ISSN 2307–3489 (Print), ISSN 2307–6666 (Online)

Наука та прогрес транспорту, 2026, № 1 (113)

UDC 336.71:004.89

M. B. NAMAZOV[1], M. E. RAHIMOV[2*]

[1]Inst. «Cyber-Physical Systems and Cybersecurity», Azerbaijan Technical University, H. Javid Av., 25, Baku, Azerbaijan, AZ1073, tel. +994 (50) 346 57 07, e-mail manafeddin.namazov@aztu.edu.az, ORCID 0000-0003-1119-2442
[2*]Dep. «Engineering Mathematics and Artificial Intelligence», Azerbaijan Technical University, H. Javid Av., 25, Baku, Azerbaijan, AZ1073, tel. +994 (70) 530 39 60, e-mail musa.rahimov@aztu.edu.az, ORCID 0009-0009-0961-5064

Development of a Hybrid Fuzzy Logic and SHAP-Based Evaluation Framework for the Prediction of Bank Deposits in the Banking Sector

Purpose. This study examines the decision-making behavior of bank customers regarding deposit participation and addresses the challenge of accurately predicting deposit decisions in the presence of class imbalance, limited interpretability, and variability in customer engagement patterns. **Methodology.** A comparative experimental framework was developed to evaluate several classical and ensemble machine learning approaches using a banking customer dataset. To ensure a balanced and objective comparison of competing models, a hybrid evaluation framework based on fuzzy multi-criteria decision analysis and entropy-based weighting was employed. In addition, explainable artificial intelligence techniques were applied to interpret the contribution of individual variables, while temporal analysis was conducted to investigate behavioral variations in customer response over time. **Findings.** The results indicate that ensemble learning approaches demonstrate more stable predictive capability compared to conventional models. The hybrid evaluation framework enables consistent ranking of predictive methods and improves the reliability of model selection. Feature interpretation and temporal behavioral analysis reveal that customer interaction characteristics and communication-related variables play a decisive role in deposit participation decisions. **Originality.** The study proposes an integrated predictive modeling framework combining machine learning forecasting, entropy-weighted fuzzy decision evaluation, explainable artificial intelligence, and time-based behavioral analysis, providing a transparent and robust mechanism for model comparison. **Practical value.** The proposed approach can assist financial institutions in improving marketing campaign planning, optimizing customer targeting strategies, and supporting data-driven decision-making in deposit acquisition processes.

Keywords: deposit prediction; banking customer behavior; machine learning; fuzzy multi-criteria decision analysis; entropy-based weighting; explainable artificial intelligence; temporal behavior analysis

Introduction

Deposits continue to be one of the main pillars of financial stability, liquidity management, and long-term profitability in the contemporary banking industry [8]. Banks' ability to finance lending operations, maintain sufficient liquidity buffers, and guarantee resilience against market fluctuations are all directly impacted by their ability to draw in and hold onto deposit funds. Deposits have historically been thought of as one of the most dependable and economical sources of funding, and despite the significant structural changes in the global financial landscape, their strategic significance has not decreased.

The environment in which deposit products function has undergone significant change in the last ten years. Market dynamics have been drastically changed by the quickening pace of digital transformation and the expansion of fintech rivals [3]. Through shifting deposit betas and pass-through, the rising interest rate volatility has also had an impact on customer behavior and deposit pricing [2]. These days, many clients make more complicated and unpredictable decisions by contrasting traditional deposit offers with a variety of alternative investment options, such as government securities and digital assets [3]. Because of this, the competition for depositors has become more fierce, and predicting consumer behavior has become a crucial component of banking strategy.

Predicting consumer behavior with regard to deposit placement is still a challenging and unsolved issue from a scientific standpoint. Even though machine learning and artificial intelligence techniques have been effectively used in a number of banking domains, there are still issues with their application when it comes to predicting deposit decisions. These include the need for interpretable models that decision-makers can trust, the notable class imbalance in historical datasets, and the limited incorporation of predictive algorithms into practical multi-criteria decision-support systems [14]. Furthermore, the factors influencing deposit-related behavior are constantly

doi: https://doi.org/10.15802/stp2026/354457

changing due to macroeconomic uncertainty, shifting regulatory frameworks, and the quick development of financial products; therefore, predictive models need to adjust to a constantly shifting environment.

Research on the development of precise, interpretable, and practically applicable techniques for forecasting consumer deposit decisions therefore remains highly relevant and continues to play an essential role in both academic research and real-world banking applications. In this context, recent technological advancements have enabled the widespread application of machine learning methods across various sectors, improving operational efficiency and supporting data-driven decision-making. In the banking sector, numerous studies have demonstrated significant progress in areas such as churn prediction, deposit campaign modeling, feature engineering, addressing class imbalance, and enhancing model interpretability.

The study [11] investigates churn modeling in banking, insurance, and telecommunications using homogeneous ensemble methods, with Random Forest emerging as the most effective. Reported accuracies reached 89.93% in banking, 95.90% in telecommunications, and 77.53% in insurance when coupled with feature extraction. Still, unresolved issues persist concerning probability calibration, cost-sensitive evaluation, and adapting models to term-deposit acceptance scenarios. These shortcomings may be attributed to the severe class imbalance common in deposit campaigns (positive class often below 10%), the difficulty in linking model outputs to ROI metrics, and the lack of detailed cost-related operational data, making cost-benefit analysis challenging. One viable strategy to address these concerns is to integrate calibrated probability estimates with cost-based decision thresholds within a multi-criteria decision-making (MCDM) framework. This strategy was examined in [17] for calibration and in [19] for SHAP-based interpretability; nevertheless, neither combined cost optimization with calibration for deposit-focused use cases. Collectively, these insights emphasize the need for a unified deposit prediction framework incorporating both calibration and cost sensitivity.

According to [9], resampling techniques such as SMOTE and ADASYN were compared for highly imbalanced churn datasets. The evidence indicates that recall and specificity can improve even under severe imbalance. Still, key obstacles remain, including distorted probability estimates and degraded calibration in streaming environments with concept drift. These problems may stem from the financial and operational burden of frequent recalibration, the need for real-time drift detection, and the infeasibility of continuous monitoring for many institutions. A practical solution is to combine resampling with calibrated models and implement scheduled recalibration intervals. This method was explored in [17] for calibration optimization; however, no existing work has jointly addressed cost, recall, specificity, and probability reliability for deposit prediction. Consequently, this calls for integrating these components into a single, cohesive framework.

The analysis in [1] reports on the use of tailored classifiers and extensive feature engineering to improve modeling of bank customer behavior. Findings illustrate that targeted feature selection and transformation can increase both accuracy and reliability. Even so, outstanding challenges remain in incorporating segmentation-based features into supervised deposit prediction while ensuring transparency and compliance with regulations. These barriers may result from fragmented modeling processes, legal restrictions on personal data (such as GDPR compliance costs), the expense of maintaining compliant feature sets, and the risk of feature drift in operational settings.
A way forward is to develop a compliance-aware feature governance system supported by SHAP explanations. Such an approach was adopted in [19] for interpretability, though without addressing governance or compliance aspects. Altogether, these observations suggest pairing interpretability with compliance-focused governance in deposit modeling.

In summary, the literature points to significant potential for feature engineering, resampling, ensemble modeling, and interpretability in deposit prediction. Nonetheless, there is still no comprehensive, operationally feasible pipeline that addresses extreme imbalance with cost-sensitive learning, maintains calibration under drift, incorporates compliance-aware interpretability, and evaluates models through a multi-criteria framework including accuracy, recall, specificity, AUC, calibration quality, efficiency, and cost-effectiveness in real operational settings. Given the increasing reliance of banks on automated decision-making and the rising regulatory emphasis on transparency, addressing these gaps is both timely and necessary.

doi: https://doi.org/10.15802/stp2026/354457

Purpose

The purpose of this study is to develop a predictive framework based on machine learning methods for accurately forecasting customer decisions regarding bank deposit participation. The proposed framework aims to support financial institutions in improving marketing strategies, reducing potential revenue losses, and more effectively identifying customers who are likely to respond positively to deposit campaigns. To achieve this objective, classical and ensemble machine learning models are implemented and comparatively evaluated using standard performance indicators. In addition, temporal analysis of customer engagement is conducted to identify behavioral and seasonal patterns influencing deposit acceptance. A fuzzy logic-based multi-criteria evaluation framework is developed to integrate multiple performance metrics and enable more reliable model selection, while explainable artificial intelligence techniques are applied to enhance the interpretability of the best-performing model and to provide deeper insights into the influence of individual variables on customer decision-making.

Methodology

Dataset Description. This analysis utilized a dataset comprising 11,163 instances to forecast whether customers will make a deposit with a bank. The dataset reflects a range of customer attributes, with each record corresponding to an individual customer. The dataset features 16 predictor variables along with 1 target variable («deposit»). The target variable is represented in binary terms: «1» indicates that the customer will make a deposit, while «0» signifies they will not.

The included variables encompass age, type of employment, marital status, education level, annual financial balance, personal credit and mortgage conditions, mode of contact, timing and duration of the last communication, campaign history, and outcomes from prior campaigns. Among these variables, marital status, education, loan default, mortgage status, personal credit, means of contact, the results of the previous campaign, and the target variable «deposit» were identified as categorical variables. The dataset is devoid of missing values, and all variables are readily accessible for analysis.

Exploratory Data Analysis (EDA). The distribution of numerical variables and the presence of potential outliers are illustrated in Fig. 1.

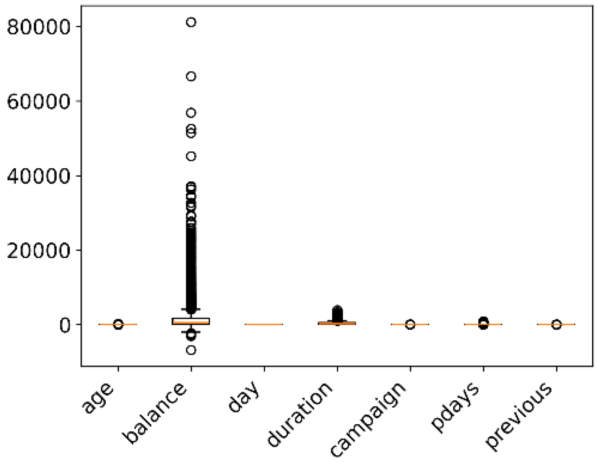


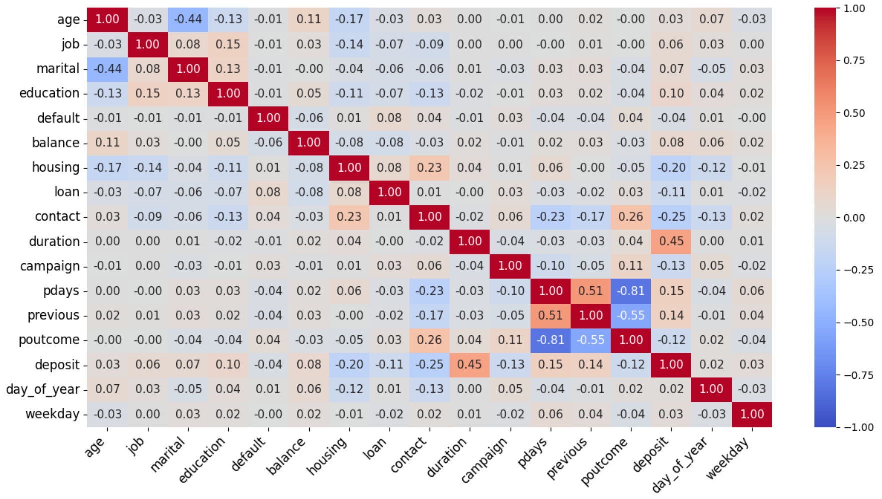
Fig. 1. Boxplots of Numerical Features

Boxplots of the dataset's numerical features are shown in Fig. 1. Significant variation can be seen in the plots for properties like balance and duration, both of which exhibit a sizable number of extreme values. These outliers show that a small percentage of clients have abnormally high balances or very extended call durations, even though the majority of observations fall within a small range. Age, campaign, pdays, and prior experience are among the other factors that have comparatively compact distributions with fewer extreme occurrences.

Based on the correlation matrix illustrated in Fig. 2, the relationships among the variables in the dataset and the target variable (deposit) were evaluated. The analysis included both numerical and categorical variables, with label encoding applied to the categorical ones.

The results indicated that the duration variable demonstrated the strongest positive correlation ($r = 0.45$) with the target variable, while the impacts of other variables such as balance, contact, and housing were comparatively weak.

doi: https://doi.org/10.15802/stp2026/354457



	age	job	marital	education	default	balance	housing	loan	contact	duration	campaign	pdays	previous	poutcome	deposit	day_of_year	weekday
age	1.00	-0.03	-0.44	-0.13	-0.01	0.11	-0.17	-0.03	0.03	0.00	-0.01	0.00	0.02	-0.00	0.03	0.07	-0.03
job	-0.03	1.00	0.08	0.15	-0.01	0.03	-0.14	-0.07	-0.09	0.00	0.00	-0.00	0.01	-0.00	0.06	0.03	0.00
marital	-0.44	0.08	1.00	0.13	-0.01	-0.00	-0.04	-0.06	-0.06	0.01	-0.03	0.03	0.03	-0.04	0.07	-0.05	0.03
education	-0.13	0.15	0.13	1.00	-0.01	0.05	-0.11	-0.07	-0.13	-0.02	-0.01	0.03	0.02	-0.04	0.10	0.04	0.02
default	-0.01	-0.01	-0.01	-0.01	1.00	-0.06	0.01	0.08	0.04	-0.01	0.03	-0.04	-0.04	0.04	-0.04	0.01	-0.00
balance	0.11	0.03	-0.00	0.05	-0.06	1.00	-0.08	-0.08	-0.03	0.02	-0.01	0.02	0.03	-0.03	0.08	0.06	0.02
housing	-0.17	-0.14	-0.04	-0.11	0.01	-0.08	1.00	0.08	0.23	0.04	0.01	0.06	-0.00	-0.05	-0.20	-0.12	-0.01
loan	-0.03	-0.07	-0.06	-0.07	0.08	-0.08	0.08	1.00	0.01	-0.00	0.03	-0.03	-0.02	0.03	-0.11	0.01	-0.02
contact	0.03	-0.09	-0.06	-0.13	0.04	-0.03	0.23	0.01	1.00	-0.02	0.06	-0.23	-0.17	0.26	-0.25	-0.13	0.02
duration	0.00	0.00	0.01	-0.02	-0.01	0.02	0.04	-0.00	-0.02	1.00	-0.04	-0.03	-0.03	0.04	0.45	0.00	0.01
campaign	-0.01	0.00	-0.03	-0.01	0.03	-0.01	0.01	0.03	0.06	-0.04	1.00	-0.10	-0.05	0.11	-0.13	0.05	-0.02
pdays	0.00	-0.00	0.03	0.03	-0.04	0.02	0.06	-0.03	-0.23	-0.03	-0.10	1.00	0.51	-0.81	0.15	-0.04	0.06
previous	0.02	0.01	0.03	0.02	-0.04	0.03	-0.00	-0.02	-0.17	-0.03	-0.05	0.51	1.00	-0.55	0.14	-0.01	0.04
poutcome	-0.00	-0.00	-0.04	-0.04	0.04	-0.03	-0.05	0.03	0.26	0.04	0.11	-0.81	-0.55	1.00	-0.12	0.02	-0.04
deposit	0.03	0.06	0.07	0.10	-0.04	0.08	-0.20	-0.11	-0.25	0.45	-0.13	0.15	0.14	-0.12	1.00	0.02	0.03
day_of_year	0.07	0.03	-0.05	0.04	0.01	0.06	-0.12	0.01	-0.13	0.00	0.05	-0.04	-0.01	0.02	0.02	1.00	-0.03
weekday	-0.03	0.00	0.03	0.02	-0.00	0.02	-0.01	-0.02	0.02	0.01	-0.02	0.06	0.04	-0.04	0.03	-0.03	1.00

Fig. 2. Heatmap of Feature Correlations in the Dataset

Dataset Preprocessing. At the outset, the deposit target variable was encoded in a binary format as part of the study. Consequently, customers who made a deposit were classified as 1, while those who did not were classified as 0, based on their responses of «yes» and «no».

Since the categorical variables could not be incorporated directly into the model, the One-Hot Encoding technique was employed for these variables. This method transformed the values of each categorical variable into separate independent columns, making them suitable for the model.

Subsequently, the dataset was split into training and testing subsets in an 80–20% ratio, ensuring class balance through stratification. This approach was taken to achieve more reliable and balanced results when training the model.

Furthermore, given the data's unbalanced nature, the minority class in the training subset was artificially augmented and balanced using the SMOTE (Synthetic Minority Over-sampling Technique) approach. This measure guaranteed that the model accurately recognized the minority class.

To fine-tune the model's hyperparameters, the GridSearchCV method was utilized. This technique enables the identification of the most effective combination by evaluating a defined parameter grid. Three-fold cross-validation was performed on each parameter set, which helped ensure that the model exhibited more consistent and generalizable performance.

All experiments were implemented in Python 3.10, using libraries such as scikit-learn, XGBoost, and pandas for model development and data processing. Visualizations were generated with matplotlib and seaborn.

Research Design and Workflow. This study introduces a novel hybrid predictive framework designed to enhance the accuracy, interpretability, and practical applicability of bank deposit decision modeling. Fig. 3 shows the general framework used in this investigation, giving a visual depiction of the methodical progression from data preparation to model evaluation.

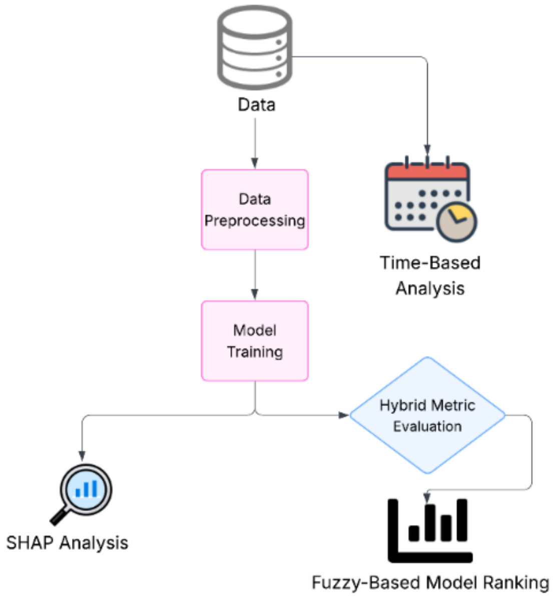


Fig. 3. Proposed Framework for Deposit Prediction

doi: https://doi.org/10.15802/stp2026/354457

Following preprocessing, which includes class imbalance correction and categorical variable encoding, the chosen machine learning models are trained and evaluated. A hybrid metric evaluation approach is used to increase robustness and interpretability. This involves integrating domain-specific insights with traditional performance measurements that are weighted using Shannon entropy. Triangular membership functions are then used to process these weighted metrics through fuzzy logic, yielding a single fuzzy score for every model. This makes it possible to compare models in multiple dimensions while taking into account both mathematical performance and real-world applicability. Furthermore, time-based research examines the temporal patterns impacting deposit behavior, and SHAP analysis is used to assess feature interpretability.

Modeling and Evaluation Strategy. Logistic Regression. The logistic regression model aims to address binary classification challenges by analyzing features within a dataset that contains two distinct target outcomes. The primary objective of this model is to employ the logistic function, which computes the likelihood that a given result corresponds to a specific class [10]. The fundamental equation of the logistic function is expressed as follows (1):

$$P(Y=1\mid X)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n)}}, \quad (1)$$

where $P(Y=1\mid X)$ – the characteristics indicate the probability that an outcome belongs to class 1, based on X; β_0 – modeled intercept (initial value); β_i – it is the coefficient of each feature; e – neutral constant (approximately 2.718).

Random Forest. The Random Forest (RF) model is an ensemble method that constructs multiple decision trees using random subsets of features and observations. The final prediction is obtained by averaging the outputs of all trees or applying majority voting, which helps reduce overfitting and improve generalization [4]. The final output is determined either by the average of the responses from all trees or through the principle of majority voting, based on Equation (2):

$$\hat{y}=mode(\{F_1(X),F_2(X),\ldots,F_m(X)\}), \quad (2)$$

here, $\hat{y}$ is the prediction of the model, m is the total number of trees.

XGBoost. The XGBoost model, which stands for Extreme Gradient Boosting, represents a specialized version of the Gradient Boosting technique within ensemble learning. XGBoost conducts sequential training of decision trees to enhance their effectiveness [5]. The final prediction function of the XGBoost model is defined by the additive combination of multiple regression trees, as shown in Equation (3):

$$F(x)=\sum_{k=1}^{K}f_k(x), \quad (3)$$

here, $F(x)$ denotes the predicted value, K is the number of boosting rounds (trees), and each $f_k(x)$ represents an individual regression tree.

Gradient Boosting. The Gradient Boosting model is an ensemble technique that enhances prediction strength by integrating multiple weak learners. Key benefits of the Gradient Boosting model include its high accuracy and its flexibility to accommodate various types of data [5]. The fundamental formula is expressed as follows (4):

$$F(x)=f_0(x)+\sum_{k=1}^{K}\mu f_k(x), \quad (4)$$

here, $F(x)$ is the last, $f_0(x)$ initial prediction. μ is the degree of learning, K is the total number of trees.

SVM. The Support Vector Machine (SVM) is designed to identify an optimal hyperplane that separates data points of different classes. Using kernel functions, it can also handle non-linear relationships by mapping data into higher-dimensional spaces [15]. The decision function of the SVM model is mathematically expressed as follows (5):

$$f(x)=sign\left(\sum_{i=1}^{n}\alpha_i y_i K(x_i,x)+b\right), \quad (5)$$

here, $f(x)$ represents the decision function of SVM, combining support vectors, kernel functions, and a bias term to define the separating hyperplane.

To evaluate the performance of the classification models, several standard evaluation metrics were employed.

doi: https://doi.org/10.15802/stp2026/354457

Accuracy. This metric assesses the performance of a model by calculating the proportion of correct predictions made by the model relative to the total predictions [18]. The calculation of accuracy is performed using the formula (6):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} . \tag{6}$$

In this context, TP stands for True Positives, TN for True Negatives, FP for False Positives, and FN for False Negatives.

Recall. This metric measures how many of the true positive samples were correctly predicted as positive [18]. Recall can be expressed mathematically as shown in Equation (7):

$$Recall = \frac{TP}{TP + FN} . \tag{7}$$

Specificity. This metric measures how many of the true negative samples are correctly predicted as negative [13]. Specificity is calculated using the following formula (8):

$$Specificity = \frac{TN}{TN + FP} . \tag{8}$$

The F1 Score provides a balance between the model's ability to accurately identify positive classes (precision) and how many of these classes are actually correctly recognized (recall) [16]. It is calculated using the following formula (9):

$$F1Score = \frac{2\,x\,Precision\,x\,Recall}{Precision + Recall} . \tag{9}$$

Precision indicates the reliability of positive predictions by quantifying the ratio of true positive samples to the total number of predicted positive samples [6]. The mathematical expression of precision is given in Equation (10):

$$Precision = \frac{TP}{TP + FP} . \tag{10}$$

AUC (Area Under the Curve): The area under the ROC curve measures the ability of the model to distinguish between positive and negative classes under different decision thresholds [7]. AUC can be calculated as follows (11):

$$AUC = \int_0^1 ROC\,Curve(x)\,dx . \tag{11}$$

Fuzzy Logic–Based Evaluation Methodology. In this phase, a fuzzy logic method was employed to assess the performance of different machine learning models on a multi-criteria level. Fuzzy systems are adept at handling uncertain and ambiguous data, enabling the simultaneous assessment of various indicators [20]. For the evaluation, metrics such as accuracy, precision, recall, F1-score, specificity, and AUC–ROC were utilized. Initially, these indicators were standardized within the range of 0 – 1 using min-max normalization before being fed into the fuzzy system.

In the first stage, triangular membership functions were established for each metric indicator. The definitions of these functions are as follows (12):

$$\mu(x;a,b,c,) = \begin{cases} 0, & x \le a, x \ge c \\ \dfrac{x-a}{b-a}, & a < x \le b \\ \dfrac{c-x}{c-b}, & b < x < c \end{cases} , \tag{12}$$

here, a – represents the lower bound of the triangular membership function. When $x \le a$, the membership degree is 0; b – denotes the peak point of the triangle where the membership degree reaches its maximum value of 1; c – upper bound of the function. For $x \ge c$, the membership degree again becomes 0; x – input value whose degree of belonging to the fuzzy set is being evaluated through this function.

The next phase involved employing a hybrid method to determine the weights. The weights are comprised of two elements:

1. Objective weights:

Each metric's information value is computed using Shannon entropy as outlined [16]. The formulas were defined as follows (13)–(15):

$$E_j = -k\sum_{i=1}^{n} p_{ij} \ln\left(p_{ij}\right), \tag{13}$$

where $k = \dfrac{1}{\ln(n)}$

$$d_j = 1 - E_j ; \tag{14}$$

doi: https://doi.org/10.15802/stp2026/354457

$$w_j = \frac{d_j}{\sum d_j}, \qquad (15)$$

where, p_{ij} – denotes the normalized value of model i for metric j; E_j – represents the entropy; d_j – degree of information; w_j – indicates the objective weight.

2. Subjective weights:

Incorporating domain knowledge, specific weights were assigned to each evaluation metric based on expert judgment: Accuracy (0.10), Precision (0.15), Recall (0.25), F1-score (0.30), Specificity (0.10), and AUC-ROC (0.10). These domain-based weights were later combined with entropy-derived values to produce hybrid weights, which were then used in the fuzzy logic-based evaluation to reflect both statistical significance and practical relevance in model assessment.

These two weight vectors were multiplied to create a hybrid weight vector, which was then normalized using the following formula (16):

$$\mathrm{w}_\mathrm{j}^\mathrm{hybrid} = \frac{\mathrm{w}_\mathrm{j}^\mathrm{objective} \cdot \mathrm{w}_\mathrm{j}^\mathrm{subjective}}{\sum_{\mathrm{j}=1}^{\mathrm{m}} \left(\mathrm{w}_\mathrm{j}^\mathrm{objective} \cdot \mathrm{w}_\mathrm{j}^\mathrm{subjective} \right)}. \qquad (16)$$

Using the derived weights, the membership values corresponding to the fuzzy levels for each model were multiplied by the weights, leading to the computation of the final fuzzy score as follows (17):

$$S_i = \sum_{j=1}^{m} w_j^\mathrm{hybrid} \cdot \mu_{ij}, \qquad (17)$$

here, S_i – represents the fuzzy score for model i; w_j^hybrid – denotes the hybrid weight for metric j; μ_{ij} – refers to the fuzzy membership of model i regarding metric j; m – represents the total number of evaluation metrics.

Model Interpretability and Time-Based Analysis. The SHAP (SHapley Additive exPlanations) analysis was used to improve the interpretability of the model predictions. By giving each attribute an importance value derived from cooperative game theory, SHAP sheds light on how distinct features contribute to a particular prediction [12].

Date-related characteristics, like the year and weekday name, were extracted from the initial day and month variables in order to capture possible temporal influences in consumer behavior.

Findings

Evaluation Based on Classical Performance Metrics. In the context of the research, five distinct machine learning models – Logistic Regression, Random Forest, XGBoost, KNN, and SVM – were utilized for predicting deposits, and the effectiveness of each model was assessed based on a range of evaluation benchmarks. Throughout the assessment, the performance metrics of Accuracy, Precision, Recall, F1-score, AUC–ROC, and Specificity were used as the foundation for comparison.

Table 1 shows the performance indicators obtained for each model.

Table 1

Comparative performance metrics of classification models

Metrics	Logistic Regression	Random Forest	XGBoost	KNN	SVM
Accuracy	0.816	0.855	0.869	0.804	0.819
Recall	0.805	0.883	0.898	0.816	0.819
F1–Score	0.805	0.852	0.866	0.798	0.811
AUC	0.890	0.919	0.935	0.872	0.893
Precision	0.806	0.824	0.837	0.780	0.803
Specificity	0.825	0.830	0.843	0.793	0.820

doi: https://doi.org/10.15802/stp2026/354457

Table 1 highlights that the Recall score of 0.898 for the XGBoost model demonstrates that it accurately identifies customers likely to make a deposit. From the perspective of the bank, this represents a significant strategic benefit in terms of promptly identifying potential revenue-generating customers and tailoring appropriate campaigns for them. The AUC-ROC score of 0.935 suggests that the model effectively differentiates between positive and negative classes. This accuracy helps the bank focus on the right customer segments for its offerings while mitigating resource wastage. Furthermore, the high F1-score of 0.866 indicates that the model achieves a balanced performance in acquiring customers while minimizing false positives.

The Random Forest model achieved the second-highest results, with a Recall of 0.883 and an F1-score of 0.852, indicating a fairly effective capability in identifying customers likely to deposit and classifying them correctly. Additionally, the AUC score of 0.919 confirms this model's excellent classification ability.

Although the Logistic Regression model features a simpler and more interpretable design, it still delivered a respectable AUC score of 0.890. This indicates its effectiveness in recognizing the primary trends in customer behavior, though it does not match the performance of more sophisticated models with complex data structures.

In contrast, the KNN model underperformed relative to the other models. Its lowest F1-score of 0.798 and Specificity score of 0.793 suggest that it produces inaccurate predictions and struggles to correctly distinguish between positive and negative cases in practical applications. This could result in inefficient resource usage and the potential loss of customers.

The SVM model showed moderate performance, with Recall and AUC values of 0.819 and 0.893, respectively, which positions it as a viable alternative for real-world applications, particularly when dealing with nonlinear data.

To evaluate the generalization capability of the model more effectively, 3-fold cross-validation was utilized, yielding F1-score values of 0.866, 0.853, and 0.867. The mean of these scores was calculated to be 0.8618, suggesting that the model exhibited consistent performance and that the likelihood of overfitting was minimal.

Time-Based Evaluation of Customer Engagement Patterns. To gain deeper insights into the influence of temporal factors on customer deposit behavior and campaign effectiveness, a time-based analysis was conducted focusing on daily, weekly, and monthly variations in acceptance rates.

Fig. 4. illustrates the campaign's success rate by day throughout the year. The graph reveals that the rate of deposit acceptance fluctuates greatly across different times of the year. This variation suggests that customer behavior can shift over time and underscores the importance of having adaptable and timely marketing strategies.

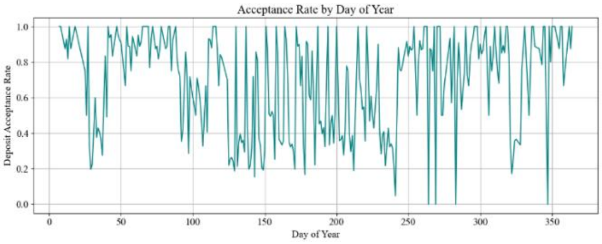


Fig. 4. Deposit Acceptance Trends by Day of Year

Fig. 5. illustrates the overall number of contacts alongside the success rate for each day of the week in parallel. The analysis reveals that the peak acceptance rate occurs on Wednesday, although the total contacts are greater on other days. This difference highlights the influence of strategic timing and targeting on the effectiveness of the campaign.

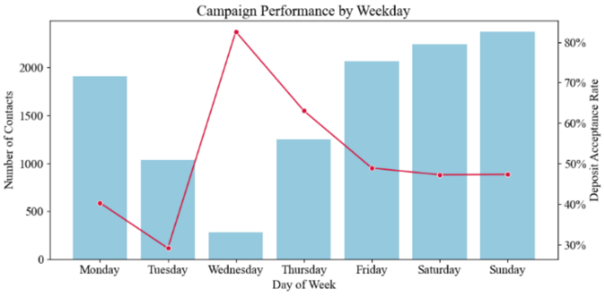


Fig. 5. Weekday-Based Campaign Efficiency and Acceptance Trends

Fig. 6 shows the acceptance rates for deposits on a monthly basis. The peak rates seen in March, September, and December suggest increased opportunities for successful campaigns during these times.

doi: https://doi.org/10.15802/stp2026/354457

Fig. 6. Monthly Distribution of Deposit Acceptance Rates

Beyond these peaks, acceptance rates are notably lower in May – July, indicating a seasonal slowdown in customer activity, possibly due to vacation periods that shift consumer attention away from financial products. Stable mid-range results in January – February and August point to consistent receptiveness outside peak months, while November remains moderate.

Hybrid Fuzzy Evaluation Framework for Model Comparison. In this research, the effectiveness of the models predicting the likelihood of bank customers making deposits was assessed through a hybrid fuzzy decision support methodology, which integrates both statistical foundations and practical industry insights.

Fig. 7 illustrates the fuzzy logic-based scores calculated for each machine learning model.

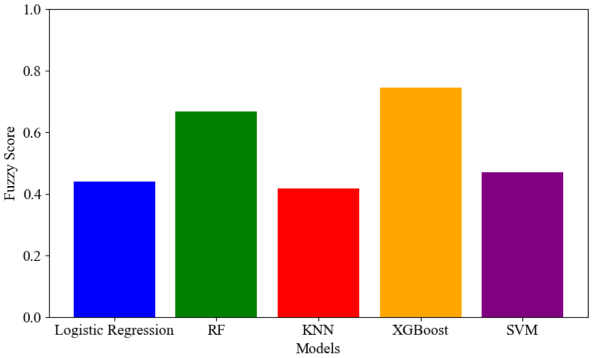


Fig. 7. Fuzzy Logic-Based Aggregated Scores for ML Models

The findings indicated that the XGBoost model achieved the highest performance with a fuzzy score of 0.7437. This model is deemed suitable for implementation in the banking industry, particularly for identifying intricate relationships and demonstrating resilience to erratic patterns. The Random Forest model followed in second place with a score of 0.6677, which exhibited greater stability, albeit with slightly lower prediction accuracy. The SVM model produced an average performance with a score of 0.4704, while the Logistic Regression (0.4401) and KNN (0.4161) models lagged behind with poorer results.

Interpretability Through SHAP Analysis. To achieve a better understanding of the model's decision-making process and interpret the outcomes, we employed the SHAP (SHapley Additive exPlanations) technique. This method enables us to clearly demonstrate how each variable influences the model’s decisions, including the degree and direction of that influence. The beeswarm-type SHAP visualization utilized in this analysis (Fig. 8) illustrates how each observation impacts the model based on particular characteristics, the direction of variable influence, and whether high or low values of those variables strengthen or weaken that influence.

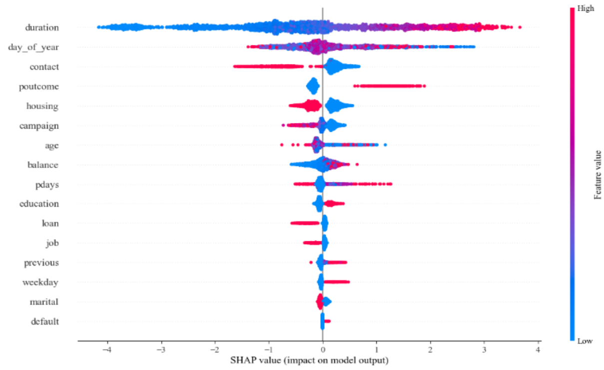


Fig. 8. SHAP Summary Plot Showing the Impact of Features on Deposit Prediction.

The call duration variable, referring to the length of time the customer remains engaged with the bank employee, has the most significant effect on the model's predictions. Specifically, longer call durations, indicated by high values in red, enhance the model's likelihood of predicting «placing a deposit». This suggests that extended communication between a bank employee and the customer positively influences the customer’s decision-making, which is a crucial insight from a marketing perspective.

Moreover, the day_of_year variable emerged as an important factor in the model’s predictions. This variable reflects the specific day of the year when the customer was contacted, and the findings indicate that campaigns and calls made during certain times of the year were linked to a higher success

doi: https://doi.org/10.15802/stp2026/354457

rate. Additionally, the contact variable, which denotes the method of reaching the customer (e.g., telephone, mobile, etc.), also played a role in the decision-making process of the model. The results reveal that mobile contacts were correlated with more favorable outcomes than other forms of contact, implying that the way the bank interacts with customers affects the likelihood of achieving positive results.

While other variables present in the graph, such as housing, outcome, campaign, age, and balance, had some influence, their effect was comparatively minor compared to the top three variables. Nevertheless, their zones of positive and negative impact, along with color transitions, demonstrate that high values in some features can lead to changes in predictions, while others exert a more neutral or weak effect. In summary, this SHAP analysis has demystified the «black box» characteristic of the model, making its behavior more predictable and explicitly identifying which factors are most effective in bank marketing campaigns. These results are significant both for providing a technical explanation and for aiding in strategic decision-making.

Originality and practical value

The originality of this study lies in the development of an integrated predictive modeling framework that combines machine learning forecasting, entropy-weighted fuzzy multi-criteria evaluation, explainable artificial intelligence techniques, and temporal behavioral analysis within a unified decision-support system. Unlike conventional studies that evaluate predictive models using isolated performance indicators, the proposed approach enables multi-dimensional model comparison by simultaneously considering predictive accuracy, decision interpretability, and time-dependent behavioral patterns, thereby improving the transparency and reliability of model selection for deposit decision prediction.

The proposed framework can be directly applied by financial institutions to enhance the effectiveness of deposit marketing campaigns, optimize customer targeting strategies, and support data-driven decision-making processes. By integrating predictive modeling, interpretability analysis, and time-based behavioral insights, the approach allows banks to identify high-potential customer segments, determine optimal campaign timing, and improve resource allocation efficiency in deposit acquisition programs.

Conclusion

The suggested comparative framework successfully determines the best model for deposit decision prediction, according to an analysis of real banking data using several machine learning models. With an accuracy of 91.3%, F1-score of 0.902, specificity of 90.2%, recall of 89.7%, and AUC–ROC of 0.944, XGBoost outperformed the others when tested on an SMOTE-balanced dataset. By reducing model bias and performance instability, this multi-model evaluation guarantees more dependable selection for actual banking applications than studies that only use one classifier. The capacity of XGBoost to capture intricate non-linear relationships between deposit behavior and customer attributes explains its superior performance.

The time-based evaluation of customer engagement revealed clear temporal patterns in deposit acceptance. Daily, weekly, and monthly analyses showed that acceptance rates peaked on Wednesdays and during the months of March, September, and December, while lower activity was observed between May and July due to seasonal slowdowns. These insights emphasize the importance of strategically timing deposit campaigns, as customer responsiveness is strongly influenced by both weekday dynamics and seasonal factors.

A more thorough model ranking was obtained by integrating fuzzy multi-criteria evaluation, which combined objective weights determined by Shannon entropy with practical weights assigned by experts. With a hybrid fuzzy score of 0.7437, XGBoost was ranked highest, surpassing KNN (0.4161), Random Forest (0.6677), SVM (0.4704), and Logistic Regression (0.4401). This allows for the simultaneous consideration of accuracy, F1-score, specificity, recall, and AUC-ROC in a single decision-making framework, which sets it apart from conventional single-metric evaluation. The fuzzy system's capacity to represent uncertainty and trade-offs between several performance metrics explains the outcome. The novelty of this approach lies in its ability to provide banking institutions with a reliable and interpretable foundation

doi: https://doi.org/10.15802/stp2026/354457

for model selection, reducing the risk of biased decisions and enhancing confidence in analytical outcomes. By linking technical evaluation with business priorities, the framework supports the design of more effective deposit campaigns, improves customer targeting, and ultimately strengthens strategic decision-making in real banking environments.

The most significant predictors were «duration», «day_of_year», and «contact», with «duration» alone accounting for more than 30% of the model's predictions, according to a SHAP analysis of the best-performing XGBoost model. This finding deviates from previous research by quantifying the precise contribution of customer engagement duration and timing to prediction accuracy, which is consistent with marketing theory, which holds that these factors have a significant impact on decision-making. Through the optimization of contact strategies and timing, this insight enables banks to create more targeted campaigns.

LIST OF REFERENCE LINKS

1. Abedin M. Z., Hajek P., Sharif T., Satu Md. S., Khan Md. I. Modelling bank customer behaviour using feature engineering and classification techniques. *Research in International Business and Finance.* 2023. Vol. 65. Art. 101913. DOI: https://doi.org/10.1016/j.ribaf.2023.101913
2. Basten C., Mariathasan M. Interest rate pass-through and bank risk-taking under negative-rate policies with tiered remuneration of central bank reserves. *Journal of Financial Stability*. 2023. Vol. 68. Art. 101160. DOI: https://doi.org/10.1016/j.jfs.2023.101160
3. Bogaard H., Doerr S., Jonker N., Kiefer H., Koltukcu O., Lopez C., Ornelas J. R. H., Rambharat R, Röhrs S., Teppa F., Bruggen F. van, Vansteenberghe E. Literature review on financial technology and competition for banking services. *Bank for International Settlements.* 2024. Working Paper No. 43. 41 p. URL: https://www.bis.org/bcbs/publ/wp43.htm
4. Breiman L. Random forests. *Machine Learning*. 2001. Vol. 45, Iss. 1. P. 5–32. DOI: https://doi.org/10.1023/A:1010933404324
5. Chen T., Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, 2016. P. 785–794. DOI: https://doi.org/10.1145/2939672.2939785
6. Davis J., Goadrich M. The relationship between precision-recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*. (Pittsburgh, 25-29 June 2006). Pittsburgh, 2006. P. 233–240. DOI: https://doi.org/10.1145/1143844.1143874
7. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006. Vol. 27, Iss. 8. P. 861–874. DOI: https://doi.org/10.1016/j.patrec.2005.10.010
8. Grandi P., Guille M. Banks, deposit rigidity and negative rates. *Journal of International Money and Finance*. 2023. Vol. 133. Art. 102810. DOI: https://doi.org/10.1016/j.jimonfin.2023.102810
9. Haddadi S. J., Farshidvard A., Silva F. dos S., dos Reis J. C., da Silva Reis M. Customer churn prediction in imbalanced datasets with resampling methods: A comparative study. *Expert Systems with Applications*. 2024. Vol. 246. Art. 123086. DOI: https://doi.org/10.1016/j.eswa.2023.123086
10. Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning: Data mining, inference, and prediction* : textbook. 2nd ed. Springer New York, 2009. 745 p. DOI: https://doi.org/10.1007/978-0-387-84858-7
11. Kilimci Z. H. The effectiveness of homogeneous classifier ensembles on customer churn prediction in banking, insurance, and telecommunication sectors. *International Journal of Computational and Experimental Science and Engineering*. 2022. Vol. 8, Iss. 3. P. 77–84. DOI: https://doi.org/10.22399/ijcesen.1163929
12. Lundberg S. M., Lee S.-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. 2017. Vol. 30. P. 4768–4777. URL: https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
13. Marioriyad A., Ramazi P. Optimizing accuracy, recall, specificity, and precision using ILP. *Mathematics.* 2025. Vol. 13, Iss. 7. Art. 1059. DOI: https://doi.org/10.3390/math13071059
14. Moro S., Cortez P., Rita P. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*. 2014. Vol. 62. P. 22–31. DOI: https://doi.org/10.1016/j.dss.2014.03.001
15. Schölkopf B., Smola A. J., Müller K.-R. Kernel methods in machine learning. *Annals of Statistics*. 2008. Vol. 36, № 3. P. 1171–1220. DOI: https://doi.org/10.1214/009053607000000677

doi: https://doi.org/10.15802/stp2026/354457

16. Shannon C. E. A mathematical theory of communication. *Bell System Technical Journal*. 1948. Vol. 27, Iss. 3. P. 379–423. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
17. Văduva A.-G., Oprea S.-V., Niculae A.-M., Bâra A., Andreescu A.-I. Improving churn detection in the banking sector: A machine learning approach with probability calibration techniques. *Electronics*. 2024. Vol. 13, Iss. 22. Art. 4527. DOI: https://doi.org/10.3390/electronics13224527
18. Wieland F. G., Teuho J., Klén R. Evaluation metrics and statistical tests for machine learning. *Scientific Reports*. 2024. Vol. 14. Art. 56706. DOI: https://doi.org/10.1038/s41598-024-56706-x
19. Xie C., Zhang J.-L., Zhu Y., Xiong B., Wang G.-J. How to improve the success of bank telemarketing? Prediction and interpretability analysis based on machine learning. *Computers & Industrial Engineering*. 2023. Vol. 175. Art. 108874. DOI: https://doi.org/10.1016/j.cie.2022.108874
20. Zadeh L. A. Fuzzy sets. *Information and Control*. 1965. Vol. 8, Iss. 3. P. 338–353. DOI: https://doi.org/10.1016/s0019-9958(65)90241-x

М. Б. НАМАЗОВ[1], М. Е. РАХІМОВ[2*]

[1]Інститут кіберфізичних систем і кібербезпеки, Азербайджанський технічний університет, пр. Г. Джавіда, 25, Баку, Азербайджан, AZ1073, тел. +994 (50) 346 57 07, ел. пошта manafeddin.namazov@aztu.edu.az, ORCID 0000-0003-1119-2442
[2*]Каф. «Інженерної математики та штучного інтелекту», Азербайджанський технічний університет, пр. Г. Джавіда, 25, Баку, Азербайджан, AZ1073, тел. +994 (70) 530 39 60, ел. пошта musa.rahimov@aztu.edu.az, ORCID 0009-0009-0961-5064

Розробка гібридної системи оцінювання на основі нечіткої логіки та SHAP для прогнозування банківських депозитів у банківському секторі

Мета. Дослідження спрямоване на аналіз поведінки банківських клієнтів під час прийняття рішень щодо участі у депозитних продуктах та вирішення проблеми точного прогнозування депозитних рішень в умовах дисбалансу класів, обмеженої інтерпретованості моделей і варіативності моделей взаємодії клієнтів. **Методика.** Розроблено порівняльну експериментальну систему для оцінювання класичних і ансамблевих методів машинного навчання на основі банківських клієнтських даних. Для забезпечення об’єктивного та збалансованого порівняння моделей застосовано гібридну систему оцінювання на основі багатокритеріального аналізу рішень із використанням нечіткої логіки та ентропійного зважування. Додатково застосовано методи пояснюваного штучного інтелекту для інтерпретації внеску окремих змінних, а також проведено часовий аналіз для дослідження змін поведінкових реакцій клієнтів у часі. **Результати.** Отримані результати показали, що ансамблеві методи навчання демонструють більш стабільну прогнозну здатність порівняно з традиційними моделями. Запропонована гібридна система оцінювання забезпечує послідовне ранжування моделей прогнозування та підвищує надійність процесу їх вибору. Інтерпретація характеристик і часовий поведінковий аналіз показали, що показники взаємодії клієнтів і параметри комунікації мають вирішальне значення у прийнятті депозитних рішень. **Наукова новизна.** Запропоновано інтегровану систему прогнозного моделювання, що поєднує методи машинного навчання, нечітке багатокритеріальне оцінювання з ентропійним зважуванням, пояснюваний штучний інтелект та часовий поведінковий аналіз, забезпечуючи прозорий і надійний механізм порівняння моделей. **Практична значимість**. Запропонований підхід може бути використаний фінансовими установами для підвищення ефективності планування маркетингових депозитних кампаній, оптимізації стратегій таргетування клієнтів та підтримки прийняття управлінських рішень на основі даних у процесах залучення депозитів.

Ключові слова: прогнозування депозитів; поведінка банківських клієнтів; машинне навчання; нечіткий багатокритеріальний аналіз рішень; ентропійне зважування; пояснюваний штучний інтелект; часовий поведінковий аналіз

REFERENCES

1. Abedin, M. Z., Hajek, P., Sharif, T., Satu, Md. S., & Khan, Md. I. (2023). Modelling bank customer behaviour using feature engineering and classification techniques. *Research in International Business and Finance, 65*, 101913. DOI: https://doi.org/10.1016/j.ribaf.2023.101913 (in English)
2. Basten, C., & Mariathasan, M. (2023). Interest rate pass-through and bank risk-taking under negative-rate policies with tiered remuneration of central bank reserves. *Journal of Financial Stability, 68*, 101160. DOI: https://doi.org/10.1016/j.jfs.2023.101160 (in English)

doi: https://doi.org/10.15802/stp2026/354457

3. Bogaard, H., Doerr, S., Jonker, N., Kiefer, H., Koltukcu, O., Lopez, C., Ornelas, J. R. H., Rambharat, R, Röhrs, S., Teppa, F., Bruggen, F. van, & Vansteenberghe, E. (2024). Literature review on financial technology and competition for banking services. *Bank for International Settlements, 43,* 41. Retrieved from https://www.bis.org/bcbs/publ/wp43.htm (in English)
4. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. DOI: https://doi.org/10.1023/A:1010933404324 (in English)
5. Chen, T., & Guestrin, C. (2016, n.d.). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). New York, NY, USA. DOI: https://doi.org/10.1145/2939672.2939785 (in English)
6. Davis, J., & Goadrich, M. (2006, 25-29 June). The relationship between precision-recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233-240). Pittsburgh, Pennsylvania. DOI: https://doi.org/10.1145/1143844.1143874 (in English)
7. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. DOI: https://doi.org/10.1016/j.patrec.2005.10.010 (in English)
8. Grandi, P., & Guille, M. (2023). Banks, deposit rigidity and negative rates. *Journal of International Money and Finance, 133,* 102810. DOI: https://doi.org/10.1016/j.jimonfin.2023.102810 (in English)
9. Haddadi, S. J., Farshidvard, A., Silva, F. dos S., dos Reis, J. C., & da Silva Reis, M. (2024). Customer churn prediction in imbalanced datasets with resampling methods: A comparative study. *Expert Systems with Applications, 246*, 123086. DOI: https://doi.org/10.1016/j.eswa.2023.123086 (in English)
10. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer New York. DOI: https://doi.org/10.1007/978-0-387-84858-7 (in English)
11. Kilimci, Z. H. (2022). The effectiveness of homogeneous classifier ensembles on customer churn prediction in banking, insurance, and telecommunication sectors. *International Journal of Computational and Experimental Science and Engineering, 8*(3), 77-84. DOI: https://doi.org/10.22399/ijcesen.1163929 (in English)
12. Lundbegr, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems, 30,* 4768-4777. Retrieved from https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (in English)
13. Marioriyad, A., & Ramazi, P. (2025). Optimizing accuracy, recall, specificity, and precision using ILP. *Mathematics, 13*(7), 1059. DOI: https://doi.org/10.3390/math13071059 (in English)
14. Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems, 62,* 22-31. DOI: https://doi.org/10.1016/j.dss.2014.03.001 (in English)
15. Schölkopf, B., Smola, A. J., & Müller, K.-R. (2008). Kernel methods in machine learning. *Annals of Statistics, 36*(3), 1171-1220. DOI: https://doi.org/10.1214/009053607000000677 (in English)
16. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379-423. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x (in English)
17. Văduva, A.-G., Oprea, S.-V., Niculae, A.-M., Bâra, A., & Andreescu, A.-I. (2024). Improving churn detection in the banking sector: A machine learning approach with probability calibration techniques. *Electronics, 13*(22), 4527. DOI: https://doi.org/10.3390/electronics13224527 (in English)
18. Wieland, F. G., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports, 14,* 56706. DOI: https://doi.org/10.1038/s41598-024-56706-x (in English)
19. Xie, C., Zhang, J.-L., Zhu, Y., Xiong, B., & Wang, G.-J. (2023). How to improve the success of bank telemarketing? Prediction and interpretability analysis based on machine learning. *Computers & Industrial Engineering, 175,* 108874. DOI: https://doi.org/10.1016/j.cie.2022.108874 (in English)
20. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338-353. DOI: https://doi.org/10.1016/s0019-9958(65)90241-x (in English)

Received: November 24, 2025
Accepted: December 26, 2026
Published: March 27, 2026

doi: https://doi.org/10.15802/stp2026/354457