Week 7 – Healthcare Project

Group Name: Cool Data Scientists Team

Team Members Details:

Name	Email	Country	College / Company	Specialization
Yousef Elbayoumi	yousefxelbayomi@gmail.com	Palestine	Bahçeşehir University	Data Science
Mukhammadjon Kholmirzev	kmukhammadjon@gmail.com	Uzbekistan	Ulsan National Institute of Science and Technology	Data Science
Jamila hamdi	jamila.hamdi90@gmail.com	Tunisia	Trainee	Data Science
H. Melis Tekin Akcin	meliss85@gmail.com	UK	Hacettepe University	Data Science

Problem Description

One of the challenges for Pharmaceutical companies is to understand the persistency of drug as per the physician prescription. This issue results in a bad impact on the pharmacies for all the categories; patients, physicians, and administration. However, the team of data scientist is capable of discovering the analyzing the dataset and detecting the factors that are impacting the primary factor which is the "persistency". By building a classification machine learning model, we will be able to classify the dataset and find the variables that affect the target variables "Persistency Flag".

Business Understanding

In this project, our aim is to examine the persistency of drug.
To have a better business understanding, our aim is to understand our dataset deeply. We will examine the features to gather some more knowledge, we will find relationships between the features and their effects on our target variable.
We will look at the demographics, clinical factors, provider attributes and disease/treatment factors.
Then our goal is to create a model to decide whether a given drug is persistent as per the physician's prescription. To do that, we will do some feature engineering and decide which of the features effects the drug persistency.
After applying different models, we will decide which model fits best to our dataset.

Project Lifecycle Along with Deadline

For working in the project we decided to use the agile methodology, the reason we decided to use it, is because the agile approach can be quickly adapted by many different type of projects, such as marketing, universities, military, software, and industrial project, the ability of quick adaptation will be useful for us because we will be working in pairs to move faster, but in the same time we need to have the ability to go backwards in case of any errors or mistakes. For the deadlines, we will be working according to the same deadlines provided in Canvas and we will

work to submit on time, but most importantly we will work to deliver everything before the final deadline.

Data Intake Report

Name: Healthcare- Persistency of a drug

Report date: <20.07.2021>

Internship Batch:<LISUM01>

Version:<1.0>

Data intake by: Yousef Elbayoumi, Mukhammadjon Kholmirzev, Jamila Hamdi, H. Melis Tekin Akcin

Data intake reviewer:

Data storage location:

Tabular data details:

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	**899 KB**

Proposed Approach

- Perform Exploratory Data Analysis on the data set
- Perform data preprocessing such as deleting null values, remove duplicates lines and checking for outliers.
- Transforming data

- Apply different machine learning models and test them in order to find the best model

GitHub Repo Link

https://github.com/jamilaHa/Healthcare---Persistency-of-a-drug