

Data Science Project: Healthcare: Persistency of a Drug

Name: Vidya Ganesan

E-mail: vidyaganesan15@gmail.com

Country: UAE

Specialization: Data Science

1. Problem Description

One of the challenges for all pharmaceutical companies is to understand the persistency of drug as per the physician prescription. With an objective to gather insights on the factors that are impacting the persistency.

2. Business Understanding

The challenges for the pharmaceutical companies are to understand the purpose of the drugs as per the physician prescription. The pharmaceutical company should provide the drug according to the indication of the persistency flag provided by the physician. The persistency flag is mostly based on clinical flags and treatment flags. We will analyze the need of drugs according to the flag indication. We will use some techniques in machine learning to obtain our end result.

3. Project Lifecycle along with the deadline.

- Data understanding, data analysis and data cleaning and transformation (March 16 2022)
- Performing Exploratory Data Analysis and EDA models and presentation (April 5-13, 2022)
- Model Building and final presentation (20 April 2022)

4. Data Intake Report

Given file format	.xls
Total Number of data's	3424
Given File	1
Number of features	26
Size of the data	899 KB

5. GitHub Repo Link

<https://github.com/vidya-ganesan/Healthcare-Persistency-of-drugs>

Week 8 Deliverables

1. Data Understanding

The given dataset is of excel(.xls) format. I imported the data set in the python notebook as pandas data frame using the python library pandas. The file contains two subfiles one with the data features and the other with the dataset which we should analyse the data. From the data we understood the target is the persistency flag and the remaining attributes are in features.

2. Data Type

Structured data type given in excel(.xls) format.

3. Data Problems

The data contains both numeric and text. There are NA values which need to be eliminated. There are outliers and skewed when we model the dataset and it should be taken care.

4. Approaches

Exploratory Data Analysis is done to eliminate null values, outliers.

5. GitHub Repo link

<https://github.com/vidya-ganesan/Healthcare-Persistency-of-drugs>

