UDC 336.713:339.138
Preliminary Communication
https://doi.org/10.2507/IJVA.9.2.5.106

Received: August 5, 2023
Accepted for publishing: October 4, 2023

BANK TERM DEPOSIT SERVICE PATRONAGE FORECASTING USING MACHINE LEARNING

Olajide, Blessing Olajide *Engr. Dr., orcid: https://orcid.org/0000-0002-6498-2475 , Computer Engineering Department, Federal University Wukari, Nigeria, olajideblessing55@gmail.com*

Andrew, Ishaku Wreford *Computer Science Department, Federal University Wukari, Nigeria, andrewwrefordishaku@outlook.com*

Abstract:
Term deposit is one of the financial services offered by the bank. An effective bank marketing campaign to forecast possible customers to engage in personal term deposit marketing interaction is vital because it's hard to stand out, considering that all banks offer similar products. Trailing to this, this study proposed the use of machine learning algorithms to develop bank term deposit patronage forecasting models, which can study the characteristics of customers to identify potential term deposit customers. Random Forest and Xtreme Gradient Boosting algorithms and the Portuguese institution marketing campaign dataset, were used to develop a bank term deposit service patronage forecasting model. The data balancing algorithm utilized is the Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors (SMOTE-ENN) and feature selection was conducted using Information Gain. The Random Forest model achieved an accuracy of 95%, recall of 92% and f1 scores of 94%. Xtreme Gradient Boosting model achieved an accuracy of 97%, recall of 97% and f1 scores of 97%. The results of the experiment revealed the Xtreme Gradient Boosting emerged as the best model

***Keywords*:**
marketing, bank, term-deposit, patronage, Random Forest, Xtreme Gradient Boosting

1. Introduction

Banks are service institution that basically offers their customers and business owners a place to stow their cash and also source loan. Other bank services include among others; savings, deposits, insurance, remittance, bill discounting, foreign exchange facilities, payroll services, loans etc (Flohr-Nielsen , 2002: 476). However, globalisation is growing faster than ever and in consequence banks are beginning to beam more effort at reinforcing the internationalisation of their services. This became imperative because banks have been encountering numerous challenges during the past two decades, including competition, recession and image problems (Czinkota and Ronkainen, 2004: 4). In order for banks to solve these problems, they embrace marketing strategies which is the direct way of reaching their public (Czinkota and Ronkainen, 2004: 7). Bank representatives conduct marketing campaigns, promoting their brands through various sources of communication. This they did through the following mix of promotion; advertising, sales promotion, personal selling, public relations and direct marketing (Dawes and Brown, 2000: 95). Despite their measures, pressure is building up for more effective marketing management of banks and banks are realising that their established marketing strategies are inadequate for new conditions as levels of customer defection is on the rise. Traditionally, banks have tried to communicate to all potential customers in their locality, but research survey revealed that banks should aim to identify and serve micro-segments (Olajide and Wreford, 2023: 86). This will help to establish and manage individual specific long-term business relationship. Also marketing of bank services is an understudied area because most banks concentrates on marketing theory more than the marketing strategy. However, effective bank marketing strategy is crucial since services are intangible products and it’s hard to stand out, considering that all banks offer similar products.

Having established that individual commercial banks are responsible for maintaining the loyalty of their existing customers and attracting new ones. To achieve this goal, it is necessary to analyse the factors that affect the involvement of new customers, including the marketing activity of the bank. It is very important for banks to always be in a prominent position in front of potential customers to get as many deposits as possible. Hence this study will make use of two machine learning models to predict potential customers to patronise the bank term deposit service while drawing inferences from some predictors of past term deposit service customers of the bank. Machine learning focused on developing autonomous systems that use statistical tools to analyse data from a variety of sources and historical occurrences to arrive at logical conclusions based on their perceptions set :) (Omanga *et al.,* 2023: 77). Machine learning uses algorithms as well as data to mimic the learning process of humans, enabling machines to become better at generating predictions over time (James *et al.,* 2023: 242). The two machine learning models proposed for use in this study are namely Random Forest and Extreme Gradient Boost. This is with the view to help the bank make more effective decisions that will result to more profit, and help to allocate appropriate resources for customer specific service marketing campaign.

2. Literature Review

(Nguyen, 2022: 270) analysed bank customer descriptions such as age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaigns, pdays, previous, outcome and deposit to choose possible deposit customer using deep machine learning. His work mustered up Long-Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional Long-Short Term Memory (BiLSTM), Bidirectional Gated Recurrent Unit (BiGRU), and Simple Recurrent Neuron Network (SimpleRNN) as an approach to developing a model for forecasting bank deposit customers. the resulted of his findings further showed that Gated Recurrent Unit (GRU) reaches the best accuracy with 90.08% at epoch 50th. (Kurapati *et al.,*2018: 6) used machine learning for analysis of bank credit defaulters based on the customer’s information. The research used the algorithms in Scikit-Learn for prediction and compared the accuracy of the developed models before feature selection and after feature selection. Their result reported Random Forest algorithm model to have performed well in its prediction in contrast to Decision Tree, Gradient Boosting, and Extra Tree Classifier. (Hossam, 2018: 416) deployed bagging and boosting of assemble machine learning to predict the likelihood of loan approval for bank loan applicants. In his approach AdaBoost, LogitBoost, Bagging, and Random Forest. were deployed and had almost comparable performances; however, boosting techniques performed slightly better than the Bagging techniques. The best performance was achieved by the LogitBoost model yielding approximately 84% accuracy and that performance is achieved when all the seven data set features were used. However, most deployed ensemble models performed better when the number of features was reduced to the balance, age, and job attributes.

(Jan, 2021: 450) developed financial distress prediction model using two deep learning algorithms: Deep neural networks (DNN) and convolutional neural networks (CNN). Chi-squared automatic interaction detector (CHAID) was used for feature selection from Taiwan's listed and OTC sample companies' dataset. The study showed that CHAID-CNN model has the highest financial distress prediction accuracy rate of 94.23%, and the lowest type I error rate and type II error rate, which are 0.96% and 4.81%, respectively. (Gregova *et al.,* 2020: 3955) used logistic regression (LR), random forests (RF), and neural networks (NN) to predict financial distress. Fourteen financial ratios are used, and the NN model has the best performance (AUC) of 0.886. (Olajide and Wreford, 2023: 88) utilized random forest algorithm, logistic regression and relevant dataset to develop two predictive models for car-loan applicant credibility. The results from the study showed that both models were efficient in predicting the status of the car-loan credibility of customers while random forest performed relatively better than logistic regression with an accuracy of 93%, precision of 92 and recall of 1. (Muhammed *et al.,* 2021: 49) experimented and observed the performance of three classification algorithms named Support Vector Machine (SVM), Neural Network (NN), and Naive Bayes (NB) to predict the possibility of a bank customer patronising the term deposit service. Then the ability of ensemble methods to improve the efficiency of basic classification algorithms was investigated and results revealed 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.

(Marc and Amir, 2022: 23) developed five models using Baseline, Linear Regression (LR), Decission Tree (DT) and Random Forest (RF) to demonstrate how to build an innovative market place within the banking system by leveraging bank transaction data. Their study compares he performance of baseline algorithm and RF algorithm to reveal the likelihood of buying in a range of selected industries ; gastronomy, shoes, sports, cosmetics and tourism. Where the average Root Mean Square Error (RMSE) result for baseline was 0.3742 and that of the RF was 0.3597. this indicated that RF performs better than Baseline. The research was deemed suitable for small merchants selling at a physical store. The study further recommends another type of analysis to be carried out to cater for online merchants. (Laksana, 2022: 35) uses direct market dataset based on 41188 records along with two machine learning methods namely: Logistic Regression (LR) and DT to analyse the performace of bankinng telemarketing campaigns. The performance of the develop models was analysed with the test dataset and it shows AUC of 0.934 with the accuracy of 0.870 and reliability of 0.869. The model is considered suitable statistically as it is not based on multicollinarity problem. (Mohammad et al., 2023: 159) carried out detailed systematic analysis of the modeling of bank customer's behaviour in order to help banking institutions to make appropraite decisions to increase bank patronage. It made use of feature engineering processes such as sine, logarithm, min-max, Z-score and cosine methods, and selected machine learning mechods as follows; DT, Extreme Lerning Machine (ELM), Gradient Boosting, K-Nearest Neigbour (KNN) and multilayer Perceptron (MLP). The results of the study indicatred that their developed knowledge mining system offers an optimal decision support system for mining bank customers behaviour. (Subramanian et al., 2023: 160) developed three predictive models to clasify bank customers to make bank telemarketing activities more effective.. the machine learning method used are the LR, KNN and RF algorithms. The results of the three models were analysed and RF performs the best with an accuracy of 95%.

3. Proposed Methodology

In an attempt to forecast customers' subscriptions to a term deposit using the Portuguese institution marketing campaign dataset. This paper devised a three-phase methodology. The first step encompasses the sourcing and reading of the dataset via Panda's Python framework. The second phase is the data preprocessing stage and it entails the conductance of data balancing considering that the adapted dataset is highly imbalance (using SMOTE-ENN (Synthetic Minority Over-sampling Technique- Edited Nearest Neighbors) data balancing techniques) the eradication of irrelevant features such as the 'id' field, scaling and encoding the dataset feature variable to enable correlated variables for the features using the Standard Scaler and Label Encoders from the Sklearn library. Furthermore, the data preprocessing stage also incorporates the use of information gain as a feature selection mechanism. The third stage involves feeding the filtered, scaled, and selected features to the proposed machine learning models, namely the Random Forest, Extreme Gradient Boost, and their hybrid (using stacking hybridization techniques) after splitting the dataset into 70:30 per cent training to test proportion, with the training data meant for training the two deep network models and the test data for validation the accuracy of the model. The third phase further incorporates the conductance of a performance evaluation scheme to evaluate the best performance models and for comparative analysis. The stepwise approach for the implementation of the proposed Random Forest, Extreme Gradient Boost, and their hybrid model for the adapted Portuguese banking institution dataset is shown in Figure 1.

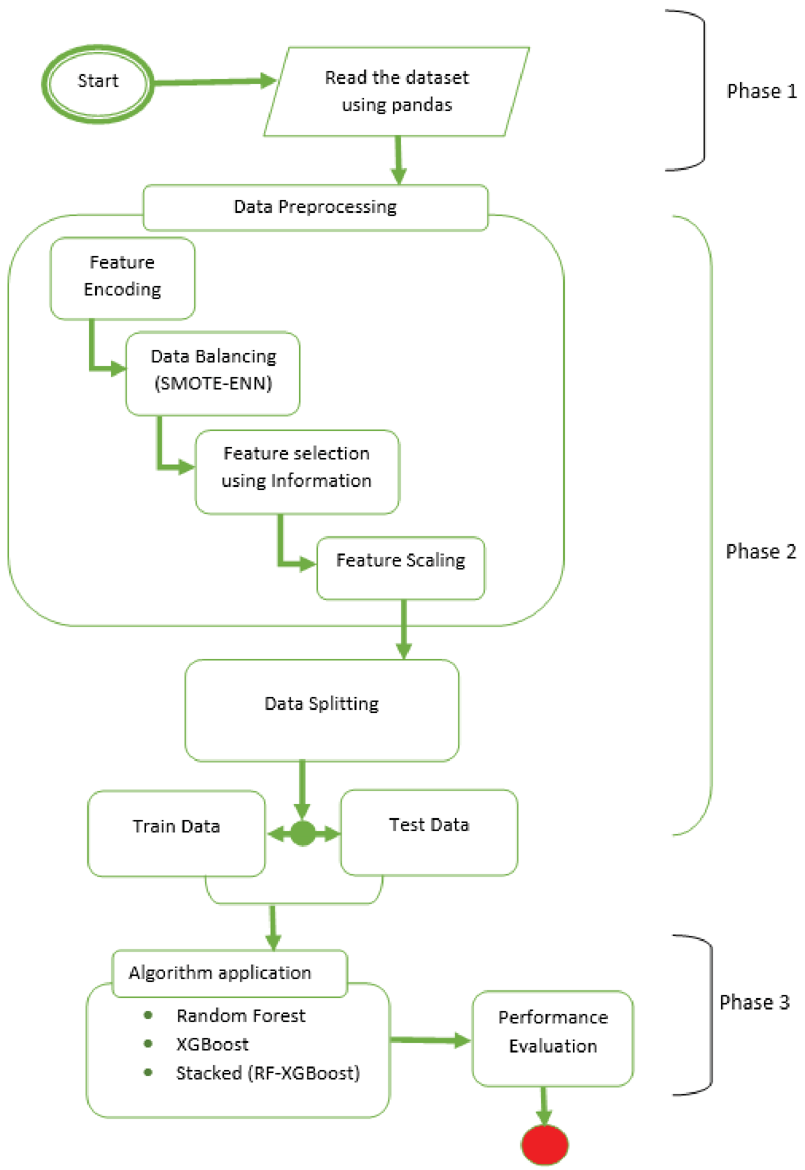

Fig 1: Proposed Methodology for Developing Bank Term Deposit Service Patronage Forecasting model

3.1 Dataset Description

The adapted dataset is the Portuguese bank marketing campaign dataset sourced from the Kaggle machine learning repository. The objective of the dataset is to classify and thus predict if a customer will subscribe for a term deposit as the dependent variable say **'y' (**Moro, 2014: 25). The marketing campaigns were based on phone calls as often customers contact the bank, to access if the product termed deposit would be granted (yes) or not (no). An overview of the dataset attributes is presented in Table 1.

Table 1: Dataset Description Portuguese bank marketing campaign dataset

Feature	**Description**
Age	The age field defines the age of the customer
Job	The job here defines the job the current customer has as an occupation
Marital	The marital status of a customer can be married, single, divorced, etc.
Education	Defines the educational level of the customer
Default	Define if a customer has credit in default
Housing	Defines if a customer has a housing loan
Balance	Customers individual balance
Loan	Define whether a customer has a pending loan or not.
Contact	Communication type for a customer.
Month	Last contact month of the year
Day	Last contact day of the week
Duration	Last contact duration, in seconds
Campaign	Number of contacts performed during the campaign and for this instance client
P-days	The number of days passed by after the client was last contacted from a previous campaign.
Previous	The number of contacts performed before this campaign and for the instant client.
P-outcome	The outcome of the previous marketing campaign.
Y	Defines if the client subscribed to a term called a deposit.

3.2 Data Balancing Techniques

As aforementioned the Portuguese banking institution dataset contains a highly imbalanced dataset and thus requires some data balancing algorithms. Hence, the data balancing algorithm utilized is the Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors (SMOTE-ENN). The algorithm combines two techniques, namely the Synthetic Minority Over-sampling Technique (SMOTE) and Edited Nearest Neighbours (ENN). The SMOTE algorithm checks the minority class label sample and randomly selects one of its nearest neighbours 'k' (usually k=5) before creating a synthetic sample by linearly interpolating between the selected sample and the original minority sample. This process is repeated until the desired balance is achieved. The ENN (Edited Nearest Neighbours) utilized an under-sampling technique aimed at removing noisy samples from the dataset by checking a particular sample class label and its k nearest neighbours (usually k=3) if the majority class samples outnumber the minority class samples among the k nearest neighbours, the sample is considered noisy and removed from the dataset, the objective is to reduce noise and potential misclassification caused by noisy samples (Muntasir, 2022: 16); (Ghorbani and Ghousi, 2020: 6909).

3.3 Information Gain for Feature Selection

For the selection of relevant features, while reducing the dimensionality of the Portuguese banking institution dataset and possibly increasing the performance of the machine learning models, the experiment adapted the Information Gain (IG) algorithm as the feature selection technique considering its suitability for the categorical problem as in the case of customer's subscriptions to a campaign or not. Information Gain (IG) measures the significance of a feature by quantifying the quantity of information provided about the class labels (called deposit). Hence, the more valuable a feature is for classification or prediction, the greater the information gain. Entropy is used to quantify the degree of uncertainty in a dataset when selecting the most pertinent characteristics. It is derived from the distribution of deposit labels within the dataset. Typically, the formula used to calculate entropy is founded on the concept of Shannon entropy and using the entropy formula, information gain is also measured as the degree of entropy reduction of the deposit variable after dataset division. This implies that the greater the information gain, the more useful its attribute is for distinguishing between classes. Information Gain is depicted by equation (1)

$$IG = E_{bs} - WE_s \tag{i}$$

Where IG represent Information Gain, E_{bs} is the entropy before separation, and the WE_s represent the weighted average entropy following splitting.

3.4 Random Forest

The proposed Random Forest algorithm is a popular ensemble learning algorithm used for classification tasks. The algorithm is based on decision trees and integrates the predictions of multiple trees to produce more precise and robust classifications. The workflow of the adapted random forest algorithm entails the creation of an ensemble of decision trees, each of which is trained on a random subset of the training data and a random subset of features, followed by the application of bootstrapped sampling techniques to each of the decision trees and the random sampling of the subsets of the decision tree with replacement. This process assures that each tree receives slightly different data, resulting in trees that are diverse and uncorrelated. Randomly selecting the subset features of the adapted dataset at each split (the splits are determined based on criteria such as Gini impurity or entropy, which measure the purity of the target class in each subset of data) increases the diversity of the trees and reduces the risk of over-reliance on any particular feature. Therefore, once all decision trees are constructed, each tree predicts, for each sample in the test set, whether the record is considered subscribed (deposit variable) by the customer or not. After that, the final prediction for a sample is determined by combining the predictions of each tree using majority voting. The class with the most votes (i.e., the most frequent prediction across all trees) is deemed the ultimate prediction for a given sample. The random forest algorithm was selected because the study demonstrated its effectiveness in classification tasks due to its capacity to handle high-dimensional data, resistance to overfitting, and ability to identify important features, making it a valuable tool for identifying customers' behaviour from a bank marketing campaign.

3.5 Xtreme Gradient Boost

XGBoost (Xtreme Gradient Boosting) is one of the most widely used ensemble machine learning algorithms for solving predictive and classification problems (Islam, 2022: 1005). Essentially, the algorithm combines multiple weak prediction models of decision trees into a robust predictive model using a gradient-boosting approach (Shehadeh et al., 2021: 1046). Gradient boosting is an iterative training procedure for decision trees. Throughout the process, successive decision trees are constructed, and each tree is trained to rectify the errors of the previous tree. The model begins with an initial decision tree and then adds additional trees iteratively to reduce error. Considering that the problem is a classification problem, the implemented Extreme Gradient Boosting algorithm employs the logistic loss function to measure the difference between the predicted and actual outcomes to monitor performance. This was accomplished using gradient descent to minimize the loss function. To prevent overfitting and improve generalization. The implemented Extreme Gradient Boosting algorithm uses regularization techniques to regulate the complexity of individual trees and the ensemble process as a whole.

3.6 Evaluation Metric

To evaluate the performance of the classifiers, some standard evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics are proposed based on some machine learning evaluation parameters such as the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) evaluation parameters.

i. **True positive (TP):** defines an instance where an observation emerged from positive classes and the classifier predicted it to be positive.
ii. **False Negative (FN):** The false-negative variable defines a condition when the actual observation from the dataset comes from a positive class label but the model predicted the label to be negative.
iii. **False-Positive (FP)** is a condition when the actual observation comes from negative classes but the model predicted the outcome to be positive.
iv. **True negative (TN)** identifies instances where observations from negative classes are predicted to be negative.

Accuracy: calculates the ratio of inputs in the test set correctly labelled by the classifier. Mathematically, accuracy can be denoted as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{(ii)}$$

Precision: defines the percentage of the number of correctly predicted positive outcomes divided by the total number of predicted positive outcomes. Thus, precision can be mathematically denoted as:

$$precision = \frac{TP}{TP+FP} \qquad \text{(iii)}$$

Recall: is the percentage of correctly predicted positive output to the actual number of positive outcomes from the dataset and can be mathematically denoted as:

$$Recall = \frac{TP}{TP+FN} Recall = \frac{TP}{TP+FN} \tag{iv}$$

F1-score is a measure that defines the harmonic mean of the model precision and recalls and thence combines the value of the recall and precision to output a single score. The F1 Score can be mathematically expressed as follows:

$$F_{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{v}$$

4. Results and Discussions

The proposed bank marketing prediction models were implemented on a Windows operating system, using the Python programming language and the Anaconda programming environment. The system setup involved a dual-core Intel Core I5 processor with 4GB of RAM. The integrated model is comprised of the Random Forest and XGBoost algorithms. In the implementation, various Python packages were used. Considering that the models are machine learning algorithms, the Sklearn Application Programming Interface (API) was employed. Additionally, the NumPy modules were utilized for numerical operations, pandas for reading the phishing dataset, and Matplotlib for visualizing the graphical behaviour of the implemented models.

4.1 Visualization of Portuguese Bank Marketing Campaign Dataset

To visualize the customers that subscribe to the campaign, the bar chart in Figure 2 was utilized via the count plot function using the seaborn packages. The y-axis shows the counts on the number of customers whereas the x-axis depicts the customers in the bar chart that subscribed as ‘yes’ or didn’t subscribe as ‘no’ to the marketing campaign. From the diagram, it can be identified that the dataset is imbalanced as the distribution has a high rate of disparity between the subscribed and not subscribed customers. The observation resulted in the need for the application of data balancing techniques.

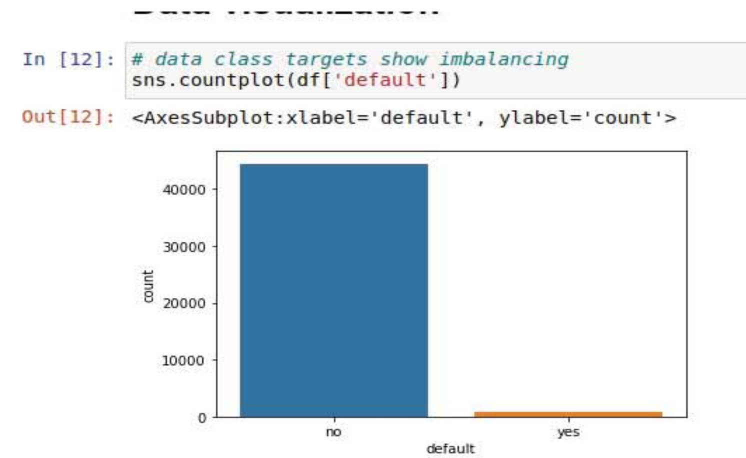

Fig 2: subscription counts

The instance correlation matrix in Figure 3, uses the correlation coefficients annotated to each cell to establish the degree to which each feature is connected in the prediction of customer deposits to a marketing campaign. From the correlation diagram, a value of 0 denotes a neutral correlation, a value of -1 denotes a weak correlation, and a value of 1 denotes a substantial influence between two factors in predicting cases whether or not an instance customer can subscribe to the term deposit after the backing marketing campaign. The diagonal axis is always equal to one because each attribute has a strong association with itself.

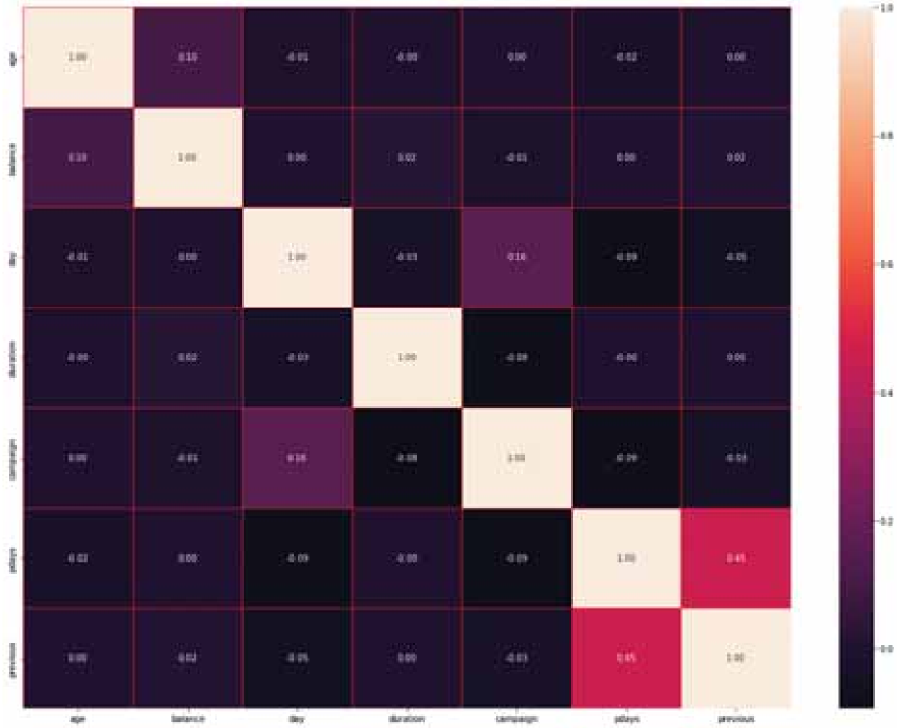

Fig 3: Correlation Diagram

Figure 4 shows the comparison of customers based on their respective occupations appended to the y-axis of the graph. The x-axis in the diagram shows the tally for the respective jobs. The jobs include management, technician, entrepreneur, blue-collar, unknown, retired, admin, services, housemaid, self-employed, unemployed, and students each having an aggregation of the numbers of individuals in an instance job category.

```
In [13]: # checking the number of job based on specialization of the customers
         sns.countplot(y='job', data=df)

Out[13]: <AxesSubplot:xlabel='count', ylabel='job'>
```

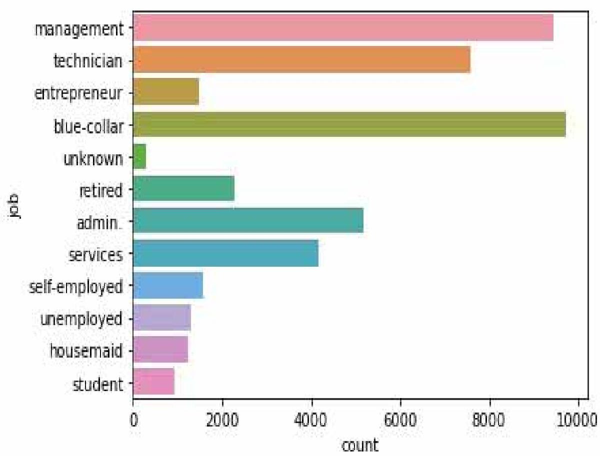

Fig 4: Correlation Diagram

4.2 Result Presentation

The experimental result of the research on the Portuguese institution marketing campaign is presented in Table 2. The headers provide details about the classification models developed, specifically the Random Forest and Xtreme Gradient Boosting, and the accuracy, precision, recall, and f1 scores for these models. The Random Forest model achieved an accuracy of 95%. When considering customers who subscribed to the term deposit (yes label), the precision was 96%, meaning it correctly identified 96% of the positive cases. For customers who did not subscribe to the campaign (no label), the precision was 94%, indicating it accurately identified 94% of the negative cases. The recall for the yes label was 92% and 97% recall for the no label. The f1 scores were 94% for the yes label and 95% for the no label, which are measures that balance precision and recall for each class.

On the other hand, the Xtreme Gradient Boosting model achieved an accuracy of 97%. Its precision for the yes-class label was 96%, and 97%, for the no-class label, the precision was 97%, correctly classifying 97% of the negative cases. The recall for both the yes and no class labels was 97%. The f1 score for both the yes and no class labels was 97%, which shows a well-balanced performance in terms of precision and recall for each class, respectively.

Table 2: Models (Random Forest, XGBoost, and Stacked Model) performance evaluation report

S. No.	Classification Model	Evaluation Metrics for the Classification Models			
		Accuracy (0/1) %	Precision (0/1) %	Recall (0/1) %	F1-Score (0/1) %
1	Random Forest (RF)	95	96/94	92/97	94/95
2	XGBoost	97	96/97	97/97	97/97

Considering Table 2, Figure 5, and Figure 6, it can be identified that the Xtreme Gradient Boost model has the best performance accuracy score, even when precision, recall and f1-score are considered. Cumulatively, it can be deduced that both the Xtreme Gradient Boost model has the best performance in the prediction of customer subscriptions to the term deposit on the Portuguese institution marketing campaign dataset.

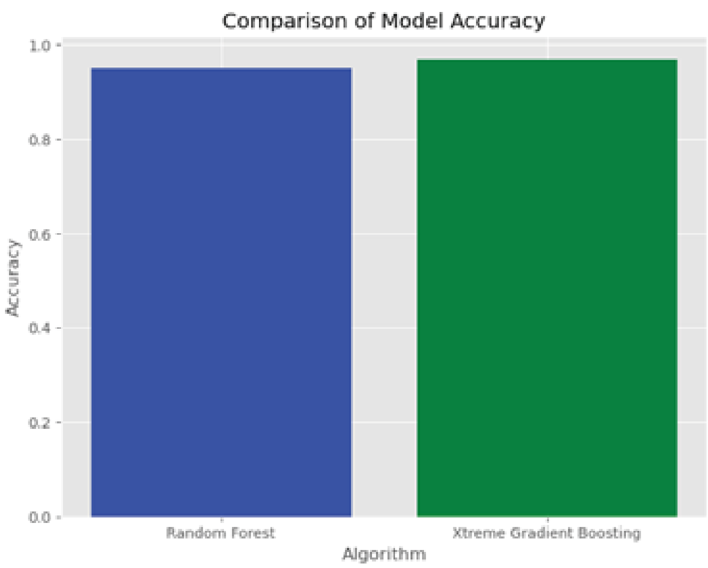

Fig 5: Model Performance Comparison

Fig 6: Model Performance Comparison based on Precision, Recall, and F1 score

5. Conclusion and Recommendation

In this paper, two machine learning algorithms namely the Random Forest and XGBoost performance were evaluated on the Portuguese institution marketing dataset. The effectiveness of the developed random forest and XGboost models was validated via some performance evaluation metrics including accuracy score, precision, recall, and f1 score. Analytically, the result of the experiment revealed the Xtreme Gradient Boosting to emerge as the best model with an accuracy of 97% while the random forest has an accuracy of 95%. The results obtained can be attributed to the application of the SMOTE-ENN which uses the viabilities of data oversampling and under-sampling and also the information gain algorithm as the feature selection techniques. Future studies can apply the viabilities of artificial intelligence methods such as the Boruta algorithm for feature selection mechanisms. Furthermore, more datasets can be employed to enhance the model's effectiveness in predicting customer's subscriptions to deposit campaigns.

Reference

1. Czinkota, M. R.; Ronkainen (2004) *International Marketing*. 7th Ed. Ohio: Harcourt. Inc.
2. Dawes, J.; Brown R. B. (2000) Postmodern Marketing: Research Issues for Retail Financial Services. Qualitative Market Research. *An International Journal*, Vol 3, No 2, , pp. 90-98.
3. Flohr-Nielsen, J. (2002) Internet Technology and Customer Linking in Nordic Banking, *International Journal of Service Industry Management*, Vol 13, No 5, pp. 475-495.
4. Ghorbani, R.; Ghousi, R. (2020) Comparing different resampling methods in predicting students' performance using machine learning techniques", *IEEE Access*, *8*, 2020, pp. 67899-67911.
5. Gregova, E.; Valaskova, K.; Adamko, P.; Tumpach, M.; Jaros, J. (2020) Predicting Financial Distress of Slovak Enterprises: Comparison of Selected Traditional and Learning Algorithms Methods. *Sustainability*, 12, pp. 3954-3959.

6. Hossam M. (2020) Predicting Loan Approval of Bank Direct Marketing Data Using Ensemble Machine Learning Algorithm. *International Journal of Circuits, Systems and Signal Processing*, Vol. 14, pp. 914-922.
7. Islam, M. D.et al. (2022) Airbnb rental price modelling based on Latent Dirichlet Allocation and MESF-XGBoost composite model. *Machine Learning with Applications*, Vol. 7, pp. 1002-1008.
8. James O.; Olajide B. O.; Omanga M. J. (2023) Over-the-top Application Traffic Analysis Model for Networks using Multilayer Perceptron (MLP) and Long Short-term Memory (LSTM). *Asian Journal of Pure and Applied Mathematics*. Vol. 5, Issue 1, pp. 242-250.
9. Jan, C. L. (2021) Financial Information Asymmetry: Using Deep Learning Algorithms to Predict Financial Distress. *Symmetry*, Vol. 13, pp. 443-453., https://doi.org/10.3390/sym13030443
10. Kurapati, N.; Kumari, P.; Bhansali. (2018) Evaluating the performance of machine learning algorithms in financial market forecasting: A comprehensive survey in http://arxiv.org.
11. Laksana V. (2022) *Analysis of Predicting the Success of the Banking Telemarketing Campaigns by Using Machine Learning Techniques*. A Published MSc. Research Project Submitted to School of Computing, National College of Ireland. pp. 1-75.
12. Marc T.; Amir T (2022) A banking Platform to Leverage Data-Driven Marketing with Machine Learning, *Entropy*, Vol. 24, pp. 1-22. https://doi.org/10.3390/e24030347
13. Maohammad Z. A.et al.. (2023) Modelling Bank Customer Behaviour Using Feature Engineering and Classification Techniques. *Research in International Business and Finance*, Vol. 65, pp. 153-164.
14. Moro, S.; Cortez, P.; Rita, P. (2014) A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, *Elsevier*, 62, pp. 22-31.
15. Muhammed J. A.; Akter S.; Bin, A.M.; Rezaul K. M. (2021) Bank Deposit Prediction using Ensemble Learning, Artificial Intelligence Evolution. Vol. 2, issue 2, pp 42-51. DOI: https://doi.org/10.37256/aie.222021880
16. Muntasir N. M.et al. (2022) A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset. *Scientific Programming*, pp. 1-17.
17. Nguyen M. T. (2022) Machine Learning Performance on Predicting Banking Term Deposit. In : *Proceedings of the 24th International Conference on Enterprise Information Systems* (ICEIS 2022) – Vol.1, pp. 267-272.
18. Olajide B. O.; Wreford A. I. (2023) A Predictive Model for Car-Loan Repayment Credibility of Customers. *Asian Basic and Applied Research Journal*, Vol. 5, No. 1, pp.81-89.
19. Omanga M. J.; Olajide, B. O.; Okpor J.; Andrew I. W. (2023) Smart User Consumption Analysis Using Support Vector Machine and Multilayer Perceptron. *Asian Research Journal of Current Science,* Vol. 5, Issue 1, pp.163-170.
20. Shehadeh, A.; Alshboul, O.; Al Mamlook, R. E.; Hamedat, O. (2021) Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. *Automation in Construction*, Vol. 129, pp. 1038- 1048.
21. Subramanian M.; Shankar N. B.; Vijaya P. (2023) Deploy Machine Learning Model for Effective Bank Telemarketing Campaign. *International Conference on Distributed Computing and Electrical Circuits and Electronics,* pp. 158- 162.