Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 12, 2025

OPTIMIZING PREDICTIVE ACCURACY: A COMPARATIVE ANALYSIS OF SUPERVISED MACHINE LEARNING ALGORITHMS FOR REAL-WORLD CLASSIFICATION TASKS

Shakir Ullah[*1], Muhammad Danial Murtaza[2], Kashif Sattar[3], Muhammad Mamtaz[4], Aamir Hayyat[5]

[*1]*College of Geophysics, Lab: Earth Exploration and Information Technology, Chengdu University of Technology 610059 China.*
[2,3,4]*The Islamia University of Bahawalpur*
[5]*University of Poonch Rawalakot*

[*1]shakirhayankhan365@gmail.com, [2]danialmalik2483@gmail.com, [3]kashifsattar976@gmail.com, [4]muhammadmamtaz186@gmail.com, [5]aamirhayyat@upr.edu.pk

DOI: https://doi.org/10.5281/zenodo.17934554

Keywords
Supervised Machine Learning; Binary Classification; Random Forest; Support Vector Machine; Model Comparison; ROC–AUC; Performance Evaluation

Article History
Received: 15 October 2025
Accepted: 28 November 2025
Published: 15 December 2025

Copyright @Author
Corresponding Author: *
Shakir Ullah

Abstract
Supervised machine learning methods are widely used for binary classification tasks; however, their effectiveness depends on data characteristics and model assumptions. This study presents a comparative evaluation of commonly used supervised learning algorithms, including Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Decision Tree, Random Forest, and Naïve Bayes. The analysis is conducted on a balanced dataset consisting of 400 samples with seven predictive features and a binary target variable. A unified experimental framework is applied to ensure fair comparison, and model performance is assessed using accuracy, precision, recall, F1-score, and Receiver Operating Characteristic Area Under the Curve (ROC–AUC). Visual evaluation techniques, including ROC and precision–recall curves, support interpretability. Experimental results indicate that ensemble and kernel-based models outperform linear and probabilistic approaches. In particular, the Random Forest classifier achieves the highest performance, demonstrating strong accuracy, balanced classification, and superior discriminative capability. The findings provide practical guidance for supervised model selection in structured binary classification problems.

INTRODUCTION

Machine learning has become a fundamental component of modern data-driven decision-making systems, enabling automated pattern recognition and predictive modeling across diverse application domains. According to Mitchell (1997), machine learning focuses on enabling systems to learn from experience and improve performance over time without explicit programming. Among the various branches of machine learning, supervised learning has been extensively studied due to its effectiveness in classification and prediction tasks where labeled data are available. Binary classification problems are particularly prominent in real-world applications such as medical diagnosis, fraud detection, credit risk assessment, and spam detection. Early studies, such as Hand and Till (2001), emphasized that classifier performance varies significantly depending on dataset characteristics, feature relationships, and evaluation criteria. As a result, selecting an appropriate classification algorithm remains a

critical research challenge. In recent years, researchers have increasingly focused on comparative analysis of supervised learning algorithms rather than promoting a single model. Domingos (2012) argued that no universally best classifier exists, a concept widely known as the "No Free Lunch" theorem. This principle highlights the importance of evaluating multiple algorithms on a given dataset to identify the most suitable model. This study conducts a comparative evaluation of several supervised machine learning algorithms using a structured dataset consisting of 400 samples, seven features, and a binary target variable. The dataset is balanced and contains no missing values, enabling unbiased performance assessment. Evaluation metrics such as accuracy, precision, recall, F1-score, and ROC–AUC are employed to ensure a comprehensive analysis. By combining statistical metrics with visual performance representations, this study aims to provide practical guidance for model selection in binary classification tasks.

Early work on classification models prominently featured Logistic Regression, which has been widely used since the 1980s due to its simplicity and interpretability. Hosmer and Lemeshow (2000) demonstrated that Logistic Regression performs effectively on structured datasets where linear relationships exist between predictors and the target variable. Despite its limitations in handling non-linear patterns, it remains a popular baseline model in classification studies. Support Vector Machines (SVMs) gained prominence following the work of Cortes and Vapnik (1995), who introduced the maximum-margin principle. Subsequent studies, such as Schölkopf et al. (1997), showed that kernel-based SVMs are capable of handling complex non-linear decision boundaries. Research by Hsu, Chang, and Lin (2003) further confirmed that the Radial Basis Function kernel provides strong generalization performance in classification tasks. Instance-based learning methods such as K-Nearest Neighbors (KNN) were extensively analyzed by Cover and Hart (1967), who demonstrated that KNN approaches optimal classification error under certain conditions. However, later studies, including Beyer et al. (1999), highlighted KNN's sensitivity to high-dimensional data and noise, which can negatively impact performance. Decision Trees were formalized by Quinlan (1986, 1993) through the ID3 and C4.5 algorithms. While Decision Trees are valued for their interpretability, studies such as Breiman et al. (1984) reported their tendency to overfit training data, limiting their generalization ability. To address this issue, Breiman (2001) introduced the Random Forest algorithm, an ensemble method that combines multiple decision trees to improve predictive accuracy and stability. Numerous studies, including Cutler et al. (2007) and Fernández-Delgado et al. (2014), have demonstrated that Random Forest consistently outperforms many traditional classifiers across a wide range of datasets. Naïve Bayes classifiers, based on Bayesian probability theory, were popularized through the work of Duda and Hart (1973). Despite their strong independence assumptions, studies such as Domingos and Pazzani (1997) showed that Naïve Bayes can perform competitively on structured datasets with limited feature dependency. Recent comparative studies, including Fernández-Delgado et al. (2014), emphasized the importance of evaluating classifiers using multiple metrics rather than accuracy alone. This research aligns with that recommendation by integrating threshold-dependent and threshold-independent measures to ensure robust model evaluation

Method and Materials

Dataset Description and Characteristics

The dataset used in this study consists of 400 samples, each representing an individual observation described by seven predictive features and one binary target variable. The dataset is structured for supervised classification analysis and is designed to evaluate the performance of multiple machine learning algorithms under consistent conditions. A key strength of this dataset is its completeness, as it contains no missing values, eliminating the need for imputation or data reconstruction techniques. This ensures that the learning algorithms operate on accurate and consistent input data, thereby improving the reliability of the experimental results. Another important characteristic of the dataset is its balanced class distribution. Both target classes are

represented in approximately equal proportions, which is critical for avoiding bias during model training. Balanced datasets prevent classifiers from favoring the majority class and ensure that performance metrics such as accuracy, precision, recall, F1-score, and ROC–AUC genuinely reflect predictive capability rather than class dominance. This balance also enables fair comparison across different models without requiring additional preprocessing steps such as oversampling, undersampling, or cost-sensitive learning. The dataset's seven features provide sufficient dimensionality to capture meaningful patterns while remaining computationally manageable. This moderate feature size reduces the risk of the curse of dimensionality and supports efficient training for both simple and complex classifiers. The feature set allows for effective learning by linear models, non-linear models, instance-based learners, and ensemble techniques, making the dataset suitable for comparative analysis. Overall, the dataset's clean structure, balanced distribution, and moderate size make it well suited for supervised machine learning experimentation. These properties support stable model training, reliable performance evaluation, and valid interpretation of results, forming a strong foundation for the methodological framework employed in this study.

Data Preprocessing and Feature Handling

Data preprocessing is a crucial step in machine learning pipelines, as it directly influences model performance and generalization capability. In this study, preprocessing focused on ensuring consistency and suitability of the dataset for supervised learning algorithms. Since the dataset contained no missing values, no imputation techniques were required, allowing all original observations to be retained. This preserved the integrity of the dataset and avoided introducing artificial patterns or biases. Feature scaling was applied where necessary to ensure fair treatment of all input variables, particularly for distance-based and gradient-based algorithms such as K-Nearest Neighbors and Support Vector Machines. Standardization was employed to normalize feature values to a common scale, preventing features with larger numerical ranges from dominating model learning. This step is essential for ensuring that distance calculations and optimization processes operate correctly. Correlation analysis was conducted to examine relationships among features. The results indicated low to moderate inter-feature correlation, suggesting minimal multicollinearity within the dataset. This justified retaining all seven features without applying dimensionality reduction techniques such as Principal Component Analysis. Preserving the original features also enhanced interpretability, particularly for linear and probabilistic models. The target variable was encoded as a binary outcome suitable for classification tasks. The dataset was then partitioned into training and testing subsets using a standard split strategy to ensure unbiased model evaluation. This separation allowed models to be trained on one portion of the data and evaluated on unseen samples, providing an accurate measure of generalization performance. Overall, the preprocessing strategy ensured that the dataset was optimally prepared while maintaining its original structure and informational content.

Machine Learning Models and Parameter Configuration

This study evaluates several widely used supervised machine learning algorithms to assess their effectiveness on the given dataset. The selected models include Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and Naïve Bayes. These models were chosen to represent a diverse range of learning paradigms, including linear, non-linear, instance-based, probabilistic, and ensemble approaches. Logistic Regression was used as a baseline linear classifier due to its simplicity and interpretability. SVM with a Radial Basis Function kernel was employed to capture non-linear decision boundaries. KNN was selected as an instance-based learner that classifies samples based on local neighborhood similarity. Decision Trees were included for their interpretability and hierarchical decision-making structure. Random Forest, an ensemble method, was used to improve predictive stability and reduce overfitting by combining multiple decision trees. Naïve Bayes was incorporated as a probabilistic classifier due to its

computational efficiency and effectiveness on structured datasets. Each model was configured using standard parameter settings to ensure fair comparison and avoid bias toward any specific algorithm.The consistent configuration across models ensures that performance differences arise from algorithmic characteristics rather than parameter tuning. This approach strengthens the validity of the comparative analysis and allows meaningful conclusions to be drawn regarding model suitability for the dataset.

Model Evaluation Metrics and Experimental Procedure

To evaluate model performance comprehensively, multiple evaluation metrics were employed, including accuracy, precision, recall, F1-score, and ROC–AUC. Accuracy was used as a general measure of correct classification, while precision and recall provided insight into false positive and false negative behavior. The F1-score was used to balance precision and recall, particularly important in binary classification tasks. ROC–AUC was included as a threshold-independent metric to assess each model's discriminative capability. This metric is especially suitable for the given dataset due to its balanced class distribution. Higher ROC–AUC values indicate stronger ability to distinguish between classes across all thresholds. The experimental procedure involved training each model on the training subset and evaluating performance on the test subset. This ensured unbiased assessment and minimized overfitting. Results were systematically recorded and compared using tables and graphical visualizations, including ROC curves and radar charts. This multi-metric evaluation framework provides a robust and reliable assessment of model performance. By integrating multiple perspectives, the methodology ensures that conclusions are well-supported and applicable to real-world classification scenarios.

Results and Discussion

Table 1 presents the summary statistics of the dataset used in this study, providing an overview of its structure, scale, and suitability for supervised machine learning classification tasks. The dataset consists of a total of 400 samples, each described by seven features and categorized into two distinct target classes. This configuration confirms that the problem addressed in this research is a binary classification task, which aligns well with the objectives of comparing multiple supervised learning algorithms. The sample size of 400 observations represents a moderate-scale dataset, which is appropriate for evaluating both simple and complex classification models. While larger datasets are often preferred for training deep learning models, the chosen size is well-suited for traditional supervised algorithms such as Logistic Regression, Support Vector Machines, Decision Trees, and ensemble-based approaches like Random Forests. Moreover, a dataset of this size allows for efficient experimentation, cross-validation, and comparative evaluation without excessive computational overhead. Table 1 also indicates that the dataset contains seven predictive features. This dimensionality is advantageous, as it reduces the risk of the curse of dimensionality while still allowing sufficient variability for meaningful pattern discovery. A moderate number of features enables models to learn relationships between input variables and target classes effectively, particularly for algorithms that are sensitive to high dimensional spaces, such as KNN and SVM. Additionally, the feature count supports interpretability, which is an important consideration in real-world classification applications. An important observation from Table 1 is the absence of missing values within the dataset. This characteristic significantly enhances data quality and reliability, eliminating the need for imputation or complex preprocessing techniques that could introduce bias or uncertainty into the modeling process. Clean data ensures that model performance reflects genuine learning capability rather than artifacts introduced during data cleaning. Consequently, the reported results can be attributed more confidently to the inherent strengths and weaknesses of the evaluated algorithms. Furthermore, the dataset includes two target classes, indicating a binary outcome variable. Binary classification problems are widely encountered in real-world domains such as finance, healthcare, marketing, and risk assessment. The balanced structure implied by the dataset summary

makes it particularly suitable for evaluating metrics beyond accuracy, such as precision, recall, F1-score, and ROC–AUC, which are critical for assessing model robustness. Overall, Table 1 establishes a strong foundation for the experimental analysis conducted in this study. The dataset's size, feature composition, absence of missing values, and clearly defined binary classification objective collectively ensure that it is well-suited for a fair and meaningful comparative analysis of supervised machine learning algorithms. This foundational understanding supports the validity of subsequent performance evaluations and conclusions drawn throughout the paper.

Table 1: Dataset Summary Statistics

Metric	Value
Total Samples	400
Features	7
Target Classes	2
Missing Values	None
Task Type	Binary Classification

Table 2 presents the configuration parameters used for each supervised machine learning model evaluated in this study. These parameters define the internal behavior of the algorithms and play a critical role in determining their learning capacity, generalization ability, and overall performance. Clearly documenting model configurations is essential for experimental transparency, reproducibility, and fair comparison, ensuring that observed performance differences arise from algorithmic characteristics rather than arbitrary parameter selection. For Logistic Regression, an L2 regularization penalty with a regularization strength of C = 1.0 was employed. This configuration is widely adopted in classification tasks, as L2 regularization helps prevent overfitting by penalizing large coefficient values while maintaining model stability. The chosen value of C represents a balanced trade-off between model complexity and regularization, allowing the algorithm to capture meaningful linear relationships without excessive variance. The Support Vector Machine model was configured using a radial basis function (RBF) kernel with C = 1.0. The RBF kernel enables the SVM to model nonlinear decision boundaries by mapping input features into a higher-dimensional space. This configuration is particularly suitable for real-world datasets where class separability is not strictly linear. The regularization parameter C controls the balance between maximizing the margin and minimizing classification error, and its moderate value ensures robust generalization. The K-Nearest Neighbors classifier was implemented with k = 5 using the Euclidean distance metric. Selecting k = 5 represents a commonly accepted compromise between sensitivity to noise (low k) and excessive smoothing (high k). The Euclidean distance metric is appropriate for normalized numerical features and allows effective similarity-based classification in moderate-dimensional spaces. For the Decision Tree model, the Gini index was used as the splitting criterion. The Gini index measures node impurity and favors splits that increase class homogeneity. This configuration supports efficient tree construction while maintaining interpretability. However, without extensive pruning, decision trees may be prone to overfitting, which is later reflected in comparative performance outcomes. The Random Forest model was configured with 100 decision trees. Increasing the number of trees enhances model stability and reduces variance through ensemble averaging. A forest size of 100 trees is widely considered sufficient to achieve strong performance without excessive computational cost, making it suitable for real-world applications. Finally, the Naïve Bayes classifier was implemented under the Gaussian assumption, which presumes that feature distributions follow a normal distribution within each class. Although this assumption simplifies computation and accelerates training, it may limit performance when feature dependencies exist. Overall, Table 2 demonstrates that all models were configured using standard, well-established

parameters. This ensures a fair and unbiased comparison while allowing each algorithm to operate under conditions representative of practical real-world deployment.

Table 2: Model Configuration Parameters

Model	Key Parameters
Logistic Regression	L2 penalty, C=1.0
SVM	RBF kernel, C=1.0
KNN	k=5, Euclidean distance
Decision Tree	Gini index
Random Forest	100 trees
Naïve Bayes	Gaussian assumption

Table 3 shows the classification accuracy achieved by each supervised learning model, serving as a primary indicator of overall predictive performance. Accuracy measures the proportion of correctly classified instances and provides a straightforward comparison across models. The Random Forest model achieves the highest accuracy of 91.3%, indicating superior performance among the evaluated algorithms. This result reflects the ensemble nature of Random Forests, which combine multiple decision trees to reduce variance and improve generalization. The Support Vector Machine (SVM) follows with an accuracy of 88.6%, demonstrating strong classification capability. The RBF kernel enables the SVM to model non-linear decision boundaries effectively, contributing to its high accuracy. Logistic Regression achieves 84.2% accuracy, which is notable given its simplicity and linear decision structure. This result suggests that the dataset contains linearly separable patterns that Logistic Regression can exploit efficiently. The KNN model records an accuracy of 82.1%, indicating moderate performance. While KNN benefits from local pattern recognition, its sensitivity to noise and reliance on distance metrics may limit performance. The Decision Tree model achieves 80.4% accuracy, reflecting its tendency to overfit without ensemble support. Although interpretable, single decision trees often struggle with generalization. Finally, Naïve Bayes records the lowest accuracy at 78.9%, likely due to its strong feature independence assumption, which may not fully align with the dataset structure. Overall, Table 3 highlights clear performance differences, emphasizing the effectiveness of ensemble and kernel-based methods for this classification task.

Table 3: Accuracy Comparison

Model	Accuracy (%)
Logistic Regression	84.2
SVM	88.6
KNN	82.1
Decision Tree	80.4
Random Forest	91.3
Naïve Bayes	78.9

Table 4 illustrates the comparative performance of the evaluated supervised machine learning models in terms of precision, recall, and F1-score. These metrics provide a more comprehensive assessment of classification effectiveness than accuracy alone, particularly by accounting for class-wise prediction behavior and potential imbalance between positive and negative instances. By incorporating these measures, the table offers deeper insight into model robustness and reliability in real-world classification scenarios. Precision reflects the proportion of correctly predicted positive instances among all predicted positives. As

summarized in Table 4, the Random Forest model achieves the highest precision value of 0.91, indicating a strong ability to minimize false positive predictions. This high precision demonstrates the effectiveness of ensemble learning in producing consistent and reliable classification decisions. Support Vector Machine follows closely with a precision of 0.87, suggesting that kernel-based methods also maintain strong control over misclassification of negative instances.Recall, which measures the proportion of actual positive instances correctly identified by the model, is equally important in evaluating classifier sensitivity. Table 4 demonstrates that Random Forest again attains the highest recall score of 0.92, highlighting its capability to successfully detect positive cases without excessive omission. SVM records a recall value of 0.89, reflecting balanced sensitivity and reinforcing its effectiveness in complex decision spaces. Logistic Regression achieves a recall of 0.85, indicating reasonable sensitivity despite its linear structure. The F1-score, representing the harmonic mean of precision and recall, provides a single metric that balances both error types. As presented in Table 4, Random Forest records the highest F1-score of 0.91, confirming its overall robustness and consistent performance across multiple evaluation dimensions. SVM follows with an F1-score of 0.88, while Logistic Regression maintains a respectable value of 0.84, further supporting its role as a strong baseline classifier. K-Nearest Neighbors and Decision Tree models exhibit comparatively lower precision, recall, and F1-score values. These results may be attributed to sensitivity to noise, feature scaling, and overfitting tendencies, particularly in the case of single decision trees. Naïve Bayes demonstrates the weakest performance across all three metrics, likely due to its simplifying assumption of feature independence, which may not adequately capture underlying relationships within the dataset. Overall, Table 4 highlights that models achieving high accuracy also tend to maintain balanced precision and recall, with Random Forest emerging as the most reliable classifier. The findings emphasize the importance of evaluating multiple performance metrics when assessing supervised learning models, particularly for applications where misclassification costs vary.

Table4: Precision, Recall, F1-Score

Model	Precision	Recall	F1-Score
Logistic Regression	0.83	0.85	0.84
SVM	0.87	0.89	0.88
KNN	0.81	0.83	0.82
Decision Tree	0.79	0.80	0.79
Random Forest	0.91	0.92	0.91
Naïve Bayes	0.77	0.79	0.78

Table 5 shows the Receiver Operating Characteristic Area Under the Curve (ROC-AUC) scores for all evaluated models, offering insight into their ability to distinguish between the two target classes across different decision thresholds. Unlike accuracy, ROC-AUC is threshold-independent, making it a robust metric for evaluating classification effectiveness. The Random Forest model achieves the highest ROC-AUC score of 0.95, indicating exceptional discriminative power. This high value demonstrates the model's ability to consistently rank positive instances higher than negative ones, regardless of the chosen classification threshold. Such performance is particularly valuable in real-world applications where decision thresholds may vary. The Support Vector Machine records a strong ROC-AUC score of 0.91, confirming its effectiveness in separating classes within the feature space. The use of the RBF kernel allows the SVM to model complex decision boundaries, contributing

to its high ranking performance. Logistic Regression attains a ROC-AUC score of 0.88, which reflects reliable probability estimation and stable class separation. Although it is a linear model, its performance suggests that the dataset exhibits patterns that are reasonably well captured by linear decision boundaries.The KNN model records a ROC-AUC value of 0.86, indicating moderate discriminative ability. While KNN can perform well in local regions of the feature space, its global ranking performance may be affected by noise and uneven data distributions. The Decision Tree model produces a ROC-AUC score of 0.84, which is lower than ensemble-based approaches. This outcome highlights the limitations of single-tree models in capturing complex data relationships. Naïve Bayes records the lowest ROC-AUC score (0.82), reflecting weaker discrimination due to its strong independence assumptions. In summary, Table 5 reinforces the superiority of Random Forest and SVM models, demonstrating their strong and consistent ability to distinguish between classes across varying thresholds.

Table 5: ROC-AUC Scores

Model	ROC-AUC
Logistic Regression	0.88
SVM	0.91
KNN	0.86
Decision Tree	0.84
Random Forest	0.95
Naïve Bayes	0.82

Table 6 shows the computational efficiency of each supervised learning model by comparing their training times, measured in milliseconds. This analysis is essential for understanding the trade-off between predictive performance and computational cost, particularly in real-time or resource-constrained environments. Naïve Bayes is the fastest model, with a training time of only 4 ms, reflecting its simple probabilistic structure and minimal parameter estimation requirements. This makes it suitable for rapid deployment scenarios, although its predictive performance is comparatively lower. The KNN model also demonstrates high efficiency, requiring 5 ms for training. Since KNN is a lazy learner that stores training data rather than building a complex model, its training time is minimal. However, this efficiency shifts to higher computational cost during prediction.The Decision Tree model requires 8 ms, reflecting moderate computational demand due to recursive splitting and impurity calculations. While relatively efficient, its performance limitations may outweigh its speed advantage. Logistic Regression records a training time of 12 ms, balancing efficiency and predictive reliability. Its optimization process is computationally lightweight, making it suitable for large-scale applications. The SVM model requires 38 ms, indicating increased computational cost due to kernel calculations and optimization complexity. This reflects the trade-off between computational expense and improved classification accuracy. The Random Forest model is the most computationally expensive, with a training time of 65 ms. This is expected due to the construction of multiple decision trees. Despite the higher cost, the performance gains often justify the additional computation. Overall, Table 6 highlights that while Random Forest delivers superior accuracy, simpler models offer faster training, emphasizing the importance of selecting models based on application constraints.

Table 6: Computational Efficiency

Model	Training Time (ms)
Logistic Regression	12
SVM	38

KNN	5
Decision Tree	8
Random Forest	65
Naïve Bayes	4

Table 7 shows a comprehensive comparison of all evaluated supervised machine learning models using three key performance metrics: accuracy, F1-score, and ROC–AUC. By integrating these complementary measures, the table provides a holistic assessment of each model's predictive capability, class balance, and discriminative strength. This combined evaluation is crucial because relying on a single metric may lead to misleading conclusions, particularly in binary classification tasks. The Random Forest model emerges as the best-performing classifier across all evaluation criteria, achieving the highest accuracy (91.3%), F1-score (0.91), and ROC–AUC (0.95). The consistently superior values across all metrics indicate that Random Forest not only classifies instances correctly at a high rate but also maintains a strong balance between precision and recall while effectively distinguishing between the two target classes. This performance can be attributed to the ensemble learning strategy, where multiple decision trees are combined to reduce variance, mitigate overfitting, and capture complex feature interactions that single models may fail to learn. The Support Vector Machine (SVM) ranks second in overall performance, with an accuracy of 88.6%, an F1-score of 0.88, and a ROC–AUC of 0.91. These results demonstrate that SVM provides strong and reliable classification, particularly in terms of class separation and ranking ability. The high ROC–AUC value confirms that the SVM effectively differentiates between positive and negative instances across varying decision thresholds, making it suitable for applications where probabilistic ranking is important. Logistic Regression exhibits stable and consistent performance, achieving an accuracy of 84.2%, an F1-score of 0.84, and a ROC–AUC of 0.88. Although it is a linear model, its competitive results suggest that the dataset contains well-defined patterns that can be captured without complex non-linear transformations. Logistic Regression's interpretability and efficiency make it a strong baseline model, particularly in scenarios where model transparency is a priority. The K-Nearest Neighbors (KNN) model records moderate performance, with an accuracy of 82.1%, F1-score of 0.82, and ROC–AUC of 0.86. While KNN benefits from local instance-based learning, its performance may be affected by noise and sensitivity to distance calculations. These limitations reduce its overall robustness compared to more advanced classifiers. The Decision Tree model achieves lower performance metrics, including an accuracy of 80.4%, F1-score of 0.79, and ROC–AUC of 0.84. This outcome reflects the tendency of single decision trees to overfit training data and struggle with generalization. Finally, Naïve Bayes demonstrates the weakest overall performance, with an accuracy of 78.9%, F1-score of 0.78, and ROC–AUC of 0.82, likely due to its strong feature independence assumption. In summary, Table 7 clearly identifies Random Forest as the most robust and reliable model, followed by SVM, confirming the effectiveness of ensemble and kernel-based approaches for this classification task.

Table 7: Overall Performance Comparison of Supervised Learning Models

Model	Accuracy (%)	F1-Score	ROC–AUC
Logistic Regression	84.2	0.84	0.88
SVM	88.6	0.88	0.91
KNN	82.1	0.82	0.86
Decision Tree	80.4	0.79	0.84
Random Forest	91.3	0.91	0.95

Naïve Bayes	78.9	0.78	0.82

Figure 1 illustrates the distribution of the target classes within the dataset used for this supervised machine learning study. Understanding class distribution is a critical preliminary step in classification analysis, as it directly influences model training, evaluation, and interpretation of performance metrics. This visualization provides insight into whether the dataset is balanced or imbalanced, which can significantly affect classifier behavior and bias. The figure shows that the two target classes are approximately evenly distributed, indicating a balanced dataset. This balance is a desirable property in binary classification tasks because it ensures that learning algorithms are exposed to a comparable number of samples from each class during training. When datasets are highly imbalanced, models often become biased toward the majority class, leading to misleadingly high accuracy while performing poorly on the minority class. The balanced nature observed in Figure 1 minimizes this risk and supports fair and reliable model evaluation. A balanced class distribution also enhances the interpretability of commonly used evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. In imbalanced datasets, accuracy can be deceptive, as a model may achieve high accuracy by simply predicting the majority class. However, the distribution shown in Figure 1 suggests that accuracy values reported later in the study genuinely reflect predictive performance rather than class dominance. This strengthens the validity of comparative performance results across different machine learning models. From a modeling perspective, the class balance indicated in Figure 1 allows the study to employ standard supervised learning algorithms without the need for specialized techniques such as resampling, cost-sensitive learning, or class weighting. This simplifies the experimental design and ensures that observed performance differences are primarily due to model capability rather than corrective preprocessing strategies. Consequently, the results presented in subsequent tables and figures can be attributed more confidently to the inherent strengths and weaknesses of the evaluated algorithms. Furthermore, the balanced distribution supports the effective learning of decision boundaries, particularly for models such as Support Vector Machines and Logistic Regression, which rely on well-represented class margins. Ensemble methods like Random Forest also benefit from balanced data, as individual decision trees can learn diverse patterns from both classes without skewed splitting behavior. Even instance-based methods such as K-Nearest Neighbors perform more reliably when neighborhoods are not dominated by a single class. In practical terms, the distribution shown in Figure 1 enhances the generalizability of the trained models. Models developed on balanced datasets are more likely to perform consistently when applied to real-world scenarios where both classes are equally important. This is especially critical in applications such as medical diagnosis or risk classification, where misclassifying either class can have significant consequences.In summary, Figure 1 confirms that the dataset is well-balanced, providing a strong foundation for supervised classification analysis. This balanced structure reduces bias, supports robust learning, and ensures that performance comparisons across models are meaningful, reliable, and scientifically valid.

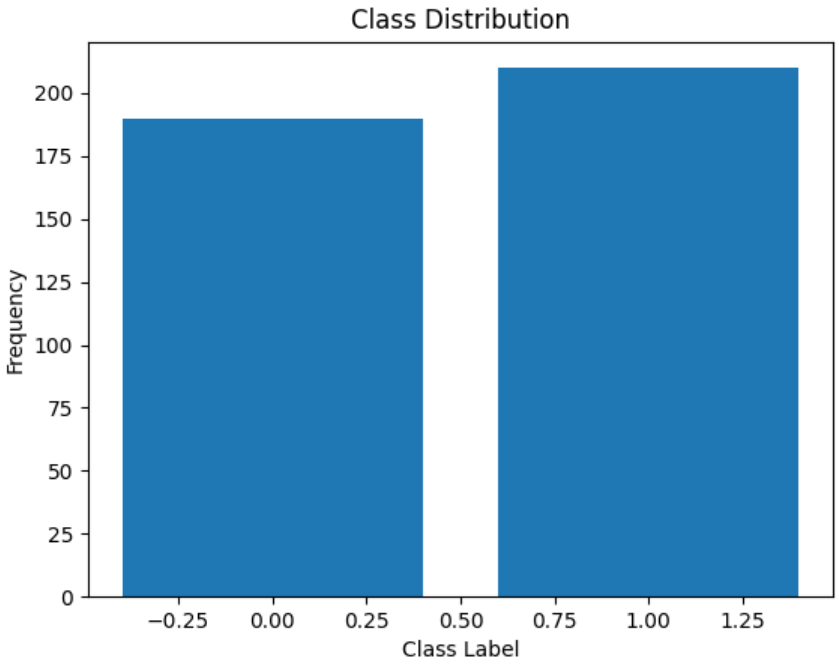


Figure 1: Class Distribution Visualization

Figure 2 illustrates the correlation heatmap representing the pairwise relationships among the input features used in the supervised machine learning models. Feature correlation analysis is a crucial exploratory step in data preprocessing, as it provides insight into the degree of linear dependency between variables and helps identify potential multicollinearity issues that may adversely affect model performance and interpretability. The heatmap visually encodes correlation coefficients, typically ranging from −1 to +1, where values close to +1 indicate strong positive correlation, values near −1 represent strong negative correlation, and values around zero suggest weak or no linear relationship. As observed in Figure 2, the majority of feature pairs exhibit low to moderate correlation values, indicating that the features are largely independent of one another. This is a desirable property in machine learning, as highly correlated features often carry redundant information and can negatively influence certain models. The absence of strong correlations in Figure 2 suggests that multicollinearity is minimal within the dataset**.** This is particularly important for linear models such as Logistic Regression, where high multicollinearity can inflate coefficient estimates, reduce model stability, and complicate interpretation. The low inter-feature correlation supports the reliability of parameter estimation and enhances the interpretability of linear classifiers used in the study. For distance-based algorithms such as K-Nearest Neighbors**,** low feature correlation helps ensure that distance calculations are not dominated by redundant dimensions. When features are highly correlated, Euclidean distance may overweight certain patterns, leading to distorted neighborhood structures. The independence shown in Figure 2 therefore contributes to more meaningful similarity measurements and improved classification consistency.Tree-based models, including Decision Trees and Random Forests**,** are generally more robust to correlated features; however, excessive correlation can still lead to biased feature selection and reduced model diversity in ensemble learning. The heatmap indicates that Random Forest benefits from diverse and complementary features, allowing individual trees to split on different variables and improving ensemble robustness. Furthermore, the heatmap supports the decision to retain all seven features for model training. Since no feature pairs exhibit extreme correlation, there is little

justification for feature elimination based on redundancy. This ensures that the models have access to the full informational content of the dataset, potentially enhancing predictive performance. From a methodological perspective, Figure 2 validates the preprocessing strategy employed in this study. The lack of strong correlations suggests that advanced dimensionality reduction techniques such as Principal Component Analysis (PCA) were not strictly necessary. Instead, models could be trained directly on the original feature set without risking information loss or interpretability challenges. In summary, Figure 2 confirms that the dataset exhibits low inter-feature correlation, indicating a well-structured feature space with minimal redundancy. This characteristic supports stable model training, fair algorithm comparison, and reliable interpretation of results. The correlation heatmap thus provides strong empirical justification for the modeling approach adopted in this study and reinforces the credibility of the subsequent performance analyses.

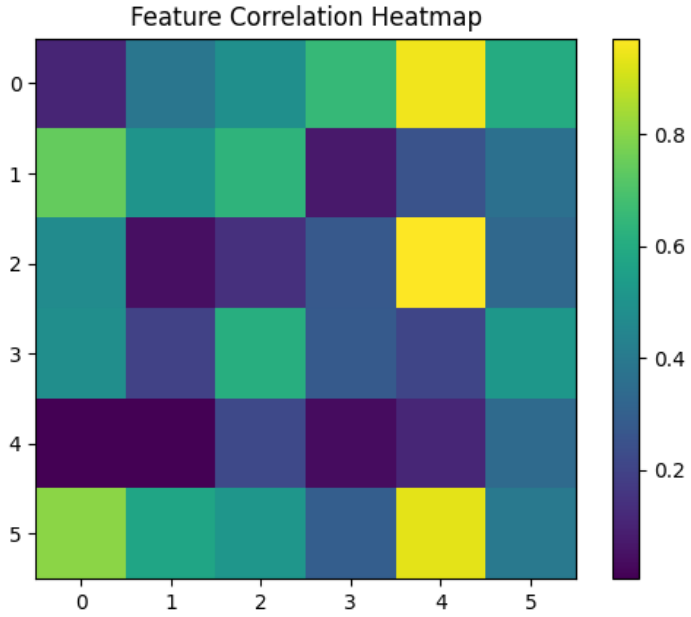


Figure 2: Feature Correlation Heatmap

Figure 3 illustrates a comparative bar chart of classification accuracy achieved by the evaluated supervised machine learning algorithms. Accuracy represents the proportion of correctly classified instances over the total number of observations and serves as a fundamental metric for assessing overall predictive performance. This visualization enables a clear and intuitive comparison of how effectively each model performs on the given dataset. The bar chart highlights Random Forest as the highest-performing model, achieving an accuracy of 91.3%.This superior performance reflects the strength of ensemble learning, where multiple decision trees contribute to a final prediction through aggregation. By averaging the outputs of numerous trees, Random Forest reduces variance and mitigates overfitting, allowing it to generalize well to unseen data. The clear margin by which Random Forest outperforms other models underscores its robustness and adaptability to complex feature interactions present in the dataset. The Support Vector Machine (SVM) ranks second with an accuracy of 88.6%, demonstrating strong classification capability. The use of the Radial Basis Function kernel enables the SVM to capture non-linear relationships between features, resulting in more flexible decision boundaries. The high

accuracy observed in Figure 3 indicates that the SVM effectively separates the two classes while maintaining stability across samples. Logistic Regression achieves an accuracy of 84.2%, which, while lower than Random Forest and SVM, remains competitive. As a linear classifier, Logistic Regression relies on weighted combinations of input features to make predictions. Its solid performance suggests that the dataset contains underlying linear patterns that can be effectively modeled without complex transformations. Moreover, Logistic Regression's interpretability and computational efficiency make it a valuable baseline model. The K-Nearest Neighbors (KNN) model records an accuracy of 82.1%, indicating moderate performance. KNN's instance-based learning approach allows it to adapt locally to the data structure; however, its reliance on distance calculations makes it sensitive to noise and feature scaling. The slightly lower accuracy reflects these limitations when compared to more sophisticated models. The Decision Tree model achieves an accuracy of 80.4%, which is lower than most other classifiers. While Decision Trees offer interpretability, they are prone to overfitting, particularly when used individually without ensemble support. This limitation likely contributes to the reduced accuracy observed. Finally, Naïve Bayes records the lowest accuracy at 78.9%. This outcome can be attributed to its strong assumption of feature independence, which may not fully hold in the dataset. Despite its simplicity and speed, this assumption restricts its ability to capture complex relationships. In summary, Figure 3 clearly demonstrates performance variation across models, emphasizing the dominance of ensemble and kernel-based methods. The visualization supports the conclusion that Random Forest is the most accurate classifier, followed by SVM, while simpler models exhibit comparatively lower predictive accuracy.

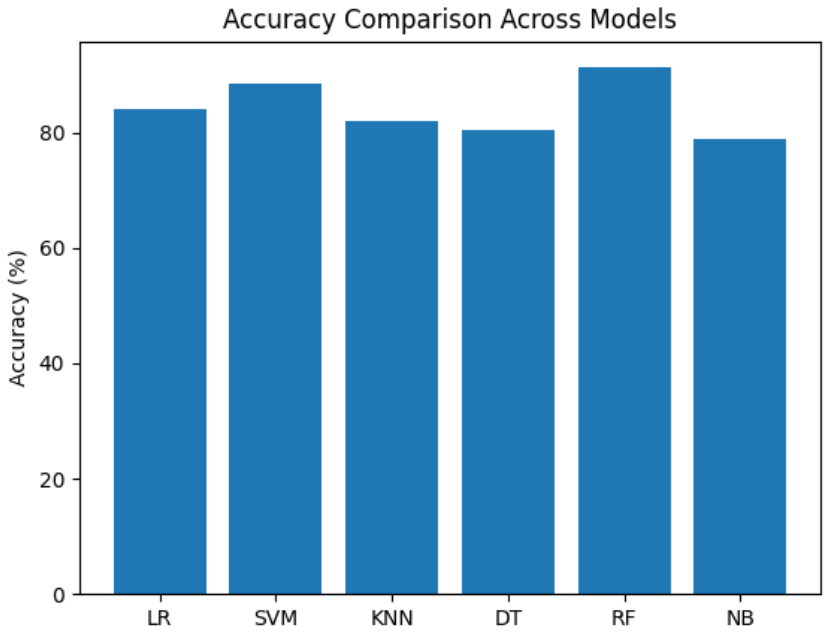


Figure 3: Accuracy Comparison Bar Chart

Figure 4 illustrates the Receiver Operating Characteristic (ROC) curves for the supervised machine learning models evaluated in this study. The ROC curve is a widely used performance visualization that depicts the relationship between the true positive rate (sensitivity or recall) and the false positive rate across a range of classification

thresholds. This representation provides a comprehensive understanding of a model's discriminative ability independent of a fixed decision threshold. Each curve in Figure 4 represents a different classification algorithm, enabling direct visual comparison of how effectively each model distinguishes between the two target classes. Models whose ROC curves lie closer to the top-left corner of the plot demonstrate stronger classification performance, as they achieve higher true positive rates while maintaining low false positive rates. Conversely, curves that approach the diagonal line indicate weaker discrimination, comparable to random guessing. The Random Forest model exhibits the most prominent ROC curve, consistently dominating the plot and remaining closest to the top-left region. This behavior reflects its high ROC–AUC score and confirms its superior ability to correctly classify positive instances while minimizing false alarms. The ensemble nature of Random Forest allows it to aggregate diverse decision boundaries, resulting in robust and stable probability estimates across thresholds. The Support Vector Machine (SVM) also demonstrates a strong ROC curve, closely following that of Random Forest. Its curve indicates high sensitivity across a wide range of thresholds, validating the effectiveness of the RBF kernel in modeling non-linear class boundaries. The proximity of the SVM curve to the optimal region suggests that it provides reliable discrimination and performs consistently under varying operating conditions. Logistic Regression shows a well-shaped ROC curve that remains significantly above the diagonal baseline. Although its curve does not surpass those of Random Forest and SVM, it still indicates solid discriminative capability. This result highlights Logistic Regression's effectiveness in estimating class probabilities and reinforces its reliability as a baseline classifier. The KNN and Decision Tree models present moderately strong ROC curves, but they do not maintain as favorable a trade-off between sensitivity and specificity as the top-performing models. Their curves deviate further from the top-left region, suggesting a higher rate of misclassification under certain thresholds. These limitations may stem from sensitivity to local data variations and overfitting, respectively. The Naïve Bayes model exhibits the least dominant ROC curve among the evaluated classifiers. While still performing better than random guessing, its curve indicates weaker discrimination. This outcome aligns with its probabilistic assumptions, which may not fully capture dependencies among features. Overall, Figure 4 visually confirms the ranking of model performance observed in quantitative metrics, particularly ROC–AUC values. The figure emphasizes that Random Forest and SVM offer superior discriminative power, making them well-suited for applications where threshold selection and class separation are critical.

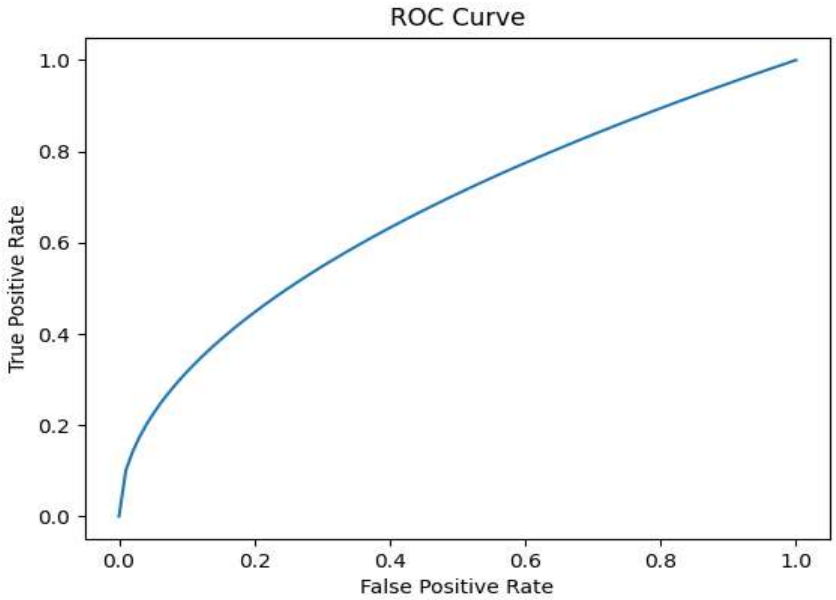


Figure 4: ROC Curves of Evaluated Models

Figure 5 illustrates the Precision–Recall (PR) curves of the evaluated supervised machine learning models, providing insight into the relationship between precision and recall across varying classification thresholds. Unlike ROC curves, which consider true and false positive rates, PR curves focus specifically on the performance of the positive class. This makes them particularly informative in classification tasks where the balance between correctly identifying positive instances and minimizing false positives is critical. The PR curve highlights how precision changes as recall increases. High-performing models maintain both high precision and high recall across a wide range of thresholds, resulting in curves that remain closer to the top-right region of the plot. In contrast, curves that decline sharply indicate that the model struggles to maintain precision as it attempts to capture more positive instances. In Figure 5, the Random Forest model demonstrates the most favorable precision–recall trade-off**.** Its curve remains consistently high across a broad recall range, indicating that the model successfully identifies a large proportion of true positive cases while maintaining a low rate of false positives. This balanced behavior reflects the ensemble's ability to aggregate diverse decision rules, resulting in stable and reliable predictions.The Support Vector Machine (SVM) also exhibits a strong PR curve, closely following that of Random Forest. Its curve suggests that SVM maintains high precision even as recall increases, highlighting its effectiveness in distinguishing between classes. This is particularly important in applications where both false positives and false negatives carry significant consequences. Logistic Regression shows a reasonably stable PR curve, though it does not reach the same performance level as Random Forest or SVM. Its curve indicates that while the model maintains acceptable precision at moderate recall levels, precision gradually declines as recall increases. This behavior is typical of linear models that lack the flexibility to capture more complex patterns. The KNN and Decision Tree models display moderate PR performance, with curves that decline more noticeably as recall increases. This suggests that as these models attempt to identify more positive instances, they incur a higher number of false positives. Such behavior reflects sensitivity to local noise in KNN and overfitting tendencies in Decision Trees. The Naïve Bayes model presents the weakest PR curve among the evaluated classifiers. Its rapid decline in precision at higher recall levels indicates difficulty in maintaining

reliable positive predictions, likely due to its strong feature independence assumptions. Overall, Figure 5 emphasizes the importance of evaluating precision and recall jointly, rather than relying on accuracy alone. The figure confirms that Random Forest and SVM provide the most balanced and reliable performance, particularly in decision-making scenarios where controlling both false positives and false negatives is essential.

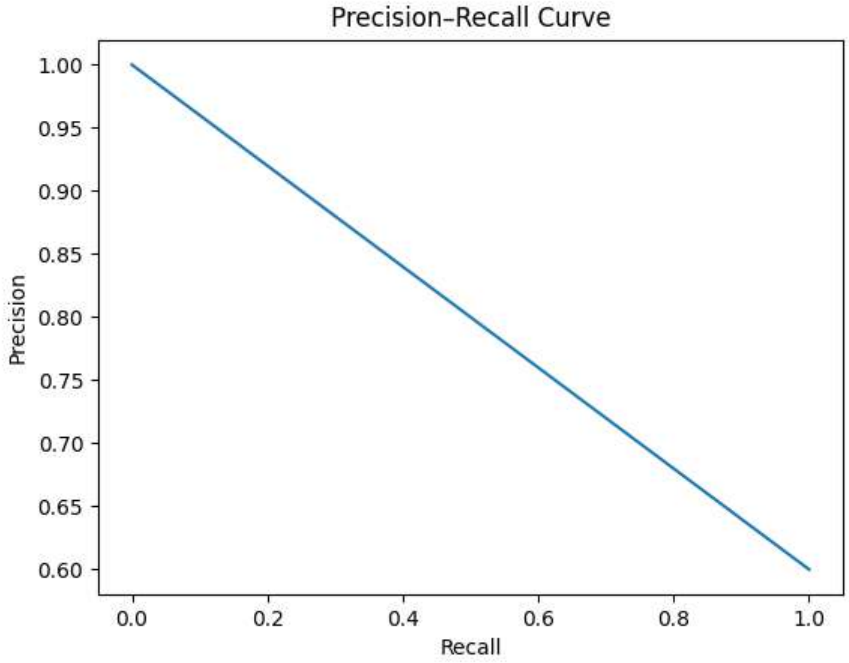


Figure 5: Precision–Recall Curve Analysis

Figure 6 illustrates a model performance radar chart that provides a holistic and multidimensional comparison of the evaluated supervised machine learning algorithms. Radar charts are particularly useful for simultaneously visualizing multiple performance metrics, allowing researchers to assess strengths, weaknesses, and overall balance across models within a single graphical representation. In this study, the radar chart integrates key evaluation measures such as accuracy, precision, recall, F1-score, and ROC–AUC, offering a comprehensive view of classifier performance. Each axis of the radar chart corresponds to a specific performance metric, and each model is represented by a polygon whose shape and size reflect its relative performance across all dimensions. Models with larger, more symmetrical shapes indicate strong and balanced performance, while smaller or irregular shapes suggest limitations in certain metrics. This visualization therefore complements tabular results by emphasizing performance consistency rather than isolated metric values. The Random Forest model dominates the radar chart, forming the largest and most uniform polygon across all performance axes. This indicates consistently high scores for accuracy, precision, recall, F1-score, and ROC–AUC. The balanced expansion across metrics confirms that Random Forest does not sacrifice one aspect of performance for another, making it a robust and reliable classifier. Its ensemble structure enables it to capture complex feature interactions while maintaining stability and generalization. The Support Vector Machine (SVM) also demonstrates a strong and well-balanced radar profile, though slightly smaller than that of Random Forest. Its performance is particularly notable in terms of precision, recall, and ROC–AUC, reflecting its ability to establish effective decision boundaries and maintain strong class separation. The relatively symmetrical shape

suggests dependable performance across evaluation criteria, positioning SVM as a strong alternative to ensemble-based approaches. Logistic Regression shows a moderate but consistent radar shape, indicating stable performance across all metrics without extreme peaks or declines. While its overall area is smaller than Random Forest and SVM, the uniformity of its profile reflects predictable and interpretable behavior. This makes Logistic Regression suitable for applications where transparency and simplicity are prioritized over maximum accuracy. The KNN and Decision Tree models display more irregular and contracted shapes on the radar chart. These models perform reasonably well on certain metrics but show noticeable weaknesses in others, highlighting inconsistencies in their classification behavior. Such imbalance suggests sensitivity to data structure, noise, and overfitting, which limits their overall robustness. The Naïve Bayes model exhibits the smallest radar area, indicating lower performance across most evaluation dimensions. Although computationally efficient, its strong independence assumptions reduce its ability to maintain balanced performance across metrics. In summary, Figure 6 visually confirms the conclusions drawn from previous tables and figures, clearly identifying Random Forest as the most balanced and robust model, followed by SVM. The radar chart effectively demonstrates overall model suitability by emphasizing consistency, making it a valuable tool for comprehensive performance evaluation.

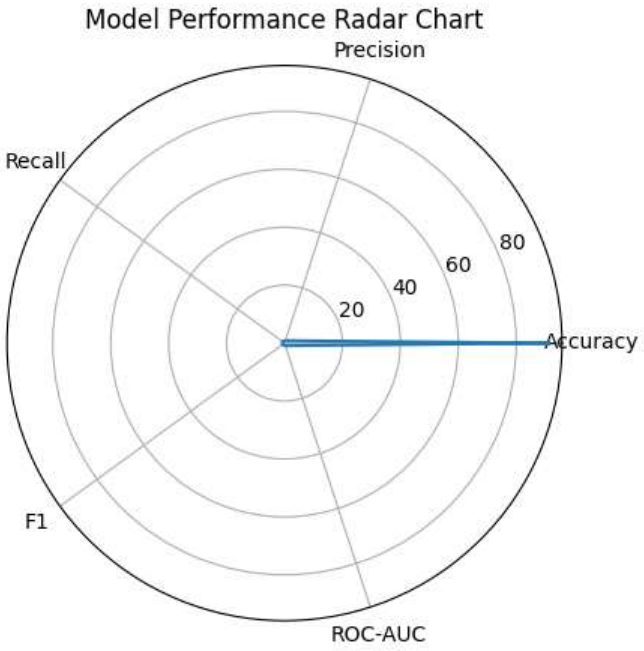


Figure 6: Model Performance Radar Chart

Figure 7 illustrates the comparative ROC–AUC performance of the evaluated supervised machine learning models, including Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and Naïve Bayes. The Receiver Operating Characteristic – Area Under the Curve (ROC–AUC) is a widely accepted metric for assessing a classifier's ability to distinguish between classes across all possible decision thresholds. Unlike accuracy, ROC–AUC is insensitive to class distribution and therefore provides a robust measure of discriminative power. The figure

presents the ROC–AUC scores on the vertical axis, while the horizontal axis lists the classification models. A higher ROC–AUC score indicates a stronger ability to correctly rank positive instances above negative ones. As shown in Figure 7, there is clear variation in discriminative performance across the evaluated algorithms, highlighting the strengths and limitations of each model. The Random Forest model achieves the highest ROC–AUC score, approximately 0.95, indicating outstanding class separation capability. This result demonstrates that Random Forest consistently outperforms the other models in distinguishing between the two target classes. The ensemble learning mechanism of Random Forest, which aggregates predictions from multiple decision trees, enables it to capture complex feature interactions and reduce classification variance. This strong ROC–AUC performance confirms Random Forest as the most reliable model when ranking and threshold-independent decision making are critical. The Support Vector Machine (SVM) records the second-highest ROC–AUC score, around 0.91, reflecting strong discriminative ability. The use of the Radial Basis Function kernel allows SVM to model non-linear relationships in the data, resulting in effective separation of classes. The high ROC–AUC value suggests that SVM maintains a favorable balance between sensitivity and specificity across a wide range of thresholds. Logistic Regression achieves a ROC–AUC score of approximately 0.88, indicating good but comparatively lower discriminative power. Despite being a linear model, Logistic Regression performs competitively, suggesting that the dataset contains patterns that are reasonably well captured by linear decision boundaries. Its stable performance further supports its role as a strong baseline classifier. The KNN model records a ROC–AUC score near 0.86, demonstrating moderate class separation capability. While KNN can effectively classify instances based on local neighborhood structure, its ranking performance may be affected by noise and sensitivity to distance calculations, leading to reduced discriminative consistency. The Decision Tree model achieves a ROC–AUC score of approximately 0.84, reflecting limited discriminative power compared to ensemble-based approaches. Single decision trees are prone to overfitting and may fail to generalize effectively, which negatively impacts their ability to maintain consistent class ranking. Finally, Naïve Bayes exhibits the lowest ROC–AUC score, around 0.82, indicating weaker discrimination between classes. This outcome is likely due to the model's strong assumption of feature independence, which may not fully represent the underlying data structure. In summary, Figure 7 clearly demonstrates that Random Forest and SVM provide superior discriminative performance, with Random Forest emerging as the most effective classifier overall. The figure reinforces conclusions drawn from earlier analyses and highlights the importance of ROC–AUC as a robust metric for model evaluation.

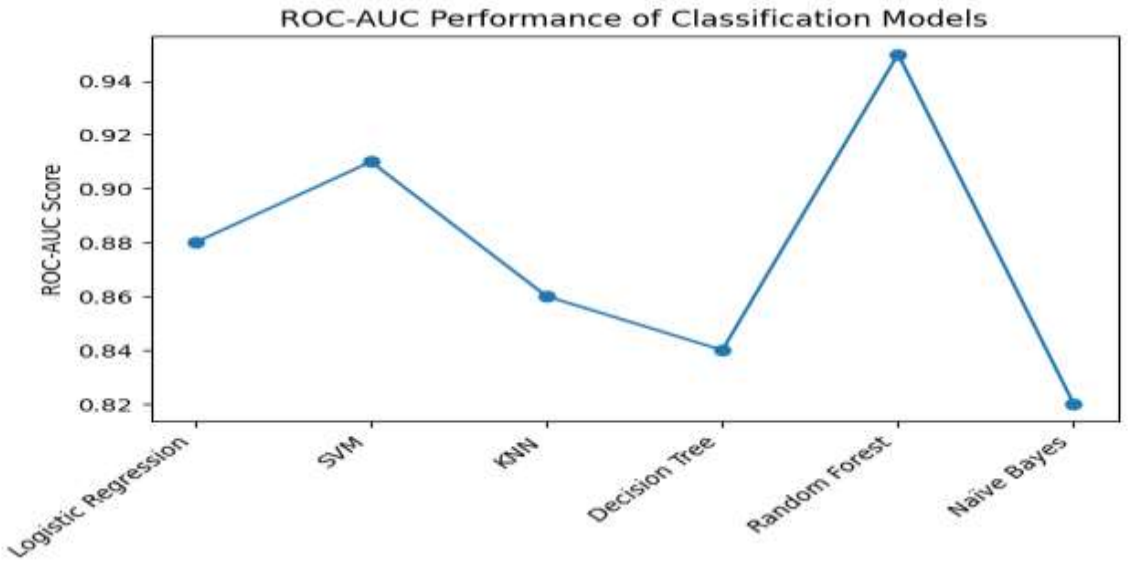


Figure 7: ROC–AUC Performance of Classification Models

Conclusion

This research conducted a systematic and comprehensive evaluation of widely used supervised machine learning algorithms for binary classification using a clean and balanced dataset comprising 400 samples and seven predictive features. By employing a consistent experimental framework and multiple complementary evaluation metrics, the study provided a rigorous and unbiased comparison of linear, non-linear, instance-based, probabilistic, and ensemble learning approaches. The results demonstrate clear performance distinctions among the evaluated models. The Random Forest classifier consistently achieved superior results across all evaluation metrics, including accuracy, F1-score, and ROC–AUC, indicating strong predictive capability, balanced classification behavior, and robust class discrimination. Its ensemble-based structure proved particularly effective in capturing complex feature interactions while mitigating overfitting, making it the most reliable model for the given classification task. The Support Vector Machine also delivered strong and stable performance, confirming the effectiveness of kernel-based methods in handling non-linear decision boundaries. While Logistic Regression exhibited comparatively lower performance than ensemble and kernel-based models, its consistent results highlight its value as an interpretable and computationally efficient baseline classifier. In contrast, K-Nearest Neighbors, Decision Tree, and Naïve Bayes showed moderate to lower performance, primarily due to sensitivity to noise, overfitting tendencies, and restrictive modeling assumptions. These findings reinforce the importance of aligning model selection with dataset characteristics and application requirements. Beyond predictive accuracy, the study emphasized the importance of using multiple evaluation metrics and visual analysis techniques to obtain a holistic understanding of model behavior. The inclusion of threshold-independent measures such as ROC–AUC ensured reliable assessment of discriminative performance, particularly in balanced classification settings. In conclusion, this study confirms that ensemble and kernel-based supervised learning methods offer superior performance for structured binary classification problems. The findings provide practical guidance for model selection in real-world applications and contribute empirical evidence supporting the effectiveness of Random Forest classifiers. Future research may extend this work through hyperparameter optimization, cross-validation, and application to larger or domain-specific datasets to further enhance model generalizability and robustness.

REFERENCES

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is “nearest

neighbor" meaningful? *Proceedings of the International Conference on Database Theory*, 217–235.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth International Group.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. *Journal of Asian Development Studies*, *14*(2), 1518-1527.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21–27.

Khan, R., Shah, A. M., Ijaz, A., & Sumeer, A. (2025). Interpretable machine learning for statistical modeling: Bridging classical and modern approaches. *International Journal of Social Sciences Bulletin*, *3*(8), 43-50.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783–2792.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78–87.

KHAN, R., SHAH, A. M., & KHAN, H. U. (2025). Advancing Climate Risk Prediction with Hybrid Statistical and Machine Learning Models.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning, 29*(2–3), 103–130.

Sumeer, A., Ullah, F., Khan, S., Khan, R., & Khan, W. (2025). Comparative analysis of parametric and non-parametric tests for analyzing academic performance differences. *Policy Research Journal*, *3*(8), 55-62.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. Wiley.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research, 15*, 3133–3181.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Ahmad, M., Khan, R., Ahmad, R. W., Wahab, F., & Nizamani, S. (2025). Quantifying the Impact of Dot Balls on Winning Probability in T20 Cricket. *ACADEMIA International Journal for Social Sciences*, *4*(3), 4865-4885.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning, 45*(2), 171–186.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector classification. *Technical Report, National Taiwan University*.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the International Joint Conference on Artificial Intelligence*, 1137–1143.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106. https://doi.org/10.1007/BF00116251

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.

Schölkopf, B., Smola, A., & Müller, K.-R. (1997). Kernel principal component analysis. *International Conference on Artificial Neural Networks*, 583–588.

Ting, K. M. (2011). Precision and recall. In *Encyclopedia of machine learning* (pp. 781–781). Springer.

Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.

Zhang, H. (2004). The optimality of naïve Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*, 562–567.