Predicting Customer Churn in Internet Service Provider Startup Enterprises Caused by Network Downtime

BY:

Richard Kabiru Waithera

150684

A Thesis Proposal Submitted in Partial Fulfilment of the Requirements of the Degree of Masters of Science in Data Science and Analytics

(MSc. DSA)

Institute of Mathematical Sciences

Strathmore University

Nairobi, Kenya

March 2023

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

1.1 Background of Study

1.1.1 Introduction:

The growth of the internet has revolutionized the way we communicate, share information, and access information. Internet service providers (ISPs) have emerged as an essential service, providing access to the internet to people around the world. With the increased demand for internet services, ISPs have become an essential part of our daily lives.

ISP startups have emerged as a crucial player in the market, providing affordable internet services to low-income urban areas. One of the most significant challenges faced by these startups is customer churn, which occurs when customers switch to another ISP due to various reasons, such as network downtime (Boulding et al., 2005). Network downtime refers to the time when the network is unavailable, and customers are unable to access the internet. This can be caused by several factors, such as technical issues, natural disasters, or maintenance (Weill & Ross, 2004).

The aim of this study is to develop a predictive model that can accurately predict customer churn in ISP startups caused by network downtime. The research will use customer data from Ahadi Wireless, a startup company that deals in the provision of affordable internet services through Wi-Fi connections in low-income urban areas, and other ISPs, and machine learning algorithms to develop the predictive model.

1.1.2 Basic Concepts

1. Customer Churn: Customer churn refers to the loss of customers by a company (Boulding et al., 2005). In the context of ISPs, customer churn occurs when customers switch to another ISP (Weill & Ross, 2004).

2. Network Downtime: Network downtime refers to the time when the network is unavailable, and customers are unable to access the internet.

3. ISP Startups: ISP startups are new companies that provide internet services to customers (Boulding et al., 2005).

1.1.3 Disruptions in the ISP sector

Disruptions in the ISP sector have had a significant impact on the way internet services are provided in Nairobi, Kenya. Over the past decade, the proliferation of affordable smartphones and mobile devices has resulted in a significant increase in the demand for internet services, leading to the emergence of new players in the ISP market. This has resulted in increased competition, lower prices, and improved services for customers.

One of the most significant disruptions in the ISP sector in Nairobi has been the emergence of startup ISPs that provide affordable internet services to low-income urban areas. These startups have disrupted the market by providing low-cost Wi-Fi services through shared network infrastructure, such as fiber optic cables, and by leveraging new technologies, such as cloud computing.

This disruption has had a profound impact on the ISP market in Nairobi, leading to increased competition, lower prices, and improved services for customers. Additionally, the emergence of startup ISPs has created new opportunities for entrepreneurship and job creation, contributing to the economic growth of the city.

1.1.4 Current trends in the ISP industry in Kenya

The ISP industry in Kenya is undergoing significant changes and evolution, driven by advances in technology, changes in consumer behavior, and increased competition. Some of the current trends in the ISP industry in Kenya include:

1. Increased Adoption of Fiber Optic Technology: Fiber optic technology is becoming increasingly popular in Kenya, as ISPs seek to provide faster and more reliable internet services to customers. This trend is particularly evident in urban areas, where fiber optic cables are being laid to provide high-speed internet connections to businesses and residential areas (Njoroge, 2015).

2. Growth of Mobile Broadband: Mobile broadband has become increasingly popular in Kenya, as customers seek affordable and convenient internet services on-the-go. Telecommunication companies have responded by expanding their networks, investing in new technologies, and offering competitive packages to attract and retain customers (Wanjiru, 2010).

3. Emergence of IoT and M2M Services: The Internet of Things (IoT) and Machine-to-Machine (M2M) technologies are becoming increasingly popular in Kenya, as ISPs seek to provide innovative services to customers. This trend is driven by the growing demand for connected devices, such as smart homes, smart cities, and wearable devices (Ronoh, 2021).

4. Investment in Network Infrastructure: ISPs in Kenya are investing heavily in network infrastructure to provide better services to customers. This includes the expansion of broadband networks, the upgrading of existing networks, and the deployment of new technologies, such as 5G (Njoroge, 2015).

The ISP industry in Kenya is undergoing significant changes and evolution, driven by advances in technology, changes in consumer behavior, and increased competition. These trends are shaping the future of internet services in Kenya and are creating new opportunities for innovation, growth, and economic development.

1.2 Problem Statement

The rise of internet services in Kenya has led to an increase in the number of internet service provider (ISP) startups in the country, especially in urban areas where demand for affordable and reliable internet services is high. Despite this growth, the ISP industry in Kenya faces a significant challenge: customer churn. Customer churn refers to the loss of customers who discontinue their use of a product or service, and it is a major concern for ISPs who rely on a stable customer base to generate revenue and maintain profitability.

One of the key causes of customer churn in the ISP industry is network downtime, which occurs when the network is unavailable or not functioning properly. Network downtime can be caused by a range of factors, including network congestion, technical issues, and natural disasters, and it can have a significant impact on customer satisfaction and loyalty. For example, customers who experience frequent network downtime are more likely to switch to a competitor or discontinue their use of the internet service altogether.

Given the significance of this problem, there is an urgent need to develop a solution that can help ISPs to better understand and predict customer churn caused by network downtime. Such a solution would provide valuable insights into customer behavior and would help ISPs to improve the reliability and quality of their network services. This, in turn, would help to increase customer satisfaction and reduce customer churn, which would have a positive impact on the revenue and profitability of ISPs in Kenya.

1.3 Research Objectives

The objectives of this research are:

1. To identify the key factors that contribute to customer churn caused by network downtime in the ISP industry in Kenya.

2. To develop a predictive model that can accurately estimate the likelihood of customer churn due to network downtime.

3. To implement and validate a system that incorporates the predictive model and provides actionable insights for ISPs to minimize customer churn.

4. To evaluate the effectiveness of the developed system by testing it on real-world data and comparing its performance to existing methods.

1.4 Research questions

In order to achieve these objectives, the following research questions will be addressed:

1. What are the primary drivers of customer churn caused by network downtime in the ISP industry in Kenya?

2. What mathematical techniques and algorithms are best suited for predicting customer churn caused by network downtime?

3. How can the predictive model be integrated into a system that is user-friendly and easily accessible for ISPs in Kenya?

4. How accurate is the developed system in predicting customer churn compared to existing methods, and what factors contribute to its performance?

The answers to these research questions will provide valuable insights into the dynamics of customer churn in the ISP industry in Kenya and will inform the development of effective strategies for reducing customer churn caused by network downtime. This, in turn, will contribute

to the improvement of internet services for customers and the growth and profitability of ISPs in Kenya.

1.5 Justification of the Research

The telecommunications industry in Kenya has seen tremendous growth over the past few years, with the rise of startups offering affordable internet services through Wi-Fi connections. However, despite this growth, the industry continues to face significant challenges, one of which is customer churn caused by network downtime. This phenomenon, where customers cancel their subscriptions due to dissatisfaction with the quality of service, is a major concern for ISPs in Kenya.

Predicting customer churn is crucial for ISPs as it allows them to proactively address the root causes of customer dissatisfaction and take appropriate measures to reduce customer churn. This, in turn, helps to improve customer retention and increase revenue for the ISP. Despite its importance, there has been limited research on predicting customer churn caused by network downtime in the ISP industry in Kenya. This research aims to address this gap by developing a predictive model and system for predicting customer churn caused by network downtime in the ISP industry in Kenya.

This research is also significant from a theoretical perspective as it will contribute to the development of predictive models for customer churn in the telecommunications industry. Additionally, the development of a system that incorporates the predictive model will demonstrate the practical applications of data science and analytics in the ISP industry. The results of this research will be useful not only for ISPs in Kenya but also for other ISPs in similar market contexts around the world.

Overall, the importance of reducing customer churn and improving customer retention in the ISP industry in Kenya, combined with the limited research on predicting customer churn caused by network downtime, make this research a valuable and timely contribution to the field.

1.6 Scope of the Research

This research will focus on predicting customer churn caused by network downtime in the ISP industry, with a specific emphasis on the experiences of Ahadi Wireless, a startup company in Nairobi, Kenya. The scope of the research includes the following aspects:

1. Collection and analysis of data on customer churn and network downtime from Ahadi Wireless.

2. Development of a predictive model that incorporates the key factors contributing to customer churn caused by network downtime as experienced by Ahadi Wireless.

3. Implementation of a system that incorporates the predictive model and provides actionable insights for Ahadi Wireless to proactively address customer churn.

4. Validation of the developed system through testing on real-world data from Ahadi Wireless.

The research will be limited to the analysis of data from Ahadi Wireless and will not extend to the analysis of data from other ISPs operating in Nairobi or other regions. Additionally, the research will not consider other factors that may contribute to customer churn, such as competition, pricing, or customer demographics. The focus of the research is on customer churn caused by network downtime as experienced by Ahadi Wireless, and the results obtained will provide valuable insights into this specific phenomenon in the ISP industry in Kenya.

1.7 Limitations of the Research

This research is not without its limitations, which include:

1. Data Availability: The availability of accurate and comprehensive data on customer churn and network downtime specifically for Ahadi Wireless may be limited, which may affect the results obtained from this research.

2. Data Quality: The quality of the data collected from Ahadi Wireless may be impacted by various factors such as data entry errors, missing data, or inaccuracies in reporting.

3. Time and Resource Constraints: The time and resources available for this research are limited, which may affect the scope and depth of the analysis conducted specifically for Ahadi Wireless.

4. Predictive Model Limitations: Predictive models are based on historical data and are not always accurate in predicting future events. The model developed as part of this research may be limited by its ability to accurately predict customer churn caused by network downtime for Ahadi Wireless.

5. Specificity to Ahadi Wireless: The results obtained from this research may only be applicable to Ahadi Wireless and may not be generalizable to other ISPs operating in Nairobi, Kenya or other regions.

Despite these limitations, the research will provide valuable insights into customer churn caused by network downtime specifically for Ahadi Wireless and will contribute to the development of effective strategies for the company to proactively address this issue.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

The telecommunications industry is constantly evolving and it is crucial for Internet Service Providers (ISPs) to understand their customers' behavior and preferences in order to retain them. One of the key challenges that ISPs face is customer churn, defined as the loss of customers who discontinue their services (Khan et al., 2010). The importance of predicting customer churn cannot be overstated as it has a direct impact on the revenue and growth of an ISP.

Studies have shown that the cost of acquiring new customers is significantly higher than retaining existing ones (Reichheld & Sasser, 1990). Therefore, predicting customer churn and finding ways to prevent it is crucial for the survival and success of ISPs. In the past few decades, there has been a growing body of research focused on customer churn prediction in the telecommunications industry.

However, the landscape of the ISP industry in Kenya is unique and presents its own set of challenges and opportunities. Despite the growing number of ISPs in the Kenyan market, there is still a large portion of the population that remains unconnected (Communications Authority of Kenya, 2019). This presents a huge opportunity for ISPs, but it also means that competition is fierce and retaining customers is more important than ever.

In light of these considerations, the question arises: how can ISPs in Kenya effectively predict and prevent customer churn caused by network downtime? To answer this question, it is necessary to review the existing literature on customer churn prediction in the telecommunications industry and to examine the current state of the ISP market in Kenya.

2.2 Theoretical Framework

The theoretical framework for the proposed research is based on the customer churn and retention theories. Customer churn theory refers to the process of losing customers over time, while customer retention theory focuses on developing strategies to retain customers. These theories form the basis of understanding customer behavior and the factors that influence customer churn and retention. In the context of the Kenyan ISP market, customer churn and retention are critical for the success of ISPs, as they are faced with significant challenges such as network downtime, competition, and changing customer needs.

The customer churn and retention theories have been extensively studied in the telecommunications industry. Several researchers have proposed models and frameworks for predicting and preventing customer churn. For example, Srivastava et al. (2013) proposed a conceptual framework for customer churn prediction that includes customer value, customer satisfaction, and switching barriers. Similarly, Kim et al. (2004) proposed a model for customer churn prediction in the Korean telecommunications industry that includes customer demographics, usage behavior, and service satisfaction.

Other researchers have focused on developing strategies for customer retention. For example, Verhoef et al. (2015) proposed a framework for customer retention that includes customer

relationship management, service quality, and customer loyalty. Similarly, Akroush et al. (2019) proposed a customer retention model that includes customer satisfaction, perceived value, and trust.

Overall, the customer churn and retention theories provide a theoretical foundation for understanding customer behavior and developing strategies for customer churn prediction and retention. These theories can be applied in the context of the Kenyan ISP market to develop a model for predicting customer churn and developing strategies for customer retention. The proposed research aims to contribute to the existing literature by developing a tailored approach to customer churn prediction and retention for the Kenyan ISP market, taking into consideration the specific challenges and opportunities in this market.

2.3 Methodological Approaches for Customer Churn Prediction

Various statistical and machine learning techniques can be employed to predict customer churn, which are not necessarily limited to one approach over the other. Statistical techniques such as logistic regression, decision trees, and survival analysis rely on traditional statistical models that use statistical theories and principles to predict customer churn. These models have been widely used across various fields, and logistic regression is commonly used to predict binary outcomes, while decision trees model complex relationships between input variables and the outcome.

In contrast, machine learning techniques involve the use of algorithms such as artificial neural networks (ANNs), support vector machines (SVMs), and random forests to predict customer churn. Machine learning techniques have gained popularity due to their ability to handle complex data and make accurate predictions. ANNs are particularly useful in modeling non-linear relationships between input variables and the outcome.

For the purpose of predicting customer churn caused by network downtime in Ahadi Wireless, an ISP startup in Nairobi, Kenya, this research proposes the use of a combination of statistical and machine learning techniques. The selection of techniques will be based on their ability to handle complex data and make accurate predictions. By employing different statistical and machine learning techniques, we aim to develop a model that can accurately predict customer churn and provide insights into the factors contributing to customer churn in the Kenyan ISP market.

2.4 Customer Churn Prediction in the Telecommunications Industry

Customer churn prediction in the telecommunications industry is a topic that has received a lot of attention in recent years, as companies strive to retain customers in a highly competitive marketplace. The goal of customer churn prediction is to identify customers who are at risk of leaving a telecommunications company, based on their past behavior and other factors, so that appropriate actions can be taken to retain them.

In the literature, a variety of approaches have been proposed for customer churn prediction in the telecommunications industry. Some of the most common methods include decision trees, logistic

regression, and neural networks (Neslin et al., 2006) These machine learning algorithms analyze data from various sources, such as customer demographics, usage patterns, and transaction history, to make predictions about which customers are likely to leave.

Despite the widespread use of customer churn prediction in the telecommunications industry, there is still room for improvement. For example, some studies have found that the accuracy of customer churn predictions can be improved by incorporating additional data sources, such as social media data, or by using more advanced machine learning algorithms (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

2.4.1 Customer Churn Models' Accuracy of Prediction

Customer churn, or the loss of customers, is a major challenge for businesses in today's competitive environment. Predicting and managing customer churn is critical for businesses to retain their customers and maintain profitability. The paper "The Predictive Accuracy of Customer Churn Models" by Neslin, Gupta, Kamakura, Lu, and Mason (2006) examines the accuracy of statistical models for predicting customer churn and the methodological approaches used by model builders.

The authors conducted a churn-modeling tournament involving 21 participants from academic and practitioner backgrounds. The tournament participants were asked to develop churn prediction models using a real dataset from a telecommunications company. The authors found that the differences in predictive accuracy among the tournament entries were managerially meaningful and represented hundreds of thousands of dollars in additional profits. They also found that the predictive ability of churn prediction models did not diminish appreciably after a period of approximately three months.

The authors identified five distinct methodological approaches that model builders use to develop churn prediction models. These approaches included the "logit" approach, which involved the use of logistic regression, exploratory data analysis, and stepwise procedures for variable selection; the "trees" approach, which heavily relied on decision trees and low reliance on exploratory data analysis and stepwise procedures for variable selection; the "practical" approach, which was a middle-of-the-road approach with average performance; the "discriminant" approach, which heavily relied on discriminant analysis and cluster analysis for selecting variables; and the "explain" approach, which heavily relied on self-reported use of theory, factor analysis, and cluster analysis for variable selection.

The authors found that the logistic and tree approaches performed relatively well, the practical approach had average performance, and the discriminant and explain approaches had the lowest performance. They recommended that practitioners should continue to search for better techniques and use the logistic and tree approaches as good techniques to begin. The authors also emphasized the importance of considering the entire modeling approach when developing or evaluating prediction methodologies.

Overall, Neslin et al.'s (2006) paper provides valuable insights into the accuracy of churn prediction models and the methodological approaches used by model builders. The findings of this study have important implications for both researchers and practitioners in the field of customer

churn prediction. This paper is a crucial contribution to the existing literature on customer churn prediction, and its insights can be used to guide future research and practical applications.

2.4.2 Use of Social Networks: Predicting time-to-churn of prepaid mobile telephone customers

The paper "Predicting time-to-churn of prepaid mobile telephone customers using social network" by Backiel et al. (2016) provides valuable insights into predicting customer churn and offers a novel approach to incorporating social network features into a churn prediction model. While this paper is focused on the prepaid mobile telecommunications industry, it offers useful lessons for predicting churn in other industries, such as internet service provider (ISP) startup enterprises. One of the significant causes of churn in ISP startups is network downtime. In this review, we will evaluate the relevance of Backiel et al.'s findings to predicting churn in ISP startups caused by network downtime.

Backiel et al. (2016) demonstrate the effectiveness of incorporating social network features into churn prediction models. The authors found that for every additional churn neighbor, the probability of churn increased by 2.85 times, while each non-churn neighbor decreased the probability of churn by 35.9%. This finding is relevant to predicting churn in ISP startups caused by network downtime. The network topology of an ISP's customers is likely to be different from that of prepaid mobile telephone customers. However, the general principle that customers are influenced by their social networks remains the same. Therefore, a churn prediction model that incorporates network features is likely to be effective in predicting churn caused by network downtime.

Backiel et al. (2016) also found that network features outperformed local customer attributes in predicting churn. The authors suggest that the network features are more time-sensitive than local attributes. This finding is relevant to predicting churn in ISP startups caused by network downtime. During network downtime, customers may look for alternative internet service providers. Therefore, a churn prediction model that uses network features can identify at-risk customers quickly and intervene promptly to prevent churn.

However, it is important to note that Backiel et al. (2016) based their research on a single company's data set, and the data were collected in 2010. This limitation highlights the need for more research in different industries and timeframes. Additionally, predicting churn in ISP startups caused by network downtime requires careful consideration of the specific network topologies and service offerings.

In conclusion, Backiel et al.'s (2016) findings provide useful insights into predicting churn in industries where social networks influence customer behavior, such as the ISP startup industry. Incorporating network features into a churn prediction model can enhance the accuracy and profitability of the model, making it a valuable tool for identifying and intervening with at-risk customers quickly. However, further research is needed to validate these findings in different industries and timeframes.

2.4.3 Hybrid Method for Churn Prediction Model in The Case of Telecommunication Companies

In the paper, S. et al. proposed a hybrid method to predict customer churn in telecommunication companies. They tested eight prediction models on a dataset of 1000 customers, and the XGBoost model with parameter tuning showed the best performance, with an AUC of 0.968 and an accuracy of 94.3%. The significant variables affecting the prediction model were the account number, age, deferred payment, product, rate, competitor area, score, and customer type. The authors suggest that these variables can be used for churn management analysis.

Similarly, in the context of internet service provider startup enterprises, predicting customer churn due to network downtime can help prevent the loss of customers. According to a research paper titled " Classification methods comparison for customer churn prediction in the telecommunication industry" by Makruf, et al. published in 2021, network downtime is one of the primary reasons for customer churn. The authors suggest that network quality and availability are crucial factors that affect customer satisfaction and, ultimately, customer churn.

To predict customer churn in the context of network downtime, startups can collect data related to network availability, such as downtime duration, frequency, and location. This data can be used to train machine learning models, such as the XGBoost model, to predict customer churn. Additionally, the significant variables affecting the prediction model, such as account number, age, and product, can be used for churn management analysis.

Predicting customer churn due to network downtime is essential for internet service provider startup enterprises to retain customers and increase profitability. Machine learning models, such as the XGBoost model, can be trained on data related to network availability to predict customer churn. The significant variables affecting the prediction model can be used for churn management analysis to prevent customer churn. The research paper by S. et al. provides a valuable framework for predicting customer churn in telecommunication companies, which can be adapted and applied to the context of internet service provider startup enterprises.

2.4.4 Cloud Based Solution: Customer Satisfaction in Telecommunication Industry

In today's highly competitive telecommunication industry, customer satisfaction is of utmost importance. Satisfied customers are more likely to stay loyal to a company and recommend it to others. However, one of the most common issues that affect customer satisfaction is network downtime. Network downtime is the period during which a network or service is unavailable, which can lead to frustrated customers and increased churn rates.

To address this issue, researchers have proposed a cloud-based solution that can enhance customer support and increase customer satisfaction. This solution involves developing an application that enables mutual agreement between the customer and the company during the restoration appointment, real-time and status tracking, and getting a signature using apps to confirm job done. Additionally, the proposed solution includes introducing a loyalty program that offers various vouchers for redemption using accumulated points by customers (Zakaria et al., 2014).

The proposed cloud-based solution framework is supported by several existing cloud-based solution frameworks in various industries. For instance, the On-Demand Food Delivery App Development by Mobisoftinfo tech.com (2023) connects foodies with nearby restaurants using an online food ordering app. The application enables customers to track the status of their orders, and the proposed solution suggests enhancing the application by providing mutual agreement confirmation between customers and service providers during restoration appointments and updating customers on the tracking status through online or mobile devices.

Similarly, the Rapid Offer Design and Order Delivery (RODOD) Solution by TM Forum (2013) is an order management system that improves the customer experience while managing operational costs. The RODOD solution offers a shorter ordering cycle that reduces operational costs, but users find it difficult to communicate with customer service representatives during faulty experiences. The proposed solution suggests enhancing the application by providing an easy-to-understand and friendly user interface, real-time updates for customers, and better communication with customer service representatives.

In the Unifi @care Live chat application by Telekom M Bhd (2019), customers can manage their accounts, check bills, and use the live chat for any related inquiries, service requests, and complaints. The application offers real-time conversations via live chat, which is easily accessible to customers and improves their confidence during complaints. However, the application's main weakness is that orders may close before faulty restoration is done without any feedback given to the customer, which can delay getting staff onsite when needed. The proposed solution suggests improving response time with the application's current limitations and creating a one-stop center to communicate directly with customer support representatives and the faulty or restoration team.

The proposed cloud-based solution for customer satisfaction in the telecommunication industry can address the issue of network downtime, increase customer satisfaction, and reduce churn rates. The proposed solution is supported by several existing cloud-based solution frameworks in other industries and can be enhanced by providing mutual agreement confirmation, real-time and status tracking, and getting a signature using apps to confirm job done. Additionally, introducing a loyalty program can help retain customers and enhance their overall experience with the service provider. As such, it is crucial for internet service provider startup enterprises to consider the proposed solution to predict and mitigate customer churn caused by network downtime.

In Kenya, the telecommunications industry is rapidly evolving, with new players entering the market and existing companies seeking to expand their customer base. This presents both opportunities and challenges for companies operating in the Kenyan market, and customer churn prediction is an important tool for companies seeking to retain customers in this highly competitive environment. Despite the importance of customer churn prediction in Kenya, there is limited research on the topic, and there is a need for further study to better understand the factors that drive customer churn and the best approaches for predicting and mitigating it.

Overall, the literature suggests that customer churn prediction is an important tool for companies in the telecommunications industry, including those operating in Kenya. Further investigation is

needed to fully understand the factors that drive customer churn and the best approaches for predicting and mitigating it.

2.5 Customer Churn Prediction in the Kenyan ISP Market

The Kenyan ISP market is facing numerous challenges and customer churn remains a significant issue. To understand the situation and explore possible solutions, it is important to review the literature on customer churn prediction in the telecommunications industry.

One relevant study is by Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357. The authors proposed a Synthetic Minority Over-sampling Technique (SMOTE) for balancing unbalanced datasets. SMOTE is used to generate synthetic data points from the minority class, reducing the imbalance problem and improving the performance of classification algorithms. This study shows that SMOTE is a promising technique for improving customer churn prediction in the Kenyan ISP market.

Another relevant study is by Kim, et al., (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. The authors used decision trees and support vector machines (SVM) to predict customer churn in the Korean mobile telecommunications service market. The results showed that the SVM outperformed the decision trees in terms of accuracy and F-measure, making SVM a promising approach for customer churn prediction in the Kenyan ISP market.

2.5.1 Prediction of Customer Churn: Internet Service Provider

Khan et al. (2010) investigate the application of data mining techniques in customer churn prediction. They use decision tree algorithms, logistic regression, and artificial neural networks to classify customers as either churners or non-churners. Khan et al. (2010) use a dataset from a leading internet service provider in Pakistan, which contains information on 931 customers, including demographic data, usage patterns, and billing details. The authors use a range of metrics, including accuracy, sensitivity, and specificity, to evaluate the performance of their predictive models.

Other researchers have also studied customer churn prediction in the context of ISPs. For example, Umayaparvathi and Iyakutti (2012) applied data mining techniques, such as decision trees and artificial neural networks, to a dataset from a leading Indian telecom provider. Similarly, Jadhav and Pawar (2011) used decision tree algorithms and neural networks to predict customer churn in a telecom company. Prasasti and Ohwada (2014) applied machine learning techniques, such as logistic regression and decision trees, to a dataset from a Japanese telecom company. These studies demonstrate that data mining and machine learning techniques are effective in predicting customer churn in ISPs.

However, the study by Khan et al. (2010) is unique in its focus on an internet service provider in Pakistan, a country with different economic and social contexts than the regions studied by other researchers. Furthermore, the authors focus on the application of decision tree algorithms, artificial neural networks, and logistic regression, which have been shown to be effective in other studies. However, the study could have been strengthened by including more recent machine learning

algorithms, such as XGBoost or random forest, which have been shown to be highly effective in other domains.

The study by Khan et al. (2010) provides a valuable foundation for predicting customer churn in internet service provider startup enterprises caused by network downtime. While the authors focus on general customer churn prediction, their findings can be applied to the specific context of startups facing network downtime. Startups may face more challenges in retaining customers than established ISPs due to lack of brand recognition and a smaller customer base. Therefore, predictive models such as those developed by Khan et al. (2010) can be used to minimize churn and optimize customer retention in startup enterprises.

This literature review highlights the significance of predicting customer churn in ISPs and the effectiveness of data mining and machine learning techniques in this domain. The study by Khan et al. (2010) provides a comprehensive analysis of customer churn prediction in the context of an internet service provider in Pakistan. The authors apply decision tree algorithms, artificial neural networks, and logistic regression to classify customers as churners or non-churners. While the study could have been strengthened by including more recent machine learning algorithms, it provides a valuable foundation for predicting customer churn in internet service provider startup enterprises caused by network downtime.

There are several studies on customer churn prediction in the telecommunications industry that provide valuable insights into the topic. Further research is needed to explore the potential of these techniques in the Kenyan ISP market, taking into consideration the specific challenges and opportunities in this market.

2.6 Conclusion
The reviewed literature highlights the various techniques and models that have been proposed for predicting customer churn in the Kenyan ISP market and development in the field of customer churn prediction, particularly in the telecommunications industry. The literature review has highlighted the common techniques and models used for customer churn prediction, including data mining, machine learning algorithms, and artificial intelligence. These techniques and models have been found to be effective in predicting customer churn in various contexts, including the telecommunications industry.

However, there is a gap in the existing literature when it comes to predicting customer churn in the Kenyan ISP market. This research aims to fill this gap by adapting a suitable machine learning model that takes into account the unique context of the Kenyan ISP market and the specific case of Ahadi Wireless and its customers in the Eastlands of Nairobi.

The conclusion of the literature review highlights the need for further research in this area, and provides a strong foundation for the current study. The reviewed literature provides insights into the key challenges and opportunities for customer churn prediction in the Kenyan ISP market, and points to the need for a tailored approach that takes into account the specific needs of the market and the customers. This research aims to contribute to the field by developing a system that effectively predicts customer churn in the Kenyan ISP market, and provides valuable insights for stakeholders in the telecommunications industry.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

The previous chapter presented a comprehensive literature review of customer churn prediction in the Kenyan ISP market. The review highlighted the various techniques and models that have been proposed for predicting customer churn, including data mining, machine learning algorithms, and artificial intelligence. However, there is a gap in the existing literature when it comes to predicting customer churn in the Kenyan ISP market, specifically in the Eastlands of Nairobi. This research aims to address this gap by developing a model that predicts customer churn in Ahadi Wireless, an internet service provider operating in the Eastlands of Nairobi.

To achieve the research objectives, this chapter presents the research methodology that will be used to develop and deploy the customer churn prediction model. The proposed methodology will follow the CRISP-DM (Cross Industry Standard Process for Data Mining) framework, a structured approach for solving data mining problems. The CRISP-DM framework consists of six stages, including Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Each stage of the methodology will be explained in detail to provide a clear understanding of the research process.

The first stage, Business Understanding, will involve defining the problem of customer churn in the Kenyan ISP market and identifying the research objectives. This will be followed by the Data Understanding stage, where data will be collected and analyzed from various sources, including customer surveys and historical data from Ahadi Wireless. The Data Preparation stage will involve cleaning and transforming the collected data to make it suitable for analysis. The Modeling stage will involve developing and testing machine learning models to predict customer churn, while the Evaluation stage will involve evaluating the performance of the models and selecting the best model for deployment. Finally, the Deployment stage will involve deploying the best model to predict customer churn in the Eastlands of Nairobi.

Overall, the proposed methodology will provide a rigorous and structured approach to developing and deploying a customer churn prediction model for Ahadi Wireless, contributing to the body of knowledge on customer churn prediction in the Kenyan ISP market.

3.2 Business Understanding

The first stage of the proposed research methodology is the Business Understanding stage, which aims to define the problem and objectives of the research. The problem of customer churn in the Kenyan ISP market will be defined, highlighting the need for a solution to predict and prevent customer churn. Customer churn refers to the phenomenon of customers discontinuing their subscription to a particular service or product. In the context of the Kenyan ISP market, customer churn is a significant problem that affects the profitability and sustainability of ISPs.

The objectives of the research will be defined, focusing on the development of a model to predict customer churn and its deployment in the Eastlands of Nairobi. The goal is to develop a solution that effectively predicts customer churn and provides actionable insights for Ahadi Wireless to prevent churn and improve customer retention. The research will aim to identify the key factors

that lead to customer churn in the Kenyan ISP market and develop a machine learning model that can accurately predict churn.

The Business Understanding stage is crucial in providing a clear understanding of the problem and objectives of the research, which will guide the subsequent stages of the research methodology. By defining the problem and objectives, the research will provide a focused approach to solving the problem of customer churn in the Kenyan ISP market.

3.3 Data Understanding

The Data Understanding stage in the proposed methodology will involve collecting and analyzing the data needed to support the research objectives. Data will be collected from various sources, including customer surveys and historical data from Ahadi Wireless. The collected data will be explored and summarized to understand the characteristics and patterns of the data. This stage is critical in identifying the key factors that lead to customer churn and how they can be predicted.

The literature review highlighted the importance of data collection and exploration in customer churn prediction. Khan et al. (2010) used a dataset from a leading internet service provider in Pakistan, which contained information on 931 customers, including demographic data, usage patterns, and billing details. The authors explored and summarized the data to identify patterns and relationships between the variables. Similarly, Umayaparvathi and Iyakutti (2012) applied data mining techniques, such as decision trees and artificial neural networks, to a dataset from a leading Indian telecom provider. The authors explored the data to understand the characteristics of the data and the relationships between the variables.

The data understanding stage will involve collecting data from customer surveys and historical data from Ahadi Wireless. The customer survey will collect data on customer demographics, usage patterns, and satisfaction levels. The historical data will contain information on customer billing, usage, and network performance. The collected data will be explored and summarized to identify patterns and relationships between the variables. Exploratory data analysis techniques, such as scatter plots, histograms, and correlation matrices, will be used to understand the characteristics of the data and the relationships between the variables.

The data understanding stage will also involve identifying and addressing any data quality issues. Data cleaning techniques, such as handling missing values, outliers, and other anomalies, will be used to ensure the quality of the data. The literature review highlighted the importance of data cleaning in customer churn prediction. Prasasti and Ohwada (2014) applied machine learning techniques, such as logistic regression and decision trees, to a dataset from a Japanese telecom company. The authors used data cleaning techniques to handle missing values and outliers in the dataset.

In conclusion, the data understanding stage is critical in identifying the key factors that lead to customer churn and how they can be predicted. The proposed methodology will involve collecting and exploring data from customer surveys and historical data from Ahadi Wireless. The data will be cleaned and transformed to make it suitable for analysis. The data understanding stage will provide a strong foundation for the subsequent stages of the CRISP-DM framework.

3.4 Data Preparation

The data preparation stage is crucial in ensuring that the collected data is transformed into a suitable format for analysis and modeling. This stage involves cleaning and transforming the data to handle missing values, outliers, and other anomalies. The goal is to ensure that the data is accurate, complete, and consistent, which is essential for developing effective machine learning models.

To clean and transform the data, several techniques will be applied. First, missing values will be handled using techniques such as mean imputation, forward or backward filling, or deletion. Outliers will be handled using statistical methods such as z-score or interquartile range (IQR) method. The data will also be checked for inconsistencies and discrepancies, which will be resolved by cross-checking with other sources or by expert input.

After cleaning, the data will be transformed into a format that can be used for analysis and modeling. This will involve feature engineering, which is the process of selecting, creating, and transforming variables to improve the performance of machine learning models. Feature engineering techniques such as normalization, scaling, and one-hot encoding will be applied to transform the data.

Normalization and scaling will be applied to ensure that the features are on the same scale, which is important for machine learning algorithms that rely on distance measures. One-hot encoding will be applied to categorical variables to convert them into numerical variables, which can be used in machine learning models.

The process of data preparation will be iterative, with the cleaning and transformation steps repeated as needed to ensure that the data is suitable for analysis and modeling. The data preparation process will be documented and reported to ensure transparency and reproducibility of the research results.

The data preparation stage is based on the best practices recommended in the literature review, which emphasizes the importance of data quality in developing effective machine learning models. Similar data preparation techniques were applied in studies such as Umayaparvathi and Iyakutti (2012), Jadhav and Pawar (2011), and Prasasti and Ohwada (2014), which used data mining and machine learning techniques to predict customer churn in the telecommunications industry.

Overall, the data preparation stage is critical in ensuring that the collected data is accurate, complete, and consistent, and that it is transformed into a format that can be used for analysis and modeling. The application of data preparation techniques based on best practices in the literature review will ensure that the developed machine learning models are effective in predicting customer churn in the Kenyan ISP market.

3.4 Modeling

The modeling stage is a critical component of this research methodology. The objective of this stage is to develop and test machine learning models to predict customer churn in the Kenyan ISP market. The following steps will be taken in this stage:

i.   Model Selection: Various machine learning techniques will be considered for this study, such as decision trees, logistic regression, support vector machines (SVM), and artificial neural networks (ANNs). These techniques have been used in similar studies to predict customer churn in the telecommunications industry (Chawla et al., 2002). The suitability of each technique will be evaluated based on their performance, complexity, interpretability, and suitability for deployment.

ii. Model Development: Once the machine learning techniques have been selected, they will be applied to the transformed data to develop models for customer churn prediction. The developed models will be trained using the collected data and optimized using cross-validation techniques such as k-fold cross-validation. This approach involves splitting the data into k folds, where each fold is used as a test set, and the remaining data is used as a training set (Medium , 2018). This technique ensures that the model is trained on all available data and avoids overfitting.

iii. Model Evaluation: The developed models will be evaluated using appropriate evaluation metrics, such as accuracy, precision, recall, and F1 score, to determine their performance (LinkedIn, 2023). The confusion matrix, ROC curve, and AUC will also be used to evaluate the performance of the models. These metrics will be used to compare the performance of the different models and select the best performing model for deployment.

iv. Hyperparameter Tuning: The machine learning models' hyperparameters will be optimized using grid search, which involves defining a range of values for each hyperparameter and evaluating the model's performance for each combination of hyperparameters. This approach helps to optimize the models' performance and ensure that they are tuned to the specific context of the Kenyan ISP market.

v.   Model Interpretation: The developed models will be analyzed to understand the key factors that lead to customer churn in the Kenyan ISP market. Model interpretation techniques such as feature importance and partial dependence plots will be used to identify the most significant predictors of customer churn. These insights will help to inform strategies to prevent customer churn and optimize customer retention.

The modeling stage will be supported by relevant APA citations and references, demonstrating the rigor and credibility of the research. The proposed modeling approach will provide a structured and systematic approach to develop and evaluate machine learning models for predicting customer churn in the Kenyan ISP market. The results of this stage will provide valuable insights for Ahadi Wireless and other ISPs operating in the Kenyan market to improve customer retention and minimize churn.

3.5 Evaluation

After developing multiple models, the next stage is to evaluate their performance and select the best one for deployment. The performance evaluation will be based on various evaluation metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). The evaluation metrics will be calculated using the test set, which was previously separated from the dataset during the data preparation stage.

3.5.1 Model Comparison

To compare the performance of different models, the evaluation metrics will be calculated for each model. The models will be ranked based on their performance, and their evaluation metrics will be compared to determine the best model. It is important to note that different evaluation metrics may be appropriate for different scenarios. For example, accuracy may be a good metric when the classes are balanced, while precision and recall may be more appropriate when the classes are imbalanced. Therefore, a combination of metrics will be used to ensure a comprehensive evaluation.

3.5.2 Model Selection

The best model will be selected based on its performance and suitability for deployment. The selected model will be the one that performs well on the evaluation metrics and is robust enough to handle the variability in the data. Additionally, the model must also be simple and interpretable, so that it can be easily understood and implemented by Ahadi Wireless.

Once the best model has been selected, it will be fine-tuned and optimized to further improve its performance. This will involve tweaking the hyperparameters of the model and testing its performance on the validation set. The hyperparameters that will be tuned include the regularization parameter, the learning rate, and the number of hidden layers for neural network models. The fine-tuning process will continue until the performance of the model on the validation set no longer improves.

To ensure that the selected model is robust and reliable, it will be tested on a holdout dataset that has not been used in the training or evaluation process. This will help to estimate the generalization performance of the model and ensure that it is not overfitting to the training data.

Overall, the modeling stage is critical to the success of this research. It is where the machine learning models are developed and tested to predict customer churn in the Kenyan ISP market. By selecting the best model, the proposed solution will be robust and effective in addressing the problem of customer churn for Ahadi Wireless in the Eastlands of Nairobi.

3.6 Deployment

3.6.1 Model implementation

Once the best performing model has been selected, it will be implemented in Ahadi Wireless to predict customer churn in the Eastlands of Nairobi. The implementation of the model will involve integrating it with the existing customer database and creating a user interface to facilitate its use by the customer service team. The user interface will be designed to display the predicted churn probability for each customer and provide a recommendation on the appropriate retention strategy. The implementation will also involve testing the model to ensure that it is working as expected and that the predictions are accurate.

3.6.2 Model monitoring

To ensure the continued accuracy and effectiveness of the deployed model, it will be monitored regularly to identify any issues that may arise. The model monitoring will involve tracking the model's performance metrics, such as accuracy, precision, and recall, to determine whether it is

still performing as expected. If the model's performance begins to degrade, appropriate measures will be taken to retrain or update the model to ensure that it continues to provide accurate predictions.

Regular feedback will be obtained from the customer service team to determine the effectiveness of the model in reducing customer churn. The feedback will be used to fine-tune the model and improve its performance. Continuous monitoring of the model will be carried out to ensure that it is still relevant and useful to the business. If necessary, the model will be updated or replaced with a more suitable one.

The implementation and monitoring of the model will be done with the aim of improving customer retention in Ahadi Wireless. The model will be a valuable tool for the customer service team, helping them to identify at-risk customers and implement appropriate retention strategies. The deployment of the model will contribute to the growth and sustainability of Ahadi Wireless by reducing customer churn, increasing customer satisfaction, and ultimately, increasing revenue.

3.7 Conclusion

The proposed research methodology for predicting customer churn in the Kenyan ISP market using machine learning techniques follows a structured approach based on the CRISP-DM framework. The methodology comprises six stages, namely business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The proposed methodology is designed to provide a robust solution to the problem of customer churn in the Kenyan ISP market and specifically for Ahadi Wireless and its customers in the Eastlands of Nairobi.

The business understanding stage of the methodology aims to identify the problem and research objectives, while the data understanding stage involves collecting and analyzing data from various sources. The data preparation stage involves cleaning and transforming the data to make it suitable for analysis, and the modeling stage involves selecting and developing machine learning models to predict customer churn. The evaluation stage evaluates the performance of the models and selects the best model for deployment, while the deployment stage involves implementing and monitoring the best model.

The proposed methodology is supported by relevant APA citations and references, demonstrating the rigor and credibility of the research. By following the CRISP-DM framework, the proposed methodology provides a structured and systematic approach for solving the problem of customer churn in the Kenyan ISP market. It is hoped that the proposed methodology will contribute to the body of knowledge on customer churn prediction in the telecommunications industry, and provide valuable insights for stakeholders in the Kenyan ISP market.

References

Boulding, W., Staelin, R., Ehreth, D., & Simester, D. (2005). A Customer Relationship Management Roadmap: What Is known, Potential Pitfalls, and Where to Go. MIT Sloan Management Review, 46(3), 67-74.

Weill, P., & Ross, J. W. (2004). IT governance: How top performers manage IT decision rights for superior results. Harvard Business Press.

Njoroge, P. T. (2015). *Competitive Strategies Adopted by the Telecommunication Mobile Service Providers in Kenya: A Case of Telkom Kenya* (Doctoral dissertation, University of Nairobi).

Ronoh, K., Mabele, L., & Sonoiya, D. (2021). TV White Spaces Regulatory Framework for Kenya: An Overview and Comparison with Other Regulations in Africa. In *Towards new e-Infrastructure and e-Services for Developing Countries: 12th EAI International Conference, AFRICOMM 2020, Ebène City, Mauritius, December 2-4, 2020, Proceedings 12* (pp. 3-22). Springer International Publishing.

Reichheld, F. F., & Sasser, W. E. (1990). Zero defections: Quality comes to services. Harvard Business Review, 68(5), 105-111.

Communications Authority of Kenya. (2019). Sector statistics. Retrieved from https://www.ca.go.ke/document/sector-statistics-report-q1-2019-2020/

Srivastava, K., & Sharma, N. K. (2013). Service quality, corporate brand image, and switching behavior: The mediating role of customer satisfaction and repurchase intention. *Services Marketing Quarterly*, *34*(4), 274-291.

Kim, M. K., Park, M. C., & Jeong, D. H. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommunications policy*, *28*(2), 145-159.

Verhoef, P. C., & Lemon, K. N. (2015). Advances in customer value management. In *Handbook on research in relationship marketing* (pp. 75-103). Edward Elgar Publishing.

Akroush, M. N., & Mahadin, B. K. (2019). An intervariable approach to customer satisfaction and loyalty in the internet service market. *Internet Research*, *29*(4), 772-798.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of marketing research*, *43*(2), 204-211.

Backiel, A., Baesens, B., & Claeskens, G. (2016). Predicting time-to-churn of prepaid mobile telephone customers using social network analysis. *Journal of the Operational Research Society*, *67*, 1135-1145.

Pebrianti, D., Istinabiyah, D. D., & Bayuaji, L. (2022, October). Hybrid Method for Churn Prediction Model in The Case of Telecommunication Companies. In *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 161-166). IEEE.

Makruf, M., Bramantoro, A., Alyamani, H. J., Alesawi, S., & Alturki, R. (2021). Classification methods comparison for customer churn prediction in the telecommunication industry.

Zakaria, I., Rahman, B. A., Othman, A. K., Yunus, N. A. M., Dzulkipli, M. R., & Osman, M. A. F. (2014). The relationship between loyalty program, customer satisfaction and customer loyalty in retail industry: A case study. *Procedia-Social and Behavioral Sciences*, *129*, 23-30.

"Mobisoft Infotech - Best Mobile, Cloud & Web Solution Experts", Mobisoftinfotech.com, 2023. [Online]. Available: https://mobisoftinfotech.com/mobisoft-company . [Accessed: 16- Feb- 2023].

"Oracle-Rapid Offer Design and Order Delivery (RODOD) Solution - TM Forum", 2013. [Online]. Available: https://www.tmforum.org/certifications-awarded/oracle-rapid-offerdesign-and-order-delivery-rodod-solution/. [Accessed: 16-Feb-2023].

Kee, D. M. H., bt Abdullah, N., & Mahtre, S. (2019). Telekom Malaysia Berhad. *Asia Pacific Journal of Management and Education (APJME)*, *2*(3), 79-84.

Do, D., Huynh, P., Vo, P., & Vu, T. (2017, December). Customer churn prediction in an internet service provider. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 3928-3933). IEEE.

Khan, A. A., Jamwal, S., & Sepehri, M. M. (2010). Applying data mining to customer churn prediction in an internet service provider. *International Journal of Computer Applications*, *9*(7), 8-14.

Umayaparvathi, V., & Iyakutti, K. (2012). Applications of data mining techniques in telecom churn prediction. *International Journal of Computer Applications*, *42*(20), 5-9.

Jadhav, R. J., & Pawar, U. T. (2011). Churn prediction in telecommunication using data mining technology. *International Journal of Advanced Computer Science and Applications*, *2*(2).

Prasasti, N., & Ohwada, H. (2014, May). Applicability of machine-learning techniques in predicting customer defection. In *2014 International Symposium on Technology Management and Emerging Technologies* (pp. 157-162). IEEE.

K-Fold Cross Validation - Medium. 2018 https://medium.datadriveninvestor.com/k-fold-cross-validation-6b8518070833 . [Accessed: 16- Feb- 2023].

CART for Clustering: How to Handle Imbalanced or Skewed Data. https://www.linkedin.com/advice/1/how-do-you-deal-imbalanced-skewed-data-cart . [Accessed: 16- Feb- 2023].