|  |
| --- |
| The Bank |
|  |

# Memorandum

**Machine Learning with Python- by Gyasi Bawuah**

## **Facts:**

## A financial institution (“The Bank”) in Portugal has adopted a marketing strategy - contacting current and prospective customers over the telephone, to sell a new product called Bank Term Deposit.

The Bank has previously established contacts with 4, 521 of these prospective customers, and it turned out that only about 12% of them purchased or subscribed. The bank understands that not every one of the customers will subscribe the Term Deposit, and does not want to invest campaign resources on those who are not likely to subscribe.

This project has three goals. To create a model that can:

* minimize the cost of making random contacts with current and prospective customers,
* target only customers who are likely to subscribe,
* and, maximize the sale of the Term Deposit.

In selecting the best model, the following conditions must be met:

* The model must adequately explain the response variations in whether customers subscribe or not (Target response) - measured as Explained Variable Score.
* The model must accurately predict the chances that a given customer will subscribe the Term Deposit- measured as Accuracy Score.
* The prediction must be highly precise, and minimize misclassifications - measured as F Score.
* Finally, the difference between the actual response and the expected response must be as low as possible- measured as Mean Score Error (M.S.E.).

**Observations:**

12 machine learning models were implemented to generate 36 different iterations (tunings). Below are the top 5 models and their performance scores:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Iteration** | **Accuracy %** | **Explained Variable %** | **F Score %** | **Mean Error** | **Comment** |
| **Extra Trees** | **Model 2** | **93** | **73** | **93** | **0.07** | **Best Model** |
| Extra Trees | Model 1 | 93 | 72 | 93 | 0.07 |  |
| Random Forest | Model 2 | 92 | 69 | 92 | 0.08 |  |
| Bagging | Model 3 | 92 | 69 | 92 | 0.08 |  |
| Random Forest | Model 1 | 92 | 68 | 92 | 0.08 |  |

**Recommendations**

Based on the defined metrics, the Extra Trees Classifier Model- (Iteration 2) is the best fit. In addition, the Stacking Voting Classifier voted 99%, 99% and 98% on all three iterations in favor of the Extra Trees Classifier Model.

This means that, if The Bank implements the Extra Trees Classifier Model – (iteration 2), it will know about 73% of factors that influence the decision of customers to subscribe the Term Deposit or not to subscribe. These factors, defined by the model, are 93% accurate and precise, and their rate of error as a model is less than 10%.

For future model improvement, the Extra Trees Classifier can be improved if a feature reduction technique is used to eliminate variables that do not significantly contribute to the overall performance of the model.

Code Reference:

Details of the python code is attached as ‘Written Work\_Machine Learning with Python’.