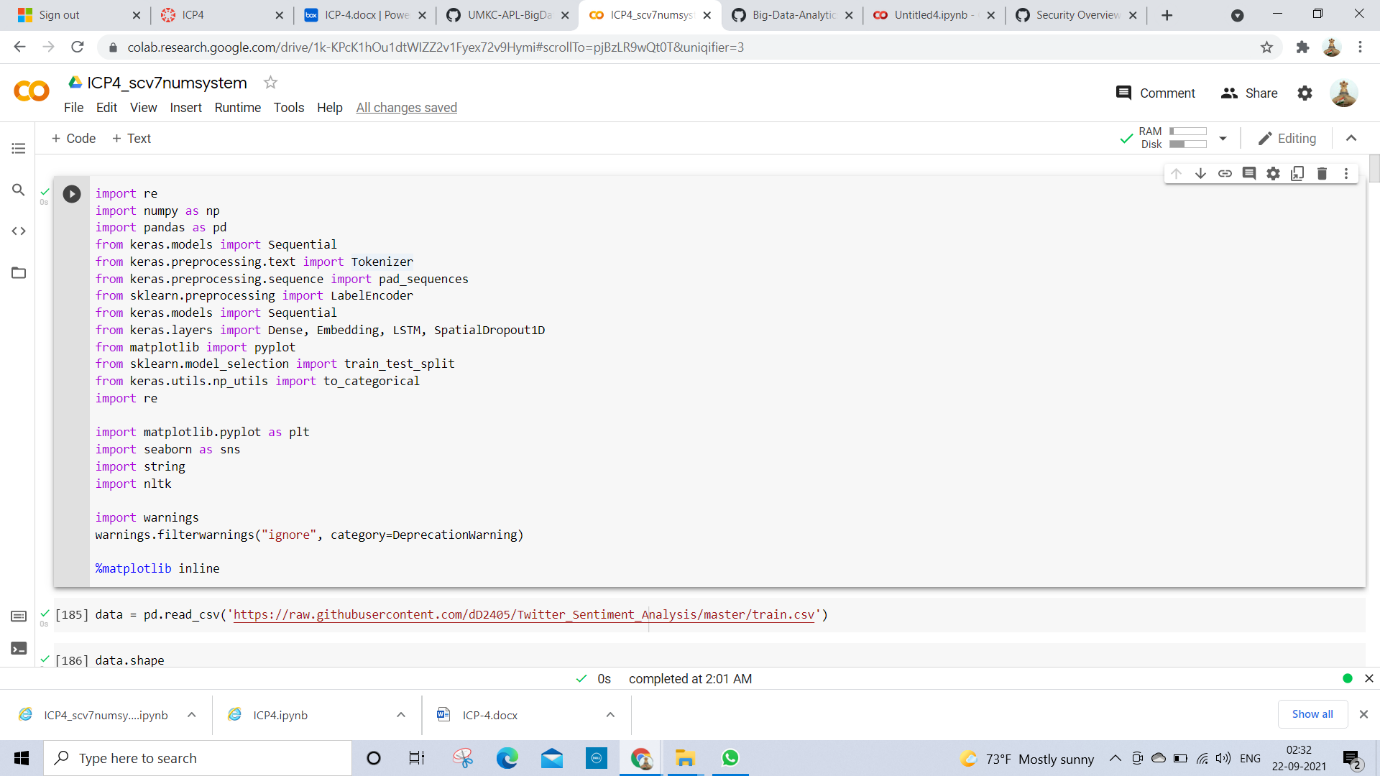
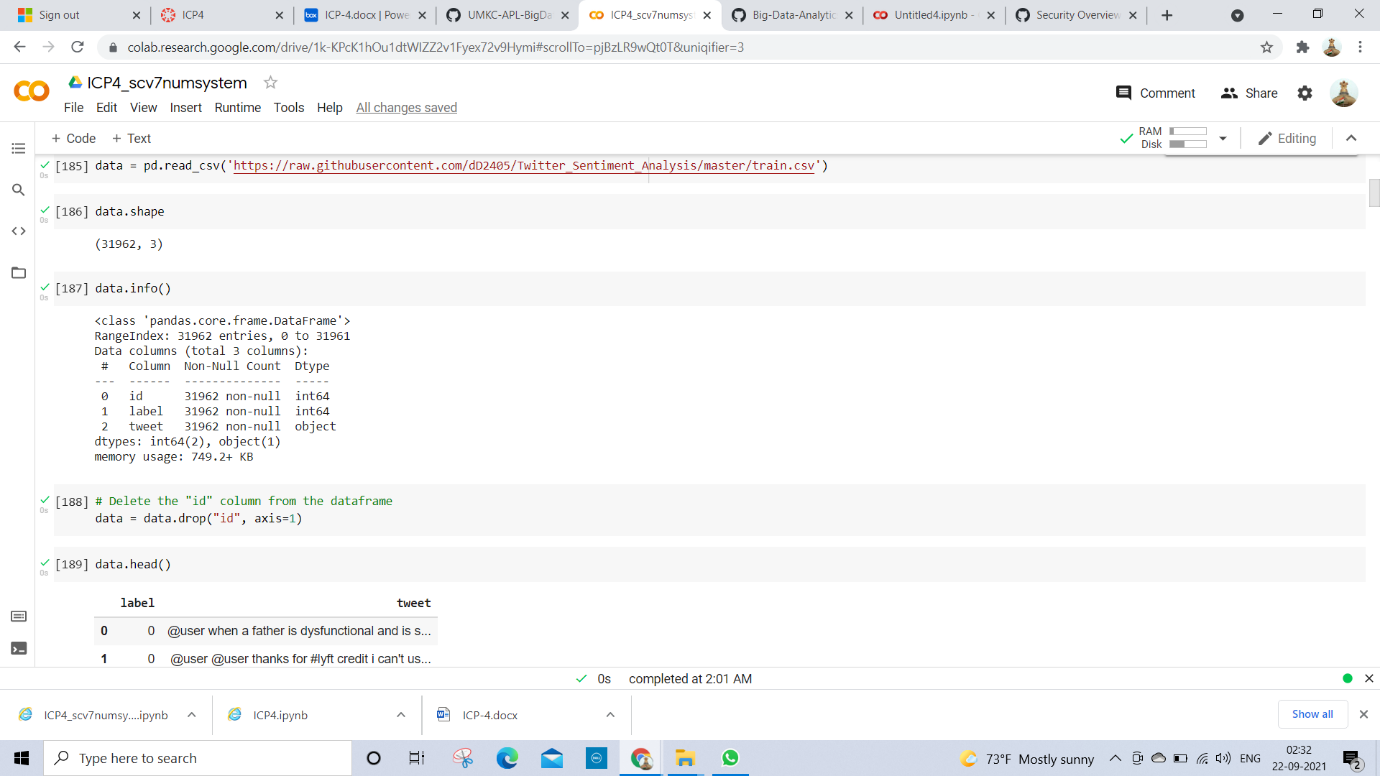
**WIKI Report – ICP4**

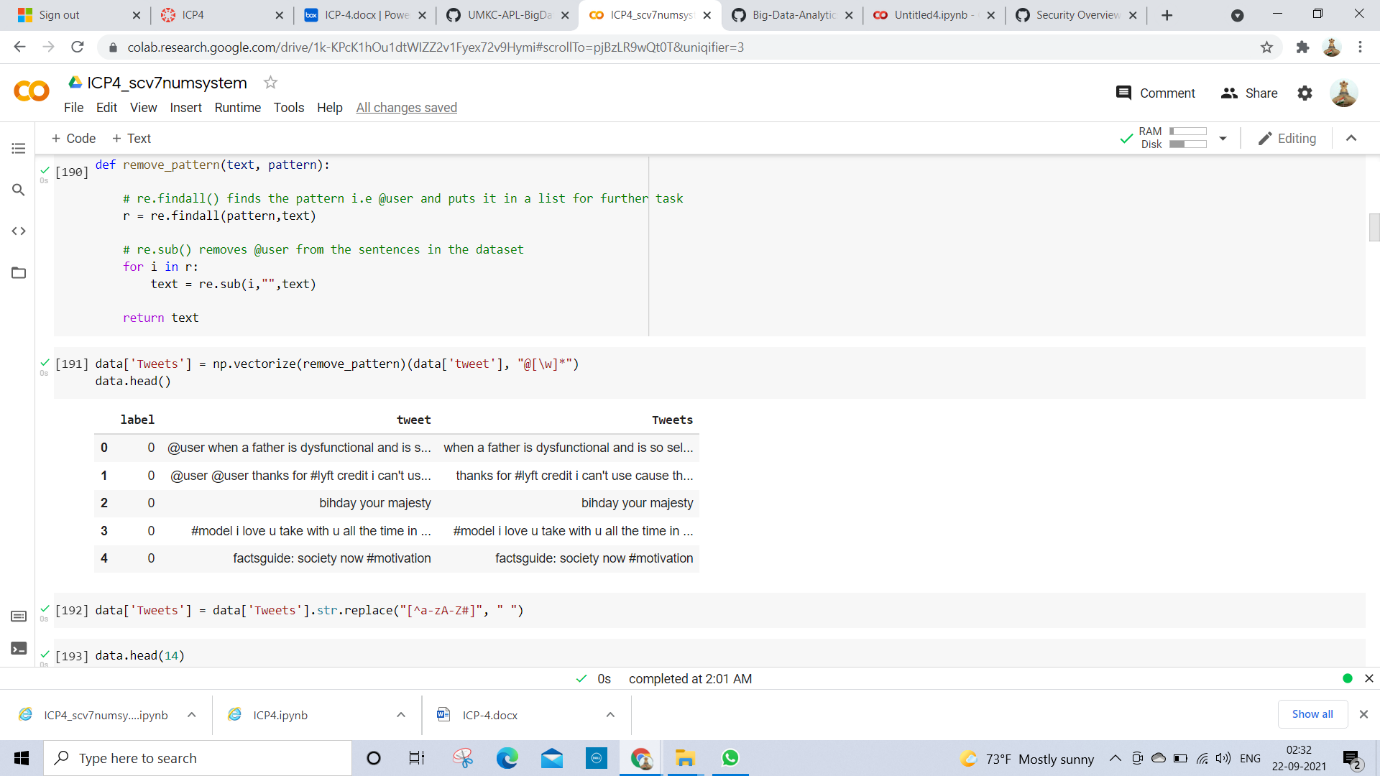
* Firstly, imported the required libraries from the given dataset.



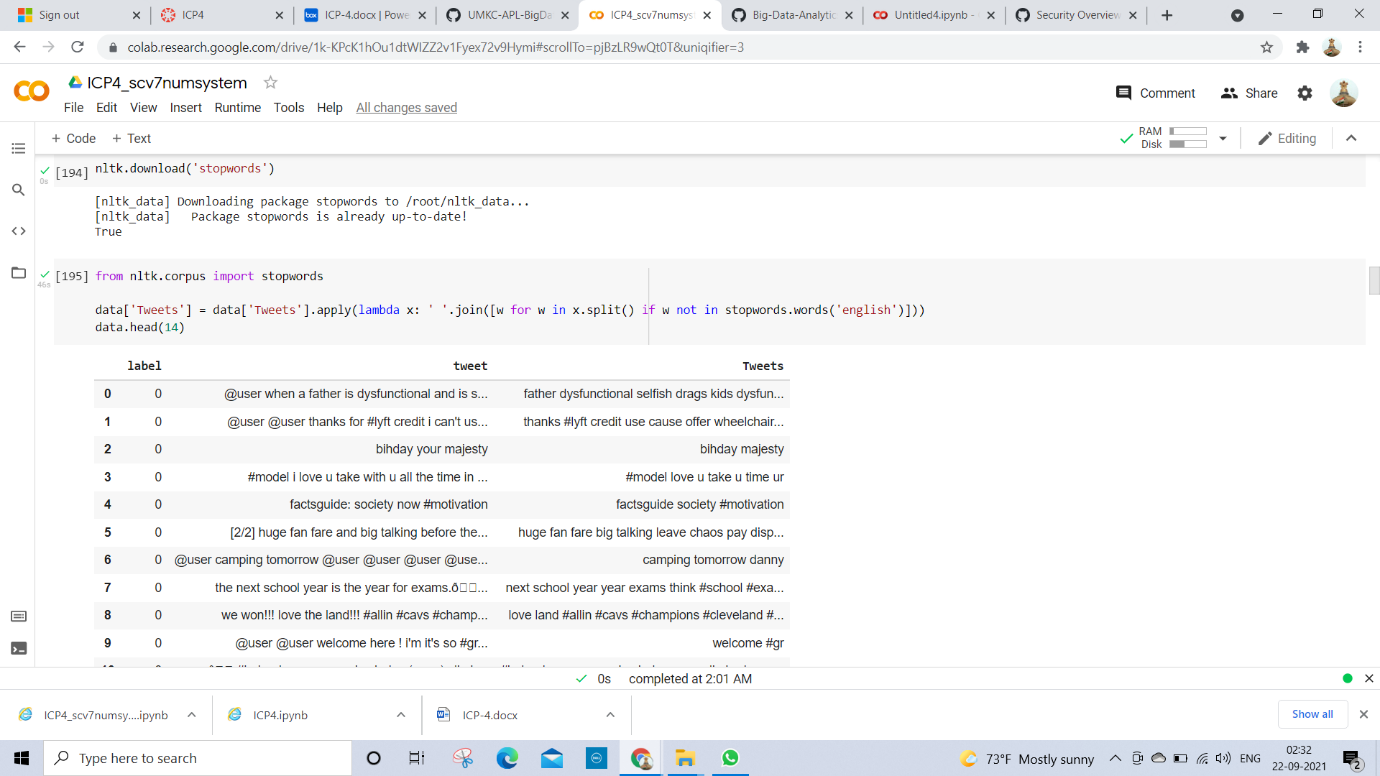
* From the given set of data deleted the columns data.



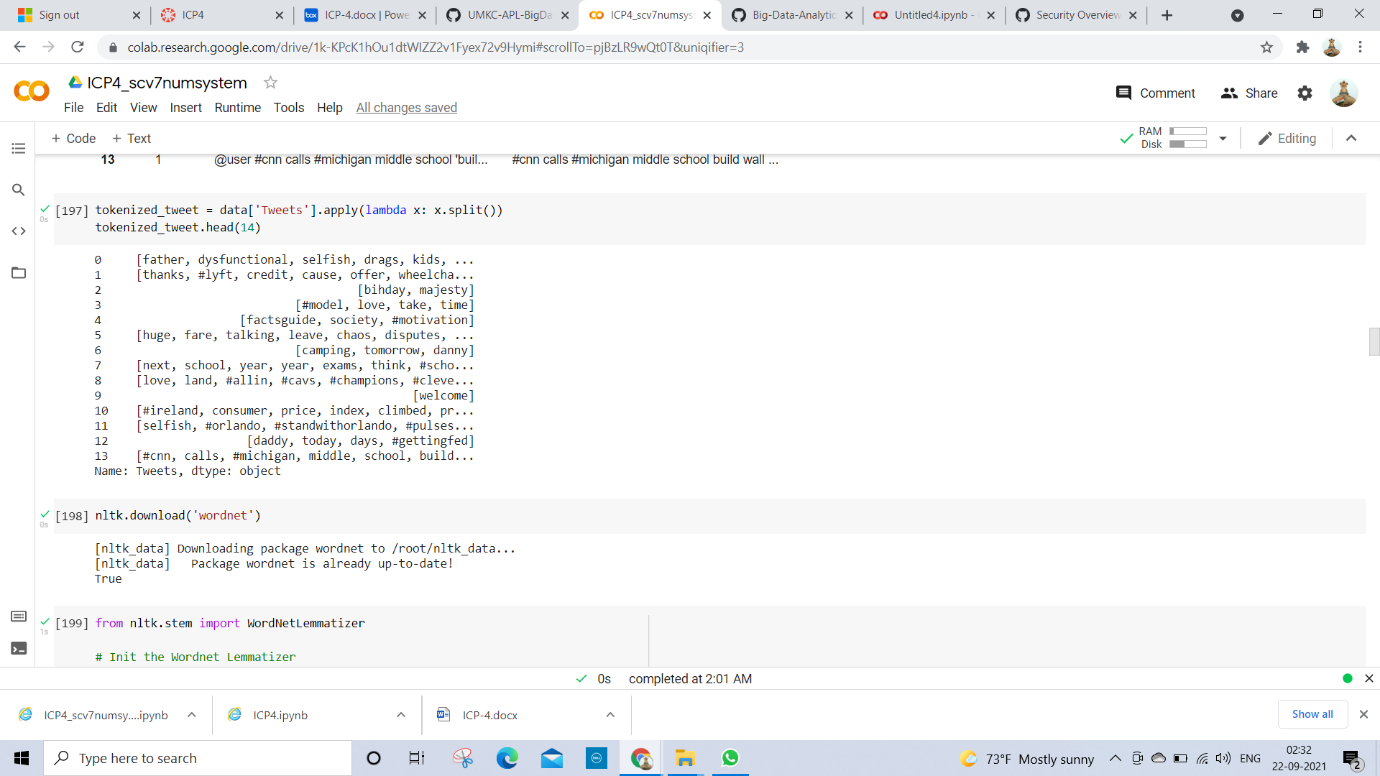
* All the data has been removed and created the new data and saved, also removed the numbers, punctuations and special characters by using str. replace function.



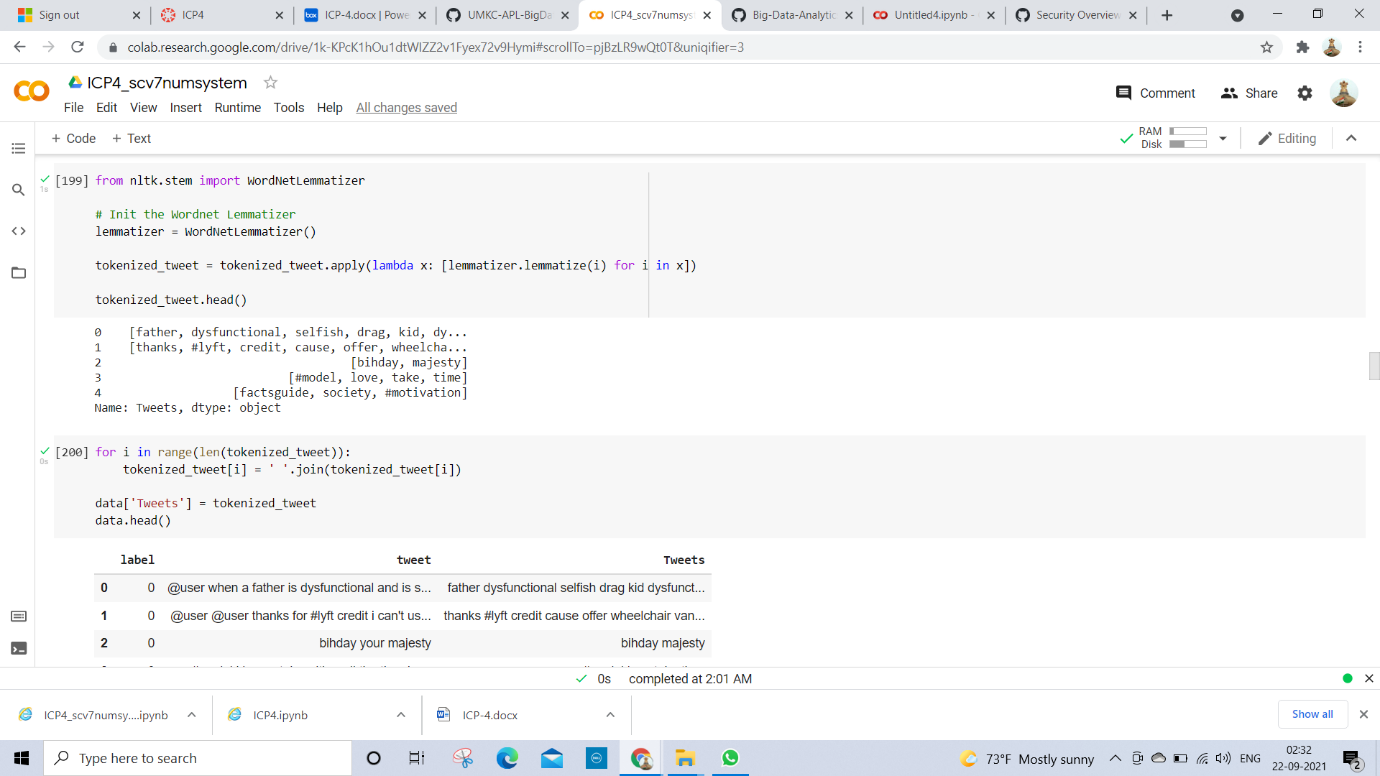
* I removed all the short English words which are not important.



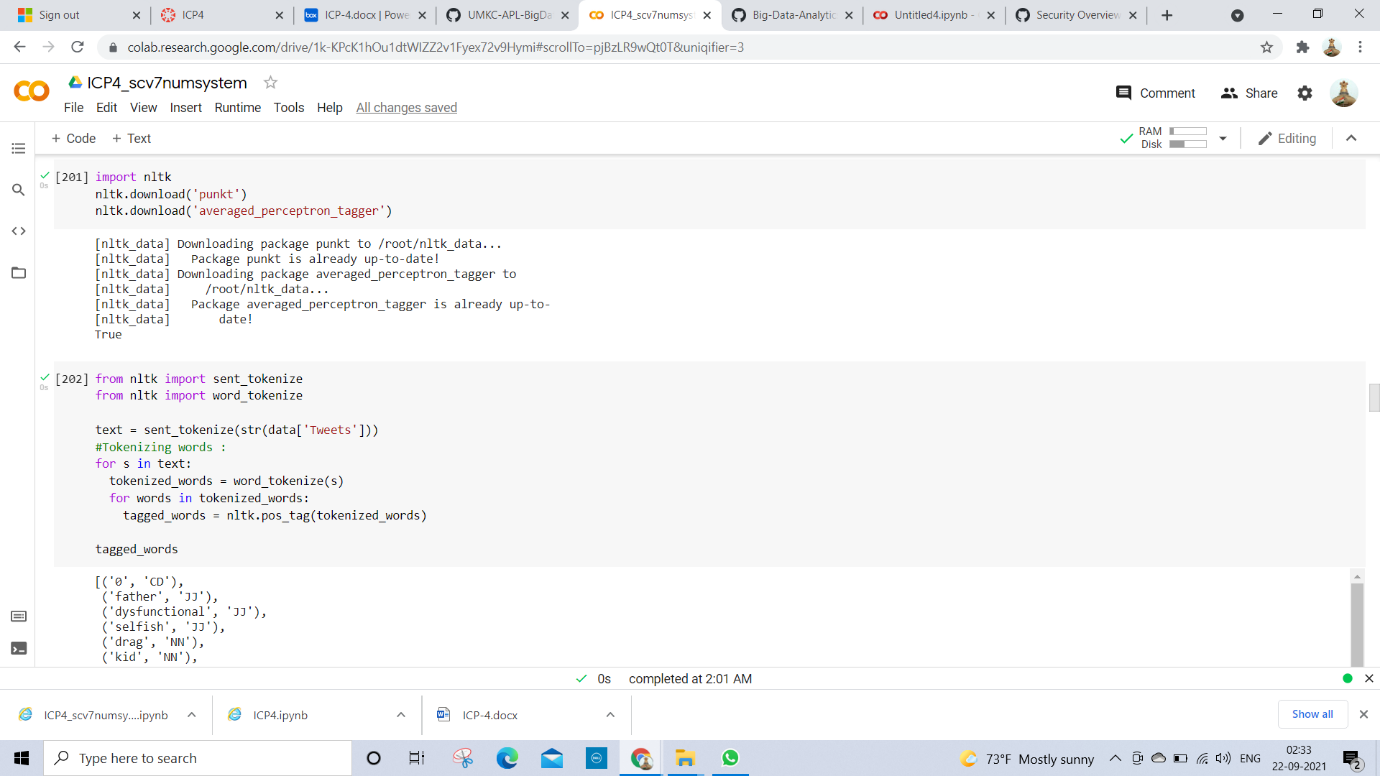
* The tokenization input has been given to split the words into different words.



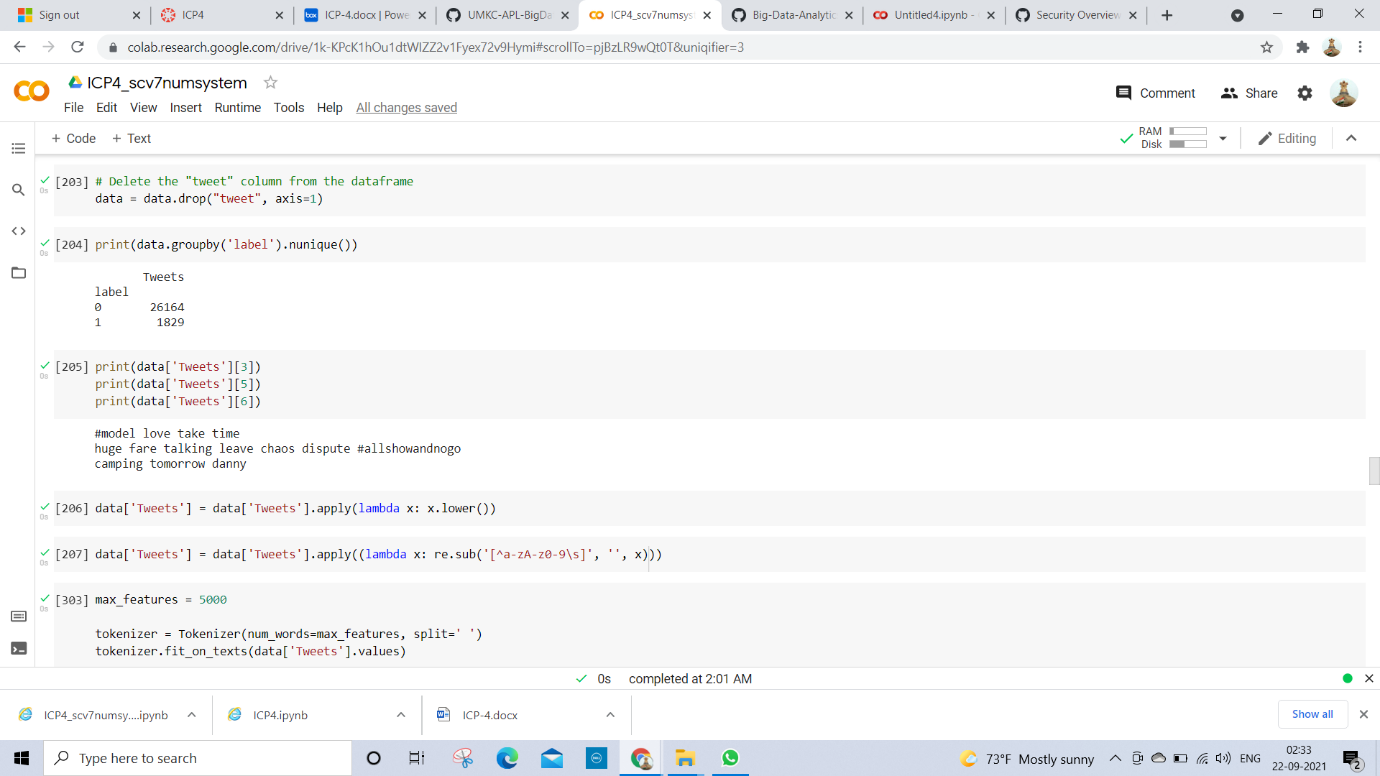
* Imported WordNetLemmatizer from the libraries called nltk.stem i.e stemming.



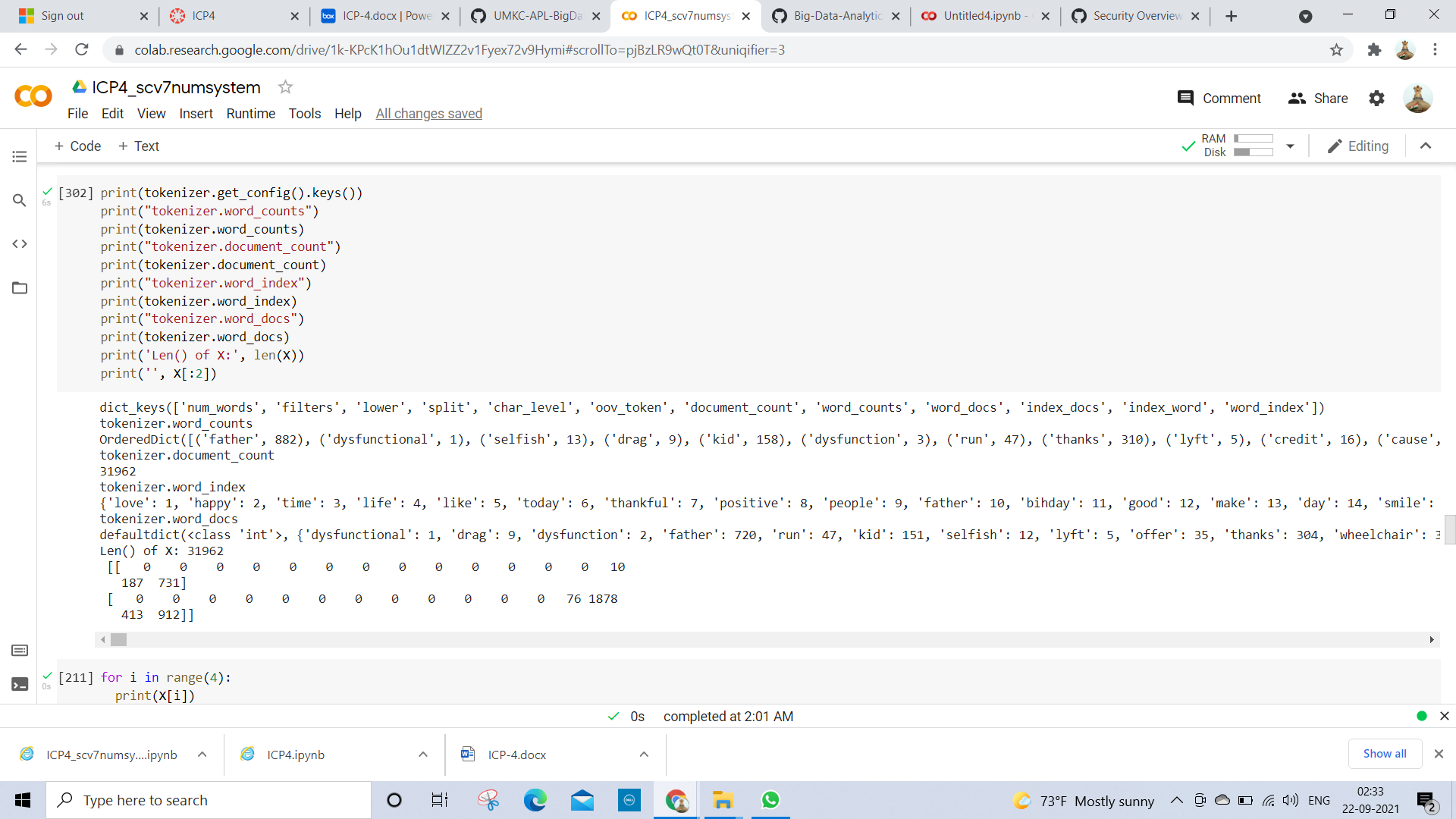
* From the nltk library, imported the sentence tokenization and word tokenization and given the for condition to tag POS for all words.



* Now, deleting the all unnecessary data and count all labels and tweets.

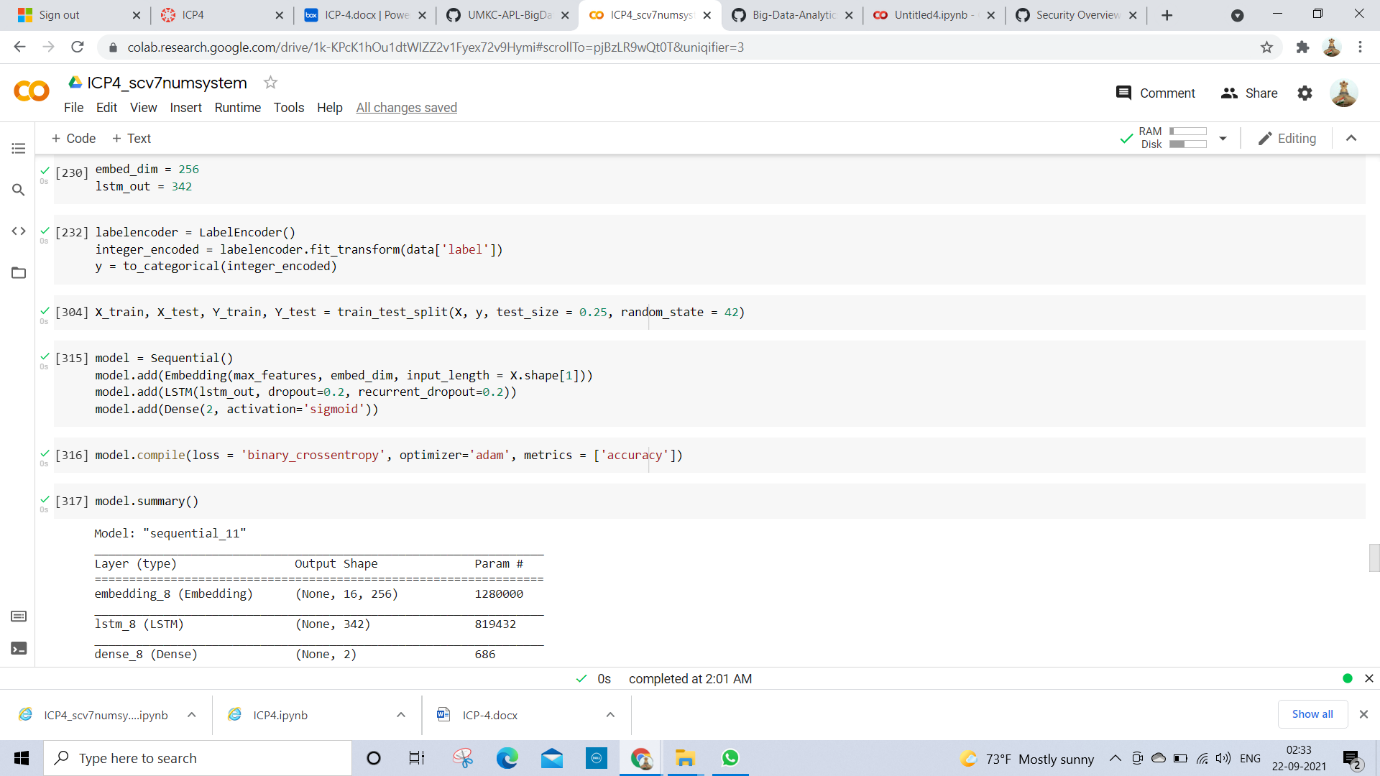


* Here with the help of tokenization, I performed the data as in order.

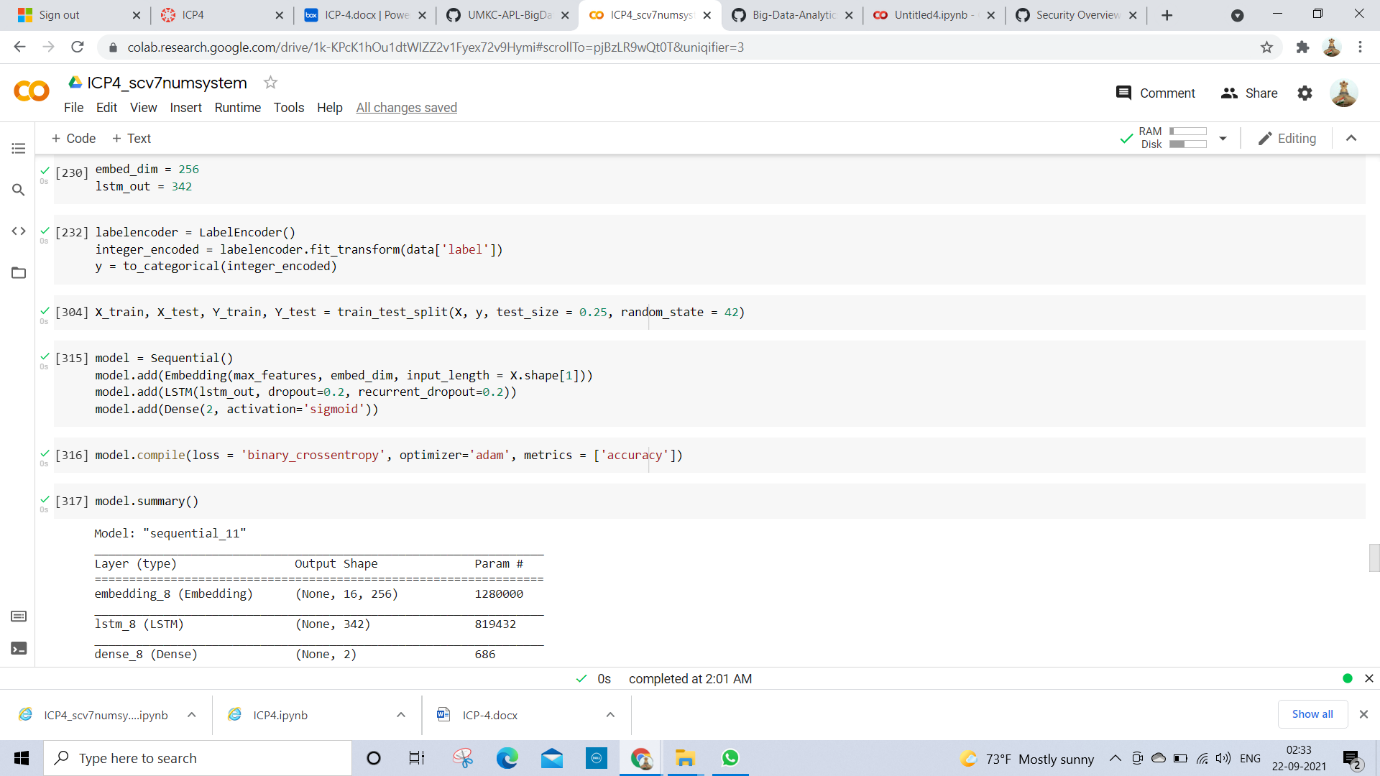


**Deep Learning Model Building:**

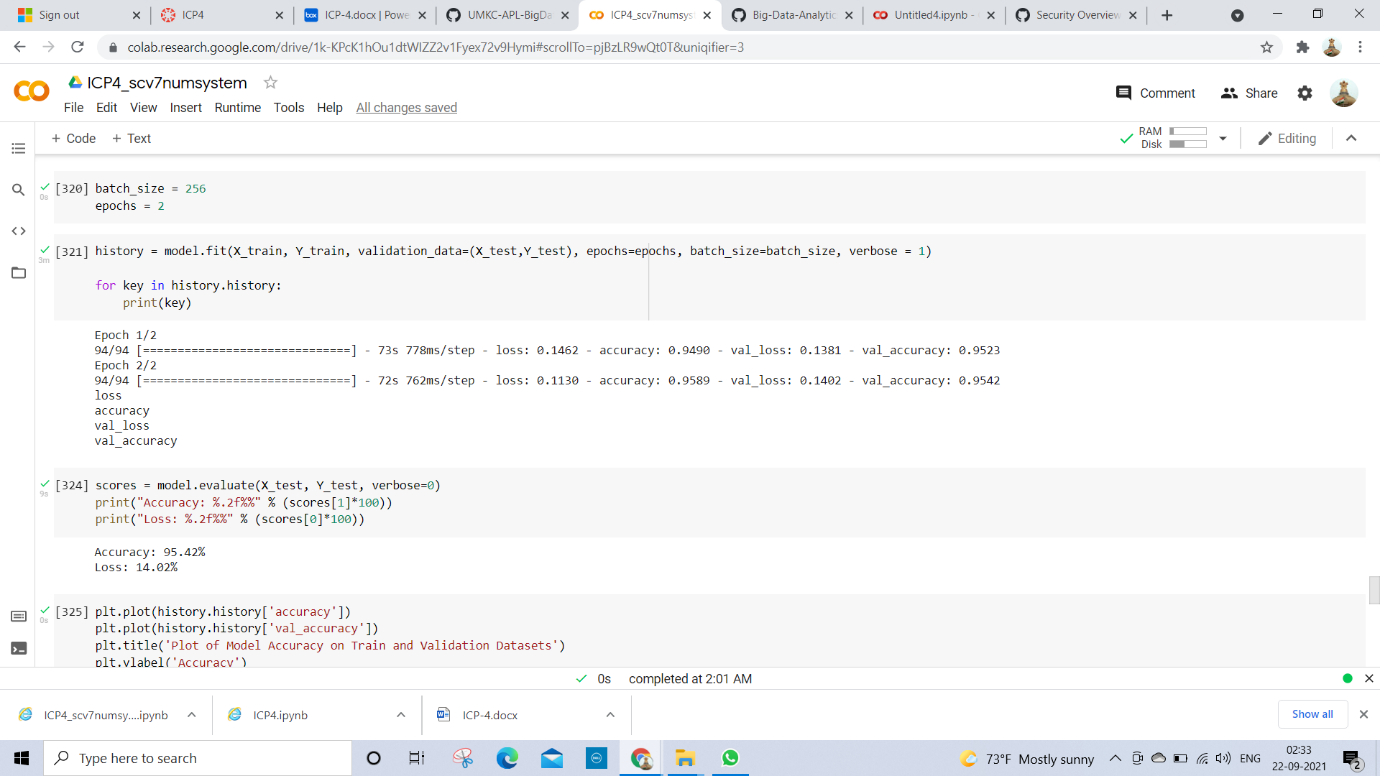
* Here I split the data by train test and split functions.



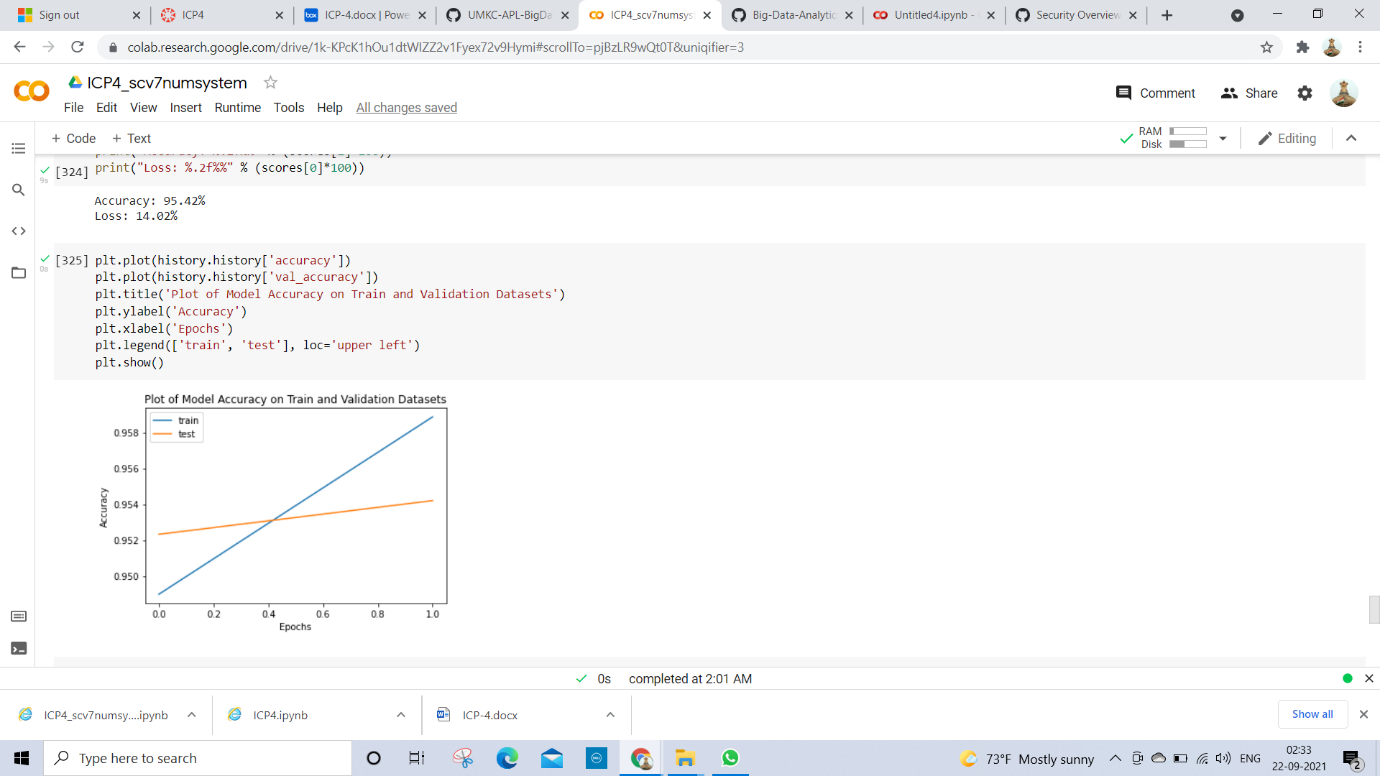
* Now I created a model by embedding ,LSTM and output layer as activation given by 2 dense size and Final summary model of the total params count.



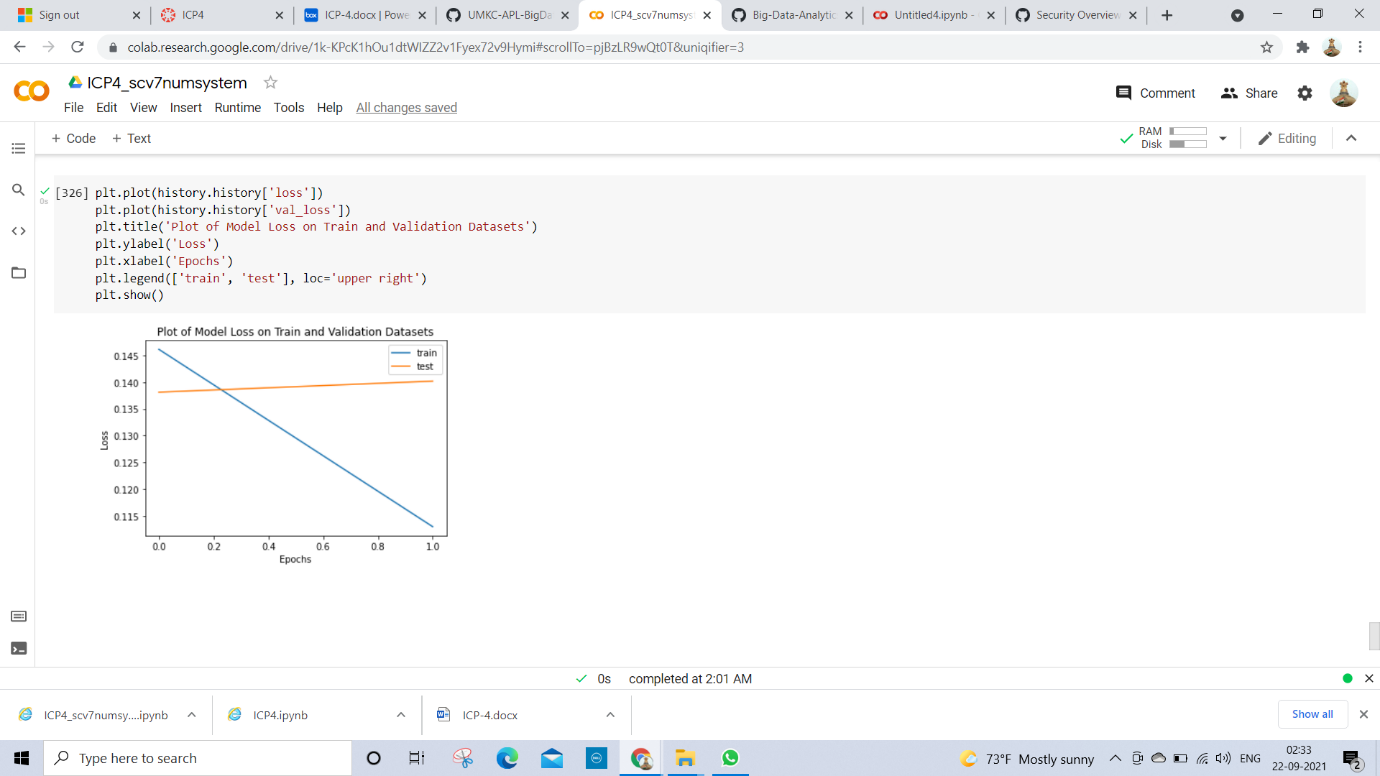
* The model is finalized and gathered all the history and performed about 95.42% accuracy.



* Plotted a graph for accuracy model in train and test from the given data.



* Plotted a graph for loss in train and test from the given data.



* I learnt new libraries like tokenization , stemming, WordNetLemmatizer, Pos tagging , some new functions, deep learning model that includes embedding layer, LSTM etc.
* I faced many challenges while running the code like removing punctuations, stop words and also while encoding the data, text sequence functions.