Data Intake Report

Name: Data Science:: Bank Marketing (Campaign) -- Group Project
Report date: 08/18/23
Internship Batch: LISUM23:30
Version:<1.0>
Data intake by: Yuheng Chen, Terry Chou, Rishi Aluri, Justine Pile
Data intake reviewer: Yuheng Chen
Data storage location: https://github.com/yuh39/Bank_Marketing_Group_Project/tree/main

Tabular data details: bank-additional-full.csv

Total number of observations	41188
Total number of files	1
Total number of features	20
Base format of the file	CSV
Size of the data	5.83MB

Tabular data details: bank-additional

Total number of observations	4119
Total number of files	1
Total number of features	20
Base format of the file	CSV
Size of the data	584KB

Tabular data details: bank-full

Total number of observations	45211
Total number of files	1
Total number of features	16
Base format of the file	CSV
Size of the data	4.61MB

Tabular data details: bank

Total number of observations	4521
Total number of files	1
Total number of features	16
Base format of the file	CSV
Size of the data	461kb

There are four datasets:
1) 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 inputs.
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).
The smallest datasets are provided to test more computationally demanding machine learning algorithms (e.g., SVM).

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

No missing values

Task:

- Business Understanding
- Data understanding
- Exploratory data Analysis
- Data Preparation
- Model Building (Logistic Regression, ensemble, Boosting etc)
- Model Selection
- Performance reporting
- Deploy the model
- Converting ML metrics into Business metric and explaining result to business
- Prepare presentation for non-technical persons.