HAL open science

Predicting Customer Churn in Banking Based on Data Mining Techniques

Wafaa A. Alsubaie, Haya Z. Albishi, Khloud A. Aljoufi, Wedyan S. Alghamdi, Eyman A. Alyahyan

▶ **To cite this version:**

Wafaa A. Alsubaie, Haya Z. Albishi, Khloud A. Aljoufi, Wedyan S. Alghamdi, Eyman A. Alyahyan. Predicting Customer Churn in Banking Based on Data Mining Techniques. 4th International Conference on Computational Intelligence in Data Science (ICCIDS), Mar 2021, Chennai, India. pp.27-39, 10.1007/978-3-030-92600-7_3 . hal-03772945

HAL Id: hal-03772945

https://inria.hal.science/hal-03772945v1

Submitted on 8 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Predicting Customer Churn in Banking Based on Data Mining Techniques

Wafaa A. Alsubaie[0000-0002-9387-780X], Haya Z. Albishi[0000-0003-3727-6273], Khloud A. Aljoufi[0000-0001-6693-7574], Wedyan S. Alghamdi[0000-0001-6155-9869] and Eyman A. Alyahyan[1][0000-0002-9272-6129]
[1]eaalyahyan@iau.edu.sa

Computer Science Department, College of Science and Humanities, Imam Abdulrahman Bin Faisal University, P.O. Box 31961, Jubail, Kingdom of Saudi Arabia

Abstract. One of the most critical challenges facing banking institutions is customer churn, as it dramatically affects a bank's profits and reputation. Therefore, banks use customer churn forecasting methods when selecting the necessary measures to reduce the impact of this problem. This study applied data mining techniques to predict customer churn in the banking sector using three different classification algorithms, namely: decision tree (J48), random forest (RF), and neural network (MLP) using WEKA. The Results showed that J48 had an overall superior performance over the five performance measures, compared to other algorithms using the 10-fold cross-validation. Additionally, the InfoGain and correlation features selection method was used to identify significant features to predict customer churn. The experiment revealed that both algorithms work better when all features are utilized. In short, the results obtained help in predicting which customer is likely to leave the bank. Furthermore, identifying these essential features will help banks keep customers from churning out and compete with rival banks.
Keywords: Classification, Customer Churn, Churn prediction, Data Mining.

1 Introduction

The tremendous development in various aspects of life has led to an increase in the growth of data in its various forms, which has increased the need for smart tools and techniques to reduce the efforts required in the search for useful information. In turn, it has helped to extract new and useful knowledge, also known as data mining techniques [1]. The same is the case in the banking industry, as it requires these technologies to improve and support their decisions in all of their activities when historically analyzing the activities of their customers [2]. One of the biggest banking concerns is customer churn, and they are waging a fierce war to preserve and gain customer loyalty [3]. Customer churn is the loss of custom when individuals change from one provider to another. The problem of customer churn can be reduced by extracting knowledge from the data that helps classify customer behavior, identifying the most important features that affect customer churn, using data mining techniques, and uncovering hidden behaviors in data sets that may not have been visible

before. Previous studies have shown the significance of predicting customers churn at an early stage to avoid losing customers, as this can have a significant effect on a bank's profits and be costly. To the researchers' knowledge, studies that used the Kaggle data set to address a customer churn problem had not previously used Neural Network (MLP), Random Forest, and Decision Trees (J48) algorithms together and compared them in terms of which performs best. Therefore, in this study, a machine learning model was built to predict customer churn in the banking sector using Neural network (MLP), random forests, and decision trees (J48), taking into consideration the optimization strategies for all models. This study's results will undeniably be beneficial to those who wish to maintain customers and discover the most important features affecting their loyalty, allowing organizations to make the right decisions at the right time to keep customers. The remainder of this paper is as follows: Section 2 discusses the literature reviewed for this work; Section 3 provides a description of the proposed techniques; Section 4 focuses on Empirical studies; Section 5 presents the results and discussion; Section 6 features further discussion highlights; and, finally, Section 7 presents the conclusion.

2 Related Work

Many studies have used various techniques for data mining to make churn predictions. This literature focuses on discussing related work that used data mining methods to apply a model of prediction.

In [3], the authors examined the problem of customer churn in banks and found that due to the intense competition between banks, they resorted to looking for intelligent ways to help them make decisions to win and maintain their customers. Consequently, the researchers proposed a model to predict customer churn using the neural network algorithm in the Alyuda NeuroInteligence software package. Using a database consisting of 1,866 customers from a small Croatian bank, they put forward the assumption that customers who use more than three products are likely to be loyal customers, whereas those who use less could pose a risk of leaving. One of the most important results they discovered was that there is the greatest risk of customer churn is among young people (students) who use less than three products.

In [4], the authors addressed the issue of customer churn. The main applied technique used data mining to extract useful data by developing the decision tree algorithm, which allows branch managers to identify who the most likely customers to leave are. They applied their study based on a dataset of random samples, consisting of 4,383 customers using electronic banking services. The technique highlights the characteristics of customers, thus enabling branch managers to introduce more marketing tools to retain them. However, the study could be more useful if it was supported by proven techniques that can be used to retain customers.

In [5], the authors discussed the problem of customer churn, which creates major issues for both enterprises and service providers, especially telecommunication companies. The research conducted a comparative study of churn prediction performance among two machine learning algorithms: K-nearest neighbor and decision tree. Their

efficiency and performance were contrasted against the prediction issue churn and used performance evaluation criteria for the algorithms by precision, F-measure, recall, and accuracy. To reduce the noise and eliminate undesired information, filters processed the input data of each of the algorithms. The cleaned data were subsequently divided into training and testing sets. Training sets were then modeled to produce the desired output using algorithms. The algorithm for the K-nearest neighbor was set to the standard number of k=5 neighbors. The standard Gini index criterion, which has a maximum depth of a size equal to 20, was used for the decision tree algorithm. There were 3,333 samples (customers) of 20 separate variables in the dataset that were used to compare all telecommunications firms' algorithms. It was found that the main outcomes were 92.6% more accurate than the K-nearest neighbor algorithm in terms of performance assessment criteria for the algorithms in the decision tree.

In [6], the authors discussed the churn that hinders the number of profitable customers from increasing, and it is the biggest challenge to sustain a telecommunication network. The key proposal included two models with a high degree of precision, and estimated customer churn. The logistic regression model was the first model, and its accuracy was improved by modifying it with the regularization parameter set to 0.01. The second model was the multilayer neural network (MLP). The dataset contained about four thousand lines. The main finding was 87.52% accurate using the logistic regression model, and it was 94.19% accurate using the neural network model. The two models' accuracy ratios were appropriate. However, some of the features were irrelevant in affecting the prediction results.

In [7], the authors considered the problem of rapid growth in Software as a Service (SaaS) companies, taking into account the existence of one source of income for the company, specifically the monthly fee for each customer. The customer behavior here needed to be analyzed to predict the factors that contribute to positive change and prevent customers from disconnection, which is essential and necessary. For companies to survive, they suggested building a predictive model using a time series perspective instead of using the traditional method of taking the variables accumulated from the moment the customer joined the company. They suggested four major algorithms: principal component analysis (PCA), logistic regression, random forest, and Extreme Gradient Boosting (XGBoost) trees using a traditional classification algorithm. They applied these algorithms to a set of data from multiple sources: the company's Microsoft SQL server, billing system, and Customer Relationship Management (CRM) platform, which contained 8,047 observations of 21 variables. The most important results came from the highest performance algorithms, namely, XGBoost and the logistic regression of 0.7257 and 0.7526, therefore they adopted the model XGBoost as the final model.

In [8], the authors explored customer churn in e-retail, building a model to make predictions using a data mining approach. They studied the problem in the context of a research community in North America of an E-retailer, investigating the characteristics of customers who reached 0.5 million in number. Furthermore, the attribute was classified into seven different categories: customer information, customer sales, demographic features, frequency, product sales, behavior, and experience. The number of all of the attributes was 35. Hadoop stack tools and classification were used with the following algorithms: logistic regression with L1 regularization, SVM, and gradient boost.

The results revealed the significant impact of the following features: click/blogging, marketing campaign, customer behavior, and experience in customer momentum, to predict customers' churn. Besides, they found that the best prediction model was the GBM, with an accuracy of 75%.

In conclusion, previous studies have shown many differences in the field of application, such as the field of banking [3][4], communications [5][6], companies [7], and e-retailers [8]. However, their goal was the same, namely, to build a model to predict customer churn using a range of different algorithms. Of these algorithms, the ones that showed the best results were the decision tree algorithm, which outperformed the bank with a rate of 99.7%, followed by the neural network algorithm, with a rate of 94.19% in the field of communications. Therefore, these results motivated us to use these two algorithms in our study.

3 Description of Proposed Techniques

3.1 Decision Tree (J48)

The decision tree is one of the supervised classification algorithms and decision analysis tools. It uses a model in the form of a tree that includes potential outcomes. Each branch represents one of the options, and each one may also subdivide into other branches of future possibilities. It helps evaluate the choice between many options available and, in turn, make the best decision. It has four different types, which are: ID3, C45, C5.0, CART[9] .

3.2 Multilayer perceptron (MLP)

A Multilayer Perceptron (MLP) is a feed-forward artificial neural network model that contains one input layer, one output layer with at least one hidden layer. MLP utilizes a supervised learning technique called backpropagation for training. ANNs significance is to mainly solve three problems which are: classification, noise reduction, and extrapolation. Besides, MLP networks are one of the widest architectures with too many applications that can solve a lot of problems in various aspects of the knowledge areas; and the most significant areas are curve fitting, pattern recognition, process identification, and control time series forecasting, and system optimization.[10][9].

3.3 Random Forest (RF)

Random Forest is a machine learning algorithm that can be used for various tasks, including regression and classification[11]. It is a mixture of tree predictors, with each tree depending on independently sampled random vector values with the same distribution of all the trees in the forest. The random forest model combines predictions of estimators to produce a more accurate prediction. This algorithm requires two parameters: the ideal number of trees and the ideal depth of trees. The advantage RF presents variable significance estimates. They also give a superior technique for dealing with

the missing data. The missing values are replaced by the variable that occurs most frequently in a specific node. RFs have the highest precision of all the classification methods available. With multiple variables running into thousands, the RF technique can also control big data. When a class is more infrequent than other data classes, it automatically balances information sets. In addition, the method also easily handles variables, making it ideal for complicated tasks.

4 Empirical Studies

4.1 Description of the Dataset

The data set contained 14 attributes and 10,000 instances and was extracted from a bank customers' data set using Kaggle, last updated was two years ago. The target variable was a binary variable reflecting whether the customer had left the bank (closed his/her account) or continued to be. Several pre-processes were followed to prepare the data, as shown in Figure No.1, where they were cleaned by deleting the unimportant attributes (customer number, customer name, and row number) and removing outliers. The dataset is imbalance; the class (No) the number of instances is 7,963, and the class (Yes) 2,037 instances. In this study, an over-sampling method using SMOTE (the technique of over-sampling of a synthetic minority), a widely used and available method of Weka as a supervised instance filter to address this problem. After the cleaning process, a transformation process was performed for some attributes. The target attribute (Exited) was converted from a numerical attribute to a categorical attribute, using an unsupervised filter (NumericToNominal). Lastly, the J48, RF, and neural network (MLP) algorithms were applied.

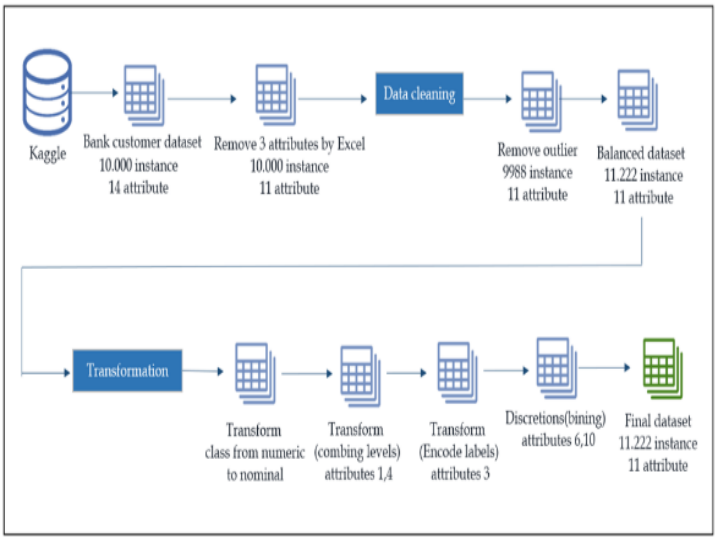

Fig. 1. Preparing Dataset Process

4.2 The Experimental Setup

Weka, a free software written under the JAVA General Public License that provides a collection of machine learning algorithms, was used in the study experiments.[13]. Initially, the dataset was processed to prepare for the experiment. Three attributes were removed from the data using Excel, and then outliers were detected and removed using two filters (Interquartile Range and RemoveWith-Values). The issue of data imbalance was resolved in the quest to optimize classification, considering the cost of errors and overfitting issues, which can also lead to suboptimal results due to the high costs associated with misclassification of the minority class. The SMOTE algorithm (as supervised instance filter) in Weka was used to balance the data, which resulted in more synthetic instances for the minority class "1". After balancing the data, the total number of cases is11222. The transformation step was critical to improving the model's performance and making the features easier to understand. The class value was transformed from a numeric value to a nominal value by WEKA using an unsupervised attribute filter (NumericToNominal). It was then combined with two attributes, and the labels for one attribute were encoded. Additionally, by binning the data, two numeric attributes were discretized; this step was completed using an unsupervised attribute filter (Discretize). Furthermore, the optimized parameters of J48, RF, and MLP were determined by setting the confidence factor of J48 and the minimum number of the object parameter. The RF number of the tree and the seed parameter were reset. In addition to adjusting MLP seed, learning rate, and the hidden layers parameter. As a consequence, the J48 performed higher, with an accuracy of 83.55 % compared to 82.39 % for the MLP, and the RF performed at 81.87 %. Moreover, correlation coefficients and information gain between the class variable and each attribute were calculated in order to rank the attributes of the selected features. The findings are shown in Tables 3 and 4. Following that, the ability to enhance the accuracy of rating efficiency has been investigated by selecting features and defining significant attributes in order to forecast consumer churn. The classifiers were evaluated on the selected attributes using the the10-fold cross-validation and optimum parameters for each classifier. Table 5 and Table 6 illustrate the results. With these results, different partition ratios were used to implement the classifiers. The highest accuracy for J48, MLP, and RF was found to be achieved with a (70:30) ratio, 70 percent for training data and 30 percent for testing data. Table 7 summarizes the findings. Finally, the final studies have been performed using the optimum options for the best subset features to achieve the best results in terms of cross-validation or partition ratio outcome.

4.3 Performance Measures

In this study, to produce more accurate results, four performance measures were considered to evaluate each classifier: accuracy, precision, recall, and f-measure. All of these measures are based on the following possibilities:

- TP: True Positive is the total of instances that a churn customer was correctly classified.

- FP: False positive is the total of instances that a churn customer was incorrectly classified.
- TN: True Negative is the total of instances that a non_churn customer was correctly classified.
- FN: False negative is the total of instances that a non_churn customer was incorrectly classified.

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

$$Recal = \frac{TP}{TP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$f - meas = \frac{2 \times precision \times recall}{precision \times recall} \quad (4)$$

5 Strategy of Optimization

In seeking to improve the classification results, the Weka meta-learner (CV Parameter Selection) search methodology was used to obtain better performance based on accuracy.[14]. Table 1 shows the default and optimum parameters for each classifier. Table 2 shows the comparison between the findings associated with optimal parameters and that of their default counterparts. It can be noticed that the algorithms accuracy increases when the optimum parameters are used in comparison to the default parameters.

Table 1. Classifiers Parameters: Default and Optimal

Model	Values of parameters		
	Parameters	**Default**	**Optimal**
J48	Confidence Factor	0.25	0.2
	Minimum number of objects	2	4
MLP	Seed	0	0
	Hidden Layers	A	a
	Learning Rate	0.3	0.116
RF	NumIterations (number of trees)	100	50
	Seed	1	32

Table 2. Classifiers Performance: Default and Optimal Parameters

Model	Performance (Accuracy)	
	Default value	**Optimal value**
J48	83.34%	83.55%
MLP	80.46%	82.39%
RF	81.66%	81.87%

6 Results and Discussion

6.1 The Impact of Features Selection on the Dataset

The Information Gain (InfoGain) and correlation-based features selection method were used to select the best performing subset, along with the most significant attributes with the highest impact on prediction of customer churn. The correlation coefficient was used to rank the attributes based on the Pearson values, from highest to the lowest variable relationship with the class variable (output), as shown in Table 3. In addition,

InfoGain was applied to classify the features based on the class information gain measure, as shown in Table 4. The backward selection method begins with building a model with all of the features required in their order of significance. The least important feature is removed. Then the next model is built using the remaining features. Features are continually removed, and models continue to be built until one feature remains. The results of the Info Gain and correlation-based features selection method are presented in Table 5 and Table 6. It was observed that the best performance required the use of all of the features (10 features). Furthermore, accuracy diminished when the number of factors was reduced.

Table 3. The correlation of each attribute and the target

Rank	Attributes	Correlation
1	Gender	0.1858
2	IsActiveMember	0.1804
3	Geography	0.1314
4	Balance	0.0812
5	NumOfProducts	0.0586
6	CreditScore	0.0197
7	Tenure	0.0137
8	Age	0.0102
9	EstimatedSalary	0.0070
10	HasCrCard	0.0003

Table 4. The information gain of each attribute and the target

Rank	Attributes	InfoGain
1	NumOfProducts	0.2075
2	Tenure	0.1813
3	IsActiveMember	0.1256
4	HasCrCard	0.0862
5	Geography	0.0332
6	Gender	0.0251
7	Balance	0.0185
8	Age	0.0026
9	CreditScore	0.0014
10	EstimatedSalary	0.0004

Table 5. Correlation-Based Feature Selection Results

Number of features	J48	MLP	RF	AVG (%)
10	83.55	82.39	81.87	82.60
9	83.38	81.78	81.09	82.08
8	83.42	81.63	80.14	81.73
7	82.97	81.58	79.66	81.41
6	83.07	81.92	82.41	82.47
5	83.24	80.86	83.57	82.56
4	77.78	77.13	78.08	77.66
3	77.07	76.43	77.06	76.85
2	76.20	70.75	76.20	74.38
1	70.95	70.95	70.95	70.95

Table 6. InfoGain Feature Selection Results

Number of features	J48	MLP	RF	AVG (%)
10	83.55	82.39	81.87	82.60
9	83.51	82.48	80.40	82.13
8	83.50	81.51	81.25	82.09
7	83.09	80.93	81.17	81.73
6	81.90	81.11	82.57	81.86
5	82.22	80.25	83.34	81.94
4	81.63	76.84	83.27	80.58
3	80.71	75.36	83.27	79.78
2	80.81	77.31	83.19	80.44
1	78.19	70.95	78.20	73.37

6.2 The Impact of different partition ratios on the dataset

After identifying the best features, it became clear that all of the features are important in both InfoGain and the correlation-based features selection method. The performance of each classifier was evaluated by performing many experiments on the data using different partition ratios ranging from 50 to 80. The results of the direct partition of each classifier are shown in Table 7.

Table 7. Results of Different Partition Ratios

Partition ratio	The Performance			
	J48	**MLP**	**RF**	**AVG**
50:50	83.1937%	80.0036%	78.2392%	80.4788%
60:40	83.2479%	80.2629%	79.1045%	80.8717%
70:30	83.4868%	80.1901%	79.4179%	81.0316%
80:20	82.3975%	80.8378%	77.6292%	80.2881%

6.3 The Comparison Between 10-Fold Cross Validation with Direct Partition Techniques

When comparing the two validation methods (10-fold cross-validation and direct partition ratio), all of the classifiers that used 10-fold cross-validation gained a higher accuracy, as shown in Table 8.

Table 8. 10-fold cross validation comparison with Direct Partition techniques

Techniques	Proposed model			
	J48	**MLP**	**RF**	**AVG**
10-fold validation	83.5502%	82.3917%	81.8749%	82.6056%
Partition ratio	83.4868%	80.1901%	79.4179%	81.0316%

7 Further Discussion

The final customer churn prediction model was built using all of the features with the optimum parameters achieved, as noted in Table 9. Using the 10-fold cross-validation method, the researchers of this study were able to produce perfect results for each of the following classifiers: RF, J48, and MLP. The J48 outperformed MLP and RF in predicting customer churn with an accuracy of 83.0957%. The classification performance was increased by choosing all the features (10 features) using the optimal criteria for each classifier, as shown in Figure 2. Through this process, we were able to identify important features that had a significant impact on the ability to predict customer churn in the banking sector, specifically: Credit Score, Geography, Gender, Age, Tenure, Balance, Number of products, Has Credit Card, Is an Active Member, and Estimated Salary. Identifying essential features can help prevent customers from leaving and f competition from rival banks. The Receiver Operating Characteristic (ROC) curve is another indicator of how the model of classification performs. The proximity and placement of this curve to the left-hand side (on the top) indicate that the experiment's accuracy is high. Overall, the area under the curve for each classifier as noted in Figure 3, shows that the most suitable classifier is determined to be J48 compared to MLP and

RF. According to table 9, J48 performed higher compared to the rest of the algorithms across all of the measures. Confusion matrices offer another look at actual and predicted classes by J48, MLP, and RF in Tables 10, 11, 12, respectively. The most important measure to check in the confusion matrices below is the rate of False Negative (FN), as an increase in its rate adversely affects the bank's profits. The lowest FN rate was using RF, J48, and MLP, respectively.

Table 9. The Performance of Proposed Model.

Proposed Model	**J48**				**RF**				**MLP**			
Performance measures	ACC%	PREC	REC	F-M	ACC%	PREC	REC	F-M	ACC%	PREC	REC	F-M
Results of final model	83.55	0.578	0.826	0.836	81.87	0.813	0.819	0.813	82.39	0.548	0.815	0.824

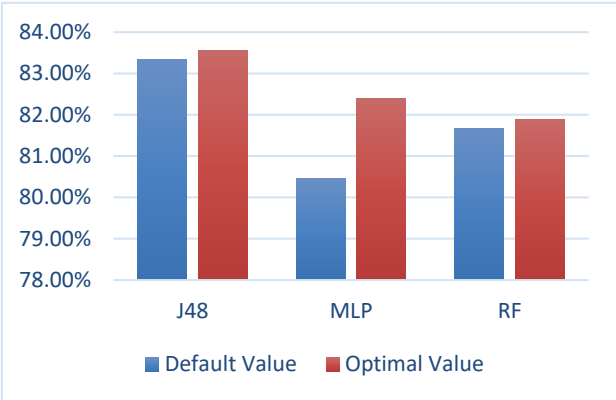

Fig. 2. The result of default and optimized model.

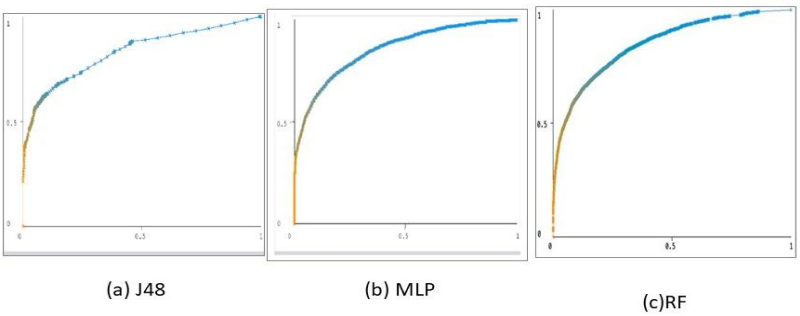

Fig. 3. The ROC curve for Yes (1) class for each classifier.

Table 10. Confusion Matrix for J48

Actual Class	Predicted Class	
	Yes (1)	No (0)
Yes (1)	1843 (TP)	1416 (FN)
No (0)	430 (FP)	7533 (TN)

Table 11. Confusion Matrix for MLP

Actual Class	Predicted Class	
	Yes (1)	No (0)
Yes (1)	1413 (TP)	1846 (FN)
No (0)	7400 (FP)	563 (TN)

Table 12. Confusion Matrix for RF

Actual Class	Predicted Class	
	Yes (1)	No (0)
Yes (1)	1967 (TP)	1292 (FN)
No (0)	742 (FP)	7221 (TN)

8 Conclusion and Recommendation

In conclusion, the customer churn of bank customers is one of the biggest causes of losses for banks, therefore maintaining existing customers is essential. In this study, one bank's database was analyzed using data mining techniques, specifically classification. Three algorithms were also applied to classify decision tree J48, MLP, and RF. Two techniques, namely cross-validation, and the partition ratio were applied and then their results were compared. We found that the accuracy of the J48 with technique cross-validation is better than MLP and RF. The algorithm J48 outperformed the rest with a classification accuracy of 83.55%. We used and compared several precision measures, namely precision, recall, and f-measure, and they outperformed the algorithm J48.

References

1. S. Agarwal, “Data mining: Data mining concepts and techniques,” in *2013 International Conference on Machine Intelligence and Research Advancement*, 2013, pp. 203–207, doi: 10.1109/ICMIRA.2013.45.
2. T. Dai, “International trade e-commerce based on data mining,” *Proc. - 2014 IEEE Work. Adv. Res. Technol. Ind. Appl. WARTIA 2014*, pp. 703–705, 2014, doi: 10.1109/WARTIA.2014.6976362.
3. A. Bilal Zoric, “Predicting customer churn in banking industry using neural networks,” *Interdiscip. Descr. Complex Syst.*, vol. 14, no. 2, pp. 116–124, 2016, doi: 10.7906/indecs.14.2.1.

4. A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, “Developing a prediction model for customer churn from electronic banking services using data mining,” *Financ. Innov.*, vol. 2, no. 1, 2016, doi: 10.1186/s40854-016-0029-6.
5. V. Lazarov and M. Capota, “Churn Prediction: A Comparative Study Using KNN and Decision Trees,” *2019 Sixth HCT Inf. Technol. Trends*, no. 1, pp. 182–186, 2019.
6. K. Kim and J. H. Lee, “Bayesian Optimization of Customer Churn Predictive Model,” *2018 Jt. 10th Int. Conf. Soft Comput. Intell. Syst. 19th Int. Symp. Adv. Intell. Syst.*, pp. 85–88, 2018, doi: 10.1109/SCIS-ISIS.2018.00024.
7. Y. Ge, S. He, J. Xiong, and D. E. Brown, “Customer Churn Analysis for a Software-as-a-service Company,” *2017 Syst. Inf. Eng. Des. Symp. SIEDS 2017*, pp. 106–111, 2017, doi: 10.1109/SIEDS.2017.7937698.
8. K. B. Subramanya and A. K. Somani, “Enhanced feature mining and classifier models to predict customer churn for an E-retailer,” *Big Data Anal. Tools Technol. Eff. Plan.*, pp. 293–309, 2017, doi: 10.1201/b21822.
9. J. Han, M. Kamber, and J. Pei, *Data mining concepts and techniques third edition*. 2011.
10. J. M. López-Gil, J. Virgili-Gomá, R. Gil, and R. García, “Method for improving EEG based emotion recognition by combining it with synchronized biometric and eye tracking technologies in a non-invasive and low cost way,” *Front. Comput. Neurosci.*, vol. 10, no. AUG, Aug. 2016, doi: 10.3389/fncom.2016.00085.
11. Y. L. Pavlov, “Random forests,” *Random For.*, pp. 1–122, 2019, doi: 10.1201/9780429469275-8.
12. “What Is a Good Credit Score? - Experian.” https://www.experian.com/blogs/ask-experian/credit-education/score-basics/what-is-a-good-credit-score/ (accessed Jan. 08, 2021).
13. S. Daw and R. Basak, “Machine Learning Applications Using Waikato Environment for Knowledge Analysis,” *Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020*, no. February, pp. 346–351, 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-00065.
14. D. Tufi, “PROCEEDINGS OF THE 10 TH INTERNATIONAL CONFERENCE ‘ LINGUISTIC RESOURCES AND TOOLS FOR PROCESSING THE ROMANI AN LANGUAGE ’ 18-19 SEPTEMBER 2014,” no. September, 2014.