

Project: Healthcare - Persistency of a drug

Week 9: Deliverables

Name: Amima Shifa

Email address: shifamima03@gmail.com

University: Osmania University

Country: India

Specialization: Data Science

Batch: LISUM12

Submitted to: Data Glacier

Problem Description:

ABC is a pharmaceutical company that wants to understand the persistency of a drug as per the physician's prescription for a patient. This company has approached an Analytics company to automate this process of identification. This Analytics company has assigned this task as part of the internship and has asked to come up with a solution to automate the persistence of a drug for the client ABC.



Data Cleaning and Transformation:

The dataset contains 70 columns and 3424 rows in total. One of the features of the given dataset is 'Ptid' which is a patient identification number and it plays no role in the prediction of the persistence of a drug on a patient and hence is removed. There are no missing values present in any column of the dataset. 'risk_segment_during_rx', 'tscore_bucket_during_rx', 'change_t_score' and 'change_risk_segment' contained unknown values, so these columns were dropped from the dataset. Binary values present in the dataset were mapped from Y, N to 1,0 respectively. Although a lot of outliers were detected in Dexa Frequency during RX which was then transformed using log transformation.

GitHub Repository:

<https://github.com/AmimaShifa/WEEKLY-TASKS/tree/main/Week-9>