

WEEK 8 REPORT

TEAM MEMBER DETAILS

Group Name: **DS HealthCare group**

Name: Bolutife Akinlawon

Email: bolu.akinlawon@gmail.com

Country: United Kingdom

College/Company: N/A

Specialization: Data Science

PROBLEM DESCRIPTION

This week's task is to perform some Exploratory Data Analysis on the data and draw several insights from it. It is a way to understand the dataset provided for future building of the classification model.

DATA UNDERSTANDING

The dataset provided is a collection of patients on a trialed drug to determine the factors that contribute to persistency of the drug. These factors can help build a model which can then be used as a framework for administration of the drug.

The dataset has 68 variables that can be used for the prediction as well as 3424 samples of individuals on the trial.

Some insights drawn from the dataset:

- Ntm_Speciality(Physician Speciality) has the most number of variability with 36
- There are more non-Persistent instances than Persistent
- More Female patients were tested than male patients
- The above 75years age group is mostly represented here and decreases drastically going down age groups.

One of the major problems of the data is the imbalanced nature of the data and it needs attention going forward.

There are no null values in the dataset.

We intend to remove the 'PtId' column because it is not relevant to the analysis and model building.

GITHUB REPO LINK: https://github.com/BoluAkinlawon/Data-Glacier-Week-8/blob/main/EDA_Data_Science.ipynb