**-Richa Swami**

**(Team- 23)**

**I follows the following pattern in my project:**

Result

Model

Creation

Feature

Engineering

EDA

Data

Preparation

Data

**Data:**

Collection of Data Sets

* Amazone\_reviews\_1.csv
* Link for the data:

<https://drive.google.com/drive/folders/1h5_9VUeCbfrF7sGuIlz1gRbPke7nuw9K?usp=sharin>

**Data Preparation:**

* Data Cleaning: Identifying and correcting mistakes or errors in the data.

In data set, I found one missing value and no duplicate rows present.

* Identifying input variables that are more relevant to the task.
* Adding new features and attributes to the data set (Amazone\_reviews\_1.csv)

**EDA:**

* Importing the Data Sets.
* See the view and shape of the data set.
* Descriptive statistics of the data set.
* Checking about the correlation between features In a data set.
* Checking about data types and missing values in the data.

**Feature Engineering:**

* Imputation: if missing values are present, impute them.
* Encoding categorical features and Standardization of data.
* Scaling: For symmetric dataset scaling is required and used in normalization and standardization.

**Training Model:**

* Train Model to kept 80% of whole data frame split train\_test\_split().
* Using Logistic Regression and more ML techniques.
* Hyperparameter Tunning.

**Test Model & Evaluation:**

* Test Data to kept 20% of whole dataframe test\_train\_split().
* Testing Will be done on test Data based on training Data.

**Model Creation:**

* It is iterative phase where a data scientist continually train and test machine learning models to discover the best one for the task.
* I have applied LogisticRegression, DecisionTreeClassifier, RandomForestClassifier and KNN techniques to sentiment analysis in model building in ML.
* And also we used keras for developing and evaluating deep learning model.

**Results:**

After successful implementation of the model we can conclude the following points:

* LogisticRegression is correct choice for this dataset.

**ML Model:**

Accuracy of LogisticRegression model:0.919223

Accuracy of DecisionTreeClassifier model: 0.793278

Accuracy of RandomForestClassifier model: 0.894180

Accuracy of KNN model: 0.813264

**DL Model:**

Accuracy of the model Before adding dropout: 0.8625

F1-score: 0.8568453930244664

Confusion matrix:

array([[902, 127],

[148, 823]])

Accuracy of the model After adding dropout: 0.88

F1-score: 0.8775510204081634

Confusion matrix:

array([[900, 129],

[111, 860]])