**Logistic Regression – Mini Project**

A bank wants to sell an investment product to its premium customers. Certain data of previous similar product offering to customers is available. The column “subscribed” is the output variable which indicates whether the customer subscribed to the product. Based on the data;

Part 1

1. Create a logistic regression model
2. Output the performance metrics such as the Confusion Matrix, Classification Report and the AUC score of the model.

Part 2

Notice there is an imbalance in classes information in the original dataset. Perform a value\_counts operation on the response variable. Notice the percentage of observation in both the classes. In order to rebalance the class information, use the following hint code to resample/rebalance the classes.

*# Importing resample from \*sklearn.utils\* package.*

*from sklearn.utils import resample*

*# Separate the case of yes-subscribers and no-subscribers*

*bank\_subscribed\_no = bank\_df[bank\_df['subscribed'] == 'no']*

*bank\_subscribed\_yes = bank\_df[bank\_df['subscribed'] == 'yes']*

*#Upsample the yes-subscribed cases.*

*df\_minority\_upsampled = resample(bank\_subscribed\_yes, replace=True, n\_samples=2000, random\_state=42)*

*# increase the number of “yes” to 2000.*

*# Combine majority class with upsampled minority class*

*new\_bank\_df = pd.concat([bank\_subscribed\_no, df\_minority\_upsampled])*

After performing the above rebalancing;

1. Create a logistic regression model
2. Output the performance metrics such as the Confusion Matrix, Classification Report and the AUC score of the new model.
3. Comment if you noticed improvement in the model metrics after the imbalance was corrected.

File: bank data.xlsx

Email: ashok.21aug@gmail.com