Data Science Intern at Data Glacier

Project: Healthcare - Persistency of a drug (Data Science)

Week 7: Deliverables

NAMES	Taiwo Akingbesote	Kerr Tan	Farzana Chowhury	
UNIVERSITY	Montclair state University	New York University	Mount Holyoke College	
EMAIL	Akingbesotet1@montclair.edu	St4153@nyu.edu	Chowd23f@mtholyoke.edu	
COUNTRY	USA	USA	USA	
SPECIALIZATION	Data Science	Data Science	Data Science	
BATCH CODE	LISUM22	LISUM22	LISUM22	
DATE	19 July, 2023	19 July, 2023	19 July, 2023	
ALL SUBMITTED TO DATA GLACIERS				

Table of Contents:

1. Project Plan

WEEKS	DATE	PROGRESS
Week 7	July 19, 2023	Problem Statement and Data Processing
Week 8	July 26, 2023	
Week 9	Aug 2, 2023	
Week 10	Aug 9, 2023	
Week 11	Aug 16, 2023	EDA Presentation and proposed model technique
Week 12	Aug 23, 2023	Model Selection
Week 13	Aug 30, 2023	Final project report and code

2. Problem Statement

The pharmaceutical industry faces numerous challenges in understanding the persistency of drugs as per physician prescriptions. Ensuring that patients adhere to prescribed medications is crucial for their health outcomes and overall treatment effectiveness. However, non-adherence to prescribed medications can lead to suboptimal results, increased healthcare costs, and potential complications. To address this critical issue, ABC pharma company has decided to take a data-driven approach and has approached us to automate the process of identifying factors impacting drug persistency.

Our task is to analyze a comprehensive dataset containing a diverse set of variables related to patient demographics, provider attributes, clinical factors, disease/treatment factors, comorbidities, concomitancy, and adherence information. The primary objective is to gather insights and build a robust classification model that predicts whether a patient will exhibit persistency with the prescribed drug or not. The target variable, "Persistency_Flag," will serve as the ground truth for this classification task, where it is coded as 1 if the patient is persistent and 0 if the patient is non-persistent between variables. Then build a model that classifies the dataset.