Bank Marketing Prediction

Hernan Trujillo
August 2021

01

Introduction

Introduction

The goal of this project is to predict if the client will subscribe a term deposit, based on phone calls of a dataset from a Bank’s marketing team.

Term Deposit Prediction

Data

Kaggle.com
41,188 Observations
20 Features

Client

Financial Institution

Campaign's Method

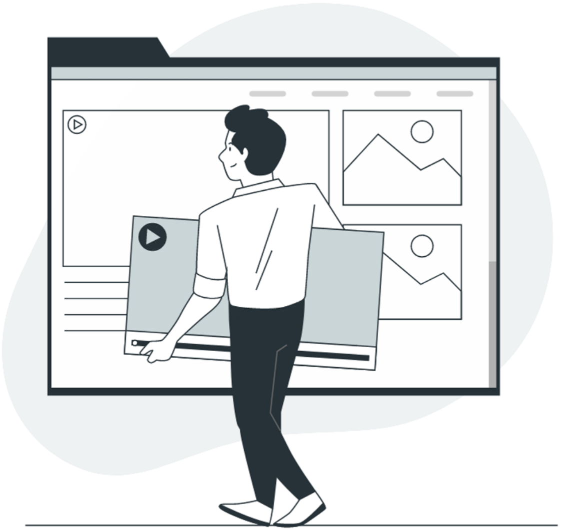
Direct marketing (phone calls) to current clients

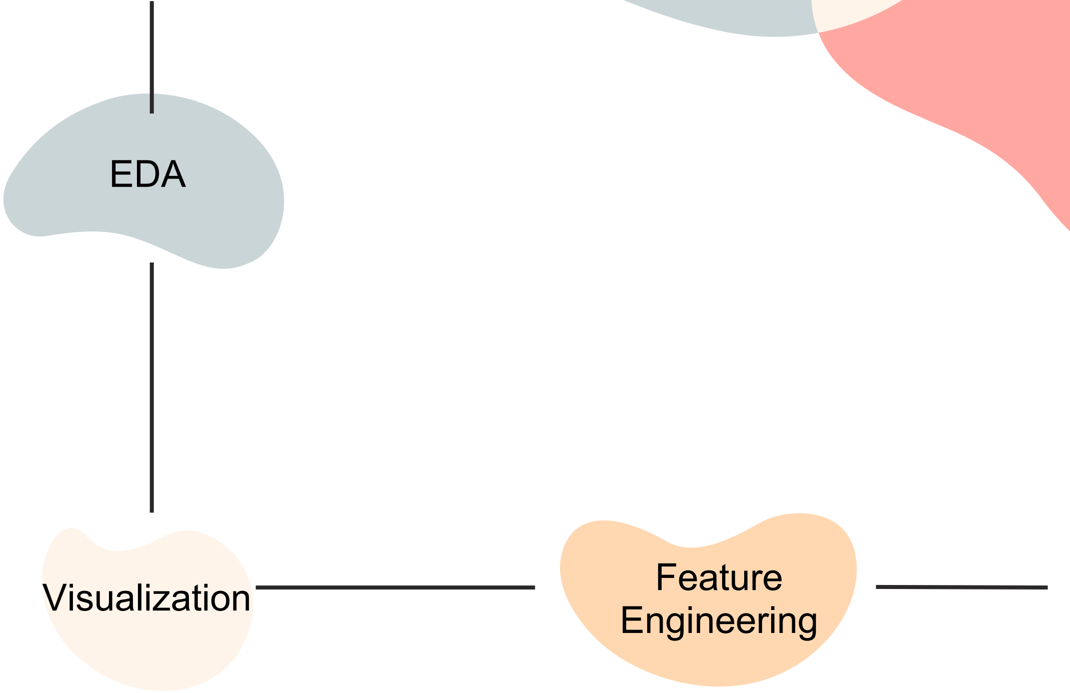
02

Approach

Classification Model

Classification Roadmap





Data Details

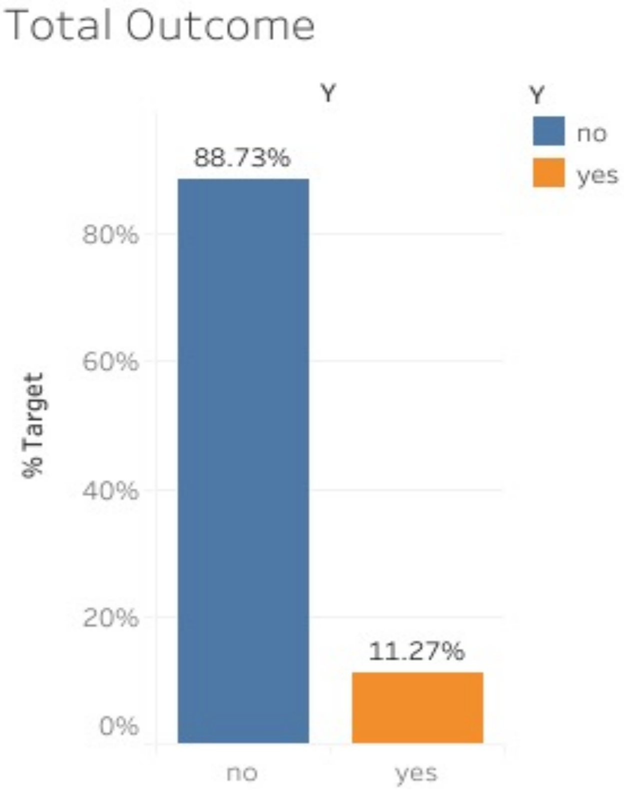
40+ Observations
20 Features

Conclusions

Processing

Modeling

Exploratory Data Analysis (EDA)



Target Variable Distribution:

Imbalanced

Metric Selection:

F1 Score-

Precision: Minimize False Positives

Predicting subscribers when they're not.

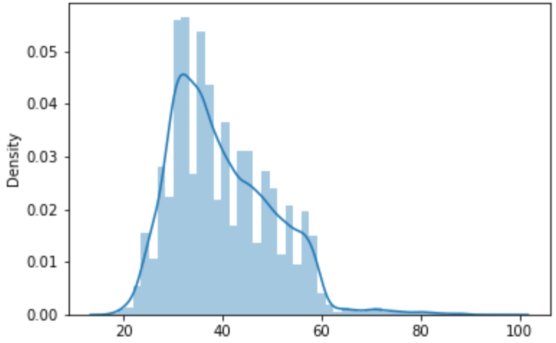
Recall: Minimizes False Negatives

Predicting not subscribers when they will

Feature Highly Correlations

Drop 'emp.var.rate' and 'nr.employed' as 'euribor' also give us the price of money in current market.

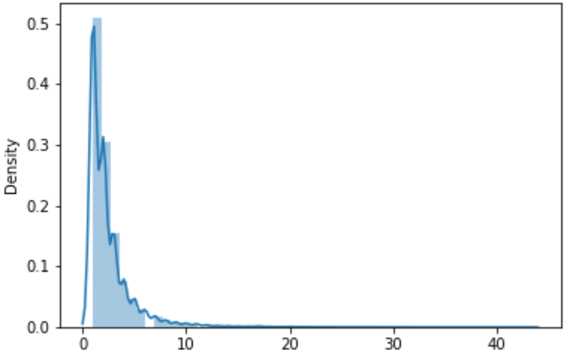
Feature Engineering
Outliers



Age

"age" and "campaign" are skewed towards right, we computed the Interquartile Range (IQR) based method and replaced the outliers with the lower and upper boundaries

Campaign



	age	campaign
count	38205.000000	38205.000000
mean	39.855935	2.567334
std	10.286042	2.768519
min	17.000000	1.000000
25%	32.000000	1.000000
50%	38.000000	2.000000
75%	47.000000	3.000000
max	98.000000	43.000000

Feature Engineering Encoding Categorical Features

```
['loan',
 'month',
 'contact',
 'job',
 'marital',
 'housing',
 'education',
 'poutcome',
 'day_of_week']
```

```
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   age              38205 non-null  int64
 1   job              38205 non-null  int64
 2   marital          38205 non-null  int64
 3   education        38205 non-null  int64
 4   housing          38205 non-null  int64
 5   loan             38205 non-null  int64
 6   contact          38205 non-null  int64
 7   month            38205 non-null  int64
 8   day_of_week      38205 non-null  int64
 9   campaign         38205 non-null  int64
 10  pdays            38205 non-null  int64
 11  previous         38205 non-null  int64
 12  poutcome         38205 non-null  int64
 13  emp.var.rate     38205 non-null  float64
 14  cons.price.idx   38205 non-null  float64
 15  cons.conf.idx    38205 non-null  float64
 16  euribor3m        38205 non-null  float64
 17  nr.employed      38205 non-null  float64
 18  y                38205 non-null  int64
```

03

Results

Modeling

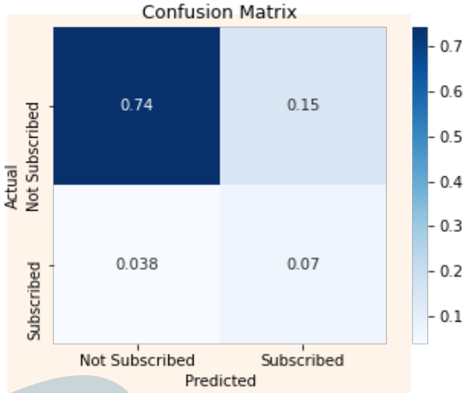
Classification Models Scores

Base Model: Logistic Regression	F1 0.32
Logistic Regression (Class Weight)	F1 0.43 (+34% improved)
Naïve Bayes	F1 0.39
Random Forest	F1 0.40

Confusion Matrix

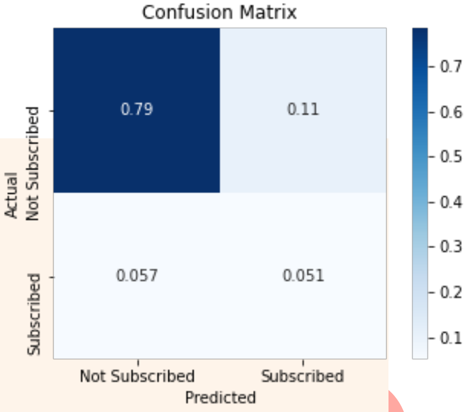
01

Logistic Regression



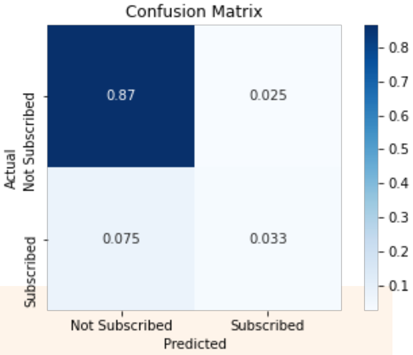
02

Naïve Bayes

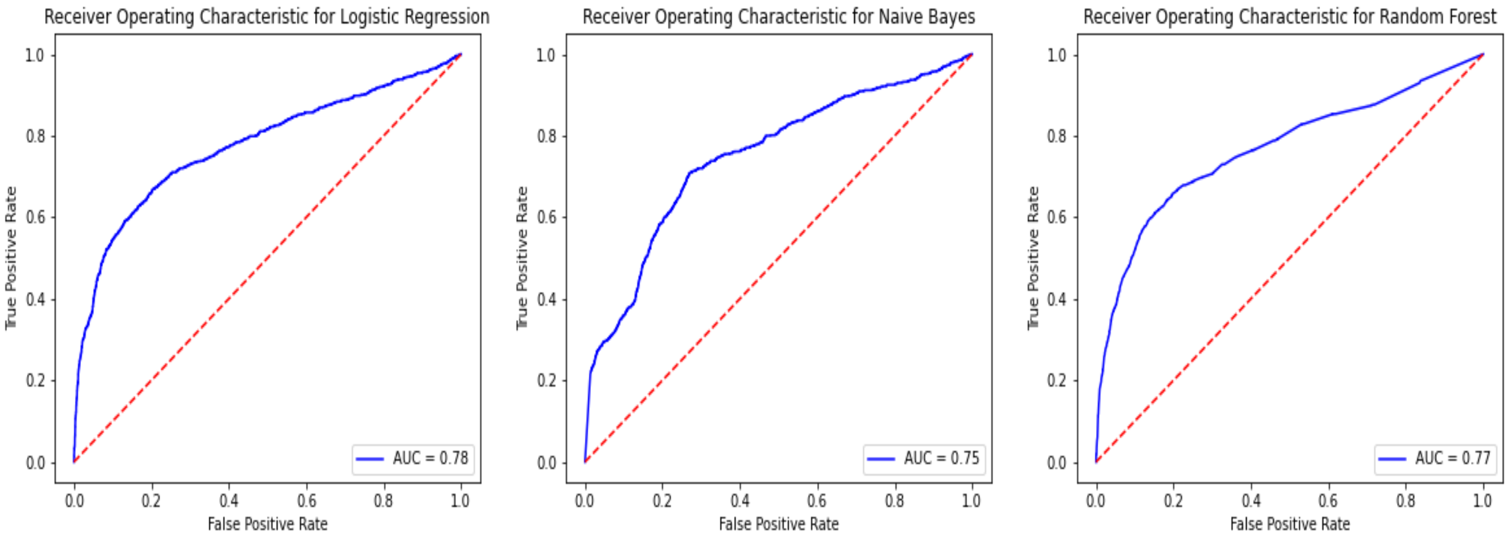


03

Random Forest



ROC Comparison



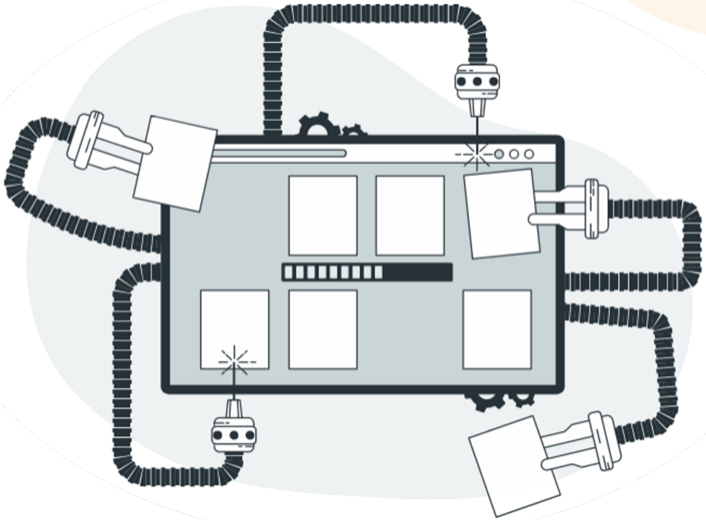
Conclusion:

Logistic Regression Model
(Class Weight)

F1=0.43

Future work

Try more sophisticated models



Model 1

XGboost

+ model performance
and computational speed

Model 2

Gradient Boosting Machines

to decrease the Bias error

Thanks!

hernantru943@gmail.com

GitHub: Hernantru943

LinkedIn: hernantrujillo/

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik and illustrations by Storyset