|  |  |
| --- | --- |
| Internship Project Title | RIO-125: Classification Model - Build a Model that Classifies the Side Effects of a Drug Batch |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | HIMALAYA AASHISH |
| Name of the Institute | ICTAK - TRIVANDRUM |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Start Date | End Date | | Total Effort (hrs.) | | Project Environment | Tools used |
| 01 March 2021 | 10 May 2021 | | 50 | | python | Pandas, Jupiter, numpy |
| Milestone #1 | 2 | Milestone: | | Day 15 ,Split the dataset into training and testing sets | | |

INTERIM PROJECT REPORT 2

**RIO-125: Classification Model - Build a Model that Classifies the Side Effects of a Drug Batch**

The project aims at building a Classification Model for the side effects of a drug.

**1. Objective**

* To build a classification model that classifies the side effects of a drug

**2. Activities upto Day 15**

* **Creating dataset** – The dataset was created using faker library
* **Cleaning of dataset –** As the dataset was created there were no null values, missing or other anomalies
* **Preprocessing - Checked for null values, missing or other anomalies**
* **Feature Engineering**
* **Splitting the dataset into training and testing sets**

**3. Approach / Methodology**

# DATA ANALYSIS

## 3.1. Importing Python libraries

The libraries and packages imported are: pandas, numpy, matplotlib and seaborn, visualization and sklearn.

## 3.2. Importing the data

The dataset data.csv is loaded to python using the read\_csv function from pandas.

data = pd.read\_csv('data.csv')

## 3.3. Data source

The data set data.csv was created using kaggle reference

## 3.4. Dataset Description

The dataset has 3108 rows and 5 columns

## 3.5 . Preprocessing

## Checking Missing Values

* There were no missing values

## 

## 3.6 Encoding

Used label encoder and one hot encoder

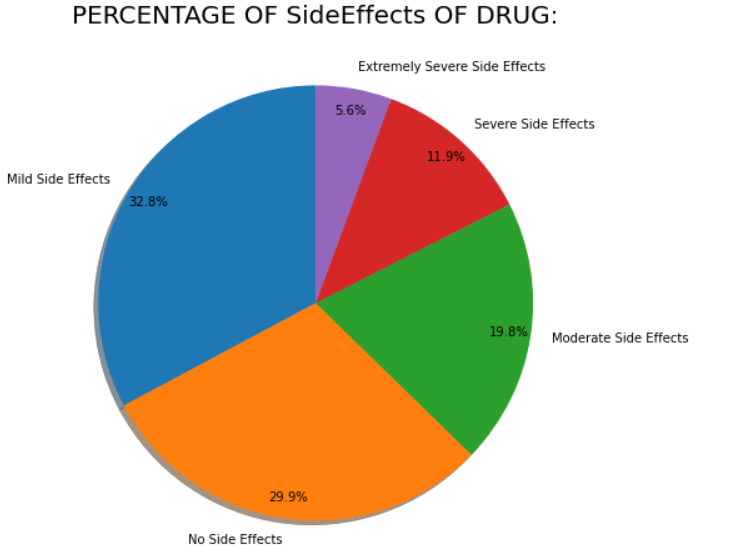
**4. Charts, Table, Diagrams**

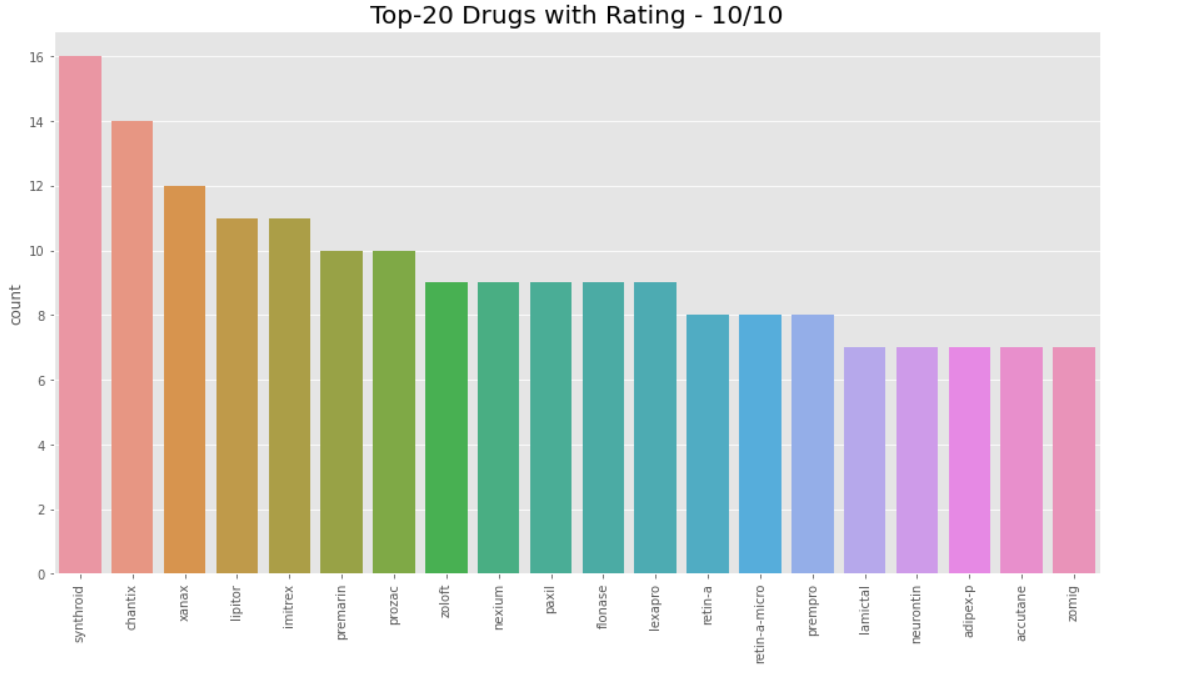
Bar Charts, Pie charts and Multiple Pie Charts were drawn for distribution of side effects with various features.

* Libraries – matplotlib and seaborn were used to plot different features.

5.1 Distribution of Side Effects of Drug

* Bar Charts and Pie charts were plotted for the distribution of Side Effects of Drug





* Outcome / Conclusion

A dataset for Classification of side effects of drug with features was created and analyzed.