Institut Riset dan Publikasi Indonesia (IRPI)
**IJATIS: Indonesian Journal of Applied Technology and Innovation Science**
Journal Homepage: https://journal.irpi.or.id/index.php/ijatis
Vol. 3 Iss. 1 February 2026, pp: 41-49
ISSN(P): 3032-7466 | ISSN(E): 3032-7474

# Classification-Based Supervised Learning Algorithms for Accurate Prediction of Customer Churn in Banking

**Nora Waningsih[1*], Alfi Surya Akbar[2], Shofia Ariska[3], Ri'lah Faizatul Husnayaini[4], Eflin Nurrin[5], Rosidur Ridho[6], Fauziah Tio Pratama Situmorang[7]**

[1,2,3]Department of Information System, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[4,5,6]Departemen of Ushuludin, Al-Azhar University, Egypt
[7]Department of Journalism, Kazan Federal University, Russia

E-Mail:[1]12350323220@students.uin-suska.ac.id, [2]12350313035@students.uin-suska.ac.id,
[3]12350322566@student.uin-suska.ac.id, [4]husnayanirilahfaizatul@gmail.com,
[5]evelynnur19@gmail.com, [6]rosidurridho@gmail.com, [7]tiopratamas2212@gmail.com

Received Jan 03rd 2026; Revised Feb 20th 2026; Accepted Mar 17th 2026; Available Online Mar 18th 2026
Corresponding Author: Nora Waningsih


**Abstract**

The banking industry has become increasingly dynamic with the emergence of financial technology (fintech) companies that have significantly changed customer behavior and expectations. As competition intensifies, customer churn has become a critical issue because it directly affects a bank's revenue, reputation, and long-term sustainability. Therefore, banks require effective analytical approaches to identify customers likely to leave and to develop appropriate retention strategies. This study aims to analyze and predict customer churn likelihood using a bank customer dataset by applying supervised machine learning classification techniques. Five algorithms were evaluated, namely Decision Tree, Random Forest, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). The models were trained and evaluated using a hold-out validation approach, and performance was assessed using accuracy as the primary evaluation metric. The experimental results show that Random Forest achieved the highest accuracy of 86%, outperforming the other algorithms, while the MLP model produced the lowest accuracy of 82%. These findings indicate that ensemble-based methods provide better performance for predicting bank customer churn. The results of this study can assist banks in identifying potential churn customers and in developing effective customer retention strategies. Future research may explore additional algorithms, advanced data preprocessing techniques, and larger datasets to further improve prediction performance.

Keywords: Churn Banking, Classification, Multi-Layer Perceptron, Random Forest, Support Vector Machine, XGBoost

## 1. INTRODUCTION

Modern banking operates in an increasingly competitive business environment, where customer retention is a strategic factor in maintaining revenue stability and long-term growth. The phenomenon of customer churn, which is when customers decide to stop using services or switch to another bank, has become a serious challenge for the industry [1], [2]. The inability to accurately predict churn can lead to significant financial losses, increased customer acquisition costs, and weakened bank competitiveness. This situation is exacerbated by competition from financial technology (fintech) companies, which offer highly flexible, digital-based services focused on customer experience. Therefore, banks' ability to anticipate churn through a data-driven analytical approach is becoming increasingly important [2].

Many banks still rely on conventional methods, such as basic statistical analysis or manual evaluation of customer behavior. These approaches are less effective at capturing nonlinear and complex relationships among variables, often resulting in low prediction accuracy, inefficient processes, and delayed detection of churn. These issues demand technology-based solutions capable of processing large and diverse amounts of data and producing consistent and reliable predictions. In the last two decades, machine learning has emerged as an effective approach to analyzing customer behavior and predicting churn.

Supervised learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost) offer great potential for improving the accuracy of churn predictions. These algorithms have been widely used in various domains due

to their ability to learn complex relationships within data. Previous studies have shown that Random Forest and SVM can achieve high accuracy in detecting churn, while Decision Tree and Naïve Bayes offer better interpretability for non-technical stakeholders. Random Forest has demonstrated strong performance on imbalanced datasets and can achieve significant improvements through parameter tuning [3]. Furthermore, previous studies have reported that Random Forest often provides higher accuracy than other machine learning methods in multi-class classification problems [4]. Previous studies reported that the Random Forest model achieved the best overall performance for bank customer churn prediction, outperforming several machine learning algorithms in terms of classification accuracy and sensitivity [5]. These findings confirm the superiority of ensemble-based algorithms in handling diverse and complex data[6], [7], [8], [9], [10].

However, even though these algorithms have been widely used, most studies are still limited to specific contexts with homogeneous datasets, so they do not provide a systematic comparative overview of algorithm performance on heterogeneous banking data. This research gap underscores the need for a comprehensive evaluation of machine learning algorithms for banking churn prediction.

This research gap underscores the need for a comprehensive evaluation of machine learning algorithms for banking churn prediction. Therefore, this study aims to compare the performance of five supervised learning algorithms Decision Tree, Random Forest, Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost) for predicting bank customer churn. The models are evaluated using accuracy, precision, recall, and F1-score to identify the best-performing approach for heterogeneous banking data, thereby supporting the development of more effective customer retention strategies in the banking industry.

## 2. MATERIAL AND METHOD

The purpose of the elements in the diagram is to structure and organize the data, to improve model accuracy and performance, and to minimize bias and error in the analysis of results. Research Methodology can see Figure 1.

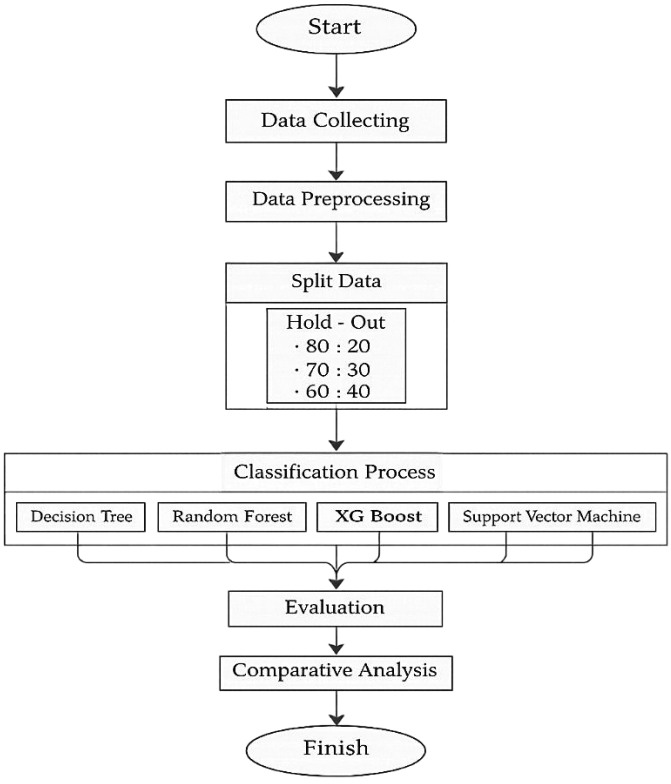


**Figure 1.** Research Methodology

### 2.1. Data Collecting

The dataset used in this study was obtained from the Kaggle repository, specifically the Bank Customer Churn dataset. The dataset contains 156,770 records with 12 attributes, including both numerical and categorical variables: customer_id, credit_score, country, gender, age, tenure, balance, products_number, credit_card, active_member, and churn. The churn variable indicates whether a customer leaves the bank's service or continues using it. This dataset is widely used in churn prediction research due to its balanced representation of customer behavioral features.

### 2.2. Data Preprocessing

Once the data has been collected, the next step is to prepare it for use in analysis with machine learning models. Before applying machine learning algorithms, several preprocessing steps were performed to ensure data quality. The preprocessing stage included identifying and handling data cleaning, Missing values, duplicate records, and inconsistencies in the dataset. Feature Selection Irrelevant attributes, such as customer_id, were removed because they do not contribute to the predictive model. Categorical Encoding: Categorical variables such as gender and country were converted to numerical values using encoding techniques. Data Transformation: The dataset was transformed into a structured numerical format suitable for machine learning algorithms. These preprocessing steps help improve model performance and reduce potential bias during training [11], [12], [13].

### 2.3. Hold Out Validation

The hold-out validation technique was applied to evaluate model performance. In this approach, the dataset is divided into two subsets: Training set (80%), Testing set (20%). The training data is used to build the classification models, while the testing data is used to evaluate the performance of the trained models.[14]. The hold-out method is widely used due to its simplicity and efficiency, particularly for large datasets where repeated resampling methods may increase computational cost

### 2.4. Classification

Classification is a supervised learning method that aims to determine the class or category of new data using algorithms such as the Multi-Layer Perceptron (MLP), Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost. The performance of these classification models is evaluated using testing metrics such as Precision, Recall, and the Confusion Matrix. These algorithms were selected because they represent different machine learning paradigms, including tree-based learning, ensemble learning, neural networks, and margin-based classifiers. Comparing these algorithms allows the study to evaluate their effectiveness in handling customer churn prediction problems[15].

### 2.5. Confusion Matrix

When assessing the performance of classification models, the Confusion Matrix is a frequently used matrix that illustrates how well the machine learning model predicts the result The Confusion Matrix serves as an evaluation method for assessing classification performance based on true and false categories. [9], [16]This matrix includes accuracy, precision, and recall metrics, calculated from four main outputs: recall, precision, accuracy, and error rate.

**Table 1.** Confusion Matrix

| | | Predicition | |
|---|---|---|---|
| | | Positive | Negative |
| Actual | Positive | TP | FN |
| | Negative | FP | TN |

The accuracy value can be computed using equations 1-3: precision and recall are calculated using equations, respectively.

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

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

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

### 2.6. Decision Tree

A decision tree is a procedure that divides a data set into branches, like a tree structure. This model is easy to understand, making it simple to explain. Although other algorithms, such as neural networks, can produce more accurate models under certain conditions, decision trees can be trained to mimic the predictions of neural networks, thereby helping to open up the "black box" of those networks[10], [17]. In addition, decision trees can model strong nonlinearities in the relationships between target and predictor variables. The splitting process is commonly determined by entropy, as shown in the equation.

$$\text{Entropy(S)} = -\text{i} = 1\sum\text{cpilog2(pi)} \quad (4)$$

The best is selected using Information Gain, as shown in equation 5.

$$\text{Gain}(S, A) = \text{Entropy}(S) - v \in \text{Values}(A)\sum | S || Sv | \text{Entropy}(Sv) \quad (5)$$

### 2.7. XG Boost

Extreme Gradient Boosting (XGBoost) is an optimized implementation of the gradient boosting framework that builds decision trees sequentially, with each tree correcting the errors of the previous one. The algorithm incorporates regularization, shrinkage, and efficient tree-pruning techniques, enabling it to model complex patterns while reducing overfitting. XGBoost is also widely utilized alongside other tree-based algorithms, such as Random Forest and Gradient Boosted Machine, in comparative studies on classification tasks, including churn-related research[4], [18]. The objective of XG Boost is defined as equation 6.

$$\text{Obj} = \sum_{i=1}^{n} i\,(yi, yi) + \sum_{k=1}^{k} \Omega\,(fk) \quad (6)$$

Regulatization term, equation 7.

$$\Omega(f) = \gamma T + 21\lambda \sum_{j=1}^{T} wj2 \quad (7)$$

### 2.7. Random Forest

Breiman introduced Random Forest (RF) as an ensemble classifier for decision trees. This method builds a number of decision trees, where each tree is trained using random vectors that are selected independently but follow the same distribution[19], [20]. This approach is effective in overcoming the tendency of single decision trees to often overfit to the training data[13], [16], [21], [22], [23]. Simply put, Random Forest is a technique that combines multiple decision trees trained on different parts of the dataset, with the main goal of reducing variance[10], [11], [24], [25]. Another advantage of RF is its ability to handle high-dimensional data without requiring dimension reduction or feature selection. In addition, the training process is relatively fast and can be easily implemented in parallel, thereby increasing computational efficiency [3], [14], [26], [27]. The final prediction is determined by majority voting, as shown in equation 8.

$$y\text{^} = \text{mode}(h1(x), h2(x), \ldots, hT(x)) \quad (8)$$

### 2.8. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that can be applied to both classification and regression tasks[28]. The main principle of SVM is to determine an optimal hyperplane capable of separating data into distinct classes[29]. This separation is achieved by maximizing the margin, which represents the distance between the hyperplane and the closest data points from each class, known as support vectors [30]. In practice, SVM can handle both linear and non-linear data by employing kernel functions such as linear, polynomial, and radial basis function (RBF), which map the input space into higher dimensions to allow more effective class separation. Previous studies also highlight that SVM performs well in high-dimensional spaces and can model complex decision boundaries, though in some comparative evaluations, its performance may appear moderate due to issues such as data imbalance. The decision function is defined as equations 8-10.

$$f(x) = w \cdot x + b \quad (8)$$

Subject to:

$$yi(w \cdot xi + b) \geq 1 \quad (9)$$

Decision function:

$$f(x) = \text{sign}(w \cdot x + b) \quad (10)$$

### 2.9. Multi-Layer Perceptron (MLP)

MLP requires normalized data as input, z-score normalization has been made prior to the algorithm training by means of a “Normalizer” node. The same technique has then been applied to test data. MLPs are a powerful class of nonlinear statistical models which consist of multiple layers of nodes in a directed graph,

with each layer fully connected to the next one. There are three different types of layers, input, hidden, and output layers. Thus, except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function [31]. The output of each neuron is calculated using equations 11 and 12.

$$z_j = \sum_{i=1}^{n} wijxi + b_j \quad (11)$$

Activation function:

$$f(z) = 1 + e - z1 \quad (12)$$

## 3. RESULTS AND DISCUSSION

In this section, the research results are explained and a comprehensive discussion is provided. Results can be presented in figures, graphs, tables, and other forms that make the reader understand easily [2], [5]. The discussion can be made in several sub-chapters.

### 3.1. Initial Data

At this stage, the dataset represents the raw data as obtained from the original source, relating to crurn banking. The dataset used in this research consists of 15677020 records and 12 attributes covering information from numerical and categorical data, including customer_id, credit_score, country, gender, age, tenure, balance, products_number, credit_card, active_member, estimated_salary, and churn. This dataset is still in its raw form and will serve as the basis for analysis, with a preprocessing plan that includes handling missing values, data type conversion, and normalization. Table 2 presents the details of each attribute, including its data type and description before preprocessing.

**Table 2.** Initial Data

| No | Attribute Name | Data Type | Description |
|---|---|---|---|
| 1 | customer_id | Numerical | Unique identifier, typically not used for training |
| 2 | credit_score | Numerical | Location of the warehouse where the product is stored |
| 3 | country | Categorical | One-Hot/Label Encoding |
| 4 | gender | Categorical | Male/Female |
| 5 | age | Numerical | Age in years |
| 6 | tenure | Numerical | Length of time as a customer (years/months) |
| 7 | balance | Numerical | Account balance |
| 8 | product_number | Categorical | Number of products (if numerical, ordinal) |
| 9 | credit_card | Categorical | Have a credit card (1/0) |
| 10 | active_member | Categorical | Active member (1/0) |
| 11 | estimated_salary | Numerical | Estimated salary |
| 12 | churn | Target | Target variable (0=Stay, 1=Churn) |

### 3.2. Data Preprocessing

The first stage in the machine learning process is data transformation. At this stage, data is transformed by removing irrelevant attributes and converting categorical values to numerical values. The purpose of data transformation is to avoid modeling errors and make it easier for the model to understand the data. The attribute removed is ID_Customer because it does not show a strong relationship with the label. Then, outliers are cleaned up. Table 3 shows the dataset after preprocessing.

**Table 3.** Data Processing

| credit_score | country | gender | …. | estimated_salary | churn |
|---|---|---|---|---|---|
| 619.0 | France | Female | …. | 101348.88 | 1 |
| 608.0 | Spain | Female | …. | 112542.58 | 0 |
| 502.0 | France | Female | …. | 113931.52 | 1 |
| …. | …. | …. | …. | …. | …. |
| 772.0 | Germany | Male | …. | 92888.52 | 1 |
| 792.0 | France | Female | …. | 38190.78 | 0 |

### 3.3. Implementation of Classification Algorithm

This study applies classification algorithms. The purpose of testing these algorithms is to evaluate their accuracy in predicting churn in the banking industry. The five algorithms tested in this study are Decision Tree, Random Forest, MLP, SVM, and XGBoost. The Hold Out method was used as a validation technique in this study. The data was divided into two main parts: the training set to train the model and the testing set to evaluate

the model's performance. The division ratios used were 90:10, 80:20, and 70:30, so that each algorithm was tested in several data division scenarios.

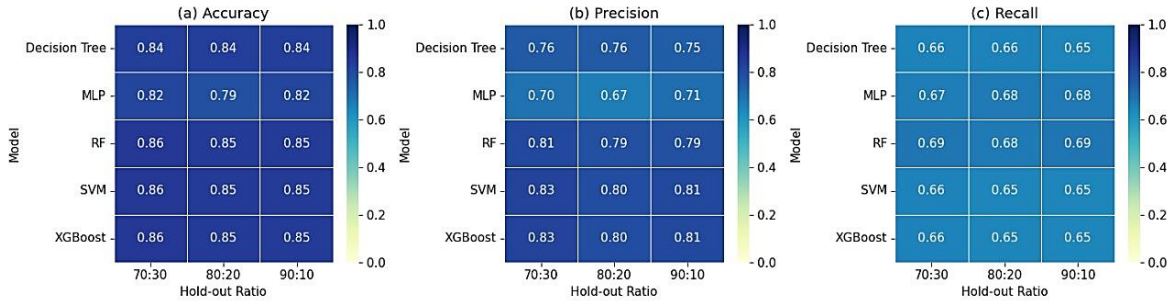


**Figure 2.** Comparison of Algorithm Classification Results

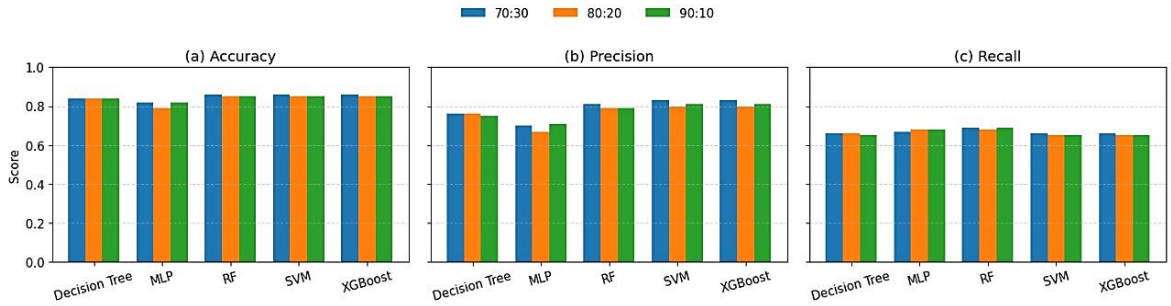


**Figure 3.** Graphic Comparison of Algorithm Classification Results

Figures 2 and 3 illustrate the comparison of classification algorithm performance using accuracy, precision, and recall metrics. The testing process was conducted in stages to ensure consistency of results. Based on the evaluation of the five classification algorithms, the 70:30 data split produced the best performance, with each algorithm achieving the highest average accuracy when tested, and the Random Forest algorithm had the best accuracy at 86%, with a precision of 81% and a recall of 69% emerged as the most effective for predicting banking churn.

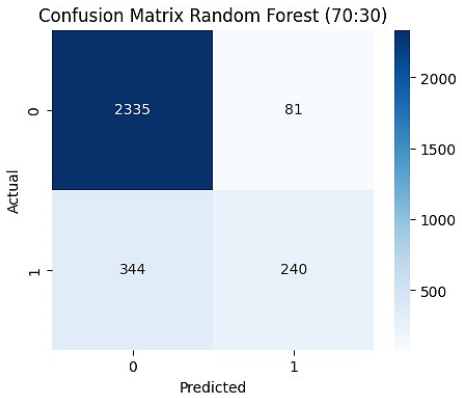


**Figure 4**. Confusion matrix of the Random Forest

Figure 4. Illustration Based on the results of the confusion matrix analysis, the Random Forest and Multi-Layer Perceptron (MLP) models show that random forest has quite good performance in predicting bank customer churn with relatively balanced levels of accuracy, precision, and recall. Random Forest has an advantage in classifying non-churn customers, as indicated by high True Negative values and low False Positives, resulting in better precision. and has a still quite large False Negative value indicating that the recall ability in detecting churn is not optimal. Conversely, the MLP model shows less optimal performance even though it is able to detect more churned customers as indicated by a higher True Positive value, thus having better recall, although accompanied by an increase in False Positives which has an impact on decreasing precision. Overall, Random Forest is more stable in producing precise predictions, while MLP is more sensitive to churn, so model selection needs to be adjusted to the objectives of the customer retention strategy.

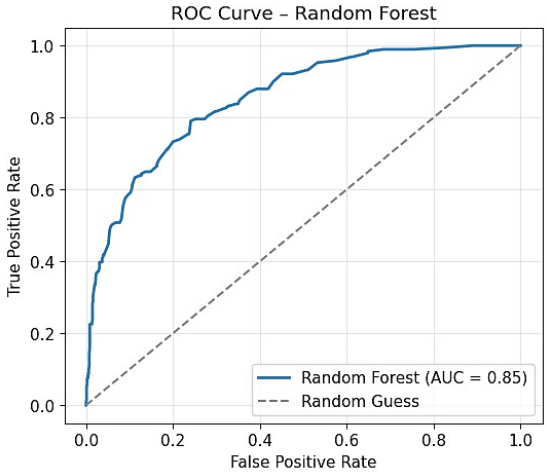


**Figure 5**. Receiver Operating Characteristic (ROC)

Figure 5 shows the Receiver Operating Characteristic (ROC) curve for bank churn classification using the One-vs-Rest approach. The ROC curve depicts the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) at various threshold values. The AUC value of 0.85 for both classes (class 0 and class 1) indicates that the model has a fairly good classification ability in distinguishing churned and non-churn customers, because its performance is far above the random guess line. The closer to the upper-left corner, the better the model's ability to correctly identify the class.

## 4. CONCLUSION

This study evaluates the performance of several classification algorithms Decision Tree, Random Forest, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and XGBoost for predicting bank customer churn. The experiments were conducted using the hold-out validation technique with three data split scenarios: 70:30, 80:20, and 90:10. The results show that the 70:30 split provided the best overall performance. Among the evaluated models, Random Forest achieved the highest accuracy of 86%, outperforming the other algorithms, while MLP achieved the lowest accuracy of 82%. These findings indicate that Random Forest is the most effective model for classifying bank customer churn in the dataset used in this study. Future research may focus on improving prediction performance through hyperparameter optimization, the use of larger and more diverse datasets, and the exploration of advanced approaches such as ensemble methods and deep learning models.

## REFERENCES

[1] B. Thenmozhi, C. Jeyabharathi, and S. Vimala, “Customer Churn Prediction Analysis in the Banking Sector Using Machine Learning,” 2024.

[2] J. Basit, A. Sheikh, N. Umer, and M. Syed, “Comparative Analysis of Deep Learning Architectures for Customer Churn Prediction in the Banking Sector,” 2025. https://journals.iub.edu.pk/index.php/JCIS/

[3] A. Raza et al., “Predicting regional-scale groundwater levels at high spatial resolution using spatial Random Forest models,” International Journal of Applied Earth Observation and Geoinformation, vol. 144, Nov. 2025, doi: 10.1016/j.jag.2025.104918.

[4] N. Simarmata et al., “Comparison of random forest, gradient tree boosting, and classification and regression trees for mangrove cover change monitoring using Landsat imagery,” Egyptian Journal of Remote Sensing and Space Science, vol. 28, no. 1, pp. 138–150, Mar. 2025, doi: 10.1016/j.ejrs.2025.02.002.

[5] P. P. Singh, F. I. Anik, R. Senapati, A. Sinha, N. Sakib, and E. Hossain, “Investigating customer churn in banking: A machine learning approach and visualization app for data science and management,” Data Science and Management, vol. 7, no. 1, pp. 7–16, Mar. 2024, doi: 10.1016/j.dsm.2023.09.002.

[6] B. Thenmozhi, C. Jeyabharathi, and S. Vimala, “Customer Churn Prediction Analysis In The Banking Sector Using Machine Learning,” 2024.

[7] C. Karunakaran, V. Niranjan, and A. S. Setlur, “Random Forest and XGBoost-based ensemble models for colorectal cancer exome variant classification and web application deployment for early prediction,” Computational and Structural Biotechnology Reports, vol. 2, p. 100063, 2025, doi: 10.1016/j.csbr.2025.100063.

[8] A. Y. Mahmoud, “Novel efficient feature selection: Classification of medical and immunotherapy treatments utilising Random Forest and Decision Trees,” Intell. Based. Med., vol. 10, Jan. 2024, doi: 10.1016/j.ibmed.2024.100151.

[9] M. Imani, A. Beikmohammadi, and H. R. Arabnia, "Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels," Technologies (Basel)., vol. 13, no. 3, Mar. 2025, doi: 10.3390/technologies13030088.
[10] T. Adugna, W. Xu, and J. Fan, "Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images," Remote Sens. (Basel)., vol. 14, no. 3, Feb. 2022, doi: 10.3390/rs14030574.
[11] M. S. Chowdhury, "Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting," Environmental Challenges, vol. 14, Jan. 2024, doi: 10.1016/j.envc.2023.100800.
[12] S. Hafsah et al., "Classification of IPB variety of cayenne pepper genotypes using physical characteristics during the growing period until harvest using machine learning," Future Foods, vol. 10, Dec. 2024, doi: 10.1016/j.fufo.2024.100500.
[13] D. C. Djarang and W. P. Sari, "An In-Depth Rainfall Classification Using Random Forest And Artificial Neural Network," Procedia Comput. Sci., vol. 269, pp. 1339–1347, 2025, doi: 10.1016/j.procs.2025.09.075.
[14] E. Helmud, E. Helmud, F. Fitriyani, and P. Romadiana, "Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix," Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 13, no. 1, pp. 92–97, Feb. 2024, doi: 10.32736/sisfokom.v13i1.1985.
[15] W. J. Sari et al., "Performance Comparison of Random Forest, Support Vector Machine and Neural Network in Health Classification of Stroke Patients," Public Research Journal of Engineering, Data Technology and Computer Science, vol. 2, no. 1, pp. 34–43, Apr. 2024, doi: 10.57152/predatecs.v2i1.1119.
[16] Y. Yang and H. Wang, "Random Forest-Based Machine Failure Prediction: A Performance Comparison," Applied Sciences (Switzerland), vol. 15, no. 16, Aug. 2025, doi: 10.3390/app15168841.
[17] M. M. Hassan et al., "A comparative assessment of machine learning algorithms with the Least Absolute Shrinkage and Selection Operator for breast cancer detection and prediction," Decision Analytics Journal, vol. 7, Jun. 2023, doi: 10.1016/j.dajour.2023.100245.
[18] O. Idemudia, J. O. Ehiorobo, O. C. Izinyon, and I. R. Ilaboya, "Evaluating the performance of Random Forest, Decision Tree, Support Vector Regression and Gradient Boosting for streamflow prediction," CTU Journal of Innovation and Sustainable Development, vol. 16, no. 2, pp. 116–130, Jul. 2024, doi: 10.22144/ctujoisd.2024.297.
[19] S. Kumar Ghosh and F. Janan, "Prediction of Student's Performance Using Random Forest Classifier," 2021.
[20] I. M. Rajagukguk, R. Hartanto, Julian, and R. Halim, "Comparative Analysis of XGBoost, Random Forest, and Logistic Regression for Classifying Jakarta's Air Pollution Index (ISPU)," in Procedia Computer Science, Elsevier B.V., 2025, pp. 108–120. doi: 10.1016/j.procs.2025.08.264.
[21] V. B. Moneravilla et al., "Random Forest Regression assisted Raman spectroscopy for authenticating the purity of virgin coconut oil," Journal of Food Composition and Analysis, vol. 149, p. 108784, Jan. 2026, doi: 10.1016/j.jfca.2025.108784.
[22] E. Meriç and Ç. Özer, "Symptom-Based Health Status Prediction via Decision Tree, KNN, XGBoost, LDA, SVM, and Random Forest," in Lecture Notes in Networks and Systems, Springer Science and Business Media Deutschland GmbH, 2023, pp. 193–207. doi: 10.1007/978-3-031-27099-4_15.
[23] S. R. Suwanlee et al., "Weed classification in sugarcane fields in Northeast Thailand from multi-temporal Sentinel-1 and Sentinel-2 data together with random forest algorithm," Science of Remote Sensing, vol. 13, p. 100352, Jun. 2026, doi: 10.1016/j.srs.2025.100352.
[24] D. N. Cosenza et al., "Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock," Forestry, vol. 94, no. 2, pp. 311–323, Apr. 2021, doi: 10.1093/forestry/cpaa034.
[25] N. Y. Nikitin and A. Stepashkin, "Classification of tensile test results of unidirectional carbon fiber-polysulfone composite material based on random forest, KNN and CNN methods," Results in Materials, vol. 28, Dec. 2025, doi: 10.1016/j.rinma.2025.100788.
[26] S. T. Hamidou and A. Mehdi, "Enhancing IDS performance through a comparative analysis of Random Forest, XGBoost, and Deep Neural Networks," Machine Learning with Applications, vol. 22, p. 100738, Dec. 2025, doi: 10.1016/j.mlwa.2025.100738.
[27] Y. Yuan, K. Wang, D. Duives, W. Daamen, and S. P. Hoogendoorn, "Machine learning-based bicycle delay estimation at signalized intersections using sparse GPS data and traffic control signals - A Dutch case study using random forest algorithm," Artificial Intelligence for Transportation, vol. 3–4, p. 100037, Nov. 2025, doi: 10.1016/j.ait.2025.100037.

[28] Y. Altork, "Comparative analysis of machine learning models for wind speed forecasting: Support vector machines, fine tree, and linear regression approaches," International Journal of Thermofluids, vol. 27, May 2025, doi: 10.1016/j.ijft.2025.101217.

[29] D. K. Murugan, Z. Said, D. Dineshbabu, S. Shankaranarayanan, G. Dhamodaran, and C. V. Dayakar, "Experimental validation and support vector machine optimization of rice husk gasification for sustainable syngas production and dual-fuel engine application," Results in Engineering, vol. 28, Dec. 2025, doi: 10.1016/j.rineng.2025.108345.

[30] A. S. More and D. P. Rana, "Performance enrichment through parameter tuning of random forest classification for imbalanced data applications," in Materials Today: Proceedings, Elsevier Ltd, 2022, pp. 3585–3593. doi: 10.1016/j.matpr.2021.12.020.

[31] X. Yu et al., "A continual-learning-based Multi-Layer Perceptronfor improved reconstruction of three-dimensional nitrate concentrations," Earth Syst. Sci. Data, vol. 17, no. 6, pp. 2735–2759, Jun. 2025, doi: 10.5194/essd-17-2735-2025.