Internship Project Title	Shakchi Prasad
Name of the Company	RIO-125: Classification Model - Build a Model that Classifies the Side Effects of a Drug
Name of the Industry Mentor	TCS iON
Name of the Institute	Debashis Roy
	PrepInsta Pvt Ltd.

Start Date	End Date	Total Effort (hrs.)	Project Environment	Tools used
05-07-2023	12-07-2023	27	Remote	Youtube, Google Sheets, Kaggle, etc.

Milestone #	2	Milestone:	Creating the dataset required for the project

TABLE OF CONTENT

Acknowledgements

I would like to thank TCS for the opportunity to be a part of the RIO – 125 Program. I am grateful for the staff at TCS for diligently curating the RIO experience and looking over my work on the same. Lastly, I am greatly thankful to my Institute – PrepInsta Pvt. Ltd. for creating this opportunity and allowing me to be go on this journey.

Objective

To build a Classification model that classifies the Side effects of a Drug.

Description of Internship

The TCS RIO – 125 is an initiative that allows people to complete an internship of 30 days over the course of 3 months, comprehensive of 125 hours. At its base it is designed to teach the intern to build an industry grade project in the above timeframe.

Internship Activities

Self-learning:

- Continued self-learning the classification algorithms.
- Decided to use ensemble model
- Learnt all the ways of data pre-processing

Dataset building:

- Curated data from various websites, specifically Kaggle.
- After data gathering, went forward with data cleaning and pre-processing
- Built a clear data set with appropriate features required for the model.

Industry Project – dataset

Dataset:

It was gathered over the course of a few days in the project and thoroughly pre-processed to convert into the dataset shown below, i.e., a form that would be most affective to train the model on its details of effectiveness and make the correct prediction of its side effects (Mild, Severe, Negligible, etc.)

The dataset used in this project has 3106 records and the following columns:

1. Drug Name
2. Rating
3. Effectiveness
4. Condition
5. Side-effects

	urlDrugName	rating	effectiveness	condition	sideEffects
0	enalapril	4	Highly Effective	management of congestive heart failure	Mild Side Effects
1	ortho-tri-cyclen	1	Highly Effective	birth prevention	Severe Side Effects
2	ponstel	10	Highly Effective	menstrual cramps	No Side Effects
3	prilosec	3	Marginally Effective	acid reflux	Mild Side Effects
4	lyrica	2	Marginally Effective	fibromyalgia	Severe Side Effects

The above snippet showcases the first 5 columns of the dataset. The entire dataset can be found at the link provided underneath.

Dataset:

https://docs.google.com/spreadsheets/d/110dJ2mMBMKkT8OWt1h1QCdBXwWTvysx93R-lcwOSQ9A/edit?usp=sharing