Data Science: Bank Marketing (Campaign)
Data Glacier - LISUM 10

Unity of One (1 team member)
Henri Edwards
henrij.edwards@gmail.com
South Africa
Graduate, Data Science

Problem Description:

Developing a model which is able to accurately predict if a client will subscribe to a term deposit.

Data Understanding:

A

Data Description:

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

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).

What are the problems in the data (number of NA values, outliers , skewed etc)

A

What approaches you are trying to apply on your data set to overcome problems like NA value, outlier etc and why?

A

Github Repo link

https://github.com/HenriEdwards/Data-Glacier/tree/main/Week%208