

Julia Donato

Proposal for Drug Persistency Classification Model

Individual Project- Julia Donato, julia.donato21@gmail.com, USA, Indiana University-Bloomington, Data Science

Problem Description:

Drug persistency is a challenge for pharmaceutical companies and understanding the factors that impact it can be difficult. My goal is to develop a machine learning model that can predict drug persistency based on physician prescription data.

Business Understanding:

ABC Pharma has approached us to automate the identification process of drug persistency. By building a classification model, ABC Pharma hopes to gain insights into the factors that impact drug persistency, which can help improve drug adherence and patient outcomes.

Project Lifecycle:

Problem Understanding/Data Understanding - 1 week

Data Cleaning and Transformation - 1 week

EDA – 1 week

Model Selection - 1 week

Model Building - 1 weeks

Preparing Final Report - 1 week

Total Project Timeline: 6 weeks

Data Intake Report:

I will be using a dataset provided by ABC Pharma that includes patient demographics, physician specialty, clinical factors, disease/treatment factors, and adherence data. The dataset contains 100,000 patient records and has been anonymized for privacy purposes.

Github Repo Link:

I will be using Github to store my project code and documentation. The repo link is:

<https://github.com/julia-donato/DG-Final-Project>

In conclusion, my goal is to build a classification model to predict drug persistency and provide ABC Pharma with insights into the factors that impact it. I anticipate completing the project in 6 weeks and will be using a dataset provided by ABC Pharma. My project code and documentation will be stored on Github for easy access and collaboration.

See next page for data intake report.

Data Intake Report

Name: ML Model Deployment on Flask

Report date: 02/22/2023

Internship Batch: LISUM17

Version: 1.0

Data intake by: Julia Donato

Data intake reviewer:

Data storage location:

Tabular data details: Healthcare_dataset.xlsx

Total number of observations	3425
Total number of files	
Total number of features	769
Base format of the file	.xlsx
Size of the data	920 KB

Proposed Approach:

Problem Understanding/Data Understanding - 1 week

Data Cleaning and Transformation - 1 week

EDA – 1 week

Model Selection - 1 week

Model Building - 1 weeks

Preparing Final Report - 1 week

Total Project Timeline: 6 weeks