

TEAM MEMBERS DETAILS

GROUP NAME: Health Research

Name	Email	Country	College/Company	Specialisation
Kelvin Mpofu	mpofukelvintafadzwa@gmail.com	South Africa	n/a	Data science
Purity Nyagweh	purityeverniter@gmail.com	Kenya	n/a	Data Science
Reshma Jayapalan	reshma.jayapalan@gmail.com	UAE	n/a	Data Science
Alhanouf Alghamdi	hanouf.haz@gmail.com	Saudi Arabia	n/a	Data Science

Problem Description

ABC pharma company has a challenge in understanding drug persistency as per physician prescription and to solve this problem it wants to automate the process of identification.

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

This will automate the process of identifying drug persistency for ABC pharma company thus helping the company to understand drug persistency as per physician prescription.

GitHub Repo Link

https://github.com/mpofukelvintafadzwa/Data_Glacier_Group_Project/tree/main/week%2010

EDA performed on the data

1. Qualitative Analysis
 - Class separation by categorical features
 - Risk ratio for categorical features
2. Quantitative Analysis
 - Mutual Information scores for categorical features, measures mutual dependence between the categorical features and the target variable.
 - Correlation for the numerical columns against target variable.
 - Chi square test (test of independence), measures how likely it is that any observed difference arose by chance.
 - Analysis of Variance (ANOVA), tells if the target variable changes to the level of the independent variable.

Final Recommendation

Only those features which are found to be significant should be used when building machine learning models.

