Applied Mathematics and Computational Intelligence
Volume 13, No. 3, 2024 [95-114]

Bank Direct Marketing Campaign Success Prediction

Lee Xuan Chian[1], Khai Wah Khaw[2*], XinYing Chew[3], Alhamzah Alnoor[4], Wei Chien Ng[5]

[1,2,5]*School of Management, 11800, Universiti Sains Malaysia, Pulau Pinang, Malaysia.*
[3]*School of Computer Sciences, 11800, Universiti Sains Malaysia, Pulau Pinang, Malaysia.*
[4]*Management Technical College, Southern Technical University, Basrah, Iraq*.

[*] Corresponding author: khaiwah@usm.my

Received: 2 April 2024
Revised: 15 July 2024
Accepted: 29 July 2024

ABSTRACT

Nowadays, there are more new and innovative bank marketing strategies available that contribute to customer engagement and acquisition. Banks must launch marketing campaigns that can successfully attract customers from other competitors. Therefore, the purpose of the research is to apply Machine Learning techniques in predicting the success of a bank's direct marketing campaign using supervised classification algorithms. Six algorithms are implemented which are Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB) and Extreme Gradient Boosting (XGBoost). In this paper, feature selection is performed, and 8 selected features are used to make the prediction. The experimental result shows that the Random Forest model has the highest accuracy and ROC AUC score which makes it the champion model in this paper. Still, these results may vary according to the nature of data preprocessing and algorithms implemented.

Keywords: Bank, Classification, Feature Selection, Machine Learning, Marketing

1 INTRODUCTION

Nowadays, there have been more new and innovative bank marketing strategies available that allure consumers without being intrusive. As individuals are spending more and depositing less due to higher incomes, lower interest rates, and fiscal policy, there is a need for the bank to execute effective marketing strategies to engage customer and acquire new customers. The bank that still sticks to traditional marketing campaigns that looks boring to the consumers will be at a disadvantage because other competitors are launching new marketing strategies that have a higher success rate. The bank needs to adapt to fast-changing technology and adjust the way its brand differentiates and engages customers. In the fast-growing banking sector, there are many options for consumers to switch rather than sticking to one bank only. Competitors may offer higher rates, better rewards, and programs to attract customers. Hence, the bank should take action to encourage existing customers to bring in more business to the bank and away from the competitors and to keep up with the inflation period that reduces deposits due to rising prices. Today's bank marketing campaigns demand data analytics for deep analysis of customer base and market trends, driving important insights into creative ideas. Research by Camilleri [1] highlights how businesses leverage data-driven marketing technologies to enhance interactive engagement with individual prospects, thereby gaining competitive advantages. Hence, effective implementation of a creative marketing strategy is essential for developing campaigns that enhance customer engagement and drive customer acquisition.

Direct marketing is a common marketing strategy implemented in business to communicate with target audiences without the involvement of third parties and intermediaries. It directly communicates to target audiences, presenting information about the company, product, or services to them with the main goal of persuading target audiences to take action. There are several types of direct marketing which include mail marketing, email marketing, mobile marketing, messenger marketing, and web push marketing. In direct marketing, customer segmentation based on demographics, customer information, and interests is critical to send the right messages only to the likeliest prospects. The less targeted campaigns may make the receiver find the direct marketing annoying and intrusive which will leave a negative impression on them and discourage them from buying any product or service from the organization. Successful direct marketing could grow a customer base through customer retention and customer acquisition.

Machine learning is one of the subfields of artificial intelligence that can uncover key insights from data through training algorithms and make a prediction. It is a robust method that is widely used in various fields and businesses to make reliable data-driven decisions with minimal human intervention. The machine learning system can learn from the data by identifying the pattern and finding statistical regularities [2]. Machine learning can deal with enormous volumes of data and even more complex data effortlessly. It quickly builds precise models to analyze and make data-driven recommendations and decisions for an organization to use. The implementation of machine learning allows an organization to work more efficiently and keep the organization ahead of the competition. Machine Learning has contributed much to financial services, government, health care, retail, and transportation. Similar to the banking sector, machine learning plays an important role in detecting illicit transactions [3]. segmenting bank customers [4] and others. In this case, machine learning is implemented to help the bank improve its direct marketing strategy.

The purpose of this paper is to understand the application of Machine Learning in predicting the success of a bank's direct marketing campaign. Six machine learning algorithms are applied to predict the dataset which includes Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), and Extreme Gradient Boosting (XGBoost). The objective of this paper is to predict the willingness of customers to subscribe to term deposits using supervised classification algorithms and to compare the performances of the machine learning algorithms implemented.

2 LITERATURE REVIEW

Referring to research on papers with marketing-related topics, machine learning techniques can provide efficient decision support in this area. [5] design an Early Purchase Prediction framework based on DT, RF, Bagging, KNN, and NB to predict purchase intention for an ongoing session on an e-commerce website. The experimental results proved that the proposed method is useful and performs great in the prediction. The study by [6] involved the implementation of DT and chi-square automatic interaction detection (CHAID) to make a prediction and understand factors influencing repeat patronage intention in the Restaurant Industry. The prediction results shown are valid and reliable. The paper by [7] examines the application of machine learning models in analyzing customer purchasing behavior. DT, KNN, and SVM along with principal component analysis (PCA) are executed in this study, and the prediction results shown are satisfactory with 99% accuracy.

Sharma and Prasann [8] proposed a decision-supportive network to anticipate a customer's purchase intention while they are surfing online. XGBoost is executed and it outperforms GBM and AdaBoost, showing a good F1- a score that proves its effectiveness in distinguishing surfing actions from buying possibilities. The paper written by [9] shows the application of the ensemble learning method (AdaBoost-FSVM) based on a fuzzy support vector machine (FSVM) to predict the purchase intention of consumers. The performance of prediction is compared with LR, SVM, FSVM, RF, and XGBoost, and results show that AdaBoost-FSVM has a more accurate prediction effect.

Chen, Diao and Zhang [10] have shown that optimizing the XGBoost model with the PSO algorithm and 10-fold cross-validation effectively predicts passengers' purchase willingness for airline services. The experimental results show that the proposed method outperforms the other methods (LR, RF, BP Neural Network, NB, DT, SVM, KNN, Long Short-Term Memory, and single-layer XGBoost). The study of [11] compared the predictive capability of the deep learning method with DT, RF, SVM, and Artificial Neural Networks, and the results show that the deep learning method outperforms the other machine learning approach in predicting purchase behavior in e-commerce platforms.

In comparison to other marketing studies, the number of studies on bank telemarketing is modest. Still, several researchers employed different types of machine learning algorithms to make predictions on the success of direct marketing recently. Below are the related works conducted by the other researchers using the same dataset. The study conducted by [12] involved three machine learning methods which are Random Subspace (RS), Multi-Boosting (MB), and Random Subspace-Multi-Boosting (RS-MB). The researchers also did interpretability analysis and selected feature importance that has a significant influence on prediction results. The result from the study shows that RS-MB performs the best among the three methods with selected independent variables.

In their study, Jin and He [13] applied SVM, Neural Network (NN), and DT machine learning models. The researchers applied classification and regression trees (CART) when using DT, Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS algorithm) when the NN model was programming, and sequential minimal optimization (SMO) algorithm when using SVM. In this study, DT has the best performance where its accuracy, precision, recall value, and AUC are all equal to 1.0. Even though NN ranked second out of the three models, it costs more time than SVM which is a great disadvantage.

The next study conducted by [14] explored the usage of Support Vector Data Description (SVDD), which is a one-class classification method inspired by SVM. The researchers make a comparison with results obtained from the other studies conducted by other researchers. It was shown that Lightly Trained Support Vector Data Description (LT-SVDD) has the highest classification accuracy compared to J48-Graft Algorithm, Least Absolute Deviation tree algorithm (LAD tree algorithm), Radial Basis Function Network, SVM, and classical Support Vector Data Description (C-SVDD).

Feng, Yin, Wang and Dhamotharan [15] introduced META-DES-AAP, a dynamic ensemble selection method for predicting the success of bank telemarketing. It is different from the mainstream machine learning methods that focus on prediction accuracy only. The proposed method is systematically evaluated from both machine learning and economic metrics. The researcher compared the proposed method with the other 10 machine learning methods which are RF, DT, KNN, NB, LR, SVM, Adaboost, XGBoost, LightBGM, and GDBT. It was shown that META-DES-AAP has the highest recall value which is 92.62%, standing out from all the 10 methods.

The study by [16] implemented a decision tree in predicting the success of bank direct marketing. They focus on a combination of resampling and feature selection approach which resulted in improved accuracy. Wrapped Subset Eval was used to do feature selection and remove extra features which results in improved accuracy and performance. Results from the study show that the J48 decision tree performs better with an accuracy of 94.39% than the other approaches reported in existing literature such as logistic regression and naïve Bayes.

In their research, Miguéis, Camanho and Borges [17] utilized RF and investigated the effects of class imbalance on prediction accuracy using SMOTE and EasyEnsemble methods. The result from the study shows that the RF technique outperforms LR, SVM, and neural networks in both balanced and imbalanced data when the statistical comparison is made. The performance of RF improved when the prediction was made based on balanced data which is when the EasyEnsemble method is used.

3 METHODOLOGY

Cross Industry Standard Process for Data Mining (CRISP-DM) methodology is applied in this paper. The six sequential steps are business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

3.1 Business Understanding

Deposits are money that is placed in a bank by a depositor for a specific period. Deposits act as one of the income sources of the bank as the bank can take control of the money on hand. The bank offers different benefits for each kind of deposit. For example, banks offer higher interest rates for deposits that are on hold with the bank for a longer period. There is a reason behind banks compensating depositors with a certain interest rate for their deposit funds. The money is considered an asset while the bank still holds the deposit. In other words, the bank may spend the money as long as they still own the asset. For instance, they make use of the money to cover the cost of consumer withdrawals. Besides that, the bank can make money by lending out the deposited funds to those who need money. The money borrower will have to pay back the loan at a higher interest rate to the bank. From here, the bank makes a profit as the interest rate paid by the bank to the depositors is less than the rate charged on the money they lend. The bank is benefiting from the economic environment where interest rates are rising as they can lock in fixed-term deposits. They have the advantage of paying depositors at a lower interest rate while earning profit by charging loan takers a higher interest rate.

It would be better if many people subscribe for deposits because that shows the number of customers and the number of funds that are being accepted. As a result, the bank will likely have more money to lend which will boost their income. If fewer people subscribe to deposits or the number of deposit subscriptions remains the same, this shows that the bank will have fewer funds available for lending. This situation is unfavorable as the bank would need to incur debt to afford the loan's demand and boost interest income. To avoid this situation, the bank has been making an effort to boost the subscription of deposits by running direct marketing campaigns. The bank wishes to retain current customers and gain more new customers. Hence, the data mining goal is to predict the subscription of deposits and gain insights from there to formulate a better marketing campaign.

3.2 Data Understanding

The dataset is obtained from kaggle.com which originates from UCI Machine Learning Repository. This data is related to direct marketing campaigns of a Portuguese Banking Institution provided [18], which data is collected from 2008 to 2013. It contains 11162 instances with 16 input variables and 1 output variable. The input variables include age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous and poutcome whereas the output variable is deposit. Table 1 shows the detailed information about the dataset.

Table 1 : Details of attributes

Attributes	**Data Type**	**Description**
Age	Numerical	Age in years
Job	Categorical	Type of job
Marital	Categorical	Marital status
Education	Categorical	Education background
Default	Categorical	Has credit in default?
Balance	Categorical	Balance of the individual
Housing	Categorical	Has a housing loan?
Loan	Categorical	Has a personal loan?
Contact	Categorical	Contact communication type
Day	Categorical	Last contact day of the week
Month	Categorical	Last contact month of the year
Duration	Numerical	Last contact duration
Campaign	Numerical	Number of contacts performed during this campaign for this customer
Pdays	Numerical	Number of days passed after the customer was last contacted from a previous campaign
Previous	Numerical	Number of contacts performed before this campaign for this customer
Poutcome	Categorical	The outcome of the previous marketing campaign
Deposit	Categorical	Has the customer subscribed to a term deposit?

3.3 Data Preparation

The data preparation as known as the data pre-processing step prepares the dataset for modeling purposes. Data cleaning, data transformation, and further data exploration will be performed in this stage. In this paper, a sequence of steps will be involved to prepare the data. The first step is the data cleaning step whereby inaccurate, duplicated, or irrelevant data will be fixed or removed. The second step is handling outliers. In this step, outliers that could distort statistical results will be handled either through the Univariate method, Multivariate method or Minkowski error method. The third step is One Hot Encoding which encodes the nominal categorical variables by creating dummy variables. This step is important to ease up machine learning models in prediction. The fourth step is feature selection whereby the irrelevant input variable is removed to increase the prediction power of machine learning models. There are three types of feature selection which include Wrapper methods, Filter methods, and Embedded methods. The last step is feature scaling to scale the attributes and normalize the range of the data.

3.4 Modeling

This section discusses the implementation of machine learning algorithms on the Portuguese bank dataset for predicting the success of a bank's direct marketing campaign. Six machine learning algorithms were applied to the processed data to predict the customer's term deposit subscription. By leveraging Python programming language and open-source Python libraries such as NumPy, Pandas and Scikit-learn, six machine learning models were implemented and utilized.

3.4.1 Logistic Regression

Logistic regression is one of the commonly used supervised classification algorithms that is used to predict the discrete variable. It is based on the concept of probability whereby it estimates the likelihood of an event taking place rather than directly predicting whether the specific event will occur [19]. This algorithm requires the dependent variable to be categorical, which is dichotomous. Therefore, the output of this algorithm will always be 0 and 1. It is not suitable for making predictions on a continuous outcome. This algorithm uses a sigmoid function which is a curve graph instead of a linear graph to make the prediction. It is commonly used in the medical field [20][21] and for fraud detection [22].

3.4.2 Support Vector Machine

SVM is one of the most influential supervised learning algorithms that is widely used for binary classification [23]. SVM works by finding the best decision boundary in categorizing n-dimensional space and separating the data into two classes [24]. This is supported by [25] who mentioned that it can minimize generalization error by constructing decision boundaries explicitly which is why SVM is often used as a binary classification. SVM is commonly implemented in image classification and text classification [26].

3.4.3 Naïve Bayes

Naïve Bayes is a probabilistic algorithm based on the Bayes theorem. This algorithm is widely used in various applications especially sentiment analysis and spam filtration [27]. Naïve Bayes assumes that the features that are used in the algorithm are independent of each other [28]. The assumption is when the value of one feature changes, the other features will not be influenced by it. Besides, it also assumes each of the features contributes equally to the outcome. There are three types of Naïve Bayes classifiers which include Bernoulli Naïve Bayes, Gaussian Naïve Bayes, and Multinomial Naïve Bayes. Naïve Bayes is widely used as it requires short computational time to deal with large datasets; however, its naivety is irrelevant to the real-world situation which may result in poor performance [29].

3.4.4 Decision Tree

The decision tree is a supervised learning algorithm that can be applied to classification and regression problems [28]. It models the decision logic into a tree-like structure by dividing a dataset into tiny sections [30]. The tree nodes represent the data attributes whereas the tree leaves represent a value that the node can assume [31]. [28] mentioned that the selection of the root node is important as the outcomes of the algorithm will highly depend on it.

3.4.5 Random Forest

Random forest is one of the most popular machine-learning algorithms [32] under ensemble methods that can be used in classification or regression tasks. It works by developing a great number of decision trees during the training of the model and combining the results of each model and finally making a prediction based on the majority votes [33]. The main advantage of random forests over decision trees is that they can reduce overfitting and bias to achieve greater accuracy. This is supported by [34], who mentioned that random forest has higher evaluation efficiency, accurate dimensionality reduction as well as better accuracy. It is particularly useful in dimensional settings and in the presence of nonadditive predictor-response relationships [35].

3.4.6 Extreme Gradient Boosting

XGBoost as known as eXtreme Gradient Boosting is one of the boosting algorithms which is an extension of the gradient boosting framework associated with two key improvements [36]. It ensembles weak prediction models, typically decision trees. The trees are built in parallel which is different from gradient-boosting decision trees. With this, the speed problem of GBM can be overcome [19]. It is one of the best-performing algorithms which adds regularization to prevent overfitting by penalizing the complexity of the model [37][38].

3.5 Evaluation

Various ways to measure the performance and effectiveness of machine learning are available. The common metric implemented to evaluate the models includes classification metrics whereby there are four types of possible outcomes.

i. True Positive (TP): A class is predicted yes deposit and is yes in reality
ii. True Negative (TN): A class is predicted no deposit and is no in reality
iii. False Positive (FP): A class is predicted yes deposit but is no in reality

iv. False Negative (FN): A class is predicted no deposit but is yes in reality

Below are the metrics that are used to evaluate the selected machine learning models.

i. Accuracy: Accuracy is a simple performance metric that can measure how accurate the model is. It is defined as the ratio of accurately predicted observations over the total observations. It is calculated as the sum of TP and TN divided by the total number of instances.

$$\frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative}$$

ii. Precision: Precision shows how many times the positive prediction is positive. It can be calculated as TP divided by the total number of all positive observations.

$$\frac{True\ Positive}{True\ Positive + False\ Positive}$$

iii. Recall: Recall is the ratio of correctly predicted positive observations out of all positive observations.

$$\frac{True\ Positive}{True\ Positive + False\ Negative}$$

iv. F1 score: F1 score find the harmonic mean of precision and recall. The highest possible value is 1 which indicates perfect precision and recall.

$$2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3.5.1 ROC and AUC

The AUC-ROC curve acts as an important evaluation tool for the performance of machine learning models. ROC can show how all classification thresholds influence the performance of classification models by plotting True Positive Rate (TPR) against False Positive Rate (FPR) in a probability curve whereas ROC AUC displays the capability of a model in distinguishing between classes. A high ROC AUC means the model is good at predicting the binary classes. When ROC AUC is equal to 1, this shows that the model can distinguish all the positive and negative classes correctly. On the contrary, when ROC AUC is equal to 0, this means that the model fails to distinguish the positive and negative classes correctly. When ROC AUC is equal to 0.5, this means that the model cannot separate the class.

The following shows the 4 types of rates, which are the measure factor of the confusion matrix.

i. True Positive Rate (TPR) as known as Recall is the proportion of positive observations that is correctly predicted.

$$\frac{True\ Positive}{True\ Positive + False\ Negative}$$

ii. False Positive Rate (FPR) is the proportion of negative observations that are incorrectly predicted as known as fall out rate.

$$\frac{False\ Positive}{True\ Negative + False\ Positive}$$

iii. True Negative Rate (TNR) as known as specificity is the proportion of negative class that is correctly predicted.

$$\frac{True\ Negative}{True\ Negative + False\ Positive}$$

iv. False Negative Rate (FNR) is the proportion of positive class that is incorrectly predicted.

$$\frac{False\ Negative}{True\ Positive + False\ Positive}$$

4 RESULTS AND DISCUSSION

4.1 Results from pre-processing steps

This section first displayed the results after the data pre-processing stage, which included removing irrelevant data and outliers, encoding variables, interpreting univariate and bivariate analysis, and performing feature selection.

4.1.1 Data cleaning, removal of outliers and One Hot Encoding

Firstly, the dataset was checked to see if it was balanced or not. Figure 1 shows that the target variable "deposit" was balanced, hence no further action was taken. The output also shows that there were no null value and duplicated rows. Secondly, the "duration" feature was removed as it was irrelevant to have realistic predictive models. Thirdly, Figure 2 reveals that there were outliers in "age", "balance", "duration", "campaign", "pdays" and "previous. The outliers were dealt with using the Z-score method. Detected outliers were removed and 1064 outliers were successfully removed. Next, the attribute columns which include "default", "housing", "loan" and "deposit" that contain "yes" and "no" were converted to Boolean variables whereby 1 represents "yes" and 0 represents "no". Categorical columns including "job", "marital", "education", "contact", "month", "poutcome" were then converted to dummies as well through One Hot Encoding method.

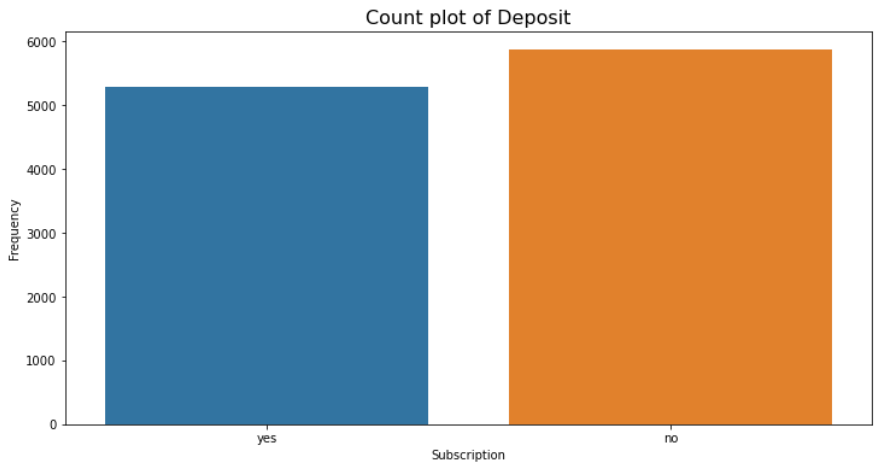

Figure 1: Count plot of target variable

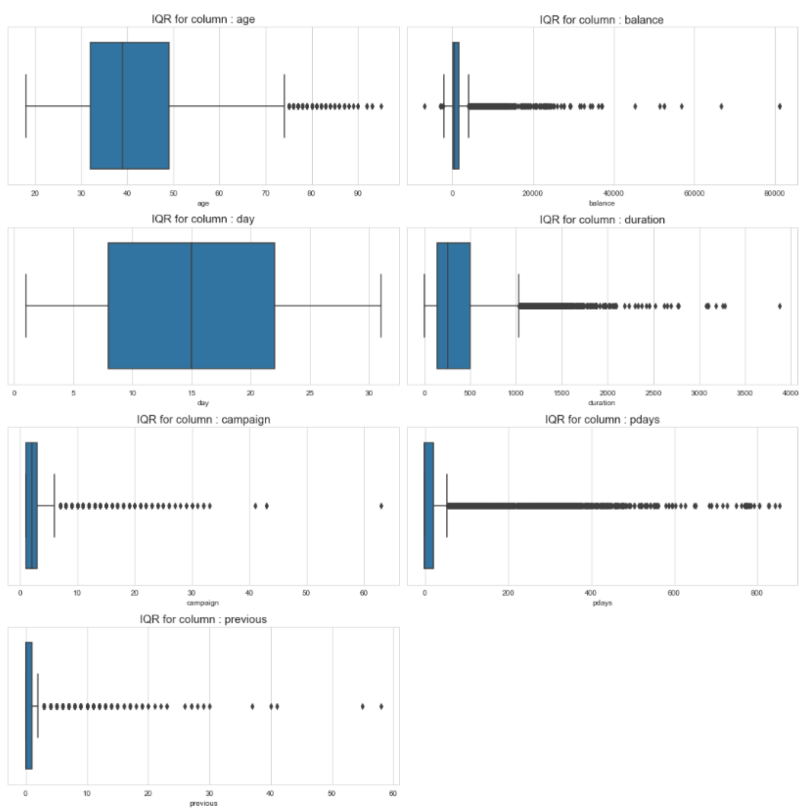

Figure 2: Result of box plots

4.1.2 Univariate and bivariate analysis

Univariate and bivariate analysis was performed to find the relationship between that single feature as well as with the target variable. It was shown that the top three professions of bank customers were working adults from a management background, blue-collar, and technicians. However, they were also the highest for not subscribing to deposits based on the bivariate analysis. Secondly, most customers were married, and the majority of the customers had a secondary education background. The same happened here where many customers who were married and from secondary backgrounds did not subscribe to deposits. Thirdly, the result showed that the majority of customers did not default on credit, and did not have housing loans, and personal loans. Customers that did not default on credit and personal loans were also the main group of customers that did not subscribe to deposits. However, the count plot of housing by deposit showed a different situation whereby most customers that subscribed to term deposit were from the category of customers that did not own housing loan. Next, the results showed that the cellular phone was the most frequently used form of contact to communicate with customers. Similarly, most customers that did not subscribe to term deposits were contacted via cellular phone. Lastly, the results showed that banks contacted customers more frequently from May to August, which was during summer. Most customers that did not subscribe to term deposits were those that were contacted in May.

4.1.3 Feature Selection

Feature selection was performed to select the most important variables and irrelevant features were removed from the dataset before modeling to improve the accuracy. In these steps, a tree-based algorithm was implemented to remove features below a certain threshold and only pick the best features. Feature selection through the Random Forest method was categorized as an embedded method where it covered the advantage of both filter and wrapper methods. After the feature selection was completed, eight features were selected which include “age”, “balance”, “day”, “campaign”, “pdays”, “housing_bool”, “contact_unknown”, and “poutcome_success”. The features “housing_bool”, “contact_unknown”, and “poutcome_success” were generated through one hot encoding of categorical variables, including both binary outcomes and categories with more than two options.

4.1.4 Feature Scaling

Feature scaling was performed to normalize the range of the dataset’s features. The step involved using ‘StandardScaler’ by fitting it to features in ‘X’ and the resulting scaled features were stored in a dataframe ‘X_sc”. This process was crucial for optimizing model performance and reducing the risk of feature bias.

4.2 Comparative performance of the machine learning models

The performance of the machine learning models was the most important as this paper aimed to predict and improve the success of bank direct marketing. This section compared the performance of six machine learning models in terms of confusion matrix, accuracy, ROC AUC score, and based on classification reports

4.2.1 Comparison of Confusion Matrix

A comparison of the machine learning model's performance was made based on the confusion matrix. Confusion matrix outputs included true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The Portuguese Bank dataset, initially containing 11,162 examples, was detected with outliers. After handling outliers, the dataset was reduced to 10,098 examples and divided into a 75:25 ratio, with 7,573 examples used for training and 2,525 for testing. The comparison was facilitated by referring to the confusion matrix output shown in Table 2. By comparing the RF model and the SVM model, the results showed 1,125 true negatives and 716 true positives for the RF model, whereas the SVM model had 1,017 true negatives and 465 true positives. In total, the RF model made 1,841 correct predictions, whereas the SVM model made 1,482 correct predictions. Additionally, the results indicated that the RF model had 418 false negatives and 266 false positives, while the SVM model had 669 false negatives and 374 false positives. This resulted in a total of 684 incorrect predictions for the RF model and 1,043 for the SVM model. The SVM model had 359 more incorrect predictions compared to the RF model.

Table 2: Confusion Matrix of six machine learning models

		Predicted	
Models	**Actual**	**No**	**Yes**
Logistic Regression	**No**	1103	288
	Yes	517	617
Support Vector Machine	**No**	1017	374
	Yes	669	465
Decision Tree	**No**	919	472
	Yes	452	682
Random Forest	**No**	1125	266
	Yes	418	716
Gaussian Naïve Bayes	**No**	1207	184
	Yes	635	499
XGBoost	**No**	1117	274
	Yes	432	702

4.2.2 Comparison of Accuracy and ROC and AUC score

Table 3 presented prediction results using accuracy and ROC AUC score as the evaluation metric. Models were ranked by accuracy, with the RF model leading (73%), followed by the XGBoost model (72%), GNB model (68%), LR model (68%), DT model (63%) and SVM model (59%). Furthermore, the ROC AUC score of the RF model was 0.73, followed by the XGBoost model at 0.72, the GNB model with a score of 0.69, the LR model with a score of 0.68, the DT model with a score of 0.63 and the SVM model with a score of 0.58. The ROC AUC score evaluated the discriminative ability of models, capturing their ability to distinguish between classes, while accuracy measured the overall correctness of predictions. Figure 3 highlighted the RF and XGBoost models as top performers with accuracies of 73% and 72%, respectively, followed closely by GNB and LR models at 68% accuracy. The DT model followed with 63%, while the SVM model showed the lowest accuracy at 59%. The RF model and the XGBoost model also exhibited excellent discrimination capabilities, as reflected in their high ROC AUC scores. The ability to discriminate for the GNB model, LR model, and DT model was considered acceptable with an ROC AUC score of more than 0.60, whereas the SVM model exhibited poor discrimination ability with an ROC AUC score of less than 0.60.

Table 3: Accuracy and ROC AUC scores of six machine learning models

Models	**Accuracy**	**ROC AUC score**
Random Forest	73%	0.73
XGBoost	72%	0.72
Gaussian Naïve Bayes	68%	0.69
Logistic Regression	68%	0.68
Decision Tree	63%	0.63
SVM	59%	0.58

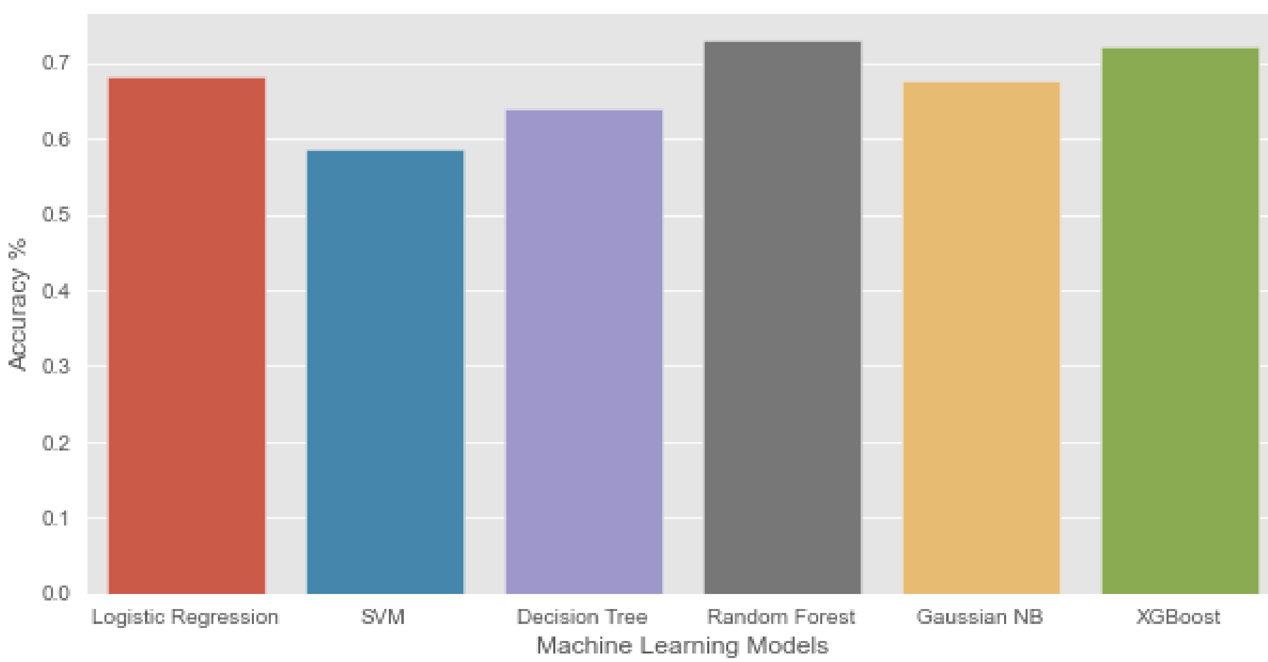

Figure 3: Bar charts of accuracy results

4.2.3 Comparison of Classification Report

Next, a comparison through the classification report of the six machine learning models was made to help evaluate the quality and efficiency of the predictions of the models. The metrics that could be visualized through the classification report included precision, recall, F1 score, and support. Referring to the classification report exhibited by Table 4, class 0 referred to the customer that did not subscribe to a term deposit whereas class 1 referred to the customer that subscribed to a term deposit. The support figures in the report indicated balanced data, with the number of class 0 being approximately the same as the number of class 1. The RF model exhibited a 0.73 precision score for both classes, indicating that 73% predictions were correctly classified for both classes. Compared with the SVM model which had the lowest accuracy, the precision for the SVM model was 0.60 for class 0 and 0.55 for class 1. This showed that 60% of them were of class 0 whereas 55% of them truly were of class 1. Next, the RF model achieved a recall value of 0.81 for class 0 and 0.63 for class 1 which was higher than the recall value of the SVM model, 0.73 for class and 0.41 for class 1. This indicated that the RF model had a greater ability to find all relevant data points by having a few false negatives only. F1-score is considered great if it is equal to 1. In this case, the RF model had an F1-score of 0.77 for class 0 and 0.68 for class 1. This F1 score was higher than the F1-score of the SVM model which was 0.66 for class 1 and 0.47 for class 0.

Table 4: Classification report of six machine learning models

Models	**Precision**	**Recall**	**F1-score**
Logistic Regression			
0	0.68	0.79	0.73
1	0.68	0.54	0.61
Support Vector Machine			
0	0.60	0.73	0.66
1	0.55	0.41	0.47
Decision Tree			
0	0.67	0.66	0.67
1	0.59	0.60	0.60
Random Forest			
0	0.73	0.81	0.77
1	0.73	0.63	0.68
Gaussian Naïve Bayes			
0	0.66	0.87	0.75
1	0.73	0.44	0.55
XGBoost			
0	0.72	0.80	0.76
1	0.72	0.62	0.67

5 DISCUSSION

From this paper, it was found that among the six machine learning models, the RF model has the highest accuracy value. The RF model can predict the subscription of deposits to determine the success of bank direct marketing with an accuracy of 73%. XGBoost, GNB, LR, DT, and SVM achieved accuracy scores of 72%, 68%, 68%, 63%, and 59% respectively. The SVM model is indeed the model that has the weakest performance. The RF model has an accuracy of 73% which indicates that 1843 correct predictions were made out of a total of 2525 examples. On the other hand, the model that gives the least accuracy is the SVM model with an accuracy of 56% only. This model only made 1414 correct predictions out of 2525 examples. Besides that, the RF model has excellent discrimination capabilities with an ROC AUC score of 0.73, which defeats the other models.

Next, based on the results from the comparison of the confusion matrix between the RF model and the SVM model, it shows that there is a huge gap difference between these two models in making the correct prediction. The RF model has the highest number of correct predictions and the lowest number of incorrect predictions on the testing set compared to the other five models. On the contrary, the SVM model has the highest number of incorrect predictions and the lowest number of correct predictions. This indeed validates the results of the accuracy score for the RF model and the SVM model. The RF model that has high accuracy produces fewer incorrect predictions whereas the SVM model that has low accuracy produces more incorrect predictions.

The classification report shows that the RF model has a higher value of precision and recall for both class 0 and class 1, which results in a higher F1 score compared to the SVM model. Overall, the accuracy and the results from the classification report clearly show that the RF model is the model that performs the best. The comparison made between the model that has the best performance and the model that has the weakest performance clearly proved that the RF model outperforms the other models. The RF model has proven the ability to correctly predict as many positive examples as possible which is important for the bank. Hence, the RF model is chosen as the champion model for the bank direct marketing campaign prediction. This finding is similar to the study conducted by [17] who found that the RF model outperforms the other machine learning models. The researcher especially suggested that the RF model should be considered as an alternative to replace traditional algorithms like the LR model.

Regardless of the findings, this study however is subject to several limitations. The potential limitations include the absence of studies that are related success prediction of bank direct marketing. There are few research studies available to study and gain a better understanding from the previous studies. The current studies implemented their own developed new classification approach and explored a dynamic ensemble framework in their study rather than the commonly used classification algorithms. As a result, this limits this paper from being comparable with studies conducted by other researchers as the method implemented is incomparable because those are of a more advanced level.

Furthermore, there are outliers detected. The approach taken to handle the outliers is through the Z-Score method. Outliers were removed and hence reduce the examples for prediction purposes. Still, there are several approaches to handling outliers and the selection of approaches depends on the causes of outliers. Different approaches work with different causes of outliers. There is a certain time when keeping outliers is important as they contain valuable information that is part of the paper. Hence, a comparison of results with and without outliers is recommended to evaluate the need of removing outliers.

Besides that, this paper has implemented feature selection to filter relevant data for prediction and remove irrelevant features to increase the prediction accuracy of the models. There are several types of filter selection methods available. The random forest method is implemented for selecting the relevant features. The number of features is reduced from 16 features to 8 features to build models. It is a concern that a dramatic reduction in features may affect the validity of the performance of models adversely. In future work, it is wiser to make a comparison of results before and after feature selection and proceed only if the accuracy can be improved.

Lastly, another limitation is the moderate performance achieved by all six models. The highest accuracy achieved is only 73% which is quite low and unsatisfactory. This limitation can be overcome through the implementation of a hyperparameter tuning method by choosing a set of optimal hyperparameters for the machine learning algorithms chosen. With this, the performance of models can be maximized, resulting in fewer errors and higher accuracy. Besides, more ensemble methods can be utilized to make comparisons and find the models that have better prediction performance.

After understanding the relationship in the dataset, two recommendations to the bank are suggested so that the bank can take them into consideration while making changes to enhance their direct marketing campaigns. Firstly, it is important to identify a target market before implementing a marketing strategy. Knowing who the potential customers are and gaining a better understanding of them will help to create marketing campaigns that look appealing to the target audience. A bank could direct its attention to retired people and students as they are likely to be the potential customers that will subscribe for a term deposit if the results are looked at proportionally. This is reasonable as a term deposit is an ideal low-risk investment that is preferable by middle-aged people as well as youngsters that have zero knowledge about investment. Besides that, the result of the bivariate analysis shows that people who are single and have a tertiary education background responded to the campaign better. On the other hand, people who do not have housing loans react better to the campaign. To conclude, the bank could target these potential customers with this profile who are more likely to subscribe to a term deposit by planning appealing marketing strategies to approach them.

Besides that, the bank should also choose the proper timing to approach their customer. Based on the bivariate analysis results, it shows that the bank should focus on February, March, April, September, October, and December as it has higher success rates. The customer seems to be approachable and is more likely to subscribe to term deposits if they are approached during these months. Moreover, the results show that banks are likely to successfully convince customers to subscribe to term deposits through cellular phones. Hence, the suggestion to a bank is to approach the potential customer through cellular phone at the right time which is the suggested months. However, the bank should collect and analyze the data to check if the trend is continuous or not before proceeding with this suggestion.

6 CONCLUSION

This paper proposed six machine learning models to predict the response of the customer to the direct marketing of a bank, whether they will subscribe to a term deposit or not. Obtaining the dataset from the Kaggle website, the dataset is prepared from scratch to be ready for model building. Data pre-processing steps and exploratory data analysis were done before building the models. Other than that, outliers detected in the dataset are handled to prevent them from affecting the performance of machine learning models. Feature selection via a tree-based algorithm is proposed to remove unnecessary features to improve the accuracy of prediction. The dataset is divided into two sets as a training and testing dataset to validate the accuracy of the models. LR, SVM, DT, RF, GNB, and XGBoost were implemented. Evaluation metrics including accuracy, ROC AUC score, precision, recall, and F1-score are utilized to measure the performance of the six models.

The experimental results show that the RF model has the best prediction effect which makes it the champion model. Overall, the RF model produces excellent results in terms of accuracy, ROC AUC score, precision, recall, and F1-score as compared to the other five machine learning algorithms. It was shown that the RF models outperform the other techniques with the trade-off of longer training time and space. This paper shows that the prediction ability of the RF model could contribute to increasing numbers of customers who respond positively to marketing campaigns. By utilizing the results predicted by the RF model, the bank could more accurately analyze the reaction of customers to their current marketing strategies and review their marketing strategy. This will guide bank companies to adjust their marketing strategies and target the right customers to increase the term deposit subscription. From here, the bank could stand out among the other competitors by retaining and acquiring more customers with improvised marketing campaigns thanks to the machine learning prediction model. Still, there is room for improvement in this paper as there are several limitations that could be addressed in future work.

REFERENCES

[1] Camilleri, M. A., "The use of data-driven technologies for customer-centric marketing", *International Journal of Big Data Management*, vol. 1, no. 1, pp. 50-63, 2020.

[2] Hew, Z. J., Johnson, O. V., Chew, X. Y., and Khaw, K. W., "Text summarization for news articles by machine learning techniques", *Applied Mathematics and Computing Intelligence*, vol. 11, no. 1, pp. 174-196, 2022.

[3] Plakandaras, V., Gogas, P., Papadimitriou, T., and Tsamardinos, I., "Credit Card Fraud Detection with Automated Machine Learning Systems", *Applied Artificial Intelligence*, vol. 36, no. 1, 2086354.

[4] Djurisic, V., Kascelan, L., Rogic, S., and Melovic, B., "Bank CRM optimization using predictive classification based on the support vector machine method", *Applied Artificial Intelligence*, vol. 34, no. 12, pp. 941-955, 2020.

[5] Esmeli, R., Bader-El-Den, M., and Abdullahi, H., "Towards early purchase intention prediction in online session based retailing systems", *Electronic Markets*, vol. 31, no. 3, pp. 697-715, 2021.

[6] Cheng, Y. S., Kuo, N. T., Chang, K. C., and Wu, H. T., "Using Data Mining Methods to Predict Repeat Patronage Intention in the Restaurant Industry", *Journal of Quality Assurance in Hospitality & Tourism*, pp. 1-27, 2021.

[7] Chi-Hsien, K., and Nagasawa, S. "Applying machine learning to market analysis: Knowing your luxury consumer", *Journal of Management Analytics*, vol. 6, no. 4, pp. 404-419, 2019.

[8] Sharma, S. N., and Prasanna, S. "A Novel Purchase Target Prediction System using Extreme Gradient Boosting Machines", *Int. J. Innovat. Technol. Exploring Eng*, vol. 8, pp. 2070-2072, 2019.

[9] Wang, P., and Xu, Z. "A novel consumer purchase behavior recognition method using ensemble learning algorithm", *Mathematical Problems in Engineering*, 2020.

[10] Chen, J., Diao, M., and Zhang, C. "Predicting Airline Additional Services Consumption Willingness Based on High-Dimensional Incomplete Data", *IEEE Access*, vol. 10, 39596-39603, 2022.

[11] Chaudhuri, N., Gupta, G., Vamsi, V., and Bose, I. "On the platform but will they buy? Predicting customers' purchase behavior using deep learning", *Decision Support Systems*, vol. 149, 113622, 2021.

[12] Xie, C., Zhang, J. L., Zhu, Y., Xiong, B., and Wang, G. J. "How to improve the success of bank telemarketing? Prediction and interpretability analysis based on machine learning", *Computers & Industrial Engineering*, 108874, 2022.

[13] Jin, W., and He, Y. "Three data mining models to predict bank telemarketing", In *IOP Conference Series: Materials Science and Engineering*, vol. 490, no. 6, p. 062075, 2019.

[14] Rekha, A. G., Abdulla, M. S., and Asharaf, S. "Artificial intelligence marketing: an application of a novel lightly trained support vector data description" *Journal of Information and Optimization Sciences*, vol. 37, no. 5, pp. 681-691, 2016.

[15] Feng, Y., Yin, Y., Wang, D., and Dhamotharan, L. "A dynamic ensemble selection method for bank telemarketing sales prediction", *Journal of Business Research*, vol. 139, pp. 368-382, 2022.

[16] Safarkhani, F., and Moro, S. "Improving the Accuracy of Predicting Bank Depositor's Behavior Using a Decision Tree", *Applied Sciences*, vol. 11, no. 19, p. 9016, 2021.

[17] Miguéis, V. L., Camanho, A. S., and Borges, J. "Predicting direct marketing response in banking: comparison of class imbalance methods", *Service Business*, vol. 11, no. 4, pp. 831-849, 2017.

[18] Moro, S., Cortez, P., and Rita, P. "A data-driven approach to predict the success of bank telemarketing", *Decision Support Systems*, vol. 62, pp. 22-31, 2014.

[19] Chung, D., Yun, J., Lee, J., and Jeon, Y. "Predictive model of employee attrition based on stacking ensemble learning", *Expert Systems with Applications*, 119364, 2022.

[20] Feng, Y., Wang, J., Shao, Z., Chen, Z., Yao, T., Dong, S., ... and Liang, X. "Predicting related factors of immunological response to hepatitis B vaccine in hemodialysis patients based on integration of decision tree classification and logistic regression", *Human Vaccines & Immunotherapeutics*, vol. 17, no. 9, pp. 3214-3220, 2021.

[21] Li, X., Yu, S., Zhang, Z., Radican, L., Cummins, J., Engel, S. S., ... and Weiner, M. "Predictive modeling of hypoglycemia for clinical decision support in evaluating outpatients with diabetes mellitus", *Current Medical Research and Opinion*, vol. 35, no. 11, pp. 1885-1891, 2019.

[22] Malik, E. F., Khaw, K. W., and Chew, X. Y. "Faruad detection by machine learning techniques", *Applied Mathematics and Computational Intelligence*, vol. 11, no. 1, pp. 88-103, 2022.

[23] Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.

[24] Shahare, K., Mitra, A., Naware, D., Keshri, R., and Suryawanshi, H. M. "Performance analysis and comparison of various techniques for short-term load forecasting", *Energy Reports*, vol. 9, pp. 799-808, 2023.

[25] Prapty, A. S., and Shitu, T. T. "An efficient decision tree establishment and performance analysis with different machine learning approaches on Polycystic Ovary Syndrome" In *2020 23rd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-5). IEEE, 2020.

[26] Khodabandehlou, S., & Rahman, M. Z. "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior", *Journal of Systems and Information Technology*, 2017.

[27] Katarya, R., and Srinivas, P. "Predicting heart disease at early stages using machine learning: A survey", *In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 302-305). IEEE, 2020.

[28] Mahoto, N. A., Shaikh, A., Sulaiman, A., Al Reshan, M. S., Rajab, A., and Rajab, K. "A machine learning based data modeling for medical diagnosis", *Biomedical Signal Processing and Control*, vol. 81, 104481, 2023.

[29] Shinde, T. A., and Prasad, J. R. "IoT based animal health monitoring with naive Bayes classification", *Int. J. Emerg. Trends Technol*, vol. 1, no. 2, pp. 252-257, 2017.

[30] Uddin, S., Khan, A., Hossain, M. E., and Moni, M. A. "Comparing different supervised machine learning algorithms for disease prediction", *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1- 16, 2019.

[31] Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., and Akinjobi, J. "Supervised machine learning algorithms: classification and comparison", *International Journal of Computer Trends and Technology (IJCTT),* vol. 48, no. 3, pp. 128-138, 2017.

[32] Zhang, H., Zimmerman, J., Nettleton, D., and Nordman, D. J. "Random forest prediction intervals", *The American Statistician*, 2019.

[33] Oh, G., Song, J., Park, H., and Na, C. "Evaluation of random forest in crime prediction: Comparing three-layered random forest and logistic regression", *Deviant Behavior*, pp. 1-14, 2021.

[34] Li, C., Liao, C., Meng, X., Chen, H., Chen, W., Wei, B., and Zhu, P. "Effective analysis of inpatient satisfaction: the random forest algorithm", *Patient preference and adherence*, vol. 15, pp. 691, 2021.

[35] Weinhold, L., Schmid, M., Mitchell, R., Maloney, K. O., Wright, M. N., and Berger, M. "A random forest approach for bounded outcome variables", *Journal of Computational and Graphical Statistics*, vol. 29, no. 3, pp. 639-658, 2020.

[36] Lu, C., Li, S., Gu, J., Lu, W., Olofsson, T., and Ma, J. "A hybrid ensemble learning framework for zero-energy potential prediction of photovoltaic direct-driven air conditioners", *Journal of Building Engineering*, vol. 64, 105602, 2023.

[37] Chen, B., Zheng, H., Luo, G., Chen, C., Bao, A., Liu, T., and Chen, X. "Adaptive estimation of multi-regional soil salinization using extreme gradient boosting with Bayesian TPE optimization", *International Journal of Remote Sensing*, vol. 43, no. 3, pp. 778-811, 2022.

[38] Wang, T., Reiffsteck, P., Chevalier, C., Zhu, Z., Chen, C. W., and Schmidt, F. "A novel extreme gradient boosting algorithm based model for predicting the scour risk around bridge piers: application to French railway bridges", *European Journal of Environmental and Civil Engineering*, pp. 1-19, 2022.