

Group Name: Individual Project - Healthcare Persistency
Name: Kirtoria Ward
Email: kirtoria@gmail.com
Country: United States
College/Company: North Central College
Specialization: Data Science

Problem Description

This project aims to predict whether a patient remains persistent with their prescribed medication or discontinues it. The goal is to provide insights for pharmaceutical companies to automate the identification process, improve patient outcomes, and optimize strategies.

Data Understanding

Data Type: The dataset includes patient demographics, clinical factors, and their persistency status. It is structured, tabular data in .xlsx format with 3,424 rows and 69 columns.

Problems Identified:

- **Missing Values:** Some columns, such as clinical factors, have missing values.
 - **Outliers:** Certain numerical columns may have extreme values (e.g., Age, Risk Scores).
 - **Imbalanced Target Variable:** The dataset has more Non-Persistent cases than Persistent, which could bias the model.
-

Approaches to Fix Problems in the Data

- **Missing Values:**
 - Plan: Fill in missing numbers with the average value, and fill in missing categories with the most common category. This keeps as much data as possible without leaving blanks.
 - **Outliers:**
 - Plan: Remove extreme values that are way too high or low. Outliers can confuse the model and make predictions less accurate.
 - **Class Imbalance:**
 - Plan: Use under sampling to reduce the number of Non-Persistent cases by randomly removing some of them, balancing the dataset. This ensures the model gives equal focus to both Persistent and Non-Persistent cases, improving its ability to predict both classes effectively.
-

GitHub Repository Link:

https://github.com/KTW04/Data_Glacier_Internship_LISUM38