Here are the results;



The Text Data has been preprocessed with PorterStemmer and WordNetLemmatizer, also stopwords has been used as hyper-parameter.

The best accuracy score  of 91% is Logistic Regression with CountVectorizer  model, followed with 90% accuracy score by Logistic Regression with TfidfVectorizer, but the fit time is considerably long of 6-7 sec.

The shortest fit time is 0.86 and 0.88 sec are done by KneighborsClassifier with CountVectorizer, and with TfidfVectorizer. The KneighborsClassifier model has lowest accuracy scores of just above 80%.

The Decision tree models have the longest fit time with 113 sec and 80 sec for CountVectorizer, and TfidfVectorizer, also the accuracy score is average between the models at about 85%.

Naïve Bayes models are second best performer between the models regarding the accuracy score of 89% , also not bad fit time between 1-2 sec.

The conclusion is the Logistic Regression Model outperforms Naïve Bayes, Decision tree and Nearest Neighbors Classification, also CountVectorizer outperforms TfidfVectorizer.

SVM models have been tried but it took too long to complete, so it has been excluded from the final result.