❏   539

Predicting customers churning in banking industry: A machine learning approach

Amgad Muneer[1], Rao Faizan Ali[1], Amal Alghamdi[2], Shakirah Mohd Taib[1], Ahmed Almaghthawi[2], Ebrahim Abdulwasea Abdullah Ghaleb[1]

[1]Department of Computer and Information Sciences, Faculty of Science and Information Technology, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia
[2]Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering University of Jeddah, Jeddah, Saudi Arabia

ABSTRACT

In this era, machines can understand human activities and their meanings. We can utilize this ability of machines in various fields or applications. One specific field of interest is a prediction of churning customers in any industry. Prediction of churning customers is the state of art approach which predicts which customer is near to leave the services of the specific bank. We can use this approach in any big organization that is very conscious about their customers. However, this study aims to develop a model that offers a meaningful churn prediction for the banking industry. For this purpose, we develop a customer churn prediction approach with the three intelligent models random forest (RF), AdaBoost, and support vector machine (SVM). This approach achieves the best result when the synthetic minority oversampling technique (SMOTE) is applied to overcome the unbalanced dataset and the combination of undersampling and oversampling. The method on SMOTED data has produced excellent results with a 91.90 F1 score and overall accuracy of 88.7% using RF. Furthermore, the experimental results show that RF yielded good results for the full feature-selected datasets.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Amgad Muneer
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS
32610 Seri Iskandar, Malaysia
Email: muneeramgad@gmail.com

1. INTRODUCTION

Every day there is much competition growing in the banking industry [1]. Thus, if any bank wants to increase its market share by acquiring new customers, it must follow customer retention strategies. It is shown that improving the retention rate by up to 5% can increase a bank's profit by up to 85% [2]. Different banks offer attractive plans like internet banking, mobile banking, debit card, credit card, savings accounts with nil balance, credit points based on the usage of the customers [3], best plans for various loans like education loan, housing loan, agricultural loan, vehicle loan, mortgage loan, and startups loan. In the group of all these facilities or plans, crediting a loan to a customer is a critical task because, in this case, each bank has to analyze the customer's capacity prior to offering that loan [4]. To complete the crediting loan process to customers, there are a number of banks that have decided to incorporate a credit card scheme that will ensure that whenever a customer applies for a credit card, his or her ability to avail of the card will be evaluated. Many banks initiate the request for providing credit cards to new customers based on their credit points [5]. However, there will be multiple opportunities for clients to churn out of a particular bank for every customer

who has more than one credit card with more than one bank [4], [6], [7]. Whenever a customer realizes that Bank A offers many facilities at a low-interest rate compared to Bank B, the customer churning prediction for Bank B is high. Therefore, it is the bank credit card account management system responsibility to ensure that the existing customers are maintained through low interest rates. Churn analysis algorithms currently exist, but they are limited by the nature of the churn prediction problem. These three features are typically associated with this problem: i) The data is imbalanced; for example, the number of churn customers represents a tiny fraction of the total samples (usually 2% of the total samples); ii) Data from large learning applications will inherently contain noise; and iii) To predict churn, it is necessary to rank subscribers according to their likelihood to churn [8], [9]. Nowadays, with the intense machine learning advancement, it is beneficial to build a prediction approach that able to predict whether a credit cardholder or a customer will churn out from a particular bank or not [4]. This prediction will be possible on previously available data collected from the old customers history records. Machine learning (ML) methods like Naive Bayes, decision trees, logistic regression, random forest, artificial neural networks, and support vector machines will determine the churn [10]. All these ML techniques are implemented not only in the banking field but also applied in various sectors like insurance [11], medical systems [12], cyberbullying [13], retail marketing [14], automobile industry, gaming industry [12]. Therefore, the contribution of this study summarizes in threefold; i) We collect credit card churn customer data of around 10,000 from Kaggle repository; ii) We have conducted an exploratory data analysis (EDA) at the first stage based on available data and employ the hybridization of SMOTE data sampling and random forest classifier to overcome inherent class imbalance problem; iii) At the final stage of model selection and evaluation, we have implemented three models (random forest (RF), AdaBoost, support vector machine (SVM)) and we have performed a detailed comparison between model results.

The remainder of this paper is organized as shown in: Section 2 discusses the background of the study and its related research. Research methodology is outlined in section 3, while experimental findings are presented in section 4. Finally, section 5 concludes the paper by describing future directions.

2.   LITERATURE REVIEW

Many data mining techniques can research credit card churn prediction systems. Related work of available methods is listed out here briefly. For example, according to Dias *et al*., [15] have predicted in advance whether a given customer will end his relationship with an organization or not. They use six different methods using machine learning like the random forest, support vector machine, logistic regression, multivariate adaptive regression splines, classification and regression techniques, and stochastic boosting applied on the retail banking customer churn prediction problem, considering predictions up to 6 months in advance. The best results are concluded from the stochastic boosting data mining technique. According to Dalmia *et al*. [16] have used a supervised machine learning technique, a proprietary algorithm has been created to predict and inform the bank about the customers at the highest risk of leaving the bank. Different classifiers are able to achieve different accuracies with different datasets. K-nearest neighbour (KNN) is a groundbreaking new approach based on weighted scales and the XGBooster algorithm for high and improved accuracy. The dataset is appropriately grouped into training and testing models based on weighted scales and the KNN algorithm. According Gholamiangonabadi *et al*. [17] proposed a study to find customer churn predictions of an Iranian bank; they introduced a new procedural approach. First, they normalize their data using data pre-processing. Then, a data cluster is formed by using a k-medoids method. The Davies-Bouldin index is used to assess clustering performance. Various neural network (NN) approaches were utilized in order to discover patterns within the data, including radial basis function (RBFNN), generalized regression (GRNN), multilayer perceptron (MLPNN), and SVM. According to the results, MLPNN and SVM models had higher precisions and lower costs. According to Ahmad *et al*. [18] have proposed three machine learning techniques to be applied to predict churn, namely, Decision trees (DT), Naive Bayes, SVM, using two benchmark datasets IBM Watson dataset, which contains 7033 observations, 21 attributes, and cell2cell dataset that contains 71,047 observations and 57 attributes. Therefore, data unbalanced is one of the key drawbacks of the aforementioned works.

The performance of the models has been measured using the area under the curve (AUC), which they scored 0.82, 0.87, 0.77 respectively for the IBM dataset and 0.98, 0.99, 0.98 for the cell2cell dataset. In [18], [19] the authors focus on applying data mining techniques in telecommunications to predict the churning behaviour of customers. In this research work, they use the CART algorithm to predict customer churning. In [20] research, they have built a computer system based on the application of artificial neural networks (ANN) and SVM approaches. According to the model, there are three different states of customers: active (*i.e.,* those that are fully engaged in business with a positive balance in their account), non-active (*i.e.,*

those with low balances in their accounts and those who do not have any investments), and churning (closed bank account). They have demonstrated excellent results with their computer software [21].

3. RESEARCH METHOD
3.1. Data collection and description

This section describes the methods used to predict customer churning within the banking industry, explain the dataset and the proposed approach utilized. The dataset used for the prediction process task is publicly available on the Kaggle website [22]. The variables included in the dataset are listed in Table 1. Of the 23 variables, the last two columns should be removed since they do not contribute to the classification process. Removing the last two columns from the dataset now contains 21 variables, 20 predictor variables, and one class variable. It contains 10,127 records, of which 8,496 (83.9%) are non-churners and 1,630 (16.1%) are churners. Therefore, the dataset is highly unbalanced in terms of the proportion of churners and non-churners. Furthermore, we conducted an exploratory data analysis to determine the percentages between genders, age groups, and so on. Before inputting the data to the classifier, it is necessary to balance the data so that the classifiers do not tend towards the majority class consisting of non-churners while predicting the future. A mixture of synthetic minority oversampling techniques (SMOTE), undersampling, and oversampling is used to achieve the balancing.

Table 1. The Description of the data

Variable	Description	Value
CLIENTNUM	Client number. Unique identifier for the customer holding the account	Positive real number
Attrition_Flag	Internal event (customer activity) variable	if the account is closed, then 1 else 0
Customer_Age	Demographic variable	Customer's Age in Years
Gender	Demographic variable	M=Male, F=Female
Dependent_count	Demographic variable	Number of dependents
Education_Level	Demographic variable	Educational Qualification of the account holder
Marital_Status	Demographic variable	Married, Single, Divorced, Unknown
Income_Category	Demographic variable	Annual Income Category of the account holder (< $40K, $40K - 60K, $60K - $80K, $80K-$120K, > $120K, Unknown)
Card_Category	Product variable	Type of Card (Blue, Silver, Gold, Platinum)
Months_on_book	Timespan	Period of relationship with the bank
Total_Relationship_Count	Product variable	Total no. of products held by the customer
Months_Inactive_12_mon	Timespan	No. of months inactive in the last 12 months
Contacts_Count_12_mon	Contact variable	No. of Contacts in the last 12 months
Credit_Limit	Credit variable	Credit Limit on the Credit Card
Total_Revolving_Bal	Credit variable	Total Revolving Balance on the Credit Card
Avg_Open_To_Buy	Open to Buy Credit Line	Average of last 12 months
Total_Amt_Chng_Q4_Q1	Change in Transaction Amount	Q4 over Q1
Total_Trans_Amt	Total Transaction Amount	Total Transaction Amount (Last 12 months)
Total_Trans_Ct	Total Transaction Count	Total Transaction Count (Last 12 months)

3.2. Exploratory data analysis

In machine learning, exploratory data analysis (EDA) is the process of analysing datasets in order to summarize their main characteristics. Data analysis is used to determine what can be learned from the data before modelling is performed [23]. It is very difficult to determine important data characteristics by reviewing a column of numbers or a whole spreadsheet. Figure 1 illustrates the distribution of customer ages as shown in Figure 1(a), and illustrates the distribution of customers for a given month as shown in Figure 1(b). Figure 2 shows the distribution of credit limits as shown in Figure 2(a), Figure 2(b) shows the distribution of total transaction amounts in the last year. Lastly, Figures 3 represent the percentage of churned and non-churned customers as shown in Figure 3(a) and the number of inactive months in Figure 3(b). The following steps will use SMOTE to up sample the churn samples in order to make them comparable with the regular customer sample size so the later selected models have a better chance of detecting small details that would be lost otherwise.

Distribution of Customer Ages

(a)

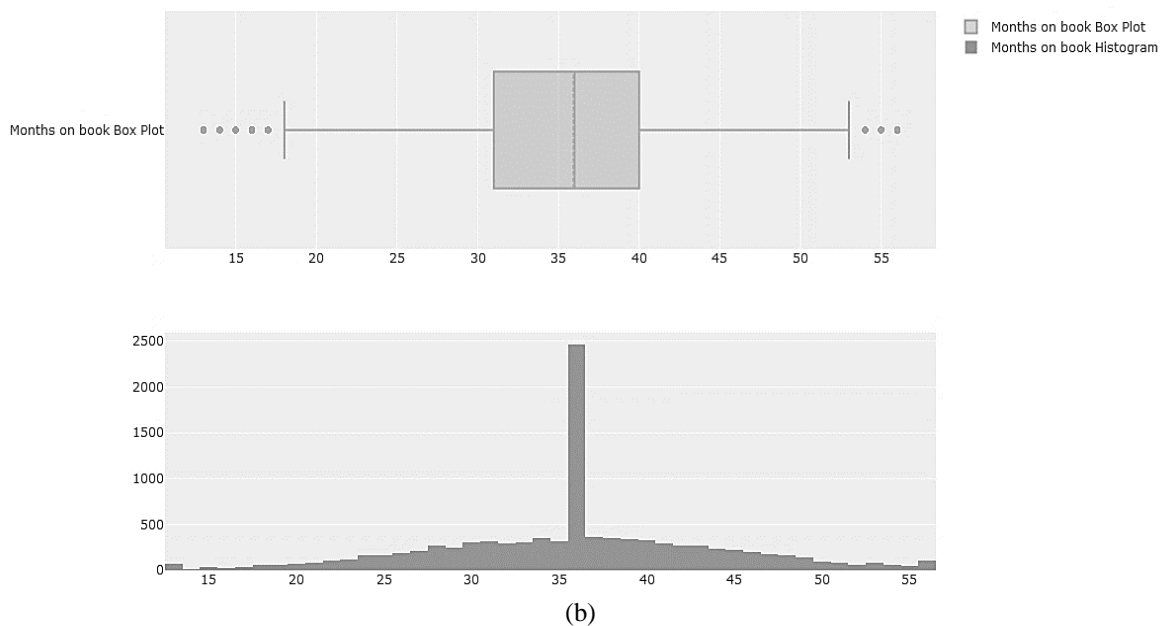Distribution of months the customer is part of the bank

(b)
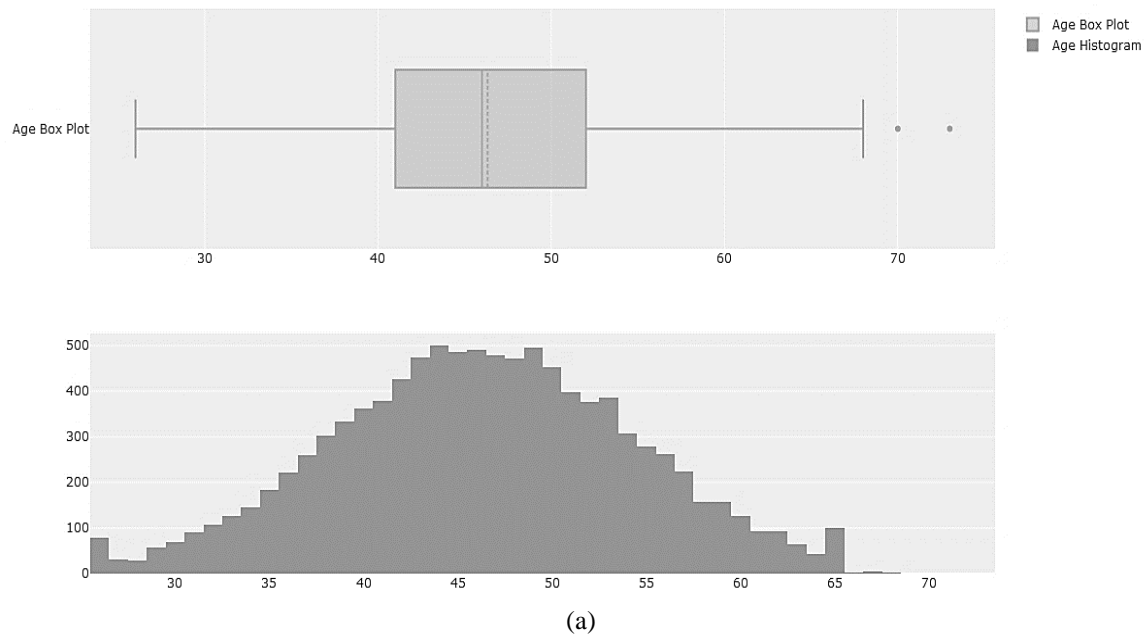
Figure 1. Illustration of (a) distribution of customer age and (b) Distribution of months the customer is part of the bank

3.3. Data pre-processing

This section pre-processed the data before introducing it to our proposed model. In the first instance, we modified the values of our class variable (Attrition_Flag). This column contains two values. The "Attrition Customer" value is changed from "1" to "0" while the "Existing Customer" value remains unchanged. The gender column is then modified. Female is replaced with 1, and male is replaced with 0. Finally, there are some Unknown values in Education_Level, Income_Category, and Marital_Status. These values have been eliminated from our dataset.

3.4. Data upsampling using SMOTE

The synthetic minority oversampling technique (SMOTE) can be described as a statistical technique. This technique aims to increase the number of cases in our dataset in a balanced manner. We generate new instances from our existing minority cases to feed our model. In this way, new instances are not simply copies of existing minority cases; instead, the algorithm takes a sample of the feature space for each target class and its nearest neighbours and creates new examples that combine features of the target case and those of its neighbours. The new approach increases the number of features available to each class and makes the samples more general. In order to increase the percentage of minority cases that are not attrited customers to twice the rate of majority cases, we use SMOTE.

(a)

(b)
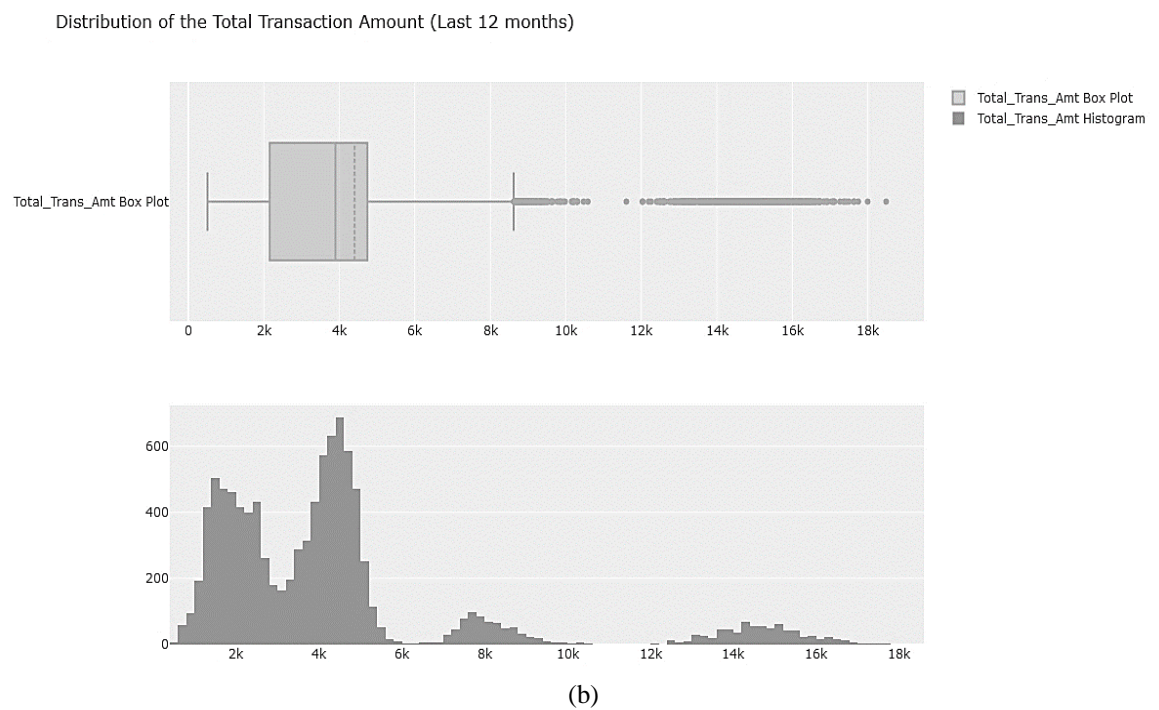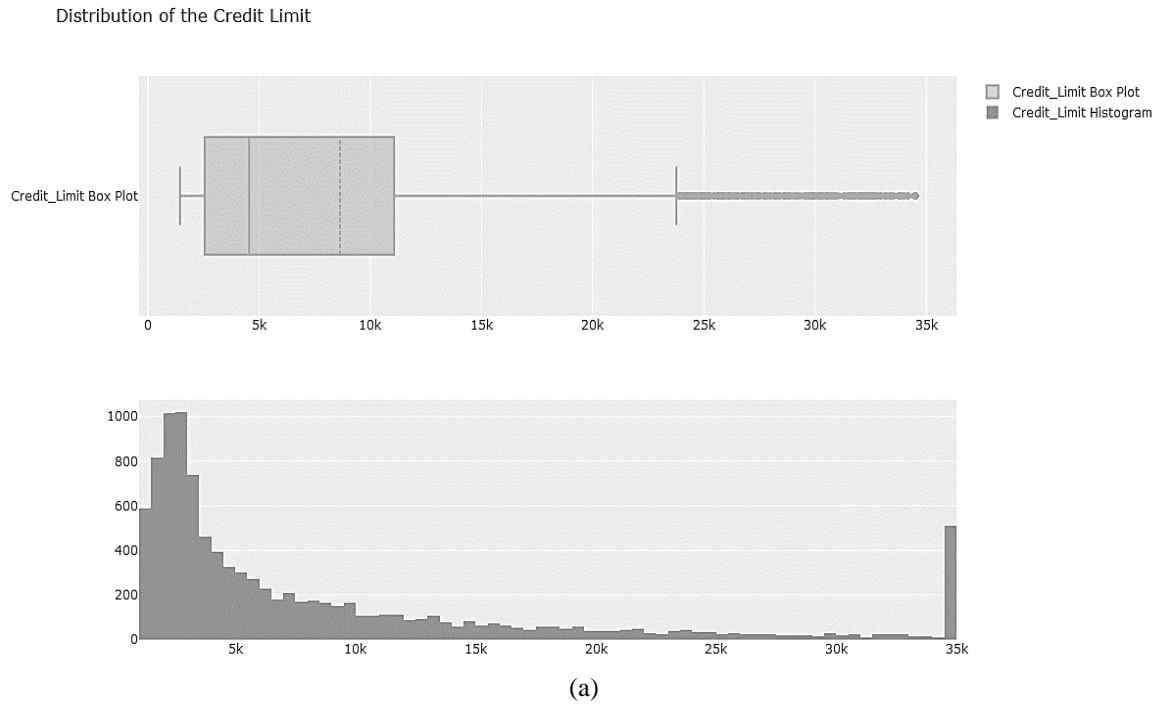
Figure 2. Illustration of (a) Distribution of the credit limit and (b) Distribution of total transaction amount

3.5. Proposed models employed in the prediction

The Random Forest method developed by Breiman and Cutler creates several classification trees. In order to classify a new object from an input vector, it must put the input vector down each tree in the forest. Every tree has a classification, and we say that its 'votes' for that classification. A forest selects the classification that has received the most votes (over all the trees in the forest).

The SVM classifies data by creating an N-dimensional hyperplane that divides it into two groups. The fundamental goal of SVM modelling is to find an ideal hyperplane that divides data in such a way that samples belonging to one category of the target variable are on one side of the plane and samples belonging to the other category are on the other side [13]. AdaBoost is one of the first boosting algorithms to be adapted to solver practices. Adaboost combines multiple "weak classifiers" into a single "strong classifier" [13].

(a)

(b)
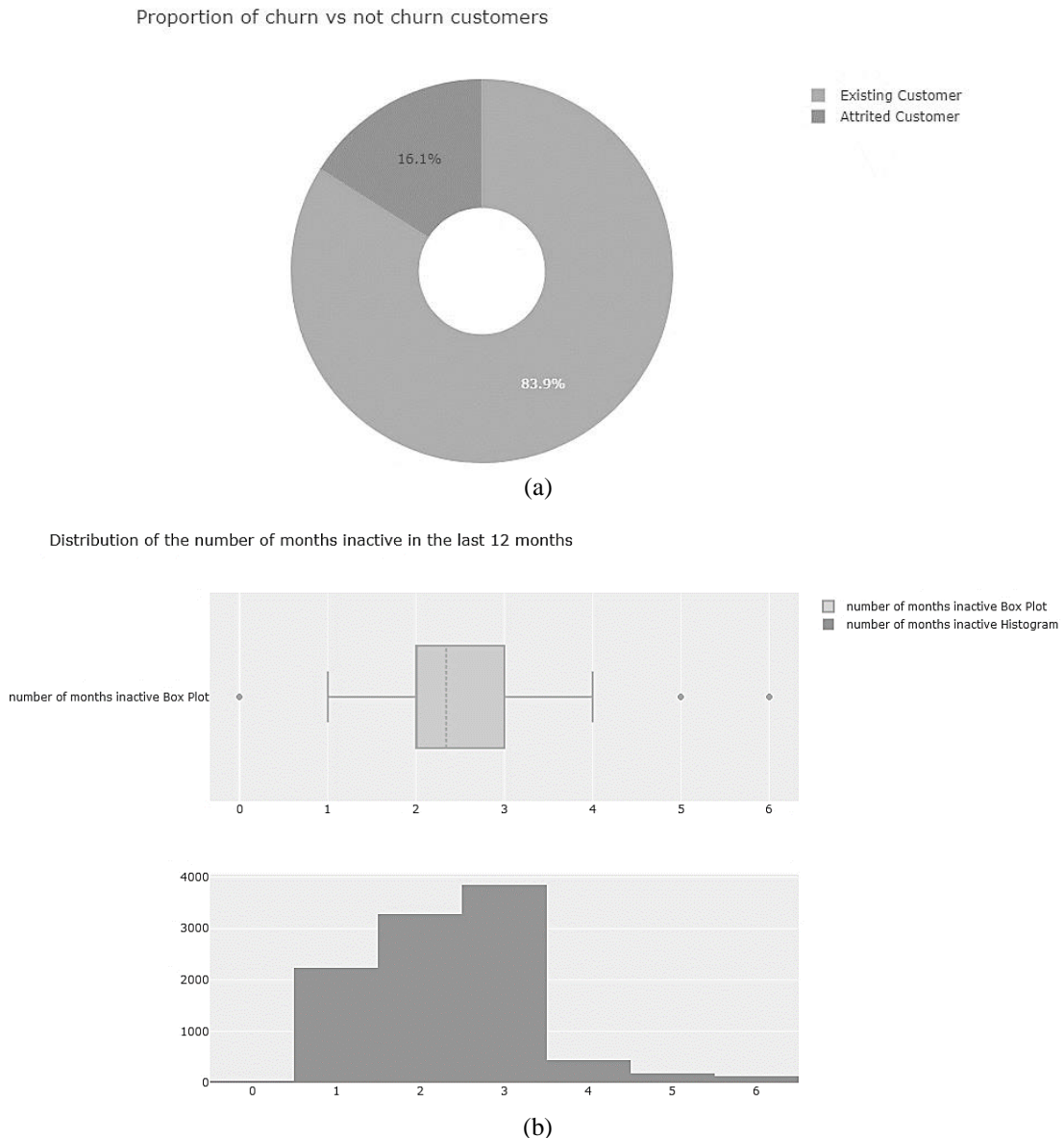
Figure 3. The results of (a) Proportion of churn vs does not churn customers and
(b) Number of inactive months

4.   RESULTS AND DISCUSSION

In the following section, we discuss the results obtained from the experiments conducted in this study. Firstly, we introduce a well-known evaluation measure to evaluate the performance and effectiveness

of the proposed classifiers. Secondly, we show the 5-corss validation and then we described the experimental results obtained in this study. Finally, the comparative analysis was provided to provide the readers a clear comparison between the proposed classifiers in this study and the state of the art.

4.1. Evaluation measures

To evaluate the effectiveness of our classifier, we used four well-known evaluation matrices since our data is balanced. These mectrics with their mathematical represntaion and difnation are discussed in this section. These metrics are as given in the follows;

4.1.1. Accuracy

Accuracy is a ratio of the true detected cases to the total cases, and it has been utilized to evaluate models on a balanced dataset [24]. Accordingly, it can be calculated as (1):

$$Accuracy = \frac{(tp+tn)}{(tp+fp+tn+fn)} \qquad (1)$$

where *tp* means true positive, *tn* is true negative, *fp* denotes false positive, and *fn* is a false negative.

4.1.2. Recall and F1-score

Recall: calculates the ratio of retrieved relevant churns over the total number of a relevant customer churning [25]. F1-score allows combining both precisions and recall into a single measure that captures both properties.

$$Recall = \frac{tp}{(tp + fn)} \qquad (2)$$

$$F\text{-}measure = \frac{2 \times precision \times recall}{recision + recall} \qquad (3)$$

4.2. 5-Fold cross-validation

We have conducted a 5-fold cross-validation of our three models. The F1 validation score for the random forest is higher than that of the Adaboost and SVM models. Figure 3 shows the performance evaluation using F1.
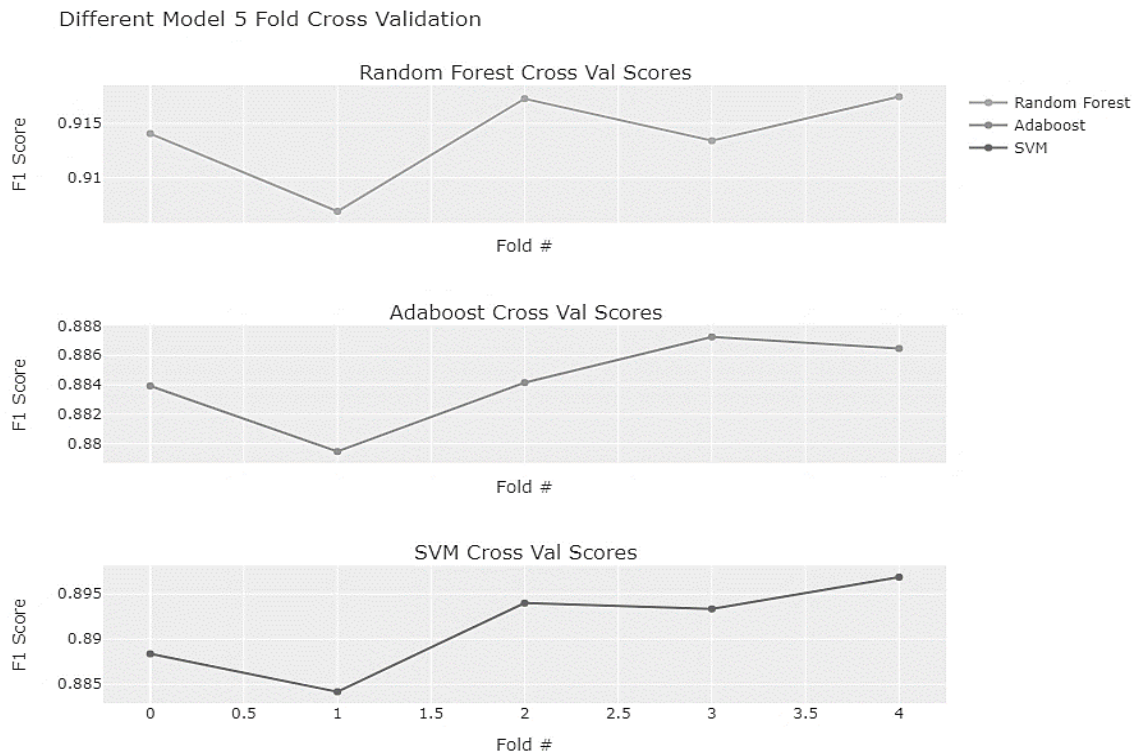

Figure 3. Performance evaluation for three proposed models using F1-score metrics

4.3. Proposed models experimental results

Table 2 presents the results of the three models proposed in this research. The results shown in Table 2 are based on upsampling the original data (SMOTE). Random forest outperforms both AdaBoost and SVM classifiers with an F1-score of 0.91 and an accuracy of 88.7. The SVM classifier has achieved the highest recall (1.00), whereas AdaBoost has achieved the lowest recall (0.87). Additionally, the proposed models were tested and evaluated using the original data before applying the SMOTE technique. These results are presented in Table 3.

Table 2. The performance of proposed three models with SMOTE technique

Proposed Model	Recall	F1 Score	Accuracy
Random Forest	0.89	0.91	0.887%
AdaBoost	0.87	0.88	0.872%
SVM	1.00	0.89	0.776%

Table 3. The performance of proposed three models on original data before applying SMOTE

Model	Recall	F1 Score	Accuracy
Random Forest	0.64	0.63	0.637%
AdaBoost	0.62	0.57	0.622%
SVM	0.75	0.55	0.562%

Table 2 and Table 3 show that the results based on random forest models are significantly higher than those based on other models. As a result, we selected the random forest model to forecast customer churning in the banking industry. The results of this prediction are presented in Figure 4.
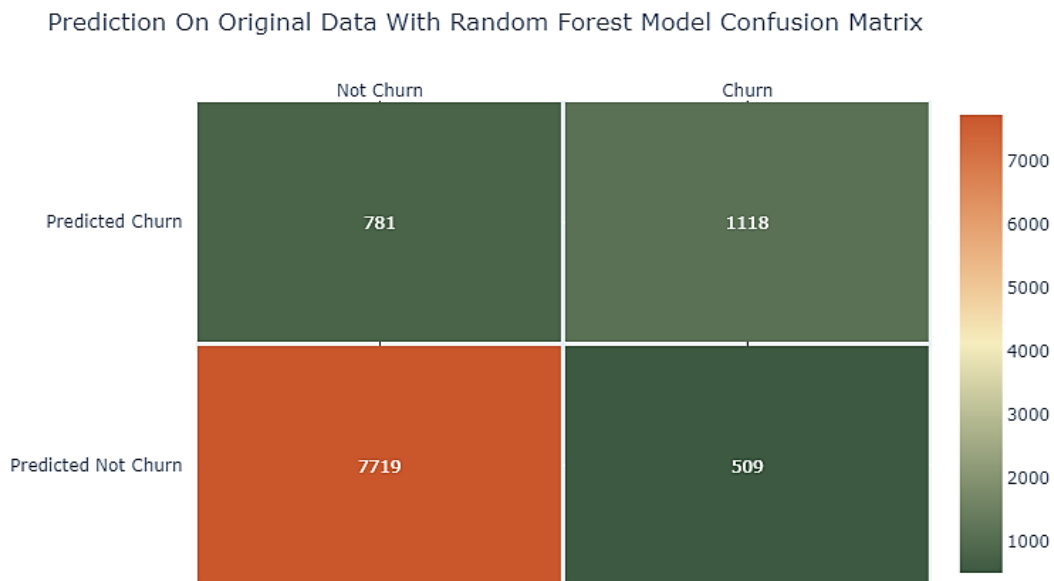

Figure 4. Confusion matrix for random forest prediction on the original data

4.4. Comparison with literature

This section compares the proposed three classifiers with the state-of-the-art methods. Several methods have been used to predict customer churn in the banking industry, including KNN, XGBoost, SVM, Naive Bayes, Decision Trees, ANN, and RF. In Table 4, we compare three proposed models with related literature contributions. The comparison is limited to the available metrics, but it essentially provides the reader with the promising results of the proposed RF predictor. Our results demonstrate that the proposed method surpasses the previous six methods for predicting customer churning in the banking industry.

Table 4. Comparison of the proposed models with related literature contributions

Prediction Model	Recall	F1 Score	Accuracy %
Proposed RF Predictor	0.89	0.91	88.7
Proposed AdaBoost Predictor	0.87	0.88	87.2
Proposed SVM Predictor	1.00	0.89	77.6
KNN [16]	Not Reported	Not Reported	83.85
XGBoost [16]	Not Reported	Not Reported	86.85
Naïve Bayes [26]	0.280	0.394	82.4
Decision Trees [26]	0.423	0.561	86.5
Random Forest [26]	0.474	0.588	86.4
ANN [26]	0.464	0.587	86.7

5. CONCLUSION

The proposed study conducted the most comprehensive investigation of the credit card churn prediction problem in banks using machine learning techniques. We proposed a customer churn prediction system with Random Forest, AdaBoost, and SVM intelligent models. The best results are achieved when the unbalanced original data is SMOTED and undersampling is combined with oversampling. When the SMOTE technique was applied to overcome the class imbalances in the data, the results revealed that RF outperformed the other two predictors with an accuracy of 88.7% and an F1 score of 0.91. The experimental results also demonstrated that RF performed well for the full feature-selected datasets. Accordingly, the proposed RF predictor can be used to calculate customer churn periodically from various perspectives. Churning can be measured in terms of the number of customers lost, the ratio of customers lost, or the percentage of customers lost compared to the total number of customers in the bank. This churning can be measured quarterly or annually. An accurate forecast provides insight into the future, which allows for developing a strategy. Lastly, in future work, we seek to implement a deep learning model in order to improve the accuracy of the proposed study.

REFERENCES

[1] I. Japparova and R. Rupeika-Apoga, "Banking business models of the digital future: The case of Latvia," *European Research Studies Journal*, vol. 20, no. 3, pp. 864–878, 2017, doi: 10.35808/ersj/749.
[2] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, "Credit card churn forecasting by logistic regression and decision tree," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15273–15285, Nov. 2011, doi: 10.1016/j.eswa.2011.06.028.
[3] R. Goel, S. Sahai, A. Vinaik, and V. Garg, "Moving from cash to cashless economy: A study of consumer perception towards digital transactions," *International Journal of Recent Technology and Engineering*, vol. 8, no. 1, pp. 1220–1226, Jun. 2019, doi: 10.17492/pragati.v7i1.195425.
[4] R. Rajamohamed and J. Manokaran, "Improved credit card churn prediction based on rough clustering and supervised learning techniques," *Cluster Computing*, vol. 21, no. 1, pp. 65–77, Mar. 2018, doi: 10.1007/s10586-017-0933-1.
[5] L. Bursztyn, B. Ferman, S. Fiorin, M. Kanz, and G. Rao, "Status Goods: Experimental evidence from platinum credit cards," *Quarterly Journal of Economics*, vol. 133, no. 3, pp. 1561–1595, Aug. 2018, doi: 10.1093/QJE/QJX048.
[6] H, Jain, G. Yadav, and R. Manoov. "Churn prediction and retention in banking, telecom and IT sectors using machine learning techniques." *Advances in Machine Learning and Computational Intelligence*. Springer, Singapore, 2021. 137-156.
[7] G. G. Sundarkumar and V. Ravi, "A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance," *Engineering Applications of Artificial Intelligence*, vol. 37, pp. 368–377, Jan. 2015, doi: 10.1016/j.engappai.2014.09.019.
[8] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, Apr. 2009, doi: 10.1016/j.eswa.2008.06.121.
[9] K. G. M. Karvana, S. Yazid, A. Syalim, and P. Mursanto, "Customer churn analysis and prediction using data mining models in banking industry," in *2019 International Workshop on Big Data and Information Security, IWBIS 2019*, Oct. 2019, pp. 33–38, doi: 10.1109/IWBIS.2019.8935884.
[10] M. A. H. Farquad, V. Ravi, and S. B. Raju, "Data mining using rules extracted from SVM: An application to churn prediction in bank credit cards," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5908 LNAI, 2009, pp. 390–397.
[11] N. A. Akbar, A. Sunyoto, M. Rudyanto Arief, and W. Caesarendra, "Improvement of decision tree classifier accuracy for healthcare insurance fraud prediction by using Extreme Gradient Boosting algorithm," in *Proceedings-2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020*, Nov. 2020, pp. 110–114, doi: 10.1109/ICIMCIS51567.2020.9354286.
[12] S. M. Fati, A. Muneer, N. A. Akbar, and S. M. Taib, "A continuous cuffless blood pressure estimation using tree-based pipeline optimization tool," *Symmetry,* vol. 13, no. 4, 2021, doi: 10.3390/sym13040686.
[13] A. Muneer and S. M. Fati, "A comparative analysis of machine learning techniques for cyberbullying detection on twitter," *Future Internet*, vol. 12, no. 11, pp. 1–21, Oct. 2020, doi: 10.3390/fi12110187.
[14] M. Al-Ghobari, A. Muneer, and S. M. Fati, "Location-aware personalized traveler recommender system (lapta) using collaborative filtering knn," *Computers, Materials and Continua*, vol. 69, no. 2, pp. 1553–1570, 2021, doi: 10.32604/cmc.2021.016348.
[15] J. Dias, P. Godinho, and P. Torres, "Machine learning for customer churn prediction in retail banking," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12251 LNCS, 2020, pp. 576–589.

[16] H. Dalmia, C. V. S. S. Nikil, and S. Kumar, "Churning of bank customers using supervised learning," in *Lecture Notes in Networks and Systems*, vol. 107, 2020, pp. 681–691.

[17] D. Gholamiangonabadi, S. Nakhodchi, A. Jalalimanesh, and A. Shahi, "Customer churn prediction using a meta-classifier approach; A case study of Iranian banking industry," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 2019, vol. 2019, no. MAR, pp. 364–375.

[18] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *Journal of Big Data*, vol. 6, no. 1, p. 28, Dec. 2019, doi: 10.1186/s40537-019-0191-6.

[19] V. K. Nijhawan, M. Madan, and M. Dave, "An analytical implementation of CART Using RStudio for Churn Prediction," *Information and Communication Technology for Competitive Strategies*, vol. 40. Springer Singapore, 2019.

[20] S. Osowski and L. Sierenski, "Prediction of customer status in corporate banking using neural networks," in *Proceedings of the International Joint Conference on Neural Networks*, Jul. 2020, pp. 1–6, doi: 10.1109/IJCNN48605.2020.9206693.

[21] K. Ebrah and S. Elnasir, "Churn prediction using machine learning and recommendations plans for telecoms," *Journal of Computer and Communications*, vol. 07, no. 11, pp. 33–53, 2019, doi: 10.4236/jcc.2019.711003.

[22] Churn for Bank Customers. (2020). Accessed: 21 March 2021. [Online]. Available: https://www.kaggle.com/mathchi/churn-for-bank-customers

[23] A. Omar and A. Almaghthawi, "Towards an integrated model of data governance and integration for the implementation of digital transformation processes in the Saudi Universities," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, pp. 588–593, 2020, doi: 10.14569/IJACSA.2020.0110873.

[24] S. Naseer, S. M. Fati, A. Muneer, and R. F. Ali, "iAceS-Deep: Sequence-based identification of acetyl serine sites in proteins using PseAAC and deep neural representations," *IEEE Access*, vol. 10, pp. 12953–12965, 2022, doi: 10.1109/access.2022.3144226.

[25] A. Muneer and S. M. Fati, "Efficient and automated herbs classification approach based on shape and texture features using deep learning," *IEEE Access*, vol. 8, pp. 196747–196764, 2020, doi: 10.1109/ACCESS.2020.3034033.

[26] S. E. Charandabi, "Prediction of Customer Churn in Banking Industry," *Age*, vol. 18, no. 92, pp. 38–92, 2020.

BIOGRAPHIES OF AUTHORS

Amgad Muneer 🆔 🔍 SC Ⓟ received the B.Eng. degree (Hons.) in mechatronic engineering from the Asia Pacific University of Technology and Innovation (APU), Malaysia, in 2018. He is currently pursuing the master's degree in information technology with Universiti Teknologi PETRONAS, Malaysia. He has authored several ISI and Scopus journal articles/conference papers. He is currently working as a Research Officer with the Department of Computer and information Sciences, University Technology Petronas, Perak, Malaysia. His research interests include machine and deep learning, image processing, the Internet of Things, computer vision, and condition monitoring. He is a Reviewer in some international impact-factor journals, and he has published more than 30 scientific publications. He can be contacted at email: muneeramgad@gmail.com.

Rao Faizan Ali 🆔 🔍 SC Ⓟ received the bachelor's degree in computer science from COMSATS University Islamabad, Pakistan, and the M.Phil. degree in computer science from the University of Management and Technology, Lahore, Pakistan. He is currently pursuing the Ph.D. degree with University Technology PETRONAS, Malaysia. He has eight years of experience in teaching and research. He has been with various computer science positions in financial, consulting, academia, and government sectors. He is currently working as a Research Officer with the Department of Computer and information Sciences, University Technology Petronas, Perak, Malaysia. He can be contacted at email: rao_16001107@utp.edu.my.

Amal Alghamdi 🆔 🔍 SC Ⓟ Currently, she is a master student in computer science and artificial intelligence at Jeddah University. She received her bachelor's degree in Computer Science from the Al-Baha University in 2014. her interests in Artificial intelligence, machine learning and deep learning. She can be contacted at email: dr.amal.alghamdi@gmail.com.

Shakirah Mohd Taib (ID) (g) (SC) (P) is a lecturer and researcher at Centre for Research in Data Science (CeRDaS) in Universiti Teknologi PETRONAS (UTP), Malaysia. She obtained a bachelor's degree in information technology from Universiti Utara Malaysia and Master of Computing from University of Tasmania, Australia. She has more than 15 years working experience at Universiti Teknologi Petronas (UTP). Her area of specialization includes data science, machine learning, knowledge discovery and information retrieval using Artificial Intelligence techniques. Shakirah is a member of international organization such as IEEE, Malaysia Board of Technologists (MBOT) and Association for Information Systems (AIS). She can be contacted at email: shakita@utp.edu.my.

Ahmad Almaghthawi (ID) (g) (SC) (P) received his bachelor's degree in Computer Science from Taibah University in 2015. He has a master's degree in the program computer science and artificial intelligence at Jeddah University. Currently, he works as adjunct lecturer at college of computer science and artificial intelligence at Jeddah university. His scientific interests are related to artificial intelligence, image and video processing, machine learning, and in IoT. He can be contacted at email: ahmed.almaghthawi.1991@gmail.com.

Ebrahim Abdulwasea Abdullah Ghaleb (ID) (g) (SC) (P) received the B.Sc. and M.Sc. Bachelor of information technology (Hons) in Networking Technology Infrastructure University Kuala Lumper, Malaysia, and He hold Master. degree in Information system from The National University of Malaysia (Malay: Universiti Kebangsaan Malaysia, abbreviated as UKM). He is a Ph.D. student on information system with UTP Universiti Teknologi PETRONAS. He has authored or coauthored more than 9 refereed journal and conference papers, with Sustainability, Journal of Theoretical & Applied Information Technology, Solid State Technology and International Congress of Advanced Technology and Engineering, IEEE and Springer. my research interests include the applications of Big Data, Healthcare evolutionary and heuristic optimization techniques to power system planning, operation, and control. He can be contacted at email: ebrahim_1800342@utp.edu.my.