Additional Model -1 Sentiment Classifier :

Sentiment analysis is the process of detecting positive or negative sentiment in text.

We will build classification algorithms to detect the text

Steps:

1. Standard NLP Preprocessing
2. Labelling sentiments 1,2,3 ratings as negative and 4,5 as positive and assign them as classes
3. Build a classifier using various algorithms and check for performance

Logistic regression: Logistic Regression is a classification that serves to solve the binary classification problem. The result is usually defined as 0 or 1 in the models with a double situation.

Performance Metrics for Logit Model:

Test accuracy 0.9120063609859528

Train accuracy 0.929074522871857

Random Forest: Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees.

Test accuracy 0.9356833642547928

Train accuracy 0.9995834595577098

Given this data set is class imbalance, the incremental advantage of performance tuning measures such as K-fold cross validation and Grid search will be of little help.

Additional Model-2: Single Product Analysis: