**WEEKLY DELIVERABLE 2**

**Group Name**: Bank Marketing DS 01

**Specialization**: Data Science

**Team Members**:

* **Ines Perko**
  + Master of Mathematics
  + Freie Universität Berlin, Germany
  + Specialization: Finite Element Methods
  + Email: ines.perko93@gmail.com
* **Suvansh Vaid**
  + Master of Data Science
  + Monash University, Melbourne, Australia
  + Specialization: Data Science
  + Email: suvanshvaid@gmail.com
* **Zeynep Başak Eken**
  + Bachelor’s in Economics, Minor: Software Development
  + Bilkent University, Ankara, Turkey
  + Specialization: Data Analysis
  + Email: zeynepbasakeken@gmail.com

**Problem description:**

**ABC Bank** wants to sell its term deposit product to customers and before launching the product they want to develop a model which helps them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

**Business understanding:**

* **Business Objective:**

The Bank wants to shortlist customers whose chances of buying the product are more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only on those customers. This will save their resource and their time (which is directly involved in the cost (resource billing).

* **Success Criteria:**

The success criteria for this business problem would be based on how much maximum number of customers we are able to predict who have subscribed to the product.

**Data Understanding:**

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls.

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

The dataset “bank-additional-full.csv” contains 20 input variables such as bank client data, the data related to the last contact of the current campaign, the data of attributes related to the campaign details and social and economic context attributes. The output variable y has two binary options: yes and no, saying if the client subscribed to a term deposit.

There are two types of variables for analysis, categorical and numerical. There are 10 categorical features: ‘job', 'marital', 'education', 'default', 'housing', 'loan', 'contact', 'month', 'day\_of\_week', 'poutcome and 10 numerical features: 'age', 'duration', 'campaign', 'pdays', 'previous', 'emp.var.rate', 'cons.price.idx', 'cons.conf.idx', 'euribor3m', 'nr.employed'. There are no NA values. In those 10 numerical features, 3 are discrete numerical features: 'previous', 'emp.var.rate', 'nr.employed’ and the rest 7 of it are continuous numerical features. Duration and campaign have heavily skewed towards left and seem to have some outliers, age as well.

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

**Project Lifecycle:**

* Business Understanding (Week 7 – 7 April 2021)
* Data Understanding (Week 8 – 14 April 2021)
* Data Cleaning (Week 9 - 20)
* Exploratory data Analysis (Week 10 – 27 April 2021)
* EDA presentation for business users (Week 11 – 1 May 2021)
* Model Selection and Building (Week 12 – 8 May 2021)
* Performance reporting (Week 13 – 15 May 2021)
* Prepare presentation for non-technical persons (Week 13 - 15 May 2021)
* Final Project Report (Week 13 - 15 May 2021)

**GitHub Repo link:**

https://github.com/SuvanshVaid27/Bank-Marketing-Project