You have to scrape at least 20000 rows of data. You can scrape more data as well, it's up to you. more the data better the model In this section you need to scrape the reviews of different laptops, Phones, Headphones, smart watches, Professional Cameras, Printers, Monitors, Home theater, Router from different e commerce websites. Basically, we need these columns 1) reviews of the product. 2) rating of the product. You can fetch other data as well, if you think data can be useful or can help in the project. It completely depends on your imagination or assumption.

We first scraped the data from different e commerce websites and make a dataframe out of it.

1)DATA CLEANING: we must check the data to be cleaned..as it contains some of the null values.We must remove the null values using the dropna()function.there are no duplicates found..we are good to go.

2)EXPLORATORY_ANALYSIS:We must check the distributions of our wanted columns or variables.we only require the ratings and reviews columns.we drop the rest of the columns.

3)DATA PREPROCESSING: since the reviews column is the string datatype,we must use the NLP preprocessing techniques.

First we convert the whole text into lower case letters,then we remove punctuations,tabs..etc

Then we tokenize the text,Remove the stopwords and Lemmatize the text.

Then we map the ratings to 0,1.

Next we use TFIDF technique to convert into machine readable texts

So our data is ready to train

4)MODEL BUILDING: we split the data into train and test.we train the data into different models such as multinomial NB,XGBclassifier,Randomforestclassifier,DecisionTreeClassifier and GradientBoostingClassifier.

5)MODEL EVOLUTION:we check the each model performance using the test data and we keep track of the model which performed well with good accuracy_score,classification_report.In our case RandomForestClassifier worked better when compared to other model interms of accuracy.

6)Selection of the best model after HyperParamter tuning is RandomForestClassifier