

DELIVERABLE WEEK - 8

Specialization : Data Science

Name : Mohammed Maqsood
Email : mohammedmaqsood48@gmail.com
Country : Germany
College : Otto von Guericke University
Specialization : Chemical and Energy Engineering

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).

Data Understanding

The objective of data understanding was to understand attributes of the data by identifying key characteristics in the data such as data volume, total number of variables in the data, understanding problems with the data such as missing values, duplicates, outliers, skewness, etc.

Type of data for analysis

The dataset consists of numerical and categorical data.

Problems in the data

The dataset was checked for missing values and duplicates, no missing values or duplicates were found. Summary statistics such as the mean, standard deviation, distribution, and kurtosis, skewness were checked. We found that age, balance, duration, campaign, p days and previous variables were positively skewed and had outliers.

Approaches to overcome problems in the data

Log transformation, replacing outliers with the median and using the inter-quartile score are common methods to deal with outliers in the data. These methods were investigated and compared. We decided to replace outliers in the dataset with median values.