Data Science Intern at Data Glacier

Project: Healthcare – Persistency of drug

Week 7

Table of contents

Performer's Details

Name: Kseniya Kudzelich

Email: ksusha.konstovich@gmail.com

Country: Belarus

Specialization: Data Science

Internship Batch: LISUM34

Problem Understanding

The objective of this project is to **understand and predict the persistency of a drug prescribed by physicians**. Persistency, in this context, refers to whether a patient, based on his/her information, will follow the prescribed medication regimen over a certain period.

Identifying factors that influence persistency is crucial for pharmaceutical companies to improve patient outcomes, reduce healthcare costs, and enhance their product offerings.

Business Understanding

Persistency of medication is a critical factor in the effectiveness of treatment plans. High persistency rates generally correlate with better health outcomes, as patients are more likely to follow their treatment plans. Conversely, non-persistence can lead to worsening health conditions, increased hospitalizations, and higher overall healthcare costs.

Pharmaceutical company ABC Pharma is interested in identifying the key factors that influence whether patients persist with their medication. By leveraging machine learning to predict persistency, the company can develop strategies to improve adherence rates.

Project Lifecycle and Deadlines

Tasks	Deadline
Problem understanding and Business understanding	19 July 2024
Data understanding	26 July 2024
Data cleaning and transformation	2 Aug 2024
EDA and Feature engineering	9 Aug 2024
EDA Presentation and Proposed modeling technique	16 Aug 2024
Model buliding, Selection, Evaluation and Deployment	23 Aug 2024
Project Report	30 Aug 2024

Data Intake Report

Name: Healthcare – Persistency of a drug
Report date: 17 July 2024
Internship Batch: LISUM34
Version: 1.0
Data intake by: Kseniya Kudzelich
Data intake reviewer: Data Glacier
Data storage location: https://github.com/kkudzelich/Data-Science-Intern/tree/main/Week%207

Tabular data details: Healthcare_dataset

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	888 KB

Proposed Approach:

- There are no missing values in dataset mentioned above, but there are categories such as 'Unknown' and 'Other/Unknown', we will handle them.
- There are no duplicated values in dataset mentioned above.

GitHub Repo Link

Project Link: https://github.com/kkudzelich/Data-Science-Intern

Submitted by: Kseniya Kudzelich
Date: 18 July 2024