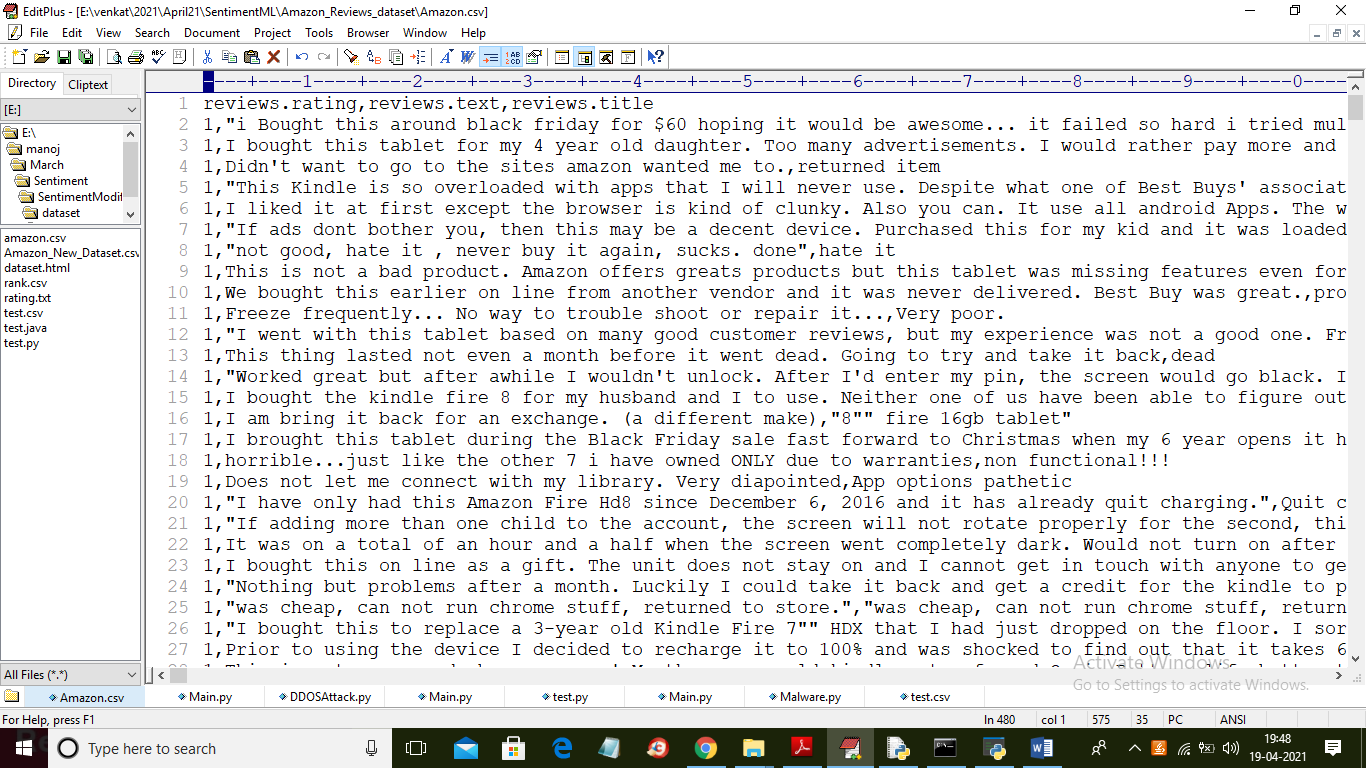
Fake Product Review Monitoring System

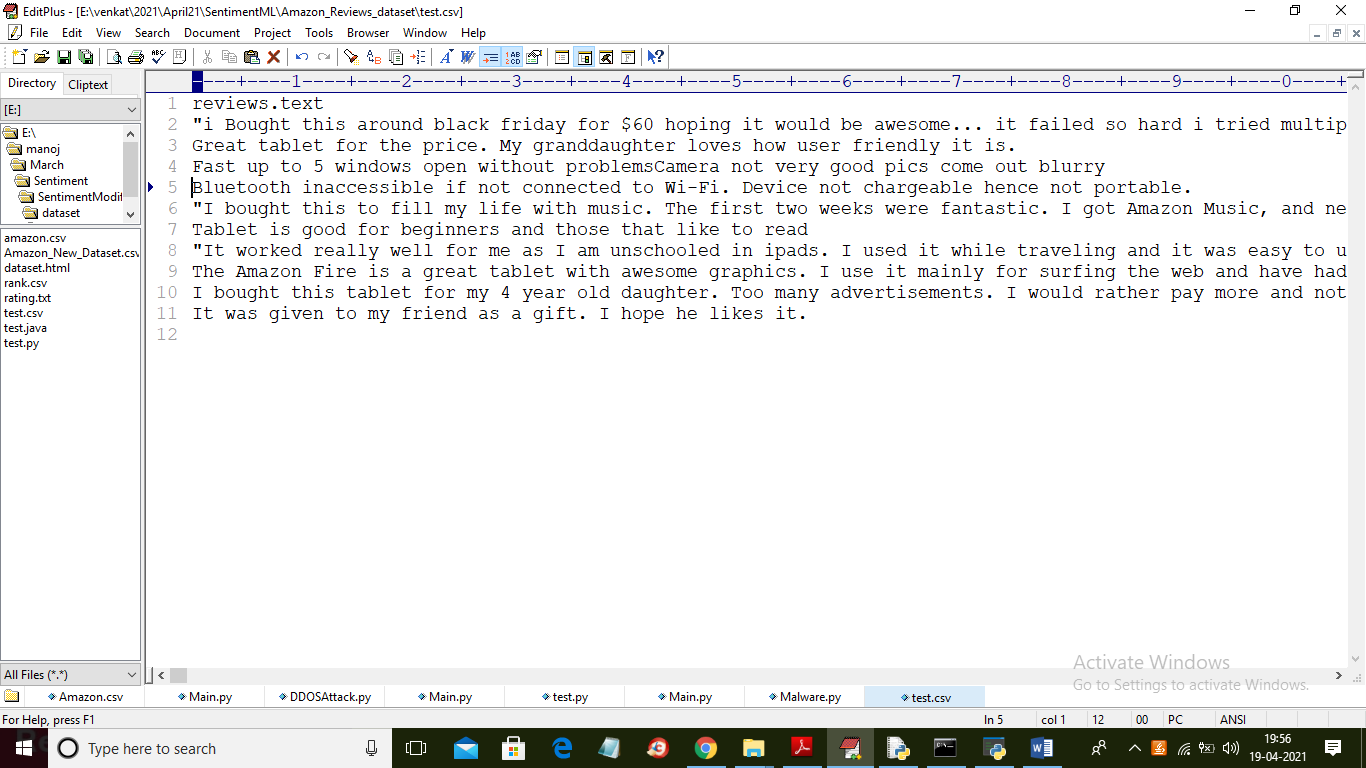
In this project author is detecting sentiments from amazon reviews by using various machine learning algorithms such as SVM, Decision Tree and Naïve Bayes. In all 3 algorithms SVM is giving better accuracy and to train this algorithms author has used AMAZON reviews dataset and this dataset is saved inside ‘Amazon\_Reviews\_dataset’ folder. Below screen shot show example reviews from dataset



In above dataset first row contains column names and remaining rows contains dataset values and in above dataset first column contains sentiment values from 1 to 5 and its associated with each review and we will use above dataset to train all 3 machine learning algorithms.

To implement this project author has used following modules

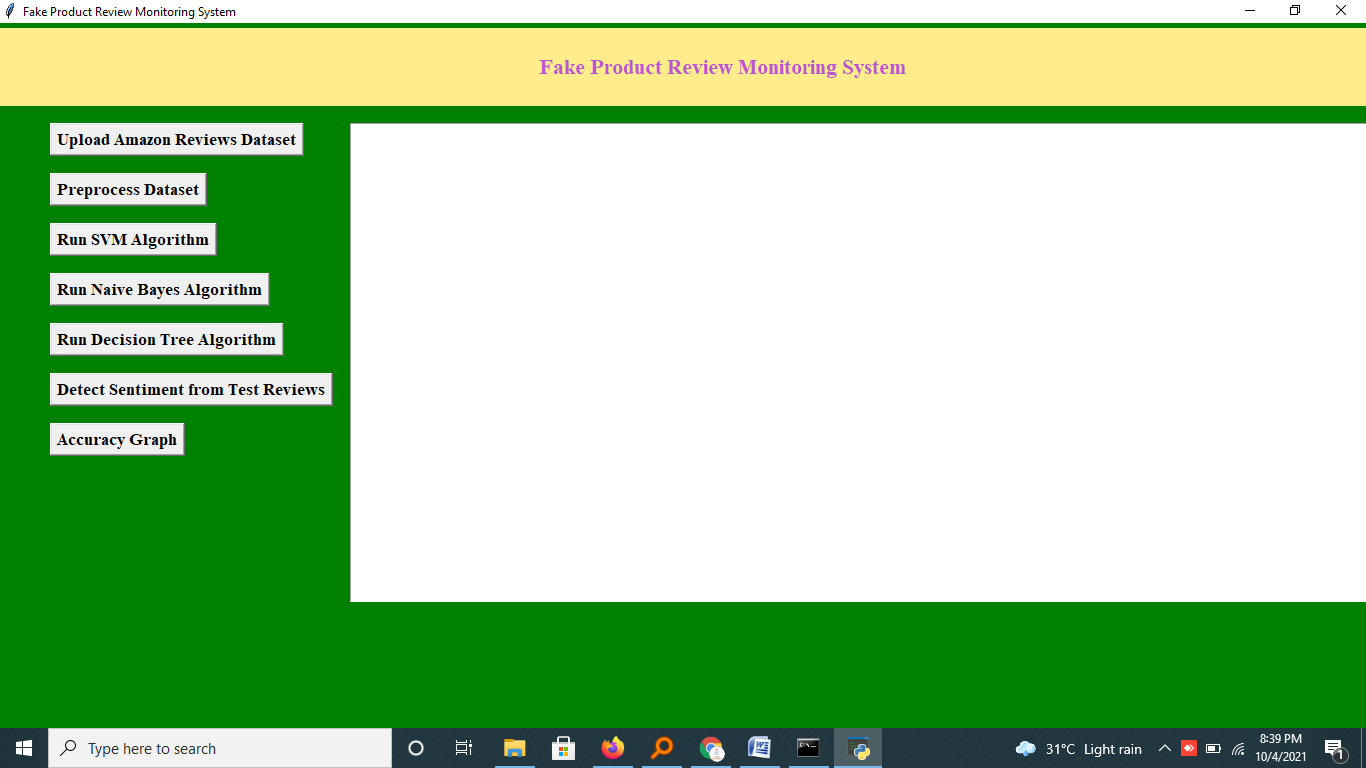
1. Data Collection: Using this module we will upload AMAZON reviews dataset to application
2. Data Preprocessing: using this module we will read all reviews and then remove stop words, special symbols, punctuation and numeric data from all reviews and after applying Preprocessing we will extract features from all reviews.
3. Features Extraction: here we will apply TF-IDF (term frequency Inverse Document Frequency) algorithm to convert string reviews into numeric vector. Each word count will be put in vector in place of words.
4. Run SVM Algorithm: We will apply SVM algorithm on TF-IDF vector to train SVM algorithm and then we apply test data on SVM trained model to calculate SVM prediction accuracy
5. Run Naïve Bayes Algorithm: We will apply Naïve Bayes algorithm on TF-IDF vector to train Naïve Bayes algorithm and then we apply test data on Naïve Bayes trained model to calculate Naïve Bayes prediction accuracy
6. Run Decision Tree Algorithm: We will apply Decision Tree algorithm on TF-IDF vector to train Decision Tree algorithm and then we apply test data on Decision Tree trained model to calculate Decision Tree prediction accuracy
7. Detect Sentiment from Test Reviews: Using this module we will upload test reviews and then ML algorithm will predict sentiment for each review and in below test reviews dataset we can see there is no sentiment value and ML will predict sentiment for each test value



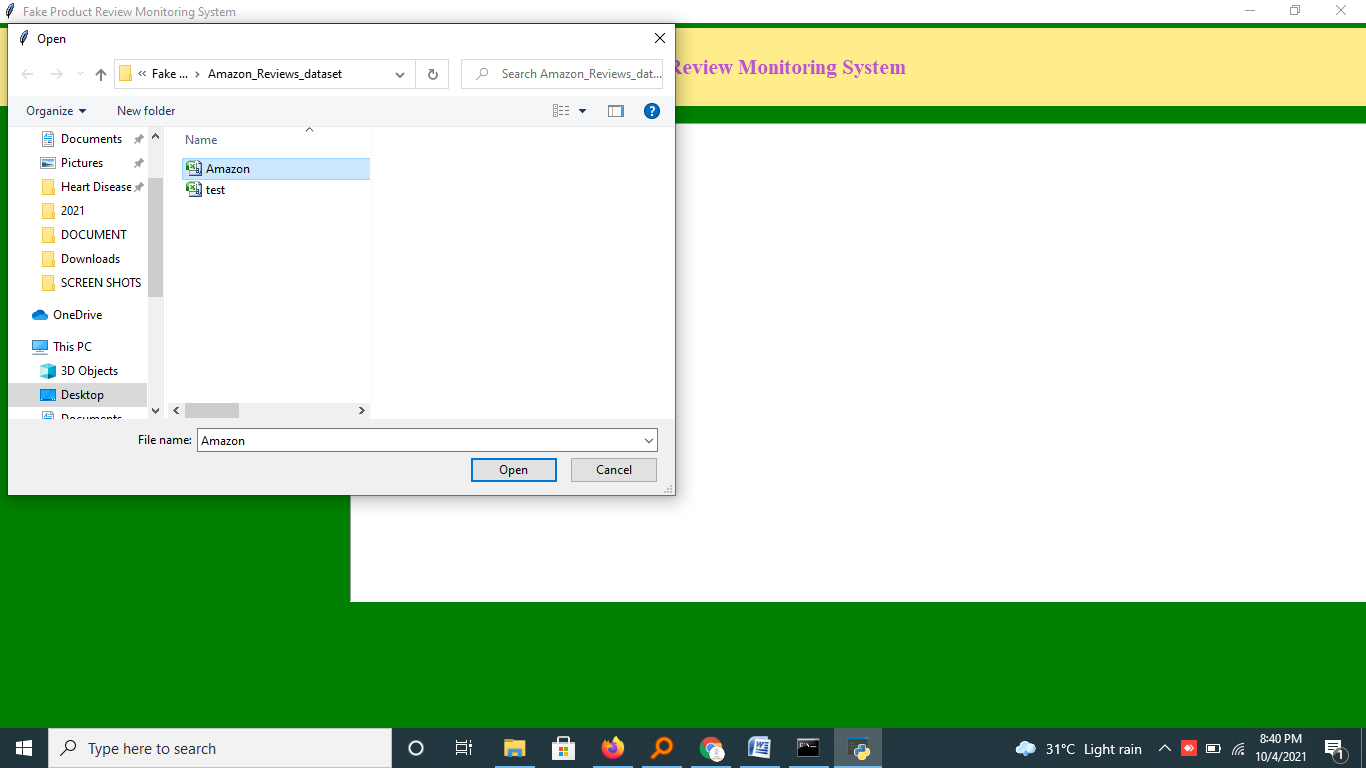
In above test data we have only test reviews and by applying ML trained model on above test data we can predict sentiment label.

SCREEN SHOTS

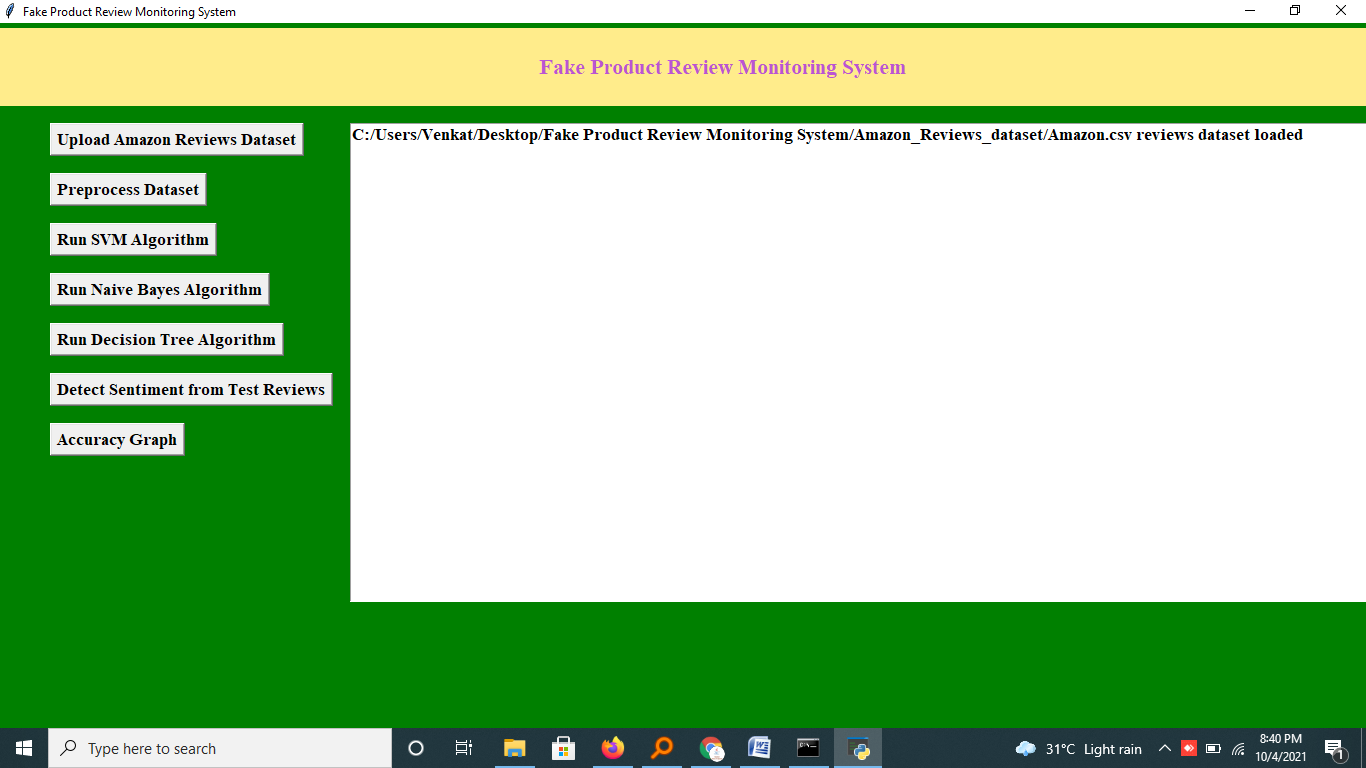
To run project double click on ‘run.bat’ file to get below screen



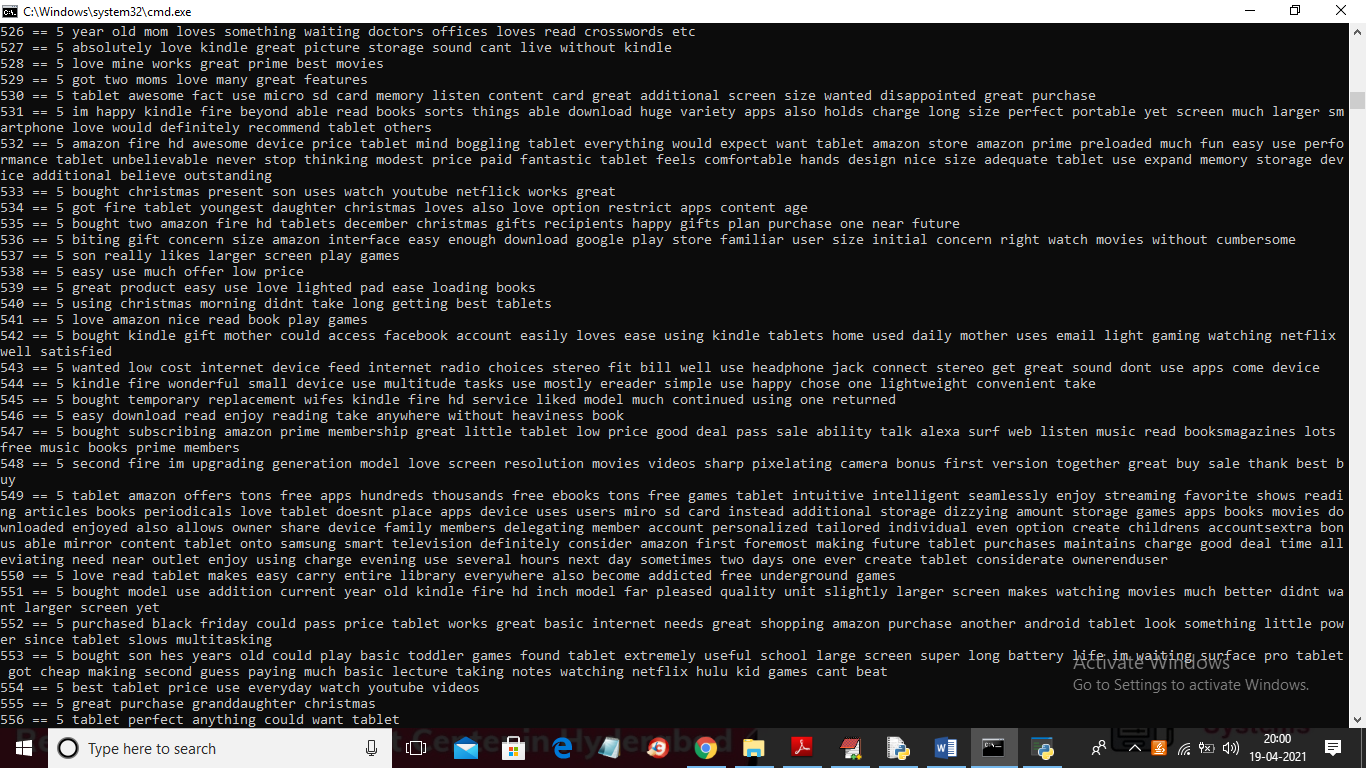
In above screen click on ‘Upload Amazon Reviews Dataset’ button to upload dataset



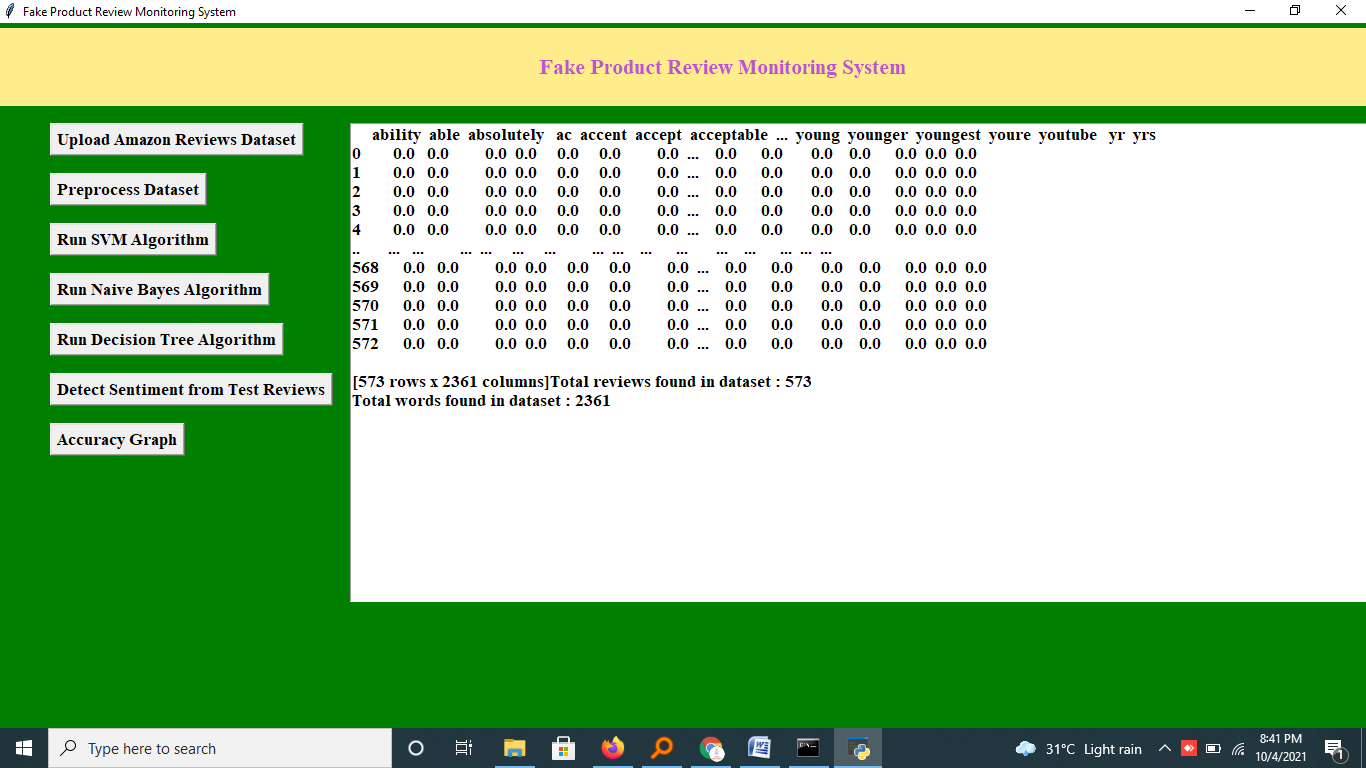
In above screen we are selecting and uploading ‘Amazon.csv’ file and then click on ‘Open’ button to load dataset and to get below screen



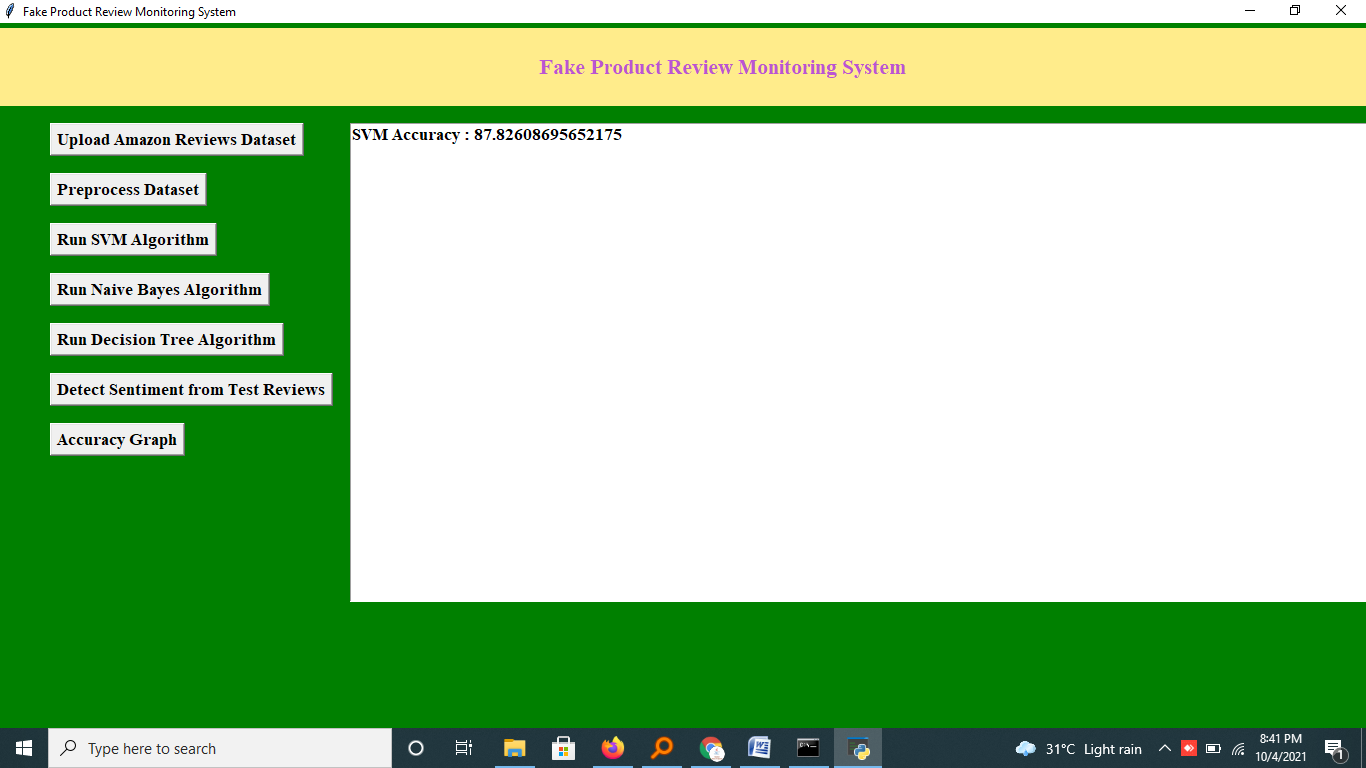
In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to read all reviews from dataset and then apply Preprocess steps to get below screen



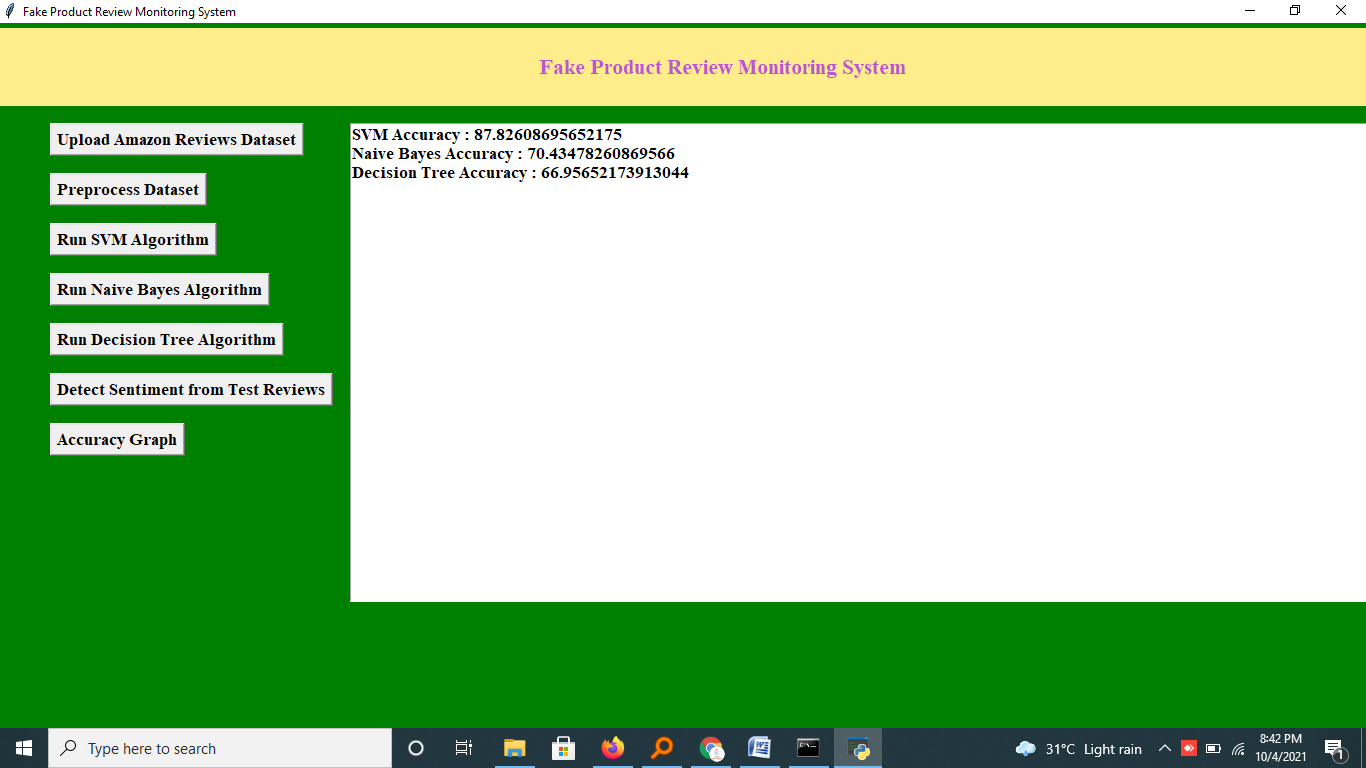
In above black console we can see application read all reviews from dataset and then generate below TF-IDF vector



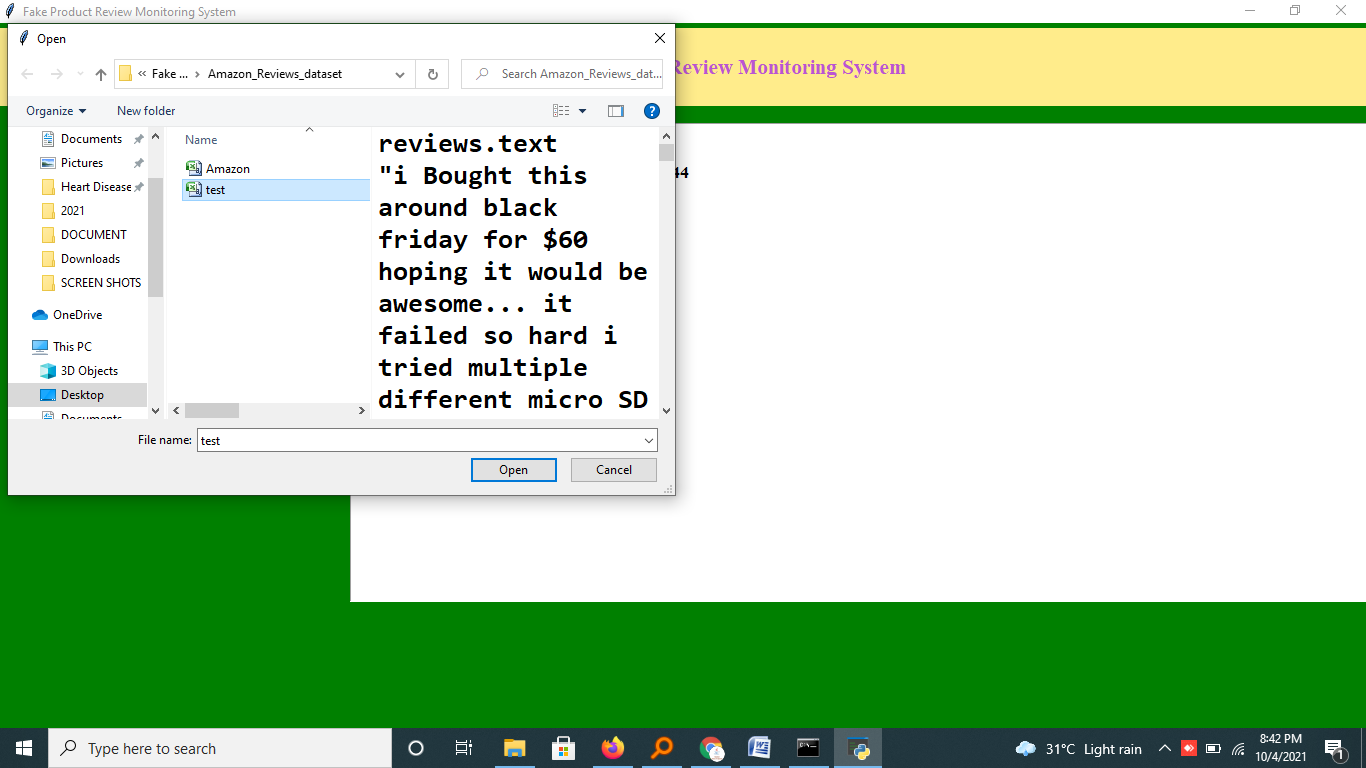
In above screen in text area we can see application extract all words from reviews and then put in top line of above test area and in remaining rows if that word appear then it put average count value of that word and if word not appear then 0 will put. In above screen vector generated and I am showing few records from that vector. In that vector total reviews are 573 and all reviews contains total 2361 unique words. Now vector is ready and now click on ‘Run SVM Algorithm’ button to train SVM with above vector



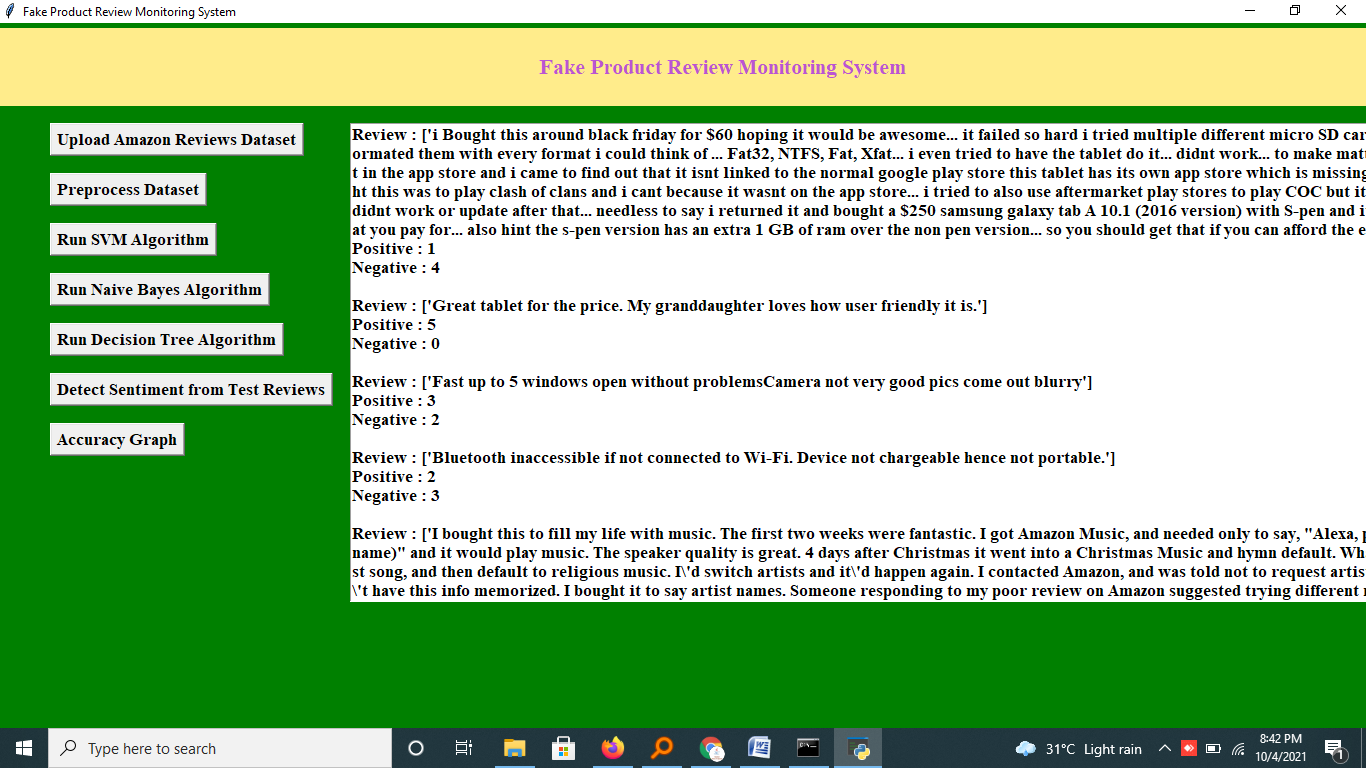
In above screen with SVM we got 82% accuracy and now click on Naïve Bayes and Decision tree button to get their accuracy



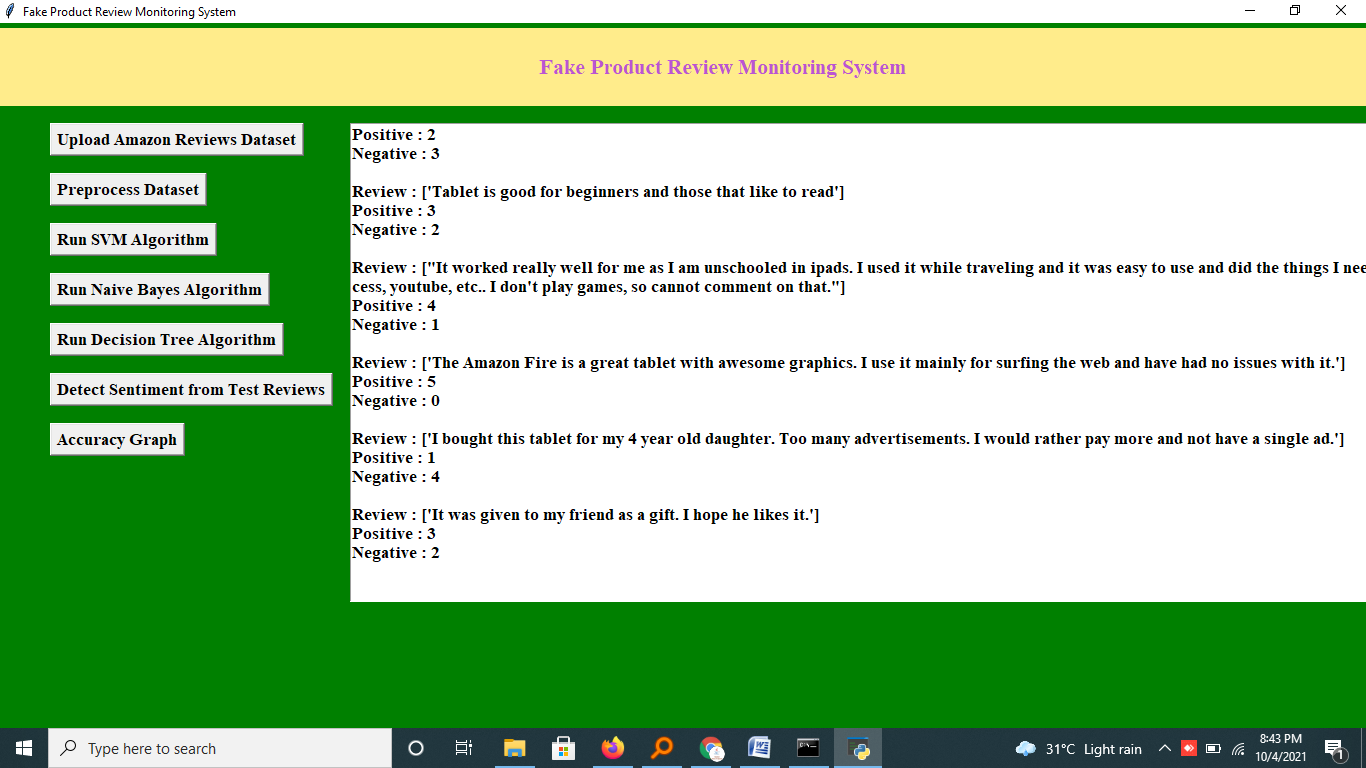
In above screen with all 3 algorithms SVM gave better prediction accuracy and now click on ‘Detect Sentiment from Test Reviews’ button to upload test reviews



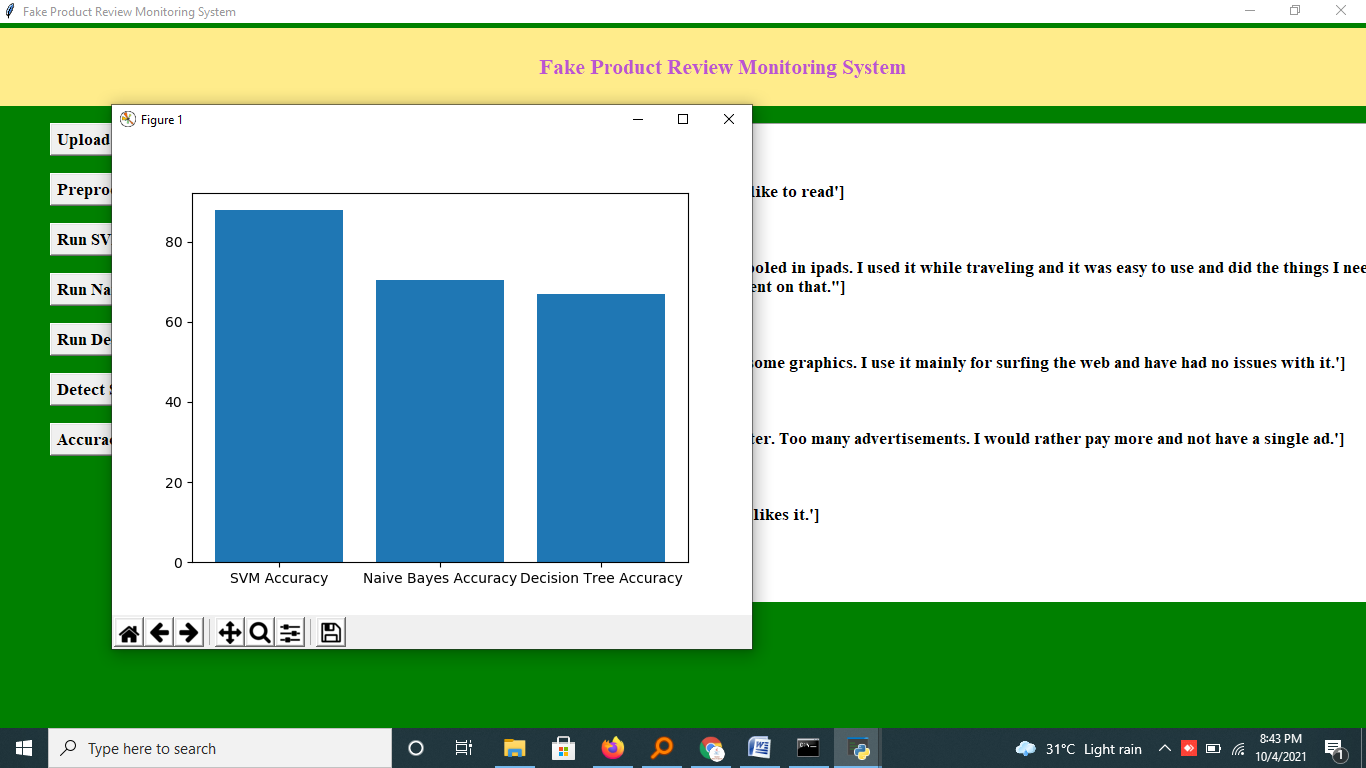
In above screen selecting and uploading ‘test.csv’ file and then click on ‘Open’ button to get below prediction result



In above screen first I am displaying reviews from uploaded test file and then predicting positive and negative sentiment for each review and you can scroll down above text area to get all outputs



In above screen we can see sentiment prediction result for all reviews and now click on ‘Accuracy Graph’ button to get below graph



In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and in all 3 algorithms SVM got higher accuracy