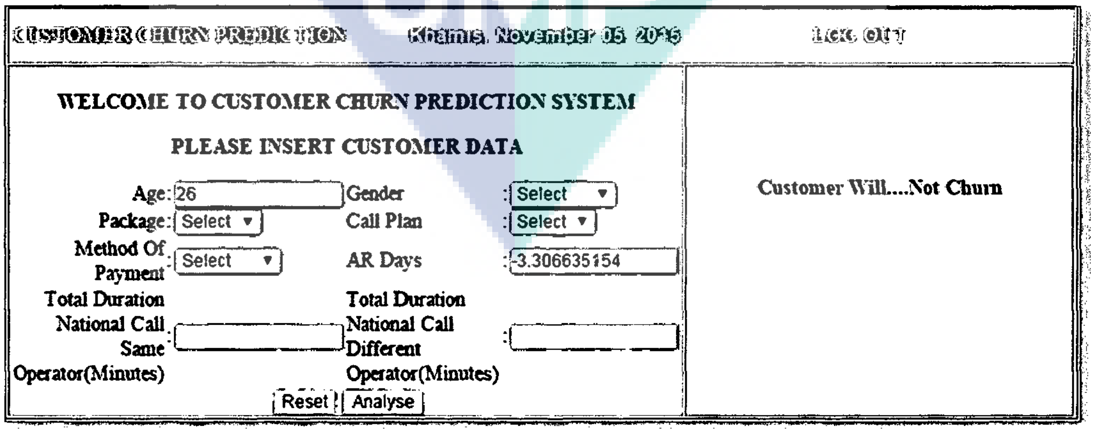


Figure 4.12: Output if Customer is classified as Non-Churner

4.4 Concluding Remarks

This chapter explained the implementations and results obtained from the classification process using the proposed model. Several experiments were conducted to test RST classification model using different datasets. The local telecommunication company dataset was experimented as the case study of the research, and the dataset retrieved from UCI Machine Learning Repository and IBM Watson Analytic Repository were used as a validation and verification for the proposed framework. Each dataset was tested by either using different splitting factors, different numbers of training and testing datasets or different reduction algorithms. All these features provided various classification accuracies. The effects of feature selection and rule filtering were also discussed in this chapter. Lastly, the end of this chapter provided some explanations on the Customer Churn Prediction System developed.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

In this chapter, the results of using the RST techniques on classifying customer churn were discussed. Research advantages, research weaknesses and research contributions were also considered. Moreover, some suggestions for future works were stated for those who are interested in continuing or improving this work in this research field.

5.2 Summary

This research was all about classifying customer churn data for a local telecommunication company. The customer churn classification model was successfully developed in this research using RST. Hence, research objective number one was achieved. Secondly, this research also evaluated the effects of feature selection prior to the classification process. Based on the experiments, feature selection does improve the classification accuracies.

In order to yield the best classification accuracy, three reduction attributes were accessed. Genetic Algorithm produced rules with the highest classification accuracies. Rules filtering also assisted in improving the classification accuracy by removing unused and insignificant rules. Hence, it reduced the processing time. Next, the second research objective was fulfilled. Several experiments were also performed to test RST-based and non-RST based classifiers. RST-based classifier performed well among the classifiers (Regression Analysis, Decision Tree and Voted Perceptron) which satisfied the third research objective.

After analysing all the results from the experiment, one decision was made to establish which of the results retrieved was more appropriate for this research. All aspects were analysed to reach the decision that the RST-based classifier was the best classifier to classify the local telecommunication company's customer churn dataset. Consequently, RST has useful methods that can help produce better results. In addition, the set of rules produced by the RST model was very informative. These rules can be converted into a predictive system to assist in identifying customer churn for local telecommunication companies.

In conclusion, this research focused on classifying churners and non-churners in a local telecommunication company. The new classification model obtained was capable in classifying customer churn with the aid of data splitting, feature selection, data discretisation, attribute reduction and rule filtering processes. Overall, this research had fulfilled the three objectives stated in Section 1.3.

5.3 Research Advantages

The following are the advantages which were identified throughout the research:

i. This research had proved that the RST technique was able to do the classification process in order to classify customer churn.

ii. It gave some exposure to the researchers on implementing RST technique in classifying customer churn in telecommunication industries.

iii. It inspired other researchers to explore more on classification tasks using the RST technique in order to improve this project for a better approach and outcome.

5.4 Research Contributions

The list below listed some of the contributions of the research:

i. This research determined whether the RST technique was truly capable to be applied on customer churn classification for the local telecommunication company which contributes the most to the data mining field.

ii. This research also contributed to the local telecommunication company whereby from this research, a new model and a new system using an integrated KDD process and RST to classify customer churn was produced. In addition, this new framework and model also contains the best feature, attribute reduction and rule filtering. Hopefully, it will be help firms to optimize their profit.

iii. A new system to assist the local telecommunication company to classify customer churns.

5.5 Research Limitations

The previous section outlined the contributions of this research to the domain of the knowledge. However, this research had some limitations:

i. Firstly, the performance of the data classification depended on the size of data set. Bigger sized data can contribute to better learning patterns. The limitation of data was due to company policies since customer data were protected under some acts.

ii. Next, the dataset used in this research was based on one the local telecommunication companies in Malaysia only. In order to obtain a good result which presented real situations of the current telecommunication industry in Malaysia, the data analysis should be collected from various providers.

5.6 Suggestion and Further Works

In the future, hopefully the weaknesses and limitations of this research will be improved and hence produces better accuracy in customer churn classification. The following is a list of suggestions that can be done to improve this work in the future:

i. Use a large number of genuine dataset from the various local telecommunication companies. Broad data coverage will assist researchers to understand the real phenomenon happen local telecommunication industry.

ii. Utilizing big portion of dataset in the experiments. For example, 500 to 1000 records and split bigger portions for training dataset for better classification accuracy. The more dataset used in the training process, the more accurate the classification task because a lot of data used will let the classifier learn the behaviour of the dataset.

iii. Use more significant attributes to conduct RST classification task. For example, telecommunication companies provide significant attributes such as usage period, time taken for a customer to churn and customer's behaviour before churn. The more significant features used, the more accurate the classification

iv. Lastly, use a combination of RST and other techniques such as Support Vector Machine (SVM) or Principle Component Analysis or Formal Concept Analysis (FCA).

REFERENCES

Abdul Aziz, A., Ismail, N. H., & Ahmad, F. (2013). Mining students ' academic performance. *Journal of Theoretical and Applied Information Technology*, *53*(3), 485–495. Retrieved from http://www.jatit.org/

Abdul Rahman, M. N., Mohd Lazim, Y., & Mohamed, F. (2011). Applying rough set theory in multimedia data classification. *International Journal onNew Computer Architectures and Their Applications*, *1*, 683–693. http://doi.org/10.1.1.233.1039

Abdul Rahman, M. N., Mohd Lazim, Y., Mohamed, F., Saany, S. I. A., & Mohd Yusof, M. K. (2013). Rules generation for multimedia data classifying using rough sets theory. *International Journal of Hybrid Information Technology*, *6*, 209–218. http://doi.org/10.1.1.396.788

Agotnes, T. (1999). *Filtering large propositional rule sets while retaining classifier performance. (Master dissertation)*. Retrieved from http://citeseerx.ist.psu.edu/

Akbar, Z. (2003). Marketing data classification using johnson's algorithm. In *Proceedings of 4th International Journees de I'Informatique Messine:Knowledge Discovery and Discrete Mathematics, 3-6 Sept.2003* (pp. 257–266). Metz: INRIA.

Al-radaideh, Q. (2008). The impact of classification evaluation methods on rough set based classifier. In *Proceedings of the 2008 International Arab Conference on Information Technology* (pp. 1–5). Tunisia: ACIT2008.

Amin, A., Shehzad, S., Khan, C., Ali, I., & Anwar, S. (2015). Churn prediction in telecommunication industry using rough set approach. In D. Camacho, S.-W. Kim, & B. Tranwinski (Eds.), *New Trends in Computational Collective Intelligence* (1st ed., pp. 83–95). Springer International Publishing.

An, A. (2005). Classification Methods. In J. Wang (Ed.), *Encyclopedia of Data Warehousing and Mining* (2nd ed., pp. 1–6). Idea Group Inc.

Anouncia, M., Clara Madonna, L., P, J., & Nandhini, R. T. (2013). Design of a diabetic diagnosis system using rough sets. *Cybernetics And Information Technologies*, *13*, 124–139. http://doi.org/10.2478/cait-2013-0030

Au, W., Chan, K. C. C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transaction On Evolutionary Computation*, *7*, 532–545. http://doi.org/10.1109/TEVC.2003.819264

Bal, M. (2013). Sets theory as symbolic data mining method : an application on complete decision table. *Information Science Letters*, *2*(1), 35–47. Retrieved from http://www.naturalspublishing.com/

Bazan, J. G. (1998). A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables. *Rough Set Knowledge Discovery*, *1*, 321–365. http://doi.org/10.1.1.36.8738

Bin, L., Peiji, S., & Juan, L. (2007). Customer churn prediction based on the decision tree in personal handyphone system service. In *International Conference on Service Systems and Service Management 2007, 9-11 June 2007* (pp. 1–5). Chengdu: IEEE.

Cunningham, S. J., & Holmes, G. (1999). Developing innovative applications in agriculture using data mining. In *The Proceedings of The Southeast Asia Regional Computer Confederation Conference* (pp. 25–29). Singapore: SEARCC.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis*, *1*(3), 131–156. Retrieved from www.elsevier.com

Dawson, A. J., Stasa, H., Roche, M. A., Homer, C. S. E., & Duffield, C. (2014). Nursing churn and turnover in australian hospitals: nurses perceptions and suggestions for supportive strategies. *BMC Nursing*, *13*, 1–10. http://doi.org/10.1186/1472-6955-13-11

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*, 37–54. http://doi.org/10.1609/aimag.v17i3.1230

Godfrey, P. B., Shenker, S., & Stoica, I. (2006). Minimizing churn in distributed systems. In *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Potocols for Computer Communications, 11-15 Sept.2006* (pp. 147–158). Pisa: ACM.

Gürsoy, Ş., & Tuğba, U. (2010). Customer churn analysis in telecommunication sector. *Journal of the School of Business Administration, Istanbul University*, *39*(1), 35–49. Retrieved from http://dergipark.ulakbim.gov.tr/

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(3), 1157–1182. Retrieved from http://www.jmlr.org/

Hadden, J. (2008). *A customer profiling methodology for churn prediction. (Doctoral dissertation)*. Retrieved from http://dspace.lib.cranfield.ac.uk/

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2006). Churn prediction using complaints data. In *The Proceedings of World Academy of Science, Engineering and Technology 19* (Vol. 13, pp. 158–163). Budapest: WASET.

Hongxia, M., & Min, Q. (2009). Research method of customer churn crisis based on decision tree. In *Management and Service Science International Conference 2009, 20-22 Sept. 2009* (pp. 1–4). Wuhan: IEEE.

Hu, D. (2008). The study on rough set theory for customers churn. In *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, 12-14 Oct. 2008* (pp. 1–4). Dalian: IEEE.

Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., … Zeng, J. (2015). Telco churn prediction with big data. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data,12 May-4 June 2015* (pp. 607–618). Melbourne: ACM.

Ismail, M. R. (2015). *Customer churn prediction in artificial neural network model. (Unpublished master dissertation)*. Universiti Sultan Zainal Abidin.

Ismail, M. R., Awang, M. K., Abdul Rahman, M. N., & Makhtar, M. (2015). A multi-layer perceptron approach for customer churn prediction. *International Journal of Multimedia and Ubiquitous Engineering*, *10*, 213–222. http://doi.org/10.14257/ijmue.2015.10.7.22

Jadhav, R. J., & Pawar, U. T. (2011). Churn prediction in telecommunication using data mining technology. *International Journal of Advanced Computer Science and Applications*, *2*, 17–19. http://doi.org/10.1.1.622.6055

Jiang, H., & Ruan, J. (2010). Applying rough set theory to evaluate network marketing performance of china's agricultural products. *Journal of Computers*, *5*, 1264–1272. http://doi.org/10.4304/jcp.5.8.1264-1272

Kamarudin, N. S. (2015). *Framework for single object images in multiple categories using ensemble method. (Unpublished master dissertation).* Universiti Sultan Zainal Abidin.

Kamarudin, N. S., Makhtar, M., Fadzli, S. A., Mohamad, M., Mohamad, F. S., & Abdul Kadir, M. F. (2015). Comparison of image classification techniques using caltech 101 dataset. *Journal of Theoritical and Applied Information Technology*, *71*(1), 79–86. Retrieved from www.jatit.org

Khan, A. A., Jamwal, S., & Sepehri, M. M. (2010). Applying data mining to customer churn prediction in an internet service provider. *International Journal of Computer Applications*, *9*, 8–14. http://doi.org/10.5120/1400-1889

Khan, I., Usman, I., Usman, T., Rehman, G. U., & Ateeq, R. U. (2013). Intelligent churn prediction for telecommunication industry. *International Journal of Innovation and Applied Studies*, *4*(1), 165–170. Retrieved from http://www.ijias.issr-journals.org/

Khoza, M., & Marwala, T. (2011). A rough set theory based predictive model for stock prices. In *Computational Intelligence and Informatics, IEEE 12th International Symposium 2011,21-22 Nov. 2011* (pp. 57–62). Budapest: IEEE.

Kirui, C., Hong, L., Cheruiyot, W., & Kirui, H. (2013). Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. *International Journal of Computer Science Issues*, *10*(2), 1694–0784. Retrieved from http://www.ijcsi.org/

Kojo, A.-A. J. (2011). *Predicting Customer Churn In The Mobile Telecommunication Industry, A "Case Study Of Mtn Ghana, Kumasi." (Master dissertation).* Retrieved from http://dspace.knust.edu.gh

Ladha, L., & Deepa, T. (2011). Feature selection methods and algorithms. *International Journal on Computer Science and Engineering*, *3*(5), 1787–1797. Retrieved from http://www.enggjournals.com/

Lazarov, V., & Capota, M. (2007). Churn prediction. *Bussiness Analyst Course*. Retrieved from http://www.vladislav.lazarov.pro/

Li, F., Lei, J., Tian, Y., Punyapatthanakul, S., & Wang, Y. J. (2011). Model selection strategy for customer attrition risk prediction in retail banking. In *Proceedings of the 9-th Australasian Data Mining Conference, 1-2Dec. 2011* (pp. 119–124). Ballarat: Australian Computer Society Inc.

Li, J. (2007). *Rough set based rule evaluations and their applications. (Doctoral dissertation)*. Retrieved from http://dl.acm.org/

Lin, C. S., Tzeng, G. H., & Chin, Y. C. (2011). Combined rough set theory and flow network graph to predict customer churn in credit card accounts. *Expert Systems with Applications*, *38*, 8–15. http://doi.org/10.1016/j.eswa.2010.05.039

Lu, J. (2002). Predicting customer churn in the telecommunications industry - an application of survival analysis modeling using SAS. In *SAS Conference Proceedings: SAS Users Group International 27* (pp. 114–27). Orlando: SAS Institute Inc.

Malaysian Communications and Multimedia Commission. (2013). *MCMC Annual Report 2013*. Retrieved from http://www.skmm.gov.my/

Mohammadi, G., Tavakkoli-Moghaddam, R., & Mohammadi, M. (2013). Hierarchical neural regression models for customer churn prediction. *Journal of Engineering*, *2013*, 1–9. http://doi.org/10.1155/2013/543940

Nath, S. V, & Behara, R. S. (2003). Customer churn analysis in the wireless industry : a data mining approach. In *Proceedings-Annual Meeting of the Decision Sciences Institute, 22-25 Nov.2003* (pp. 505–510). Washington: DSI.

Nguyen, H. S. (1998). Discretization problem for rough sets methods. In *Proceedings of the First International Conference on Rough Sets and Current Trends in Computing, 22-26 June 1998* (pp. 545–552). Warsaw: Springer Berlin Heidelberg.

Ohrn, A. (2000a). *Discernibility and rough sets in medicine : tools and applications. (Doctoral dissertation)*. Retrieved from http://citeseerx.ist.psu.edu/

Ohrn, A. (2000b). *ROSETTA technical reference manual*. ROSETTA Development Team.

Ong, D. L. T., Tan, M. S. L., & Andrews, E. (2012). A study of customer retention and churn rate management through data mining and customer profiling of malaysian mobile users. In *Academy of World Business, Marketing & Management Development Conference Proceedings, 16-19 July 2012* (pp. 90–105). Budapest: Journal of Management and World Business Research.

Oseman, K., Mohd Shukor, S., Abu Haris, N., & Abu Bakar, F. (2010). Data mining in churn analysis model for telecommunication industry. *Journal of Statistical Modeling and Analytics*, *1*(19), 19–27. Retrieved from http://www.academia.edu/

Ozer, P. (2008). *Data mining algorithms for classification. (Bachelor dissertation)*. Retrieved from http://citeseerx.ist.psu.edu/

Pawlak, Z. (1996). Rough sets and data analysis. In *Soft Computing in Intelligent Systems and Information Processing. Proceedings of the 1996 Asian Fuzzy Systems Symposium, 11-4 Dec.1996* (pp. 1–12). Keting: IEEE.

Perumal, K., & Bhaskaran, R. (2010). Supervised classification performance of multispectral images. *Journal Of Computing*, *2*(2), 124–129. Retrieved from https://www.researchgate.net/

Prasad, U. D., & Madhavi, S. (2012). Prediction of churn behavior of bank customers using data mining tools. *Business Intelligence Journal*, *5*(1), 96–101. Retrieved from http://www.indianjournalofmarketing.com/

Rashid, T. (2008). Classification of churn and non-churn customers for telecommunication companies. *International Journal of Biometrics and Bioinformatics*, *3*(5), 82–89. Retrieved from https://www.researchgate.net/

Sarkar, S. (2014). An improved rough set data model for stock market prediction. In *Business and Information Management (ICBIM), 2014 2nd International Conferencee,9-11 Jan. 2014* (pp. 96–100). Durgapur: IEEE.

Senthilkumaran, N., & Rajesh, R. (2009). A study on rough set theory for medical image segmentation. *International Journal of Recent Trends in Engineering*, *2*(2), 236–238. Retrieved from http://ijrte.academypublisher.com/

Sharma, A., & Panigrahi, P. K. (2011). A neural network based approach for predicting customer churn in cellular network services. *International Journal of Computer Applications*, *27*, 26–31. http://doi.org/10.5120/3344-4605

Simon, G. J., Kumar, V., & Li, P. W. (2011). A simple statistical model and association rule filtering for classification. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 21-24 August 2011* (pp. 823–831). San Diego: ACM.

Soeini, R. A., & Rodpysh, K. V. (2012). Applying data mining to insurance customer churn management. In *IACSIT Hong Kong Conference, 5-7 Jan.2012* (pp. 82–92). Hong Kong: IACSIT Press, Singapore.

Suraj, Z. (2004). An introduction to rough set theory and its applications. In *The 1st International Computer Engineering Conference New Technologies for the Information Society, 27-31 Dec,2004* (pp. 1–39). Cairo: ICENCO.

Suraj, Z., Gayar, N. El, & Delimata, P. (2006). A rough set approach to multiple classifier systems. *Fundamenta Informaticae*, *72*, 393–406. http://doi.org/10.1016/S0893

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Classification : basic concepts , decision trees , and model evaluation. In *Introduction to Data Mining* (1st ed., pp. 145–205). Pearson Addison Wesley.

Tian, R., Yang, Z., & Zhang, M. (2010). Method of road traffic accidents causes analysis based on data mining. In *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference, 10-12 Dec.2010* (pp. 1–4). Wuhan: IEEE.

Tripathy, B. ., Acharjya, D. ., & Cynthya, V. (2011). A framework for intelligent medical diagnosis using rough set with formal concept analysis. *International Journal of Artificial Intelligence & Applications*, *2*, 45–66. http://doi.org/10.5121/ijaia.2011.2204.

Tsai, C.-F., & Chen, M.-Y. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems with Applications*, *37*, 2006–2015. http://doi.org/10.1016/j.eswa.2009.06.076

Tuv, E., Borisov, A., Runger, G., & Torkkola, K. (2009). Feature selection with ensembles , artificial variables , and redundancy elimination. *Journal of Machine*

Learning Research, *10*, 1341–1366. http://doi.org/10.1.1.193.5851

Van den Poel, D., & Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, *157*, 196–217. http://doi.org/10.1.1.62.8919

Venkateswara Reddy, E., Suresh, G. ., Reddy, E. ., & Shaik, S. S. (2011). Rough set analysis for uncertain data classification. In *Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference, 8-9 Dec.2011* (pp. 22–29). Chennai: IEEE.

Xu, E., Liangshan, S., Xuedong, G., & Baofeng, Z. (2006). An algorithm for predicting customer churn via BP neural network based on rough set. *Science And Technology*, *1*, 6–9. http://doi.org/10.1109/APSCC.2006.23

Yahia, M. E., & El-Taher, M. E. M. (2010). A new approach for evaluation of data mining techniques. *International Journal of Computer Science Issues*, *7*(5), 1694–814. Retrieved from www.IJCSI.org

Zhang, L., Li, Y., Sun, C., & Nadee, W. (2013). Rough set based approach to text classification. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences,17-20 Nov.2013* (pp. 245–252). Atlanta: IEEE.

Zhang, M., & Yao, J. T. (2004). A rough sets based approach to feature selection. In *Fuzzy Information, 2004. Processing NAFIPS '04. IEEE Annual Meeting, 27-30 June 2004* (pp. 434–439). Alberta: IEEE.

Zhoa, W., & Zhang, Z. (2005). An email classification model based on rough set theory. In *Proceedings of the 2005 International Conference on Active Media Technology, 19-21 May 2005* (pp. 403–408). Kagawa: IEEE.

UMP

APPENDICES

Appendix A: RulesGenerated From Reduction Set

% Rules/patterns generated by ROSETTA.

% Exported 2015.10.04 13:10:12 by Administrator.

%

% DGRSES-R5(F2)

% 258 rules.

1. A3([3, 5)) AND A8([17.14100, 52.46820)) => D(1)
2. A3([*, 2)) AND A8([52.46820, *)) => D(1)
3. A2(0) AND A6([314, 1382)) => D(1)
4. A2(1) AND A6([14728, *)) => D(0)
5. A2(0) AND A6([14, 314)) => D(1)
6. A2(0) AND A6([1382, 14728)) => D(1)
7. A6([314, 1382)) AND A8([5.52027, 17.14100)) => D(1)
8. A6([14728, *)) AND A8([1.26161, 5.52027)) => D(0)
9. A6([14728, *)) AND A8([17.14100, 52.46820)) => D(0)
10. A6([314, 1382)) AND A8([52.46820, *)) => D(1)
11. A6([1382, 14728)) AND A8([17.14100, 52.46820)) => D(1)
12. A6([314, 1382)) AND A8([*, 1.26161)) => D(0)
13. A6([14, 314)) AND A8([52.46820, *)) => D(1)
14. A6([14, 314)) AND A8([17.14100, 52.46820)) => D(1)
15. A6([1382, 14728)) AND A8([1.26161, 5.52027)) => D(1)
16. A2(0) AND A8([5.52027, 17.14100)) => D(1)
17. A2(0) AND A3([*, 2)) => D(1)
18. A3([2, 3)) AND A7([1, 901)) => D(1)

19. A6([14728, *)) AND A7([901, 6746)) => D(0)

20. A6([14728, *)) AND A7([20519, *)) => D(0)

21. A6([314, 1382)) AND A7([901, 6746)) => D(1)

22. A6([314, 1382)) AND A7([6746, 20519)) => D(0)

23. A6([14, 314)) AND A7([*, 1)) => D(1)

24. A6([14, 314)) AND A7([20519, *)) => D(1)

25. A1([46, 53)) AND A2(0) AND A7([20519, *)) => D(1)

26. A1([*, 29)) AND A2(1) AND A7([901, 6746)) => D(1)

27. A1([*, 29)) AND A2(1) AND A7([20519, *)) => D(0)

28. A1([36, 46)) AND A2(0) AND A7([20519, *)) => D(0)

29. A3([5, *)) AND A6([14728, *)) => D(0)

30. A3([5, *)) AND A6([314, 1382)) => D(0)

31. A3([2, 3)) AND A6([14, 314)) => D(1)

32. A3([2, 3)) AND A6([14728, *)) => D(0)

33. A3([*, 2)) AND A6([14728, *)) => D(0)

34. A3([3, 5)) AND A6([1382, 14728)) => D(1)

35. A1([46, 53)) AND A4(2) AND A7([20519, *)) AND A8([5.52027, 17.14100)) => D(1)

36. A1([46, 53)) AND A6([14728, *)) => D(0)

37. A1([36, 46)) AND A6([14728, *)) => D(0)

38. A1([*, 29)) AND A6([314, 1382)) => D(1)

39. A1([53, *)) AND A6([314, 1382)) => D(1)

40. A1([36, 46)) AND A6([314, 1382)) => D(0)

41. A1([*, 29)) AND A6([14, 314)) => D(1)

42. A1([29, 36)) AND A6([14, 314)) => D(1)

43. A1([53, *)) AND A6([14728, *)) => D(0)

44. A1([46, 53)) AND A7([901, 6746)) => D(0)

45. A1([29, 36)) AND A7([20519, *)) => D(1)

46. A1([29, 36)) AND A7([1, 901)) => D(1)

47. A1([53, *)) AND A7([1, 901)) => D(0)

48. A1([53, *)) AND A7([20519, *)) => D(0)

49. A1([29, 36)) AND A7([901, 6746)) => D(1)

50. A2(0) AND A3([5, *)) AND A7([901, 6746)) => D(0)

51. A2(1) AND A3([5, *)) AND A7([6746, 20519)) => D(0)

52. A2(1) AND A3([5, *)) AND A7([901, 6746)) => D(1)

53. A2(0) AND A3([2, 3)) AND A7([901, 6746)) => D(1)

54. A2(1) AND A3([*, 2)) AND A7([1, 901)) => D(0)

55. A2(0) AND A3([3, 5)) AND A7([6746, 20519)) => D(0)

56. A2(0) AND A3([5, *)) AND A7([6746, 20519)) => D(1)

57. A1([46, 53)) AND A3([5, *)) AND A8([1.26161, 5.52027)) => D(0)

58. A1([46, 53)) AND A3([3, 5)) AND A8([5.52027, 17.14100)) => D(0)

59. A1([53, *)) AND A3([2, 3)) AND A8([52.46820, *)) => D(1)

60. A1([53, *)) AND A3([*, 2)) AND A8([*, 1.26161)) => D(1)

61. A1([36, 46)) AND A3([5, *)) AND A8([5.52027, 17.14100)) => D(0)

62. A1([*, 29)) AND A3([2, 3)) AND A8([5.52027, 17.14100)) => D(1)

63. A1([36, 46)) AND A3([2, 3)) AND A8([17.14100, 52.46820)) => D(1)

64. A1([53, *)) AND A3([2, 3)) AND A8([5.52027, 17.14100)) => D(1)

65. A1([*, 29)) AND A3([3, 5)) AND A8([*, 1.26161)) => D(0)

66. A1([36, 46)) AND A3([2, 3)) AND A8([5.52027, 17.14100)) => D(1)

67. A1([*, 29)) AND A3([5, *)) AND A8([52.46820, *)) => D(1)

68. A1([46, 53)) AND A3([5, *)) AND A8([5.52027, 17.14100)) => D(0)

69. A1([*, 29)) AND A3([5, *)) AND A8([17.14100, 52.46820)) => D(0)

70. A1([53, *)) AND A3([3, 5)) AND A8([5.52027, 17.14100)) => D(0)

71. A1([46, 53)) AND A3([5, *)) AND A8([17.14100, 52.46820)) => D(1)

72. A1([*, 29)) AND A3([2, 3)) AND A8([*, 1.26161)) => D(1)

73. A1([53, *)) AND A3([2, 3)) AND A8([17.14100, 52.46820)) => D(0)

74. A2(0) AND A7([901, 6746)) AND A8([1.26161, 5.52027)) => D(0)

75. A2(1) AND A7([901, 6746)) AND A8([*, 1.26161)) => D(1)

76. A2(0) AND A7([901, 6746)) AND A8([*, 1.26161)) => D(0)

77. A2(1) AND A7([1, 901)) AND A8([1.26161, 5.52027)) => D(0)

78. A2(0) AND A7([6746, 20519)) AND A8([52.46820, *)) => D(0)

79. A2(0) AND A7([6746, 20519)) AND A8([17.14100, 52.46820)) => D(1)

80. A2(0) AND A3([5, *)) AND A8([1.26161, 5.52027)) => D(0)

81. A2(1) AND A3([5, *)) AND A8([*, 1.26161)) => D(0)

82. A2(0) AND A3([2, 3)) AND A8([1.26161, 5.52027)) => D(0)

83. A2(1) AND A3([*, 2)) AND A8([1.26161, 5.52027)) => D(0)

84. A2(0) AND A3([2, 3)) AND A8([52.46820, *)) => D(1)

85. A2(0) AND A3([2, 3)) AND A8([17.14100, 52.46820)) => D(1)

86. A3([3, 5)) AND A7([*, 1)) AND A8([52.46820, *)) => D(1)

87. A3([3, 5)) AND A7([1, 901)) AND A8([5.52027, 17.14100)) => D(0)

88. A3([5, *)) AND A7([20519, *)) AND A8([5.52027, 17.14100)) => D(0)

89. A3([2, 3)) AND A7([20519, *)) AND A8([5.52027, 17.14100)) => D(1)

90. A3([3, 5)) AND A7([20519, *)) AND A8([52.46820, *)) => D(1)

91. A3([5, *)) AND A7([6746, 20519)) AND A8([52.46820, *)) => D(0)

92. A3([5, *)) AND A7([1, 901)) AND A8([*, 1.26161)) => D(1)

93. A3([*, 2)) AND A7([901, 6746)) AND A8([1.26161, 5.52027)) => D(0)

94. A3([3, 5)) AND A7([20519, *)) AND A8([5.52027, 17.14100)) => D(0)

95. A3([3, 5)) AND A7([6746, 20519)) AND A8([52.46820, *)) => D(0)

96. A3([*, 2)) AND A7([6746, 20519)) AND A8([1.26161, 5.52027)) => D(0)

97. A3([5, *)) AND A7([901, 6746)) AND A8([17.14100, 52.46820)) => D(1)

98. A4(2) AND A7([*, 1)) AND A8([52.46820, *)) => D(1)

99. A4(1) AND A7([6746, 20519)) AND A8([17.14100, 52.46820)) => D(1)

100. A4(2) AND A7([901, 6746)) AND A8([5.52027, 17.14100)) => D(1)

101. A4(2) AND A7([6746, 20519)) AND A8([1.26161, 5.52027)) => D(1)

102. A1([29, 36)) AND A2(1) AND A8([52.46820, *)) => D(1)

103. A1([53, *)) AND A2(1) AND A8([*, 1.26161)) => D(1)

104. A1([*, 29)) AND A2(0) AND A8([52.46820, *)) => D(1)

105. A1([36, 46)) AND A2(0) AND A8([1.26161, 5.52027)) => D(0)

106. A1([29, 36)) AND A2(0) AND A8([*, 1.26161)) => D(1)

107. A1([53, *)) AND A2(0) AND A8([52.46820, *)) => D(1)

108. A1([36, 46)) AND A2(0) AND A8([52.46820, *)) => D(1)

109. A1([29, 36)) AND A2(0) AND A8([52.46820, *)) => D(0)

110. A1([53, *)) AND A2(0) AND A8([*, 1.26161)) => D(0)

111. A1([46, 53)) AND A2(0) AND A8([17.14100, 52.46820)) => D(1)

112. A1([29, 36)) AND A7([*, 1)) AND A8([52.46820, *)) => D(1)

113. A1([46, 53)) AND A7([6746, 20519)) AND A8([5.52027, 17.14100)) => D(0)

114. A1([46, 53)) AND A7([1, 901)) AND A8([5.52027, 17.14100)) => D(0)

115. A1([53, *)) AND A7([6746, 20519)) AND A8([*, 1.26161)) => D(1)

116. A1([36, 46)) AND A7([6746, 20519)) AND A8([17.14100, 52.46820)) => D(1)

117. A1([36, 46)) AND A7([1, 901)) AND A8([5.52027, 17.14100)) => D(1)

118. A1([*, 29)) AND A7([*, 1)) AND A8([52.46820, *)) => D(1)

119. A1([*, 29)) AND A7([*, 1)) AND A8([17.14100, 52.46820)) => D(1)

120. A1([53, *)) AND A7([*, 1)) AND A8([52.46820, *)) => D(1)

121. A1([36, 46)) AND A7([1, 901)) AND A8([52.46820, *)) => D(1)

122. A1([29, 36)) AND A7([6746, 20519)) AND A8([52.46820, *)) => D(0)

123. A1([53, *)) AND A7([*, 1)) AND A8([17.14100, 52.46820)) => D(0)

124. A1([46, 53)) AND A3([3, 5)) AND A7([6746, 20519)) => D(0)

125. A1([*, 29)) AND A3([3, 5)) AND A7([901, 6746)) => D(1)

126. A1([*, 29)) AND A3([5, *)) AND A7([20519, *)) => D(0)

127. A1([36, 46)) AND A3([5, *)) AND A7([6746, 20519)) => D(0)

128. A1([36, 46)) AND A3([2, 3)) AND A7([6746, 20519)) => D(1)

129. A1([53, *)) AND A3([2, 3)) AND A7([901, 6746)) => D(1)

130. A1([*, 29)) AND A3([2, 3)) AND A7([*, 1)) => D(1)

131. A1([36, 46)) AND A3([2, 3)) AND A7([20519, *)) => D(0)

132. A1([53, *)) AND A3([5, *)) AND A7([6746, 20519)) => D(0)

133. A1([46, 53)) AND A3([5, *)) AND A7([1, 901)) => D(0)

134. A1([53, *)) AND A3([3, 5)) AND A7([901, 6746)) => D(0)

135. A1([36, 46)) AND A3([3, 5)) AND A7([1, 901)) => D(0)

136. A1([46, 53)) AND A3([*, 2)) AND A7([6746, 20519)) => D(0)

137. A1([46, 53)) AND A6([*, 14)) AND A8([5.52027, 17.14100)) => D(0)

138. A1([46, 53)) AND A6([*, 14)) AND A8([17.14100, 52.46820)) => D(1)

139. A1([46, 53)) AND A6([*, 14)) AND A8([1.26161, 5.52027)) => D(0)

140. A1([46, 53)) AND A4(2) AND A6([*, 14)) AND A7([6746, 20519)) => D(0)

141. A1([46, 53)) AND A2(1) AND A6([*, 14)) AND A7([6746, 20519)) => D(0)

142. A1([36, 46)) AND A8([*, 1.26161)) => D(0)

143. A1([29, 36)) AND A8([17.14100, 52.46820)) => D(1)

144. A1([*, 29)) AND A8([1.26161, 5.52027)) => D(1)

145. A1([53, *)) AND A8([1.26161, 5.52027)) => D(0)

146. A1([29, 36)) AND A8([1.26161, 5.52027)) => D(1)

147. A1([46, 53)) AND A8([*, 1.26161)) => D(0)

148. A7([901, 6746)) AND A8([52.46820, *)) => D(1)

149. A7([20519, *)) AND A8([1.26161, 5.52027)) => D(0)

150. A7([*, 1)) AND A8([1.26161, 5.52027)) => D(1)

151. A7([1, 901)) AND A8([17.14100, 52.46820)) => D(1)

152. A7([20519, *)) AND A8([*, 1.26161)) => D(0)

153. A1([*, 29)) AND A4(2) AND A7([901, 6746)) => D(1)

154. A1([53, *)) AND A4(2) AND A7([6746, 20519)) => D(1)

155. A1([*, 29)) AND A4(2) AND A7([20519, *)) => D(1)

156. A1([36, 46)) AND A4(1) AND A7([6746, 20519)) => D(1)

157. A1([53, *)) AND A4(2) AND A8([52.46820, *)) => D(1)

158. A1([53, *)) AND A4(1) AND A8([*, 1.26161)) => D(1)

159. A1([*, 29)) AND A4(2) AND A8([5.52027, 17.14100)) => D(1)

160. A1([36, 46)) AND A4(1) AND A8([17.14100, 52.46820)) => D(1)

161. A1([46, 53)) AND A4(1) AND A8([5.52027, 17.14100)) => D(0)

162. A1([*, 29)) AND A4(2) AND A8([17.14100, 52.46820)) => D(1)

163. A1([*, 29)) AND A4(1) AND A8([17.14100, 52.46820)) => D(0)

164. A1([36, 46)) AND A4(2) AND A8([52.46820, *)) => D(1)

165. A3([5, *)) AND A4(2) AND A6([*, 14)) AND A7([901, 6746)) => D(1)

166. A3([2, 3)) AND A4(2) AND A6([*, 14)) AND A8([52.46820, *)) => D(1)

167. A3([5, *)) AND A4(2) AND A6([*, 14)) AND A8([*, 1.26161)) => D(1)

168. A3([2, 3)) AND A4(2) AND A6([*, 14)) AND A8([1.26161, 5.52027)) => D(1)

169. A3([5, *)) AND A4(2) AND A6([*, 14)) AND A8([17.14100, 52.46820)) => D(1)

170. A3([5, *)) AND A6([1382, 14728)) AND A7([6746, 20519)) => D(0)

171. A3([5, *)) AND A6([1382, 14728)) AND A7([20519, *)) => D(1)

172. A3([*, 2)) AND A6([*, 14)) AND A7([1, 901)) => D(0)

173. A3([*, 2)) AND A6([*, 14)) AND A7([*, 1)) => D(1)

174. A3([2, 3)) AND A6([1382, 14728)) AND A7([6746, 20519)) => D(1)

175. A3([2, 3)) AND A6([314, 1382)) AND A7([20519, *)) => D(1)

176. A3([5, *)) AND A6([*, 14)) AND A7([20519, *)) => D(0)

177. A3([2, 3)) AND A6([*, 14)) AND A7([20519, *)) => D(0)

178. A3([5, *)) AND A4(2) AND A7([6746, 20519)) => D(0)

179. A3([*, 2)) AND A4(1) => D(1)

180. A6([*, 14)) AND A7([6746, 20519)) AND A8([17.14100, 52.46820)) => D(1)

181. A6([*, 14)) AND A7([1, 901)) AND A8([1.26161, 5.52027)) => D(0)

182. A6([*, 14)) AND A7([6746, 20519)) AND A8([1.26161, 5.52027)) => D(0)

183. A6([*, 14)) AND A7([20519, *)) AND A8([5.52027, 17.14100)) => D(0)

184. A3([*, 2)) AND A6([*, 14)) AND A8([*, 1.26161)) => D(1)

185. A3([5, *)) AND A6([1382, 14728)) AND A8([5.52027, 17.14100)) => D(0)

186. A3([*, 2)) AND A6([*, 14)) AND A8([1.26161, 5.52027)) => D(0)

187. A3([2, 3)) AND A6([*, 14)) AND A8([*, 1.26161)) => D(1)

188. A4(1) AND A7([20519, *)) => D(0)

189. A4(4) AND A7([20519, *)) => D(0)

190. A4(1) AND A7([*, 1)) => D(0)

191. A2(1) AND A4(1) AND A8([5.52027, 17.14100)) => D(0)

192. A2(0) AND A4(2) AND A8([17.14100, 52.46820)) => D(1)

193. A2(1) AND A3([5, *)) AND A4(1) => D(0)

194. A2(0) AND A3([2, 3)) AND A4(2) => D(1)

195. A1([*, 29)) AND A3([5, *)) AND A4(1) => D(0)

196. A1([*, 29)) AND A3([2, 3)) AND A4(2) => D(1)

197. A1([*, 29)) AND A3([5, *)) AND A4(2) => D(1)

198. A1([53, *)) AND A3([5, *)) AND A4(1) => D(0)

199. A1([*, 29)) AND A2(0) AND A4(2) => D(1)

200. A1([46, 53)) AND A2(1) AND A4(1) => D(0)

201. A1([*, 29)) AND A2(0) AND A4(1) => D(0)

202. A1([*, 29)) AND A2(0) AND A3([2, 3)) => D(1)

203. A1([*, 29)) AND A2(0) AND A3([3, 5)) => D(1)

204. A1([46, 53)) AND A2(1) AND A3([5, *)) => D(0)

205. A1([*, 29)) AND A2(0) AND A3([5, *)) => D(0)

206. A1([53, *)) AND A2(0) AND A3([3, 5)) => D(0)

207. A1([46, 53)) AND A2(1) AND A3([*, 2)) => D(0)

208. A3([2, 3)) AND A4(2) AND A8([5.52027, 17.14100)) => D(1)

209. A3([5, *)) AND A4(2) AND A8([52.46820, *)) => D(1)

210. A2(0) AND A7([1, 901)) => D(1)

211. A2(0) AND A7([*, 1)) => D(1)

212. A1([46, 53)) AND A6([1382, 14728)) AND A7([6746, 20519)) => D(1)

213. A1([46, 53)) AND A6([*, 14)) AND A7([20519, *)) => D(0)

214. A1([36, 46)) AND A6([*, 14)) AND A7([20519, *)) => D(0)

215. A2(1) AND A3([3, 5)) AND A6([*, 14)) AND A8([*, 1.26161)) => D(0)

216. A2(0) AND A3([3, 5)) AND A6([*, 14)) AND A8([52.46820, *)) => D(0)

217. A2(1) AND A3([5, *)) AND A6([*, 14)) AND A8([17.14100, 52.46820)) => D(1)

218. A2(1) AND A3([3, 5)) AND A7([*, 1)) AND A8([*, 1.26161)) => D(0)

219. A1([29, 36)) AND A3([2, 3)) => D(1)

220. A1([29, 36)) AND A3([5, *)) => D(1)

221. A1([29, 36)) AND A3([*, 2)) => D(1)

222. A1([46, 53)) AND A2(1) AND A6([314, 1382)) => D(0)

223. A2(0) AND A6([*, 14)) AND A8([1.26161, 5.52027)) => D(0)

224. A2(1) AND A6([14, 314)) AND A8([1.26161, 5.52027)) => D(0)

225. A1([46, 53)) AND A3([*, 2)) AND A6([*, 14)) => D(0)

226. A1([36, 46)) AND A3([2, 3)) AND A7([*, 1)) AND A8([52.46820, *)) => D(0)

227. A1([46, 53)) AND A3([2, 3)) AND A7([6746, 20519)) AND A8([1.26161, 5.52027)) => D(1)

228. A1([36, 46)) AND A3([2, 3)) AND A4(2) AND A6([*, 14)) => D(1)

229. A3([3, 5)) AND A4(2) AND A6([14, 314)) => D(1)

230. A4(4) AND A8([1.26161, 5.52027)) => D(0)

231. A4(1) AND A8([52.46820, *)) => D(0)

232. A2(0) AND A4(4) => D(0)

233. A1([36, 46)) AND A4(4) => D(0)

234. A1([29, 36)) AND A4(1) => D(1)

235. A4(2) AND A6([14, 314)) AND A7([901, 6746)) => D(1)

236. A2(0) AND A3([5, *)) AND A4(2) AND A8([*, 1.26161)) => D(1)

237. A5(0) AND A8([1.26161, 5.52027)) => D(0)

238. A3([*, 2)) AND A5(0) => D(0)

239. A1([46, 53)) AND A5(0) => D(0)

240. A5(0) AND A7([6746, 20519)) => D(0)

241. A2(1) AND A5(0) => D(0)

242. A4(5) => D(0)

243. A2(0) AND A6([*, 14)) AND A7([901, 6746)) => D(0)

244. A2(1) AND A6([*, 14)) AND A7([20519, *)) => D(0)

245. A2(0) AND A4(1) AND A7([6746, 20519)) => D(1)

246. A3([5, *)) AND A4(1) AND A5(1) AND A8([*, 1.26161)) => D(0)]

247. A1([36, 46)) AND A2(1) AND A3([2, 3)) AND A8([52.46820, *)) => D(0)

248. A1([36, 46)) AND A2(1) AND A3([3, 5)) AND A6([*, 14)) => D(0)

249. A1([36, 46)) AND A2(1) AND A3([5, *)) AND A6([*, 14)) => D(1)

250. A3([3, 5)) AND A6([*, 14)) AND A7([1, 901)) AND A8([*, 1.26161)) => D(0)

251. A1([53, *)) AND A4(2) AND A6([*, 14)) AND A8([*, 1.26161)) => D(0)

252. A1([53, *)) AND A3([*, 2)) AND A4(2) AND A7([901, 6746)) => D(0)

253. A1([36, 46)) AND A2(1) AND A6([*, 14)) AND A8([1.26161, 5.52027)) => D(1)

254. A1([46, 53)) AND A5(1) AND A7([6746, 20519)) AND A8([1.26161, 5.52027)) => D(1)

255. A1([53, *)) AND A2(1) AND A3([*, 2)) AND A4(2) => D(0)

256. A2(0) AND A4(2) AND A6([*, 14)) AND A7([6746, 20519)) => D(0)
257. A1([53, *)) AND A2(1) => D(1)
258. A2(1) AND A3([*, 2)) AND A4(2) => D(1)

UMP

LIST OF PUBLICATIONS

PROCEEDINGS:

1. **Syafiqah, N.,** Mokhairi, M., Awang, M. K., A.Rahman, M. N., & Mustafa, M. D. (2015). The study of Feature Reduction Based on Rough Set Theory for Customer Churn Classification. *In UniSZA Research Conference*.MJAS.

2. **Syafiqah, N.,** Mokhairi, M., Awang, M. K., A.Rahman, M. N., & Mustafa, M. D. (2015). Predictive Modeling for Telco Customer Churn Using Rough Set Theory. *In International Conference on Telecommunications, Electronic and Computer Engineering (ICTEC2015).*

JOURNALS:

1. **Syafiqah, N.,** Mokhairi, M., Awang, M. K., A.Rahman, M. N., & Mustafa, M. D. (2015). The study of Feature Reduction Based on Rough Set Theory for Customer Churn Classification.*Malaysian Journal of Analytical Sciences.*(In press)

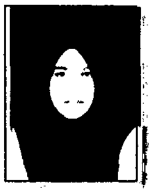
2. **Syafiqah, N.,** Mokhairi, M., Awang, M. K., A.Rahman, M. N., & Mustafa, M. D. (2015). Predictive Modeling for Telco Customer Churn Using Rough Set Theory. *ARPN Journal of Engineering and Applied Sciences.*

3. **Syafiqah, N.,** Mokhairi, M., Awang, M. K., A.Rahman, M. N., & Mustafa, M. D. (2015). Feature Selections and Classification Model for Customer Churn. *Journal of Theoretical and Applied Information Technology,* 75(3), 356-365.

BOOK CHAPTER:

1. **Syafiqah, N.,** Mokhairi, M., Awang, M. K., A.Rahman, M. N., & Mustafa, M. D. (2016). Meningkatkan Nilai Projek Ilmiah: Pengintegrasian Dengan Kaedah Sains Komputer Lanjutan, *Aplikasi Teori Set Kasar Terhadap Model Pengkelasan*, Universiti Sultan Zainal Abidin. (Accepted)

CANDIDATES BIODATA



Nur Syafiqah binti Mohd Nafis was born May 16, 1989 in Terengganu, Malaysia. She attended Mara Junior Science College Besut. She performed well during Malaysian Certificate Exam and was selected to complete her foundation at the Pahang Matriculation College before graduating in 2008.

In 2012, Syafiqah graduate from Universiti Sultan Zainal Abidin in Bachelor of Science in Computer Science with Honours and awarded *'Pingat Ibnu Sina'* for her excellence performance during undergraduate studies. Later, she worked as an Application Developer at CSC Malaysia Sdn.Bhd until Febuary 2014 she was then invited into the graduate program in UniSZA. Her research includes data mining and classification.

During her graduate studies, Syafiqah is a scholar of Skim Latihan Akademik Bumiputera (SLAB) for Universiti Malaysia Pahang and was selected as Fellow Assistant of The Residential Collages of Tembila Campus. After finishing her Master programme, she planned to pursue PhD and Syafiqah planned to continue her work on data mining and classification research using medical data.