

# **Week 9 Deliverables**

**Group Name: Carpe-Diem group**

**Specialization: Data Science**

**Project Name: Bank Marketing (Campaign) --  
Group Project**



## **Team Members:**

**1.Name: Mohini Kalbandhe**

- **Email:** amohini099@gmail.com
- **Country:** Canada
- **Company:** Happy Orchard
- **Specialization:** Data science

## **2.Name: Kashish Joshipura**

- **Email:** kashishjoshipura@gmail.com
- **Country:** Canada
- **Collage:** The University of British Columbia (UBC)
- **Specialization:** Data Science

## **3.Name: Amir Shahcheraghian**

- **Email:** Amir.shahcheraghian@gmail.com
- **Country:** Canada
- **Collage:** University of Quebec , Canada
- **Specialization:** Data Science, Energy Management analysis

# **Bank Marketing (Campaign)**

## **Problem Statement:**

ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help 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).

## **Project Objective:**

By converting this problem into a machine learning classification problem we will build a model to predict whether a client will subscribe a term deposit or not so that the banks can arrange a better management of available resources by focusing on the potential customers “predicted” by the classifier .

**Technique to be used: Classification**

## **Data cleansing and transformation:**

Dataset has been checked and reviewed by all the peers and concluded that there were no missing values found in the dataset however we found some “unknown” and “other” values which needs to be converted to numerical values or needs to be removed in order to clean the dataset. No “skewness” in dataset was found, data seems symmetrical dataset. Statistics summery such as slandered deviation, distribution, kurtosis and skewness has been checked.

## **Approaches Reviewed by every peers:**

### **Kashish:**

- As there are over 40000 data points in dataset, data set has no null values but found some unknown values, that values are been dropped to clean the data set
- He found some negatively and positively related columns, he found out some positively correlated columns are irrevellent, so thought to drop it off which provides a neat data.

- EDA is performed for continuous and categorical variable
- flooring and clapping using interquintile range(IQR)  
outliers are removed by dropping values that is below 25% and 75% percentile.

## **Amir:**

- the columns which has two values('yes' and 'no') and slightly imbalanced such as default, loan, y, has been converted to (1,0) numerical values. rest are continuous variable were binned so that outliers value are converted into count values.
- He has decided to drop the outliers and ambiguous values, such as "others" and "unknown".
- Skewness doesn't provides much insights in data, as values of columns are nearly zero apart from 'previous'. data seems symmetrical.
- flooring and clapping using interquintile range(IQR)  
Outliers are removed by dropping values that is below 25% and 75% percentile.
- Amir used undersampling method to delete the samples in majority class as There are huge difference between  $y=1$  and  $y=0$ , So the ML algorithm will omit the smaller value, which may affect the performance of algorithm.
- Overall visualization provides beautiful insights and cleaned data.

## **Mohini:**

- From correlation on dataset she found out positive correlated values which needs to be removed from data.
- While handling unknown and NAN values she found out 'pdays' has 39673 values are 999 that shows client was not previously contacted, this can add bias to our model hence converting this column to categorical column. other values 3-27 are low in number. so decided to use one hot encoding on "pdays" and converted into numerical value.
- She decided to keep unknown values as to use minskowski error which reduces impacts of outliers on model and used one hot encoding on categorical variable convert it into numerical value.
- She identified the variable that has predictive power using weight of evidence and information value.

## **Github Repo link:**

“<https://github.com/amohini099/Banco-de-portugal-marketing/tree/main/week%209>”