Addis Ababa Institute of Technology
School of Electrical and Computer Engineering
Telecommunication Engineering Graduate Program

Machine Learning Based Mobile Airtime Credit Risk Prediction using Customer Profile and Loan Information

By: Shashu Brhane Berhe

Advisor: Surafel Lemma (PHD)

A THESIS SUBMITTED TOTHE SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING PARTIAL FULFILLMENT FOR THE DEGREE OF MASTERS OF SCIENCE IN TELECOMMUNICATION ENGINEERING

ADDIS ABABA, ETHIOPIA

December 2021

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

By: Shashu Brhane Berhe

Advisor: Surafel Lemma (PHD)

This is to certify that the thesis prepared by Shashu Brhane Berhe, titled: Machine Learning Based Mobile Airtime Credit Risk Prediction using Customer Profile and Loan Information submitted in partial fulfillment of the requirements for the degree of master of science in Telecommunication Engineering complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

DECLARATION

I, the undersigned, declare that this MSc thesis involves my original work, has not been presented before in any other university; and all sources and materials used for the thesis have fully acknowledged.

Shashu Brhane Berhe ____________________________

Full Name Signature

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

Machine Learning Based Mobile Airtime Credit Risk Prediction using Customer Profile and Loan Information

Signed by the Examining Committee:

Internal Examiner_____________ Signature _____________Date______________

External Examiner _____________ Signature ______________Date______________

Advisor Surafel Lemma (PHD) Signature ______________ Date_____________

Co-Advisor _____________ Signature ______________Date_____________

____

Dean, School of Electrical and Computer Engineering

DEDICAT ION

This thesis is dedicated to my immensely proud family Ato Brhane Berhe (My Dad), Wro. Alemtsehay Zemicheal (My Mom), Muruts Brhane, Yirgalem Brhane, Desta Brhane, Lezeba Brhane, Mulu Brhane & Goitom Brhane My Respected Brothers and Sisters.

ACKNOWLEDGMENT

Frist I would like to thank God and his holy virgin mother St. Mary, who has given me the strength to complete my thesis work, even if there are different struggles faced in my life. Next I would like to express my deepest gratitude to my advisor, Surafeal Lemma(PhD) for his valuable comment and advice that helps me throughout my thesis work. He is strong and punctual advisor in my education journey. I would like to thank my examiners Ephrem Teshale (PhD) and Fitsum Assamnew (PhD) for their helpful comments and feedbacks during my thesis progress and evaluations.

I would like to express my immense gratitude to my respected family for their encouragement and follow up's the progress of my work. In addition to this my gratitude goes to my respected one Halefom Gebremedihin, my best friends Senait Kiros and Winta Teshale and my classmates Gebremedihin Weldemariam and Desalegn Medhin for their support for the completion of this study. Finally, I would like to thank Ethio Telecom domain experts especially Tamirat Tesfaye, Gebremicheal Zeray and Selamawit Tadesse for their kind support during the data collection of the study.

ABSTRACT

Airtime credit service is usually used by prepaid mobile subscribes when they cannot recharge their service number. The credit service helps to increase service subscribers' satisfaction and revenue of the telecom operator. However, the service has its own risk as there is no guarantee that the prepaid subscribers would pay back the credit they took. To solve this issue, several researchers propose to predict airtime credit risk using machine learning based approaches. To build the prediction models, these approaches use either customer or customer activity related information as features. These features cover only part of the customer related information captured by telecom companies. Other potentially relevant information such as education of the customer, subscriber loan amount and frequency, and recharge frequency are not taken into consideration to predict airtime credit risk. In this research work, we propose an approach that takes in to consideration education and recharge frequency from customer profile and subscriber loan amount and loan frequency from the loan information to predict airtime credit risk.To assess the impact of the proposed approach, we conducted an experiment using 90,000 mobile subscriber’s data. The experiment uses four machine learning algorithms: decision tree (DT), logistic regression (LR), random forest (RF) and (MLp). The results show that the combination of the existing and new features improves airtime credit risk prediction for all algorithms. The highest improvement, i.e., 6.91%, in accuracy while using the existing customer profile and usage, and new features is observed by LR. For the existing loan and new features, the highest improvement, i.e., 7.8%, in F-measure is observed by RF. We ranked the features based on their importance using feature ranking algorithms. Subscriber loan amount is the most important feature. Loan frequency and recharge frequency are also in the top seven features that contributed to the risk prediction. This shows that the newly added features have helped to enhance the airtime credit risk prediction.

Keywords: Airtime credit risk, Airtime credit risk prediction, Machine learning algorithms.

Contents

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

ACS	Airtime Credit Service
AI	Artificial Intelligence
ARFF	Attribute Relation File Format
ANN	Artificial Neural Networks
BSS	Business Support System
CBS	Convergent Billing System
CDR	Call Detail Record
CRM	Customer Relation Management
CSV	Comma Separated Value
DT	Decision Tree
ESB	Enterprise Service Bus
GUI	Graphical User Interface
LR	Logistic Regression
MLP	Multi-layer Perceptron
MNO	Mobile Network Operator
RF	Random Forest
SMS	Short Message Service
SMSC	Short Message Service Center
SOM	Service Organization Map
USSD	Unstructured Supplementary Service Data
VAS	Value Added Service
VC	Voucher Card
WEKA	Waikato Environment For Knowledge Analysis

CHAPTER ONE: INTRODUCTION

1.1 Background

Communication technology has become essential, and it has aided in the creation of a society in which anyone, no matter how far away, can be reached in a matter of seconds. Due to corporate requirements and technological developments, the number of mobile subscribers is steadily increasing in the telecommunications industry, which requires a high level of multi-service delivery. One of the services included in this package is a mobile airtime credit [1].

Airtime is increasingly becoming a basic commodity among the rapidly growing middle class in developing nation. Several mobile network operators (MNOs) engaged in emerging markets have recognized this and are investing in new services such as short term airtime loans at a moderate interest rate [1].

Short term airtime loans would be useful when subscribers run out of airtime and cannot top up their service number. In some situations, it may be difficult for a customer to purchase a recharge card. particularly late at night or while traveling. Short term airtime loans will help in reducing the inconvenience for the customer and loss of revenue for the telecom operators [2].

The airtime loan service is implemented following two approaches. In the first approach, mobile network operators (MNOs), provide clients with airtime loans and face the risk of non-performing loans. Safaricom, through its Okoa Jahazi service, is an example of a corporation that employs such an approach. When a subscriber runs out of airtime, they can borrow money equal to the amount they had topped up in the previous seven days, with the expectation of repaying it within five days [1].

The second approach entails collaboration between the MNO and a third-party lender, in which the MNO offers access to clients and the mobile network while the risk is shifted to the third party. ComzAfrica is an example of a lender that relies on such collaboration. The third-party lender pays the MNO the credit amount before the customer repays it under this process. As a result, in the event of a default, the third party loses the entire sum [1].

The addition of this new service has the potential to boost MNO's average revenue per user. In most parts of the country, the airtime loan is repaid once the subscriber's account has been credited.

The time period in which the loan has to be paid, however, varies based on the rules set by MNO. For example, some MNOs set the period to be five days, while others set it to be 72 hrs. If the loan is not settled the loan in the given time period, the subscribe who took the loan will be considered as a default. To early identify subscribers who default and minimize the loss incurred due to them, MNOs implement credit scoring algorithms [3]. The credit scoring algorithms allow MNOs to estimate the subscriber's ability to pay the loan in the given time period.

In Ethiopian context airtime credit service is introduced in August 31,2018 by Hikma electronic PLC and Credok communication technology partners using a contractual agreement with Ethio telecom. Three parties are involved in the provisioning of airtime credit service in Ethiopia. These are the mobile service operator, a VAS (Value Added Service) Provider and the subscriber.

Each of these parties, play different roles in the airtime credit service. The subscriber is the end- user interested in the credit service, the VAS provider is the middleman who provides the credit (the airtime) and bears the risk of loss if the subscriber defaults in the repayment of the credit, while the mobile operator is the one that provides the platform that links the other two parties in the value chain. To settle the airtime loan, Ethio telecom gives 180 days.

The airtime loan service is accessed using SMS (by sending one of the key words (A, L or C) to 810 or dialing *810#. In order to activate the service and give loan to the users Ethio telecom uses credit scoring method to check the eligibility of the user.

The criteria defined in the credit scoring method are the following: the user the user must have an active prepaid service number, must have been on the Ethio telecom network for a minimum of 3 months and with a minimum top-up of 30-birr airtime per month with 10% service fee.

1.2 Statement of the problem

One of the most critical risks that a financial organization must manage is credit risk. Because there is no profit without loan repayment, the issue of credit risk management affects all financial organizations that lend to individuals and legal companies [3].

The telecom sector offers a variety of value-added services (VAS) to customers and service providers. Airtime credit service is one of the VAS offered by MNO, allows prepaid mobile customers to top up their mobile airtime credit at any time and from any location [4].

This service, however, has its own challenges when it comes to loan repayment. Many subscribers usually do not pay back their loans in time or they default. For instance, the Ethio telecom airtime credit service report shows that there is birr 291 million bad debt on April/2020 (see Figure 1.1).

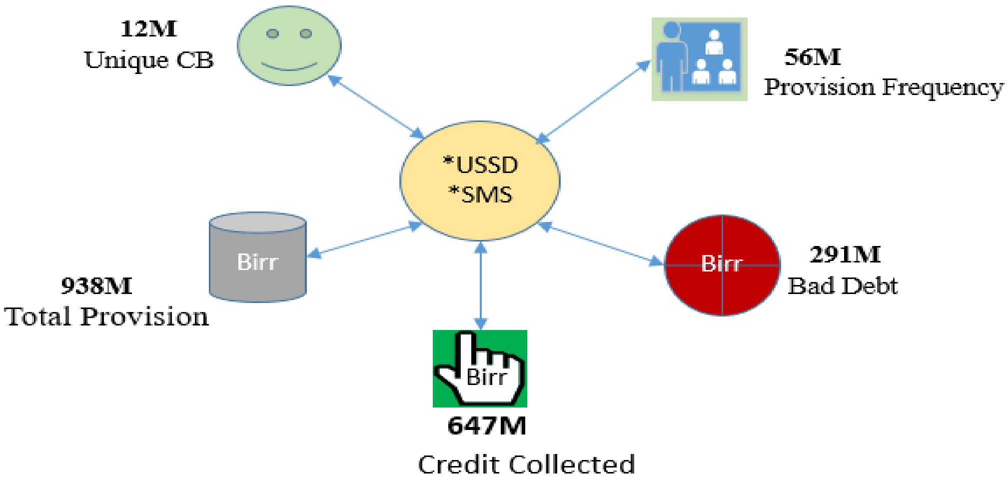


Figure 1 1: Ethio telecom airtime credit service report

To minimize the loss due to bad airtime debt, different researchers propose to use prediction algorithms to estimate the liability of the credit airtime service requester. To perform the prediction, the algorithms use different sets of features computed from customer profile, loan information and usage. These features, however, capture only limited aspects of the service requester, and, hence, affect the performance of the prediction algorithms. In this research, we study the customer and customer activity data to identify features that better help to predict the liability of airtime loan service requesters.

1.3 Objective

The general and specific objective of this research is described below: -

1.3.1 General objective

The general objective of this research was mobile airtime credit risk prediction using customer profile and loan information using the machine learning algorithms.

1.3.2 Specific objective

The following specific objectives are identified in order to achieve the specified general objective:

- To identify new features which could help to improve airtime credit risk prediction
- To study and select machine learning algorithms that are suitable for airtime credit risk prediction
- To evaluate and compare the performance of airtime credit risk prediction models built using existing features and proposed feature

1.4 Scope and limitation of the study

The study covers only the mobile airtime credit service using customer profile and loan information of the subscribers and not mobile package credit.

1.5 Contribution of the study

The main contributions of the study are: -

- Characterize the subscribers on the airtime credit service
- Allows to decide based on the previous knowledge of the subscriber as defaulters and non-defaulters before activating the credit to the subscriber
- It leads to reduces the risk of non-repayment of the credit and increase the revenue of the company

1.6 Research methodology

The methodology of the study that we followed is briefly listed below:

- First, we select the related papers, books and other resources that help us to find and to shape the general objectives, and to design the methodologies of the study.
- In order to do the experiment data was collected from Ethio telecom
- Data preprocessing like data cleaning, data integration and data transformation, feature extraction and feature selection techniques are performed
- Classification machine learning algorithms Random forest, logistic regression, J48 Decision tree and deep learning MLP are selected based on their prediction accuracy in WEKA
- The developed models' performances were evaluated using performance metrics: accuracy, precision, recall, and F-measure
- Finally, results and findings are discussed

1.7 Thesis organization

This study consists of five chapters. Chapter One presents the overall overview of the study, including background of the study, a statement of the problem, objectives, scope and limitation and the contribution of the study. The literature review on airtime credit service, machine learning and deep learning algorithms are detailed in Chapter Two. Chapter Three discusses about the proposed approach. Chapter Four presents the dataset used in the experiment, the experiment setup and the result with discussion. Chapter Five concludes the paper and provide a recommendation along with future works.

CHAPTER TWO: LITERATURE REVIEW

2.1 Mobile airtime credit service

An airtime credit is a credit that can be used to grant a certain amount of speaking time. The amount of time spent chatting on a mobile phone is referred to as "airtime." Service providers typically track time to determine billing expenses. Sending and receiving calls, as well as other wireless transmissions such as email and data files, are all examples of usage. Usage is typically measured in minutes for voice calls. The value of airtime is offered at the present time to a customer to use, and the customer is expected to repay the value of the airtime at a future time based on a set rate of interest[1]. Airtime credit make it easier for the user to make phone calls, send SMS messages, and get bundles for browsing. When phones are loaded with airtime credit, they have numerous benefits, including strengthening connections, facilitating clear communication, and serving as a marketing tool [2].

Using airtime credits, subscribers can receive airtime at any time and from anywhere, make calls at any time, feel secure (can contact emergency services at any time), and have a quick and safe option to get airtime when traditional channels like VC (scratch-able voucher cards) are unavailable [3].

2.2 Airtime credit service flow

Figure 2.1 shows the service flow of airtime credit service implemented in Ethio telecom.

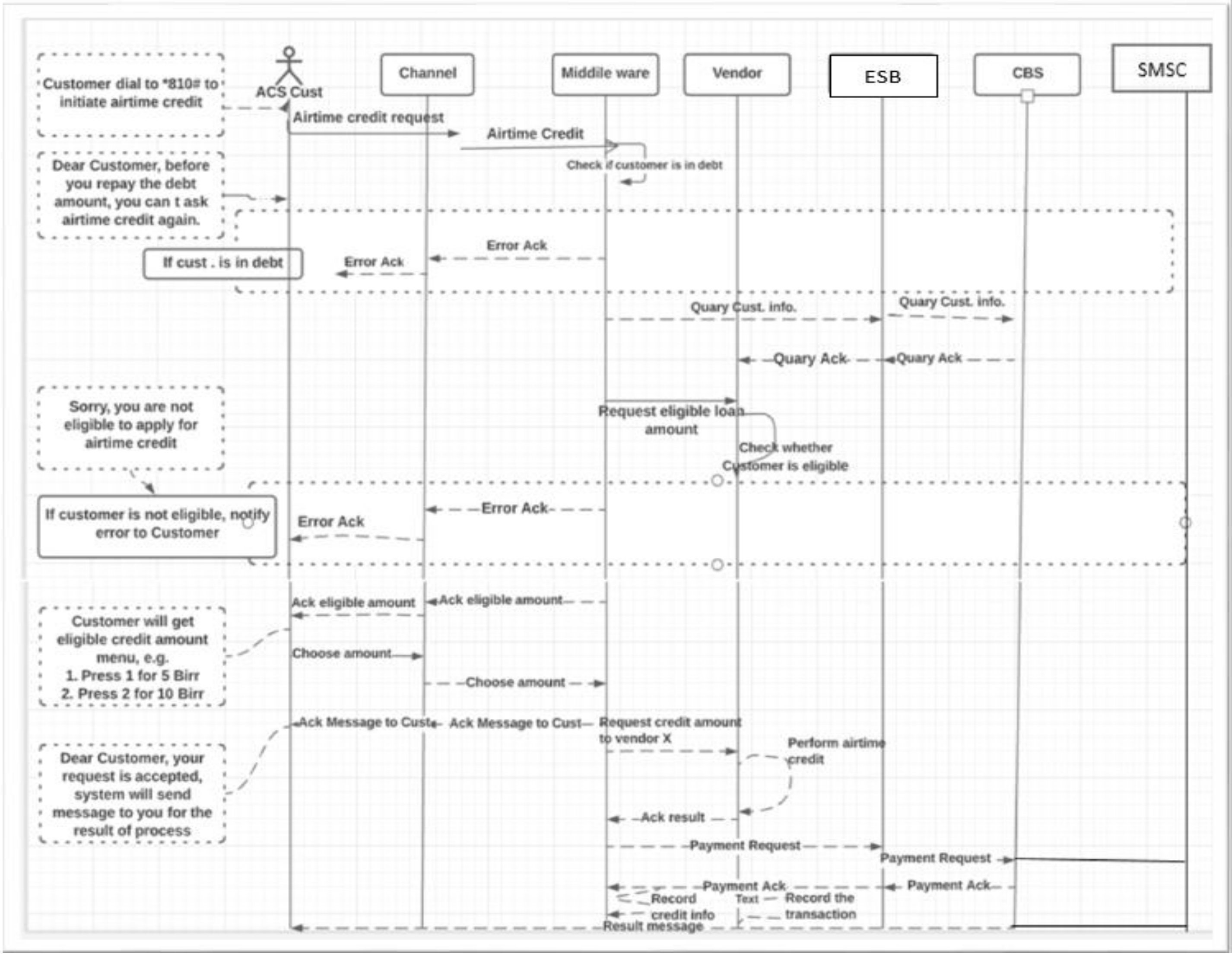


Figure 2. 1: Ethio telecom airtime credit service flow

Steps of the airtime credit service:

1) The request will be forward to Middleware system

2) The Middleware system will check whether customer reach the limit of debts or not. If it reaches, the middleware will response error message to channel, then channel forward to customer are not able to take credit and it terminate from the system.

3) If able to take credit, the middleware system sends "Query Customer info" request to ESB.

4) Then ESB forward request to CBS.

5) Then CBS retrieve subscriber's eligibility information and response to ESB.

6) ESB forward response message to Middleware.

7) Then middleware will send the customer information to the vendor X based on robin algorithm. 8) Then if customer is not eligible, vendor will response to middleware, and middleware will response error message to channel, then channel forward it to customer; whereas if customer is eligible, go to step 8)

8) Middleware system send eligible credit amount to channel

9) Channel show eligible credit amount to customer

10) Then customer choose the credit amount and the channel forward customer's chosen to Middleware

11) Middleware will response to the customer the request is accepted, and the result will be sent upon the process is completed. Then middleware system sends credit request to the same vendor X.

12) Then, the Vendor will confirm the requested credit amount to middleware.

13) The middleware system sends "Payment" request to ESB

14) Then ESB forward request to CBS and CBS perform the payment, then response the result to ESB. After that, ESB forward response to Middleware.

15) Middleware inform the payment result to the same vendor X.

16) Vendor X do the record for the credit request, and response to the middleware.

17) Finally, middleware record credit information and send short message to SMSC, and SMSC deliver the short message to customer for the result.

2.3 Airtime credit service risk

Credit risk, also known as credit default, refers to the likelihood of a customer's failure to repay a loan taken as part of a service. For telecom companies, the service they usually provide is airtime. The credit risk related to the airtime service is called airtime credit risk. Airtime credit risk happens when the subscribers are unable to return the airtime credit they have taken in time. The value added tax for airtime credit is paid as the service is provided. Hence, if unpaid, it will have the additional cost to the telecom company [4].

Credit risk management entails forecasting the likelihood of a loan default. Credit risks are managed differently by different financial institutions, but the goal is always the same minimizing the risk.

Credit risk has long been a heavily researched subject in bank lending choices, with the major goal of the credit evaluation process being to compare an applicant's characteristics to those of past applicants who have repaid the loan amount[5].

For telecom companies, the airtime credit has the potential to boost their revenue per customer. The benefit, however, is true only if the customer repays the debt within a defined time limit. To minimize the risk of a default and increase revenue, many financial institutions use credit scoring. Credit scoring models are used to estimate the customer's ability to pay a particular amount within a certain length of time period in order to limit this risk[6].

Credit scoring is a tool used by financial institutions to help them decide whether to approve or reject a loan application. Credit scoring determines the probability of a customer's ability to repay a loan by looking at different features related to the customer. The features include the customer's historical and financial data such as loan history, number of transactions, and source of income [7]. The benefits of credit scoring include the ability to make faster credit judgments, lower the cost of credit analysis, and keep track of an existing account selection.

2.4 Machine learning

Machine learning is a branch of artificial intelligence (AI) and computer science that focuses on the use of data and algorithms to mimic how humans perceive things and improve their accuracy over time. ML is the science of providing data and information to computers without explicitly programming them to learn and act like humans [8].

Depending on the behavior of the problem and data, machine learning uses different types of algorithms. The datasets are fed to the algorithm and the system learn from each pattern of the data, then the algorithm can predict when new data is fed to the system. This indicates that, first the machine learning can learn using historical data and then make decision for new input data [14].

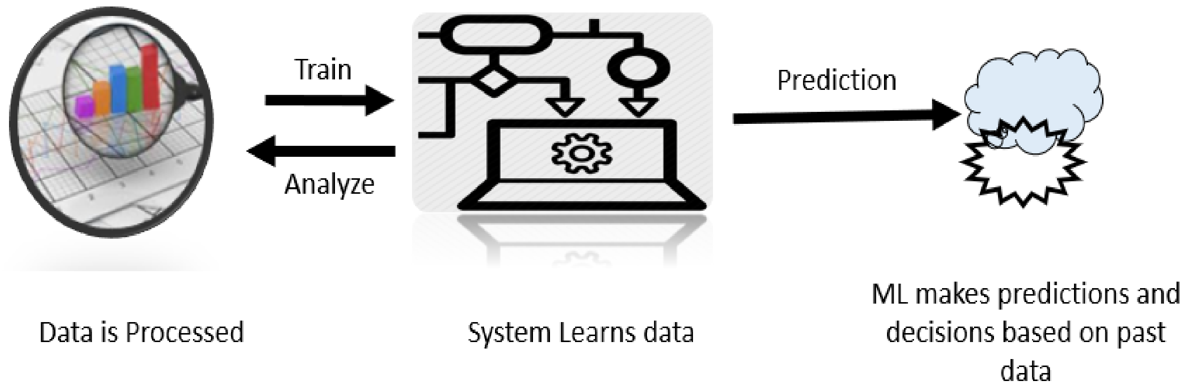


Figure 2. 2: Machine learning working process [8]

Machine learning algorithms allow the systems to make decisions autonomously without any external support. Such decisions are made by finding valuable underlying patterns within complex data[8].

Generally machine learning is specified into descriptive and predictive analytics. The descriptive analytics are used to describe the general property and aspects of the data, whereas, predictive analytics are used when you need to know anything about the future.

Based on the learning capability, type of input and output data, and the behavior of the problem machine learning algorithms are categorized in to supervised, unsupervised, reinforcement and semi supervised learning [15].

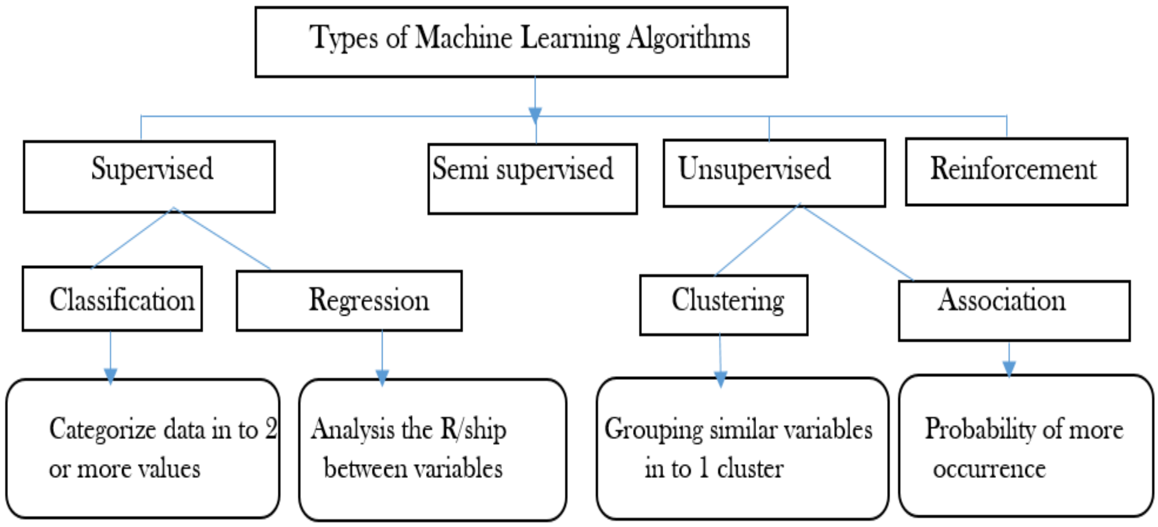


Figure 2. 3: Types of machine learning algorithms

2.4.1 Supervised learning algorithm

Supervised learning is a sort of machine learning that is used when data has an input variable or attributes and produces a target or output value. The algorithm does this by learning from the input data and predicting the output value. Data pre-processing in supervised learning includes operations such as data cleaning, normalization, transformation, feature extraction and selection, and so on. The final training set is the result of data pre-processing [16]. In general, there are two types of supervised learning: (1) classification, and (2) regression [10].

Classification

Classification is a machine learning function that uses a labeled data to train and classify new data into two or more categories. The main goal of the classification is to accurately classify the target class from the input data. The simplest type of classification is the binary classification.

This thesis uses a binary classifier on a data that has two classes: good customers (Non defaulter) and bad customers (defaulters). Non defaulters are customers who repay their loan in the time given to settle the loan, while defaulters are those who cannot repay their loan in the time given to settle the loan.

Some common classification algorithms are the logistic regression, J48 decision tree types, gradient boosting machines, Naïve Bayes, random forest, SVM (support vector machine), Multi-layer perceptron and K-Nearest Neighbor. From these algorithms, we selected to use random forest, J48 decision tree, Multi-layer perceptron and logistic regression in these thesis.

✓ **J48 Decision Tree**

Decision Trees (DT) are a tree-like structure that sorts instances based on feature values to classify them. Each branch of a decision tree indicates a value that the node can accept, and each node represents a feature in an instance to be classified. In a group of observations that make up a data set, decision trees try to establish a strong association between input values and goal values. When a set of input values is found to have a strong link to a target value, all of these values are grouped into a bin, which creates a decision tree branch. It begins with a single root node that divides into several branches, each of which leads to other nodes, each of which can split further or terminate as a leaf node [18].

✓ **Logistic Regression**

One of the supervised categorization machine learning algorithms used in the likelihood of the goal value is logistic regression. Logistic regression is similar to linear regression in that it uses a binomial response variable to model the likelihood of a given result depending on individual characteristics [19].

✓ **Random Forest**

Another supervised machine learning technique for classification and regression is RF. RF produces from many decision trees, as the name implies. A random tree is one that is generated at random from a set of possible trees, each with K random features at each node [18].

As a result, each tree has an equal probability of being sampled. The general RF structure was described in the below Figure 2.4.

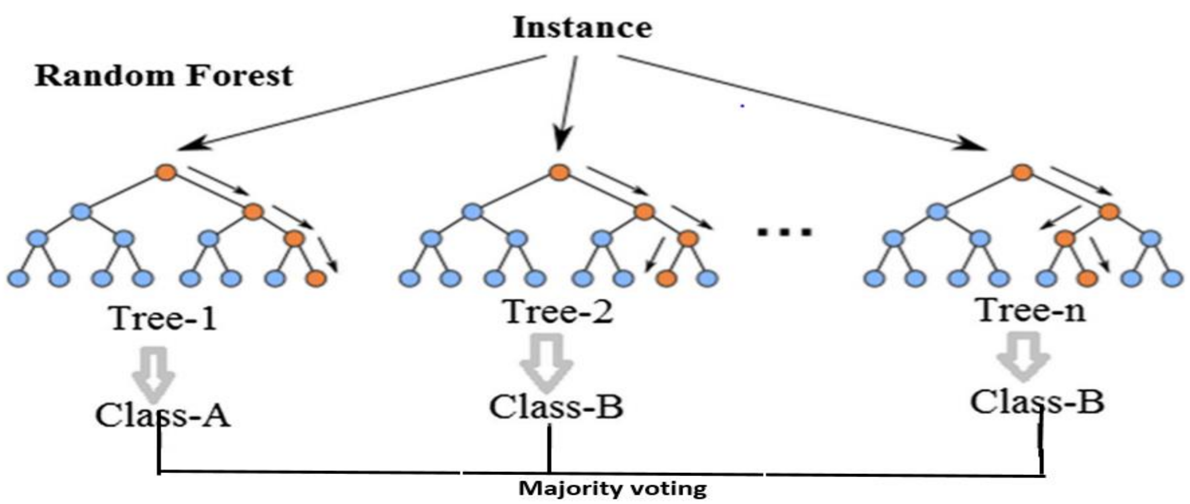


Figure 2. 4: General RF algorithm[20]

As shown in figure 2.4, the tree consists of one root node, several internal and leaf nodes. The internal node consists several leaf node and this leaf node correspond to decision result. The final model of the random forest is decided by the majority votes produced by all individual decision trees.

✓ **MLP** (Multi-layer perceptron**)**

Deep learning offers a variety of methods; in this study, we employ the MLP (Multi-layer Perceptron) which is available as a package on WEKA. WEKADeeplearning4j (MLp classifier) is a WEKA back end deep learning package. It was created in order to incorporate modern deep learning techniques in WEKA [11]. The major purpose of this technology is to make deep learning accessible without needing users to write code through WEKA's GUI. The GUI allows users to do experiments using the following simple steps at the most basic level: (1) Importing data into the Attribute-Relation File Format (ARFF), (2) Creating a neural network architecture, (3) Selecting an experimental procedure, and (4) Executing the experiment [21]. WEKA employs a variety of splitting techniques. K-fold Cross-validation is a popular data resampling method for estimating model true prediction error and tuning model parameters [17].

Regression

Regression function is used to find the relationships between two or more features and use this relationship to classify the target value. For example, when there are two variables and one variable increases the other variable may also increase or decrease, or vice versa. Based on this, each variable has a positive or negative relationship. The regression can be grouped into linear and logistic regression.

✓ **Linear regression**

Linear regression can be used to find a linear relationship between one or more variables. Simple regression and multiple regression are the two types of linear regression. One of the most basic and widely used machine learning methods is linear regression. It is a method for performing predictive analysis that is based on mathematics. Linear regression allows for projections of continuous/real or mathematical variables [22].

2.4.2 Unsupervised learning algorithm

Unsupervised learning, in contrast to supervised learning, is a type of machine learning that can find new patterns in unlabeled data. and is used when there is only input data and no output. To make advantage of the vast amount of unlabeled data, unsupervised learning techniques are employed to learn complicated, highly non-linear models with millions of parameters [8].

The goal of unsupervised learning algorithms is to learn and cluster unlabeled datasets. These algorithms realize hidden patterns or data groupings without the interaction of human intervention. Unsupervised ML algorithms are classified into clustering and association as shown in Figure 2.3

A. Clustering

Clustering is the process of grouping comparable things into a single cluster. To group the comparable things into a single cluster, the algorithms learn the patterns in the input data. Cluster analysis is the formal study of methods and algorithms for grouping, or clustering, objects based on similarities and similar traits [8]. The most common machine learning clustering algorithms are: K-Means Clustering, Mean-Shift Clustering, Hierarchical Clustering and Spectral clustering.

B. **Association**

Associative learning is a type of unsupervised rule-based machine learning that identifies significant relationships between features in a dataset. If someone burns themselves on a hot stove, they may learn to link hot stoves with pain and avoid touching them. K-means clustering, hierarchical clustering, and Self-Organizing Map(SOM) are the most prevalent unsupervised Machine learning techniques.

2.4.3 Semi supervised learning algorithm

Semi-supervised learning is a type of machine learning that combines labeled and unlabeled data sets. Semi-supervised learning is a term that refers to a combination of supervised and unsupervised learning. There is a limited amount of labeled data and a huge number of unlabeled data [7]. The semi-supervised approach aims to solve the scarcity of labelled data by first creating clusters of the unlabeled data using the unsupervised learning technique, and then labeling the clusters using the labeled dataset and supervised learning.

2.4.4 Reinforcement learning algorithm

Reinforcement learning is a sort of machine learning that creates a training sequence based on rewarding selected behaviors and punishing those that aren't. An artificial agent receives either rewards or penalties for the behaviors it does during the learning process. The objective is to maximize the total reward [7]. Figure 2.5 shows the reinforcement information exchange.

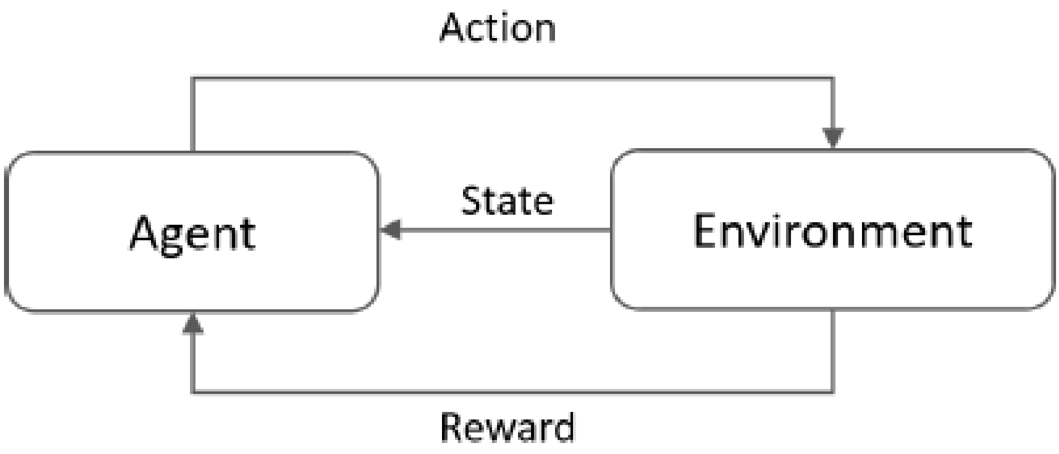


Figure 2. 5: Overview of reinforcement learning[8]

2.3 Deep learning

Deep learning is a machine learning and artificial intelligence (AI) technique that mimics how humans acquire knowledge. Deep learning algorithms are built in a hierarchy of increasing complexity and abstraction, unlike typical machine learning algorithms, which are linear.

Deep learning refers to a class of machine learning techniques that employ numerous layers to extract higher-level features from raw data. Lower layers in image processing, for example, may recognize edges, whereas higher levels may identify human-relevant concepts like numerals, characters, and faces [9].

2.4 Related Work

Credit risk has been the subject of studies in financial institutions. Financial loans are based on a collateral agreement, a relatively large sum of money, and are not risk free. Hence, the loan needs to be assessed physically and financially before it is granted.

Financial institutions are increasingly using AI and machine learning to manage other risks such as financial fraud, money laundering, the risk of not complying with regulations (potential risk of being fined by the regulator), and client behavior that could result in potential income loss to the financial institution. [5,6].

When comparing the present scoring model to the machine learning model established and explained here, the machine learning approach has substantial advantages. The scoring model has a 37.9% accuracy, whereas the machine learning model (neural network) has an 82.1% accuracy. This means that the model is correct in more than 82 cases out of 100. The scoring model has a recall of 1.80%, whereas the machine learning model has a recall of 52.6%.

This means that the scoring model properly predicts that just 1.8% of default loans would be defaulted, whereas the neural network model predicts 52.6 % of defaulted loans. The scoring model's precision is 75.2%, and the model's recall is 15.5%, although for most classification issues, precision is less significant

than recall. It is less expensive to anticipate a loan that will be good as bad (false positive) than it is to predict a loan that will be bad as good (false negative). In the first example, some potential clients may be lost, but the opportunity cost of doing so is likely smaller than the cost of giving away a loan that will not be repaid or will take a lengthy time and effort to repay [5].
Unlike airtime credit, which is technology-dependent, a micro loan, or a relatively small amount of money, is granted at the convenience of the subscriber and solely on the assumption that the subscriber will not fail.

For telecom companies, however, this approach does not work mainly because their airtime credit service has to be given in real time. Usually, customers ask for airtime credit when they want to make a phone call and realize that they do not have a balance. Hence, to minimize the airtime credit risk and give the service, a different risk management approach needs to be followed. In this regard, different researchers proposed different approaches. In this section, we discuss machine learning based approaches.

The role of data mining in predicting airtime credit risk was investigated and the experiment was carried out using WEKA, an open source data mining tool. To find the best performing model, a variety of classification techniques were used. J48 Decision Tree, Nave Bayes, and Logistic Regression. The algorithms were built and tested using telecom prepaid subscriber usage data, which included 86, 024 cases and eleven attributes. The models' performance was further assessed using a confusion matrix. The J48 decision tree model outperformed the other classifiers with a precision, recall and F-measure values of 98.6% and a ROC area threshold of 99.6%. The authors used 10-fold cross validation to avoid bias. The model created using logic regression has a 97.1717 percent accuracy. The accuracy of the Multilayer Perceptron and Nave Bayes classifiers was 96.7622 percent and 94.6355 percent, respectively. Some significant rules and parameters are derived from the selected classifier, which can support in the decision-making process for airtime credit [3].
The problem with this study is that it only uses customer profile and usage features to make predictions, which reduces the accuracy of its results.

The authors employ machine learning approaches to develop a credit score model for airtime loans and identify defaulters and non-defaulters in this research. Credit scoring techniques are reviewed, and this knowledge is used to construct a machine learning model that is suitable for airtime lending. Over three million loans belonging to 41 thousand customers with a three-month repayment period have been examined. Logistic Regression, Decision Trees and Random Forest are evaluated for their ability to classify defaulters and they use the cross-validation approaches to avoid the bias. They found that Random Forest was the best classifier, with an accuracy of 80%. Finally, they decide on a base rule: if the default rate is less than 2%, it is better to offer everyone a loan [7].

The problem with this research is that they generalize the scoring method based on some specific customer information and they cannot clearly identify the performance evaluation of the algorithms. This study aids in understanding the usage of other important airtime credit customer information, which is useful in improving prediction accuracy and scoring method.

The purpose of this study is to investigate the trend and pattern of advanced airtime lending, recovery, and probability distribution in Nigeria. Because of its ability to handle the large volume of data generated by subscribers each day, the Audit Command Language (ACL) was chosen (over 10milllion rows). For their analysis, they used 40 million customers, which included both active and inactive subscribers. In this analysis, 5 million active subscribers were used as a sample population for airtime gross lent per lending period. The average daily loan amount was N118, and the maximum loan amount was N142. On a daily, the average amount borrowed by a customer was N128, while the company recovered a minimum of N78 and a maximum of N98. According to the findings of the analysis, the total amount recovered by telecommunications operators on a daily basis is N88. From the analytical result they suggested that more airtime data lending occurs than recovery [13].

This research is used for our study to understand what appears to be the trend of lending and recovery for airtime credit service, and from this we conclude that there is a risk for repayment.

CHAPTER THREE: SYSTEM MODEL

3.1 Proposed Features

The proposed features are newly extracted features that are utilized to improve the model's prediction accuracy. The proposed features in this study are derived from the Ethio telecom database and are those that are used to understand the subscriber's loan information behavior. Customer profile, loan information, and usage are the feature categories in which we added new features. We add an education feature from the customer profile to understand the subscriber's educational level and to ensure that the subscriber is aware of the credit and has a high likelihood of repaying the credit. We add the subscriber loan amount, loan frequency, and repayment amount of the loan information from the loan information category. These loan related features are used to have good information about the subscriber's loan usage, and the last was the recharge frequency, which indicates how many times the subscriber's behavior with respect to loan. The usage category captures information about the subscriber's phone usage by looking at the number recharge frequency. Table 3.1 provides detailed information about the proposed features.

Table 3.1: Proposed features

Features	Description	Rational	Measurement
Education	Educational level of the subscriber	People who are educated are usually expected to be more responsible and would pay their loan in time.	Categorical value collected from the database.
Recharge frequency	The number of frequencies to recharge the service number per month.	Subscribers who recharge their number more frequently are more likely to pay their loan in time.	Summation of the number of recharges per month.
Subscriber Loan amount	Total subscriber loan amount taken per month	Subscribers who have less loan amount are mostly pay their loan on in time.	Sum of loans taken per month
Repaid amount	The amount of repaid loan per month	When Subscribers have more repay amount they are more likely expect to pay their loan.	Sum of money paid for loan per month
Loan frequency	Number of loans taken per month	Subscribers are not able to use loan when they cannot repay their previous loan. So, when they use loan frequently this indicates, they are more likely pay their loan in time.	Number of loans taken per month

3.2 Methodology

To assess the effect of the newly introduced features in airtime credit risk, we adopted a methodology that is used in similar studies [3,7]. The overview of the methodology is shown in Figure 3.1. Below we detail each step in the methodology.

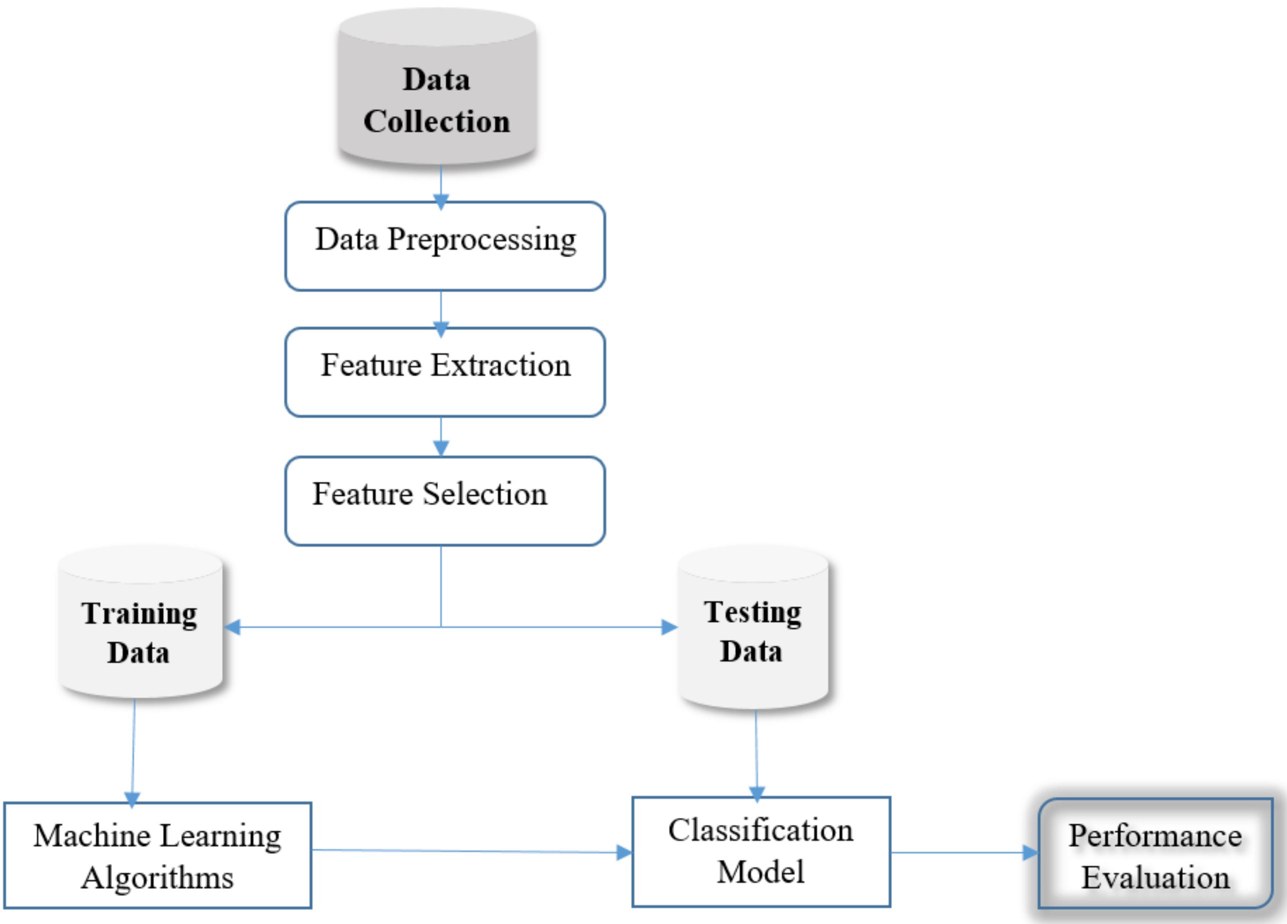


Figure 3. 1: General Methodology of the Study

3.2.1 Data Collection

The data for the study is collected from telecom prepaid mobile service subscribers' business support system (BSS), which included the customer relationship management (CRM) system and the convergent billing system (CBS).

3.2.2 Data Preprocessing

Data preprocessing refers preparing the dataset to make it suitable for the next steps, i.e., feature extraction and training. knowledge discovery during the training phase is more challenging if there is a lot of irrelevant and redundant information. In this study, we performed data cleaning, integration and transformation.

Data cleaning: The process of preparing data for prediction by removing the incomplete data, inconsistence data and organizing the raw data in order to increase the performance of the models.

Data integration: Data preprocessing technique that involves combining data from various data sources.

Data transformation: The process of changing the format, structure, or values of the data. Data was collected from different data source and these data source have their own data format. in order to come up the different data source in to the same format data transformation is one of the best data preprocessing technique.

3.2.3 Feature Extraction

At this stage, we extracted features that we think are helpful to predict airtime credit risk. The features extracted are classified into two groups: existing and new. The existing features are features that are used in previous state-of-the-art risk prediction studies [3,7], while the new features are those that we propose to include for the prediction. Table 3.2 summarizes the existing and new features extracted from the dataset collected in the previous step.

Table 3. 1: Extracted Features of the Study

Existing Features		**Proposed Features**	
Features	**Description**	**Features**	**Description**
Service Number	Unique identity of the Subscriber	Education	Educational level of the subscriber
Customer Age	The age of the customer	Recharge frequency	The number of frequencies to recharge the number per month
Network Age	Network age of the service number	Subscriber loan amount	Total subscriber loan amount taken per month
Recharge Amount	The total amount of recharge per month	Repaid amount	The amount of repaying loan per month
Voice usage log	The amount used for voice	Loan frequency	Number of loans taken per month
Data usage log	The amount used for data		
SMS usage log	The amount used for SMS		

Status	The status of the number		

3.2.4 Feature Selection

Feature selection is the process of identifying the most significant features in order to improve prediction accuracy. Some features could have less impact on the prediction performance probably because they provide duplicate information. To identify such features, we computed Pearson correlation coefficient. One of the features is selected from those features that correlated strongly.

3.2.5 Data Splitting

Data splitting is used in machine learning to divide data into training and testing datasets. The models learned from the training dataset, and then we fed the testing data into them and evaluated their performance. Different data splitting strategies exist. In this study, we used the K-fold cross validation. In cross-validation the sampling is done so that no two test sets are the same. The available learning set is partitioned into k disjoint subsets of approximately equal size. The term "fold" refers to the number of subgroups that result. Especially for small dataset k-fold validation is more useful and avoids bias.

3.2.6 Model Building

To build the prediction model, we used machine learning algorithms. The machine learning algorithms use as an input the training data and build a prediction model.

3.4 Performance Evaluation Metrics

On this study, to assess the performance the models built, we use standard evaluation metrics such as accuracy, precision, recall and F-measure (see Section 4.5).

CHAPTER FOUR: EXPERIMENT

4.1 Dataset

This subsection discusses how the data is collected from different systems of the company and integration techniques used to integrate the data. In this study, we take three-month average data on the airtime credit service. This research uses Ethio telecom prepaid mobile service subscriber's data and the data is collected from business support system (BSS) which includes the customer relationship management (CRM) system and convergent billing system (CBS).

Customer profile information Data

The customer profile data were collected from CRM database which consists of information such as customer name, gender, birthdate, age, place of birth, address, service number, subscription date, billing information, customer type, language, offer type, payment type, education and income. In this study, we used the information in the CRM database as features. To protect the personal privacy of the customers', we omitted the customer name feature from the dataset. In the database, 75% of gender and income features do not have values, and hence, removed.

As the study concerns about the residential customers that are prepaid mobile users the customer type and payment type are omitted from the study. Gender, address, birthdate, language and place of birth features do not have any contribution to the goal of the study and are not selected on this study.

Customer Loan Information Data

The loan information is collected from the CBS database of the company. Loan information consists list of customers who have taken airtime credit. The data has different fields which include service number, subscriber loan amount, loan frequency, a loan taken date, repaid amount, repaid date, and loan access channel. For this study, we use subscriber loan amount, repayment amount and loan frequency features.

Customer Usage Information Data

The call record detail (CDR) consists of customer information data such as recharge amount, recharge frequency, recharge date, recharge channel, voice usage log, data usage log, SMS data usage log and date of each log. From these features recharge date and recharge channel are not used in this study. For this study, we use three months (from August to October 2020) Ethio telecom prepaid mobile customers' data. The total number of subscriber records collected for the three months' period is 150,000.

After collecting the data, we performed data cleaning which involves deleting undesired data, correcting incomplete data and resolving inconsistencies (see Section 3.2.2). The cleaning step has reduced the data to a total of 90,000 subscribers' records (see Table 4.1).

Table 4. 1: Show the Preprocessed Data Records

Collected Dataset		**Final Dataset**	
150,000 subscribers		90,000 subscribers	
Defaulters(Bad)	Non Defaulters(Good)	Defaulters(Bad)	Non Defaulters(Good)
65,000	85,000	45,000	45,000

Because the data was gathered from various secure databases, it was required to combine them in order to use the integrated dataset. It is necessary to have an attribute to link the collected data from many sources in order to integrate the data. To achieve this, the service number is utilized as a key to integrate the data from various sources. However, due to subscriber's privacy, the service number was removed after the data was integrated. The integrated data is saved in a CVS file.

4.2 Experimental Setup

To assess the performance of the proposed features in airtime credit risk prediction, we conducted two experiments. The first experiment uses a dataset containing the features used in baseline studies [3, 7] as an input, while the other experiment uses a dataset containing the existing and proposed features as an input. The existing and new features are shown in Tables 3.2 and 3.1, respectively

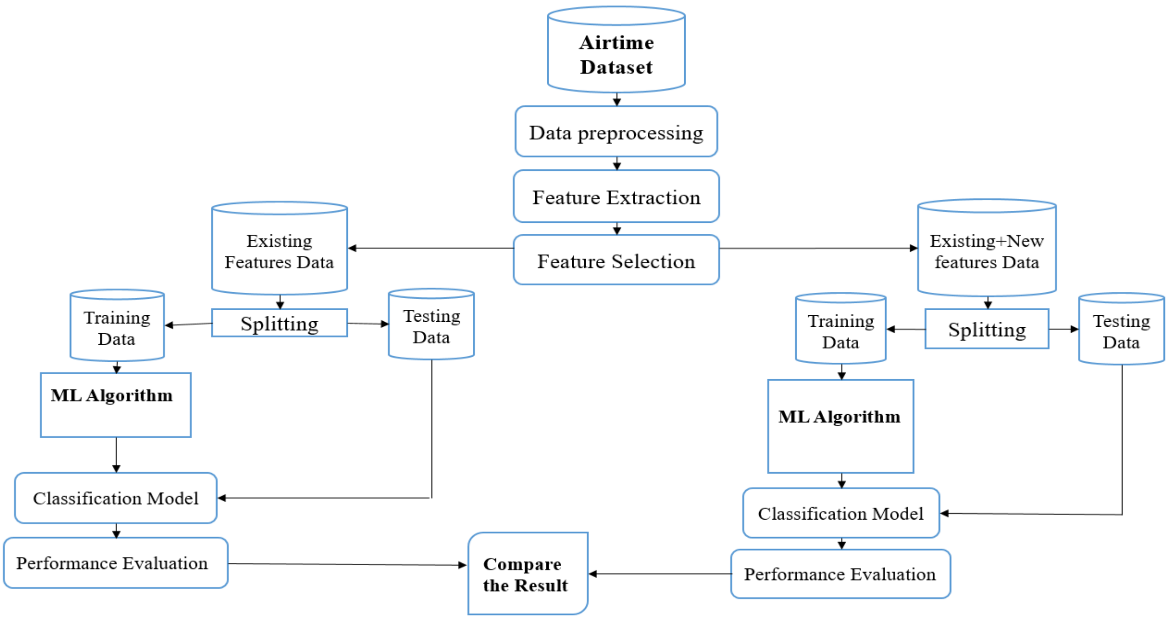


Figure 4. 1:Experimental Design of the Study

4.3 Feature selection

The original dataset contains 28 features. To identify and select from features that correlate, we computed the correlation coefficient between the features using Pearson correlation method. The correlation of the features is shown in Table 4.2.

Table 4. 2: : Pearson correlation coefficient values between the features. The bold values show strong correlation.

Features	Loan Amount	Subscriber Lo Frequency	Repaid Amount	Recharge Amount	Recharge Frequency	Voice	Data	SMS	Network Age
Subscriber loan Amo	1								
Loan frequency	0.693806	1							
Repaid Amount	**0.997814**	0.693608	1						
Recharge Amount	0.70071	0.534737	0.701604	1					
Recharge Frequency	0.462447	0.621977	0.46247	0.732563	1				
Voice	0.471515	0.375243	0.472496	0.591152	0.452874	1			
Data	0.53215	0.385252	0.532389	0.624752	0.456208	0.181965	1		
SMS	0.099006	0.11065	0.098753	0.110622	0.088041	0.091356	0.072213	1	
Network Age	0.126752	0.119273	0.129623	0.131276	0.01878	0.146545	0.028668	0.05852	1

The features with a correlation value of more than 95% are strongly correlating features, and hence, considered to have the same information.

Features that strongly correlate usually increase the model's complexity and lower its performance, and hence, its recommended to select and keep only one of the features. In our dataset, loan and repayment amounts are found to strongly correlate. From the two features, we randomly selected the subscriber loan amount to be part of the features used in the next step. The final list of features selected for the experiments are shown in Table 4.3.

Table 4. 3: The final selected features

Feature	Description
Customer Age	The age of the customer
Network Age	Network age of the service number
Education	Educational level of the subscriber
Recharge Amount	The total amount of recharge per month
Recharge frequency	Number of frequency to recharge the number per month
Voice usage log	The amount used for voice
Data usage log	The amount used for data
SMS usage log	The amount used for SMS
Subscriber Loan amount	Total loan amount taken by subscriber per month
Loan frequency	Number of loan taken per month
Status	The status of the number

4.4 Model building

On this study, four machine learning models are used. From the machine learning the classification algorithms: logistic regression, random forest and j48 decision tree, and from the deep learning algorithms MLP algorithms are selected for the experiment.

4.5 Evaluation metrics

Performance evaluation metrics are used to determine how effectively a machine learning model performed with the test data was given. A learning model's fundamental purpose is to generalize successfully on data that has never been seen before. On specific learning models, specific metrics must be utilized, and not all metrics may be employed in a single model. To compute the metrics, we use a confusion matrix as shown in table 4.4 below.

Table 4. 4: Confusion matrix

		Actual value	
		Defaulter subscribers (Bad)	**Non-Defaulter subscribers(Good)**
Predicted value	**Defaulter subscribers (Bad)**	**TP**	**FP**
	Non-Defaulter subscribers(Good)	**FN**	**TN**

In the confusion matrix:

- True Positive(TP) refers to defaulters who are predicted as defaulters.
- True Negative (TN) refers to non-defaulters who are predicted as non-defaulters.
- False Positive (FP) refers to non-defaulters who are predicted as defaulters.
- False Negative (FN) refers to defaulters who are predicted as non-defaulters.

A. Accuracy

Accuracy is calculated by dividing the correct predictions to the overall prediction value (see Equation 1).

$$Accuracy = \frac{TN+TP}{TP+FP+TN+FN} \quad \text{---}(1)$$

B. Precision

Precision is the ratio of correct prediction to the sum of true and false positive prediction (see Equation 2).

$$Precision = TP/(TP+FP) \quad \text{--}(2)$$

C. Recall

Recall is the ratio of correct prediction to the sum of true positive and false negative prediction (see Equation 3)

$Recall = TP/(TP+FN)$ ---(3)

D. F-Measure

The F-score is a way of combining the precision and recall of the model. It is calculated as the harmonic mean of precision and recall (see Equation 4).

F1-Measure=2*((precision*recall)/(precision+recall)) --(4)

4.6 Tool

WEKA is an open source powerful tool for developing several most widely used machine learning algorithms and contains tools for data preprocessing, classification, regression, clustering, association rules and visualization [12]. WEKA was user friendly to use and detect various hidden patterns in the dataset. So, on this study, we use the classification function to conduct the experiment.

4.7 Results and Discussion

Table 4.5 shows accuracy, precision, recall and F-measure values of the two types of datasets containing only existing features and existing plus new features. The metrics are computed for the four selected algorithms.

Table 4. 5: Performance evaluation result using [3,7] dataset and new +existing dataset

Algorithm	**Dataset**	**Performance Evaluation Metrics**			
		Accuracy	**Precision**	**Recall**	**F-Measure**
J48 DT.	Customer profile features [3].	81.71%	0.799	0.848	0.823
LR.		73.61%	0.719	0.774	0.746
RF.		82.18%	0.805	0.849	0.827
MLP		73.55%	0.748	0.710	0.729
J48 DT.	Loan information features [7].	83.53%	0.891	0.765	0.823
LR.		78.07%	0.738	0.870	0.799
RF.		82.60%	0.863	0.775	0.817
MLP		77.52%	0.832	0.686	0.750
Existing and new features					
J48 DT.	[3,7] +New features	86.20%	0.868	0.854	0.861
LR.		80.02%	0.779	0.852	0.814
RF.		87.71%	0.895	0.854	0.874
MLP		80.44%	0.829	0.767	0.797

Accuracy

Table 4. 6: Delta accuracy of Existing +New dataset with customer profile and usage [3] dataset

Algorithms	Accuracy of Existing +New dataset	Accuracy of Cust.profile& usage [3] dataset	Δ Accuracy
J48_DT	86.20%	81.71%	4.49%
LR	80.52%	73.61%	6.91%
RF	87.71%	82.18%	5.53%
MLP	80.44%	73.55%	6.89%

Table 4. 7: Delta Accuracy of Existing +New dataset with loan information [7]dataset

Algorithms	Accuracy of Existing +New dataset	Accuracy of loan information[7] dataset	Δ Accuracy
J48_DT	86.20%	83.53%	2.67%
LR	80.52%	78.07%	2.45%
RF	87.71%	82.60%	5.11%
MLP	80.44%	77.52%	2.92%

Tables 4.6 and 4.7 show the improvements achieved in accuracy while using the existing + new features to predict airtime credit risk. For all the algorithms, the combined feature set shows an improvement over the existing features (see Figure 4.2, Tables 4.6 and 4.7). For the existing features, a better performance in accuracy is observed while using customer profile and usage features when compared to loan features. When comparing the accuracy achieved while using customer profile and usage with the existing and new features, the highest improvement, i.e., 6.91%, is shown for LR. LR also gives the highest accuracy measure for the existing and new features.

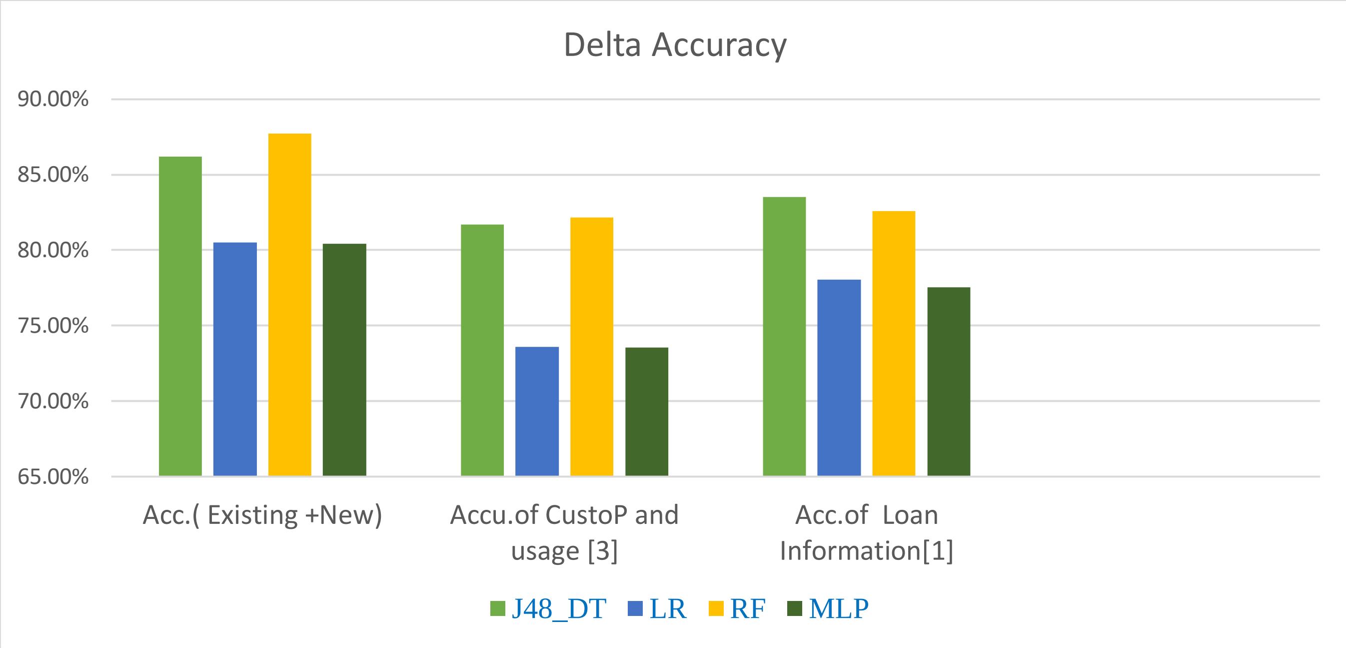


Figure 4. 2: The overall accuracy using all dataset

F-measurement

Table 4. 8: Delta F-measure of Existing +New with Customer profile and usage [3] dataset

Algorithm	F_M of Existing +New data set	F_M Customer Profile and Usage[3] dataset	Δ F-Measure
J48_DT	0.861	0.823	3.8%
LR	0.814	0.746	6.8%
RF	0.895	0.827	6.8%
MLP	0.797	0.729	6.8%

Table 4. 9: Delta F-measure Existing +New with loan information [7] dataset

Algorithm	F_M of Existing +New dataset	F_M of Loan Information[7] dataset	Δ F-Measure
J48_DT	0.861	0.823	3.8%
LR	0.814	0.799	1.5%
RF	0.895	0.817	7.8%
MLP	0.797	0.750	4.7%

Tables 4.8. and 4.9 show the improvements achieved in F-measure while using the existing + new features to predict airtime credit risk. For all the algorithms, the combined feature set shows an improvement over the existing features (see Figure 4.3, Tables 4.8 and 4.9). For the existing features, a better performance in F-measure is observed while using loan information features when compared to customer profile and usage features. When comparing the F-measure achieved while using loan information with the existing

and new features, the highest improvement, i.e.7.8%, is shown for RF. RF also gives the highest f-measure for the existing and new features.

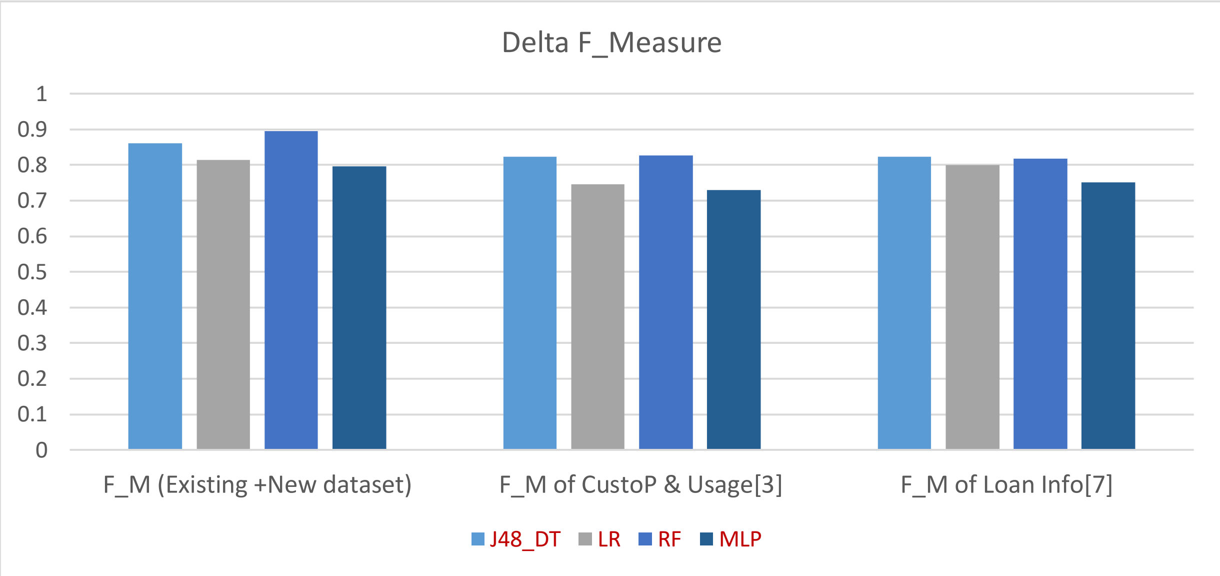
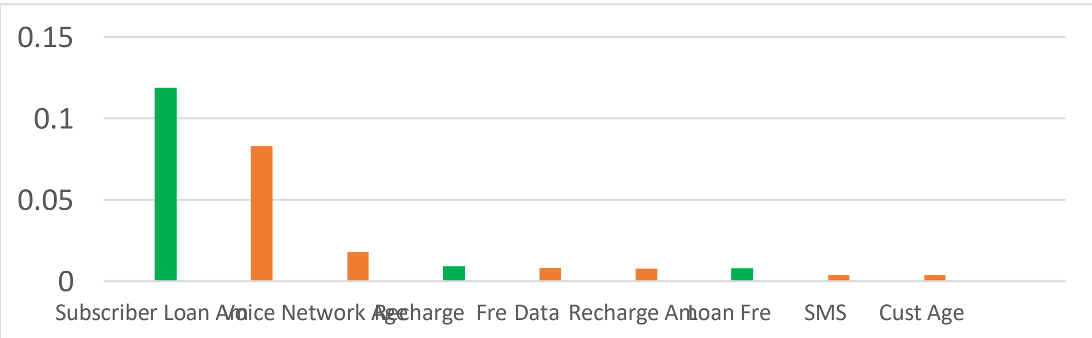


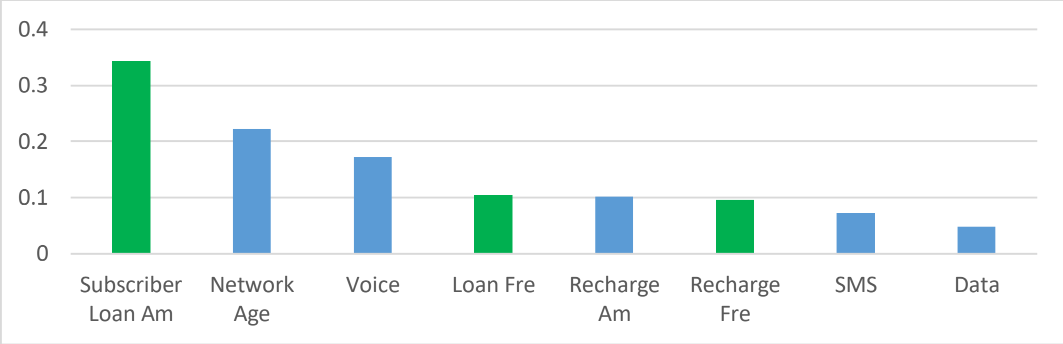
Figure 4. 3: The overall F-measure using all dataset

4.7.1 Feature Ranking

The experiments conducted in Section 4.6 show that the new + existing features give a better performance in predicting airtime credit risk. To see which features contributed to this improvement in performance, we ranked the features using two different feature ranking algorithms found in Weka: CorrelationAttributeEval and GainRatioAttributeEval. The results of the ranking algorithms are shown in Figure 4.4. In the figure, the green bars show the rank of the newly added features. In both algorithms, the newly added feature subscriber loan amount is ranked first. Loan frequency and recharge frequency are ranked 4th and 6th when the CorrelationAttributeEval algorithm is used. For the GainRatioAttributeEval algorithm the new features recharge frequency and loan frequency are ranked 4th and 7th respectively. This shows that the newly added features, in particular subscriber loan amount, loan frequency and recharge frequency have helped to improve the prediction of airtime risk.



1: CorrelationAttributeEval



2: GainRatioAttributeEval

Figure 4. 4: Rank of the features used on the study

4.7.2 Discussion

The accuracy and f-measure of machine learning algorithms on airtime credit risk prediction using existing and existing+ new feature dataset are shown in the results. For all algorithms used in the experiment, the combined feature set shows an improvement in accuracy and F-measure performance over the existing features. When comparing the accuracy of existing +new with existing features using loan information, customer profiles and usage features, the highest improvement is seen when we use the customer profile and usage feature. The improved result was 6.91 percent for LR (as shown in Figure 4.2, Tables 4.6 and 4.7). The highest accuracy improvement is indicated by LR.

When comparing the F-measure with existing with existing + new features loan information, customer profiles and usage features the highest improvement is 7.8% when we use the loan information (as shown in Figure 4.3, Tables 4.8 and 4.9). RF has the highest f-measure for existing+ new features. Finally, the proposed customer profile and usage features, as well as loan information, have a significant impact on the models' prediction accuracy, as show in the findings. On this study, the highest accuracy was 87.7% using the random forest machine learning algorithm.

Finally, we compare the results to the scoring mechanism that Ethio telecom currently employs. To activate the airtime credit service, Ethio telecom currently uses the monthly recharge amount and the subscribers' network age. The data we collected for this study shows that 850,000 birr is not repaid at the time of settlement using this scoring approach.

The highest machine learning algorithm, the random forest, has a 60 percent accuracy using the current scoring technique of the airtime credit service. On the other hand, on this study has an accuracy rate of 87.7%.

As a result of this, the company was able to activate the airtime credit with a 40% error rate. whereas when the company used the model developed from this study, the inaccuracy was reduced to 13%. The correctly activated airtime credit of 276,000 birr was reduced. This means that an amount of 850,000 birr was incorrectly activated to the subscribers. However, when Ethio telecom used the model developed from this study, the airtime credit risk was reduced, and the company received 574,000 birr from defaulters.

From this result, we understand the company's current scoring method (monthly recharge amount and network age of subscribers) was insufficient to reduce risk and enable subscribers to activate airtime credit service. Finally, we conclude the subscriber's profile, usage (voice, data, and SMS), and loan information are more powerful features that we should consider before activating airtime credit service in order to reduce the risk.

CHAPTER FIVE: CONCULUTION AND RECOMMENDATION

5.1 Conclusion

When prepaid mobile subscribers are unable to recharge their service number, they can use the airtime credit service. The credit service helps in increasing the satisfaction of service subscribers as well as the telecom operator's revenue. However, the service comes with its own set of risks, as there is no guarantee that prepaid customers will repay the amount they have taken. Some researchers have proposed prediction algorithm approach to address this problem. These approaches are better in terms of customer profile, usage, and loan information. However, they did not take into account the subscribers' educational background, subscriber loan amount, loan frequency, or recharge frequency.

In this research work, we propose an approach that takes into consideration education and recharge frequency from customer profile and subscriber loan amount and loan frequency from loan information aspects of the prediction. We conducted this thesis with a set of 90,000 subscriber data for the experiments using four machine learning algorithms: decision tree (DT), logistic regression (LR), random forest (RF), and MLP. To analyze the impact of the proposed approach the accuracy of the existing +new dataset was enhanced by 6.91 percent using the LR when we use customer profile and usage features.

As a result, when we used the loan information, the delta of F -measure of the existing + new dataset improved by 7.8% using RF. Finally, all models achieve satisfactory performance result and the RF have the highest prediction accuracy.

5.2 Recommendations

This study contribution could be used to develop and execute the airtime credit service in order to improve customer satisfaction and revenue for the service provider. In addition, the study's finding help to give direction on how telecom companies reduce the number of customers who do not pay their loan in time. Finally, based on the study's findings, we have the following recommendations.

- This study used different classification algorithms in which random forest performed better than others logistic regression, J48 decision tree and MLP.
- Based on the feature ranking algorithms the subscriber loan amount, voice usage, loan frequency, network age and card frequency features can be used at a time of decision making as they had shown strong prediction power which can help to reduce the risk of airtime credit defaulting.
- The findings of this study can be used as input by the company, which can then be integrated with its airtime credit system to improve service delivery while reducing the risk of default.
- Because the data preparation stage takes a long time, it is recommended that large companies, such as Ethio telecom, implement or own a data warehouse where all of their data can be formally stored

for a longer period of time and where different machine learning techniques can be used to simplify day-to-day service provisioning.

5.3 Future Works

- ✓ In the future we recommend for the researchers to integrate different features such as package usage of the subscriber in order to clearly understand loan usage of the subscribers and enhance performance of the models.
- ✓ Other classification deep learning algorithms might reveal a better accuracy. Therefore, further research must be conducted using other algorithms.

References

[1] K. Garba and S. Sa, “Airtime Credit Loan and Service Charge / Fee by Telecommunications Service Providers in Nigeria : Islamic Law Perspective,” vol. 21, no. 8, pp. 8–14, 2019, doi: 10.9790/487X-2108040814.

[2] I. Aamo, A. Myom, and Y. I. Shehu, “Airtime Credit Banking : From Two Applications to One Application,” pp. 10–15, 2017, doi: 10.4236/jcc.2017.510002.

[3] O. Tarekegn, “Application of Data Mining Technique for Predicting Airtime Credit Risk : The Case of Ethio Telecom,” 2019. Jun-2019

[4] S. Madise, “Mobile money and airtime : emerging forms of money.” February 2015

SSRN Electronic Journal Forthcoming: DOI:10.2139/ssrn.2589058.

[5] Z. Ereiz, “Predicting Default Loans Using Machine Learning (OptiML),” no. November 2019, pp. 3–7, 2020, doi: 10.1109/TELFOR48224.2019.8971110.

[6] I. J. M. Education, C. Science, D. K. Gupta, and S. Goyal, “Credit Risk Prediction Using Artificial Neural Network Algorithm,” no. May, pp. 9–16, 2018, doi: 10.5815/ijmecs.2018.05.02.

[7] B. Dushimimana, Y. Wambui, T. Lubega, and P. E. Mcsharry, “Use of Machine Learning Techniques to Create a Credit Score Model for Airtime Loans,” no. Mode 2017, 2020, doi: 10.3390/jrfm13080180.

[8] S. Sah, “Machine Learning: A Review of Learning Types,” *ResearchGate*, no. July, 2020, doi: 10.20944/preprints202007.0230.v1.

[9] Q. Liu and Y. Wu, “Supervised Learning,” no. April, 2015, doi: 10.1007/978-1-4419-1428-6.

[10] V. Nasteski, “An overview of the supervised machine learning methods,” *Horizons.B*, vol. 4, no. December 2017, pp. 51–62, 2017, doi: 10.20544/horizons.b.04.1.17.p05.

[11] E. Frank, M. Hall, G. Holmes, R. Kirkby, and I. H. Witten, “WEKA-A Machine Learning Workbench for Data Mining WEKA A Machine Learning Workbench for Data Mining,” no. May 2014, pp. 0–10, 2010, doi: 10.1007/978-0-387-09823-4.

[12] I. WEKA, G. N. U. General, P. License, N. Zealand, W. Environment, and K. Analysis, “Introduction to WEKA- A Toolkit for Machine Learning.” **DOI:** 10.1109/ISACV.2017.8054931

[13] U. Journal and O. F. Business, “Unilag journal of business vol. 6 no. 1 2020,” vol. 6, no. 1, pp. 96–113, 2020.

[14] D. Björkegren and D. Grissen, “Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment,” pp. 1–28. Daniel Björkegren1 and Darrell Grissen

[15] T. O. Ayodele, “Types of Machine Learning Algorithms.” Taiwo Oladipupo Ayodele University of Portsmouth ,United Kingdom

[16] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, “Data Preprocessing for Supervised Leaning,” no. June 2014, 2006. S. B. Kotsiantis, D. Kanellopoulos and P. E. Pintelas

[17] D. Berrar, “Cross-Validation Cross-validation,” no. January 2018, pp. 0–8, 2019, doi: 10.1016/B978-0-12-809633-8.20349-X.

[18] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, “Random Forests and Decision Trees,” *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 272–278, 2012. Jehad Ali1 , Rehanullah Khan2 , Nasir Ahmad3 , Imran Maqsood4

[19] S. Sperandei, “Lessons in biostatistics Understanding logistic regression analysis,” no. February, 2014, doi: 10.11613/BM.2014.003.

[20] A. K. Mishra, S. V Ramteke, P. Sen, and A. Kumar, “Random Forest Tree Based Approach for Blast Design in Surface Mine,” *Geotech. Geol. Eng.*, 2017, doi: 10.1007/s10706-017-0420-8.

[21] S. Lang, F. Bravo-marquez, C. Beckham, M. Hall, and E. Frank, “Knowledge-Based Systems WEKADeeplearning4j : A deep learning package for WEKA based on,” *Knowledge-Based Syst.*, vol. 178, pp. 48–50, 2019, doi: 10.1016/j.knosys.2019.04.013.

[22] D. Maulud and A. M. Abdulazeez, “A Review on Linear Regression Comprehensive in Machine Learning,” J. Appl. Sci. Technol. Trends, vol. 1, no. 4, pp. 140–147, 2020, doi: 10.38094/jastt1457.

Appendix

Machine Learning Based Mobile Airtime Credit Risk prediction Using Customer Profile and Loan Information

Shashu Brhane Berhe
Addis Ababa Institute of Technology
Addis Ababa Universtiy, Addis Ababa, Ethiopia

sisethiob@Gmail.com

Surafel Lemma Abebe
Addis Ababa Institute of Technology
Addis Ababa University, Addis Ababa, Ethiopia
surafel.lemma@aait.edu.et
surafellemma.a@gmail.com

Abstract-
Airtime credit service is usually used by prepaid mobile subscribes when they cannot recharge their service number. The credit service helps to increase service subscribers' satisfaction and revenue of the telecom operator. However, the service has its own risk as there is no guarantee that prepaid subscribers would pay back the credit they took. To solve this issue, several researchers propose to predict airtime credit risk using machine learning based approaches. To build the prediction models, these approaches use either customer or customer activity related information as features. Potentially relevant information such as education of the customer, subscriber loan amount and frequency, and recharge frequency are not into consideration to predict airtime credit risk. In this research work, we propose an approach that takes in to consideration education and recharge frequency from customer profile and subscriber loan amount and loan frequency from the loan information to predict airtime credit risk. We conducted an experiment using 90,000 mobile subscriber's data. The experiment uses four machine learning algorithms: decision tree (DT), logistic regression (LR), random forest (RF) and (MLp). The results show the combination of the existing and new features improves airtime credit risk prediction for all algorithms. The highest improvement is 6.91%, in accuracy while using the existing customer profile and usage by LR. For F-measure the highest improvement is 7.8% by RF. We ranked the features based on their importance using feature ranking algorithms and shows that the newly added features have helped to enhance the airtime credit risk prediction.

Keywords: Airtime credit risk, Airtime credit risk prediction, Machine learning algorithms.

Introduction

Communication technology has become essential, and it has aided in the creation of a society in which anyone, no matter how far away, can be reached in a matter of seconds. Due to corporate requirements and technological developments, the number of mobile subscribers is steadily increasing in the telecommunications industry, which requires a high level of multi-service delivery.

One of the services included in this package is a mobile airtime credit. Airtime is increasingly becoming a basic commodity among the rapidly growing middle class in developing nation [1].

Short term airtime loans would be useful when subscribers run out of airtime and cannot top up their service number. In some situations, it may be difficult for a customer to purchase a recharge card. particularly late at night or while traveling. Short term airtime loans will help in reducing the inconvenience for the customer and loss of revenue for the telecom operators [2].

The airtime loan service is implemented following two approaches. In the first approach, mobile network operators (MNOs), provide clients with airtime loans and face the risk of non-performing loans. Safaricom, through its Okoa Jahazi service, is an example of a corporation that employs such an approach. When a subscriber runs out of airtime, they can borrow money equal to the amount they had topped up in the previous seven days, with the expectation of repaying it within five days. The second approach entails collaboration between the MNO and a third-party lender, in which the MNO offers access to clients and the mobile network while the risk is shifted to the third party. ComzAfrica is an example of a lender that relies on such collaboration. The third-party lender pays the MNO the credit amount before the customer repays it under this process. As a result, in the event of a default, the third party loses the entire sum [1].

The time period in which the loan has to be paid, however, varies based on the rules set by MNO. For example, some MNOs set the period to be five days, while others set it to be 72 hrs. If the loan is not settled the loan in the given time period, the subscribe who took the loan will be considered as a default. To early identify subscribers who default and minimize the loss incurred due to them, MNOs implement credit scoring algorithms [3].

In Ethiopian context airtime credit service is introduced in August 31,2018 by Hikma electronic PLC and Credok communication technology partners using a contractual agreement with Ethio telecom. Three parties are involved in the provisioning of airtime credit service in Ethiopia. These are the mobile service operator, a VAS (Value Added Service) Provider and the subscriber.
The airtime loan service is accessed using SMS (by sending one of the key words (A, L or C) to 810 or dialing *810#. In order to activate the service and give loan to the users Ethio telecom uses credit scoring method to check the eligibility of the user. With settle time of 180 days.
The criteria defined in the credit scoring method are the following: the user must have an active prepaid service number, must have been on the Ethio network for a minimum of 3 months and with a minimum top-up of 30-birr airtime per month.

Literature Review

Credit risk has been the subject of studies in financial institutions. Financial loans are based on a collateral agreement, a relatively large sum of money, and are not risk free. Hence, the loan needs to be assessed physically and financially before it is granted. Unlike airtime credit, which is technology-dependent, a micro loan, or a relatively small amount of money, is granted at the convenience of the subscriber and solely on the assumption that the subscriber will not fail.

For telecom companies, however, this approach does not work mainly because their airtime credit service has to be given in real time. Usually, customers ask for airtime credit when they want to make a phone call and realize that they do not have a balance. Hence, to minimize the airtime credit risk and give the service, a different risk management approach needs to be followed. In this regard, different researchers proposed different approaches. In this section, we discuss machine learning based approaches.

The role of data mining in predicting airtime credit risk was investigated and the experiment was carried out using WEKA, an open source data mining tool. To find the best performing model, a variety of classification techniques were used. J48 Decision Tree, Nave Bayes, and Logistic Regression. The algorithms were built and tested using telecom prepaid subscriber usage data, which included 86, 024 cases and eleven attributes. The models' performance was further assessed using a confusion matrix. The J48 decision tree model outperformed the other classifiers with a precision, recall and F-measure values of 98.6% and a ROC area threshold of 99.6%. The authors used 10-fold cross validation to avoid bias. The model created using logic regression has a 97.1717 percent accuracy. The accuracy of the Multilayer Perceptron and Nave Bayes classifiers was 96.7622 percent and 94.6355 percent, respectively. Some significant rules and parameters are derived from the selected classifier, which can support in the decision-making process for airtime credit [3].

The authors employ machine learning approaches to develop a credit score model for airtime loans and identify defaulters and non-defaulters in this research. Credit scoring techniques are reviewed, and this knowledge is used to construct a machine learning model that is suitable for airtime lending. Over three million loans belonging to 41 thousand customers with a three-month repayment period have been examined. Logistic Regression, Decision Trees and Random Forest are evaluated for their ability to classify defaulters and they use the cross-validation approaches to avoid the bias. They found that Random Forest was the best classifier, with an accuracy of 80%. Finally, they decide on a base rule: if the default rate is less than 2%, it is better to offer everyone a loan [7].

The purpose of this study is to investigate the trend and pattern of advanced airtime lending, recovery, and probability distribution in Nigeria. Because of its ability to handle the large volume of data generated by subscribers each day, the Audit Command Language (ACL) was chosen (over 10milllion rows). For their analysis, they used 40 million customers, which included both active and inactive subscribers. In this analysis, 5 million active subscribers were used as a sample population for airtime gross lent per lending period. The average daily loan amount was N118, and the maximum loan amount was N142. On a daily, the average amount borrowed by a customer was N128, while the company recovered a minimum of N78 and a maximum of N98. According to the findings of the analysis, the total amount recovered by telecommunications operators on a daily basis is N88. From the analytical result they suggested that more airtime data lending occurs than recovery [2].

Statement of the problem

One of the most critical risks that a financial organization must manage is credit risk. Because there is no profit without loan repayment, the issue of credit risk management affects all financial organizations that lend to individuals and legal companies [3].

The telecom sector offers a variety of value-added services (VAS) to customers and service providers. Airtime credit service is one of the VAS offered by MNO, allows prepaid mobile customers to top up their mobile airtime credit at any time and from any location [4]. This service, however, has its own challenges when it comes to loan repayment. Many subscribers usually do not pay back their loans in time or they default. For instance, the Ethio telecom airtime credit service report shows that there is birr 291 million bad debt on April/2020 (see Figure 1.1).

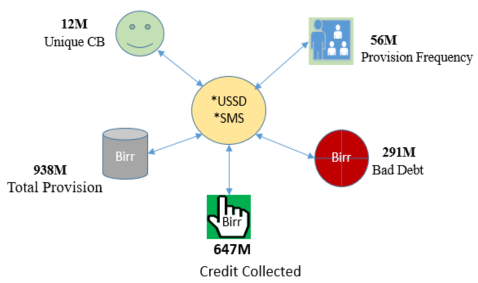


Figure 1: Ethio telecom airtime credit service report

To minimize the loss due to bad airtime debt, different researchers propose to use prediction algorithms to estimate the liability of the credit airtime service requester [5]. To perform the prediction, the algorithms use different sets of features computed from customer profile, loan information and usage. These features, however, capture only limited aspects of the service requester, and, hence, affect the performance of the prediction algorithms. In this research, we study the customer and customer activity data to identify features that better help to predict the liability of airtime loan service requesters.

Methodology

The proposed features are newly extracted features that are utilized to improve the model's prediction accuracy. The proposed features in this study are derived from the Ethio telecom database and are those that are used to understand the subscriber's loan information behavior. Customer profile, loan information, and usage are the feature categories.
The overview of the methodology is shown in Figure 4.1. Below we detail each step in the methodology.

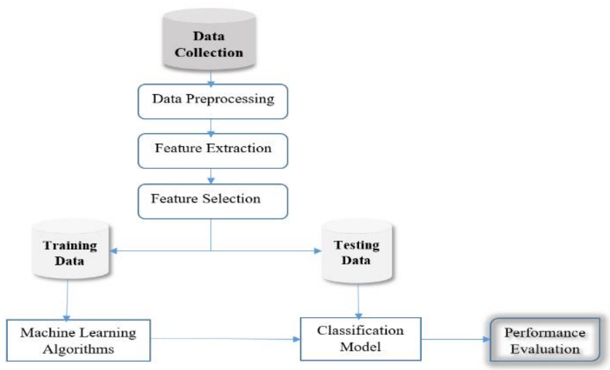


Figure 2: General Methodology of the Study

Data Collection

The data for the study is collected from telecom prepaid mobile service subscribers' business support system (BSS), which included the customer relationship management (CRM) system and the convergent billing system (CBS).

Table10: Preprocessed Data records

Collected Dataset		**Final Dataset**	
150,000 subscribers		90,000 subscribers	
Defaulter s(Bad)	Non Defaulters (Good)	Defaulters (Bad)	Non Defaulters (Good)
65,000	85,000	45,000	45,000

Data Preprocessing

Data preprocessing refers preparing the dataset to make it suitable for the next steps, i.e., feature extraction and training. knowledge discovery during the training phase is more challenging if there is a lot of irrelevant and redundant information. In this study, we performed data cleaning, integration and transformation.
Data cleaning: The process of preparing data for prediction by removing the incomplete data, inconsistence data and organizing the raw data in order to increase the performance of the models.
Data integration: Data preprocessing technique that involves combining data from various data sources.
Data transformation: The process of changing the format, structure, or values of the data. data was collected from different data source and these data source have their own data format. in order to come up the different data source in to the same format data transformation is one of the best data preprocessing technique.

Feature Extraction

At this stage, we extracted features that we think are helpful to predict airtime credit risk. The features extracted are classified into two groups: existing and new. The existing features are features that are used in previous state-of-the-art risk prediction studies [3,7], while the new features are those that we propose to include for the prediction. Table 2 summarizes the existing and new features extracted from the dataset collected in the previous step.
Table 2: Proposed and Existing Features of the Study

Existing Features	
Features	**Description**
Service Number	Unique identity of the Subscriber
Customer Age	The age of the customer
Network Age	Network age of the service number
Recharge Amount	The total amount of recharge per month
Voice usage log	The amount used for voice
Data usage log	The amount used for data
SMS usage log	The amount used for SMS
Proposed Features	
Education	Educational level of the subscriber
Recharge frequency	The number of frequencies to recharge the service number per month.
subscriber Loan amount	Total subscriber loan amount taken per month
Repaid amount	The amount of repaid loan per month
Loan frequency	Number of loans taken per month

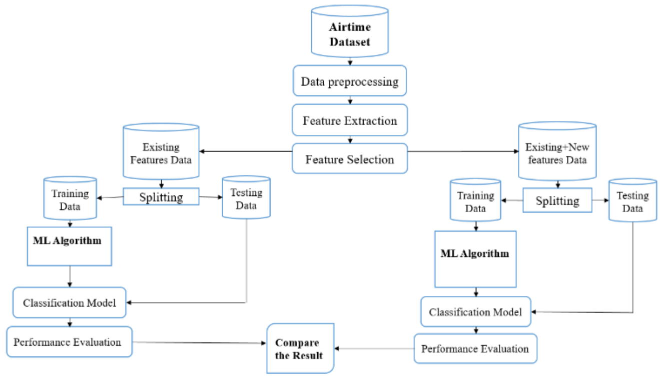


Figure 3:Experimental Design of the Study

Feature Selection

Feature selection is the process of identifying the most significant features in order to improve prediction accuracy. Some features could have less impact on the prediction performance probably because they provide duplicate information. To identify such features, we computed Pearson correlation coefficient. One of the features is selected from those features that correlated strongly. On this study, data splitting is used in machine learning to divide data into training and testing datasets. The models learned from the training dataset, and then we fed the testing data into them and evaluated their performance. Different data splitting strategies exist. In this study, we used the K-fold cross validation. In cross-validation the sampling is done so that no two test sets are the same. The available learning set is partitioned into k disjoint subsets of approximately equal size. The term "fold" refers to the number of subgroups that result. Especially for small dataset k-fold validation is more useful and avoids bias.

Experimental Setup

A.Dataset

This subsection discusses how the data is collected from different systems of the company and integration techniques used to integrate the data. In this study, we take
three-month average data on the airtime credit service. This research uses Ethio telecom prepaid mobile service subscriber’s data and the data is collected from business support system (BSS). The data was gathered from various secure databases; it was required to combine them in order to use the integrated dataset. It is necessary to have an attribute to link the collected data from many sources in order to integrate the data. To achieve this, the service number is utilized as a key to integrate the data from various sources. However, due to subscriber’s privacy, the service number was removed after the data was integrated. The integrated data is saved in a CVS file.

The below figure 3, indicates the general experimental design which is used on this study.

The original dataset contains 28 features. To identify and select from features that correlate, we computed the correlation coefficient between the features using Pearson correlation method. The features with a correlation value of more than 95% are strongly correlating features, and hence, considered to have the same information.

Features that strongly correlate usually increase the model's complexity and lower its performance, and hence, its recommended to select and keep only one of the features. In our dataset, loan and repayment amounts are found to strongly correlate. Then the repayment feature was discarded from the experiment.

B.Model building

On this study, four machine learning models are used. From the machine learning the classification algorithms: logistic regression, random forest and j48 decision tree, and from the deep learning algorithms MLP algorithms are selected for the experiment.

C. Evaluation metrics

Performance evaluation metrics are used to determine how effectively a machine learning model performed with the test data was given. A learning model's fundamental purpose is to generalize successfully on data that has never been seen before. On specific learning models, specific metrics must be utilized, and not all metrics may be employed in a single model. To compute the metrics, we use a confusion matrix as shown in table 3 below.

Table 3: Confusion matrix

Actual value

Predicted value

	Defaulter (Bad)	**Non-Defaulter (Good)**
Defaulter (Bad)	**TP**	**FP**
Non-Defaulter (Good)	**FN**	**TN**

In the confusion matrix:

- ✓ True Negative (TN) refers to non-defaulters who are predicted as non-defaulters.

Alg.	Dataset	Performance Evaluation Metrics			
		Accuracy	Precision	Recall	F-Me
RF.	informatio n features [7].	82.60%	0.863	0.775	0.817
MLP		77.52%	0.832	0.686	0.750
Existing and new features					
J48 DT.	[3,7] +New features	86.20%	0.868	0.854	0.861
LR.		80.02%	0.779	0.852	0.814
RF.		87.71%	0.895	0.854	0.874
MLP		80.44%	0.829	0.767	0.797

Algo.	Accuracy of Existing + New dataset	Accu.of Cust.profile& usage [3] dataset	Δ Accuracy
J48_DT	86.20%	81.71%	4.49%
LR	80.52%	73.61%	6.91%
RF	87.71%	82.18%	5.53%
MLP	80.44%	73.55%	6.89%

- ✓ True Positive(TP) refers to defaulters who are predicted as defaulters.
- ✓ False Positive (FP) refers to non-defaulters who are predicted as defaulters.
- ✓ False Negative (FN) refers to defaulters who are predicted as non-defaulters.

- **Accuracy**

Accuracy is calculated by dividing the correct predictions to the overall prediction value (see Equation 1).

$$Accuracy = \frac{TN+TP}{TP+FP+TN+FN} \text{ ----------------------------(1)}$$

- **Precision**

Precision is the ratio of correct prediction to the sum of true and false positive prediction (see Equation 2).

$$Precision=TP/(TP+FP) \text{ -------------------------------(2)}$$

- **Recall**

Recall is the ratio of correct prediction to the sum of true positive and false negative prediction (see Equation 3)

$$Recall=TP/(TP+FN) \text{ ----------------------------------(3)}$$

- **F-Measure**

The F-score is a way of combining the precision and recall of the model. It is calculated as the harmonic mean of precision and recall (see Equation 4).

F1-Score=2*((precision*recall)/(precision+recall)) -(4)

Results and Discussion

Table 4 shows accuracy, precision, recall and F-measure values of the two types of datasets containing only existing features and existing plus new features. The metrics are computed for the four selected algorithms.

Table 4: Performance evaluation result using [3,7] dataset and new +existing dataset

Accuracy

Table 5: Delta Accuracy of Existing +New dataset with of Customer profile& usage [3] dataset

Table 6: Delta Accuracy of Existing +New dataset with loan information [7] dataset

Algo.	Accuracy of Existing + New dataset	Accu. of loan information [7] dataset	Δ Accuracy
J48_DT	86.20%	83.53%	2.67%
LR	80.52%	78.07%	2.45%
RF	87.71%	82.60%	5.11%
MLP	80.44%	77.52%	2.92%

Tables 5 and 6 show the improvements achieved in accuracy while using the existing + new features to predict airtime credit risk. For all the algorithms, the combined feature set shows an improvement over the existing features (see Figure 4 Tables 5 and 6). For the existing features, a better performance in accuracy is observed while using customer profile and usage features when compared to loan features. When comparing the accuracy achieved while using customer profile and usage with the existing and new features, the highest improvement, i.e., 6.91%, is shown for LR. LR also gives the highest accuracy measure for the existing and new features.

F-measurement

Table 7: Delta F-measure of Existing +New with Customer profile and usage [3] dataset

Algo.	F_M of Existing + New data set	F_M Customer Profile and Usage [3] dataset	Δ F- Me
J48_DT	0.861	0.823	3.8%
LR	0.814	0.746	6.8%
RF	0.895	0.827	6.8%
MLP	0.797	0.729	6.8%

able 8: Delta F-measure Existing +New with loan information [7] dataset

Algo.	F_M of Existing +New dataset	F_M of Loan Information[7] dataset	Δ F- Me
J48_DT	0.861	0.823	3.8%
LR	0.814	0.799	1.5%
RF	0.895	0.817	7.8%
MLP	0.797	0.750	4.7%

Tables 7 and 8 show the improvements achieved in F-measure while using the existing + new features to predict airtime credit risk. For all the algorithms, the combined feature set shows an improvement over the existing features (see Figure 4, Tables 7 and 8). For the existing features, a better performance in F-measure is observed while using loan information features when compared to customer profile and usage features. When comparing the F-measure achieved while using loan information with the existing and new features, the highest improvement, i.e.7.8%, is shown for RF.

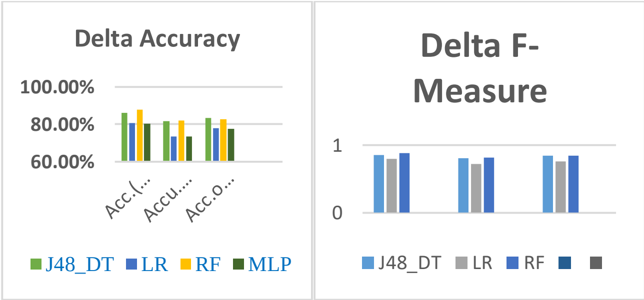
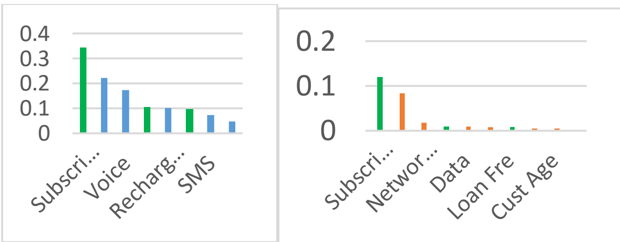


Figure 4: The overall accuracy and f-measure using all dataset

Feature Ranking

The experiments conducted in Section 4.6 show that the new + existing features give a better performance in predicting airtime credit risk. To see which features contributed to this improvement in performance, we ranked the features using two different feature ranking algorithms found in Weka: CorrelationAttributeEval and GainRatioAttributeEval. The results of the ranking algorithms are shown in Figure 3: In the figure, the green bars show the rank of the newly added features.



1: CorrelationAttributeEval 2: GainRatioAttributeEval

Figure 5: Rank of the features used on the study

Conclutsion

When prepaid mobile subscribers are unable to recharge their service number, they can use the airtime credit service. The credit service helps in increasing the satisfaction of service subscribers as well as the telecom operator's revenue. However, the service comes with its own set of risks, as there is no guarantee that prepaid customers will repay the amount they have taken. Some researchers have proposed prediction algorithm approach to address this problem. These approaches are better in terms of customer profile, usage, and loan information. However, they did not take into account the subscribers' educational background, subscriber loan amount, loan frequency, or recharge frequency.

In this thesis work, we propose an approach that takes into consideration education and recharge frequency from customer profile and subscriber loan amount and loan frequency from loan information aspects of the prediction. We conducted this thesis with a set of 90,000 subscriber data for the experiments using four machine learning algorithms: decision tree (DT), logistic regression (LR), random forest (RF), and MLP. To analyze the impact of the proposed approach the accuracy of the existing +new dataset was enhanced by 6.91 percent using the LR when we use customer profile and usage features.

As a result, when we used the loan information, the delta of F - measure of the existing + new dataset improved by 7.8% using RF. Finally, all models achieve satisfactory performance result and the RF have the highest prediction accuracy of 87.7%.

References

[1] B. Dushimimana, Y. Wambui, T. Lubega, and P. E. Mcsharry, “Use of Machine Learning Techniques to Create a Credit Score Model for Airtime Loans,” no. Mode 2017, 2020, doi: 10.3390/jrfm13080180.

[2] U. Journal and O. F. Business, “Unilag journal of business vol. 6 no. 1 2020,” vol. 6, no. 1, pp. 96– 113, 2020.

[3] O. Tarekegn, “Application of Data Mining Technique for Predicting Airtime Credit Risk : The Case of Ethio Telecom,” 2019. Jun-2019

[4] Z. Ereiz, “Predicting Default Loans Using Machine Learning (OptiML),” no. November 2019, pp. 3–7, 2020, doi: 10.1109/TELFOR48224.2019.8971110.

[5] I. J. M. Education, C. Science, D. K. Gupta, and S. Goyal, “Credit Risk Prediction Using Artificial Neural Network Algorithm,” no. May, pp. 9–16, 2018, doi: 10.5815/ijmecs.2018.05.02.

[6] T. O. Ayodele, “Types of Machine Learning Algorithms.” TaiwoOladipupo Ayodele University of Portsmouth ,United Kingdom

[7] B. Dushimimana, Y. Wambui, T. Lubega, and P. E. Mcsharry, “Use of Machine Learning Techniques to Create a Credit Score Model for Airtime Loans,” no. Mode 2017, 2020, doi: 10.3390/jrfm13080180.