Team member's details: Group Name: Healthcare Team

Emily Yao, yaowemily@gmail.com, United States of America, University of Florida, Data Science
Quynh Nguyen, qtn2@case.edu, United States of America, Case Western Reserve University, Data Science

Problem description:

Persistence of drugs, otherwise known as the duration between initial drug use to its discontinuation, is a big challenge for Pharmaceutical companies. However, with the use of machine learning, we can automate this process of identification by gathering insights on the factors impacting drug persistence.

Business understanding:

Making medicine is expensive, difficult, and time-consuming. Some medicines can take decades to develop and cost billions to make. However, successes are always worth it and change lives all over the world. Making drugs typically takes three steps: research, development, and approval and the project we are doing heavily involves the development stage to analyze the massive amounts of data used during clinical trials. In order to determine how long drugs should be prescribed for or even if they are “successful”, we need to analyze the data. Doing so will save the company money, time, and resources.

Project life cycle along with deadline:

> Initial Project Report - 12/19/22
Analyze and figure out initial approaches for the project (understand its importance and reasoning behind analysis).

> Understanding the Data - 12/26/22
Initial cleaning and quality control for data and considering methods for analysis.

> Data Cleaning & Transformation - 1/02/23

> EDA -1/09/23

> EDA Presentation (for business users) + proposed modeling - 1/16/23

> Model building - 1/23/23

>Final Project Report and Code - 1/30/23