Data Intake Report

Name: Bank Marketing (Campaign)
Report date: 18-06-2024
Internship Batch: LISUM33
Version: 1.0
Data intake by: Team BRAVO
Data intake reviewer: Balamurugan Purushothaman
Data storage location:

Tabular data details:

bank-additional-full.csv

Total number of observations	41188
Total number of files	
Total number of features	21
Base format of the file	.csv
Size of the data	5.56MB

bank-additional.csv

Total number of observations	4119
Total number of files	
Total number of features	21
Base format of the file	.csv
Size of the data	570kB

bank.csv

Total number of observations	4521
Total number of files	
Total number of features	17
Base format of the file	.csv
Size of the data	450kB

bank-full.csv

Total number of observations	45211
Total number of files	
Total number of features	17
Base format of the file	.csv
Size of the data	4.39MB

Proposed Approach:

Various ML models would be compared to perform binary classifications. Each team member would be taking care of each model and in the end, a comparative study is taken to understand which model executes better for classification tasks. After finding the best classification model, we would be deploying with the help of Flask pipeline. Some columns in the dataset were entered as ‘unknown’ , data analysis is to be carried out by replacing values that are similar enough i.e mean, median and mode. After careful consideration, we found that there are no unique or primary keys present. So we can assure that the possibility of duplicate values is 0%. Most of the features do not have any missing values which would be a blessing in model development. Some classification models cannot learn when the features were not in numerical order. For those models, we need to carry out Data Preprocessing steps like Categorical encoding.