ISSN: 2093-5374 / E-ISSN: 2093-5382

A Hybrid Approach for Customer Segmentation Using Ensemble Methods: Bagging and Voting Classifiers

Sufaira Shamsudeen[1*], and Dr.K. Ranjith Singh[2]

[1*]Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. sufaira@mesmarampally.org, https://orcid.org/0009-0003-3783-9117

[2]Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. ranjithsingh.koppaiyan@kahedu.edu.in, https://orcid.org/0000-0003-2651-2509

Received: June 10, 2025; Revised: July 21, 2025; Accepted: September 03, 2025; Published: September 30, 2025

Abstract

Customer Classification is the approach of categorizing diverse customers into interrelated classes and thereby helping banks to know loyal customers, by offering, more tailored products and services, ultimately increasing business revenue. The study explores the evolution of Machine Learning (ML) and Ensemble Learning (EL) techniques in the banking domain, progressing from basic to advanced classification methods for customer categorization based on credit information. The objective is to identify profitable customers and classify them into five categories like Outstanding, Excellent, Good, Satisfactory, and Bad. Using algorithms such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN), the researcher finds that, the Decision Tree classifier underperforms in comparison to others. The main aim of the study is to strengthen the classifiers using the ensemble learning method- Voting and Bagging approaches for the multi-class customer segmentation. Initially, the study integrates the strengths of neural network optimizers with the DT classifier, and ensembles DT with the base models using a Voting Classifier. A neural network solver such as Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (Adam), and Limited–memory Broyden–Fletcher–Goldfarb–Shannon (L-BFGS) is used to find the best local minima for the Decision Tree algorithm. Then crafting an ensemble base model using Voting Classifier, aims to enhance the Decision Tree classifier's performance despite its initial shortcomings. The efficiency of the proposed ensemble Decision Tree models using different optimizers such as VC-DT-ADAM, VC-DT-SGD, and VC-DT-L-BFGS, and ensemble with the base models such as VC-DT-KNN, VC-DT-SVM, and VC-DT-RF compared for better selection. Both hard and soft voting classifier methods are computed and analyzed. The result demonstrates that all the ensemble models are fine in their performance. Despite the neural network solver VC-DT-SGD delivering better classification performance with an accuracy score of 0.98 and all the base ensemble models scoring 0.98, VC-DT-RF offers superior performance with an accuracy score of 0.99 and auc value of 1.000 using both hard and soft voting classifiers. Furthermore, the study enhances the DT classifier performance based on the Bagging Classifier (BC) approach, i.e., BC-DT. Amongst all hybrid models, bagging classifier with DT i.e., BC-DT is superior with the score 0.99 and with auc value of 0.999. Various evaluation metrics are measured here to highlight the performance of the models and compared to assess the efficacy of the proposed models. Taking into account all evaluation metrics beyond accuracy, the decision tree classifier combined with a neural network optimizer excelled all other hybrid models. The

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 16, number: 3 (September), pp. 414-432. DOI: *10.58346/JOWUA.2025.I3.025*

*Corresponding author: Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.

simulation results show that the hybrid model produces superior performance compared to the base classifier models across all key metrics.

Keywords: Machine Learning, Artificial Neural Network, Decision Tree Classifier, Ensemble Learning, Voting Classifier, Bagging Classifier.

1 Introduction

Customer analytics is essential for driving the growth of businesses within the banking sector (Das, 2015). The modern world is handling big data from diverse sources, such as the Internet of Things (IoT), smart devices, social media, financial transactions, etc. (Sehgal & Soni, 2024). Massive unstructured and structured data is stored in the database; many businesses have analyzed this data for better decision-making using advanced analytical techniques. Customer analytics is used in many sectors mainly in online shopping, social media, websites, and customer-centric business organizations, etc. Business organizations around the world are approaching their businesses with a customer-centric view, connecting the large amount of customer data to make an intelligent, predictable system. The proposed study deals with the bank customer data to analyze the customer credit information for identifying the potential customer. In the banking business world nowadays, we are focusing more on customer-centric data than product-oriented data to develop potential business growth. As mentioned in (IBM Corporation, 2012), acquisition of big data from various sources, and in the banking sector, more transactional data is generated daily, which has been recycled for many purposes leading the business's potential revenue. Banks are storing a rising amount of data, such as customer behavior, transactional data, credit, and demographic information, which gives them the chance to perform predictive analytics and thereby develop the growth of the business. However, managing the vast volume of data effectively and producing insights with actual business value present significant problems for data scientists (Sun et al., 2014). Hence, using advanced analytical tools such as machine learning techniques, unstructured and structured data in the database may be analyzed systematically to foreknow the better decision.

Predicting potential customer through customer predictive analytics and segmentation using Machine Learning (ML), plays a pivotal role in business area. For the growth of the business in the banking sector, the availability of unstructured data through many transactions and procedures are vast. Managing this huge information and discovering unknown patterns in a large dataset is absolutely essential. To analyze and discover the pattern from this volume of data, the organization needs powerful techniques like machine learning, which has been used in areas such as financial services, healthcare, hotel operations management, customer relationship management and more for making data-driven decisions (Talón-Ballestero et al., 2018). Machine learning techniques are classified into supervised, unsupervised and reinforcement techniques (Noori, 2021). The article focused on supervised learning algorithms to classify and predict the potential customer. These techniques are often used to classify the customers based on their demographic and credit information. The classification level in the proposed work is categorized into five distinct classes. Through classification models, different classes are identified and well defined distinctly. The purpose of the classification is to understand the distinct classes of future data objects, which are recycled for many decision-making systems, based on historical information. In classification approach, the model learned using training set of the data, and the test data is used to evaluate the information gained model. Although, the real dataset for the proposed study is classified using different machine learning classifier such as K-Nearest Neighbor, Support Vector Machine, Random Forest, Decision Tree and Artificial Neural Network, out of which DT underperformed than others (Vasuki et al., 2025). Since no technique is ideal for every data set, various classification algorithm is developed in the literature review (Gulsoy & Kulluk, 2019). It may vary, based on the dataset size, features and many other aspects. However, in the proposed dataset decision tree accuracy score is less expectancy with 0.81 than others which have more than 0.90 measure. The aim of the study is to improve the DT model by adopting the two ensemble approaches – Voting and Bagging Classifiers.

1.1 Literature Review

A voting classifier is an ensemble approach that combines the multiple classifiers to make a better prediction, which is the final one. This method is used for both binary and multiclass classification. The proposed study emphasizes multiclass classification; a voting classifier used to fuse the base models with the DT model and enhance the measures we adopted. Moreover, the voting classifier is used in the proposed study is to ensemble the DT model with neural network optimizers to demonstrate the result, which meets the mark of the individual DT model. In (Jindal et al., 2022), the voting strategies are two: soft and hard voting. In the soft voting approach, each classifier's prediction is made based on the class probability, and the final prediction is concluded with the highest average probability class. Therefore, the models will be more effectively predicted if the classifiers are well calibrated. In the hard voting classifier (Jindal et al., 2022), the final prediction is determined by the majority vote of all individual classifiers. In multiclass classification, it opts for the class that has the most votes over the others. According to the scikit-learn documentation, an ensemble meta-estimator known as a bagging classifier fits basic classifiers independently on arbitrary subsets of the original dataset, combining their unique predictions into a final prediction through voting or averaging (Santhosh et al., 2016). By adding randomization to the creation process of a black-box estimator such as a decision tree and creating an ensemble from it, a meta-estimator of this type may usually be used to lower the variance of the estimator. Bagging (Singh & Singh, 2019), also known as Bootstrap Aggregating (Alfaro et al., 2013), is a powerful ensemble technique that has outperformed individual models by reducing variance and obtaining higher predictive performances. This is more effective in the models that are prone to overfitting, such as decision trees. Overfitting and underfitting categorize the models with less accuracy. For the large and complex datasets, the simple machine learning models may be weak learners (Datta et al., 2023). To improve the performance of the models, the combined approach or ensemble methods are required. The voting approach is a more robust and reliable approach to create a strong model to learn accurately. (Mungra et al., 2020) Voting-based ensemble model Majority Voting (MV) using five supervised machine learning classifiers used for better sentimental analysis. Not only in sentiment analysis but also in various fields strengthen the decision-making process is strengthened using ensemble learning approaches, obtaining better predictions or outcomes than the individual model (Tripathi et al., 2022). In the financial system, the development of credit scoring systems has a major role; therefore, a new hybrid ensemble model (Zhang et al., 2021) is introduced here for better prediction (Narang et al., 2025). Based on the survey, the researcher realized the significance of the ensemble models and the ease of better prediction. The main advantage of the ensemble models is the ability to reduce the generalization errors for better classification performance. The combination of the outcomes of multiple models through an ensemble approach like voting and bagging methods for multiclass classification strengthens the performance of algorithms to learn and predict while reducing their weaknesses. As a result, the accuracy, robustness, and handling of overfitting were enhanced.

In the present research initiative, the researcher proposes to study the application of artificial intelligence in banks for customer profiling and identifying profitable customers. Machine learning and deep learning approaches enable the banking industry to know the customers in depth, what they are interested and ensure customer needs, and help the customers to fulfil their expectations from the banks. The experimentation is done on the real dataset from the bank because of confidentiality, the name of the bank is concealed. The analysis of the proposed system focuses on customer demographics and credit information, aiming to filter the profitable customers by categorizing the customers into five distinct classes: Class 0–Class 4. This article focuses more on improving the simple decision tree classifier using the ensemble method for multiclass classification. In short, the researcher proposes voting and bagging classifier approaches to multiply the decision tree classifier's performance. The main highlight of the proposed work lies in the following aspects:

A neural network optimizer's ensemble with a decision tree classifier to enhance the classification performance and thereby predict the profitable customer class accurately.

Ensemble technique, voting classifier used here to integrate the simple decision tree classifier with an optimizer, which underperformed than other models.

Using the voting classifier, the base models are combined with a decision tree to demonstrate an accuracy score higher than simple models.

Bagging Classifier, a bootstrap technique is also defined here to highlight enhanced performance when combined with a Decision Tree classifier.

The rest of this article is systematized in a way an overview of the decision tree classifier, and ensemble methods - voting classifier and bagging classifier are described in section II. Section III, explains the proposed hybrid approach of decision tree with neural network optimizers - VC-DT-ADAM, VC-DT-SGD, and VC-DT-L-BFGS, ensemble approach of decision tree with base models - VC-DT-KNN, VC-DT-SVM, and VC-DT-RF, and decision tree with bagging classifier BC-DT frameworks (Petrovic & Alvarez, 2025). Section IV outlines the experiments conducted in these ensemble models and the results will be compared (Kavitha, 2025). Final remarks and future scope are explained in section V.

2 Materials & Methods

This section discusses datasets and methodologies implemented in the research. The dataset's dimensions, standardization of datasets, and various ensemble learning techniques are discussed. Especially, the voting classifier and the bagging classifier are explained in detail.

2.1 Dataset

The dimensions and characteristics of the dataset determine the categorization accuracy. Finding the bank's real-time data, however, was an extremely time-consuming process. The primary obstacle in obtaining information from the bank is its reluctance to investigate client details because of confidentiality concerns. The dataset used in this work includes almost one lakh records, 43 labelled attributes that were gathered from the Indian Bank. The bank's name will not be revealed due to privacy concerns. Data pre-processing, including data analysis, denoising, imputation, data reduction, integer encoding categorical variables, oversampling, normalization, dividing the dataset into training and test sets, and feature scaling, is likely necessary because the actual dataset is untidy. The algorithm is then trained using the training set, and evaluated using cross-validation techniques and other non-parametric measures. The bank customers are classified into five categories: Class0 – Class1, based on their credit history and thereby these classes of customers are cost-effective customers for the bank.

2.2 Methods

Ensemble learning (Tripathi et al., 2022; Polikar, 2012; Kumari et al., 2021) is a machine learning technique that combines different simple models to produce better predictions to improve the overall performance of the system. Each model in the ensemble may have different strengths and weaknesses, so by combining them, the ensemble can figure make predictions. Here, the methods used for the classification are ensemble learning techniques – Voting and Bagging approaches.

Voting Classifier

A Voting Classifier is a machine learning model that trains on a set of models and predicts the outcome based on the class with the best chance of being chosen. It simply aggregates the results of each classifier that is given into the Voting Classifier and predicts the output class based on the greatest number of votes (El-Kenawy et al., 2020; Leon et al., 2017). The formula for calculating the majority vote V for a given sample x is given by (Pratama et al., 2024),

$$V(x)=\text{mode}\{C1(x), C2(x), C3(x)\} \text{ ……………………} (1)$$

Where C1, C2, and C3 represent the predictions and mode refers to the most frequent prediction.

The Voting Classifier supports two types of voting: hard voting and soft voting. Hard voting is often referred to as majority voting. In this case, each classifier votes on a class, and the majority wins. In statistical terminology, the anticipated target label of the ensemble is the mode of the distribution of independently predicted labels (Vujović, 2021). In Soft Voting, base classifiers produce probabilities or numerical scores. It calculates the average score (or likelihood) and compares it with a threshold number, which means the output class is the prediction based on the average of probabilities.

Bagging Classifier

Bagging is a prominent ensemble learning strategy used to increase the performance and durability of machine learning models, particularly those that are unstable, such as decision trees (Alqahtani & Ilyas, 2024). In this approach, the numerous models (base learners) are trained independently and their predictions are combined to improve the model's overall performance. The main idea underlying bagging is to reduce model variance and boost robustness by training several learners on slightly different subsets of data. To create diverse models, a bootstrap sample is created by randomly selecting data points from the original dataset with replacement. This randomness allows each base learner to have a unique perspective on the data. Once all of the base learners have been trained, their predictions are merged to generate a final prediction.

3 Proposed Methodology

The proposed methodology focuses on improving the performance of traditional classification algorithms by taking advantage of ensemble learning and hybrid models, particularly focusing on the improvement of Decision Tree (DT) classifier performance in customer classification algorithms. The objective is to discover profitable customers and categorize them into five distinct classes: Outstanding, Excellent, Good, Satisfactory, and Bad, based on their credit information. This methodology integrates advanced techniques from Machine Learning (ML) and Ensemble Learning (EL) to create a robust, accurate, and efficient customer classification system.

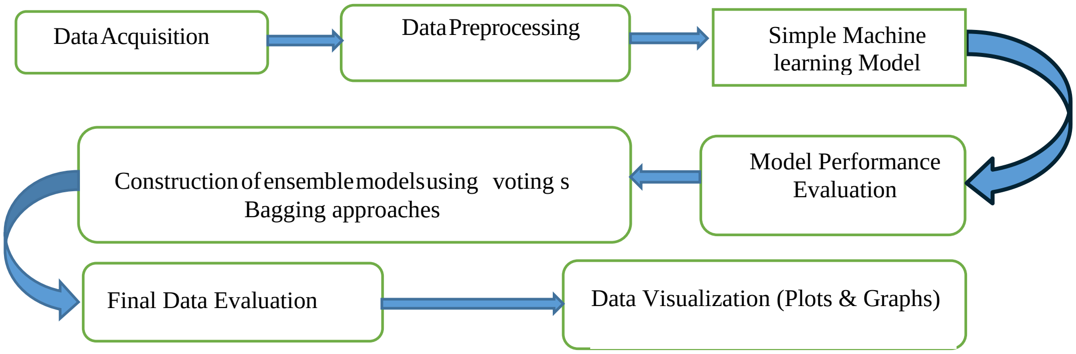


Figure 1: Framework for Proposed Methodology

Figure 1 illustrates the key components of the proposed methodology. The following steps explain each component in detail.

3.1 Dataset and Preprocessing

Gather credit-related information from the real bank consisting of over one lakh records and 43 labeled features. Due to privacy concerns, the name of the bank will not be revealed. The real bank dataset is most likely messy and requires data pre-processing such as data analysis, denoising, imputation, data reduction, integer encoding of categorical variables, oversampling, normalization, separating the dataset into training and test sets, and feature scaling. The algorithm is then trained using the training set, then evaluated using cross-validation techniques.

3.2 Overview of Traditional Classifiers and Evaluation Metrics

Initially, train and evaluate the performance of different machine learning base models including KNN, SVM, DT, RF, and ANN. Using evaluation metrics, measures the performance of base models. The experiments show moderate performance in simple classifiers and that the Decision Tree classifier underperforms compared to others with an accuracy score of 0.81.

3.3 Proposed methods – Voting and Bagging Classifiers

The proposed methodology for this study is to improve the performance of client categorization models in the banking area by using Machine Learning (ML) techniques, specifically Ensemble Learning (EL) methods. The purpose is to discover and categorize clients based on their credit information into five different categories. The study begins by categorizing clients based on credit data using five well-known classification algorithms. While all of these methods are tested, the Decision Tree (DT) classifier is found to perform worse than the other classifiers. To address the Decision Tree model's weaknesses, the study uses neural network optimizers to improve its performance. To optimize the DT classifier's training process, three solvers are used: stochastic gradient descent (SGD), adaptive moment estimation (Adam), and limited-memory Broyden-Fletcher-Goldfarb-Shannon (L-BFGS). These optimizers assist in locating the optimum local minima, hence enhancing the Decision Tree's classification accuracy. Next, an ensemble technique is used to merge many classifiers into a more robust model. An ensemble approach voting classifier is used here to combine the Decision Tree classifier with various basic models like KNN, SVM, and Random Forest. Each model's predictions are aggregated using both hard and soft voting classifiers. The voting method is expected to overcome the constraints of individual classifiers by exploiting their combined capabilities.

3.4 Model Evaluation and Selection

The performance of the ensemble models, including the various Voting Classifier versions, is assessed using a variety of measures, with a focus on accuracy and AUC. The models being evaluated are VC-DT-ADAM (DT Ensemble with Adam Optimizer), VC-DT-SGD (DT Ensemble with SGD Optimizer), VC-DT-L-BFGS (DT Ensemble with the L-BFGS Optimizer), VC-DT-KNN, VC-DT-SVM, and VC-DT-RF (DT-Integrated Base Models). Among them, VC-DT-RF performs best, with an accuracy score of 0.99 and an AUC value of 1.000 utilizing both hard and soft voting methods. For further study of the ensemble learning technique, a Bagging Classifier (BC) is applied to the Decision Tree model. This method, known as Bagging Decision Tree (BC-DT), is found to outperform other hybrid models, achieving an accuracy score of 0.99 and an AUC value of 0.999.

The study measures a variety of evaluation metrics, including accuracy, precision, recall, F1-score, and AUC, to assess the performance of the models comprehensively. The results suggest that the hybrid models excel over traditional classifiers.

4 Evaluation Metrics

The model evaluation is done by measuring the performance factors such as accuracy (Vujović, 2021), Cohen’s Kappa (Naidu et al., 2023), F-score, Precision, Recall, ROC-AUC (Yang et al., 2021) score, TPR, and FPR. The percentage of accurate classifications is what is used to calculate the accuracy score. A higher accuracy rating means that more of the model's predictions were accurate.

$$\text{Accuracy}=(TP+TN)/(TP+TN+FP+FN) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

Whereas, TP-True Positive, TN- True Negative, FP- False Positive, and FN- False Negative. These are all the predicted classes used in classification to gauge how accurate the model is.

Cohen’s Kappa is suggested as a way to gauge how well two classifiers agree. The chance of agreement (total accuracy) minus the possibility of disagreement (random accuracy) divided by 1 minus the probability of disagreement is used to determine the Kappa coefficient. This score is positive, indicating that the parties have reached some sort of understanding. Higher values of the Kappa coefficient signify stronger agreement or better performance, with values ranging from -1 to 1.

The F score is a single score formed by combining recall and precision. Consequently, a higher F score is preferred since it denotes a better balance between recall and precision. F-score = 2 * (precision * recall) / (precision + recall). In most circumstances, the F1 score is more useful than the accuracy, especially if your class is unevenly distributed. When false positives and false negatives are almost equal in cost, accuracy performs best. It is desirable to include both Precision and Recall if the costs of false positives and false negatives vary significantly.

Precision measures how well a model can forecast a particular classification. A greater precision indicates that a model provides more relevant results than irrelevant ones.

Recall quantifies the proportion of pertinent data points that the model accurately identifies. As it shows a reduced rate of false negatives and an outstanding ability to record all positive cases, a higher recall value is typically recommended.

ROC-AUC weighed the performance of a multi-class classification model using a characteristic curve, which is a probability graph, demonstrating the curve between two parameters tpr (true positive rate) and fpr (false positive rate) at various threshold levels.

A classification model's accuracy in accurately classifying actual positive event is measured by a performance parameter called TPR. The ratio of true positives to the total of true positives and false negatives is used to compute it. FPR is a performance indicator that quantifies the percentage of real negative instances that classification models mistakenly classify as positive. The ratio of false positives to the total of false positives and true negatives is used to compute it.

5 Results and Discussions

Table 1: Comparison of Simple Machine Learning Models based on Performance Metrics

Algorithm	Precision	Recall	F1-Score	TPR	FPR	AUC	Kappa Value	Accuracy
SVM	1.00	0.89	0.94	0.891	0.001	0.997	0.95	0.96
KNN	0.94	0.86	0.90	0.860	0.014	0.991	0.94	0.96
DT	1.00	1.00	1.00	0.999	0.000	0.963	0.77	0.81
RF	0.95	1.00	0.98	0.954	0.000	0.999	0.95	0.96
ANN	1.00	1.00	1.00	0.999	0.000	1.000	0.97	0.98

The comparison of simple machine learning models according to the performance metrics is shown in Table 1. With an accuracy score of 0.81, the decision tree (DT) underperformed others. Due to the large and complex dataset, the decision tree can easily overfit, the model can capture noise in the training data instead of the general pattern, and it is difficult to learn complex patterns. The splitting of data in the decision tree is recursive, which makes it computationally expensive, especially in a large dataset with a large number of features. In addition, the usage of memory is substantially greater if the dataset size is large and time-consuming. By the ensemble learning approach, the model can improve its performance and ease the bias towards complex features. The ensemble methods combine the predictions of multiple trees and strengthen various machine learning tasks. Hence, the overfitting issues generated in the single decision tree are reduced and better generalization is achieved by the ensemble approach.

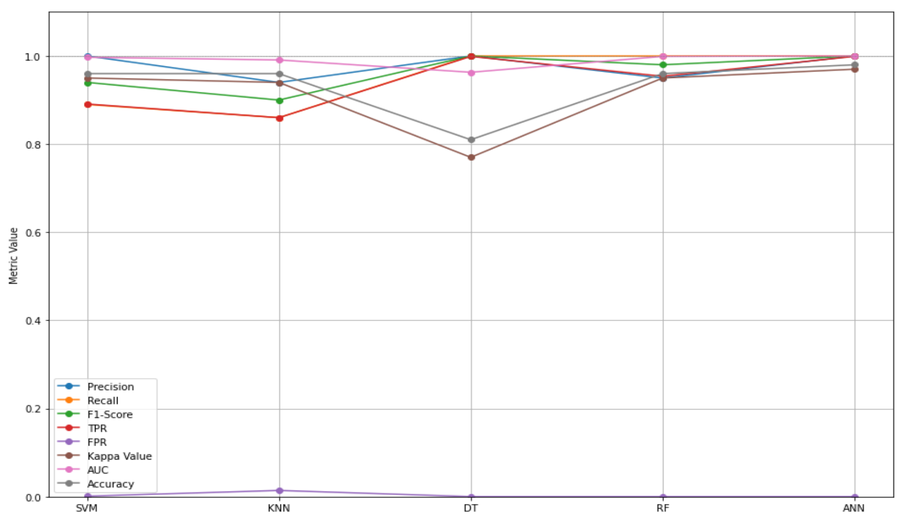


Figure 2: Comparison of ML Models Based on Performance Metrics

In the above figure, Figure 2 displays the model’s performance based on metrics. The comparative study of the different models was weighed, based on the metrics gauged. At this model evaluation time, the indicators like precision, recall, f1-score, TPR, FPR, Kappa value, AUC, and accuracy are measured. Among of these, ANN and RF show exceptional performance for identifying true positive classes with precision and recall value 1.00. The metric f1-score of ANN and RF indicates the optimal and strong performance with a score of 1.00, specifying the robustness and good generalization. The kappa value, the measure of more accurate agreement for categorical data, representing ANN and RF is strong agreement with values 0.97 and 0.95 respectively. For a more comprehensive evaluation, the AUC metric is measured here along with precision, recall, and f1-score. Based on this metric, ANN and RF measure demonstrate excellent analytical capability with the score 1.000 and 0.999 respectively. According to the accuracy, among all models, ANN resulted in outstanding execution with a measure of 0.98. The RF, KNN, and SVM display an accuracy score of 0.96 compared with models, DT underperformed with an accuracy of 0.81, although considering other metrics such as precision and recall, it displays excellent performance in classifying all true classes positively. However, the kappa value is 0.77, indicating that the class is imbalanced because of overfitting. Lower accuracy score compared to others, inferring probable issues with generalization.

Based on the unseen data, the decision tree (DT) model's performance needs to be fine-tuned due to overfitting and generalization. Implementing ensemble methods like Voting Classifier and Bagging Classifier, along with base models and neural network optimizers, can excellently tackle the overfitting and generalization issues of the Decision Tree model, thereby enhancing the stability of the model. Evaluating the models by all metrics provides insights into the strengths and weaknesses of each model comprehensively, thereby selecting a better model and enhancing the weaker model.

Table 2: Comparison of Performance of Ensemble Models Using Soft Voting Classifier (MLP + DT+ Optimizers)

Algorithm	Precision	Recall	F1-Score	TPR	FPR	AUC	Kappa Value	Accuracy
SOFT VOTING CLASSIFIER- VC-MLP-DT-ADAM								
Class 0	0.98	1.00	0.99	0.982	0.000	1.00	0.89	0.92
Class 1	0.99	1.00	1.00	0.991	0.000	1.00		
Class 2	0.97	1.00	0.98	0.969	0.000	1.00		
Class 3	0.76	0.93	0.84	0.758	0.021	0.97		
Class 4	0.97	0.65	0.78	0.971	0.076	0.99		
SOFT VOTING CLASSIFIER- VC-MLP-DT-LBFGS								
Class 0	0.71	1.00	0.83	0.705	0.000	0.96	0.88	0.90
Class 1	0.99	0.90	0.94	0.993	0.025	1.00		
Class 2	0.99	1.00	1.00	0.994	0.000	1.00		
Class 3	1.00	0.73	0.84	1.000	0.072	0.96		
Class 4	0.94	0.93	0.93	0.941	0.017	0.96		
SOFT VOTING CLASSIFIER- VC-MLP-DT-SGD								
Class 0	0.98	1.00	0.99	0.982	0.000	1.00	0.98	0.98
Class 1	0.99	1.00	1.00	0.991	0.000	1.00		
Class 2	1.00	1.00	1.00	0.997	0.000	1.00		
Class 3	1.00	0.92	0.96	1.000	0.024	0.96		
Class 4	0.94	1.00	0.97	0.937	0.000	0.99		

Table 2 lists the various metrics evaluated on the decision tree (DT) classifier using an ensemble approach such as a soft voting classifier to enhance the performance of the decision tree. The neural network optimizers such as Adam, LBFGHS, and SGD, were combined with DT and evaluated the performance. The entire hybrid model has increased the performance of DT, especially the hybrid model VC-MLP-DT-SGD which is superior to others. The inference on the VC-MLP-DT-Adam model implies that the performance measure of this soft voting classifier performs exceptionally well for 0, 1, and 2 classes with high precision and perfect recall. Despite the fact, classes 3 and 4 restricted classification performance. The model VC-MLP-DT-LBFGS struggles with classes 0 and 3, signifying the enhancement in the performance metrics precision and recall. On the other hand, the model performs well for classes 1, 2, and 4. Excellent performance of the ensemble model VC-MLP-DT-SGD with balanced metrics leads to a superior model. The value of AUC in all hybrid models indicates that a multiclass classification model with five classes is doing well. The strong results imply reliability in expectations and accurate classification.

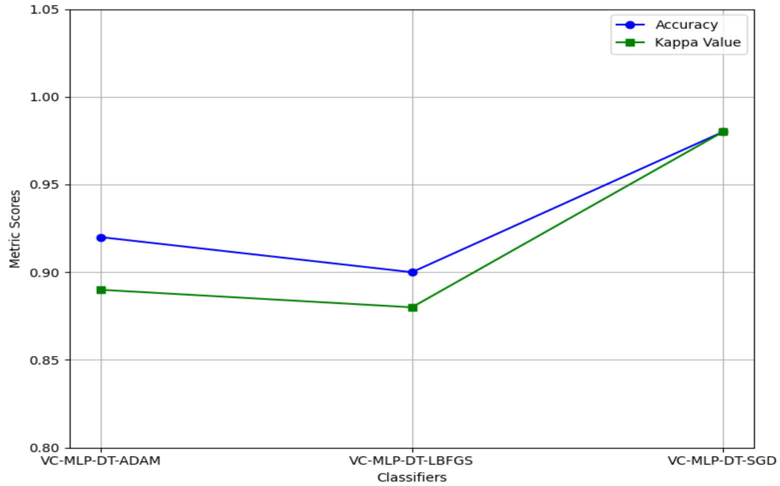


Figure 3: Comparison of Ensemble Models (DT + Neural Optimizers) Using Soft Voting Classifier Based on Accuracy $ Kappa value

Figure 3 displays the comparison of ensemble models using the voting classifier to combine the decision tree with a neural network based on the accuracy score. Here, it is noticed that the optimizer SGD with DT outperformed well than others. Stochastic Gradient Descent (SGD) is more suited for large data sets, by enhancing the learning rate.

Table 3: Comparison of Performance of Ensemble Models using Hard Voting Classifier (MLP + DT+ Optimizers)

Algorithm	Precision	Recall	F1-Score	TPR	FPR	Kappa Value	Accuracy
HARD VOTING CLASSIFIER - VC-MLP-DT-ADAM							
Class 0	0.98	1.00	0.99	0.982	0.000	0.88	0.91
Class 1	0.99	1.00	1.00	0.991	0.000		
Class 2	0.96	1.00	0.98	0.963	0.000		
Class 3	0.74	0.93	0.82	0.738	0.022		
Class 4	0.97	0.61	0.75	0.974	0.084		
HARD VOTING CLASSIFIER - VC-MLP-DT- LBFGS							
Class 0	0.71	1.00	0.83	0.705	0.000	0.88	0.90
Class 1	0.99	0.90	0.94	0.993	0.025		
Class 2	0.99	1.00	1.00	0.994	0.000		
Class 3	1.00	0.73	0.84	1.000	0.072		
Class 4	0.94	0.93	0.93	0.941	0.017		
HARD VOTING CLASSIFIER- VC-MLP-DT-SGD							
Class 0	0.98	1.00	0.99	0.982	0.000	0.98	0.98
Class 1	0.99	1.00	1.00	0.991	0.000		
Class 2	1.00	1.00	1.00	0.997	0.000		
Class 3	1.00	0.92	0.96	1.000	0.024		
Class 4	0.94	1.00	0.97	0.937	0.000		

The comparison of ensemble models using a hard voting classifier is shown in Table 3. Comparing all the metrics of soft and hard voting classifiers, both models perform similarly. The only the way of selecting the prediction method varies. Both soft and hard voting classifiers are ensemble techniques that combine multiple models and improve the accuracy score. However, they vary in aggregating the predictions of the individual classifiers. Hard voting classifiers are most probably used in scenarios where individual models show different strengths and weaknesses, thereby resulting in more accurate predictions and classifications. In the hard voting classifiers, each classifier votes for a class, and the class with the majority votes becomes the final forecast. Each vote is counted first and the highest vote count is selected. If multiple classes are tied with count, another method is used to break the tie.

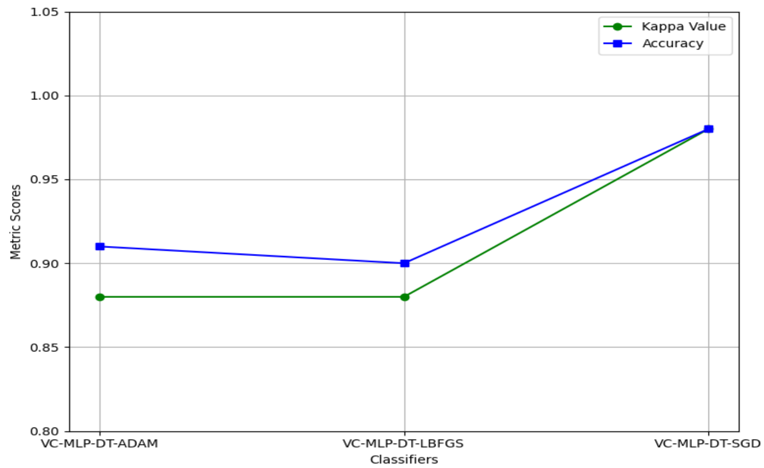


Figure 4: Comparison Of Ensemble Models (DT + Neural Optimizers) Using Hard Voting Classifier Based On Accuracy $ Kappa Value

Figure 4 highlights the superior performance of a specific ensemble model configuration - a hard voting classifier that combines a decision tree with neural networks, using the SGD optimizer. According to Figure 4, this combination is shown to achieve the best results across both the accuracy score and the kappa value. This suggests that the SGD optimizer, when integrated with the decision tree within this ensemble framework, effectively powers the strengths of both components to produce a more accurate and robust model compared to other configurations.

Table 4: Comparison of Performance of Ensemble Models Using Soft Voting Classifier (DT + Base Models)

Algorithm	Precision	Recall	F1-Score	TPR	FPR	AUC	Kappa Value	Accuracy
SOFT VOTING CLASSIFIER- VC-DT-KNN								
Class 0	0.95	1.00	0.97	0.947	0.000	1.00	0.97	0.98
Class 1	0.97	1.00	0.99	0.973	0.000	1.00		
Class 2	0.99	1.00	1.00	0.991	0.000	1.00		
Class 3	1.00	0.90	0.95	1.000	0.029	1.00		
Class 4	0.97	1.00	0.99	0.975	0.000	1.00		
SOFT VOTING CLASSIFIER- VC-DT-SVM								
Class 0	0.98	1.00	0.99	0.982	0.000	1.00	0.98	0.98
Class 1	0.99	1.00	1.00	0.991	0.000	1.00		
Class 2	1.00	1.00	1.00	0.997	0.000	1.00		
Class 3	1.00	0.92	0.96	1.000	0.024	0.98		
Class 4	0.94	1.00	0.97	0.937	0.000	0.99		
SOFT VOTING CLASSIFIER- VC-DT-RF								
Class 0	0.98	1.00	0.99	0.982	0.000	1.00	0.98	0.98
Class 1	0.99	1.00	1.00	0.991	0.000	1.00		
Class 2	1.00	1.00	1.00	0.997	0.000	1.00		
Class 3	1.00	0.92	0.96	1.000	0.024	0.98		
Class 4	0.94	1.00	0.97	0.937	0.000	0.99		

Table 4 shows the performance comparison of ensemble models using Soft Voting Classifier (DT + Base Models). Here the base models used are KNN, SVM, and RF with DT to enhance the decision tree classifier performance. Soft voting classifiers based on SVM and RF outperform KNN in most metrics, especially for Classes 0 and 4. All classifiers have high accuracy and great predictive capabilities, making them ideal for jobs requiring multi-class categorization. Further optimization and validation may improve performance, particularly in Class 3, where KNN demonstrated slightly worse recall. Based on the performance metrics for the soft voting classifiers using different algorithms such as KNN, SVM, and RF, the performance of all metrics is demonstrated in the above table. All three-hybrid soft voting classifiers namely, VC-DT-KNN, VC-DT-SVM, and VC-DT-RF exhibit high performance across the classes, as indicated by their high precision, recall, and F1-scores. The AUC values for all classifiers are perfect (1.00), signifying that the models have excellent discriminatory ability across the classes. The overall accuracy remains high at 0.98 for all classifiers, affirming that the models are performing well in terms of correct classifications. The Kappa values are also high with 0.97 for KNN and 0.98 for SVM and RF, suggesting strong agreement between predicted and observed classifications, reinforcing the reliability of these models.

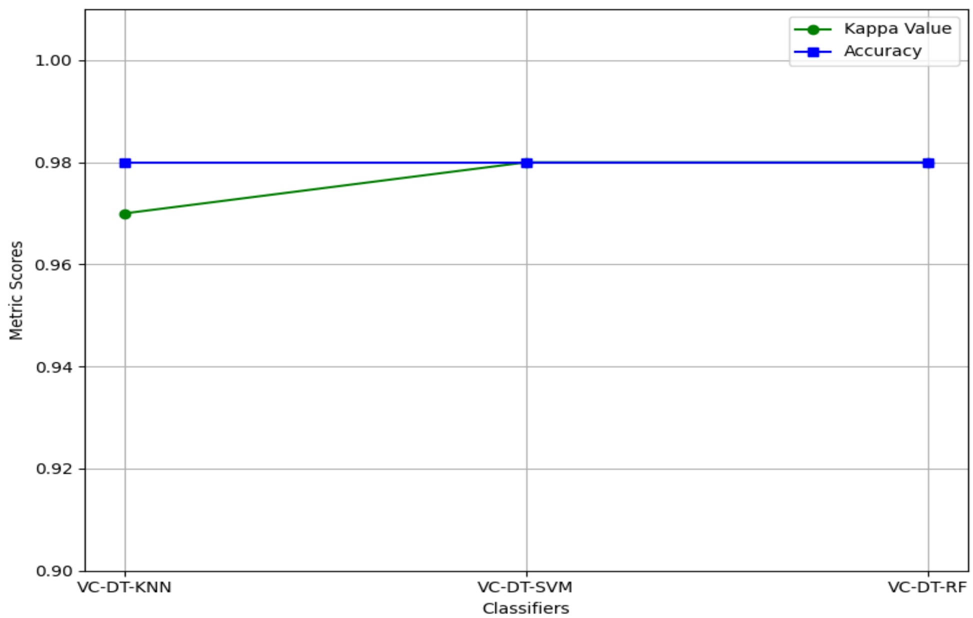


Figure 5: Comparison of Ensemble Models (DT + Base Models) Using Soft Voting Classifier Based on Accuracy & Kappa Value

Figure 5 demonstrate that all three soft voting classifiers achieve outstanding performance, with very high accuracy and kappa values, indicating reliable and robust multiclass classification. Among them, VC-DT-SVM and VC-DT-RF show a slight edge in kappa value, suggesting marginally better consistency in predictions compared to VC-DT-KNN. The performance metrics indicate that these ensembles can handle multiclass classification tasks effectively, maintaining high performance across all classes.

Table 5: Comparison of Performance of Ensemble Models Using Hard Voting Classifier (DT + Base Models)

Algorithm	Precision	Recall	F1-Score	TPR	FPR	Kappa Value	Accuracy
HARD VOTING CLASSIFIER - VC-DT-KNN							
Class 0	0.93	1.00	0.96	0.947	0.000	0.97	0.97
Class 1	0.96	1.00	0.98	0.973	0.000		
Class 2	0.99	1.00	0.99	0.987	0.000		
Class 3	1.00	0.88	0.93	1.000	0.030		
Class 4	0.99	1.00	1.00	0.975	0.000		
HARD VOTING CLASSIFIER - VC-DT-SVM							
Class 0	0.92	1.00	0.96	0.922	0.000	0.69	0.75
Class 1	0.63	1.00	0.78	0.634	0.000		
Class 2	0.62	0.85	0.72	0.617	0.038		
Class 3	0.91	0.52	0.66	0.910	0.123		
Class 4	0.94	0.40	0.56	0.939	0.125		
HARD VOTING CLASSIFIER - VC-DT-RF							
Class 0	0.98	1.00	0.99	0.982	0.000	0.99	0.99
Class 1	0.99	1.00	1.00	0.991	0.000		
Class 2	1.00	1.00	1.00	0.997	0.000		
Class 3	1.00	0.97	0.99	1.000	0.008		
Class 4	1.00	1.00	1.00	1.000	0.000		

The hard voting classifier using Random Forest outperforms both KNN and SVM in most metrics, particularly in handling Class 1 and Class 4 effectively. KNN performs well across all classes but struggles with Class 3's recall. SVM displays inconsistent performance, especially for Classes 1, 2, 3, and 4, which may necessitate further tuning or consideration of alternative algorithms. Overall, RF and KNN are more robust choices for multi-class classification in this context. The Kappa values reflect a strong agreement between predicted and actual classifications, particularly for KNN (0.97) and RF (0.99). SVM's lower Kappa (0.69) suggests less reliability in its predictions. Overall accuracy remains high for KNN (0.97) and RF (0.99), while SVM shows a significantly lower accuracy of 0.75, indicating potential issues with generalization across classes.

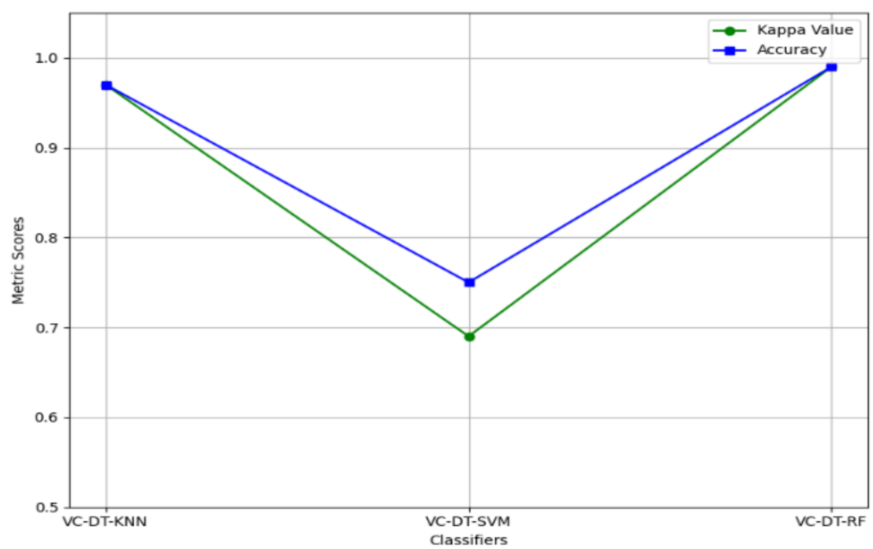


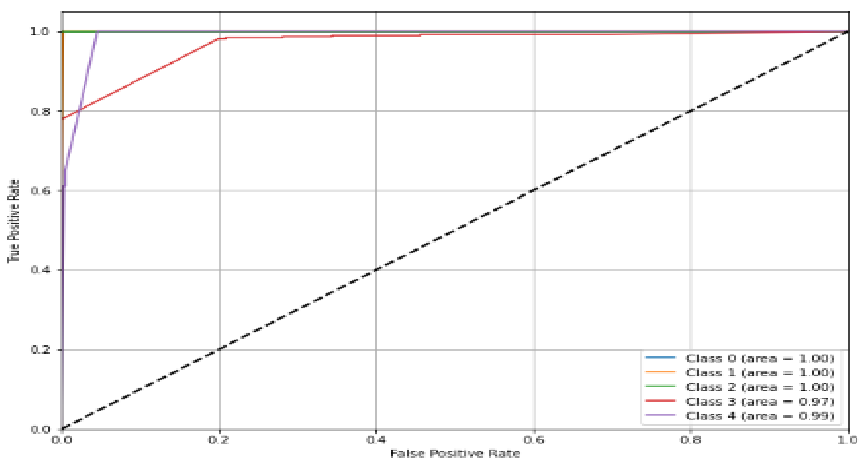
Figure 6: Comparison of Ensemble Models (DT + Base Models) based on Accuracy & Kappa Metric

Figure 6 indicates that among the hard voting ensembles, VC-DT-RF delivers the best performance across both accuracy and kappa value, demonstrating strong reliability for multiclass classification tasks. VC-DT-KNN performs well but slightly below RF, while VC-DT-SVM exhibits significantly weaker agreement and accuracy, suggesting it may not be optimal for this dataset.

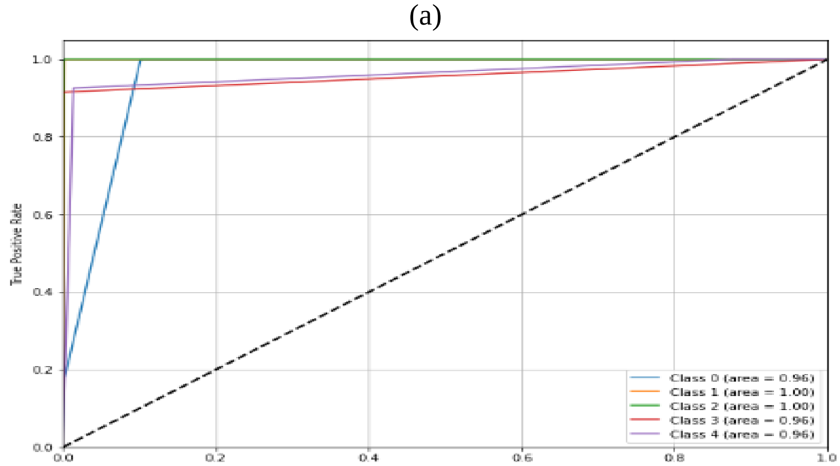
Table 6: Performance Metrics of Ensemble Model: Bagging Classifier + Decision Tree (BC-DT)

Algorithm	Precision	Recall	F1-Score	TPR	FPR	AUC	Kappa Value	Accuracy
Class 0	0.98	1.00	0.99	0.982	0.000	1.00	0.999	0.99
Class 1	0.99	1.00	1.00	0.994	0.000	1.00		
Class 2	1.00	1.00	1.00	1.000	0.000	1.00		
Class 3	1.00	0.97	0.98	1.000	0.009	1.00		
Class 4	0.99	1.00	0.99	0.987	0.000	1.00		

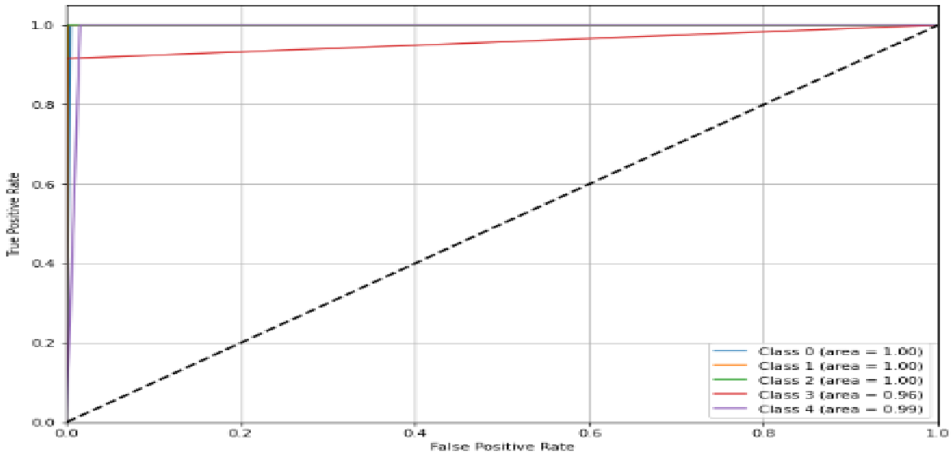
The ensemble method, bagging classifier is used here with the decision tree. In Table 6, the evaluation of the performance metrics of the ensemble model BC-DT (Bagging Classifier + Decision Tree) is displayed. The simple decision tree used for multiclass classification performed with 0.81 score, while the hybrid model of decision tree with bagging classifier outperformed with a score of 0.99. By considering the other metrics such as precision, recall, and F1-score, high measures are compared with the individual DT. In all classes, the TPR rate is approximately near to one indicating that the model can accurately identify the positive instances. Similarly, the FPR rate is low, i.e., the model is good at avoiding false instances.



(a)

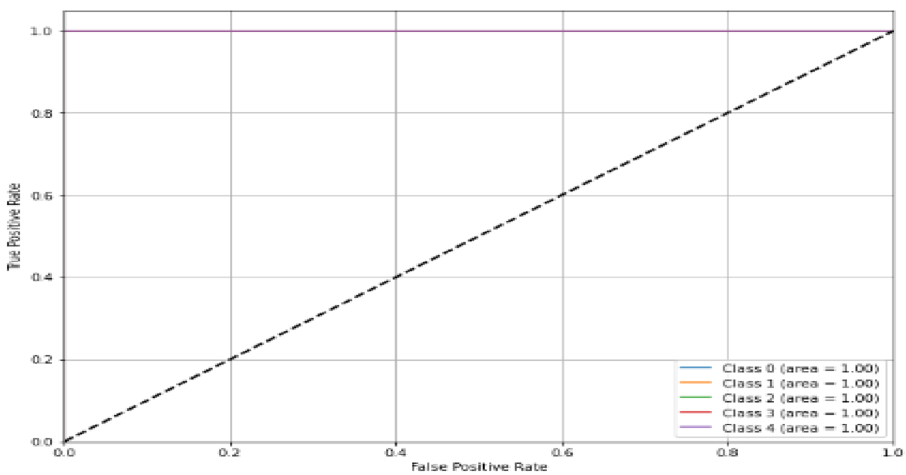


(b)

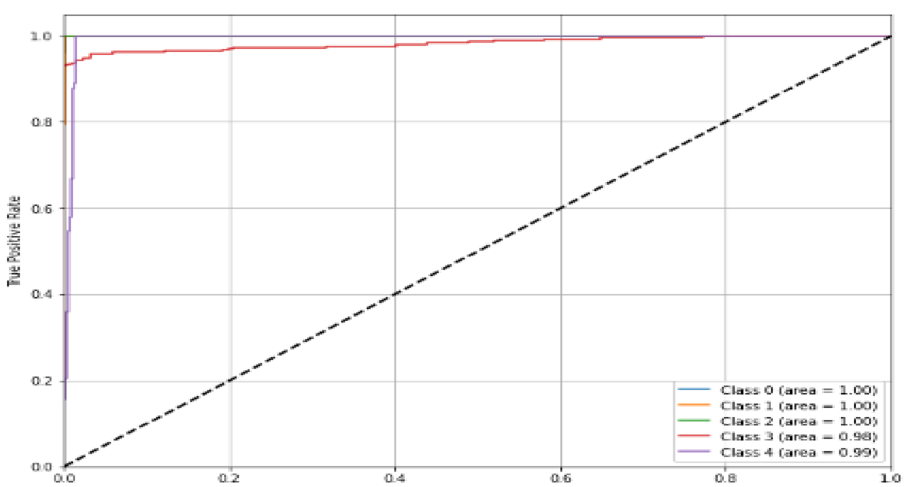


(c)

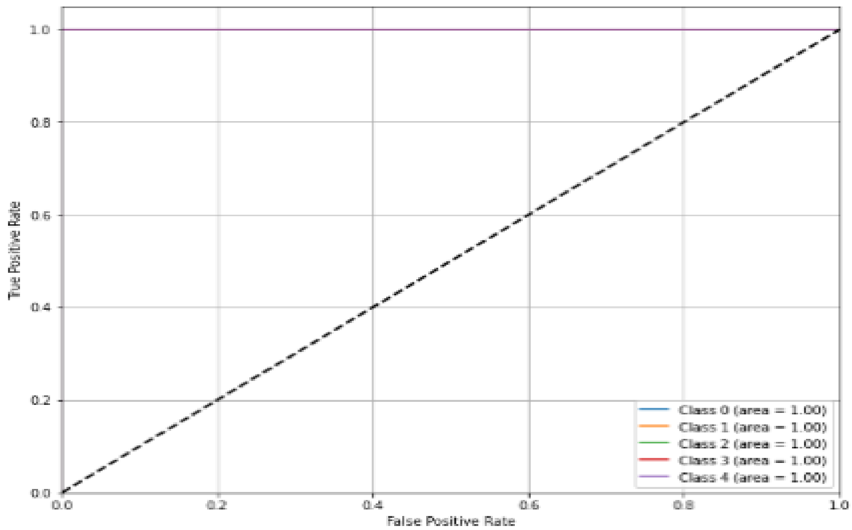
Figure 7: ROC Curve of Ensemble Model Using Voting Classifier (a) Decision Tree with ADAM Optimizer (b) Decision Tree With L-B-FGHS Optimizer (c) Decision Tree With SGD Optimizer



(a)



(b)



(c)

Figure 8: ROC Curve of Ensemble Model using Voting Classifier (a) Decision Tree with KNN (b) Decision Tree with SVM (c) Decision Tree with RF

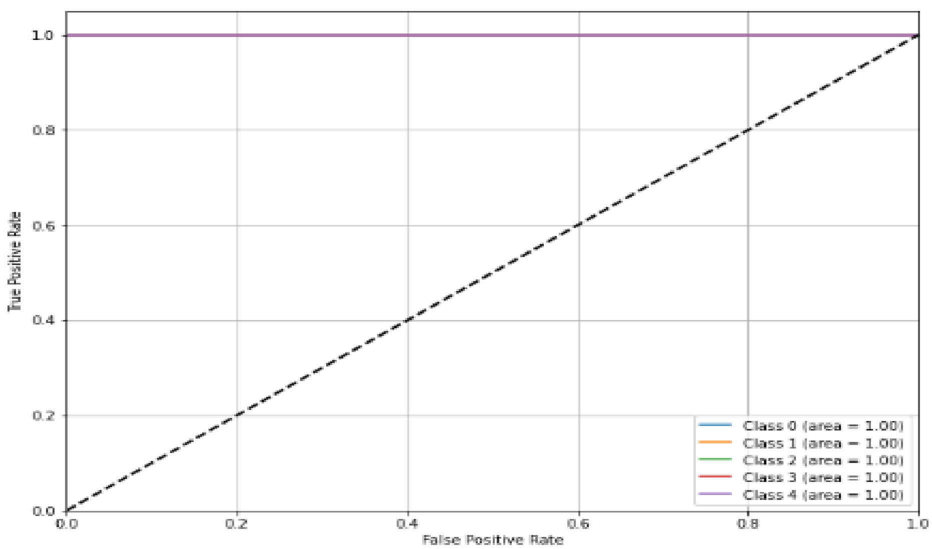


Figure 9: ROC Curve of Ensemble Model Using Bagging Classifier (BC + DT)

From an ROC perspective, in Figures 7, 8, and 9 all the ensemble models exhibit near-perfect curves with high TPR and low FPR across all classes, confirming excellent classification capability than the simple classification techniques.

6 Conclusion and Future Scope

To sum up, the study offers a detailed examination of customer segmentation in the banking domain, aiming to enhance the decision tree classifier's accuracy and other measures using the ensemble learning techniques. The notable element is that while considering the basic classification algorithms KNN, SVM, and RF, their performances are better than the DT classifier. Based on this, the work focused on maximizing the performance metrics of DT by the ensemble method. While integrating the DT classifier with the base models and neural network optimizers' using the voting classifier and combining the DT with the bagging classifier increases the classification outcomes. Additionally, among the proposed models, VC-DT-SGD and VC-DT-RF outperformed. A significant enhancement in classification performance was observed across various metrics while implementing both hard and soft voting classifiers. The Bagging Classifier approach (BC-DT) also provided promising results in improving the DT classifier's robustness and effectiveness. In brief, the hybrid DT models, which are enhanced with neural network optimizers and ensemble techniques, display superiority over the traditional classifiers in the multiclass customer segmentation, presenting better prediction and classification.

The aforementioned changing aspects imply that future studies can focus on additional neural network optimizers and deep learning techniques to explore the impact of ensemble learning method advantages on classification. The future study will also incorporate the additional classifiers that are more advanced, such as gradient boosting and XGBoost and their variants, which provide deeper insights into ensemble learning for improving the DT performance. To summarize future research aims to extend the understanding of machine learning and ensemble learning techniques in customer classification and their potential for more personalized and profitable banking services.

References

[1] Alfaro, E., Gamez, M., & Garcia, N. (2013). Adabag: An R package for classification with boosting and bagging. *Journal of Statistical Software*, *54*, 1-35. https://doi.org/10.18637/jss.v054.i02.

[2] Alqahtani, A. F., & Ilyas, M. (2024). An ensemble-based multi-classification machine learning classifiers approach to detect multiple classes of cyberbullying. *Machine Learning and Knowledge Extraction*, *6*(1), 156-170. https://doi.org/10.3390/make6010009.

[3] Das, T. K. (2015, October). A customer classification prediction model based on machine learning techniques. In *2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)* (pp. 321-326). IEEE. https://doi.org/10.1109/ICATCCT.2015.7456903

[4] Datta, D., Bhattacharya, M., Rajest, S. S., Shynu, T., Regin, R., & Priscila, S. S. (2023). Development of predictive model of diabetic using supervised machine learning classification algorithm of ensemble voting. *International Journal of Bioinformatics Research and Applications*, *19*(3), 151-169. https://doi.org/10.1504/IJBRA.2023.133695.

[5] El-Kenawy, E. S. M., Ibrahim, A., Mirjalili, S., Eid, M. M., & Hussein, S. E. (2020). Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. *IEEE access*, *8*, 179317-179335. https://doi.org/10.1109/ACCESS.2020.3028012.

[6] Gulsoy, N., & Kulluk, S. (2019). A data mining application in credit scoring processes of small and medium enterprises commercial corporate customers. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3), e1299. https://doi.org/10.1002/widm.1299

[7] Patel, S., Shah, C., Patel, P., & Rathod, D. (2022, August). Diagnosis of Denizen Cirrhosis Disorders Using Supervised Machine Learning Classifiers. In *Proceedings of Seventh International Congress on Information and Communication Technology: ICICT 2022, London, Volume 4* (pp. 809-815). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-2397-5_72

[8] Kubina, M., Varmus, M., & Kubinova, I. (2015). Use of big data for competitive advantage of company. *Procedia Economics and Finance*, *26*, 561-565. https://doi.org/10.1016/S2212-5671(15)00955-7

[9] Jindal, S., Sachdeva, M., & Kushwaha, A. K. (2022). Performance evaluation of machine learning based voting classifier system for human activity recognition: 10.48129/kjs. splml. 19189. *Kuwait Journal of Science*. https://doi.org/10.48129/kjs.splml.19189

[10] Kavitha, M. (2025). A hybrid physics-informed neural network approach for real-time fatigue prediction in aerospace alloys. *Advances in Mechanical Engineering and Applications*, *1*(1), 50–58.

[11] Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, *2*, 40-46. https://doi.org/10.1016/j.ijcce.2021.01.001.

[12] Leon, F., Floria, S. A., & Bădică, C. (2017, July). Evaluating the effect of voting methods on ensemble-based classification. In *2017 IEEE international conference on INnovations in intelligent Systems and applications (INISTA)* (pp. 1-6). IEEE. https://doi.org/10.1109/INISTA.2017.8001122

[13] Mungra, D., Agrawal, A., & Thakkar, A. (2020). A voting-based sentiment classification model. In S. Choudhury, R. Mishra, R. Mishra, & A. Kumar (Eds.), *Intelligent communication, control and devices* (Vol. 989, pp. 543–550). Springer. https://doi.org/10.1007/978-981-13-8618-3_57

[14] Naidu, G., Zuva, T., & Sibanda, E. M. (2023). A review of evaluation metrics in machine learning algorithms. In R. Silhavy & P. Silhavy (Eds.), *Artificial Intelligence Applications in Networks and Systems. CSOC 2023. Lecture Notes in Networks and Systems* (Vol. 724). Springer. https://doi.org/10.1007/978-3-031-35314-7_2.

[15] Narang, R. V., & Chatterjee, M. (2025). AI-Powered Knowledge Management Systems: A Hybrid Model for Smart Organizations. *International Academic Journal of Innovative Research*, *12*(3), 14-19. https://doi.org/10.71086/IAJIR/V12I3/IAJIR1220

[16] Noori, B. (2021). Classification of customer reviews using machine learning algorithms. *Applied Artificial Intelligence*, *35*(8), 567-588. https://doi.org/10.1080/08839514.2021.1922843.

[17] Petrovic, T., & Alvarez, R. (2025). *AI-driven content recommendation in learning management systems (LMS): A hybrid filtering approach*. *International Academic Journal of Science and Engineering, 12*(2), 20–24. https://doi.org/10.71086/IAJSE/V12I2/IAJSE1213.

[18] Polikar, R. (2012). Ensemble learning. In *Ensemble machine learning* (pp. 1-34). Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9326-7_1.

[19] Pratama, I. P. A., Atmadji, E. S. J., Purnamasar, D. A., & Faizal, E. (2024). Evaluating the performance of voting classifier in multiclass classification of dry bean varieties. *Indonesian Journal of Data and Science*, *5*(1), 23-29. https://doi.org/10.56705/ijodas.v5i1.124

[20] Santhosh, M., Kavitha, S., Keerthana, R., Suganya, L., & Krishnakumar, S. (2016). Electronic voting machine using internet. *International Journal of Communication and Computer Technologies*, *4*(2), 72-75.

[21] Sehgal, R., & Soni, S. (2024). The Impact of Digital Transformation on Customer Experience: A Study of How Firms Use Digital to Enhance CX. *Indian Journal of Information Sources and Services*, *14*(4), 54–59. https://doi.org/10.51983/ijiss-2024.14.4.09.

[22] Singh, N., & Singh, P. (2019). A novel Bagged Naïve Bayes-Decision Tree approach for multi-class classification problems. *Journal of Intelligent & Fuzzy Systems*, *36*(3), 2261-2271. https://doi.org/10.3233/JIFS-169937

[23] Sun, N., Morris, J. G., Xu, J., Zhu, X., & Xie, M. (2014). iCARE: A framework for big data-based banking customer analytics. *IBM Journal of Research and Development*, *58*(5/6), 4-1. https://doi.org/ 10.1147/jrd.2014.2337118.

[24] Talón-Ballestero, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S., & Rojo-Álvarez, J. L. (2018). Using big data from customer relationship management information systems to determine the client profile in the hotel sector. *Tourism Management*, *68*, 187-197. https://doi.org/10.1016/j.tourman.2018.03.017.

[25] Tripathi, D., Shukla, A. K., Reddy, B. R., Bopche, G. S., & Chandramohan, D. (2022). Credit scoring models using ensemble learning and classification approaches: a comprehensive survey. *Wireless Personal Communications*, *123*(1), 785-812). https://doi.org/10.1007/s11277-021-09158-9

[26] Tripathi, D., Shukla, A. K., Reddy, B. R., Bopche, G. S., & Chandramohan, D. (2022). Credit scoring models using ensemble learning and classification approaches: a comprehensive survey. *Wireless Personal Communications*, *123*(1), 785-812.

[27] Vasuki, P., Sripriya, N., Jacindha, S., Abishek, G., & Supriya Abirami, A. (2025). Neural program synthesis for automatic image enhancement. *Journal of Internet Services and Information Security*, *15*(2), 791–806. https://doi.org/10.58346/JISIS.2025.I2.052.

[28] Vujović, Ž. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, *12*(6), 599-606.

[29] Yang, Z., Xu, Q., Bao, S., Cao, X., & Huang, Q. (2021). Learning with multiclass AUC: Theory and algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(11), 7747-7763. https://doi.org/10.1109/TPAMI.2021.3101125.

[30] Zhang, W., Yang, D., & Zhang, S. (2021). A new hybrid ensemble model with voting-based outlier detection and balanced sampling for credit scoring. *Expert Systems with Applications*, *174*, 114744. https://doi.org/10.1016/j.eswa.2021.114744

Authors Biography

Sufaira Shamsudeen received her Master’s Degree in MCA in the year 2006 from Indira Gandhi National Open University (IGNOU), New Delhi, India. During 2008-2010, she was with the Mahatma Gandhi Off-Campus Centre, Kottappuram, Thrissur as Lecturer in Computer Science. During 2011-2015, she was with the College of Applied Sciences (IHRD), Kodungallur, Thrissur as Lecturer in Computer Science. Currently, she is working as an Assistant Professor in the Department of Computer Applications at MES College Marampally, Aluva, Ernakulum, Kerala, affiliated with Mahatma Gandhi University, for 10 years and continues till this day. She has published and presented several research articles in International Conferences and Scopus-indexed book chapters. Her main research interests include Machine Learning, Deep Learning, and Image Processing.

Dr.K. Ranjith Singh obtained his Ph.D. Degree from Periyar University, Salem, Tamil Nadu, India. He is working as Head of the Department, Department of Computer Technology, Faculty of Arts, Science, Commerce and Management, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, Tamil Nadu, India. He has more than 21 years of experience in the academic field. He has published several research articles in reputed International Journals (Scopus) and International Conferences (Scopus) so far. His research areas of interest include Cryptography and Network Security, Mobile Ad-hoc Networks, and Wireless Sensor Networks.