

10.

215

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
FACULTY OF INFORMATICS
DEPARTMENT OF INFORMATION SCIENCE

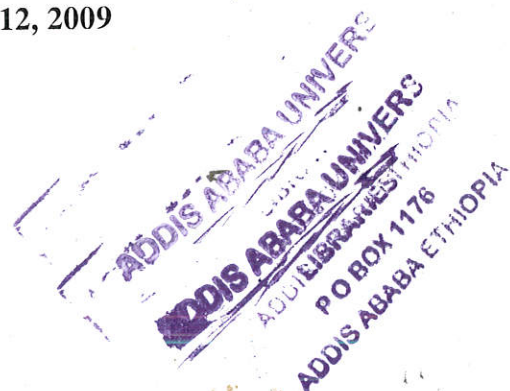
**APPLICABILITY OF DATA MINING TECHNIQUES TO CUSTOMER RELATIONSHIP
MANAGEMENT (CRM):**
**The case of Ethiopian Telecommunications Corporation's (ETC) Code Division Multiple
Access (CDMA) Telephone Service**

**A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDIS
ABABA UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENT
FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION SCIENCE**

By

Melaku Girma

January 12, 2009



DEDICATION

This research is dedicated to the “Almighty God” who is always there for me in all the hard times and challenges of my life.

ACKNOWLEDGMENT

There are many individuals that have really contributed directly or indirectly for the successful accomplishment of this research. Although nothing can actually spells out and expresses their greatness and enormous cooperation, all of them deserve special appreciations and acknowledgements for being with me in all those challenging times of the study.

First and foremost I am grateful to the almighty God for his kindness and blessings of me with the courage and endurance to successfully complete the research. Next to this, I would like to express my gratitude and heart felt thanks to my advisor, Dr. Million Meshesha, for his unreserved and diligent advising. I really appreciate his genuine, enlightened, and constructive comments and critical readings of the study. I admire his patience and in-depth knowledge.

I would like to extend my appreciation to my families for being with me in all the challenging times while undertaking the research. Particularly my sister, Liyuwork Girma, is always there for me when I am in need of something. This research could not haven been completed without the help and kindness of my colleague, Ato Ketema Dillebo. He has supported me with the necessary materials to successfully complete the study.

My special thanks also go to staffs of ETC for providing me with pertinent data and other necessary information for the research.

Last but not least I remain in debt to my colleagues, Ato Meshesha Legesse and Ato Fikiru Amanuel, at Unity University for their encouragement and invaluable advice while undertaking the study.

LIST OF ABBREVIATED WORDS

ANN:	Artificial Neural Network
AU:	Africans Union
CAAZ:	Central Addis Ababa Zone
CCB:	Customer Caring and Billing
CCSD:	Corporate Customer Support Division
CDMA:	Code Division Multiple Access
CDR:	Call Detail Record
CEO:	Chief Executive Officer
CIO:	Chief Information Officer
CMO:	Chief Marketing Officer
CRISP:	CRoss Industry Standard Process
CRISP-DM:	CRoss Industry Standard Process of Data Mining
CRM:	Customer Relationship Management
DCEO:	Deputy Chief Executive Officer
EAAZ:	East Addis Ababa Zone
ETC:	Ethiopian Telecommunications Corporation
EU:	Europeans Union
GSM:	Global System for Mobile
IT:	Information Technology
KDD:	Knowledge Discovery in Databases
MCSD:	Market Customer Support Division
NGO:	Non-Government Organizations
NPTSD:	Network Planning and Technical Support Division
OLAP:	Online Analytical Processing
OLTP:	Online Transaction Processing
OMC:	Operating and Maintenance Center
RDBMS:	Relational Database Management Systems
SBU:	Strategic Business Unit
SQL:	Structured Query Language
SVM:	Support Vector Machine
UN:	United Nations
WANW:	Wireless Access Network

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGMENT	iii
LIST OF ABBREVIATED WORDS	iv
TABLE OF CONTENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF APPENDICES	x
ABSTRACT	xi
CHAPTER ONE	1
INTRODUCTION	1
1.1. Background	1
1.1.1. CRM Overview	2
1.1.2. Data Mining Overview	5
1.1.3. ETC Overview	8
1.1.3.1. Organizational structure Description	9
1.1.3.2. CDMA Telephone Service	14
1.1.3.3. The "USHACOM" Database	16
1.2. Statement of the problem	17
1.3. Justification of the study	21
1.4. Objectives of the Study	22
1.4.1. General objective	22
1.4.2. Specific Objectives	22
1.5. Research Methodology	23
1.5.1. Review of related literature	23
1.5.2. Fact Finding	24
1.5.3. Data Identification, collection, and preparation	25
1.5.4. Training, Model Building and Testing	25
1.5.5. Tools	26
1.6. Scope and Limitation of the Research	26
1.7. Application of the Results	27
1.8. Thesis organization	27
CHAPTER TWO	29
LITERATURE REVIEW	29
2.1 DATA MINING CONCEPTS	29
2.1.1 Overview	29
2.1.2. Data Mining and Knowledge Discovery in databases	32
2.1.2.1 Business understanding	33
2.1.2.2 Data understanding	34
2.1.2.3 Data preparation	34
2.1.2.4 Modeling	36
2.1.2.5 Evaluation	36
2.1.2.6 Deployment	37
2.1.3 Data mining and related Fields	37
2.1.3.1 Data Mining vs. Databases	37
2.1.3.2 Data Warehousing vs. Data Mining	38

2.1.3.3 OLAP vs. Data Mining.....	39
2.1.3.4 Data Mining, Machine Learning and Statistics	39
2.1.4 Data Mining Tasks	40
2.1.4.1 Data description and summarization	41
2.1.4.2 Segmentation	42
2.1.4.3 Concept descriptions	42
2.1.4.4 Classification	43
2.1.4.5 Prediction.....	43
2.1.4.6 Dependency analysis	44
2.1.5 Application Areas of Data Mining	44
2.1.5.1 Application of Data Mining in Telecommunications	45
2.2 CRM AND SEGMENTATION CONCEPTS	47
2.2.1 Customer Relationship Management	47
2.2.1.1 Overview	47
2.2.1.2 Reasons for adopting CRM	50
2.2.1.3 Effective CRM.....	51
2.2.2 Customer Segmentation.....	51
2.2.2.1 Bases of Segmentation	54
2.2.3 Customer Segmentation at ETC	56
2.3. DATA MINING IN CRM AND CUSTOMER SEGMENTATION	57
2.4 DATA MINING METHODS FOR CUSTOMER SEGMENTATION	58
2.4.1 Overview	58
2.4.2 Clustering Techniques	58
2.4.2.1 The K-Means Method.....	61
2.4.3 Decision Trees	62
2.4.3.1 Tree Induction.....	64
2.4.3.2 Different Decision algorithms	65
2.4.4 Artificial Neural Networks (ANNs)	67
2.4.4.1 A multilayer feed-forward neural network.....	69
2.4.4.2 Defining a network topology.....	70
2.4.4.3 Backpropagation.....	71
CHAPTER THREE	77
DATA PREPARATION	77
3.1. Overview	77
3.2. Data understanding	77
3.3. Data processing	79
3.3.1. Data Cleaning	79
3.3.2. Data Integration and Data Transformation	80
3.3.3. Data Reduction Methods	81
CHAPTER FOUR	83
EXPERIMENTATION	83
4.1. Overview	83
4.2. Business Understanding	84
4.3. Data Understanding	85
4.3.1. Data Selection Processes	85
4.3.2. Description of data collected	86
4.4. Data preprocessing	86

LIST OF APPENDICES

Appendix 1 The Original Collected Sample Data	133
Appendix 2. The Weka 3.5.8 Interface with the Dataset Opened to Start the First Clustering Run	135
Appendix 3. Sample of the Decision Tree Generated With 10-Fold Cross-Validation Technique	136
Appendix 4. Summary of the ANN's multilayerperceptron algorithm classification models at different values of hidden layer nodes and learning rate.....	137
Appendix 5. The decision tree and the rules derived from the tree.....	139

ABSTRACT

In this research the applicability of clustering and classification techniques of data mining on CRM the case of CDMA telephone service of ETC have been explored within the framework of CRISP-DM model. The CDMA CDR data along with billing information and the customers' profiles are collected, cleansed, transformed and integrated for experimenting with the clustering models. The final dataset consists of 10, 090 records on which different clustering models at K values of 6, 5, and 4 with different seed values have been experimented and evaluated against their performances. Hence, the cluster model at K value of 6 has shown a better performance. Consequently, its output is used as an input for the decision tree and ANN classification models.

First the different classification models with J48 decision tree algorithm are experimented with the 10-fold cross-validation, and splitting the dataset into 80 % training and 20 % testing, techniques by setting the cluster index formed by the cluster model as dependent variable and the rest as independent variables. Among these models a model that showed a classification accuracy of 98.97% is selected. Similarly, different classification models of multilayerperceptron ANN algorithm are carried out by changing its hidden layer number of nodes and learning rate parameters' value. A model with a classification accuracy of 98.62 % is chosen. Finally a comparison of decision tree and ANN models in terms of the overall classification accuracy, accuracy in classifying high value customers, and accuracy in classifying low value customers have been undertaken. Hence, the decision tree model has excelled in these evaluation parameters and therefore selected as the best classifier for CRM applications.

The result of this research is really encouraging as very high classification accuracy has been obtained. Besides, high precision and recall in classifying high and low value customers correctly have been achieved.

CHAPTER ONE

INTRODUCTION

1.1. Background

Information is the key resource, in parallel with land, human, and capital, to the success of any organization. It has become an invaluable input to any sound and acceptable strategic decision-making processes in the twenty-first century. Many individuals, government, business organizations, managers and, generally, individuals involved in decision-making responsibilities, increasingly depend on information to successfully accomplish and discharge their duties and responsibilities. Information has been tightly integrated with the day-to-day life of almost all individuals. For example, human beings are uncertain of what they are doing without the right information. It enables, particularly business organizations, to take proactive measures in dealing with uncertainties of the future. Generally, information becomes an integral part of every walks of human life.

Nevertheless, not all individuals and organizations in the world are using information in an equitable way. Based on the usage of information, the world nowadays is divided into agrarian, industrial, and information societies. Among them, information societies use information intensively in their day-to-day activities. The majority of these societies are engaged in the collection, processing, storage, and dissemination of information.

Information, however, entails substantial amount of money for its production. It has sources that pass through several different processes. Capturing the relevant data for further processing is a

costly step. Besides, processing information embraces application of sophisticated and very expensive Information Technologies (IT). Due to these and other costs, the information is buried deeply under a massive data. Here, it has to be noticed that data is the ultimate source of information.

Accordingly, the question, "From where do organizations gather data?" calls for the various sources of data. Today, in many business areas, detailed customer interaction data is abundant. Organizations uphold data about customer purchasing behavior, returns, complaints, wishes, CDR (in the case of telecommunications) and more. Particularly, Telecommunications gather data from various sources such as the customer CDR, the billing history of customers, the network fault alarms, and the network status reports generated automatically from the various interconnected switches. Although these voluminous data are kept in business organizations' legacy and current databases, it has been quite difficult to analyze them manually; thereby it is difficult to understand what information and/or knowledge are deeply hidden in them. The reason for this is that the technology for generating, capturing, and storing data has far outpaced than the human capacity to understand, analyze, and exploit it for maximum impact.

1.1.1. CRM Overview

The focus of many service-providing industries has been towards customer orientation. While competition is increasing, there is a need to better understand customers, and to quickly respond to their individual needs and wants. Thus, the objective of the customer-centered industry becomes increasingly the share of wallet for each individual customer and costs by focusing on more targeted promotions.

Customers are the life blood of any successful venture in the face of today's fierce competition both from the local and international markets. Accordingly, organizations, particularly service-oriented ones, have shifted towards a long-lasting customer relationship to gain a competitive advantage over their rivalries. These organizations exert their maximum effort to retain their existing customers as well as attracting new potential ones.

Thus, Customer Relationship Management (CRM) is the only way for businesses to create a prolonged tie with their customers. CRM, according to Bigus [4], is a process that manages the interaction between a company and its customers. The purpose of CRM initiatives is to allow customers to be treated individually during the marketing process. This differentiated marketing improves marketing results by making each marketing touch more appropriate and effective than is possible under any one-size-fits-all marketing scheme. To this end, the products and services of the organization are tailored to the needs and wants of individuals. The idea here is that an organization cannot go after the whole market with the same strategy, rather identify a niche in it to effectively and efficiently address the products and services it intends to offer. This change in focus from broad market segments to individual customer requires changes through the enterprise, but nowhere more than in CRM, marketing, and sales [4].

CRM is an integration of technologies and business processes used to satisfy the needs of a customer during any given interaction. More specifically, it involves acquisition, analysis and use of knowledge about customers in order to sell more goods or services and to do it more efficiently [5]. To this end, the customers' data alone do not have any implication to the business unless it is backed by the appropriate tools and technologies. These tools and IT are integrated to uncover relevant knowledge from the sheer volume of customers' data.

According to Srivastava [40], CRM functions include *marketing automation* (e.g., campaign management, cross- and up-sell, customer segmentation, customer retention), *sales force automation* (e.g., contact management, lead generation, sales analytics, generation of quotes, product configuration), and *contact center management* (e.g., call management, integration of multiple contact channels, problem escalation and resolution, metrics and monitoring, logging interactions and auditing), among others.

On the other hand, Kotler [30] states that, CRM principally revolves around marketing and begins with a deep analysis of customer behavior. It uses IT to gather data, which can then be used to develop information required to create a more personal interaction with customer. As stated in the aforementioned discussion, the focus of the business nowadays shifted towards individual customer service usage behaviors and trends. Particularly, industries like telecommunications whose activities mainly engage in the highly sensitive and customer-centric business should pay considerable attention to their existing and prospective customers. This in turn, calls for a strategy amenable to individual needs and wants.

Marketing is the process of planning and executing the conception, pricing, promotion and distribution of ideas, goods and services to create exchanges that satisfy individual and organizational objective (s). Nowadays, customers that have real value to a company are the center of marketing strategies. Accordingly, businesses have found it essential to acquire new customers as well as to retain those that have high value. Hence, one of the major focuses in marketing is customer segmentation [38].

Customer segmentation refers to an attempt to categorize customers into several homogeneous groups for the purpose of designing and executing a certain business strategy. Customers belonging to the group share similar characteristics. This in turn helps marketers to design similar promotion

campaigns for customers that fall in a specific segment. However, segmentation requires the collection, organization and analysis of customer data [6] [38]. Customer segmentation also makes the customer management easy.

The telecommunication industry in particular, should deal with individual level market segments in order to increase sales of its various services and gain insight into the individual needs and wants. To this end, the telecom operator cannot go after the whole market for its various services with a single generic strategy. Yet, customer segmentation is at the heart of its effective and efficient marketing.

Edelstein [15], outlined the following as CRM phases.

- Acquiring customers
- Increasing the value of the customer
- Retaining good customers

In all of the three phases, data mining plays an important role. For example, a certain telecommunication company can analyze its customers' database to identify their patterns and trends so as to target its particular service to a certain market group (s). It has become evident that data mining has a lot to contribute from the acquisition to retention of customers to business organizations achievement.

1.1.2. Data Mining Overview

Telecommunication companies maintain a great deal of data about their customers. In addition to the general customer data that most businesses collect, telecommunication companies also store CDR, which precisely describe the calling behavior of each customer. This information can be used to profile the customers and these profiles can then be used for marketing and/or forecasting purposes [47]. From this, it

is easy to infer that data mining has a great role to play, particularly in telecommunication industries, to uncover knowledge that is deeply buried under these massive data.

The foregoing discussion has clearly depicted that organizations collect data from different sources. Mainly in their interactions with their customers to exchange goods and services, values, and many others have led to the sheer volume of information to be retained in the organization legacy and operational databases. For example, customers profile, their order for products or services, and/or any interaction with the business are among the major sources of data. This in turn has created information overload in the sides of organizations. It is impossible to manually trace and see into the relationships within various data without an automated and/or semi automated tools. Especially, in the telecommunication industry where there are millions and billions of data available about:

- i. Call detail data, which describes the calls that traverse the telecommunication networks,
- ii. Network data, which describes the state of the hardware and software components in the network, and
- iii. Customer data, which describes details of telecommunication customers.

These data are hourly generated and kept in different heterogeneous database and flat files out of which finding relevant patterns and knowledge manually is unthinkable. The amount of data is so large that manual analysis of the data is difficult, if not impossible [47]. However, recent breakthrough, especially, in the area of data mining and artificial intelligence, are very promising in handling and extracting valuable information for effective planning and decision making purpose [48].

Hence, data mining technology, which focuses on identifying interesting patterns and developing predictive models from data, has the greatest potential enabling business to control data resources

Marketing and Customer Support Division performs marketing functions to the Fixed Line SBU. It identifies and reports to problems that had been observed in sales; reports summarized daily, weekly, monthly, quarterly, semiannually and annually fixed and CDMA services sales; it translates and executes the marketing strategies articulated and devised by the CMO, which is under the direct supervision of DCEO Operations, and performs other marketing related functions as required.

1.1.3.2. CDMA Telephone Service

CDMA is a form of spread-spectrum, a family of digital communication techniques that has been used in military applications for many years. The core principle of spread spectrum is the use of noise-like carrier waves, and, as the name implies, bandwidth is much wider than that required for simple point-to-point communication at the same data rate. Originally, there were two motivations: to resist enemy efforts to jam the communications (anti-jam, or AJ), and to hide the fact that communication was even taking place, sometimes called low probability of intercept (LPI) [43].

CDMA is a recent Wireless telephone technology that has been adopted by different telecom companies. Commercially introduced in 1995, CDMA quickly became one of the world's fastest-growing wireless technologies. In 1999, the International Telecommunications Union selected CDMA as the industry standard for new "third-generation" (3G) wireless systems. Many leading wireless carriers are now building or upgrading to 3G CDMA networks in order to provide more capacity for voice traffic, along with high-speed data capabilities [43].

As per the interviews and discussions with the senior management of ETC, the CDMA telephone service has been introduced by ETC in 2004 with 50, 000 exchange capacity. It was initially planted in the EAAZ of the corporation. The corporation has introduced with the intention that it could be used as a substitute for the old copper wire fixed telephone network. As per the discussion

1.3. Justification of the study

CRM is at the heart of keeping existing customers of an organization satisfied while attracting and retaining new customers. Particularly, organizations that provide different type of services should identify the niche market for its various services thereby flourish a good CRM in order to get a competitive advantage over the competitors; Vis-à-vis ETC should do the same. But segmentation of customers requires a thorough analysis of customers' databases in order to identify valuable information that has deeply hidden in these massive databases. *Thus, the theme of the research is to investigate the applicability of data mining to customers' segmentation and classification.* The result is useful to create an effective CRM in ETC.

On the other hand, ETC has already finished expanding the CDMA exchange capacity from 50,000 to 2.4 million. However, to market this much capacity requires the identification and understanding of the pervious customers who have subscribed the same. Vis-à-vis this, data mining is useful to discover trends and patterns of customers from the customers' databases. Consequently, the information mined can be useful for strategic market formulation and decision making, and enables to flourish sound CRM at ETC. Again the same information enables the corporation to understand the behavior of customers thereby increase its sales while maximizing its profit margin.

Fekadu [21] has done similar research on the post-paid mobile CRM. He used the K-means clustering algorithm for clustering and decision tree algorithm for classification. He found a 94.78 % accuracy of prediction. He stated that similar researches could be conducted on the different services of the corporation. He further recommended that the application of artificial neural network may give better accuracy result.

Similarly, Deneke [14] has undertaken a research on CRM in Ethiopian Airlines. He used K-means and decision tree algorithms for clustering and classification respectively. He recommended that clustering and classification functions of data mining are applicable to CRM. Deneke further pointed out that the application of artificial neural network may result in better classification accuracy.

This research complements with the research work mentioned above. But it differs in the area of application, variables used for model building, the theme it contains, and data mining methods.

1.4. Objectives of the Study

1.4.1. General objective

The general objective of this research is to support CRM activities at ETC by employing appropriate data mining techniques (clustering and classification) on the customers' database to uncover patterns/knowledge so as to identify different segments of them. This objective is pursued to test whether data mining techniques can be applied in the behavioral customers segmentation in the context of ETC. On the other hand the result of successfully achieving this objective enables the corporation to identify, create, and maintain a good relationship with customers. In addition to these, this study can be a source of methodological approach for future researches.

1.4.2. Specific Objectives

The specific objectives of the study are the following:

- Identify the type of customers' data residing in the customers database ("USHACOM").
- Collect relevant data for the research from the "USHACOM" database.

- Select the data mining clustering and classification algorithms to be used based on the type of data and objectives of the present study.
- Prepare the data for model building through cleaning, extracting and transforming the data into a format required for the clustering and classification algorithms.
- Apply clustering algorithm to test behavioral segmentation on the CDMA subscribers.
- Apply the classification algorithms to build and train the classification models that classify new instances of customer into one of the class labels identified by the clustering algorithm.
- Test and compare the resulting performances of the clustering and classification models and recommend the overall best results of the clustering and classification models.
- Suggest further research areas in the future

1.5. Research Methodology

Methodology defines the steps and procedures one follows to achieve the objectives and research questions of a certain research. It is a road-map that shows the direction a research follows to attain the end. To this end, in order to achieve the objectives of the present research, the researcher has adopted the CRISP-DM model. This is because the model has been widely applied for data mining studies. Besides, it is open-source and industry standard data mining processes model. Accordingly, the present study has followed the following methodologies in order to develop good customer segmentation and classification models.

1.5.1. Review of related literature

Thorough review of literatures in the area of data mining and KDD has been conducted. Particularly, clustering and classification tasks of data mining, and their application in the area of CRM are deeply analyzed and assessed. Various books, journals, magazines, articles, and papers from the World Wide Web have been consulted in order to get clear insight into the concepts and

techniques of data mining, KDD, and customers' segmentation in the framework of CRM. These sources have given firm theoretical and practical background to the current undertaking.

1.5.2. Fact Finding

To clearly identify, understand, and analyze the business problems, the primary (observation and interview), and secondary (database analysis) data collection methods are employed. Interview is employed to define marketing concepts and features selection with the domain experts while observation is conducted to understand some complex business processes. Further, databases such as CDR along with billing information and customers' profiles are consulted to gather the pertinent data for the present research.

Thus, the researcher has been conducting a thorough survey (using an in-depth interviews and observations) at different offices of ETC to get a thorough understanding of the business\domain. The researcher has been conducting the survey in the following offices of the corporation.

- CCB: This office manages the CDR and customer profiles data in the "USHACOM" database. The data for this researcher has been collected from this particular office.
- CCSD: This division manages the corporate customers as stated in section 1.1.2.1.
- The fixed Line SBU, which is located in the CAAZ, is one business unit per se and manages the fixed and CDMA telephone services of the corporation.
- CDMA telephone administration office, which is located in EAAZ of the corporation. This office is entrusted to manage, control, maintain, and run the CDMA telephone service of the corporation. And
- The Marketing Division of the corporation: is responsible to undertake activities related to marketing. All the market strategies are designed by this particular office.

1.5.3. Data Identification, collection, and preparation

The CDMA CDR along with the billing data, and the customers' profiles data are the primary data used for analysis in this research work. This data has been collected from the corporation database, "USHACOM." The target populations are all customers of CDMA service of ETC.

After all relevant attributes are derived from the CDMA CDR, the total number of individual call records are summarized into 23, 172 records that describe the calling behavior of each customer. This figure represents the total population of CDMA customers. However the stratified sampling technique has been applied to reduce the data size almost by half. Hence, the final dataset used for model building consisted of 10,090 records.

The data has been cleansed, edited, integrated, and normalized before any model building commencement. Accordingly, it has been prepared in such a way that it becomes appropriate for the clustering and classification algorithms.

1.5.4. Training, Model Building and Testing

The K-means clustering algorithm has been experimented to segment the customers' data. The clusters at K values of 6, 5, and 4 have been carried out. The models of the segments at these different values of K have been evaluated together with the domain experts and finally the best cluster model, which is at K value of 6, is selected as an input for the J48 decision tree and multilayerperceptron ANN algorithms models building.

Next, different classification models of the J48 decision tree have been experimented through changing some of its parameters. First, the algorithm is trained and validated with the 10-fold cross validation learning technique, a technique that uses 90 % of the data set for training and the remaining 10 % for the model for testing. In addition to this, other models have been

experimented with 80 % (8072 records) of the dataset for training and the rest 20 % (2018 records) for testing. The overall best model of J48 decision tree algorithm is selected.

Similarly models of the multilayerperceptron ANN algorithm have been built through changing some of its parameters such as learning rate, and number of hidden nodes. Here also the best performed model is selected. Finally, the best model of this algorithm has been compared with that of the decision tree model.

1.5.5. Tools

In this research, SQL Server 2000, Oracle 9i, and MS-Access 2003 RDBMSs, and MS-Excel 2003 are intensively used for the data preparation and preprocessing. For clustering and classification experimentations of the present research, weka 3.5.8 has been used. It is selected mainly due to the following reasons:

- It implements all algorithms, K-means for clustering, and J48 algorithm of decision tree and feed forward back propagation algorithm of ANN for classification, which are required for the present research.
- It has an easy interface to work with
- It is open source software and is freely available
- Many data mining researches and projects have been implemented with it

1.6. Scope and Limitation of the Research

The scope of this research is limited to development of clustering and classification models building of ETC's CDMA telephone customers. The research has been tried to explore the applicability of the clustering and classification function of data mining to behavioral customer segmentation, which is the main function of CRM. The clustering is employed to identify class

labels, while classification has been used to predict the class label of new instances of the data. Besides, the best model for each of the clustering and classification techniques is chosen.

Although there are other services of the corporation such as fixed telephone line, mobile (postpaid and prepaid), Internet (broadband and dialup) and other value added services, the research limits itself to CDMA service due to the availability of CDR, which is not readily available for some of these services.

The classification model prototype could not be implemented mainly due to time constraints.

1.7. Application of the Results

The result of this research can be used by ETC to identify trends and patterns of CDMA customers thereby tailoring its promotion and sales strategy based on the particular cluster of customers that will be identified in the course of the study. It also helps the CCSD that is solely established to identify, maintain good relationships, and deliver all the necessary facilities and special services to both the existing and prospective corporate customers, to easily undertake its duties and responsibilities. Besides, it will initiate other researchers to further undertake research on similar problem domain.

1.8. Thesis organization

This study is organized into five chapters. The first chapter comprises of the general background, overview of ETC, statements of the problem, objectives, scope, and application of the result of this research. The second chapter lays the foundation for other parts of this research. It thoroughly discusses important literatures in the area of data mining and customer segmentation in the framework of CRM, and existing CRM practice in ETC. This chapter also describes both the

conceptual and empirical researches, which are related to this research, in greater depth. The data collection and preparation for the final experimentation/models building are discussed with greater depth in chapter three.

The research questions are answered and its objectives are achieved in the experimentation part, which is chapter Four. Chapter five builds up from the experimentation part of the research. It provides the conclusion and recommendations based on the findings of the research. The conclusion section of this chapter summarizes the major points of the experimentation and concludes with the identification of problems, which are beyond the present research. The recommendations are forwarded for further research areas on the ground of the research results.

CHAPTER TWO

LITERATURE REVIEW

2.1 DATA MINING CONCEPTS

2.1.1 Overview

As clearly stated in the background part of this research, information is the life-blood of contemporary organizations, particularly those that operate businesses in a global-basis. These organizations use it to maximize their market share and gain a competitive-advantage. In addition, today's fierce competition among business organizations calls for information in the right time, at the right place and to the right person. Information is the corner-stone of any development and successful venture, particularly in the developed countries. Many renowned scholars in the field of information and communications technologies defined the term information from different perspectives. In the nutshell, it has been defined as power, energy, something useful for decision-making, the back-bone of the developed countries, information societies', economy, and others related to these. Witten, Bentley, and Dittman [50] define information as data that has been refined and organized by processing and purposeful intelligence.

Information is basically extracted from the huge size of data. Now-a-days, we are generating huge amount of data from different sources, which in turn has made data processing as a costly effort. Similarly, Han & Kamber [23] have clearly stated that our capabilities of both generating and collecting data have been increasing in the last several decades contributing factors include the wide spread use of bar codes for most commercial products, the computerization of many businesses, scientific and governmental transactions, and advances in data collection tools ranging

from scanned text and image platforms to satellite remote sensing systems. In addition, popular use of the World Wide Web as global information system has flooded us with a tremendous amount of data and information. This explosive growth in stored data has generated an urgent need for new techniques and automated tools that can assist us in transforming the vast amount of data into useful information and knowledge.

Han & Kamber [23], further explained that the fast-growing tremendous amount of data, collected and stored in large and numerous databases, has far exceeded our human ability for comprehension without powerful tools. Hence, to generate useful information and knowledge from a large store of databases, data mining tools and techniques play a pivotal role. Similarly Fayyad, Piatetsky-Shapiro and Smyth [20] have explained that manual analysis and extraction of information and knowledge from today's unprecedented growing size of data is almost impossible and always inefficient. Thus, there is an urgent need of a new generation of computational theories and tools to assist humans in extracting useful information/knowledge from the rapidly growing volumes of digital data. It is in here that the data mining tools and techniques come into play.

On the other hand, databases' sizes have significantly increased into terabytes bytes of data. Within these masses of data lies hidden information of strategic importance. To uncover these hidden information and knowledge data mining, which is being used both to increase revenues and to reduce costs, is at the heart. The potential returns are enormous. Innovative organizations worldwide are already using data mining to locate and appeal to higher-value customers, to reconfigure their product offerings to increase sales, and to minimize losses due to error or fraud [45].

Srivastava [40] has elucidated the contribution of data mining in identifying interesting patterns. Data Mining has enjoyed great popularity in recent years, with advances in both research and commercialization. The first generation of data mining research and development has yielded

several commercially available systems, both stand-alone and integrated with database systems; produced scalable versions of algorithms for many classical data mining problems; and introduced novel pattern discovery problems.

Many renowned scholars in the field have come up with numerous definition of data mining. Among them, those that are related to the research are mentioned and thoroughly discussed.

The term data mining is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. The goal of data mining is to allow the corporation to improve its marketing, sales, and customer support operations through a better understanding of its customers. This is customer relationship management oriented definition. In the course of dealing with customers and their tests and preferences, a business can unlock the potential of data mining to improve sales and revenue while keeping them satisfied [3].

The task of discovering interesting patterns from large amounts of data where the data can be stored in databases, data warehouses, or other information repositories [23]. The data stored in legacy databases and data warehouses do not give the patterns and deeply hidden knowledge in it per se. Rather it must be further analyzed to uncover and investigate relevant information with the help of data mining.

Data mining is the extraction of implicit, previously unknown, and potentially useful information from data [49]. Had the data stored in the databases revealed the interesting patterns per se, data mining would not have been invented. Data in themselves do not carry any meaning unless they are further manipulated and analyzed for a certain purpose. Hence, data mining tools and techniques make the implicit and previous unknown information explicit and known.

2.1.2. Data Mining and Knowledge Discovery in databases

The phrase knowledge discovery in databases was coined at the first KDD workshop in 1989 (Piatetsky-Shapiro (cited in Fayyad, Piatetsky-Shapiro and Smyth, [20]) to emphasize that knowledge is the end product of a data-driven discovery. It has been popularized in the AI and machine-learning fields.

Different authors take data mining and KDD as one and the same. However, there is a significant difference between the two. Data mining is one step in the overall knowledge discovery processes. Nevertheless Fayyad, Piatetsky-Shapiro and Smyth [20] elucidate KDD as the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process.

The term data mining and KDD are used interchangeably in this research. This is because the KDD elucidates the overall phases of pattern extraction from large databases. Although data mining is one process in the course of KDD, it provides all the techniques and tools to investigate meaningful patterns and knowledge from the database. Besides, data mining requires a data that is cleansed, integrated, and transformed before the mining process begins, which are included in the overall KDD processes.

The data mining phases embraces defining business problem, building marketing database, exploring the data, preparing data for modeling, building model, evaluating model, and deploying model results [46]. However, CRISP-DM defines the phases as business understanding, data understanding, data preparation, modeling, evaluation, and deployment (see **Figure 2.1.**).

The Cross-Industry Standard Process for Data Mining (CRISP-DM) was developed in 1996 by analysts representing DaimlerChrysler, SPSS, and NCR. CRISP provides a nonproprietary and

freely available standard process for fitting data mining into the general problem-solving strategy of a business or research unit [35].

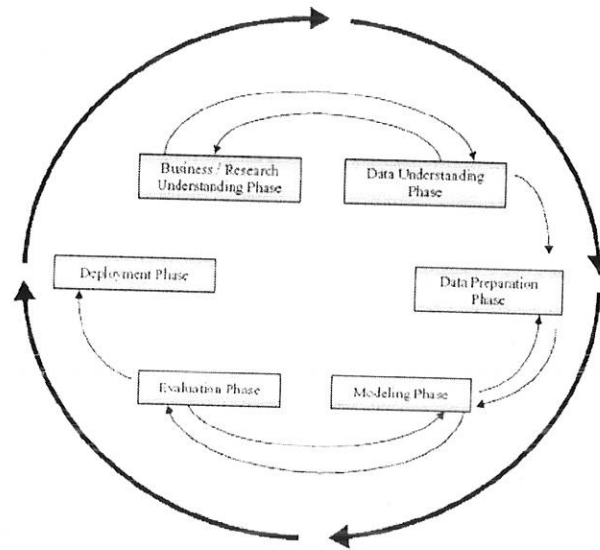


Figure 2. 1 : CRISP-DM process model.

Note that the phase sequence is adaptive. That is, the next phase in the sequence often depends on the outcomes associated with the preceding phase. The arrows indicate the most significant dependencies between phases. Here under CRISP-DM phases are presented [35].

2.1.2.1 Business understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

2.1.2.2 Data understanding

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.

2.1.2.3 Data preparation

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record and attribute selection as well as transformation and cleaning of data for modeling tools.

i. Data Cleaning

Data cleaning is an attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data [23] [3].

Missing Values: The case here is tuples have no recorded values. Different methods to handle missing values [23] [3] such as:

- Ignore the tuple (usually done when the class label is missing)
- Fill in the missing value manually (time consuming and may not be feasible)
- Use global constant to fill in the missing values (Replace all missing values by the same constant)
- Use attribute mean to fill in the missing values (especially for numeric values)
- Use the most probable value to fill in the missing value

Detecting Noisy Data and Outliers: - Han & Kamber [23] describe noise as a random error or variance in a measured variable. Outliers are unusual data values that are not consistent with most observations. Commonly, outliers result from measurement errors, coding and recording errors, and, sometimes, are natural, abnormal values. Such non representative samples can seriously affect the model produced later”.

There are two strategies for dealing with outliers and noisy values:

- Detect and eventually remove outliers as a part of the preprocessing phase, or
- Develop robust modeling methods that are insensitive to outliers.

ii. Data Integration and Transformation

Data mining often requires data integration- the merging of data from multiple data stores. The data may also need to be transformed into forms appropriate for mining [3] [23].

As Han & Kamber [23] clarified, data are transformed or consolidated in to forms appropriate for mining and involves many activities such as Smoothing (to remove noise from data), Aggregation (to apply summarization option), Generalization (to replace low level data by higher-level concepts), Normalization (to scale attributes such that they fall within a specified ranges), Attribute Construction (to construct new attributes from features).

iii. Data Reduction

Data reduction techniques can be applied to obtain a reduced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data. That is, mining on the reduced data set should be more efficient yet produce the same (or almost the same) analytic results [23]

From many strategies available for data reduction, Han & Kamber [23], presents the following: data cube aggregation(applying aggregation operations for data cube construction), dimension reduction(detection and removal of irrelevant, weakly relevant, or redundant attributes or dimension), data compression(reduce datasets through compression mechanisms), numerosity reduction(replace data by alternative smaller data representation), discretization and concept hierarchy generation(replace raw data values for attributes by ranges).

2.1.2.4 Modeling

A model is a high-level, global description of a data set. It takes a large sample perspective. It may be descriptive- summarizing the data in a convenient and concise way- or it may be inferential, allowing one to make some statement about the population from which the data were drawn or about likely future data values [24].

In this phase, various modeling techniques such as clustering, classification, association, and the like are selected and applied and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often necessary.

2.1.2.5 Evaluation

Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model and review the steps executed to construct the model and to be certain that it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

2.1.2.6 Deployment

Creation of the model is generally not the end of the research. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. It often involves applying 'live' models within an organization's decision making processes, for example in real-time personalization of Web pages or repeated scoring of marketing databases.

However, depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. In many cases it is the customer, not the data analyst, who carries out the deployment steps. However, even if the analyst will not carry out the deployment effort it is important for the customer to understand up front what actions need to be carried out in order to actually make use of the created models.

2.1.3 Data mining and related Fields

2.1.3.1 Data Mining vs. Databases

Data mining is applicable and feasible as long as there are huge data collocated and stored in a certain format. The information/data repository includes relational databases, data warehouses, transactional databases, advanced database systems, flat files and the World Wide Web. Particularly the relational databases are one of the most popularly available and rich in information repositories, and thus they are at the heart of data mining processes [23]. Relational databases store data in a structured table upon which mining can be performed. To this end the data mining must be integrated with databases to identify interesting patterns and uncover hidden knowledge [32].

2.1.3.2 Data Warehousing vs. Data Mining

A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, which usually resides under a single site. Data warehouse are constructed via process of data cleaning, data transformation, and data integration, data loading and periodic data refreshing [23]. Place [33] defined it as a centralized data management and retrieval. These definitions clearly state that data warehouse integrates and consolidates data from multiple sources thereby allowing enterprise to remember what it has noticed about its customers over the long period of time.

According to Two Crows Corporation [46], data to be mined is first extracted from an enterprise data warehouse into a data mining database. There is some real benefit if your data is already part of a data warehouse. This is due to the fact that the problems of cleansing data for a data warehouse and for data mining are very similar. If the data has already been cleansed for a data warehouse, then it most likely will not need further cleaning in order to be mined. Furthermore, you will have already addressed many of the problems of data consolidation and put in place maintenance procedures.

In the nutshell data warehouse provides the enterprise with a memory. But, memory is of little use without intelligence. Hence, data mining provides different tools and techniques that add intelligence to the data warehouse [3]. Nevertheless, data warehouse is not a prerequisite for data mining since building a data warehouse requires huge investment and long period of time, and the benefit of data mining projects should exceed the cost to be incurred [22]. Setting up a large data warehouse that consolidates data from multiple sources, resolves data integrity problems, and loads the data into a query database can be an enormous task, sometimes taking years and costing

millions of dollars. You could, however, mine data from one or more operational or transactional databases by simply extracting it into a read-only database [46].

2.1.3.3 OLAP vs. Data Mining

One of the most common questions from data processing professionals is about the difference between data mining and OLAP (On-Line Analytical Processing). As discussed in the following manner, they are very different tools that can complement each other.

According to Two Crows Corporation [46], OLAP is part of the spectrum of decision support tools. Traditional query and report tools describe what is in a database. OLAP goes further and can be used to answer why certain things are true. The user forms a hypothesis about a relationship and verifies it with a series of queries against the data. In other words, the OLAP analyst generates a series of hypothetical patterns and relationships and uses queries against the database to verify them or disprove them. OLAP analysis is essentially a deductive process. In OLAP, when the number of variables being analyzed is in the dozens or even hundreds, it becomes much more difficult and time-consuming to find a good hypothesis.

Data mining is different from OLAP because rather than verify hypothetical patterns, it uses the data itself to uncover such patterns. It is essentially an inductive process than OLAP which is a deductive process. However, OLAP is complementary in the early stages of the knowledge discovery process because it can help you explore your data, for instance by focusing attention on important variables.

2.1.3.4 Data Mining, Machine Learning and Statistics

Data mining is a multidisciplinary field. It draws works form different disciplines. According to Han and Kamber [23], it encompasses concepts from database technology, artificial intelligence,

machine learning, neural networks, statistics, pattern recognition, knowledge-based systems, knowledge acquisition, information retrieval, high-performance computing, and data visualization.

According to Two Crows Corporation [46], data mining takes advantage of advances in the fields of artificial intelligence (AI) and statistics. Both disciplines have been working on problems of pattern recognition and classification. Both communities have made great contributions to the understanding and application of neural nets and decision trees.

Data mining does not replace traditional statistical techniques. Rather, it is an extension of statistical methods that is in part the result of a major change in the statistics community. Accordingly, data mining is the application of AI and statistical techniques to common business problems in a fashion that makes these techniques available to the skilled knowledge worker as well as the trained statistics professional. Data mining is a tool for increasing the productivity of people trying to build predictive models.

2.1.4 Data Mining Tasks

Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks. In general data mining tasks can be classified into two categories: prediction and description. Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest, and description focuses on finding human-interpretable patterns describing the data. Although the boundaries between prediction and description are not sharp (some of the predictive models can be descriptive, to the degree that they are understandable, and vice versa). The relative importance of prediction and description for particular data-mining applications can vary considerably. The goals of prediction and description can be achieved using a variety of particular data-mining methods [20]

Literatures describe the data mining tasks/methods/functionalities in a slightly different ways. For example, Fayyad, Piatetsky-Shapiro and Smyth [20] have put it as classification, regression, probability density estimation, summarization, dependency modeling, and change and deviation and detection modeling. On the other hand, Han and Kamber [23] described the data mining tasks as concept/Class description: characterization and description, association analysis, classification and prediction, cluster analysis, outlier analysis, outlier analysis, and evaluation analysis. Berry and Linoff [3] summarize the mining tasks as classification, estimation, prediction, affinity grouping, clustering, and description and profiling. The data mining tasks as stated by [13] are data description and summarization, segmentation, concept descriptions, classification, prediction, and dependency analysis.

Despite the different views and summarizations of the data mining tasks as depicted in the aforementioned discussions, the researcher prefers to describe it from CRISP perspective. This is due the fact that it is the industry standard process of the data mining project or research. The following are the tasks of data mining [13].

2.1.4.1 Data description and summarization

Data description and summarization aim at the concise description of characteristics of the data, typically in elementary and aggregated form. This gives the user an overview of the structure of the data. Sometimes, data description and summarization alone can be an objective of a data mining project. This kind of problem would be at the lower end of the scale of data mining problems.

However, in almost all data mining projects data description and summarization is a sub goal in the process, typically in early stages. It is advisable to carry out data description and summarization before any other data mining problem type is addressed. To this end, it can help to understand the

nature of the data and to find potential hypotheses for hidden information. Data description and summarization typically occurs in combination with other data mining problem types.

Summarization also plays an important role in the presentation of final results. The outcomes of the other data mining problem types may also be considered summarizations of data, but on a higher conceptual level.

2.1.4.2 Segmentation

The segmentation data mining task aims at the separation of the data into interesting and meaningful subgroups or classes. All members of a subgroup share common characteristics. For instance, in shopping basket analysis one could define segments of baskets depending on the items they contain.

Segmentation can be performed manually or (semi-) automatically. The analyst can hypothesize certain subgroups as relevant for the business question based on prior knowledge or based on the outcome of data description and summarization. However, there are also automatic clustering techniques that can detect previously unsuspected and hidden structures in data that allow segmentation.

2.1.4.3 Concept descriptions

Concept description aims at an understandable description of concepts or classes. The purpose is not to develop complete models with high prediction accuracy, but to gain insights. For instance, a company may be interested to learn more about their loyal and disloyal customers. From a concept description of these concepts (loyal and disloyal customers) the company might infer what could be done to keep customers loyal or to transform disloyal customers to loyal customers.

Concept description has a close connection to both segmentation and classification. Segmentation may lead to an enumeration of objects belonging to a concept or class without any understandable description. Typically, there is segmentation before concept description is performed.

2.1.4.4 Classification

Classification assumes that there is a set of objects – characterized by some attributes or features – which belong to different classes. The class label is a discrete (symbolic) value and is known for each object. The objective is to build classification models (sometimes called classifiers), which assign the correct class label to previously unseen and unlabeled objects. Classification models are mostly used for predictive modeling.

The class labels can be given in advance, for instance defined by the user or derived from segmentation. Classification is one of the most important data mining problem types that occur in a wide range of various applications. Classification has connections to almost all other tasks of data mining. For example, in the course of building a good classification model, segmentation can either provide the class labels or restrict the dataset.

2.1.4.5 Prediction

Prediction is very similar to classification. The only difference is that in prediction the target attribute (class) is not a qualitative discrete attribute but a continuous one. The aim of prediction is to find the numerical value of the target attribute for unseen objects. In the literature, this problem type is sometimes called regression. If prediction deals with time series data then it is often called forecasting. Prediction is usually used to predict some missing values rather than class labels [23].

2.1.4.6 Dependency analysis

Dependency analysis consists of finding a model that describes significant dependencies (or associations) between data items or events. Dependencies can be used to predict the value of a data item given information on other data items. Although dependencies can be used for predictive modeling, they are mostly used for understanding. Dependencies can be strict or probabilistic.

Associations are a special case of dependencies. Associations describe affinities of data items (i.e., data items or events which frequently occur together). A typical application scenario for associations is the analysis of shopping baskets.

Dependency analysis has close connections to prediction and classification, where dependencies are implicitly used for the formulation of predictive models. There is also a connection to concept descriptions, which often highlight dependencies.

In applications, dependency analysis often co-occurs with segmentation. In large datasets, dependencies are seldom significant because many influences overlay each other. In such cases it is advisable to perform a dependency analysis on more homogeneous segments of the data.

2.1.5 Application Areas of Data Mining

The potential areas of application of data mining techniques are practically without limit. It is integrated and is applicable in different sectors. The key areas for the application of data mining include banking and financial services, retail and packaging, direct marketing, quality and warranty management, insurance, etc. [3] [12] [23].

2.1.5.1 Application of Data Mining in Telecommunications

The telecommunications industry generates and stores a tremendous amount of data. These data include call detail data, which describes the calls that traverse the telecommunication networks, network data, which describes the state of the hardware and software components in the network, and customer data, which describes the telecommunication customers. The amount of data is so great that manual analysis of the data is difficult, if not impossible. To this end, the advent of data mining technology promised solutions to these problems and for this reason the telecommunications industry was an early adopter of data mining technology [47].

It is critical for telecommunications operators to have a solid understanding of their market, customers, and networks. Traditional expert systems have drawbacks to fulfill this requirement in that they often depend on experts with obvious natural limitations. Data mining could dynamically determine various rules directly from the data in databases. It has been widely applied in telecommunication industry in recent years [41].

According to Weiss [47], one of application of data mining in telecom includes fraud data mining. This can be used to identify telecommunication fraud, improve marketing effectiveness, and identify network faults.

Related Literature Review

Different research works in the area of customer segmentation, which is the major task of CRM, have been done. They have tried to investigate the applicability of data mining in the different areas of telecommunications. Fekadu [21] has conducted his research on Customer Relationship Management in case of the post-paid service of ETC. He has used the K-Means and decision tree algorithms for customer segmentation and classification. Jember [27] has undertaken his research

on mobile fraud detection, and Tamene [42] conducted a descriptive data mining research on telephone fault.

However, the present study investigates the applicability of data mining techniques, particularly clustering and classification techniques, to the behavioral customer segmentation. It is different from the existing body of knowledge, research works, in the following points:

- The research is conducted on CDMA data, which has not been done so far
- The methods employed; the data sources and collection methods are different
- The objectives of the research; one of the objective of this research is to compare, contrast and recommend the best classification algorithm for behavioral customer segmentation
- The tools employed to build the clustering and classification models
- The attributes derived to describe the calling behavior of customers and the like.

Telecommunication companies maintain a great deal of data about their customers. In addition to the general customer data that most businesses collect, telecommunication companies also store call detail records, which precisely describe the calling behavior of each customer. It is often useful to profile customers based on their patterns of phone usage, which can be extracted from their call detail data. These customer profiles can then be used for marketing purposes, or to better understand the customer, which in turn may lead to better forecasting models. In order to effectively mine the call detail data, it must be summarized to the customer level. Then, a classifier induction program can be applied to a set of labeled training examples in order to build a classifier [47].

2.2 CRM AND SEGMENTATION CONCEPTS

2.2.1 Customer Relationship Management

2.2.1.1 Overview

Customers are the heart of any successful businesses, particularly large-scale and multinational ones. To this end, many forward-looking companies are creating a long-lasting relationship with them. Many companies recently are adopting customer centric strategies programs, tools and technology for efficient and effective CRM. Building a business around a customer relationship is a revolutionary change for most companies. In every industry, forward-looking companies are moving toward the goal of understanding each customer individually and using that understanding to make it easier for the customer to do business with them rather than with competitors [3]

CRM has become number one focus in today's competitive market. More than ever, the ability to understand and manage close relationship with customers has become a prerequisite to achieve business goals. Past and present trends imply the relevance of customer information and knowledge to build strong relationship with their customers over long period of time through providing customers satisfaction, as the same time, earn business value [29].

CRM helps companies improve the profitability of their interactions with customers while at the same time making the interactions appear friendlier through individualization. To succeed with CRM, companies need to match products and campaigns to prospects and customers in order to intelligently manage the customer life cycle [15]. The key factors of successful CRM are interaction with and identification of the customer [8].

CRM is the broad category of concepts, tools, and processes that allows an organization to understand and serve everyone with whom it comes into contact. CRM is about gathering

information that is used to serve customers—basic information, such as name, address, meeting and purchase history, and service and support contacts. In a supplier relationship it might be procurement history, terms and conditions, or contact information. This information is then used to better serve the clients [1].

CRM is a multifaceted process, mediated by a set of information technologies that focuses on creating two-way exchanges with customers so that firms have an intimate knowledge of their needs, wants, and buying patterns [51].

Kalakota & Robinson (cited in Chen Ruey-Shun, Wu and. Chen J. Y [10]) defines CRM as the strategy integrating sales, marketing and service, which unites operating procedures and technology to better understand customers from different perspectives. Kandell (cited in Chen Ruey-Shun, Wu and. Chen J. Y [10]) views CRM as a customer centric initiative that regards customer lifecycle as an important business asset and aims to retain customers and enhance customer satisfaction. According to Caldwell (cited in Chen Ruey-Shun, Wu and. Chen J. Y [10]), CRM is not a new concept. Many businesses have practiced it for a long time, for instance, by memorizing customer's background and spending habits and introducing promotions targeting certain customers based on the information obtained.

Mckinsey & Company [31] sees CRM as the target application of customers' data to critical business problems. It in turn provides 10 to 20 percent performance improvements to many businesses in less than a year.

Srivastava [40] explains CRM emerged in the last decade to reflect the central role of the customer for the strategic positioning of a company. CRM takes a holistic view over customers. It encompasses all measures for understanding the customers and for exploiting this knowledge to

design and implement marketing activities, align production and coordinate the supply-chain. CRM puts emphasis on the coordination of such measures, also implying the integration of customer-related data, meta-data and knowledge and the centralized planning and evaluation of measures to increase customer lifetime value. CRM gains in importance for companies that serve multiple groups of customers and exploit different interaction channels for them. This is due to the fact that information about the customers, which can be acquired for each group and across any channel, should be integrated with existing knowledge and exploited in a coordinated fashion.

CRM is a broadly used term, and covers a wide variety of functions, not all of which require data mining. These functions include *marketing automation* (e.g., campaign management, cross- and up-sell, customer segmentation, customer retention), *sales force automation* (e.g., contact management, lead generation, sales analytics, generation of quotes, product configuration), and *contact center management* (e.g., call management, integration of multiple contact channels, problem escalation and resolution, metrics and monitoring, logging interactions and auditing), among others [40].

Edelstein [15] states that CRM is about managing all customers' interactions. However, effective interactions require information about customers and prospects in all stages of the customers' life cycle.

- ✓ Acquiring customers
- ✓ Increasing the value of customers
- ✓ Retaining good customers

The purpose of CRM is to improve marketing productivity, which is measured in terms of efficiency, effectiveness and economy. This can be achieved through creating cooperative and collaborative process that help reduce transaction costs, increase revenue, and finally create value to business during the lifetime of a customer. It is an integrated effort to identify, maintain and

build up a network for the mutual benefit of both sides through interactive, individualized and value added contacts over a long period of time. In such cases, CRM is regarded as a dominant/core paradigm of marketing. It is regarded as a shift of marketing role from manipulating the customer to genuine involvement with customer through appropriate communication and sharing of knowledge [34].

CRM is regarded as business strategy aimed at gaining a long-term competitive advantage of optionally delivering customer value and extracting business value simultaneously. For the vast majority of business, the ability to acquire, retain and enhance customer relationship is the last place left to find as advantage [9].

As it has clearly been described above, CRM is the core of any customer-centric organization. The aforementioned different definitions give a solid understanding, concepts, and issues of CRM. More or less, all of the definitions state that CRM is the integration of information and communication technologies in every business interaction with customers to understand their preferences and contribution to the business thereby tailoring the services and products accordingly.

2.2.1.2 Reasons for adopting CRM

According to SPARTAN INFORTECH [39], CRM helps:

- Develops better communication channels
- Collects vital data such as customer details and order history
- Creates detailed profiles such as customer preferences
- Delivers instant access to customer history
- identifies new sales opportunities
- Simplifies marketing and sales processes by understanding customer needs.

The principal benefits of CRM Gray and Byun (cited in SPARTAN INFORTECH [39]) are to

- Improve the organization's ability to retain and acquire customers
- Maximize the lifetime value of each customer (share of wallet)
- improve service without increasing cost of service. CMG (cited in SPARTAN INFORTECH, 2008)

2.2.1.3 Effective CRM

Effective CRM consists of the following three components Gray, Byun (cited in SPARTAN INFORTECH [39]):

- **Customer:** The customer is the only source of the company's present profit and future growth.
- **Relationship:** The relationship between a company and its customers involves continuous bi-directional communication and interaction. The relationship can be short-term or long-term, continuous or discrete, and repeating or one-time.
- **Management:** CRM is not an activity only within a marketing department. Rather it involves continuous corporate change in culture and processes. The customer information collected is transformed into corporate knowledge that leads to activities that take advantage of the information and of market opportunities.

Thus, an effective CRM puts together these components, equipped with tools, to put together a long term beneficial strategy of growth for the organization.

2.2.2 Customer Segmentation

Kelly [28] explains the importance of customer segmentation; increasingly companies realize that consumers differ in their needs, preferences, sensitivities, opinions and behaviors. The transition from mass marketing to target marketing that is currently in progress creates demands for more

sophisticated customer segmentation techniques. The key enabler of any segmentation strategy is customer data.

Customer data are the raw material that must be captured, integrated and effectively analyzed in order to achieve the goal of profiling customers. Effective segmentation requires both static customer profile data (e.g. demographic, lifestyle, preferences) and customer behavior data (e.g. usage, loyalty, profitability). Once these data are integrated they can be interrogated to discover groups of customers sharing the same characteristics and needs.

The focus of many industries has changed towards a more clearly customer-orientation. While competition is increasing, there is a need to better understand customers, and to quickly respond to their individual needs and wants. Thus, the objective is to increase the share of wallet for each individual customer and save costs by focusing on more targeted promotions [36].

All customers are not equally important to a business. Some of them may be big spenders, others average/medium spenders, and the rest low income generators to the organizations. Accordingly, segmentation enables to identify different group of customers.

Segmentation and customer profiling are two of the most well-known application of CRM [6]. They further defined segmentation as a term used to describe the process of dividing customers into homogeneous groups on the basis of shared or common attributes (habit, test etc.). However, customer profiling is describing customers by their attributes, such as age, income, and life style. This is done by building a customer's behavior model and estimating its parameters. Customer profiling is a way of applying external data to a population of possible customers. Depending on the data available, they can be used to prospect new customers or to "drop out" existing bad customers. The goal is to predict behaviors based on the information we have in each customer

Thearling (cited in Bounsaythip and Rinta-Runsala [6]). Profiling is performed after customer segmentation.

Market segmentation is the process of partitioning the heterogeneous market into separate and distinct homogeneous segments. A segment consists of a group of consumers who react in a similar way to a given set of marketing stimuli [28].

According to Bounsaythip and Rinta-Runsala [6], customer segmentation is a way to have more targeted communication with the customers. The process of segmentation describes the characteristics of the customer groups (called segments or clusters) within the data. Segmenting means putting the population into segments according to their affinity or similar characteristics. Customer segmentation is the preparation step for classifying each customer according to the cluster groups that have been defined.

Segmentation is a way to have more targeted communication with the customers. The process of segmentation describes the characteristics of the customer groups (called segments or clusters) within the data. Segmenting means putting the population into segments according to their affinity or similar characteristics. Customer segmentation is a preparation step for classifying each customer according to the customer groups that have been defined. Segmentation is essential to cope with today's dynamically fragmenting consumer marketplace. By using segmentation, marketers are more effective in channeling resources and discovering opportunities.

The construction of user segmentations is not an easy task. Difficulties in making good segmentation are the following Verhoef et al. (cited in Jansen [26]):

Relevance and quality of data are essential to develop meaningful segments. If the company has insufficient customer data, the meaning of customer segmentation is unreliable and almost worthless. Alternatively, too much data can lead to complex and time-consuming analysis. Poorly organize data (different formats, different source systems) makes it also difficult to extract interesting information. Furthermore, the resulting segmentation can be too complicated for the organization to implement effectively. In particular, the use of too many segmentation variables can be confusing and result in segments which are unfit for management decision making. On the other hand, apparently effective variables may not be identifiable. Many of these problems are due to an inadequate customer database.

Intuition: Although data can be highly informative, data analysts need to be continuously developing segmentation hypotheses in order to identify the 'right' data for analysis.

Continuous process: Segmentation demands continuous development and updating as new customer data is acquired. In addition, effective segmentation strategies will influence the behavior of the customers affected by them; thereby necessitating revision and reclassification of customers. Moreover, in an e-commerce environment where feedback is almost immediate, segmentation would require almost a daily update.

Over-segmentation: A segment can become too small and/or insufficiently distinct to justify treatment as separate segments. One solution to construct segments can be provided by data mining methods that belong to the category of clustering algorithms.

2.2.2.1 Bases of Segmentation

Many segmentation approaches have been devised and each of these has merits and demerits. Experience suggests that most enterprises use a combination of approaches to deliver maximum

benefit. The ability to deliver sophisticated segmentation techniques to the business enhances the ability of the enterprise to deliver more tailored marketing programs, to identify segments that are more important to the business, to identify segments that have been neglected and to become more attentive to previously unrecognized consumer needs [28].

According to Kelly [28] customer segmentation can be achieved two ways-a priori approach and cluster segmentation. A priori approach to segmentation defines, in advance, a framework or system that describes characteristics of customers or prospects based on information that is known about those individuals. Some common a priori approaches to looking at segments include loyalty, profitability, sensitivity, usage, demographics, psychographics and attitude.

In addition to the a priori approach, data-mining techniques make possible a different approach to segmentation— namely, cluster segmentation. The cluster segmentation approach, in direct contrast to the a priori method, seeks to discover naturally occurring clusters of customers who share common characteristics or behave in the same way. Regardless of the segmentation technique used, the starting point is the collection of the data that provide the variables to construct the segments.

Berry and Linoff [3] state that the purpose of segmentation is to tailor products, services, and marketing message to each segment. Customer segments have traditionally been based on market research and demographics. The problem with this segments based on market research is that it is hard to know how to apply them to all of the customers who were not part of the survey. The problem with the customer segments based on demographics is that not all of the customers actually have the test and product affinities ascribed to their segment. Consequently, undirected (clustering) data mining technique can be used for behavioral segmentation, which is the division

of a market into groups based on consumer knowledge, attitude, use of or response to a product. Behavioral data for this segmentation is summarized from transactions and billing histories.

2.2.3 Customer Segmentation at ETC

Customer segmentation is at the heart of CRM applications. Therefore, ETC has classified customers into big business, ETC employee, ETC service, government, international organization, payphone, reseller, public organizations, small business, and residential. Besides, the CCSD has tried to classify customers into corporate and non-corporate customers. CCSD considers the following list of customers as corporate customers that are to be provided with special support [18].

1. Private enterprises
2. Public enterprises
3. Financial Institutions
4. Major Governmental Offices,
 - Ministries, Commissions, authorities, agencies, Federal Offices and their branches, Government Universities and hospitals,
 - Regional government offices
 - Beneficiaries of school net, agri net, woreda net, etc. via broadband
5. All embassies, international organizations (UN, EU, AU, etc.), international and local non-governmental organizations (NGOs), religious institutions and associations.

These corporate customers are expected to fulfill at least one of the requirements listed below in order to be beneficiaries of the special support rendered by the corporation.

1. Those that use broadband data and Internet services of 128KB per second and above
2. All embassies, international organizations (UN, AU, EU, etc.) and International NGOs

3. Those organizations that have more than two branches under their head office
4. Those customers that rank within the first fifty in their monthly average payment in all services in the respective zonal or regional offices of the corporation.

Leaving this as it is, as a result of the nature of their work, exporters (coffee, flower, leather and hides, etc.), investors, embassies, and international organizations such as UN, EU, AU and International NGOs are corporate customers of the corporation.

The foregoing discussion has elucidated that there is the concept of customer segmentation in the corporation side although it is not based on the service usage patterns of customers. Besides, this kind of segmentation does not provide a basis to understand customers and manage relationships with them for the purpose of creating value to the business. It is the behavioral segmentation that provides useful information on usage patterns of customers to design and implement sound marketing strategies and programs.

2.3. DATA MINING IN CRM AND CUSTOMER SEGMENTATION

Edelstein [15] elucidates the difference between operational and analytical CRM in that the operational CRM was limited to capturing, storing, and retrieving customers profile without telling what hidden information it entails. However, data mining has enabled a business to turn its customers-base into a useful knowledge through discovering non-trivial and interesting patterns. Similarly, Berry and Linoff [3] note that, data mining is an intelligent approach that comb through the enterprise customers data warehouse, which allows an enterprise to remember what it has noticed about its customers.

Data mining is being applied to help manage all phases of a customer lifecycle, including acquiring new customers, increasing revenue from existing customers, and retaining good customers [45].

Hence, data mining enables to differentiate, create long-lasting relationships with customers through tailoring services and products, incentives, promotions, and other special services.

Data mining is a process that applies the techniques of artificial intelligence to the task of discovering useful patterns in data, and is proving particularly powerful in the identification of customers sharing the same characteristics. This segmentation of customers into affinity clusters presents new possibilities for customer segmentation [28].

Han and Kamber [23] believe that data mining is being adopted in different business sectors to help them provide more customized, personal service to their customers in lieu of mass marketing. By studying the browsing and purchasing patterns, companies can learn more about individual customers or customer groups.

2.4 DATA MINING METHODS FOR CUSTOMER SEGMENTATION

2.4.1 Overview

The data mining techniques mostly used for customer segmentation are clustering and classification. Saarevirta [37] elucidate that customer clustering and classification are two of the most important data mining methodologies used in marketing and CRM. Furthermore, Saarevirta beliefs that the business can use the data to divide customers into segments based on such variables as current customer profitability, a measure of the life time value of a customer, and the retention probability, which highlights visible marketing opportunity.

2.4.2 Clustering Techniques

In this research, supervised (decision tree and Artificial Neural Network) and unsupervised (clustering) data mining tasks are experimented. The unsupervised clustering task is carried out to investigate the behavioral and optimal segments of customers. Clustering is a widely used

technique in data mining applications for discovering patterns in underlying data. According to Kelly [28] customer segmentation can be achieved into ways-a priori approach and cluster segmentation. In marketing clusters formed for a business purpose are usually called "segments," and customer segmentation is the popular application of clustering [3]. Segmentation and customer profiling are two of the most well-known application of CRM (Bounsaythip and Rinta-Runsala [6]. Clustering is an attempt to identify natural clusters of records that share similar characteristics. In early days market segmentations were done not based on customers' product or service usages, but now a day clustering technique of data mining is at the heart of identifying the natural customer segmentation.

Han and Kamber [23] define the clustering as the process of grouping a set of physical or abstract objects into classes of similar objects. A cluster is a collection of data objects that are similar to one another within the same cluster and dissimilar to the objects in another cluster.

Clustering technique considers data tuples as objects. They partition the objects into groups or cluster, so that objects within a cluster are "similar" to one another and "dissimilar" to objects in other clusters. Similarity is commonly defined in terms of how "close" the objects are in space, based on distance function. The quality of the cluster may be represented by its diameter, the maximum distance between any two objects in the cluster. Centroid distance is an alternative measure of cluster quality and is defined as the average distance of each cluster object from cluster centroid [23].

There are many different algorithms to do the clustering task of data mining. Han and Kamber [23] categorize them into partitioning, hierarchal, Density-based, Grid-based, and model-based methods. Despite the availability of these different methods, the most widely used one is K-means algorithm, which is the partitioning method [6] [23] [28]. Different researchers have successfully

been using the K-means algorithm to customers' segmentation [14] [21] [25] [36]. The present study also uses the K-means algorithm to undertake the clustering task.

Why K-means?

1. The K-Means algorithm is the best known and easiest algorithm. Customer segmentation is best achieved using clustering algorithms—the best-known and most commonly utilised is the k-nearest neighbour technique. This approach is also known as k-Means, case-based learning or automatic cluster detection [28] [6].
2. The K-Means algorithm is relatively scalable and efficient in processing large data sets because the computational complexity of the algorithm is $O(nkt)$, where n is the total number of objects, k is the number of clusters, and t is the number of iterations. Normally, $k \ll n$ and $t \ll n$ [23].
3. Han and Kamber [23] further explains that the K-means algorithm is applicable if it is possible to determine the mean of the cluster. Usually it works well if the dataset is numeric. Hence in this research almost all attributes are numeric except category, which was normalized into binary form, the K-means was found the natural choice of the clustering task carried out.
4. The researcher had also tried to study the other clustering algorithms, particularly the EM (Expectation Maximization). As Han and Kamber [23] clearly explained, EM extends from the K-means paradigm. It is recommended when there are large amount of missing data.
5. Finally, K-means algorithm is implemented in the WEKA data mining tool, which has been applied in this research. Besides, the researcher found it easy to interpret the clustering results obtained from the K-means algorithm.

2.4.2.1 The K-Means Method

The K-means algorithm divides a data set into a predefined number of 'K' clusters. 'K' in the K-means refers to the number of segments to partition the dataset, while 'means' refers to the average location of all of members (which are records from a database) of a particular cluster. The K-means algorithm "self-organizes" to create clusters.

The K-means algorithm basically has three steps to do the clustering task [3]. The first step, the algorithm randomly selects K data points to be the seeds. Each of the seeds is an embryonic cluster with only one element. The second step assigns each record to the closest seed. One way to do this is by finding the boundaries between two clusters. The boundaries between two clusters are the points that are equally close to each cluster. Finally, the third step is to calculate the centroid of the clusters; these now do a better job of characterizing the clusters than the initial seeds finding the centroid is simply a matter of taking the average value of each dimension for all the records in the cluster. The centroids become the seeds for the next iteration of the algorithm (K-means).

Step 2 is repeated, and each point is assigned to the cluster with the closest centroid. The process of assigning points to a cluster and then recalculating centroids continues until the cluster boundaries stop changing. In practice, the K-means algorithm usually finds a set of stable clusters after a few iterations.

In order to form clusters, each record from a database is mapped as a point in 'record space.' The number of dimensions contained in the space corresponds to the number of fields in the records. The value of each field can be geometrically interpreted as a distance from the origin along the corresponding axis of the space. In addition, to ensure the usefulness of this interpretation, the fields must all be converted into numbers and the numbers must be normalized so that a change in one dimension is comparable to a change in another.

According to Bounsaythip and Rinta-Runsala [6], K-means is based on a concept of distance, which requires a metric to determine distances. Euclidean distance can be used for continuous attributes, while for categorical variables; one has to find a suitable way to calculate the distance between attributes in the data; in which case a business expert is needed to help determine a good metric.

The original choice of a value for K determines the number of clusters that is to be found. In addition, if this number does not match the natural structure of the data, the technique will not obtain good results. Unless the data miner suspects the existence of a certain number of clusters, the experimenter will have to experiment with different values for K.

Berry and Linoff [3] elucidate that, in general, the best set of clusters is the one that does the best job of keeping the distance between members of the same cluster small and the distance between members of adjacent clusters large. They further state that the best set of clusters in the descriptive data mining may be the one showing unexpected pattern in the data.

2.4.3 Decision Trees

Once the clusters have been created using clustering algorithms, they need to be interpreted. Though there are several approaches to perform this, one of the approaches widely used for understanding clusters is building a decision tree with the cluster label as the target variable and using it to drive rules explaining how to assign new records to the correct cluster [3].

According to Two Crows Corporation [45], decision trees are a way of representing a series of rules that lead to a class or value. There are two types of decision trees, namely classification and regression trees. Classification trees label records and assign them to the appropriate class. Classification trees can also provide the confidence that the classification is correct. In such cases

the classification tree reports the class probability, which is the confidence that a record belongs to a given class. Regression trees estimate the value of a target variable that takes on numeric values. Trees can grow in any forms. They could be binary trees of non-uniform depth, that is, each node has two children and the distance of a leaf to the root varies. In **Figure 2.2.**, each node represents a 'yes' or 'no' question, the answer to which determines through which of the two paths a record proceeds to the next level of the tree.

According to Bounsaythip and Rinta-Runsala [6] decision tree learning, is a method for approximating discrete-valued target functions, in which the learned function is represented by a decision tree. Decision trees perform many tests and then try to arrive at the best sequence for predicting the target. Each test creates branch that leads to more tests, until testing terminates in a leaf node (**Figure 2.2**). The path from the root to the target leaf is the rule that classifies the target. The rules are expressed in if-then form.

According to Han and Kamber [23] a decision tree is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, the leaf nodes represent class or class distributions (see **Figure 2.2**). In order to classify an unknown sample, the attribute values of the sample are tested against the decision tree. A path is traced from the root to a leaf node that holds the class prediction for that sample. Decision trees can easily be converted to classification rules.

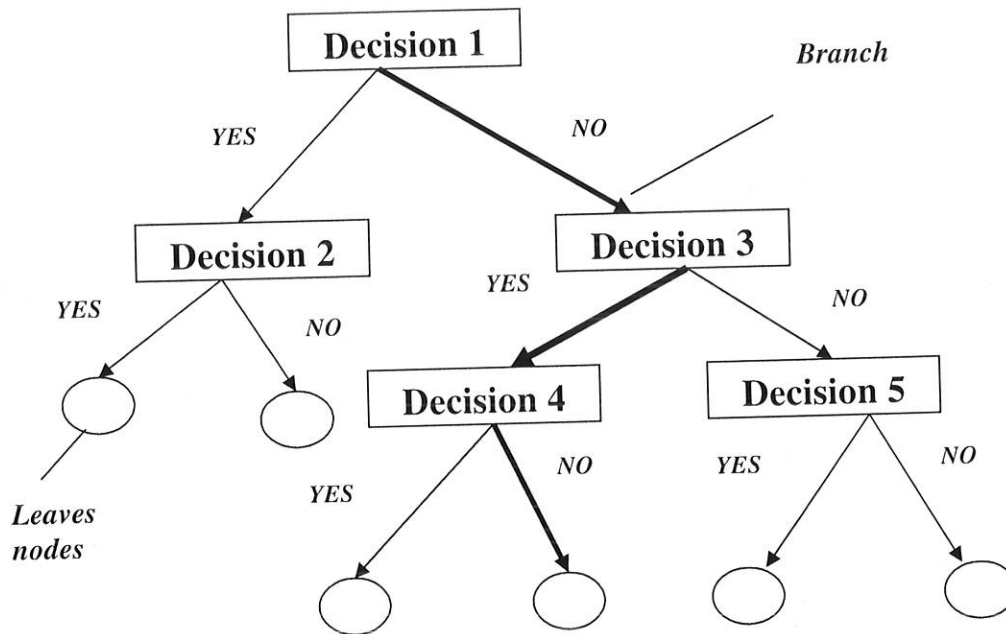


Figure 2. 2: A Scenario that shows how decision tree is constructed

2.4.3.1 Tree Induction

The training process that creates the decision tree is called induction and requires a small number of passes through the training set. Most decision tree algorithms go through two phases: a tree growing (splitting) phase followed by a pruning phase [6].

Splitting: The tree growing phase is an iterative process which involves splitting the data into progressively smaller subsets. The first iteration considers the root node that contains all the data. Subsequent iterations work on derivative nodes that will contain subset of the data. at each split, 94 e variables are analyzed and the best split is chosen. One important characteristics of splitting is that it is greedy, which means that the algorithm does not look forward in the tree to see if another decision would produce a better overall result.

Stopping Criteria: Tree-building algorithms usually have several stopping rules. These rules are usually based on several factors including maximum tree depth, minimum number of elements in a node considered for splitting, or its near equivalent, the minimum number of elements that must be

in a new node. In most implementations the user can alter the parameters associated with these rules. Some algorithms, in fact, begin by building tree to their maximum depth. While such a tree can precisely predict all the instances in the training set (except conflicting records), the problem with such a tree is that, more than likely, it overfits the data.

Pruning: After a tree is grown, one can explore the model to find out nodes or subtrees that are undesirable because of overfitting, or rules that are judged inappropriate. Pruning removes splits and the subtrees created by them. Pruning is a common technique used to make a tree more general. Algorithms that build trees to maximum depth will automatically invoke pruning. However, users have also the ability to prune the tree interactively.

2.4.3.2 Different Decision algorithms

Decision tree algorithms commonly implemented include Chi-squared Automatic Interaction Detection (CHAID), Classification and Regression Trees (CART), C4.5 and C5.0. All are well suited for classification; some are also adaptable for regression. The distinguishing features between tree algorithms include [6]:

Target variables: Most tree algorithms require the target (dependent) variable be categorical. Such algorithms require that continuous variables are binned (grouped) for use with regression.

Splits: Many algorithms support only the binary splits, that is, each parent node can split into at most two child nodes. Others generate more than two splits and produce a branch for each value of categorical variables.

Split Measures: help select which variables to use to split at a particular node. Common split measures include criteria based on gain, gain ratio, GINI, chi-squared and entropy.

Rule generation: algorithms such as C4.5 and C5.0 include methods to generalize rules associated with a tree; this removes redundancies. Other algorithms simply accumulate all the tests between the root node and the leaf node to produce the rules.

Why Decision Tree?

The decision classification algorithm was selected due to the following reasons:

- Decision trees are easy to understand
- Decision trees are easily converted to a set of production rules
- Decision trees can classify both categorical and numerical data, but the output attribute must be categorical
- There are no a priori assumptions about the nature of the data.

As stated in the literature review under section 2.4.3.2, different decision tree algorithms were discussed and C4.5 is among one of the algorithms that includes methods to generalize rules associated with a tree; this removes redundancies. J48 algorithm is the weka's implementation of the C4.5 (Witten and Frank, 2000). In fact, J48 implements a later and slightly improved version of which is known as C4.5 Revision 8, which was the last public version of this family version of algorithms before C5.0 (new version of C4.5); a commercial implementation was released [48].

The process of the J48 algorithm to build a decision tree:

1. Choose an attribute that best differentiates the output attribute values.
2. Create a separate tree branch for each value of the chosen attribute.
3. Divide the instances into subgroups so as to reflect the attribute values of the chosen node.
4. For each subgroup, terminate the attribute selection process if:

- a. All members of a subgroup have the same value for the output attribute, terminate the attribute selection process for the current path and label the branch on the current path with the specified value.
 - b. The subgroup contains a single node or no further distinguishing attributes can be determined. As in (a), label the branch with the output value seen by the majority of remaining instances.
5. For each subgroup created in (3) that has not been labeled as terminal, repeat the above process.

The algorithm is applied to the training data. The created decision tree is tested on a test data set, if available. If test data is not available, J48 performs a cross-validation using the training data itself.

2.4.4 Artificial Neural Networks (ANNs)

ANN has been initially designed to let machines think and reason out as human beings do [6]. It mimics the natural human being's neuron system. According to Edelstein [15], neural networks offer a means of modeling large and complex problems in which there may be hundreds of predictor variables that have many interactions. Neural nets are used in classification problems. ANNs have also successfully been used in many classification and prediction tasks of data mining such as fraudulent call prediction, customer profiling, financial predictions, etc. [3].

Neural networks are best distinguished from other intelligent techniques in that they are non-rule-based and can additionally be made stochastic so that the same action does not necessarily take place each time for the same input. A stochastic behaviour allows a neural network to explore its environment more fully and potentially to arrive at a better solution than linear methods might allow [11].

As depicted in **Figure 2.3**, a neural net starts with an input layer, where each node corresponds to a predictor variable. These input nodes are connected to a number of nodes in a hidden layer. The nodes in the hidden layer may be connected to nodes in another hidden layer, or to an output layer. The output layer consists of one or more response variable.

After the input layer, each node takes in a set of inputs, multiplies them by a connection weight, adds them together, applies a function (called the activation or squashing function) to them, and passes the output to the node(s) in the next layer.

The goal of training is defined as the process of iterating through the training set to adjust the weights. Each iteration is sometimes called generation, the neural net is to estimate the connection weights so that the output of the neural net accurately predicts the test value for a given input set of values. The most common training method is backpropagation [3] [23]. Each training method has a set of parameters that control various aspects of training such as avoiding local optima or adjusting the speed of conversion.

In neural networks, data of any quality can be used to provide reasonable predictions and they will sift through it to find the truth. However, neural nets require very careful data cleansing, selection, preparation, and pre-processing.

In **Figure 2.3.**, a training sample, $X = (x_1; x_2; \dots; x_i)$, is fed to the input layer. Weighted connections exist between each layer, where w_{ij} denotes the weight from a unit j in one layer to unit i in the previous layer.

Neural network is a set of connected input/output units where each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to

be able to predict the correct class label of the input samples. Neural network learning is also referred to as connectionist learning due to the connections between units.

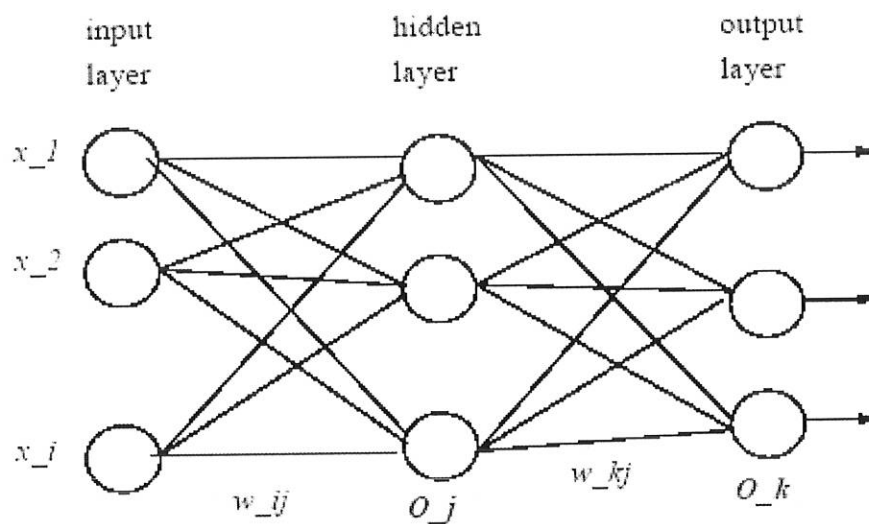


Figure 2. 3: A multilayer feed-forward neural network [23].

Advantages of neural networks, however, include their high tolerance to noisy data as well as their ability to classify patterns on which they have not been trained. In addition, several algorithms have recently been developed for the extraction of rules from trained neural networks. These factors contribute towards the usefulness of neural networks for classification in data mining.

2.4.4.1 A multilayer feed-forward neural network

One of the common types of neural network is the feed-forward backpropagation network [46]. According to Han and Kamber [23], the backpropagation algorithm performs learning on a multilayer feed-forward neural network. An example of such a network is shown in **Figure 2.3**. The inputs correspond to the attributes measured for each training sample. The inputs are fed simultaneously into a layer of units making up the input layer. The weighted outputs of these units are, in turn, fed simultaneously to a second layer of “neuron-like” units, known as a hidden layer. The hidden layer's weighted outputs can be input to another hidden layer, and so on. The number

of hidden layers is arbitrary, although in practice, usually only one is used. The weighted outputs of the last hidden layer are input to units making up the output layer, which emits the network's prediction for given samples.

The units in the hidden layers and output layer are sometimes referred to as neurodes, due to their symbolic biological basis, or as output units. The multilayer neural network shown in **Figure 2.3** has two layers of output units. Therefore, we say that it is a two-layer neural network. Similarly, a network containing two hidden layers is called a three-layer neural network, and so on.

Multilayer feed-forward networks of linear threshold functions, given enough hidden units, can closely approximate any function [23]. The network is feed-forward in that none of the weights cycle back to an input unit or to an output unit of a previous layer. It is fully connected in that each unit provides input to each unit in the next forward layer [6].

The feedforward network has two common types of architectures. According to Two Crows Corporation (2005), these are Multilayer perceptrons (MLPs) and Radial Bases Function (RBF). The only fundamental difference is the way in which hidden units combine values coming from preceding layers in the network-MLPs use inner products, while RBFs use Euclidean distance

2.4.4.2 Defining a network topology

The architecture (or topology) of the neural network is the number of nodes and hidden layers, and how they are connected. In designing the network either the user or the software must choose the number of hidden nodes and hidden layers, the activation function, and the limit on the weight [46]. Similarly, Han and Kamber [23] state that before training can begin, the user must decide on the network topology by specifying the number of units in the input layer, the number of hidden layers (if more than one), the number of units in each hidden layer, and the number of units in the output layer.

Normalizing the input values for each attribute measured in the training samples will help speed up the learning phase. Typically, input values are normalized so as to fall between 0.0 and 1.0. Discrete-valued attributes may be encoded such that there is one input unit per domain value. One output unit may be used to represent two classes (where the value 1 represents one class and the value 0 represents the other). If there are more than two classes, then 1 output unit per class is used [23].

Han and Kamber [23], states that there are no clear rules as to the “best” number of hidden layer units. Network design is a trial by error process and may affect the accuracy of the resulting trained network. The initial values of the weights may also affect the resulting accuracy. Once a network has been trained and its accuracy is not considered acceptable, then it is common to repeat the training process with a different network topology or a different set of initial weights. Besides, Berry and Linoff [3] elucidate that a neural network can have any number of hidden layers, but in general, one hidden layer is sufficient. The wider the layer (that is, the more units it contains) the greater the capacity of the network to recognize patterns. This greater capacity has a drawback, although, because the neural network can memorize patterns-of-one in the training examples. We want the network to generalize on the training set, not to memorize it. To achieve this, the hidden layer should not be too wide; in other words the number of hidden layer nodes should be kept small.

2.4.4.3 Backpropagation

Han and Kamber [23] explain that Backpropagation learns by iteratively processing a set of training samples, comparing the network’s prediction for each sample with the actual known class label. For each training sample, the weights are modified so as to minimize the mean squared error between the network’s prediction and the actual class. These modifications are made in the

“backward” direction, i.e., from the output layer, through each hidden layer down to the first hidden layer. Although it is not guaranteed, in general the weights will eventually converge, and the learning process stops.

Backpropagation follows the following three steps [3] [23]:

1. initialize the weights in the network to small random numbers
2. The network gets a training example and, using the existing weights in the network, it calculates the output and outputs.
3. Back propagation then computes the error by taking the difference between the calculated result and the expected one.
4. The error is feed back through the network and the weights are adjusted to minimize the error.

Initialize the weights: The weights in the network are initialized to small random numbers (e.g., ranging from -1.0 to 1.0, or -0.5 to 0.5). Each unit has a bias associated with it, as explained below. The biases are similarly initialized to small random numbers.

Each training sample, X , is processed by the following steps.

Propagate the inputs forward. In this step, the net input and output of each unit in the hidden and output layers are computed. First, the training sample is fed to the input layer of the network. The net input to each unit in the hidden and output layers is then computed as a linear combination of its inputs. To help illustrate this, a hidden layer or output layer unit is shown in **Figure 2.4**. The inputs to the unit are, in fact, the outputs of the units connected to it in the previous layer. To compute the net input to the unit, each input connected to the unit is multiplied by its corresponding weight, and this is summed. Given a unit j in a hidden or output layer, the net input, I_j , to unit j is:

$$I_j = \sum_i w_{ij} O_i + \theta_j \quad (2.1)$$

Where w_{ij} is the weight of the connection from unit i in the previous layer to unit j ; O_i is the output of unit i from the previous layer; and θ_j is the bias of the unit. The bias acts as a threshold in that it serves to vary the activity of the unit.

Each unit in the hidden and output layers takes its net input, and then applies an activation function to it, as illustrated in **Figure 2.4**. The function symbolizes the activation of the neuron represented by the unit. The logistic or sigmoid function is used.

Berry and Linoff [3] describe the activation function as having two parts. The first part is the combination function that merges all the input into a single value. Each input to the unit has its own weight. The most common function is the weighted sum, where each input is multiplied by its weight and these products are added together as indicated in equation (2.1). Other combination functions are sometimes useful and include the maximum of the weighted input, the minimum, and the logical AND and OR of the values. Although there is a lot of flexibility in the choice of combination function, the standard weighted sum work well in many situations.

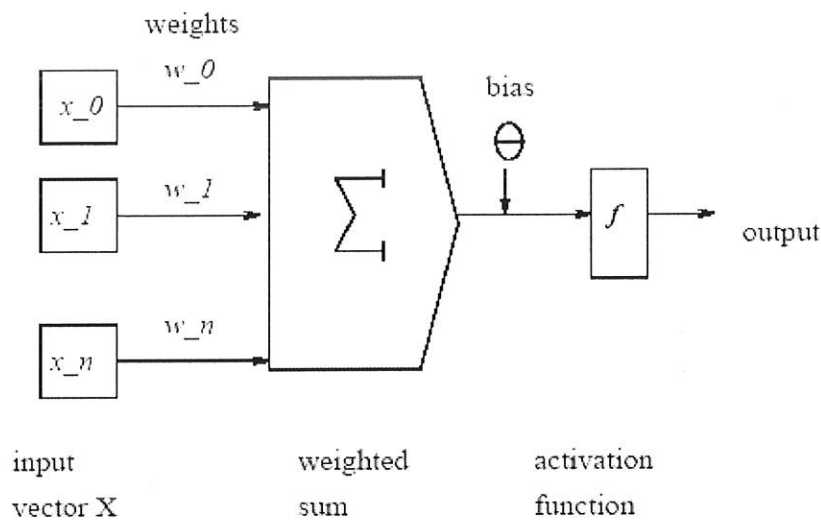


Figure 2.4 A hidden or output layer unit: The inputs are multiplied by their corresponding weights in order to form a weighted sum, which is added to the bias associated with the unit. A nonlinear activation function is applied to the net input [23]

The second part of the activation function is the transfer function, which gets its name from the fact that it transfers the value of the combination function to the output of the unit. Three typical transfer functions: sigmoid (logistic), linear, and hyperbolic tangent functions.

Feed-forward neural network consisting only of units with linear transfer functions and weighted sum combination function is really just doing a linear regression. Sigmoid functions are S-shaped functions of which the two most common for neural networks are the logistic and the hyperbolic tangent. The major difference between them is the range of their outputs, between 0 and 1 for the logistic and between -1 and 1 for the hyperbolic tangent.

The logistic and hyperbolic tangent transfer functions behave in a similar way. When the weighted sum of all the inputs is near 0, then these functions are a close approximation of a linear function. As the magnitude of the weighted sum gets larger, these transfer functions gradually saturate (to 0 and 1 in the case of the logistic; to -1 and 1 in the case of hyperbolic tangent). This behavior corresponds to a gradual movement from a linear model of the input to a non-linear model. In short, neural networks have the ability to do a good job of modeling on three types of problems: Linear problems, near-linear problems, and nonlinear problems.

Given the net input I_j to unit j , then O_j , the output of unit j , is computed as:

$$O_j = \frac{1}{1 + e^{-I_j}} \quad (2.2)$$

This function is also referred to as a squashing function, since it maps a large input domain onto the smaller range of 0 to 1. The logistic function is nonlinear and differentiable, allowing the backpropagation algorithm to model classification problems that are linearly inseparable.

Backpropagate The Error The error is propagated backwards by updating the weights and biases to reflect the error of the network's prediction. For a unit j in the output layer, the error Err_j is computed by:

$$Err_j = O_j(1 - O_j)(T_j - O_j) \quad (2.3)$$

Where O_j is the actual output of unit j , and T_j is the true output, based on the known class label of the given training sample. Note that $O_j(1 - O_j)$ is the derivative of the logistic function.

To compute the error of a hidden layer unit j , the weighted sum of the errors of the units connected to unit j in the next layer is considered. The error of a hidden layer unit j is:

$$Err_j = O_j(1 - O_j) \sum_k Err_k w_{jk} \quad (2.4)$$

Where w_{jk} is the weight of the connection from unit j to a unit k in the next higher layer, and Err_k is the error of unit k .

The weights and biases are updated to reflect the propagated errors. Weights are updated by Equations (2.5) and (2.6) below, where Δw_{ij} is the change in weight w_{ij} .

$$\Delta w_{ij} = (l) Err_j O_i \quad (2.5)$$

$$w_{ij} = w_{ij} + \Delta w_{ij} \quad (2.6)$$

The variable l in **Equation 2.5** is the learning rate, a constant typically having a value between 0.0 and 1.0. Backpropagation learns using a method of gradient descent to search for a set of weights which can model the given classification problem so as to minimize the mean squared distance between the network's class predictions and the actual class label of the samples. The learning rate helps to avoid getting stuck at a local minimum in decision space (i.e., where the weights appear to converge, but are not the optimum solution), and encourages finding the global minimum. If the

learning rate is too small, then learning will occur at a very slow pace. If the learning rate is too large, then oscillation between inadequate solutions may occur. A rule of thumb is to set the learning rate to $1/t$, where t is the number of iterations through the training set so far.

Biases are updated by Equations (2.7) and (2.8) below, where Δ_j is the change in bias θ_j .

$$\Delta\theta_j = (t)Err_j \quad (2.7)$$

$$\theta_j = \theta_j + \Delta\theta_j \quad (2.8)$$

Note that here we are updating the weights and biases after the presentation of each sample. This is referred to as case updating. Alternatively, the weight and bias increments could be accumulated in variables, so that the weights and biases are updated after all of the samples in the training set have been presented. This latter strategy is called epoch updating, where one iteration through the training set is an epoch. In theory, the mathematical derivation of backpropagation employs epoch updating, yet in practice, case updating is more common since it tends to yield more accurate results.

Terminating condition: Training stops when

1. all Δw_{ij} in the previous epoch were so small as to be below some specified threshold, or
2. the percentage of samples misclassified in the previous epoch is below some threshold, or
3. a prespecified number of epochs has expired.

In practice, several hundreds of thousands of epochs may be required before the weights will converge.

CHAPTER THREE

DATA PREPARATION

3.1. Overview

This chapter thoroughly discusses about the data preparation or preprocessing tasks of the data mining. This embraces detail description of the database from which the research data has been collected. The data cleaning, transformation, and integration activities are explained in this chapter. The forthcoming discussions present issues of data in consistent with the CRISP-DM.

3.2. Data understanding

Data is at the heart of data mining tasks, methods and tools. In the best of all possible worlds, the required data would already be resident in a corporate data warehouse, cleansed, available, historically accurate, and frequently updated [3].

As stated in the section 1.1.2.3 of chapter one, ETC stores the CDR along with the billing information and Customers' profile in a single database ("USHACOM"). The "USHACOM" database has been in operation since 2004. The back-end (database) is implemented with oracle 9i and the front-end (interface) is designed with Java computer programming language. As it has been stated in the aforementioned discussions, "USHACOM" is a giant database that hosts billions of records for all telephone services' CDR along with the billing information and customers' profile since its introduction. Despite its scalability for large size records, the corporation has limited online storage capacity. Accordingly, the CDR of all services for the year 2004 and 2005 were moved to offline storage as the records exceeded the ETC's storage capacity.

The researcher initially planned to take all CDMA telephones' CDR starting from the year the service has been introduced (i.e. 2004). Nevertheless, the CDR is available online only for the year

2006 to 2008. In addition, the database administrators of the corporation advised and recommended the researcher that it was unmanageable to process five years CDR that contains more than 200 million rows.

Hence, for the present research three months (April, May, and June, 2008) CDR has been considered. These three months data accounts for more than 8 million record sizes. This data is partitioned into 18 tables when it was imported into MS-Access 2003 database. Finally, the data was exported to SQL Server 2000 for preprocessing the data (both the CDR and Customers' profile). This is because the three tables' records could be merged in one table in SQL Server 2000, which is impossible in access. In addition to this, merging the table in one table made the data preprocessing easy and efficient.

The researcher employed purposive sampling in selecting the three months (April, May, June, 2008) CDR due to the following reasons:

- The database did not contain all CDR data due to the abovementioned reasons
- The Servers computers have been busy in tabulating and calculating the monthly services bill of customers at the time of collecting the research CDR
- MS-Access 2003 could not handle large size of data. Here it should be clear that the data was directly imported into MS-Access from the ETC's database, oracle 9i, as the corporation is not using SQL Server database. And
- The database administrators recommended that relatively these three months CDR is clean and contains relatively less outliers and missing values, and

The researcher discussions and survey of the data with the domain experts and database administrators while selecting the data were constructive and made the data analysis manageable. The purposive sampling, which is discussed above, is also chosen due to the database

administrators and domain experts' advices and recommendations. The purposive sampling technique revealed quality data, one that has less outliers and missing values.

3.3. Data processing

The data preprocessing task is one of the challenging parts of this research. Its preparation has taken much of the research time and effort. As described by Han and Kamber [23] the data preprocessing task of data mining includes data cleaning, data integration, data transformation, and data reduction. Each of these tasks has been performed by the researcher in the course of preparing the data for model building.

3.3.1. Data Cleaning

The data mining algorithms perform quality mining as long as there is a cleansed data, which is free from outliers, noise, and with no missing values for an attribute (s). Many of the attributes used in this research are derived from the base attributes. For example, Total Number of Calls was derived through the SQL COUNT aggregate function. Similar SQL aggregate functions were employed while deriving attributes directly used for model building. However the researcher tried to do the data cleaning such as handling missing attributes values, and noise removal before deriving the attributes that are used to build the clustering and classification models. Accordingly, there are only two attributes that had missing values. The 'termination Location' column from the CDR and 'Rent' field from the customers' profile contained many missing values. The following points briefly describe the methods employed in handling missing and outlier data.

- After the thorough look at the CDR data, the researcher identified many values are missing in termination location attribute, which is practically difficult to fill in. Hence, it has been deleted from further considerations.

- The rent column again also has some missing values. However, the researcher has learnt from the database administrators that those missing values depict a zero rent. Hence, the missing values of the rent attribute are filled with zero.
- 373 ETC's CDMA Service and 1 payphone telephone records, a total of 374 records, were deleted from the datasets. This was because the domain experts stated that these telephones were not meant for profit rather used for the daily normal operations of the business. Therefore, there was no meaning to include them in the investigation of natural segmentation of customers based on their usage behavior.
- 392 ETC Employees' CDR records are also deleted from the dataset because the domain experts advised that these telephone numbers are not meant for profit rather they are given to some employees of the corporation as one of the facilities. As these employees are given certain range of free calls, their CDR statistics may create outliers, or dominate the total calls variable, as there may be more total calls than the other categories of customers.
- The residential customers category field comprises of 18,476 records out of 22, 406 dataset. This may dominate/influence the quest for natural and best segmentation of customers. Vis-à-vis this, the researcher in collaboration with the domain experts reduced its size to 6,000 by using stratified sampling technique. For the present study, every third (18,476/6,000) residential customers' records has been chosen for inclusion to the dataset.

3.3.2. Data Integration and Data Transformation

Data mining often requires data integration (merging of data) from multiple data stores. The data may also need to be transformed into forms appropriate for mining [23].

The datasets, which are used for the subsequent model building, are prepared and derived from the two integrated data sources, the CDR along with the billing information and the customers' profile.

However, the merging of these databases are been done after all relevant attributes have been derived from the CDR. The service numbers, which represent each customer telephone number, is used as key in the merging processes. To ease the data manipulation using MS-Access 2003, the CDMA CDR is partitioned into 18 tables. All of the 18 tables have been imported in SQL Server 2000 and they are finally consolidated into one table using the SQL insert command. It is to be noted that the CDMA CDR that is taken to this research contains only local calls. This is due to the fact that ETC has banned almost all customers from international calls as it had been suffering from uncollected billing for this service (CDMA). Accordingly the CDR for international calls could not be used.

After all relevant attributes are derived from the CDMA CDR, the total number of individual call records are summarized into 23, 172 records that describe the calling behavior of each customer. This figure represents the total population of CDMA customers. However the stratified sampling has reduced the data size by half. Hence, the final dataset used for model building consisted of 10,090 records.

The data transformation task includes smoothing, aggregation, generalization, normalization and attributes selection [23]. The dataset that has been used for modeling building are smoothened and normalized. For instance, all attributes values have been mapped to in the range $[-1,1]$ while experimenting with the ANN's models.

3.3.3. Data Reduction Methods

Some attributes which are assumed not relevant, as judged by the domain experts together with the researcher, to the present research have been eliminated through the application of select SQL command on the oracle 9i DBMS. Many values are missing in termination location attribute, which

is practically difficult to fill in. Hence, it has been deleted from further considerations. Other attributes, which describe the switches code, are eliminated as they do not carry any information about customers calling behavior. These attributes have been eliminated at the time of data collection and in the preprocessing steps.

In all the data preprocessing tasks, Oracle 9i, SQL Server 2000, MS-Access 2000, and MS-Excel software have been intensively employed and used.

CHAPTER FOUR

EXPERIMENTATION

4.1. Overview

This chapter presents steps and procedures followed during experimentations. Besides, description and evaluation of the performances of the clustering and classification models developed are presented. It applies the methods, techniques and algorithms of data mining that are thoroughly discussed in the research methodology under section 1.5 of chapter one to accomplish the objectives of the research. All processes of the experimentation are described inline with the CRISP-DM references guide.

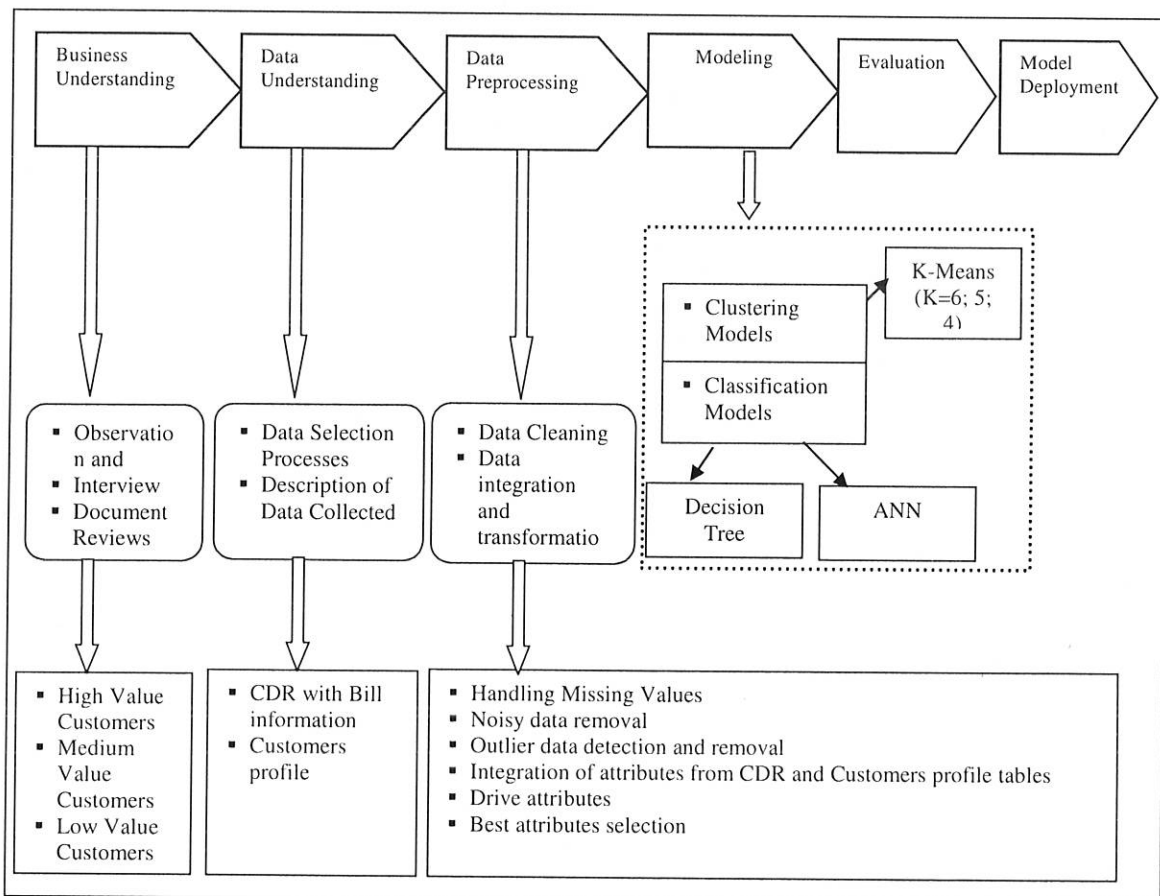


Figure 4. 1. The experimentations flow diagram

4.2. Business Understanding

As stated in chapter one section 1.5, the researcher has applied different techniques to understand the business/domain of this research. With the help of these methods, the researcher has tried to thoroughly investigate and understand the business. The corporation rate the value of customers depending on their usage, vis-à-vis revenue generated.

Customers that generate high revenue to the corporation with little use of the network resources are considered as high value (corporate customers). Little use of the network resources refers to minimum number of telephone calls, short duration per call and very short overall call durations. This category of customers creates little network congestion but generate substantial amount of revenue. However customers who congest and create high network traffic but generate high revenue to the corporation are still treated as high value customers next to the previous corporate customers.

Customers that highly congest or create high telephone network traffic with least revenue are taken as very low value customers. In addition, those customers that do not create the network traffic and contribute very low revenue to the corporation are still under the list of low value customers. The intermediate between high and low value customers are medium value customers. These customers reasonably use the corporation telephone networks' resources with reasonable revenue contribution to the business.

In addition to the identification and understanding of the different types of customers that constitute the business, ETC's CCSD considers all International and local NGOs, embassies, governmental organizations, big business customers as corporate customers. It offers special services to this segment of customers. However, these customers are not identified and segmented

based on their usage behavior, rather it is solely based on the premises that they would potentially generate high revenue to the corporation as they are big in size or have international base.

Thus, the theme of this research is to thoroughly see into the customers' transactional data, CDR along with the bill information and their profile, to investigate whether high, medium, and low revenue generating customers can be differentiated based on their telephone usage behaviors. In addition, the relevance of identifying them is two fold. In one hand, it enables corporation to tailor, design and offer different marketing strategies for different segments. This in turn enables the corporation to focus on the existing high value customers while attracting new one with little cost. On the other hand, the corporation can create a positive relationship with its high spending customers while maintaining them satisfied with the services it offers.

4.3. Data Understanding

As stated in the methodology (chapter one section 1.5), different methods have been employed to understand the content and the structure of the data available. All data, which were judged relevant by the researcher together with domain experts, are collected and further analyzed inline with the objectives of the research at hand. The data understanding task has been accomplished through data selection processes, description of data collected, data exploration and data quality verification. The following sections thoroughly describe them.

4.3.1. Data Selection Processes

The researcher has faced really challenging situations at the initial collection of the data, which are relevant to the research. The data collection is really one of the biggest challenges of the research. It took the researcher almost two months to get the permission of accessing the CDR and Customers profile data. This is, as justified by top management of the ETC, due to customers'

privacy reasons. Hence, the researcher must have devised a mechanism to access the pertinent data in a way that could not reveal the callers'/customers' service/telephone number and persuade the top management.

Accordingly, the researcher recommended the top management that it could be possible to replace or map caller telephone numbers to some other intentionally generated numbers. Finally, they were convinced with this idea and permitted access to the pertinent data. All caller telephone numbers are mapped into numbers in the range 200001 to 232682 with the first caller telephone number being 200001. As stated in the data preprocessing, attributes both from the CDMA and Customers services tables have been selected with the help of domain experts. On the top of this, the research objectives have guided the selection of these attributes. The data size and selection methods are thoroughly described in the data preprocessing chapter.

4.3.2. Description of data collected

The pertinent data to carry out the research have been collected from two different tables of the oracle 9i database. These sources are CDMA CDR and the customers' services tables as depicted in **Table 4.1.** and **Table 4.2.** respectively. The tables depict only the initial set of attributes that are basically used to derive many other attributes along with their data types and descriptions.

4.4. Data preprocessing

The methods used for data preparation/preprocessing involves data cleaning, data integration, and cleaning as described below.

4.4.1. Data cleaning

The researcher has tried to do the data cleaning such as handling missing attributes values, and noise removal before deriving the attributes that are used to build the clustering and classification

models. Accordingly, there are only two attributes that have missing values. The termination Location column from the CDR and Rent field from the customers' profile contained many missing values.

Attributes	Data type	Description
SERVICENO	Number	The customers telephone number. This attribute initially was sting data type. Because of the customers' privacy reason, it must have been mapped into other arbitrary numbers in the range 200001 to 232682.
CATEGORY	String	The classification code/name assigned to each customer by ETC. It comprises of the values BigBusiness, ETCEmployee, ETC Service, Government, InternationalOrganization/Religious, PayPhone, PublicOrganization, Reseller, Residential, SmallBusiness.
RENT	Number	The monthly service subscription fee of customers

Table 4. 1. Customers Services Table

- After a thorough look at CDR data, the researcher identified many of the values missing in termination location attribute. The researcher then deleted the whole column as it is difficult to apply other methods such as replacing it with the most common values as significant portion of it contained null values (see **Appendix 1, A**).
- The rent column again has some missing values. However, the researcher has learnt from the database administrators that those missing values depict a zero rent. Hence, the missing values of the rent attribute have been filled with zero (see also **Appendix 1, B**)
- 373 ETC's CDMA Service and 1 payphone telephone records, a total of 374 records, are deleted from the datasets. This is because the domain experts stated that these telephones

Attributes	Data type	Description
SERVICENO	Number	The caller telephone numbers, which are replaced/mapped to other numeric values in range 200001 to 232682. Each Caller's telephone number in the CDR has been replaced by matching the SERVICENO in the Customers table (see Table 4.1.).
CALLEDNO	String	The called parties telephone numbers
CALLDATE	Date/Time	The customers call date
CALLTIME	String	The time at which a customer has placed the telephone call. It is expressed in the form HH:MM:SS, where HH refers to call hour, MM indicates the minutes of the call, and SS connotes the seconds of the call had been made
CALLDURATION	Number	The length/duration of customers call. It is the call duration in terms of seconds.
BILLABLEAMOUNT	Number	The revenue generated per each individual call
CALLTYPECD	String	The Call Type Code. The assumes different values such as WLWZN (Wire Line Within Zones or Addis Ababa), which refers to the customer call made to fixed telephone line within Addis Ababa, WZN (Wireless/Mobile Within Zones), which indicates the call made to Mobile cell phone within Addis Ababa, etc.
TERMINATINGLOCATION	String	The Place of the exchange machine in which the call has terminated.

Table 4. 2. CDMA CDR Table

are not meant for profit rather used for the daily normal operations of the business. Therefore, there is no meaning to include them in the investigation of natural segmentation of customers based on their usage behavior.

- There are also 1756 duplicate records in the customer service table. They have been eliminated by using the delete SQL command in MS-Access 2003.

TNCFixedAA	Number	Total number of each customer call to fixed telephone within Addis Ababa
TNCMobile	Number	Total number of each customer call to mobile telephones
DCMOM	Number	Duration of each customer call during the middle of the three months, which includes the days in the range 11 to 20
DCFixedAA	Number	Duration of calls made by each customer to fixed telephone within Addis Ababa
TNSCall	Number	Total Number of calls made by each customer during the weekend, Sunday
TNCWLOC	Number	Total Number of calls made by each customer to fixed telephone line within Addis Ababa and within the same exchange machine or area
TNCFixedReg	Number	Total Number of calls made by each customer to fixed telephone line from Addis Ababa to different Regions
TNCMOM	Number	Total Number of calls made by each customer during the middle of the months, which includes the days in the range 11 to 20
TNCWeekday	Number	Total Number of Calls made by each customer during the week days, which embrace Monday to Saturday
Trevenue	Number	Total Revenue earned from each customer
RperCall	Number	Revenue per each customer Call
RperDuration	Number	Revenue generated Per Duration
DperCall	Number	Duration per each Individual Call
AvgCallPerThreeMonth	Number	Average Call made by each customer Per Three Months
AvgRperThreeMonth	Number	Average Revenue generated from each customer per three months
AvgDPerThreeMonth	Number	Average Duration calls by each customer Per Three Months Period
Tduration	Number	Total Duration of calls by each customer
TNCEOM	Number	Total Number of Calls made by each customer at end of the three months, which includes the days in the range 21 to 31
TNNightCall	Number	Total Number of calls made by each customer during the night time, which includes times in the range 18:00:01PM to 23:59:59 PM and 00:00:00 AM to 05:59:59 AM
DCFixedReg	Number	Duration of calls made by each customer to fixed telephone from Addis Ababa to regions
DCWeekday	Number	Duration of Calls made by each customer during the week days, which embrace Monday to Saturday

DCSCall	Number	Duration of calls made by each customer during the weekend, Sunday
TNDayCall	Number	Total Number of calls made by each customer during the day time, which includes times in the range 06:00:00 AM to 18:00:00 PM
DCBOM	Number	Duration of each customer call during the beginning of the three months, which includes the days in the range 01 to 10
DdayCall	Number	The duration of each customer call during the Day time, which includes times in the range 06:00:00 AM to 18:00:00 PM
RENT	Number	The Service Rent a customer is required to pay at every another month

Table 4. 3. List of all attributes together with their type and descriptions

After importing all tables, the next task is to combine all tables into one table based on which the researcher derive relevant attributes, as listed in **Table 4.3**. For each derived attribute, a table must have been created to contain the derived values with the create SQL statement. The insert into table SQL command has been used to insert the derived attributes' values into the table. **Figure 4.3.** shows the creation of a table TNCall to hold the Total Number of Calls of individual customers along with its derivation and insert SQL statement. This process is carried out repeatedly to all other attributes derivation.

Finally as part of data integration, individually derived attributes, which have been stored in different tables, must have been brought together to form the final dataset for the clustering, decision tree, and Neural network models building. **Figure 4.4.** exhibits the SQL Select command used to bring the attributes together.

```

insert into cdma_processed (serviceno , TNumberofcall , NBOM , DNightCall , DNEOM , DWZN ,
TCWLWZN , WZN , DMEOM , DWLWZN , NSCall , WLLOC ,
WLAZN , NMOM , NWeekday , TRevenue , NEOM ,
NightCall , DWLAZN , NDWeekday , NDSCall , DayCall ,
DBEOM , DDayCall )
SELECT      dbo.tblTotalCall.serviceno, dbo.tblTotalCall.Tnumber, dbo.tblNBOM.NBOM,
            dbo.tblDNightCall.DNightCall, dbo.tblDNEOM.DNEOM, dbo.tblDWZN.DWZN,
            dbo.tblTCWLWZN.TCWLWZN, dbo.tblWZN.WZN, dbo.tblDMEOM.DMEOM,
            dbo.tblDWLWZN.DWLWZN, dbo.tblSCall.NSCall, dbo.tblWLLOC.WLLOC,
            dbo.tblWLAZN.WLAZN, dbo.tblNMOM.NMOM, dbo.tblWeekday.NWeekday,
            dbo.tblTRevenue.TRevenue, dbo.tblNEOM.NEOM,
            dbo.tblNightCall.NightCall, dbo.tblDWLAZN.DWLAZN,
            dbo.tblDWeekday.DWeekday, dbo.tblDSCall.DSCall, dbo.tblDayCall.DayCall,
            dbo.tblDBEOM.DBEOM, dbo.tblDDayCall.DDayCall
FROM        dbo.tblDayCall LEFT OUTER JOIN
            dbo.tblTotalCall ON dbo.tblDayCall.serviceno = dbo.tblTotalCall.serviceno LEFT OUTER JOIN
            dbo.tblDBEOM ON dbo.tblTotalCall.serviceno = dbo.tblDBEOM.serviceno LEFT OUTER JOIN
            dbo.tblDDayCall ON dbo.tblTotalCall.serviceno = dbo.tblDDayCall.serviceno LEFT OUTER JOIN
            dbo.tblDNEOM ON dbo.tblTotalCall.serviceno = dbo.tblDNEOM.serviceno LEFT OUTER JOIN
            dbo.tblDNightCall ON dbo.tblTotalCall.serviceno = dbo.tblDNightCall.serviceno LEFT OUTER JOIN
            dbo.tblDSCall ON dbo.tblTotalCall.serviceno = dbo.tblDSCall.serviceno LEFT OUTER JOIN
            dbo.tblDWeekday ON dbo.tblTotalCall.serviceno = dbo.tblDWeekday.serviceno LEFT OUTER JOIN
            dbo.tblDWLAZN ON dbo.tblTotalCall.serviceno = dbo.tblDWLAZN.serviceno LEFT OUTER JOIN
            dbo.tblDWLWZN ON dbo.tblTotalCall.serviceno = dbo.tblDWLWZN.serviceno LEFT OUTER JOIN
            dbo.tblDWZN ON dbo.tblTotalCall.serviceno = dbo.tblDWZN.serviceno LEFT OUTER JOIN
            dbo.tblNBOM ON dbo.tblTotalCall.serviceno = dbo.tblNBOM.serviceno LEFT OUTER JOIN
            dbo.tblNEOM ON dbo.tblTotalCall.serviceno = dbo.tblNEOM.serviceno LEFT OUTER JOIN
            dbo.tblNightCall ON dbo.tblTotalCall.serviceno = dbo.tblNightCall.serviceno LEFT OUTER JOIN
            dbo.tblNMOM ON dbo.tblTotalCall.serviceno = dbo.tblNMOM.serviceno LEFT OUTER JOIN
            dbo.tblSCall ON dbo.tblTotalCall.serviceno = dbo.tblSCall.serviceno LEFT OUTER JOIN
            dbo.tblTCWLWZN ON dbo.tblTotalCall.serviceno = dbo.tblTCWLWZN.serviceno LEFT OUTER JOIN
            dbo.tblTRevenue ON dbo.tblTotalCall.serviceno = dbo.tblTRevenue.serviceno LEFT OUTER JOIN
            dbo.tblWeekday ON dbo.tblTotalCall.serviceno = dbo.tblWeekday.serviceno LEFT OUTER JOIN
            dbo.tblWLAZN ON dbo.tblTotalCall.serviceno = dbo.tblWLAZN.serviceno LEFT OUTER JOIN
            dbo.tblWLLOC ON dbo.tblTotalCall.serviceno = dbo.tblWLLOC.serviceno LEFT OUTER JOIN

```

MELAKU_737249\cdma_thesis (8 MELAKU_737249\Administrator \ cdma_db 0:00:00 0 rows Ln 1, Col

Figure 4. 4. The SQL command to bring derived attributes under one table

describing the profitability of customers. Sample values of the final dataset are depicted in Appendix 1, C.

4.5. Modeling

The task of customer segmentation is done with the K-means algorithm. The challenging task in the K-means algorithm is determining the value of K, which determines the optimal clustering

model that creates dissimilar segments of customers. The optimal number of cluster size, i.e. the value of K, is obtained through trial of different models [3] [23]. This is because K is user-defined number. Saarevirta [40] advises that the number of clusters chosen should be driven by how many clusters the business can manage. Accordingly, the business experts have been consulted in setting the optimal value for the same. They have suggested the K values to be in the range 4-6. Vis-à-vis this four cluster models are experimented and evaluated against their performance in creating dissimilar clusters/segments. The classification models have been experimented with the J48 decision tree learner and the back propagation neural network algorithms.

4.5.1. Cluster Modeling

After the preparation of the dataset, the next step is to build the clustering model using the selected tool, weka version 3.5.8. The basic parameter in weka 3.5.8 explorer for simple K-means clustering embraces:

- The number of clusters: the number of clusters (K in K-means) that need to be created. This value was specified by the user manually.
- Seed size: The random number seed to be used. This defines the number of data tuples the cluster must start with.
- The variables to be used for building the clusters

Once the necessary parameters are set, experiments are undertaken in a step wise manner.

4.5.1.1. Experiment 1

All the attributes are input to the first cluster run except the SERVICENO. This is due to the fact that it does not provide any useful information but it may reduce the accuracy of the algorithm. In addition to this, it is meant to be used as a primary key to manipulate, derive and aggregate attributes values for individual customer. The weka 3.5.8 explorer's remove facility is used to

avoid the SERVICENO attribute from the cluster run. **Table 4.4.** depicts the list of input for the first cluster run.

Cluster Run	Variables	Number of records	Number of Clusters
1 st	All attributes listed under Table 4.3. Except SERVICENO	22798	6

Table 4. 4. Summary of Input parameters for the first cluster run

The weka 3.5.8 explorer window, as depicted in **Appendix 2**, lists all attributes ready for the first cluster run. In addition, it is possible to know the minimum, maximum, mean, standard deviation, data type, number of missing values, number of distinct values, and number of unique values of each currently selected attribute. One can see these values for all other attributes by selecting each of them one-by-one. These values in turn enable the researcher together with domain experts to determine the threshold values of each variable for the analysis of the result.

Table 4.5. exhibits the result of the cluster run and the resulting segments. The algorithm is instructed to segment the dataset into six clusters. However, the second cluster has only 22 instances, i.e. almost 0% of the dataset. The researcher then suspected that there would be outliers in the training dataset, as the cluster two shares very few records with the other clusters. The algorithm might have treated these outliers as records behaving differently from other clusters and put them in a separate cluster.

Cluster Distribution					
Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
4871 (48%)	22 (0.0001 %)	3170 (31%)	716 (7 %)	743 (7%)	568 (6%)

Table 4. 5. Training the First Cluster with the default seed (10)

Then the class assigned dataset is assessed visually for the presence of outliers. And it is found that there are extreme values on the TNumberofCall, Tduration, TRRevenue. Accordingly, these outliers have been replaced with the mean values of their fields.

After removing all outliers manually, the next task is to identify the most statistically significant attributes from the dataset and to use these attributes in model building process to increase the performance of the algorithm. Han and Kamber [23] state that the attributes that appear at the top of the decision tree provide more information to classify the instances of the test dataset. Accordingly in this research, attributes are selected with the aid of the J48 decision tree algorithm. The decision tree is built by selecting the cluster index as the dependent variable and the rest as determinant or independent variables. Attributes that appear at the very top of the decision tree and those attributes upon which many records are predicted into different class labels are taken for the next experiment.

The attributes found statistically relevant for the second experimentation are:

- Category, DCallMobile, DPerCall, Rent, RperCall, Tduration, TNumberofCall, and TRRevenue.

4.5.1.2. Experimentation 2

After all the outliers and attributes that are selected with the help of the J48 decision tree, as listed above, the second experiment with the same number of K (6) and default seed (10) is conducted. Different values of the seed are experimented in the quest for effective and normally distributed segment of clusters. **Table 4.6.** summarizes the records distribution in segments of the second cluster run with K=6 and the default value of the seed (10).

Cluster Distribution					
Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
4690 (46%)	86 (1 %)	3180 (32%)	713 (7 %)	751 (7%)	670 (7%)

Table 4. 6. Cluster distribution with k=6 and the algorithm converged after 23 iteration with the default seed (10)

In an attempt to improve the data distribution of the segments, different seed values have been tried. But the seed value at 1000 gives a better data distribution in the segments. When seed was set ten, cluster2 has only 86 (1%) instances as shown in **Table 4.6.**, whereas with the value 1000, its size has drastically changed to 416 (4%) as exhibited in **Table 4.7.** The optimal seed value has been obtained after a number of experimentations. Similarly, many values of the seed have been attempted for the other K values (4, 5). However, the default seed value, i.e. 10, has somewhat uniformly distributed the data in these segments.

Cluster Distribution					
Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
746 (7%)	416(4 %)	4871 (48%)	716 (7 %)	2769 (27%)	572 (6%)

Table 4. 7. Cluster distribution with k=6 and the iteration with the seed (1000)

As indicated above three cluster models are built for K (6, 5, and 4). Besides, different seed sizes are tested on each of this cluster formation to see whether the distribution of the segments could be improved or normally distributed. Yet, only the cluster with K=6 has showed significant difference on the segments data distribution. It fairly distributed the data tuples in each cluster when the value of the seed was set to 1000. The other cluster models with K (4 and 5) performed well with the default seed value.

The next task of the research is to set threshold value for each numeric attributes used to build the clustering model. The threshold value for each attribute has been determined with the marketing experts and with the aid of the weka's minimum, maximum, and mean values display for each attribute. The need for determining the threshold values is solely to determine what patterns are

discovered for each subsequent cluster models with K (6, 5 and 4) and the best cluster seed. **Table 4.8.** depicts the threshold values for each variable suggested by the domain experts.

TnumberofCall (TC)	DCallMobile (DCM)	Trevenue (TR)	RperCall (RPC)	DperCall (DPC)	Tduration (TD)
TC≤300 Very low	DCM≤4500 Very short	TR≤150 very Low	RPC≤0.50 very low	DPC≤90 very short	TD<4.00 very short
300<TC≤800 low	4500<DCM≤11000 short	150<TR≤250 Low	0.50<RPC≤0.75 Low	90<DPC≤105 short	4.00<TD≤10 short
800<TC≤1200 medium	11000<DCM≤23000 medium	250<TR≤550 medium	0.75<RPC≤1.00 medium	105<DPC≤121 medium	10<TD≤16 medium
1200<TC≤1900 High	23500<DCM≤48500 Long	550<TR≤631 High	1.00<RPC≤1.50 high	121<DPC≤151 Long	16<TD≤39 Long
TC>1900 very high	DCM>48500 very Long	TR>631 very high	RPC>1.50 very high	DPC>151 Very Long	TD>39 very Long

Table 4. 8. List of Range of conditions by which a cluster result is assessed

In the following discussions the abbreviation of terms as listed in **Table 4.9.** are used in clusters models analysis and comparison.

List of Abbreviated Terms used in the each cluster analysis		List of Attributes of the dataset	
Abbreviated Words	Description	Abbreviated Attributes	Description
VH	Very High	ATNC	Average Total Number of Call
H	High	ADCM	Average Duration of Call to Mobile
M	Medium	ATR	Average Total Revenue
VL	Very Low	ARPC	Average Revenue Per Call
L	Low	ADPC	Average Duration Per Call
VLG	Very Long	ATD	Average Total Duration
LG	Long	Category	Most frequent category of customers
S	Short	Rent	Monthly Service Rent
VS	Very Short		

Table 4. 9. List of abbreviated words and attributes of the dataset along with their description

The result of the second cluster run is presented in **Table 4.10.**

First Segmentation (K=6, Seed=1000)									
Cluster #	Freq. of records/(share in %)	ATNC	ADCM	ATR	ARPC	ADPC	ATD	Category	Rent
1	746 (7%)	626.1398	45372.3862	637.1721	1.091	117.839	19.314	Government	17
2	416 (4%)	1893.56	135521.353	1898.774	1.113	119	60.781	BigBusiness	17
3	4871 (48%)	150.0597	4765.0386	78.3886	0.55	99.5569	4.1686	Residential	8
4	716 (7%)	110.3087	8378.3198	125.958	1.304	255.762	7.5135	Residential	8
5	2769 (27 %)	322.7625	19052.5574	275.1076	0.933	116.603	9.6966	BigBusiness	17
6	572 (6 %)	1629.963	33123.0717	691.3307	0.426	97.7969	44.233	Residential	8
Total	10090 (100%)								
Cluster #	Freq. of records/(share in %)								
1	746 (7%)	L	LG	VH	H	M	LG		
2	416 (4%)	H	VLG	VH	H	M	VLG		
3	4871 (48%)	VL	S	VL	L	S	S		
4	716 (7%)	VL	S	VL	H	VLG	S		
5	2769 (27 %)	L	M	M	M	LG	M		
6	572 (6 %)	H	LG	VH	VL	VS	VLG		
Total	10090 (100%)								

Table 4. 10. Clustering Result of the second experiment

The upper part of **Table 4.10.** exhibits the attributes' average values for each segment (1-6) and bottom part shows the corresponding mapping of these values into the discrete values. The mapping is performed to facilitate the comparison among the segments generated from the clustering algorithm.

After the average value of each attribute in each cluster has been replaced with the corresponding discrete values, a description for each segment of cluster has been done as exhibited in **Table 4.11.** The ranking is determined based on the profitability or the amount of revenue of customers' in different segments.

Cluster	Description	Possible Rank
1	Low number of calls, long duration of mobile calls, very high revenue, high revenue per call, medium duration per call, Long total duration, Government business category with high frequency, and monthly rent of 17 birr.	1
2	High number of calls, Very Long duration of mobile calls, very high revenue, and Medium revenue per call, medium duration per call, very long total duration, Big business category, and monthly service rent of 17 birr.	2
3	Very low number of calls, short duration of mobile calls, Very Low revenue, Low revenue per call, short duration per call, short total duration, residential business category, and monthly service rent of 8 birr	6
4	Very low number of calls, short duration of mobile call, very low revenue, high revenue per call, very long duration per call, short total duration, residential business category, and a monthly service rent of 8 birr	5
5	Low number of calls, medium duration of mobile calls, medium revenue, medium revenue per call, Long duration per call, medium total duration, Big business category, a monthly service rent of 17 birr	4
6	High number of calls, Long duration of mobile calls, very high revenue, Very Low revenue per call, very short duration per call, very long total duration, residential business category, a monthly service rent of 8 birr	3

Table 4. 11. Cluster summary for K=6 and seed=1000 and the possible rank of clusters.

As stated in the business understanding of the corporation, high value customers are those that generate high revenue to the business with little use of the network resources such as low total number of calls and short duration of calls, of the corporation. Those customers with high number of total calls and long duration of call but generate high revenue to the corporation are still considered to be high value customers next to the formerly described customers.

By considering the abovementioned facts of the business, each cluster has been assigned rank. As can be clearly seen from **Table 4.11.**, the first cluster is ranked first. This is due to the fact that customers in this cluster use little network resources, very low total number of calls, medium duration per call, and reasonably long duration of mobile calls, of the corporation while generating very high revenue and high revenue per call. Cluster one comprises of Government category customers with a monthly rent of birr 17.

The second cluster is ranked second as customers in this segment generate very high revenue and medium revenue per call but with high total number of calls and very long duration of mobile calls.

This cluster consists of high number of business category customers with a monthly service rent of birr 17.

Cluster 6 is similar to cluster 2 in TNumberofCall and Trevenue values, which are high number of calls, and very high revenues respectively. Again customers in this group generate very high revenue but very low revenue per call. Hence, it has been ranked third. Residential customers constitute major portion of this segment. Customers in cluster 5 generate medium revenue, and medium revenue per call with low total number of calls, medium duration of mobile calls and medium total duration. There are large numbers of Big business category customers in this segment. The cluster is ranked fourth.

Customers under cluster 3 and 4 are low value customers as they generate very low revenue. However customers categorized under cluster four generate high revenue per call, therefore, ranked fifth. The third cluster is ranked last. Both cluster 3 and 4 share significant portion of residential customer category.

This experimentation with $K=6$ and seed=1000 seem to have created dissimilar clusters or segments of customers. It has sufficiently differentiated among the high, medium, and low value customers. In addition to this, different categories of customers have been identified in different segments. As depicted in **Table 4.11.**, clusters 1, 2, and 6 embrace high-value customers with first, second, and third rank respectively, whereas cluster 5 embraces medium-level customers, which are next to high-value customers. Government customers constitute major portion of cluster 1, while Big business category of customers are contained in higher proportion in cluster 2. Cluster 6 contains residential customers in higher proportion. Clusters 3 and 4 contain low-value customers. Residential customers constitute the major portion of both clusters.

Although it seems that the second experimentation resulted in very dissimilar clusters, another cluster run have been conducted with K=5 and seed=10 in the quest for a better segmentation model.

4.5.1.3. Experimentation 3

The third clustering model experimentation was conducted with K=5 and seed=10. The result of the cluster run is depicted in **Table 4.12**. The description of the resulting cluster model is interpreted and briefly explained in **Table 4.13**.

First Segmentation (K=5, Seed=10)									
Cluster #	Freq. of records/(share in %)	TNC	DCM	TR	RPC	DPC	TD	Category	Rent
1	4872 (48%)	150.1951	4772.74	78.4773	0.55	99.5511	4.171	Residential	8
2	571 (6%)	1631.399	33107.0245	691.6472	0.425	97.843	44.282	Residential	8
3	3180 (32%)	508.5001	32828.0527	467.2263	0.953	116.746	15.599	BigBusiness	17
4	716 (7%)	110.3087	8378.3198	125.958	1.304	255.762	7.5135	Residential	8
5	751 (7%)	707.75	51382.1666	720.6576	1.103	118.553	22.554	Government	17
Total	10090 (100%)								
Cluster #	Freq. of records/(share in %)								
1	4872 (48%)	VL	S	VL	L	S	S		
2	571 (6%)	H	LG	VH	VL	S	VLG		
3	3180 (32%)	L	LG	M	M	M	M		
4	716 (7%)	VL	S	VL	H	VLG	S		
5	751 (7%)	L	VLG	VH	H	M	LG		
Total	10090 (100%)								

Table 4. 12. Clustering result of the third experimentation

In the third experimentation five clusters are formed. This experiment has resulted in similar segment formation with the second experimentation's result. The interpretation of the result is

previous cluster run with K=6, this cluster run could not differentiate between the high value big business customers but the former does.

4.5.1.4. Experimentation 4

Best clustering scheme is usually found after a number of trial with different segment sizes.

Accordingly the researcher has tried the fourth experimentation with K=4 and seed size of 10.

Table 4.14. summarizes the result obtained for the fourth experimentation. The detail description of each of the segments is presented in **Table 4.15.**

First Segmentation (K=4, Seed=10)									
Cluster #	Freq. of records/(share in %)	TNC	DCM	TR	RPC	DPC	TD	Category	Rent
1	4872 (48%)	150.1951	4772.74	78.4773	0.55	99.5511	4.171	Residential	8
2	571 (6%)	1631.399	33107.0245	691.6472	0.425	97.843	44.282	Residential	8
3	3931 (39%)	546.5659	36372.7333	515.6432	0.982	117.091	16.928	BigBusiness	17
4	716 (7%)	110.3087	8378.3198	125.958	1.304	255.762	7.5135	Residential	8
Total	10090 (100%)								
Cluster #	Freq. of records/(share in %)								
1	4872 (48%)	VL	S	VL	L	S	VS		
2	571 (6%)	H	LG	VH	VL	S	VLG		
3	3931 (39%)	L	LG	M	M	M	LG		
4	716 (7%)	VL	S	VL	H	VLG	M		
Total	10090 (100%)								

Table 4. 14. Clustering result of the fourth experimentation

Cluster	Description	Possible Rank
1	Very Low number of calls, Short duration of mobile calls, Very Low revenue, Low revenue per call, short duration per call, very short total duration, residential business category, and monthly service rent of 8 birr	4
2	High number of call, Long duration of mobile calls, very high revenue, very low revenue per call, short duration per call, very long total duration, residential business category, and monthly service rent of 8 birr	1
3	Low number of calls, Long duration of mobile calls, Medium revenue, Medium revenue per call, Medium duration per call, Long total duration, Big business category, and monthly service rent of 17 birr.	2
4	Very low number of calls, short duration of mobile call, very low revenue, high revenue per call, very long duration per call, Medium total duration, residential business category, and a monthly service rent of 8 birr	3

Table 4. 15. Cluster summary for K=4 and seed=10 and the possible rank of clusters.

Cluster 2 customers generate very high revenue and very low revenue per call with high number of call, long duration of mobile calls, short duration per call, and very long total duration. This segment constitutes residential customers. Cluster 3 constitutes medium revenue and medium revenue per call. This cluster generally embraces medium-level profitable customers. Big business customers share large percentage of the total customer categories. However cluster 1 and 4 contain generally least profitable customers. They differ in that the later contains customers that generate high revenues per call while the former embraces customers with low revenue per call. Residential customers' category constitutes the higher percentage as compared to other customers. However, this cluster model could not identify government category of customers. Besides the segments formed do not really exhibit the customers with high revenue as cluster model of K=6 does.

4.5.1.5. Choosing the best clustering model

In the foregoing experimentations, four clusters model experimentations have been conducted in the quest for appropriate segmentation model. The experimentations are carried out with different values of K (6, 5, and 4) and seed. Despite experimentation with different values of the seed size, only the seed size that resulted in fair distribution of segments is reported. In the first experimentation very few instances of the dataset is assigned in cluster 2. The researcher then

suspected outliers in this cluster (cluster 2). Hence outliers are detected and replace with average attributes values. Besides, the patterns discovered are not meaningful and satisfactory.

The second experimentation with the same number of K (6) but with replaced outliers is conducted and dissimilar clusters are formed. All segments in this cluster model have different characteristics, which is desirable property in inter-cluster relationship. Although this experiment seems to have created segregated segments, it is natural to try other cluster models, particularly K=5 and K=4 are tried in this research.

Finally, the best cluster model has been chosen based on the following parameters:

- Number of iteration the algorithm has undergone to converge. This shows the algorithm has relocated all misplaced data items in their correct classes within a few looping. The minimum value exhibits K-Means algorithm has converged very soon.
- Within cluster sum of squared errors. This is a measure of the “goodness” of the clustering and tells how “tight” the clustering is overall. Lower values are better.
- The domain experts’ judgment.

Number of cluster (K value)	Number of Iterations	Within cluster sum of squared errors
6	21	510.29
5	27	549.30
4	27	1185.80

Table 4. 16 summary of the number of iterations and within cluster sum of squared errors

As described in **Table 4.16**, cluster model at K=6 exhibits the least within cluster sum of squared errors and number iterations as compared to cluster models at K values of 5 and 4 respectively. This shows the cluster model formed at this value of K is a good model. Besides, the domain

experts judged that this cluster model to be the best as compare to other models experimented in this research. This is due to the fact that it could able to create dissimilar segments of customers.

4.5.2. Classification Modeling

Clustering model could identify segments of customers that share high intra-class similarity and low inter-class similarity. Yet, it does not classify new instance of customers into a certain segment. Hence, classification using the decision tree algorithm, particularly J48, and the ANN algorithm (Back propagation algorithm) are further investigated.

In this section the different decision tree and neural network models are built separately by changing different parameters. The classification accuracy of each of these models is reported and their performance is compared in classifying new instances of records.

However, before embarking upon the classification models building, it is worthwhile to discuss about the default parameter for the model training and testing, which is the 10-fold cross-validation. In 10-fold cross-validation, the initial data are randomly partitioned into 10 mutually exclusive subsets or "folds," 1, 2, 3, ..., 10, each approximately equal size. Training and testing is performed 10 times. In the first iteration, the first fold is reserved as a test set, and the remaining 9 folds are collectively used to train the classifier; the classifier of the second iteration is trained on folds 1, 3, 4, ..., 10 and tested on the second fold; and so on. The accuracy estimate is the overall number of correct classifications from the 10 iterations, divided by the total number of samples in the initial dataset [23]. Although the use of such techniques to estimate classifier accuracy increases the overall computation time, yet it is useful for selecting among several classifiers [23].

4.5.2.1. Decision Trees Model Building

In this research work for building decision tree model J48 algorithm is used. All attributes, which are selected for the cluster models building, are fed as independent variables and the cluster labels, which are assigned by the clustering algorithm, as dependent variable for the algorithm.

J48 algorithm contains some parameters that can be changed to further improve classification accuracy. Initially the classification model is built with the default parameter values of the J48 algorithm. **Table 4.17** summarizes the default parameters' values for J48 decision tree algorithm.

Parameter	Description	Default Value
confidenceFactor	The confidence factor used for pruning (smaller values incur more pruning).	0.25
minNumObj	The minimum number of instances per leaf.	2
Unpruned	Whether pruning is performed.	False
10-folds cross-validation		

Table 4. 17. Summary of some of the parameters and their default values in weka's J48 algorithm

Using the default values of the parameters the learning scheme has been conducted and resulted in a decision tree containing 119 nodes and 65 leaves. The confusion matrix generated from this learning scheme is exhibited in **Table 4.18**.

Actual	Predicted						Total	Sore (Accuracy Rate)
	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6		
Cluster1	734	4	0	0	8	0	746	98.39 %
Cluster2	5	398	0	0	13	0	416	95.67 %
Cluster3	0	0	4849	14	0	8	4871	99.55 %
Cluster4	0	0	19	697	0	0	716	97.35 %
Cluster5	6	16	1	0	2746	0	2769	99.17 %
Cluster6	0	0	9	1	0	562	572	98.25 %
Total	745	418	4878	712	2767	570	10090	98.97 %

Table 4. 18. Output from the J48 Decision tree learner with default parameters' values

From the above discussion, it is easy to infer that classification model built with the default parameters' values resulted in better classification accuracy although it has too lengthy tree to weed through and drive all useful rules.

Again as stated in the foregoing discussions, the next task is to build a decision tree with 80 % of dataset for model training and the remaining 20% for test set. This experiment was carried out to see whether better classification accuracy could be achieved than the first and second experimentations. First this experiment has run with default value of the parameters of the J48 decision tree learner algorithm. The result of this experiment is summarized **Table 4.20**.

Actual	Predicted						Total	Sore (Accuracy Rate)
	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6		
Cluster1	151	2	0	0	0	0	153	98.69 %
Cluster2	0	77	0	0	3	0	80	96.25 %
Cluster3	0	0	945	6	0	1	952	99.26 %
Cluster4	0	0	3	143	0	0	146	97.95 %
Cluster5	3	6	0	0	558	0	567	98.41 %
Cluster6	0	0	2	1	0	117	120	97.5 %
Total	154	85	950	150	561	118	2018	98.66 %

Table 4. 20. Summary of the confusion matrix with default parameters with 80 % for training and 20 % for testing dataset

This model is built with 80 % (8072 records) for training and 20 % (2018 records) for validation/testing of the total 10,090 dataset. Out of 2018 (20 %) the dataset used for testing, this algorithm has classified 98.66 % (1991) of it correctly, while misclassifying 1.34 % (27) of the testing dataset. As compared with the decision tree model built in the first experiment with all default parameter values, this algorithm has relatively showed lesser accuracy, even in each individual class prediction accuracy.

Another classification model similar to the above experimentation has been carried out by changing only the default parameter value of the minNumObj from 2 to 20. Out of the 2018 (20 %

of the whole dataset) test set 98.02 % (1978) are correctly classified. In other words, 1.98 % (40) the test set is wrongly classified. However, the tree size has reduced into 37 with leaf nodes of 34. But still the overall accuracy could not be improved as compared to the first model of the decision tree constructed with default parameters of the weka's J48 decision tree learner algorithm.

In the nutshell, the model built with default parameters' values and 10-fold cross-validation has shown better classification accuracy than splitting of 80 % and 20 % of the dataset for training and testing of the model respectively. Hence, among the different decision tree models built in the foregoing experimentations, the first model, with the default parameters' values and 10-fold cross-validation, has been chosen due to its better both overall and individual cluster classification accuracy. A tree and rules generated from this model are depicted in **Appendix 5, A**, and **Appendix 5, B** respectively.

4.5.2.2. ANN Classification Model

After experimenting with many J48 decision tree models, the best model that has shown better overall classification accuracy has been chosen. The next task is to normalize the data in the range $[-1, 1]$, which is suitable for the neural net algorithm. ANN model learn very fast as long as the attributes' values fall in the range $[-1, 1]$ [3] [23]. As stated in the literature and in the specific objectives part of this research, the researcher is opted to build an ANN model for categorical label classes', which are derived through the application of clustering, prediction of unseen dataset. To this end, the researcher has used the WEKA preprocessing facilities to normalize the attributes' values to fall in the range $[-1, 1]$.

Consequently, all numeric and categorical attributes are normalized. Category, which shows the coding scheme the corporation is using to differentiate its customers, is the only categorical attribute in the dataset. There are seven distinct values in this attribute and each of them is assigned

numeric value 1-7 arbitrarily. After mapping these nominal values into these numeric values, the weka's explorer normalize preprocessing facility has been used to normalize all values to fall in the range [-1, 1].

The same attributes that are used to build the decision tree models, are also used in the neural net modeling. With all preprocessing takes in place, the experimentation proceeded with the different neural network models by changing the hidden layers and learning rate parameters. **Table 4.21.** shows some of the neural net classifier algorithm parameters along with their default values.

Parameter	Description	Default Value
hiddenLayers	This defines the hidden layers of the neural network. This is a list of positive whole numbers. 1 for each hidden layer.	'a' = (attributes + classes) / 2 a= (8 (number of attributes +6)/2=7
learningRate	The amount the weights are updated.	0.3
10-fold cross-validation		

Table 4. 21. Parameters and their initial/default values of the neural network classifier

Despite the different neural net models built based on the experiments by changing the parameters' values, only the model that has showed the best result in classification accuracy on the 10-fold cross-validation and 80-20 training and testing model are described here. For full list of the neural networks' different models experimented in this research refer **Appendix 4.**

The model with the learning rate and number of hidden layer nodes of 0.5 and 8 respectively and 10-fold cross-validation, has resulted in a 98.62 % overall classification accuracy. In other words, out of the 10,090 dataset given, it has correctly classified 9951 of them. The rest 1.38 % (139) records are placed in wrong classes. For example, out of 746 high revenue generating customers, the model has correctly classified 725 (97.18 %) in their correct cluster, i.e. cluster 1, and wrongly

Consistent with the domain experts' parameters, the decision tree and neural net models are evaluated. Finally the best classification model is chosen. The comparison is presented in terms of accuracy in **Table 4.22**.

Classification Model (10-fold classification model)	Overall Accuracy (10,090 Records)		High Value customers, Cluster 1, accuracy (746 records)		Low Value customers, Cluster 3, accuracy (4871 records)	
	Correctly Classified	Wrongly Classified	Correctly Classified	Wrongly Classified	Correctly Classified	Wrongly Classified
Decision Tree Model	9986 (98.97 %)	104 (1.03 %)	734 (98.39 %)	12 (1.61 %)	4849 (99.55 %)	22 (0.45 %)
Neural Net Model	9951 (98.62 %)	139 (1.38 %)	725 (97.18 %)	21 (2.82 %)	3,4821 (98.97 %)	50 (1.03 %)

Table 4. 22. Summary of the accuracy level of the decision tree and neural net classification models

As can be clearly shown in the table above, the overall performance level of the decision tree model (98.97 %) exceeds the accuracy level of the neural net model (98.62 %). The decision tree wrongly classified only 104 (1.03 %) of the whole dataset, while the neural net misclassified 139 (1.38 %) of the dataset (10,090). Again with regard to cluster 1, high value customers, classification accuracy, the decision tree model is better/accurate than the neural net. As can be seen from the **Table 4.22.**, the accuracy level of classifying high value customers in their cluster label, cluster 1, of the decision tree is 98.39 %, which is again far better than that of the neural net model that has an accuracy of 97.18 %. Wrongly classified high-value customers are small in decision tree model, which is 1.61 %, than the neural net model, which is 2.82 %.

Furthermore, the J48 decision tree model is better in correctly classifying low-value customers than neural net algorithm. The decision tree model has an accuracy of 99.55 % in correctly classifying low-value customers in cluster 3; whereas the neural net model has only 98.97 % accuracy. In

other words, the misclassification error of the former is 0.45 %, which is less than the later model (1.03 %).

The misclassifications of the models may be the result of the substitution of outliers' data by their fields mean values in the course of clusters model building. Again this can be attributed to the fact that some customers' categories extremely have very high values of attributes, which are used for the model building. Besides 6000 records, which accounts for more than 60 % of the 10,090 dataset, are residential customers' category. Consequently, these values might have dominated other values. This could have contributed for the misclassification errors.

Based on the above comparisons, the J48 decision tree has shown better performance. Hence, it is really plausible to conclude that the decision tree model is the best classifier model for CRM applications.

4.5.4. Evaluation

The core of any data mining task is data. However, unfortunately often than not data required for effective mining is not readily available in a format that the data mining algorithm required it. Besides, these data are usually stored in different format and heterogeneous databases. To make things worse some of the fields may contain outliers, missing values, inconsistent data types with in a single field and many other possible anomalies. But this must be cleansed, integrated and transformed in a format suitable for the data mining task to be undertaken.

Accordingly the researcher has taken considerable time for the data preprocessing task. Data cleaning (handling missing values, and outlier detection and removal), and data integration and transformation tasks are carried out in a format suitable for the clustering and classification. The data preprocessing phase culminated with the selection of the final attributes of dataset for cluster model building.

Next to the data preprocessing followed the building of models. The model-building phase is divided into two. The first phase is comprised of building a clustering model using the K-means algorithm. This task, which is iteratively conducted, resulted in various cluster models that segmented customers. Cluster models that are formed with the K values of 5 and 6 have shown better segmentation of customers. Particularly, the model with K=6 has differentiated the value of customers into high and medium even customers within the same category. For example, big business customers are segmented into high and medium revenue contributing customers. Consequently, this cluster model is chosen by domain experts.

The classification models are built with the weka's J48 decision tree and multilayerperceptron neural net algorithms. First different models of the decision tree have been carried out with 10-fold cross-validation and splitting of the dataset into 80 % for model training and 20 % for testing or validation. Similarly in the second experimentation of the neural net, the input variables to multilayerperceptron algorithm, weka implementation of the feed forward back propagation algorithm, are normalized within the range [-1, 1]. After the preparation of the dataset, different models of the neural nets have been developed and the best one is chosen. Finally, the decision and neural net models are compared based on the three parameters, overall classification accuracy, accuracy in classifying high value customers, and accuracy in classifying low value customers. The decision tree model scored better in all the parameters than the neural net model. Hence, it is selected as the best classification model for this research. Its overall classification accuracy is 98.97 %. The detail of the comparison among the models of the neural network and decision tree is presented in section 4.5.3.

However the classification accuracy of the neural network is very near to the decision tree model. Hence, one can easily guess that by changing some parameters of the algorithm with many trails,

the accuracy can significantly be improved and even may result in better accuracy than the decision tree algorithm.

4.5.5. Model Deployment

There is the concept of customer segmentation at ETC. As stated in section 2.2.3 of chapter two, the corporation has already setup a division, Corporate Customers Support Division (CCSD), which identifies, manages, and delivers special services to what they call “corporate customers”. Basically these customers are identified on the premises that if they are big organizations such as governmental offices, big business category customers, NGOs, International organizations and others, they may generate high revenue to the corporation. This is not based on the customers calling or services usage behavior. However, the result of this research proved that behavioral segmentation is possible and is best to identify and create a long lasting relationship with customers. Hence, this model can be deployed with further modification and evaluation to identify and provide special services to corporate customers. It can also enable to identify potential market place to CDMA telephone service after analyzing further the segments investigated in this research. In addition to these, the model can have a paramount help in the marketing strategy formulation.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

This research attempted to study the application of data mining techniques, clustering and classification techniques, to support CRM at ETC. It has been conducted particularly on the CDMA telephone services of the corporation. The service has been introduced in the country in 2004. Initially it was implemented with the 50, 000 exchange capacity for testing the technology. Currently, the corporation has already expanded the exchange capacity to 2.4 million. However, it has faced with the problem of marketing this service to customers as the mobile telephone has dominated the market. To this end the theme of this research is two fold. In one hand, it is intended to thoroughly analyze the CDMA's CDR along with the bill information and customers' profile data so as to investigate different segments of customers. This in turn enables the corporation to see which type of customers intensively use the CDMA telephone thereby find possible market for the same. On the other hand, it is destined to investigate high, medium, low value customers based on their calling behavior and profitability (revenue). The identification of these different types of customers enables the CCSD of ETC to create good customer relationship management while delivering special services to high value customers, called "corporate customers."

Accordingly, to achieve the abovementioned goals and the specific objectives of this research, related literature reviews in the area of data mining, CRM and customer segmentation have been thoroughly carried out. In addition to this, CRISP-DM process model has strictly been followed while undertaking the experimentation. CRISP-DM process model embraces business

understanding, data understanding, Data preparation, modeling, evaluation, and deployment phases.

The data, CDR and Customers' profiles, used in this research has been gathered from the corporation's database, "USHACOM". The processing to change this data into a format suitable for the data mining tasks to be performed for this research has been carried out. This phase took considerable time of the research.

Next to the data preparation and preprocessing, different clustering models are built using the K-means algorithm with K values of 4, 5, and 6. These values of K are chosen and experimented after a thorough discussion with domain experts. The variables/attributes are selected with the help of the J48 decision tree algorithm and those that are judged relevant by domain experts. The clustering models built at 4, 5, and 6 values of K are evaluated and interpreted. Among these three models, a model at K value of 6 has showed overall better segmentation of customers. It could able to create dissimilar clusters of high, medium, and low value of customers. In addition to this, the model has appropriately segregated high and medium values of big business category customers. The model also identified government category customers as the highest value customers. This result complies with the corporation assumptions that Government organizations are by default corporate customers. Finally this model has been accepted and approved by the domain experts.

The classification models are experimented with the J48 decision tree and multilayerperceptron neural net algorithms. Different models for each of these algorithms are built. The best overall accurate classifier models from the decision tree and neural net models are selected. Finally, the overall classification accuracy, accuracy in classifying high value customers, and accuracy in classifying low value customers with minimum error parameters have been defined together with domain experts and used to compare the results of the best models from decision tree and neural

nets. Accordingly, the decision tree model has scored better performance in all these evaluation parameters.

The result of this study is really encouraging as it could able to differentiate the high, medium, and low value customers properly. In addition to this, the classification model has showed best classification accuracy of 98.97 % with the decision tree model. Hence, it is the researcher belief that the model built in this study will contribute a lot for better CRM and marketing strategy formulation, if it is going to be deployed by the corporation with further improvements and modifications. This research has proven that data mining is applicable in the area of CRM.

In the course of undertaking this research, considerable time of the research has been spent on the data preparation as ETC does not have data warehouse. In addition to this, a lot of time is also spent to consult the domain experts on the appropriateness of the models, which are built for clustering and classification tasks of data mining. Again the ANN's models are built by changing only the multilayerperceptron algorithm's number of hidden layer nodes and learning rates. However, other parameters such as number of iteration or learning time of the algorithm have not been tried.

5.2. Recommendations

Based on the findings of this study, the researcher makes the following recommendations.

Develop Corporate Data Warehouse

Data is at the heart of any data mining task. The data mining algorithms' performance and scalability in uncovering interesting patterns, which are previously unknown knowledge, rest up on the quality of data. To this end, the data must be cleansed, integrated and transformed into a format suitable for the data mining task (s) to be undertaken. Nevertheless, the data preprocessing/preparation usually take much of the time for a data mining research or project. This time can be shortened if the data warehouse has been established before hand.

Unfortunately, ETC does not have a data warehouse. Due to this reason, much of the research time has been spent on the data preprocessing/preparation. The researcher strongly recommends ETC to develop customers' data warehouse, which may contain the CDR and the customers' profiles. Here, it should be noticed that the development of the data warehouse not only useful for data mining purposes, but also for other statistical analysis such as OLAP and OLTP.

Integration of Knowledge-Based Systems

The result and relevance of any data mining model is usually judged and evaluated by the human domain experts. It is time consuming, laborious, and tedious processes in consulting the human domain experts. Moreover, the chance of correctly evaluating the model highly depends on a domain expert's level of understanding, knowledge and experiences in the domain in which the data mining tasks are carried out.

Considerable time of this study has been spent for intensively interacting with the domain experts to clarify the business understanding, and the models evaluation. It would have been easy and less time consuming task, if an expert system that automatically evaluates the worthiness of the models generated were in place. Hence, there is a need to integrate the expert system with the data mining tools.

Model Performance Improvement

Although this study resulted in better classification accuracy, 98.97 %, with the J48 decision tree algorithm, the neural network model has shown very near performance in class label prediction accuracy, which is 98.62 %. In this study, only learning rate and hidden nodes of multilayerperceptron neural network algorithm's parameters are tried at different values to improve the classification accuracy of the resulting model. Despite the changing of these parameters in the quest for better neural network classification model, its overall accuracy could not exceed the decision tree model. Hence, the J48 decision tree model is chosen as better classifier for CRM area. Nevertheless this generalization and conclusion should be substantiated with further researches by changing other parameters such as momentum, number of iteration (training time), validation set size, and validation threshold parameters of multilayerperceptron of the neural network algorithm. To this end, experimenting with the different values of these parameters may result in better classification accuracy.

Data mining requires large dataset both for training and testing. Hence, increasing the dataset size may yield a better clustering and segmentation models. In addition to this, further attributes that describe the service usage behavior of a customer can be derived to generate a better clustering and classification models. Again other clustering algorithms such as hierarchical, density-based, grid-

based, and model-based and classification models such as SVM (Support Vector Machine) needs to be integrated to come up with a practical system.

CRM and Data Mining Awareness

CRM is customer-centric approach of managing the interaction with customers. Individuals at different levels in the organization should be aware of the preconditions towards creating a long lasting relationship with customers. In addition to this, data mining is at the heart of CRM. Hence, it should be integrated in an attempt to identify, manage, and create relationship with profitable (big spenders) customers. To utilize data mining capabilities, awareness on its merits should be created among employees of the corporation. On the top of this, knowledge of data mining techniques, customers' value, and the corporation's business processes should be integrated to successfully implement CRM.

Other Areas Amenable For Data Mining Research in ETC

ETC generates and stores very large amount of data such as the network switches' alarms, which are automatically generated and stored in the log files when network errors or devices failure occur. To this end, Data mining can potentially be applied to investigate meaningful patterns of the same. This discovered pattern or knowledge can in turn be used by the corporation to easily pinpoint, troubleshoot, and repair the network and different devices in it.

The other important data maintained in the corporation and is amenable for data mining research is the web log files. These files are generated and stored in the logs when internet subscribers are connected to the World Wide Web. They maintain information about the web browsing habit/behavior of customers. These files can be mined to identify the customers' web sites look-and-feel preferences and web surfing habits. This may be useful to tailor the websites design

according to the customers' needs and preferences. In addition to this, the knowledge is useful to determine where and when to expand internet infrastructure facilities.

Other Organizations

CRM encompasses many functions. Customer segmentation, customer churn analysis, cross-selling, and the like are among the well known ones. In this research, only the customer segmentation has been carried out. To this end, data mining applicability on the rest functions of CRM can be further researched in other organizations, which highly interact with customers and have strong competitors, such as Banking, Insurance, Retail, Hospitality, and the like industries.

References:

- [1] ALBERTA efuturecentre n.d., Customer Relationship Management, viewed 11 November 2008,
< http://www.e-future.ca/pdf/efc_customer_relationship_management.pdf>
- [2] Berry, M. & Linoff G 1997. Data Mining Techniques for Marketing Sales and Customer Support. New York: John Wiley and Sons.
- [3] Berry, M. & Linoff G 2004, Data mining techniques for marketing, sales and customer relationship management, 2nd edn, Wiley Publishing, Inc., Indianapolis, Indiana.
- [4] Bigus, Joseph P 1996. Data Mining with Neural Networks: Solving Business Problems for Application Development to Decision Support. New York: McGraw-Hill.
- [5] Bose, R. 2002. Key Components for Information Technology Success, Industrial Management and Data System 102/2, 89-97. Available At URL: <http://www.emeraldinsight.com>
- [6] Bounsaythip, C. & Rinta-Runsala, E. 2001, Overview of data mining for Customer behavior modeling, Version 1, VTT Information Technology, viewed 6 July, 2008,
< <http://www.vtt.fi/inf/julkaisut/muut/2001/customerprofiling.pdf>>
- [8] Breu C., Meckl N. & Sametinger J. n.d., Project-Based Customer Relationship Management in Virtual Enterprises, D-93040 Regensburg, Germany, viewed 6 November 2008,
< <http://www.swe.uni-linz.ac.at/publications/pdf/tr-se-02.02.pdf> >
- [9] Bull, C. 2003, 'Strategic Issues in Customer relationship management implementation', Business Process management journal, vol. 9, no. 5.

- [10] Chen, R., Wu, R. & Chen J. Y. 2005, 'Data Mining Application in Customer Relationship Management of Credit Card Business', Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC'05) 0730-3157/05, Institute of Information Management, National Chiao Tung University, Taiwan, viewed 6 November 2008, <<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01508080>>
- [11] Clarkson, T. n.d., Applications of Neural Networks in Telecommunications, King's College London.
- [12] COMSOL AG n.d., Application of data mining technology in key business areas, viewed 6 November 2008, <http://www.comsol.ch/comsol/pub/media/pdf/Angoss%20Applications_of_Data_Mining_in_Key_Businesses.pdf>
- [13] CRISP-DM. 2000, CRISP 1.0: Step-by-Step data mining guide, viewed 6 November 2008, <<http://www.crisp-dm.org>>
- [14] Deneke, A. 2004, 'Application of data mining to support customer relationship management (CRM) at Ethiopian Airlines', Department of Information Science, M.Sc. Thesis, Addis Ababa University.
- [15] Edelstein, H. (president) n.d., Building profitable customer relationships with data mining, Two Crows Corporation, viewed 23 July 2008, <<http://www.twocrows.com/crm-dm.pdf>>
- [16] ETC 2008, 'Ethiopian Millennium: Special Issue', *ETC Magazine*, September 2008, Corporate Communication Division.
- [17] ETC 2008, 'Ethiopian Telecommunications Corporation Company Profile', *ETC Magazine*, Corporate Communication Division.
- [18] ETC n.d., *Special support to be provided to corporate customers*, Author, Fixed Line Strategic Business Unit.

- [19] ETC Online Report n.d. available at URL: <http://www.ethionet.et.net>
- [20] Fayyad, U, Piatetsky-Shapiro, G, & Smyth, P 1996, 'From data mining to knowledge discovery in databases', AI Magazine, 17(3), pp. 37-54,, Viewed 30 July 2008,
< <http://www.daedalus.es/fileadmin/daedalus/doc/MineriaDeDatos/fayyad96.pdf>>
- [21] Fekadu, M. 2004, 'Application of data mining techniques to support customer relationship management at Ethiopian Telecommunications Corporation', Department of Information Science, M.Sc. Thesis, Addis Ababa University.
- [22] Forcht, K.A & Cochran, K 1999, Using data mining and data warehousing techniques, Industrial management and data systems, pp.189-196, viewed 20 September 2008,
<<http://www.emeraledinsight.com>>
- [23] Han, J, & Kamber, M 2001, Data Mining: concepts and technique, Morgan Kaufmann Publishers, San Francisco.
- [24] Hand, D, Mannila, H, & Smyth, P 2001. Principles of Data Mining. MIT Press.
- [25] Hencok, W. 2002, 'Application of data mining techniques to support customer relationship management at Ethiopian Airlines, Department of Information Science, M.Sc. Thesis, Addis Ababa University.
- [26] Jansen, S.M.H. 2007, 'Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior: A Vodafone Case Study', Department of Mathematics of University of Maastricht (UM) and at the Department of Information Management of Vodafone Maastricht, Master Thesis, UM, viewed 8 October,2008,
<http://www.math.unimaas.nl/personal/ronaldw/PhDMaBa-teaching/GraduationStudents/StephanJansen2007/Stephan_Jansen2007.pdf>

- [27] Jember, G. 2005, 'Application of Data Mining to support Fraud Detection on mobile- the case of Ethio-Mobile', Department of Information Science, M.Sc. Thesis, Addis Ababa University.
- [28] Kelly, S. 2002, 'Mining data to discover customer segments', HENRY STEWART PUBLICATIONS 1463-5178, Interactive Marketing, vol.4, no. 3, PP 235-242, viewed 9 15 October 2008, < <http://www.seankellyassociates.com/PDDocs/Kelly.pdf>>
- [29] Kim, J., Suh, E. & Hwang, H. 2003, 'A model for evaluating the effectiveness of customer relationship management using the balanced scorecard', Journal of Interacting Marketing, vol. 17, no. 2.
- [30] Kotler, P. 1998. Marketing Management: Analysis, Planning, Implementation and Control. 9th edn. New Delhi, Prentice Hall of India.
- [31] Mckinsey Marketing Solutions (MKMS) 2001, Tactical CRM: Three step profits, not Data, viewed 14 November 2008, <http://www.mckinsey.com/practices/marketing/ourknowledge/pdf/Solutions_TacticalCRM.pdf>
- [32] Netz, A., Chaudhuri, S, Bernhardt, J, & Fayyad, U 2000, 'Integration of Data Mining and Relational Databases', Proceedings of the 26th International Conference on Very Large Databases, Cairo, Egypt, pp. 719-722, viewed 2 November 2008, <<http://www.tcs-trddc.com/tecs/integration-of-data-mining.pdf>>
- [33] Palace, B.1996, Data Mining, Technology Note prepared for Management 274A, Anderson Graduate School of Management at UCLA, viewed 6 November 2008, <<http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>>
- [34] Parvatlyar, A & Sheth, N.J. 2002, Customer relationship management emerging practice, process and discipline.
- [35] Peter, C, Julian, C, Randy, K, Thomas, K, Thomas, R, Colin, S, & Rudiger, W 2000, CRISP-DM Step-by-Step Data Mining Guide, The CRISP-DM consortium, viewed 4, September, 2008, <<http://www.crisp-dm.org/>>

Appendix 2. The Weka 3.5.8 Interface with the Dataset Opened to Start the First Clustering Run

The screenshot shows the Weka 3.5.8 Explorer window. The 'Preprocess' tab is active. The 'Filter' section shows 'Choose' selected. The 'Current relation' is 'cdma_cdi_user_pro 241208-weka.filters.unsupervised.attribute.Remove-R1' with 22798 instances and 32 attributes. The 'Attributes' list on the left includes 'Numberofcalls' (1), 'TNBOM' (2), 'DNWghtCall' (3), 'DCEOM' (4), 'DCallMobile' (5), 'TNCFixedAA' (6), 'TNCMobile' (7), 'DCMOM' (8), 'DCFixedAA' (9), 'TNSCal' (10), 'TNCWLOC' (11), 'TNCFixedReg' (12), 'TNCMOM' (13), 'TNCWeekday' (14), 'TRRevenue' (15), 'RperCall' (16), 'RperDuration' (17), and 'DperCall' (18). The 'Selected attribute' section shows 'Name: TNumberofcall', 'Missing: 0 (0%)', 'Distinct: 2232', and 'Type: Numeric Unique: 746 (3%)'. A table of statistics for 'TNumberofcall' is displayed:

Statistic	Value
Minimum	1
Maximum	24606
Mean	344.446
StdDev	713.65

The 'Class: RENT (Num)' is selected, and a visualization of the data is shown at the bottom right. The status bar at the bottom indicates 'Status OK'.

Appendix 3. Sample of the Decision Tree Generated With 10-Fold Cross-Validation Technique

J48 pruned tree

```
RENT <= 8
| TNumberofcall <= 911
| | DperCall <= 167.591837
| | | TRevenue <= 407.544
| | | | RperCall <= 1.272093: cluster2 (4471.0/1.0)
| | | | RperCall > 1.272093
| | | | | DperCall <= 159.06
| | | | | RperCall <= 2.113439: cluster2 (152.0/1.0)
| | | | | RperCall > 2.113439
| | | | | | TNumberofcall <= 8: cluster2 (3.0)
| | | | | | TNumberofcall > 8: cluster3 (5.0)
| | | | | | DperCall > 159.06
| | | | | | RperCall <= 1.378073
| | | | | | DperCall <= 162.278107: cluster2 (3.0)
| | | | | | DperCall > 162.278107: cluster3 (2.0)
| | | | | | RperCall > 1.378073: cluster3 (20.0)
| | | | TRevenue > 407.544
| | | | | TNumberofcall <= 809
| | | | | RperCall <= 1.99569
| | | | | | DCallMobile <= 52946: cluster2 (54.0/1.0)
| | | | | | DCallMobile > 52946
| | | | | | | TNumberofcall <= 644: cluster2 (5.0/1.0)
| | | | | | | TNumberofcall > 644: cluster5 (6.0/1.0)
| | | | | | | RperCall > 1.99569: cluster3 (3.0)
| | | | | | | TNumberofcall > 809: cluster5 (13.0)
| | | | DperCall > 167.591837
| | | | | DperCall <= 198.620939
| | | | | RperCall <= 0.877009
| | | | | | DperCall <= 182.876728: cluster2 (113.0)
| | | | | | DperCall > 182.876728
| | | | | | | RperCall <= 0.564071
| | | | | | | DperCall <= 193.704403: cluster2 (40.0)
| | | | | | | DperCall > 193.704403
| | | | | | | | RperCall <= 0.308105: cluster2 (4.0)
| | | | | | | | RperCall > 0.308105
| | | | | | | | | DCallMobile <= 3994: cluster3 (5.0)
| | | | | | | | | DCallMobile > 3994: cluster2 (3.0/1.0)
| | | | | | | | RperCall > 0.564071
| | | | | | | | DperCall <= 186.894737
| | | | | | | | RperCall <= 0.710783: cluster2 (6.0)
| | | | | | | | RperCall > 0.710783: cluster3 (8.0)
```

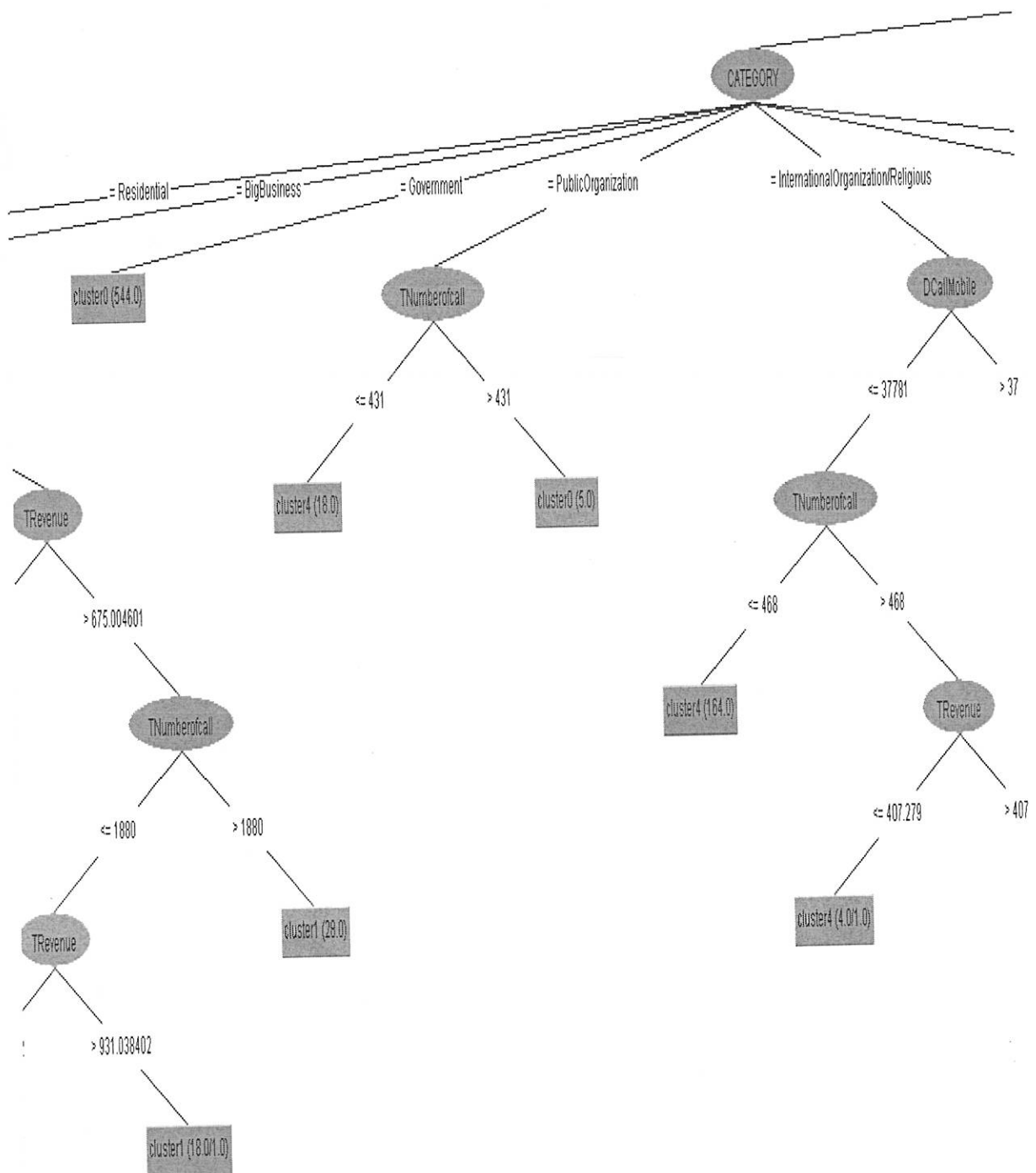
Appendix 4. Summary of the ANN's multilayerperceptron algorithm classification models at different values of hidden layer nodes and learning rate

Neural Network Parameters	Records count in each cluster	Clusters						Total
		Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	
Learning Rate =0.3 # Nodes in the Hidden Layers= 7 10-fold cross validation	# Records in the cluster (A)	746	416	4871	716	2769	572	10090
	# Records correctly Classified (B)	695	395	4836	689	2747	562	9924
	# Records Wrongly Classified (C=(A-B))	51	21	35	27	22	10	166
	% Correctly Classified (accuracy) (D=(B/A)*100)	93.16 %	94.95 %	99.28 %	96.23 %	99.21 %	98.25 %	98.35 %
	% of Wrongly Classified instances (E=(C/A)*100)	6.84 %	5.05 %	0.72 %	3.77 %	0.79 %	1.75 %	1.65 %
Learning Rate =0.4 # Nodes in the Hidden Layers= 7 10-fold cross validation	# Records in the cluster (A)	746	416	4871	716	2769	572	10090
	# Records correctly Classified (B)	721	392	4808	694	2753	563	9931
	# Records Wrongly Classified (C=(A-B))	25	24	63	22	16	9	159
	% Correctly Classified (accuracy) (D=(B/A)*100)	96.65 %	94.23 %	98.71 %	96.93 %	99.42 %	98.43 %	98.42 %
	% of Wrongly Classified instances (E=(C/A)*100)	3.35 %	5.77 %	1.29 %	3.07 %	0.58 %	1.57 %	1.58 %
Learning Rate =0.5 # Nodes in the Hidden Layers= 7 10-fold cross validation	# Records in the cluster (A)	746	416	4871	716	2769	572	10090
	# Records correctly Classified (B)	725	399	4826	693	2737	563	9943
	# Records Wrongly Classified (C=(A-B))	21	17	45	23	32	9	147
	% Correctly Classified (accuracy) (D=(B/A)*100)	97.18 %	95.91 %	99.08 %	96.79 %	98.84 %	98.43 %	98.54 %
	% of Wrongly Classified instances (E=(C/A)*100)	2.82 %	4.09 %	0.92 %	3.21 %	1.16 %	1.57 %	1.46 %
Learning Rate =0.6 # Nodes in the Hidden Layers= 7 10-fold cross validation	# Records in the cluster (A)	746	416	4871	716	2769	572	10090
	# Records correctly Classified (B)	728	394	4805	702	2747	557	9933
	# Records Wrongly Classified (C=(A-B))	18	22	66	14	22	15	157
	% Correctly Classified (accuracy) (D=(B/A)*100)	97.59 %	94.71 %	98.65 %	98.04 %	99.21 %	97.38 %	98.44 %
	% of Wrongly Classified instances (E=(C/A)*100)	2.41 %	5.29 %	1.35 %	1.96 %	0.79 %	2.62 %	1.56 %
Learning Rate =0.2 # Nodes in the Hidden Layers= 7 10-fold cross validation	# Records in the cluster (A)	746	416	4871	716	2769	572	10090
	# Records correctly Classified (B)	723	386	4823	682	2742	557	9913
	# Records Wrongly Classified (C=(A-B))	23	30	48	34	27	15	177
	% Correctly Classified (accuracy) (D=(B/A)*100)	96.92 %	92.79 %	99.01 %	95.25 %	99.02 %	97.38 %	98.25 %
	% of Wrongly Classified instances (E=(C/A)*100)	3.08 %	7.21 %	0.99 %	4.75 %	0.98 %	2.62 %	1.75 %
Learning Rate =0.5	# Records in the cluster (A)	746	416	4871	716	2769	572	10090

# Nodes in the Hidden Layers= 8 10-fold cross validation	# Records correctly Classified (B)	725	399	4821	701	2745	560	9951
	# Records Wrongly Classified (C=(A-B))	21	17	50	15	24	12	139
	% Correctly Classified (accuracy) (D=(B/A)*100)	97.18 %	95.91 %	98.97 %	97.91 %	99.13 %	97.9 %	98.62 %
	% of Wrongly Classified instances (E=(C/A)*100)	2.82 %	4.09 %	1.03 %	2.09 %	0.87 %	2.1 %	1.38 %
Learning Rate =0.5 # Nodes in the Hidden Layers= 9 10-fold cross validation	# Records in the cluster (A)	746	416	4871	716	2769	572	10090
	# Records correctly Classified (B)	733	386	4822	697	2743	563	9944
	# Records Wrongly Classified (C=(A-B))	13	30	49	19	26	9	146
	% Correctly Classified (accuracy) (D=(B/A)*100)	98.26 %	92.79 %	98.99 %	97.35 %	99.06 %	98.43 %	98.55 %
	% of Wrongly Classified instances (E=(C/A)*100)	1.74 %	7.21 %	1.01 %	2.65 %	0.94 %	1.57 %	1.45 %
Learning Rate =0.5 # Nodes in the Hidden Layers= 8 80 % Training and 20 % Testing	# Records in the cluster (A)	145	81	990	134	551	117	2018
	# Records correctly Classified (B)	138	74	977	134	536	117	1976
	# Records Wrongly Classified (C=(A-B))	7	7	13	0	15	0	42
	% Correctly Classified (accuracy) (D=(B/A)*100)	95.17 %	91.36 %	98.69 %	100 %	97.28 %	100 %	97.92 %
	% of Wrongly Classified instances (E=(C/A)*100)	4.83 %	8.64 %	1.31 %	0 %	2.72 %	0 %	2.08 %

Appendix 5. The decision tree and the rules derived from the tree

A. The Partial View of the J48 Decision Tree Model

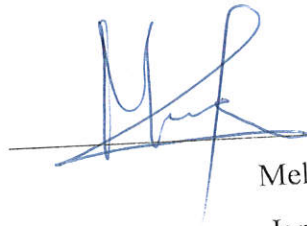


B. Sample Rules to Predict New Instances of Records Customers into Their Corresponding Cluster

- Rule # 1: If Rent ≤ 8
 Tnumberofcall less or equal to 911
 DperCall is less than or equal to 167.59
 TRevenue ≤ 407.544 and
 RperCall ≤ 1.27 , then the customer is classified in **cluster 3**.
- Rule # 2: If Rent ≤ 8
 Tnumberofcall > 911 and
 TRevenue > 307.922 , then a customer is classified into **cluster 6**.
- Rule # 3: If Rent ≤ 8
 $911 < \text{Tnumberofcall} \leq 976$ and
 TRevenue ≤ 307.922
 DperCall ≤ 103.33 , then a customer is classified into **cluster 3**.
- Rule # 4: If Rent > 8
 DCallMobile $\leq 81,064$ and
 Category is "Government", then a customer belongs to **cluster 1**.
- Rule # 5: If Rent > 8
 DCallMobile $\leq 81,064$ and
 Category is "Big Business",
 Tnumberofcall > 1297 ,
 TRevenue ≤ 675 , then a customer belongs to **cluster 5**.
- Rule # 6: If Rent > 8
 DCallMobile $\leq 81,064$ and
 Category is "Big Business",
 Tnumberofcall ≤ 1297 ,
 TRevenue ≤ 1104.78 , then a customer belongs to **cluster 5**.
- Rule # 7: If Rent > 8
 DCallMobile $\leq 81,064$ and
 Category is "Big Business",
 Tnumberofcall ≤ 1297 ,
 $1104.78 < \text{TRevenue} \leq 1785.19$, then a customer belongs to **cluster 4**.

DECLARATION

The thesis is my original, has not been presented for a degree in any other university and that all sources of material used for the thesis have been duly acknowledged.

A handwritten signature in blue ink, appearing to be 'Melaku Girma', written over a horizontal line.

Melaku Girma

January 2009

The thesis has been submitted for examination with my approval as university advisor.

A handwritten signature in blue ink, appearing to be 'Dr. Million Meshesha', written over a horizontal line.

Dr. Million Meshesha

January 2009