Artificial Intelligence Evolution
https://ojs.wiserpub.com/index.php/AIE/

UNIVERSAL WISER PUBLISHER

Research Article

Bank Deposit Prediction Using Ensemble Learning

Muhammed J. A. Patwary[1*], S. Akter[1], M. S. Bin Alam[2], A. N. M. Rezaul Karim[1]

[1]Department of Computer Science and Engineering, International Islamic University Chittagong, Chittagong, Bangladesh
[2]Department of Computer Science, Asian University for Women, Chittagong, Bangladesh
Email: jamshed@szu.edu.cn

Received: 19 April 2021; **Revised:** 21 August 2021; **Accepted:** 21 August 2021

Abstract: Bank deposit is one of the vital issues for any financial institution. It is very challenging to predict a customer if he/she can be a depositor by analyzing related information. Some recent reports demonstrate that economic depression and the continuous decline of the economy negatively impact business organizations and banking sectors. Due to such economic depression, banks cannot attract a customer's attention. Thus, marketing is preferred to be a handy tool for the banking sector to draw customers' attention for a term deposit. The purpose of this paper is to study the performance of ensemble learning algorithms which is a novel approach to predict whether a new customer will have a term deposit or not. A Portuguese retail bank data is used for our study, containing 45,211 phone contacts with 16 input attributes and one decision attribute. The data are preprocessed by using the Discretization technique. 40,690 samples are used for training the classifiers, and 4,521 samples are used for testing. In this work, the performance of the three mostly used classification algorithms named Support Vector Machine (SVM), Neural Network (NN), and Naive Bayes (NB) are analyzed. Then the ability of ensemble methods to improve the efficiency of basic classification algorithms is investigated and experimentally demonstrated. Experimental results exhibit that the performance metrics of Neural Network (Bagging) is higher than other ensemble methods. Its accuracy, sensitivity, and specificity are 96.62%, 97.14%, and 99.08%, respectively. Although all input attributes are considered in the classification method, in the end, a descriptive analysis has shown that some input attributes have more importance for this classification. Overall, it is shown that ensemble methods outperformed the traditional algorithms in this domain. We believe our contribution can be used as a depositor prediction system to provide additional support for bank deposit prediction.

***Keywords*:** bank deposit prediction, Naive Bayes, SVM, NN, ensemble methods, bagging, boosting, stacking

1. Introduction

In all economic systems [1] banks have the leading role in planning and implementing financial policies. The clients make deposits in banks to gain from the annual interest offered by the banks. These deposits [2] are used in turn to lend to the disparity of customers for a variety of loans. The profit for the banks is generated from the difference between the two different interest rates. Banks play a significant role in the economy by offering people a service to let them deposit. These loans and business investments are vital for enabling economic growth, which signifies a country's development. Therefore, bank deposits essentially participate in the aggregation of the funding structure of economic

DOI: https://doi.org/10.37256/aie.222021880

growth. Besides, the customer's database contains many features, which immensely influence a depositor's potentiality. But it is not an easy task for a human expert to analyze the customer's database to find a good depositor. If it is possible to figure out the remarkable input attributes, the target can be easily achieved.

Data mining [3-5] is the method of extracting hidden information from a wide variety of datasets. In recent times, data mining is used in different sectors such as banking and finance, where it is used to develop highly efficient classification models. In bank marketing, various data mining algorithms can be used for classifying as well. As we have a shortage of human experts and significant features may have been missed, the automated system can be used as an additional choice to predict potential depositors by analyzing raw data. In this paper, our main objectives are to build the best model for identifying the potential depositor and the relationship between the input attributes and the output attribute. By applying the different classification algorithms of data mining, we will identify the most probable customers who would declare a positive response. This work will also help to find out the significant attributes which successfully affect the output.

The proposed system is developed by using different classification algorithms, namely Naive Bayes (NB), Support Vector Machine (SVM), and Neural Network (NN). Then different Ensemble Methods are applied, such as Bagging, Boosting, and Stacking, to improve the efficiency of basic NB, SVM, and NN algorithms. Implementing ensemble models is a novel technique for bank deposit prediction. In addition, we performed an analysis to find out the dominant input attributes.

This paper is structured and organized as follows: related work is presented in the next section. In section 3, we introduce our methodology and focus on describing our datasets. In section 4, we express our experimental environment. Finally, section 5 concludes the paper and investigates directions for future work.

2. Related works

Different types of algorithms are proposed for the prediction of deposits. The most used methods are Naive Bayes, MLPNN, TAN, Logistic Regression, Decision Tree (C5.0, J48), LADT, RBFN, SVM, and Artificial Immune Network. Elsalamony et al. [6] conducted a comparative study on Multilayer Perception Neural Network (MLPNN) with different classifications, e.g., Tree Augmented NAIVE BAYES (TAN), Logistic Regression (LR), and C5.0, is used for analyzing the results of accuracy, sensitivity, and specificity. MLPNN shows the best percentage for accuracy, which is 90.49%. LR shows the best sensitivity ratio, which is 65.53%, and C5.0 shows the best percentage for specificity, which is 93.23% for the bank term deposit prediction. Wisaeng et al. [7] analyzed the value of the lifetime of customers and neural networks to improve the prediction of bank deposit subscriptions in telemarketing campaigns. Moro et al. [8] assessed and compared the classification performance of four different classification algorithms, i.e., Decision Tree (DT), Neural Network (NN), Logistic Regression (LR), and Support Vector Machine (SVM) on the bank direct marketing dataset to classify the bank deposit prediction. The area of the cumulative lift curve (ALIFT) and Receiver Operating Characteristic Curve (ROC) was used to compare these models. NN provides the best prediction performance, which produced the resulting ROC of 0.80 and ALIFT of 0.67.

Three iterations of the CRoss Industry Standard Process for Data Mining (CRISP-DM) methodology were performed by Moro et al. [9] to achieve the tuned DM model results using the dataset related to the direct marketing campaigns of a Portuguese bank institution. This work discovered that call duration is the most dominant predictor for this problem of classification. This paper shows that the best model, materialized by the neural network (NN), can achieve higher predictive performances.

3. Methods

Machine learning [10-15] techniques can be broadly categorized into two groups, such as supervised learning and unsupervised [16] learning. Supervised learning is the most popular approach, where the class label of each training tuple is known. In this work, the supervised learning approach [17] has been used. This paper has analyzed the effectiveness of four data mining techniques for deposit prediction on direct bank marketing.

3.1 *Support vector machine*

Support Vector Machine (SVM) [18] is used for classification and regression problems and yields excellent results in various domains. The SVM classifier aims to separate the target classes from the other ones with the widest possible margin. Different kernel functions handle different types of data: Linear kernel, Radial Basis Function (RBF), and polynomial kernel. In this work, the RBF kernel has been used with C = 0.5 and gamma = 0.01. Every data element in the n-dimensional feature space is separated using SVM. Then the hyperplane created by SVM separates specific items into their appropriate groups.

3.2 *Neural network*

The Artificial Neural Network [19-21] is a relatively crude electronic model based on biological neurons. It's a collection of nodes that remain connected through a directed link. The Artificial Neural Network (ANN) approach has many benefits over the other classification algorithms. An ANN aims to build a model with an input value that should predict a target value. The building process of the ANN in Waikato Environment for Knowledge Analysis (WEKA) is called the Multilayer perception function. The Back Propagation Algorithm (BPA) trains the Multilayer Perceptron (MLP). It is the most commonly used NN learning technique that consists of the input layer, the hidden layer, and the output layer. In the following steps, the Back Propagation Algorithm can be briefly described:

(1) Randomly initialize all weights and biases.

(2) As long as the error is too large, keep repeating the following steps.

(a) Present a training example and propagate the input vector over the network to get the output. (Forward pass)

(b) Compare the error with the real output to the desired output at each node.

(c) Adjust weight starting from the output layer and working backward. (Backward Pass)

(3) Evaluate performance using the test set.

In this paper, after tuning the parameters finally we use 3 hidden layers consist of 4, 6 and 2 neurons respectively, 1 input layer containing 8 neurons, and 1 output layer with 1 neuron. The learning rate was 0.3.

3.3 *Naive Bayes classifier*

Naive Bayes (NB) [22] is one of the most popular classifiers that is used in the banking sector for classification. Naive Bayes is a classifier that performs probabilistic prediction and predicts probabilities of class membership. Naive Bayes classifier, also known as a statistical classifier, is based on Bayes Theorem with independent assumptions between the predictors.

The Bayes Theorem is:

$$P(h \setminus x) = \frac{P(x \setminus h) * P(h)}{p(x)} \tag{1}$$

Where,

$P(x)$ = Prior probability of x.

$P(h)$ = Prior probability of h.

$P(h \setminus x)$ = Posterior probability of h condition on x.

$P(x \setminus h)$ = Posterior probability of x condition on h.

3.4 *Ensemble methods*

Ensemble methods refer to the powerful learning model that combines multiple algorithms to produce better classification performance. There are three types of Ensemble methods:

Bagging:

Bagging [23] stands for Bootstrap Aggregation. It is one of the most straightforward and most intuitive ensemble-based algorithms [6, 24] that creates separate samples of the training dataset. Each training dataset is used to train a different classification.

Algorithm:

Input:

Data Set $D = \{(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)\}$

Number of iteration T

Process:

Step 1: for i = 1 to T

(a) Through sampling data points with replacement, create a dataset sample Sm.

(b) From each dataset sample, Sm learns a classifier Cm.

Step 2: for every test example.

(a) Try all classifiers Cm.

(b) Estimate the class that earns the largest number of votes.

Boosting:

Boosting [22] is an interactive technique that calculates the output using many different models created by the different algorithms. Using the weighted sum of the weak classifier finds the final result.

Algorithm:

Input:

Data set $D = \{(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)\}$

Number of iteration T

Process:

Step 1: Initialize Weight: Each case receives the same weight.

$Wi = 1/N$, where $i = 1, 2, 3 \ldots N$.

Step 2: Construct a classifier using current weight,

Compute its error:

$$Em = \frac{\sum wi \times I\{Yi \neq gm(xi)\}}{\sum wi} \tag{2}$$

Step 3: Get a classifier influence and update example weight.

$$am = \log\left(\frac{1 - Em}{Em}\right) \tag{3}$$

Step 4: Go to step 2.

AdaBoost [25] was the first boosting technique and is still now widely used in several domains. AdaBoost, in theory, is not prone to overfitting. Stage-wise estimation may slow down the learning process since parameters aren't jointly optimized. AdaBoost may be used to increase the accuracy of the weak classifiers, allowing it to be more flexible. It requires no normalization and has a low generalization error rate. However, training the algorithm takes enormous time. The method is also susceptible to noisy data and outliers. Therefore, removing them before employing them is strongly advised.

Stacking:

Stacking [20] is a method similar to boosting. Stacking is an interesting way of combining different models where multiple different algorithms are applied to the training dataset to create a model. The Meta classifier is used to predict unseen data accurately.

4. Results and discussion

4.1 *Dataset description*

This paper used a bank marketing dataset [26] gathered from a Portuguese retail bank. This dataset was created by P. Rita, P. Cortez, and S. Moro from actual bank data. The dataset contains information regarding a Portuguese banking institution's direct marketing campaigns, and it contains 45,211 phone contacts. Each contact has 16 input attributes and one decision attribute. The types of input attributes are different. Three attributes are binary type, seven attributes are numeric, and six attributes are categorical. The target attribute is binary type. It has two results; the client subscribes to a term deposit or not. The two available datasets are:

(1) bank-full.csv, which contains all the examples regarding the 45,211 contacts, ordered by dates. This particular dataset serves as the training dataset to build the model.

(2) bank.csv contains 10% of the examples (4,521) selected randomly from bank-full.csv to serve the purpose of the test dataset.

4.2 *Data preprocessing*

Data preparation is essential since preprocessing and transformational approaches affect predictive data-mining techniques differently. For obtaining good accuracy, the data must be preprocessed or prepared, and transformed. In this case, we have used oversampling method and the Discretization technique.

Oversampling: The training dataset (bank-full.csv) was imbalanced. The number of "NO" classes was 39,922 and the frequency of the "YES" class was 5,289. To overcome the imbalanced problem we applied oversampling method and increased the frequency of the "YES" class to 42,312.

Discretization: Some Data mining algorithms cannot sufficiently handle numerical attributes but can handle categorical/Nominal attributes and produce the best accuracy. Discretization is a process [27] that transforms a Numerical attribute into a Categorical/Nominal Attribute by dividing the range of attributes into intervals/sub-ranges called buckets or bins. In this dataset, we have split each of these into six bins or Intervals. There are five such attributes in this dataset: "Age", "Duration", "campaign", "p-days", and "previous".

WEKA is a platform for data mining developed by the University of Waikato, New Zealand. We used three data mining classification algorithms in the WEKA tool for bank term deposit prediction.

4.3 *Performance measurement*

A confusion matrix is a visualization tool or contingency table that describes the performance of a classification algorithm. The confusion matrix is made up of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). True positive is the number of correct predictions whether an instance is true, and false positive is the number of wrong predictions of whether an instance is true. The true negative is the number of correct predictions whether an instance is false, and the false negative is the number of wrong predictions of whether an instance is false. Sensitivity, specificity, and accuracy have been used to appraise the result.

Sensitivity: Sensitivity is characterized as the percentage of correctly classified positive, and it is also defined as a True Positive Rate (TPR) and recall.

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

Specificity: Specificity, also known as a True Negative Rate (TNR), is defined as the percentage of correctly classified negative.

Specificity is equivalent to 1 minus False Positive Rate.

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

Accuracy: Accuracy refers to the total number of prediction that was correct.

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

Error-rate: Error-rate refers to the total number of prediction that was incorrect.

$$Error\text{-}rate = \frac{FP + FN}{TP + FN + TN + FP}$$

$$= 1 - Accuracy \tag{7}$$

4.3.1 *Result for classification using Neural Network (NN)*

From the experiment, we find that the neural network's accuracy, sensitivity, specificity, and error rate are 94.87%, 95.52%, 98.23%, and 5.13%, respectively. For neural network (Bagging), accuracy, sensitivity, specificity, and error rate are 96.62%, 97.14%, 99.08%, and 3.38%, respectively, where neural network (AdaBoost) shows accuracy, sensitivity, specificity, and the error rate of 94.87%, 94.21%, 98.28%, and 5.13%, respectively. Figure 1 shows that the accuracy, sensitivity, and specificity of Neural Network (Bagging) are greater than the Neural Network (AdaBoost) and Neural Network.

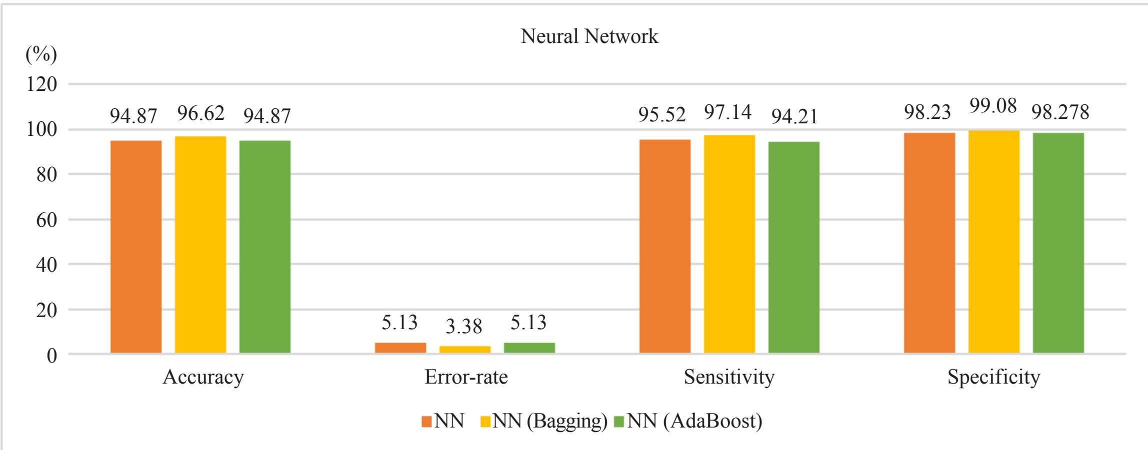

Figure 1. Performance comparison of a Neural Network with Ensemble methods

4.3.2 *Result for classification using Support Vector Machine (SVM)*

We observe that the accuracy, sensitivity, specificity, and error rate of the Support Vector Machine are 89.76%, 85.2%, 96.88%, and 10.24%, respectively. SVM (Bagging) shows accuracy, sensitivity, specificity, and the error rate of 89.80%, 89.84%, 96.95%, and 10.19%, respectively. For SVM (Boosting), accuracy, sensitivity, specificity, and error rate are 90.11%, 90.5%, 97%, and 9.89%, respectively. Figure 2 shows that the accuracy, sensitivity, and specificity of SVM (Boosting) are greater than SVM, and Error-rate is less than the SVM.

4.3.3 *Result for classification using Naive Bayes (NB)*

Figure 3 compares the performance of Naïve Bayes with ensemble methods. The accuracy, sensitivity, specificity, and error rate of the Naive Bayes (NB) are 88.23%, 84.32%, 94.20%, and 11.77%, respectively. For NB (Bagging), accuracy, sensitivity, specificity, and error rate are 88.37%, 87.4%, 94.33%, and 11.63%, respectively. NB (Boosting) has the accuracy, sensitivity, specificity, and error rate of 88.72%, 89.7%, 95.03%, and 11.28%, respectively. Therefore, the overall performance of Naive Bayes (Boosting) is more significant than Naive Bayes.

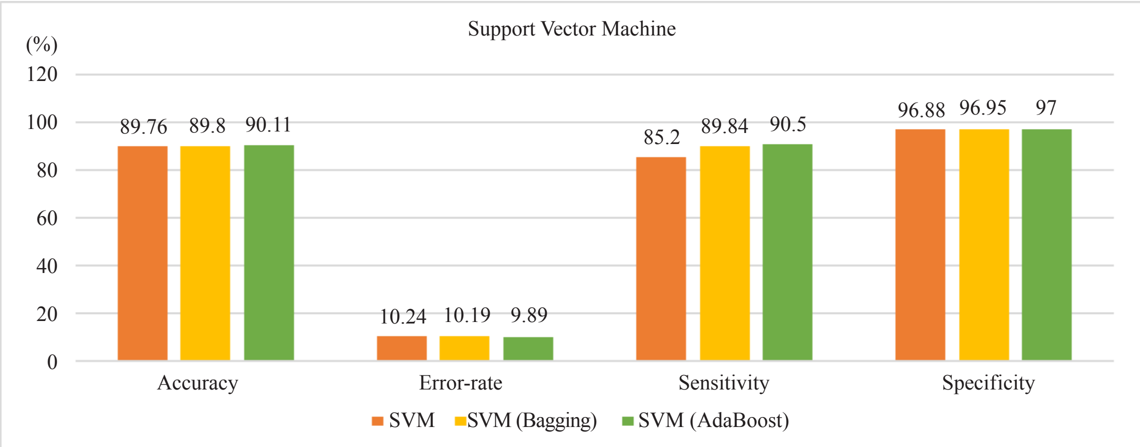

Figure 2. Performance comparison of SVM with Ensemble methods

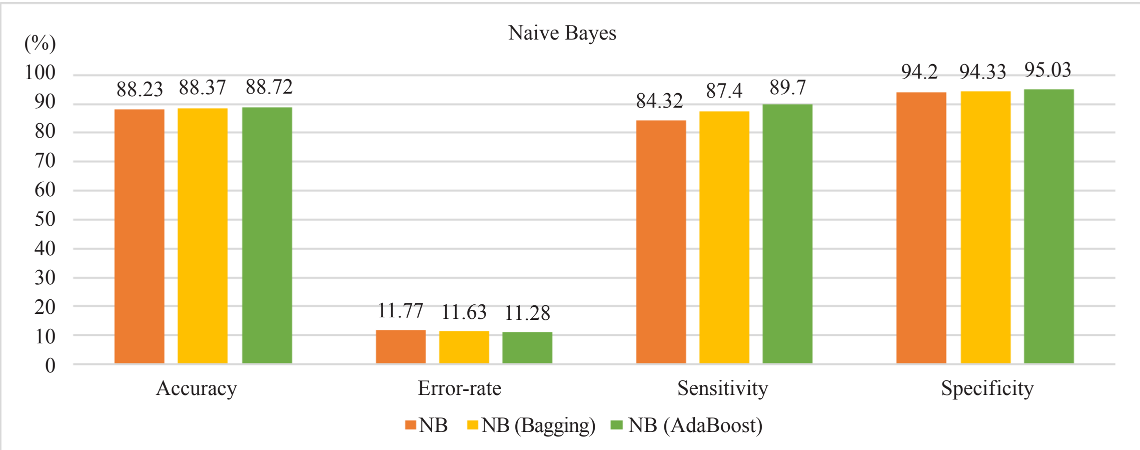

Figure 3. Performance comparison of a Naïve Bayes with Ensemble methods

4.3.4 *Result for classification using the Stacking method*

Figure 4 shows the results for the stacking method. From the figure, it is seen that the accuracy and sensitivity of Neural Network are higher than Stacking, but Stacking has higher specificity than the other three methods.

4.3.5 *Comparison*

The summary of the results is shown in Table 1 where it shows the comparison result of Support Vector Machine, Neural Network, Naive Bayes, and Stacking with Ensemble methods. It is seen from the table that the Bagging and Boosting methods outperformed the conventional classification algorithms.

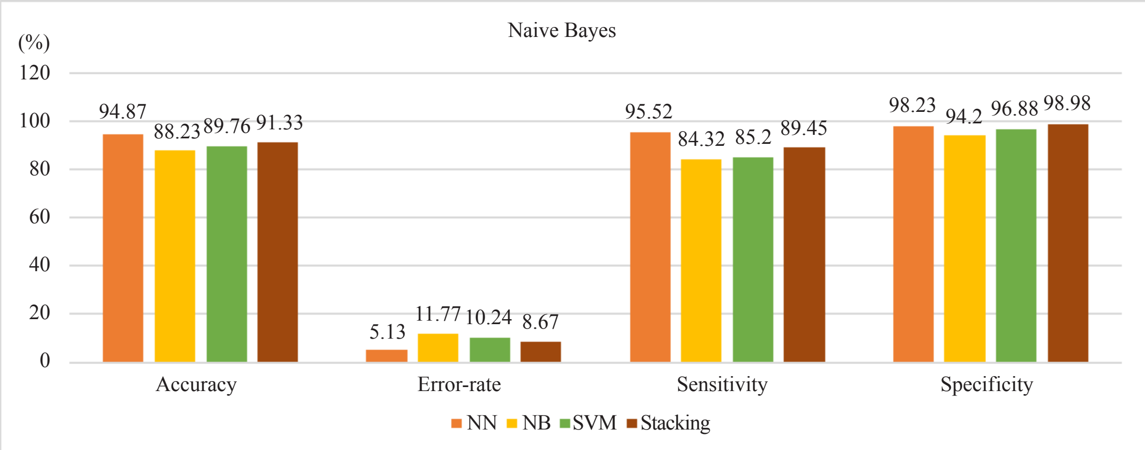

Figure 4. Performance comparison of NN, SVM, and NB with the Stacking method

Table 1. Comparison result of Neural Network, SVM, Naïve Bayes and Stacking with Ensemble methods

		Accuracy (%)	Error-rate (%)	Sensitivity (%)	Specificity (%)
Neural Network	NN	94.8684	5.1316	95.5175	98.225
	NN (Bagging)	96.6158	3.3842	97.14	99.075
	NN (Boosting)	94.8684	5.1316	94.21	98.275
SVM	SVM	89.7589	10.2411	85.20	96.875
	SVM (Bagging)	89.8031	10.1969	89.84	96.95
	SVM (Boosting)	90.1198	9.8872	90.50	97.00
Naive Bayes	NB	88.2327	11.7673	84.32	94.2
	NB (Bagging)	88.3654	11.6346	87.40	94.325
	NB (Boosting)	88.7193	11.2807	89.70	95.025
Stacking	Classifier (SVM, NN) Meta classifier (SVM)	91.3294	8.6706	89.45	98.975

In addition, we conducted a descriptive analysis that demonstrates a relationship between the input variables and the output variable. It shows that some input variables (job, education, marital status, month, loan, balance, and housing) have a notable impact on the output variable and thus perform highly to derive successful output. Table 2 shows the accuracy of the algorithms with and without these input variables.

5. Conclusion

The bank is one of the financial institutions responsible for the economic health of a country. Machine learning techniques can be a hope for the banking domain because of having the potentiality of extracting knowledge from raw data. In this paper, different machine learning classifiers, namely, NN, SVM, and Naive Bayes, have been assessed, which are commonly used to develop the bank term deposit prediction model. After discussing the different classifiers, an attempt was made to increase the efficiency of the SVM, NN, and Naive Bayes algorithm on the Bank marketing dataset, by using different ensemble methods, which is a novel technique in this domain. The decision has been reached that the ensemble methods produce more efficient results than the traditional algorithms in this experiment. In this study, Neural Network (Bagging) shows the best accuracy, sensitivity, and specificity. The other ensemble techniques performed better than SVM, NN, and Naive Bayes algorithms. Moreover, the obtained results conclude that the depositor's performance depends on the input features and some influential features, and it is very effective in finding out potential depositors. Therefore, it will be very significant to use a depositor prediction system as a part of a banking management support system and create a new database containing real depositors.

Table 2. The accuracy of the algorithms with and without significant attributes

		Accuracy (without the significant attributes)	Accuracy (with the significant attributes)
Neural Network	NN	68.42%	94.8684%
	NN (Bagging)	69.51%	96.6158%
	NN (Boosting)	68.89%	94.8684%
SVM	SVM	59.79%	89.7589%
	SVM (Bagging)	60.11%	89.8031%
	SVM (Boosting)	62.00%	90.1198%
Naive Bayes	NB	52.44%	88.2327%
	NB (Bagging)	53%	88.3654%
	NB (Boosting)	53.85%	88.7193%
Stacking	Classifier (SVM, NN) Meta classifier (SVM)	60.90%	91.3294%

References

[1] Hodgson GH. The great crash of 2008 and the reform of economics, Cambridge. *Journal of Economics*. 2009; 33(6): 1205-1221.

[2] Keeley MC. Deposit insurance, risk, and market power in banking. *The American Economic Review*. 1990; 1183-1200.

[3] Uddin MT, Patwary MJ, Ahsan T, Alam MS. Predicting the popularity of online news from content metadata. In *2016 International Conference on Innovations in Science, Engineering and Technology (ICISET)*. 2016. p.1-5.

[4] Cao W, Patwary MJ, Yang P, Wang X, Ming Z. An initial study on the relationship between meta features of dataset and the initialization of nnrw. In *2019 International Joint Conference on Neural Networks (IJCNN)*. 2019. p.1-8.

[5] Patwary MJ, Liu JN, Dai H. Recent advances of statistics in computational intelligence (RASCI). *International Journal of Machine Learning and Cybernetics*. 2018; 9: 1-3.

[6] Elsalamony HA. Bank direct marketing analysis of data mining techniques. *International Journal of Computer Applications*. 2014; 85(7): 12-22.

[7] Wisaeng K. A comparison of different classification techniques for bank direct marketing. *International Journal of Soft Computing and Engineering (IJSCE)*. 2013; 116-119.

[8] Moro S, Cortez P, Rita P. Using customer lifetime value and neural networks to improve the prediction of bank deposit subscription in telemarketing campaigns. *Neural Computing and Applications*. 2015; 26(1): 131-139.

[9] Moro S, Cortez P, Rita P. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*. 2014; 62: 22-31.

[10] Patwary MJ, Wang XZ. Sensitivity analysis on initial classifier accuracy in fuzziness based semi-supervised learning. *Information Sciences.* 2019; 490: 93-112.

[11] Patwary MJ, Wang XZ, Yan D. Impact of fuzziness measureson the performance of semi-supervised learning. *International Journal of Fuzzy Systems.* 2019; 21(5): 1430-1442.

[12] Liu J, Patwary MJ, Sun X, Tao K. An experimental study on symbolic extreme learning machine. *International Journal of Machine Learning and Cybernetics*. 2019; 10(4): 787-797.

[13] Patwary MJ, Akter S, Mahmud T. An expert system to detect uterine cancer under uncertainty. *IOSR Journal of Computer Engineering (IOSR-JCE)*. 2014; 16(5): 36-47.

[14] Hassan M, Alam MS, Ahsan T. Emotion Detection from Text Using Skip-thought Vectors. In *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)*. 2018. p.501-506.

[15] Hossain S, Hasan MZ, Patwary MJ, Uddin MS. An expert system to determine systemic lupus erythematosus under uncertainty. In *Proceedings of International Joint Conference on Advances in Computational Intelligence*. Singapore: Springer; 2021. p.117-130.

[16] Hastie T, Tibshirani R, Friedman J. *Unsupervised Learning*. New York: Springer; 2009.

[17] Kotsiantis SB. Supervised machine learning: A review of classification techniques. *Informatica*. 2007; 31(3): 3-24.

[18] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intelligent Systems*. 1998; 13(4): 18-28.

[19] Hansen LK, Salamon P. Neural network ensemes. *IEEE Transections on Pattern Analysis & Machine Intelligence.* 1990. p.993-1001.

[20] Breiman L. Bagging predictors. *Machine Learning*. 1996; 24: 123-140.

[21] Meir R, Rätsch G. An introduction to boosting and leveraging. *Advanced lectures on machine learning*. Berlin Heidelberg: Springer; 2003. p.118-183.

[22] McCallum A, Nigam K. A comparison of vent models for neive Bayes text classification. *AAAI Technical Report WS-98-05*. 1998; 41-48.

[23] Graczyk M, Lasota T, Trawinski B, Trawiński K. Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal. *Intelligent Information and Database Systems*. Berlin, Heidelberg: Springer; 2010. p.340-350.

[24] Alam MS, Patwary MJ, Hassan M. Birth mode prediction using bagging ensemble classifier: A case study of bangladesh. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. 2021. p.95-99.

[25] Rätsch G, Onoda T, Müller K. Soft Margins for AdaBoost. *Machine Learning*. 2001; 42(3): 287-320.

[26] The UCI Machine Learning Repository is a collection of databases. Available from: http://archive.ics.uci.edu/ml [Accessed 20th July 2021].

[27] Kerber R. Chimerge: Discretization of numeric attributes. *Proceedings of the tenth national conference on Artificial intelligence*. 1992. p.123-128.