

Week 8 Deliverables

Team member's details :

Group Name: Hossain_Hubiak

- Alyssa Hubiak, alyssa.hubiak@gmail.com, United States, currently a student at Maryville University, Data Science
- Nasibul Hossain, Nasibulh@gmail.com, US, currently seeking a role in Data Science

Problem description

Develop a model to predict whether a customer is likely to buy the bank's product or not. This will help ABC Bank determine where to focus their marketing for their new term deposit product. This will need to be a binary classification model to predict whether the customer will subscribe (1) or not (0).

Data understanding

This data contains call data for 41188 calls, and includes information on the client, previous campaigns, and social/economic factors which may play a role in the final outcome of whether the client subscribes to the new term deposit. Several of the columns display evidence of skew and outliers, but many do not have evidence of either.

What type of data you have got for analysis

Our data is comprised of 41,188 rows that are related to individual calls in attempts to subscribe clients to a new term deposit. This dataset includes both categorical and numeric columns.

What are the problems in the data (number of NA values, outliers , skewed etc)

We have no missing/NA values in our dataset. There are several columns which have skew and outliers. Some have expected skew due to the large number of new clients affecting outcomes, and others may need to be handled prior to model selection. At least one column will need to be dropped due to its lack of use when creating a predictive model.

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

We have no issues with missing values. For handling outliers, some of them will need to be dropped, others may need to be normalized, and others are expected and will need to remain a part of the dataset otherwise it may affect the predictive capabilities of our model.

Github Repo link

https://github.com/ahubiak/Data_Glacier_DS_Project