

Assignment 02

Use case: Bank Marketing Dataset

The dataset is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The goal is to predict if the client will subscribe a term deposit.

<https://archive.ics.uci.edu/ml/datasets/bank+marketing#>

Produce Python Notebook, which performs the following task

- (a) Preprocess the dataset as specified in the data mining process.
 - I. Handle Missing Values and Outliers if any
 - II. Produce Q-Q Plots and Histograms of the features, and apply the transformations if required.
 - III. If it is required, apply suitable feature coding techniques.
 - IV. Scale and/or standardized the features, produce relevant graphs to show the scaling/ standardizing effect.
 - V. If necessary, apply feature discretization, and produce a relevant graph to show the discretization
- (b) Perform Feature Engineering by executing the following task:
 - I. Appropriately use PCA (Principal Component Analysis) or SVD (Singular Value Decomposition) for feature reduction.
 - II. Identify significant and independent features using appropriate techniques. Show how you selected the features using suitable graphs.
- (c) Apply the following techniques to predict the value of Y for the test dataset (Training Dataset to Test Dataset is 0.8 to 0.2 ratio)
 - I. Support Vector Machines (SVM) with kernels

The Assignment Report (Submit as a Report / Submit as a Blog)

You need to formulate solutions for each of parts (a) through (e) above, clearly explaining your Python code and specifying the outputs produced by the code for the dataset given in a *Jupyter Notebook* named *Solution_StudentNumber.ipynb*. For each such part, a descriptive summary with an interpretation should be given for the output obtained after each executable cell.