



Project: Healthcare - Persistency of a drug

Week 9 Deliverables

Batch code: LISUM20

Submission date: 2nd June 2023

Submission to: Data Glacier

Reviewed by: Nwankwo Ezinne Anasthecia

Project Team:

| Group Name: Ensemble Elites | | | | |
|-----------------------------|--|---------|-----------------------------------|----------------|
| Name | Email | Country | College/Company | Specialization |
| Nwankwo Ezinne Anasthecia | nwankwoanasthecia@gmail.com | Nigeria | Freelance | Data Science |
| Bindu Musham | bindu.musham@gmail.com | USA | The University of Texas at Dallas | Data Science |

Problem Description:

ABC Pharma is facing a significant challenge in understanding the persistency of drug usage as per physician prescriptions. The persistency of drug usage refers to the extent to which patients continue to take their prescribed medications over a specific period of time.

Currently, ABC Pharma Company relies on manual methods to track and analyze drug persistency. This involves reviewing patient records, conducting surveys, and



relying on self-reporting, which can be time-consuming, prone to errors, and lack real-time insights.

To address this challenge, ABC Pharma Company has decided to approach an analytics company to automate the process of identifying drug persistency. The goal is to develop a data-driven solution that can accurately and efficiently track the usage of prescribed medications by patients, enabling ABC Pharma Company to gain valuable insights into the patterns of medication adherence.

Data Understanding:

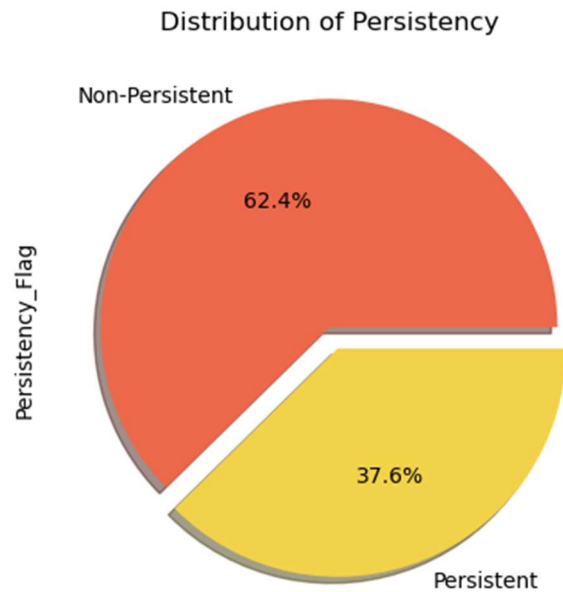
The target variable in the dataset is the persistency flag. This flag indicates whether a patient was persistent or not in taking their prescribed medication. It serves as the target variable for the analysis, with values such as 'persistent' or 'non-persistent'. Hence the persistency flag is the dependent variable which means other factors determine whether a patient was persistent or non-persistent with the drug.

Adherence refers to the extent to which patients follow their prescribed medication regimens. It is an important factor in assessing the effectiveness of treatments and patient outcomes. The adherence-related features in the dataset provide insights into the patient's medication adherence behaviors.

Understanding adherence patterns and factors affecting adherence is crucial for pharmaceutical companies like ABC Pharma to evaluate the effectiveness of their drugs and develop strategies to improve patient adherence. By analyzing the adherence-related features, ABC Pharma can gain insights into medication-taking behaviors, identify potential barriers to adherence, and tailor interventions to improve patient compliance and persistency.

Target Analysis before transformation:

There are 37.6% of patients are persistent in taking prescription drugs while 62.4% are non-persistent.



Data Cleansing:

Null values:

There are no null values in the dataset. So, no need to handle NA or missing values.

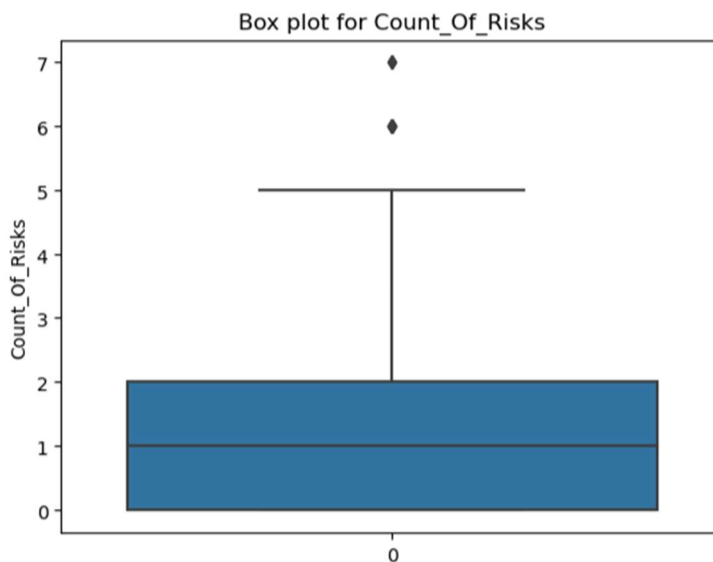
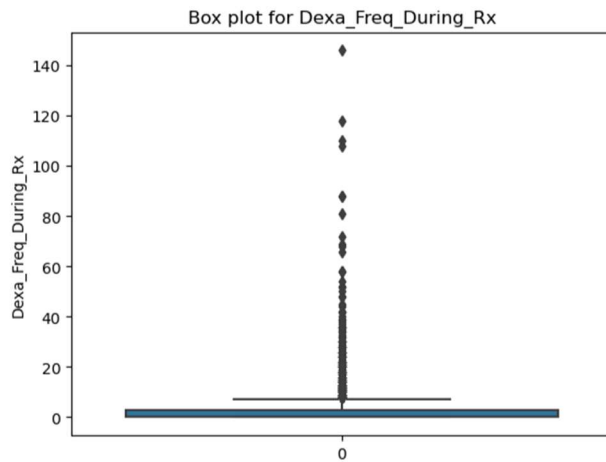
Duplicate values:

There are no duplicate values in the dataset.

Outlier Analysis:

There are outliers for DEXA_Freq_During_Rx(number of DEXA scans taken during Rx) and Count_Of_Risks(count of risks).

Identified Outliers using Inter Quartile range and Box plot and removed the outliers.



Data Transformation:

Deleted Ptid (Patient id) which is not an important feature to understand factors affecting persistency of a drug.

Mapping Categorical values to numeric:

We have converted all the categorical values to numerical values for model training.

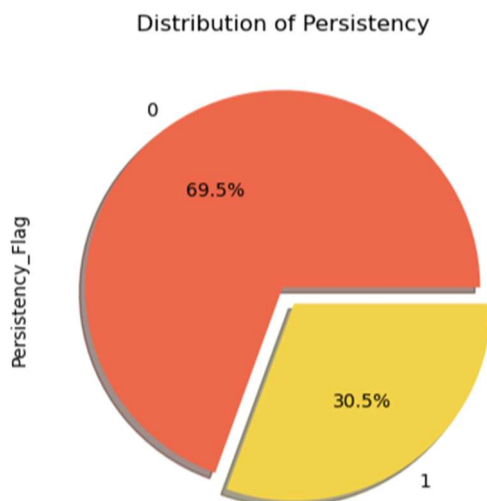
1. Converted all Y and N values to 1 and 0.



2. Converted Gender, Ntm_Specialist_Flag, Adherent_Flag and Persistency_Flag to 0 and 1.
3. Converted categorical features with multiple values to numeric using One Hot encoding.

Target Analysis after transformation:

There are 30.5% of patients are persistent in taking prescription drugs while 69.5% are non-persistent.



After data transformation, the observations of data set are 2956 with 68 columns.

Standardization:

Standardized values using MinMaxScaler(), So there will be no bias in the model.



Proposed Approach:

- There are no missing values in the dataset.
- There are no duplicate values in the dataset.
- Identified outliers and removed using the Inter Quartile range and box plot.
- Mapped Categorical features to numeric values.
- Standardized values using MinMaxScaler().

GitHub Repo link:

<https://github.com/bmusham/Ensemble-Elites-Healthcare-Persistency-of-drug/tree/main/Week%209>