Data Glacier

Your Deep Learning Partner

PERSISTENCY OF A DRUG: GATHERING INSIGHTS FOR A PHARMA COMPANY

Name: Shruthi Madgi
Email:Shruthi.madgi05@gmail.com
Country: USA
University: Visvesvaraya Technological University
Specialization: Data Science
Batch no:LISUM17
Reviewer: Data Glacier

Problem Description:

One of the challenge for all Pharmaceutical companies is to understand the persistency of drug as per the physician prescription, i.e whether a patient will be persistent in completing his/her dose. To solve this problem, ABC pharma company has approached an analytics company to automate the process of identification.

Objective is to build a classification model for drug persistency identification.

Exploratory Data Analysis:

1.What are the most common risk factors?

Most of the patients already hold comorbidity factors, while having risk factors is much less common.

- The main comorbidity factor is related to lipoproteins, metabolism and cholesterol.
- The main risk factor is deficiency in Vitamin D.
- More than one third have been found to be taken narcotics.

2. How do risk factors relate to demographics?

There are some significant differences between genders:

- Women seem to be more affected by vitamin D deficiency,
- There is more than a twofold difference in the number of women and men who have passed a screening for malignant neoplasms, with women passing in greater numbers.
- Four times as many men than women suffer from Hypogonadism (untreated).
- Patients older than 65 are affected by the mentioned factors in a higher proportion.
- There are some risks and other factors that seem to be significantly higher in South and West regions.
- There seem to be some remarkable differences between Asian and other races(due to either culture or race).

3.What is the proportion of patients who were affected by the treatment and had a fracture?

Of the total number of patients, 8.38% of people were affected by the treatment, weakening their bones.

4.Does the specialty of the person who prescribed the drug have any effect on the persistent rate?

The distributions of frequency for the target variable by specialty are pretty similar. There does not seem to be any effect of specialty.

5.Does 'Ntm_Specialist_Flag' and 'Ntm_Speciality_Bucket' variables have useful information for the classification task?

Variables that are recorded during the treatment have more useful information for the classification than others. It can be checked with the percentages shown by Dexa_During_Rx variable.

6.What is the proportion of patients who were affected by the treatment, decreasing their t-score?

There is 10.31% of people with treatment who had a decrease in the t-score

7. Does the gender play a role in the chances of a drug being persistent or not?

60.31% of males were flagged as non-persistent.

62.48% of females were flagged as non-persistent.

Both the genders seem to experience same result with the drug persistency.

Recommendations:

1. Decision Tree as a model is a perfect fit since the main predictors are comorbidity factors and Rx. Since the data is clean and free of outliers, a decision tree can predict accurately about the adherence of the drug with patients. A decision tree as a model is also fast, simple to execute and use.
2. A Random Forest will also work and gives greater accuracy, however; it is slow and uses lot of computation power. It is also not easy to interpret end result of a RF model.

In conclusion, a Decision tree will work for the as per the EDA.

GitHub repo:https://github.com/shru0405/DataGlacierInternship/tree/main/Week_10