**TEAM MEMBERS DETAILS**

GROUP NAME: PACHACUTEAM

SPECIALIZATION: Data Science

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Email | Country | College/Company |
| Andersson Romero | andromdez@gmail.com | Perú | Freelancer |

**PROBLEM DESCRIPTION**

Classification the process of identification of the persistency of the drug as per prescription provided by the physician.

**BUSINESS UNDERSTANDING**

Understanding the persistency is an issue for pharmaceutical companies and so, ABC pharma company has approached XYZ analytics company to provide insights into the same.

The role of XYZ analytics company is to undertake this project and provide a detailed understanding regarding the drug persistency.

**PROJECT LIFECYCLE**

* Step 1

Problem Description

Business understanding

Project lifecycle along with deadline

* Step 2

Data understanding

What are the problems in the data (number of NA values, outliers, skewed etc)

* Step 3

Data cleansing and transformation done on the data.

Try at least 2 techniques to clean the data

* Step 4

EDA performed on the data

* Step 5

Model Selection and Model building

* Step 7

Final Report, Code, Ppt submission

**DATA INTAKE REPORT**

Name: Data Science – Healthcare project – Persistency of a drug

Report date: 07/04/2022

Internship Batch: LISUM07

Version:1.0

Data intake by: PACHACUTEAM

Data intake reviewer:<intern who reviewed the report>

Data storage location: GitHub

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 3424 |
| **Total number of files** | 1 |
| **Total number of features** | 69 |
| **Base format of the file** | .xlsx |
| **Size of the data** | 888 KB |

**GITHUB LINK:** https://github.com/And2300/Healthcare\_PersistencyDrug