|  |  |
| --- | --- |
| Name of the Student | Bryan Menezes |
| Internship Project Topic | Classification Model for Drug trial |
| Industry Mentor | Dr. Pradheep |
| Institute Name | CSMU-Chhatrapati Shivaji Maharaj University |

TCS iON Remote internship

RIO-210

|  |  |  |
| --- | --- | --- |
| Date | Day | Hour Spent |
|  | DAY 3 | 3.5 |
| Activities done today—Finding a Dataset for the classification model of drug trial:  Finding the drug trial: -  We are to find a data set at least 4,00,000 patients containing the following details for each patient:   * Name * Age * Gender * Race * Side effect   To find such a data set we will use the website Kaggle is a data science platform and online community of data scientist and machine learning professionals used for helping individuals for create projects to enhance their capabilities. It mainly contains datasets that are useful to help individuals like me to find datasets for projects.  After searching through Kaggle I have found the dataset I am to use to build the classification model:  About the dataset: WebMD Drug Reviews Dataset  The dataset provides user revies on specific drugs along with related condition, side effects, age, sec and rating reflecting overall patient satisfaction.  The data acquired by scrapping WebMD site. There are around 0.36 million or 3.6 lakh rows of unique reviews and is updated till march 2020.  The dataset: [WebMD Drug Reviews Dataset (kaggle.com)](https://www.kaggle.com/datasets/rohanharode07/webmd-drug-reviews-dataset) | | |