**Report**

Data extraction technique used: This assignment is a bit challenging because I have to extract the data from websites like youtube. youtube is not s static website, data present in such website comes from API call and that part makes scarping pain like a hell.

you can also extract data from Youtube data API but to a certain limit.

In order to extract data from youtube, I had spent hours to understand the structure in order to find a loophole to get data. I noticed that after scrolling down to the end of the page, you can see all required information in inspect element

from there I just save that HTML structure in my local machine.

Modeling:

category 1: naive Bayes

the reason why i have chosen this model because in naive Bayes because such model works better when we have a dataset in which we have the number of features more than the number of data. in Naive Bayes we make an assumption that every feature is independent of others.

Category 2: Booting algorithm

In the second category I have tried on both bagging and booting algorithm, but the results are pretty same.

it may be because i did not use sufficient data to train model.

category 3: CNN

all other models that are mentioned in category 3 are feedback system. such models are used when during training model need to remember the previous input like in case of text generation and this is not our requirement.

CNN is a feedforward algo mostly used in case of image processing when we have a larger dimension of the input.