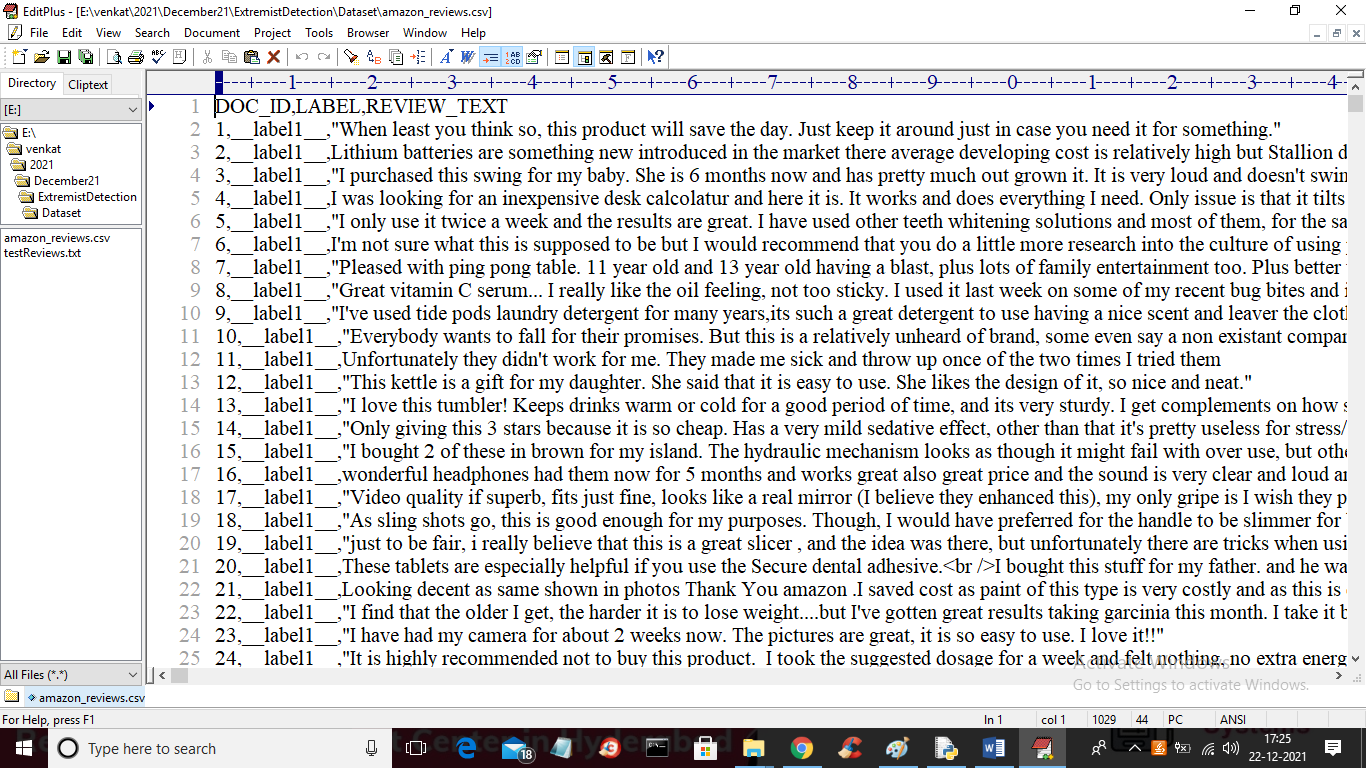
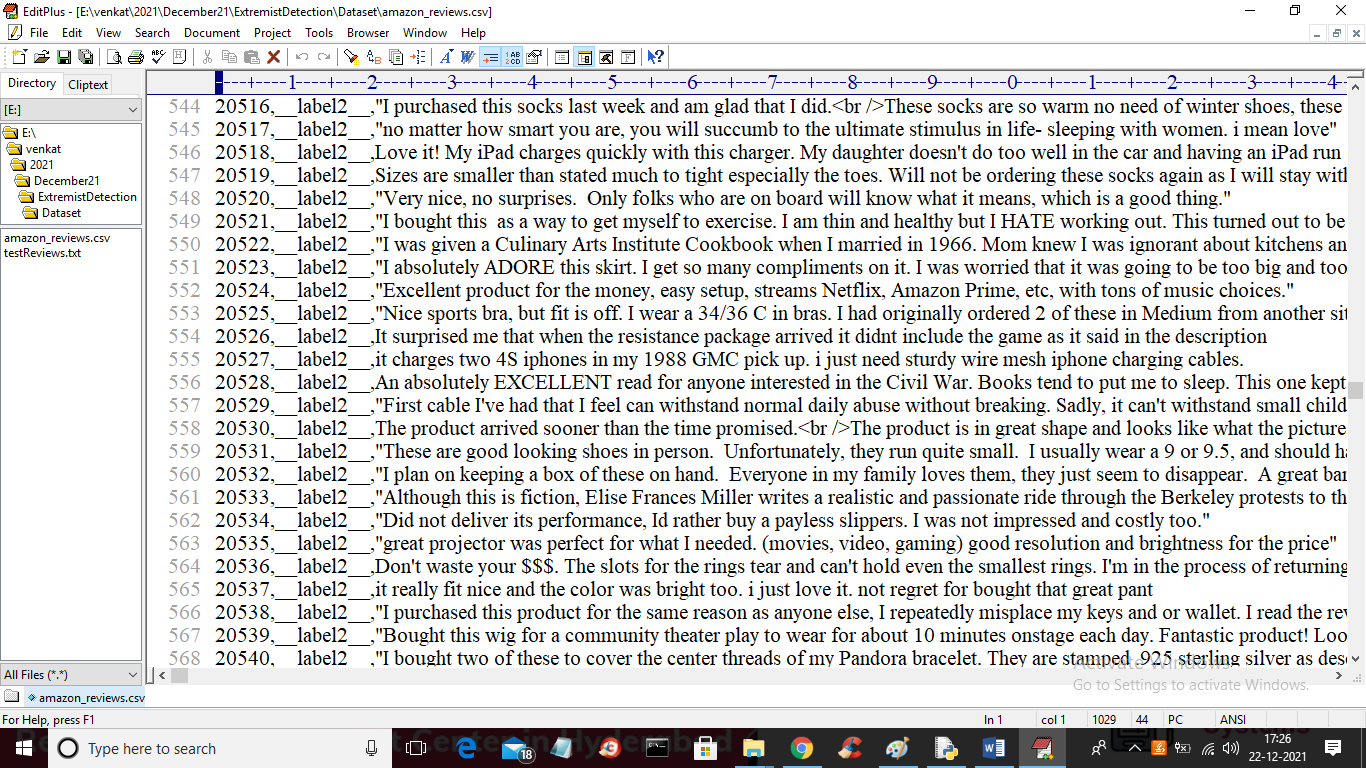
Detecting and Characterizing Extremist Reviewer Groups in Online Product Reviews

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.

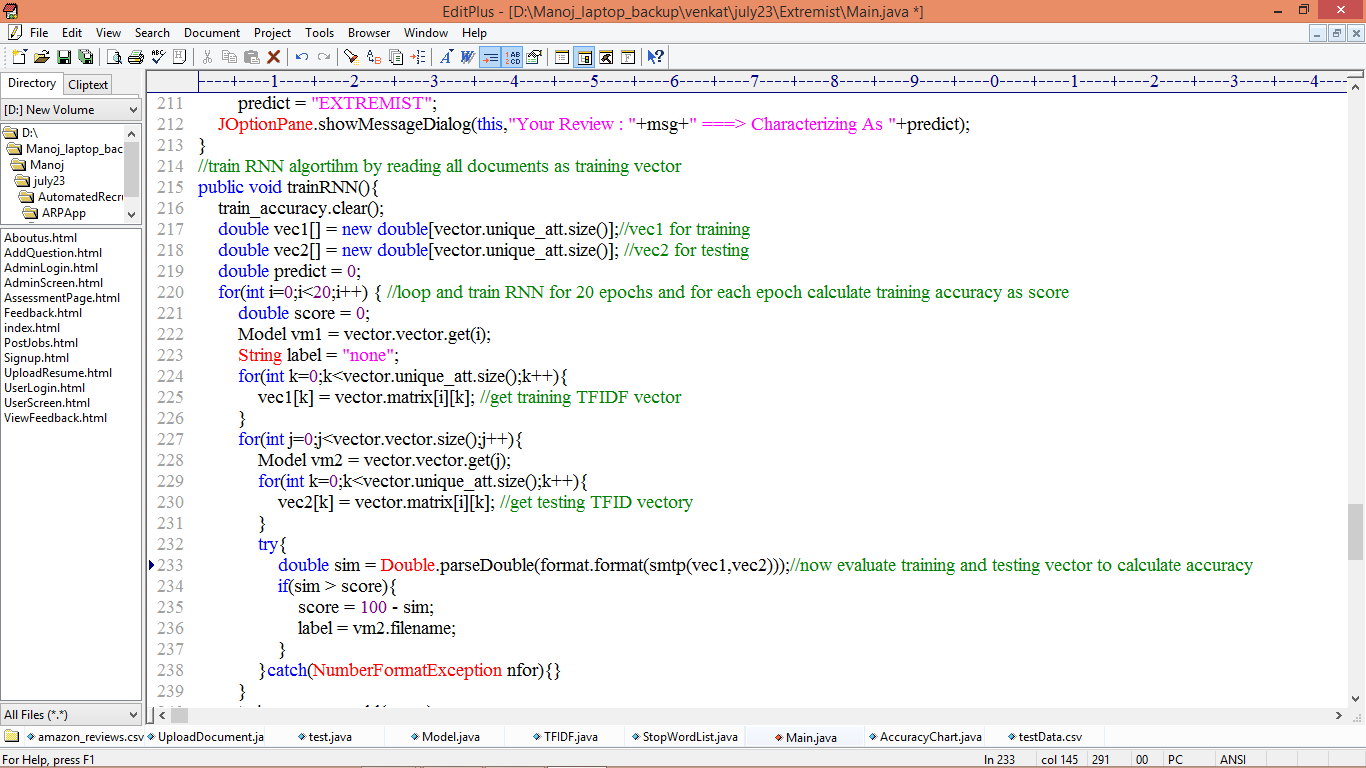
We have implemented this algorithm using JAVA

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.





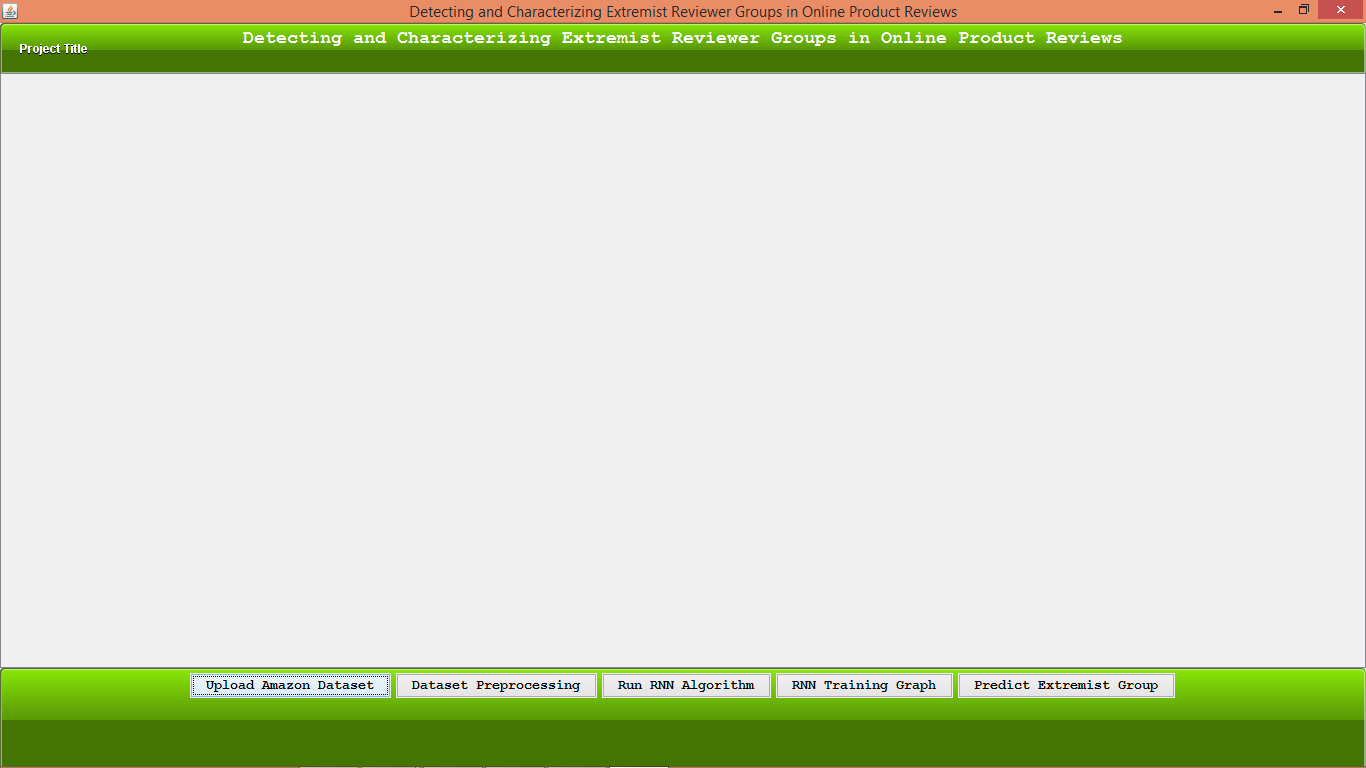
In above screen label1 refers to MODERATE and label2 refers to EXTREMIST and after label we can find text review column. In below screen you can see we input training X (review TF-IDF features) and Y (class labels) as input to LSTM to train model.



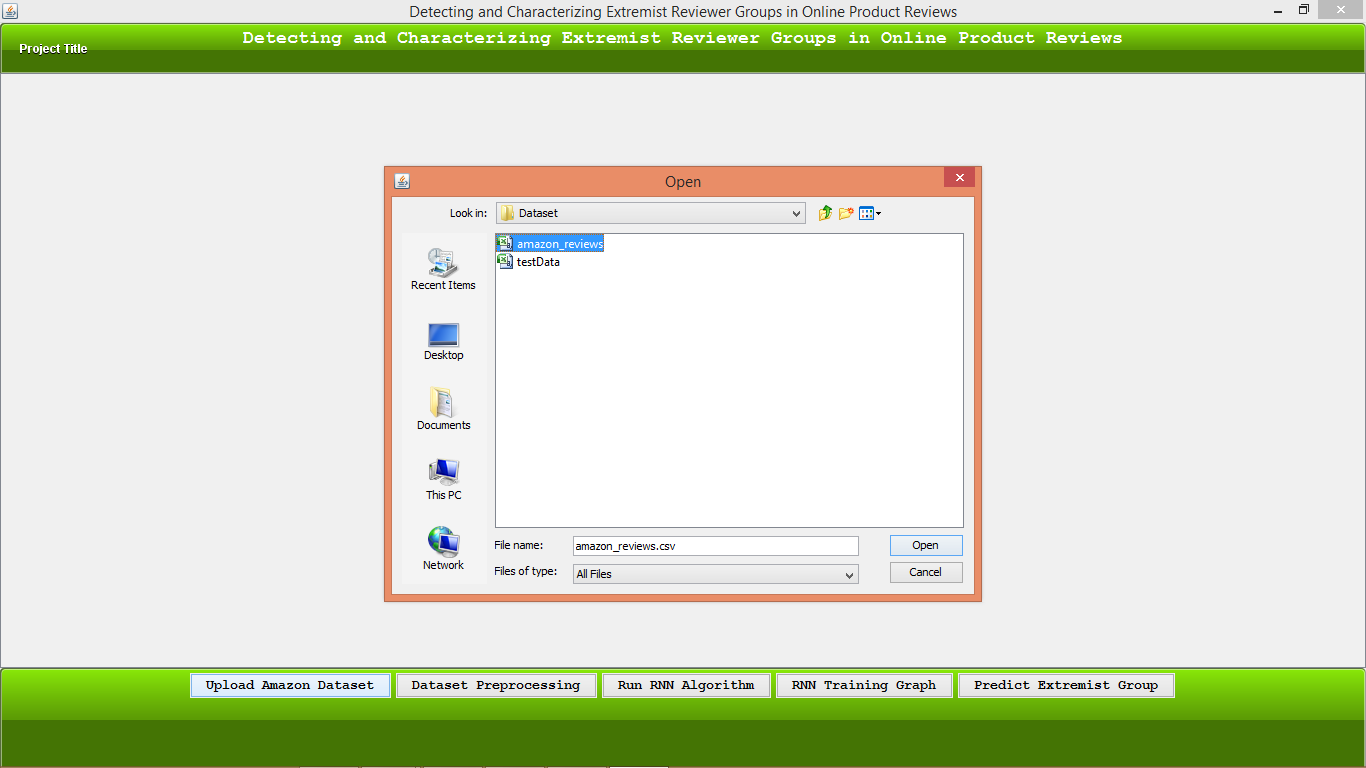
In above screen read green colour comments to know about RNN LSTM training.

SCREEN SHOTS

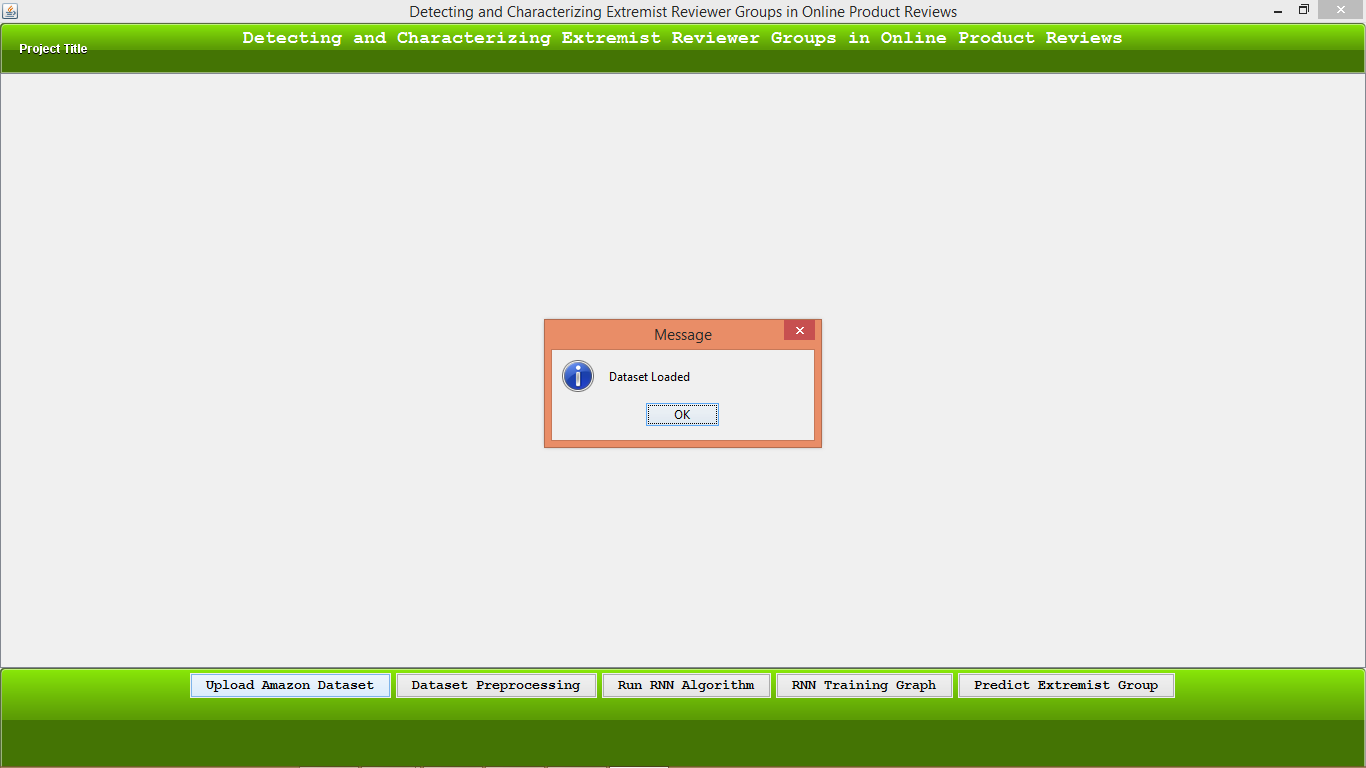
To run project double click on ‘run.bat’ file to get below screen



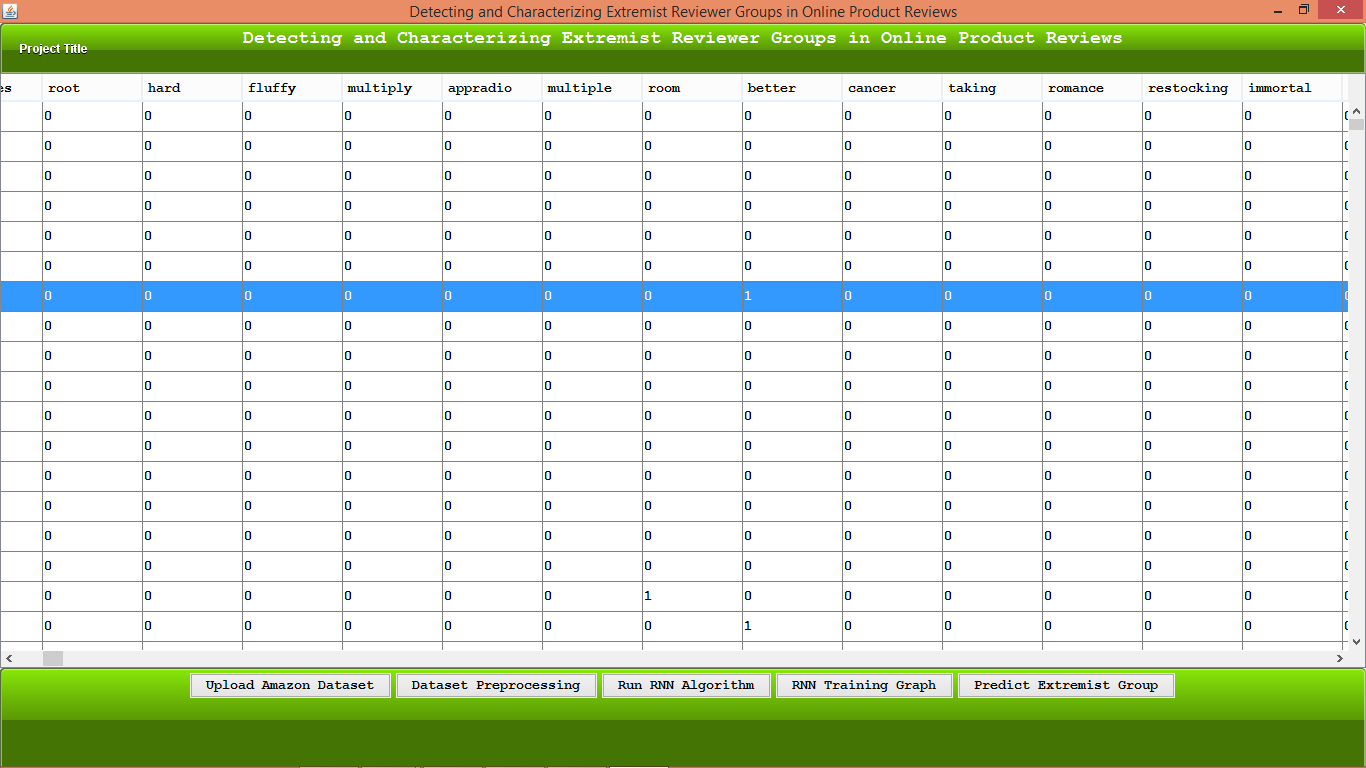
In above screen click on ‘Upload Amazon Dataset’ button to load dataset and to get below screen



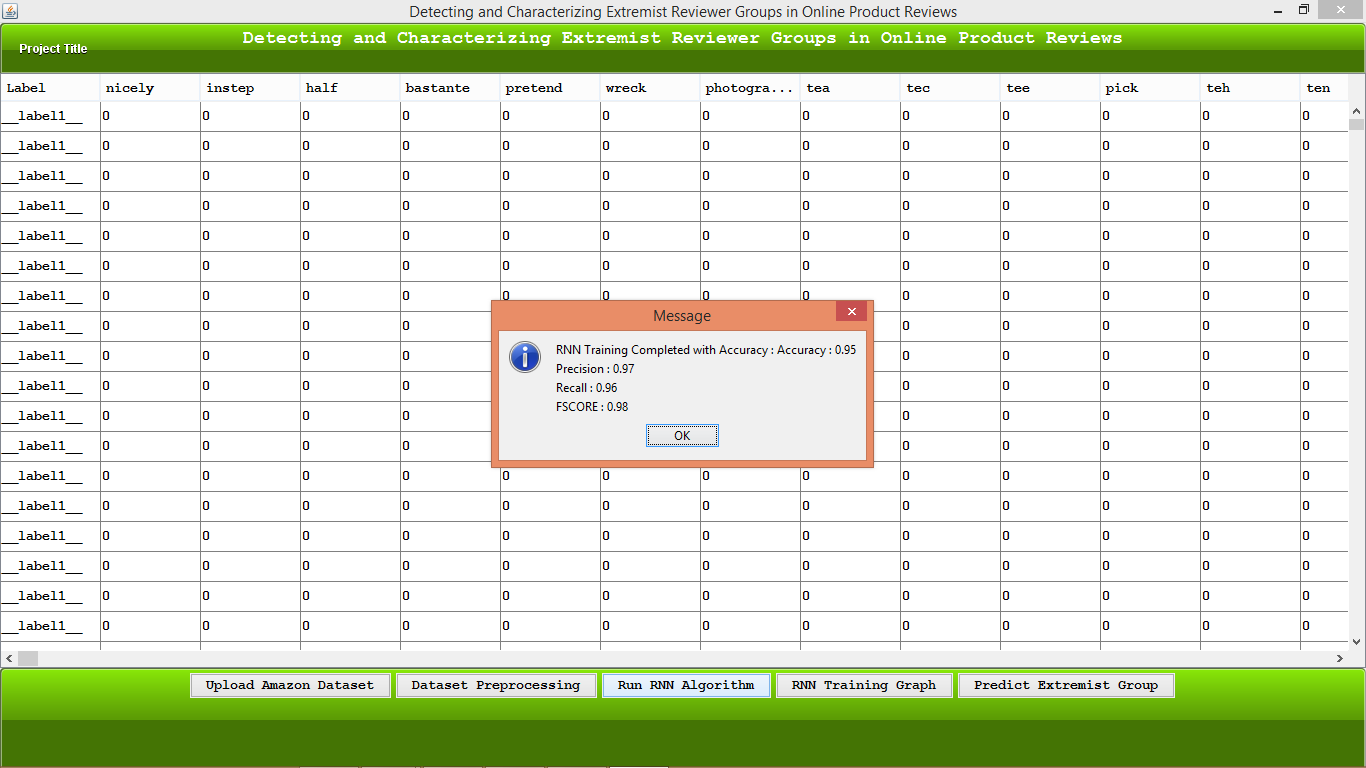
In above screen selecting and uploading amazon\_reviews.csv file and then click on ‘Open’ button to load dataset and to get below screen



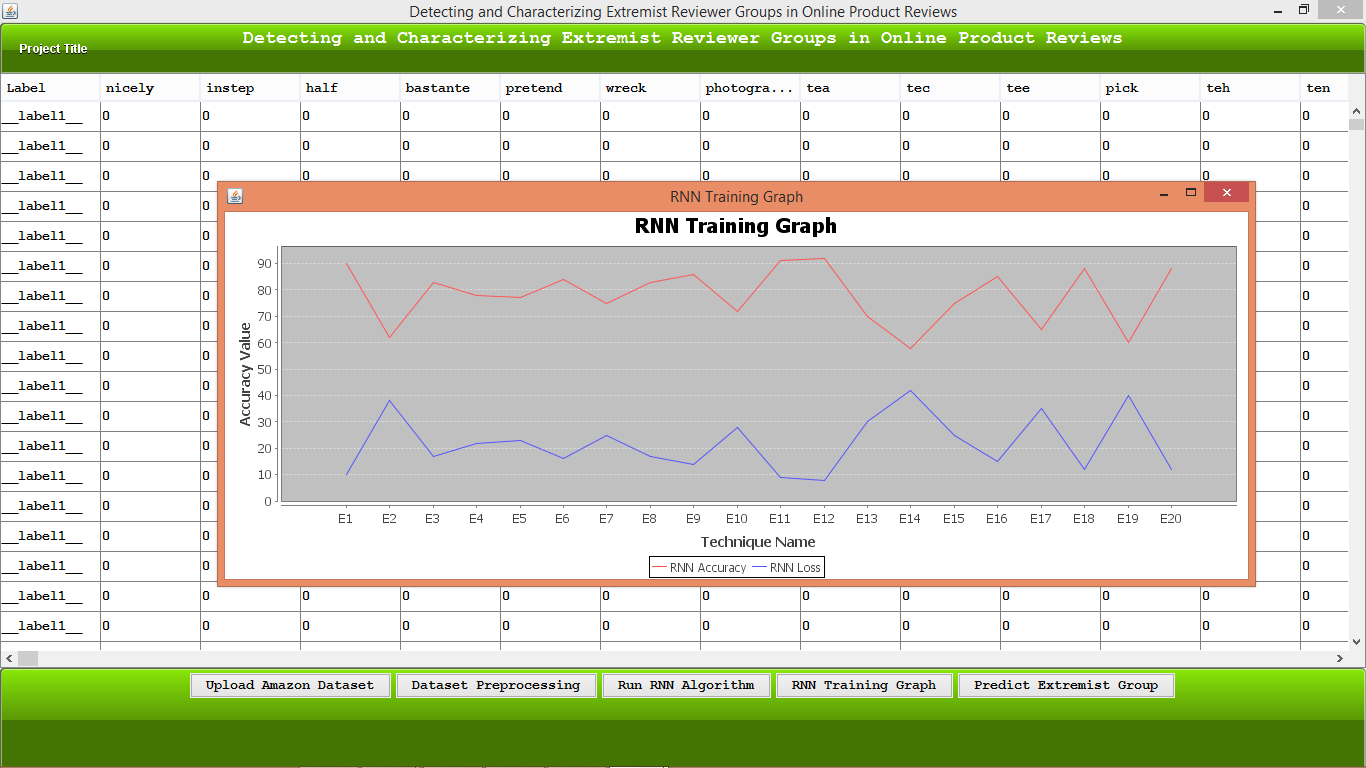
In above screen dataset loaded and reviews contains special symbols and stop words like (the and or where etc.) and we can remove such words by pre-processing reviews and then split dataset into train and test and then convert entire dataset into TFIDF vector by clicking on ‘Data Preprocessing’ button



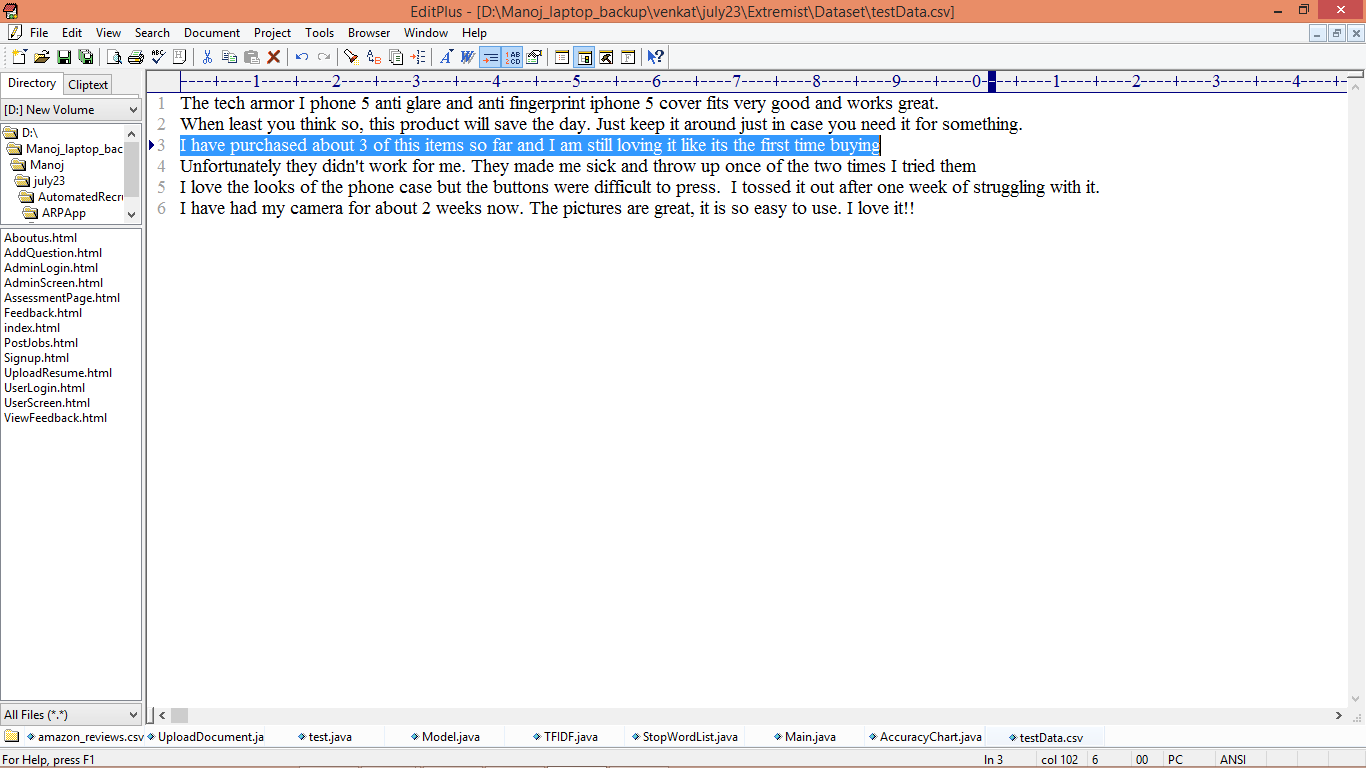
In above screen each review is converted to TFIDF vector where column header represents word names from all reviews and column rows represents word occurrence count and if 0 means word not occur in that review and now cleaned reviews are and now click on ‘ Run RNN Algorithm’ button to train data with LSTM and to get below screen



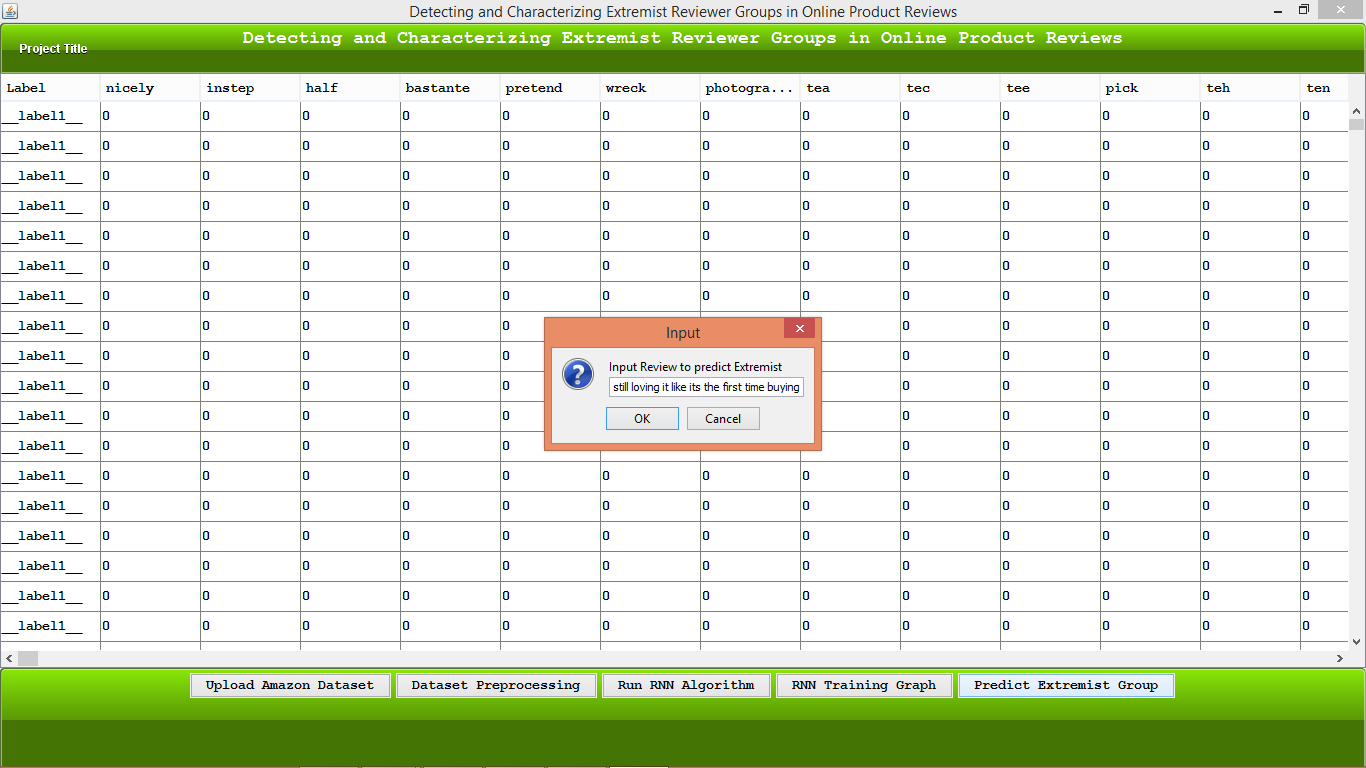
In above screen we got RNN accuracy as 95% and we can see other metrics like precision, recall and FSCORE and now click on ‘RNN Training Graph’ button to get below graph.



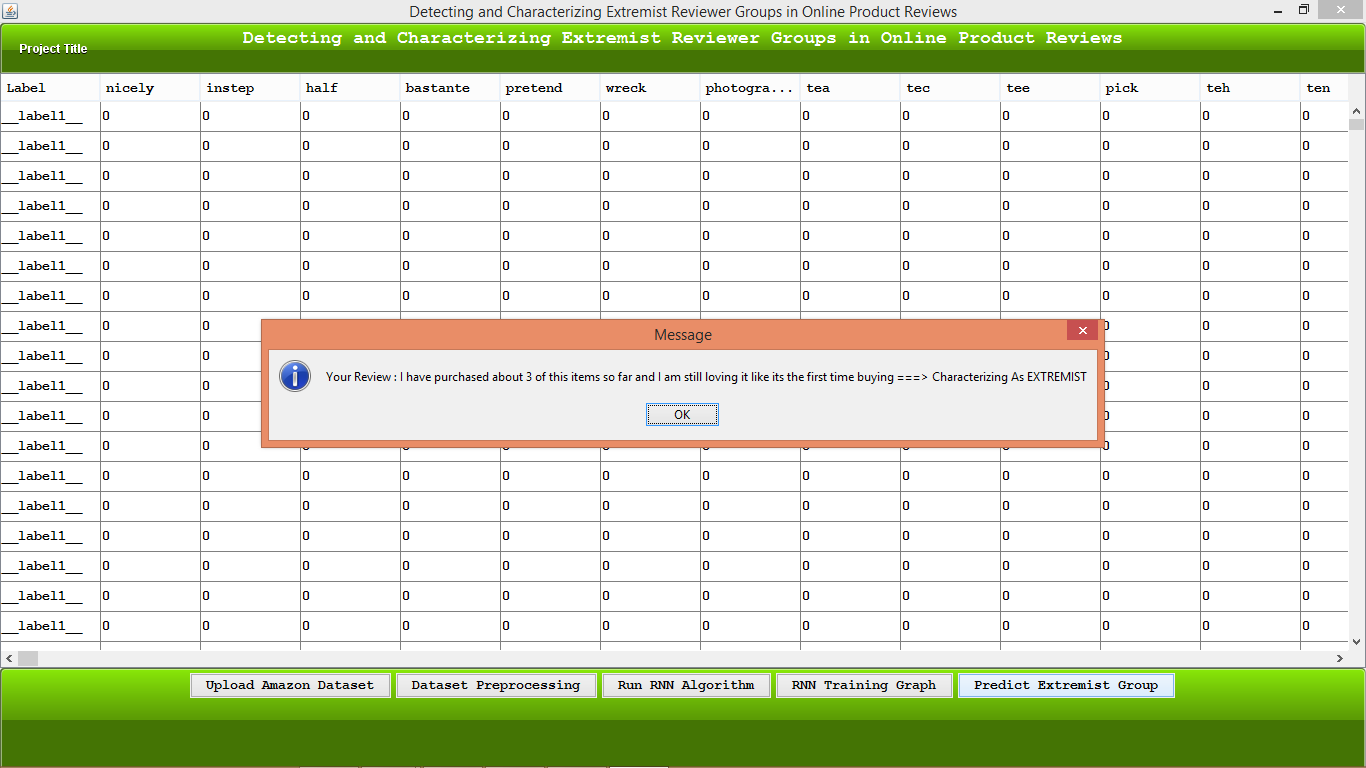
in above graph red line represents ACCURACY values and blue line represents loss values and to train LSTM we took 20 epoch and x-axis represents epoch and y-axis represents accuracy and loss values. In above graph with each increasing epoch accuracy got increase and loss got. Now close above graph and then enter some reviews by copying from ‘TestData.csv’ file and then click on ‘Predict Extremist Group’ button to predict review as extremist or moderate.



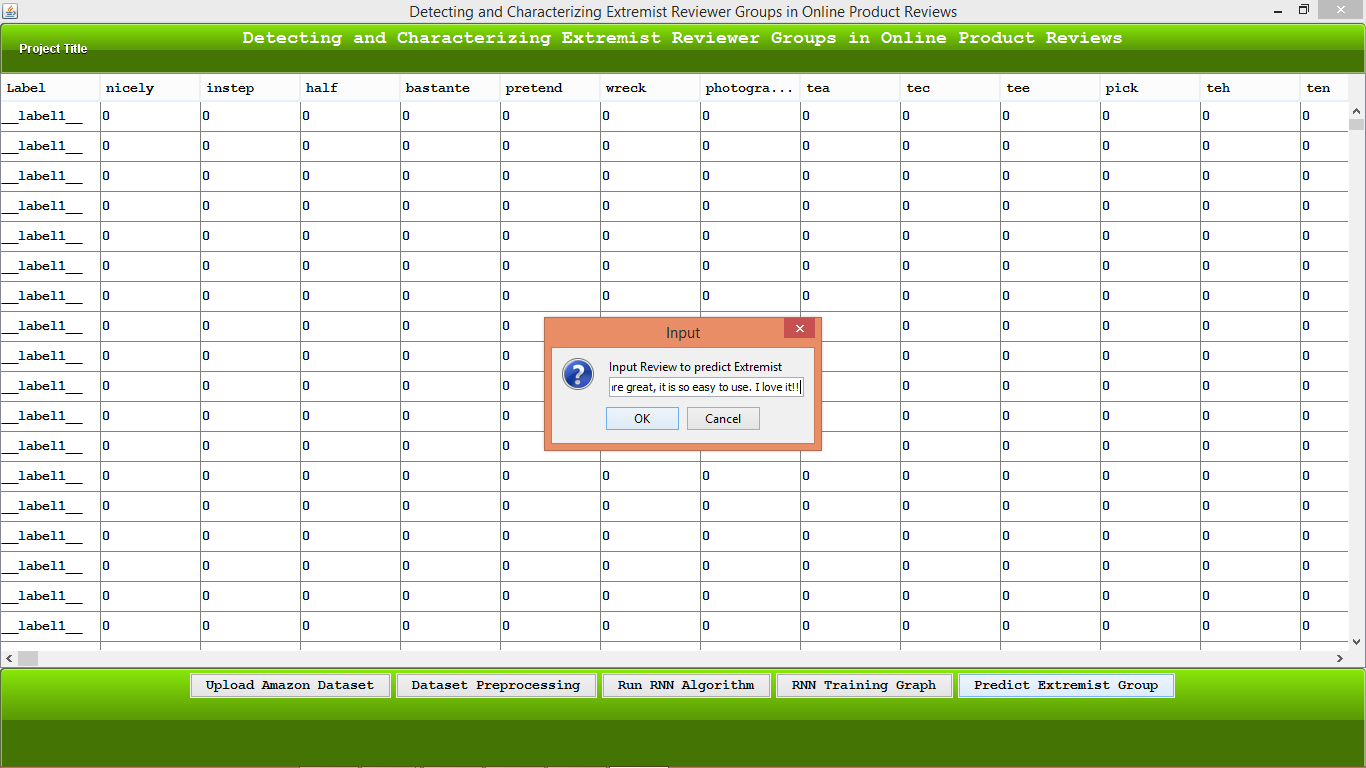
From above test reviews screen I am selecting and copying one review and paste in application dialog box by clicking ‘Predict Extremist Group’ button like below screen



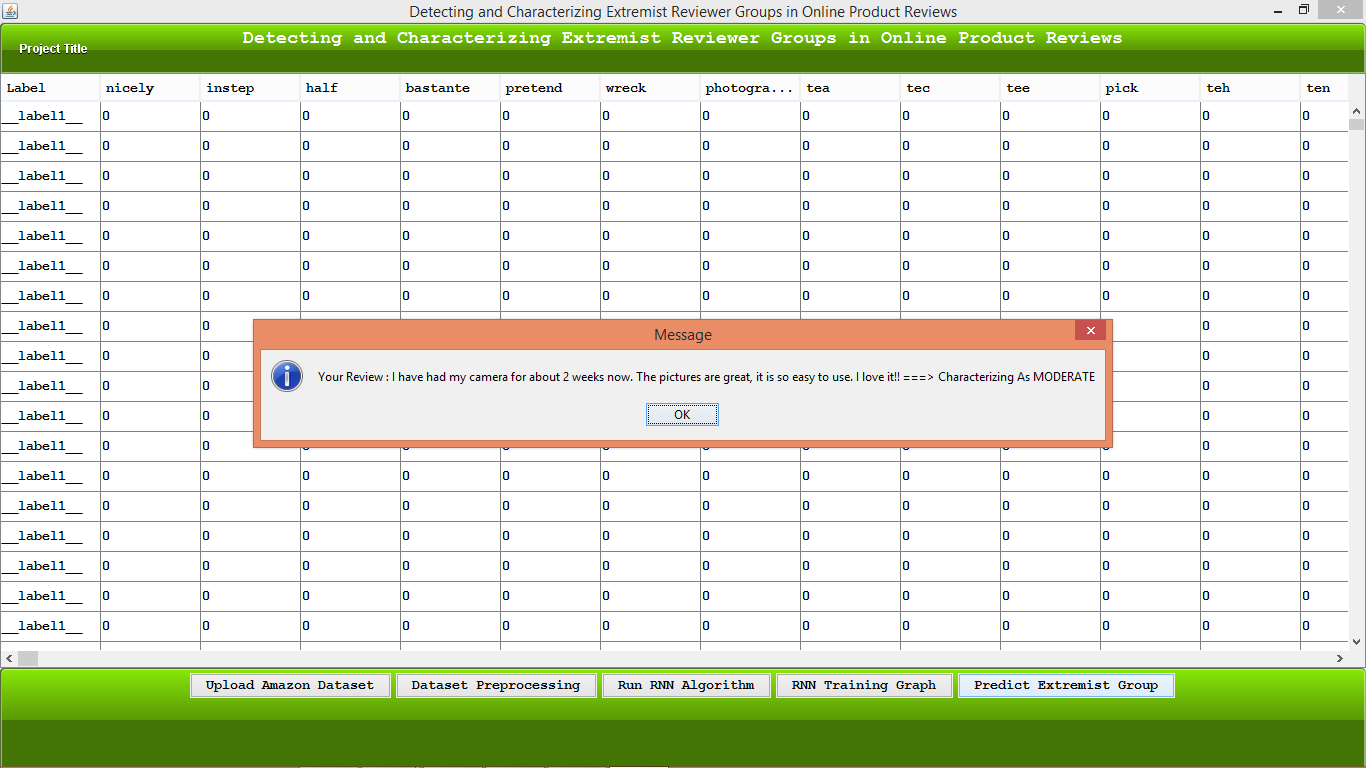
Now in above screen in dialog box I paste some review and then click on ‘OK’ button to get below output



In above screen in output dialog box after arrow symbol given review predicted as ‘EXTREMIST’ and now test another review



In above screen pasted another review and now click OK button to get below output



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

Similarly you can enter any review and predict as Extremist or moderate