Drug Persistence Classification

Created by Pharmalytics

Week 10: EDA

Noah Igram
noahigram2023@u.northwestern.edu
Northwestern University, USA

ABC Pharma is a pharmaceutical company looking to improve their patient identification process. They hope to improve their reach to customers by targeting more specific groups of people who they believe would be potential drug patients. We hope to automate the process of predicting drug persistence by creating a classification model which uses past patient data to predict whether or not a patient was persistent in taking a drug for its prescribed course.

This week was focused on doing EDA on the data to try and gain a better background understanding of the data. The following strategies were implemented:
- File reading using Pandas - Reading from an excel file into a Pandas DataFrame for easier use
- Dealing with missing values- Finding missing values in the dataset, or values which have odd or incorrectly read values. Also attempting to impute these values using different strategies (Done with Data Cleansing and Transformation)
- Data type analysis - understanding the correct datatype for each column of data and whether or not Pandas read this in correctly and fixing if not.
- Analysis of categorical features - Plotting, contingency tables to try and understand which values are related for better modeling. Main goal was to understand which features most affect drug persistence.

Our final recommendation is to continue with our classification model and use this data set to continue tuning it for the best possible accuracy.

GitHub Repo: https://github.com/noahigram/Pharmalytics