Capstone Project-3

Bank Marketing Effectiveness Prediction

Team Members:
V. Bhavya Reddy

Problem Solution Brief

Our model predicts 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.

Approach

Data Set Overview

In this first step we will load our dataset, observe the dataset shape, brief analysis of the dataset, removing the NaN values, observe the variables data types, etc.

EDA

Target distribution, importance of various numeric and categorical features, heatmap for correlation, vif, etc.

Feature Engineering

Drop unnecessary features, removing unwanted observations, encoding, and train test split.

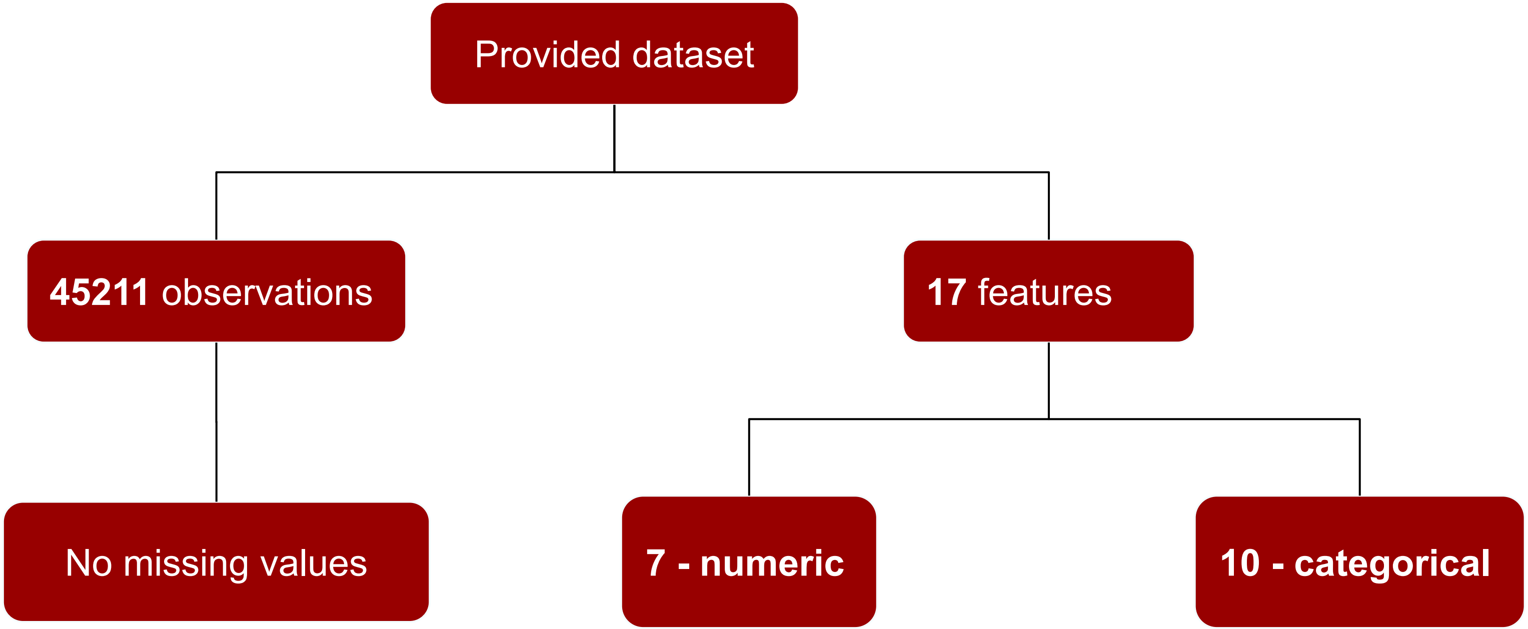
Modelling

Several models are designed using Logistic regression, Decision Tree, Random Forest Classifier, K Nearest Neighbours, Naive Bayes algorithms. Several hyperparameters are tuned to optimise the performance.

Model Comparison

In model comparison section, we try to find the most suitable model by comparing the evaluation metrics of various models.

Data Set Overview



Exploratory Data Analysis

These are highly imbalanced dataset!!

Target Variable - Term Deposit → probability(‘yes’)= 8 x probability(‘no’)

Numeric feature

Age:

Age

- Old people >= 60
  - Subscribed to Term Deposit = 33.63%
- Adults < 60
  - Subscribed to Term Deposit = 10.79%

Impact on old people is more than on adults.

Exploratory Data Analysis Cont..

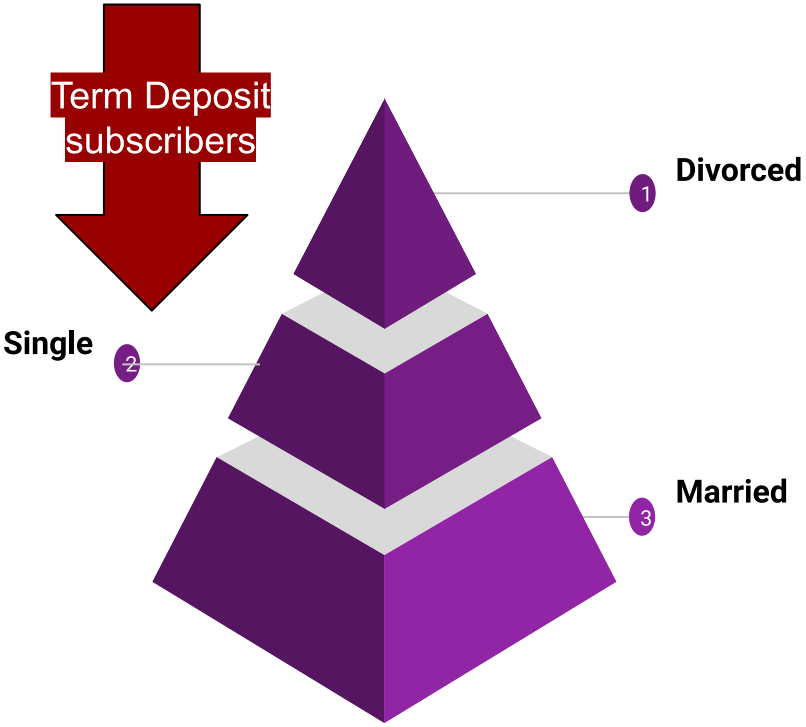
Categorical feature

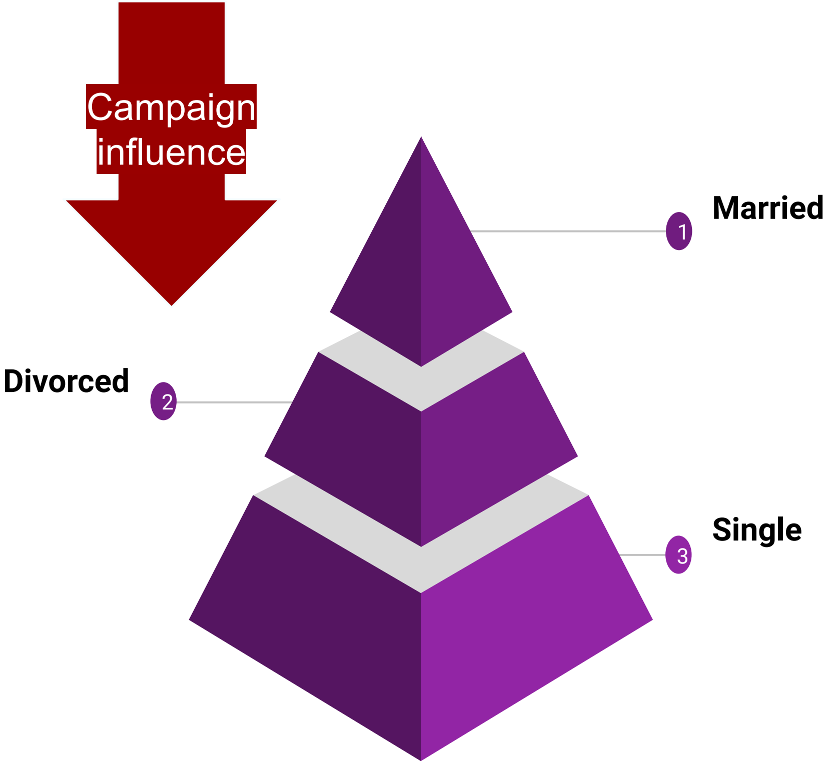
Job:

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.

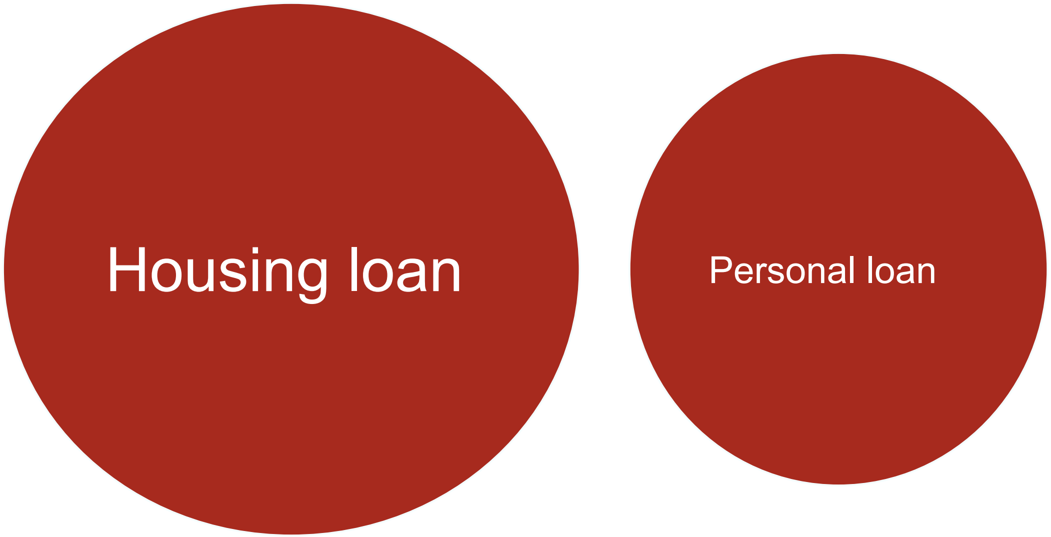
Exploratory Data Analysis Cont..

Marital :





Exploratory Data Analysis Cont..



Those with housing loans over personal loans are most likely to be influenced by the campaign.

Feature Engineering

1. **Dropping 'duration', 'campaign', 'month', 'day' columns.**
   We could never know how many calls it takes (campaign) and how long it takes(duration) to make target value to yes. Effect of month ,and day seems random.

2. **Delete observations having 'unknown' .**
   Columns for job, education, contact, and outcome contain a 'unknown' parameter, which should be removed.

3. **One-Hot encoding**
   One-hot encoding on default, housing, loan, contact. Dummies encoding on job, poutcome, education, and marital.

4. **Test Train Split**
   80-20 split with 80% of the rows belonging to training data.

Feature Engineering Cont..

	age	default	balance	housing	loan	contact	pdays	previous	job_admin.	job_blue-collar	...	job_self-employed	job_services	job_student	job_technician	poutcome_failure
31160	35	0	2823	1	0	1	1	1	0	0	...	0	0	0	1	1
34803	40	0	-606	1	0	1	1	1	0	0	...	0	0	0	0	1
40055	42	0	2665	1	0	1	1	11	0	0	...	0	0	0	0	1
40103	34	0	303	0	0	1	1	3	1	0	...	0	0	0	0	1
44773	31	0	147	0	0	1	1	5	0	0	...	0	0	0	0	0
36698	48	0	162	1	0	1	1	4	1	0	...	0	0	0	0	1
30525	30	0	436	1	0	1	1	8	0	0	...	0	0	0	1	0
33978	57	0	478	0	0	1	1	1	0	0	...	0	0	0	0	1
42827	65	0	1973	0	0	0	1	3	0	0	...	0	0	0	0	0
43224	30	0	201	1	0	1	1	13	0	1	...	0	0	0	0	0

Train- Test split:
This is train data.

Train Set : (6273, 24)
Test set: (1569, 24)
Response: 0/1

During train-test split the term deposit was stratified.

Evaluation Metrics

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

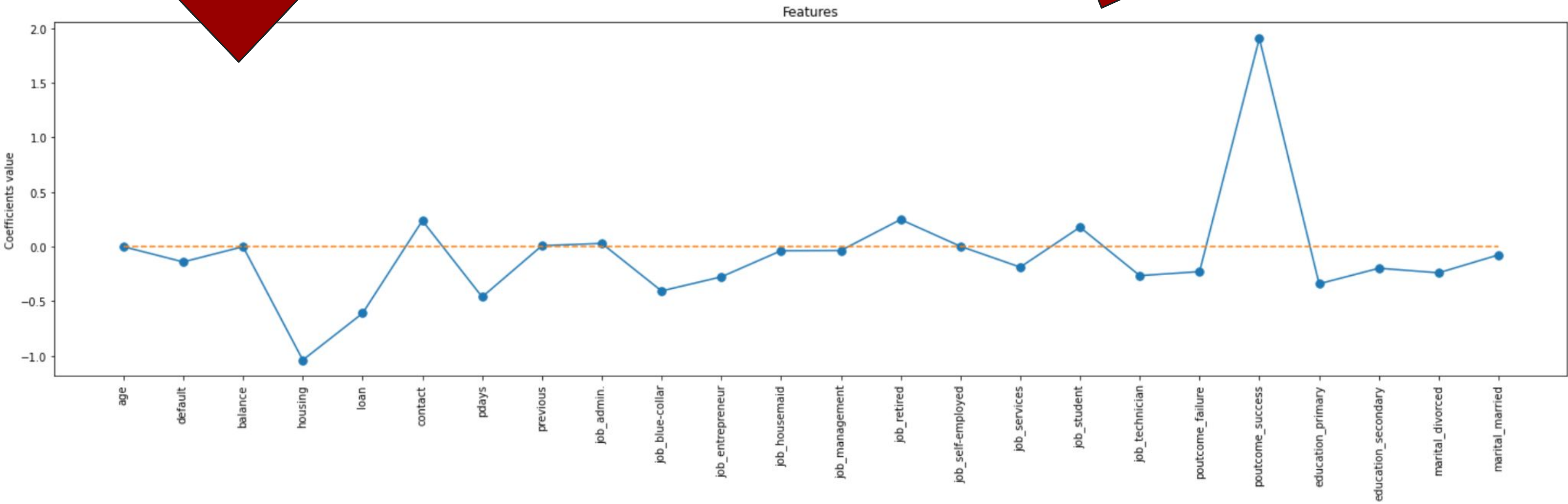
1. Accuracy = (TP+TN)/(TP+FP+TN+TP)
2. Precision = TP/(TP+FP)
3. Recall = TP/(TP+FN)
4. F1_Score = 2*Recall*Precision/(Recall + Precision)

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.

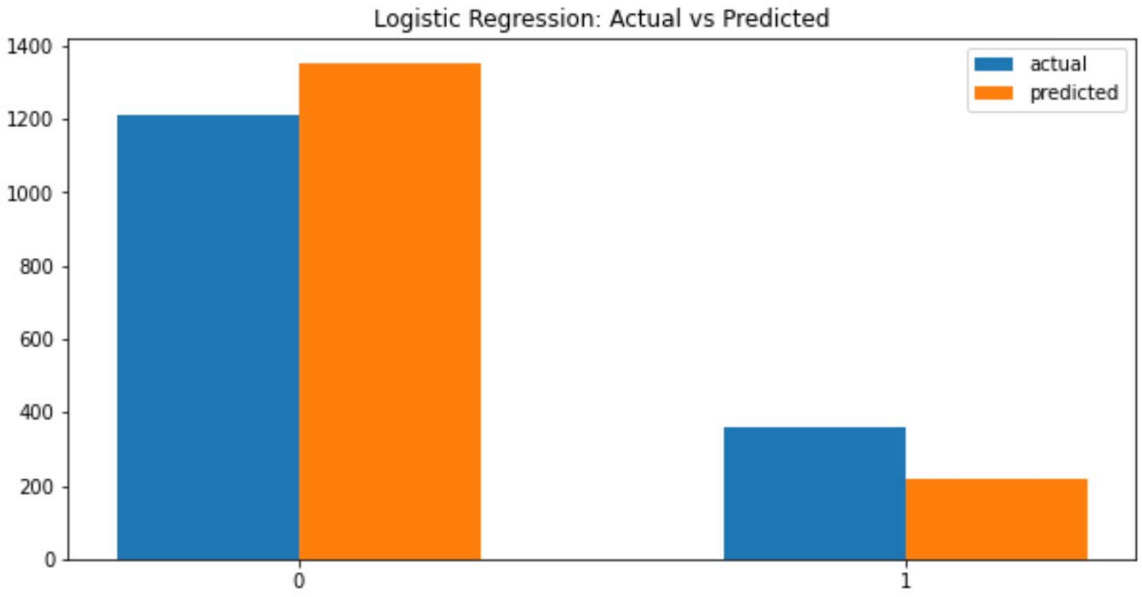
Logistic Regression

Based on the plot, we can see that age, married people, self-employed, housemaid, admin, balance, etc., have very little effect on term deposits.

The coefficients- w0, w1, w2, … for every feature.



Logistic Regression Cont..



Those who are likely to subscribe term deposit are predicted to be unwilling. So, recall will be poor.

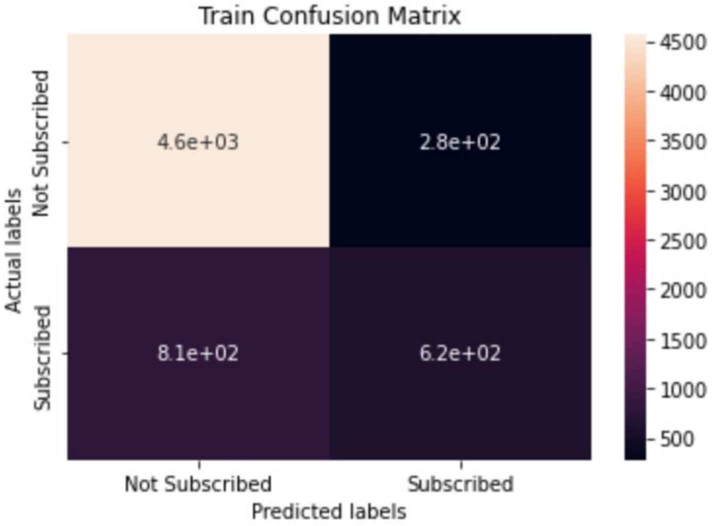
Logistic Regression Cont..

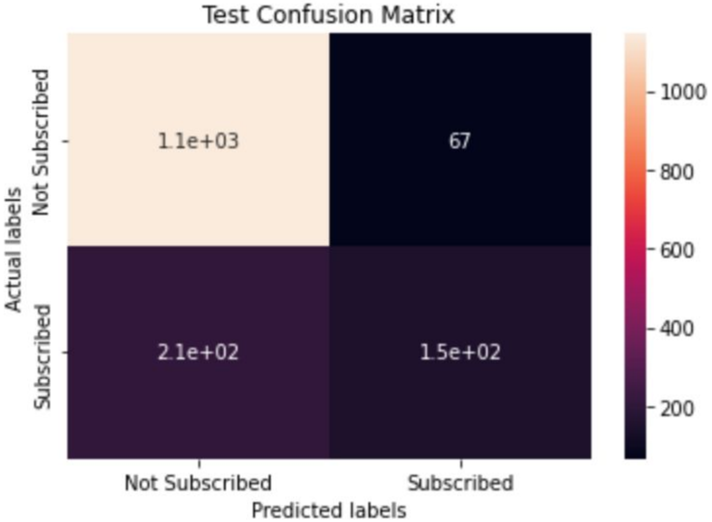
Evaluation Metrics	Value
Train Accuracy:	0.8260800255061375
Train Precision:	0.6886160714285714
Train Recall:	0.43177046885934217
Train F1 Score:	0.530752688172043
Train Confusion Matrix:	[[4565 279] [812 617]]
Test Accuracy	0.8253664754620778
Test Precision:	0.6912442396313364
Test Recall:	0.42016806722689076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

Test AUC: 0.796
Train AUC: 0.813

Although it appears that the model did a good job predicting non-subscribers, it failed to get the correct number of predictions for subscribers.

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.

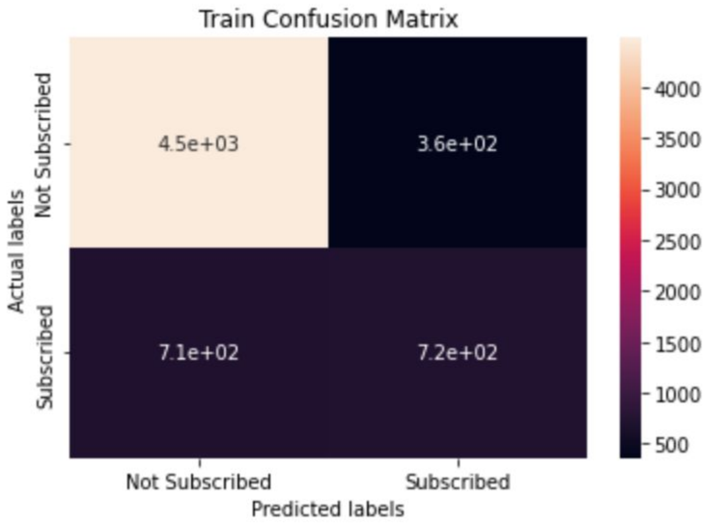




Decision Tree

Best estimators:
1. Criterion : ‘entropy’
2. Max depth : 5

Test AUC: 0.786
Train AUC: 0.817



OBSERVATION:
1. Recall for decision tree has surely improved over logistic regression. However, precision has reduced.
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.

Random Forest Classifier

BEST PARAMETERS:

criterion: entropy
max_depth: 6
n_estimators: 200

Test AUC: 0.796
Train AUC: 0.836

Evaluation Metrics	Value
Train Accuracy:	0.8289494659652479
Train Precision:	0.7230576441102757
Train Recall:	0.40377886634009796
Train F1 Score:	0.5181859003143242
Train Confusion Matrix:	[[4623 221] [852 577]]
Test Accuracy	0.8183556405353728
Test Precision:	0.6836734693877551
Test Recall:	0.3753501400560224
Test F1 Score:	0.484629294755877
Test Confusion Matrix:	[[1150 62] [223 134]]

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.

KNN

Best parameter:

N_neighbors = 40

Test AUC: 0.611
Train AUC: 0.672

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 worst one so far.

Evaluation Metrics	Value
Train Accuracy:	0.7725171369360753
Train Precision:	1.0
Train Recall:	0.0013995801259622112
Train F1 Score:	0.0027952480782669456
Train Confusion Matrix:	[[4844 0] [1427 2]]
Test Accuracy	0.7724665391969407
Test Precision:	0.0
Test Recall:	0.0
Test F1 Score:	0.0
Test Confusion Matrix:	[[1212 0] [357 0]]

Naive Bayes

Test AUC: 0.766
Train AUC: 0.786

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

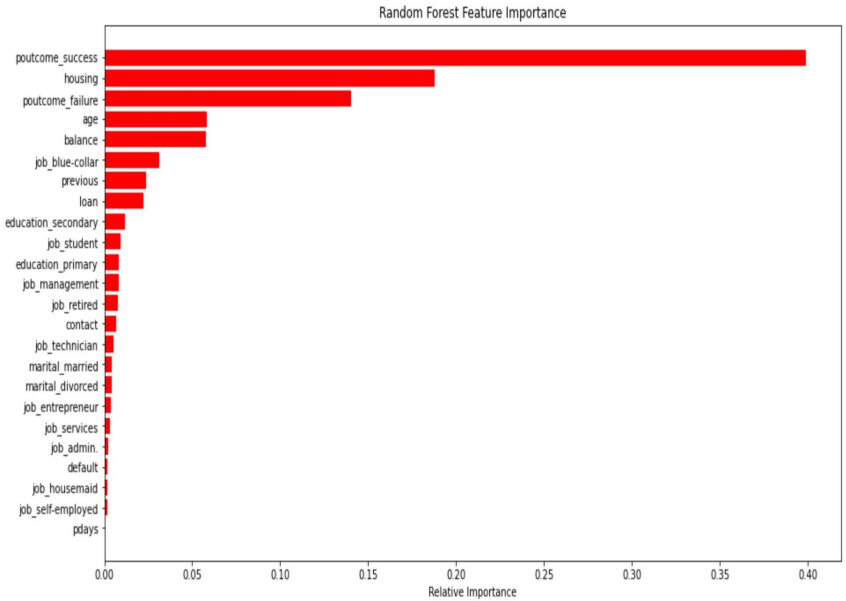
Evaluation Metrics	Value
Train Accuracy:	0.779212498007333
Train Precision:	0.5126728110599078
Train Recall:	0.622813156053184
Train F1 Score:	0.5624012638230647
Train Confusion Matrix:	[[3998 846] [539 890]]
Test Accuracy	0.7807520713830465
Test Precision:	0.5158150851581509
Test Recall:	0.5938375350140056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

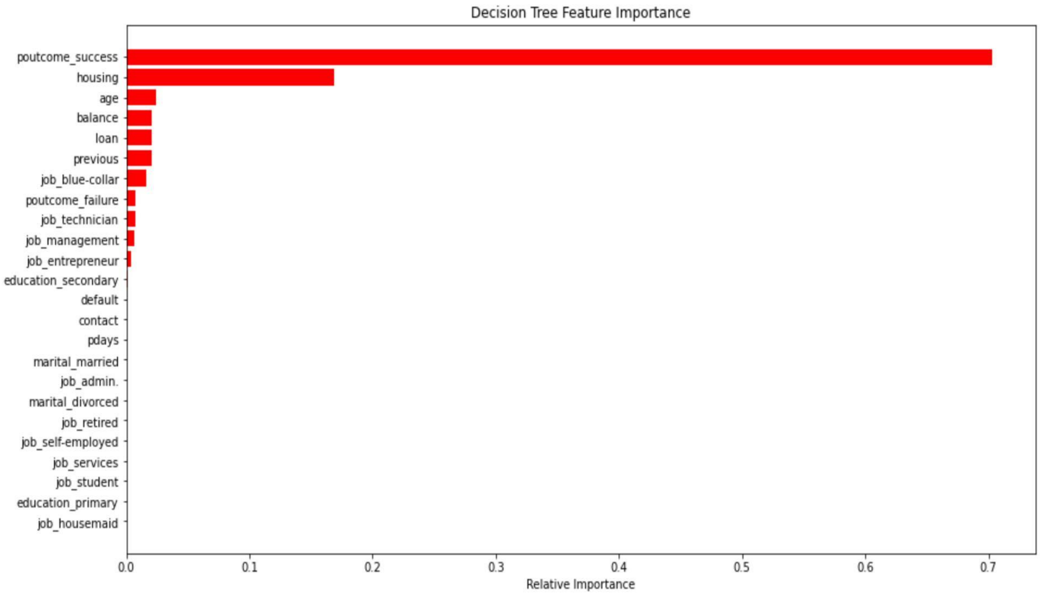
Comparison among several models using evaluation metrics

Models	Accuracy	Precision	Recall	F1 Score
Logistic Regression	✓	✓	✗	✓
Decision Tree	✓	✓	✗	✓
Random Forest	✓	✓	✗	✗
KNN	✓	✗	✗	✗
Naive Bayes	✓	✓	✓	✓

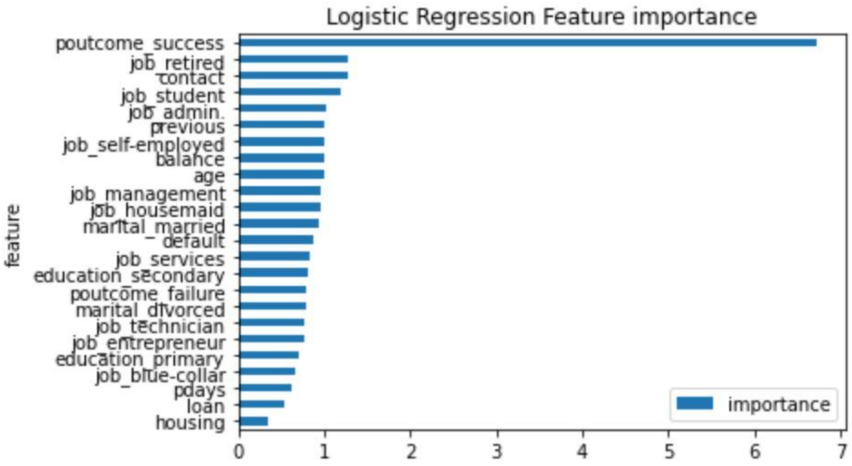
✓ - acceptable
✗ - not acceptable

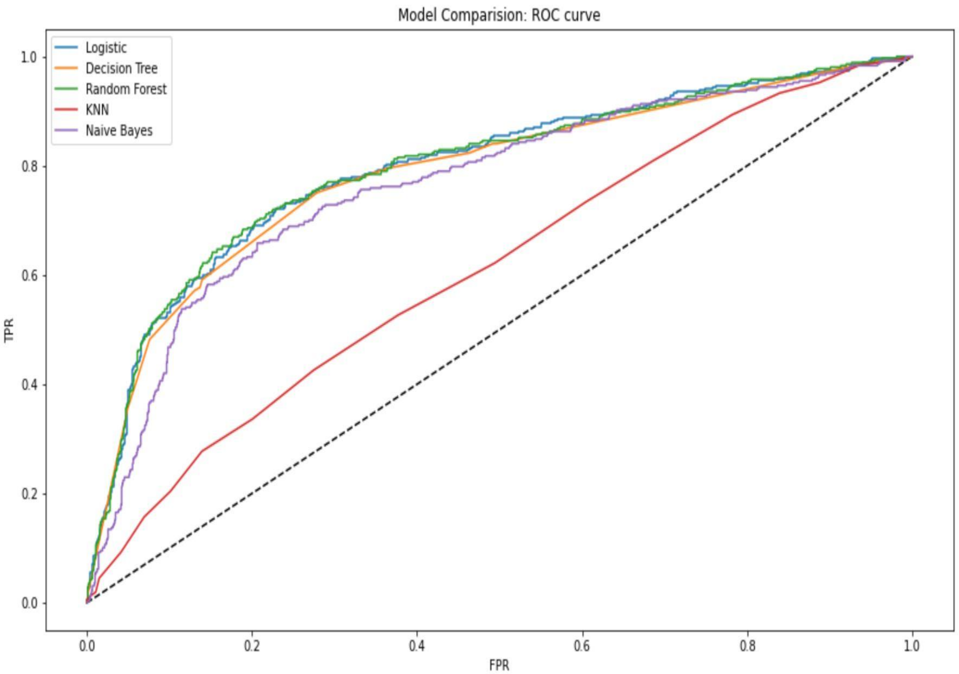
Evaluation Metrics	Logistic Regression	Decision Tree	Random Forest	KNN	Naive Bayes
Train Accuracy:	0.8260800255061375	0.8300653594771242	0.8289494659652479	0.7725171369360753	0.779212498007333
Train Precision:	0.6886160714285714	0.667590027700831	0.7230576441102757	1.0	0.5126728110599078
Train Recall:	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]]

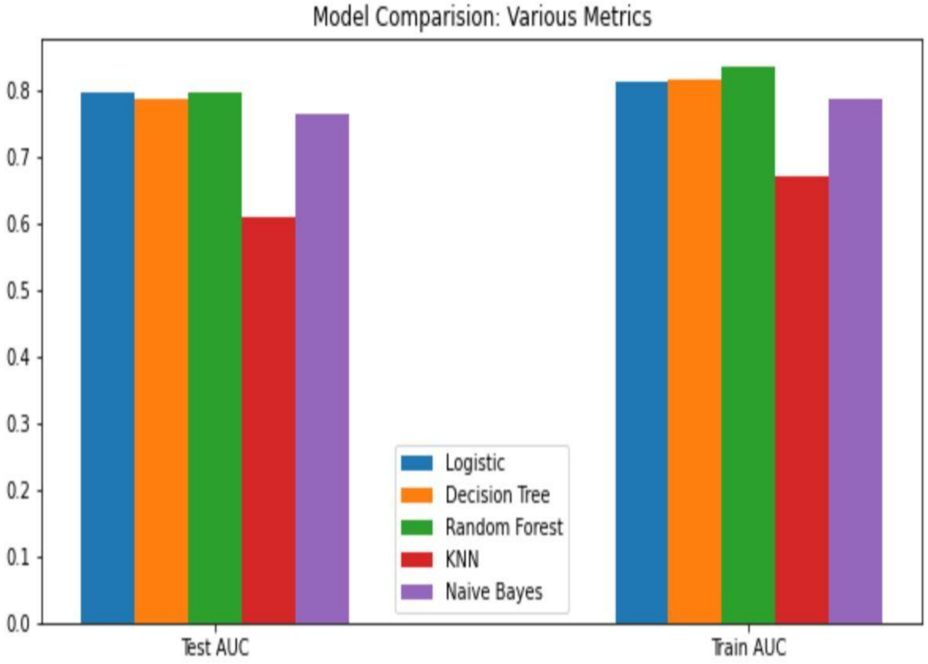




Attributes such as poutcome, loan, age, balance, job are important in predicting the term deposit.







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

Challenge

- 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

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