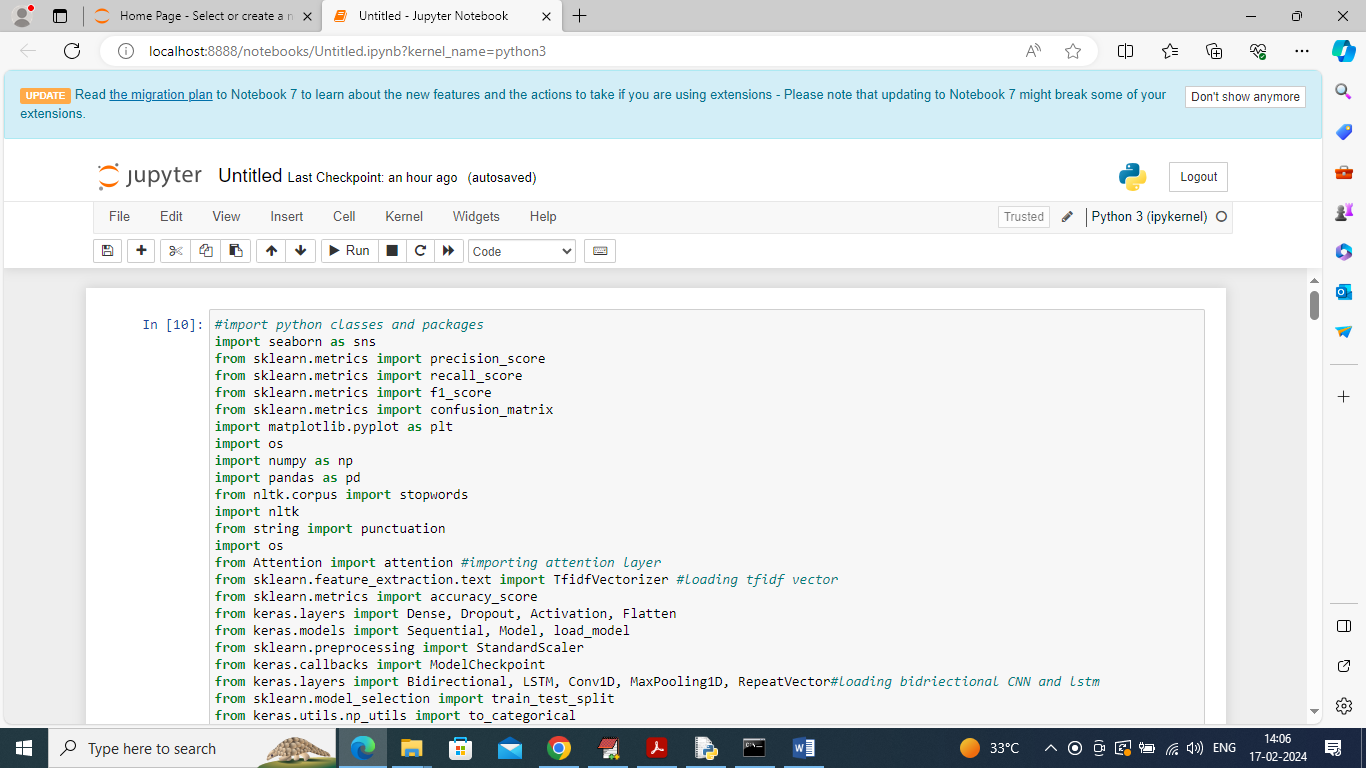
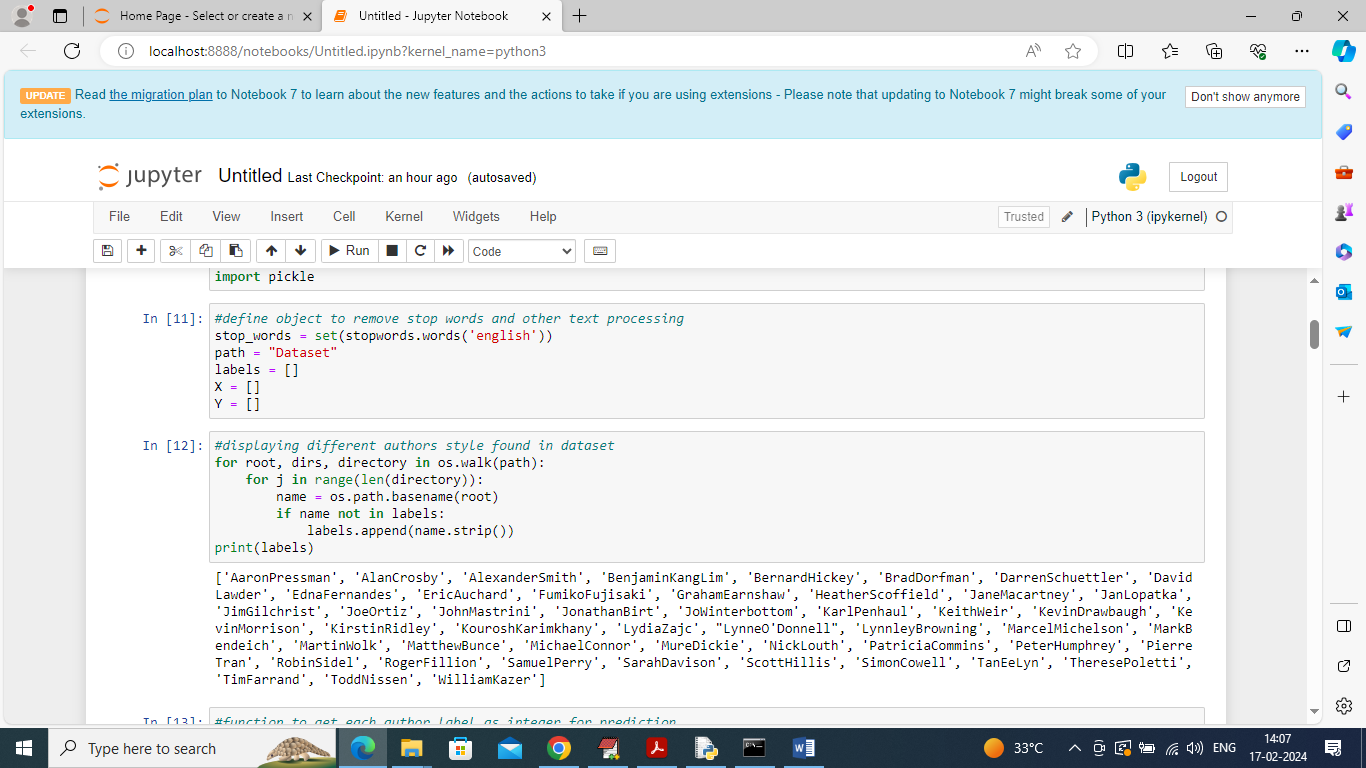
Authorship Attribution using Deep Learning

In this project as per your request we have develop LSTM model to predict author from his text style. We have utilize given based paper to extract Syntactic and structural information from given Reuters\_50\_50 dataset and then text data will be converted to numeric vector and then train with combination of CNN bidirectional LSTM to predict author from given style. Structural and syntactic information contains all hidden and minute features from each author which can help in easily distinguish between different author styles.

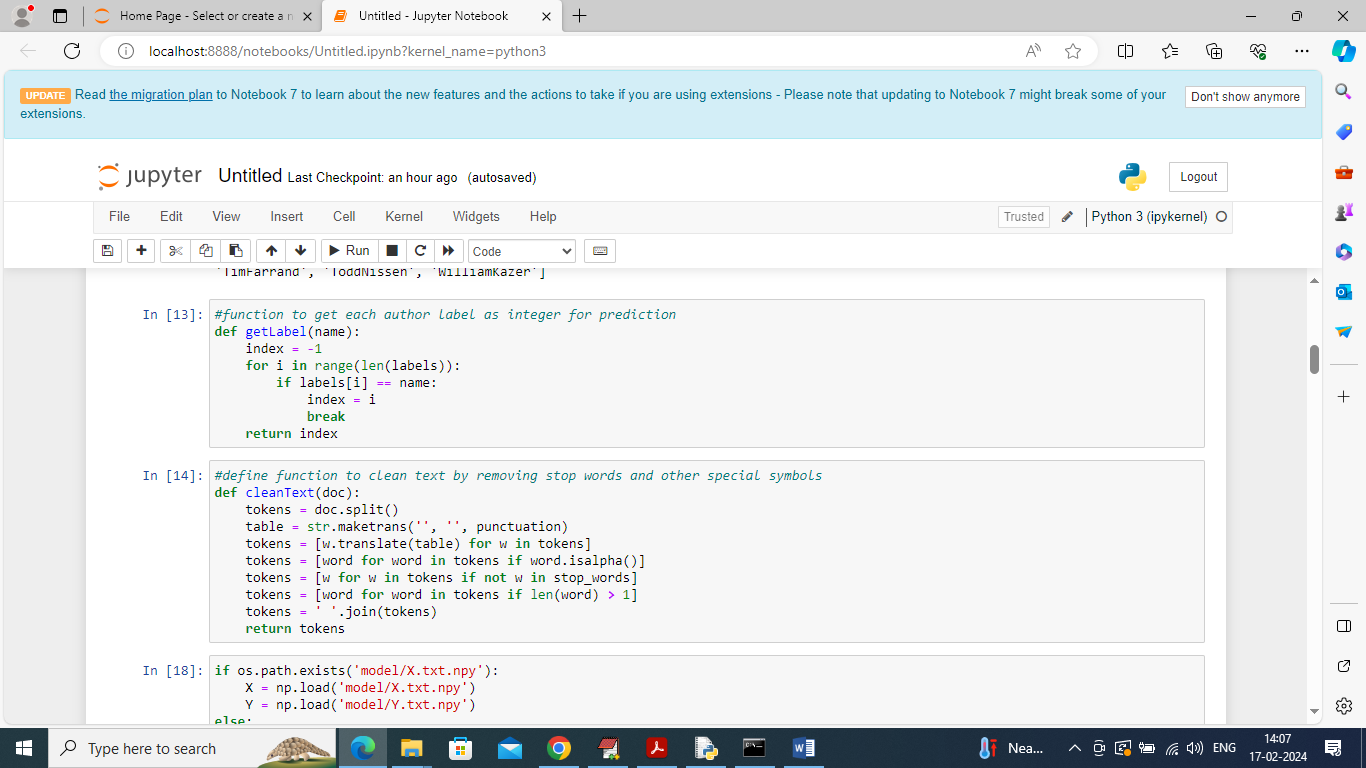
We have coded this project using JUPYTER notebook and below are the code and output screens with blue colour comments



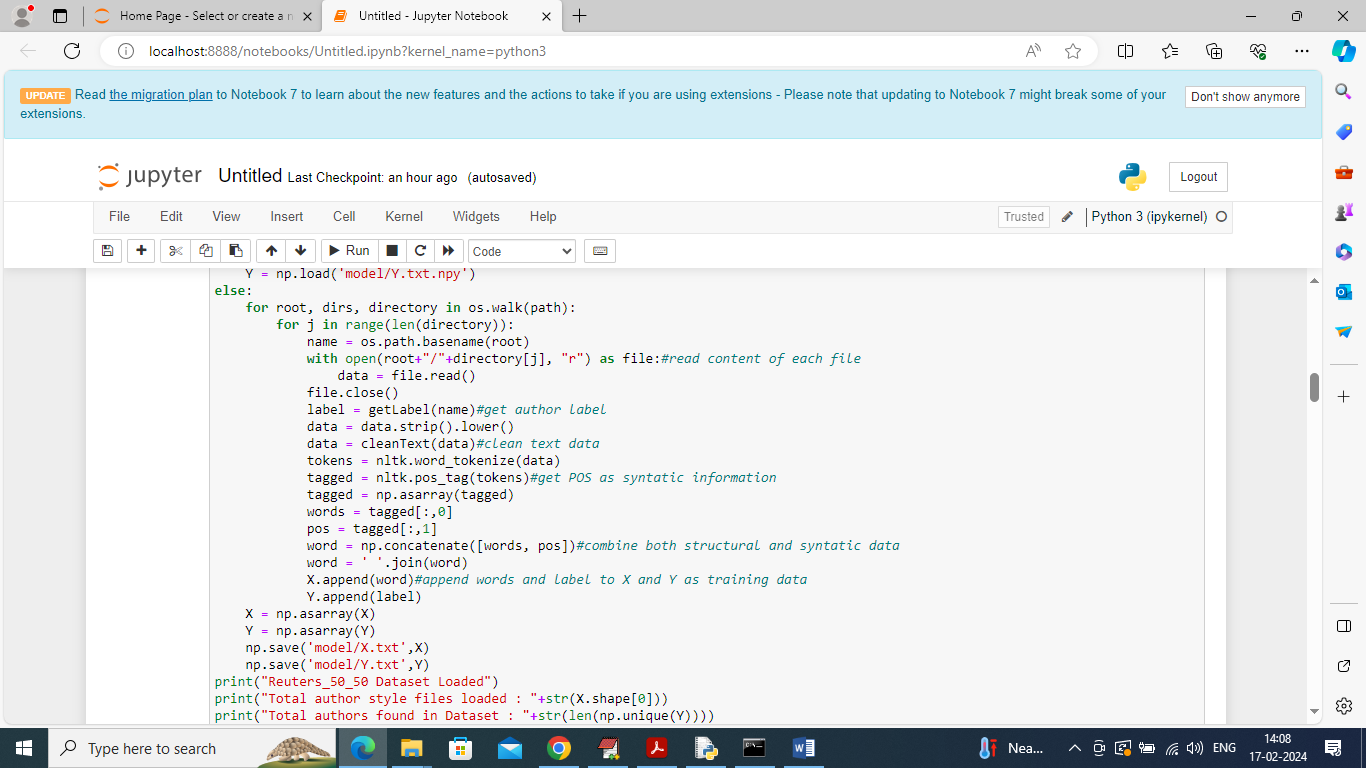
In above screen importing required python classes and packages



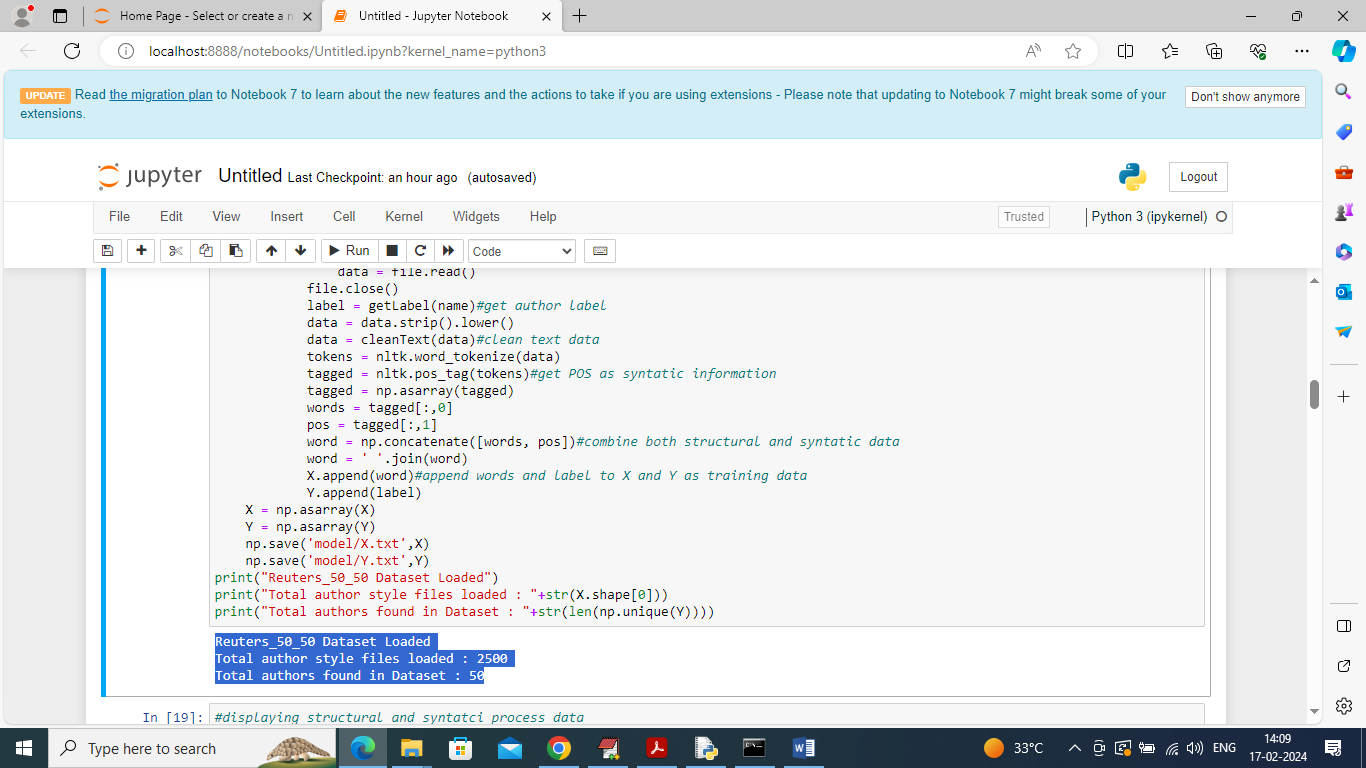
In above screen defining function to remove stop words and to clean text and then displaying all author names available in dataset



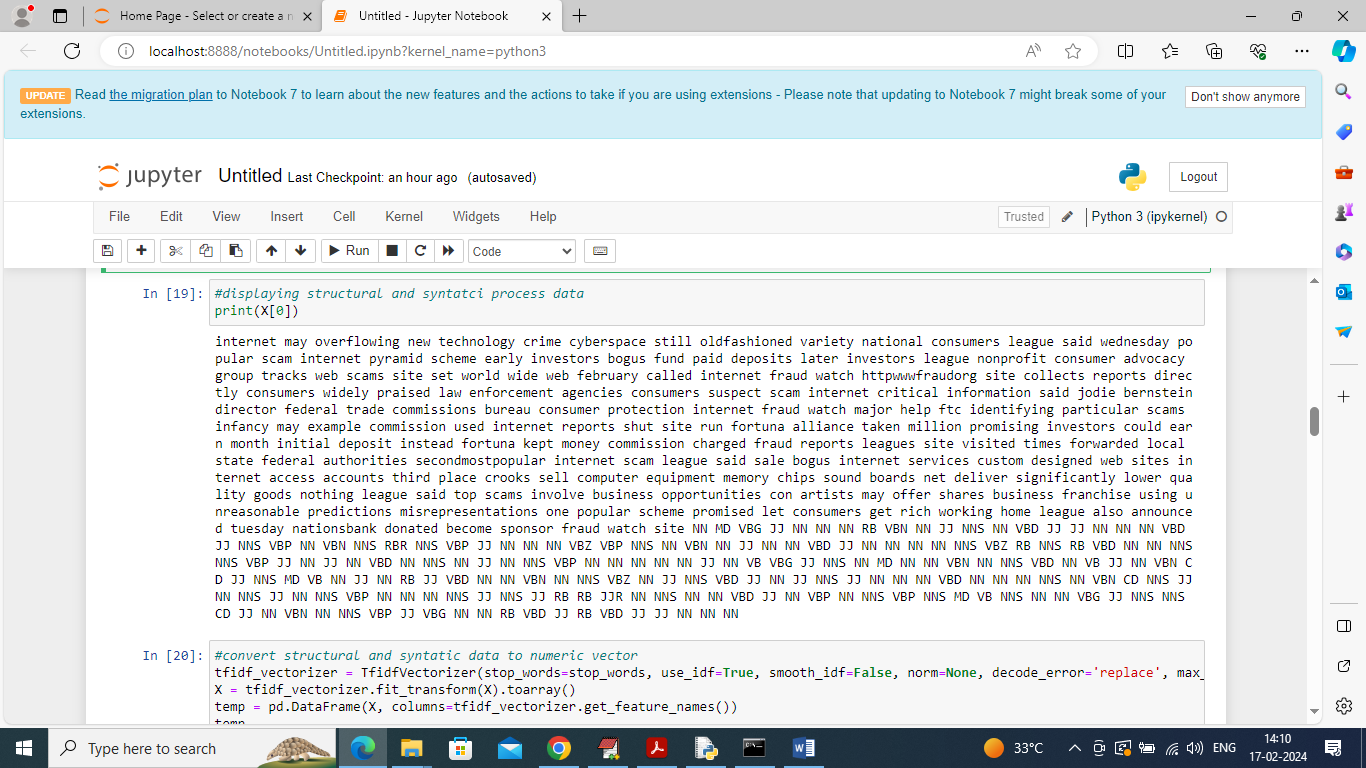
In above screen defined function to clean text like stop words removal, special symbols and numbers etc.



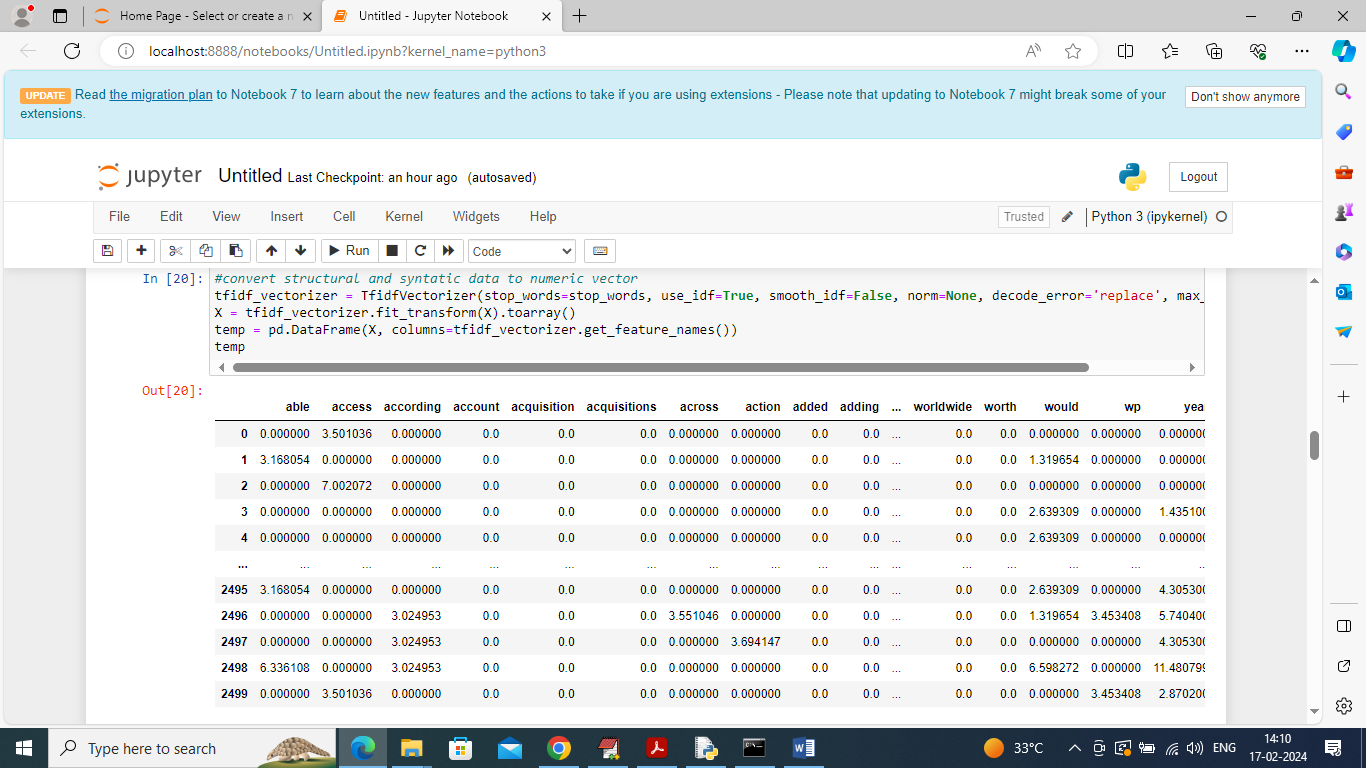
In above screen looping and reading all files from dataset and then cleaning and converting all text files into structure and syntactic information and then adding all process data into X and Y array



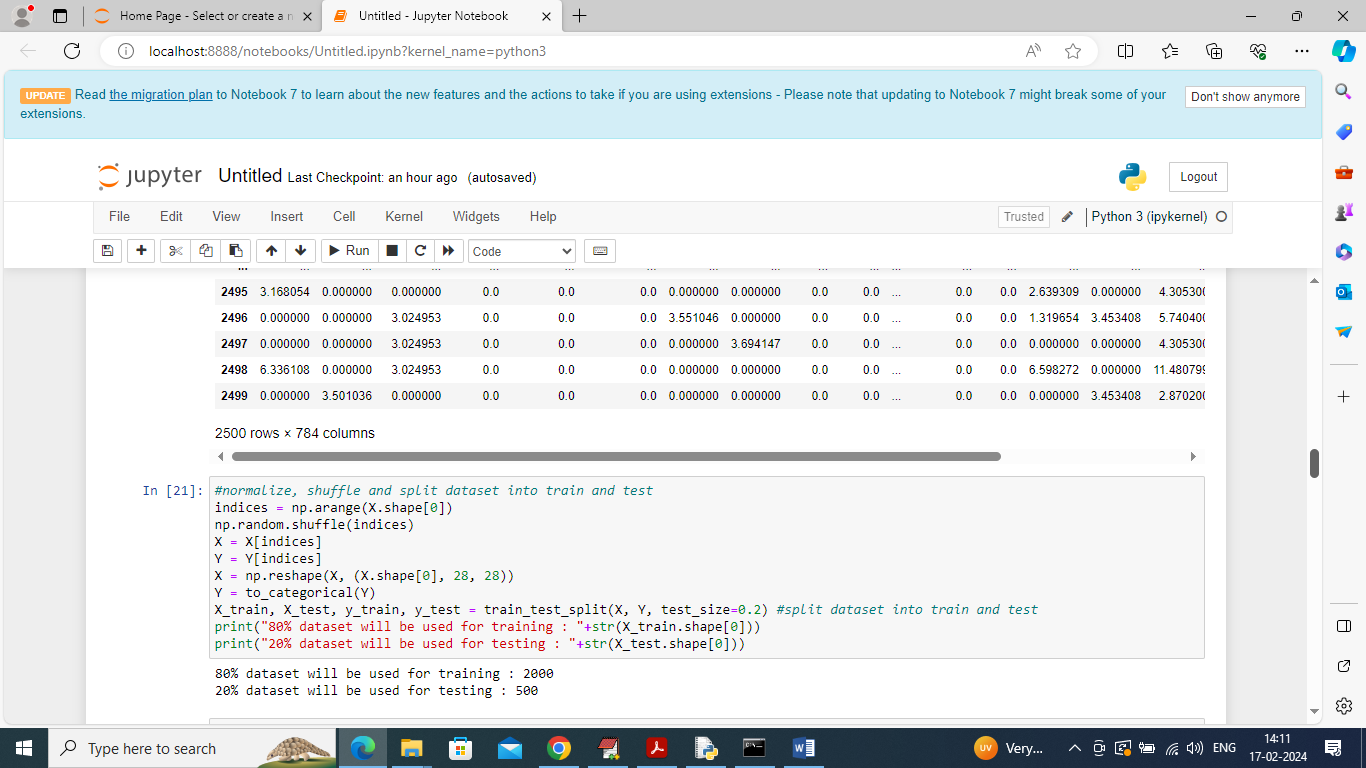
In above screen displaying total files and author available in dataset



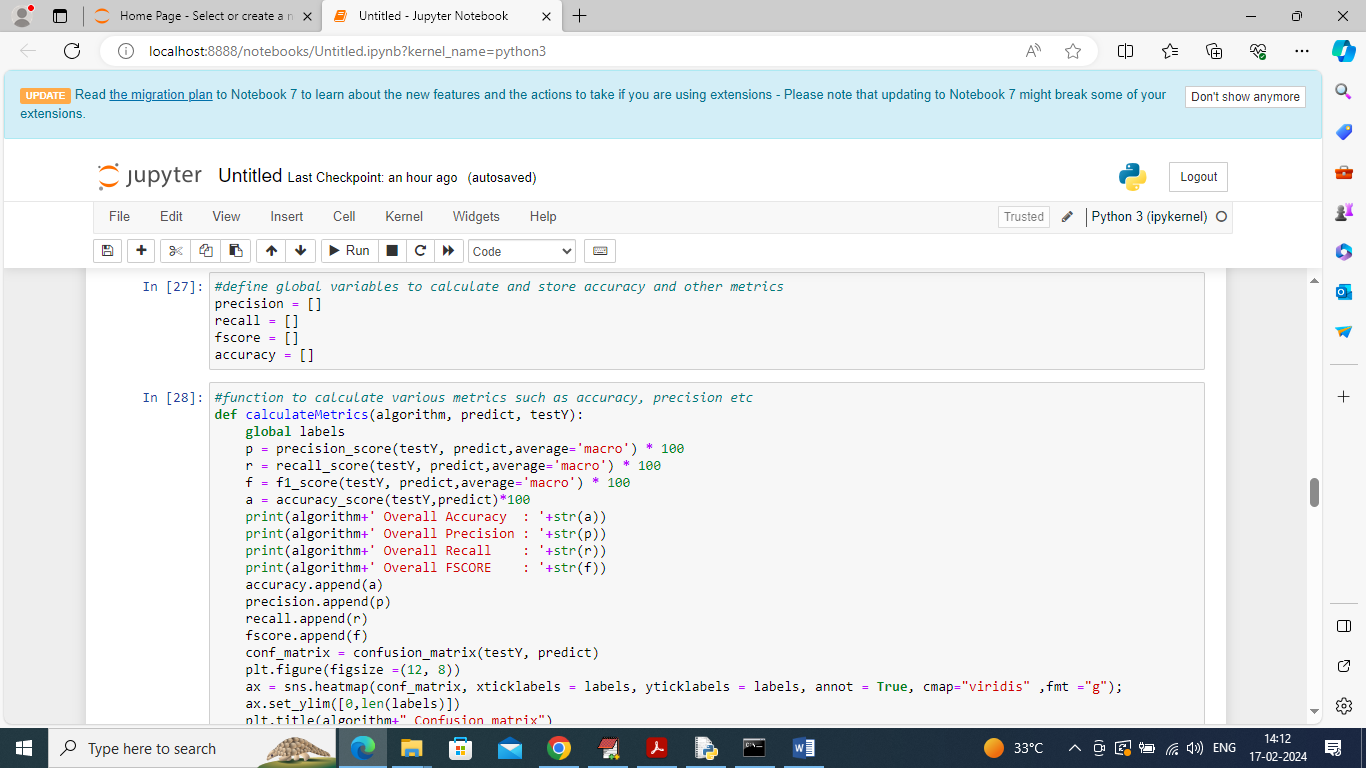
In above screen displaying processed text from one author where we can see structure and syntactic POS values



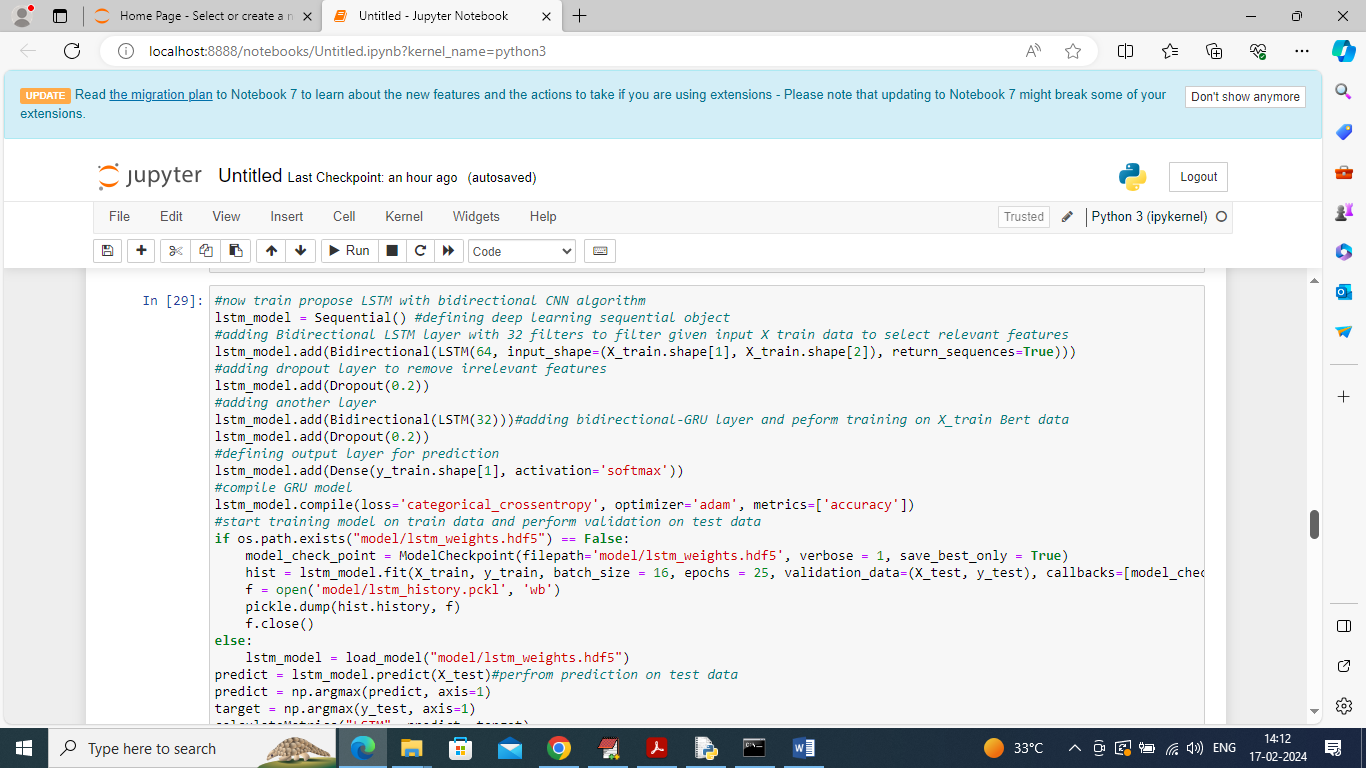
In above screen converting all text into numeric vector where table column contains word names and table rows contains average frequencies of those words



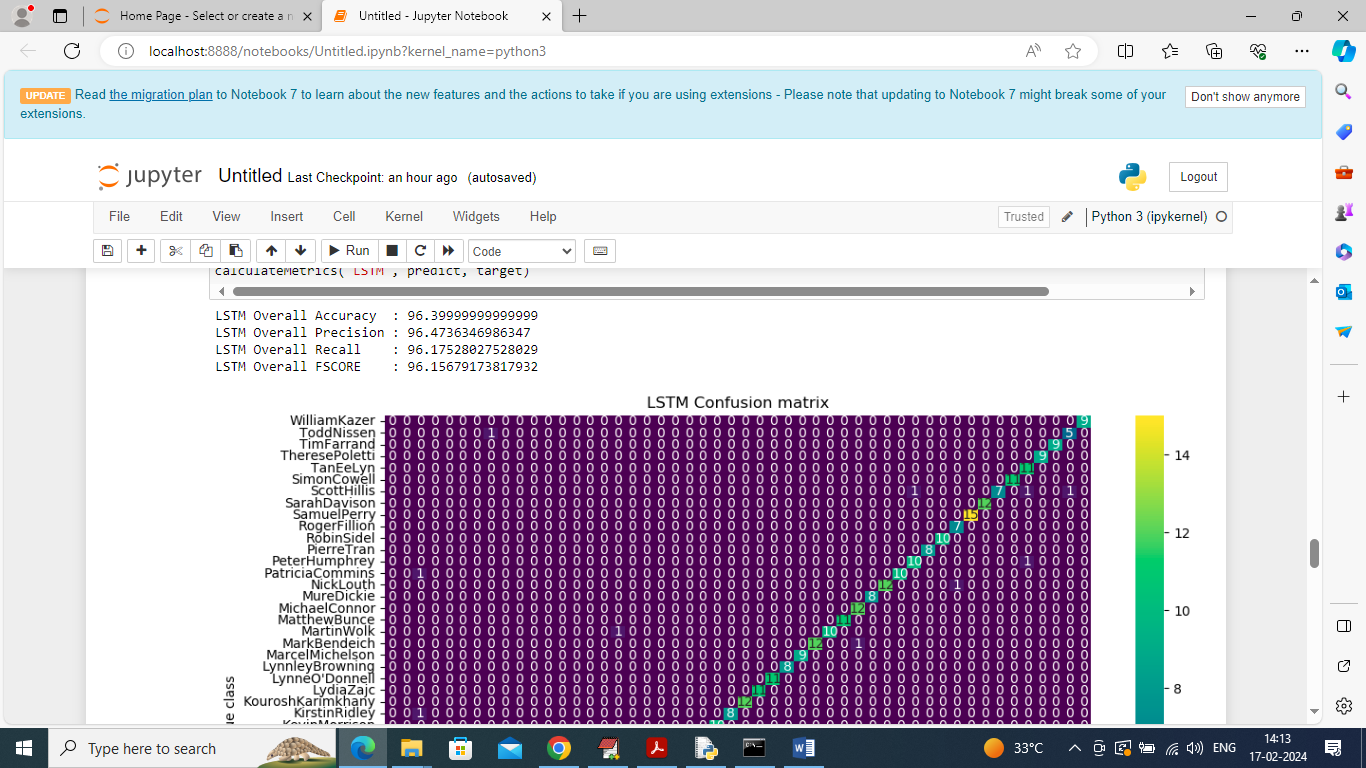
In above screen entire numeric vector is normalizing, shuffling and splitting into train and test where application using 80% for training and 20% for testing



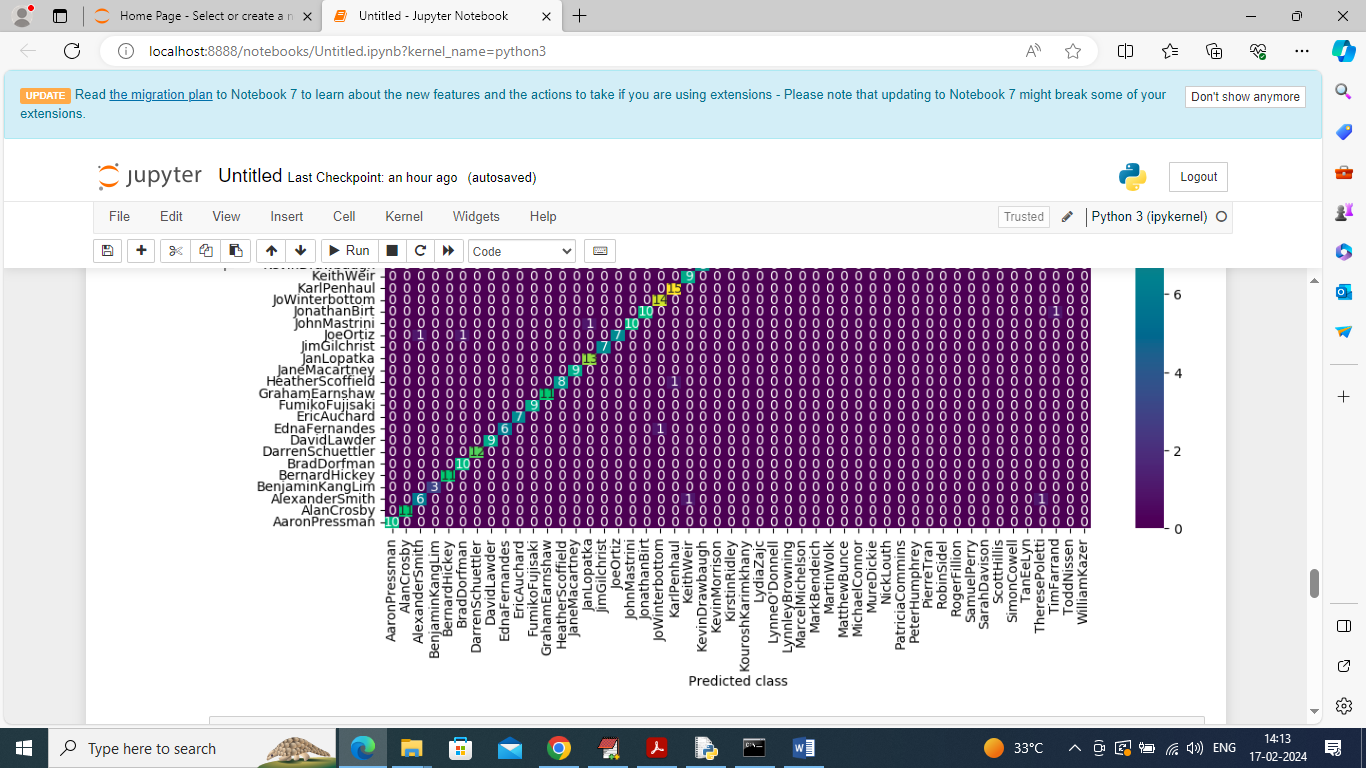
In above screen defining function to calculate accuracy and other metrics



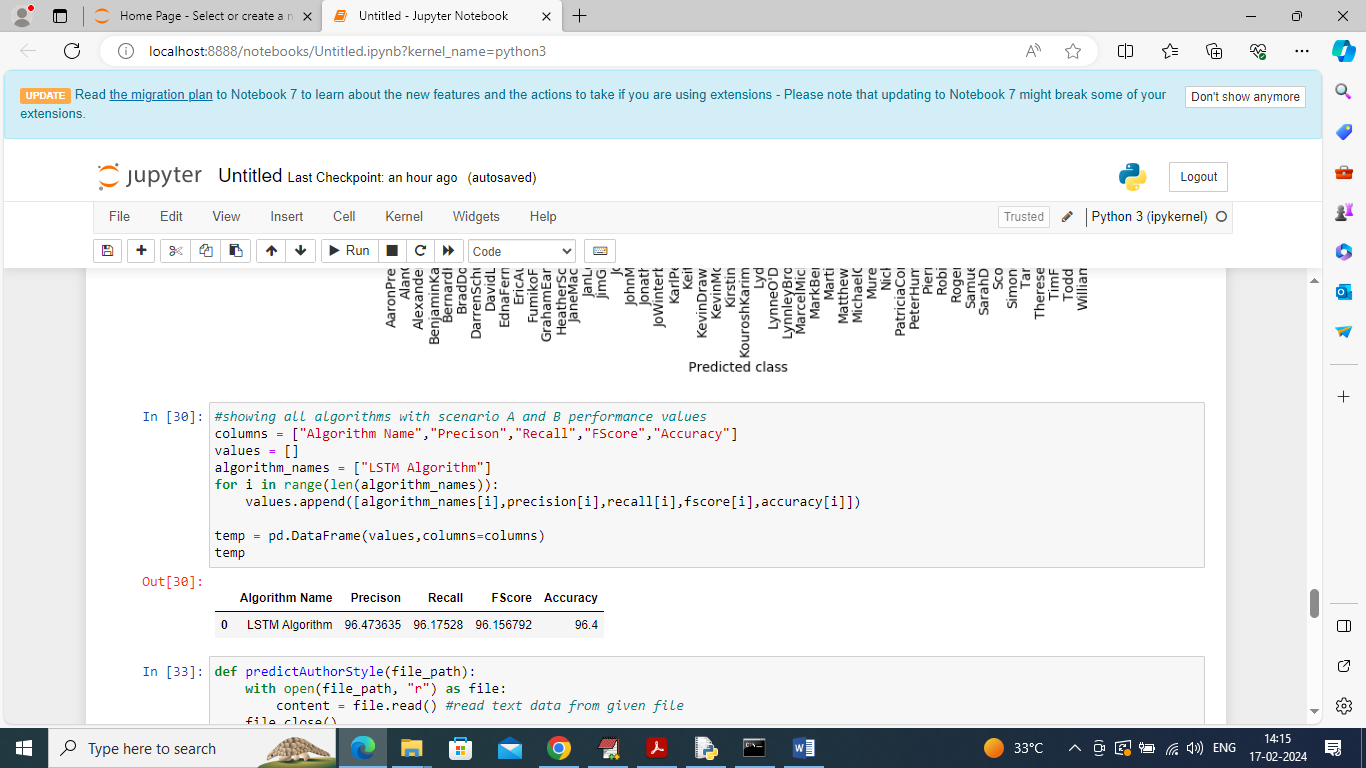
In above screen training LSTM on 80% training data and then performing prediction on 20% test data to calculate prediction accuracy



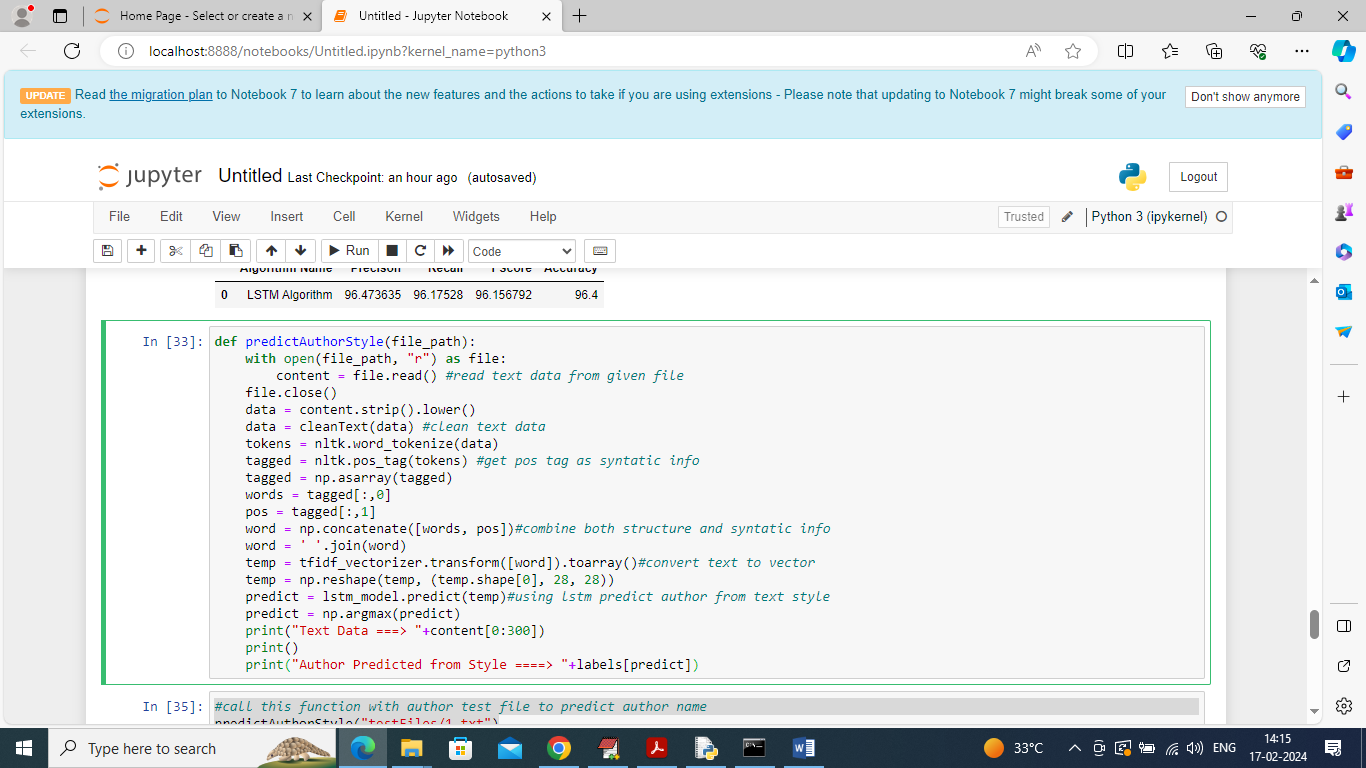
In above screen LSTM got 96% accuracy on 20% test data and can see other metrics like precision, recall etc.



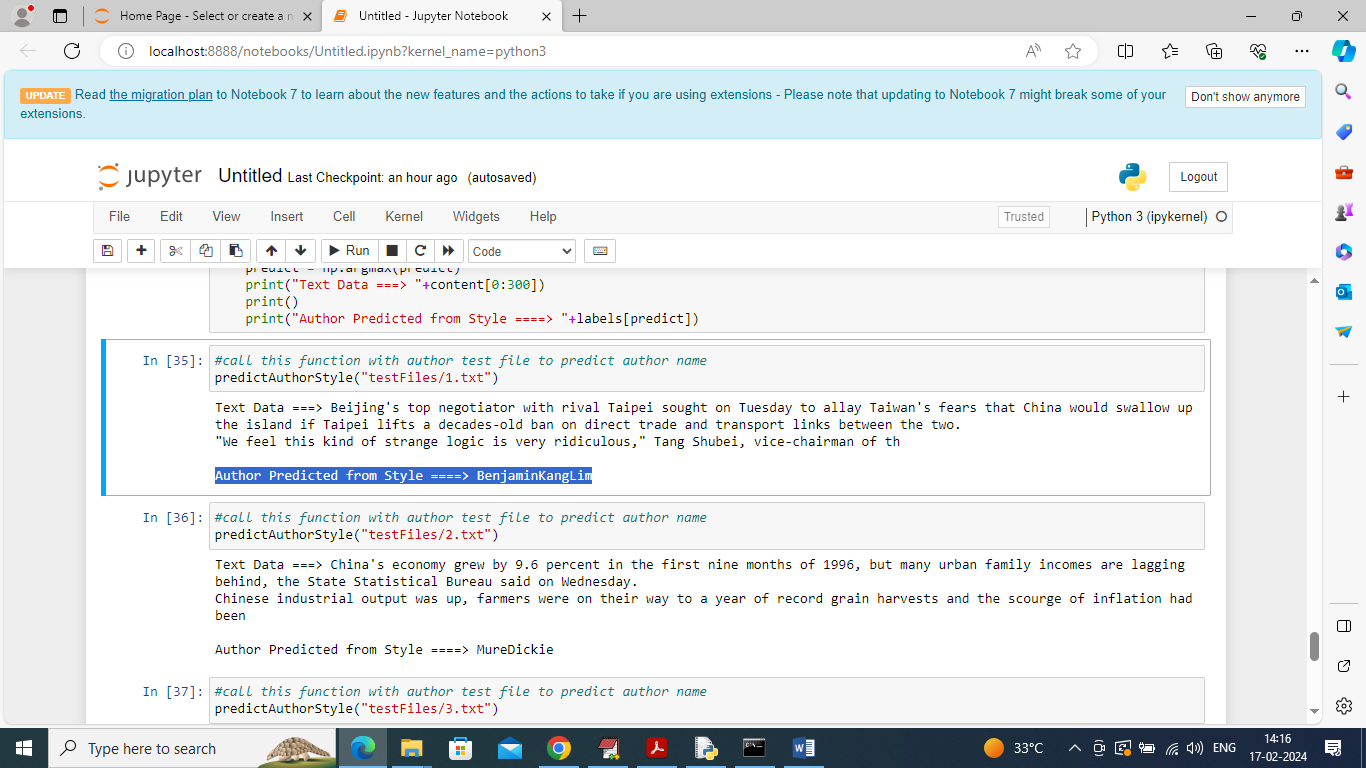
In above confusion matrix graph x-axis represents Predicted Author names and y-axis represents True author names and then all different colour boxes in diagnol represents correct prediction count and remaining blue boxes represents incorrect prediction count which is 0 or very few.



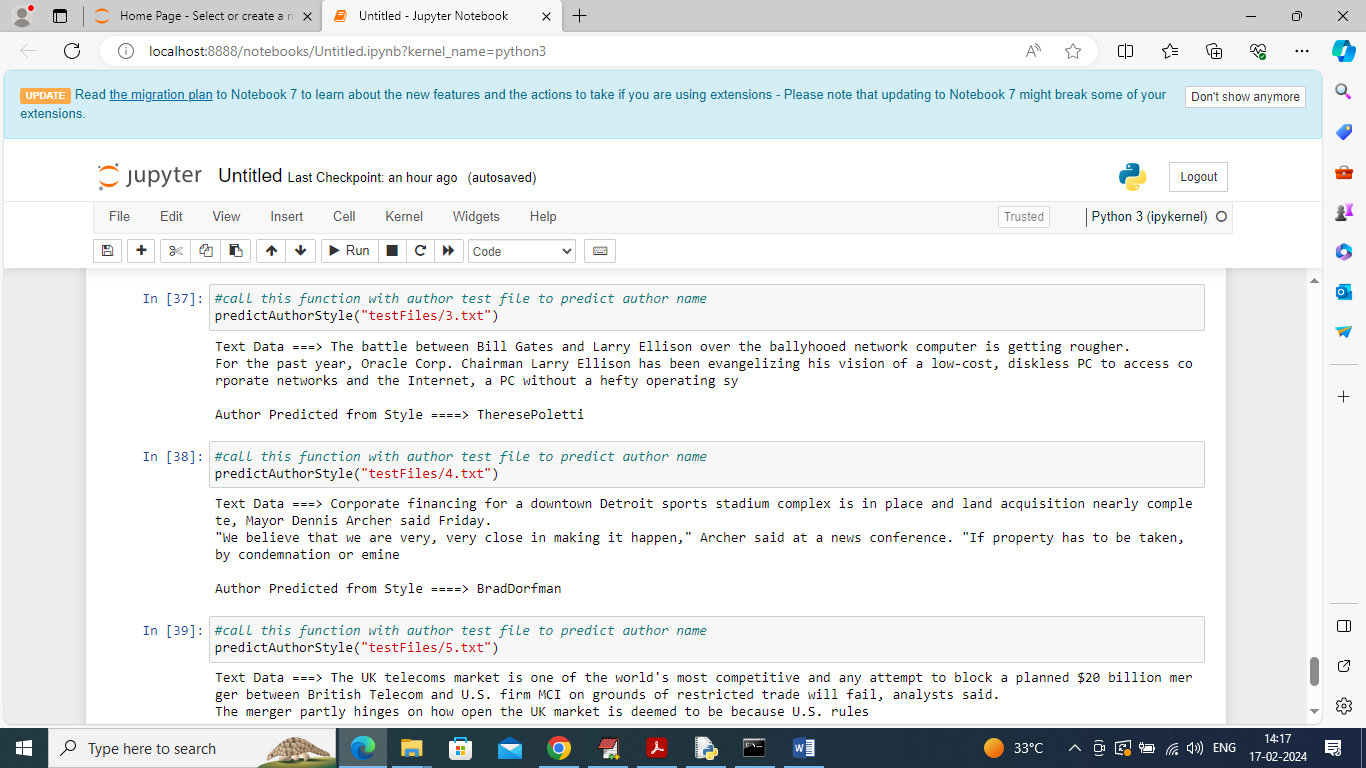
In above screen in table can see LSTM performance



In above screen defining predict function which will take author TEXT file as input and then process text data and then predict author using LSTM based on author style



In above screen calling Predict function with author TEXT data and then displaying some content form TEXT data and then after ARROW =🡺 symbol can see predicted author name from given STYLE TEXT



In above screen predicting author from different TEXT style. Similarly by giving any text file you can predict author from given TEXT style