Tourist Place Reviews Sentiment Classification Using Machine Learning Techniques

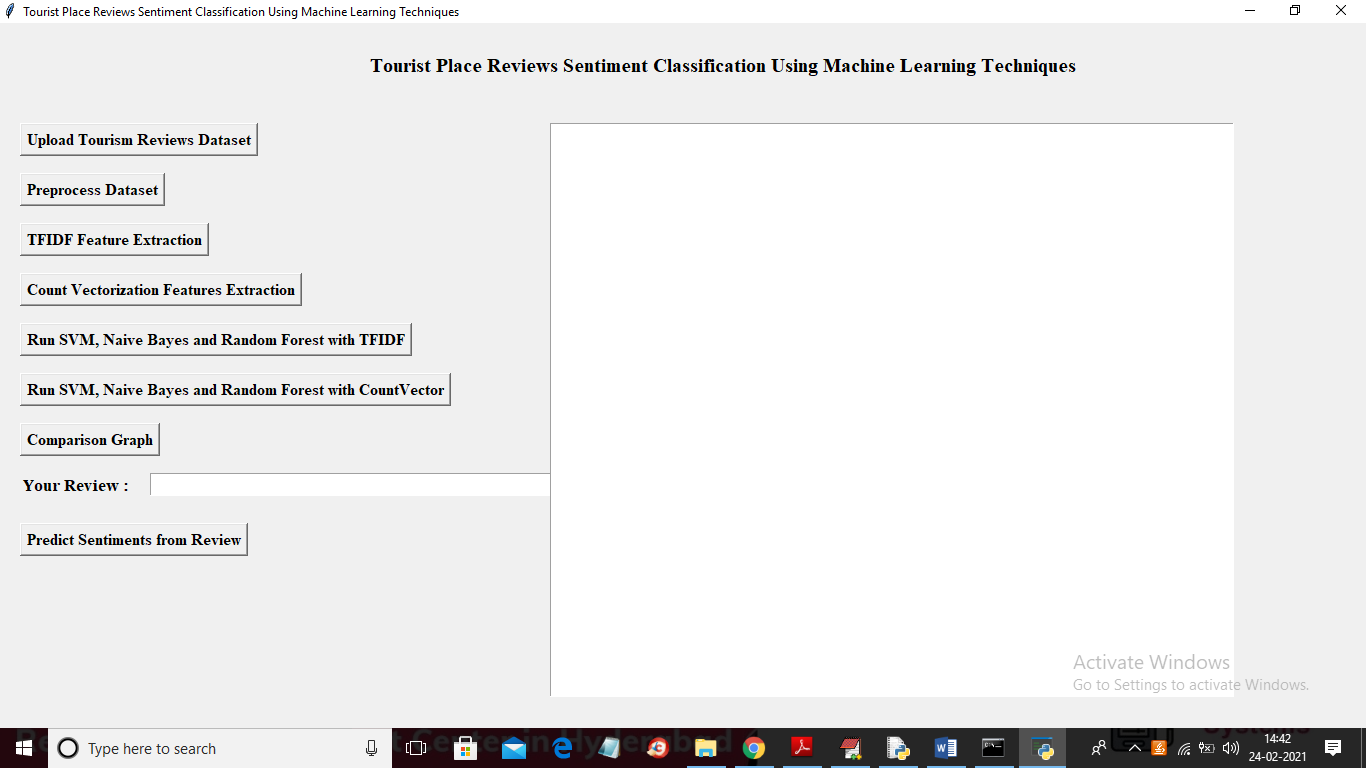
In this paper author using machine learning algorithms such as SVM, Naïve Bayes and Random Forest to predict sentiments from tourist reviews dataset and then evaluating performance of CountVectorizer and TFIDFVectorizer features extraction algorithms. In this paper author is extracting features from reviews by using both CountVectorizer and TFIDFVectorizer and then applying this features on machine learning algorithms and then calculating accuracy, precision, recall and F!SCORE between both feature extraction algorithms.

To implement this project author has designed following modules

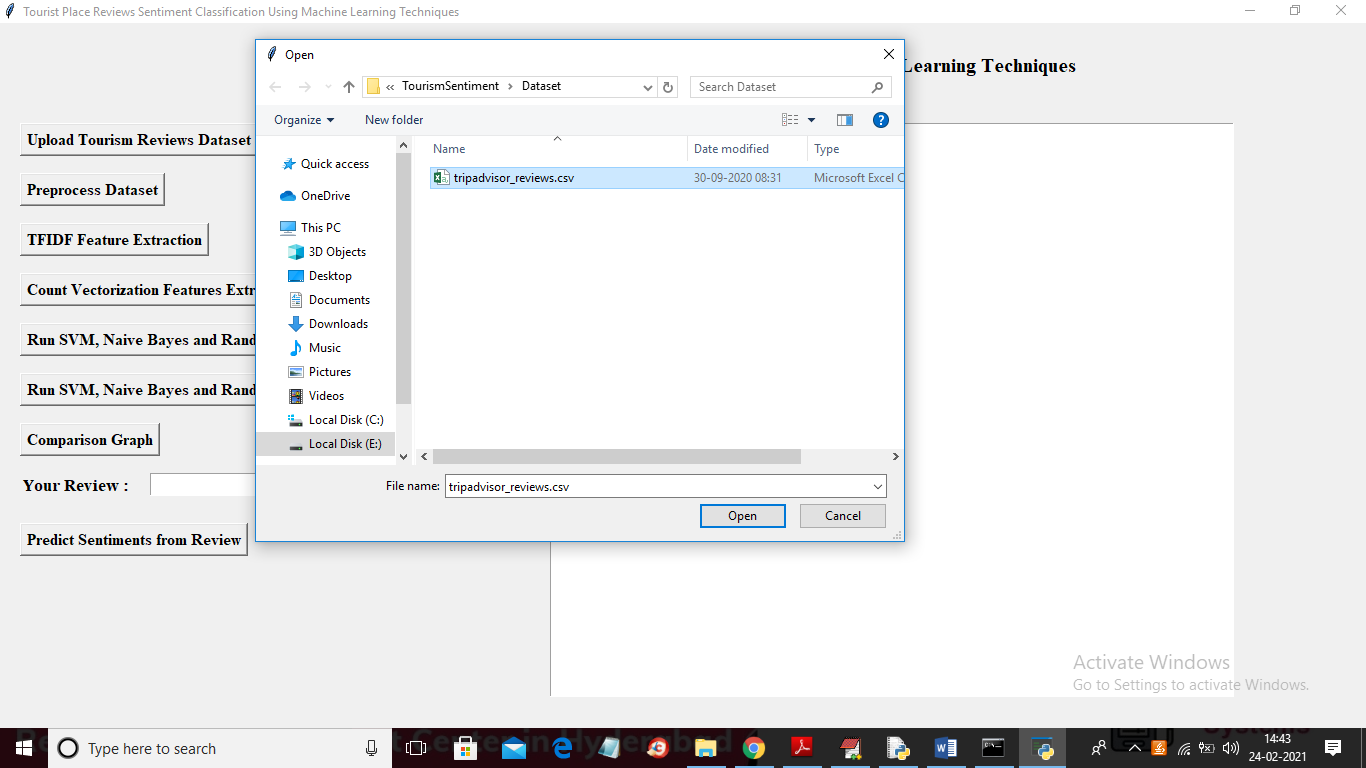
1. Upload Dataset: using this module we will upload reviews dataset to application
2. Data Preprocessing: using this module we will read all reviews and then remove stop words, special symbols and apply NGRAM techniques on review clean text.
3. CountVectorization: using this module clean text will be converted to count vector where each word count will be calculated and then features vector will be generated.
4. TFIDFVectorization: Using this module TF (Term Frequency) and IDF (Inverse Document Frequency) will be calculated and then generate features vector.
5. RUN SVM, Naïve Bayes and Random Forest with TFIDF: Using this module we will train all 3 algorithms with TFIDF features and then calculate execution time, accuracy, precision, Recall and FSCORE.
6. RUN SVM, Naïve Bayes and Random Forest with Count Vector: Using this module we will train all 3 algorithms with Count Vector features and then calculate execution time, accuracy, precision, Recall and FSCORE.
7. Comparison Graph: using this we will visualize performance graph of both feature extraction algorithms with various machine learning algorithms
8. Predict Sentiments from Review: using this module user can enter his review and then application calculate sentiments from that review.

SCREEN SHOTS

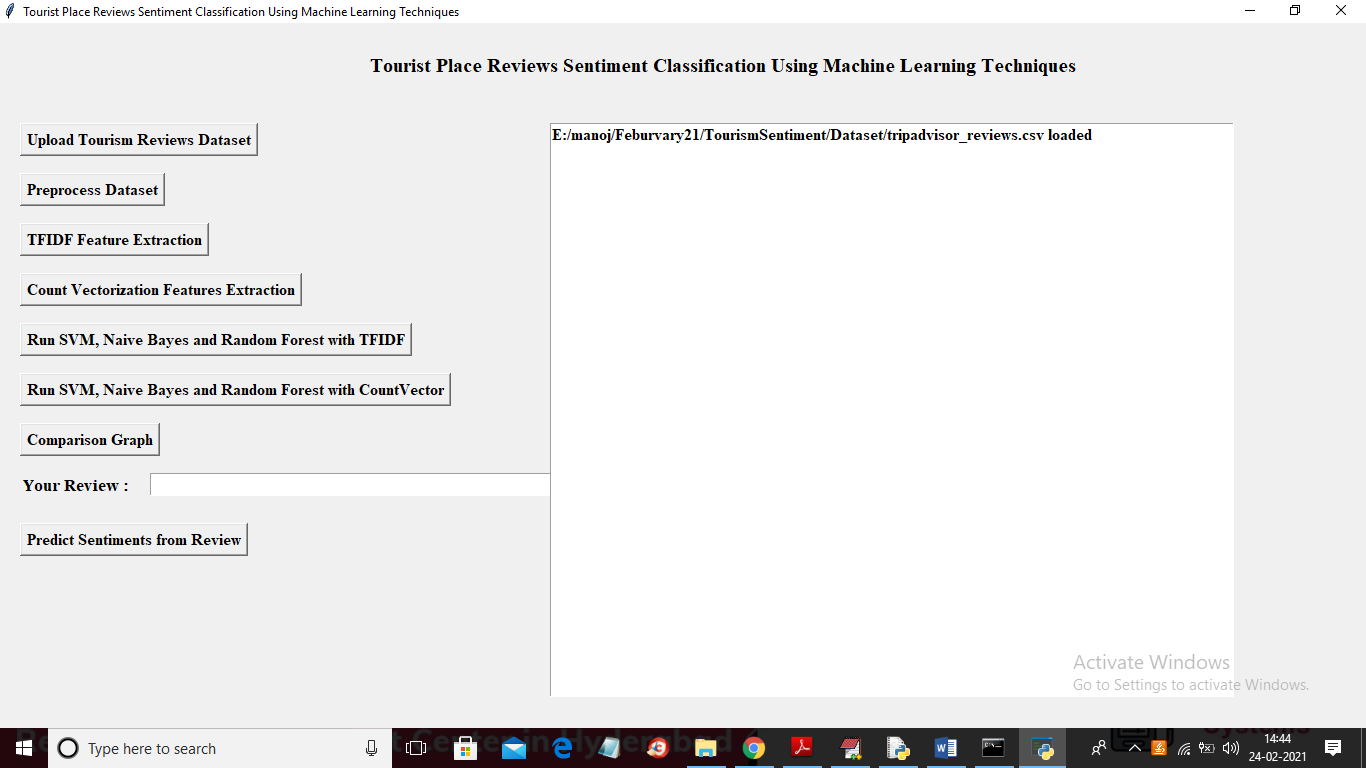
To run project double click on ‘run.bat’ file to get below screen



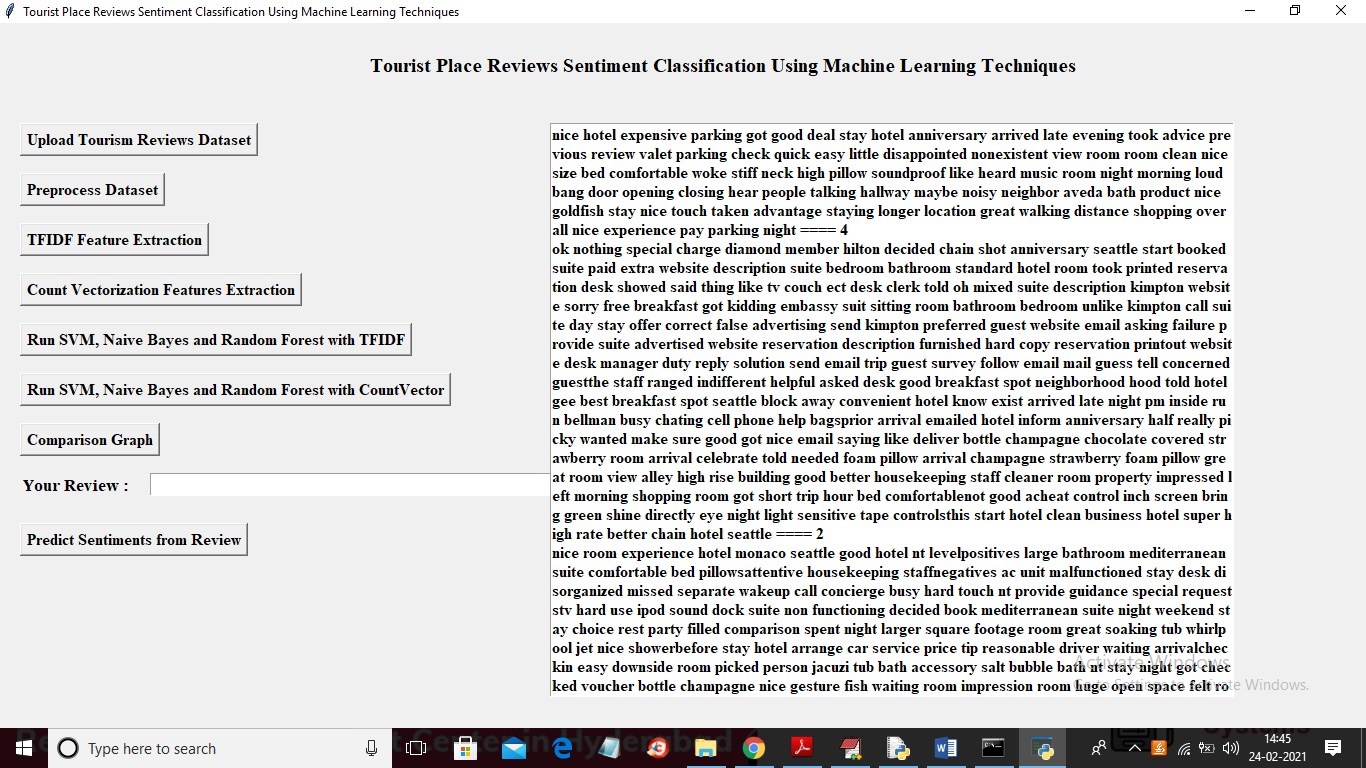
In above screen click on ‘Upload Tourism Reviews Dataset’ button to upload dataset and to get below screen



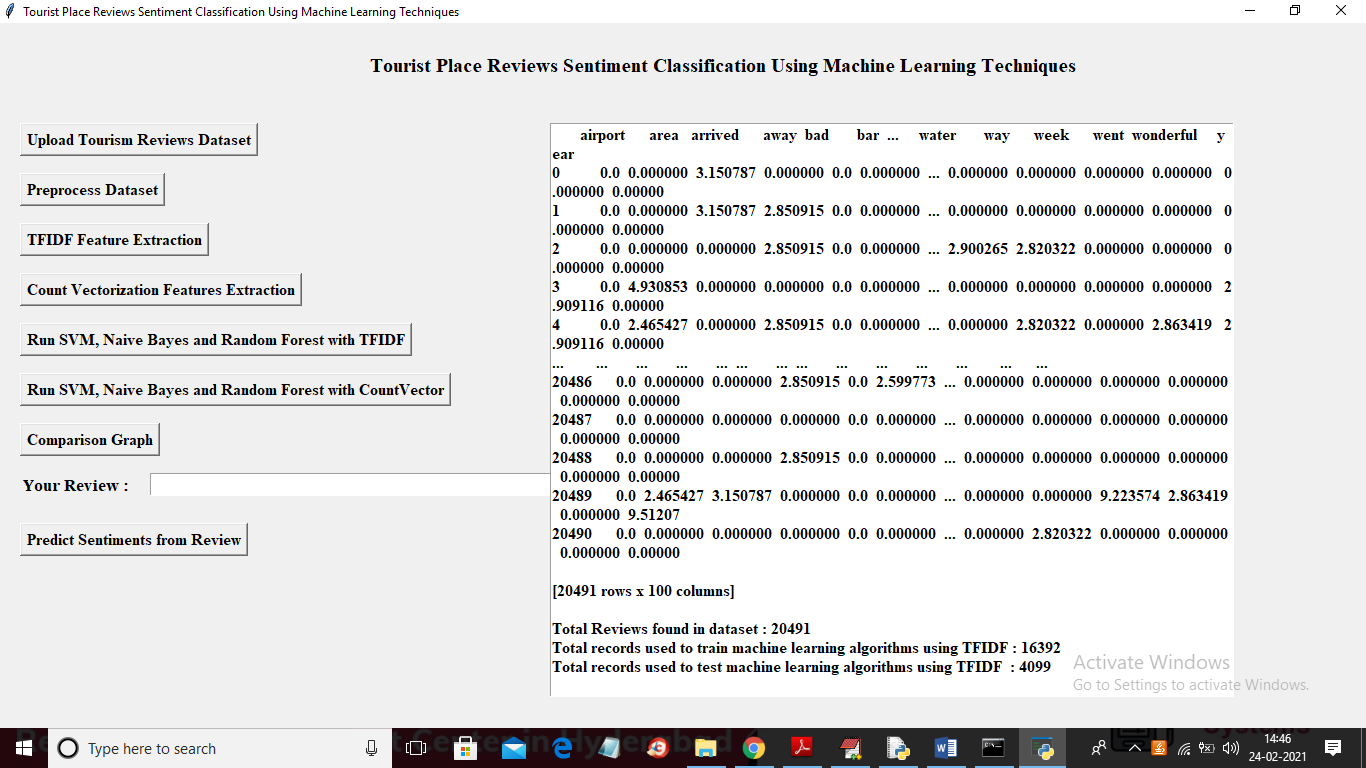
In above screen selecting and uploading ‘trioadvisor\_reviews.csv’ file and then click on ‘Open’ button to load dataset and to get below screen



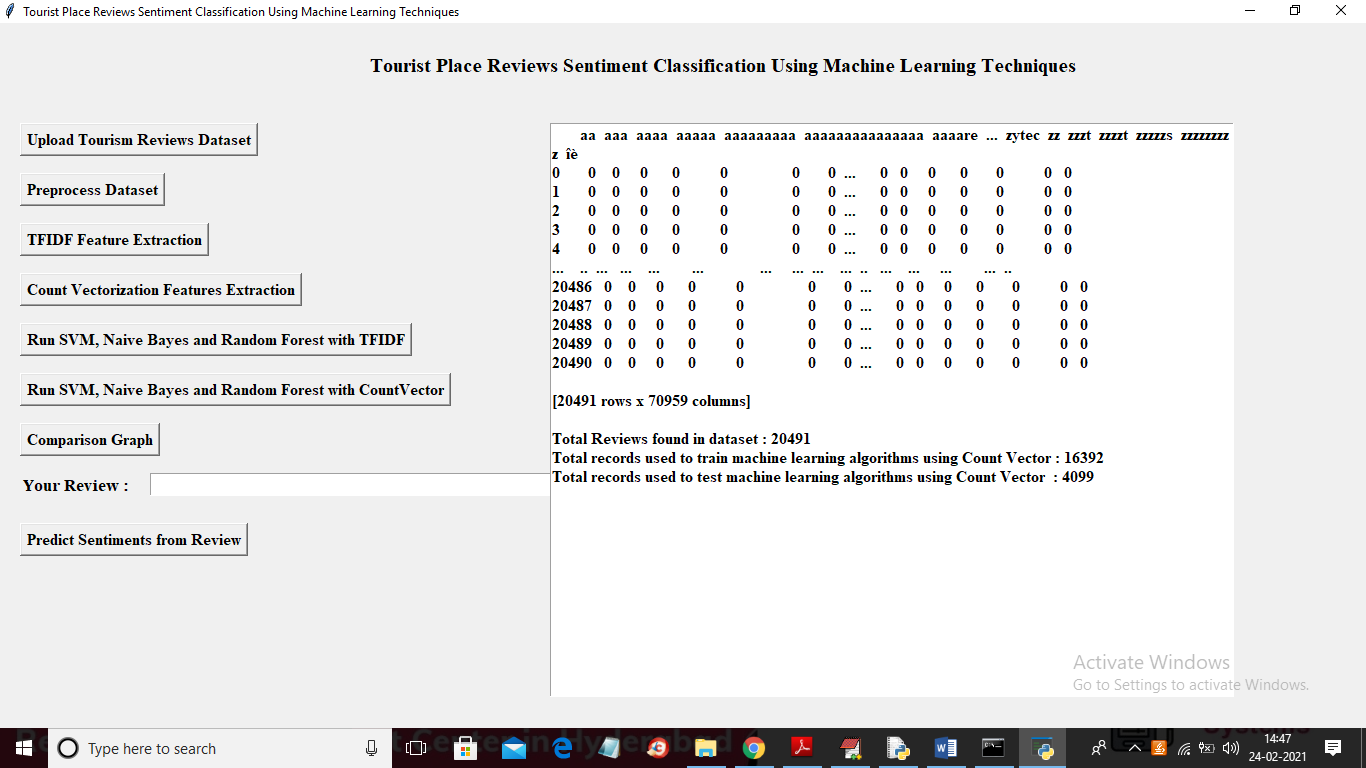
In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to read dataset and then clean it



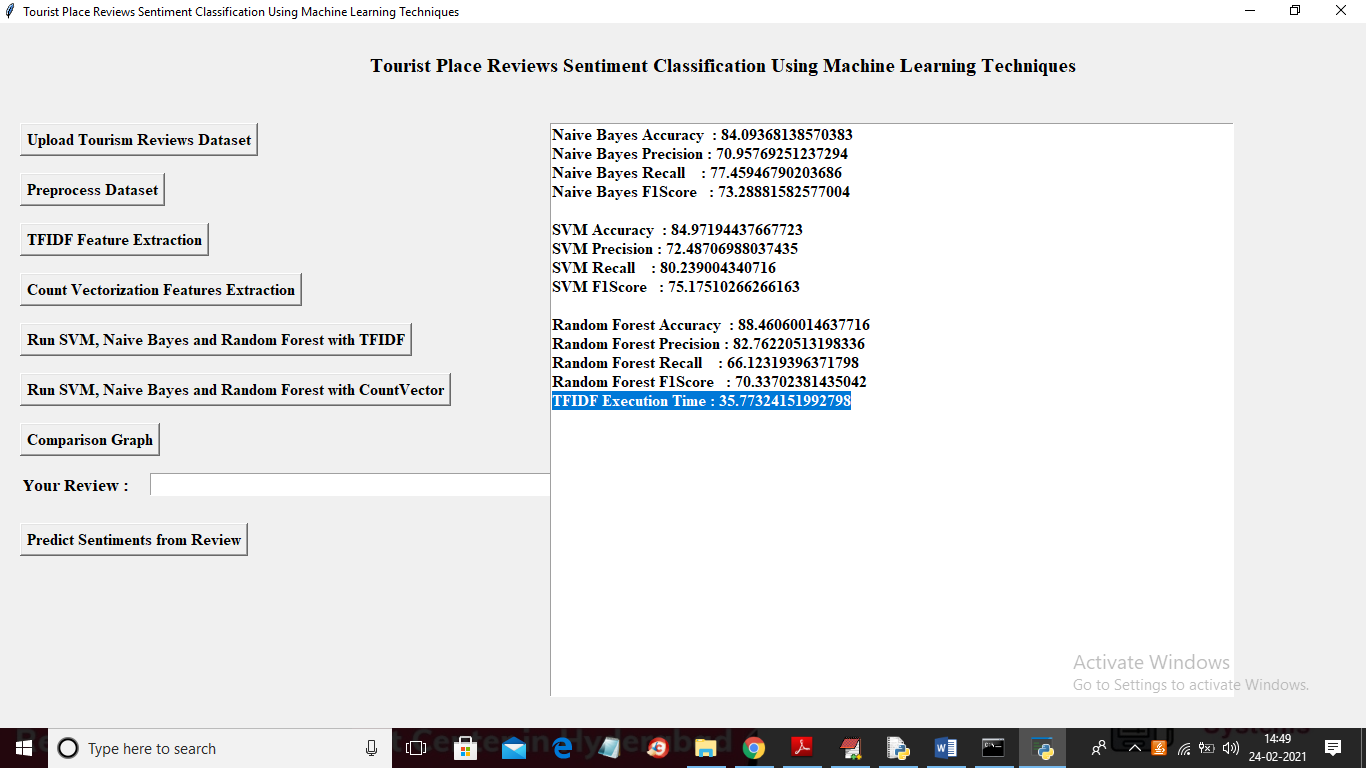
In above screen all reviews read from dataset and then displaying in text area and now click on ‘TFIDF Feature Extraction’ button to extract features using TFIDF technique



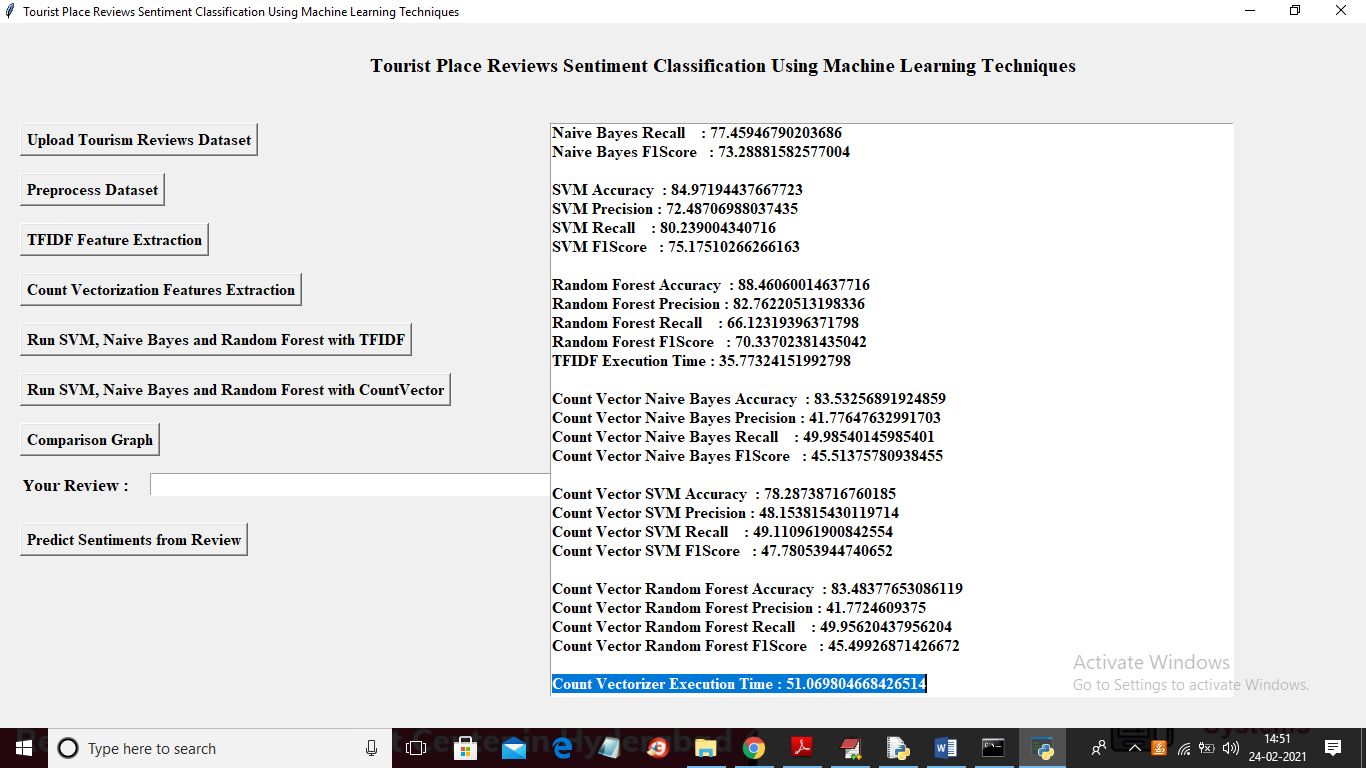
In above screen application calculated TFIDF for each word and we can see the numeric TFIDF values for each word and in below we can see total reviews found in dataset and application using 80% reviews to train machine learning algorithms and 20% reviews to test them. Now Click on ‘Count Vectorization Feature Extraction’ button to count each word



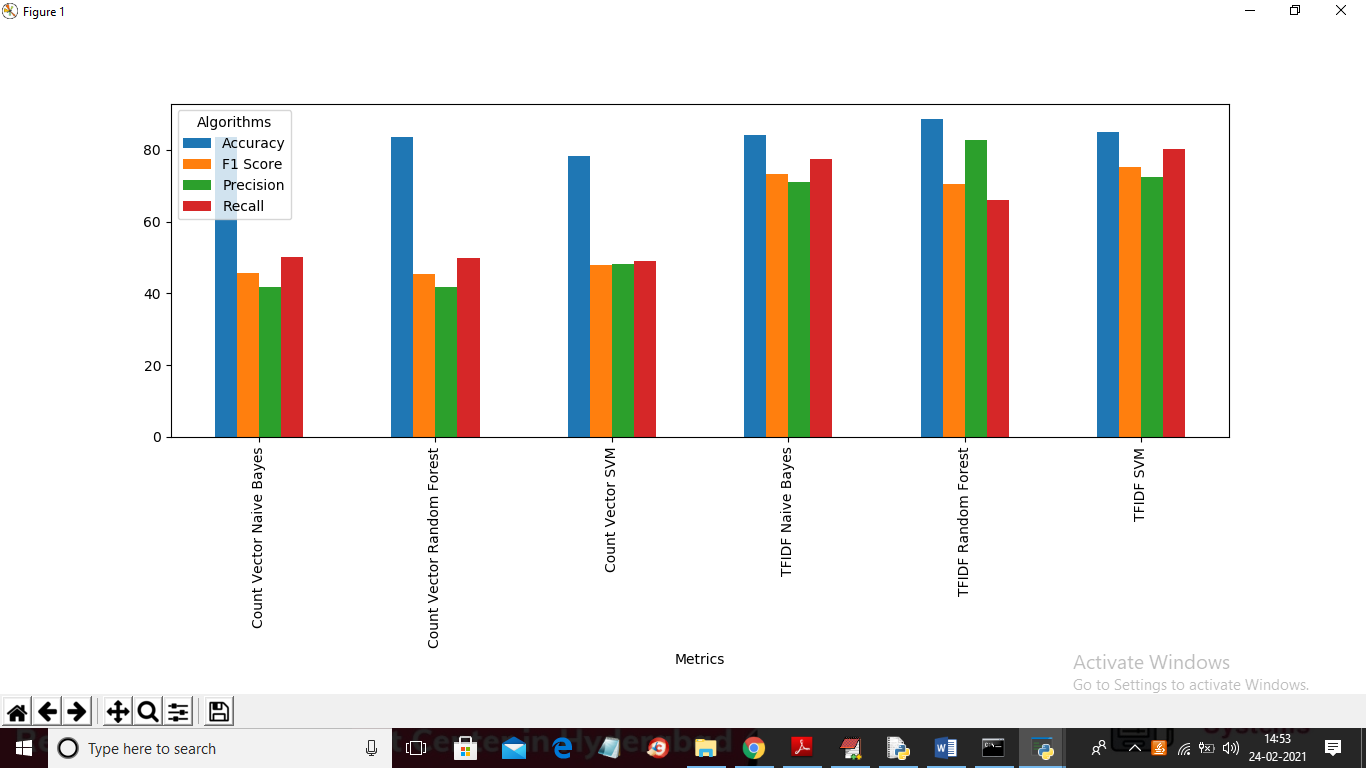
In above screen got count vectorizer and now both features TFIDF and count vector is ready and now click on ‘Run SVM, Naive Bayes and Random Forest with TFIDF’ button to train 3 machine learning algorithms with TFIDF features and to get accuracy and other details



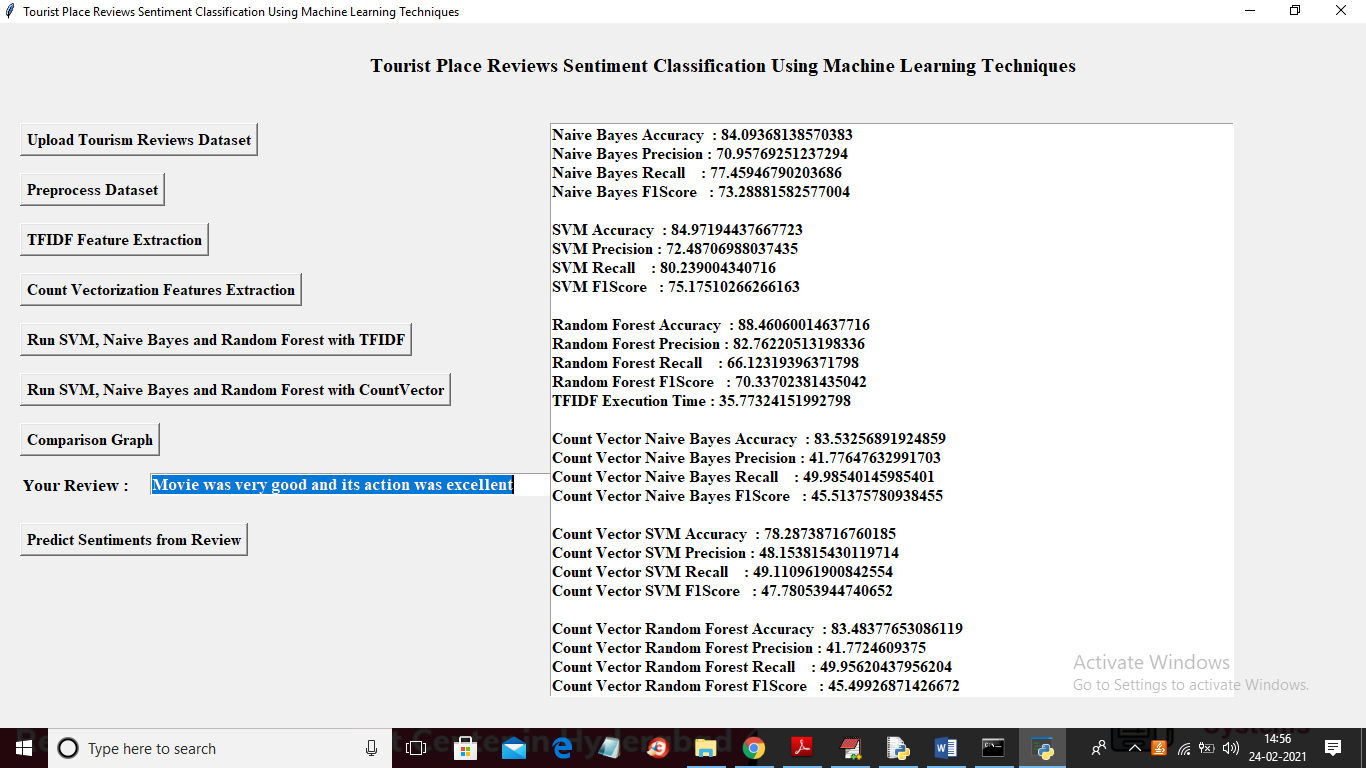
In above screen for all 3 algorithms we got accuracy, precision, recall and FSCORE with TFIDF features and in above screen in selected text TFIDF took 35.77 milli seconds for execution and now click on ‘Run SVM, Naive Bayes and Random Forest with CountVector’ button to get accuracy details with count vector



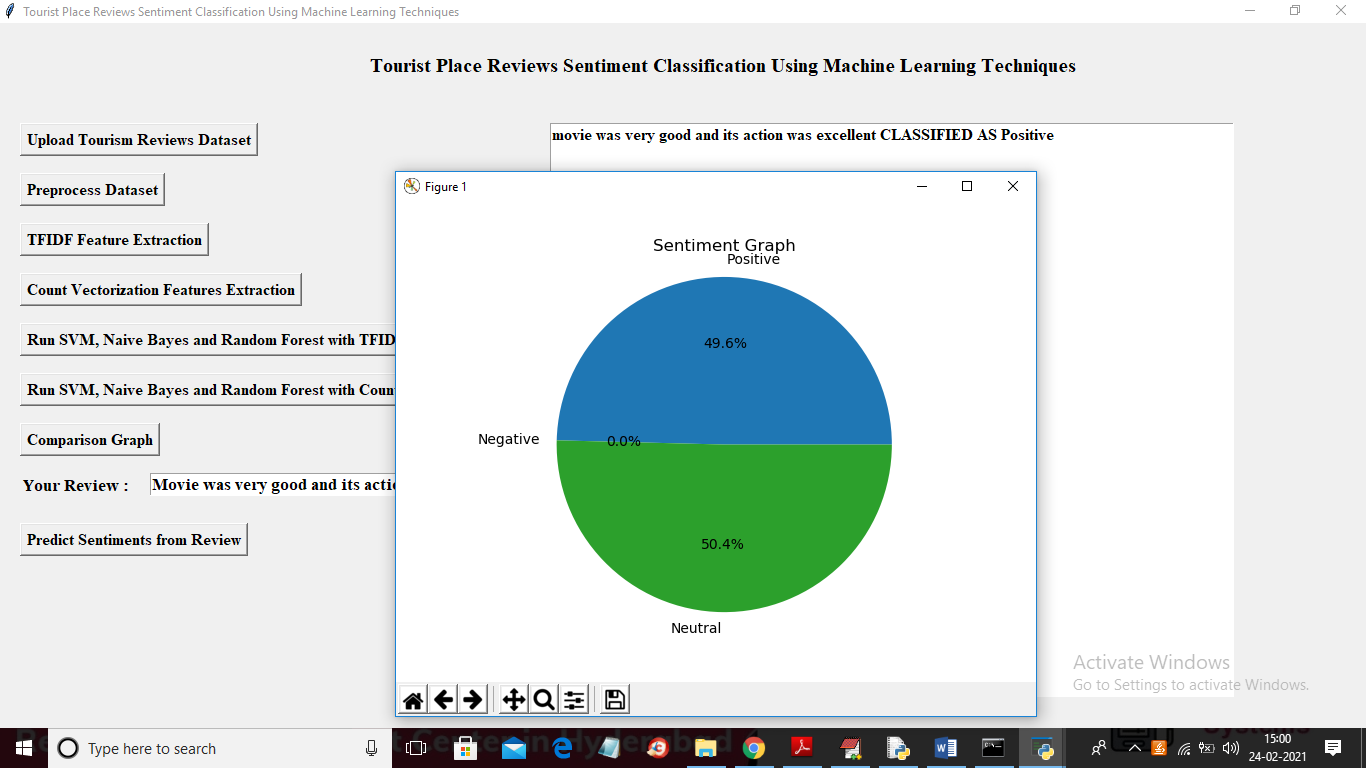
In above screen we can see count vector took 51 milli seconds and its accuracy, precision, recall and FSCORE is also less compare to TFIDF. Now click on ‘Comparison Graph’ button to get below graph



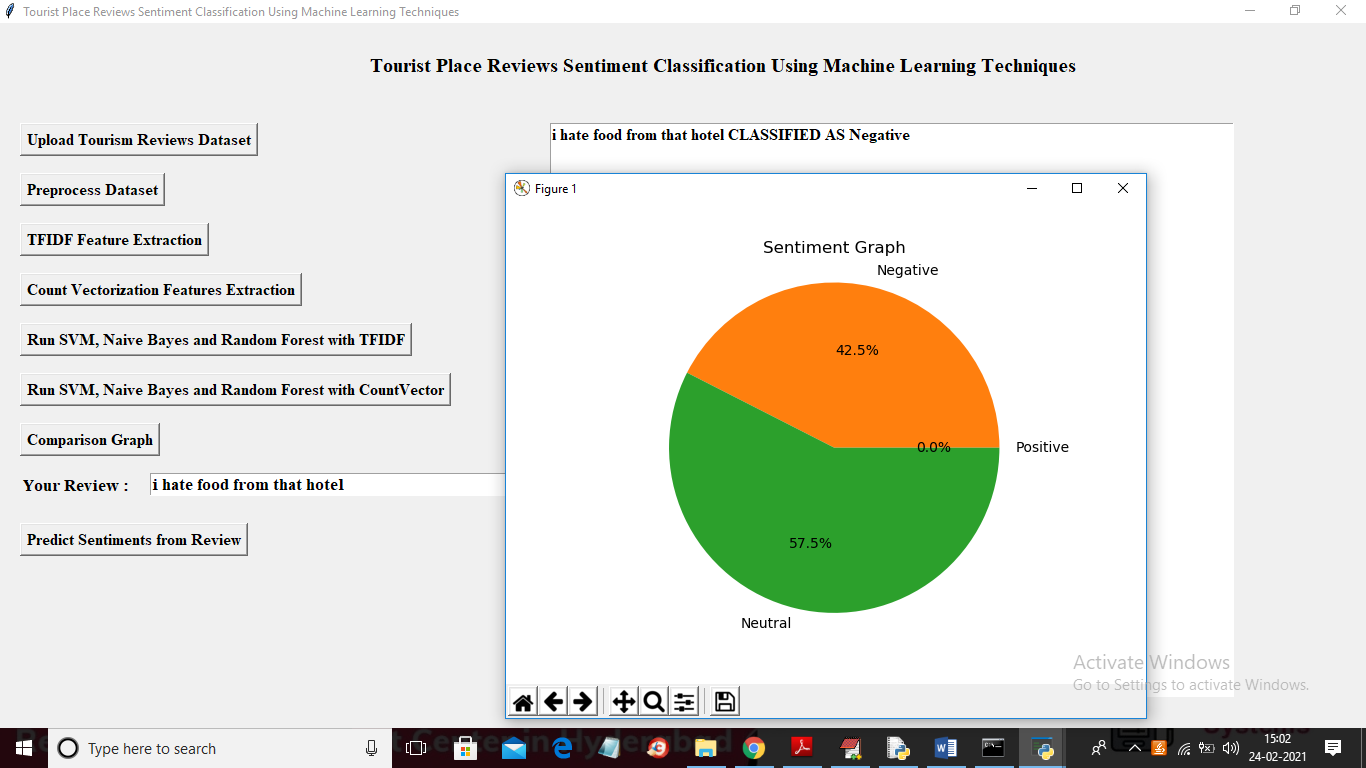
In above graph x-axis represents algorithm name and y-axis represents accuracy, precision, recall and fscore and in above graph blue line is for accuracy, yellow for FSCORE, green for precision and red for recall. Between count and TFIDF we can see TFIDF got better prediction result. Now enter any review in text field and click on ‘Predict Sentiments from Review’ button to get below result



In above screen in text field I entered some review and then click last button to get below result



In above screen in text area we can see sentiment result and in graph we can see positive, negative and neutral value



Similarly u can enter any review and get result