FINAL REPORT

30-June-2022

BANK MARKETING - CAMPAIGN

ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

The bank wants to use the Machine Learning model to shortlist customers whose chances of buying the product is more so that their marketing channel (telemarketing, SMS/email marketing etc) can focus only on those customers whose chances of buying the product is more.

Our Goal was and to look for a way to solve this issue that was presented to us.

We are to;

Develop a model with Duration and without duration feature and report the performance of the model.

The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

Data Set Information

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Bank-additional-full.csv with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014]

2) Bank-additional.csv with 10% of the examples (4119), randomly selected from 1, and 20 puts.

3) Bank-full.csv with all examples and 17 inputs, ordered by date (older version of this dataset with less inputs).

4) Bank.csv with 10% of the examples and 17 inputs, randomly selected from 3 (older version of this dataset with less inputs).

Analysis

Our first task was looking for abnormalities with our data, elements that could affect our model and result.

We encountered some values that are unknown in the dataset. Some of these include the columns for “Default”, “Housing”, “Education”, and “Loan”. In addition we encountered problems in the ‘pdays’ column which had a mass value of -1, which means it was either the customers were not contacted at all or they were missing values.

For the Unknown values this was sorted out using the oneHotEncoding. For the elements in the pdays column we decided to convert each of the ‘-1’ values to the mean of values in the pdays column and after we realised we had so much values in the pdays with the same value. In order to handle this before building out model we converted the converted values to Zero if they were equals to the mean and then One if they were not equal to the mean in the pdays column.

During our Analysis our aim was to see how Age, The number of Contacts made to the Customer, Job type, Marital Status all had effects and impacts on the Deposits. And from our Analysis we found that;

1. Customers with 'blue-collar' and 'services' jobs are less likely to subscribe for term deposit.
2. Married customers are less likely to subscribe for term deposit.
3. Customers with 'cellular' type of contact are less likely to subscribe for term deposit.
4. People who subscribed for term deposit tend to have greater balance and age values.
5. People who subscribed for term deposit tend to have fewer number of contacts during this campaign.

Model

The team carried out the following Machine learning Model algorithms and below is their accuracy score;

1. Logistic Regression - Accuracy Score: 91.38%
2. Random Forest Classifier - Accuracy Score: 91.16%
3. SVC (Support Vector Classifier) - Accuracy Score: 89.61%
4. Decision Tree Classifier - Accuracy Score: 87.29%

Conclusion

From the Model carried out we Picked the Logistic Regression ML Model as it has the highest accuracy score of 91.38%.

THANK YOU

Your Deep Learning Partner