**UNIVERSITY OF NAIROBI**

**SCHOOL OF COMPUTING & INFORMATICS**

# REDUCING CUSTOMER CHURN IN THE TELECOMMUNICATION INDUSTRY BY USE OF PREDICTIVE ANALYTICS

**By**

**MARTIN MATHU**

**(P52/12707/2018)**

**Supervisor**

**DR. EVANS MIRITI**

A RESEARCH SUBMITTED AS PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE COMPUTATIONAL INTELLIGENCE IN THE SCHOOL OF COMPUTING AND INFORMATICS OF UNIVERSITY OF NAIROBI

**AUGUST 2020**

# DECLARATION

I proclaim that this research and prototype is my own unique work and has not been distributed or submitted elsewhere for the honor of a degree or other purposes. I likewise declare that this contains no material composed or distributed by others with the exception of where due reference is made.

Student Name: **Martin Mathu** Reg.No: **P52/12707/2018**
Sign …… Date: ……

Name: **Dr. Evans Miriti**
Sign …… Date: ……

# Abstract

Customer churn is a big problem in various businesses and especially so in the telecommunication industry. When a business loses its customers, it loses the revenue that was being generated from the customers and possibly revenue from potential customers who receive negative marketing from customers who churn. Managing customer churn in the Kenyan telecommunication industry has been largely ineffective due to the reactive approach where by churn is just a metric that is reported by the business after a certain period.

The objective of this study is to show how we can use predictive analytics to proactively identify customers who are about to churn. By doing so businesses can take measures to prevent or reduce churn and therefore increase their customer retention. This was done by identifying features that are most important in predicting churn, developing, implementing and testing a churn prediction models and evaluating the performance of the models.

While there exist different approaches to solving the churn problem, machine learning was used to do the churn prediction based on various customer attributes such as age, usage, gender, etc. Since there exists multiple algorithms to do this kind of machine learning, this research implements four of them and does a comparison to see which one would be the most suited based on their performance.

The final result shows which features can be used for churn prediction which were Registration Document, Age on Network, Subscriber age and Talk Time. The importance of each of these features was also shown. Classification algorithms that were used are Random Forest, K Nearest Neighbors, Naïve Bayes and Neural Network. The end results show how the algorithms perform in terms of accuracy and execution time.

# Acknowledgements

Most importantly, thanks be to God for the endowment of life. This exploration would not have been conceivable were it not for the gifts and openings He has granted me.

Furthermore, my profound appreciation is reached out to my administrator Dr. Evans Miriti. His direction and management guaranteed I finished my exploration on schedule and his huge information in the AI space was key in the improvement of my model for the investigation. I might likewise want to express gratitude toward him for his time and responsibility, it was an extraordinary benefit to concentrate under him

At long last, I am amazingly appreciative to my family for the love, backing and inspiration they have offered me to guarantee I complete my examination.

Table of Contents

# List of Tables

# List of Figures

# ACRONYMS

ARPU – Average Revenue Per User

AUC – Area Under Curve

CRM - Customer Relationship Management

NPS - Net Promoter Score

NSGA II - Nondominated sorting genetic algorithm II

PII - Personally identifiable information

# 1 Introduction

## 1.1 Background

The current rule in business growth is that customer retention is the most important part of the equation, it doesn't matter how many customers you acquire if you can't retain them (Alex Birkett, 2017). Customers are the most vital asset any business relies on and there is no business that can achieve success without first having to establish a solid customer base. Mahatma Gandhi in 1890 said: "*A customer is the most important visitor on our premises. He is not dependent on us but rather we are dependent on them. They are not an interruption of our work, they are the purpose of it, they are not an outsider of the business but a part of it, We are not doing them a favor by serving them, but they are doing us a favor by giving us the opportunity to do so.*" (Steve Shellabear, 2017).

In the pursuit to get more customers, businesses tend to focus more on acquisition of new customers while putting less effort to retain current customers. Acquisition of new customers is critical, but retaining them fastens profitable growth (Larry Myler, 2016). Customers who are dissatisfied or neglected might stop using the products or services of the business in favor of other alternatives. This results in a phenomenon called churn. Churn quantifies the number of customers who no longer use a company's service and are no longer revenue generating. Churn rate is basically the percentage of customers that cease using your service and is usually measured as an annual percentage but can also be measured quarterly, monthly, or even weekly (Baremetrics, 2011).

Safaricom PLC (2019) reported that for the year ended March 2019, its market share stood at 63.5% down from 65.4% the previous making it the lowest ever. Another key number to note was the churn rate which stood at 23.45%. All this coming at the back of sustained cheaper tariffs from the competition, Airtel and Telkom. To quantify the magnitude of the churn problem, Safaricom has 31.85M customers each with a monthly average revenue per user (ARPU) of 658.30. 23.45% churn rate translates to 7.4M customers having left the business during that financial year. Based on the ARPU, the calculated loss of revenue was 4.8B per month.

As per Airtel Africa's report whose Kenyan market share stood at 21.4% their monthly churn rate was 5.7% (Airtel Africa, 2019). The company posted a Sh2.89 billion for the previous year despite its subscriber base having risen by 45 percent to more than 13 million. Airtel Kenya is now focusing on its subscriber numbers, churn rate and revenue per user in order to increase profitability (Business Daily, 2019).

Customers turning their back to your service or product can be a major headache to the business. Not only is it very hard and expensive to win them back once lost, but they can also do more harm to the business through negative marketing by word of mouth. Hence it is now accepted that the best marketing strategy is retention of existing customers or more simply to avoid customer churn (Golshan Mohammedi et al., 2013)

The foundation of churn prediction is based on one of the most important assets a company can have - data. For this case the data that is required is customer related such as customer profile, revenue, usage, etc. This data forms the basis for understanding the customer, of which Safaricom

is at an advantageous position, as it has a lot of data/ big data which would be very useful for observing various customer related patterns and trends.

An example of how predictive analytics has been applied to prevent churn was Sprint (USA) using artificial intelligence to improve customer experience. In 2014, sprint had a customer churn rate of 2.3%, almost twice that of its main competitors. They decided to implement a solution that used predictive and self-learning analytics to help find subscribers at risk of churn and proactively providing personalized retention measures to such customers. The results were amazing, Sprint decreased its customer churn by ten percent which was the lowest historically and increased its Net Promoter Score (NPS) by forty percent. Sprint also managed to boost customer upgrades by eight times resulting in 40% more customers performing a replacement business line and improving the overall customer service satisfaction.

## 1.2 Problem Statement

According to the strategies of the three main telecommunication companies in Kenya, they all strive to increase their market share which translates to increased customers and revenue. This has led to cut-throat competition in acquiring new customers.

However, the overall number of customers has not been growing as expected despite acquisition of new customers. This is due to customers who leave/ churn from the business because of various factors such as customer dissatisfaction, competition and other reasons. These customers then seek the services of an alternate service provider which leads to the positive impact of new customers being canceled out by those leaving.

## 1.3 Objectives

**Main objective:**

The primary purpose of this research was to use predictive analytics to establish customers who have a high probability of churning from a telecommunication firm.

**Specific objectives:**

- To identify customer features that are most relevant in predicting churn.
- To develop, implement and test a model for predicting customer churn using specific classification algorithms.
- To evaluate and compare the performance of different algorithms used.

## 1.4 Significance of the research

Customer churn is like a leaky bucket being filled with water. You add water into the top, but it keeps pouring from the sides. If you can plug a few of the holes, your bucket will fill faster (Zac Harris, 2019). A leaky bucket refers to a business with bad customer retention. For a business to succeed greater emphasis must be placed on retaining customers once you acquire them. Thus, the telecommunication firms or any other business that implement the findings of this study can benefit by increasing their customer retention. Increased customer retention leads to the following benefits:

**Reduction in customer acquisition costs**. According to various studies, acquiring a brand-new customer could be up to 25 times costlier than retention of an existing one. Furthermore, the longer a customer stays with the business the more revenue they're likely to give the business more revenue.

**Improved Net Promoter Score.** This measures the willingness of a businesses' customers to recommend the company's products or services to other people. In addition to the customer being better satisfied, they are likely to bring in more customers through their positive recommendations which leads to further growth of the business.

**Increased profitability**. According to research done by Frederick R. of Bain & Company and also the inventor of the online net promoter score, the research shows that increasing customer retention rates by 5% increases profits by 25% to 95%. Retained customers are loyal, buy more often and spend over newer customers. They've learned the worth of a product or service and keep returning, again and again.

**Long-term success**. Businesses that can retain customers for a long period are likely to survive longer than those that don't. This is due to the loyalty and recurring business from such customers.

# 2 Literature Review

This section reviews literature regarding predictive analytics in customer churn and its application in churn management

## 2.1 Customer Churn

Churn is the number of customers who stop using or purchasing products or services from a company (Kenneth, Jane and Ahmed, 2013). In the telecommunication industry churn is when a customer stops using the operator's services such a  s calling, SMS, data, mobile money.

The cost of losing profitable customers in very competitive markets to rivals is making a lot of companies to change their goals from massive capture of new customers to the retention of existing ones. However, not all cases churn types are equally important nor are they all predictable (David et al, nod). Based on the reasons why it occurs, churn can be classified into:

**Involuntary cancellation**: This happens to customers whereby the service provider withdraws the service for example due to fraud or payment arrears. Most companies do not usually consider these as churn for their services.

**Voluntary cancellation**: This is a result who customers who engage in actions to change their current service provider. This type has two categories:

- Circumstantial: A customer's circumstances might change which prevent them from continuing with their patronage. For example, due to change of physical location a might not be able to use the company's services whether they want to or not. This type of cancellation is inherently unpredictable.
- Deliberate: The customer deliberately and voluntarily decides to change from their current service provider in favor of a competitor.

## 2.2 Churn Management

There is proof that the amount of crucial customer data an organization has, is vital for businesses to become customer-centric (O'Halloran, 2003). One major use case of such data is churn management. Churn management is the process of identifying customers who plan to move to a competing service provider (Hadden et al., 2007). Churn management is basically the ability to discover early warning indicators from potential churners and taking proactive steps to stop it from occurring.

It is important to understand the causes and triggers that make customers leave the company in order to be able to develop strategies against it (Kumar & Petersen, 2012). Some of these triggers can be identified through complaints, reduction in usage and surveys among other things. Once identified, the business can then take preventative measures such as giving the customer discounts, refunds and resolving any other customer pain points.

While previous work has been done on churn management in telecoms and other businesses, a major issue with successful churn management is the high rate of false positives and false negatives in churn prediction. The false positives can result in severe negative impact on measures undertaken to resolve churn because they increase the cost of the programs. This is because the company is essentially spending money trying to retain customers that were never really going to churn. In some cases, such false positives can often be higher than 50%.

By building a structured, proactive and 360-degree churn management process, an organization can be able to continually boost its customer retention over time while keeping its customers happy and the revenues high.

## 2.3 Related Work

Halim Joseph (2015) carried out a research titled *The Causes of Churn in the Telecommunication Industry: A Case Study of various Kenyan Service Providers*. The research question was framed into three specific questions. First specific question that was investigated was the behavioral patterns of customers that churned. The second question looked into was the economic patterns leading to churn, and the third question was in regard to the policies and regulations that influence churn. Interviews with at least one manager from each service provider were conducted and data collection was gathered from subscribers who use multiple SIM cards or have switched their service provider. The study identified causes of churn based on the different perspectives forming a basis on which churn prediction software could be created. Limitations of the study included lack of a proposed churn management solution and the fact that sample customers used were only from Nairobi.

Buckley (2011) undertook a study of *Customer churn prediction in telecommunications*. The study presented a set of subscriber attributes for churn prediction of land line subscribers, including 2 six-month Henley segmentations, accurate four-month call data records, line details, bill and payment data, demographic data, account details, service orders and complaints interactions. With the new features, the researcher used seven prediction techniques: Decision Trees, Logistic Regression, Linear Classification, Multilayer Perceptron Neural Network, Naive Bayes, Support Vector Machines and the Evolutionary Data Mining Algorithm. In conclusion, experiments were carried out to compare and evaluate the subscriber attributes and the seven modelling techniques for customer churn prediction. The outcome showed that the new attributes and the seven modelling techniques were more efficient than other existent ones for customer churn prediction in the telecommunication companies. This research was very informative regarding features that could be used and modelling techniques for churn prediction. However, its main limitation is that the research was done for land line customers. Land lines are not as widely used as before and not all the characteristics of land line customers might apply to mobile customers.

Bingtuan Huang (2009) carried out a research for *Multi-objective feature selection through use of NSGA-II in churn prediction of subscribers in the telecommunications industry*. Nondominated sorting genetic algorithm II is a very popular algorithm applied in multi-objective Optimization

Problems. The research proposed a new multi-objective feature selection methodology for predicting churn in the telecommunication industry, through optimization by use of NSGA-II. The main idea of this methodology was to improve the approach of NSGA-II to select local features subsets of varying sizes and then apply searching nondominated solutions to select the global non-dominated features subset. The proposed solution was evaluated by conducting experiments and the experimental outcomes showed that the proposed feature selection methodology was efficient for churn prediction with multi-objectives. This research was very useful for identifying features that could be used for churn prediction and managed to show the effectiveness of Nondominated sorting genetic algorithm II. The limitations were that only one algorithm was explored in identifying the features Call data records, contract tenure, age, gender among others and the research did not rank the importance of the features.

Huang et al. (2015) undertook a study on the problem of customer churn in the big data platform. The researcher's main objective was to prove that big data significantly improved the process of predicting churn. The big data depends on the quantity of data, its variety, completeness and velocity. The data used was from the Operation and Business Support department at one of China's largest telecommunications firm that needed a big data platform to engineer the solution. Random Forest was implemented and evaluated using Area Under Curve (AUC). This research was very useful in showing that the more data you have the more robust churn prediction model you can build. One of the limitations of the algorithm was that only one algorithm was used for classification and therefore no comparison on how different algorithms perform. Another limitation was that it focused only on using big data platform for churn prediction thus its application is only limited to businesses that have big data platforms.

## 2.4 Predictive Analytics

Predictive analytics is a branch of data analytics which is part of business intelligence, therefore to understand it we need to understand data analytics and business intelligence. Business intelligence is a wide category of applications, technologies, and methodology used for gathering, storing, accessing and analyzing data to enable the business to make better decisions, (Watson, 2009). Business intelligence is therefore a collection of tools utilized to improve the decision-making process in an organization through transformation of data into useful business information and knowledge. This is achieved by utilizing data mining tools and analytical techniques. Data analytics is a branch of business intelligence and consists of:

- Descriptive analytics: Describes the present situation and answers the question what is happening currently?
- Diagnostic analytics: attempts to explain why something is happening.
- Predictive analytics: the objective is to show what could happen in the future.
- Prescriptive analytics: tries to establish the right decision or solution

Predictive analytics can also be defined as attempting to predict the future through analysis of past performance and study of historical data to discover relationships and patterns in these data.

Predictive analytics will not tell you what happens in the future but instead tries to forecast what might happen in the future with an acceptable threshold of certainty and includes risk assessment for various outcomes. Churn prediction can therefore be defined as the use of customer related data to forecast the probability of a customer or group of customers ceasing use of the company's services in the future.

Apart from telecom industry predictive analytics is extensively used in various industries such as banking, insurance, manufacturing, government, health, among others. Other use cases in addition to churn prediction include risk assessment associated with a particular cause of action, marketing by customer targeting and cross selling, sales forecasting by predicting future sales, fraud detection by checking anomalies in various transactions and improving operations by forecasting supply and demand.

While largely beneficial and effective, predictive analytics has its share of challenges. Even if a company has enough data, it could be argued that when predicting human behavior, computers and algorithms may not consider some variables such as emotions, changing weather, spontaneity and other relationships that might affect the customer. Time also affects how well these techniques work. Even though a model can perform well at a specific point in time, customer behavior changes with time and therefore a model should continuously be updated to remain relevant. Data privacy and protection issues may also arise when doing analytics on customers data. Previous research shows that while doing such analytics it is important to also analyze the digital data privacy in order to ensure trust through sound business practices in data analytics and to enhance marketing activities (Leonard 2014; Martin and Murphy 2017).

## 2.5 Classification Algorithms

Classification is a kind of supervised learning whereby the algorithm learns from training data given to it and then uses the skills learnt to classify a new observation. Classification may be bi-class for example identifying whether a human is male or female or it may be multi-class for example animal classification into Mammal, Bird, Reptile, Fish, Amphibian. Churn classification was bi-class with the target classes being Yes and No. Yes representing those who churned and No representing those who did not churn.

In machine learning there is a theorem known as No Free Lunch theorem which means no one algorithm works best for every situation. This is widely applicable in classification models where we train a model on the dataset and later use the trained model for predictions on new observation using one of the many classification algorithms. It is therefore recommended to try a variety of algorithms for your problem, while using a hold-out test set of data to evaluate performance and select a winner (Raheel, 2018).

There are many classification algorithms, but we shall only discuss 6 of the most common.

### 2.5.1 K-Nearest Neighbors

K-Nearest Neighbors works by calculating the distance of the test observation from the known values of some training observations. The group of training observations that would result in the shortest distance from the training observations and the test observation is the class that is selected. K represents the number of neighbors (training points) that were used to classify the test point. These neighbors vote and the majority wins i.e. the test point was placed in the class that majority of the neighbors belongs to.

### 2.5.2 Decision Trees/Random Forests

A decision tree implements classification to mimic a tree structure. The data set is split into smaller subsets in a logical manner as the it develops a decision tree incrementally until the final result is a tree containing decision nodes and leaf nodes. A decision tree node usually has two or more branches while a leaf node has no branches and represents a decision(classification).

Random forests work by creating multiple decision trees to be used during training time and outputting the class that is the most common of the classes (for classification) or mean prediction (for regression) of the individual trees. By doing so, random forests minimize the habit of decision trees whereby they overfit to their training set.

### 2.5.3 Naive Bayes

A Naive Bayes Classifier shows the likelihood that a given observation belongs to a particular class. It calculates probability that an event will happen given that some specific conditions have been observed. Naive Bayes operates under the assumption that a feature is independent from any of the other features the features independently contribute to the probability. Not only is naïve Bayes simple but it has been known to outperform even complex classification methods.

### 2.5.4 Logistic Regression

Logistic Regression works by making predictions for test data points using a binary scale that is zero or one. If the output of the prediction is 0.5 or above, the observation is classified as belonging to class 1. When output of the prediction is below 0.5 if is classified as belonging to class 0. All the features will also have a label of either 0 or 1. It is a linear classifier and most suited when the data has linear relationships.

### 2.5.5 Support Vector Machines

Support Vector Machines operate through creating a line to separate different clusters of data points into their respective classes. Observations found on one side of the line will classified in same class while those on the other side will belong to the other class. The algorithm attempts to maximize the distance between the line created and the observations on either side of it so as to increase the confidence of which observations belong to which class. During testing observations are plotted and the side of the line they will fall on is the class which they belong to.

### 2.5.6 Neural Network

Neural network is a set of algorithms combined to model the human brain. They consist of neurons organized in various layers which convert the given inputs into an output. A neuron unit takes inputs and applies a nonlinear function after which it forwards the output on to the next layer. The

networks are usually set up so as to be feed-forward meaning a neuron feeds its output to all the neurons on the next layer without feedback to the previous layer. Specific weights will be applied to the signals moving from one neuron to another and the weightings could be tuned during the training to improve the neural network to the specific problem being solved.

## 2.6 Summary of identified gaps

While a lot of research has been done on customer churn before there still exists some gaps particularly in proactively managing the problem.

These gaps include: how predictive analytics can be applied in churn management, the role/benefits of using machine learning in solving the churn problem, which features to use and their importance, which machine learning algorithm should be used and the reliability and accuracy of the machine learning models.

## 2.7 Proposed Solution

The figure below illustrates how the proposed solution works

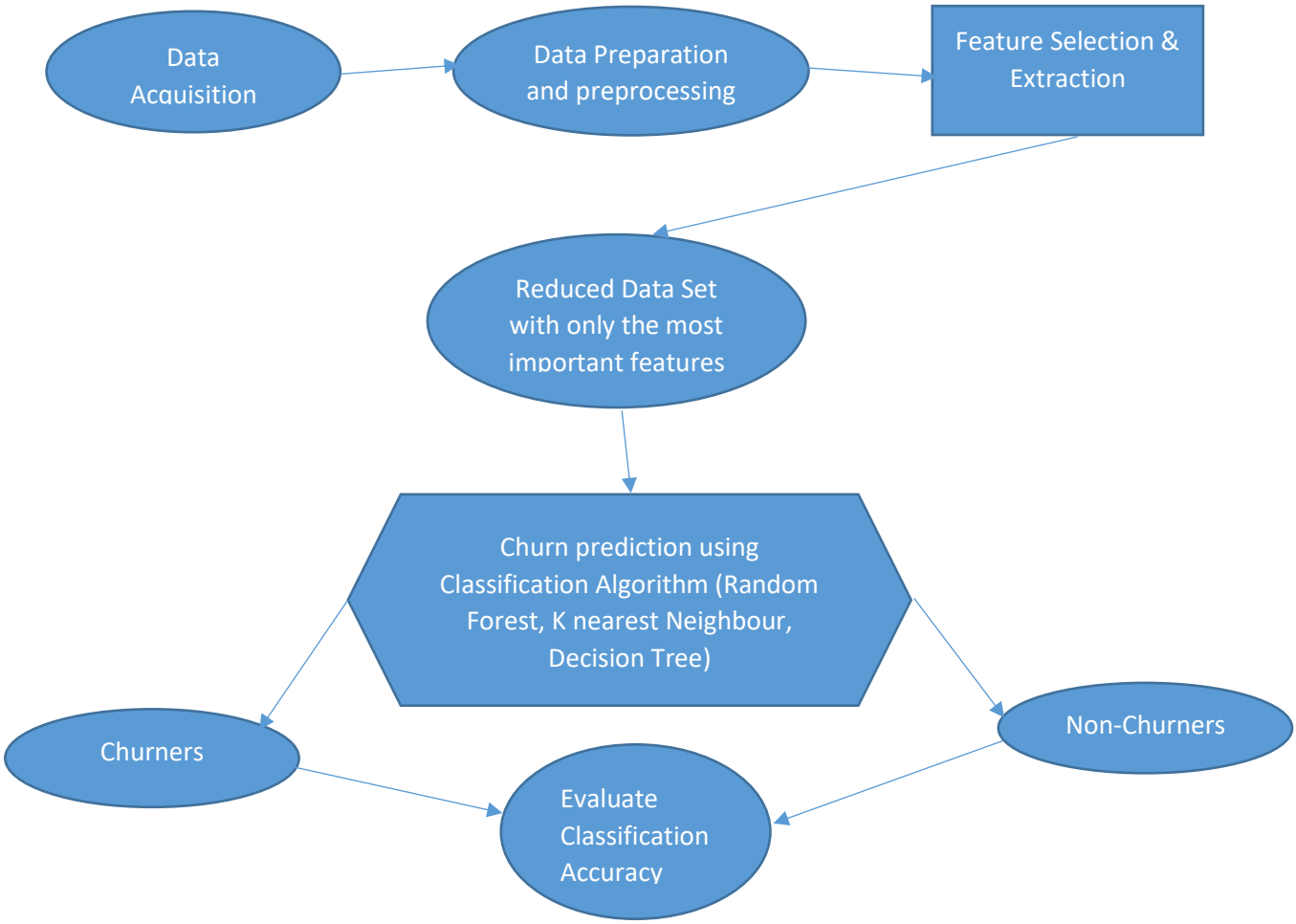


Figure 2. 1 Illustration of proposed solution

### 2.7.1 Data Acquisition

Data acquisition is the process of getting a dataset like configuring an API, internet, querying a database, etc. It also involves understanding the data-set, the definitions and contents of all variables. In this case data was retrieved from Safaricom's data warehouse, a relational database, by writing queries.

### 2.7.2 Data Preparation and preprocessing

This is the most important part of building a robust prediction model. When data is incomplete, or some values are missing, the performance of the model will be affected in a negative way. After checking accuracy of the data, completeness of the data and any other anomalies, corrective action was taken such as filling null values, conversion of data types, date formats, etc. The aim was to have the cleanest data possible before using it in the prediction algorithm.

### 2.7.3 Feature Selection & Extraction

Feature selection works by reducing the number of features or variables when implementing a predictive model. Reducing the number of features helped reduce the computational cost of modeling resulting in improved performance of the model. The feature selection is done by examining the relationship between each input variable and the output variable by use of statistical methods that selecting those input variables that have the highest correlation with the target output variable.

Feature extraction involved deriving new and more useful features from existing features that would have a higher importance in the model.

### 2.7.4 Reduced Data Set

This dataset was a subset of the original dataset but was smaller as we had only kept the most important attributes/features. Out of all features that were obtained from the customer only the most important features were fed to the algorithm for predicting churn.

### 2.7.5 Prediction using Classification Algorithm

Classification, a type of supervised learning was used for prediction with targets being churner or non-churner. Churners were where our focus lied on and that's where the company can implement a retention strategy. Four of the algorithms discussed earlier were implemented, the four were chosen because of their ease of use, high accuracy in previous implementations and known practicality.

### 2.7.6 Evaluate Classification Accuracy

Evaluating performance of a machine learning algorithm is an essential part when creating any model. A model may give satisfying results when evaluated it using a metric such as classification accuracy (correct predictions / total number of samples) but it may perform poorly under other performance metrics such as Area under Curve or F1 Score. Since we were more interested in the actual positives (churners) and not negative (non-churners), we used f1 score to evaluate the

performance of the algorithms. This helped choose the best algorithm among the four that were implemented.

# 3 Research Methodology

## 3.1 Introduction to Research Methodology

Research Methodology is a procedure for conducting your research. Leady & Ormrod (2001) define it as "the general approach a researcher uses in carrying out the research project". The methodology that was used originated from the exploration question and not from my own inclination for one design or another.

Research methodology is a logical approach to solving an issue. It's a science that shows how a research shall be administered. Basically, it is the procedures which enable researchers carry out their work of describing, explaining and predicting various occurrences. It's also the methods by which knowledge is acquired with a purpose of coming up with a work plan for the research. Research can be grouped into three major categories: quantitative, qualitative and mixed methods research (Creswell et al., 2008).

### 3.1.1 Quantitative Research

Quantitative implies using numerical data to quantify. Quantitative methods usually rely on experiments and surveys to collect measurable data such that statistical processes can be applied (Creswell, 2003). Measuring of variables is done through use of a numerical system, analyzing the measurements using a range of various statistical methods and reporting relationships and associations between the variables under study.

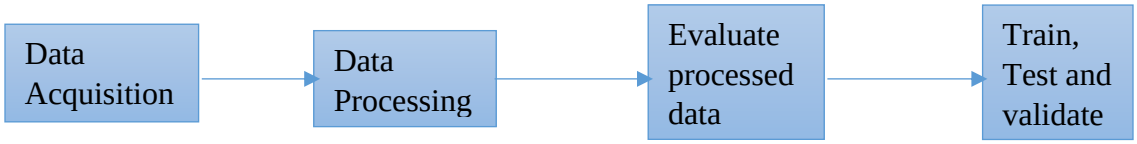
### 3.1.2 Qualitative Research

Qualitative methodologies are utilized to analyze and evaluate non-numerical information like emotion and behavior. It consists of open-ended information that the researcher usually gathers through interviews, focus groups and observations. Applicable examples are studies that involve relationships between individuals, individuals and their environments, and motives that drive individual behavior and action.

### 3.1.3 Mixed Methods Research

By utilizing mixed methods research, researchers make use of methods of collecting or analyzing data from the quantitative and qualitative research approaches in the same research (Creswell, 2003; Johnson & Onwuegbuzie; Tashakkori & Teddlie). Its central premise is that the employment of quantitative and qualitative approaches together gives a deeper understanding of the research question than either of the approaches separately.

This research used mixed methodology by trying to understand qualitative data such as behaviors of customers who churn vs those who do not and quantitative data by numerical measurement of various variables. Below is an outline of how the research was conducted:




## 3.2 Research design

Research design is a plan for selecting subjects, research sites and the data collection procedure to answer the research questions (MacMillan Schumacher, 2001). Research design is the logical framework for implementation that acts as a bridge connecting research question to the implementation of the actual research plan.

The research design was the systematic approach that we would use to carry out the study. It consisted of the general strategy that we preferred to integrate the various aspects of the study in a continuous and logical way therefore making sure we effectively addressed the research problem. It was made up of the blueprint for the gathering, measurement, and analysis of knowledge.

This research used experimental research design, a quantitative method. The research was a collection of research designs which used manipulation followed by controlled testing to establish the causal processes. Either one or more variables under consideration were manipulated to determine their effect on a dependent variable (Oskar Blakstad, 2008). Experimental Research is preferred in situations where:

- Cause comes before the effect that is there is time priority in the causal relationship.
- Presence of consistency during the causal relationship whereby the cause always leads to the same the identical effect.
- The level of the correlation is high.

Experiments are conducted in order to be able to predict some phenomenon. Typically, the experiment is constructed so as to explain some kind of causation(Oskar Blakstad, 2008).

This study explains how various customer variables such as age, revenue, usage can have a causal effect on churn. These attributes were then used to predict future outcome, that is customers who churned.

## 3.3 Data Acquisition

Data acquisition was the process used for collecting and organizing information. The first step for every data science project is data collection, that is, getting the actual raw data (Tomi, 2016).

Telcom companies have huge datasets regarding their customers which is usually stored in their data warehouse. However only a small portion of it was needed for this exercise, hence the need to extract only the relevant data in a structured format.

### 3.3.1 Types of data

There are 3 categories of data that were acquired and used in our machine learning model.

- **Demographic data** – This helped identify customers based on information like age, gender, geographic location, etc. Demographic data is crucial to understanding what groups may be likely to perform in what way.
- **Behavioral data** – This data helped to uncover underlying patterns that customers reveal during their interactions with the company. It included transactional data and usage data. We looked into customer consumption of products and services and usage by using call data records (CDRs).
- **Engagement Data** – This data told us how customers interact with the company brand via various avenues. The telecommunication company stores data for such interactions including calls and SMS to the call center, interaction on social media and chatbot interactions.

### 3.3.2 Acquisition process

Subscriber data was acquired from the telecommunication company by querying from their data warehouse for only what we need from. The variables required were identified in advance and queries developed to fetch this specific data from the various tables in the data warehouse. Measures were taken to ensure this was in accordance with the data protection policies. These measures included masking subscriber phone numbers and ensuring no Personally identifiable information (PII) was queried. This data was then saved in a single file of csv format which is what the solution consumed as the input.

## 3.4 Data Processing

Data collected and compiled from any source should be accurate and complete and therefore must be checked for accuracy and adequacy before proceeding further, in this case before training the machine learning algorithm. Data processing refers to a logical set of procedures performed on data so as to validate it, organize, transform, integrate and extract information in a logical output form for further use.

The major objective of data preprocessing was to resolve data quality issues. As illustrated in fig 2.5.1 of the proposed solution, this was among the major steps. Data was processed as below:

### 3.4.1 Exploratory data investigation

Exploratory data investigation involves examining the data for errors and describing data using summary statistics and visualizations. We can also detect anomalies such as outliers using plotted graphs.

### 3.4.2 Data Noise

This includes Missing or incorrect values. This is a common problem in data whereby due to various reasons the data is missing and incorrect. If used without correction this usually leads to poor performance of the model. Incorrect values were imputed using various methods such as mode and mean. For example, subscribers whose age was missing had the mean age assigned to them.

### 3.4.3 Redundancy

This anomaly involves duplication of data e.g. the same subscriber and the related attributes appear more than once in the same dataset. This was be resolved by selecting distinct data records from the dataset.

### 3.4.4 Outliers

An outlier is an object whose values deviate significantly from the remainder of the objects. This is usually caused by measurement or execution error for example due to mistakes in data entry and data processing errors. Z-score was used to detect the outliers and some of the outliers were deleted while others had their values updated with imputed values. For example, subscribers who had abnormal usage or had abnormal ages were removed from the data.

### 3.4.5 Handling Categorical Variables

Categorical variables are explained as variables that are discrete and not continuous in nature. An example is phone type: dual vs non-dual. Most of the categorical variables were handled using One-Hot Encoding works by x columns where x is the number of unique values that the categorical variable holds. For the x columns, only one column could have a value of 1 and the rest all had their value as 0. For the column of whether a phone is dual or non-dual, if dual column is 1 then non-dual is 0 and vice versa.

### 3.4.6 Standardization/ Scaling

During Standardization we transformed all values resulting in the mean of all values becoming 0 and standard deviation being 1. For example, some of the variables being used, Age which will usually be 2 digits and revenue which can even reach 5 digits. The two are not on the same scale and revenue will usually be higher scaled than Age. This would make our model to give higher weightage to revenue which is not the ideal scenario as age is also an important feature. We standardized using minmax scaler which had 0 as the minimum and 1 as the maximum. All values were scaled to fit to fit within the scale of 0 to 1.

### 3.4.7 Feature selection/ prominent feature identification.

This was the process of reduction of the number of input variables by checking which variables have the strongest relationship with the outcome. This was done by using statistical methods such as chi2, which showed us how important the variable's relationship was to the output. Some of the features that were tested include: Biographical Age, Age on network, revenue, phone type, location, interactions such as calls to call center, social media.

### 3.4.8 Feature Creation and Extraction.

This involved combining existing features to create new, more useful features that didn't already exist in the dataset which had a higher importance in the model. One such feature was date of birth which on its own was not useful, however a new more powerful feature called age was created.

## 3.5 Interpretation and evaluation of the Processed data

This was the last step and generally involved examining quality of the processed data and whether it was sufficient enough for building the model. We checked whether the anomalies that existed

before processing had been fixed. The processed data was sufficient to be used in our models and we proceeded to the next step.

## 3.6 Training, Validation and Testing

Various datasets were used at different stages while training the models and doing the actual churn predictions. Below is how the dataset was split.

### 3.6.1 Training

The Training dataset was the sample of data that was used to fit the model. The models learnt from the training data and usually the more the data that is available the better skilled the models became. The data used consisted of three months historical subscriber data which was used to train the classification model. Since this data was historical the outcome of whether the subscriber would churn or not was already known.

### 3.6.2 Validation

Validation dataset was the sample of data that was used to give an impartial assessment of how the models fitted on the training data while optimizing various features and model hyperparameters.

Cross Validation, specifically K fold, was used for validation in order to minimize over fitting or under fitting during training. This was done by randomly splitting the set of training observations into k groups of the same size. The training data was divided into 10 folds meaning value for k was fixed to 10. 10 was chosen because it is a value that has been found through previous experimentations to majorly give a model skill estimate with low bias and modest variances. The first fold was set aside as a validation set and the algorithm was fitted on the remaining k-1 folds, this was done for all the 10 folds.

### 3.6.3 Testing

Test Dataset is part of the data that was used to provide a true evaluation of a final model fit based on the skills acquired from the training dataset.

Once the models had completely trained (using the train and validation sets), their skills were tested by predicting the outcome of the test dataset. The test data used one months' data, the month succeeding the three months of training data.

Evaluating the performance of the models for prediction accuracy was done using F1 score. This was the preferred metric as it is most suited where we have an imbalanced dataset. The dataset we had was highly imbalanced, subscribers who churned were 2% while those who did not churn were 98%. F1 score conveys a balanced average between the precision and recall and is given by 2*((precision * recall) / (precision + recall)).

# 4 Data Presentation, Analysis And Discussions

## 4.1 Introduction

In this chapter we analyze the outcomes from the research and the prototype built, discuss them and present. We look at the extent of customer churn, analyze various attributes of churn customers, present the features that were used for prediction and their importance, discuss implementation of the machine learning model and analyze performance of various classification algorithms. The total customer base that was used is 20M subscribers.

## 4.2 Extent and Impact of Customer Churn

Customer churn happens once the subscribers cease consuming services from the company or stop purchasing its products. A customer was considered to have churned for a particular month if they were not active in the succeeding month. A customer is considered active if they have usage or generate revenue.

While the telecom company under research reports its churn rate annually this research looked at the monthly churn rate that is the number of subscribers lost compared to the company's overall subscribers count per month as a percentage.

The below chart gives a view of customer churn for the 6 months September 2019 to February 2020. The churn is shown as a percentage and was calculated by dividing the number of customers who are no longer active in the succeeding month by those who are still active then multiplying by 100. This was based on the entire dataset of 18M subscribers. The revenue loss was estimated by multiplying number of customers who have churned by the Average Revenue Per User (ARPU).

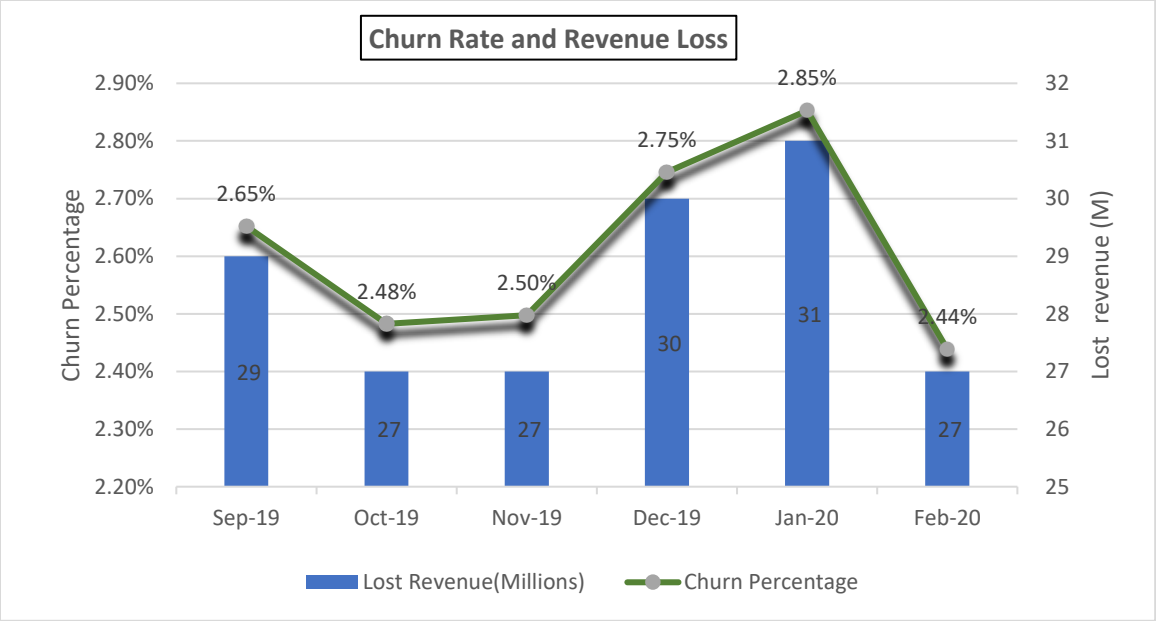



Figure 4. 1: Monthly Churn Rate and Lost Revenue

**Churn Rate Summary** – January had the highest churn rate at 2.85 % with February being the lowest at 2.44%. The average churn rate for the six months was 2.6%.

**Lost Revenue Summary** – The lost revenue is directly proportional to the Churn Rate. January 2020 had the highest lost revenue standing at 31M while October 2019, November 2019 and February 2020 had the least lost revenue at 27M. Average monthly loss stood at 28 Million KES while the cumulative loss for the 6 months totaled to 173 Million.

## 4.3 Exploratory Data Analysis

Data exploration is a method of analyzing data sets in order to summarize the main attributes and it is typically done with visual strategies. Since we are attempting to classify our subscribers into either churned or Not churned we shall visualize the data by each characteristic while comparing and contrasting the two groups.

### 4.3.1 Churn Rate by Gender

In this section we look at how churn rate compares across the genders that is male and female. The chart below shows this comparison.

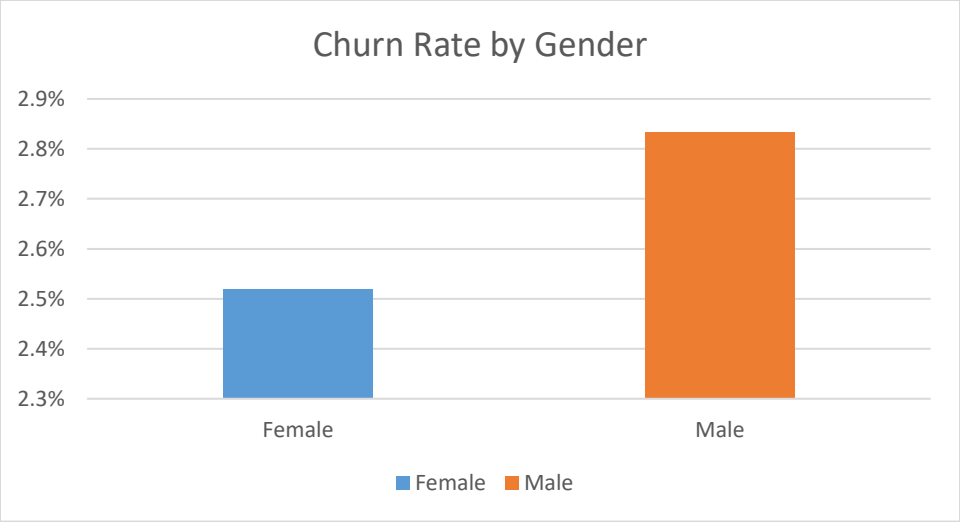



Figure 4. 2: Churn Rate by Gender

From the analysis it is clear that males are more likely to churn with a churn rate of 2.8% as compared to women who have a churn rate of 2.5%.

### 4.3.2 Distribution of Subscribers and Churn Rate by Age

The subscribers ages were grouped into 4 age groups: under 35, 35 – 50, 50 -65, 66 and above. These age groups were chosen based on similar researches that had been done before. The chart below shows a comparison of churn rates and total subscribers across various age groups

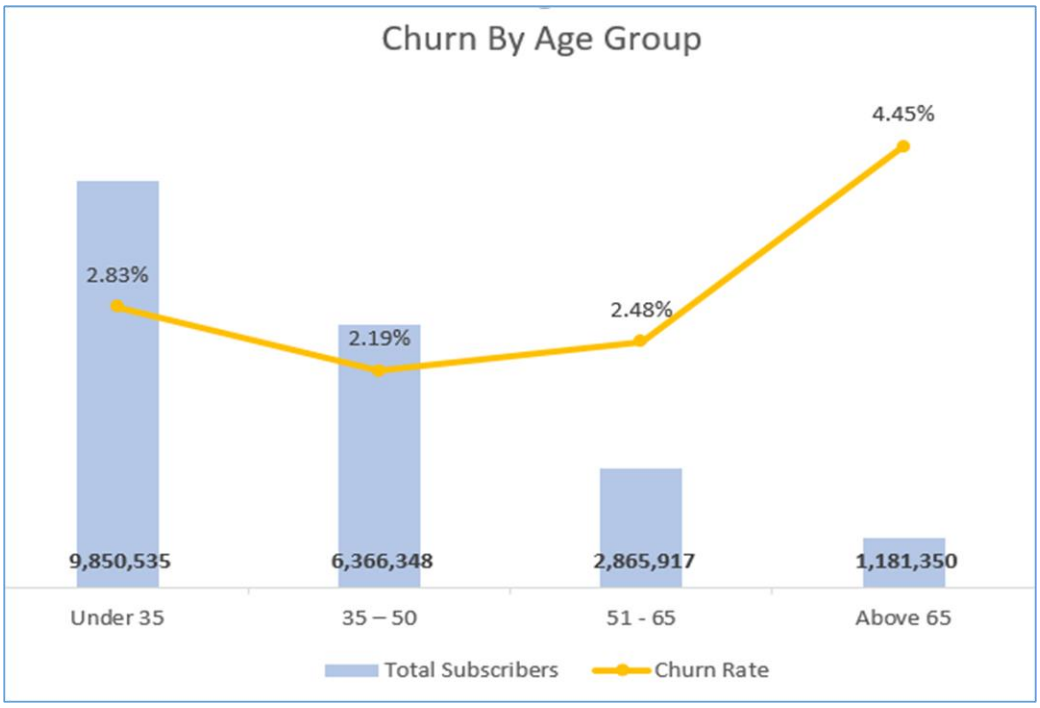


Figure 4. 3: Churn by Age Group

From the analysis it was seen that the distribution of customers above the age of 65 have the highest chance of churning with a churn rate of 4.45% while customers aged between 35 to 50 had the lowest churn rate at 2.19%

### 4.3.3 Distribution of Churn Rate by Number of Services Consumed

We checked whether there was a relationship between the number of services a subscriber consumes and their likelihood to churn. These services include: Calls, Data, Message, MPESA, Loan, Personalized Offers and Subscriptions. The chart below shows the relationship.

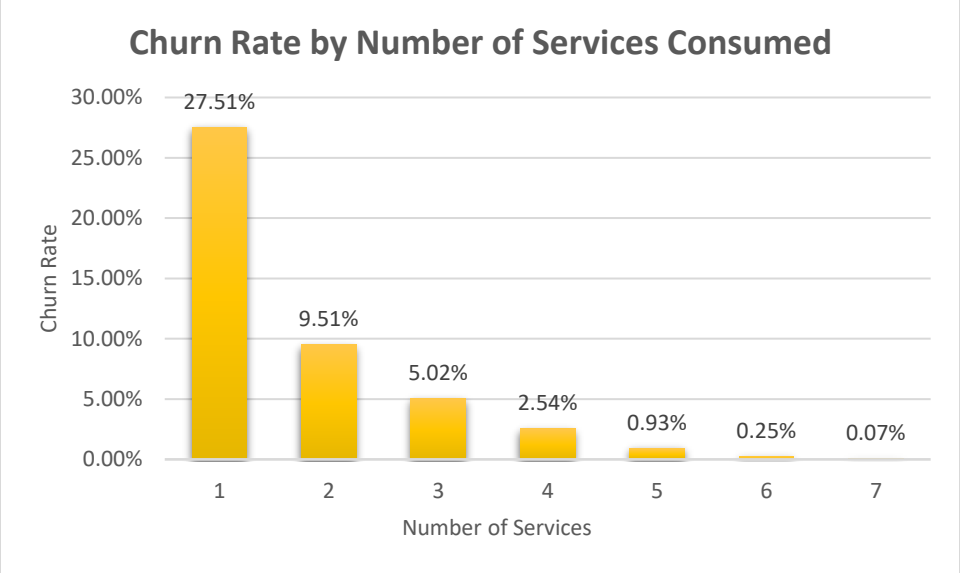



Figure 4. 4: Churn Rate by Number of Services Consumed

It was seen that subscribers who only consume a single service are most likely to churn and the churn rate decreases as the number of services consumed increases. This is in line with our expectations whereby it was more difficult for a customer to leave a business if they consumed a lot of products or services. Customers who consume fewer services also tend to be less loyal.

### 4.3.4 Churn rate based on registration document

In this section we analyze churn based on the document that a subscriber used to register their line. These documents were broadly classified into 4:

- Kenyan – This is any Kenyan document used to register a line such as Kenyan national ID or passport.

- Foreigner – This is any foreign document for example foreign ID, foreign passport used to register a line.
- Organization – This is the document when an organization registers lines under its name.
- Other – These are special documents used to register a line such as special requests from the government, emergency documents, among others

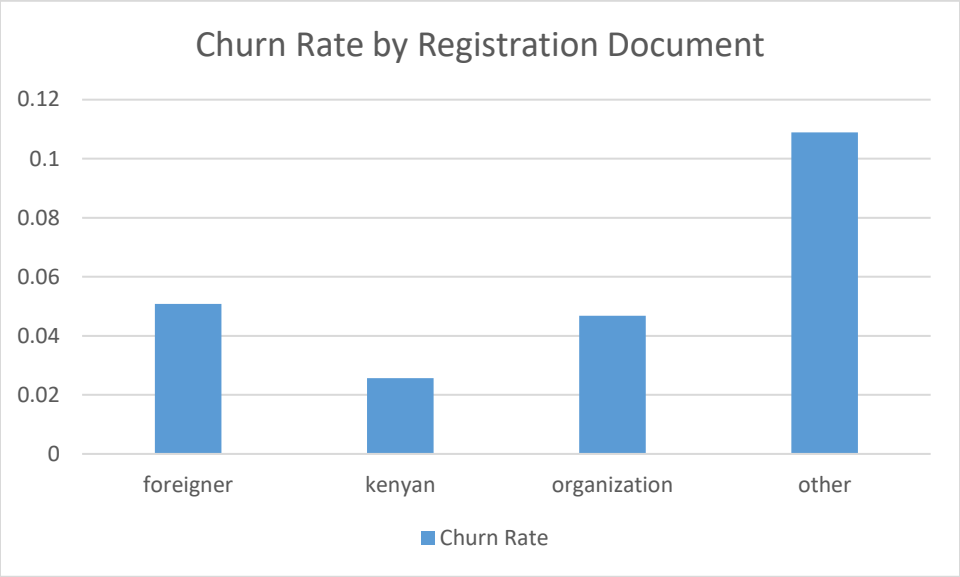



Figure 4. 5: Churn Rate by registration Document

As it can be seen from the analysis people who register with documents classified as other have the highest churn rate. This is mainly because such lines are usually required to be used for a short period. People who register lines using foreign documents have the second highest rate, this was attributed to their temporary stay in the country. Lines registered under companies come third in terms of churn rate while lines registered under individual citizens have the least churn rate.

## 4.4 Data Preprocessing

### 4.4.1 Missing values

Taking care of missing values is a very important preprocessing task that can drastically affect the models when not done with sufficient care. Example of missing values can arise due to data entry mistakes such as not entering a subscribers age into the system. Before preceding to other steps an analysis was ran on the data we obtained to check for missing values. The results were as below:

Table 4. 1 Missing Feature Values

| Feature | Total Missing Values | Missing% |
|---|---|---|
| Subscriber age | 36 | 0 |
| Number Of Services Consumed | 0 | 0 |
| Age on Network | 0 | 0 |

| | | |
|---|---|---|
| Registration Document | 0 | 0 |
| Talk Time | 0 | 0 |
| Total Revenue | 0 | 0 |
| Data Usage | 0 | 0 |
| Number of Lines | 0 | 0 |
| Gender | 0 | 0 |

From the results we can see only age had few missing values. Out of the entire base of 1M subscribers only 36 had their age missing. The missing values were corrected by imputation whereby mean age of all subscribers was used.

### 4.4.2 Categorical features

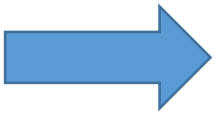
A categorical feature is any feature that is categorical in nature and represents discrete values which belong to a limited set of categories or classes. Unfortunately, machine learning algorithms don't support handling categorical data and it is therefore necessary to convert categorical features to a numerical representation. This conversion is usually done through an encoding scheme.

While there are several encoding schemes for dealing with categorical data, it is important to establish if the feature being used is ordinal or nominal. Ordinal features was explained as features with natural ordered categories and the distances between those categories is unknown.

There were two categorical features in the data we obtained that is Gender and Registration. Both of these features were nominal making the preferred method to be one hot encoding

Gender was encoded as below

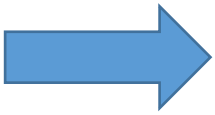
| **Gender** |
|---|
| Male |
| Female |



| **MALE** | **FEMALE** |
|---|---|
| 1 | 0 |
| 0 | 1 |

Registration Document was encoded as below

| **Registration Document** |
|---|
| Local |
| Foreign |
| Other |



| **Local** | **Foreign** | **Other** |
|---|---|---|
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 1 |

### 4.4.3 Feature Scaling

Feature scaling refers to putting the feature values in the same range or same scale so that no variable is dominated by the other. When a feature's variance is larger than that of other features, it might dominate the objective function therefore making the models unable to learn from other features sufficiently.

MinMaxScaler was chosen as the preferred method for scaling. It transformed the features by scaling them to the range of 0 to 1. This range was set by specifying the feature range parameter with the default values (0,1). By applying this, the data was scaled such that the lowest value for any feature was 0 while the highest value was 1. Below is a comparison of the data before and after scaling.

Table 4. 2 Sample data before scaling

| DOC_TYPE | IS_MALE | QT_LINES | TOTL_REV | DATA_USAGE | CALL_SECS | REVN_TYPES | Age |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0.008375 | 0 | 1 | 51 |
| 1 | 1 | 1 | 830.2361 | 0.739605 | 8756 | 6 | 48 |
| 1 | 1 | 1 | 243.9729 | 0 | 4181 | 4 | 62 |
| 1 | 1 | 1 | 3514.425 | 1283.651 | 18918 | 7 | 56 |
| 0 | 1 | 1 | 0 | 0.2023 | 75 | 3 | 30 |

Table 4. 3 Sample data after scaling

| DOC_TYPE | MALE | QT_LINES | TOTL_REV | DATA_USAGE | CALL_SECS | REVN_TYPES | age |
|---|---|---|---|---|---|---|---|
| 0.333333 | 1 | 0.0000000 | 0.0000017 | 0.00000 | 0.00000 | 0.00000 | 0.53691 |
| 0.333333 | 1 | 0.0000000 | 0.0000724 | 0.00000 | 0.00018 | 0.83333 | 0.51678 |
| 0.333333 | 1 | 0.0000000 | 0.0000225 | 0.00000 | 0.00009 | 0.50000 | 0.61074 |
| 0.333333 | 1 | 0.0000000 | 0.0003010 | 0.00005 | 0.00039 | 1.00000 | 0.57047 |
| 0 | 1 | 0.0000000 | 0.0000017 | 0.00000 | 0.00000 | 0.33333 | 0.39597 |

## 4.5 Features Influencing Churn

In this section we look at what's predicting the telecom's churn which are also known as features predicting. These were the measurable properties and characteristics of the churn phenomenon. From previous researches, related work and industry practices there are multiple features that could be used to predict churn. Some of these features include: Device model and brand, Customer age, Carrier tenure/Network Age, Subscriber Revenue, Subscriber usage, days of usage, location, mode of payment (pre-paid or post-paid), Time spent on network, etc.

The data that was obtained had 9 features that could help predict how likely a mobile subscriber is to churn to a different mobile network carrier.

### 4.5.1 Feature Definitions

The below table describes the features that were in the data obtained and what each feature means.

Table 4. 4 Churn features and their descriptions

| Feature | Feature Description |
|---|---|
| Registration Document | The document the subscriber used to register their line |
| Number Of Services Consumed | Number of services the subscriber uses e.g. calling, data, SMS, Subscription services, personalized offers and mobile money. |
| Age on Network | How long a subscriber has been using the Telco's services in years |
| Subscriber age | Actual age of the subscriber in years |
| Talk Time | Amount of time spent on calls |
| Total Revenue | Amount of money spent to consume various services |
| Data Usage | Amount of data used for internet |
| Number of Lines | Number of lines the subscriber owns |
| Gender | Gender of the subscriber |

### 4.5.2 Feature Importance

Statistical tests were used to show the relationship between the features and the output variable. The specific test chosen was chi-squared (chi²) statistical test because our model was being used for classification and the data did not contain negative values. This test gave us a score for each of the features and the bigger the score the more important a feature was towards influencing our output variable of churn.

The chart below shows how the features compared in order of their significance

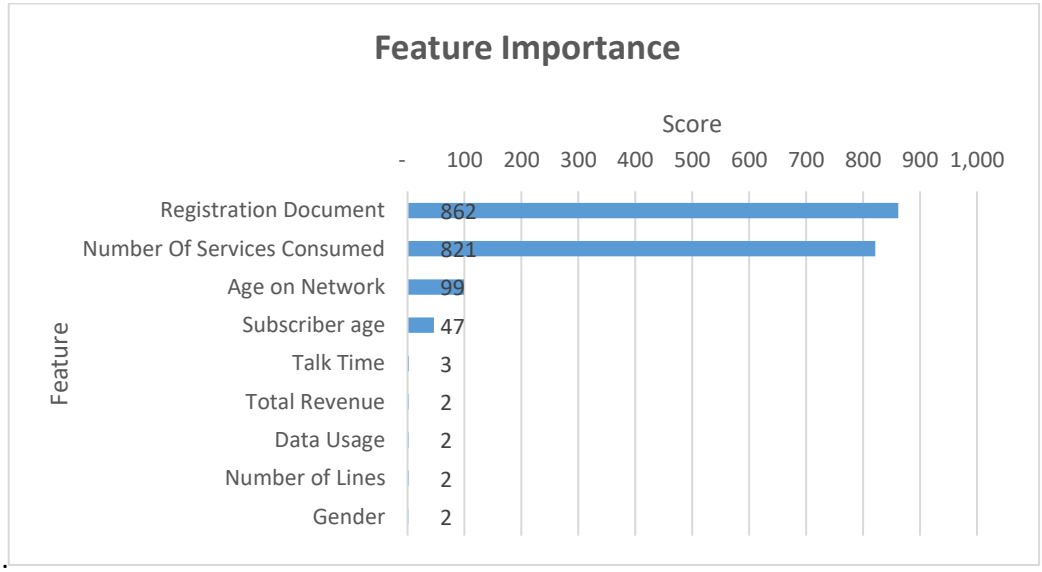



Figure 4. 6: Feature Importance

### 4.5.3 Feature selection

Feature selection was a crucial part in the machine learning and it significantly impacted the skill of the models. It was the process where we automatically chose the variables which contributed the most to our prediction output.

Benefits of performing feature selection prior to modeling our data included:

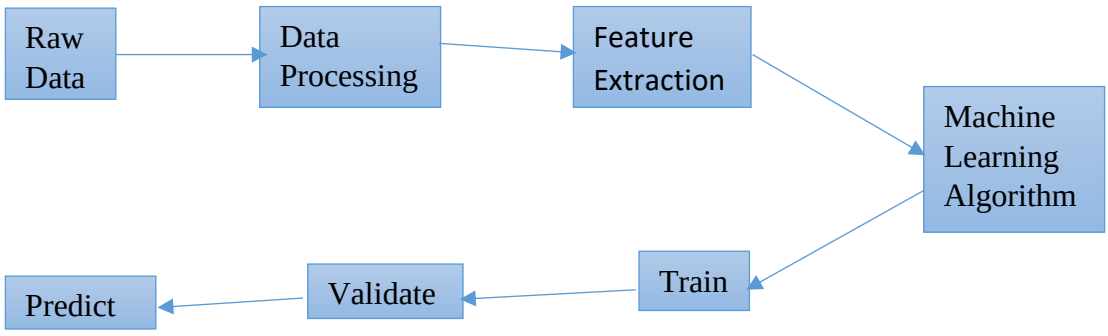
- Overfitting reduction due to less redundancy which meant lower bias to make decisions affected by noise in the data.
- Improved Accuracy: Less misleading data meant our modeling accuracy could improve, therefore churners vs non-churners were more accurately predicted.
- Lower Training Time: By reducing the features algorithm complexity reduced and the algorithms trained faster and used less computational resources.

Having performed tests for feature importance on all the features that were available, 5 out of 9 features were selected to be used in the prediction of churn. The 5 features that were selected were: Registration Document, Number Of Services Consumed, Age on Network, Subscriber age and Talk Time.

## 4.6 Implementation of the Classification Algorithms

With a clear understanding and knowledge of the problem and the data available, the next milestone was identifying the algorithms that were relevant and practical to implement. Some of the elements affecting the choice of the classification algorithm were: accuracy, interpretability, complexity, scalability of the model and how long does it take to build, train, and test the model.

Four suitable machine learning classification algorithms were identified and implemented. The chosen algorithms were: Random Forest, K Nearest Neighbors, Naïve Bayes and Neural Network. High-level architecture of how the algorithms were implemented to work is as below:



## 4.7 Performance of the Classification Algorithms

Every model that we use for predicting the true class of the outcome variable (those who churned and those who didnt) is bound to have errors. This result in incorrect classification whereby the subscribers are incorrectly classified, also known as false positives and false.

After implementation the algorithms were tested by checking their performance during cross validation and actual predictions on test data. In addition to this the execution time was measured to ensure that the algorithm runs in a reasonable amount of time.

### 4.7.1 Performance Metrics

#### *4.7.1.1 Confusion Matrix*

Even though confusion matrix was not a metric to measure performance, most of the performance metrics were calculated based on confusion matrix and the respective values inside it. The confusion matrix can be visualized as a table with two rows and columns for actual versus predicted values. It takes the format below:

| | | Actual | |
|---|---|---|---|
| | | Positive | Negative |
| Predicted | Positive | TP | FP |
| | Negative | FN | TN |

True Positive means that the customer actually churned and the model predicted they churned, True Negative is where a customer actually churned and the model predicted correctly that the customer did not churn, False Positives are where a customer who did not churn is incorrectly predicted by our models as having churned and finally false negatives are the cases when the customer actually churned but was falsely predicted as not churned.

#### *4.7.1.2 Recall/Sensitivity*

Recall is used to show what size of subscribers who truly churned were classified by the algorithm as having churned. It aims to establish what proportion of actual positives were classified correctly. Using values from the confusion matrix, recall is calculated as below:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall is not only about capturing cases correctly like in classification accuracy but rather measuring all cases predicted as churned and comparing to those that have actually churned.

#### *4.7.1.3 Precision*

Precision shows what proportion of subscribers that were classified as churned that had actually churned. It can be calculated as below.

$$\text{Precision} = \frac{TP}{TP + FP}$$

#### 4.7.1.4 F1 Score

Precision and recall will usually be in tension. Meaning that improving precision will result in a reduction of recall and vice versa. To fully evaluate effectiveness of the models it was necessary we examine both precision and recall.

$$\textbf{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1 score was therefore the metric that was used to measure the performance of the algorithms that were implemented. F1-score was the preferred metric because of the imbalanced classes in the case study, Customers who churn per month average at about 2% while those who do not churn around 98%.

### 4.7.2 Cross Validation

Cross-validation is a statistical way of estimating the skill of a machine learning model (Jason Brownlee, 2018). The method had a parameter denoted by k that represent the quantity of groups which the data sample was divided into, hence it's also known as k-fold cross-validation. In the models that were implemented the value for k that was chosen was 10 therefore making our k-fold cross validation a 10-fold cross validation.

The table below shows how the models performed in validation

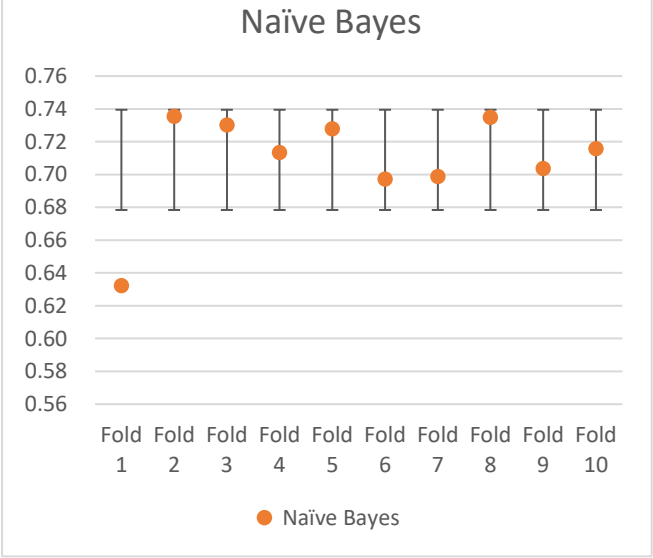
Table 4. 5 Algorithms performance during validation

| Classification Model | Random Forest | Naïve Bayes | K Nearest Neighbor | Neural Network |
|---|---|---|---|---|
| Fold 1 | 0.69 | 0.63 | 0.64 | 0.61 |
| Fold 2 | 0.75 | 0.74 | 0.69 | 0.64 |
| Fold 3 | 0.73 | 0.73 | 0.68 | 0.63 |
| Fold 4 | 0.74 | 0.71 | 0.62 | 0.65 |
| Fold 5 | 0.74 | 0.73 | 0.69 | 0.62 |
| Fold 6 | 0.71 | 0.70 | 0.66 | 0.68 |
| Fold 7 | 0.70 | 0.70 | 0.66 | 0.66 |

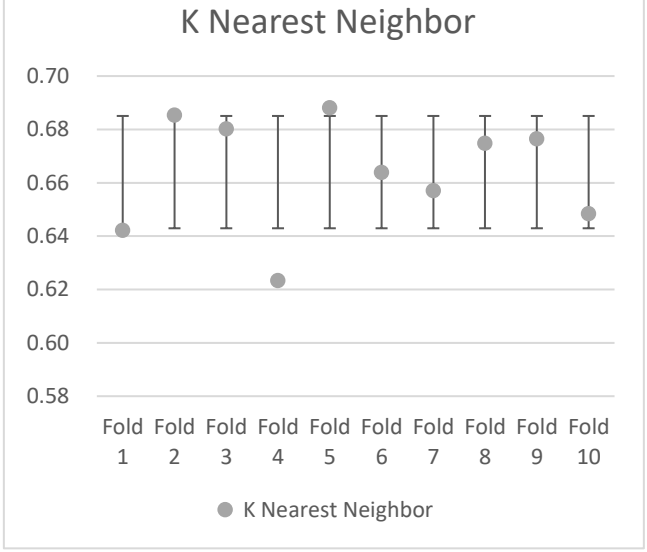
| | | | | |
|---|---|---|---|---|
| Fold 8 | 0.73 | 0.73 | 0.67 | 0.65 |
| Fold 9 | 0.73 | 0.70 | 0.68 | 0.64 |
| Fold 10 | 0.71 | 0.72 | 0.65 | 0.62 |
| **Mean** | **0.73** | **0.71** | **0.66** | **0.64** |
| **Standard Deviation** | **0.0200** | **0.0305** | **0.0211** | **0.0218** |

The mean shows the average f1 score of the model across the 10 folds, the higher the score the better. Random forest had the highest accuracy at 0.73, followed by naïve Bayes at 0.71, K Nearest Neighbor was third at 0.66 while neural network was last at 0.64

Standard deviation measures the amount of variance in a set of values. In our case, a lower standard deviation is preferred across the outcome for the different folds. The plots below explain further the standard deviation by model across all folds.




**Standard Deviation** = 0.0305




**Standard Deviation** = 0.0211

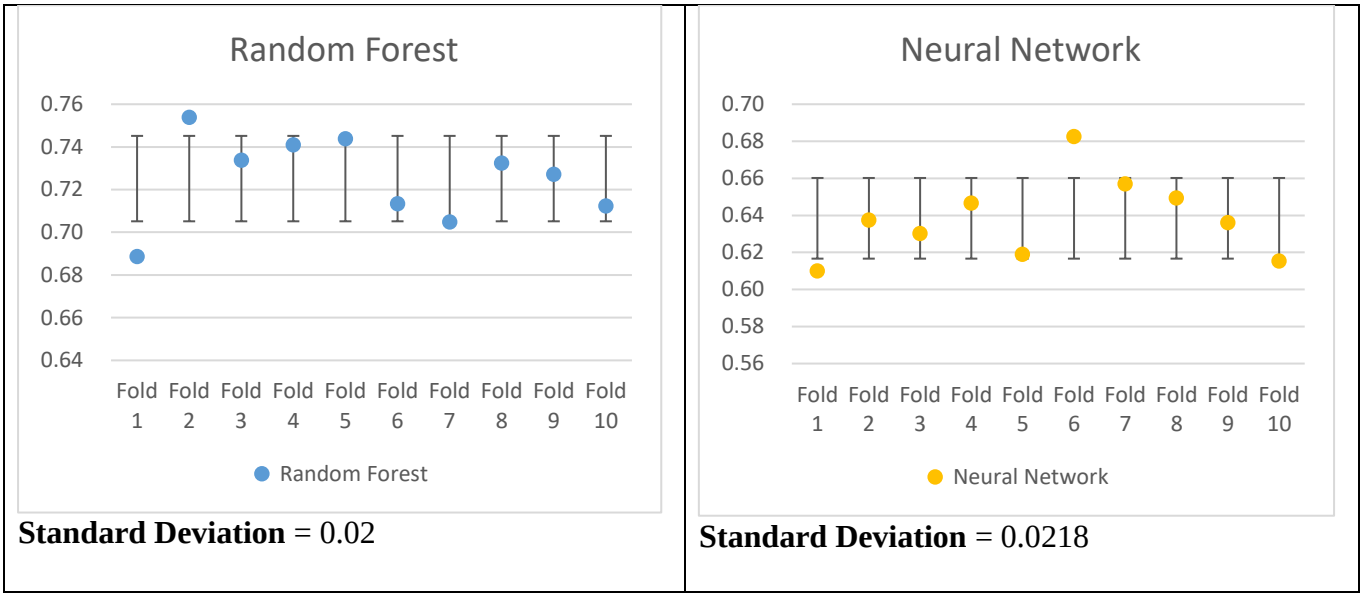



Figure 4. 7: Standard Deviation During cross Validation by Algorithm

A low standard deviation indicates that the values in the observation are near to the mean of the set. The lower the standard deviation the better as it means the model is performing similarly across the different folds. High standard deviation would mean that the values are scattered out over a large range and that our machine learning model is overfitting or underfitting in some of the folds. Random forest had the lowest standard deviation at 0.02, K Nearest Neighbor at 0.0211, neural network at 0.0218 and Naïve Bayes at 0.0305.

Generally, all models had a very low standard deviation in the cross validation which combined with the good f1 scores was an indicator that all the models performed fairly well.

### 4.7.3 Performance of the Algorithms on Actual Test Data

Having validated the four algorithms and with a clear understanding of the skills of our model, it was now time to evaluate how well our algorithm performs on the actual test data. For this we used the whole training data set to train and tested using the test data for which accuracy was measured using F1 score. The execution time was also measured for each algorithm which includes the duration it took to train the model and the duration to predict outcomes for the test data. The results are summarized by the table below:

Table 4. 6 Algorithm accuracy scores and Execution Time

| Model | Test F1 Score | Execution Time (Sec) |
|---|---|---|
| naive Bayes | 0.69 | 1 |
| knn | 0.65 | 19 |

| | | |
|---|---|---|
| neural network | 0.63 | 15 |
| random forest | 0.71 | 14 |

While the accuracy score did not vary greatly across the four algorithms, execution time had huge variations.

From the outcome we can see we are getting very good accuracy across all the algorithms as we are the f1-score is above 0.6 for all models. However, in terms of execution time, the results vary greatly with Naïve Bayes being the fastest in execution completing at only 1 second and KNN being the slowest at 19 seconds.

## 4.8 Summary of findings

The extent of the customer churn in the telecom company under research stood at 23.45% annually and averaged at about 2 percent monthly from the data analyzed. Even though the figure seems small, the equivalent revenue that is lost is staggering, billions of shillings every month. Cumulatively, the company lost more than 50B in the year 2019 due to churn. This showed huge impact of churn and highlights the need for churn management in the company.

Features influencing churn (Churn indicators) were also identified. These are the subscriber attributes that were used to show whether a subscriber will churn or not. These features were very crucial as they helped us filter out unwanted data and also influenced the prediction accuracy. The top 5 features influencing churn were identified as: Registration Document, Number Of Services Consumed, Age on Network, Subscriber age and Talk Time.

Since churn prediction could be done using several classification algorithms, the research tried to establish which algorithm would be best. In deciding the most suitable algorithm we looked at its performance during validation, its execution time and accuracy during prediction on actual test data. The algorithm that performed best under the 3 metrics was random forest, it had the highest accuracy and execution time even though not the fastest was within a reasonable range.

# 5 Achievements, Limitations, Recommendation and Further Research

In this chapter we make a summary of findings, highlight research limitations, make suggestions for further research and conclude.

## 5.1 Achievements

Prior to the research we set the main objective as to use predictive analytics to identify subscribers who are likely to churn from a telecommunication firm. With Specific objectives being to identify customer features that are most relevant in predicting churn, to develop, implement and test a model for predicting customer churn using four classification algorithms and to evaluate and compare performance of the different algorithms used.

This study managed to fulfill all the stated objectives. This was done through acquisition of relevant data, processing the data and finally developing a prototype for churn prediction. This prototype showed importance of various features, implemented several classification algorithms and compared the performance of this algorithms. The results of which were presented.

## 5.2 Research Limitations

During the research some limitations were encountered. These were the shortcomings and conditions that could not be controlled by the researcher which resulted in some restrictions on the methodology and outcomes.

The findings of this study therefore have to be seen in light of the limitations discussed below:

### 5.2.1 Data availability

Due to data protection policies, the telecom company only gave out partial data. In terms of features that were to be used by the model, the telecom only shared 11 features out of more than 50 features / subscriber attributes it has. The company highlighted that they believed these were the most important features they could share without compromising data privacy. While these features were able to serve as good indicators of churn, elimination or inclusion of any features should be purely scientific.

### 5.2.2 Classification algorithms

Classification algorithms that can be used for churn prediction are many. To narrow down on which four will be implemented and evaluated the researcher relied on similar work that had been done before and research on which algorithms tend to perform better in general. In machine learning, there's a theorem known as No Free Lunch theorem which dictates that no single algorithm performs best in all situations. This means that an algorithm that was best for another situation might not necessarily be the best for this problem. However due to constraints such as time, it becomes necessary to filter out which algorithms are worth implementing and evaluating.

### 5.2.3 Computational Resources

Machine Learning and big data require enormous amounts of computing power. The prototype that was built for this research was done on a personal computer as opposed to a more powerful machine due to the costs associated with getting such power. Even though the personal computer is fairly powerful, it meant we could only train and test the models using 1m million subscribers out of a possible 18m that were obtained. The limitation was mainly due to the memory and processing power.

## 5.3 Recommendations

If the company is to resolve the churn problem, below are some of the recommendations it should implement:

### 5.3.1 Change approach to churn from reactive to proactive

Currently churn is just a metric that is reported after it has happened. After feeling the impact of customers who have been lost due to churn, the company usually tries to apply some reactive measures. One of such reactive strategies being applied by the telecommunication company under research is titled 'win back'. Using this strategy, the company tries to win back customers who have already churned and are using the competitors' products. This is mostly done by targeted advertising online through platforms such as Facebook, Google, among others. Such methods are not only costly but also ineffective. These measures are usually reactive, ineffective and cost much more than what the company would have used to retain them. Trying to fix your churn after things have already gone wrong is a bit like playing the lottery, the odds of success are slim to none. The company should therefore become proactive by using churn prediction.

### 5.3.2 Integrate churn prediction into the existing customer retention framework.

The telecom company under study has a very extensive customer retention framework. However, the framework does not show how to identify potential churners so that customer retention policies can be applied to such customers proactively. By adding churn prediction to the framework, not only does the framework become more comprehensive but it also helps to increase customer retention by preventing customers from churning.

### 5.3.3 Monthly analysis and prediction of churn.

Churn is currently reported and tracked at the end of every financial year. The churn rate is reported but little to no action is taken to address the problem. From research, churn has huge impact on the business and should therefore be tracked more frequently as opposed to reporting it at the end of the year. The churn rate should be analyzed monthly in order for the business to take quick action. Potential churners should also be predicted monthly so that the business can take proactive measures to prevent them from churning.

### 5.3.4 Conduct surveys on customers who are about to churn.

After identifying the customers who are about to churn the company should conduct unobtrusive surveys which do not involve direct elicitation of data from the research subjects. These surveys should be carried out with an objective of finding out what is making the customer dissatisfied.

With such data the company can then be able to resolve customer pain points and therefore increase the chances of retaining such customers.

#### 5.3.5 The telecom company should invest more on retention and less on acquisition.

This is on the basis of various studies done on customer acquisition that show acquisition of new customers could be up to 25 times costlier than retention of an existent one. While trying to gain more market share the company should invest more in customer retention as it is more cost effective than acquiring a new customer.

## 5.4 Suggestions for Further Research

It is not feasible to investigate all aspects related to customer retention and churn in a single research. This is due to paucity of time, personal limitations and the scope of work that would be required.

Therefore, this study opens up more avenues for further research some of which are briefly discussed below:

#### 5.4.1 Cost benefit analysis of implementing churn prediction.

This study has given a high-level estimate of the revenue lost through customers who churn. Even though preventing loss of revenue from churn is a major benefit to the business, there are other benefits to implementing churn prediction. Some of these benefits include avoiding negative marketing from customers who churn, increased customer satisfaction, reduction of acquisition costs, etc. Future research can look at the impact of churn prediction on other aspects of the business

#### 5.4.2 How to prevent customers from churning

While this study looks at one important aspect of churn management which is churn prediction, there exists several other aspects e.g. what happens once you've identified potential churners. For future research one can look at measures that can be used to prevent customers who have predicted as likely to churn, from not churning. Such measures may include things such as giving incentives, resolving customer pain points, etc. The research can also try to establish the effectiveness of the different measures.

#### 5.4.3 Classification algorithms.

While this study has attempted to compare four of the most common algorithms, there exists numerous algorithms that can be used for such a classification problem. The algorithms also have hyper parameters which if tuned differently can result in different results. For further research it can be suggested to try more classification algorithms, use different hyper parameters and also use an ensemble of the algorithms e.g. through voting where each algorithm gets to vote on whether a customer will churn.

## 5.5 Conclusion

The current approach by telecommunication companies in Kenya for churn management is very ineffective. This research has shown how these organizations can change their reactive approach to a proactive approach by using machine learning to predict subscribers who intend to churn. The predictions from the prototype that was created were fairly accurate and such a model can be taken into production. The company would therefore largely benefit by incorporating churn prediction to their customer retention framework.

# 6 References


1) Abbott, D. (2014). Applied predictive analytics: principles and techniques for the professional data analyst: John Wiley & Sons
2) Agarwal, R., & Dhar, V. (2014). Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research.
3) Bingo, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intel. 35(8), 1798–1828 (2013)
4) Buckinx, W., Verstraeten, G., & Van den Poel, D. (2007). Predicting customer loyalty using the internal transactional database. Expert Systems with Applications, 32(1), 125-134. doe: 10.1016/j.eswa.2005.11.004
5) David L. Garcia, Angela Nebot, Alfredo Vellido (n.d). Intelligent Data Analysis Approaches to Churn as a Business Problem: a Survey
6) Gareth James, Daniela Witten, Trevor Hastie& Robert Tibshirani (2013) An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)
7) H. Lee, Y. Lee, H. Cho, K. Im, Y.S. Kim (2011) "Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model", Decision Support System 52.
8) Kenneth C. Laudon, Jane P. Laudon, Ahmed A. Elraga(2013) Managing The Digital Firm
9) Kumar, V. & Petersen, J.A.(2012). Statistical Methods in Customer Relationship Management.
10) Leedy, P. & Ormrod, J. (2001). Practical research: Planning and design (7th ed.). Upper Saddle River, NJ: Merrill Prentice Hall. Thousand Oaks: SAGE Publications.
11) M. A. Waller and S. E. Fawcett. Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. Journal of Business Logistics, 34(2):77–84, 2013.
12) Thearling, K. (1999). An introduction of data mining. Direct Marketing Magazine.
13) Zhou, L., Pan, S., Wang, J., Vasilakos, A.V.: Machine learning on big data: opportunities and challenges. Neurocomputing 237, 350–361 (2017)
14) Alex Birkett (2017, August 15). *A 7-Point Guide to Reducing Customer Churn with Customer Research*. Retrieved from https://www.growthmanifesto.com/reducing-customer-churn
15) Steve Shellabear (2017, June 1). *Mahatma Gandhi And Customer Retention In Contact Centres*. Retrieved from https://contact-centres.com/dancing-lion-mahatma-gandhi-and-customer-retention-in-contact-centres/
16) Jason Brownlee (2018, May 23). *A Gentle Introduction to k-fold Cross-Validation*. Retrieved from https://machinelearningmastery.com/k-fold-cross-validation/
17) Larry Myler (2017, June 1). *Acquiring New Customers Is Important, But Retaining Them Accelerates Profitable Growth*. Retrieved from https://www.forbes.com/sites/larrymyler/2016/06/08/acquiring-new-customers-is-important-but-retaining-them-accelerates-profitable-growth/#59a1a6ff6671
18) Baremetrics (2011). *How Churn Prediction Can Improve Your Business*. Retrieved from https://baremetrics.com/academy/churn-prediction-can-improve-business

19) Safaricom PLC (2019). *News Release*. Retrieved from https://www.safaricom.co.ke/images/Downloads/Resources_Downloads/FY2019/FY2019_Press_Commentary.pdf
20) Business daily (2019, July 31). Airtel sinks deeper as losses pile up to Sh68bn. Retrieved from https://www.businessdailyafrica.com/corporate/companies/Airtel-sinks-deeper-as-losses-pile-up-to-Sh68bn/4003102-5217242-l6mco2/index.html
21) Airtel Africa (2019). Quarterly report on the results for the fourth quarter and year ended March 31, 2019. Retrieved from https://airtel.africa/assets/pdf/pdf1.pdf
22) Zac Harris (2019, July 24). What Is Customer Churn? Everything You Need to Know To Decrease Customer Attrition [Blog post]. Retrieved from https://www.spyfu.com/blog/customer-churn/
23) Mary Woodley (2019). Research help – Types of resources. Retrieved from https://libguides.merrimack.edu/research_help/Sources
24) Jeff Simpson(2015). Data Masking and Encryption[Blog post]. Retrieved from https://www.iri.com/blog/data-protection/data-masking-and-data-encryption-are-not-the-same-things/
25) Oskar Blakstad (Jul 10, 2008). Experimental Research. Retrieved Jan 09, 2020 from Explorable.com: https://explorable.com/experimental-research
26) Tomi Mester (Nov 13, 2016). How data collection works. Retrieved from: https://data36.com/data-collection/
27) Raheel Shaikh (2015). Choosing the Best Algorithm for your Classification Model. Retrieved from https://medium.com/datadriveninvestor/choosing-the-best-algorithm-for-your-classification-model-7c632c78f38f