Drug Persistence Classification
Created by Pharmalytics

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

This proposal outlines a data science project which aims to aid ABC Pharma in understanding patient drug persistence. 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. Below describes the timeline of the project and the final vision for what we hope to achieve.

Week 1-2: Data intake
During this time the data will be initially viewed in an effort to better understand the source of the data and the study that was completed which produced the dataset. We also hope to extract basic statistical information from the data and make any observations that might be helpful. Lastly, we will be doing a more in depth analysis of the features of the data, trying to get a better idea of which features might be the most relevant and helpful for our model.

Weeks 3-4:
Here we will be data cleansing to try and deal with any NA values or outliers found in the dataset. We will also create an EDA notebook to get a more in depth understanding of our dataset.

Week 5:
This week will be dedicated to creating the classification model to predict drug persistence. We hope to also be able to make significant progress in tuning the model for best performance.

Week 6:
This week we will be deploying our model to the web, and presenting it to ABC Pharma in hopes that they will be able to use it successfully.

The following link leads to a github repository which will eventually contain all the project deliverables and relevant documents/files.
https://github.com/noahigram/Pharmalytics

Data Intake Report

Name: Drug Persistency Classification
Report date: 9/1/2022
Internship Batch: LISUM11
Version:<1.0>
Data intake by: Noah Igram
Data storage location: https://github.com/noahigram/Pharmalytics

Tabular data details:

Total number of observations	3424
Total number of files	1
Total number of features	69
Base format of the file	.xlsx
Size of the data	920 KB

Proposed Approach:

- Pandas will be used to better understand the data and its features.
- We assume that the patients in the dataset are from the same study.
- We will later create a model in an effort to automate ABC's Pharma patient identification process.