Direct Bank Marketing

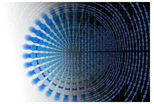
Jessica How
JinYao Huang
Aswathy Surendran

Overview

1. Problem Statement

2. Project Vision

3. Approach & Solution



4. Dataset

5. Implementation

6. Conclusion & Next steps

7. Github Accounts

Traditional Bank Marketing

Person to person

Prints

Broadcast

Direct mail

Phone

Outdoor advertising like billboards

Problem Statement

Does the Banks know who is the right target customer for a particular banking service or offering?

Key issues

Communication strategy

Customer morale

End results

Project Vision

Can Machine learning models predict the success of a customer conversion?

Can we suggest which communication method and product work for each customer?

Say for example deposits for customer 1 or other bank services for customer 2.

Proposed Model can increase the campaign efficiency:

Identifying the main characteristics that affect success

Better management of the available resources like effort, communication methods, time etc

Selection the right target customers who have the potential for conversion

Our Approach & Proposed Solution

- Optimize marketing campaigns steps:
    - Import data from dataset and perform initial high-level analysis:
        - number of rows,
        - missing values,
        - dataset columns
        - values respective to the campaign outcome
    - Clean the data:
        - Remove irrelevant columns
        - Deal with missing and incorrect values
        - Turn categorical columns into dummy variables
    - Use machine learning techniques to predict the marketing campaign outcome and to find out factors, which affect the success of the campaign

Our Approach & Proposed Solution contd..

Solution

Machine Learning and Classification model for campaign outcome prediction

Approach

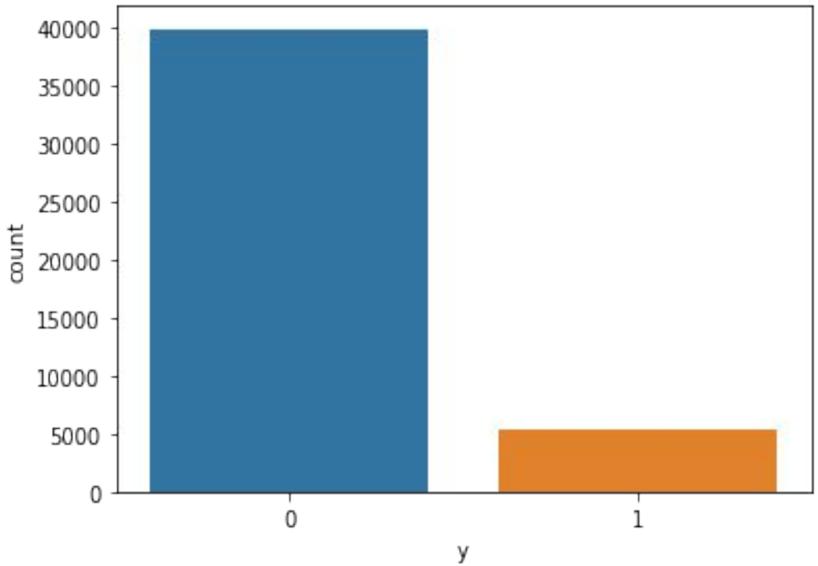
Dataset - We cleaned the datasets for prediction of campaign outcome with help of machine learning classification models

Library - XGBoost, machine learning library was used for modelling

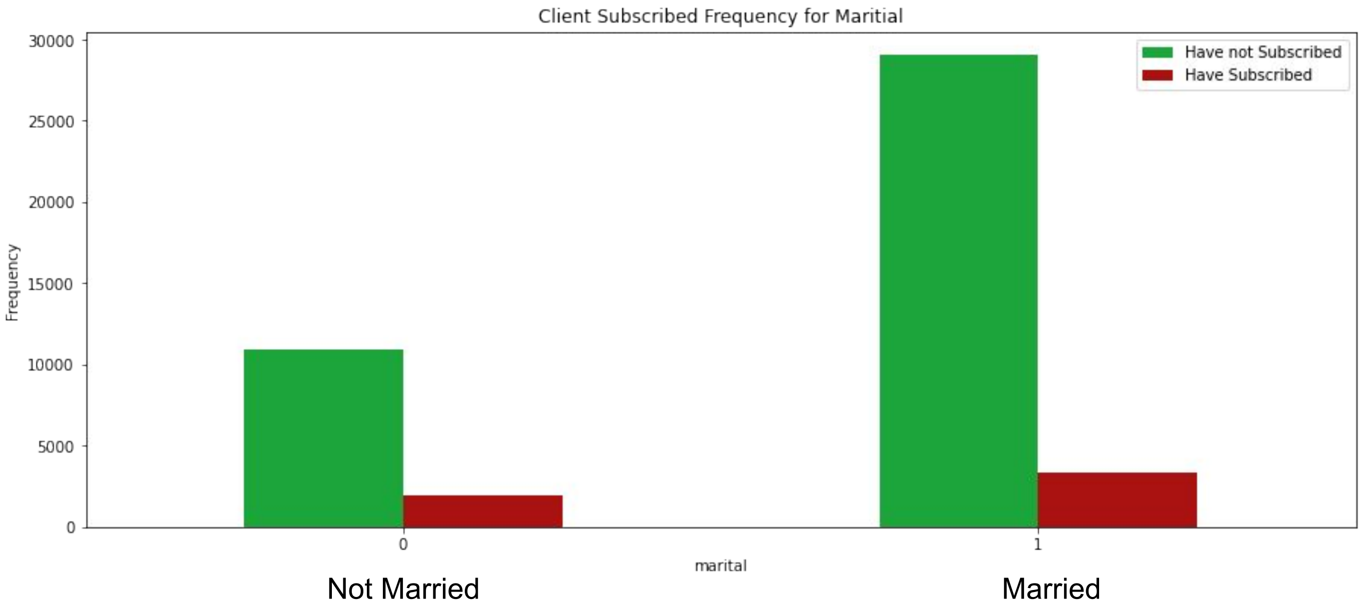
Resulting model - help to understand, which features have the greatest importance for the predicting the results of the campaign

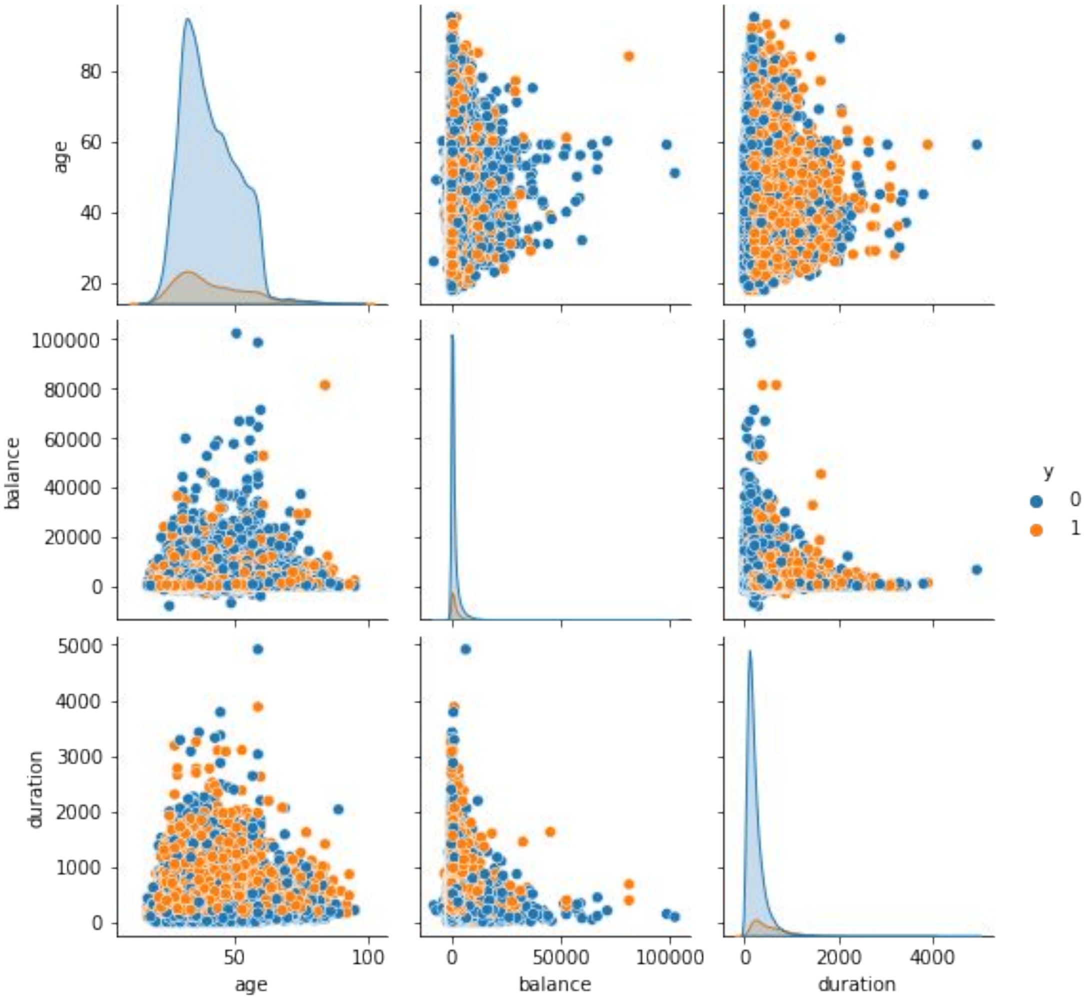
Dataset

- Used dataset from Kaggle
  - Contains 45,211 customers
  - 16 Output Attributes (X-Variables)
    - age, job, marital, education, default balance, housing loan, contact, day, month, duration, campaign, pdays, previous outcome, deposit
  - Y variable: The classification goal is to predict if the client will subscribe a term deposit (yes or no)
  - Information about a marketing campaign of financial institutions



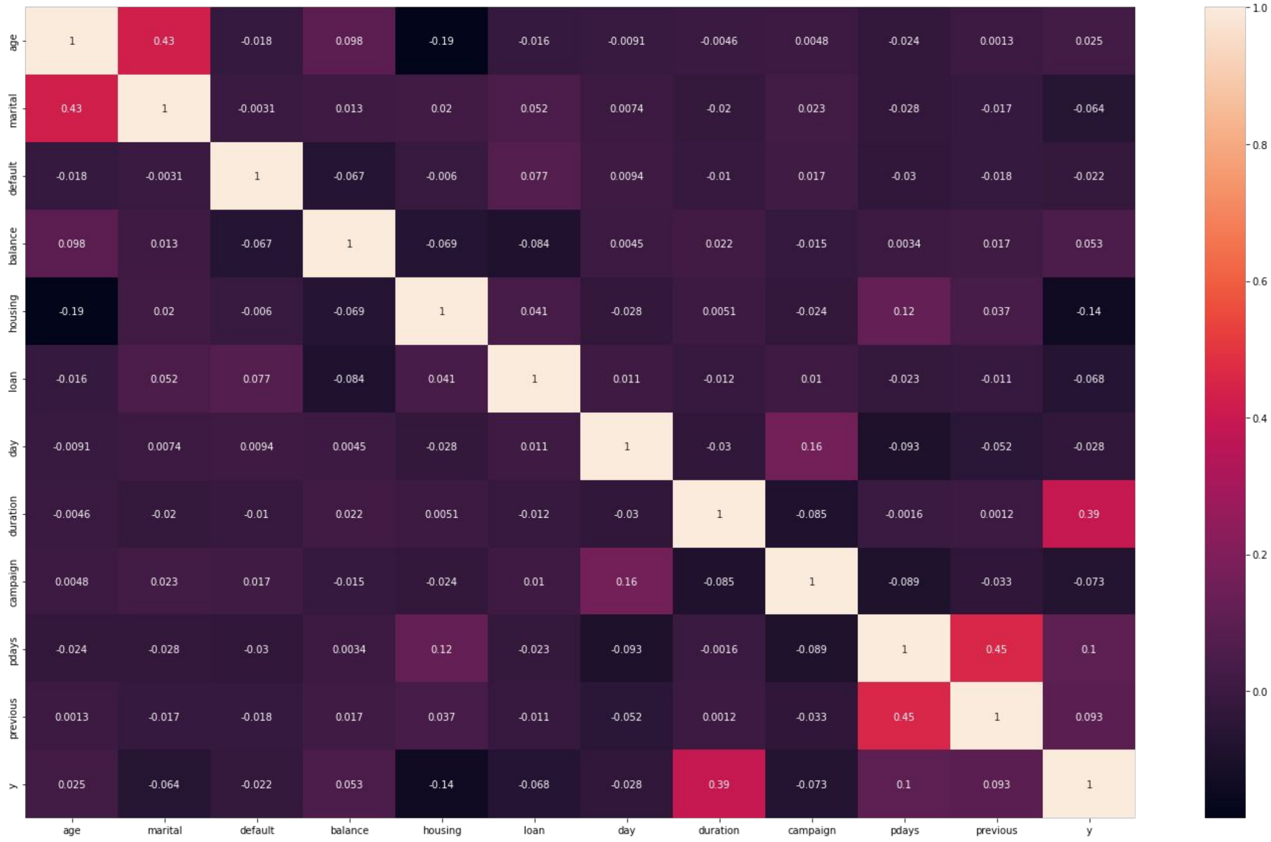
0 = Have not subscribed
1 = Have subscribed


Client Subscribed Frequency for Maritial
Have not Subscribed
Have Subscribed
Frequency
30000
25000
20000
15000
10000
5000
0
0
1
marital
Not Married
Married



age
balance
duration
y
0
1

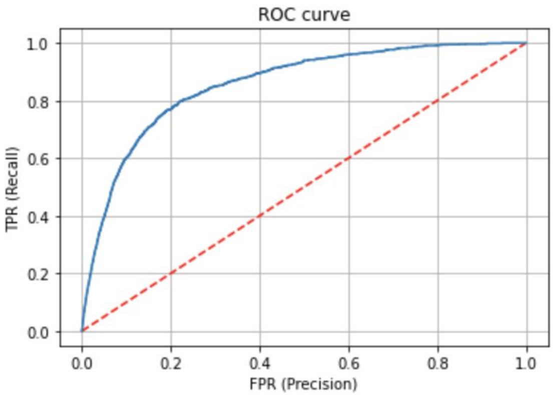
Correlation



Results - Logistic Regression vs. Decision Tree

```
Confusion matrix:
[[11759   247]
 [ 1231   327]]
```

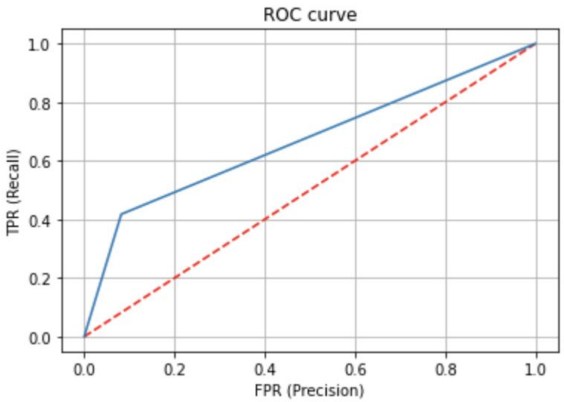
	precision	recall	f1-score	support
0	0.91	0.98	0.94	12006
1	0.57	0.21	0.31	1558
accuracy			0.89	13564
macro avg	0.74	0.59	0.62	13564
weighted avg	0.87	0.89	0.87	13564




```
Area under curve (AUC):  0.8570352714100801
```

```
Confusion matrix:
[[11019   987]
 [  907   651]]
```

	precision	recall	f1-score	support
0	0.92	0.92	0.92	12006
1	0.40	0.42	0.41	1558
accuracy			0.86	13564
macro avg	0.66	0.67	0.66	13564
weighted avg	0.86	0.86	0.86	13564




```
Area under curve (AUC):  0.6678172467039908
```

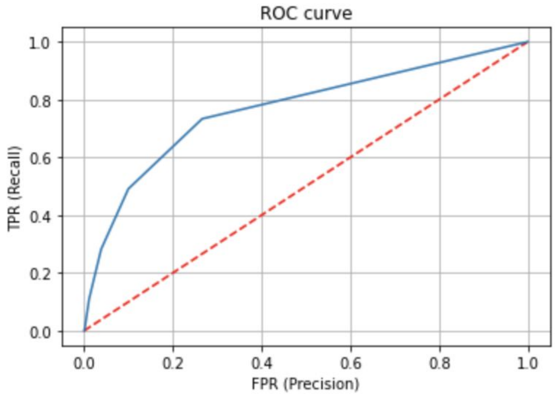
Results - KNN vs. Naive Bayes

```
Confusion matrix
 [[11544   462]
 [ 1119   439]]
```

```
              precision    recall  f1-score   support

           0       0.91      0.96      0.94     12006
           1       0.49      0.28      0.36      1558

    accuracy                           0.88     13564
   macro avg       0.70      0.62      0.65     13564
weighted avg       0.86      0.88      0.87     13564
```



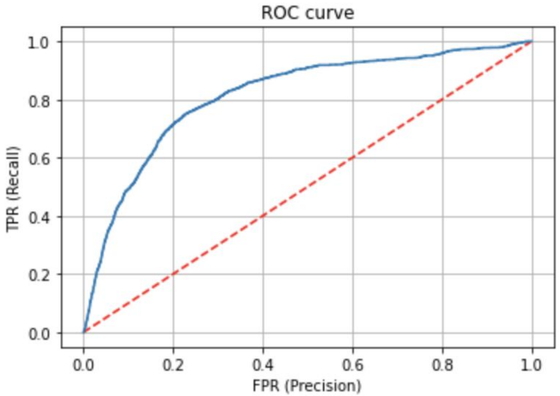

```
Area under curve (AUC):  0.7674716877761376
```

```
Confusion matrix
 [[11149   857]
 [  917   641]]
```

```
              precision    recall  f1-score   support

           0       0.92      0.93      0.93     12006
           1       0.43      0.41      0.42      1558

    accuracy                           0.87     13564
   macro avg       0.68      0.67      0.67     13564
weighted avg       0.87      0.87      0.87     13564
```




```
Area under curve (AUC):  0.8152877455153467
```

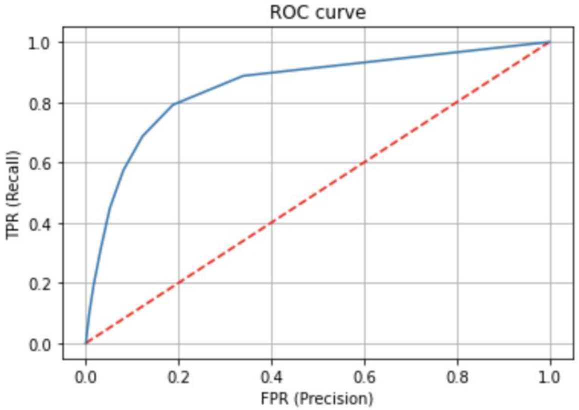
Results - Random Forest (Best model)

```
Confusion matrix
 [[11622   384]
 [ 1076   482]]

              precision    recall  f1-score   support

           0       0.92      0.97      0.94     12006
           1       0.56      0.31      0.40      1558

    accuracy                           0.89     13564
   macro avg       0.74      0.64      0.67     13564
weighted avg       0.87      0.89      0.88     13564
```




```
Area under curve (AUC):  0.8529759243185425
```

	Model	Accuracy
0	Logistic Regression	0.891035
1	Decision Tree	0.860366
2	KNN	0.883441
3	Naive Bayes	0.869213
4	Random Forest	0.892362

Conclusion & Next steps

Key outcomes of the analysis are the recommendations for future marketing campaigns:

The customer's account balance

has a huge influence on the campaign's outcome. People with account balance above 1490$ are more likely to subscribe for term deposit, so in future the bank needs to address those customers.

The customer's age

affects campaign outcome as well. Future campaigns should concentrate on customers from age categories below 30 years old and above 50 years old.

Number of contacts

with the customer during the campaign is also very important. The number of contacts with the customer shouldn't exceed 4.

Thanks

Github Accounts

GitHub Accounts

Jessica How - https://github.com/jessicahowzy

JinYao Huang - https://github.com/JinyaoHuang

Aswathy Surendran https://github.com/ASurendran