BANK MARKETING

PRESENTATION

Tejas Mirashi
Mingjie Zhou
Rubina Pradhan

Contents

Research Overview

Methodology and Analysis

Recommendations & Conclusions

Background

- Increasing number of **marketing campaigns** over time has reduced their effects on the general public.
- High competition results lower subscription for a term deposit, i.e. less than **1%** of the contacts.
- Direct marketing has drawbacks, such as causing negative attitude towards banks due to **intrusion of privacy**.
- Portugal **interest rate** (Jan 2003 - Dec 2020)
  - The Average was **1.76%**,
  - The Lowest was **0.06%** (in October 2020)
  - The highest was **4.65%** (in October 2008)

Research Scope

- Predict the **accuracy** rate of term deposit subscription.
- Help banks to manage customers and **improve their efficiency**.
- Understand which factors are influential to customers' decision to design efficient and precise campaign strategy.
- All credit institutions operating in Portugal are obliged to participate in the Portuguese Deposit Guarantee Scheme. The Deposit Guarantee Scheme protects 100,000 euros per customer if the bank fails to meet its obligations to customers.
- Help reduce **costs** and improve the **profits**.

Research Objective

Prediction

Accuracy Testing

Compare Model

Data Description

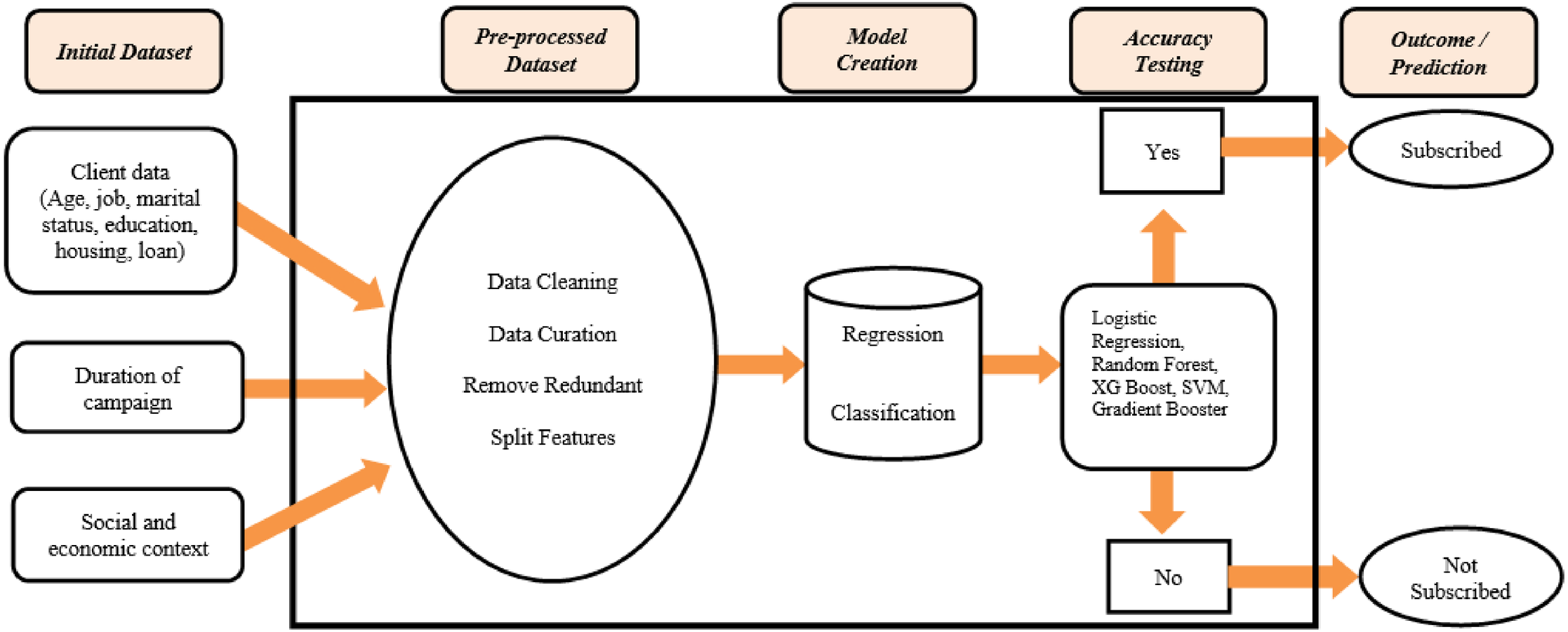
- Portuguese marketing campaign related to bank deposit subscription.
- **45211** clients and **17** columns
  The response is whether the client has subscribed a term deposit.
- Source: http://archive.ics.uci.edu/ml/datasets/Bank+Marketing
- The marketing campaigns were based on **phone calls.**
  Sometimes more than one contact to the same client was required.

Variables	Data Description
Age	Job type
Job	marital status
Marital	Education level
education	yes, no
Default	Yes, no
Housing	have house loan or not
Loan	Have personal loan or not
Contact	contact communication type
Month	last contact day of the Month
day_of_week	last contact day of the week
Duration	Last contact duration, in seconds

Variables	Data Description
Campaign	number of contacts performed during this campaign and for this client
Pdays	number of days that passed by after the client was last contacted from a previous campaign
Previous	number of contacts performed before this campaign and for this client
Poutcome	outcome of the previous marketing campaign
Emp.var.rate	employment variation rate
y	has the client subscribed a term deposit?

Data Source

Data Modelling



Data Source

Machine Learning Algorithms

As our data are imbalanced, we use resampling methods before building models. After preprocessing the data, we build five models:

01 Logistic Regression

02 Random Forest

03 Gradient Boosting

04 XG Boosting

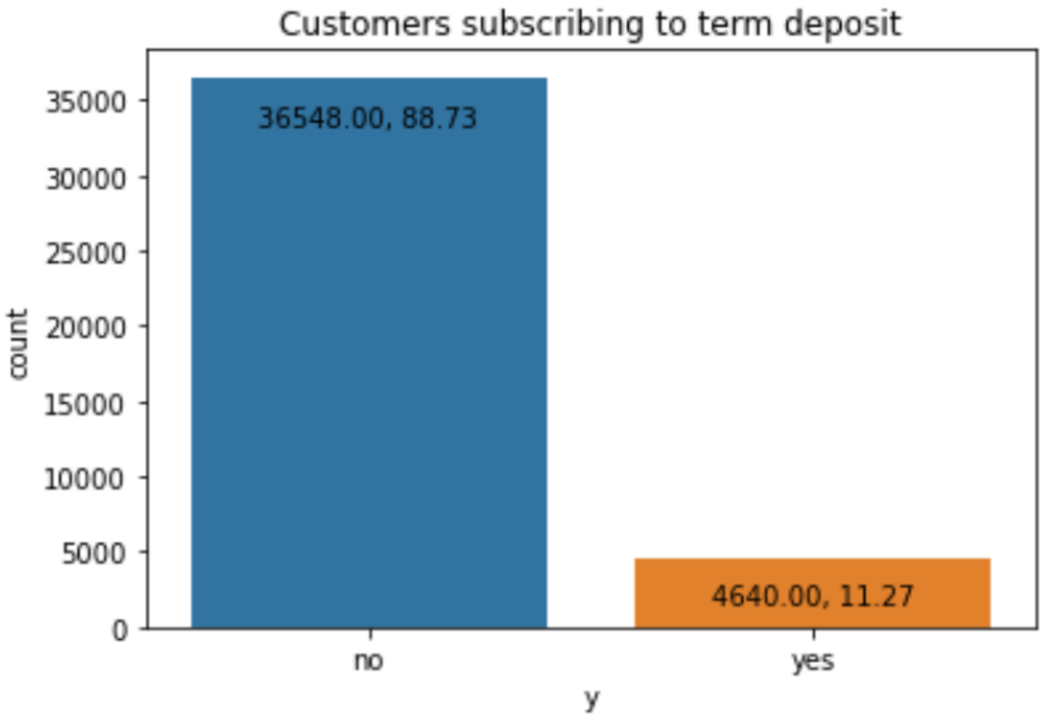
05 Support Vector Machine

Source

Analysis and Interpretation

Balancing and Splitting the Data

- Imbalanced dataset
- MinMaxScaler to standardize the dataset which handles the outliers
- Split into train and test data in the ratio of 70:30



70%

Train Dataset

30%

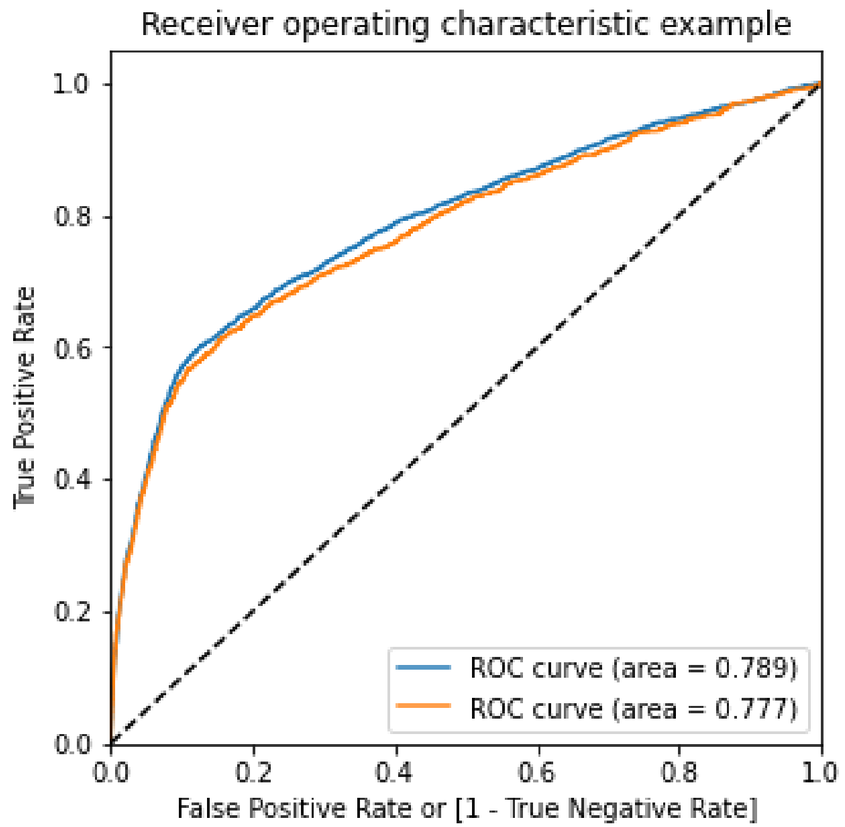
Test Dataset

Subscribed	11.27%
Not Subscribed	88.73%

Logistic Regression

Running the model in the Train Dataset

Train Model



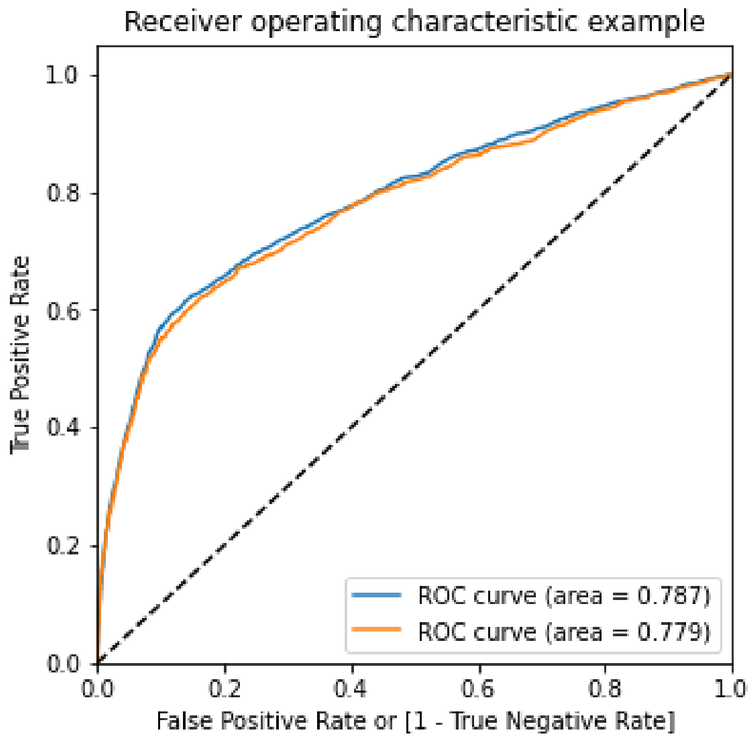
Accuracy	Rate (%)
Train Dataset	89.92%
Test Dataset	90.20%

The model accuracy of while running the model in the train dataset is **89.92%.**

Logistic Regression

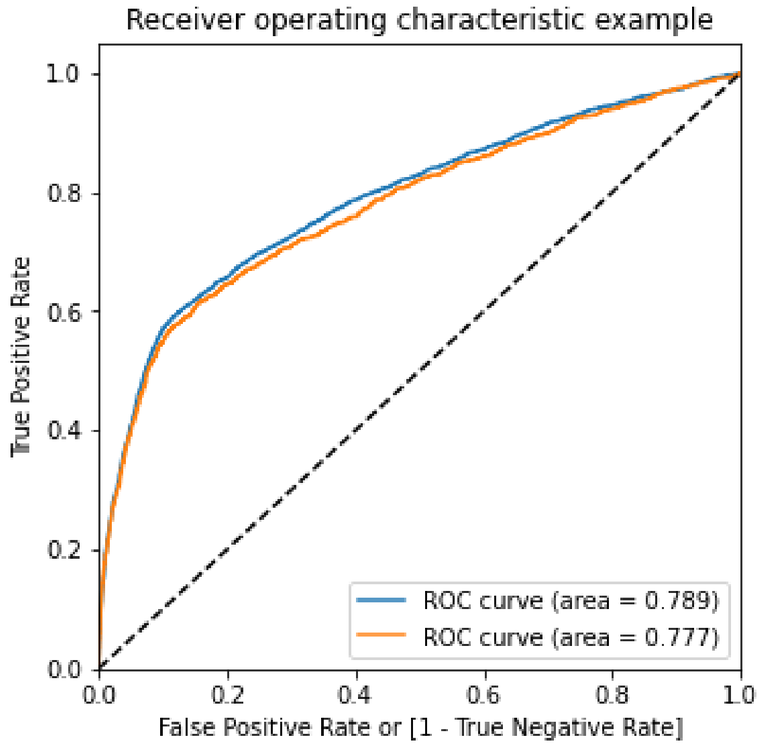
Using logistic regression using Feature selection and hyperparameter tuning:

Feature Selection Using RFE with 20 Features




```
Accuracy train:   0.8993313660304061
Accuracy test:   0.9022137005403521
```

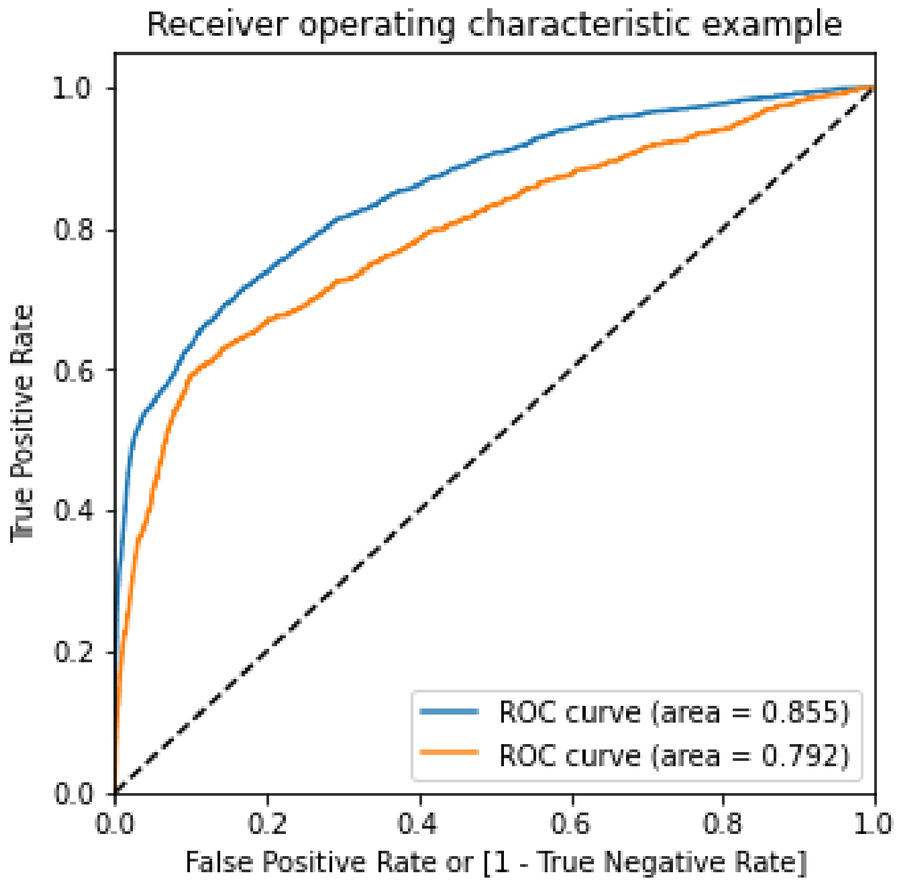
Training model with hyperparameter tuning




```
Accuracy train:   0.8991819506182063
Accuracy test:   0.9016907791528673
```

Random Forest

Random Forest Classifier model with hyperparameter tuning

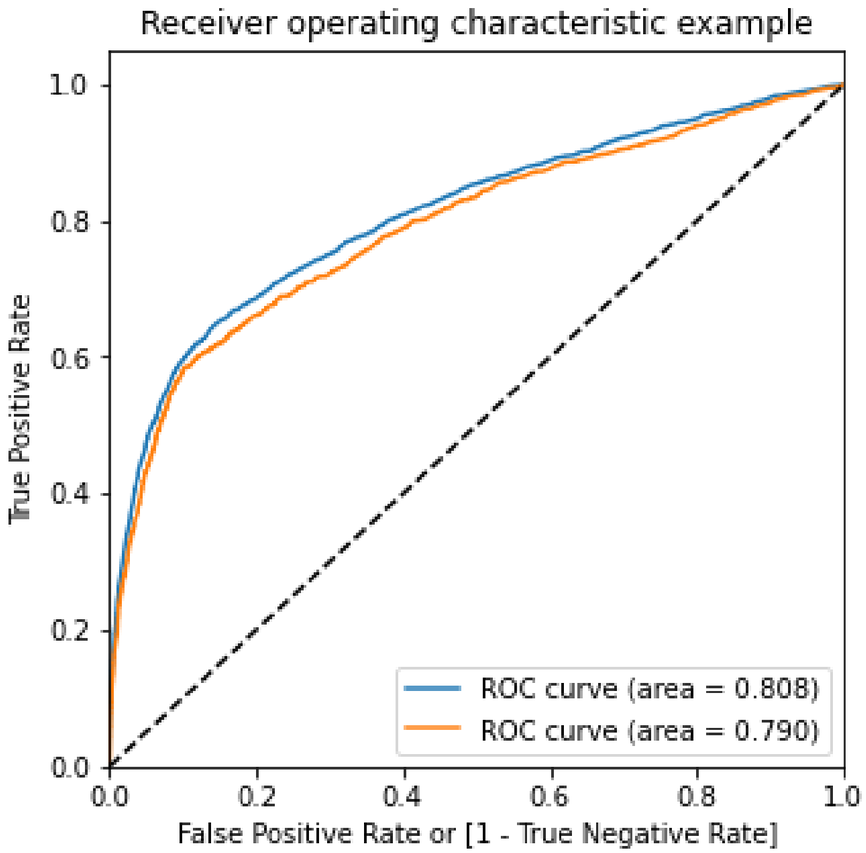


Accuracy	Rate (%)
Train Dataset	91.74%
Test Dataset	90.38%

The model accuracy of Random Forest Classifier model with hyperparameter tuning in train dataset is **91.74%**.

Gradient Boosting

Train Gradient Boosting Classifier model with hyperparameter tuning



Accuracy	Rate (%)
Train Dataset	90.50%
Test Dataset	90.51%

The model accuracy of Train Gradient Boosting Classifier model with hyperparameter tuning in train dataset is **90.50%**.

Support Vector Machine

:

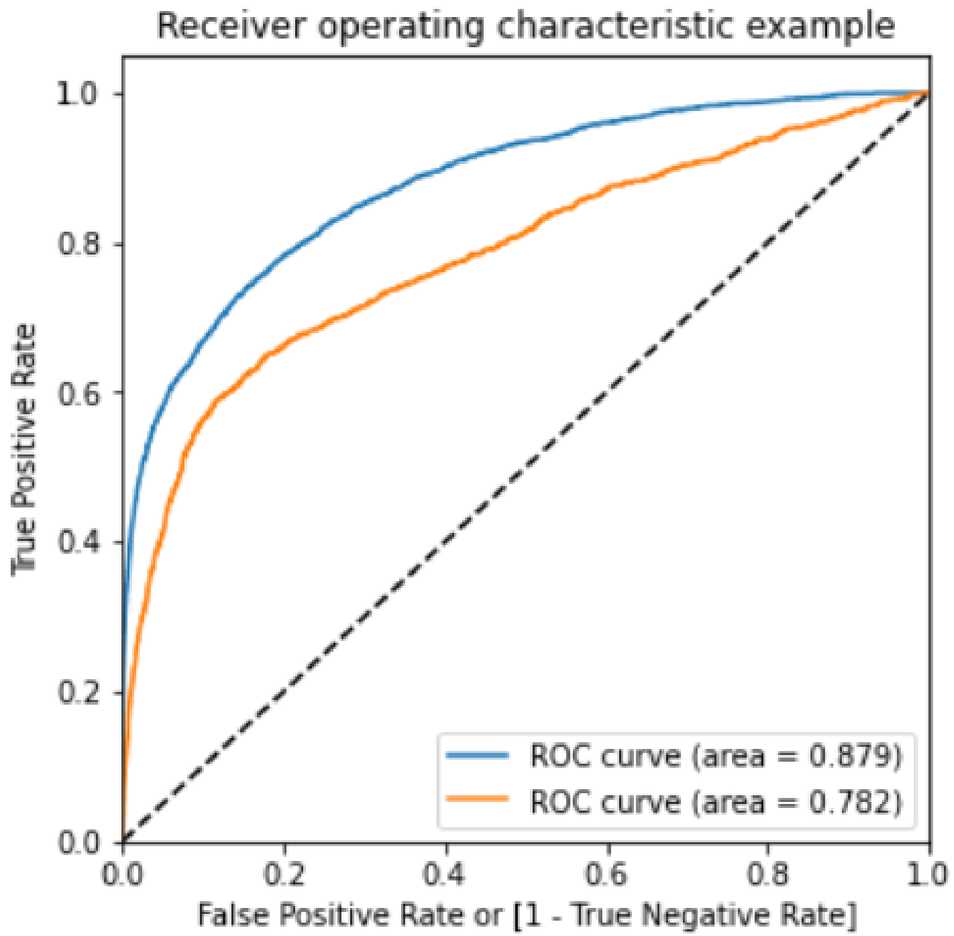
	param_kernel	param_degree	mean_train_score	mean_test_score
0	linear	1	0.596636	0.592088
1	rbf	1	0.862975	0.713407
2	linear	2	0.596636	0.592088
3	rbf	2	0.862975	0.713407

Accuracy	**Rate (%)**
Train Dataset	90.04%
Test Dataset	90.25%

The model accuracy of Support Vector Machine (SVM) is **90.04%.**

XG Boosting

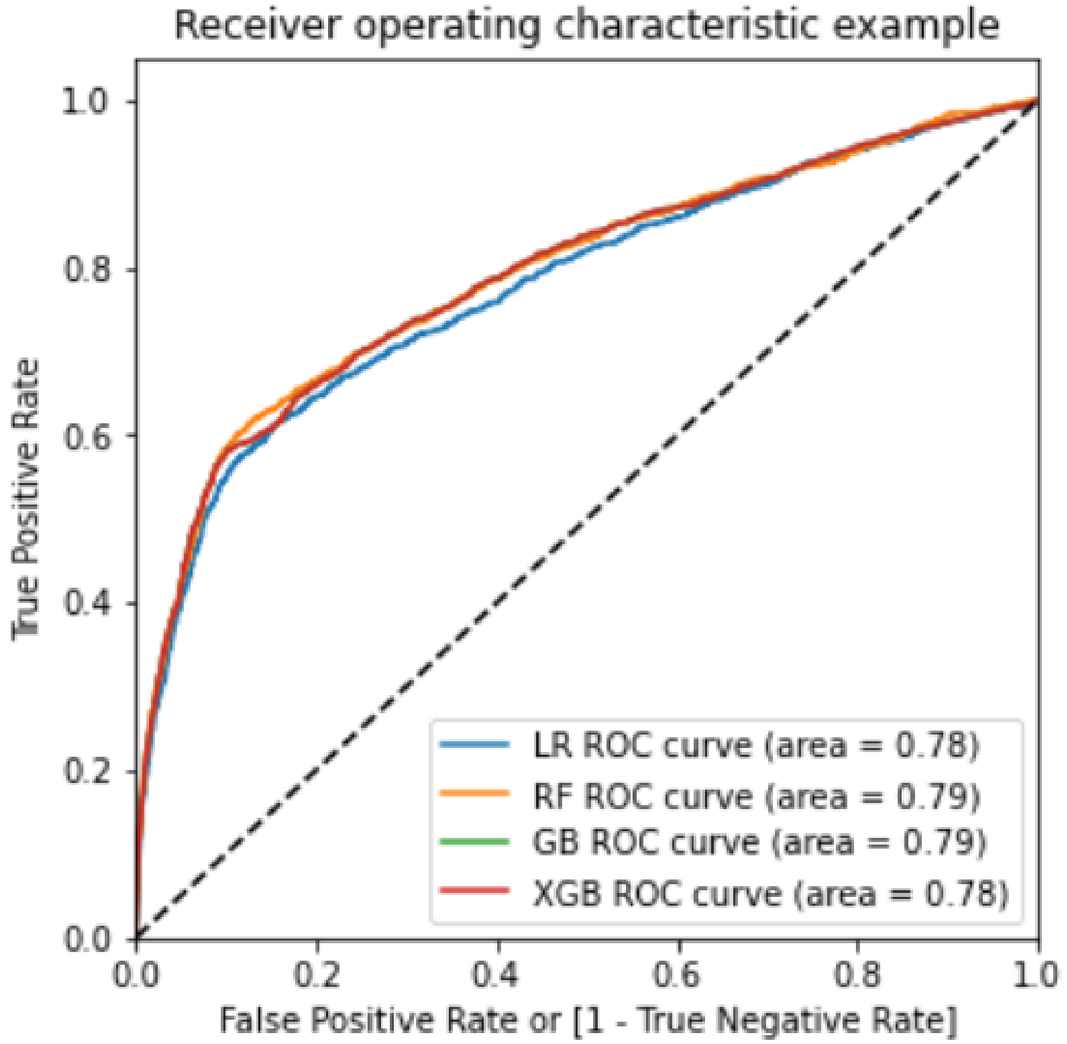
XG Boosting Classifier



Accuracy	Rate (%)
Train Dataset	92.36%
Test Dataset	90.26%

The model accuracy of XG Boosting Classifier is **92.36%**.

Comparing the Models



Model	Area (ROC)
Logistic Regression	0.78
Random Forest	0.79
Gradient Boosting	0.79
XG Boosting	0.78

In the light of overall ROC test, the most accurate prediction ability is Random Forest and Gradient Boosting which covers area of 0.79.

Conclusion

Accuracy of the various algorithms are as follows:

Algorithms	Logistic Regression	Random Forest	Gradient Boosting	Support Vector Machine	XG Boosting
Accuracy Test	89.92%	91.74%	90.50%	90.04%	92.36%

As per test accuracy,

the best model is **XG Boosting** with **92.36%** of accuracy rate followed by **Random Forest** with **91.74%**.

Comparing the Models

- As per ROC, most accurate prediction ability is Random Forest and Gradient Boosting.

- As per test accuracy, the best model is XG Boosting (92.36%).

Limitations

- No enough information to predict the customer potential factors, such as the interest rate during the period.

- Lack of information, preference and marketing campaign of other banking systems.

- As the data were collected from phone call interviews, many clients may refuse to provide their personal information due to the privacy issue.

Findings

- Most employees from Administration department.
- Employment variation rate has positive correlation.
- A stable employment rate denotes a stable economic environment in which people are more confident to make their investment.
- To improve the lead generation: hire more people, improve the quality of phone calls and run their campaigns.
- Multiple Regression Models maybe used to compare relationship with multiple variables and give more accurate suggestions.
- The bank can use features selection using machine learning algorithms to understand target customer-base for subscription of term deposit.

THANK YOU !