Bank Marketing Effectiveness Prediction

V. Bhavya Reddy

***Abstract*—We are given the data of direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe to a term deposit (target variable y). This project will determine the success of Bank Telemarketing.**

The goal of this project is to predict whether someone will make a deposit based on the given attributes. We will try to build five models using different algorithms - Logistic Regression, Decision Tree, Random Forest, Naive Bayes, and K-Nearest Neighbors. The hyperparameters will then be tuned using GridSearch to optimise the model. Our next step will be to evaluate the metrics and compare each model to determine which model is most effective.

***Index Terms*—Logistic Regression, Decision Tree, Random Forest, Naive Bayes, and K-Nearest Neighbors.**

I. Problem Statement

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed. The classification goal is to predict if the client will subscribe a term deposit (variable y).

A. Input Variables

1) Bank Client data:

- age (numeric)
- job : type of job(categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
- marital : marital status(categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)
- education(categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
- default: has credit in default?(categorical: 'no', 'yes', 'unknown')
- housing: has housing loan?(categorical: 'no', 'yes', 'unknown')
- loan: has personal loan?(categorical: 'no', 'yes', 'unknown')

2) Related with the last contact of the current campaign:

- contact: contact communication type (categorical: 'cellular', 'telephone')
- month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
- day_of_week: last contact day of the week (categorical: 'mon', 'tue', 'wed', 'thu', 'fri')
- duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

3) Related with the last contact of the current campaign:

- contact: contact communication type (categorical: 'cellular', 'telephone')
- month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
- day_of_week: last contact day of the week (categorical: 'mon', 'tue', 'wed', 'thu', 'fri')
- duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

B. Other attributes

- campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
- pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
- previous: number of contacts performed before this campaign and for this client (numeric)
- poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

C. Output variable (desired target)

- y - has the client subscribed a term deposit? (binary: 'yes', 'no')

The provided data consists of over 45211 observations with 17 column variables.

II. Introduction

This project is approached in the following way:

We begin with the data set overview, in which we briefly analyze the observations and note several characteristics of numeric and categorical features. We take note of all the unique values present in every feature. In addition, we find the number of NULL values - none were found.

The next step is Exploratory Data Analysis (EDA) on the dataset, where we analyze numeric and categorical features through bar charts, box plots, countplots, etc. We also use heatmaps to analyze correlation, and VIF calculations are carried out to check for multicollinearity.

In feature Engineering, we drop unnecessary columns, modify the dataset, and remove unwanted observations. Later, Data encoding followed by train-test split is performed on the data set.

With the train data we train our models using Logistic regression, Decision Tree, Random Forest Classifier, K Nearest Neighbours, Naive Bayes algorithms. Several hyperparameters are tuned to optimize the performance. We try to find the most suitable model by comparing the evaluation metrics of various models. We found that the Naive Bayes Model had the best performance. Worst performance is seen by Random Forest Classifier. In order to gain a better understanding of the data, several plots are drawn (Accuracy, Precision, Recall, ROC-AUC, etc.).

III. Exploratory Data Analysis

Before we do the modelling, let's look at the important features and their influence on the Term Deposit.

A. Target Variable - Term Deposit

Signifies if the client subscribed a term deposit or not. The target column (Term Deposit) is a binary categorical column with 'yes' or 'no' values. The number of 'no' is 8 times more than 'yes'. The output 'term deposit' is moderately imbalanced.

B. Numeric features

Out of the 17 features 7 are numeric. We shall discuss the important numeric features.

1) Age: Adults under 60 are only 10.79% likely to subscribe compared to those over 60 with 33.63%.

C. Categorical features

Out of the 17 features 10 are categorical. We shall discuss the important categorical features.

1) Job: Below is a table that shows the impact of various jobs on term deposits.

01	Admin., blue-collar, management, services, technician	• More subscribers. • salaried positions
02	Student, Retired	• More likely to subscribe
03	Entrepreneur, Self-employed, housemaid, unemployed	• Entrepreneur, self-employed - not interested in term deposit as they like to spend it in their own business.

2) Marital: The campaign did better with single clients, despite the fact that there are more married clients.

	marital	y
0	divorced	11.945458
1	married	10.123466
2	single	14.949179

3) Housing & loan: Those with housing loans over personal loans are most likely to be influenced by the campaign.

D. Observations so far

- People in salaried positions such as admin, service, technicians, blue collar jobs, etc. subscribe to term deposits more than entrepreneurs and self-employed individuals.
- More than married people, single people are proportionally more interested.
- People having personal loans subscribe lesser than people with housing loan.
- Retired people and students are proportionally more inclined to the subscribe for the term deposit.

IV. Feature Engineering

Feature engineering involves the following steps.

A. Dropping columns

- We could never know how many calls it takes (campaign) so we will drop the campaign column.
- We could also not know how long it takes(duration), hence, we drop duration column.
- The effect of month is random so we drop 'month'.
- Effect of day is also random so we drop 'day'.

B. Delete observations

Columns for job, education, contact, and outcome contain a 'unknown' parameter, which should be removed.

C. One-Hot encoding

- One-hot encoding is performed on default, housing, loan, contact.
- Dummies encoding is performed on job, poutcome, education, and marital.

D. Test Train Split

80-20 split with 80% of the rows belonging to training data.

V. Evaluation Metrics

Let's have a brief idea on the several evaluation metrics used in the project.

- $Accuracy = \frac{TP+TN}{TP+FP+TN+TP}$
- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+NP}$
- $F1score = \frac{2xRecallxPrecision}{Recall+Precision}$

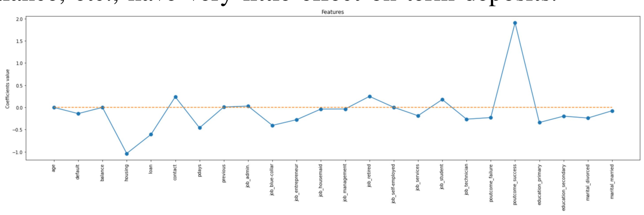
I consider AUC metric to be the most reliable metric to predict term deposit performance. This is because we have more 'no's than 'yes'es in our data, so both TPR and FPR are taken into account.

VI. Modelling

After the feature engineering, we split the data into Train set and Test set. Using various evaluation metric, we compare models and select the algorithm based on the AUC score on the Test data.

A. Logistic Regression

Age, married people, self-employed, housemaid, admin, balance, etc., have very little effect on term deposits.



Those who are likely to subscribe term deposit are predicted to be unwilling. So, recall will be poor. Although it appears that the model did a good job predicting non-subscribers, it failed to get the correct number of predictions for subscribers.

Evaluation Metrics	Value
Train Accuracy:	0.8260800255061375
Train Precision:	0.6886160714285714
Train Recall:	0.43177046885934217
Train F1 Score:	0.530752688172043
Train Confusion Matrix:	[[4565 279] [812 617]]
Test Accuracy	0.8253664754620778
Test Precision:	0.6912442396313364
Test Recall:	0.42016806722689076
Test F1 Score:	0.5226480836236934
Test Confusion Matrix:	[[1145 67] [207 150]]

Recall of 0.42 indicates that the model failed to predict subscribers properly. The train data or the test data do not differ greatly in their metrics.

1) Area Under Curve: In this project we consider AUC to be an important metric.

AUC on train data	0.813
AUC on test data	0.796

AUC of 0.796 is quite decent for the first model.

B. Decision Tree

We tune the hyperparameter Max_depth to 5 and train the model.

The AUC performance metrics is given in the below table.

AUC on train data	0.817
AUC on test data	0.786

We shall notice that AUC didn't improve.

1. Recall for decision tree has surely improved over logistic regression.

2. Decision trees still outperforms logistic regression models even though precision has decreased because recall has improved and I believe recall should be more important than precision.

C. Random Forest Classifier

We tune the hyperparameter criterion to 5, max_depth to 6, n_estimators to 200 and train the model.

The AUC performance metrics is given in the below table.

AUC on train data	0.836
AUC on test data	0.796

1. While AUC is the same as decision tree, other metrics such as recall and precision have decreased.

2. Random Forest is surely not the right model for this project.

D. KNN

We tune the hyperparameter N_neighbors = 40 with GridSearchCV of 5-fold cross validation to train the model with training dataset.

The AUC performance metrics is given in the below table.

AUC on train data	0.672
AUC on test data	0.611

1. As we know that our data is imbalanced and this is a classic example to show that the performance of k-NN classifiers will be significantly impacted by the imbalanced class distributions of data.

2. This model is by far the worst one.

E. Naive Bayes

Evaluation Metrics	Value
Train Accuracy:	0.779212498007333
Train Precision:	0.5126728110599078
Train Recall:	0.622813156053184
Train F1 Score:	0.5624012638230647
Train Confusion Matrix:	[[3998 846] [539 890]]
Test Accuracy	0.7807520713830465
Test Precision:	0.5158150851581509
Test Recall:	0.5938375350140056
Test F1 Score:	0.5520833333333334
Test Confusion Matrix:	[[1013 199] [145 212]]

The AUC performance metrics is given in the below table.

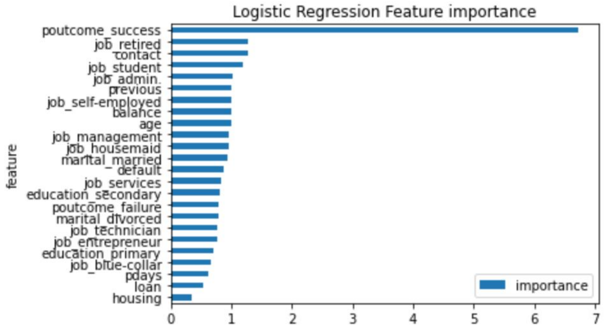
AUC on train data	0.786
AUC on test data	0.766

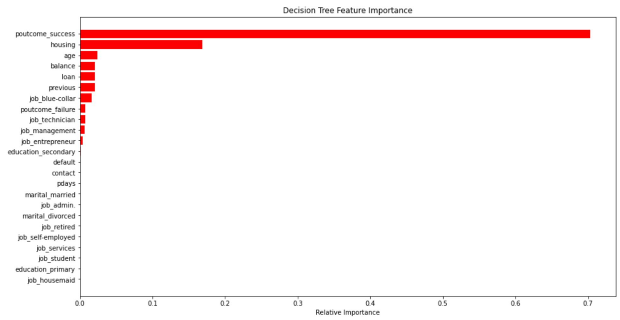
Even though AUC has not improved, we see an improvement in recall and precision which makes Naive Bayes the best model.

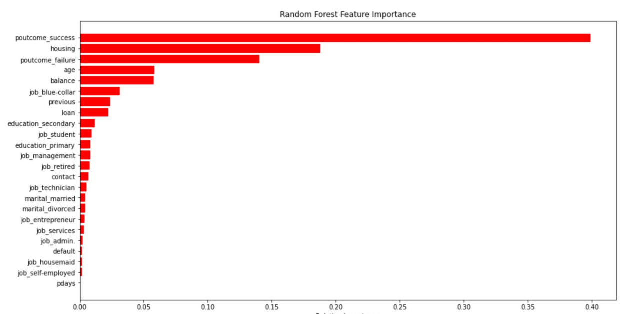
VII. Model Comparison

Attributes such as poutcome, loan, age, balance, job are important in predicting the term deposit.

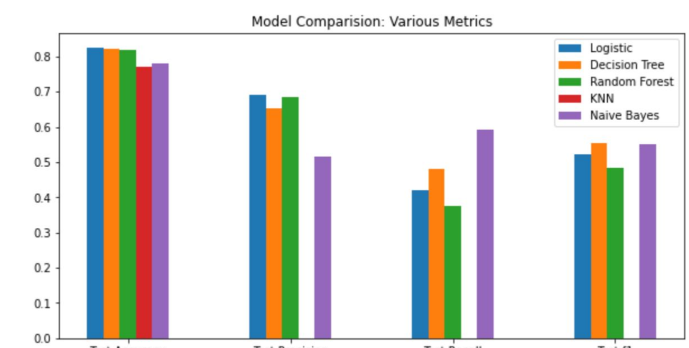
Evaluation Metrics	Logistic Regression	Decision Tree	Random Forest	KNN	Naive Bayes
Train Accuracy:	0.8260800255061375	0.8300653594771242	0.8289494659652479	0.7725171369360753	0.779212498007333
Train Precision:	0.6886160714285714	0.667590027700831	0.7230576441102757	1.0	0.5126728110599078
Train Recall:	0.43177046885934217	0.5059482155353394	0.40377886634009796	0.0013995801259622112	0.622813156053184
Train F1 Score:	0.530752688172043	0.5756369426751592	0.5181859003143242	0.0027952480782669456	0.5624012638230647
Train Confusion Matrix:	[[4565 279] [812 617]]	[[4484 360] [706 723]]	[[4623 221] [852 577]]	[[4844 0] [1427 2]]	[[3998 846] [539 890]]
Test Accuracy	0.8253664754620778	0.8234544295729764	0.8183556405353728	0.7724665391969407	0.7807520713830465
Test Precision:	0.6912442396313364	0.6515151515151515	0.6836734693877551	0.0	0.5158150851581509
Test Recall:	0.42016806722689076	0.48179271708683474	0.3753501400560224	0.0	0.5938375350140056
Test F1 Score:	0.5226480836236934	0.5539452495974235	0.484629294755877	0.0	0.5520833333333334
Test Confusion Matrix:	[[1145 67] [207 150]]	[[1120 92] [185 172]]	[[1150 62] [223 134]]	[[1212 0] [357 0]]	[[1013 199] [145 212]]

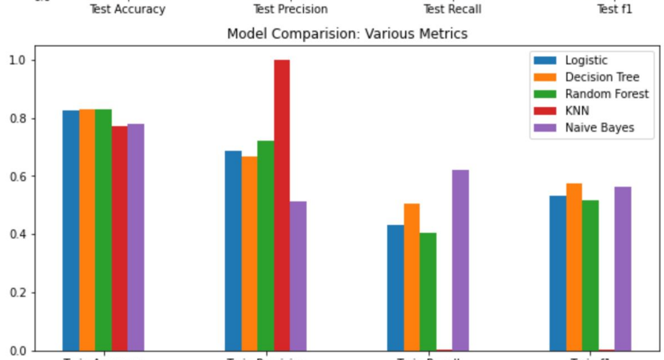






The KNN model has the lowest AUC score, and the other models exhibit similar results.





VIII. Challenges

- Huge chunk of data was to be handled keeping in mind not to miss anything of even little relevance.
- Feature engineering was quite challenging.
- Certain models took a long time to optimize hyperparameters.

Conclusion

- The key features or attributes that helped in the prediction of the term deposit were - poutcome, age, balance, previous, loan.
- KNN's prediction is heavily influenced by the majority class, so it seems to be the poorest model for imbalanced data.
- As for this project, I have considered the AUC parameter to be significant since it considers TPR and FPR. AUC scores are similar for other models, except for KNN. AUC for Naive Bayes is slightly lower than that of logistic regression, decision trees, and random forests.
- Naive Bayes remains the right fit for term deposits due to the recall score (59.4
- In this project, I considered recall to be more significant than precision. This assumption was made based on intuition.