# Plagiarism Report

*by* Turnitin Plagiarism Check

---

**Submission date:** 08-Sep-2023 12:44PM (UTC-0400)
**Submission ID:** 2160859697
**File name:** vansh_paper.docx (449.36K)
**Word count:** 6048
**Character count:** 33307

# Bank Churn Prediction: Leveraging Advanced Analytics for Customer Retention in the Banking Sector

Vansh
Department Name
College Name, City
State
India
email@gmail.com

Guide Name
Department Name
College Name, City
State
India
Guideemail@gmail.com

***Abstract*— Consumers are more likely to switch to better alternatives in today's fiercely competitive market if they believe they are available. If consumers have easy access to information, incur minimal switching costs, and are convinced that better options are available, they are more likely to switch. Particularly in markets with an abundance of information, this is true. As a result, businesses must keep a close eye on their clients in order to identify any potential problems. As a result of significant advancements in business intelligence, numerous information discovery and predictive analytics techniques have been developed. This study proposes a novel monitoring and forecasting paradigm for client attrition using data from ABC Bank's website. Historical churning behavior patterns were analyzed using data mining techniques to generate predictor variables for machine learning-based predictive models. To validate the efficacy of the classifiers, a sample of consumers who were about to depart was utilized. Businesses must plan for customer churn because it allows them to figure out customers who might stop using their services. Getting more customers is comparatively more expensive then retaining the old customers. As a result, it is critical to understand the precise marketing strategies that will increase the likelihood of retaining a customer who is about to leave. Businesses must consider a variety of customer habits, preferences, and cancellation factors in order to retain customers. Maintaining constant communication with each customer and determining when and how to run campaigns that will be most effective for them is critical. In order to develop effective marketing retention strategies, this study will introduce a novel method for identifying prospective customers who might leave anytime soon.***

***Keywords*— *Customer Attrition, Data Mining, Predictive Analytics, Machine Learning, Classifier Validation, Customer Churn, Retention Strategies.***

## I. INTRODUCTION

### A. *Background and Problem Identification*

The ECB implemented unconventionally accommodating monetary policies (such as quantitative easing as well as negative interest rate policies) during the subprime mortgage crisis that occurred in 2008. These policies were designed to boost economic growth, inflation, and unemployment rates (An & Arch, 2014). The primary source of income for retail banks is highly dependent on the interest rate set by the ECB; consequently, retail banks' earnings decreased significantly despite the fact that the economies of nations that broke the rules improved. This stance has been maintained by the ECB for a very long time, and it is not very likely that it will shift in the near future. In order to combat these initiatives, financial institutions were required to investigate other potential sources of revenue. This typically entailed reducing operating expenses while also modifying pricing strategies by increasing commissions. These operations led to an increase in additional margins (commission income) for the banks, but at the same time, they led to an increase in both customer retention and customer churn. The rate at which a company loses customers, measured as the percentage of customers who stop doing business with the company within a given period of time, is known as customer churn, and it has become a major concern for many businesses all over the world (Chandar, Laha, & Krishna, 2006).

This study will focus on a bank that is both one of the most well-known brands in the public consciousness and a popular corporation in the Portuguese business sector. The bank in question is the Banco do Porto. It was for a very long time one of the most important pillars of Portugal's financial system and a major driver of the economic growth of the country. In recent years, however, it has fallen on hard times. Due to the fact that it has been around for such a long time, it has one of the largest and most diverse customer bases in the country. Because of the need to protect customers' privacy, in this research, this organization has been addressed as bank XYZ. Because of the proliferation of the internet and social media, the retail banking industry has undergone significant transformations in recent years, making it simpler and more expedient for customers to access the information they require. It seems that the movement toward digitizing the processes involved in banking will continue. Bank XYZ is in the process of redesigning its branch network in order to accommodate this new reality. This effort, which may result in overall reductions, is intended to save money. The company generates leads for daily operations based on predetermined marketing strategies by utilizing its business intelligence architecture. These leads are then used in the company's operations. An increase in overall commissions has been associated with an increase in customer attrition, which has implications for churn management, according to recent marketing research that has come to the attention of Bank XYZ. A separate organization was responsible for carrying out the investigation. The findings of the study were based on data collected in the United States of America. In spite of these findings, the company has shown that it is committed to addressing customer churn in a proactive manner by continuing to prioritize this issue and launching marketing recovery initiatives. In order to maintain their

status as customers, bank XYZ delays dealing with the issue until after the client has already vacated their account.

### *B. Study Objectives and Importance*

The objective of this study is to use data mining and machine learning algorithms to lower the customer churn rate at bank ABC by devising a proactive strategy. The study aims to forecast the amount of customers who will leave the company by calculating the likelihood of departure for each individual customer in the company's database. The long-term objective is to establish a new paradigm for churn control so that marketing efforts can be directed toward retaining current customers rather than regaining former ones. The findings of this study are important for financial institutions because machine learning algorithms are very good at predicting which bank customers will stop using their services in the near future. It enables financial institutions to keep their existing customers and grow their profits. The predictive accuracy of churn models needs to be improved through additional research, and effective retention strategies need to be developed. The study addresses challenges encountered during the study, such as providing a challenging method for grading papers in accordance with bank specifications, to avoid making overly pessimistic predictions regarding the performance of the model. The findings of this study may have significant repercussions for the business world of finance.

## II. METHODOLOGY

Attrition, also known as customer attrition, is the tendency of a consumer to switch from one brand of goods or services to another and become a paying customer of a different business or organisation. Customer churn rate refers to the proportion of customers who stop using a company's services or products within a given time frame. After multiple negative interactions, a customer may decide to stop doing business with your company (or just one company). In addition, if a substantial number of dissatisfied customers leave abruptly, there will be substantial financial and reputational losses.
The "ABC BANK" is attempting to predict the Churn rate. Using various machine learning techniques, develop a predictive model of the optimal outcome.

### *A. Dataset Details*

The goal of this report is met by utilizing an existing dataset from the Kaggle repository [23]. The dataset contains 10,000 consumer records with 14 attributes. Among the many features of the dataset, the following stand out.:
In this dataset, the Exited attribute is the dependent variable or target variable, while the remaining attributes are independent variables. However, some independent variables have little impact on the prediction of attrition. RowNumber, CustomerId, and Surname are removed from the set of independent variables due to their lack of significance. In the subsequent pre-processing step, numeric attributes such as tenure, expected income, age, and credit score are scaled to a range between 0 and 1. The training dataset and testing dataset are then split 7:3:1. The classifier model creates a knowledge base and patterns using training data. This knowledge base is then utilized by the evaluation procedure. The presence of the dependent or target variable differentiates the training dataset from the testing dataset. The training dataset contains the independent and dependent variables required for building the knowledge base. However, the testing dataset will only contain known independent variables. Figure 1 depicts the dependent variable of the dataset, which is the distribution of customers who have abandoned purchases.

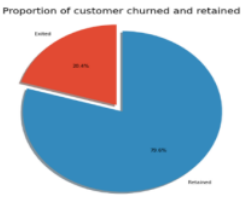



Figure 1 Proportion of Customer Churned and Retained

### *B. Data Pre-Processing*

Figure 1 depicts the four data preparation components. The sample data is then cleansed of elements (Row Number, Customer ID, and Surname) that have no relevance on the attrition forecast and are therefore omitted. String and categorical attributes were converted to numeric values using one-hot encoding because K-means can only understand numerical values. Only numerical values can be understood by K-means.
Using data visualization, determine which characteristics have the strongest correlation with attrition. Append a few features to the churn classification to improve it in light of the findings of the visualization. 1,000 samples with seventeen qualities are used out of a total of 10,000 data entries. From the preprocessed data, training data (80%) and assessment data (20%) are then extracted.

a. Scaling and One-hot Coding:

i. Scaling:

Min-max scaling is a normalisation technique used to convert numerical feature values to a standard scale, typically between 0 and 1.

The formula for min-max scaling is:

$$Scaled\ value = (Original\ value - Minimum\ value) / (Maximum\ value - Minimum\ value)$$

where "minimum value" and "maximum value" refer to the minimum and maximum values of the feature in the dataset. In the our code, the min() and max() functions are used to calculate the minimum and maximum values of each feature in the continuous_vars list.

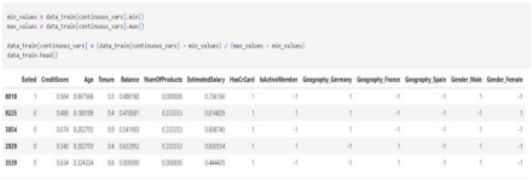


Figure 2 Scaling of the dataset using min() and max()

This technique can enhance the efficiency of machine learning algorithms when all features are measured on the same scale.

ii. One-hot Encoding:

To make categorical variables into numerical variables, label encoding assigns distinct integer values to them. Categorical data sets are transformed in this way. Labels can be encoded using the LabelEncoder class from scikit-learn. class of coders.

In our code, the 'Geography' and 'Gender' columns are selected from the input data X, and then label encoding is applied to each column using the fit_transform() method of the LabelEncoder object. The resulting encoded values are then assigned back to the respective columns of the input data X as shown below:

| | CreditScore | Geography | Gender | Age | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember | EstimatedSalary |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 619 | 0 | 0 | 42 | 2 | 0.00 | 1 | 1 | 1 | 101348.88 |
| 1 | 608 | 2 | 0 | 41 | 1 | 83807.86 | 1 | 0 | 1 | 112542.58 |
| 2 | 502 | 0 | 0 | 42 | 8 | 159660.80 | 3 | 1 | 0 | 113931.57 |
| 3 | 699 | 0 | 0 | 39 | 1 | 0.00 | 2 | 0 | 0 | 93826.63 |
| 4 | 850 | 2 | 0 | 43 | 2 | 125510.82 | 1 | 1 | 1 | 79084.10 |

Figure 3 Encoded Values using Label Encoder

One-hot encoding is commonly used for machine learning models as they typically work with numerical data. By using one-hot encoding, we can convert categorical data into a numerical format that can be used in these models.

*C. Training, Validation and Test phases:*

A statistical model can be created by using a specific method and historical data. This model must be trained using a training set, which involves adjusting the classifier to match the available data. During training, a loss function is optimized to demonstrate the model's inability to match the training data. It is not a given that a model's predictions will be accurate when used to predict new data, even if it performs exceptionally well on the training set of data. An overfitted model performs well with the data it was trained on but poorly with fresh data. Overfitting may be influenced by both disorder and a small training set. It is necessary to evaluate a model using data that was not used during training in order to determine whether it is overfit or not. If a model can identify non-random patterns, its error rate when applied to fresh data should be similar to its error rate during training. To accurately assess the performance of a model, a validation set helps quantify error. The model's parameter calibration is crucial for minimizing error. The Holdout method and k-Folds cross-validation are two common validation techniques. The process concludes with testing is used to evaluate the performance of the model. Underfitting can also occur if a model is too simple and cannot accurately represent the fundamental trend of the data. To prevent underfitting, new features must be added or alternative classifiers must be used.

*D. Evaluation Metrics:*

Before optimising a classifier, it is necessary to select the metric to be used to ensure that the optimisation objective is maintained throughout the entire analysis. Sometimes, but not always, design guidelines provide the metric. If no metric is specified, data analysts must select the most appropriate metric based on the issue at hand.

*1) Accuracy:*

Due to its simplicity and intuitiveness, accuracy is one of the metrics utilized most frequently by data scientists. The classifier's accuracy is determined by the number of correctly predicted records. The technique is straightforward, and its application to problems involving multiple classes is even simpler.

$$\text{Accuracy} = \frac{\text{number\_of\_correct\_predictions}}{\text{total\_number\_of\_records\_to\_be\_predicted}}$$

We can conclude that all classes have measurements with the same degree of precision using this equation. Accuracy is not a useful performance metric when some classes are more crucial to predict than others.

*2) Area Under the Curve:*

In data mining and ML, the Curve (AUC) is a familiar metric. It can only be used for tasks that require binary classification, in which only two outcomes are possible. In this context, the two classes are frequently referred to as positive and negative classes. As shown in Table 2.2, it is possible to specify exact values based on the projected class as opposed to the actual class.

In particular, we define:

$$\text{True\_Positive\_Rate} = \frac{TP}{TP+FN}$$

$$\text{False\_Positive\_Rate} = \frac{FP}{FP+TN}$$

The term "True_Positive_Rate" refers to the percentage of "positive" records that the model accurately classified as "positive." The percentage of records that should have been interpreted as negative but were, in error, interpreted as positive is referred to as the "erroneous positive rate." (Real) Positive Having a negative mental attitude The actual positive evidence pertaining to (TP) (Predicted) (FP), the positive evidence concerning (TP) Invalid Positive False pessimism "Genuine Negative" is what "FN" stands for as an acronym.

By contrasting the proportion of true positives to the proportion of false positives at each of a number of threshold levels, the ROC curve is created. The ROC curve's shape is assessed using the AUC. The ability of a classifier to distinguish between two groups is shown by the 12 AUC graph. For datasets with an uneven distribution of class members, the AUC calculation is required. The bias in the data is 95 percent negative and 5 percent positive, making it unbalanced. We will talk about a well-known problem in machine learning in this section. The area under the curve (AUC) can be used to determine how well the model understands the uncommon and challenging positive class. Due to the fact that a classifier can achieve high accuracy by ignoring the sparse positive class, accuracy alone cannot determine a model's efficacy.

*3) F-measure:*

The F-measure is an indispensable evaluative metric. Recall and precision are two additional metrics included in Fmeasure. Keep in mind that the true positive rate is specified. In contrast, precision is determined according to Table 2.2:

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

The model produces relatively few errors in the positive class because almost all of the projected positive records will almost certainly be accurate and positive. Watch this. High-precision models can only find a small percentage of positive records because they are designed to minimize error to the greatest extent possible. A high recall shows that the model, despite making some mistakes in its attempt to include all potentially positive data, is able to correctly classify a significant number of positive records. A sizable portion of positive records are correctl categorised by a model with a high recall. These theories contend that recall and precision have an inverse relationship, with high recall implying low precision and low recall implying high precision. When both are required, high precision and recall can be combined into one metric to speed up the optimization process. It makes more sense to maximize one function as opposed to two or more.

The harmonic mean of recall and precision, or F-measure, is the measurement in question. Here, recall is referred from 'r' and 'p' is for precision.

Then,

$$\text{F\_measure} = \frac{2*r*p}{r+p}$$

It is critical to remember that a measure's F-value can only be considered high if its precision and recall are both high. This metric, in addition to the AUC, is useful for evaluating models that handle unbalanced datasets.

*E. Model Building:*

The current study proceeds through a multi-step classification process. We implement single learner models such as Logistic Regression classifier, SVM, along with the ensemble learning paradigm like Random Forest, Sstochastic Gradient Descent

and XGBoost. We even apply deep learning Neural Network algorithms (NN). The best predictive model is chosen by comparing all the models. Fig. 2 shows the workflow for the methodology that was used.

*F. Algorithms Used:*

*1) Logistic Regression:*

A classification algorithm will be used in this case to distinguish between customers who have left and those who have not. This is one of the most straightforward algorithms. Its values range from 0 to 1, depending on whether you use the sigmoid or logit function.

$$\partial(z) = \frac{1}{(1 + e - z)}$$

Here, $\partial$ is sigmoid function.

- z is function value.

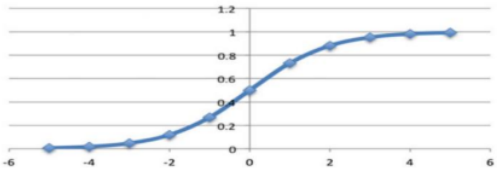
The following example illustrates the logistic function graphically: The application of logistic regression offers the following advantages:

- It is simple to update and rearrange the data to avoid overfitting (Figs. 1 and 2).
- Due to its adaptability, stochastic gradient descent is incredibly useful when adding new data, updating existing data, or updating existing data (in our case, when we will obtain a live data stream). One of the drawbacks of logistic regression is that it does not always function as expected.

Figure 4 Logistic Function

*2) Random_Forest*

In the realm of machine learning, random forest ranks highly. One example of an ensemble technique is bagging (bootstrap aggregation). Because random forests generate and combine multiple decision trees, they produce more accurate and trustworthy results than decision trees. A disadvantage of decision trees is that their predictions are highly interdependent. The predictions of each subtree are less correlated, which is an improvement for random forest. The




CART algorithm considers all variables and their associated values when figuring out the ideal data split point, in contrast to the random forest method, which can only search a randomly chosen subset of features. This makes it possible for the CART algorithm to choose the ideal data split point. The number of attributes that the algorithm must look at in order to find a split point is specified by a parameter (m).

For the purpose of classification, 'm' has been computed as follows:

$$m = \sqrt{p}$$

Calculating 'm' in the context of regression can be done as follows:

$$m = p/3$$

where 'p' is the total number of variables being input.

The following are some of the benefits:

- Random forests are useful in banking because of their ability to generate accurate and reliable forecasts as well as their robustness. It is used to distinguish between trustworthy and dangerous customers.
- Multiple decision trees averaged together significantly reduce overfitting. The greatest disadvantage of random forest is the difficulty in evaluating the model and determining the basis of the forecast.

*3) SVM:*

SVM is a classification-based algorithm that uses kernels to calculate the degree of separation between two variables. The goal of dividing two classes with a boundary or hyperplane is to minimize the distance between their closest value points. In our scenario, we distinguish customers who have left from those who have not. Maximum-margin This is the definition of a hyperplane. The following function is used to calculate the distance among 2 points using vector dot product:

$$S(x) = B_0 + \sum_{i=1}^{n} \left(a_i \times (x \times x_i)\right)$$

The primary advantages of using SVM are that it is resistant to overfitting, particularly when applied to high-density

datasets; in our case, the polynomial kernel was the best choice for the dataset. SVM may not perform as well with larger datasets due to its reliance on previously stored information of the support vectors.

*4) XGBoost:*

Unlike previous gradient boosting algorithm iterations, XGBoost employs a complex model to prevent overfitting and improve performance. XGBoost can be used for regression and ranking in addition to classifying data as "churn" or "no churn," as we have done in this article. XGBoost's feature selection process can be automated, and its randomization feature reduces tree correlation. The gradient boosting machine (GBM) approaches the optimization problem in two steps:

Step-1: Determine the first phase of action.

Step-2: While XGBoost uses a single equation to determine the step directly, the second stage optimizes the step length.

*5) Stochastic Gradient Descent:*

Stochastic Gradient Descent (SGD) is a popular optimization algorithm used in machine learning to train models, particularly when dealing with large datasets. It updates the model parameters on a subset of the data (mini-batch) instead of the entire dataset at once.

To predict customer churn using SGD, a binary classification model must first be defined that can predict the likelihood of a customer churning based on their characteristics (demographics, behavior, usage patterns, etc.). The output is a probability between 0 and 1, indicating the likelihood of churn.

A loss function, such as binary cross-entropy, should then be defined to measure the difference between the predicted churn probabilities and the actual churn labels.

The data must be divided randomly into training, validation, and test sets. The training set is used to train the model, with SGD updating the model's parameters after each mini-batch of data. The model's performance on the validation set is evaluated, and hyperparameters, including the learning rate, regularization strength, and mini-batch size, are adjusted. Finally, the model's performance on unknown data is evaluated using the test set.

*6) Neural Networks:*

Using neural networks, financial institutions can forecast the loss of customers. A neural network is made up of many layers of connected nodes that can all learn from the input data and make precise predictions.

The mathematical formula for a neural network is complex and depends on the specific architecture and activation functions used. However, the basic idea is that each node in the network receives inputs from the nodes in the previous layer, performs a mathematical transformation on these inputs, and passes the result on to the nodes in the next layer. During training, specific node weights and biases are changed to improve the agreement between predicted and actual outputs.

*a) Neural Network classifier with 1 Hidden layer - with Early Stopping:*

Early stopping is employed to prevent model overfitting by terminating the training process as soon as the model's performance on the validation set begins to decline. This method requires a validation set, which is a subset of the training data that is not used to train the model. Throughout training, the performance of the model on the validation set is evaluated at regular intervals; training is terminated if it fails to improve after a predetermined number of epochs.

*b) Neural Network Architecture with multiple Hidden layers*

The network architectures of deep neural networks (DNN) contain a large number of hidden layers. There could be one, two, or more hidden layers depending on the subject's complexity. Here's an illustration of a structure with three hidden levels:

- Input Layer: The input characteristics are stored in this layer. The number of nodes in this layer is proportional to the number of input features
- Hidden Layer 1: This layer contains the same number of nodes as the hidden units. Each node in this layer receives input from nodes below it and nonlinearly transforms it.
- Hidden Layer 2: Like Hidden Layer 1, this layer is made up of a fixed number of nodes that represent hidden units. Each node in this layer receives input from nodes below it and nonlinearly transforms it.
- Hidden Layer 3: Like Hidden Layer 2, this layer contains the same number of nodes as hidden units. Each node in this layer receives input from nodes below it and nonlinearly transforms it.
- Output Layer: The output nodes in this layer represent the projected class or value. The number of nodes in this stratum is determined by the nature of the problem being addressed.

To train the network, we use backpropagation to compute the gradients of the loss function with respect to the weights and biases, and update them using gradient descent or a variant such as Adam or RMSprop. Early stopping can also be used to prevent overfitting.

*c) Neural Network Architecture with Early Stopping*

A common neural network architecture with early stopping involves the following steps:

1. Input layer: The input layer receives the input data.
2. Hidden layers: The hidden layers perform nonlinear transformations on the input data to extract features and representations.
3. Output layer: The output layer produces the final predictions.
4. Early stopping: During the training process, a portion of the data is reserved for validation. The performance of the model on the validation data is monitored, and training is stopped when the validation performance stops improving.

## III. EXPERIMENTAL RESULTS AND DISCUSSIONS

After that, by comparing the data to the suggested models, the classification accuracy of churners and non-churners is evaluated. The test results indicate that only one customer may experience a defect while everyone else does not.

### A. BoxPlots

Box plots show outliers in a dataset. With the exception of "outliers" that have been mathematically identified and are depicted in a different colour, the whiskers in a box plot represent the remainder of the distribution, while the central rectangle represents the interquartile range.

Each continuous variable in the dataset gets a box plot, one for each level of the target variable (for example, one for customers who left and one for those who stayed). As a result, we can compare similar data points side by side to find outliers. Finding instances where the box plot strays from its whiskers is the next step.

These points are considered outliers and may need to be removed or treated in some way before modelling.

The boxplot above provides an in-depth understanding of the following characteristics:

- Credit Score: It is obvious that having a low credit score has little effect on customer loss or customer churn.
- Age: The bank might lose some of its more senior customers. It's conceivable that this is being done to keep younger workers.
- Balance: The bank is losing a significant number of customers with high account balances. This may have an impact on the bank's capitalization.
- Estimated Salary: It does not appear that Estimated Salary has a significant effect on customer churn.
- Number of Products: This factor does not appear to have a substantial impact on customer churn.
- Tenure: This example demonstrates how new hires and older customers are more likely to churn.

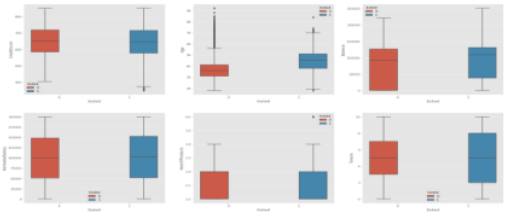


Figure 5 Box Plot Different Features

### B. ROC Curve

The ROC curve is an effective tool for displaying how well a classification model performs at different levels of categorization. Most frequently, the ratio of true positives to false positives is used to evaluate the affiliation among sensitivity as well as the specificity. If the AUC is equal to or greater than 0.80, it is assumed that a randomly selected churner (with DepVar=1) will perform better than a randomly selected non-churner (with DepVar=0). This remains true even if the area under the curve is not equal to or greater than 0.80. When there is a significant difference between the distributions of two groups, the area under the curve will be one. In other words, the test's accuracy increases as the graph approaches either its upper or lower limit, but decreases as it approaches the diagonal.

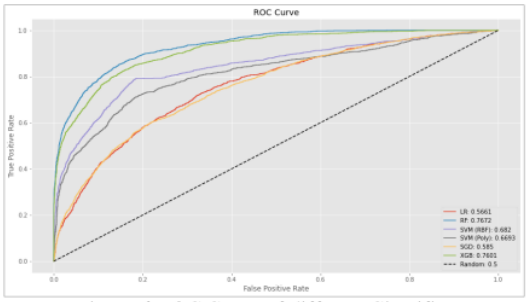



Figure 6 ROC Curve of different Classifiers

The Random Forest (RF) classifier has the highest ROC score (0.7705) and, as a result, the largest area under the curve, as depicted in the preceding graph.

From this, we finally choose the Random Forest classifier as our final machine learning classifier model.

On our test data, we apply the Random Forest Classifier and the results are depicted by the ROC curve below:

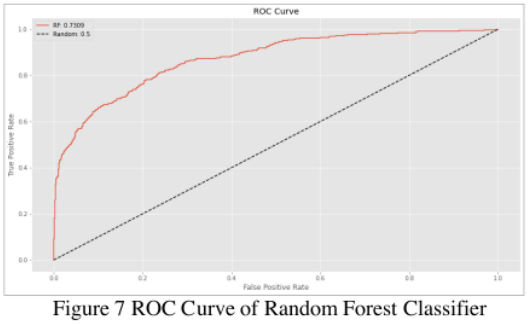



Figure 7 ROC Curve of Random Forest Classifier

On previously unreleased test data, the model's accuracy in predicting 1s, or customers who leave, is marginally higher. Despite its high accuracy, the model misses more than half of the people who ultimately churn.

This could be enhanced by continuously feeding the model with new training data and using the model to retain the 61 percent of data that would otherwise be lost.

## IV. DISCUSSION

To achieve the study's goal, a great deal of emphasis has been placed on applying and comparing all models. The primary focus of this study is the predictive model's performance. For the following reasons, the models are contrasted:

• According to Table 9, the XGBoost Classifier and Random Forest had accuracy of 88.81% and 89.53%, respectively. However, the combined accuracy of these two models has increased to 87%. A significant improvement was discovered when comparing the efficacy of different models.

| Models | Training Accuracy Score | Testing Accuracy Score |
|---|---|---|
| Logistic Regression | 88.00 % | 81.41 % |
| Random Forest | 94.42 % | 89.53 % |
| Support Vector Machine (SVM) (RBF kernel) | 90.00 % | 86.00 % |
| Support Vector Machine (SVM) (Poly kernel) | 91.53 % | 85.32 % |
| Stochastic Gradient Descent (SGD) | 90.47 % | 81.78 % |
| XGBoost Classifier | 96.45 % | 88.81 % |
| Neural Network classifier | 89.67 % | 84.35 % |
| Neural Network classifier with Early Stopping | 89.67 % | 84.35 % |
| Neural Network Architecture with multiple Hidden layers | 86.67 % | 86.20 % |
| Neural Network Architecture with Early Stopping | 86.65 % | 86.23 % |

Figure 8 Comparison of Accuracy of all the models

Precision calculations do not take into account the potential for false-positive or false-negative results. Metrics for accuracy and recall concentrate, respectively, on false positive and false negative results. The f1-score is strongly advised because it can be used for a variety of things. The proportion of correctly predicted positive observations to all predicted observations is referred to as the "precision" ratio. The Random Forest Classifier model has the highest precision score, achieving a 90% accuracy rate. because the model's 90% classification accuracy. Precision lowers the percentage of false-positive findings. The percentage of correctly predicted positive observations to all of the observations in a class is known as recall. The algorithm that comes closest to 98.00, the Random Forest Classifier, has the highest recall score.

| Models | Precision for Retained (0) | Precision for Exited (1) |
|---|---|---|
| Logistic Regression | 82.00 % | 69.00 % |
| Random Forest | 90.00 % | 89.00 % |
| Support Vector Machine (SVM) (RBF kernel) | 86.00 % | 84.00 % |
| Support Vector Machine (SVM) (Poly kernel) | 86.00 % | 81.00 % |
| Stochastic Gradient Descent (SGD) | 83.00 % | 67.00 % |
| XGBoost Classifier | 89.00 % | 85.00 % |
| Neural Network classifier | 87.00 % | 69.00 % |
| Neural Network classifier with Early Stopping | 87.00 % | 69.00 % |
| Neural Network Architecture with multiple Hidden layers | 88.00 % | 75.00 % |
| Neural Network Architecture with Early Stopping | 87.00 % | 83.00 % |

Figure 9 Comparison of Precision of all Models

- When the f1-scores of the constituent parts of the hybrid model were compared, the XGBoost model had a significantly lower f1-score than the Random Forest model. Both models have improved accuracy, according to XGBoost, with a f1-score of 93%. The Random Forest model also performed well, scoring 94% on the f1 test. In this case, it is critical to outperform Random Forest and XGBoost in terms of f1-score.
- It is critical to assess the significance of various characteristics in order to make the most accurate forecast. Relevant factors such as expected salary, credit card ownership, and others appear to be major impediments to customer retention strategies. It is critical to evaluate these factors in order to attract customers.

Figure 10 Comparison of Recall of all models

Figure 11 F1-Score for all the models

## V. CONCLUSION

In conclusion, churn prediction is crucial for businesses to develop customer retention strategies and retain loyal customers. False positive rates should be kept low to prevent giving promotions to customers who are not actually churning, which can lead to significant losses for the

| Models | Recall for Retained (0) | Recall for Exited (1) |
|---|---|---|
| Logistic Regression | 98.00 % | 15.00 % |
| Random Forest | 98.00 % | 55.00 % |
| Support Vector Machine (SVM) (RBF kernel) | 98.00 % | 38.00 % |
| Support Vector Machine (SVM) (Poly kernel) | 98.00 % | 36.00 % |
| Stochastic Gradient Descent (SGD) | 98.00 % | 19.00 % |
| XGBoost Classifier | 98.00 % | 55.00 % |

| Models | F1 Score for Retained (0) | F1 Score for Exited (1) |
|---|---|---|
| Logistic Regression | 89.00 % | 25.00 % |
| Random Forest | 94.00 % | 68.00 % |
| Support Vector Machine (SVM) (RBF kernel) | 92.00 % | 53.00 % |
| Support Vector Machine (SVM) (Poly kernel) | 91.00 % | 50.00 % |
| Stochastic Gradient Descent (SGD) | 90.00 % | 30.00 % |
| XGBoost Classifier | 93.00 % | 66.00 % |
| Neural Network classifier | 91.00 % | 53.00 % |
| Neural Network classifier with Early Stopping | 91.00 % | 53.00 % |
| Neural Network Architecture with multiple Hidden layers | 92.00 % | 59.00 % |
| Neural Network Architecture with Early Stopping | 92.00 % | 57.00 % |

business. Precision is the most significant criterion because of its relationship to the quantity of false-positive results. An important performance indicator is the accuracy score, which is determined by dividing the total number of observations by the number of accurate predictions. For accuracy, precision, recall, and F1 (89.53 percent, 90.00 percent, 98.00 percent, and 94.000 respectively), the Random Forest Classifier model received the highest scores.

## REFERENCES


[1] Khine, Thazin and Win Win Myo. "Customer Churn Analysis in Banking Sector." (2020).

[2] Chiang, D., Wang, Y., Lee, S., & Lin, C. (2003). Goal-oriented sequential pattern for network banking churn analysis, 25, 293–302. https://doi.org/10.1016/S0957-4174(03)00073-3

[3] Zhao, Y., Li, B., Li, X., Liu, W., & Ren, S. (2005). One-Class Support Vector Machine, 300–306

[4] B. He, Y. Shi, Q. Wan, X. Zhao, Prediction of customer attrition of commercial banks based on SVM model, "2nd International Conference on Information Technology and Quantitative Management, ITQM (2014)".

[5] A. P. Patil et.al, "Customer Churn Prediction for Retail Business", "(ICECDS-2017)".

[6] Fa-Gui LIU et.al. "Using Combined Model Approach for Churn Prediction in Telecommunication (FKP-SVM)", "3rd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2017)".

[7] A. O. Oyeniyi & A.B. Adeyemo, Customer Churn Analysis In Banking Sector Using Data Mining, "African Journal of Computing & ICT (2015)".

[8] F. Abdi and S. Abolmakarem, "Customer Behavior Mining Framework (CBMF) using clustering and classification techniques", "Journal of Industrial Engineering International (2018)".

[9] H. Sayed, Manal A. Abdel-Fattah, S. Kholief, Predicting Potential Banking Customer Churn using Apache Spark ML and MLlib Packages: a Comparative Study, "

International Journal of Advanced Computer Science and Applications, (IJACSA), Vol. 9, No. 11, (2018)".

[10] A. B. Zoric, Predicting Customer Churn in Banking Industry using Neural Network, " Interdisciplinary Description of Complex Systems 14(2), 116-124, (2016)".

[11] W. Etaiwi, M. Biltawi and G. Naymat, Evaluation of classification algorithms for banking customer's behaviour under Apache Spark Data Processing System, " The 4th International Symposium on Emerging Information, Communication and Networks, (EICN 2017)".

[12] S. Singhal and V. Passricha, K-means based SVM for Prediction Analysis, " International Journal of research in electronics and computer engineering, (IJRECE), Vol. 6 issue 3 (2018)".

[13] Albu, L. L., Lupu, R., Calin, A. C., & Popovici, O. C. (2014). Estimating the Impact of Quantitative Easing on Credit Risk through an ARMA-GARCH Model. Romanian Journal of Economic Forecasting, 17(3), 39-50.

[14] Bilal Zoric, A. (2016). Predicting Customer Churn in Banking Industry using Neural Networks. Interdisciplinary Description of Complex Systems, 14(2), 116–124. https://doi.org/10.7906/indecs.14.2.1

[15] Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: Partial defection of behaviorallyloyal clients in a non-contractual fmcg retail setting. European Journal of Operational Research, 164(1), 252–268.

[16] Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. Expert Systems with Applications, 36(3 PART 1), 4626–4636. https://doi.org/10.1016/j.eswa.2008.05.027

[17] Chandar, M., Laha, A., & Krishna, P. (2006). Modeling churn behavior of bank customers using predictive data mining techniques. In National conference on soft computing techniques for engineering applications (SCT-2006), March 24–26,2006.

[18] Chiang, D., Wang, Y., Lee, S., & Lin, C. (2003). Goal-oriented sequential pattern for network banking churn analysis, 25, 293–302. https://doi.org/10.1016/S0957-4174(03)00073-3

[19] Chitra, K., & Subashini, B. (2011). Customer Retention in Banking Sector using Predictive Data Mining Technique. The 5th International Conference on Information Technology, 4. Retrieved from http://icit.zuj.edu.jo/icit11/PaperList/Papers/E-Commerce/612_Chitra.pdf

[20] Coussement, K., & Poel, D. V. d. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter selection techniques. Expert Systems with Applications, 34, 313–327. Coussement, K., Lessmann, S., & Verstraeten, G. (2016).

[21] Here is the reference list in APA format:

[22] Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning - ICML '06, 233–240. https://doi.org/10.1145/1143844.1143874

[23] Dekimpe, M. G., & Degraeve, Z. (1997). The attrition of volunteers. European Journal of Operational Research, 98(1), 37–51.

[24] Gardner, M. ., & Dorling, S. . (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric Environment, 32(14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0

[25] Glady, N., Baesens, B., & Croux, C. (2008). Modeling churn using customer lifetime value. European Journal of Operational Research. doi:10.1016/j.ejor.2008.06.027.

[26] Gür Ali, Ö., & Aritürk, U. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. Expert Systems with Applications, 41(17), 7889–7903. https://doi.org/10.1016/j.eswa.2014.06.018

[27] Hayati, M., & Mohebi, Z. (2007). Application of Artificial Neural Networks for Temperature Forecasting, 22(July), 275–279.

[28] He, B., Shi, Y., Wan, Q., & Zhao, X. (2014). Prediction of customer attrition of commercial banks based on SVM model. In Procedia Computer Science (Vol. 31, pp. 423–430). https://doi.org/10.1016/j.procs.2014.05.286

[29] Hecht-Nielsen, R., 1990. Neurocomputing. Addison-Wesley, Reading, MA.

[30] Ismail, M. R., Awang, M. K., Rahman, M. N. A., & Makhtar, M. (2015). A multi-layer perceptron approach for customer churn prediction. International Journal of Multimedia and Ubiquitous Engineering, 10(7). Retrieved from http://www.sersc.org/journals/IJMUE/vol10_no7_2015/22.pdf

[31] Jinbo, S., Xiu, L., & Wenhuang, L. (2007). The Application of AdaBoost in Customer Churn Prediction, 0.

[32] Karahoca, A., & Karahoca, D. (2011). GSM churn management by using fuzzy cmeans clustering and adaptive neuro fuzzy inference system. Expert Systems with Applications, 38(3), 1814–1822.

[33] Kaur, M. (2013). Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers.

[34] Kumar, D. A., & Ravi, V. (2008). Predicting credit card customer churn in banks using data mining, 1(1), 4–28.

# Plagiarism Report

ORIGINALITY REPORT

| 10% | 7% | 5% | 6% |
| --- | --- | --- | --- |
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
| --- | --- | --- |
| 1 | www.mdpi.com<br>Internet Source | <1% |
| 2 | Submitted to Berlin School of Business and Innovation<br>Student Paper | <1% |
| 3 | Submitted to University of Zakho<br>Student Paper | <1% |
| 4 | Submitted to Technological University Dublin<br>Student Paper | <1% |
| 5 | Submitted to Deakin University<br>Student Paper | <1% |
| 6 | Submitted to University of Computer Studies<br>Student Paper | <1% |
| 7 | Submitted to The University of the West of Scotland<br>Student Paper | <1% |
| 8 | Submitted to University of Salford<br>Student Paper | <1% |
| 9 | Submitted to Aston University | |

| | | |
|---|---|---|
| | Student Paper | <1% |
| 10 | Submitted to National College of Ireland<br>Student Paper | <1% |
| 11 | Submitted to National School of Business Management NSBM, Sri Lanka<br>Student Paper | <1% |
| 12 | Submitted to London Metropolitan University<br>Student Paper | <1% |
| 13 | Owusu F. Aidoo, Philipe G.C. Souza, Ricardo S. Silva, Paulo A.S. Júnior et al. "Modeling climate change impacts on potential global distribution of Tamarixia radiata", Science of The Total Environment, 2022<br>Publication | <1% |
| 14 | William W. Clark. "Noise and Ultrasound", Wiley, 2001<br>Publication | <1% |
| 15 | scholarworks.uttyler.edu<br>Internet Source | <1% |
| 16 | Submitted to Liverpool John Moores University<br>Student Paper | <1% |
| 17 | core.ac.uk<br>Internet Source | <1% |
| | Submitted to CSU, Fullerton | |

| | | |
|---|---|---|
| 18 | Student Paper | <1% |
| 19 | Submitted to UNITEC Institute of Technology<br>Student Paper | <1% |
| 20 | Submitted to University of East London<br>Student Paper | <1% |
| 21 | Submitted to University of Northumbria at Newcastle<br>Student Paper | <1% |
| 22 | Submitted to University of Teesside<br>Student Paper | <1% |
| 23 | Waleed Salehi, Preety Baglat, Gaurav Gupta, Surbhi Bhatia Khan, Ahlam Almusharraf, Ali Alqahtani, Adarsh Kumar. "An Approach to Binary Classification of Alzheimer's Disease Using LSTM", Bioengineering, 2023<br>Publication | <1% |
| 24 | Submitted to Kirsehir Ahi Evran Universitesi<br>Student Paper | <1% |
| 25 | Xin Huang, Yuanping Zhu, Shuqin Wang. "An Efficient Dynamic Regulated Fuzzy Neural Network for Human Motion Retrieval and Analysis", Symmetry, 2021<br>Publication | <1% |
| 26 | assets.researchsquare.com<br>Internet Source | <1% |

| | | |
|---|---|---|
| 27 | www.ijraset.com<br>Internet Source | <1 % |
| 28 | Tewodros Megabiaw Tassew, Nie Xuan, Bosong Chai. "PDDS: A software for the early diagnosis of Parkinson's disease from MRI and DaT scan images using detection and segmentation algorithms", Biomedical Signal Processing and Control, 2023<br>Publication | <1 % |
| 29 | coek.info<br>Internet Source | <1 % |
| 30 | www.worldscientific.com<br>Internet Source | <1 % |
| 31 | "Recent Advances on Soft Computing and Data Mining", Springer Science and Business Media LLC, 2017<br>Publication | <1 % |
| 32 | Matthew Casey, Nianjun Zhou. "Chapter 42 Analysis of Viability of TCGA and GTEx Gene Expression for Gleason Grade Identification", Springer Science and Business Media LLC, 2020<br>Publication | <1 % |
| 33 | Submitted to University of Ulster<br>Student Paper | <1 % |
| 34 | bibliotecadigital.fgv.br<br>Internet Source | <1 % |

| | | |
|---|---|---|
| 35 | lib.ugent.be<br>Internet Source | <1 % |
| 36 | ujcontent.uj.ac.za<br>Internet Source | <1 % |
| 37 | www.iujharkhand.edu.in<br>Internet Source | <1 % |
| 38 | www.research-collection.ethz.ch<br>Internet Source | <1 % |
| 39 | Swati Prakash Gawand, Meesala Sudhir Kumar. "A Comparative Study of Cyber Attack Detection & Prediction Using Machine Learning Algorithms", Research Square Platform LLC, 2023<br>Publication | <1 % |
| 40 | iieta.org<br>Internet Source | <1 % |
| 41 | opus.lib.uts.edu.au<br>Internet Source | <1 % |
| 42 | pdfs.semanticscholar.org<br>Internet Source | <1 % |
| 43 | www.research.manchester.ac.uk<br>Internet Source | <1 % |
| 44 | M Sowmiya Manoj, M Sheik Misbar Ali, M Shyam Kumar, K Sudharsan, D Gokulavani, G Kalpanadevi. "A Novel Strategy to Segment | <1 % |

the Retinal Layers using Convolutional Neural Network", 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), 2023

Publication

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | Off |
| Exclude bibliography | On | | |