

# Bank Marketing Data Classification Using Machine Learning

Dr. Smitha Shekar B<sup>1</sup>, Pooja A<sup>2</sup>

Associate Professor, Department of CS&E, Dr.A.I.T, Bengaluru, India<sup>1</sup>

M.Tech Student, Department of CS&E, Dr.A.I.T, Bengaluru, India<sup>2</sup>

**Abstract:** In today's world, where huge amount of data is generated in every field of day to day activities, banking sector is one of them. As an outcome of work, various machine learning concept are studied with respect to Bank marketing data classification. Banking is a provision of the services by bank to an individual customer. The dataset is originally collected from UCI Machine learning repository and Kaggle website. The data is related to bank marketing campaigns of banking institution based on phone call. In this work, Python is used as a coding language and Machine learning concept is used as statistical learning for data analysis. The main reason of using machine learning is to build a predictive model to produce the better prediction. The outcome of the result is analyzed with supervised Naïve Bayes algorithm for classification purpose. The main objective of building the model is to describe whether the customer has opted for term deposit. The bank should target the potential customer with considerable amount of time responding to the phone calls. The work implemented resulted in measuring accuracy, precision, recall and F1 score, towards term deposit prediction.

**Keywords:** Bank marketing, Customer, Machine learning, Deposit prediction

## I. INTRODUCTION

A Bank is a financial institution, which provide various service to the customer which perform deposit and providing a loan at an interest rate to the various customer. Banks store massive amount of information about their customer to improve the banking strategies and to maintain good relationship between the customers. Customer are the main asset of the bank. Usually, the selected customer are contacted directly through mail, email, personal contact, telephone cellular or any other contact to advertise the new service this kind of marketing called direct marketing. The objective of marketing in banking is to attract the new customers[1]. The collected data from UCI machine learning repository[2], is related to bank marketing campaign of banking institution the classification goal is to predict if the customer will subscribe the term deposit.

In this work Python used as a programming language[12], high-level, interpreter and extensive standard library are freely available source for all major platform from the Python web site and Machine learning technique for data analysis method and automates analytical building model to predict the accuracy of the bank customer data. Where each instance in a dataset is described by a set of attributes and classification algorithm used such as Naïve bayes classifier algorithm gave the best performance measure accuracy of the data. The bank should target the potential customer who have spent considerable amount of time responding the bank calls.

The main object of this work is to find how to use machine learning technique[6], for analysis and making the prediction using existing dataset in banking marketing for creating effective decision making knowledge and to build a machine learning model using classification algorithm to predict the accuracy of the data.

## II. LITERATURE REVIEW

The author in [3] had used the machine learning techniques for analysis and making prediction using existing data in banking marketing. The success rate of banking marketing depend on the result and decision in order to make more accurate prediction statistical tool and methods are used. A different stage for data analysis and to find, how they can be used together in a process converting raw data to effective decision making knowledge and building the predictive model in this work used decision tree algorithm will help to predict the customer will subscribe the term deposit.

Elsalamony et all. discussed all bank marketing campaign are depend on customer large data, the size of data source is impossible for human to analyst to come up with satisfying information that will help in decision making process. Data mining model are helping in the performance of the campaigns, in this work used most important data mining technique Multilayer Perception Neural Network(MLPNN),Naïve bayes, logistic regression, and decision tree, the purpose is increasing the campaign effectiveness and identifying the characteristics that effect a success[4].

A data driven approach was suggested in [5], to predict the success of bank telemarketing used data mining approach to predict the success telemarketing call for term deposits, data related to Portuguese retail bank it include the effect of financial crisis, analyzed large set of feature related to bank client, social and economic characteristics and product. In the modelling phase a semi-automatic feature had selected, performed with the data prior and reduce set of the feature. Compare data mining model super vector machine, decision tree, logistic regression and neural network, using two metrics, the four models were tested and neural network present the best result, decision tree is a knowledge extraction method were applied to neural network to predict the several key attribute. Finally, the selected model as credible and valuable for telemarketing campaign.

Bank direct marketing is an interactive process[6], for building the good relationship among customers, to study the customer characteristics and behavior use an effective multi-channel communication. A part from profit growth, which may raise customer positive response, the goal of bank marketing is to increase the customer response of direct marketing campaign.

Customer profiling in [7], using classification approach for bank telemarketing, data mining approaches started by many companies to restore the customer profiling. Decision tree, random forest, and Naïve Bayes were used, for predicting the customer profiles and increasing the telemarketing sales classification is useful for measured accuracy percentage, precision and recall rates. Before evaluating the classifiers preprocessing and normalization were conducted for conducting the experiments and evaluation process RapidMiner tool was used. Finally, result show that decision tree is the best classifier for predicting the customer profile and behavior.

### III. DESIGN AND METHODOLOGY

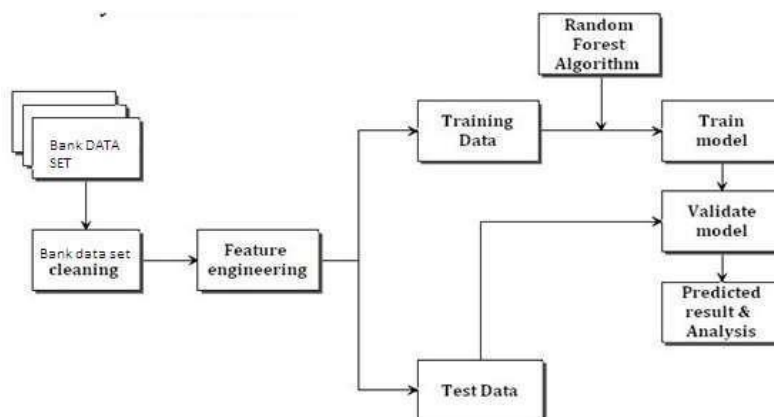


Fig.1 System Design

The system design include various stages like data collection, pre-processing, making training, testing data, implementing algorithm and last stage is predicted result. The collected raw data may be incomplete or noisy. The data must go through pre-processing phase to clean the data before using the data for learning, another step for training model is feature extraction. In machine learning feature engineering is a critical aspect it includes feature selection and extraction it boost the model performance. When we trying to predict the output first we need to train the model using a dataset and model try to learn the data to make accurate prediction. Test data is independent of training data, if a model fit to the train data then it also fit to the test data, least overfitting has taken place. Random forest is most used algorithm in ML most of time this algorithm give the best accuracy result. Training the model using ML algorithm with the training data. The process of trained model is evaluated with the test data is validate data, finally, the model give the best predicted result.

**A. Python:** Python is a high-level, interpreted and general purpose programming language[12]. Python is used by a software developers as a support language, it is easy to learn and its syntax is very easy code and consists of lot of code library, easy to build models for machine learning. This program consists of fewer lines of code than the other programming language. Various companies used Anaconda, is the most popular Python distribution widely used for machine learning and data science.

**B. Machine Learning:** Machine learning is an application of artificial intelligence[8], is a method for data analysis and automates building model it learn from the previous data based on the ideas it identify data pattern and take decision on minimal human intervention. Machine learning mainly concerned with pattern and accuracy. Most industries working on machine learning technique to analyze large amount of data such as financial service, Government, Healthcare, Retail, and Transportation.

**C. Supervised:** The majority of machine learning uses supervised learning [9]. In supervised learning it allows to

collect the data and produce the output data based on previous experience. The task of learning function that maps an input and output variable and use an algorithm to learn the mapping function from the input to the output, the process of an algorithm learning from the training data. Supervised learning classified into two groups, classification and regression.

**D. Unsupervised:** Unsupervised learning algorithm are used when the information used to training the machine that is neither classified nor labelled and algorithm allowed to act on the data without guidance, this algorithm mainly deals with hidden structure from unlabelled data and this algorithm does not give the right output[10]. Unsupervised learning algorithm are less accurate compared to supervised learning, Unsupervised learning classified into two groups, clustering and association problems [11].

**E. Random Forest:** Random Forest [13] is a supervised, flexible, straightforward learning algorithm used for classification and regression. This random forest consists of multitude of decision tree and results are aggregated, random forest collect the classification and select the most voted prediction as the result, this algorithm reduce the risk of overfitting. Random forest algorithm is reduce overfitting, high accuracy and estimates missing data.

**F. Naïve Bayes:** It is most effective classification technique[14] based on Bayes theorem used for solving classification problem with an assumption of independent between predictors and calculate the probability of an event related to previous knowledge. A Naïve Bayes classifier assumes that the presence of a particular features in a class is unrelated to the presence of any other feature.

#### IV. RESULT

The Customer bank dataset is used for term deposit prediction. This dataset is publicly available at UCI machine learning repository. It contains customer information. The dataset has two type of prediction either Yes or No. There are 16 input feature and 1 output. After implementing Supervised Naïve Bayes algorithm used for classification purpose, the algorithm gives 82.65 accuracy for dataset by measuring accuracy, precision, recall, f1-score.

ID	Attributes	Type	Values	Descriptions
1	Age	Numeric	Real	Age at the contact date (≥18)
2	Job	Categorical	Admin, Unknown, Unemployed, Management, Housemaid, Entrepreneur, Student, Blue-collar, Self-employed, Retired, Technician, Services	
3	Marital	Categorical	Married, Divorced, Single, widowed	
4	Education	Categorical	Unknown, Secondary, Primary, Tertiary	
5	Default	Binary	Yes, No	Yes or No
6	Balance	Numeric	Real	In euro currency
7	Housing	Binary	Yes, No	Yes or No
8	Loan	Binary	Yes, No	Yes or No
9	Contact	Categorical	Unknown, Telephone, Cellular	
10	Day	Numeric	Real	Referring to when the contact was made
11	Month	Categorical	Jan, Feb, mar, ..., Nov, Dec	
12	Duration	Numeric	Real	Of the contact (in seconds)
13	Campaign	Numeric	Real	
14	Pday	Numeric	Real	
15	Previous	Numeric	Real	
16	Poutcome	Categorical	Unknown, Failure, Success	

Fig.2 Attributes of bank dataset

```
In [17]: acc_nb = accuracy_score(y_true=y_test, y_pred= y_pred)
print("Overall accuracy of NB model using test-set is : %f" %(acc_nb*100))
Overall accuracy of NB model using test-set is : 82.651934
```

Fig.3 NB model accuracy

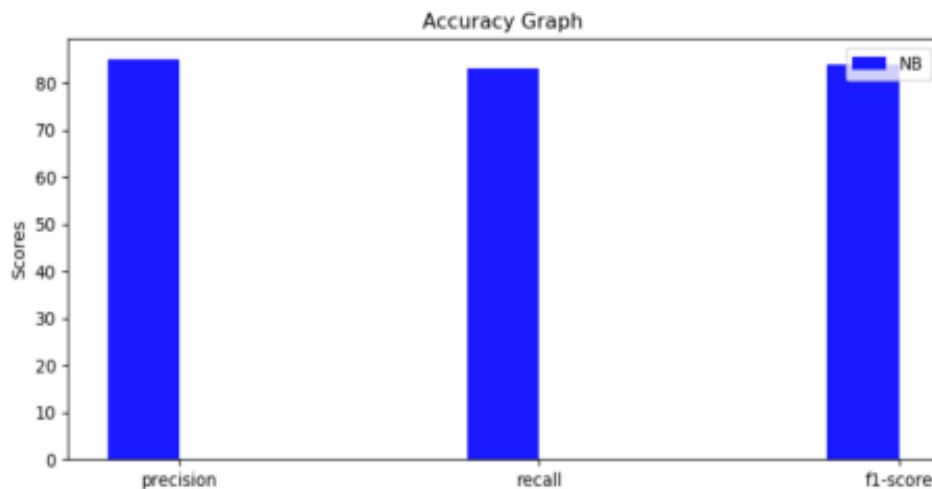


Fig.4 Accuracy Graph

## V. CONCLUSION

In banking field huge amount of data is generated continuously and this data can be used to extract meaningful information. The main objective of this work to predict whether a customer will subscribe to a term deposit. The work in this paper have used bank dataset from UCI machine learning repository or Kaggle website to make classification. After implementing, the result obtained was satisfactory.

## REFERENCE

- [1]. <https://www.technofunc.com/index.php/domain-knowledge/banking-domain/item/what-is-a-bank>
- [2]. [archive.ics.uci.edu/ml/datasets/Bank+Marketing](https://archive.ics.uci.edu/ml/datasets/Bank+Marketing).
- [3]. Mihova, Yana. (2019). Knowledge creation in banking marketing using machine learning techniques. 10.13140/RG.2.2.31756.16007.
- [4]. Elsalamony, Hany. (2013). Bank Direct Marketing Analysis of Data Mining Techniques. International Journal of Computer Applications. 85. 10.5120/14852-3218. [5]. Moro, S et al., (2014). A data-driven approach to predict the success of bank telemarketing. doi:10.1016/j.dss.2014.03.001
- [6]. Miguéis, V.L., Camanho, A.S. & Borges, J. Predicting direct marketing response in banking: comparison of class imbalance methods. *Serv Bus* **11**, 831–849 (2017). <https://doi.org/10.1007/s11628-016-0332-3>
- [7]. S. Palaniappan, A. Mustapha, C. F. M. Foozy, and R. Atan, “Customer profiling using classification approach for bank telemarketing,” JOIV: International Journal on Informatics Visualization, vol. 1, no. 4-2, pp. 214–217, 2017.
- [8]. [https://www.sas.com/en\\_in/insights/analytics/machine-learning.html](https://www.sas.com/en_in/insights/analytics/machine-learning.html)
- [9]. <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>
- [10]. <https://www.expert.ai/blog/machine-learning-definition>
- [11]. <https://www.geeksforgeeks.org/supervised-unsupervised-learning/>
- [12]. Akshansh Sharma et al, May(2020). Python: The Programming Language of Future, IJIRT | Volume 6 Issue 12 | ISSN: 2349-6002
- [13]. W.T. Aung, K.H. Hla, Random forest classifier for multi-category classification of web pages, in IEEE Asia-Pacific Conference on Service Computing, Biopolis, Singapore (2009), pp. 372–376 Google Scholar
- [14]. Kaviani, Pouria & Dhotre, Sunita. (2017). Short Survey on Naive Bayes Algorithm. International Journal of Advance Research in Computer Science and Management. 04.