Data Science Healthcare Project

Details of Team Member:

Course ID: LISUM04

Name: Zain Ul Haq

Zainulhaq904@gmail.com

Marzie Hossein
shosseini@uni-potsdam.de

Bao Nguyenkhan
nguyenkhanhbao8695@gmail.com

Country: Germany

Specialization: Data Science

Problem Description:

One of the fundamental challenges in the pharmaceutical industry is to comprehend the persistency of drugs as per the physician's prescription. To address this problem “ABC Pharma Company” approached the Data Analytics Company to automate this process and identify the persistency of drugs among patients. ABC Pharma Company provided the recorded data with several attributes contained in an Excel file. The dataset contains four independent features such as for each patient (Unique patient ID), patient Demographics (Provider: Doctor/Medical Staff/ Nurse) attributes, Clinical Factors, and Treatment factors. In this dataset, the target variable (independent variable) is classified as a persistency flag for the patient that means understanding whether the patient is persistent with their medication or not.

The dataset contains a few demographic features such as Age, Race, Region, and Ethnicity. Attributes of the physician who prepared the prescription for the patient and this independent variable is supposed to be an important predictor. The disease which is considered in this project is Nontuberculous Mycobacterial (NTM). Numerous scans have been performed for NTM and the evaluation metric is used as T-score for this disease. Clinical factors have also been considered such as results of these tests during the Rx and the performance changes in the last one or two years, along with other factors of Risk segment, multiple risk factors, and tracking of risk segment after getting therapies. Other treatment factors have been considered such as comorbidity disease i.e. other diseases along with NTM and concomitancy of multiple drugs that have been used for curing the NTM disease. All of these features and attributes help in building up the automated Machine learning model to correctly identify the patients based on their “Persistency Flag”. Efforts have been invested to identify the most influential features which help in understanding the person’s choice for continuing the medicine.

Business Understanding:

According to WHO, non-persistent long-term medication for diseases such as hypertension and diabetics is a common problem that leads to health benefits and economic consequences such as waste of money and time, uncured disease [1]. The persistence of drugs is an open topic of research and researches is conducted to analyze what factors are contributing to the non-compliance behavior of patients towards

drugs. As per the article on DTC perspectives, it is shown that most of the people in the US are not persistent with the drugs. Pharmaceutical companies, Medical organizations, and hospitals lose $100 billion per year along with 125,000 deaths per year due to the non-persistent behavior of patients [2].

Data Intake Report

Data Science Project: Persistency of Drugs

Report Date: 10/11/2021

Data Intake by: Zain Ul Haq

Total number of observations	3424
Total number of Files	1
Total number of features	69
Format of the File	.xlsx
Size of the Data	899 KB

Proposed Approach:

- We will perform Exploratory Data Analysis on the given dataset
- Implement data preprocessing pipeline such as filtering null values, looking for outliers in the data, checking the datatypes.
- Transform the data into their correct data format
- Apply different ML models and test them with different scenarios and find the optimal model for this dataset.

References

[1]. Lee, Ka & Fader, Peter & Hardie, Bruce. (2007). How to Project Patient Persistency. Foresight: The International Journal of Applied Forecasting. 31-35.

[2]. Cramer, J., Benedict, Á., Muszbek, N., Keskinaslan, A. and Khan, Z., 2021. *The significance of compliance and persistence in the treatment of diabetes, hypertension and dyslipidaemia: a review*.