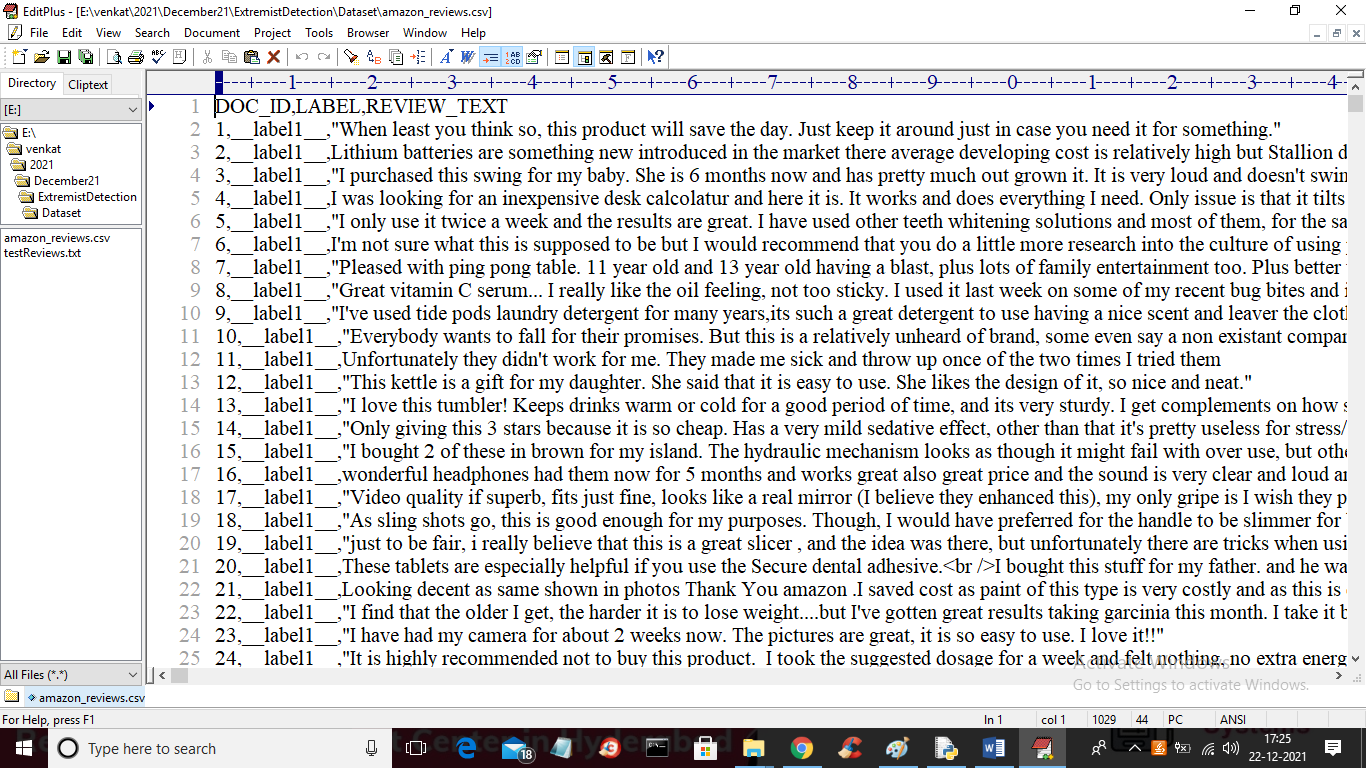
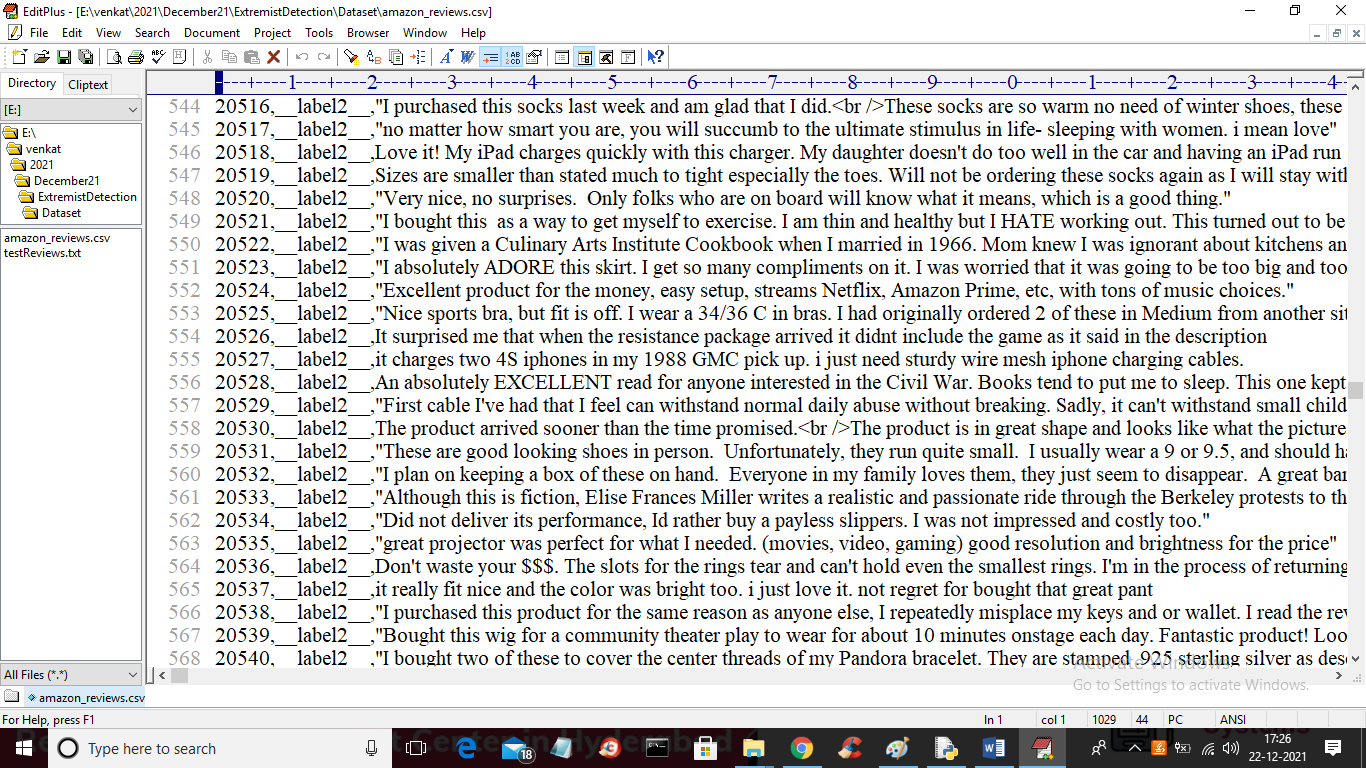
Detecting and Characterizing Extremist Reviewer Groups in Online Product Reviews

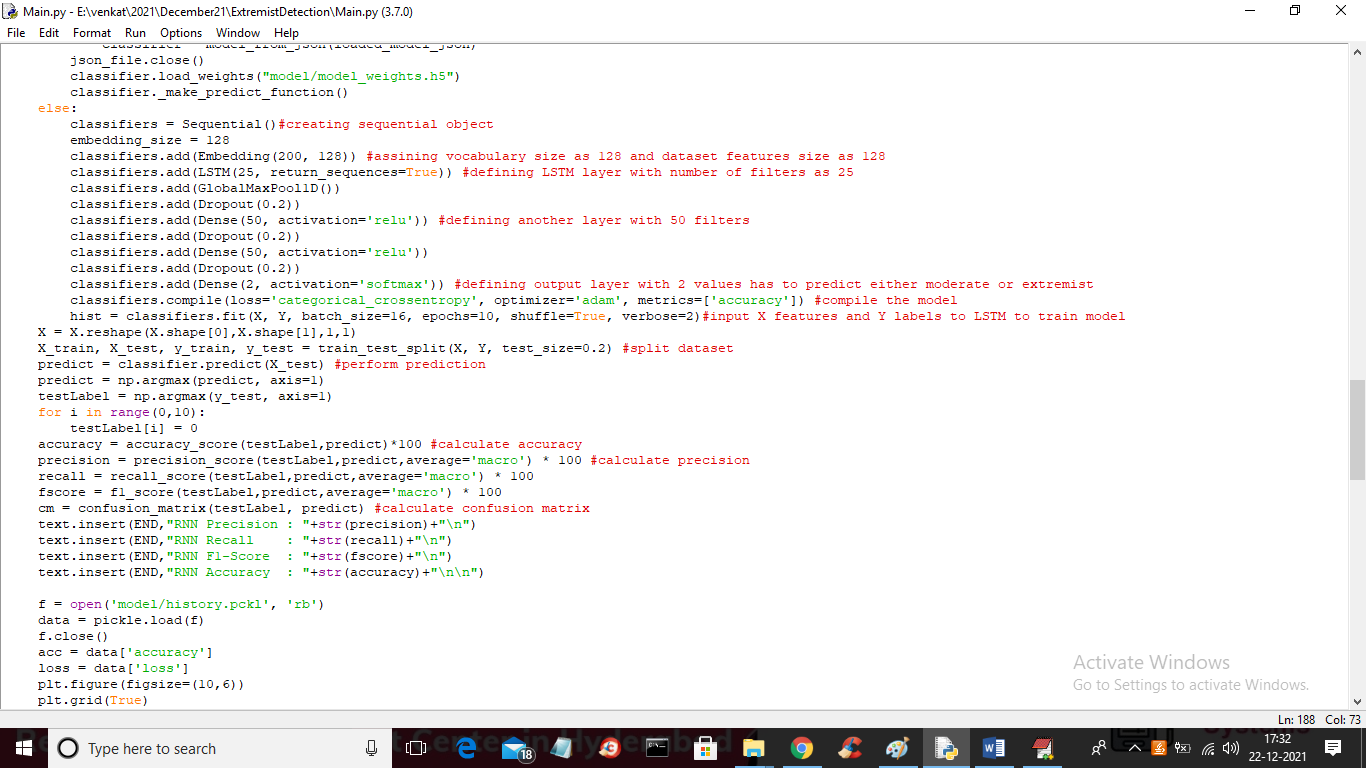
In this project we are using LSTM (type of RNN algorithm) to train Extremist Reviews detection and to train LSTM we have used AMAZON product reviews dataset. To clean dataset by removing stop words and special symbols we have applied NLTK (natural language Tool Kit) technique. After cleaning text we have applied TF-IDF (term frequency inverse document frequency) algorithm to convert reviews in to numeric vector. TF-IDF replace each words with its average frequency and this TF-IDF vector will be input to LSTM to train model.

After training model we can input any review then LSTM will predict whether that input review is EXTREMIST of MODERATE. Below is the AMAZON dataset used in this project.





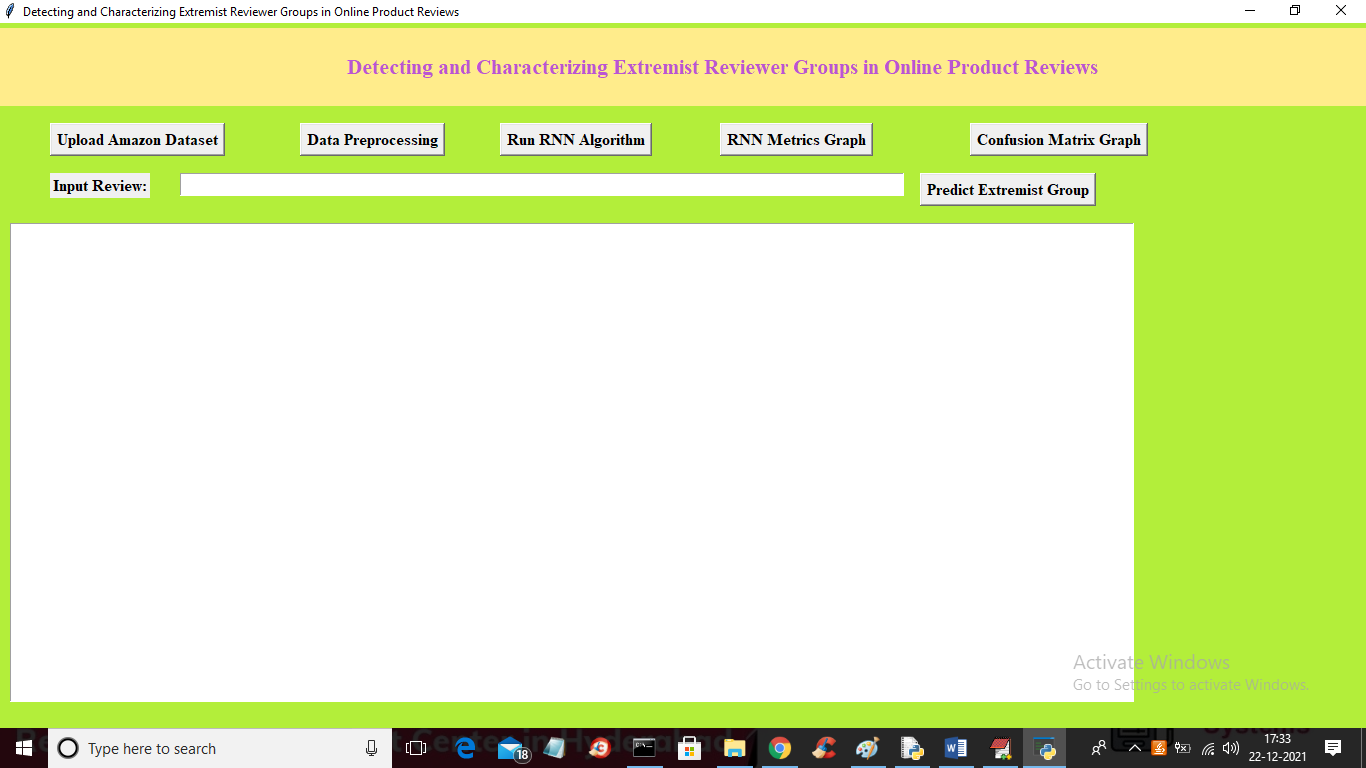
In above screen label1 refers to MODERATE and label2 refers to EXTREMIST and after label we can find text review column. In below screen you can see we input training X (review TF-IDF features) and Y (class labels) as input to LSTM to train model.



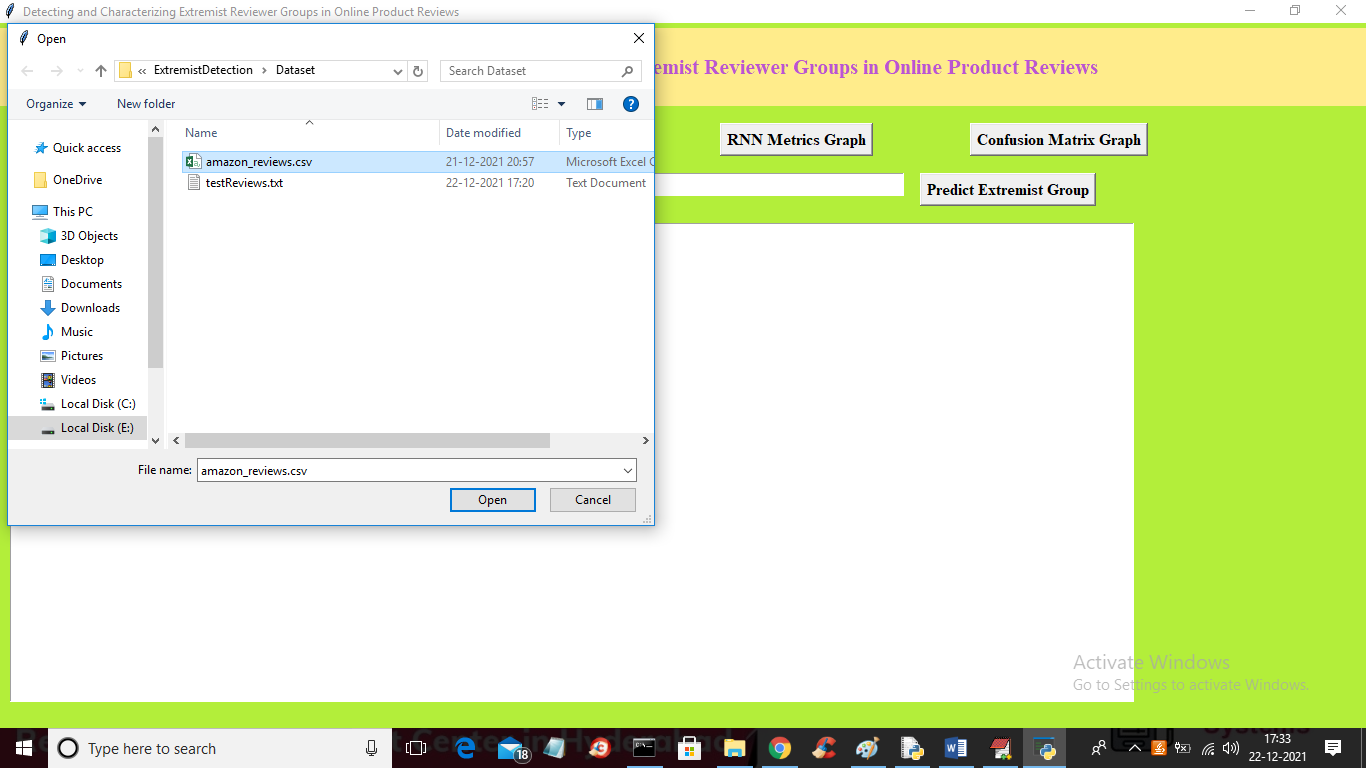
In above screen read red colour comments to know about RNN LSTM training.

SCREEN SHOTS

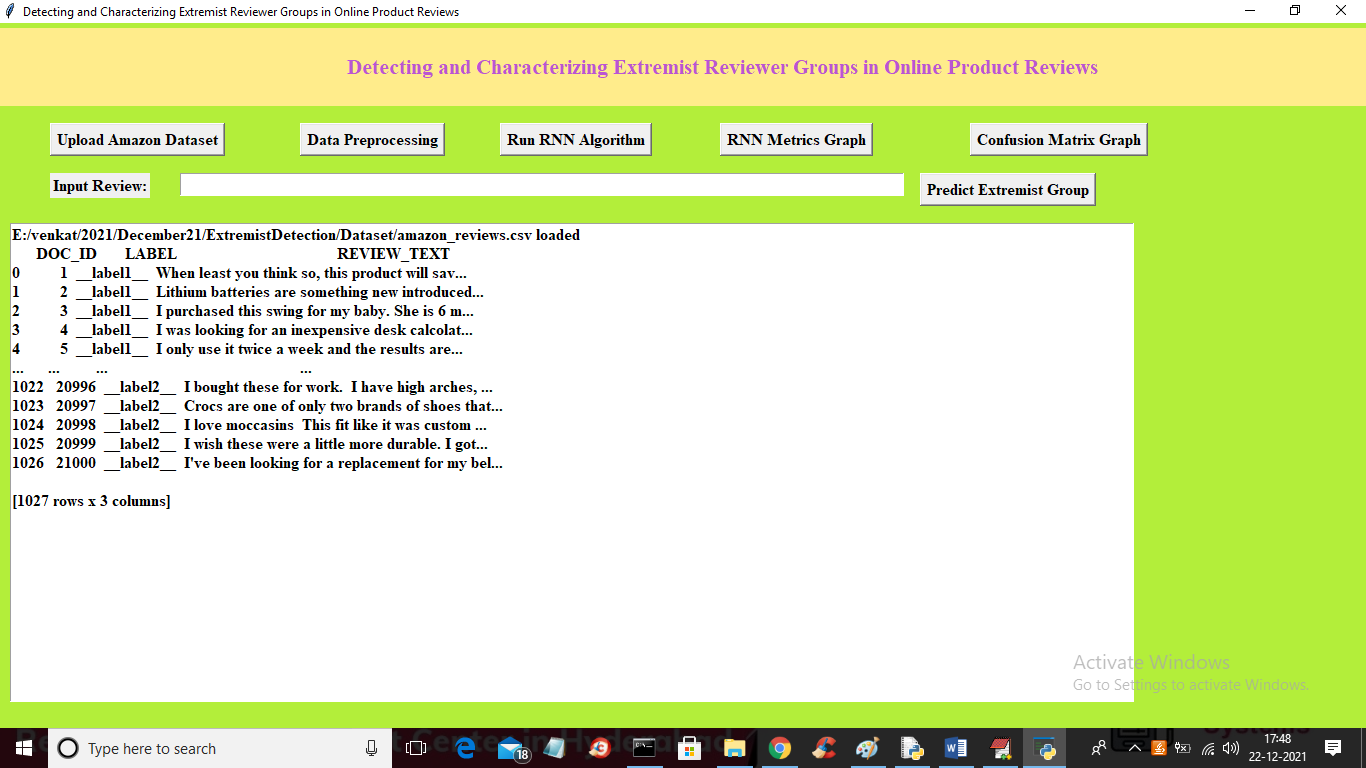
To run project double click on ‘run.bat’ file to get below screen



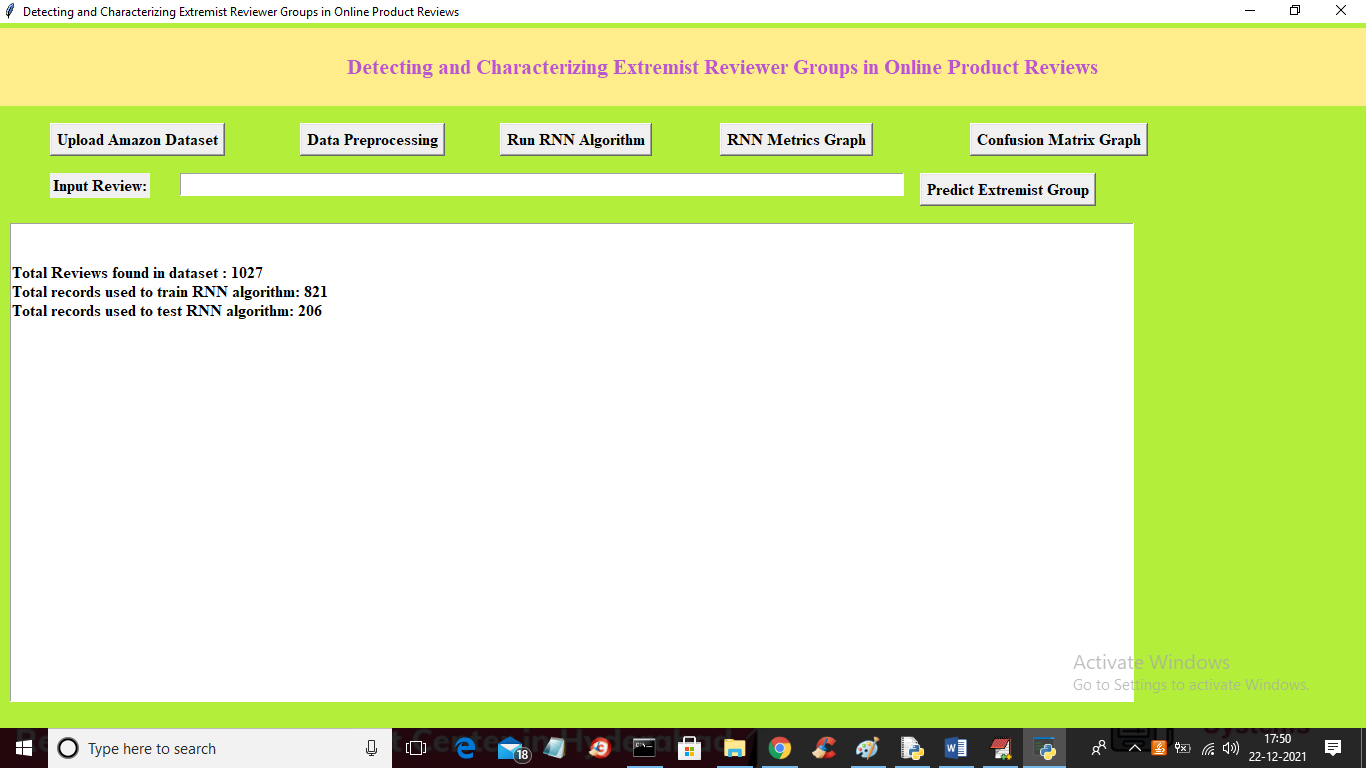
In above screen click on ‘Upload Amazon Dataset’ button to load dataset and to get below screen



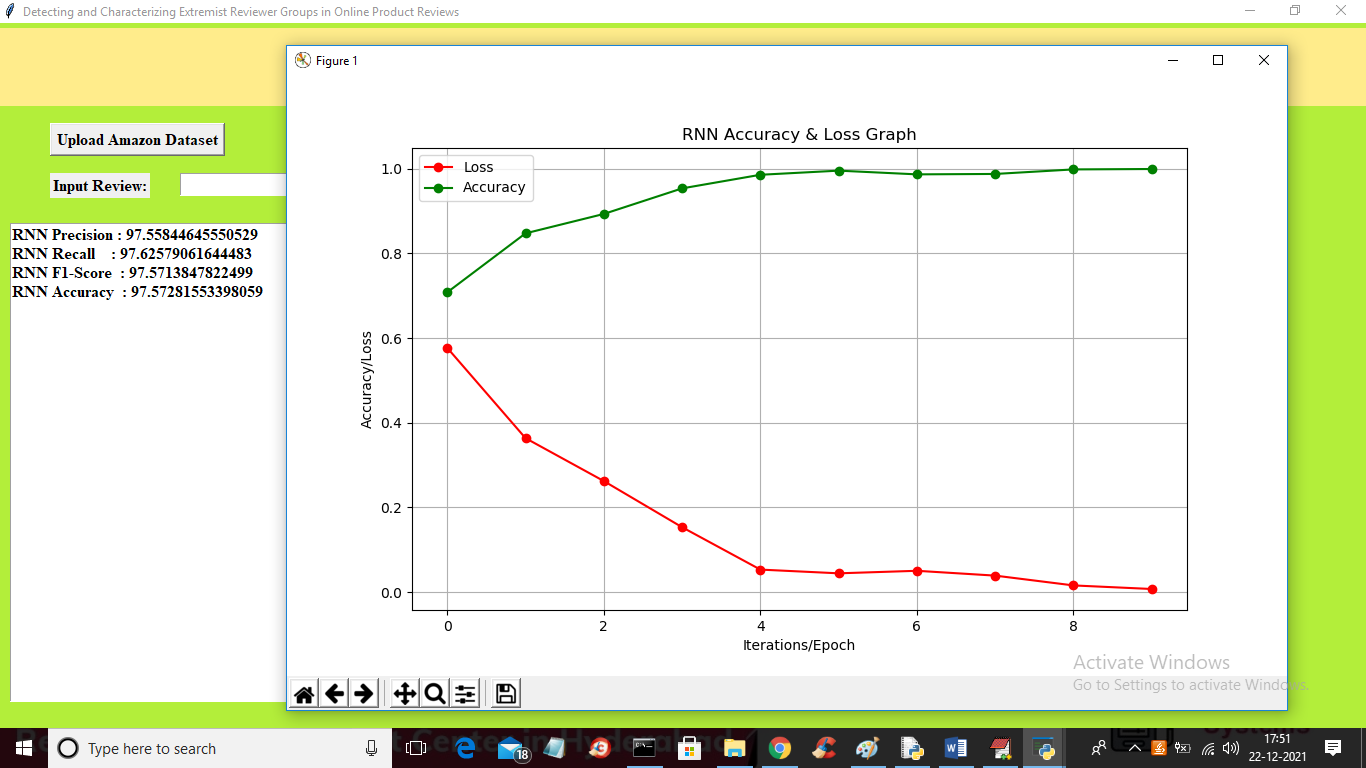
In above screen selecting and uploading amazon\_reviews.csv file and then click on ‘Open’ button to load dataset and to get below screen



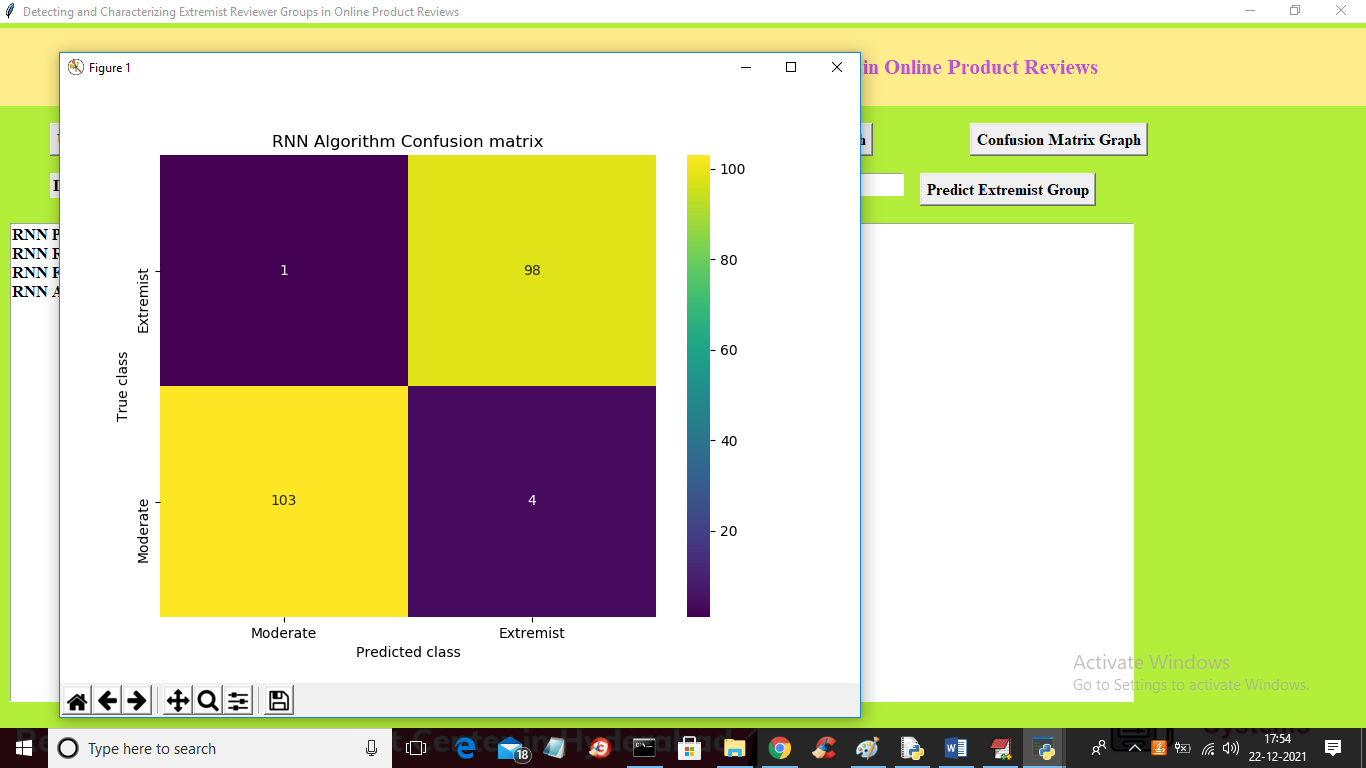
In above screen dataset loaded and reviews contains special symbols and stop words like (the and or where etc.) and we can remove such words by pre-processing reviews and then split dataset into train and test by clicking on ‘Data Preprocessing’ button



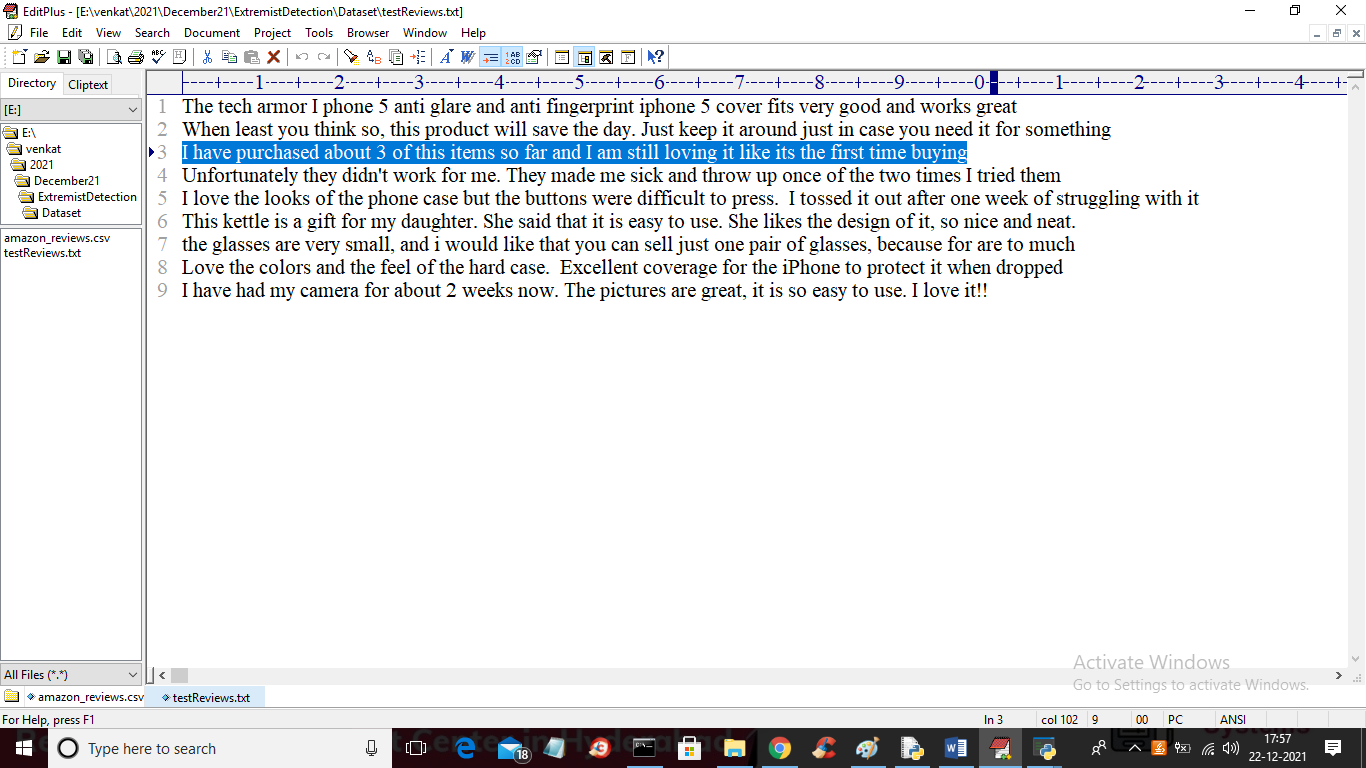
In above screen after cleaning reviews application displaying total reviews as 1027 and then using 821 reviews for LSTM training and 206 reviews for testing and now cleaned reviews are and now click on ‘ Run RNN Algorithm’ button to train data with LSTM and to get below screen



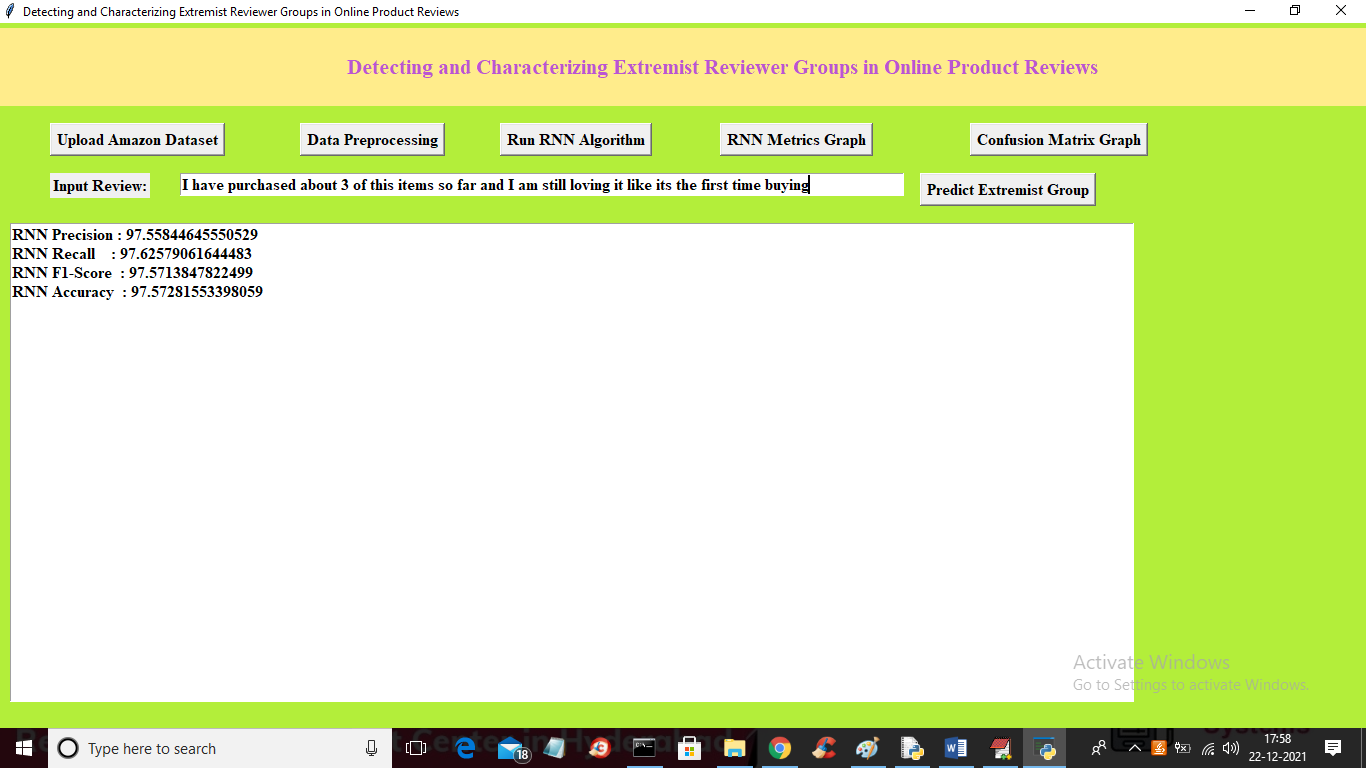
In above screen we got RNN accuracy as 97% and in above graph red line represents LOSS values and green line represents accuracy values and to train LSTM we took 10 epoch and x-axis represents epoch and y-axis represents accuracy. In above graph with each increasing epoch accuracy got increase and loss got decrease and accuracy reached closer to 100%. Now close above graph and then click on ‘Confusion Matrix Graph’ button to get below graph



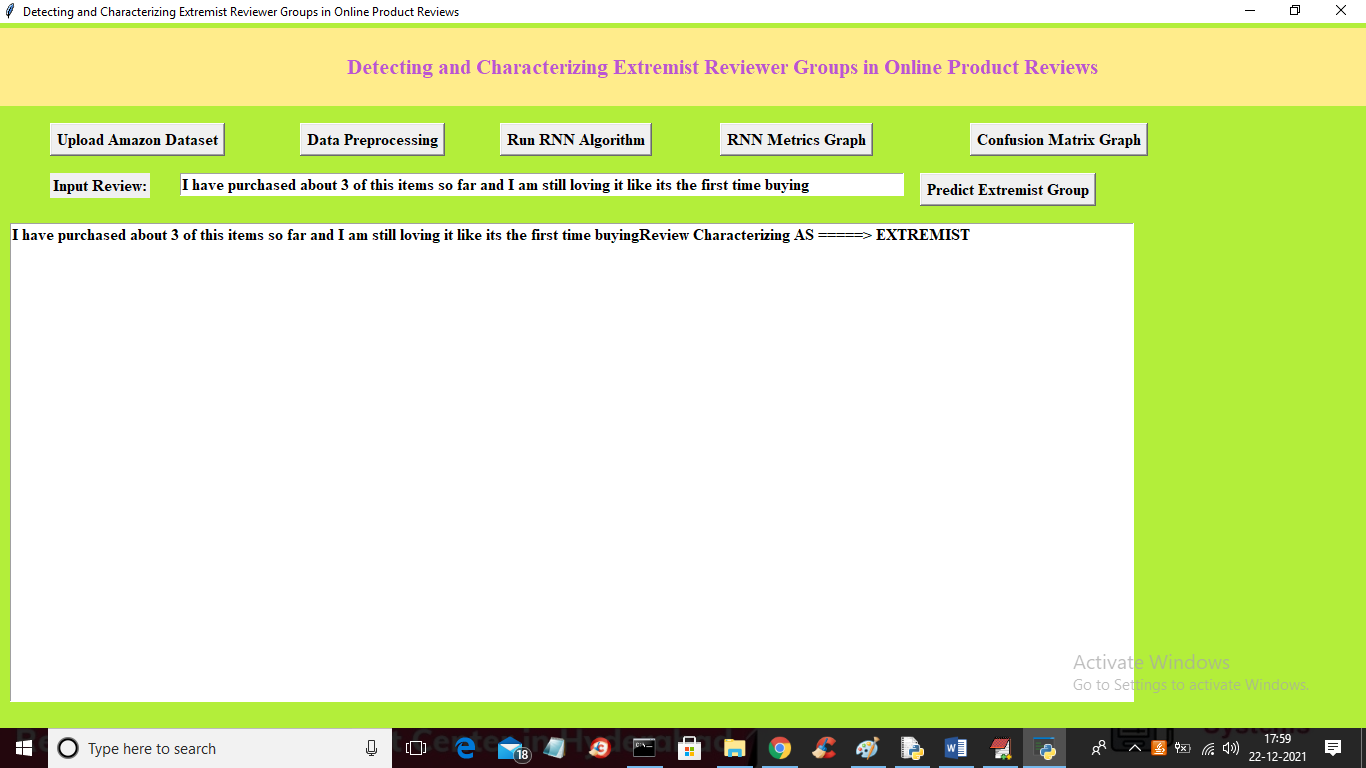
In above confusion matrix x-axis represents Predicted class and y-axis represents TRUE or original class values and in above 103 records are correctly predicted as MODERATE and 1 class is incorrectly predicted as EXTREMIST. For Extremist prediction also 4 records are incorrectly predicted and ‘MODERATE’ and 98 correctly predicted as EXTREMIST. Now close above graph and then enter some reviews by copying from ‘TestReviews.csv’ file and then click on ‘Predict Extremist Group’ button to predict review as extremist or moderate.



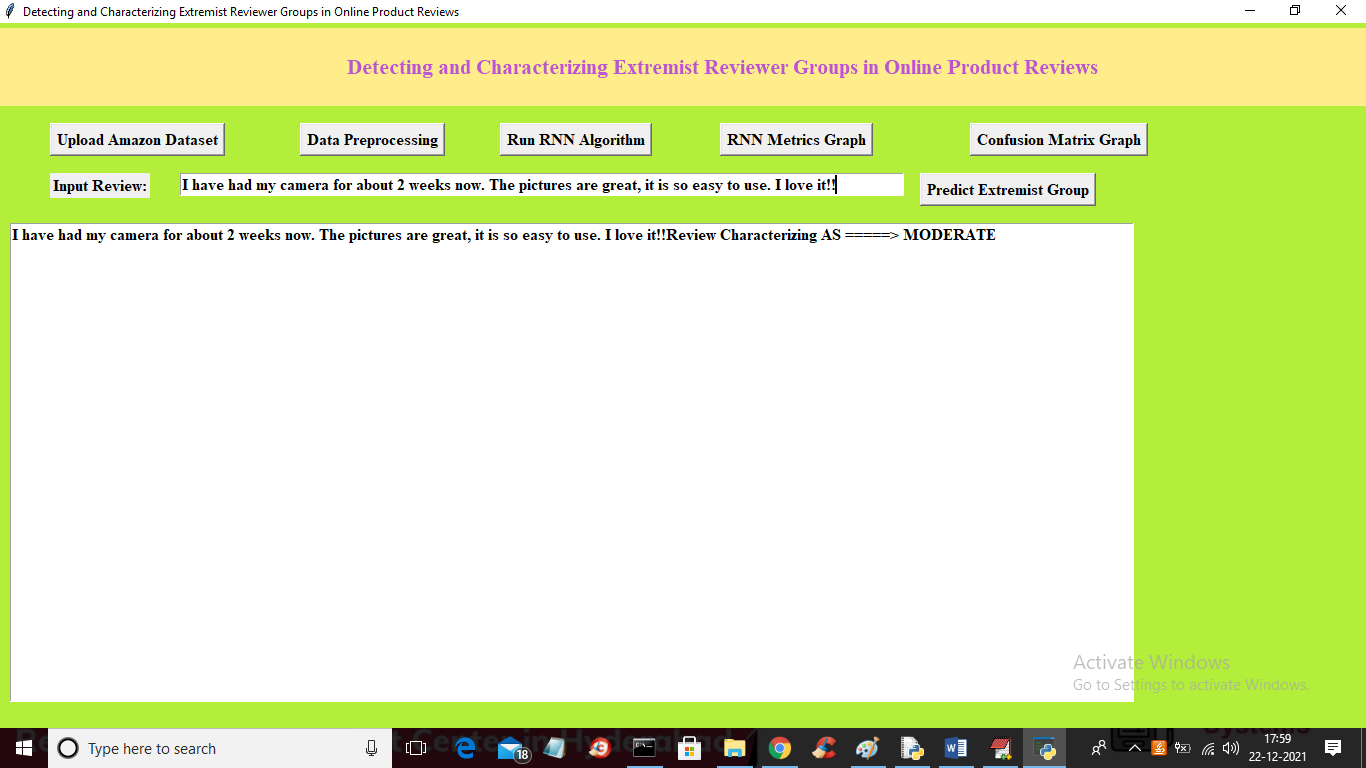
From above test reviews screen I am selecting and copying one review and paste in application field



Now in above screen in text field I paste some review and then click on ‘Predict Extremist Review’ button to get below output



In above screen in text area after arrow symbol given review predicted as ‘EXTREMIST’ and now test another review



In above screen new review predicted as ‘MODERATE’ and similarly you can input any review and get prediction result