ACTIVITY DETAILS OF PARTICIPANT

Course Name : RIO-125: Classification Model - Build a Model that Classifies the Side Effects of a Drug Batch 01

Course Duration : 90 Days

Course Type : Self Paced

Participant Name : Vardaan Vishnu

Report Generated At : 02 Jan 2023 05:44 PM

Participant Email ID : vardaan209@gmail.com

Activity Type	Activity Name	Attempt No	Total Marks	Pass Marks/CutOff	Marks Obtained	Grade	Activity Attempted	Activity Status	Assignment Link
Assessment	RIO Pre-Assessment	1	40.0	20	28	NA	Completed	Pass	N.A.
Report	Viva Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Project Report	1	30.0	0	24.5	A	Completed	NA	N.A.
Report	Day 1 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 2 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 3 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 4 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 5 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 6 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 7 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 8 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 9 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 10 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 11 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 12 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.

Activity Type	Activity Name	Attempt No	Total Marks	Pass Marks/CutOff	Marks Obtained	Grade	Activity Attempted	Activity Status	Assignment Link
Report	Day 13 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 14 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 15 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 16 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 17 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 18 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 19 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 20 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 21 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 22 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 23 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 24 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 25 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 26 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 27 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 28 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 29 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Day 30 - Activity Report	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Interim Project Report - 1	1	NA	NA	NA	NA	Completed	NA	N.A.
Report	Interim Project Report - 2	1	NA	NA	NA	NA	Completed	NA	N.A.
Assessment	Project Test	1	10.0	5	6	NA	Completed	Pass	N.A.

INTERNSHIP: PROJECT REPORT

Internship Project Title	Classification Model - Build a Model that Classifies the Side Effects of a Drug
Name of the Company	TCS iON
Name of the Student	Vardaan Vishnu
Name of the Industry Mentor	Debashis Roy
Name of the Institute	VIT BHOPAL UNIVERSITY

Start Date	End Date	Total Effort (hrs.)	Project Environment	Tools used
7-11-22	2-01-23	125	VS Code, Chrome, Windows 10, GitHub	Python, Colab, libraries

TABLE OF CONTENT

- Acknowledgements
- Objective
- Introduction / Description of Internship
- Internship Activities
- Approach / Methodology
- Charts, Table, Diagrams
- Algorithms
- Challenges & Opportunities
- Reflections on the Internship
- Outcome / Conclusion
- Enhancement Scope
- Link to code and executable file

Acknowledgements

I thank TCS iON for the guidance and support as well as for providing necessary information regarding the project. I also thank my industry guide, Mr. Debashish Roy, and the other mentors for being helpful and co-operative during this internship.

Objective

The objective of this project is to develop a classification model that classifies the side effects of a particular drug by age, gender, and race.

Introduction / Description of Internship

The Purpose of this internship is to build a Classification Model. A Classification models are a subset of supervised machine learning. A classification model reads some input and generates an output that classifies the input into some category. Here we intend to use these tools to create a model that can categorize different side effects that can occur due to age, gender and race.

We use various tools and setup environments for the development of this model.

Since the goal here is to develop a classification model that classifies the side effects of a particular drug by age, gender, and race, we start with gathering data that will help us model the algorithm. We then proceed to clean, sanitize, and order that data into a concise and clean dataset. Afterwards of this we perform certain exploratory analysis in order to gather an idea about the data we have and make sense of it. Then we use Machine Learning algorithms to create models that classify the side effects caused due to age, gender, and race.

Internship Activities

Some of the activities to be done during the internship include Pre-project test, activity report, interim project reports, final project report, project test.

Approach / Methodology

1. Create a dataset. Here we use the WebMD dataset from Kaggle.

2. Clean the dataset and sanitize the data. Datasets usually contain raw and disoriented data and we need to clean it and make it proper for better performance with our classification model.

3. We pre-process the data according to our needs and utilize.

4. So, we remove the outliers, remove missing values, and then use the dataset.

5. We carry out exploratory data analysis.

6. We create various bar plots and pie charts to understand and have firm grasp about our dataset.

7. We then proceed to use only those parameters that help us achieve our goals,

8. Thus, irrelevant columns are not utilized further for data analysis and are dropped instead.

9. We Split the dataset into training and testing sets for the purpose of having different values when training our model and different values when testing it for real world.

10. We the create a Classifier that trains on our training dataset and make a model that fits the data.

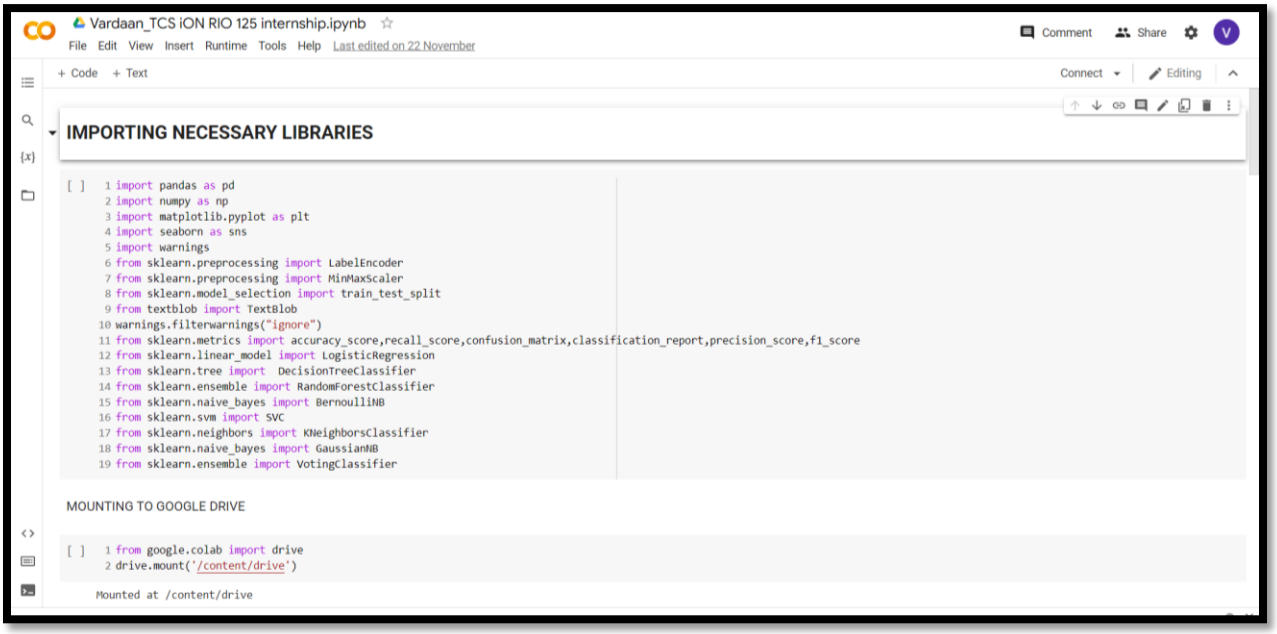
11. Then we run our classification model that on our testing dataset and can make proper inferences from it too.

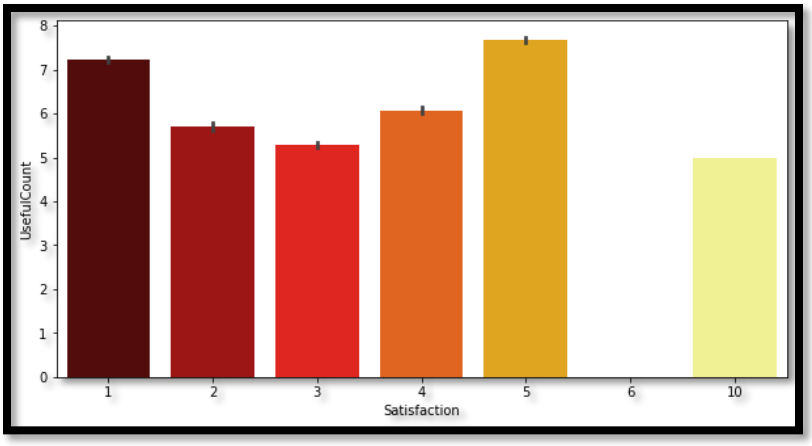
12. We use different algorithms to build our model so that we can gauge which one of the listed methods is best for creating our desired classification model.

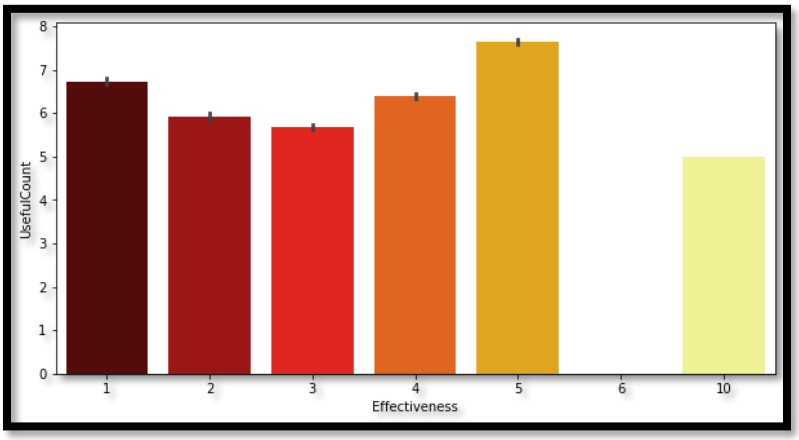
13. We use logistic regression, Naïve-bayes, Random Forest Classifier, Decision Tree and K-Nearest neighbor algorithms to carry out the functionality.

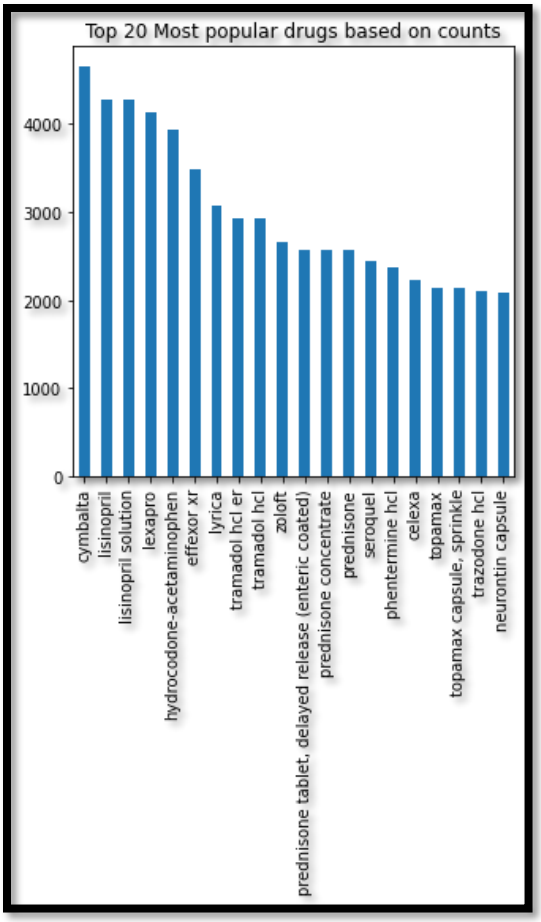
14. We have thus created our classification model and tested it too. Now we make the final analysis and the conclusion and our project is completed.

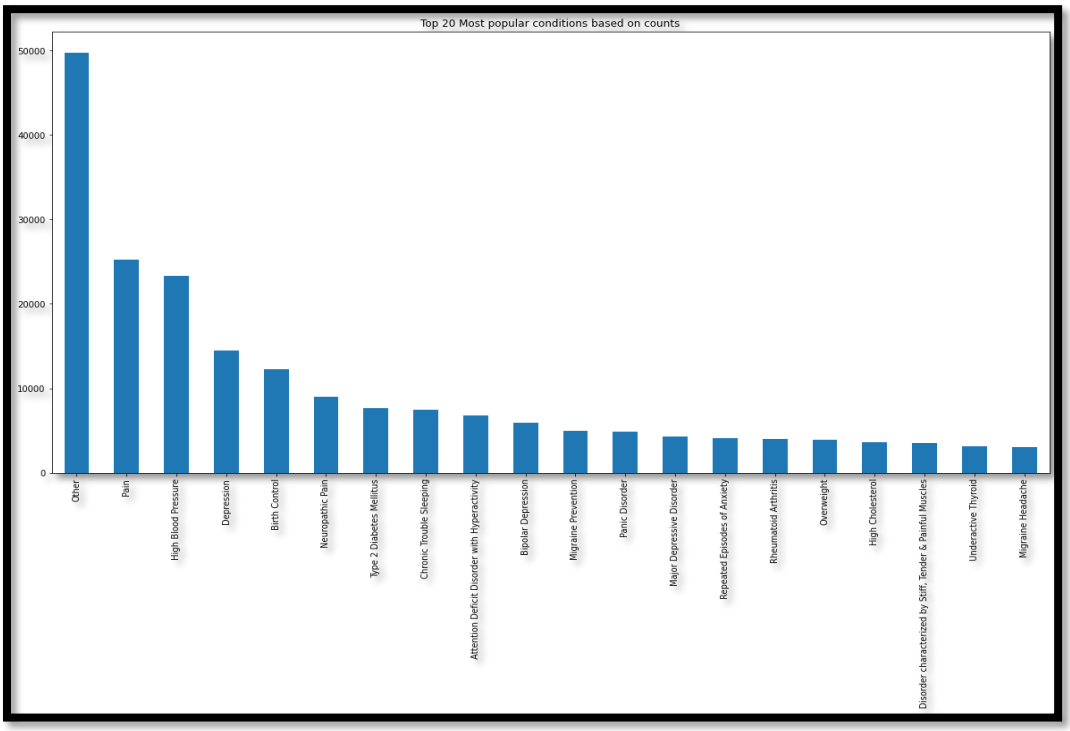
Charts, Table, Diagrams

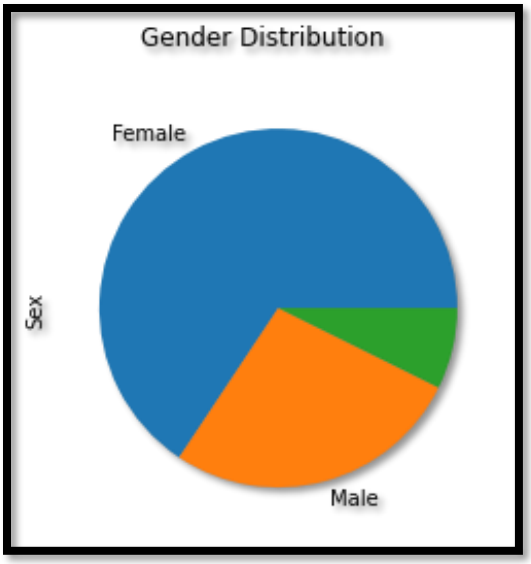


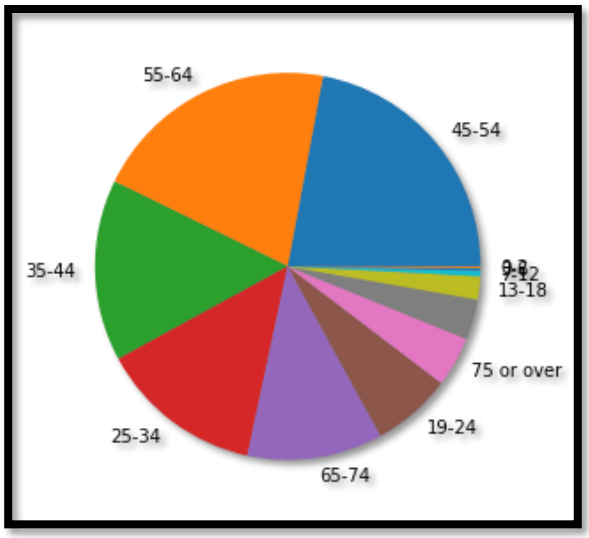

UsefulCount
0
1
2
3
4
5
6
7
8
1
2
3
4
5
6
10
Satisfaction

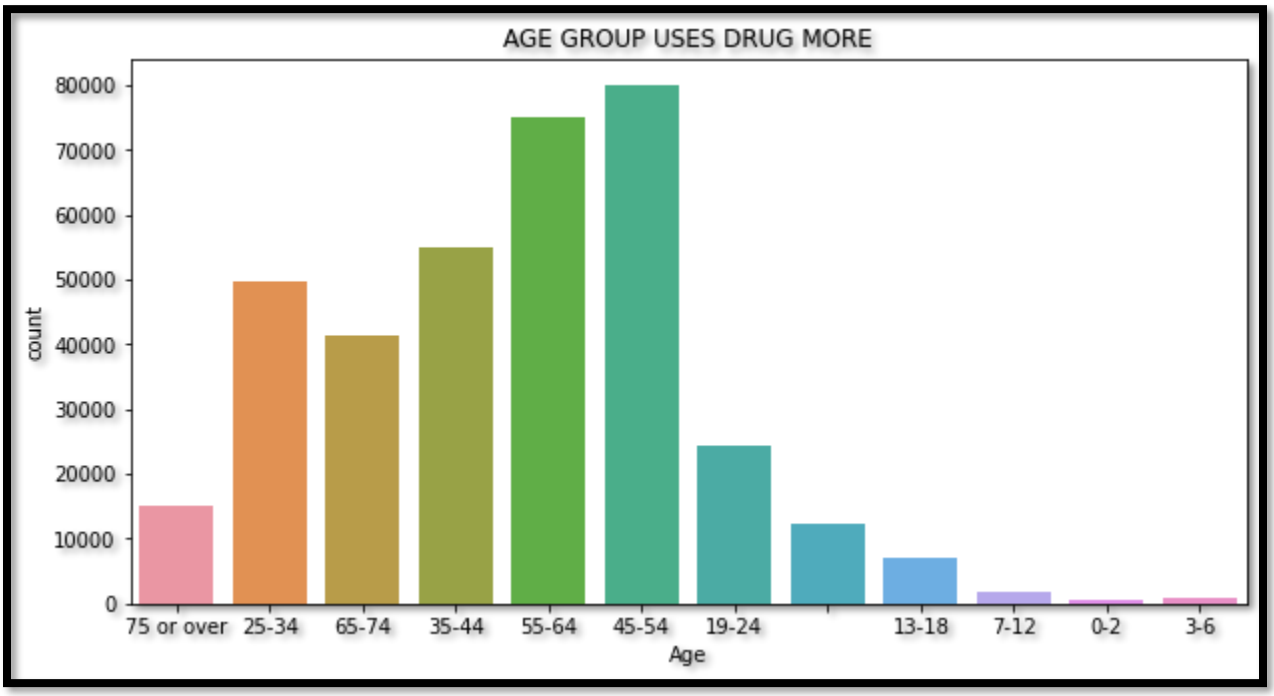

UsefulCount
0
1
2
3
4
5
6
7
8
1
2
3
4
5
6
10
Effectiveness


Top 20 Most popular drugs based on counts
4000
3000
2000
1000
0
cymbalta
lisinopril
lisinopril solution
lexapro
hydrocodone-acetaminophen
effexor xr
lyrica
tramadol hcl er
tramadol hcl
zoloft
prednisone tablet, delayed release (enteric coated)
prednisone concentrate
prednisone
seroquel
phentermine hcl
celexa
topamax
topamax capsule, sprinkle
trazodone hcl
neurontin capsule


Top 20 Most popular conditions based on counts
50000
40000
30000
20000
10000
0
Other
Pain
High Blood Pressure
Depression
Birth Control
Neuropathic Pain
Type 2 Diabetes Mellitus
Chronic Trouble Sleeping
Attention Deficit Disorder with Hyperactivity
Bipolar Depression
Migraine Prevention
Panic Disorder
Major Depressive Disorder
Repeated Episodes of Anxiety
Rheumatoid Arthritis
Overweight
High Cholesterol
Disorder characterized by Stiff, Tender & Painful Muscles
Underactive Thyroid
Migraine Headache







Algorithms

Some of the ML Algorithms used are Naïve Bayes Classification, K-nearest neighbor, Decision Tree, Random Forest, and Logistic Regression. They help us to build the classifier model.

Challenges and Opportunities

It is a big challenge to learn the basics of exploratory data analysis since the field is so wide and needs to be understood thoroughly to have favorable outcomes, it is also an opportunity to learn more and expand my capabilities through this learning experience.

Reflections on the Internship

I am hoping that working as an intern with TCS ion will allow me to gain greater knowledge of classification models and data analytics.

Outcome / Conclusion

As an outcome, we have completed the importing of various libraries, cleaning and sanitizing of dataset, exploratory analysis of the dataset and further splitting of dataset into the training and testing sets. Then we created a simple classifier using logistic regression and trained it on our training set. Finally, we tested our model on the testing set and we got the accuracy of 69%.

Enhancement Scope

There is a lot that can be added to this classification model. We can add more parameters to this model to gauge in other external factors into consideration that we have not till now. We can also focus on adding an interface/GUI for normal users who do not possess the knowledge about data science, so that they can make use of our model with relative ease.

Link to code and executable file

https://github.com/vardaan11/TCS-iON-RIO-125-Internship

Link to Project Video

https://www.loom.com/share/dee4386bbaf5401392aa6fed5e265967