

Group Name: DataLex
Name: Ovbude Uankhehi
Email: ovatedom@gmail.com
Country: Nigeria
Specialization: Data Science

Problem Description:

One of the challenges for all pharmaceutical companies is to understand the persistency of drug as per the physician prescription. To solve this problem ABC pharma company approached an analytics company to automate this process of identification.

Data Understanding:

In the Dataset sheet of the excel document, after dropping the “Id” variable, it contains data with 3424 observations and 68 features.

Note:

1. There are no missing values, but there are unknown categories, relevant to the analysis.
2. There are 2 numeric features, and 66 categorical features.

Numeric Features (Handling Outliers):

I plotted the distribution of both numeric features, to spot for outliers, and removed the 99th percentile of both features.

Categorical Features:

I used one-hot encoding to prepare dummy variables for the machine learning model, using the get dummies function, and standardized all feature names.