# Data Intake Report

Name: Bank Marketing Campaign – Customer Product Purchasing Classification
Report date: 11/18/2022
Internship Batch: LISUM14
Version: <1.0>
Data intake by: Islom Pulatov and Ammar Sidhu
Data intake reviewer: <intern who reviewed the report>
Data storage location: https://github.com/Islompulatov/Projects

**Tabular data details:** Name – bank-full.csv

| | |
|---|---|
| **Total number of observations** | 45211 |
| **Total number of files** | 1 |
| **Total number of features** | 17 |
| **Base format of the file** | .csv |
| **Size of the data** | 4.503 MB |

**Proposed Approach:**

- Read dataset into Pandas DataFrame by Comma Separation
- Check for Missing Values, and Duplicates
- Assess Normality and Kurtosis of Numerical Variables
- Explore Outliers for Target Variable
- Create Univariate and Bivariate Data Visualizations for Features and Target Variable
- Convert Relevant Categorical Features into Numerical Variables with OneHotEncoding
- Conduct Correlation Analysis with Correlation Matrix of All Features
- Split Preprocessed Data into X (Features) and Y (Target Variable) then Split X and Y into 80% Training Data and 20% Test Data
- Create a Function to Train Multiple Classifiers as Suggested by Scikit-Learn Algorithm Selection Sheet (https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html) and acquire Training/Testing Accuracies
- Hyperparameter Tune the 3 Best Performing Algorithms with GridSearchCV
- Acquire, for the best performing model, the following classification metrics: Cross-Validated (5-Folds) Accuracy, Classification Report, Precision and Recall, F1-Score, AUC-ROC, and Confusion Matrix as Evaluation Metrics; Visualize Evaluation Metrics as Bar Plot to Compare
- Acquire Feature Importance and Visualize as Bar Plot
- Save the Best Model using Joblib
- Create Flask App with Relevant HTML and CSS Scripts for Model Deployment
- Deploy Model onto Heroku

**Assumptions:**

- Binary Classification Problem (Classes – Yes (1) and No (2))
- Not an Imbalanced Classification Problem