**Part A: IMDb Movie Review Sentiment Analysis**

*Explanation video link- https://drive.google.com/file/d/1Xc7E5EcbUsgbgGMBhoSv59yAApL0UU17/view?usp=sharing*

1. After loading data I count the distribution of data into negative and positive value using value\_counts() function after that I print the missing values if they exist and then I analyze review lengths using histogram and plot a boxplot to analyze outliers.
2. Then I created a preprocess\_text(text) function to convert text to lowercase and tokenize it using word\_tokenize(text.lower()) function and remove non-alphabetic token using word.isalpha() function and then remove stop words and lemmatize words and then apply the preprocessing by using preprocess\_text method to review.
3. Then I i split data training and resting phase for full data and created a dictonary models\_full contains logistic regression and naïve bayes model method
4. Then I evaluate model for full data I calculated matrices like accuracy, F1socre,precision etc for each model in models\_full dictonary and calculate confusion matrix and display it in the form of heatmap.
5. Then for sample data I split data for sample data or we can say small part of data then I created models\_sample dictonary which contain SVM and Random Forest model methods.
6. After that I evaluate the SVM and Random Forest model for sample data as it require a lot of time to process with full data and calculate etrices for model like accuracy, F1socre,precision etc after that calculate confusion matrix and display it in the form of heatmap.
7. Ater that test each model with new and unseen data as the result except Naïve Bayes Model all three model making proper prediction for new and unseen data

*Explanation video link- https://drive.google.com/file/d/1Xc7E5EcbUsgbgGMBhoSv59yAApL0UU17/view?usp=sharing*