Regression-Based Machine Learning Framework for Customer Churn Prediction in Telecommunication Industry

Sylvester Igbo Ele [1], Uzoma Rita Alo [2,*], Henry Friday Nweke [3], and Ofem Ajah Ofem [1]

[1] Department of Computer Science, University of Calabar, Nigeria;
Email: myyrs2015up@gmail.com (S.I.E.), ofemofem2019@unical.edu.ng (O.A.O.)
[2] Computer Science Department, Alex Ekwueme Federal University, Ndufu-Alike, Abakaliki, Nigeria
[3] Centre for Research in Machine Learning, Artificial Intelligence and Network Systems, Computer Science Department, Ebonyi State University, Abakaliki, Nigeria; Email: henry.nweke@ebsu.edu.ng (H.F.N.)
*Correspondence: alo.uzoma@funai.edu.ng (U.R.A.)

***Abstract*—Customers' movement from one telecom provider to the other has become a foremost issue in the telecommunication industry. This exacting issue has engendered stiff competition among vendors in the telecommunication industry to retain their customers. This competition is consequent upon the fact that it is more costly to acquire new customers than it takes to maintain the existing ones. The ability to make an accurate prognosis about customers who are likely to churn, and to offer incentives to retain them, places such telecom providers on a foundational platform to stand in the market. Recent studies in churn prediction utilized a single machine learning model that the results cannot be easily generalized to a new dataset or new scenario. In addition, these machine learning models are complex and with high computational time. In this study, we propose a comprehensive and computationally efficient Regression-Based Machine Learning Framework for Customer Churn Prediction in Telecommunication Industry. We evaluated nine different Regression Models and compare their performances. Moreover, we evaluated and determine which model is best suited to the proposed approach. The models were evaluated using four commonly used Regression based metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). The best accuracy was recorded by Lasso Regression, with MAE, MSE, RMSE, and R^2 of 7.77E−02, 1.21E−02, 1.10E−01, and 0.981407 (98%), respectively. This result shows that the Lasso Regression-based model performed better, realized the line of best fit, fit well with the observed data, and guarantee better predictions when deployed using the proposed approach.**

***Keywords*—customer Churn, machine learning, regression models, telecommunication, multiple classifier systems**

I. Introduction

Recently, the telecommunication industry has proved to be one of the fastest-growing industries and has evolved to encompass various components of customer satisfaction that require improvement. Today companies under this subdivision of the economy are doing everything in their efforts to maintain and retain their teeming customers by satisfying them through offers and rebates. Telecommunication industries operate based on selling certain products to customers. Finding the target group that is more interested in a certain product is the main task of a company. Creating, maintaining, and retaining relationships with a company's customer base is one of the crucial business and marketing tasks in heavily competitive markets such as telecommunications or subscription-based models. The process of determining the target group interested in a certain product is at the top of the agenda of every business company. Technical implementation of these methods is widely spreading in various countries' telecommunication industries with limited impact in African countries. It is a consequence of this fact that the acquisition of a new customer is far more costly than the absorbent of an existing one [1]. To retain existing customers, companies including telecommunication sectors provide incentives and special packages to them. Therefore, determining the customers that need these special packages and making sure the customer did not churn or move to competitors' services is very important [2].

Customer churn is a major problem in the telecommunication industry, and one of the most significant concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecommunication field, companies are seeking to develop means to predict potential customers that might churn in the future [3]. Customer churn could be the result of low-level customer satisfaction, aggressive competitive strategies, new products, regulations, etc. Consequently, churn models aim to identify early churn signals and recognize customers with an increased likelihood to leave voluntarily [4].

Companies use various prediction tools, such as descriptive statistics, rule-based, and machine learning to predict customer churn. However, the statistics method is too simplistic and analyzes data in small quantities,

Manuscript received November 21, 2022; revised March 1, 2023; accepted June 19, 2023; published October 8, 2023.

doi: 10.12720/jait.14.5.1046-1055

therefore, it is difficult to use the method to accurately predict customer churn [5]. The rule-based method expresses data patterns and uncovers meaning in data using an if-then rule. Nonetheless, rule-based methods lack scalability and robustness to uncover the pattern that might show customer churn. The machine learning model seeks to improve churn prediction models and ascertain customers that might leave the company using past experiences and packages. The recent implementation of machine learning models for customer churn prediction in telecommunication industries includes Linear Regression [6, 7], Neural Network Model [8, 9], Decision Tree [10, 11], Support Vector Machine [12, 13], etc. The use of these machine learning algorithms is useful for the correct identification of customer retention and the type of products that are essential for telecommunication companies [1]. However, some of the machine learning models implemented for customer churn prediction in telecommunication industries are unsuitable due to their computational complexity, prone to model over-fitting, and require lots of training data [14].

For instance, artificial neural networks although produce high-performance accuracy when used for churn predictions, require large amount of training data which might be difficult to obtain. Moreover, the support vector machine algorithm is sensitive to missing data even after preprocessing. Therefore, this paper proposes multiple regression models to investigate and predict customer churn in the telecommunication industry using a publicly available dataset with varied customer features and data formats.

Regression models are supervised machine learning methods that evaluate the relationship between variables and emphasize the association between the target (dependent) variable and predictor variable (independent) [15, 16]. In addition, it indicates the line or curve that passes along the data point and regression line with the lowest minimum thereby demarcating customers that may likely leave the services of the telecom industry if adequate and better packages are not provided for them [17]. As recently noted by Shin [18], regression analysis remains the most versatile tool with the widest applications, less black box, easy to understand and communicate, and helps to determine the most important factors that will cause the customer to churn. Considering the strengths of regression analysis-based machine learning, this paper comprehensively evaluates nine regression models for comparison and chooses the best model. The regression models evaluated include the linear regression model, Lasso regression, elastic net regression, polynomial regression, stochastic gradient descent regression, ridge regression, robust regression, support vector regression, and random forest regression. A detailed explanation of these regression models is explained in Section III.D of the paper.

In the nutshell, the contributions of this paper are as follows:

- Comprehensive analysis of recent studies on machine learning algorithms for telecommunication-based customer churn prediction;
- Develop and evaluate machine learning-based regression model for customer churn prediction in the telecommunication industry using publicly available datasets;
- Evaluate nine regression models which include linear regression model, Lasso regression, elastic net regression, polynomial regression, stochastic gradient descent regression, ridge regression, robust regression, support vector regression, and random forest regression methods for telecommunication-based customer churn prediction and ascertain the best model;
- The proposed evaluation was implemented using thirteen features/attributes extracted from customer details;
- All-encompassing evaluation of the proposed system using various regression performance metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and cross-validation.

The remainder of the paper is organized as follows. Section II discusses recent literature on churn prediction using telecommunication data. Section III presents the methodology used to develop the churn prediction system. Section IV discusses the experimental setup, results, and discussion, while section V concludes the paper and outlines future research directions.

II. Literature Review

In the past decades, various studies have utilized classification and clustering algorithms for the identification of customers that have the likelihood to leave services of telecommunication firms due to low-services satisfaction, regulations, and lack of aggressive marketing strategies to attract customers. These machine-learning models include support vector machines, logistic regression, tree-based algorithms, and neural networks. In this section, we review recent studies that utilize various machine learning algorithms for customer churn prediction in the telecommunication industry. Coussement and Poel [19] applied Support Vector Machines (SVM) in a newspaper subscription context to construct a churn model with higher predictive performance. The study carried out a comparison between two SVM parameter-selection techniques, required to implement support vector machines. These techniques were based on grid search and cross-validation. The researchers, through their study, show that support vector machines can outperform traditional pattern recognition when the optimal parameter selection procedure is applied. However, when compared with other machine learning models such as random forest and logistic regression, the proposed support vector machine obtained lower performance accuracy. Furthermore, the support vector machine model is sensitive to data with missing values.

In a similar study, Vafeiadis and Diamantaras *et al.* [12] conducted a comparative study on the use of standard machine learning methods to resolve the challenging issue of customer churning prediction in the telecommunication industry. In the first phase of their experiments, they made use of all models and evaluated them using cross-validation on a popular, and public-domain dataset.

To determine the most effective parameter combinations, they carried out a sequence of Monte Carlo simulations for each method applied, and for a host of parameters. The study results show a clear superiority of the boosted versions of the models against the non-boosted versions. SVM-POLY using AdaBoost yielded the overall best classifier with an accuracy of close to 97% and an *F*-measure of more than 84%. To reduce the computational overhead inherent in some machine learning models, studies have utilized logistic regression for customer churn prediction. Logistic regression helps to model the relationship between various packages provided by telecommunication firms and customer satisfaction. Zhang and Moro *et al.* [2] proposed a fusion of discriminant analysis and logistic regression and customer segmentation method for customer churn prediction. The study achieved a prediction accuracy of 93.94% accuracy. The author utilized factor analysis to investigate the business of characterization of telecom clients and to build a discriminant analysis model and logistic regression model to predict client churn segmentation data from three major Chinese telecommunication companies. Implementation of churn prediction in telecommunication industries using mass data was presented by Gursoy [7]. The study deployed logistic regression and decision tree machine learning models, however, low-performance accuracy was obtained. Consequently, further improvements such as feature selection and evaluation of simpler machine learning models such as regression required obtained higher performance accuracy.

Recently, studies have also evaluated multiple classifier systems for customer churn prediction. These studies either evaluated multiple classification algorithms or integrate different algorithms to ascertain the best model for customer churn prediction.

For example, Khodabandehlou and Zivari [20] developed a predictive framework for customer churn via six stages for accurate prediction and inhibiting customer churn in business. These six stages, according to the study, involve: 1) the collection of customer behavioral data and preparation of the data; 2) the formation of resultant variables and selection of important variables, using the discriminant analysis method; 3) selection of training and testing data and studying their proportion; 4) building of the prediction models by applying simple, bagging and boosting forms of supervised machine learning; 5) comparative analysis of the churn prediction models using various machine-learning approaches and selected variables; and 6) providing suitable strategies based on the proposed model. Their findings revealed that their model with an accuracy of 97.92%, performed much better in churn prediction, in comparison to Recency, Frequency, Monetary (RFM) value. Among the supervised machine learning methods, according to the study, Artificial Neural Networks (ANN) perform better with the highest accuracy, and the least accuracy was recorded by Decision Trees (DT). Jain [21] evaluated different machine learning models to assess which of the algorithms are good at discriminating customers that might churn in long run. The algorithms implemented include the k-Nearest Neighbors, fuzzy cluster, and convolutional neural networks. Even though experimental results show that the convolutional neural network proposed superior results compared to k-NN and fuzzy clusters, the model has a higher computation time.

Hadden and Rahman *et al.* [22] analyze the impact of different machine learning algorithms on the performance of the churn prediction model. The study evaluated three machine learning models such as neural networks, regression tree, and simple regression which showed that the decision tree has higher performance among other machine learning models analyzed in their study.

Ahmed and Khunteta *et al.* [23] developed multiple classification algorithms by bagging for customer churn prediction. Experimental results obtained using two publicly available datasets shows that the proposed system obtained higher accuracy and can easily generalize to the new dataset. However, the model has a high computation time. Bilal and Almazroi *et al.* [1] proposed the integration of clustering and ensemble bagging and voting algorithms for churn prediction. In the study, different clustering algorithms were evaluated separately and combined with voting, bagging, and stacking ensemble methods to improve the performance. The integration achieved 93.6 average accuracy on publicly available data. However, multiple classifier systems have high computation time and increased complexity as the algorithms require larger training data to produce good performance accuracy.

Some studies have implemented tree-based algorithms and artificial neural networks for customer churn prediction. Tree-based algorithms combine multiple decision trees on independent subsects of customer churn data while artificial neural networks utilize multiple layers of neurons to train data collected from telecom customers. For instance, Adwan and Tiwari *et al.* [24] applied different tree-based machine learning models such as Glioblastoma (GBM), decision tree, random forest, and XGboost for customer churn prediction. In the evaluation, XGBoost outperformed other tree-based algorithms in terms of AUC and accuracy. In addition, Idris and Iftikhar *et al.* [3] proposed the use of Genetic Programming (GP)-Adaboost method to establish a telecommunication companies' customer churn prediction model with good understandability and intuitiveness. However, it is difficult to deal with the linear relationship of variables in the decision tree, and also challenging to handle the interaction impact of variables in logistic regression

One of the major issues in churn prediction using a machine learning model is data imbalance. In recent studies, Bruez and Afzal *et al.* [25] proposed the integration of logistic regression, random forest, and re-sampling approach to reduce the class imbalance in customer churn and improve prediction performance. In addition, the study also applied to boost algorithms to improve prediction accuracy. Gajowniczek and Ząbkowski *et al.* [26] implemented an artificial neural network embedded with an entropy cost function for the prediction of customer churn. However, the proposed system produce poor accuracy when compared with other numerical methods such as decision tree, and support vector machine.

Although a lot of machine learning models have been proposed for churn prediction in telecommunication sectors, these studies utilized limited numbers of machine learning models which makes it difficult to assess the best model for customer churn prediction implementation. Moreover, these machine learning models have high computation time and require a large number of datasets to obtain optimal performance results. There is a need for comprehensive implementation of different models, especially regression models that have shown increased performances. To bridge this gap in the literature, this paper presents an implementation of nine regression models for the churn prediction task. In addition, these models were evaluated using five performance metrics to ensure an all-inclusive evaluation of the proposed methods.

III. Materials and Methods

The proposed regression-based machine learning for customer churn prediction in the telecommunication industry is shown in Fig. 1. The steps involved include data collection, preprocessing, extraction of relevant features, building effective telecommunication customer churn prediction, and evaluation of the built model using appropriate evaluation metrics. These steps are explained below.

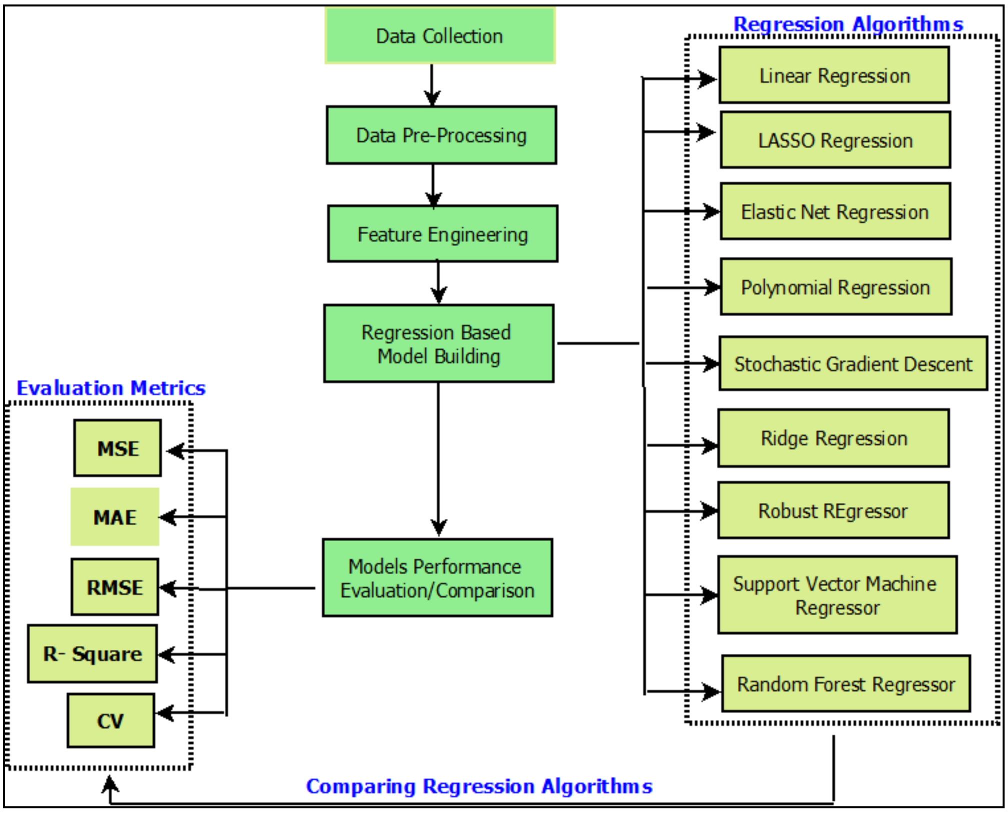

Figure 1. Proposed regression-based framework for the telecom customer churn prediction.

A. *Data Collection*

The first step in telecommunication customer churn prediction is the collection of appropriate customer data. In this study, we utilized a publicly available dataset to develop the proposed customer churn prediction model. The essence is to ensure the reproducibility and generalization of the proposed model. The dataset used is available on Kaggle.com (https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets). The dataset termed “churn-big-80” contains nineteen feature variables that show the characteristics of clients of the telecommunication corporation. The class label indicates whether the customer has left the telecommunication corporation within the last month or not. The class labels include “*No*” which indicates the customer did not leave the organization last month while “*Yes*” indicate the customer left the organization last month. The essence of generating a big machine learning dataset is to analyze the relationship between the customer’s characteristics and the churn. The dataset contains customer information grouped into three subgroups and includes demographic information, customer account information, and services provided by the telecommunication company. The demographic information includes customers’ gender, senior citizen, relationship status, and whether the customer has dependents or not. Customer account information provides details such as the period the customer has stayed with the company, contact details, whether paperless billing or not, payment details, etc. Service information contains details of phone services, multiple lines, types of internet services, online security status, online backup, etc. This information constitutes the feature variable to be analyzed to determine if a customer will leave the telecom company or not. The dataset information and their characteristics are shown in Table I below.

TABLE I. MODEL COEFFICIENTS TABLE

Attributes	Coefficient
Total day minutes	6.832052e–14
Total day calls	4.523422e–16
Total day charge	–6.785133e–14
Total eve minutes	1.020983e–15
Total eve calls	1.580627e–16
Total eve charge	–1.124510e–15
Total night minutes	–1.742031e–14
Total night calls	–2.570126e–18
Total night charge	1.744569e–14
Total intl minutes	–4.374194e–15
Total intl calls	2.256892e–17
Total intl charge	4.507358e–15
Customer service calls	–2.242435e–16
Churn	7.046660e–01

B. Data Preprocessing

In most cases, the collected customer data are affected by noise, anomalies, data duplication, and missing values. Therefore, data preprocessing is used to remove data duplication, noise, and biases anomalies, and improve the correlation and linear relationship between each data point [1]. In extracting data, we performed various data preprocessing steps. First, we replaced the missing values with the mean of the data. All the data that contain zero values were removed as these variables are fixed, which makes prediction difficult. Second, data with duplicate values, noise, and errors in the downloaded dataset were identified and removed. The preprocessed was saved for the identification of relevant attributes to develop the churn prediction model.

C. Feature Extraction and Selection

The feature extraction process is the identification of relevant and discriminative attributes to describe whether a telecommunication customer would leave the company or not. In addition, the identification of relevant attributes from the churn prediction dataset would help to improve prediction accuracy and reduce computation time. In the study, we utilized all nineteen attribute variables in the dataset. Some of the attribute variables deployed to implement the proposed churn prediction models include Total day minutes, Total day calls, Total day charge, Total eve minutes, Total eve calls, Total eve charge, Total night minutes, Total night calls, Total night charge, Total international minutes, Total international calls, Total international charge, Customer service calls, Churn. The churn variable is the target (dependent) variable or the predictor variable and while the rest were used as the predictor variables.

To fully explored and understand the telecommunication churn dataset, we carried out data exploration and visualization, by creating a simple plot to check the data, and a correlation matrix of the heatmap to graphically visualize our dataset, measure their dependencies between the different variables, as illustrated in Fig. 2. The predictor variable in our dataset is the 'Churn' with two classes, 0 representing the non-churn class and 1 representing a churned class.

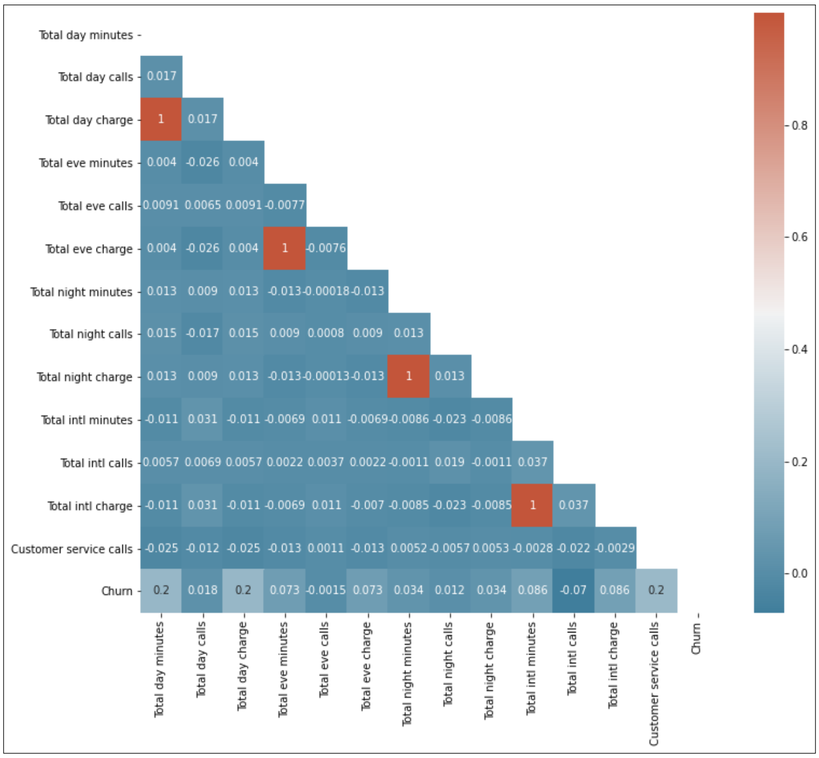

Figure 2. Correlation Heatmap showing correlation between different variables.

D. Regression Model Building

The techniques and methodology used for the development of this model are shown in Fig. 1. We extensively evaluated nine for their performance and accurate prediction of churn in mobile telecommunication. The models evaluated include Linear Regression (LRM), Lasso Regression, Elastic Net Regression (ENR), Polynomial Regression (PR), Stochastic Gradient Descent (GDS), Ridge Regression (RR), Robust Regression (RoR), Support Vector Regression (SVR), and Random Forest Repressor (RFR). The description of these models is explained below.

1) Linear Regression Model (LRM)

In the Linear Regression model, a dataset of previous observations is used to understand upcoming values of descriptive and numerical targeted variables [27]. The formula for LRM is in Eq. (1)

$$prob(y=1)\frac{e^{\beta_0+\sum_{k=1}^{k}\beta_k y_k}}{1-e^{\beta_0+\sum_{k=1}^{k}\beta_k y_k}} \tag{1}$$

where y represents the dependent variable, B_0 is the intercept, B_1 is the regression coefficient, ε is the model error.

2) Lasso regression

Lasso regression is a kind of linear regression that uses shrinkage. Data values like the mean are shrunk towards a point using shrinkage. The procedures of Lasso encourage models with fewer parameters [28]. The cost function for the lasso, or what is referred to as the least absolute shrinkage repression is given in Eq. (2)

$$\sum_{i=1}^{k}=(y_i-x_i)^2=\sum_{j=1}^{k}(y_i-\sum_{j-1}^{q}M_1 X N_{ij})^2+\lambda\sum_{j=0}^{q}\left|M_j\right| \tag{2}$$

where y represent dependent variable, k and q is the data point, λ is the tuning parameter and M is the coefficient

3) Elastic net regression

The elastic net regression algorithm applies a weighted grouping of L_1 and L_2 regularization. The goal of this algorithm is to minimize loss function as shown in Eq. (3).

$$L_{rnrt}(\beta)=\frac{\sum_{n=1}^{n}(y_i-\alpha_i^t\beta)^2}{2n}+\lambda(\frac{1-\alpha}{2}\sum_j^n\beta_j^2+\alpha\sum_j^n|\beta_j|) \tag{3}$$

where y is the observation, λ is positive regularization parameter, B_0 and B_1 are intercept and coefficient of interpolation.

4) Polynomial regression model

This regression model involves regression analysis where the relationship that exists between the independent variables and dependent variables is analyzed in a certain gradation, say the nth degree polynomial. Polynomial regression for a model for a single predictor, X, is represented in Eq. (4). Here, B_0 and B_1 are intercept and coefficient of interpolation and ε error rate

$$Y=\beta_0+\beta_1+\beta_0 X+\beta_2 X^2+\ldots+\beta_n X^n+e \tag{4}$$

5) Stochastic Gradient Descent (SGD) model

Stochastic Gradient Descent (SGD) is an optimization technique for the unrestrained optimization process. It estimates the true gradient of E (w,b) such that it considers a particular training instance at a time. This model iterates over the training instance, and for every instance, it updates the model parameters base on the update protocol Eq. (5) [29]. Where, w, b represent the weight and biases, and y is the dependent variable.

$$E(w,b)=\frac{1}{n}\sum_{i=1}^{n}L(y_i,f(x_i))+\alpha R(w) \tag{5}$$

6) Ridge regression model

In the ridge regression algorithm, the cost function is changed by the addition of a penalty comparable to the square of the scale of the coefficients. It minimized the cost function as shown in Eq. (6) [30], where y is the dependent variable, w is the weight, and X is the input variable.

$$\sum_{i=1}^{k}=(y_i-x_i)^2=\sum_{i=1}^{k}(y_i-\sum_{j=1}^{k}w_i X n_{lj})^2+\lambda\sum_{j=1}^{k}|w_j^2| \tag{6}$$

7) Robust regression model

The robust Regression algorithm offers an alternative to least squares regression by needing less restrictive assumptions. These approaches try to diminish the influence of outlying observations such as to provide an improved fit to a good number of the data. This algorithm down-weights the effect of outliers, which causes their residuals to become larger and easier to identify. Assuming there is n size of the dataset, the Robust Regression is shown in Eq. (7)

$$y_i=X_i^T\beta+e_i \tag{7}$$

This equation becomes as described in Eq. (8)

$$e_i(\beta)=y_i=X_i^T\beta \tag{8}$$

Here, the value of i = 1… n, and the error term have been rewritten $e_i(\beta)$ to reflect the dependency of the error term on the regression coefficients.

8) Support Vector Regression (SVR)

SVR is a supervised learning model deployed to predict discrete values. The Support Vector Regression (SVR) uses of similar principle as the SVMs. The fundamental idea of SVR is to discover the best-fit line. In this algorithm, the best-fit line is the hyperplane, which is the maximum number of points [31].

9) Random Forest Regression (RFR)

Random forest regression is an ensemble of the decision tree. It is quite a robust algorithm. It grows a forest of many trees. It grows each tree on an independent bootstrap sample from the training data [32].

E. Model Evaluation

1) Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) represents the mean of the absolute difference between the actual value in the dataset and the predicted value of the model. Eq. (9) depicts the formula for the MAE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) \tag{9}$$

2) Mean Square Error (MSE)

Mean Square Error (MSE) is scale–dependent metric that measures the differences between the predicted values and the actual values of the data being computed [15]. Eq. (10) illustrates the mathematical formula for the Mean Square Error. The best way to measure how well the prediction matches the observed data (model performance), is to use the MSE. The closer the model prediction to the observations, the smaller the MSE becomes.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \tag{10}$$

3) Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) evaluates the ratio of the variance of the dependent variable described by the target variable. It is used to find the accuracy of the model. The RMSE is modeled mathematically as illustrated in Eq. (11).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)}^{2} \tag{11}$$

4) Coefficient of Determination (R^2)

R square has values ranging from 0 to 1. The nearer the *R* square is to 1, the more accurate the model is depicted in Eq. (12)

$$R^2 = \frac{SSE}{SST} \tag{12}$$

where

Sum of Square of Error (SSE) is the square of the difference between the actual value and the mean of all the actual values of the dataset, defined as in Eq. (13). Here, $\hat{y}_i$ is the predicted values of the dependent variable.

$$SSE = \sum_{i=1}^{M}(y_i - \hat{y}_i)^2 \tag{13}$$

Sum of Square of Total (SST) is the sum of the difference between the actual value and the mean of all the actual values in the dataset, defined as in Eq. (14).

$$SST = \sum_{i=1}^{M}(y_i - \bar{y})^2 \tag{14}$$

5) Cross Validation (CV)

The CV is a machine learning method for estimating a model by training different models on subclasses of the available input data and estimating them on the complementary subclass of the data. The chances of detecting over-fitting are very high when using cross-validation, hence the reason it is applied in this study. Our cross-validation was performed on the training set. The commonest kind of cross-validation is k-fold cross-validation. The accuracy of a model is the average of the accuracy of each fold. The mathematical expression for the k-fold prediction error is illustrated in Eq. (15).

$$CV(f) = \frac{1}{K}\sum_{i=1}^{K}L(y_i, f(x_i)) \tag{15}$$

IV. Experimental Setup, Result, and Discussion

A. Experimental Setup

This section describes the experimental implementation of the proposed machine learning-based telecommunication churn prediction model. The experiment was performed in the *churn-bigml-80 dataset* as described in Section III.A. The data were preprocessed and the relevant attributes were given to the regression models discussed in Section III.B for churn prediction. The data preprocessing and model building were done using Python programming with the relevant libraries as explained below.

To train the models, we started by splitting the dataset into the train features (*x array*) and the target features (*y array*). Furthermore, we split the dataset into a training set (to train the model), and a testing set (to evaluate the model). We adopted train-test split for the data splitting due its reduced computation complexity and efficient generalization error. The percentage of the split is 70% for training the model and 30% for testing the trained model. The model was trained using the sci-kit-learn python library in Jupyter Notebook, while the model was evaluated using the regression model evaluation metrics listed in Section III.E.

To be able to evaluate the model and interpret it, its coefficient and intercept were checked. The intercept is the mean value of the regression model when the predictor was zero. The intercept (constant) for the model is 2.290460878885316, indicating the average dependent variable was 2.2905 when the predictor (Churn) was zero, which is a substantial effect for incentives. The coefficients of the variables are shown in Table II.

TABLE II. MODEL PERFORMANCE RESULTS

Model	**MAE**	**MSE**	**RMSE**	**R^2 Square**	**Cross Validation**
Linear Regression	6.46E−16	7.46E−31	8.64E−16	1	1
Lasso Regression	7.77E−02	1.21E−02	1.10E−01	0.975684	0.031222
Elastic Net Regression	6.79E−02	9.29E−03	9.64E−02	0.981407	0.032726
Polynomial Regression	4.68E−15	4.67E−29	6.83E−15	1	0
Stochastic Gradient Descent	6.20E−02	4.90E−03	7.00E−02	0.990198	0
Artificial Neural Network	6.79E−02	9.29E−03	9.64E−02	0.981407	0
Random Forest Regression	0.00E+00	0.00E+00	0.00E+00	1	0
SVM Regression	2.98E−02	7.00E−03	8.36E−02	0.985991	0

B. Results and Discussion

For each regression model evaluated, all the common regression metrics as described in (Section II.E) were applied as common evaluation criteria. These metrics are the loss functions that we want to minimize using our model. The result obtained from the model evaluation is indicated in Table II, while Figs. 3–7 depict how the different regression models performed in each metric.

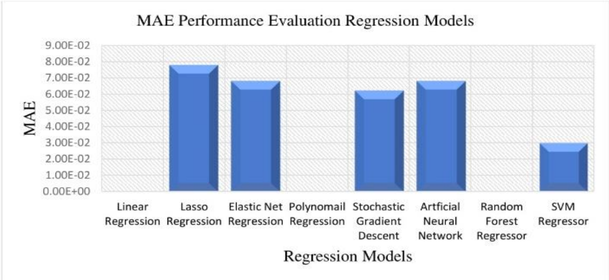

Figure 3. Performance of Regression Models in terms of MAE.

From the results, as illustrated in Figs. 3–7, Lasso regression recorded the best (minimal) MAE, MSE, and RMSE of 7.77E−02, 1.21E−02, and 1.10E−01, respectively, and Regression Coefficient (R^2) of 0.981407, and Cross-Validation of 0.031222. The best regression coefficient of determination (R^2) was recorded by Linear Regression (LR), Polynomial regression, and Random Forest Regression (RFR) with an R^2 of 1 (100%). This result was closely followed by stochastic gradient descent, Support Vector Regression, Artificial Neural Network (ANN), Elastic Net Regression (ENR), and Lasso regression which recorded R^2 of 0.990198, 0.985991, 0.981407, 0.981407, and 0.975684, respectively. The Linear regression model achieved a perfect Cross-Validation of 1 (100%). It was followed by Elastic Net and Lasso, with cross-validation of 0.032726 and 0.031222, respectively, while the rest of the models had zero (0).

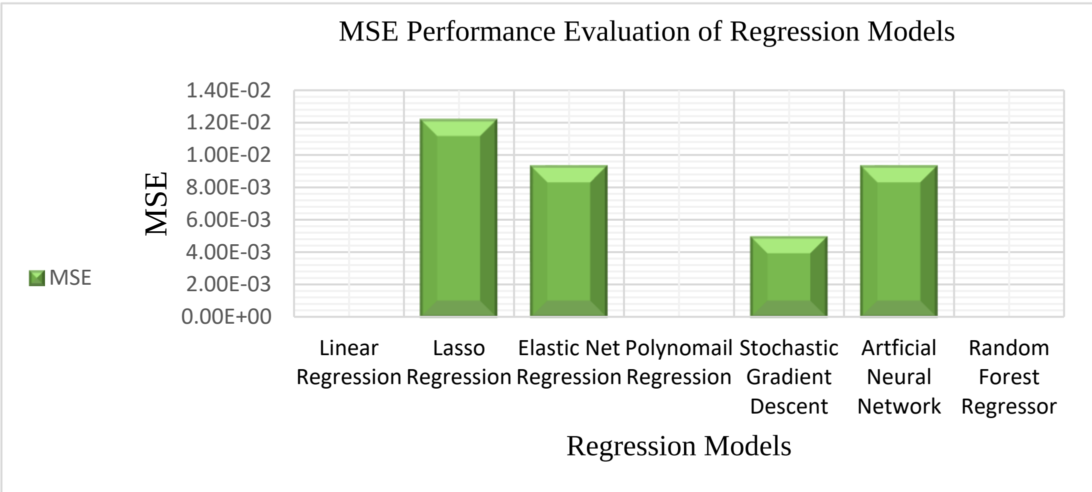

Figure 4. Performance of Regression Models in terms of MSE.

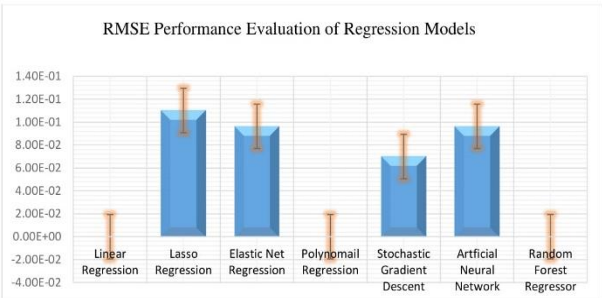

Figure 5. Performance of Regression Models in terms of MSE.

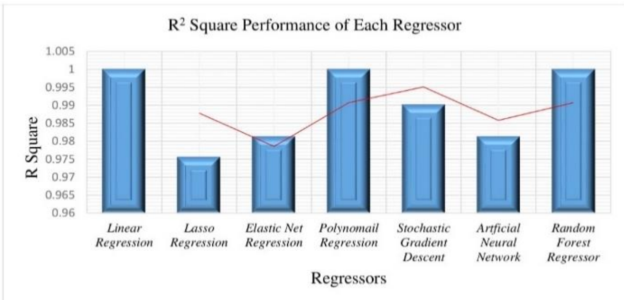

Figure 6. Performance of Regression Models in terms of R-Square.

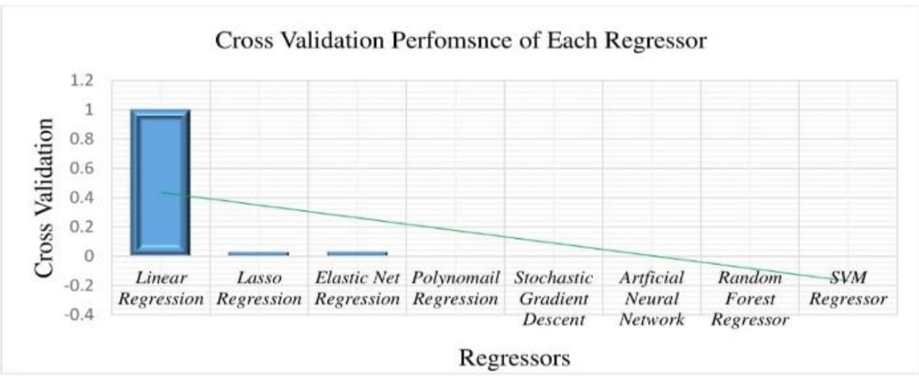

Figure 7. Performance of Regression Models in terms of cross-validation.

The essence of this study is to find the line of best fit by minimizing the loss function and obtaining a larger and positive correlation in our model framework. Overall, the study hopes to determine which of the regression models in our framework can best minimize or reduce the loss function and achieve the line of best fit. Therefore, in our evaluation, we take the models with the least error estimated by MAE, MSE, RMSE, and the largest R^2 (R-squared), which indicates how well our model fits the observed data. The result obtained from our experiment indicates that Lasso Regression achieved the lowest (best) estimates (MAE, MSE, and RMSE), and a very large correlation coefficient (R^2). Although Linear Regression, Polynomial, and Random Forest Regressions achieved an R^2 of 1, the respective error estimates were poor in comparison with Lasso. Also, significant to this study is the accuracy and efficiency of the model on unseen data are of. That is, building a model that achieves better forecasts on unseen data and eludes under-fitting and overfitting, as such we performed Cross-Validation (CV) on the training data. The result shows that the best cross-validation (100%) was obtained when the model was trained using Linear Regression. Other regression models recorded zero (0) CV, except Lasso and Elastic Net Regressions which recorded (0.031222) and (0.032726), respectively. The findings of the study indicate that the proposed Regression-Based model would perform better, realize the line of best fit, and would fit well the observed data when deployed using Lasso Regression. Lasso Regression recorded the least error estimated by MAE, MSE, and RMSE and a very large R^2. The limitation, however, is that the model would not have the ability to guarantee better forecasts on unseen data and handle well the problem of under-fitting and overfitting since Lasso obtain an insignificant CV accuracy of 3.12% (0.031222).

V. Conclusion and Future Work

The purpose of this study was to develop a machine learning-based framework for customer churn prediction in the telecommunication industry, using different regression models, in a view to determine which model(s) is best suited to predict customer movement from one telecom provider to the other. Several Regression models such as Linear Regression (LRM), Lasso Regression, Elastic Net Regression (ENR), Polynomial Regression (PR), Stochastic Gradient Descent (GDS), Support Vector Regression (SVR), Artificial Neural Network (ANN) and Random Forest Repressor (RFR), were applied to determine which of the model can best be deployed in the proposed framework. To achieve this goal, several regression-based metrics like MAE, MSE, RMSE, R^2, and Cross-Validation (CV) were used to evaluate the performance of each regression model. Overall, the result indicates that the proposed framework works better when deployed using the Lasso Regression model, based on the performance of the. Lasso model in minimizing and reducing the loss functions, and attaining the line of best fit on the observed data.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization of idea: S.I.E and U.R.A, Methodology: H.F.N and U.R.A, software: S.I.E and O.A.O, Implementation: S.I.E and O.A.O, Writing: S.I.E, H.F.N and U.R.A, Supervision: U.R.A and H.F.N. All authors have read and agreed to the published version of the manuscript.

References

[1] S. F. Bilal, A. A. Almazroi, S. Bashir, F. H. Khan, and A. A. Almazroi, "An ensemble-based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in telecom industry," *PeerJ Computer Science*, vol. 8, p. e854, 2022.

[2] T. Zhang, S. Moro, and R. F. Ramos, "A data-driven approach to improve customer Churnn Prediction based on telecom customer segmentation," *Future Internet*, vol. 14, no. 94, 2022. https://doi.org/10.3390/fi14030094

[3] A. Idris and A. Iftikhar, "Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO under-sampling," *Cluster Computing*, vol. 22, pp. 7241–7255, 2017.

[4] P. Lalwan, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: A machine learning approach," *Computing*, vol. 104, no. 2, pp. 271–294, 2022.

[5] M. Al Amin, M. M. Jewel, and M. H. Fouji, "Influencing factors of customer attitude towards SMS marketing—A case of mobile telecommunication industry in Bangladesh," *Jagannath University Journal Business Studies*, vol. 1, pp. 65–78, 2019.

[6] G. Kraljević and S. Gotovac, "Modeling data mining applications for prediction of prepaid churn in telecommunication services," *Automatika*, vol. 51, no. 3, pp. 275–283, 2010.

[7] U. S. Gürsoy, "Customer churn analysis in telecommunication sector," *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, vol. 39, no. 1, pp. 35–49, 2010.

[8] E. Shaaban, Y. Helmy, A. Khedr, and M. Nasr, "A proposed churn prediction model," *International Journal of Engineering Research and Applications*, vol. 2, no. 4, pp. 693–697, 2012.

[9] C. Kirui, L. Hong, W. Cheruiyot, and H. Kirui, "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining," *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 2, p. 165, 2013.

[10] K. Kim, C. H. Jun, and J. Lee, "Improved churn prediction in telecommunication industry by analyzing a large network," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6575–6584, 2014.

[11] D. Radosavljevik, P. van der Putten, and K. K. Larsen, "The impact of experimental setup in prepaid churn prediction for mobile telecommunications: What to predict, for whom and does the customer experience matter?" *Transaction Machine Learning and Data Mining*, vol. 3, no. 2, pp. 80–99, 2010.

[12] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simulation Modelling Practice and Theory*, vol. 55, pp. 1–9, 2015.

[13] T. V. Ly and D. V. T. Son, "Churn prediction in telecommunication industry using kernel Support Vector Machines," *PloS One*, vol. 17, no. 5, p. e0267935, 2022.

[14] C. Vercellis, *Business Intelligence Data Mining and Optimization for Decision Making*, Wiley, 2009.

[15] U. R. Alo, S. I. Ele, and H. F. Nweke, "A conceptual framework for network traffic control and monitoring using artificial neural networks," *Journal of Theoretical and Applied Information Technology*, vol. 97, no. 22, pp. 3396–3412, 2019.

[16] R, Panchotia. (2020). *Predictive Modelling Using Linear Regression.* [Online]. Available: https://medium.com/swlh/predictive-modelling-using-linear-regression-e0e399dc4745#:~:text=Linear%20regression%20is%20one%20of,given%20predictor%20variable(s)

[17] S. Gawali. (2021). *Linear Regression in Machine Learning*. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/06/linear-regression-in-machine-learning/
[18] T. Shin. (2021). 3 reasons why you should use linear regression models instead of neural networks. Toward data science. [Online]. Available: https://towardsdatascience.com/3-reasons-why-you-should-use-linear-regression-models-instead-of-neural-networks-16820319d644
[19] K. Coussement and D. van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *Expert Systems with Applications*, vol. 34, no. 1, pp. 313–327, 2008.
[20] S. Khodabandehlou and M. Zivari-Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior," *Journal of Systems and Information Technology*, vol. 19, no. 12, pp. 65–93, 2017.
[21] H. Jain, A. Khunteta, and S. Srivastava, "Telecom churn prediction and used techniques, datasets and performance measures: A review," *Telecommunication Systems*, vol. 76, pp. 1–18, 2020. doi: 10.1007/s11235-020-00727-0
[22] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Churn prediction: Does technology matter," *International Journal of Intelligent Technology*, vol. 1, no. 2, pp. 104–110, 2006.
[23] M. Ahmed, H. Afzal, I. Siddiqi, M. F. Amjad, and K. Khurshid, "Exploring nested ensemble learners using overproduction and choose approach for churn prediction in telecom industry," *Neural Computing and Applications*, vol. 32, no. 8, pp. 3237–3251, 2020. doi: 10.1007/s00521-018-3678-8
[24] O. Adwan, H. Faris, K. Jaradat, O. Harfoushi, and N. Ghatasheh, "Predicting customer churn in telecom industry using multilayer preceptron neural networks: Modeling and analysis," *Life Science Journal*, vol. 11, no. 3, pp. 75–81, 2014.
[25] J. Burez and D. van den Poel, "Handling class imbalance in customer churn prediction," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4626–4636, 2009.
[26] K. Gajowniczek, A. Orłowski, and T. Ząbkowski, "Insolvency modeling with generalized entropy cost function in neural networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 526, 120730, 2019.
[27] N. A. Naz, U. Shoaib, and M. S. Sarfraz, "A review on customer Churn prediction data mining modeling techniques," *Indian Journal of Science and Technology*, vol. 11, no. 27, pp. 1–6, 2018. doi: 10.17485/ijst/2018/v11i27/121478
[28] C. Wheelan. (2014). Lasso Regression: Simple Definition. [Online]. Available: https://www.statisticshowto.com/lasso-regression/
[29] S. Ruder. (2016). An overview of gradient descent optimization algorithms. [Online]. Available: https://ruder.io/optimizing-gradient-descent/
[30] S. Bhattacharyya. (2018). *Ridge and Lasso Regression: L_1 and L_2 Regularization*. [Online]. Available: https://towardsdatascience.com/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b
[31] A. Raj. (2020). Unlocking the true power of support vector regression. Towards data science. [Online]. Available: https://towardsdatascience.com/unlocking-the-true-power-of-support-vector-regression-847fd123a4a0#:~text=Support%20Vector%20Regression%20is%20a,the%20maximum%20number%20of%20points
[32] A. Cutler. (2020). Random forests for regression and classification. [Online]. Available: https://www.usu.edu/math/adele/RandomForests/Ovronnaz.pdf