ANALYSIS ON TERM DEPOSIT SUBSCRIPTION

*Tejas Mirashi, Rubina Pradhan, Mingjie Zhou*

Introduction

Our project aims to increase campaign efficiency by identifying the main factors that affect the success of a campaign and predicting whether the campaign will be successful to a certain client, namely, whether the client will subscribe a term deposit or not. A term deposit is a fixed-term investment that requires the deposit of money at a financial institution for a fixed period, ranging from one month to a few years. Banks allocate a lot of expenses on their campaigns to promote their customer to subscribe to their term deposit which contributes a big fraction of their revenues. Increasing number of marketing campaigns over time has reduced their effects on the general public.

Data Cleaning

Objectives

- Predict accuracy of customers subscribing a term deposit to increase revenue.
- Recommend strategies to increase campaign efficiency based on predictions.

Data Description
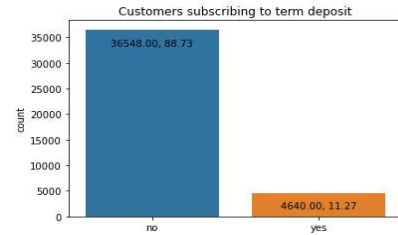
45211 Clients

17 Features

Customers subscribing to term deposit

Based on 3 years data (from May 2008 to November 2010) from a Portuguese banking institution to predict the client subscription for term deposit. As our dataset is imbalanced, we use MinMaxScaler to standardize the dataset before splitting into train and test data which handles the outliers as well.

Analyzing Correlated Variables

Variable	Definitions
emp.var.rate	A statistical index derived by the Employment Agencies to trigger the possibility of change in employment.
nr.employed	number of days employed
euribor	Euro Interbank Offered Rate is the interest rate at which credit institutions lend money to each other, which is often referred to as "the price of money".
previous	number of clients contacted before this campaign
p.days	number of days that passed by after the client was last contacted from a previous campaign
nonexistent	The contacted number is unknown.
success	The customer did subscribe for a term deposit.
failure	The customer did not subscribe for a term deposit.

Highly Positive correlated variables

Highly correlated variables
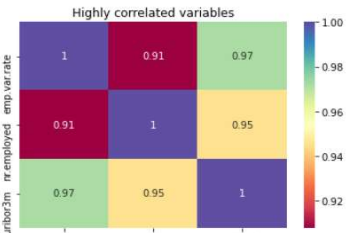
From the positive correlation chart, we can say that 'emp.var.rate' (employment variation rate) and 'nr.employed' (number of days employed) are positively correlated with Euribor.
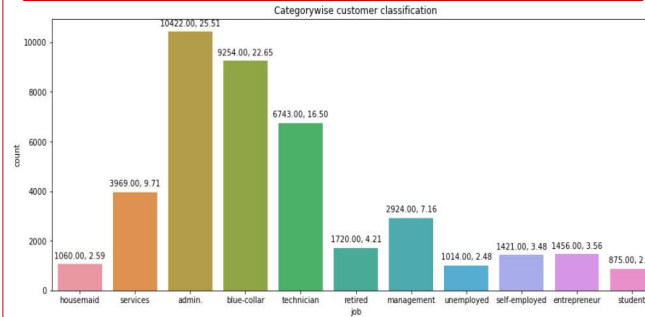
Highly Negative correlated variables
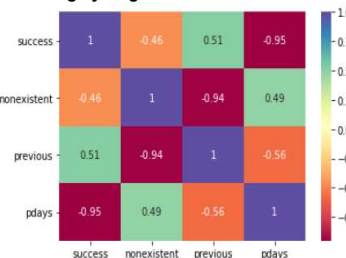
From the negative correlation chart, we can say that Previous and Nonexistent have a negative correlation. Also, the Pdays and Success have negative correlation.

Methodology & Analysis

Job-wise Customer Distribution

Categorywise customer classification

Around 25% of the client are from Administrative field and around 61% of the customers are married. During the month of May there is higher likely chance of an increase in term deposit.

Machine Learning Algorithms

After preprocessing and balancing the data, we build five different models:
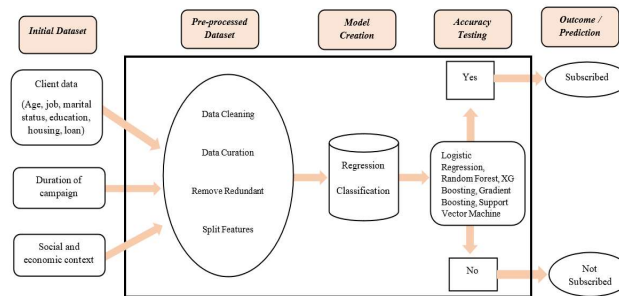
01 Logistic Regression

02 Random Forest

03 Gradient Boosting

04 XG Boosting
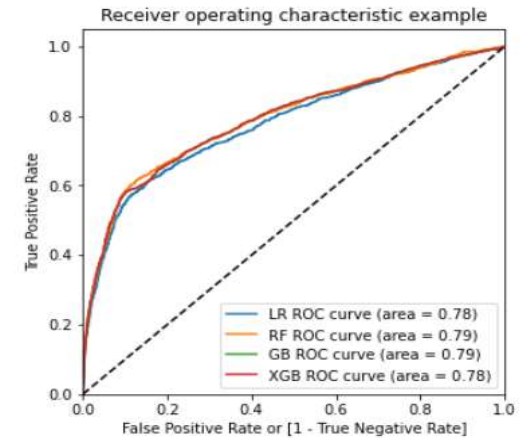
05 Support Vector Machine

System Modelling

Model Accuracy Rate

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

Limitations & Conclusions

Model Comparisons

Receiver operating characteristic example

Findings

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

- As per test accuracy, the best model is XG Boosting with 92.36% of accuracy followed by Random Forest with 91.74%.

- There is a positive response from the customer in the administration department as per the dataset.

- Employment variation rate has negative influence.

- In conclusion, a stable employment rate denotes a stable economic environment in which people are more confident to make their investment. Suggestion for the bank to improve their lead generation: hire more people, improve the quality of phone calls and run their campaigns.

Recommendations

- Due to data restriction, there is not enough information about customer's characteristics to predict the customer preference. In addition, preference and marketing campaign of other banking systems are also not included. If such information is available in the future, we could further compare and analyze the customer satisfaction with rivalry companies in banking industry.

- If we have more detailed dataset, we can used Multiple Regression Models 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.

References

(UCI Machine Learning Repository: Bank Marketing Data Set, 2022) http://archive.ics.uci.edu/ml/datasets/Bank+Marketing
(Bank Marketing, 2022), https://www.kaggle.com/henriqueyamahata/bank-marketing
(Term Deposit Definition, 2022), https://www.investopedia.com/terms/t/termdeposit.asp
(Machine Learning System Design, 2022), https://becominghuman.ai/machine-learning-system-design-f2f4018f2f8

Supervised By: Professor Yue Gao