Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 8, Issue 3, Pages 85-99, July-September 2024

Loan-eligibility prediction for Airtime Credit Service Subscribers using Machine Learning: The case Ethio Telecom

Megersa Jambo Degefa[1*] & Lemlem Kassa (PhD.)[2]

[1,2]Department of Software Engineering, School of Post Graduates' Studies, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia. Corresponding Author (Megersa Jambo Degefa) Email: megersa.jambo111@gmail.com*

DOI: https://doi.org/10.46759/IIJSR.2024.8309

Article Received: 20 July 2024 Article Accepted: 25 September 2024 Article Published: 29 September 2024

ABSTRACT

Ethio telecom is largest telecommunication companies in Ethiopia. Ethio telecom begun airtime credit service in 2010 to its above 70 million mobile subscribers. The service aims to maintain customer relationships, minimize churn, and generate additional income. However, collecting service money from subscribers is a challenge, resulting in losses of around 2.7 million birrs. To address this, a machine learning-based technology solution was developed to predict eligible customers based on historical data usage. The model uses an Ethio telecom dataset with 114871 rows and 18 columns. Five machine learning classification algorithms were selected for the proposed solution based on the research behavior and literature review: Random Forest, Logistic Regression, Gradient Boosting machines, Naïve Bayes, and Support vector machines. These algorithms are implemented on the prepared dataset and evaluated using model evaluation metrics like accuracy, precision, recall, F1-score are applied. In addition, the confusion matrix table and receiver operating characteristics-area under curve are used to evaluate performance. The accuracy of Random Forest was 90.40%, Logistic Regression was 75%, Gradient boosting classifier was 90.42%, Naïve Bayes was 89.8%, and the Support vector machines was 73.1%. We performed comparative analysis between the models to select the robust model. So that, the Gradient Boosting classifier model provide an outstanding result in predicting eligibility for the airtime credit service.

Keywords: Credit score; Machine learning; Airtime credit; Ethio telecom; Digital lending; Eligibility; Credit risk; Mobile subscribers; FICO.

1. Introduction

In Ethiopia, telecommunication services like all-mobile services have been provided by the Ethio telecom organization. This organization were started provide service since 1894 by Emperor II. After that the organization did many improvements to extend the services they provide to the customers. Building telecommunication infrastructure, extending services availability between regions, and other relevant jobs that enable the telecom for the users have been performed in the organization. In 2010, the Ethiopian government absolute to consider on the development of telecommunication industries, considering them as a key target in the development of country's economy. November 29, 2010, Ethio telecom was begun, from this ambition of supporting the solid evolution of our country, within the growth transformation plan (GTP), with ambitious objectives for 2015 [1],[2].

From the Ethio telecom services mobile communication is one of the most popular and significant services with over 70 million users utilizing mobile networks [1]. Furthermore, the advancement of technology in the communication industry plays significant role to the general growth of all fields about social, political, and economic matters. The telecom business grew as a result of technological improvements; as a result, the number of subscribers increased as well as their demand increased, which in turn encouraged the interest acquiring airtime service providers to stay connected on the call opportunity with customers [3]. The organization uses several approaches to keep the customers always connected to the company's services. The major approach performed by the Ethio telecom, increase operational excellence, implementing new systems and building infrastructure, increase service availability, expanding revenue sources, and attracting service quality and price to meet expectations, are used to achieve its planning.

Uncollateralized and digital lending is a new telecom services sector that has emerged in Ethiopia in recent years to rise the advantage of using credit for the country's underprivileged citizens. Fintech is a technology platform that can be used for digital lending [18]. This type of services remotely managed, automated, and instantaneous loan appraisal and distribution characteristics set digital credit apart. Digital lending operations depend on different activities of the customers like usage history, credit score system, and dependable features of the customers [1],[2].

Ethiopia started a digital lending that called airtime credit system in 2018 to provide airtime credit service, data, and other services to mobile network subscribers. This approach provides different lending options for the subscriber based on the usage history of the customers [1],[4]. However, the amount of money borrows using these services limited for the mobile service like for call, data, and SMS. The mobile subscribers can borrow up to 100 Birr from the mobile operators of the Ethio telecom. The mobile operators ensure the customers' requests for airtime credit service, and identifies the customers based on the usage history and frequency of prepaid amount on the mobile subscribed identifications. The amount of credit is calculated by the user's activity on the Ethio telecom networks, and the repayment of the service on the given time and but charged a blanket 10% interest fee for all options of digital credits. Currently, telecommunication operators design strategies that enable the company to overcome the challenges and mitigate the risks that come with embedded technologies. The operators record the customer data such as customers' name, customers address, contractual details, usage details, service details of the customers. So, by using cutting-edge technology like machine learning, we can clearly identify the eligible customers for the airtime credit service. Machine learning is a major type of artificial intelligence that learns from previous activity of the customers and develop predicting model that able to predict for eligible customers. In addition, it can minimize the effort required and time wasted to memorize the previous activity of the customers.

This research paper proposes to predict customers' loan eligibility for Ethio telecom airtime credit service using machine learning algorithms. This model can identify the eligible and non-eligible customers based on their usage history dataset collected from the Ethio telecom corporate mobile airtime credit service operators.

1.1. Statement of the problem

The subscribers' credit usage history, and availability of the subscribers on the network of Ethio telecom are the major elements used to identify the customers for the airtime credit services. Also, the subscribers must use prepaid accounts to get the airtime credit service offered by the organization. This type of activities enables subscribers to borrow airtime credit whenever they need without disturbing about running out of balance in their account. Many time the subscribers late at night and early in the morning are unable to recharge their subscriber identity module (SIM) to stay connected with Ethio Telecom network services.

Despite the advantages of the airtime credit, there are a lot of risks and challenges occur after the service was provided to the end users. The main issues were collecting the repayment loan they provide to the subscriber and identify the eligible customers for the services. For example, Ethio Telecom provides digital lending airtime service for all mobile subscribers which satisfies the requirements specified by the service providers. This service solves the problems and limitations of disconnected when running out of balance and unable to purchase the credits from where they are located. According to Ethio Telecom reports [5], millions of subscribers (4.5 million) are unable to repay loans they get from the service providers. This degrades the trust between customers and providers and it increases the worrying on the providers as the debt on the customers is not returned on time is increasing [6].

So, properly consider and understanding the issues is needed to overcome the risks comes with the airtime credit service. However, if the problem is not properly mitigated the organization faces a revenue loss, degrades the service quality and customer relation with organization which is the reverse of the airtime credit service missions. To solve these challenges, many approaches are implemented in the area of Ethio Telecom services and another related organization. The researcher [5], developed data mining application solution to identify the airtime credit risk for the Ethio Telecom Corporation. They use the mobile subscriber's dataset and extract the hidden patterns of the subscriber's that identify the credit risks in the corporation by using data analysis algorithms. J48 Decision Tree, Naïve Bayes, Multilayer Perceptron, and Logistic Regression data analysis algorithms are used to analysis the risks level of airtime credit service users. These algorithms implemented using dataset contains 86,024 instances and eleven (11) attributes were used for developing and testing the algorithms. J48 decision trees were performs better accuracy for identifying the airtime credit risks. In addition, the researcher [4], designed credit score model for airtimes loans using machine learning algorithms. Logistic Regression, Decision tree, and Random Forest were used to develop the model and evaluated based on the dataset collected from the ComzAfrica company. However, these researches have the limitation in the way they approach to solve the problem. Dataset used to evaluate the algorithms are specifically based on the identified company and the number of selected features to develop the model is small in numbers. This causes for the low performances of the algorithms, limits effectiveness of the model to identify the customers activities and eligibility for the airtime credit services [7].

So, to fill the gaps we propose a solution by using cutting-edge technology that supports the service provider to know their customer's activities. The machine learning-based model is designed to properly identify the defaulter's customers from the non-defaulters by developing the model that trained on the characteristics and activity of the customers on the mobile network services.

1.2. Study Objectives

The general and specific objectives of the research are described below.

General Objective of the study: The general objective of this study is to build machine learning based model that can predict the loan-eligibility of customers for airtime credit services.

Specific Objectives of the study: The specific objectives of the research of the study listed in below: (1) Conduct inclusive literature review to analysis the research gaps; (2) Gathering important data and information that used to provide airtime credit services (ACS) and select relevant features from the gathered data; (3) Conduct complete preprocesses on the gathered data from the Ethio Telecom corporations; (4) To design classification model based on machine learning algorithms in order to develop a predictive model that skilled to identify loan eligibility and ineligibility for Ethio Telecom airtime credit services (ACS) consumers based on the usage history and consistently involved in the networks services of mobile network operators (MNO); (5) To training and testing the designed models; (6) To evaluate the performance of the proposed machine learning model using basic measure metrics such as F1-score, Precision, and Accuracy methods; and (7) To prepare finding report on the results and recommends the best machine learning algorithms to predict the eligible customers, and identify the important features during the research processing that mostly correlated with other features.

2. Literature Review

2.1. About Credit score

Credit score is a collection of users' activities from specified networks or companies for different purposes. By using this credit score, the organization can provide several services, such as financial support, gifts, and another that improve the customer satisfaction within the organization services [6],[8]. Organization like banks and financial organizations can use credit scores values and implement in their business closer to their customers. Credit score become an alternative solution for financial institutions to clarify the ability of customers to get financial services since 1989 [8],[9]. A credit score of persons represented by the worthiness they have in the organization that can be calculated and convert into numerical values represented between 300-850. Lenders, financial institutions, and other organizations use it as a standard measurement to determine the risk involved in giving someone credit or a loan. A person's credit history which includes things like payment patterns, credit utilization, account types, duration of credit history, credit scoring approach, and recent credit inquiries is usually used to compute credit scores [10]. In the US, a credit score normally falls between 300 and 850; higher scores are indicative of lesser credit risk and vice versa [9]. To prepare these ratings, the providers frequently apply credit scoring approach developed by the organization like financial organizations, telecommunication industries.

2.2. Traditional and Modern credit scoring approach

As per [9],[11], Fannie Mae and Freddie Mae established the credit score role in 1989. They released their first credit score, called the Fair Isaac Cooperation (FICO) score, and it is applied in the United States to calculate the individual's credits to determine the eligibility of consumers for different purposes such as financial loans, mortgages, evaluating businesses, and decision-making in the organization [11]. Traditional credit score algorithms, like the Fair Isaac Cooperation (FICO) score, based on the listed criteria and weightings to assess creditworthiness based on usage history data like payment history, credit usage, length of credit history, categories of credit, and recent credit inquiries [9],[12].

The FICO score is prepared for largest financial industries and makes them trusted by lenders for different purposes, such as decision-making, reducing loan defaulters, and fair distribution between consumers. However, this score model may not capture all the complex relationships and patterns of individual data, potentially leading to a less accurate risk assessment, especially for individuals with limited credit histories or non-traditional financial behavior [8].

The traditional and rule-based approaches have been undertaken a significant transformation due to the recent trend of technological advancement. The emerging technologies that improve the weakness of the traditional approach like machine learning approach which uses a robust statistical algorithm to assess huge volumes of data and identify spot patterns are become popular in the development of the credit score [13]. These algorithms may be able to provide a better risk assessments for the service offers; that more closure to reality since they consider a wider factor and capture complex correlations among variables. Because machine learning algorithms are flexible and can learn from new data, they may ultimately perform better. Many researchers explore machine learning applications that able to predict credit scoring risks using different organizations' datasets [14].

2.3. Airtime Credit service

Airtime credit services can be called mobile airtime credit or mobile airtime lending service, began emerging in various regions around the world during the mid-2000s [15]. However, the specific launch dates may vary depending on the country and the service provider. Ethio Telecom Corporation is one of the leading telecom industries in East Africa, and it launched the airtime credit service in 2018. Mobile network operator's subscribers are increasing every year, according to reports from telecoms [16]. MNO stores an individual's usage history like personal information, customer loan information. Using this MNO's subscriber data, the Ethio telecom offer mobile service delivery that enables customers to stay connected within the Ethiopian telecom networks consistently. Airtime credit service streams through a diverse process to offer a service to the customers. The service offered by the telecom is managed by the mobile network operators that identifies the detailed usage history of the subscribers.

2.4. Machine Learning Based Techniques

The trend of leveraging machine learning benefits in different activities across industries have been increasing from time to time. Especially for the jobs of processing huge amounts of data for decision-making in companies, machine learning plays a significant role in multiple ways. With the help of machine learning (ML), the telecom sector can have a better sympathetic of competitive tactics, business markets, subscriber management and monitoring, strategic planning for ongoing development, and decision-making techniques [13],[17]. Recently, there has been a lot of needs in the field of research on creating machine-learning algorithms for loan eligibility prediction. By utilizing data-driven models to evaluate applicants' creditworthiness, these strategies seek to increase the precision and effectiveness of loan approval procedures [17]–[18]. The researcher [17] found that logistic regression is a more superior choice in machine learning technique to predict whether loan eligibility is approved or denied using customer creditworthiness collected from different financial areas. Machine learning based solution provides many opportunities for the organizations to process their daily works, meet their customer needs, and improve their work efficiency [19]. Organizations like Ethio telecom deal with huge amount of data processing in their daily activity. Extracting an important insight from those huge datasets is important for the organization to design the strategic planning and evaluate their customer services [20]. Performing this type of task in manual way is time-consuming and tedious for the organizations. There are three categories of machine learning are commonly known; Supervised, Unsupervised, and Re-enforcement machine learning method. In the study supervised machine learning techniques are used to develop the predicting model for the identified problems.

2.5. Related Work of the Research

In this study, several research reviews have been conducted and reviewed to identify works performed around the airtime credit service and financial credit score. The main objective of this research review is to recognize the research areas and gather the required insights from various researchers' findings for the airtime credit service user's identification and prediction by developing model using machine learning model. In order to enhance and integrate technology into the telecommunication services, we reviewed papers that mostly focused on the telecommunications business. Ethio telecom is one of the largest telecommunication industries. The industries process a huge amount of data to operate their daily services. So, they become attractive area for researches and it need comprehensive solution using emerging technology like machine learning. In this study, Ethio telecom airtime credit service is the center of the study. Airtime Credit score is one of the research topics that requires an integrated solution to properly identify and manage the service users.

According to the paper [14], "Mobile Airtime Advances for the Financially Excluded" studies the impact of the mobile based airtime services in the financial sector for the excluded customers from the financial services. This paper describes the advance of the mobile in the financial sectors. For those without access to typical financial services like credit cards or bank accounts. This service is very crucial in order to make calls, send messages, or access internet services. Customers can borrow unsure amounts of airtime credit from their mobile network operators. This paper identifies additional financial services and products for the individuals based on the usage score of the customers. The study stresses mobile airtime advancements' potential as a financial service for those who are financially excluded overall. In order to fully comprehend its implications, difficulties, and prospects for advancing financial inclusion, it demands more investigation and study in this field.

The paper [12] study on "Mobile Airtime Credit and Microfinance: this paper identify opportunities and Challenges between mobile airtime credit services and microfinance initiatives. The paper describes the overall concepts of mobile airtime credit services and microfinance. And highlights the potential interactions between these two parts, both of which are interested to provide financial services to underserved populations, particularly in developing countries. This authorizes individuals to meet their financial needs, build credit histories, and improve their economic well-being. Despite the opportunities, the paper also addresses challenges and considerations associated with mobile airtime credit and microfinance integration. These may include regulatory hurdles, technological barriers, affordability concerns, and the risk of over-indebtedness among users.

In the dissertation [5] "Application of Data Mining Techniques for Predicting Airtime Credit Risk: The Case of Ethio Telecom," Tarekegn study on the application of data mining techniques to predict airtime credit risk, focusing on the context of Ethio Telecom. The primary objective of the papers was to develop a predictive model that assess the likelihood of credit default among airtime users. The papers explore the feasibility and effectiveness of using data mining techniques to predict airtime credit risk in the Ethio Telecom. The researcher analyzes historical transactional data from the airtime credit system of Ethio Telecom using data mining techniques such as association rule mining, clustering techniques, and classification algorithms. Using a variety of data mining techniques, including association rule mining, logistic regression, and decision trees, Tarekegn searching patterns and connections in the airtime credit service data. The results of the study for airtime credit risk are included along with performance indicators including area under the curve (AUC), accuracy, precision, and recall.

According to the paper [21], and explores the development of the application that based on machine learning approach for predicting loan eligibility. The paper performs comparative approach between different machine learning such as logistic regression, Decision tree, and random forest, and identify the effectiveness of those machine learning algorithms to predict the loan eligibility. The authors compare the performance of the machine learning models and select the most accurate and reliable approach for assessing loan applicants' eligibility. And the models implemented on the preprocessed dataset prepared by the researcher. The author most likely employed a set of evaluation criteria, such as accuracy, precision, recall, F1 score, and area under the curve (AUC) of the receiver operating characteristic (ROC) curve, to measure the performance of each machine learning model. The comparison analysis's findings are presented in the report, the random forest model was better performing algorithms which score 81.1% accuracy and selected for loan eligibility prediction model.

3. Research Methodology

In this study a various research design frameworks and models are identified to develop a solution that capable to determine consumers based on their credit scoring techniques. The Fair Isaac Corporations (FICO) Score is one of the earliest models developed to support lenders in determining whether their consumers are able to repay the loan and mitigate the risk associated with defaulting consumers [9]. This model is the origins for all other credit scoring based frameworks. Five factors are determined in the FICO frameworks. These are payment history, credit utilization, credit history, credit use, and new credit.

These factors have their own defined weightiness in manipulative the credit score of consumers. Based on the concepts behind the FICO model, this study aims to develop a model that can predict a customer's eligibility for the airtime credit services in the Ethio telecom company, by using customers past usage history, and consistently connected with the Ethiopian mobile telecom.

This developed model could be useful for the organization to mitigate and assessment risk occur by the credit-based service Numerous data credit score domains are appropriate for FICO. Depending on the scoring model, FICO scores can range from 300 to 850 or 250 to 900 [9],[11].

This study based on the experimental and comparative approach in terms of research methodology. In this research we focus on the cross-industry standards for machine learning methodologies are used for the proposed machine learning algorithms, and includes the other methodologies from the reviewed papers. The best and robust algorithms is selected based on the experiments we performed in the research. In addition, certain characteristics that are adjusted to create the model are considered in this research. To identify the best loan eligibility prediction model, this study work designed into several phases to provide solution for the proposed research.

3.1. Business Understanding

In order to provide a proper solution clearly understanding the business area of the research is significant. The trend of leveraging technology in telecommunication plays a significant role in different dimensions to improve their business strategies and analysis of their customer satisfactions. Clearly identifying the machine learning project’s feasibility, performance measurable criteria, and clearly identifying the requirements and data in the telecommunication industry provides a clear vision and common understanding in the development of machine learning based solutions.

Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 8, Issue 3, Pages 85-99, July-September 2024

3.2. Data Collection

The dataset used in this study was gathered from Ethio telecom. The Ethio telecom's business support system collects the company's data raw and provide dataset required for this research. The business support system including several components such as customer relationship, convergent billing system, and other components, that are used to support Ethio telecom airtime credit service provider to understand their customers personal information, customer knowledge and experience on the Ethio telecom networks, and describes value added service.

To develop the proposed machine learning model, we feed secondary datatypes to the model that collected from the Ethio telecom databases. These data classified into three different datasets. These are customer personal information, customer loan information, and customer usage history in the Ethio telecom mobile network.

3.3. Data Preprocessing

In the research methodologies, data preprocessing is the critical process and it has a direct impact on the model if data preprocessing is not properly performed. The dataset contains customer individual information, loan information, and customer usage history on the telecom networks. So, properly merging those data types, and preprocess them as one dataset is important for the development of the models. The collected dataset is very huge, contains unstructured data, missing values, uncleaned data, and mixing data values, and unnecessary data was included. So, we use data preprocessing techniques such as screening methods, wrapper methods, or embedding methods to get an important and required features for the developing model. also, we select dataset by reducing samples that don't meet machine learning standards for data quality.

In machine learning data preprocessing several techniques are suggested to identify the required data feature. Among these, we used a technique like data selection, data cleaning, and feature selection techniques.

3.4. Data Selection

This study uses Ethio Telecom prepaid airtime credit service subscriber's dataset. The dataset used in this study contains 114871 rows and 18 features. The feature of the dataset has different types of data types, 8 number of features are classified as categorical data types and the remaining features are classified as numerical class. To select the proper data for the proposed machine, we performed analysis on the available features and it is impacts on the model, use correlation coefficient matrix to identify the relation between the dataset features, and reducing the null values, missing values during the data preprocessing.

Table 1. Data selected feature with their data types

Features selected	**Their data types (class)**
service_number	Numeric
cust_type_name	Categorical
net_busi_type_name	Categorical

status_name	Categorical
cust_age	Numeric
gender_name	Categorical
per_type	Categorical
init_loan_amt	Numeric
loan_amt	Numeric
repay_amt	Numeric
loan_grade	Numeric
recharge_category	Categorical
recharge_amnt	Numeric
sms_local_usage	Numeric
sms_local_fee_etb	Numeric
data_usage_mb	Numeric
data_revenue_etb	Numeric
check_eligibility	Categorical

3.5. Data Cleansing

In the development of machine learning based solution cleansing and preparing data to feed into the model is significant steps. In the selected dataset from the Ethio telecom many missing, duplicate, irrelevant, and null value dataset are available. To fix these issues and improve the quality of the dataset, we performed data cleansing by removing unwanted data, reducing duplicated data, and handling missing data by using methods like imputation, deletion, or replacing the missing data values.

3.6. Feature Engineering

In the development predictive model feature engineering technique is significant in order to identify the measurable feature to provide a robust solution in the machine learning algorithms. After the data selection and data cleansing, feature engineering techniques are crucial to select the data feature that had impact on the performance of the machine learning model. Feature engineering includes various techniques like feature scaling, feature selection, normalization, and generalization techniques. If the data feature of the model can't be selected properly, it affects the performance, quality, accuracy of the model.

Feature scaling: In machine learning the impact of each feature compared based on the scaling standards used in the range between two numbers that represented between [0,1], or [-1,1]. This technique minimizes the variance occurred by the models and make standards the impacts of feature by scaling between 0 and 1 or -1 and 1.

We are using StandardScalar() and OneHotEncoder () techniques to identify the impacts of each feature for the models developed.

Feature selection: Selecting a relevant feature when developing machine learning model is the significant step. It increases the model's quality and minimizes the bias occurred in the model prediction. Feature selection aims to improve interpretability, decrease resource consuming and complexity, and balance the model to become overfitting or underfitting [22]. In the model development to select the dependency and relation between the we used Pearson correlation coefficient (r) techniques that measure the negative which in numeric values -1 and positive relation which relation numeric values represented in 1.

3.7. Selecting Machine learning algorithms

This study proposes machine learning based solution, which provides a strong solution that capable to predicts the loan eligibility of the customers for the airtime credit service subscribers. This study focuses on the binary classification algorithms that are used to develop models that are capable of predicting the eligibility of the customers for the airtime credit service in Ethio telecom. The most popular binary classification algorithms selected for the identified problem are logistics regression, random forest regression, support vector method, gradient boosting classifier, and Naive Bayes algorithms based on the reviewed literature.

3.8. Model Evaluation metrics

In the classification machine learning algorithms several evaluation metrics are applied to evaluate the model's performance in predicting class labels properly. The most popular of the evaluation metrics we are using in this study for the classification algorithms are accuracy, precision, recall, and F1-score. In addition, the Confusion matrix and Receiver operation characteristics-area under the curve (ROC_AUC) performance evaluation metrics are used to measure the performance of each model.

3.9. Apply comparative method between the models

Based on the model evaluation metrics such as accuracy, precision, recall, and F1-score, we identify the model that perform better to identify the loan-eligibility for the airtime credit service. In this study, we implemented five machine learning algorithms for the identified problem in the research and apply comparative analysis between each model.

4. Result

To answer the research questions of the study, the required dataset is collected from the Ethio telecom airtime credit subscribers, and the dataset contain 114871 rows and 18 columns after preprocessing the collected dataset. The raw data collected from the Ethio telecom database are unstructured and contains many missing values. We applied dataset preprocessing to make clean the dataset using machine learning techniques like fill the missing data using mean and media and feature engineering to identify the better features. The dataset contains seven categorical datatypes values and the other remaining columns are numerical data types values. In order to make the dataset understandable by machine learning algorithms and scale between the identified features, the data transformation techniques such as StandardScalar() and OneHotScalar() functions are performed on the prepared dataset. The

Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 8, Issue 3, Pages 85-99, July-September 2024

dataset target contains an imbalanced value. The target class of the dataset contains 82748 eligible values and 32123 non-eligible values. So, this target value variance causes for the data imbalance between the eligible and non-eligible class. The imbalance dataset causes for the underfitting or overfitting models and impacts the performance of the models that capturing noise data rather than true patterns of the learning data. The SMOTE oversampling techniques are applied to balance between the target class is become 66166 for both eligible and non-eligible values, and the total dataset is 132,332. To identify the relation between the selected feature we applied Pearson Correlation analysis. According to the correlation analysis report Figure 1, there is 53% correlation between Loan amount and initial loan amount, and 51% correlation between repay amount and initial amount. So, based on this value the correlation between the variable can be defined as good correlation between variables.

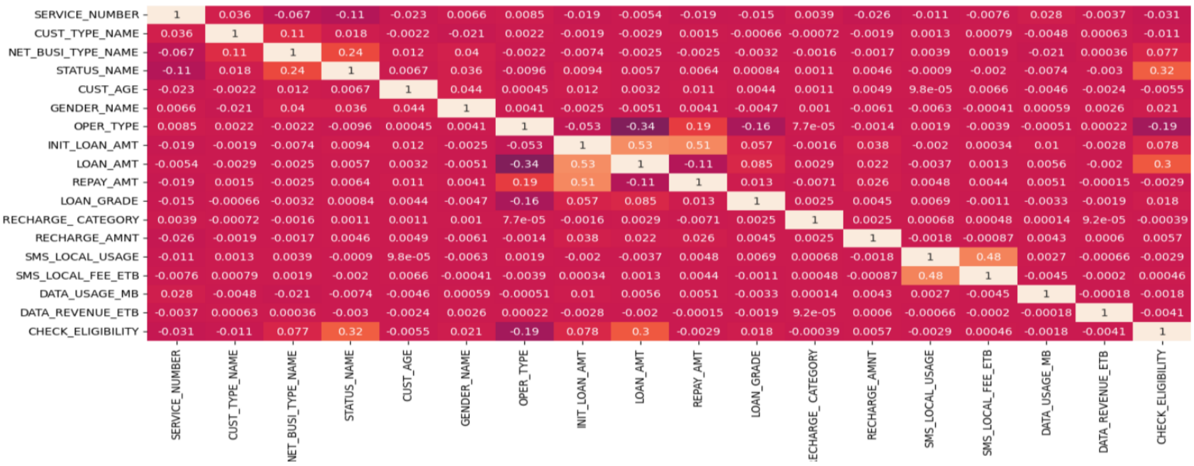

Figure 1. Correlation values of each feature

5. Conclusion and Recommendation

In the Ethio telecom, airtime credit service is popular service that enables the customers to borrow an airtime credit while their prepaid account balance is low. The airtime credit service must be managed properly to identify the customers repay back the loan they borrow from the Ethio telecom company and minimizes the risks occur because of the airtime credit service. Unless the organization can lose money and impacted by financial, and unable to compete in the business market. To minimize the impacts and identify the customers from the defaulters, supporting the airtime credit service by the cutting-edge technology like machine learning algorithms are significant for the organization. Such type of solution primarily based on the data usage and past activity of the customers subscribes the airtime credit service. The Ethio telecom mobile network has the ability to record the customers detail information and their activities on the organizations network. By using this customers data usage and activity, this study provide solution that based on the machine learning algorithms for classifying the eligible customers from the defaulters. The dataset collected from the Ethio telecom organization. However, getting proper dataset for the machine algorithm are not easy task. The collected data contain missing values and unnecessary data that can't be used for develop predicting model using machine learning. The collected data was must be preprocessed and cleaned by using sklearn data analysis tools. The data preprocessing includes process such as cleaning data, handle missing values, feature scaling and selection, and encoding data for the trained machine learning model. The raw dataset contains many unwanted features for the developed model, it is excluded from the preprocessed data. So, data preprocessing is the basic for developing machine learning and converting to understandable dataset input for the machine learning algorithms.

The data imbalance issues have been occurred in the target class values. So, this problem causes for developed model to be overfitting, improper model performance, and inaccurate prediction values. SMOTE random over sampling techniques are used in the research paper to make the dataset balanced and fair class distribution for training model. The balanced dataset was containing 132332 values and 66166 for eligible customers and 66166 for ineligible customers.

Feature selection applied on preprocessed dataset by using the Pearson Correlation analysis. In this study, five machine learning algorithms was selected for predicting the eligibility of the airtime credit users. And the developed model has been evaluated by using confusion matrix and other machine learning classification measurement metrics like accuracy, Precision, Recall, and F1-score.

By training the selected model on the prepared Ethio telecom prepaid subscribers, their performance accuracy is identified, the random forest algorithm was scored 90.4%, the gradient boosting classifier algorithm 90.42%, the Naïve Bayes algorithm was scored 89.8%, the logistic regression was scored 75%, and the support vector machine was scored 73.1%. based on this performance accuracy, the random forest and gradient boosting classifier perform better prediction result relatively while comparing with other machine learning model.

6. Recommendations

The major aim of this study was to design and develop predictive model to classify the eligible and ineligible airtime credit service subscribers. Overall, by analyzing the developed model results, the study has the following recommendation:

❖ Five classification algorithms such as Random Forest, Gradient boosting classifier, Naive Bayes, logistic regression, and Support Vector Machine are implemented. from this algorithm, we recommend that the Gradient Boosting model is perform better to build the predicting eligible customers from the defaulters.

❖ The company can take the findings of this research as input for extra insights and analysis for the business planning and reducing the risk of defaulters in airtime credit service.

❖ Since the number of mobile subscribers increasing from time-to-time for Ethio telecom services, managing their customer data and building their raw data keeping warehouse and properly recording the data can help the company to develop the effective and robust machine model that can simplifies and improves their customer daily service and improve their relationship between customers. Unless preparing effective data can be take long time and impacts on the performance of the models.

❖ The mobile service subscribers' behaviors and their characteristics on the network are dynamically changing. This increases the complexity to identify the eligible customers easily. So, the company must record the customers daily activity on the company's network properly. And it is important keeping the up-to-date records about the customers to properly identify the eligible data.

7. Future Works

In this study, Ethio telecom prepaid customer dataset was explored and analyzed by using machine learning model that capable to predicting customer eligibility for airtime credit service. Today, the monolithic telecom services provider was ended after Safaricom telecom joined telecom industry in Ethiopia.

- By adding Safaricom airtime credit service subscribers' data in the research for further exploration can be a future task.
- Improve the performance of the model and building a robust model that minimizes the risk occur because of the airtime credit service.
- Training and building model on the dataset by adding other features that predicting the eligibility of customers.
- Building a predictive model by using deep learning algorithms using customers airtime credit service and further analysis based on the algorithms.

Declarations

Source of Funding

This research did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests Statement

The authors declare no competing financial, professional, or personal interests.

Consent for Publication

The authors declare that they consented to the publication of this research work.

References

[1] Ethio Telecom, Addis Ababa (2022). Ethio telecom, annual-business-performance.

[2] Lin, L., & Huang, N. (2009). Credit Risk Assessment Using BP Neural Network with Dempster-Shafer Theory. International Conference on Artificial Intelligence and Computational Intelligence, Pages 397–401.

[3] Mohammadi, S., & Pirmohamadiani, R. (2015). Behavioral scoring model using Data mining approach and analytic hierarchy process. Journal of Soft Computing and Information Technology, 4(3): 52–65.

[4] Dushimimana, B., Wambui, Y., Lubega, T., & Mc Sharry, P.E. (2020). Use of Machine Learning Techniques to Create a Credit Score Model for Airtime Loans. Journal of Risk and Financial Management, 13(8): 180. https://doi.org/10.3390/jrfm13080180.

[5] St. Mary's University (2019). The Grants Register 2020. Pages 771–772. https://doi.org/10.1057/978-1-349-95943-3_816.

[6] Shafei, I., & Tabaa, H. (2016). Factors affecting customer loyalty for mobile telecommunication industry. Euro Med Journal of Business, 11(3): 347–361. https://doi.org/10.1108/emjb-07-2015-0034.

[7] Hoyle, B., Rau, M.M., Zitlau, R., Seitz, S., & Weller, J. (2015). Feature importance for machine learning redshifts applied to SDSS galaxies. Monthly Notices of the Royal Astronomical Society, 449(2): 1275–1283. https://doi.org/10.1093/mnras/stv373.

[8] McCorkell, P.L. (2002). The Impact of Credit Scoring and Automated Underwriting on Credit Availability. The Impact of Public Policy on Consumer Credit, Pages 209–227. https://doi.org/10.1007/978-1-4615-1415-2_8.

[9] Moldovan, D. (2023). Algorithmic Decision Making Methods for Fair Credit Scoring. IEEE Access, 11: 59729–59743. https://doi.org/10.1109/access.2023.3286018.

[10] Amat, O., Manini, R., & Antón Renart, M. (2017). Credit concession through credit scoring: Analysis and application proposal. Intangible Capital, 13(1): 51. https://doi.org/10.3926/ic.903.

[11] Fico Score (2015). The SAGE Encyclopedia of Economics and Society. https://doi.org/10.4135/9781452206905.n273.

[12] Abdou, H.A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. Intelligent Systems in Accounting, Finance and Management, 18(2–3): 59–88. https://doi.org/10.1002/isaf.325.

[13] Mahmoud, H.H.H., & Ismail, T. (2020). A review of machine learning use-cases in telecommunication industry in the 5G Era. 16th International Computer Engineering Conference (ICENCO), IEEE.

[14] Isaac, E. (2023). Credit Management Practices and Financial Performance of Commercial Banks in Rwanda: A Case of Access Bank Plc. Journal of Finance and Accounting, 7(3): 119–138. https://doi.org/10.53819/81018102t2156.

[15] Aggarwal, N. (2018). Machine Learning, Big Data and the Regulation of Consumer Credit Markets: The Case of Algorithmic Credit Scoring. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3309244.

[16] Ethio telecom (2023). https://www.ethiotelecom.et/.

[17] Hand, D.J. (2018). Data Science. Wiley StatsRef: Statistics Reference Online. Portico, Pages 1–5. https://doi.org/10.1002/9781118445112.stat08150.

[18] Ma, Z., Hou, W., & Zhang, D. (2021). A credit risk assessment model of borrowers in P2P lending based on BP neural network. Plos One, 16(8): e0255216. https://doi.org/10.1371/journal.pone.0255216.

[19] Smirnov, V.S., & Stupnikov, S.A. (2023). A Deep Learning Approach to Credit Scoring Using Credit History Data. Lobachevskii Journal of Mathematics, 44(1): 198–204. https://doi.org/10.1134/s1995080223010365.

[20] Studer, S., Bui, T.B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K.R. (2021). Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. Machine Learning and Knowledge Extraction, 3(2): 392–413. https://doi.org/10.3390/make3020020.

[21] Kiran, V.S., Reddy, B.T., Kumar, D.U., Varma, K.S.A., & Kiran, T.S. (2023). Loan Eligibility Prediction Using Machine Learning. International Journal for Research in Applied Science and Engineering Technology, 11(8): 55–60. https://doi.org/10.22214/ijraset.2023.55132.

[22] Pallavi, M., & Sachin, S. (2016). Survey on Methodologies and Techniques Involved in Feature Selection. International Journal of Science and Research, 5(2): 948–952. https://doi.org/10.21275/v5i2.nov161320.