

Deliverable Week 8

Group Name: Destined Data Team

Specialization: Data Science

Team members:

1. **Name:** Surya Chandra
Email: ksuryachandra619@gmail.com
Country: Germany
College: Otto von Guericke University
Specialization: Electrical Engineering and Information Technology
2. **Name:** Praneetha Rajupalepu
Email: praneetha.724@gmail.com
Country: Canada
Company: Modest Tree
Specialization: Data Science
3. **Name:** Selaelo Kholofelo Ramokgopa
Email: sly.kholo@gmail.com
Country: South Africa
College: University of Johannesburg
Specialization: Chemical Engineering
4. **Name:** Cagla Yucel
Email: caglasipahi@gmail.com
Country: Switzerland
College: Georgia Institute of Technology
Specialization: Analytics

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 will help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank and other Financial Institutions).

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, pdays 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-quantile 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.