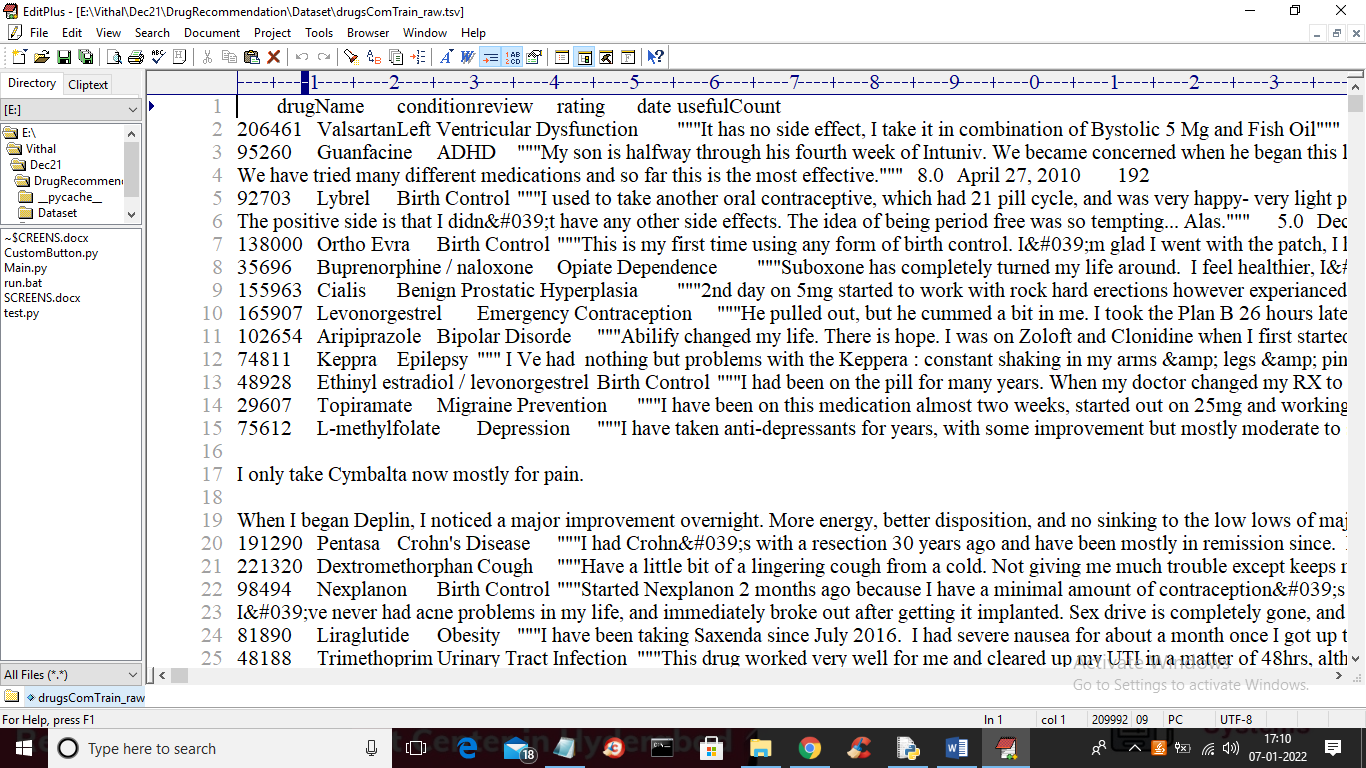
Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning

Now-a-days new diseases are attacking human world and corona virus is such disease and this diseases require lots of medical systems and medical human experts and due to growing disease medical experts and systems are not sufficient and patients will take medicines on their risk which can cause serious death or serious damage to patient body.

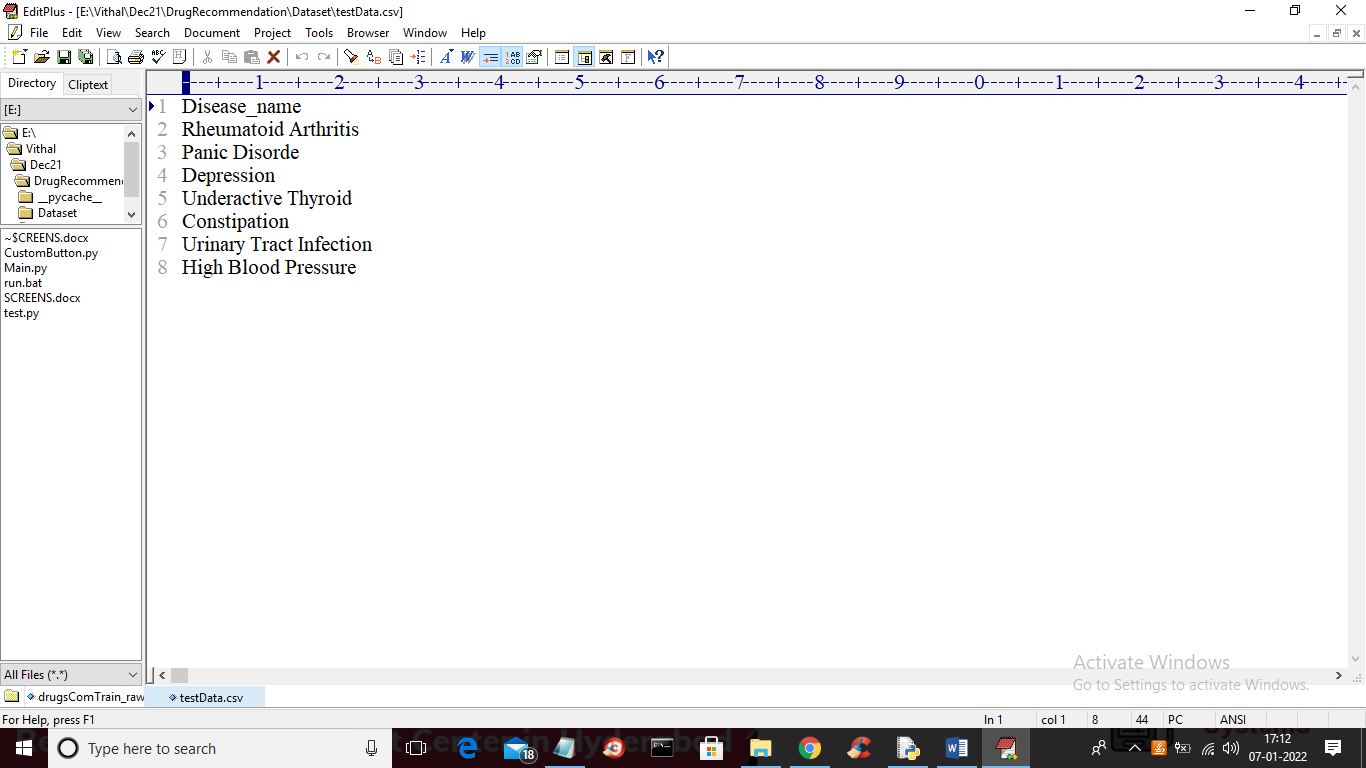
To overcome from above problem author of this paper introducing sentiment and machine learning based drug recommendation system which will accept disease names from patient and then recommend DRUG and simultaneously display SENTIMENT rating based on reviews given by old users based on their experience. If predicted rating is high then patient can trust and took recommended drug.

In propose paper author has used various features extraction algorithms such as TF-IDF (term frequency – inverse document frequency), BAG of WORDS and WORVEC and this extracted features will be applied on various machine learning algorithm such as Logistic Regression, Linear SVC, Ridge classifier, Naïve Bayes, Multilayer Perceptron classifier, SGD classifier and many more. Among all algorithms TF-IDF is giving better performance so we are using TF-IDF features extraction algorithm with above mention algorithm.

To implement this project author has used DRUGREVIEW dataset from UCI machine learning website and below is the dataset screen shots



In above screen first row represents dataset column names such as drug name, condition, review and rating and remaining rows contains dataset values and we will used above REVIEWS and RATINGS to trained machine learning models. Below is the test data which contains only disease name and machine learning will predict Drug name and ratings.

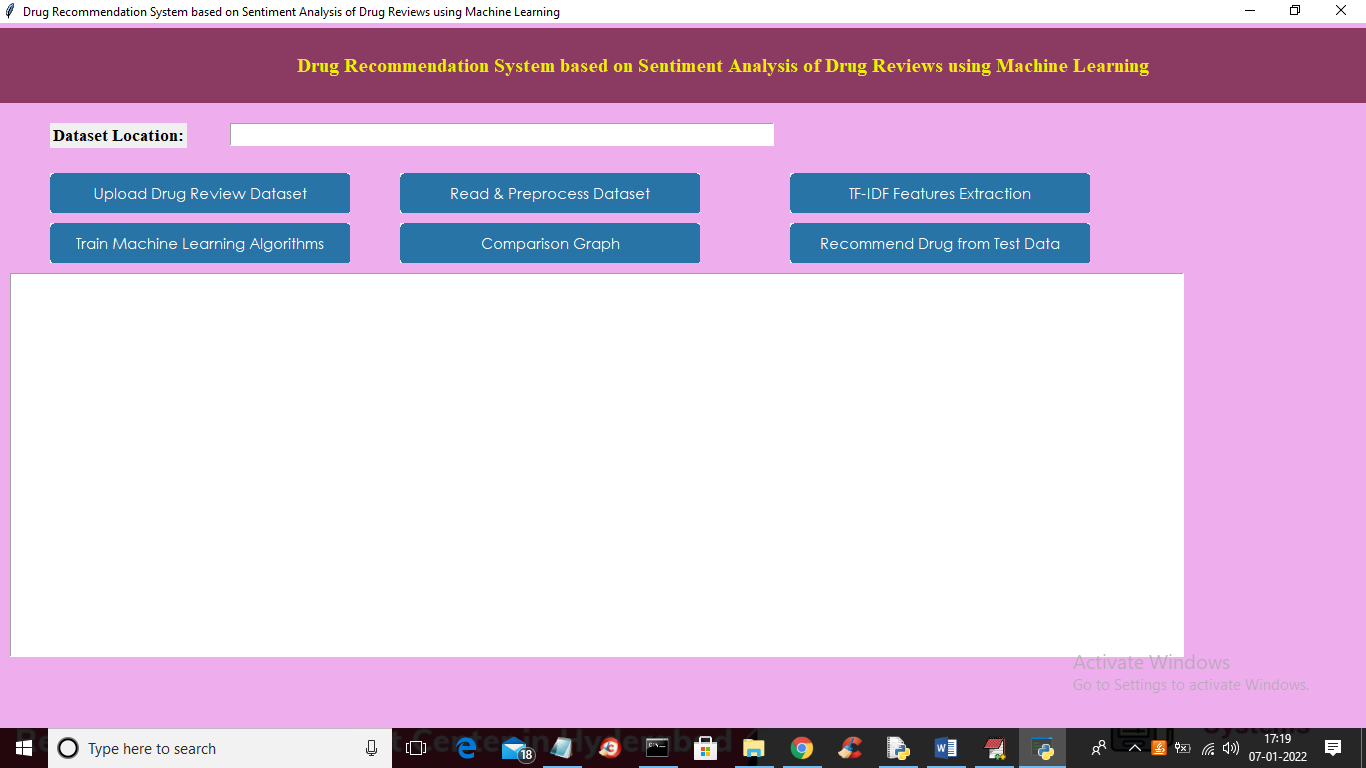


In above test data we have only disease name and machine learning will predict ratings and drug names. To implement this project we have designed following modules

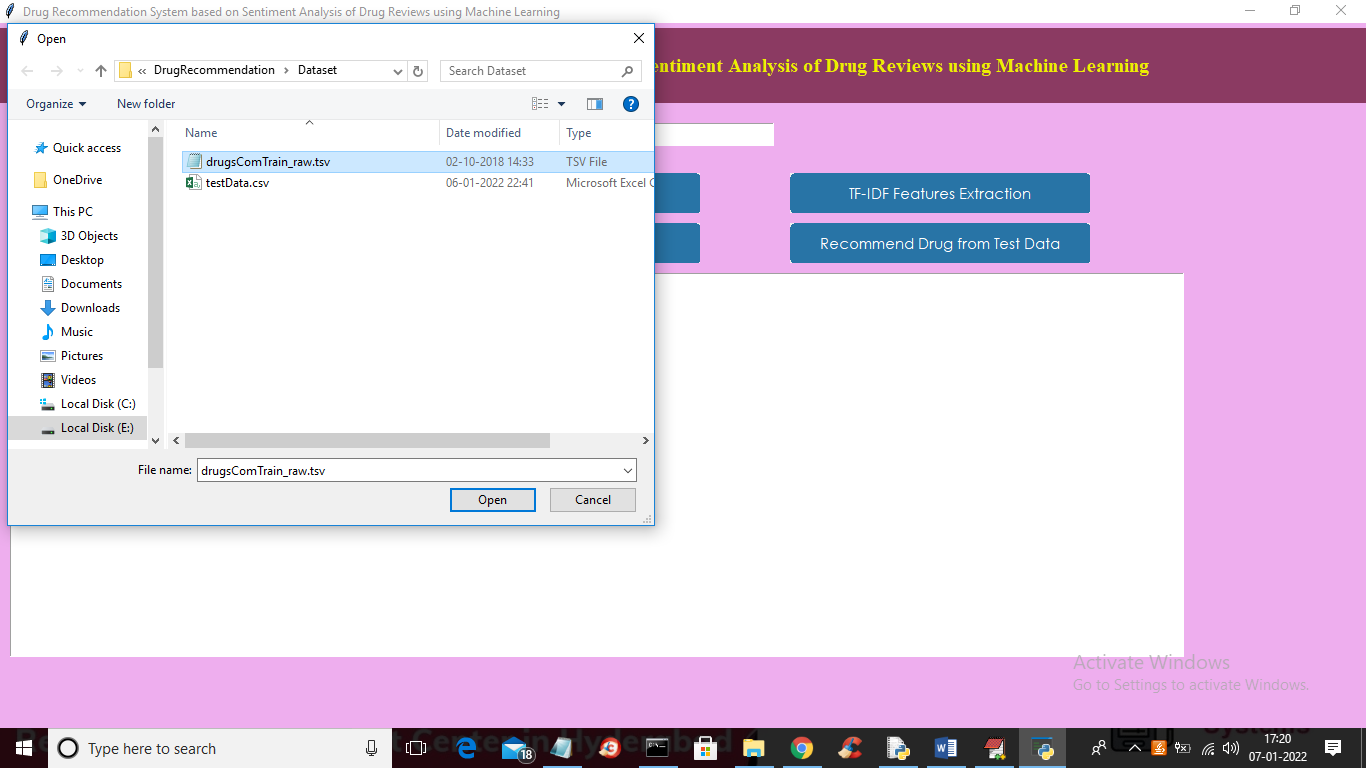
1. Upload Drug Review Dataset: using this module we will upload dataset to application
2. Read & Preprocess Dataset: using this module we will read all reviews, drug name and ratings from dataset and form a features array.
3. TF-IDF Features Extraction: features array will be input to TF-IDF algorithm which will find average frequency of each word and then replace that word with frequency value and form a vector. If word not appear in sentence then 0 will be put. All reviews will be consider as input features to machine learning algorithm and RATINGS and Drug Name will be consider as class label.
4. Train Machine Learning Algorithms: using this module we will input TF-IDF features to all machine learning algorithms and then trained a model and this model will be applied on test data to calculate prediction accuracy of the algorithm.
5. Comparison Graph: using this module we will plot accuracy graph of each algorithm
6. Recommend Drug from Test Data: using this module we will upload disease name test data and then ML will predict drug name and ratings.

SCREENSHOTS

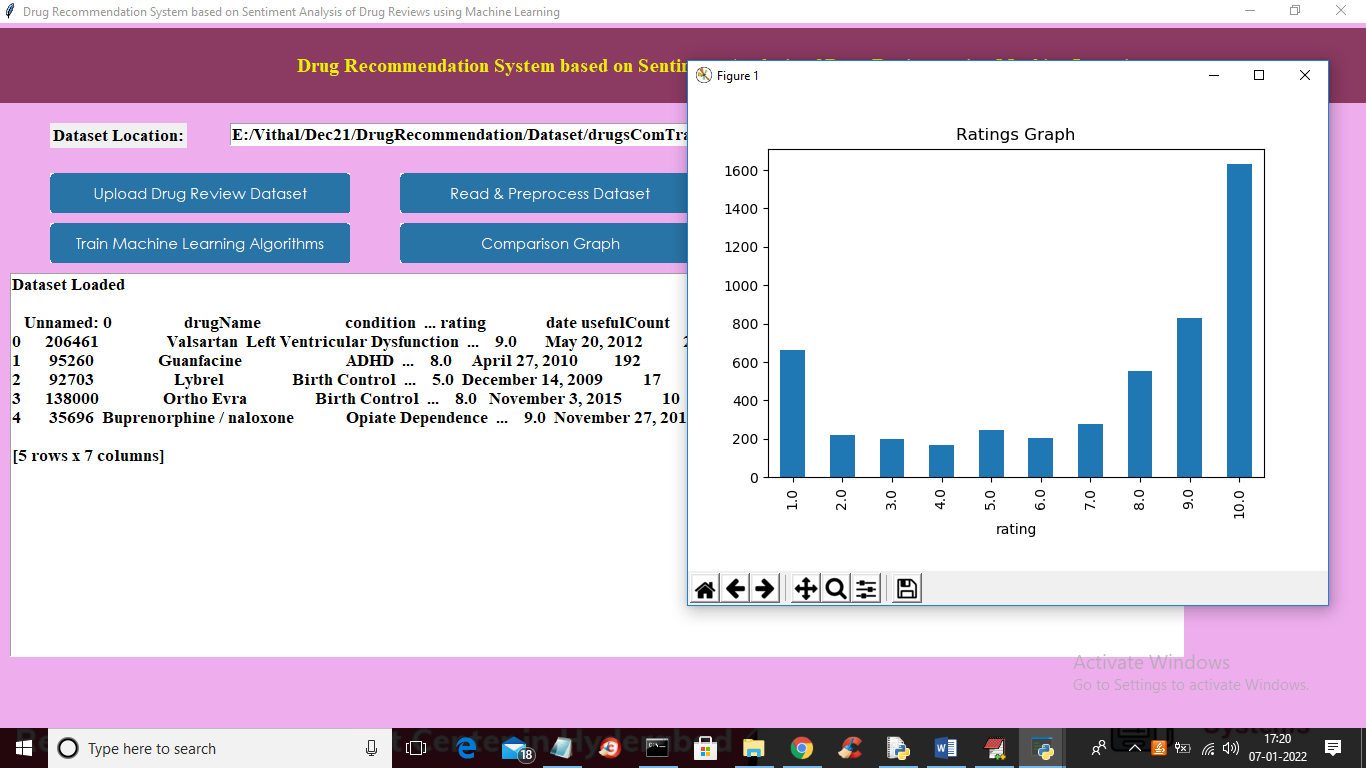
To run project double click on ‘run.bat’ file to get below screen



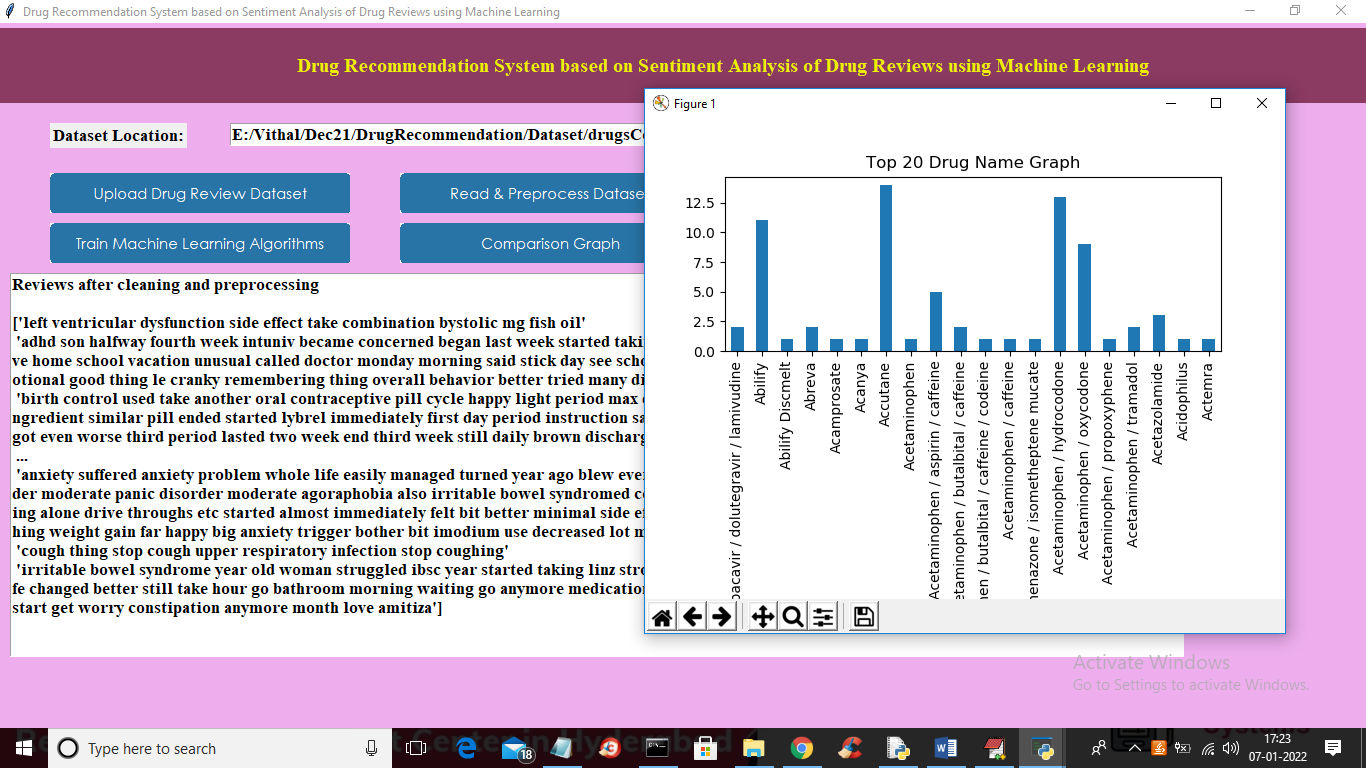
In above screen click on ‘Upload Drug Review Dataset’ button to upload dataset to application and to get below screen



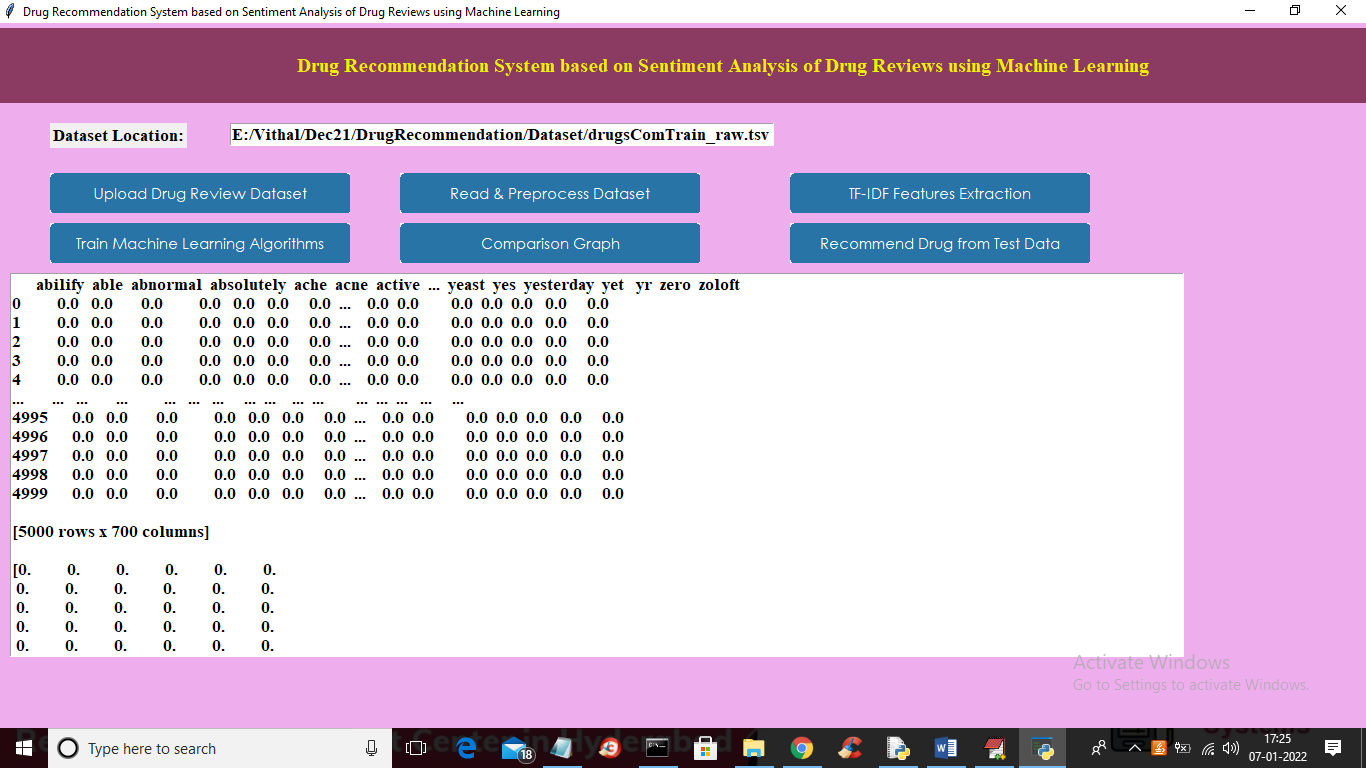
In above screen selecting and uploading DRUG dataset and then click on ‘Open’ button to load dataset and to get below screen



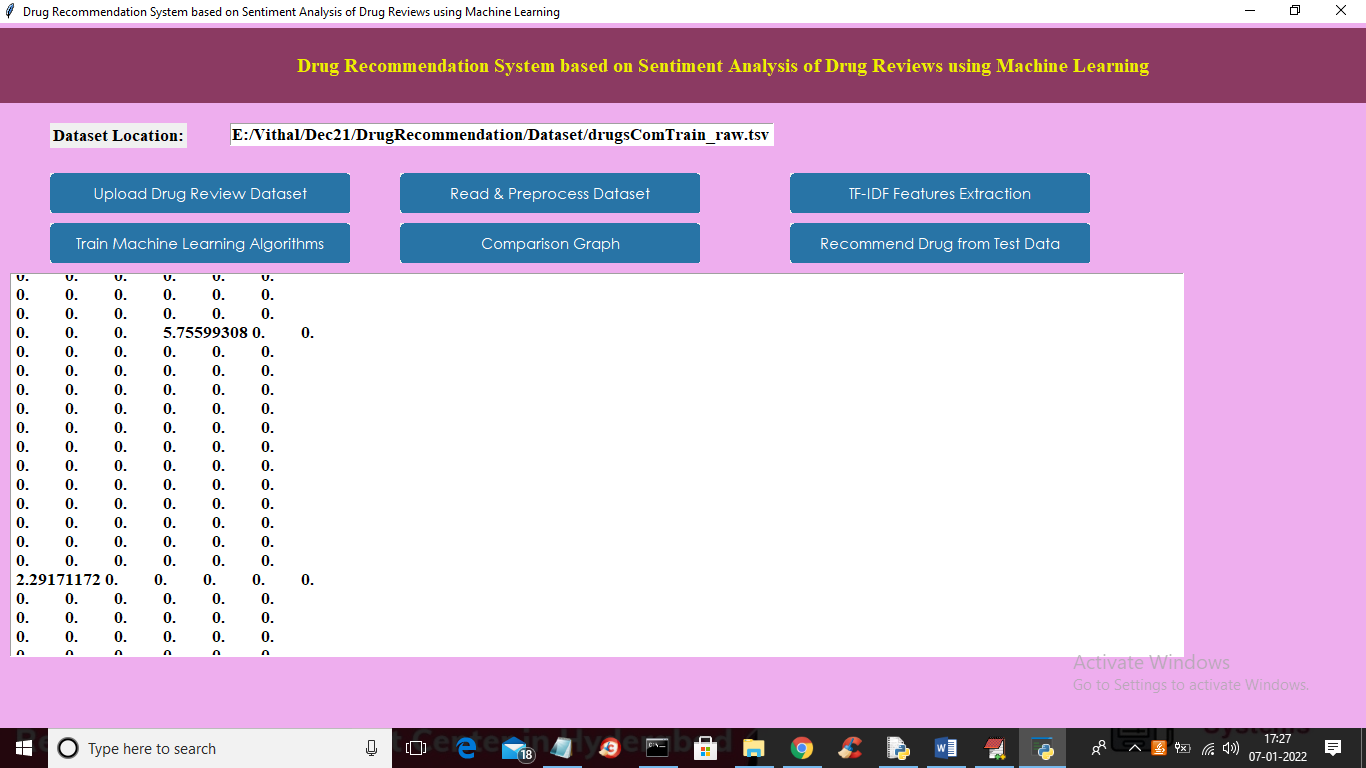
In above graph we can see dataset loaded and in graph x-axis represents ratings and y-axis represents total number of records which got that rating. Now close above graph and then click on ‘Read & Preprocess Dataset’ button to read all dataset values and then preprocess to remove stop words and special symbols and then form a features array.



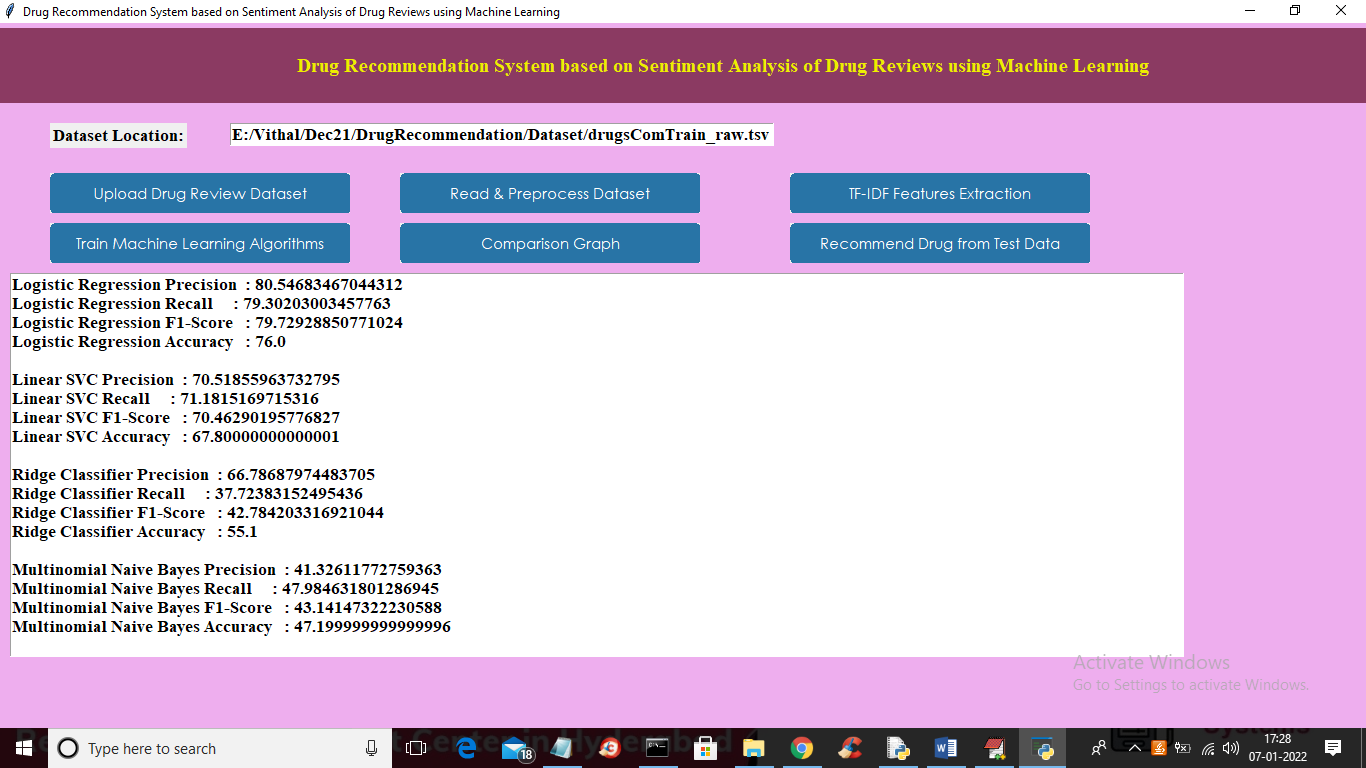
In above screen we can see from all reviews stop words and special symbols are removed and in graph I am displaying TOP 20 medicines exist in dataset. In above graph x-axis represents drug name and y-axis represents its count. Now close above graph and then click on ‘TF-IDF Features Extraction’ button to convert all reviews in to average frequency vector

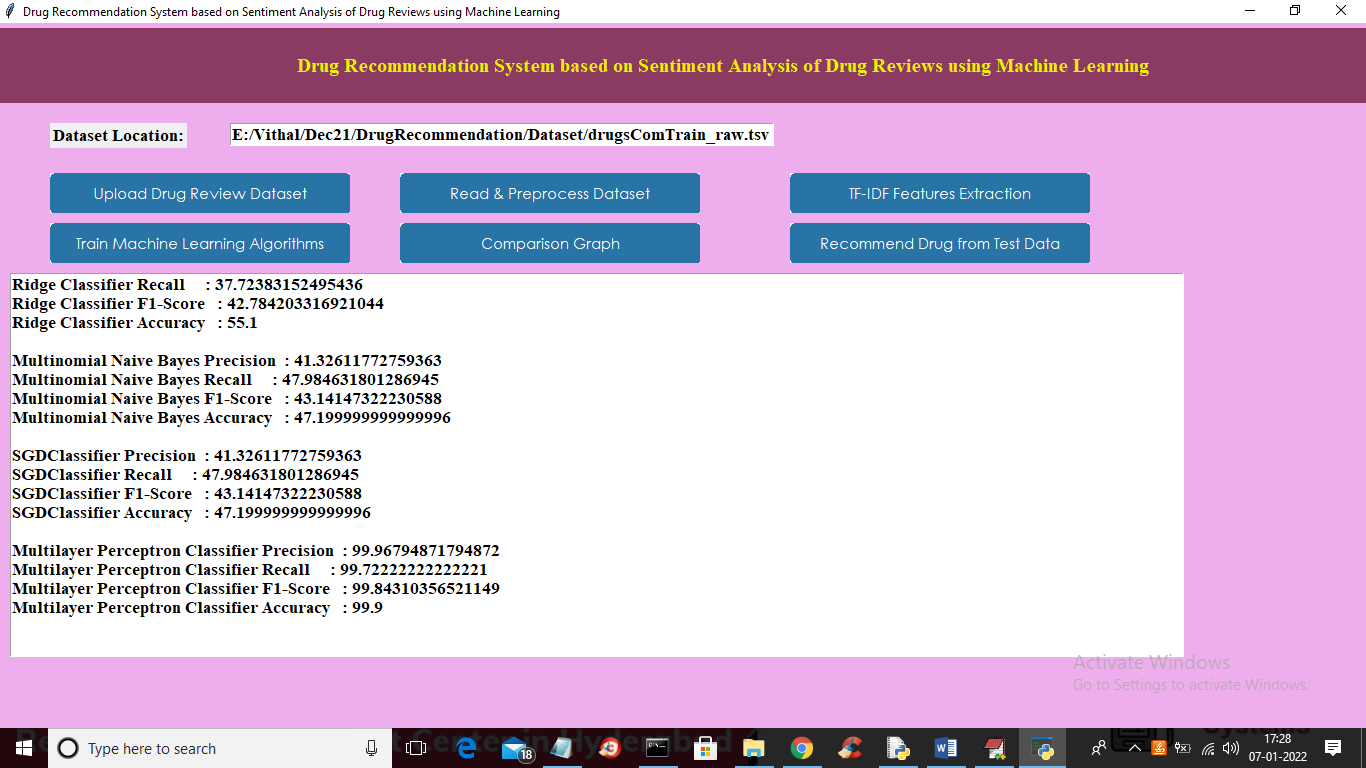


In above graph all reviews converted to TF-IDF vector where first row represents review WORDS and remaining columns will contains that word average frequency and if word not appear in review then 0 will put. Now scroll down above screen to view some non-zero frequency values

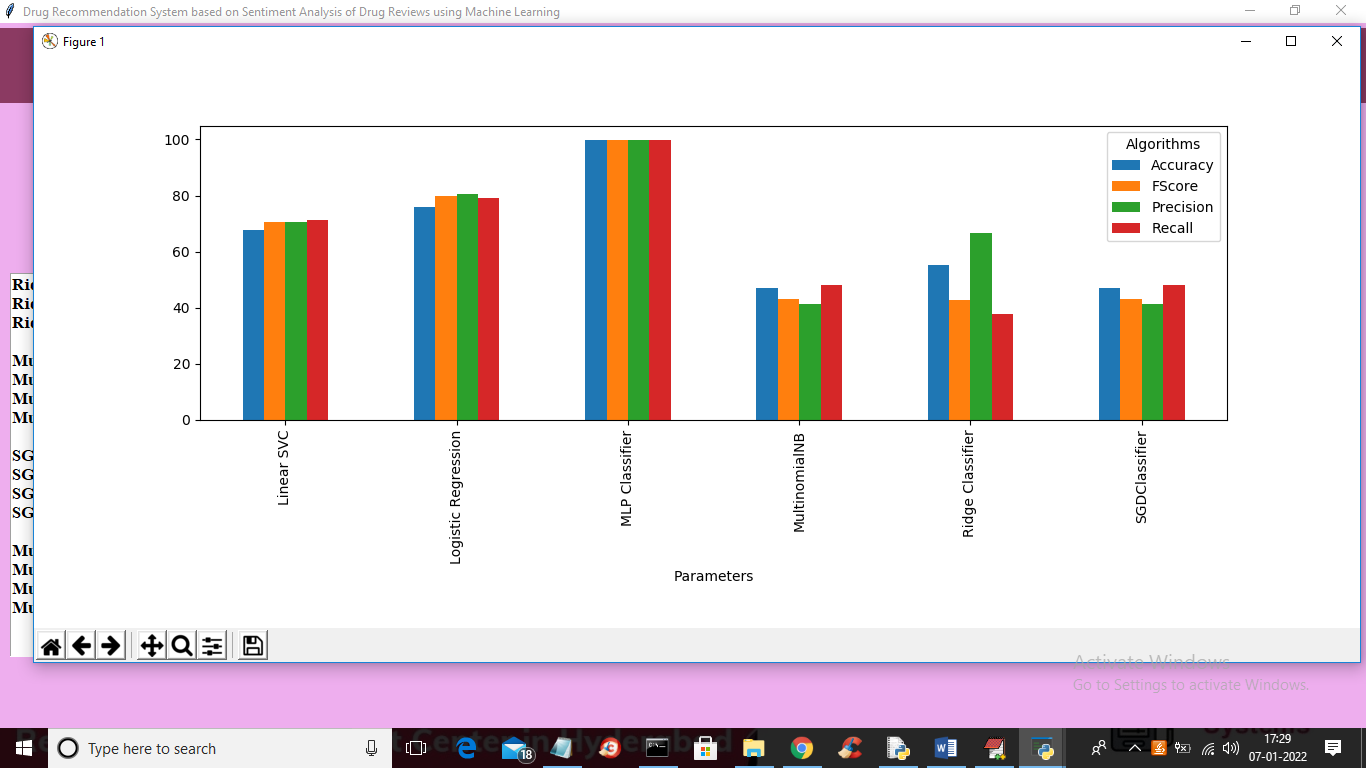


In above screen you can see some columns contains non-zero average frequency values and now TF-IDF vector is ready and now click on ‘Train Machine Learning Algorithm’ button to train all algorithm and get below output

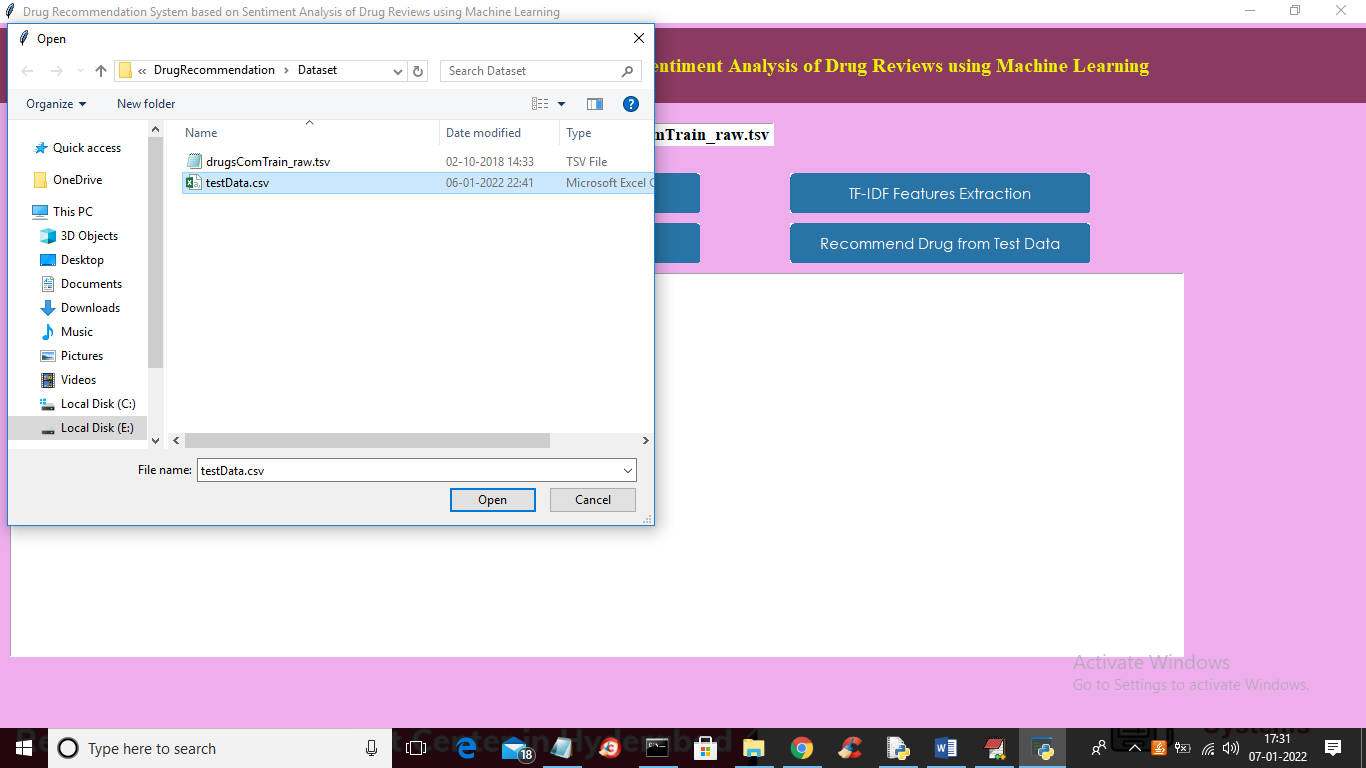




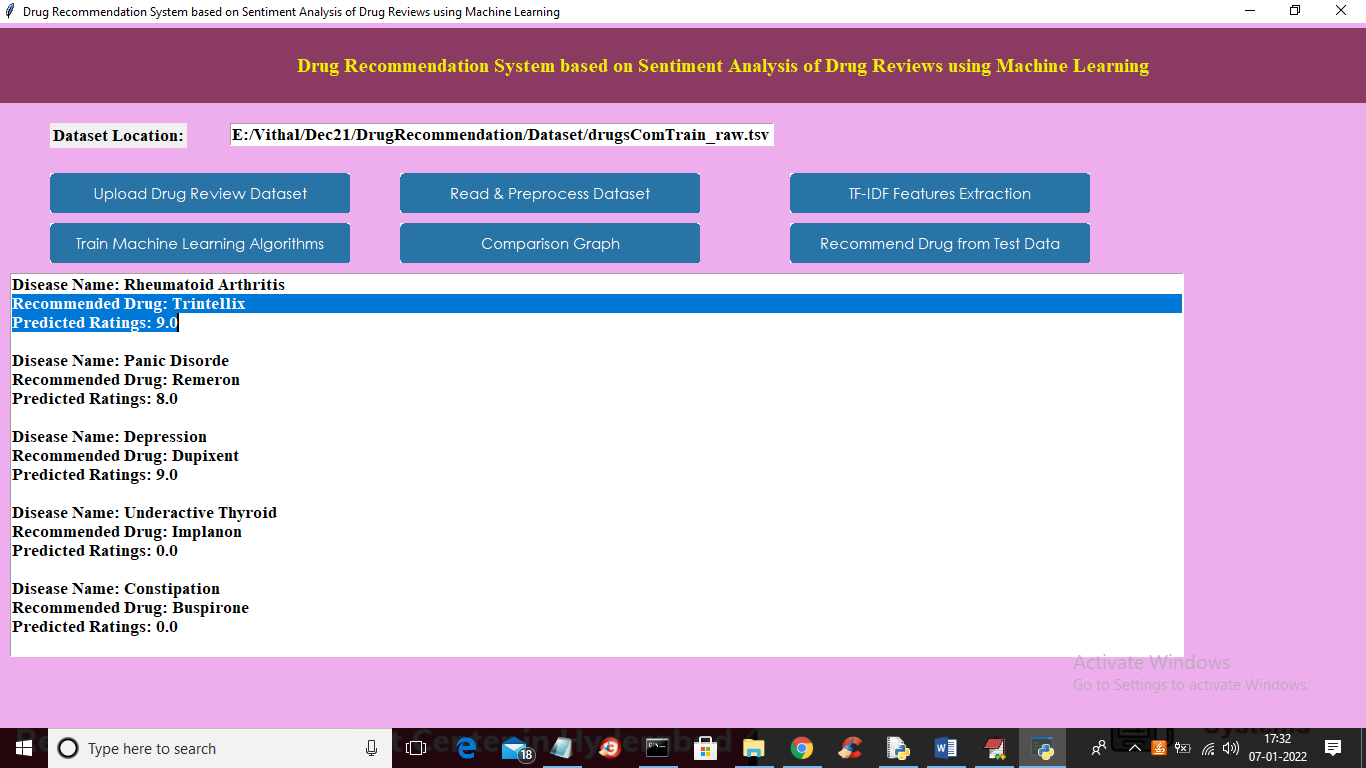
In above screen for each algorithm we calculate accuracy, precision, recall and FSCORE and in all algorithms MLP has got high performance and now click on ‘Comparison Graph’ button to get below graph



In above graph x-axis represents algorithm name and y-axis represents accuracy, precision recall and FSCORE where each different colour bar will represents one metric and in above graph we can see MLP got high performance. Now close above graph and then click on ‘Recommend Drug from Test Data’ button to upload test data and to get predicted result as drug name and ratings.



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



In above screen for each disease name application has predicted recommended drug name and ratings