

Name: Arian Moslehi

Email: arianmoslehi@gmail.com

Country: USA

College/Company: SLU

Specialization: Data Science

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 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).

Data understanding: Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc.) can focus only to those customers whose chances of buying the product is more. This will save resource and their time (which is directly involved in the cost (resource billing)). Develop model with Duration and without duration feature and report the performance of the model. Duration feature is not recommended as this will be difficult to explain the result to business and also it will be difficult for business to campaign based on duration.

What type of data you have got for analysis: CSV

What are the problems in the data: There are some outliers, but no missing values.

What approaches you are trying to apply on your data set to overcome problems like NA value, outlier etc and why? Remove outliers because outliers increase the variability in your data, which decreases statistical power. Consequently, excluding outliers can cause your results to become statistically significant.