**Details**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Email | Country | College/Company | Specialization |
| Kelvin Mpofu | mpofukelvintafadzwa@gmail.com | South Africa | n/a | Data Science |
| Purity Nyagweth | Purityeverniter@gmail.com | Kenya | n/a | Data Science |
| Reshma Jayapalan | reshma.jayapalan@gmail.com | UAE | n/a | Data Science |
| Hanouf Hazza | 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>

**Data cleansing and treatment methods**

1. **For Outlier detection and treatment**

* Boolean Method
* Boxplots
* Z-score

1. **Identifying Skewed data and treatment**

* Histograms – Identifying
* Skewness and kurtosis measurement – Identifying
* Boxplots – Identifying

**For treatment**

* Log transformation
* Reciprocal transformation
* Square root transformation
* Min-max transformation
* Trimming outliers

1. **Identifying and treating class imbalance**

Used the method in python ‘dataframe[col].value\_counts()’ to identify the imbalance.

Oversampled the minority class to treat class imbalance.