



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: IV Month of publication: April 2022

DOI: <https://doi.org/10.22214/ijraset.2022.41713>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Predicting the Probability of Bank Deposit Subscription

Prof. Pallavi Patil¹, Sanket More², Askaan Shaikh³, Hitesh Patil⁴, Rohit Patil⁵

¹Professor, Dept. of Computer Engineering, Genba Sopanrao Moze College of Engineering, Balewadi

^{2, 3, 4, 5} Dept. of Computer Engineering, Genba Sopanrao Moze College of Engineering, Balewadi

Abstract: Recently, economic depression, which scoured all over the world, affects business organizations and banking sectors. Such an economic pose causes severe attrition for banks and customer retention becomes impossible. Accordingly, marketing managers are in need to increase marketing campaigns, whereas organizations evade both expenses and business expansion. To solve such a riddle, data mining techniques are used as an uttermost factor in data analysis, data summarizations, hidden pattern discovery, and data interpretation. In this paper, rough set theory and decision tree mining techniques have been implemented, using real marketing data obtained from a Portuguese marketing campaign related to bank deposit subscriptions.

I. INTRODUCTION

Business intelligence is a recent term that concerns using the information space and intelligent mechanisms to support business managers' decisions. Since business organizations including banking sectors yield tones of records and transactions every day, the most suitable intelligent mechanisms that can handle such vast growth of data set and information is data mining techniques.

Data mining is known as the process of monitoring new and innovative information from the vast amount of data sets by discovering hidden and unknown relationships between features that are entailed in the data records, spotting the interesting events and buried patterns, summarizing the information space to extract predictive decision rules, discriminating the information space into sets of objects and minimizing the features the describes the information space. Accordingly, DM can be used to aid decision-makers in the banking sector to confront the economical pretence by avoiding risky transactions that cause bank attrition and increasing the customer retention incentives to raise the bank revenues.

II. DATA SET TERMINOLOGY

In this research, we use a real dataset that was collected from a Portuguese bank that used its contact centre to do direct marketing campaigns to motivate and attract deposit clients. The dataset is related to 17 marketing campaigns and corresponds to 79354 contacts. The telephone and the internet were the central marketing channel, in which, an attractive long-term deposit application, with good interest rates, was offered.

There are two datasets:

- 1) Bank-full.csv that contains various examples corresponding to 45211 objects and ordered by date.
- 2) Bank.csv that holds 10% of the examples (4521 records), randomly selected from bank-full.csv.

However, it contains almost all possible varieties for the attributes' values and object instances. The bank.csv data set was firstly used in the implementation phase as a test database; however, it has been implemented in the form of a relational database as seen below in the database implementation subsection.

III. DATA SET DESCRIPTION

The dataset consists of one table with 16 non-empty conditional attributes and one decision attribute, where:

- 1) Age: the age of the customer
- 2) Job: type of job (categorical)
- 3) marital: marital status (categorical)
- 4) education: the education level (categorical)
- 5) Default: has credit in default?
- 6) Balance: average yearly balance
- 7) Housing: has a housing loan?
- 8) Loan: has a personal loan?
- 9) Contact: last contact of the current campaign (categorical)

- 10) day: last contact day
- 11) month: last contact month
- 12) duration: last contact duration in seconds
- 13) campaign: number of contacts performed during this campaign for this client includes the last contact.
- 14) pdays: number of days that passed by after the client was last contacted from a previous campaign
- 15) previous: number of contacts performed before this campaign for this client
- 16) poutcome: outcome of the previous marketing campaign (Categorical)

Output attribute (desired target):

Deposits: has the client subscribed to a term deposit?

The attributes types are various continuous, categorical, binary and discrete. Where categorical type means that its value is limited between several choices

IV. LITERATURE SURVEY

According to findings based on the paper of Advanced Data Analysis, Department of Statistics, Columbia University: Based on signs of coefficients of variables in logistic regression, “duration” has a positive effect on people saying “yes”. This is because the longer the conversations on the phone, the higher interest the customer will show to the term deposit. “nr.employed”, which is the number of employees in the bank, has a positive effect on turning people to subscribe the term deposit. This can be because the more employees the bank has, the more influential and prestigious this bank is. “Euribor” is another important variable, which denotes the Euribor 3-month rate. This indicator is based on the average interbank interest rates in Eurozone. It also has a positive effect since the higher the interest rates the more willingly customer will spend their money on financial tools. Employment variation rate (emp.var.rate) has a negative influence, which means the change in the employment rate will make customers less likely to subscribe to a term deposit. This makes sense because the employment rate is an indicator of the macroeconomy. A stable employment rate denotes a stable economic environment in which people are more confident to make their investments. Therefore, if banks want to improve their lead generation, what they should do is hire more people to work for them, improve the quality of conversation on the phone and run their campaigns when interest rates are high and the macroeconomic environment is stable.

According to the findings based on the paper in the International Journal of Computer Applications: The set of features that describes the data set can be discriminated from 16 features into the 3 features predominant set {Age, Duration, Balance}, which is considered as a CORE of that dataset. Thus, a huge number of more valuable decisions and predictive rules can be extracted based on such a CORE set, these rules, in turn, help decision-makers to have a provision about acquiring and targeting customers, make faster and better decisions about loan approvals, and minimize the management risk by the accuracy associated with the extracted rules. So, instead of analysing the data set based on 16 features, which yield an intractable process for extracting meaningful patterns, only the CORE set that entails 3 features is used. Moreover, the Decision tree approach has been implemented on the same data set, the gain ratio for each attribute has been calculated and the C4.5 classifier has been used in the classification process. The gain ratios showed that the “Duration” feature has the maximum gain ratio, however, the “age” feature was the 8th in the gain, and the “balance” was the 10th. DT also provides a huge number of decision and predictive rules using the 16 features associated with accuracy. Some of the DT extracted rules are more summarized than those rules extracted by the RST and others are not. Therefore, despite the decision tree being easy to be implemented as a classifier, and the hardness of the rough set theory implementation, RST yields a better summarization of the data set due to the feature reduction process that achieves the best minimal set of features to describe the information space and preserve its approximations.

V. CONCLUSIONS

By taking the help of these previous research papers we got to know that all the attributes were not necessary for the prediction analysis. The Decision Tree Classifier helps us get better results compared to the other machine learning algorithms. We also added an extra feature in our project where we provided the user with a simple website where the user can put the important attributes of the customer and the machine learning algorithm which is used as a pickle file at the backend, helps the user the predict whether the customer will subscribe to the term-deposit or not.

REFERENCES

- [1] International Journal of Computer Applications (0975 – 8887), Volume 110 – No. 3, January 2015 Deposit subscribe Prediction using Data Mining Techniques based Real Marketing Dataset.
- [2] Advanced-Data Analysis, Department of Statistics, Columbia University, Who Will Subscribe a Term Deposit?



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)