



Comparative Performance Study of Machine Learning Algorithms for Marketing Campaign Response Prediction

Zhila Rahimi^{1*}, Fitria Nurhayati², Muhamad Faiz³,
Laksana Pratama Jaya Ningrat⁴, Sri Rezeki Utami⁵, Nopri Darusman⁶

^{1,2,3}Department of Information System, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

⁴Department Mukala Hadramaut, Universitas Al-Iman, Yemen

⁵Department of Arabic, Faculty of Dirasat Islamiyyah wal 'Arabiyyah, Al-Azhar University, Egypt

⁶Department of Shariah and Law, Faculty of Shariah, International University of Africa, Sudan

E-Mail : ¹12350320226@students.uin-suska.ac.id , ²12350321057 @students.uin-suska.ac.id,
³12350312931@students.uin-suska.ac.id, ⁴laksanaduri@gmail.com,
⁵sri.rezekiutami100702@gmail.com, ⁶novridarusman01@gmail.com

Received Aug 18th 2026; Revised Sep 25th 2026; Accepted Sep 28th 2026; Available Online Oct 05th 2026

Corresponding Author: Zhila Rahimi

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

This study presents a comparative analysis of five Machine Learning (ML) classification algorithms, namely Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), XGBoost, and Gradient Boosting Classifier (GBC), for predicting customer responses to marketing campaigns. The main challenge in this task is the highly imbalanced distribution of customer response data, which may cause biased model predictions. Therefore, this study evaluates model performance using precision, recall, and F1-score metrics to provide a more reliable assessment. Experiments were conducted using the Customer Personality Analysis dataset obtained from Kaggle. The results show that ensemble-based algorithms achieved more consistent performance than individual classifiers. XGBoost obtained the highest F1-score of 84.85% with an 80:20 data split, followed by Random Forest with 84.72%. In contrast, SVM showed lower stability, with its F1-score decreasing from 83.59% to 78.66% on the 90:10 data split. This study provides empirical evidence on the effectiveness of ensemble learning approaches, particularly XGBoost, in handling the complexity of customer behavior and supporting more targeted, data-driven marketing strategies.

Keywords: Classification, Customer Response Prediction, Machine Learning, Marketing Analytics, Predictive Analysis

1. INTRODUCTION

Global digital transformation has triggered a fundamental paradigm shift in marketing. Companies are no longer using traditional mass marketing methods, but are shifting to a more personalized and data-driven approach. In today's business ecosystem, data generated by customer interactions (Big Data) has become a highly valuable strategic asset, requiring companies to significantly enhance their analytical capabilities to remain competitive. The cornerstone of today's marketing strategy is efficient Customer Relationship Management (CRM). The primary goal of CRM is to optimize customer lifetime value by gaining a deep understanding of each individual's profile. Consumer profiles, which include demographic information, transaction records, and psychological traits, serve as the basis for developing effective marketing strategies [2], [3]. Recent industry reports indicate that organizations are increasingly adopting artificial intelligence (AI) and predictive analytics to enhance customer engagement, personalize marketing, and optimize campaign effectiveness. These data-driven approaches enable organizations to analyze customer behavior patterns, improve decision-making processes, and allocate marketing resources more efficiently [4].

Despite the abundance of available information, many companies still face inefficiencies due to unsegmented marketing practices. Strategies that are not tailored to individuals often lead to significant budget waste and increase the likelihood of customer burnout, as promotions do not align with their interests or needs [5]. The inability to transform raw data into accurate predictive analysis is a major obstacle in identifying the most promising market segments. The technical challenges are further complicated by the fact that customer response behavior data is typically characterized by a highly imbalanced class distribution [6], [7], [8]. The proportion of customers who respond positively to a campaign is usually much lower than that of those who

decline, causing machine learning models to become biased toward the majority class [9]. This condition may lead to misleading evaluation results because models can achieve high accuracy by primarily predicting the majority class while failing to identify potential customers from the minority class.

When faced with imbalanced data, model evaluation requires performance metrics beyond accuracy. The goal is to reduce false positives (sending expensive promotions to non-respondents) and false negatives (failing to identify potential buyers). Therefore, metrics such as precision, recall, and F1-score are more suitable for evaluating marketing response prediction because they provide a more balanced assessment of minority-class classification performance, particularly when dealing with imbalanced datasets [10], [11]. To address these prediction challenges, Machine Learning (ML) with Supervised Learning methods provides a robust framework. Supervised Learning algorithms are designed to understand the mapping from input characteristics (customer profiles) to target variables (response or non-response). The application of ML enables the automation of customer classification, eliminating subjective heuristic assumptions [12]. Preliminary studies often favor simple, linear models. Logistic Regression [13], for example, offers a simple, interpretable probability model. On the other hand, Support Vector Machine (SVM) [14], [15], [16], [17] is renowned for its ability to identify optimal hyperplanes that separate classes, even in high-dimensional feature spaces, making it a powerful choice for complex customer data.

However, individual models often fail to capture complex nonlinear interactions in customer data, particularly in imbalanced datasets. Therefore, focus has shifted to ensemble approaches. Random Forest (RF) [18], [19], which uses bagging to reduce variance and overfitting, has shown highly stable results. In contrast, boosting methods, including Gradient Boosting Classifier (GBC) [20] and XGBoost [7], work sequentially to correct errors from previous prediction models, often resulting in better classification performance, particularly when evaluated using imbalance-sensitive metrics such as precision, recall, and F1-score.

In line with the development of machine learning approaches in marketing analytics, recent studies have demonstrated the ability of these techniques to extract customer insights and support data-driven marketing decisions [23]. El-Hajj and Pavlova [1] developed a customer response prediction model for marketing campaigns using a Decision Tree approach. The study evaluated model performance using accuracy, precision, recall, and F1-score and demonstrated that resampling techniques improved the model's ability to handle class imbalance. However, the study focused on a single classification algorithm, limiting comparison with other machine learning approaches.

Gülder and Cevher [21] compared several classification algorithms, including Random Forest, Gradient Boosting, Logistic Regression, Support Vector Classifier, and Decision Tree, using data from digital marketing campaigns. The models were evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis, with Gradient Boosting achieving the highest prediction performance. Similarly, Lin [12] evaluated SVM, XGBoost, CatBoost, and BPANN for consumer behavior prediction using the Online Shoppers Purchasing Intention Dataset. The results showed that XGBoost and CatBoost achieved strong predictive performance, as measured by F1-score and ROC-AUC.

Furthermore, Boozary et al. [28] conducted a comparative analysis of Logistic Regression, KNN, Random Forest, Gradient Boosting, and XGBoost for customer retention prediction using a telecommunications dataset. The study found that ensemble models achieved superior performance across accuracy, precision, recall, and F1-score. Chen et al. [25] also demonstrated the effectiveness of XGBoost compared with Random Forest for churn prediction using precision and recall metrics. However, these studies mainly focused on churn prediction or customer retention rather than on predicting marketing campaign response.

Based on previous studies, machine learning algorithms have demonstrated promising capabilities in customer analytics and marketing prediction. However, existing studies still lack a comprehensive evaluation of individual and ensemble classifiers under the same experimental setting. Many previous studies evaluate only a limited number of algorithms, rely mainly on accuracy metrics, or do not examine model robustness across different data-partitioning scenarios. Therefore, a direct comparison between individual classifiers (LR and SVM) and ensemble-based approaches (RF, XGBoost, and GBC) remains necessary, particularly for imbalanced customer response datasets.

Thus, this study specifically proposes a comparative analysis of five classification algorithms (Logistic Regression, SVM, Random Forest, XGBoost, and Gradient Boosting Classifier). This study aims to empirically evaluate the performance of algorithms for recognizing patterns in customer profile data using precision, recall, and F1-score across different data split configurations. The research provides empirical insights into selecting reliable machine learning models for predicting marketing campaign responses under imbalanced data conditions.

2. MATERIAL AND METHOD

The research methodology was designed to provide a systematic framework for data preparation, model development, evaluation, and comparison of machine learning algorithms. The stages consist of data collection, preprocessing, data partitioning, model training, and performance evaluation, as illustrated in Figure 1.

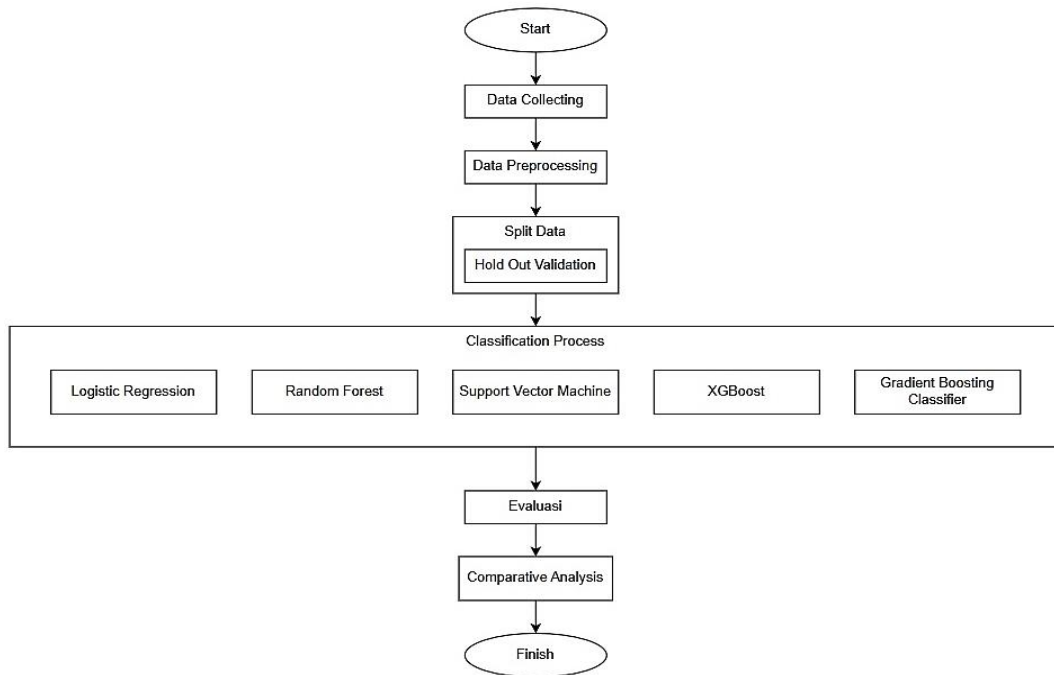


Figure 1. Research Methodology

2.1 Data Collecting

The data in this study is drawn from the Customer Personality Analysis dataset obtained from Kaggle. The dataset consists of 2,240 customer records with numerical and categorical attributes, including customer demographic, behavioral, and purchase-related information. The target variable represents customer responses to marketing campaigns. The dataset presents an imbalanced class distribution, with non-responsive customers outnumbering responsive customers. Therefore, this study evaluates model performance using precision, recall, and F1-score to provide a more balanced assessment than accuracy alone. The class distribution of the dataset is presented in Table 1.

Table 1. Class Distribution of Customer Response Dataset

Class	Samples	Percentage
Non-response (0)	1,906	85.09%
Response (1)	334	14.91%

2.2 Data Preprocessing

Once data collection is complete, the next step is to prepare the data for analysis with machine learning models. This preprocessing stage includes filling in missing data, removing duplicates, checking for inconsistencies, cleaning the data, and correcting any detected errors [4]. In this study, the data used were organized and had no missing values, so no further cleaning was required. Data transformation was performed by removing irrelevant attributes and converting categorical data to a numeric format. Categorical variables were encoded as numerical values to enable processing by machine learning algorithms. Numerical features were maintained in their original format, while irrelevant identifiers that did not contribute to prediction were removed. The preprocessing stage was performed consistently before model training to ensure comparable evaluation among all algorithms.

2.3 Hold Out Validation

The hold-out method divides the data into two parts: training data and testing data. The training data is used to design the model, while the testing data is used to evaluate the performance of the model generated from the training data. However, this method has a weakness: the evaluation results can vary significantly depending on how the data is split into training and testing sets [19]. In this study, two data partitioning configurations were evaluated: 80:20 and 90:10 training-testing splits. The training data was used to construct the classification models, while the testing data was used to evaluate their generalization performance.

2.4 Validation Procedure

Each algorithm was trained and evaluated using the same data partition configurations to ensure a consistent comparison. The performance of each model was measured using precision, recall, and F1-score.

The model with the highest F1-score across different split configurations was considered the best-performing classifier.

2.5 Algorithm Parameter

To ensure a fair comparison among the machine learning algorithms, each model was implemented using predefined parameter configurations. The parameter settings used for each classifier are presented in Table 2.

Table 2. Algorithm Parameter Configuration

Algorithm	Parameter
Logistic Regression	max_iter=1000, random_state=42
Random Forest	n_estimators=100, random_state=42
SVM	kernel=RBF, C=1.0
XGBoost	n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42
GBC	n_estimators=100, learning_rate=0.1, random_state=42

2.6 Customer Personality Dataset

Consumer personality, from a customer behavior perspective, encompasses specific traits, characteristics, and ways of thinking, feeling, and acting that influence how individuals interact with the marketplace. It encompasses both enduring qualities and situational factors that influence consumer preferences, decision-making processes, and brand choices. Companies can use this understanding of customer characteristics to tailor marketing strategies, create tailored brand experiences, and design appropriate communication approaches. Personality traits significantly influence how consumers make choices, communicate with brands, and express their interest [5].

2.7 Confusion Matrix

A confusion matrix is a table that details how test data are classified as accurate or inaccurate. An example of a confusion matrix for a two-category classification is shown in Figure 2 [8].

		Predicted Class	
		1 (Positive)	0 (Negative)
Actual Class	1 (Positive)	TP (True Positive)	FN (False Negative)
	0 (Negative)	FP (False Positive)	TN (True Negative)

Figure 2. Confusion matrix

The confusion matrix formula for calculating accuracy, precision, and recall uses Equations (1), (2), (3), and (4) [20].

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total}} \quad (1)$$

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

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

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

2.8 Logistic Regression

Logistic Regression is a machine learning algorithm used to measure the relationship between one or more independent variables (features) and a categorical dependent variable. Unlike linear regression, which predicts continuous values, Logistic Regression is used to estimate classes, usually binary (0 or 1), such as spam and non-spam, positive and negative, or agree and disagree. Logistic Regression estimates the probability that an input belongs to a particular class using the sigmoid function, which maps the output to the range (0, 1). This probability value is then compared with a threshold (typically 0.5) to determine the category.

In Natural Language Processing (NLP), Logistic Regression is often used for text classification, such as news theme classification, sentiment analysis, and spam detection, using numerical features generated by

methods like Bag-of-Words or TF-IDF [10]. For two-class classification, the Logistic Regression model is formulated as in equation 5.

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n = w^T x \quad (5)$$

Where: z is Linear score (linear combination of features); $x = [x_1, x_2, \dots, x_n]$ is the input features; and $w = [\beta_0, \beta_1, \dots, \beta_n]$ is Weights.

2.9 Random Forest

Random Forest, an ensemble method known for its reliability in handling large datasets and for reducing overfitting through random feature selection, was first introduced by Ho (1995) and further developed by Breiman (2001) [15], [16]. Random Forest effectively handles high-dimensional data and is well-suited to complex applications such as landslide vulnerability prediction and health condition classification [15]. The formula used in the Random Forest method is given by equation 5.

$$H(X) = \arg \max_c \sum_{k=1}^K I(h_k(x) = c) \quad (5)$$

This equation specifies the final prediction $H(x)$ of the Random Forest for input x . The $\arg \max$ function determines the class c with the highest number of votes, where $h_k(x)$ is the class prediction of the k th tree k_1 , and $I(\cdot)$ is an indicator function that has a value of 1 if the k th tree chooses class c , or 0 otherwise. The summation is performed for each tree in the forest, and the class with the most votes becomes the final prediction.

2.10 Support Vector Machine

Support Vector Machine (SVM) is a machine learning method for classification and regression that utilizes kernel functions to optimize the margin between classes using an appropriate hyperplane. SVM works effectively on high-dimensional data [11], [12]. SVM determines the best hyperplane that separates the data classes, defined by the support vectors, thereby increasing the margin and improving the model's generalization [11]. SVM accuracy results are influenced by the parameters and kernel functions used [21]. SVM is considered effective at handling difficult classification tasks, especially with high-dimensional data, by using an optimal hyperplane to enlarge the margin, even when the data is not linearly separable [13], [16]. SVM also offers the advantage of clear visualization of data separation, making it useful in applications such as pattern recognition and satellite image classification [13], [14]. Equation 6 describes the SVM formulation for obtaining the hyperplane with maximum margin [12].

$$\min \frac{1}{2} \|w\|^2, y_1(w \cdot x_1 + b) \geq 1 \quad (6)$$

This SVM equation minimizes the objective function $\frac{1}{2} \|w\|^2$, to enlarge the margin between classes, where $\|w\|$ is the norm of the weight vector w that determines the direction of the hyperplane. The constraint $y_1(w \cdot x_1 + b) \geq 1$ guarantees that the data is on the correct side of the hyperplane with a distance of at least 1.

2.11 XGBoost

XGBoost was introduced by Chen and Guestrin as an optimized implementation of gradient boosting decision trees, designed to improve computational efficiency and predictive performance. XGBoost is a machine learning library for tree-based classification and prediction tasks. This algorithm allows optimization that can be 10 times faster than other Gradient Boosting Machines (GBMs). The accuracy of the classification results is determined by the applied parameters. Both XGBoost and random forests are algorithms consisting of several decision trees. However, unlike random forests, which use bagging, XGBoost relies on boosting to compile its algorithm [6]. XGBoost has been extensively applied in marketing analytics and customer behavior prediction, demonstrating superior performance on large-scale and imbalanced datasets [22].

2.12 Gradient Boosting Classifier

Gradient Boosting is a machine learning method that uses ensemble techniques. The ensemble approach combines two machine learning techniques to achieve superior performance compared to a single method. The goal of ensemble learning is to overcome the weaknesses of a particular classifier by utilizing the strengths of another classifier. Ensemble learning has been shown to improve classification results. Kang and colleagues applied ensemble learning with Hidden Markov Models to improve text-based classification results, demonstrating that the accuracy achieved in this study exceeded that of individual classifiers. Zhang and

colleagues proposed ensemble learning for sentiment classification, demonstrating that ensemble methods can integrate predictions from multiple classifiers to achieve better performance, thereby ensuring generalization and stability in classification results. Based on these results, this study also applied a well-known ensemble method, namely Gradient Boosting [23].

2.11 Generative AI Declaration

Generative AI tools were used only for language improvement and assistance with manuscript formatting. The research methodology, data processing, model implementation, result interpretation, and scientific conclusions were independently conducted and verified by the authors.

3. RESULTS AND DISCUSSION

In this phase, the dataset used in this study was obtained from Kaggle and comprises 2,240 customer records with 10 attributes related to customer characteristics, purchasing behavior, preferences, and segmentation for more effective marketing. This dataset is still in its original form and will serve as the basis for the analysis. The dataset was then prepared through preprocessing steps, including attribute selection and categorical data transformation. Table 3 outlines the details of each attribute, including its data type and description, prior to the initial process.

Table 3. Initial Data

No	Attribute Name	Data Type	Description
1	ID	Numerical	Customer's unique identifier
2	Year_Birth	Numerical	Customer's birth year
3	Education	Categorical	Education Qualification of the customer
4	Marital_Status	Categorical	Marital Status of the customer
5	Income	Numerical	Customer's yearly household income
6	Kidhome	Numerical	Number of children in the customer's household
7	Teenhome	Numerical	Number of teenagers in the customer's household
8	Dt_Customer	Numerical	Date of the customer's enrollment with the company
9	Recency	Numerical	Number of days since the customer's last purchase
10	MntWines	Numerical	Amount Spent On Wine

3.1 Data Preprocessing

The first step in machine learning is data preprocessing. At this stage, the data is transformed by removing irrelevant attributes and converting categorical data into numeric values. The purpose of this data transformation is to avoid errors during model building and help the model better understand the data. The attributes removed were ID and birth year, because they were considered irrelevant identifiers and demographic attributes that could introduce noise into the prediction process. Table 4 shows the dataset that has undergone preprocessing.

Table 4. Preprocessing Data Result

Income	Kidhome	Teenhome	...	Complain	Z_CostContact	Z_Revenue
2.216.000.000	2.240.000.000	2.240.000.000	...	2.240.000.000	2240.0	2240.0
52.247.251.354	0.444196	0.506250	...	0.009375	3.0	11.0
25.173.076.661	0.538398	0.544538	...	0.096391	0.0	0.0
1.730.000.000	0.000000	0.000000	...	0.000000	3.0	11.0
35.303.000.000	0.000000	0.000000	...	0.000000	3.0	11.0
51.381.500.000	0.000000	0.000000	...	0.000000	3.0	11.0
68.522.000.000	1.000.000	1.000.000	...	0.000000	3.0	11.0
666.666.000.000	2.000.000	2.000.000	...	1.000.000	3.0	11.0

3.2 Logistic Regression

In this study, the classification process was implemented using the Logistic Regression algorithm model. Table 5 shows the evaluation results for the LR algorithm.

Table 5. LR Performance Evaluation

Split Data	Precision	Recall	F1-Score
80:20	83.45%	85.93%	83.13%
90:10	82.96%	84.82%	82.03%

Table 6 displays the performance assessment of the LR algorithm based on Precision, Recall, and F1 Score for two data proportions, 80:20 and 90:10. With the 80:20 split, the algorithm achieved a Precision of 83.45%, a Recall of 85.93%, and an F1 Score of 83.13%, indicating well-balanced performance. Meanwhile,

with the 90:10 split, the algorithm achieved a Precision of 82.96%, a Recall of 84.82%, and an F1 Score of 82.03%, slightly lower than the 80:20 split. These results indicate that the 80:20 split provides better performance due to the larger training data. Figure 3 shows that the F1-score of the LR algorithm with the 80:20 split reached 83.13%, higher than that of the 90:10 split, which achieved 82.03%.

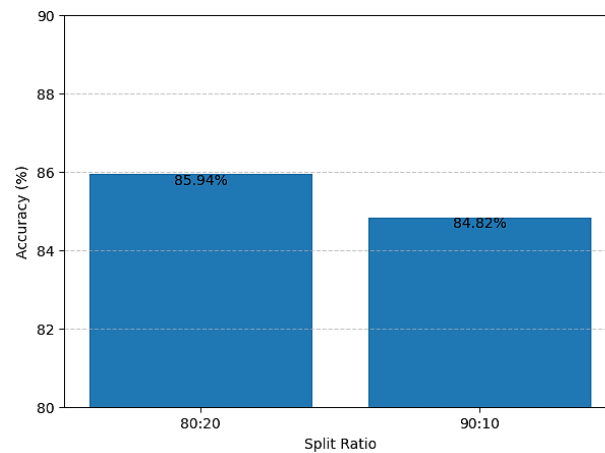


Figure 3. F1-score Performance of Logistic Regression

3.3 Random Forest

In this study, the Random Forest algorithm was used to predict customer responses toward marketing campaigns based on customer characteristics. Random Forest is ideal for this type of structured customer data because it can handle many features with varying levels of importance, such as income, age, total spending, and product choices. Table 6 shows the results of the standard evaluation of the RF algorithm.

Table 6. RF Performance Evaluation

Split Data	Precision	Recall	F1-Score
80:20	84.94%	86.83%	84.72%
90:10	84.20%	85.26%	82.01%

Table 6 displays the performance assessment of the RF algorithm based on Precision, Recall, and F1 Score for two data proportions: 80:20 and 90:10. With the 80:20 split, the algorithm achieved a Precision of 84.94%, a Recall of 86.83%, and an F1 Score of 84.72%, indicating well-balanced performance. Meanwhile, with the 90:10 split, the algorithm achieved a Precision of 84.20%, a Recall of 85.26%, and an F1 Score of 82.01%, slightly below those of the 80:20 split. These results indicate that the 80:20 split provides better performance due to the larger training data. Figure 4 shows that the F1-score of the RF algorithm with the 80:20 split reached 84.72%, higher than that of the 90:10 split, which achieved 82.01%.

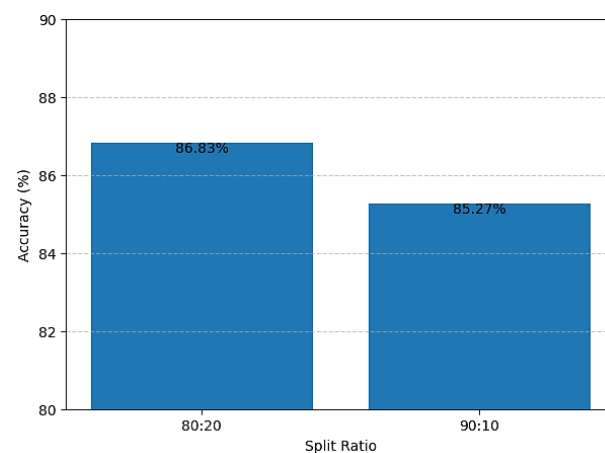


Figure 4. F1-score Performance of Random Forest

3.4 Support Vector Machine (SVM)

In this study, the Support Vector Machine (SVM) algorithm was applied to classify customer responses toward marketing campaigns. The Customer Personality dataset often has overlapping patterns between classes, for example, customers with the same income level but different spending habits. SVM is effective in such situations because it focuses on finding the ideal boundaries that most effectively separate these groups. Table 7 shows the evaluation results for the SVM algorithm.

Table 7. SVM Performance Evaluation

Split Data	Precision	Recall	F1-Score
80:20	82.96%	84.82%	83.59%
90:10	83.83%	83.92%	78.66%

Table 8 displays the performance assessment of the SVM algorithm based on Precision, Recall, and F1 Score for two data proportions, 80:20 and 90:10. In the 80:20 split, the algorithm achieved a Precision of 82.96%, a Recall of 84.82%, and an F1 Score of 83.59%, indicating well-balanced performance. Meanwhile, in the 90:10 split, the algorithm achieved a Precision of 83.83%, a Recall of 83.92%, and an F1 Score of 78.66%, slightly lower than the 80:20 split. These results indicate that the 80:20 split provides better performance due to the larger training data. Figure 5 shows that the F1-score of the SVM algorithm with the 80:20 split reached 83.59%, higher than the 90:10 split, which only achieved an F1-score of 78.66%.

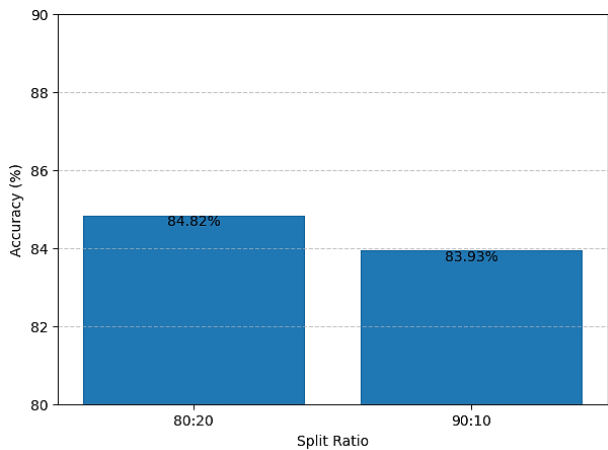


Figure 5. F1-score Performance of Support Vector Machine

3.5 XGBoost

In this study, the XGBoost model was applied to predict customer responses toward marketing campaigns. XGBoost is particularly effective for this dataset because consumer behavior often involves complex interactions, such as how income affects spending on luxury goods or how family size affects purchase frequency. Table 8 shows the evaluation results of the XGBoost algorithm.

Table 8. XGBoost Performance Evaluation

Split Data	Precision	Recall	F1-Score
80:20	84.70%	86.60%	84.85%
90:10	83.63%	85.26%	82.72%

Table 8 displays the performance assessment of the XGBoost algorithm based on Precision, Recall, and F1 Score for two data proportions, 80:20 and 90:10. With the 80:20 split, the algorithm achieved a Precision of 84.70%, a Recall of 86.60%, and an F1 Score of 84.85%, indicating well-balanced performance. Meanwhile, with the 90:10 split, the algorithm achieved a Precision of 83.63%, a Recall of 85.26%, and an F1 Score of 82.72%, which was slightly below that of the 80:20 split. These results indicate that the 80:20 split provides better performance due to the larger training data. Figure 6 shows that the F1-score of the XGBoost algorithm with the 80:20 split reached 84.85%, higher than that of the 90:10 split, which achieved only 82.72%.

3.6 Gradient Boosting Classifier

In this study, a Gradient Boosting Classifier is applied to predict customer responses toward marketing campaigns. Gradient Boosting develops the model in stages, with each stage aiming to correct the errors of the previous one, making it suitable for datasets with subtle variations in customer behavior. Gradient Boosting models are effective at capturing complex non-linear relationships by sequentially correcting prediction errors,

thereby improving generalization performance. Table 9 shows the evaluation results of the GBC algorithm implementation.

Table 9. GBC Performance Evaluation

Split Data	Precision	Recall	F1-Score
80:20	83.07%	85.71%	82.96%
90:10	85.96%	86.16%	83.10%

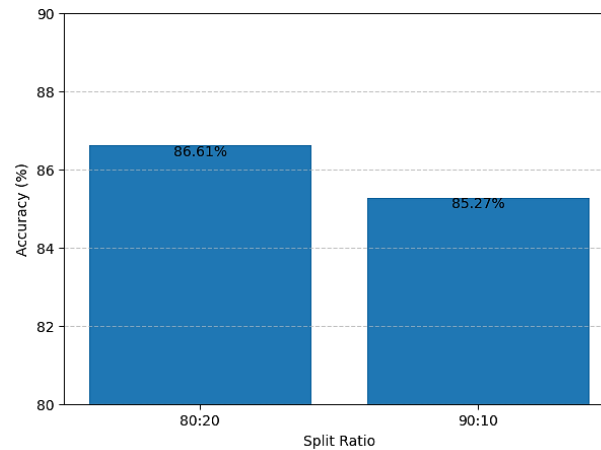


Figure 6. F1-score Performance of XGBoost

Table 9 displays the performance assessment of the GBC algorithm based on Precision, Recall, and F1 Score for two data proportions, 80:20 and 90:10. With the 80:20 split, the algorithm achieved a Precision of 83.07%, a Recall of 85.71%, and an F1 Score of 82.96%, indicating well-balanced performance. Meanwhile, with the 90:10 split, the algorithm achieved a Precision of 85.96%, a Recall of 86.16%, and an F1 Score of 83.10%, slightly exceeding the results of the 80:20 split. These results indicate that GBC maintained relatively stable performance across both split configurations, with a slightly higher F1 Score on the 90:10 split. Figure 7 shows the F1-score of the GBC algorithm with the 90:10 split reached 83.10%, higher than the 80:20 split, which only achieved an F1-score of 82.96%.

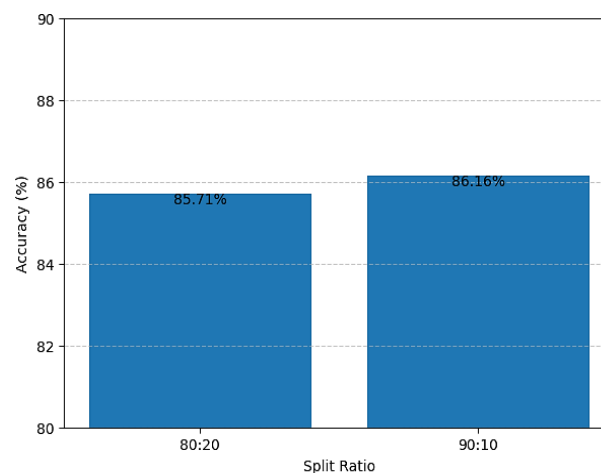


Figure 7. F1-score Performance of Gradient Boosting Classifier

3.7 Confusion Matrix

The evaluation of five classification algorithms LR, RF, SVM, XGBoost, and GBC was conducted using a dataset consisting of 2,240 entries with two data split scenarios (80:20 and 90:10). The performance of each algorithm was summarized using a confusion matrix, which provides information on the number of correctly classified instances (True Negatives and True Positives) and the number of incorrectly classified instances (False Positives and False Negatives). Comparative presentation of these results provides a better understanding of the strengths and weaknesses of each algorithm. Precision, recall, and F1-score are widely recommended metrics for evaluating classification performance in imbalanced datasets, as they provide a more reliable assessment than accuracy alone [24]. Figures 8 and 9 present the confusion matrix visualization of the

best-performing model, while the complete confusion matrix results for all algorithms are summarized in Table 10.

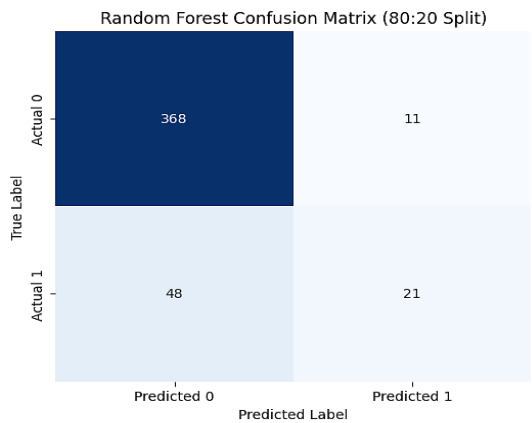


Figure 8. Confusion Matrix 80:20

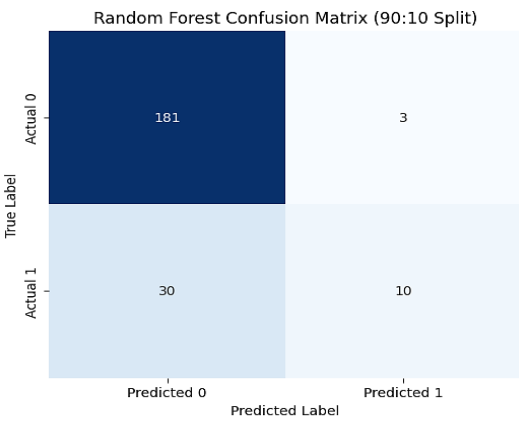


Figure 9. Confusion Matrix 90:10

Table 10. Confusion Matrix

Algoritma	Split Data	TN	TP	FP	FN
LR	80:20:00	369	16	10	53
	90:10:00	179	11	5	29
RF	80:20:00	368	21	11	48
	90:10:00	181	10	3	30
SVM	80:20:00	356	24	23	45
	90:10:00	183	5	1	35
XGBoost	80:20:00	365	23	14	46
	90:10:00	179	12	5	28
GBC	80:20:00	368	16	11	53
	90:10:00	182	11	2	29

The confusion matrix results in Table 10 show the classification performance of Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), XGBoost, and Gradient Boosting Classifier (GBC) in 80:20 and 90:10 data split scenarios. In general, algorithms using ensemble approaches such as RF, XGBoost, and GBC demonstrated more consistent performance, as evidenced by higher True Negative (TN) and True Positive (TP) values and a relatively low False Positive (FP) rate.

In the 80:20 split, RF and XGBoost achieved an optimal balance between TP and FN, demonstrating a reliable ability to recognize customer responses while reducing misclassification errors. SVM produced higher TP values but also recorded more FPs, indicating a tendency to misclassify unresponsive customers as responsive. Logistic Regression and GBC showed high TN values but relatively high FN rates, indicating that some positive responses were misclassified. At a 90:10 split, most algorithms showed an increase in False Negatives, while SVM showed the most significant decrease in detecting positive cases. However, RF, XGBoost, and GBC remained fairly stable, demonstrating their robustness to variations in data proportions. Overall, confusion matrix analysis showed that ensemble methods generally demonstrated more stable performance compared with single classifiers in terms of error control, although reducing False Negatives remained a challenge across all models.

3.8 Comparative Analysis

The comparative analysis was conducted to evaluate the performance differences among the five classification algorithms using consistent evaluation metrics. Precision, recall, and F1-score were analyzed across different data split configurations to determine the most reliable model.

Table 11. Overall Performance Comparison of Classification Algorithms

Algorithm	Split	Precision	Recall	F1-score
Logistic Regression	80:20	83.45%	85.93%	83.13%
Random Forest	80:20	84.94%	86.83%	84.72%
SVM	80:20	82.96%	84.82%	83.59%
XGBoost	80:20	84.70%	86.60%	84.85%
GBC	80:20	83.07%	85.71%	82.96%
Logistic Regression	90:10	82.96%	84.82%	82.03%

Algorithm	Split	Precision	Recall	F1-score
Random Forest	90:10	84.20%	85.26%	82.01%
SVM	90:10	83.83%	83.92%	78.66%
XGBoost	90:10	83.63%	85.26%	82.72%
GBC	90:10	85.96%	86.16%	83.10%

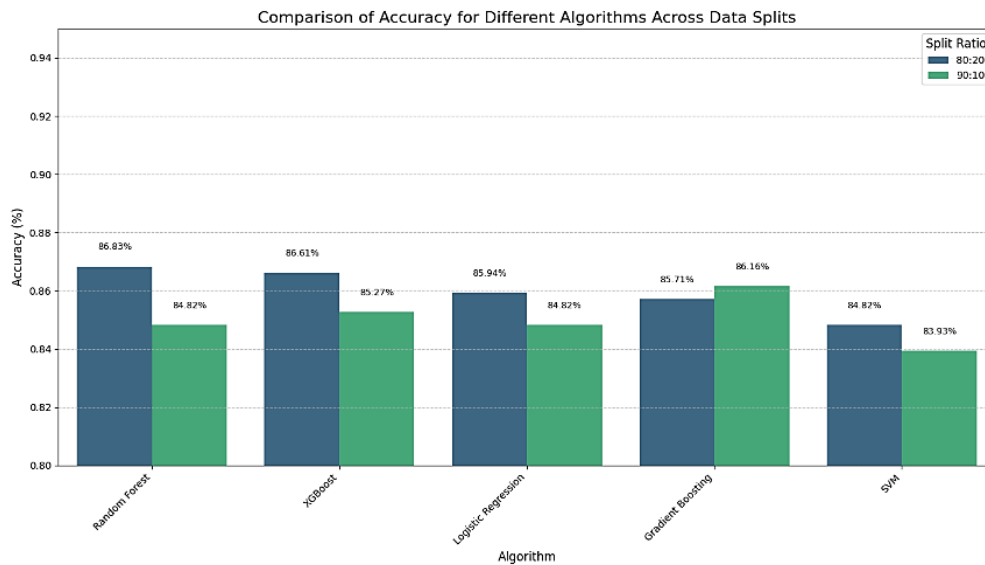


Figure 10. Comparison of Classification Performance Across Different Data Splits

Figure 10 illustrates the comparison of F1-score performance among five machine learning algorithms under two data split configurations (80:20 and 90:10). In the 80:20 split, XGBoost achieved the highest F1-score of 84.85%, followed by Random Forest with 84.72%. Meanwhile, in the 90:10 split, the Gradient Boosting Classifier achieved the highest F1-score of 83.10%. These results indicate that ensemble-based models generally exhibit more consistent performance than individual classifiers across different data-splitting configurations.

These findings are consistent with previous studies reporting that ensemble-based algorithms such as Random Forest and XGBoost generally provide better performance for complex customer behavior prediction tasks due to their ability to capture nonlinear relationships. However, the performance differences observed in this study depend on the data partition configuration, indicating that model stability is an important consideration in selecting a classifier.

3.9. Discussion

The findings indicate that XGBoost demonstrated the best performance among the evaluated algorithms. XGBoost achieved a maximum F1-score of 84.85% with an 80:20 data split and maintained a consistent F1-score of 82.72% with a 90:10 split, demonstrating a good balance between Precision and Recall across multiple data partitions. This indicates that XGBoost achieved competitive performance in this experimental setting, likely due to its ability to capture nonlinear relationships among customer attributes.

The Random Forest (RF) algorithm also demonstrated competitive and consistent performance, with F1-scores of 84.72% (80:20) and 82.01% (90:10). The bagging and random feature-extraction strategies help reduce variance and overfitting, making RF ideal for structured customer data. In contrast, Logistic Regression and Support Vector Machine (SVM) yielded inferior and unstable results. Logistic Regression achieved F1 scores of 83.13% and 82.03%, while SVM decreased from 83.59% (80:20) to 78.66% (90:10), indicating sensitivity to data splitting and decreased robustness to imbalanced data. The Gradient Boosting (GBC) classifier performed well, with F1 scores of 82.96% (80:20) and 83.10% (90:10), demonstrating good generalization ability through sequential error correction. Overall, these findings indicate that ensemble-based algorithms, particularly XGBoost, Random Forest, and Gradient Boosting, generally provide more consistent performance than individual classifiers in predicting marketing campaign responses. Similar findings have been reported in recent studies, in which ensemble-based classifiers consistently outperform individual models on customer response and marketing prediction tasks [25].

4. CONCLUSION

This study examines the effectiveness of five machine learning algorithms, Logistic Regression, Support Vector Machine (SVM), Random Forest, XGBoost, and Gradient Boosting Classifier, in estimating

responses to marketing campaigns using a customer personality analysis dataset sourced from Kaggle. The findings indicate that XGBoost performs best, achieving the highest F1-score of 84.85% with an 80:20 data split, followed by Random Forest at 84.72%. Algorithms that adopt ensemble approaches, such as XGBoost and Gradient Boosting, have demonstrated better, more consistent classification performance under imbalanced data conditions in this experimental setting compared with single models such as Logistic Regression and SVM.

While these algorithms yield positive results, challenges remain in reducing the number of False Negatives, which can lead to missed customer opportunities. This study highlights the importance of employing machine learning methods to analyze consumer behavior and enhance the effectiveness of marketing campaigns. By leveraging more advanced, data-driven techniques, companies can develop more personalized and efficient marketing strategies, thereby gaining an advantage in an increasingly complex market.

This study has several limitations. The evaluation was conducted using a single dataset obtained from Kaggle, which may limit the generalizability of the findings to other marketing environments. Future research should consider external validation using real-world company datasets and explore alternative approaches for handling imbalanced data, such as advanced resampling techniques, cost-sensitive learning, and hybrid ensemble methods.

REFERENCES

- [1] M. El-Hajj and M. Pavlova, "Predictive Modeling of Customer Response to Marketing Campaigns," *Electronics (Switzerland)*, vol. 13, no. 19, Oct. 2024, doi: 10.3390/electronics13193953.
- [2] T. Davenport, A. Guha, D. Grewal, and T. Bressgott, "Cite this article 148k Accesses 141 Citations 33 Altmetric Metrics," 2020, doi : 10.1007/s11747-019-00696-0.
- [3] B. Vlačić, L. Corbo, S. Costa e Silva, and M. Dabić, "The evolving role of artificial intelligence in marketing: A review and research agenda," May 01, 2021, Elsevier Inc. doi: 10.1016/j.jbusres.2021.01.055.
- [4] McKinsey & Company, "What is personalization?," 2023. [Online]. Available: <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-personalization>.
- [5] H. Langen and M. Huber, "How causal machine learning can leverage marketing strategies: Assessing and improving the performance of a coupon campaign," *arXiv preprint arXiv:2204.10820*, 2022.
- [6] T. A. Smith, "The role of customer personality in satisfaction, attitude-to-brand and loyalty in mobile services," *Spanish Journal of Marketing - ESIC*, vol. 24, no. 2, pp. 155–175, Oct. 2020, doi: 10.1108/SJME-06-2019-0036.
- [7] M. Syukron, R. Santoso, and T. Widiyarih, "perbandingan metode smote random forest dan smote xgboost untuk klasifikasi tingkat penyakit hepatitis c pada imbalance class data", [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>, 2020, doi: 10.14710/j.gauss.9.3.227-236.
- [8] I. Domor Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects", 2022, doi: 10.1109/ACCESS.2022.3207287.
- [9] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.
- [10] Kulatilleke, G. K., & Samarakoon, S. (2022). Empirical study of machine learning classifier evaluation metrics behavior in massively imbalanced and noisy data. *arXiv*. <https://doi.org/10.48550/arXiv.2208.11904>
- [11] Opitz, J. (2024). A closer look at classification evaluation metrics and a critical reflection of common evaluation practice. *Transactions of the Association for Computational Linguistics*, 12, 1161–1183. https://doi.org/10.1162/tacl_a_00675
- [12] J. Lin, "Application of machine learning in predicting consumer behavior and precision marketing," *PLoS One*, vol. 20, no. 5 May, May 2025, doi: 10.1371/journal.pone.0321854.
- [13] C. Ali, "klasifikasi topik berita politik menggunakan model logistic regression dan fitur bag of words," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 3, pp. 282–291, Sep. 2025, doi: 10.69916/jkbt.v4i3.375.
- [14] Y. Wang, W. Liao, H. Shen, Z. Jiang, and J. Zhou, "Some notes on the basic concepts of support vector machines," *J Comput Sci*, vol. 82, Oct. 2024, doi: 10.1016/j.jocs.2024.102390.
- [15] S. A. Salleh et al., "Support Vector Machine (SVM) and Object Based Classification in Earth Linear Features Extraction: A Comparison," *Revue Internationale de Géomatique*, vol. 33, no. 1, pp. 183–199, 2024, doi: 10.32604/rig.2024.050723.
- [16] I. T. Akinola, Y. Sun, I. G. Adebayo, and Z. Wang, "Daily peak demand forecasting using Pelican Algorithm optimised Support Vector Machine (POA-SVM)," *Energy Reports*, vol. 12, pp. 4438–4448, Dec. 2024, doi: 10.1016/j.egyr.2024.10.017.
- [17] Y. tao Zhu, C. shi Gu, and M. A. Diaconeasa, "A missing data processing method for dam deformation monitoring data using spatiotemporal clustering and support vector machine model," *Water Science and Engineering*, vol. 17, no. 4, pp. 417–424, Dec. 2024, doi: 10.1016/j.wse.2024.08.003.

-
- [18] R. Chintalapudi, G. N. Kannaiyan, B. Pappula, and S. Makgato, "The novel Vogel's approximation method integrated with a random forest algorithm in the vibration analysis of a two-directional functionally graded taper porous beam: Assessment," *Sci Afr*, vol. 26, Dec. 2024, doi: 10.1016/j.sciaf.2024.e02397.
 - [19] G. Samodra, Ngadisih, and F. S. Nugroho, "Benchmarking data handling strategies for landslide susceptibility modeling using random forest workflows," *Artificial Intelligence in Geosciences*, vol. 5, Dec. 2024, doi: 10.1016/j.aiig.2024.100093.
 - [20] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif Intell Rev*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.
 - [21] E. Gültür and M. F. Cevher, "Evolution of digital marketing campaigns with artificial intelligence and machine learning: Analysing success prediction capabilities," *Business & Management Studies: An International Journal*, vol. 13, no. 2, pp. 478–493, Jun. 2025, doi: 10.15295/bmij.v13i2.2498.
 - [22] Z. Liu, P. Jiang, K. W. De Bock, J. Wang, L. Zhang, and X. Niu, "Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction," *Technological Forecasting and Social Change*, vol. 198, 2024, doi: 10.1016/j.techfore.2023.122945.
 - [23] M. Mustak, H. Hallikainen, T. Laukkanen, L. Plé, L. D. Hollebeck, and M. Aleem, "Using machine learning to develop customer insights from user-generated content," *Journal of Retailing and Consumer Services*, vol. 81, 2024, doi: 10.1016/j.jretconser.2024.104034.
 - [24] A. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Progress in Artificial Intelligence*, vol. 12, pp. 1–23, 2023. doi: 10.1007/s13748-023-00295-5.
 - [25] H. Chen, Q. Tang, Y. Wei, and M. Song, "Churn Prediction Model of Telecom Users Based on XGBoost," *Journal on Artificial Intelligence*, vol. 3, no. 3, pp. 115–121, 2021, doi: 10.32604/jai.2021.026851.
 - [26] U. Laili, Y. #1, A. Purwarianti, and I. +3, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Tuning Hyperparameter pada Gradient Boosting untuk Klasifikasi Soal Cerita Otomatis," 2022, doi : 10.26418/jp.v8i1.50506.
 - [27] Y. Hu, L. Tong, Y. Gao, Y. Zeng, B. Yan, and Z. Li, "GazeViT: A gaze-guided hybrid attention vision transformer for cross-view matching of street-to-aerial images," *Pattern Recognit Lett*, vol. 191, pp. 80–88, May 2025, doi: 10.1016/j.patrec.2025.03.012.
 - [28] P. Boozary, S. Sheykhan, H. GhorbanTanhaei, and C. Magazzino, "Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction," *International Journal of Information Management Data Insights*, vol. 5, no. 1, Jun. 2025, doi: 10.1016/j.jjime.2025.100331.