Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **3**, September **2025**

Enhancing Bank Customer Churn Classification using Wrapper Feature Selection Methods

Nurain Ibrahim[1,2*], Nurul Hidayah Khamsani[1], Nurul Izzah Ali[1],
Siti Aminah Mohamad Zali[1], Siti Khadijah Arafin[1], Azalina Farina Abd Aziz[2],
Muhammad Haziq Jaffri[3]

[1]School of Mathematical Sciences, Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2]Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[3]Institute of Business Excellence, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

*Corresponding author: Nurain Ibrahim (nurain@tmsk.uitm.edu.my)
Received June 1, 2025, revised July 5, 2025, accepted July 6, 2025.

Abstract. *In highly competitive service sectors, customer churn is a major problem. A lot of businesses have prioritized customer retention in their management and marketing strategies. Logistic regression (LR), decision tree (DT), and support vector machine (SVM) were used to determine the influence of the features on customer churn through the utilization of different wrapper feature selections methods: forward selection (FS), backward elimination (BE) and optimize selection (OS). The result found that forward selection with decision tree (FS+DT) is the best model as it has the best accuracy (77.46%), precision (77.29%), sensitivity (77.78%) and area under the ROC curve (0.844). Credit score, gender, age, tenure, balance, product number, credit card, active member and estimated salary were found as important features in predicting customer churn. Companies dealing with such issues should start developing comparable models immediately to be able to identify earlier on whether a customer churn the company. The findings can help businesses create more precise models to identify churn at an early stage, enabling them to implement focused retention strategies that boost customer loyalty and minimize turnover over time.*

Keywords: Bank Customer Churn Prediction; Machine Learning Models; Feature Selection; Wrapper Methods

1. Introduction.

In highly competitive service sectors, customer churn is a major problem [1]. These days, those rivals are willing to offer the same products and services at cheaper costs and superior quality [2]. As a result, customers are changing their bank loyalty. An analysis of the potential for a customer to stop using a product or service is known as a customer churn analysis [3]. The most basic definition of this phrase is that customers are left to choose the company due to competition. Banks have come to understand how crucial good customer relations are to their success [4]. How to keep the most profitable customers is the problem that banks must overcome. If done early on, predicting which customers are most likely to leave the business could be a significant additional revenue source [1]. Since the last decade, a lot of businesses have prioritized customer retention in their management and marketing strategies. It can cost five times as much to get a new customer as it does to keep an old one [2]. Selling to an existing customer has a success rate of 60–70% whereas selling to a new customer has a success rate of 2–5% only [5]. For this reason, businesses adopt policies that favour keeping their current customers above acquiring new ones. A proper statistical analysis is needed to increase knowledge of the features that influence the prediction of customer churn in the banking industry. Hence, in this study, a churn prediction model of classifying bank customers is developed using Logistic Regression (LR), Decision Trees (DT), and Support Vector Machine (SVM) while optimizing the feature selection methods.

2. **Related work.** In today's increasingly technological society, banks will find it very challenging to compete effectively without constant technology advancements. Additionally, bank customers are growing increasingly educated, and their expectations are too high towards these technical advancements. A previous study on customer churn in electronic banking is examined using the binomial logistic regression technique [6]. Due to the intense competition in electronic banking services, banks must comprehend the circumstances that cause customer churn. The findings revealed that customer churn is influenced by factors such as the amount of mobile banking transactions, age, gender, and length of customer connection. Another study indicated that the jobs in the banking industry with the highest churn rate are those related to food services, such as fast-food restaurants and merchants [7]. The sports facilities and households follow them in the order of churn from the bank services. They also found that the least risky corporate customers of the bank were kindergartens and governmental institutions, in contrast to counselling clinics. Additionally, among retail customers, those between the ages of 30 and 40 experienced the most service churn. In order to recommend an effective model to forecast customer churn in the banking business and factors affecting the customer churn, [8] compared the effectiveness of supervised classification techniques. Decision trees' strategy showed that younger consumers (under 41) and older consumers who are active members are more likely to stay with banks. Generally, customers with one or two accounts are more likely to stay with the bank, whereas creating more accounts may lead to churning behaviour. A previous study claimed that customer churn could be predicted with the use of data mining [9]. The aim of the study is to find the best data mining learning model that XYZ Bank could use to keep customers from leaving. Result from this study, the number of data samples utilized for learning has a significant impact on modelling outcomes. If the distribution of the data is significantly imbalanced, accuracy values cannot be considered as a reference for comparison in their entirety. Despite the fact that LR produces smaller losses, this study found that SVM provided the most accurate modelling for its research. A study [10] compared Random Forest (RF), AdaBoost, and SVM for predicting customer churn for unbalanced data. When under sampling is combined with oversampling and the unbalanced original data is SMOTED, RF was found as the best model with 91.90% of F1 score and an overall accuracy of 88.7%. Another study done by [11], the bank dataset is used to use a variety of machine learning models to estimate the possibility that a customer will churn, including LR, DT, K-nearest neighbours (KNN) and RF. The data was divided into training and testing data. 70% of the data set used for model construction is made up of training data, while the remaining 30% is made up of testing data, which is used to assess how well the trained model performed. The outcomes demonstrate that, among all classifiers, stratified and cross validation perform best in each condition. In addition, RF classifier outperforms others with a 0.3929 recall value and 85.23% accuracy. Feature selection is a technique for minimizing the input variable to the model by using only relevant data and eliminating irrelevant data [12]. Finding correct data models can be done with feature selection. There are three main techniques under feature selection which are wrapper method, filter method and embedded method.

Wrapper methods use a particular machine learning algorithm to attempt to fit on a given dataset for the feature selection process. This method has the advantages of being able to account for feature dependencies, as well as interaction between feature subset search and model selection [13]. Forward selection, backward elimination, and stepwise selection are the most often used approaches under wrapper methods. According to [14], by focusing simply on the data's inherent properties, filter approaches evaluate the relevance of features. Filtering techniques frequently have the issue of ignoring the relationship with the classifier. The feature selection process is carried out using embedded methods while the modelling algorithm is executing. As a result, these techniques are either part of the algorithm's standard functionality or its extended capabilities. A previous study [15] investigated how to execute experiments with a dataset of diabetes diagnoses by modifying an existing wrapper and filter approach. The wrapper performed excellent in this study, as was to be expected, and far better than filter approaches generally. It can be said that filter methods are more advantageous in practice because of their much-reduced complexity compared to wrapper methods and the small amount of information they use. The wrapper has the ability to make a correct choice, but even if the classifiers utilize the same training sets, their performance or even their chosen features may vary.

3. **Methodology.** This study focuses on the important features influencing customer churn and examines changes in customer loyalty. Section 3.1 presented the research flowchart for this study. Section 3.2 displays the data acquisition and section 3.3 shows the data preparation. Meanwhile, Section 3.4 mentioned three data modelling that were used in this study. These three methods are compared, and the best classification method is determined to predict customer churn in the banking industry using performance indicators mentioned in Section 3.5.

3.1. **Research Flowchart.** This section presented the flowchart being used in this study as shown in Fig. 1. The process starts with the data acquisition, followed by the data preparation and model preparation. Then, the next step is the constructed models were assessed, and the result and discussion were explained in this study. All steps in the flowchart is being explained in the subsequent sections.

3.2. **Data Acquisition.** This dataset is a secondary data taken from Kaggle website with the link access to the dataset of `https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset`, where the data was made public by Gaurav Topre and published in 2022, entitled "Bank Customer Churn Dataset". This dataset is related to Sustainable Development Goals (SDGs) which is decent work and economic growth. The data were obtained from 10000 records and consist of 12 variables which are customer ID, credit score, country, gender, age, tenure, balance, product number, credit card, active member, estimated salary and churn. This dataset is also being used in the previous study [16, 17]. The data description is displayed in Table 1.

3.3. **Data Preparation.** Account Balance and Estimated Salary features are converted from real to numerical using a data type conversion tool. Set role is used to change the role of Churn feature from regular to label. Next, the Churn feature is very imbalanced since there are 7963 customers that churn, while only 2037 customers do not churn. Therefore, SMOTE is used to balance the Churn feature. After SMOTE, Churn attribute is balance (7963 customers churn and 7963 customers do not churn). Any attributes that have no impact on the subject under discussion are considered irrelevant. The effectiveness of classifiers may occasionally be impacted by maintaining such attributes. The Customer ID and Country attributes have no impact on the prediction when the churn dataset

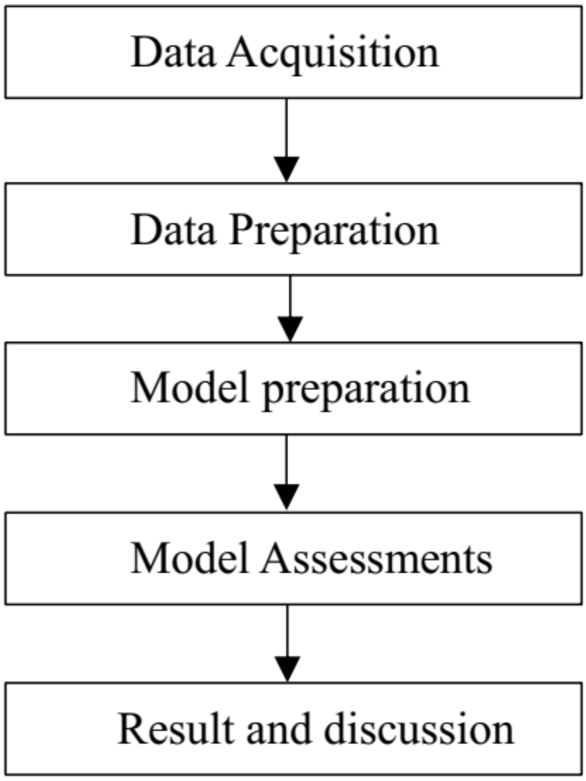

Figure 1. Research Flowchart

is taken into account. Furthermore, this study was focus on other variables that were deemed more directly related to the research objectives and it was not influenced by geographical location. Therefore, these variables were manually discarded in the study by using the select attribute tool in Rapidminer software. The process of converting data from one form to another is known as data transformation. Data transformation is done on Gender attribute (male = 0, female = 1). Moreover, there are some outliers in Age and Credit Score attributes. For Age attribute in the churn category have 486 outliers while in the not churn category have 13 outliers. The Credit Score attribute has only 11 outliers. In this study, all the outliers will be ignored since it consists only 0.05% of the full data. There are several tuning parameters in SVM such as kernel, regularization, gamma and margin. In this study, researchers only focus on the kernel dot. For DT, the splitting criteria that researchers use is gain ratio. Moreover, feature selection is used to minimize the number of input variables when building the predictive model. For this study, feature selection under wrapper methods such as backward elimination, forward selection and optimize selection is used. This study used a split data operator for data partitioning. The dataset is partitioned into two parts which are train data and test data. The ratio of the partition should be between 0 and 1. The sum of all ratios should be 1. Therefore, researchers will partition the dataset into 0.8 for train data and 0.2 for test data. In terms of feature selection methods, three wrapper feature selection methods were used: forward selection (FS), backward elimination (BE) and optimize selection (OS).

3.4. **Model Preparation.** Logistic regression is more flexible and has advantages compared to discriminant analysis since it does not require rigid assumptions. The probability of the dependent variable in logistic regression will be 1 (probability of success, p) or 0 (probability of failure, 1-p). This analysis generates a predictive equation, where the coefficient is used to assess the predictive capability of independent variables. This study

Table 1. Data Description

Variable Name	**Description**
Customer ID	Account number
Credit Score	The credit score
Country	Country of residence 0: France 1: Germany 2: Spain
Gender	Customer gender 0: Male 1: Female
Age	Customer age (years)
Tenure	How many years the customer is having bank account in ABC Bank (years)
Balance	Account balance (EUR)
Product Number	Number of product from bank 0: 1 1: 2 2: 3 3: 4
Credit Card	Whether customer have credit card or not 0: No 1: Yes
Active Member	Whether customer is active member of bank or not 0: Not Active Customer 1: Active Customer
Estimated Salary	Estimated Salary per year
Churn	Churn Status 0: Not Churn 1: Churn

is used to achieve the objective of predicting customer churn in the banking industry. Logistic regression equation is as Eq. (1)[18]:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_9 X_9 \tag{1}$$

As a starting point, the two possible outcomes of customer churn were given the values 1 (churn) and 0 (not churn). Then, Eq. (1) expresses logistic regression in terms of the natural log of its probabilities, where $X_1, X_2, \ldots, X_9$ are the input features (credit score, country, gender, age, tenure, balance, product number, credit card, active member and estimated salary). In contrast, the 10 parameters, β (the constant parameter), $\beta_1, \beta_2, \ldots, \beta_9$ will be estimated. Decision Tree (DT) is a non-parametric supervised learning technique that can be employed for classification and regression. The objective is to construct a model capable of predicting the value of a target property using simple rules drawn from the dataset. The decision tree is constructed using a recursive partitioning approach. It is a flowchart structure in which each internal node represents a test on an attribute, and the paths from the root to the leaf reflect the categorization rules. There are numerous criteria for dividing a large, heterogeneous population into smaller, more homogeneous groups based on the specific target in a decision tree model, such as CART

(Classification and Regression Trees), C4.5, CHAID (Chi-Squared Automatic Interaction Detection), and QUEST (Quick, Unbiased, Efficient, Statistical Tree) [19]. Researchers employ Decision Tree as a churn prediction model for two reasons: first, as they may produce findings that are simple to understand, and second, due to the type of data they use. The decision Tree approach is appropriate for this type of data since the data used are both category and numerical. This study will concentrate entirely on the default splitting criterion of gain ratio. Typically, the decision tree is used to (1) Calculate the likelihood that a given record belongs to each category and (2) Assign the record to the most likely class classification (or category). It is straightforward to comprehend and decipher a decision tree. Furthermore, this application requires no data preparation, whereas others usually require data normalization, creating fake characteristics, and eliminating blank entries. It is important to note, however, that this module does not support missing data. Support vector machine (SVM) is a model for supervised machine learning or data mining mainly used to sort problems into groups [20]. SVM has been widely utilized for feature reduction, regression, novelty detection, and classification models. Both linearly separable and non-linearly separable datasets can yield excellent results when using the SVM method and it also can work even with insufficient data. In SVM, there are two-label classifications, and it can deal with problems involving more than two classification levels. SVM is better than other classification algorithms due to their ability to select a decision boundary that optimizes the distance from all classes nearest data points. The maximum margin classifier or maximum margin hyperplane is the name of the decision boundary produced by SVM. Multiple automatic iterations can be produced using SVM to determine and select the optimal hyperplane since manual identification of the hyperplane is highly challenging in real-world conditions. Data can be input and then transformed into the format needed for processing by using a kernel function [21]. The mechanism for manipulating the data in a Support Vector Machine is provided by a set of mathematical operations. In order for a non-linear decision surface to turn into a linear equation in a higher number of dimension spaces, Kernel function often changes the training set of data. The dot product of the changed feature vectors is the only part of the objective function that is optimizing to match the higher dimensional decision boundary.

3.5. **Model Assessments.** Using a confusion matrix, a technique used to summarize the performance of a classification algorithm is one of the methods for assessing the predictive accuracy of any model. However, classification accuracy alone can be misleading when the number of observations in each class is equal. Ahmed et al. [22] says that making a confusion matrix could show how often predictions are correct and how often they are wrong compared to accurate data. The assessment measures can be evaluated using the metrics mentioned above. The proportion of actual positive cases correctly recognized is referred to as "sensitivity," and the proportion of actual negative cases correctly identified is known as "specificity." If the model predicts more than 70% of the time correctly and gets it wrong less than 30% of the time, it is considered to have a good prediction. The area under the Receiver Operating Characteristic (ROC) curve is a measurement of a two-dimensional area spanning from (0, 0) to (1, 1) underneath the complete ROC curve (1, 1). It plots the actual positive rate, also known as sensitivity, and the false positive rate, known as 1 - specificity. The area below the ROC curve represents the performance over all possible categorization levels. It is the likelihood that a model will rank a random positive case higher than a random negative case.

4. **Result and Discussion.**

4.1. **Descriptive Analysis.** According to Table 2, the skewness values of the continuous attributes fall within the range of -1 to 1 [23], and only estimated and credit score attributes were normally distributed. Thus, the mean value of estimated, tenure, age, balance and credit score can be used to represent the measure of the central tendency. The average of estimated salary per year, tenure, customer credit score, customer's age and account balance were 100182.69 EUR, 4.98 years, 648.19, 41 and 81952.12 EUR. They were also complete and free from any missing value. Therefore, the corrective action of removing or imputing a value for the missing record was not carried out.

Table 2. Descriptive Statistics for Continuous Variable

Variable Name	Skewness	Mean	Missing value
Age	0.563	41.08	0
Balance	-0.330	81952.12	0
Estimated	-0.001	100182.69	0
Credit Score	-0.084	648.19	0
Tenure	0.030	4.98	0

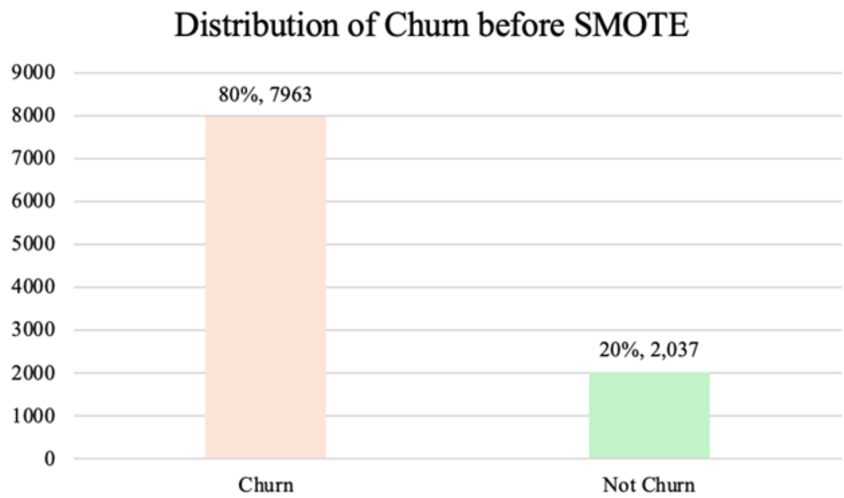

Figure 2. Distribution of Churn before SMOTE

4.2. **Predictive Analytics.** Table 3 shows a summary of the results of the logistic regression model's performance measures while employing various feature selection methods. Based on the table, logistic regression using backward elimination was chosen as the best model as it had the highest accuracy (72.16%), precision (72.34%), specificity (72.57%), and area under the ROC curve (0.782), with the lowest misclassification rate (27.84%). Even so, logistic regression with forward selection had a better performance in terms of sensitivity (72.44%). However, the sensitivity of logistic regression using backward elimination was still good enough as they were nearly identical to those of logistic regression using forward selection, with 71.75%, respectively. Hence, the logistic regression model with backward elimination was chosen as the best logistic regression model since it has most of the best performance measures. In spite of that, the decision tree with optimize selection had better performance in terms of sensitivity (85.56%). However, the sensitivity of decision tree using forward selection was still good enough as they were nearly identical to those of decision tree using optimize selection, with 77.78%, respectively. Hence, the decision tree model with forward selection was chosen as the best decision tree model

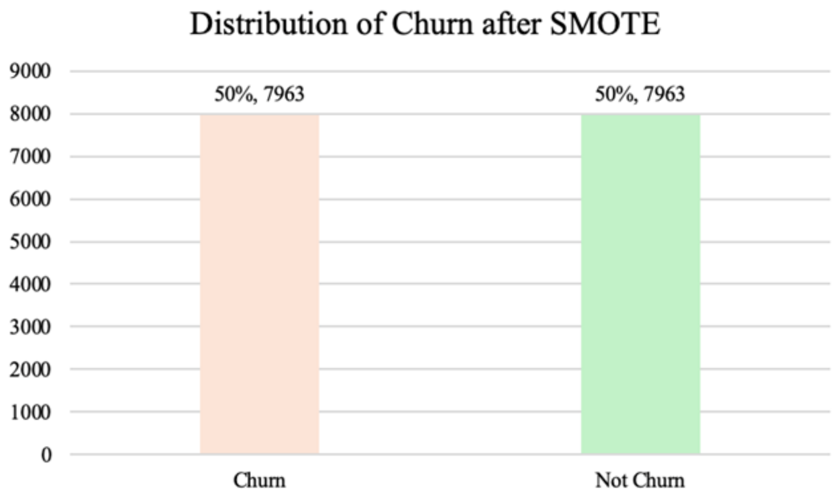

Figure 3. Distribution of Churn after SMOTE

since it has the majority of the best performance measures. Forward selection, backward elimination and optimize selection were added in the SVM model and the results showed the comparison measure for each selection method towards SVM model. In this study, forward selection has the highest accuracy (71.09%) and sensitivity (77.97%) compared to backward elimination and optimize selection. The SVM model with forward selection, which shows the highest sensitivity value, effectively identifies bank customers who are likely to churn and accurately predicts their actual churn behavior. Although SVM has shown the highest value accuracy and sensitivity to classify bank customer churn, it still has some disadvantages with the lower value of precision, specificity and area under the ROC Curve. However, backward elimination shows that its precision (73.76%), specificity (77.46%) and area under the ROC curve (0.772) have the highest percentage compared to forward selection and optimize selection. The model performs better at differentiating between the positive and negative classes when the ROC value is higher. Hence, a support vector machine with backward selection was chosen as the best model since most of the performance measures in backward elimination have the highest percentage compared with the performance measures in forward selection.

Table 3. Performance measures for each feature selection methods integrated with classification model

Methods	Acc(%)	Pre(%)	Sen(%)	Spe(%)	AUROC
FS + LR	71.25	70.75	72.44	70.06	0.756
BE + LR	72.16	72.34	71.75	72.57	0.782
OS + LR	70.40	69.99	71.44	69.37	0.740
FS + DT	77.46	77.29	77.78	77.15	0.844
BE + DT	77.34	76.22	79.47	75.20	0.842
OS + DT	77.28	73.40	85.56	68.99	0.829
FS + SVM	71.09	68.54	77.97	64.22	0.759
BE + SVM	70.40	73.76	63.34	77.46	0.772
OS + SVM	70.68	70.18	71.94	69.43	0.752

In order to find the best method for bank customer churn prediction, the findings of the best model with feature selection was tabulated in Table 4. In comparison to logistic regression using backward elimination and support vector machine using backward

Table 4. Performance measures for the best method

Methods	Acc(%)	Pre(%)	Sen(%)	Spe(%)	AUROC
BE + LR	72.16	72.34	71.75	72.57	0.782
FS + DT	77.46	77.29	77.78	77.15	0.844
BE + SVM	70.40	73.76	63.34	77.46	0.772

elimination, it was found that decision tree with forward selection had the best accuracy (77.46%), precision (77.29%), sensitivity (77.78%), and area under the ROC curve (0.844). A support vector machine with backward elimination, however, has the highest value of specificity, which is 77.46%, indicating that this model has the ability to correctly identify bank clients who will not churn. Additionally, compared to other measures, the accuracy of the decision tree with the forward selection model was more effective in determining the correct classification of bank customer churn which makes it the best model to predict bank customer churn. Additionally, this model is better at separating the positive and negative classes when the ROC curve value, which is 0.844, is higher compared to other models. Therefore, ROC curves are useful in determining the best model to predict bank customer churn. The findings in this study were not in line with a few previous studies that have been done using the same dataset from Kaggle. This is because most studies did not use the same feature selection which are backward elimination, forward selection and optimize selection for comparison classification methods. However, artificial neural networks appeared to do better overall performances, according to Chandarabi [24] who employed the same dataset to compare supervised classification techniques but decision trees can be considered as the second best model. Based on feature weight from each selection, it will determine important attributes. From Table 5, it shows that different feature selection methods choose different important features. Based on the best model, which was a decision tree using forward selection, the features of Credit score, gender, age, balance, product number, active member, estimated salary were important in predicting customer churn. As a result, if a researcher from banks or finance institutions is planning to build models to predict the customer churn, the variables of country and tenure should not be prioritized when collecting data and should not be considered in the model. This suggestion would reduce the burden of researchers on collecting specific data that was proven to be not important in explaining the presence of customer churn.

Table 5. Important variables for the best method

Methods	Important variables
BE + LR	Credit score, gender, age, tenure, balance, product number, active member
FS + DT	Credit score, gender, age, balance, product number, active member, estimated salary
BE + SVM	Credit score, gender, age, tenure, balance, product number, credit card, active member, estimated salary

5. **Conclusion.** In conclusion, decision tree using forward selection was found to be the best model for predicting customer churn in the banking industry. since it performs the best in terms of accuracy, precision, sensitivity, misclassification rate, and area under the ROC curve in comparison to other models. It also was the best model for predicting

customer bank churn cases. The SVM using backward elimination was the best model for predicting customer bank does not churn cases since it performs the best in terms of specificity.

Acknowledgment. The authors also acknowledge Faculty of Computer and Mathematical Sciences, UiTM, Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Research Nexus UiTM (ReNeU) and Institute of Business Excellence (IBE).

REFERENCES

[1] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, "Customer churn prediction in telecom sector using machine learning techniques," *Results in Control and Optimization*, vol. 14, p. 100342, 2024, doi: https://doi.org/10.1016/j.rico.2023.100342.

[2] H. Guliyev and F. Yerdelen Tatoğlu, "Customer churn analysis in banking sector: Evidence from explainable machine learning models," *Journal of Applied Microeconometrics*, vol. 1, no. 2, pp. 85–99, 2021, doi: 10.53753/jame.1.2.03.

[3] O. O. Osmanoglu, "Comparing to Techniques Used in Customer Churn Analysis," *Journal of Multidisciplinary Developments*, vol. 4, no. 1, pp. 30–38, 2019.

[4] A. Bilal Zoric, "Predicting Customer Churn in Banking Industry using Neural Networks," *Interdisciplinary Description of Complex Systems*, vol. 14, no. 2, pp. 116–124, 2016, doi: 10.7906/indecs.14.2.1.

[5] B. Trinh, "How Are Novel Technologies and Advanced Analytics Utilized to Deliver Best-In-Class Customer Experience and Prevent Customer Churn and Loss of New Businesses?," no. October, 2020.

[6] A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, "Developing a prediction model for customer churn from electronic banking services using data mining," *Financial Innovation*, vol. 2, no. 1, p. 10, 2016, doi: 10.1186/s40854-016-0029-6.

[7] S. H. Iranmanesh, M. Hamid, M. Bastan, G. Hamed Shakouri, and M. M. Nasiri, "Customer churn prediction using artificial neural network: An analytical CRM application," *Proceedings of the International Conference on Industrial Engineering and Operations Management*, no. July, pp. 2214–2226, 2019.

[8] A. Y. W. Chong, K. W. Khaw, W. C. Yeong, and W. X. Chuah, "Customer Churn Prediction of Telecom Company Using Machine Learning Algorithms," *Journal of Soft Computing and Data Mining*, vol. 4, no. 2, pp. 1–22, 2023, doi: 10.30880/jscdm.2023.04.02.001.

[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, pp. 33–38. doi: 10.1109/IWBIS.2019.8935884.

[10] A. Muneer, R. F. Ali, A. Alghamdi, S. M. Taib, A. Almaghthawi, and E. A. Abdullah Ghaleb, "Predicting customers churning in banking industry: A machine learning approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 539–549, 2022, doi: 10.11591/ijeecs.v26.i1.pp539-549.

[11] I. Kaur and J. Kaur, "Customer Churn Analysis and Prediction in Banking Industry Using Machine Learning," in *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2020, pp. 434–437. doi: 10.1109/PDGC50313.2020.9315761.

[12] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014, doi: https://doi.org/10.1016/j.compeleceng.2013.11.024.

[13] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Oct. 2007, doi: 10.1093/bioinformatics/btm344.

[14] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, pp. 1200–1205. doi: 10.1109/MIPRO.2015.7160458.

[15] M. Rahman and V. Kumar, "Machine Learning Based Customer Churn Prediction In Banking," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 1196–1201. doi: 10.1109/ICECA49313.2020.9297529.

[16] R. Bhuria et al., "Ensemble-based customer churn prediction in banking: a voting classifier approach for improved client retention using demographic and behavioral data," *Discover Sustainability*, vol. 6, no. 1, Dec. 2025, doi: 10.1007/s43621-025-00807-8.

[17] H. Guliyev and F. Yerdelen Tatoğlu, "Customer churn analysis in banking sector: Evidence from explainable machine learning models," *Journal of Applied Microeconometrics*, vol. 1, no. 2, pp. 85–99, Dec. 2021, doi: 10.53753/jame.1.2.03.

[18] S. Sarah et al., "Hybrid embedded and filter feature selection methods in big-dimension mammary cancer and prostatic cancer data," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 3, pp. 3101–3110, 2024, doi: 10.11591/ijai.v.

[19] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.

[20] S. Lu, Y. Chen, X. Zhu, Z. Wang, Y. Ou, and Y. Xie, "Exploring Support Vector Machines for Big Data Analyses," in *Proceedings of the 4th International Conference on Computer Science and Software Engineering, CSSE '21*, New York, NY, USA: Association for Computing Machinery, 2021, pp. 31–37. doi: 10.1145/3494885.3494891.

[21] J. Hair and A. Alamer, "Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example," *Research Methods in Applied Linguistics*, vol. 1, no. 3, p. 100027, 2022, doi: `https://doi.org/10.1016/j.rmal.2022.100027`.

[22] N. Ahmed et al., "Machine learning based diabetes prediction and development of smart web application," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 229–241, 2021, doi: `https://doi.org/10.1016/j.ijcce.2021.12.001`.

[23] J. Hair and A. Alamer, "Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example," *Research Methods in Applied Linguistics*, vol. 1, no. 3, Dec. 2022, doi: 10.1016/j.rmal.2022.100027.

[24] S. E. Charandabi, "Customer Churn Prediction in Banking Industry," *California Business Review*, vol. 3, no. 1, pp. 41–46, 2015, doi: 10.18374/cbr-3-1.6.