

Group 374: Bank Deposit Prediction Model

Your submissions:

- Group374_Report.pdf
- Group374_Codes.ipynb (with necessary comments)

Notes

- No extension to the deadline
- Each team can only submit one copy by a single member, just list all of your members in the report
- use RED font for the parts that you revised according to the feedbacks in your presentation

First Name	Last Name	Email address
Aarish Salam	Memon	amemon2@hawl.iit.edu

Table of Contents

1. Introduction	2
2. Data	2
3. Problems and Solutions	2
4. KDD	2
4.1. Data Processing	2
4.2. Data Mining Methods and Processes	2
5. Evaluations and Results	3
5.1. Evaluation Methods	3
5.2. Results and Findings	3
6. Conclusions and Future Work	4
6.1. Conclusions	4
6.2. Limitations	4
6.3. Potential Improvements or Future Work	4

1. Introduction

There has been a revenue decline for the Portuguese bank and they would like to know what actions to take. After investigation, we found out that the root cause is that their clients are not depositing as frequently as before. I had to create prediction model to analyze the data and suggest them future work about what customers to focus on. **It will create an automated process for them to predict which customer is going to subscribe a deposit in the future.**

2. Data

I've picked a dataset from Kaggle Titled "Bank marketing campaigns dataset | Opening Deposit" which has 21 columns having different data types, The dataset has almost 42,000 records. Label is named as 'y' which defines if the deposit has been 'yes' or 'no'.

3. Problems and Solutions

Problem is that the bank is not getting deposits and they want to focus in the right direction for their marketing channeling process.

Solution would be to create an automation model, like whenever the new person comes in to the bank, they can enter the information which is the features and they can predict if the person would subscribe for a deposit or not so they can offer them royalty programs, discounts, and etc.

4. KDD

4.1. Data Processing

Encoding, Frequency Encoding, Ordinal Number Encoding, Dummies Creation for Data Type Conversion

Standard Scaler for Normalization

Feature Selection and Reduction, Extra Trees Classifier Used as a method and Impurity Metric Used to measure the quality of a split. Supported criteria are "gini" for the Gini impurity

Outlier Detection and Removal, Local Outlier Factor used for Outlier Detection and Removal.

Imbalance Issue and Solution, Dataset has extreme imbalance and issue and I solved it by using Random Over Sampler.

4.2. Data Mining Methods and Processes

KNN Classifier

Logistic Regression

SVM Classifier

Naïve Bayes Classifier

Decision Trees

Random Forest

Neural Networks

K-Means Clustering

5. Evaluations and Results

5.1. Evaluation Methods

Accuracy

Precision

Recall

F-1 Score

N-Folds Evaluation

5.2. Results and Findings

KNN Classifier – Accuracy of 88%

Logistic Regression – Accuracy of 84%

SVM Classifier – Accuracy of 61%

Naïve Bayes Classifier – Accuracy of 73%

Decision Trees – Accuracy of 73%

Random Forest – Accuracy of 49%

Neural Networks – Accuracy of 82%

KNN Classifier after Clustering – Accuracy of 90%

6. Conclusions and Future Work

6.1. Conclusions

Best Model and Setting would be KNN Classifier after the clustering process which gave us an accuracy of 90%. Any future data we add in the dataset and don't have the label for would be predicting the label for us though it won't be 100% correct but in the field of science 90% is a really good accuracy to predict.

6.2. Limitations

When it comes to data and prediction of real humans there is always a chance of error as human psychology is involved and there are always chances of errors in the prediction, as of now the limitation would be that but as the time passes and dataset increases the chances of error would be reduce as we would prune the model accordingly.

6.3. Potential Improvements or Future Work

Future work would be to connect the models with marketing platforms, ads and campaigns so when the predictions are out and if the prediction is for deposit as 'yes' they are included in the list of potential customers or targeted audiences, also as a part of this project we did the analysis on the past data which was in our training and that will help the bank to overcome the challenges they had in past.