**DATA SCRAPPING**

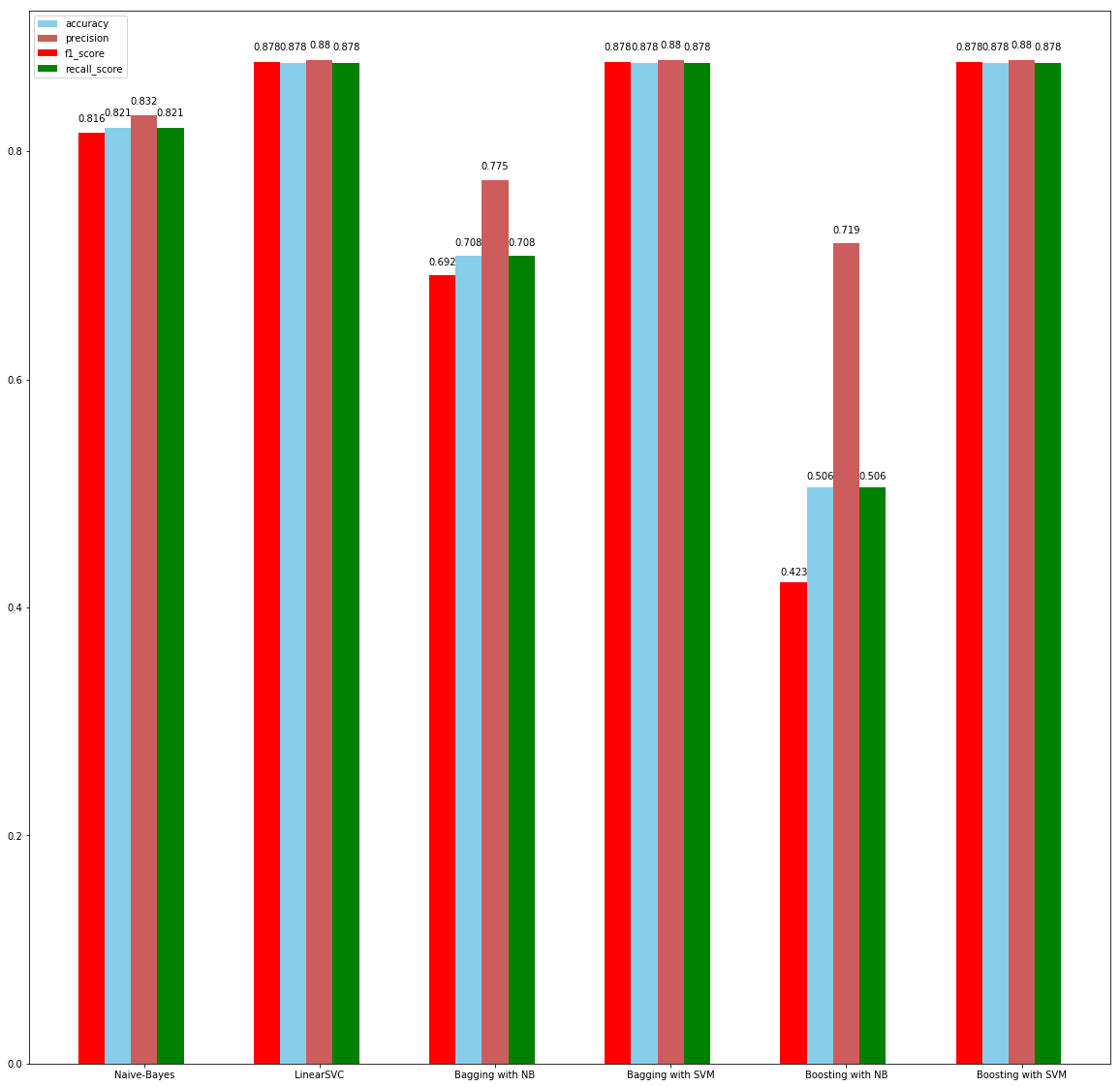
I scraped the data using selenium library of python. Since not much data was available through YouTube alone, I scraped DailyMotion also using the same technique for both of them. I used multithreading to increase the performance and decrease the time taken by the program. Since the internet connection at my place was not good because of technical problems in the hostel, I was unable to scrap all of the link but I am sure if you run the script with a good internet connection, you will be able to scrap all the links.

**MODEL SELECTION**

I have previously worked on sentiment analysis through textual data of comments and during that period I observed that SVMs and Naive Bayes work best for text-based data classification. Since I didn’t know which algorithm would have been better for this kind of dataset, I used both of them!

For the second category I used Advanced Boost and Bagging classifier. I these two because I recently studies them but never got a chance to use them. Bagging is used for reducing the variance of the model whereas boosting is used for reducing high bias.

I currently have no experience in Deep Learning so I didn’t choose any DL based algorithm.



**RESULT**

The reason for SVM performing better than Naive-Bayes is the independence of words with each other. In NB we ‘assume’ that the features will not be co-related but as we can know that text in the description of a video is co-related to each other and since NB is not a weak learner, the after-effect of boosting iss this reduce in metrics.