CUSTOMER SUBSCRIPTION

Machine Learning Model

ADITYA TEWARI

INTRODUCTION

It is a known fact that term deposit helps the bank to invest better to make profit. This also allows the bank to insist customer to get insurance, loans, and other services.

PROBLEM STATEMENT

One of the Portuguese banking institutions held a direct marketing campaign. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

The information about the same is available in the dataset.

OBJECTIVE

- Gather available customer's data, age, campaign outcome's data and usage pattern
- Convert structured and unstructured data/information into meaningful insights
- Utilize these insights to predict clients who are likely to subscribe

TECHNOLOGY USED

Windows 11, Jupyter Notebook

DATASET

The dataset comprises of:

- 45211 rows
- 17 columns

There are a total of 16 features and 1 target variable.

Sample Dataset

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
1	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
4	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no

METHODOLOGY

This work aims to predict whether a customer will subscribe to term deposit or not.

- **IMPORTING NECESSARY LIBRARIES**
  Necessary libraries are imported that are required for data analysis, visualization, and model building.

- **IMPORTING DATASET AND PREPARATION**
  The 'bank-full' dataset is imported. The dataset is cleaned by removing missing values and outliers. Some of the categorical variables are converted to numerical values.

- **EXPLORATORY DATA ANALYSIS**
  Visualization and analysis are performed on the cleaned dataset. Insights are drawn out that will help in feature selection.

- **FEATURE SELECTION**
  Features and the target variable are set by dropping and assigning the columns respectively.

- **TRAIN-TEST SPLIT**
  Train-Test split is performed. The entire dataset is split into train and test dataset.

- **MODEL BUILDING**
  After having a good understanding of data, a model is built to predict the outcome.

- **MODEL EVALUATION**
  The performance of the model is evaluated. Metrics such as as accuracy, precision, and recall are used.

QnA

1. **What is the distribution of the customer ages?**
   Most of the customers' ages are in the range of 30-40.
2. **What is the relationship between customer age and subscription?**
   The percentage of subscription taken among customers of ages 30-60, but it increases significantly above the ages of 60.
3. **Are there any other factors that are correlated with subscription?**
   Some of the features that are correlated with subscription taken or not are duration, month, age, etc.
4. **What is the accuracy of the logistic regression model?**
   88.39%
5. **What are the most important features for the logistic regression model?**
   Logistic regression model follows classification according to the sigmoid function.
6. **What is the precision of the logistic regression model?**
   0.98
7. **What is the recall of the logistic regression model?**
   0.89
8. **What is the f1-score of the logistic regression model?**
   0.93
9. **How can you improve the performance of the logistic regression model?**
   Classifying the categorical data into numerical data, correct data preprocessing, removing unimportant features or scaling the data.
10. **What are the limitations of the logistic regression model?**
    The limitation of logistic regression can be is that it assumes a linear relationship between independent(x) and dependent variables(y), whereas sometimes the dependency can be of higher orders.

REFERENCES

[1] UC Irvine Machine Learning Repository,
     https://archive.ics.uci.edu/dataset/222/bank+marketing

[2] https://github.com/adititewari13/Adult-Data_ML/blob/main/Adult-Data/income_analysis.ipynb