CHURN PREDICTION IN TELECOMMUNICATION INDUSTRY IN KENYA USING DECISION TREE – CASE STUDY ORANGE KENYA

BY

PATRICIA KEMUNTO NYAMBANE

KCA/13/01935

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN DATA COMMUNICATIONS

FACULTY OF COMPUTING & INFORMATION MANAGEMENT

KCA UNIVERSITY

OCTOBER, 2017

DECLARATION

I hereby declare that this Dissertation is my original work and has not been published or submitted elsewhere for the award of a degree. I also declare that this contains no material written or published by other people except where due reference is made and author duly acknowledge.

Student Name: **Patricia Kemunto Nyambane** Reg.No: **13/01935**

Sign ………………………………………………………. Date: …………………

I do hereby confirm that I have examined the master's dissertation of
Patricia Kemunto Nyambane

Sign: ……………………………………………………… Date: ……………………

Mr. Samuel Matende.

Dissertation Supervisor

ABSTRACT

Customer churn in the telecommunication industry is still a big problem because emerging new technologies, lower costs, among other factors. Right now Kenya is developing at a very fast rate and the market is getting new players leaving the only way to gain customers is by winning them over from the competitors. The retention of customers is becoming a huge challenge and the cost of acquiring new customers is more expensive, therefore by collecting information already available to the telecom industry can go a long way to helping them. The best way for the telecommunication industries to address this is to develop precise and reliable predictive models so as to identify potential churners by understanding their behavior and trends beforehand so as to introduce them to the programs suited to their needs in a bid to retain them. Orange Kenya was used as a case study to develop a predictive a model in an aim to reduce the customer churn rate. The objective of the study was to find the extent of the customer churn in Orange Kenya and make it obligatory for the telecommunication industry to do a churn prediction analysis. Also use the available resources to design a model for customer churn prediction for the pre-paid users. Questionnaires use was to find out customer's perspective. The project provides a framework for churn prediction model and implemented using data mining. The final results show the prediction rate for the model used.

ACKNOWLEDGEMENT

First and foremost I want to thank the Almighty God for the gift of life and for granting me this wonderful opportunity and capability to start and be able to finish this project successfully.

This work is part of the contribution towards the fulfillment of my master's degree which I render my humble gratitude to my supervisor, Mr. Samuel Matende who has supported me throughout my project with his patience and knowledge while I stalked him regularly with my many phone calls.

Writing this project has been an exercise in sustained patience which is thanks to my twin, who in the final steps of my project decided that at two years of age thought they knew the laptop better than their mother and decided to press everything and accidentally formatted it making me lose everything including the project and had to start again. If the law would have permitted it, you would have seen me enroll them direct into university to do this project.

Also can't forget my boss who gave me time to be able to do my research in the office and colleague who encouraged and cheered me on. I thank you from the bottom of my heart.

TABLE OF CONTENTS

Table of Content Page

List of tables

List of Figures

List of Figures Page

ACRONYMS

ARPU – Average Revenue per User

CAK – Communications Authority of Kenya

Weka - Waikato Environment for Knowledge Analysis

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

Major source of profit are customers, hence customer churn play a huge role in the survival of an industry and its development. Customers today have unlimited supply of information at their fingertips. For example smart phones enable faster access to products, price comparison and brands available at a cheaper price. With the technology advances in our country, an increase in competition in the market, introduction of different players in the market and customer demands it has become a challenge for telecommunication operators.

Statistics report by the Communications Authority of Kenya (CAK), "at the end of the 2016 quarter, mobile penetration stood at 88.1per cent with 37.8million subscribers up from 36.1 million in the previous quarter. It showed that pre-paid subscriptions continue to dominate the mobile telephony sector, registering 36.8 million subscribers, accounting for 97.3 per cent of the total subscriptions. Post-paid subscriptions saw a marginal increase to 989,889 up from 963,684 in the previous quarter. The number of Internet users grew to 31.9 million from 29.6 million in the previous quarter. As a result, the portion of the Kenyan population accessing Internet services reached 74.2 per 100 inhabitants up from 69.0 per 100 inhabitants recorded in the previous quarter. a significant increase in mobile subscriptions and Minutes of Usage in voice as a report from the period of April to June. The sector report for the last quarter of the financial year 2015/2016 indicates that, mobile subscription grew by 9.9 per cent to 39.7 million. The mobile penetration level hit 90 per cent, having grown by 6.1 percentage points. Total traffic originating from mobile networks also increased by 22.5 per cent during the financial year to post 41.6 billion minutes up from 34.0 billion minutes posted in the previous financial year." With these statistics we can see the telecommunication sector in Kenya has seen a big technology explosion with strong demographics and this is leading to a strong growth and investments

Also it's past time the telecommunication industries concentrate not only on adding more customers to their base they also have to think of customer retention. For them to be able to sustain growth in the industry they have to find a way of sustaining them. And also with this age of information where customers have it on the go, they are well informed of better prices and products from other firms that they might suit their needs. For the telecommunication industries to be able to handle these they need to be able to understand customer behavior so as to predict the future and prevent them from moving with the competing firms.

Customer churn therefore means switching of a customer from one provider to another over a given period of time. When an existing customer leaves a provider for another, he is referred to as voluntary churner while involuntary churner is one who the provider disconnects due to non-payment of costs like the okoa jahazi.

Churn management is a term used by the telecommunication companies to retain profitable customers like in our case the post paid customers. Kentrias (2001) in his study explains churn management in the telecommunication industry as the procedure of retaining the most important customers for the company. He also emphasized to predict how each customer will react to specific offers and predict which customers will be positively influenced.

With a practical and well operated churn management practice, it can solve issues and help the firm with increased average revenue per user (ARPU). ARPU in the telecommunication lingo is code for operator subscriber base. This is useful because it helps in measuring performance growth. ARPU level is able to withstand competition when churn decreases which brings an upward growth in the firm. Also the firm can use resources available at hand to prevent churn by using better packaged offers to the consumer. With these resources they can predict early enough the causes of churn and stop the consumer from migrating, basically proactive customer care. And have intelligent data analysis that are up-to-date awareness and have quick reaction times

For a firm to be able to achieve the above they have to entice their mobile subscribers by anticipating their customer's needs and also understanding them by having a business process tool with a quick reaction time on customer fundamentals and concentrate on churn prevention. The firm can also benefit by moving the churn management from a Business Intelligence platform to the daily performance.

Telecommunication can be thought of it as the world's biggest computer, connected together by complicated network, telephone lines, wireless mobile phones and internet linked PCs. It enables to connect with anyone anywhere in the world and also do business internationally or nationally. With the change in technology, telecommunication has become more than just voice but also text, images and videos. Over the years the telecommunication industry has moved from government monopoly to an oligopsony market which is privatized and face a flood of competitors in the market. These markets have been upset as the landlines are replaced by wireless and transfer of images and videos. With different players in the market it brings about customer loyalty. Customer churn is a huge issue facing this industry which is proving to be costly if not managed carefully. Costs like loss of revenue, customer retention, and advertising are due to customer churn. There are also organizational costs due to planning and budget chaos. Surveys across industries show that it is 5-7 times more costly to acquiring new customer compared to retaining the old ones. With churn management the firms can make timely offer to customers at the right time. This helps to increase customer loyalty to the firm.

Many telecommunications have used much of their resources in customer acquisition ignoring customer retention. Competition in the telecommunication industries makes them market more on wining more customers and making them lag behind on customer satisfaction of the already acquired customers. These show that retention strategies are not being given priority in the firm. Because management have not analyzed how much it would cost to maintain an existing loyal customer versus acquiring new customer. According to Orange's marketing department, they will delete a subscriber who has not been active in the system for six months. After that they will recycle the number and resell it to another new customer and while doing all this they are using money to repackage it. This cost can be avoided if they had a way or tool to understand the customer's needs and expectation so that they can provide service and products perfectly suited for him/her to prevent them from going to another firm. With no retention strategy in place a firm cannot be able to prevent customer churn. There are two types of churners, voluntary and involuntary. With voluntary the customer willingly leaves the firm and moves to another firm while in involuntary churn, the customer is asked by the firm to leave. Most churners fall in the voluntary churn because they are partial churners. They haven't really left the firm but they are using other network s apart from the one they are already using. This affects the ARPU and the profit margins of the firm but doesn't affect the shares in the market. Churn prediction helps with this loss by predicting future churners and understanding their behavior so that they can anticipate a customer's needs.

1.2 Statement of the Problem

For an industry to apply data mining it depends on two things; availability of data and the problems facing the firm. For the telecommunication company availability of data is not a problem. They maintain data of phone calls on the call detail records which all the information for each phone call made by the customers. Data from the call detail records is used mainly for marketing and fraud detection. They also store extensive detailed customer information e.g. personal information and also more information from other parties like the registration bureau (RB). This information is useful during data mining.

This industry faces many challenges since it produces and keep enormous amount of data which is related to the operation of their company thus one of the main challenge is scalability of the data for data mining. Another challenge is that data is mostly on transaction form and is not at well-formed level for data mining for example if one wants to mine call history details on the customer level, the raw data available is for an individual customer. Therefore it is a must to gather data to appropriate acceptable level before we mine the data.

The major challenge that Orange Kenya is facing is how to know the number of customers who have churned. With losing customers, it brings a lot of problems which even the managers don't want to admit they are losing. CCK on their January 2017 report called out Telekom Kenya for falsifying their report on the number of subscribers. On their report, Telkom Kenya had said they have 5.2 million customers as at June but upon investigation they had 2.9 million customers. By falsifying there report by not cleaning out the churners, it misled the investors and the authorities. This leads to a loss in revenue coming in from the investors.

1.3 Operational Definition

Customer churn means switching of a customer from one provider to another over a given period of time. When an existing customer leaves a provider for another, he is referred to as voluntary churner while involuntary churner is one who the provider disconnects due to non-payment of costs like the okoa jahazi.

1.4 Objectives

The main objective is to make it obligatory for the telecommunication industry to do a churn prediction analysis. Specifically the research sought to;

i. Identify the churn rate in Orange Kenya
ii. Identify the causes of customer churn
iii. Analyze the applications of data mining techniques to a telecommunication industry
iv. Design a model for customer churn prediction for the pre-paid users.

1.5 Justification of the Study

According to the Communications Authority's report of 2012/13 "an operator lost 28.6% of its customer base in half a year indicating the significance of the problem. The study would help the Orange Company to know what really makes the customers decide to move to another firm. With the predictive model they are able to know the customers behavior and know who is about to churn, find ways to provide the customer with tailor made products and service to prevent them from shifting and thus avoid wasting resources on gaining new customers.

The significance of the study to me as a researcher is to improve on what is already done, by adding more information and providing a study to be improved on in the future. Gain more insight on why a customer would choose to leave one service provider for another.

Other thing is help Orange Kenya have a churn prediction model so as to know the non-churners and the churners so as not to be caught pants down again by CCK when they are giving out the reports of the number of subscribers. And also to help fight customer churn.

Orange Kenya has a marketing incentive to attract new customers but still don't know how to retain the existing ones or to know the potential churners.

1.6 Motivation

Retaining one customer costs an organization from 5 to 10 times more than gaining a new one. Predictive models can provide correct identification of possible churners in the near future in order to provide a retention solution.

Another motivation is customer behavior; In the CA's report of 2012/13 "an operator lost 28.6% of its customer base in half a year indicating the significance of the problem. The study would help the Orange Company and other telecommunication firms to know what really makes the customers decide to move to another firm. With the predictive model they are able to know the customers behavior and know who is about to churn, find ways to provide the customer with tailor made products and service to prevent them from shifting and thus avoid wasting resources on gaining new customers.

Other thing is help Orange Kenya have a churn prediction model so as to know the non-churners and the churners

1.7 Scope of the Study

The study is limited to Orange customers in Kenya. Data collected was from a sample of 60 and both primary and secondary data was used. Population selected was Orange subscribers both subscribers and non-subscribers around Nairobi area.

1.8 Limitation of the Study

i. One of the major issues faced was confidentiality which prevented me to gain access to some of the customer information in their database like the billing data. Also this made me leave out some features when building the predictive model

ii. Customer demographic was a problem when doing the research

iii. Enough time and resources to properly conduct this study

CHAPTER TWO: LITERATURE REVIEW

2.1. Introduction

This chapter presents background of related works. It includes recent ant prominent publications of literature review of various papers, journals and articles in the field of churn prediction, data preparation and data mining techniques. The main focuses are the approaches and techniques applied to churn prediction in telecom businesses, as well as some other related works

2.2 Customer Concept

Today customers have and endless supply of information on their hand. For example a smart phone makes it possible to gain faster access. This makes it harder for the telecommunication industries to retain the current customers they already have. The concept of customer has changed over the years; decades ago that the customer needed the firm but nowadays things have changed. Firms need the customers for the production strategy. Today the market is more globalized and governments being less monopoly to liberated. With more customers on the market it brings about a fierce competition which the firms are facing. Kotler *et al.,* (2002) on their paper stated that the view point about consumers has changed from production to societal marketing concept period. And because of this the telecommunication companies need to take a leaf from Donio *et al (*2006) that it is cheaper to retain existing customers than enrolling new customers.

2.3. Customer Churn

Like I said earlier on this paper that telecommunication is like the biggest computer. Hadden *et el* (2007) goes on to state that the dominant medium worldwide is the mobile telecommunication over the last decade. Our market in Kenya has not yet reached saturation like the developed countries that have many competitors playing in the market. Because of the public regulations and standardization that the CAK authorized, allowed customers to easily migrate from one provider to another which leads to a very lymphatic market.

The cost of wining new customer out ways the cost of retaining existing and it is being witnessed that now the telecommunication industries are concentrating more on customer retention and moving from acquiring more customers (Fildes, 2002). Churn prediction has moved to be the first priority on the BI application platform, making itself useful in identifying the customers who are about to move to another carrier.

Song *et al* (2007) stated that a good churn prediction system should look at the future and further on in the future and I quote “good churn prediction systems should not only pinpoint potential churners successfully, but further provide a sufficiently long horizon forecast in its predictions.” On his research he detailed out that one the potential churner is identified, the firm has to call up the churner and get details on how the reason of his wanting to move and after that the retention department can take appropriate measures to provide an incentive to the customer or negotiate terms. And the advantage of the telecommunication having being able to predict they can forecast further on and anticipating the future needs of the customers and providing for them their needs is cheaper. Their limited allocation of resources to retention department and few churners are being followed up compared to the hundreds intending to move which does not make it an easy task for those working there. Like in the Orange Company, they have not yet created a retention department, all this it taken care of by the customer care department which is overwhelmed by customer complaints. If the firm cannot handle customer complaints and offer them good services they tend to abandon ship and move to another carrier.

Loyalty to a firm goes hand in hand with customer churn and also customer retention rate of a firm. Customer defection is an issue in the highly competitive wireless telecommunication industry (Hwang *et al.* 2004). On their research they pointed out that customer churn rate affects the long term value hugely because it affects the firms revenue and service length of a customer.

2.3.1. Causes Why Customers Churn

The telecommunication industries are battling with huge numbers of customer churn because of the intense competitive market. Customers are always bombarded with better offers from the competition and they face many reasons to leave their carrier. Geppert (2002) named so of the causes of churn;

i. Price; carriers offer pricing promotions like low monthly fees, high-volume offerings and low rates per minute and these price incentives can make provided permanent customers and attract other customers from the competitor.
ii. Service quality or network quality; most customers who have smart phones move to another carrier if they offer better coverage and better access to the network where ever they are.
iii. Fraud; customers can borrow airtime and fail to pay up and move to another carrier
iv. Lack or poor customer care (carrier responsiveness); slow or no response to customers complaints
v. Brand loyalty; issues may arise with a particular brand due to service or other issues experienced over time involving the incumbent carrier, or entry into the market of a new player with strong brand recognition and reputation.

vi. Privacy concern; Consumers are aware that companies they deal with have a lot of information about them, including their spending habits, personal financial information, health information, and the like.

vii. New technology introduced by the competitor, for example the introduction of 4G internet speed by Safaricom

viii. New players in the market; the existence of better competitors may cause certain disloyal customers to churn.

ix. Billing disputes; disputes about service disruptions can cause customers to switch carriers

2.3.2 Customer Churn in Banking and Insurance

The financial markets are also seeing a high competition in the market with many players saturated on the platform because of the new regulations that have changed over the years and also the changing nature of the sector. This leads to low profit margins on the sectors. There is a blurry line between the banks, insurance companies and brokerage firms because a firm can offer a large number of services venturing into all three areas. With all these pressure the companies are under, they have come to realize the importance of maintaining a loyal customer. Prinzie *et al,* (2005) on his paper on Decision Support Systems, said that, "the substantive relevance of attrition modeling comes from the fact that a bank is able to increase profits by 85% due to a 5% improvement in the retention rate." Glady, Baesens & Crou (2009) designed a churn rate model that used customer lifetime value established on classification technique based on a financial company. The churn prediction was on the basis of threshold value whereby an increase in the value may lead to an inconsequential churn factor.

2.3.3. Customer Churn in ISP

The internet is growing at a very fast rate. Reports from the CAK showed that number of internet users grew to 31.9 million from 29.6 million on the last quarter report. The report showed that the Kenyan population accessing the internet reached 74.2 per 100n inhabitants. With these numbers, customer has a choice of choosing an ISP provider. ISPs experience a higher churn rate which is 10% monthly (Minguel, 2005). On the journal, Au *et al,* (2003) commented that nearly half of all internet subscribers leave their providers every year. Madden, *et al* (1999) developed a probability model for ISP subscriber churn. Their model laid it on the line the probability of subscribers churn to various service attributes and subscriber characteristics.

2.4. Effects of Churn

Customer churn is a real-time challenge in the firm and to prevent it is crucial to the survival of it. For them to remain viable in the market they need to hold on to the customers they have. So preventing them from shifting to the competitor would cost them less because the cost estimate for acquiring a new customer is much costly, and things like advertising, marketing and technical support have to be taken into consideration. When we look at what it entails to retain an existing customer is just to call them up and knowing what they would be done to prevent them from shifting (Berson *et al* 2002). The telecommunication companies have to find a way to be able to predict churn behavior of the customer and come up with proactive actions. The solution also to all these is for the company to have a data analytics system that can use real-time integration and dynamic real time responses which can detect churn risks and enhance revenue opportunities and they can see a reduction in the churn rate and can increase the ARPU.

Another way that a firm can benefit from reduced churn rate is through advanced network analytics which help with the providers differentiate themselves with other subscribers. With advanced network analytics it enables a subscriber to be empowered and to provide improved customer experience which in returns helps with the firms' revenue and both are satisfied, the firm and the customer.

When the telecommunication companies deactivate and disconnection a customer, take like the example of Orange where they do that when a customer is not active for six months, they are in risk of revenue and margin deterioration.

Ahn *et al*, (2006) comments that the potential impacts on profitability that come from inactive, underutilized and unprofitable network facilities must be considered.

2.5. Data Warehousing Concept in Relation to Churn Prediction

Three main factors that affect the vitality and have converged to propel the telecommunication into a competitive industry are rapid growth of modern technology, market (mostly user) demand and competition and these factors create new technologies and products, which open a series of options and offers to customers so as to satisfy their needs and requirements (Bose and Chen, 2009). As we discussed earlier, the customer plays a crucial role in commercial companies and the telecommunication companies. In a market place where the customer is fundamental, price, products and customer service are leaden with necessity. New services have emerged, users demanding ever-increasing quality and deregulations have made the markets open up to competition. What telecommunication companies have been doing in developing countries to set up a customer churn relationship management center that can contact customers who are about to churn and offer choice and respond to the individual needs of the customer making them be ahead of their competitor.

The main problem with churn is that customers don’t announce their intentions in advance (Kim and Yoon, 2009). It becomes the job of the company to come up with evidence to show that a potential churner before the customer finally migrates. Comparing ten years ago and know, it was a hopeless quest but nowadays with improved technology it is feasible. Today, effective business process enabled by technology can help reveal customer behavior patterns and aid in assessing the profitability of various customer segments, what is important to them, and how the carrier can build loyalty within the most valued customer sets (Baurdeau *et al*, 2005).

"Putting the customer at the center of the business is one of the key trends in the industry," said Stewart Meyer, telecommunications industry analyst for Micro Strategy. "The best way to do this is through a data warehouse." He also noted that, as in other industries, there is a need to integrate customer input from numerous channels, including call centers and the growing number of online transactions. And what makes telecommunication more advantageous than other companies is the outstanding sheer volume of data.

Data warehousing holds all the information coming through for an organization and creates a central archive that contains all important data from relevant systems and outside sources in a steady format to help in modeling. Ahn *et al*, (2006) comments that by having all such data available simultaneously and uniformly, analysts can uncover relationships between customer characteristics, customer value, and churn likelihood. With data warehousing it presents a platform to turn a gold mine of raw data into valuable information through data mining and data visualization.

Hung *et al* (2006) defined data mining as the process of combing through a large number of data to revile the relationship and patterns of a customer’s activity. On the development of customer profile and discovering the pattern from the customer’s history and there reason to churn, data mining helps to unearth all this information. Demographics, period of service, usage patterns and credit details are related variables that data mining tools use.

2.6. Data Mining Steps

The telecommunication basically is a utility type industry not that I am undermining it but for such an industry it has redefined competition into a whole new level which has given rise to many issues and situations that were not foreseen, uncommon and unimaginative. What the telecommunication had to consider previously was to satisfy the needs of their customers, understand their behavior, and predict their needs and design products and services tailored to their needs. With the industry's redefinition of competition, it has become fiercer as mobile penetration has reached its peak and still continuing. This has attracted new customers and the telecommunication industry has realized that it is easy to attain new customers but very difficult to maintain them and very easy to lose existing ones.

For an industry to methodically strategize data mining analysis, it has to follow certain steps. Şimşek Gürsoy (2010) gave the following as the steps in data;

Business understanding; this is the first phase in data mining and it includes determining business objectives, taking measure of current situation, set down data mining goals and develop a exploitable plan. Important information taken from here are the length of period the subscriber has been with the company and who are likely to churn and weighing whether to find ways to retain the customer or just let go. Therefore it is important to develop a very accurate churn model during the detention studies according to Simsek Gursoy (2010).

Data understanding; after the plan has been developed and objectives set, then data can be gathered. In this phase it includes data collection, data description, data is explored and verified for quality. Exploration of data can be done at the end of this phase. On this step it is important to know the data quality problems and discovering the similar subsets for formation of hypothesis from their characteristics. Tasks are gathering data, describing, exploring and verifying quality.

Data preparation; in this step data is selected, cleaned, design into specific order and then formatted in preparation of data modeling. An in-depth look at data exploration is looked at; more models used so as to see patterns based on business understanding. Tasks that are done in this phase are tables, record, attribute, transformation and elimination of data.

Modeling; first of all here a more in-depth understanding of the business is seen and after it is gained other details for modeling that are appropriate to the data type is applied. Mostly, tools such as visualization for plotting and cluster analysis for identifying which variables merge are some of the useful tools that can be used for the analysis. Also rule induction tool can be used to build an association rule. In this phase, there is always a repeat visit to the data preparation phase. Tasks involved in this phase are evaluating results, reviewing the process and determining the next step.

Evaluation; here model will be matched together with the business objectives from the first phase. This will bring an emergence of other things like pattern recognition. Peppard and Rylander (2006) warn though that the key objective is to determine if there is a business issue that has not been sufficiently looked at.

Deployment; this is the last step but not necessarily the final step because it is not the end of the project. The deployment stage can be generating a report or implementing a data mining process but all this depends on the requirement and most cases the outcome is determined by the customer and not the data analyst. Tasks include planning deployment, reporting final results and reviewing results.

The above steps are not fixed to a once-all-procedure; there is always a review of previous stages (back-tracking). Results obtained can be used by the CRM department to establish whether to maintain a potential churner or let go. And if it is to maintain what can they offer the customer to retain them.

Figure 2.1: an overview of the data mining steps

2.7. Data Mining Concept

Data mining is also known as knowledge discovery or knowledge discovery in databases (KDD), and is defined as “the process of exploration and analysis of large quantities of data in order to discover meaningful patterns and rules."(Berry & Linoff, 2004, p.7). The taxonomy of data mining is shown below in Figure 1 (MO Zan, ZHOA Shan, LI Li and LIU Ai-Jun, 2007)

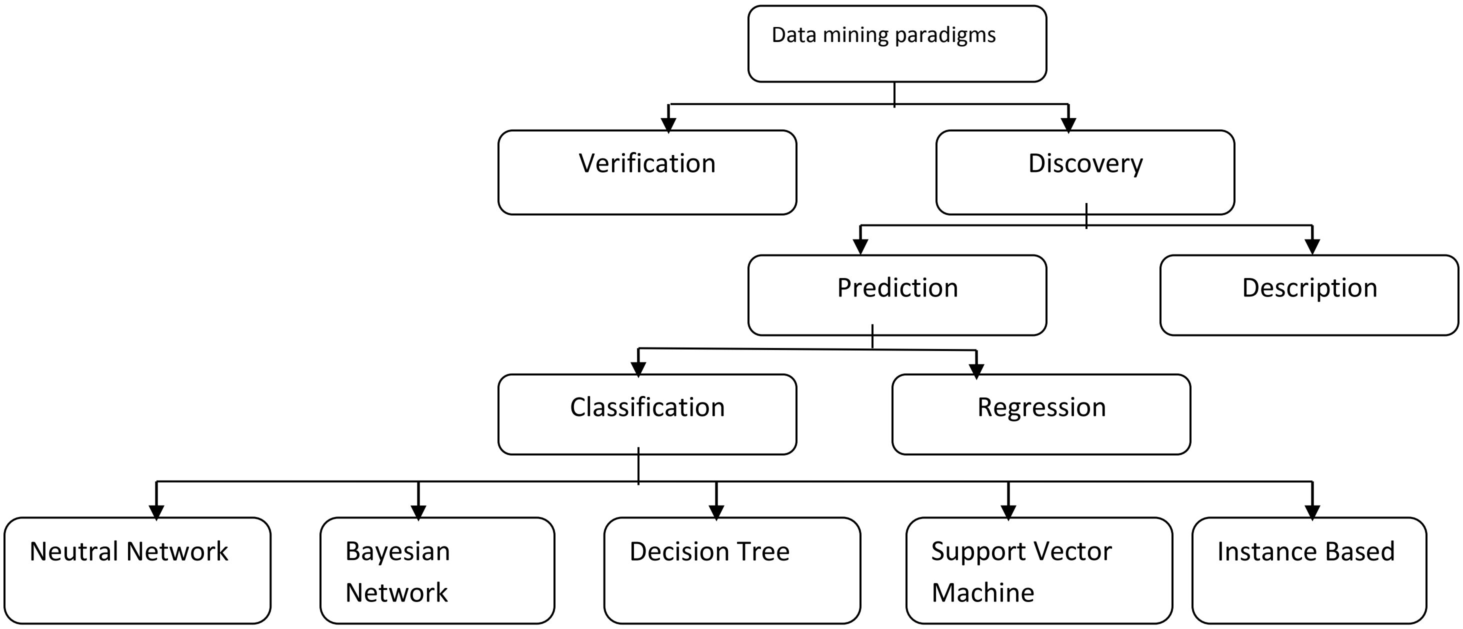

Figure 2.2: The data mining paradigms

Verification method is a traditional statistical method and deals with evaluation of a hypothesis proposed by an external source. It is less associated with data mining, it's more of discovery-oriented. And as we know data mining problems are more associated with selecting a hypothesis rather than testing. (MO Zan, ZHOA Shan, LI Li and LIU Ai-Jun, 2007). Discovery methods identify patterns in the data and it consists of prediction and description methods. With Description it focuses on understanding the way data operates and Prediction method aims to build a behavioral model that can get new and unseen samples so that it can able to predict one or more variables to the samples. Predefined groups are divided in the classification method into five branches; neutral networks, Naïve Bayesian networks, decision trees, support vector machine and instance based methods.

According to Witten and Frank (2005, p. 5), the data mining process must be automatic or (more usually) semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities.

2.7.1. Background of Data Mining Usage

The telecommunication industries in Kenya have a large amount of data in store but what mostly are learnt from those data is only presumptions. The most continuous use of these huge amounts of data is mostly for statistics and few for data mining. More factors have kept coming up that have made it possible for data mining to become prevalent and easily to adopt it. Increased growth in data collection, storing of data in warehouses, increase in access to data from the WEB and intranets, competitive pressure to increase market share, development of data mining software suites and growth in the computing power and storage capacity are some on the few factors.

The telecommunications industry was the first to adopt data mining techniques. These because telecommunication companies generate and store enormous amounts of high-quality data, have a very large customer base, and operate in a rapidly changing and highly competitive environment (Weiss, 2009).Most of data mining application has been used in the telecommunication industry but it has mostly been concentrated around three areas; fraud detection, marketing and network fault isolation and prediction.

Data mining for marketing strategies by the telecommunication has been used to identify and retain customers and maximize profit base. Han, Altman, Kumar, Mannila & Pregibon (2002) on their work did a market research and identified many small but well-connected sub graphs in the calling activity. This was one of the most famous use of data mining to acquire new customers, was called the MCI's Friends and Family Program. Getoor & Diehl (2005) was one of the early uses of data mining on social network and linking.

With fraud detection most researchers compared the customer's current calling behavior with his past usage deviation detection techniques and also anomaly detection techniques. In Cortes & Pregibon (2001) research they generated a signature from data streams of the call records to briefly describe the calling behavior of the customers. Then they used the anomaly detection to admensurate the eccentricity of a new call relative to a particular account. Recent work on fraud detection have used dynamic clustering together with deviation detection to detect fraud (Alves, 2006)

As the telecommunication networks become more complex monitoring and maintain them becomes a challenge. Weiss, Ros & Singhal (1998) developed an expert system to handle alarms created by the network elements. But the problem was the systems were expensive to develop and maintain them with the growing technological advancement so data mining applications came in. Data mining application used to aid with fault identification is the Telecommunication Alarm Sequence Analyzer (TASA) (Klemettinen, Mannila &Toivonen, 1999)

2.8. Data Mining Techniques

Data mining helps in churn analysis to predict when and why the customer will churn or not and the telecommunication companies can use it to reduce the churn rate in their organizations. There are two types of data mining techniques that are used in practice according to Coussement and Poel (2008)

Supervised learning which requires data sets that contain target variables which show the classes of data or behavior that will be predicted.

Unsupervised learning doesn't require the data sets to get the target variables like in the supervised learning. Example is clustering method that is used to calculate the natural structure by exploring the data sets.

2.8.1. Decision Tree

Decision tree is a well-known technique and has been used for; describing sequence of interested decisions or predicting future data trends in Berry and Linoff (2004), Chen, Hsu and Chou (2003), also to some real life world problems (Tsai & Chiou, 2009). Decision tree is a learning technique that details information from a trained dataset in an ordered system having the root nodes, internal nodes and leaf or terminal nodes. Since the decision tree can be shown in form of a tree, it is the easiest algorithm to understand the results gained. Lorena & Carvalho (2007) say that decision trees have the ability to build models using numerical and categorical datasets.

Osei-Bryson (2004) classifies decision tree into two types;

Classification trees; here each leaf nodes on the decision tree partners with a value for each class and the target variables get there values from a discrete domain. Examples include ID3, C4.5 and CART. They are mostly used to predict a categorical outcome.

Regression trees are used in cases where there will be a continuous outcome.

My research intends to use the already existing decision tree algorithm and does not intend to improve on it. I will use Weka (Witten & Frank, 2005) which is open-sourced and is well known.

With decision trees they are accurate for forecasting because they enable long-term forecasting and early detection of customer value loss.

Decision tree is a supervised model and therefore requires a labeled training set. The outcomes of the observation of ‘churn’ or ‘non-churn’ will be indicated by a 1 or 0 categorically. Optimal size will be achieved by over-fitting to get artifacts and noise in the dataset, but predictive power will be lost. So there will be pre-pruning and post-pruning. With oversampling, will alter the proportion in the training set

2.8.2. Neural Network

With neutral networks we have to go way back to the 1930’s and 1940’s even before the existence of digital computers. They were viewed as biological neurons as the original working functions of neurons used to understanding the workings of the brain. But this provided the basis for problem solving outside the sphere through more insight on research in the field of artificial intelligence (Rygeilski *et al,* 2002)

They became to be popular in the 1980’s after the availability of computers in the business world where data was available and data analysts were able to relate because they closely related to statistical methods. Basically neutral networks are defined as mathematical models which process information through computation based on biological neural networks.

They are referred to as black boxes. With trained neural networks, they have several optimized parameters and weighs that cannot be interpreted easily therefore it is not possible to understand why they give a particular outcome.

The main advantages about neutral networks it that they have a way of taming noisy data, because they have many nodes that learn to work around these noisy data in the datasets.

While decision tree are understandable to the amateur analysts, neutral networks are non-transparent to easy interpretation.

2.9. Customer Retention in Churn Analysis

With this high competition among the mobile telecommunication industry, customers have a variety of product and services to choose from which are suited to their needs. One of the major reasons that made the customers have freedom to choose was the mobile portability. Therefore in this competitive industry customers choose products and services that are tailor made to their needs. So the service providers are left to come up with ways on how they can acquire new customers. This is a huge fit because the churn rate experienced by the telecommunication industry is about 25-30 percent annually. It cost less to retain existing customers than to acquire new customers (Roberts, 2000) it costs up to five times as much to make a sale to a new customer as it does to make an additional sale to an existing customer (Dixon, 1999; Floyd 2000).

Drucker (1973) on one of his statements stated that the main reason of a business was to gain a customer. But things have changed now because retention of a customer has become of high priority. Many studies carried out have proved the power of customer retention.

2.10. Opportunities and Challenges in Data Mining

Data mining enables businesses to acquire business intelligence that they can use to address a myriad of problems affecting them. Data mining likewise helps analysts recognize significant facts, relationships, trends, patterns, exceptions, and anomalies that might otherwise go unnoticed (Wang, 2003).

Challenges of data mining include larger databases that require methods for dealing with lump sum data volumes like efficient algorithms and high dimension due to the numerous numbers of fields thus increasing the chances that a data mining algorithm will find bogus patterns that are not valid in general. Other challenges are missing and noisy data and the problem of integration with other systems, since here a standalone discovery system might not be very useful. Changing data and knowledge is a particularly troublesome issue in data mining, as rapidly changing (non-stationary) data can make previously discovered patterns invalid (Fayyad, et al., 1996).

2.11. Gaps

Over the years, many organizations have been saving large volumes of data because of their operations, products and customers. With all these digitized information has made it easy to capture and inexpensive to store due to the technology improvement on the data processing and storage. Organization store data because they would want to use it in future but rare is raw data useful or of direct benefit. And also the raw data may become valuable if they are able to extract information useful for decision support or exploration. Also the raw data extracted would be used to understand the development controlling the data source. Data in the telecommunication industry is still unrecognized, not accessible and not properly utilized. The traditional method of turning data into knowledge relies on manual analysis and interpretation (Fayyad and Uthurusamy, 1996, Fayyad, Piatetsky-Shapiro, and Smyth, 1996).

The organization’s ability to scrutinize and understand an elephantine datasets dawdles far behind their ability to gather and store the data. A lot of studies have been done on data mining and have listed the advantages and described the data mining process, but there is scant research on the success aspect of data mining. The field of data mining is still practically new and the benefits of a data mining project are difficult to quantify in advance because of their exploratory nature

2.12. Background of Orange Kenya

Telkom Kenya was once part of the Kenya Posts and Telecommunications Corporation (KPTC) and back then was the sole provider for both postal and telecommunications. It was the only company then that offered landline phone services in Kenya. In 1999 were split into; CCK, Postal Corporation of Kenya (POSTA) and Telkom Kenya. Telkom Kenya now could offer a wide range of data and voice services and also network facilities after its establishment under the Companies Act to all residential and business customers.

In 2007 at an acquiring rate of 51% of Telkom Kenya's shares it partnered with Orange S.A (formerly France Telecom Group), which saw the launch of the Orange brand in Kenya in 2008.

Orange Kenya offers both postpaid and prepaid service for the different segments of the market. This study only focuses on the prepaid segment of the market. The company offers different call and data plans to suit different customer needs. Orange has one of the cheapest products and service than any other mobile network in the country. Due to the intense competition in the industry, the company now sells customized and subsidized items like Modems, smart phones, laptops and low end handsets to provide one-stop communication facility for customers, moving farther away from its core business. Orange Kenya used to operate Orange Money which gives customers the chance to receive and transfer money but didn’t pick up because it wasn’t able to gunner more customers.

Orange Kenya being a small company is faced with customer churn because of the competitive nature of the telecommunication industry in Kenya and the availability of Mobile Number Portability which allow customer to maintain their numbers on another networks.

Orange Kenya has only 2 service centers that offer customer service and serves as sales points for customers.

Like the rest of the companies in the telecommunication market, Orange Kenya has to deal with the problem of customer churn in the stiff competition.

2.13. Conceptual Framework

Managing customer churn is of great concern for the Orange marketing department. Retaining the customers in the firm is the major challenge faced by the telecom company. As earlier explained in this chapter, it elicits the concepts and reviews related to the factors of customer retention and the customer churn.

On the research we used the dataset from the Orange marketing department. The attribute whose value has to be predicted is known as dependent variable. Its value is decided by value of other attributes. These attributes that predict the value of the dependent variable are known as independent variables.

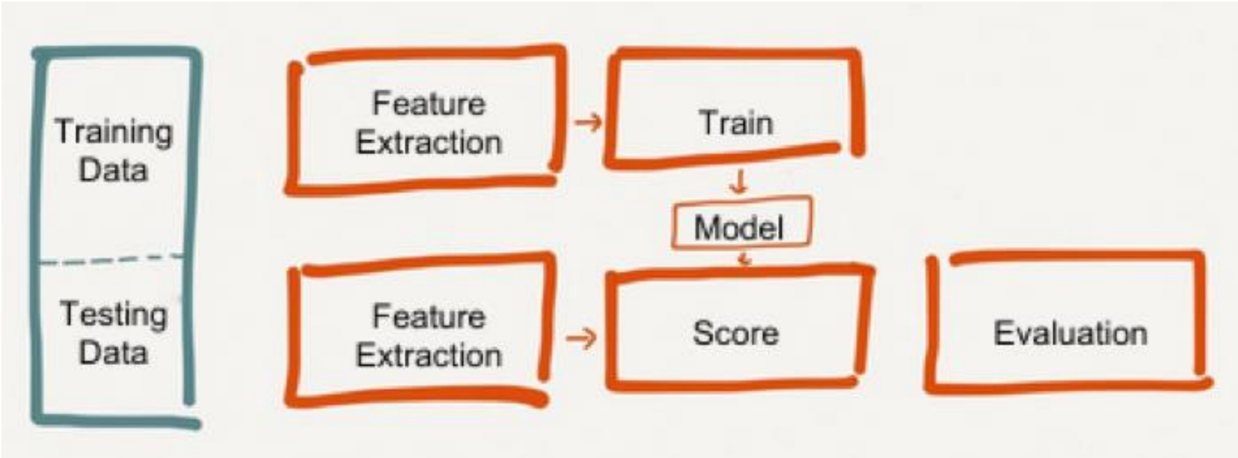

Figure 2.3 conceptual model

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the methodologies and the techniques that were available for research. Various methodologies were identified and discussed and then all the methodologies will be evaluated to select the best method to be used for this project. The section, therefore, will delve into the research methodology used and how it is applied in solving the research problem.

3.2. Research Design

i. Population; the population for the study was all Orange users and former users in Nairobi. There are about 30% of 4.8 million Orange subscribers in Nairobi alone according to the CCK 2016 report. Doing the calculation the population size was 1.44 million.

ii. Sample; there were a number of factors that made it impossible to get information from the entire population. The first one was time and budget constraints. And the other is impossible to get information from the entire 1.44million subscribers in the area so a sample had to be selected hence the sampling technique comes in.

iii. Techniques; In random sampling technique every subscriber (in this case respondent) is chosen entirely by chance and each member of the population has an equal chance of being included in the sample while in snowball sampling technique, is a technique in which research participants are asked to assist in identifying other potential subjects in this case the subscribers who had already churned were selected to know the reason why they churned.

iv. Sample size; on this research, for primary data, I settled for 50 respondents who represented the entire population. The 600 records obtained for this research were gathered from the Marketing department of Orange randomly and were used to create the predictive model. The manager and the supervisor from the Marketing Department were interviewed to know the extent of customer churn and how it affected the company and also what was done about the problem.

3.3. Research Strategies

Research methodologies can serve different purposes such as exploratory, descriptive, explanatory and improving. When using the exploratory approach the purpose is to describe what is happening and when the researcher is searching for new insights, ideas and hypotheses for new research. When serving the purpose of portraying a situation or phenomenon then the descriptive approach comes in. Explanatory research tries to look for an explanation to a situation or problem. The improving approach endeavors to improve an aspect of a studied phenomenon (Runeson and Host, 2009). My thesis has an exploratory approach purpose.

Precisely defined steps within a methodological framework and actions to be undertaken at each stage contribute to revealing the process of Data Mining application development. Also, big project teams cannot perform properly without clearly defined process of application development, modeling and experimentation since it considerably facilitates project design, time scheduling and project supervision.

3.3.1 Research Strategy Framework

Data Acquisition; Churn analysis is done on the basis of historical data which is available from the telecommunication company. Getting data from the telecommunication industries is a big task because of fear of privacy and misuse of it. The data set for this study was acquired from a selected group and willing participants. Also the data used for the predictive model was acquired from the marketing department at Orange, Kenya.

Data preparation; the data acquired cannot be applied directly to the model, so a collection of data is required hence new variables are added to the existing ones by viewing the seasonal usage behavior of the customers. We need the variables to predict the behavior of the customers beforehand since they contain essential information used in the prediction model.

Data preprocessing; this is the most important stage because the data coming in has errors, redundancy and vagueness which can be cleaned before the process starts. Data gathered from different sources is combined first and then cleaned since all data collected will not be satisfactory for the modeling purposes. For example, records with unique values or null values can be discarded.

Data extraction; the variables established for the classifying process. In my research I have used numeric and categorical values.

Decision; after assigning threshold to each attribute will let the subscribers identify and classify in the different categories of churners and non churners. Here basically is identifying the patterns.

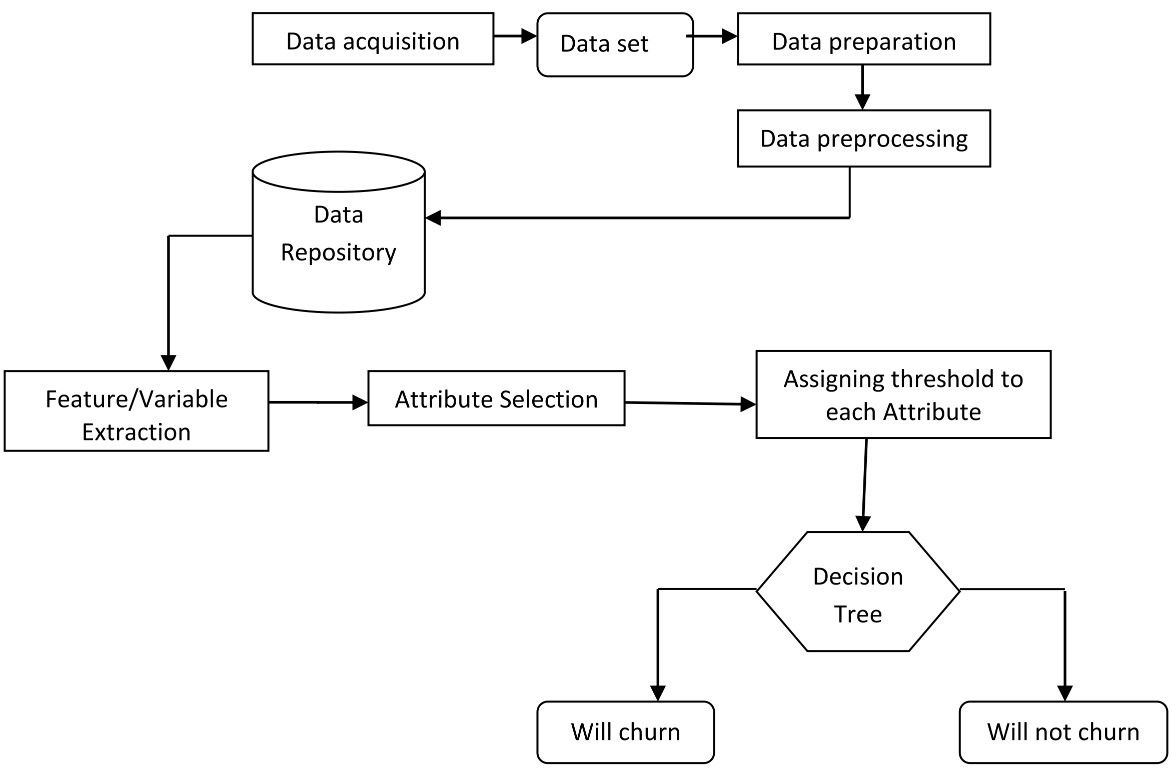

Figure3.1. Pictorial research design framework

3.4. Data Needed

This indicates to the different methods used in doing this study. The time frame of this study was from July 2016 to December 2016. The monthly churn rate for the months of July through December of 2016 was used to know the extent of customer churn in Orange. Questionnaires were used to analyze customer perception and after gathering the information, it was analyzed with SPSS. The Manger from Orange Kenya at Telekom Kenya house was interviewed to get the churn effects on Orange. On the development of the predictive model used dataset of 3000 records from 20 fields that were already defined on the months under observation. With the help of Neutral Network to put together the capacity for a customer to churn and the Decision Tree comes in to show the behavior of the potential churners.

3.5 Data Collection Methods

Primary and secondary data were gathered in stages. In primary data collection, it was gathered through close and open ended questionnaires in the basis of; network quality, customer care and affordability and other causes not on the questionnaire papers. The questionnaires were hand-delivered to the selected participants and collected after a few days. While secondary data was from the CCK figures and reports. Internal data was from the Orange database to help build the model.

What was considered during both data collection was the quality of network, customer care and affordability. If more information was needed to be added by the customer, there were open ended questionnaires to gain more data on which to evaluate from for the churn prediction model. The 600 records used to develop the predictive model were provided by the marketing department of Orange. The datasets were randomly selected from the subscriber list.

3.6 The Churn Analysis Process

For the churn predictive to be built the first thing was for data mining techniques had to be used so as to know whether a particular customer will churn or not and why they churned and for those who decided to stay, what was there reason for staying

i. **Data selection;** The datasets gotten from Orange had already a churn variable for the current customers who had churned. It contained 600 records of both churned and active subscribers. Data was divided into two groups training and validation.

ii. **Data Pre-processing;** The stages consist of;
   - Describing the dataset
   - Removing statistically insignificant fields
   - Define and introduce the target field
   - Exploratory analysis and the target variable

iii. **Data Analysis;** The primary data collected was processed with data mining

iv. **Data presentation;** the data was presented in tables, pie charts and graphs and the tree format.

CHAPTER FOUR: DATA PRESENTATION, ANALYSIS AND DISCUSSION OF RESULTS

4.1 Introduction

Data collected from the primary and secondary sources will help to answer the research objective; what is the extent of customer churn in the company, why the customers churn and what effect they have on the company. And afterward was to build a predictive model for churn.

4.2Extent of the Customer Churn at Orange Mobile

So far the marketing department in Orange is using the Oracle software to know the number of existing subscribers still using their network. For example, those who are still using their Orange line for the past one year are known as the loyal customers and for those who haven't used for the past one month are referred to as potential churners. The potential churners are the customers who haven' called or texted during the months the study was being taken.

During the data processing, I used the different customer demographics to train the data sets; length of service, area and total number of calls.

The table below shows the extent of the churn rate from the month of July to December 2016.

Table 4.1: churn rate

Month	**Churn rate (%)**
July	3
August	2.3
September	2.5
October	3
November	2.4
December	2.1
average	**2.55**

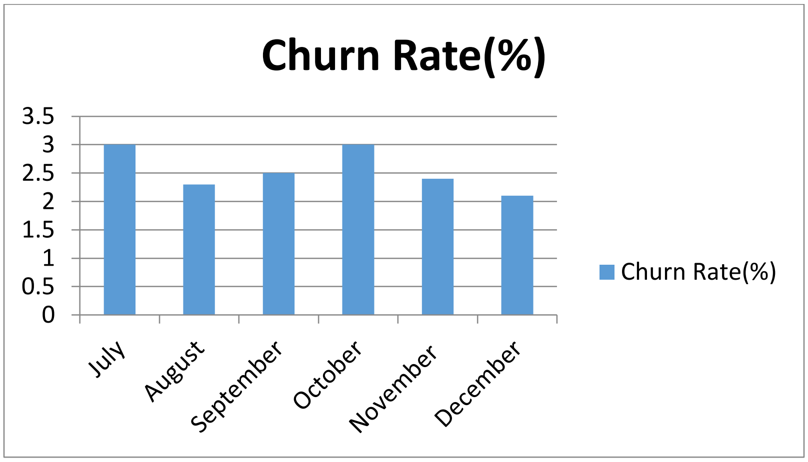

Figure 4.1: churn rate

4.3 Causes of Customer Churn

The 50 respondent of my study were contacted so as to find out the reason why they would move to either Safaricom or Airtel. From the 50 respondents only 45 were using Orange while the other 5 were not on Orange.

The table below shows why some respondents stopped using Orange

Table 4.2: Causes of churn

	Frequency	Percent
Poor network quality	5	11.1
Money transactions	30	66.7
Bad customer care	5	11.1
transparency	2	4.4
Special calls	2	4.4
cost	1	2.3
Total	45	100

From the table above, many of the respondents when asked why they would stop using Orange, most of them went with money transaction. Poor network and others took 11.1%. Also the respondents had an issue with the customer care. They didn't even know that it existed and they have never called to ask for assistance.

Also a number of the 50 respondents had churned partially due to a number of reasons. They still have their Orange lines but stilled used other subscribers for other purposes. The major cause was money transactions service.

Table 4.3: causes of churn

	Frequency	**Percentage (%)**
Money transactions	40	80
Customer care	1	2
Network quality	5	10
Promotion and bonus	**1**	2
Cost	1	2
More transparent	1	2
Special calls	1	2

This shows that most customers consider the money transaction of great importance when deciding to partially churn. Compared to other subscribers cost was not an issue because it was cheaper to make calls compared to other subscribers. The cost charged for calling is not high and it seems to be cheaper than other subscribers. With customer service, the customers declared that they have never called to the customer care service to ask for any help or have never received any call from them.

4.4 Effect of Customer Churn on Orange

Information being readily available at any one time, customers is always on the search for services that suit their needs. Intense competition; there is always a new product that is being launched or bonus by another subscriber and this is a huge problem of losing customers to the competitors.

With losing customers, it brings a lot of problems which even the managers don't want to admit they are losing. CCK on their January 2017 report, called out Telekom Kenya for falsifying there report on the number of subscribers. On their report, Telkom Kenya had said they have 5.2 million customers as at June but upon investigation they had 2.9 million customers. By falsifying there report by not cleaning out the churners, it misled the investors and the authorities. This lead to a loss in revenue coming in from the investors

On a one on one interview with the Regional Manger at Ex-telecom house, shares that the loss of customers has brought a huge loss on revenue because of many investors choosing subscribers with large market share.

Telkom's, market share as at September 2016 was at 7.6% compared to Safaricom 69% and Airtel 17.5%. Their market share is not going to be rising soon because of the new competitor on the block, Equitel which is providing money transaction services that Orange is still unable to offer.

On the matter of operational cost specially marketing, it has to come up with other services to offer customers so as to not lose more clients to the competitors. 10% of the company's expenditure is used for marketing, trying to retain there already existing customers.

4.5 Implementation

The little data that obtained from Orange Kenya had 600 customers and from that I built the Decision Tree. Created two datasets; small and large. For the small datasets there are 10 numeric variables and 50 instances while for the large datasets have 100 variables and 600 instances.

In the making of a predictive model, the data has to be partitioned accordingly so that we can avoid over fitting/under fitting issues.

Data partitioning

Table 4.4: data partitioning

Phase	% of Data	Used for	Remarks
Training	60	Training a Model	60% data for training
Validation	30	Model Validation	30% data for measuring the effectiveness
Test	10	Test data for the model	10% used for testing

Important variables picked up by the decision tree.

Table 4.5: variables

OBS	NAMES	NRULES	IMPORTANCE	VIMPORTANCE	RATIO
1	Total_rev_m	2	1	1	1
2	Usage_frequency	4	0.73	1.00	1.00
3	Total_topup	8	0.27		0.288
4	Avg_min_ob	1	0.13	0.12	0.93
5	Lgth_serv	3	0.10	0.09	0.96

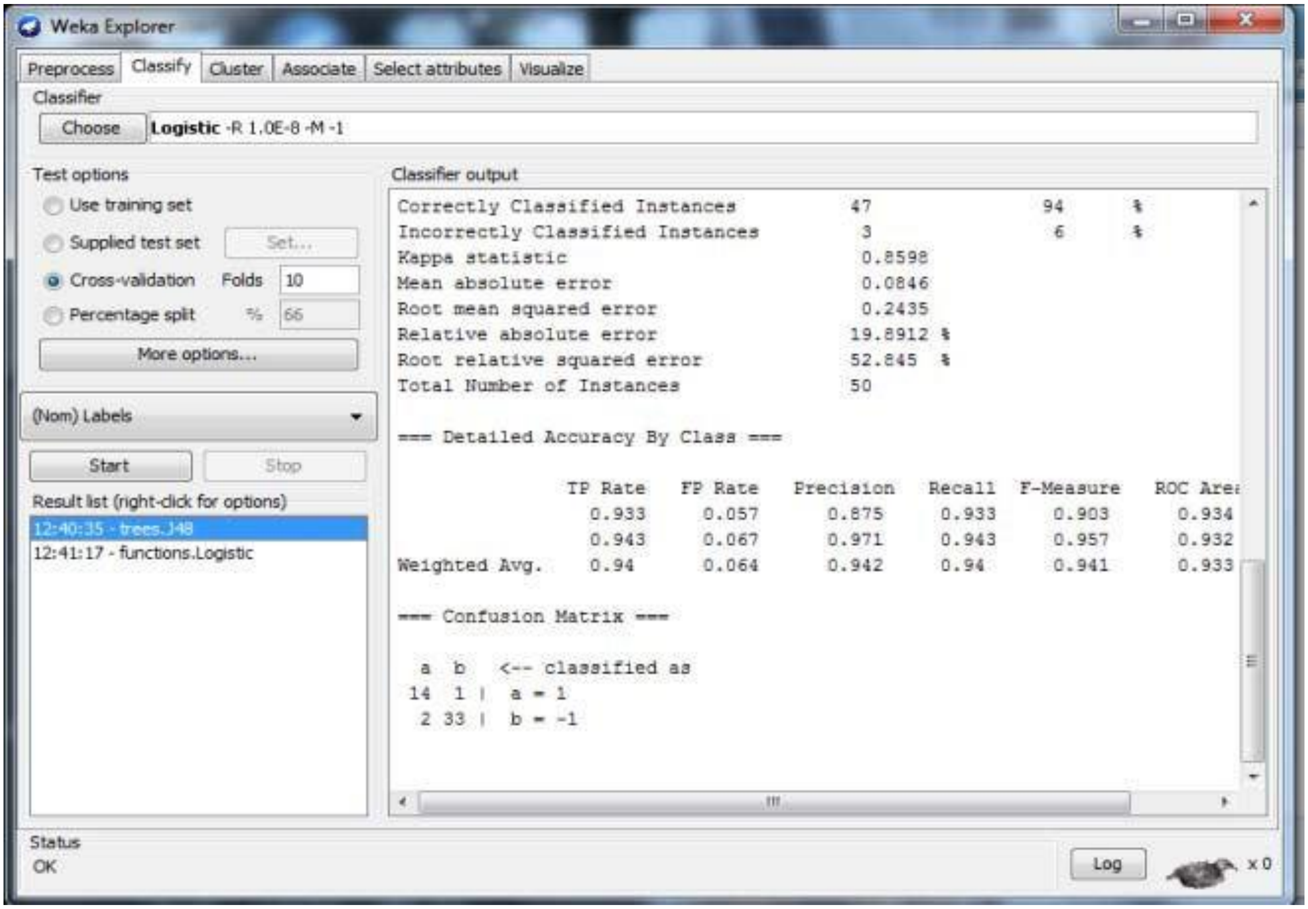

Figure 4.2: Weka J-48 classifier for small datasets

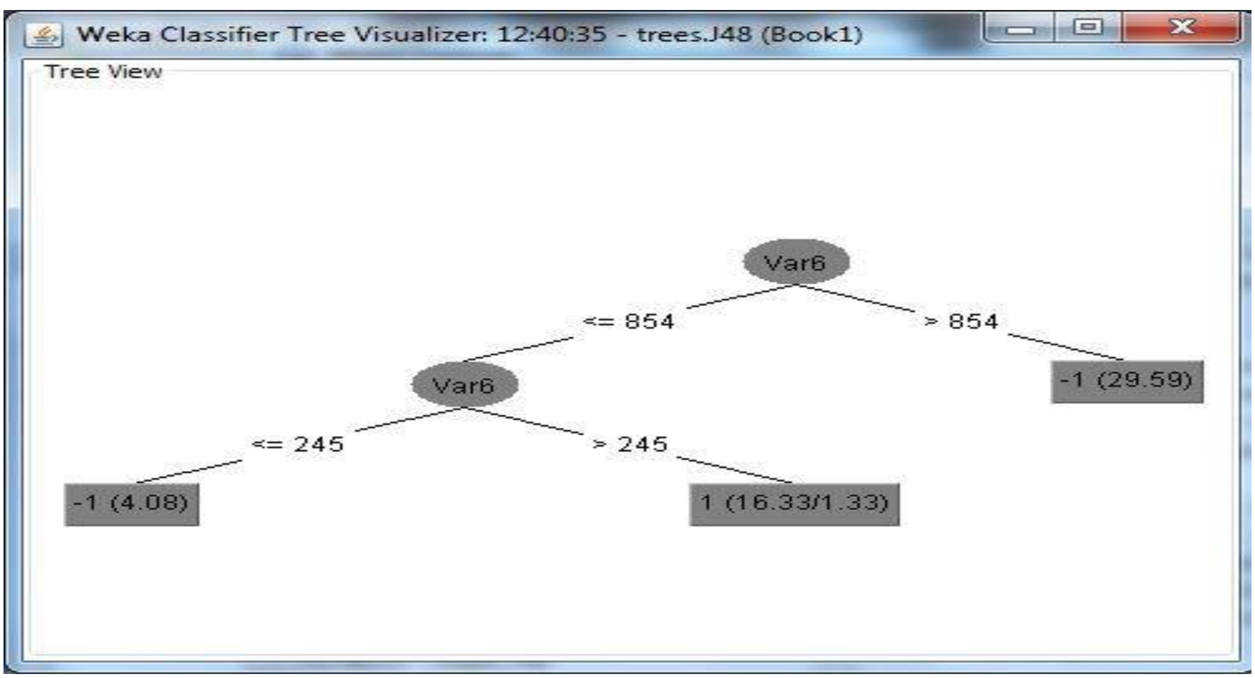

Figure 4.3: Weka J-48 (small datasets)

The correctly and incorrectly classified instances show the percentage of test instances that were correctly and incorrectly classified. The raw numbers are shown in the confusion matrix, with a and b representing the class labels. Here there were 50 instances, so the percentages and raw numbers add up, aa + bb = 16+34 = 50, ab + ba = 2+1 = 3

With the ROC area measurement an "optimal" classifier will have ROC area values approaching 1, with 0.5 being comparable to "random guessing" (similar to a Kappa statistic of 0). Kappa in general is a chance-corrected measure of agreement between the classifications and the true classes.

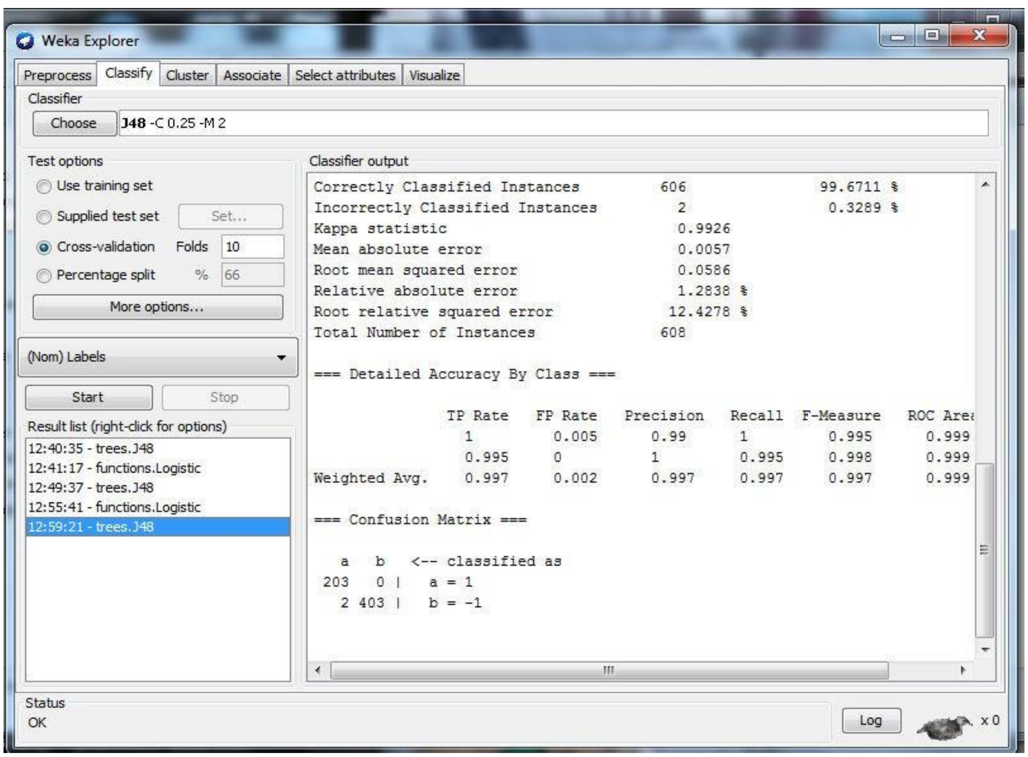

Figure 4.4: Weka J-48 classifier for large datasets

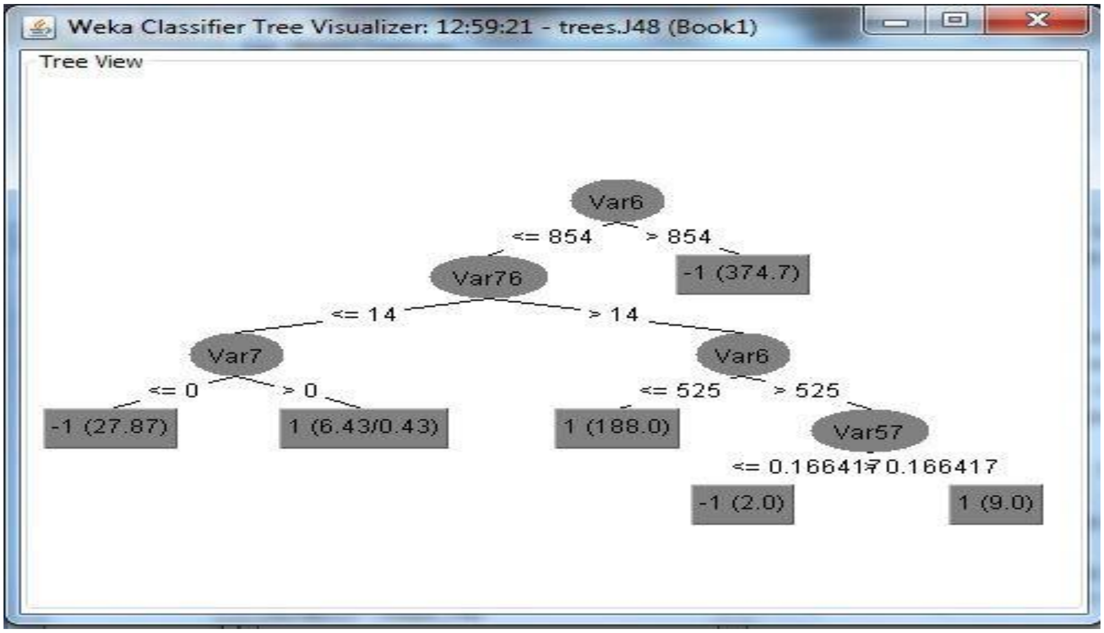

Figure 4.5: Weka J-48 tree visualize (large datasets)

Table 4.6: analysis outcomes using data mining

Data Mining Techniques	Small Data set		Large Data set	
	Churned	**Not Churned**	**Chur ned**	**Not Chu rned**
J-48 Decision Tree	47	3	598	2

After analyzing the results from the decision tree I realized if you grouped the attributes correctly and setting up proper threshold values, I can get accurate results.

The performance of the decision tree was evaluated by calculating the accuracy and the error rate by the below formula;

Accuracy = Number of true outcomes/Total number of predictions

Error Rate = Number of false outcomes/Total number of predictions

CHAPTER FIVE: SUMMARY FINDINGS, CONCLUSION AND RECOMMENDATION

5.1 Introduction

This chapter reports the final findings of the study, whether the objectives were met and what the conclusions are.

5.2 Current Churn Management at Orange Kenya

When conducting the study, Orange' marketing team had not yet built a model to predict churn. So far they only had an automatic predictive tool which only alerted them when a customer was inactive for the past 10-60 days. And so far the only thing the marketing team had to do was to send messages to the inactive members that there lines will be deactivated if they did not top of their lines with airtime. And if a customer did nothing after the end of the six months, they would be deleted and the number recycled.

From the survey taken, it was realized that like 10% of the people whose lines were deleted came back for the same line and found that their original lines are not available so they had to buy new lines or just left. The reason they gave for not using their lines was that they had traveled to the rural areas and the network quality was not good so they had to abandon the lines and buy the lines that had better quality.

Banks and the telecommunication are joining together to make money transactions easy for their customers. And what the banks can offer, for example loans, the operators are also offering. With the money transactions moving to the mobile platform making many customers prefer a mobile operator with the option of moving money. In this case Orange is disadvantaged greatly because their money platform failed greatly. Now customers prefer to do everything at the touch of a button.

5.5 Summary of the Findings

The extent of the customer churn in Orange was between 2 to 3 percent from the data analyzed. This figure might seem small but the customer churn rates are rising.

The causes of customer churn were poor network quality and mostly mobile money. More than half of the respondents stated that they have no problem with their subscriber but were unable to stick to it because of the mobile money. Nobody nowadays wants to walk around with money and there is a subscriber who has that money platform available to them where they can do business with their phones

Effects of customer churn have some serious consequences on Orange Kenya because it frustrates its efforts to achieve projected revenues and have a low market share. Also it increases their operational costs because of marketing and rebranding itself to win back customers who have churned.

5.4 Conclusions

The telecommunication industries in Kenya are undergoing a tremendous transformation and the number portability is not making it easy for the subscribers to move to the other competitors. They have suffered from high churn rates and huge churning loss which though is unavoidable, but can still be managed and reeled in to an acceptable level.

With Orange the problem they had was to know the true number of active customers in the network and had not used data mining techniques to be able to predict churners in the near future and find ways to retain the potential churners in the case of already customers. The company's way was achieving new customers by doing a lot of marketing offering products that was affordable than the competitors which was not helping them with the total collection of revenue.

To help them out, they have to find a way to know the churners and the on-churners which will help them to know the right people to target with the necessary incentives and offers. The model was to help the company know the real churners.

By studying the customer behavior they are able to offer them packages suited to their needs. The Decision tree model applied is to interpret and give us the factors causing customers to churn. The rules used were to explain the reason of customer churn.

5.5 Recommendation

After the research the author, would recommend that Orange build a model for churn prediction to help them in managing the churners. The company is wasting money doing mass marketing, trying to entice new customers and all they can do is try retaining the already existing customers.

The number one product that should be returned by Orange Kenya is the mobile money transfer which would be a huge boost to their company. They have been able to lure the matatu industry in Nairobi by providing them with free Wi-Fi that is cheap and affordable. Such incentives to their customers go a long way to retaining their customer base.

Further studies should be done using other data mining models besides Decision Tree to find out their accuracy levels in prediction.

5.6 Contribution to Knowledge

A study done by Karianga compared two methods Decision tree and Cox proportional hazard method. And his study concluded that decision tree performed better than the Cox proportional method. His case study was on Safaricom Limited doing a churn prediction model. He concluded that for the company to utilize the model, they have to run the models monthly. His would assist in continuously tracking the behavior of the subscribers as the behavior patterns are affected by many occurrences that cannot be controlled.

Mwangi’s 2015 study took a more direct approach on the organizational strategies for customer retention in the mobile telecommunication sector in Kenya.

Adding this study to some of the authors who concentrated on churn analysis, other studies should use this as a guideline to do more than what is presented and should use the other data mining models besides Decision Tree to find out if they will give a more accurate churn prediction than what was used. Also further studies should be done on postpaid to find out if they provide a high propensity to churn than the prepaid or vice versa.

References

Ahmad, R. and F. Buttle (2003), ―Customer Retention Management: A Reflection of Theory and Practice‖, *Journal of Marketing Intelligence and Planning,* Vol.20, No. 3.

Ahn, J., Han, S. and Lee, Y. (2006) Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry, *Telecommunications Policy*, 30(10-11)

Bateson, J.E.G. and Hoffman, K.D. (1999), *Managing Services Marketing: Text and Readings*, 4th ed., Dryden Press, Orlando, FL,

Ballings, M. and Van Den Poel, D. (2012). Customer event history for churn prediction: How long is long enough? *Expert Systems with Applications*, 39(18)

Baurdeau, A., Liebenau, J. and Savin, B. (2005), A New View of Scale and Scope in The Telecommunications Industry; Implications For Competition And Innovation, *International Journal of Digital Economics*. 60(4).

Berson, A., Smith, S., & Therling, K. (1999). *Building data mining applications for CRM.* New York: McGraw-Hill.

Bhattacharyya, S, & Pendharkar, P. (1998). Inductive, evolutionary and neural techniques for discrimination: A comparative study. *Decision Sciences*, *29*.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees.* New York: Chapman & Hall / CRC.

Carrier, C., & Povel, O. (2003). Characterising data mining software. *Intelligent Data Analysis , 7*.

CRISP-DM: Cross Industry Standard Process Model for Data Mining (http://www.crisp-dm.org), 2010.

Coussement, K., & Van den Poel, D. (2008a). Churn prediction in subscription services: An applicayion of support vector machines while comparing two parameter-seceltion techniques. *Expert Systems with Applications , 34*.

Coussement, K., & Van den Poel, D. (2008b). Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management*, *45*.

Han, J., & Kamber, M. (2006). *Data Mining: concepts and techniques* (2nd Edition ed.). San Francisco: Elsevier.

Hung, S., Yen, D., & Wang, H. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications , 3*.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics , 29* (2).

Lee, S., & Siau, K. (2001). A review of data mining techniques. *Industrial Management and Data Systems , 101* (1).

Madden, G., Savage, S., & Coble-Neal, G. (1999). Subscriber churn in the Australian ISP market. *Information Economics and Policy , 11*.

Montgomery, A., Li, S., Srinivasan, K., & Liechty, J. (2004). Modeling online browsing and path analysis using click stream data. *Marketing Science , 23* (4).

Mwangi L.N. (2014); Organizational strategies for customer retention in the mobile telecommunication sector in Kenya;

Neslin, S., Gupta, S., Kamakura, W., Lu, J., & Mason, C. (2006). Defection Detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research , XLIII*.

Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications , 36*.

Njenga, B.M., Mohamed M.H.& Wayu S. ,(2012) ; The application of data mining in the mobile telecommunication Industry in Kenya.

Reichheld, F., & Sasser, W. (1990). Zero defection: Quality comes to services. *Hravard Business Review.*

Rygielski, C., Wang, J., & Yen, D. (2002). Data mining techniques for customer relationship management. *Technology in Society , 24* (4).

Shaw, M., Subramaniam, C., Tan, G., & Welge, M. (2001). Knowledge management and datamining for marketing. *Decision Support Systems , 31*.

SPSS Inc. (2007). *Clementine 11.1 Algorithms Guide.* Integral Solutions Limited.
Swait, J., & Andrews, R. (2003). Enriching scanner panel models with choice experiments. *Marketing Science , 22* (4).

Wei, C., & Chiu, I. (2002). Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications , 23*.

WEKA (Open source, Data Mining software in Java), University of Waikato, New Zealand (http://www.cs.waikato.ac.nz/ml/weka), Weka 3.7.0., 2010.

Yan, L., Fassino, M., & Baldasare, P. (2005). Predicting Customer Behavior via Calling Links. *Proceedings of International Joint Conference on Neural NEtworks*, (pp. 2555-

2560). Montreal.

Zhao, Y., Li, B., Li, X., Liu, W., & Ren, S. (2005). Customer churn prediction using improved one-class support vector machine. *Lecture Notes in Computer Science , 3584*, 300-306.

Appendix

Appendix One: Questionnaire

Customer Interviews

The aim of this interview is to understand the problem of churn in the telecommunication industry from the point of view of the customer and what causes them to churn.

Screening Question

Qa: Do you use more than one SIM card? Which operators do you use?

Qb: what is your age? (If less than 18 years old cancel the interview)

If yes and over 18 respectively then please answer the following:

Demographic Questions

Q1: Gender: M/F

Q2: In which part of Nairobi do you live?

Q3: Do you have a family/married/or in a relationship?

Q4: What is your education level?

Questionnaire

Q5: a) Are you on post-paid or pre-paid subscription plan?

b) Which tariff are you subscribed to?

c) What is your average usage for each operator/operator (price/month)?

d) Which network services and bundles do you use?

e) What promotions do you subscribe to?

Q6: How does your income affect your calling patterns?

Q7: Is the price a major concern or are there other concerns you would consider when using a certain operator? What are these concerns?

Q8: If a new operator provides a cheaper tariff or a promotion will you consider buying their SIM as well? What kind of offer would you consider?

Q9: Have you ever contacted customer support? What do you think of their service? How does this affect your calling patterns?

Q10: Does the quality of network (network services, excluding customer support services, for example, the signal strength, sound quality, the connectivity, and the call-drop rates) make you consider changing the service provider? Why?

Q11: Number portability has been introduced in Kenya. How does this affect your decision in changing the operator?

Q12: Does your calling patterns differ during the day or year around? (for example, do you make fewer calls in the morning or during vacations), please explain the pattern?

Q13: Would you prefer to make your own customized calling plan/ bundle? If yes, what bundles would you choose?

Q14: How do you measure the value of the service provided by the operator?

Q15: Why would you consider switching to another operator?

Q16: What other factors do you think would support your decision?

Q17: What action if done by the operator would let you re-consider your decision?

Q18: How does the brand and marketing affect your decision of changing your operator?

Appendix Two: Customer analysis sample

Customer Usage Analysis **Customer Code**	**Primary Operator**	**Average Usage (Ksh/Mon)**	**Secondary Operator**	**Average Usage (Ksh/Mon)**	**Other Operators**
CI001	Orange	1000	Airtel	500	Safaricom (pre-paid),
CI002	Orange	1000	Safaricom	500	Yu (short while)
CI003	Orange	3000	Airtel	700	Not Applicable
CI004	Orange	3000	Airtel	when required	Yu (short while)
CI005	Orange	1000	Airtel	300	Yu (short while)
CI006	Orange	500	Safaricom	300-1000	Not Applicable
CI007	Orange	2000	Used to use YU	Not Applicable	Safaricom(business internet)
CI008	Orange	500	Safaricom	1500	Airtel (short while)
CI009	Orange	6000	Airtel	1000	Not Applicable
CI010	Orange	600	Used to use YU	300	Not Applicable
CI011	Orange	800	Safaricom	500	Not Applicable
CI012	Orange	300	Safaricom	100	Unassigned
CI013	Orange	3000	Used to use Airtel	Not Applicable	Not Applicable
CI014	Orange	4000	Airtel	2000	Not Applicable
CI016	Orange	1000	Safaricom	100	Yu (short while)
CI017	Orange	900	Safaricom	700	Safaricom (for business)
CI018	Orange	1000	Safaricom	1000	Yu (short while)
CI019	Orange	2000	Safaricom	500	Not Applicable
CI020	Orange	1500	Airtel	1000	Yu (short while)
CI021	Orange	2500	Airtel	1200	Yu
CI022	Orange	4000	Airtel	2000	Unassigned
CI023	Orange	5000	Airtel	500	Not Applicable
CI024	Orange	7000	Airtel	100	Not applicable
CI027	Orange	4000	Safaricom	3000	Not Applicable
CI028	Orange	2000	Airtel	1000	Not Applicable
CI029	Orange	1500	Airtel	1000	Unassigned