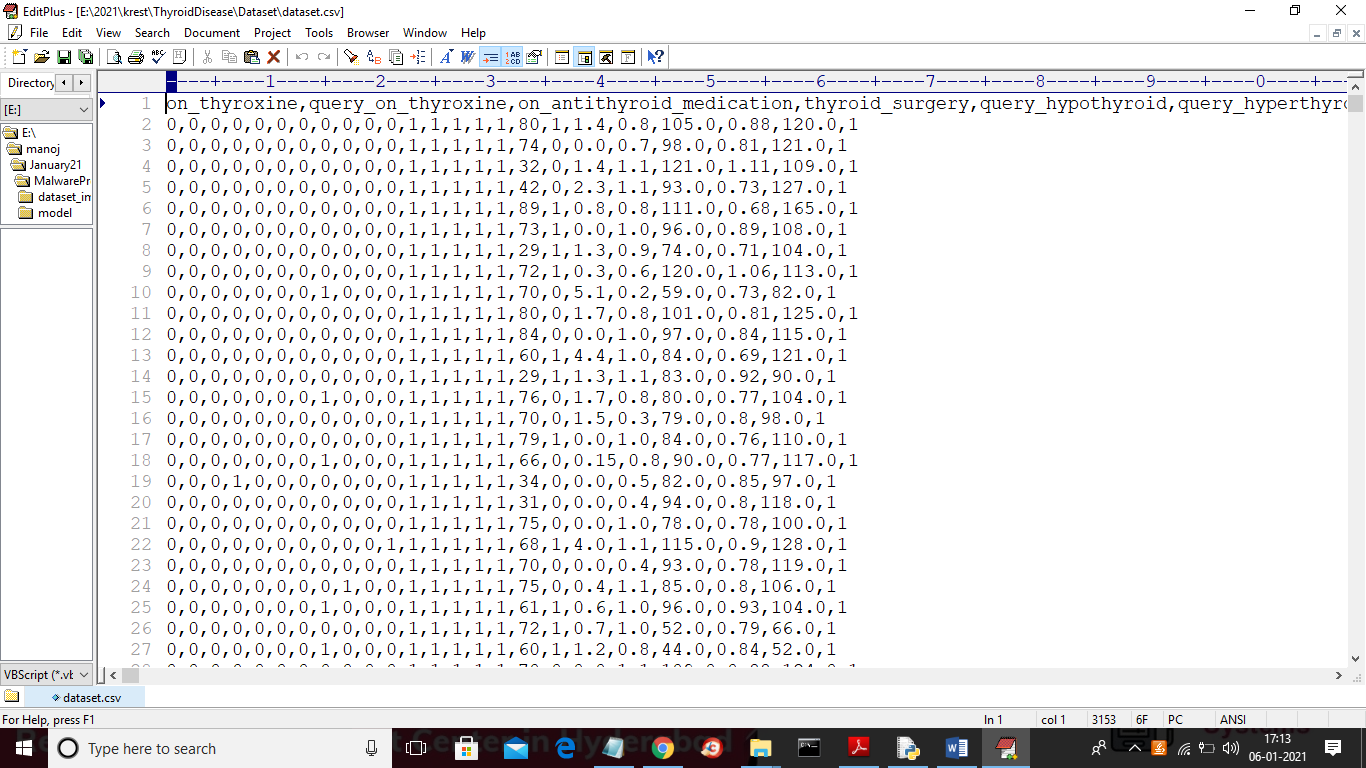
Prediction and Providing Medication for Thyroid Disease Using Machine Learning Technique (SVM)

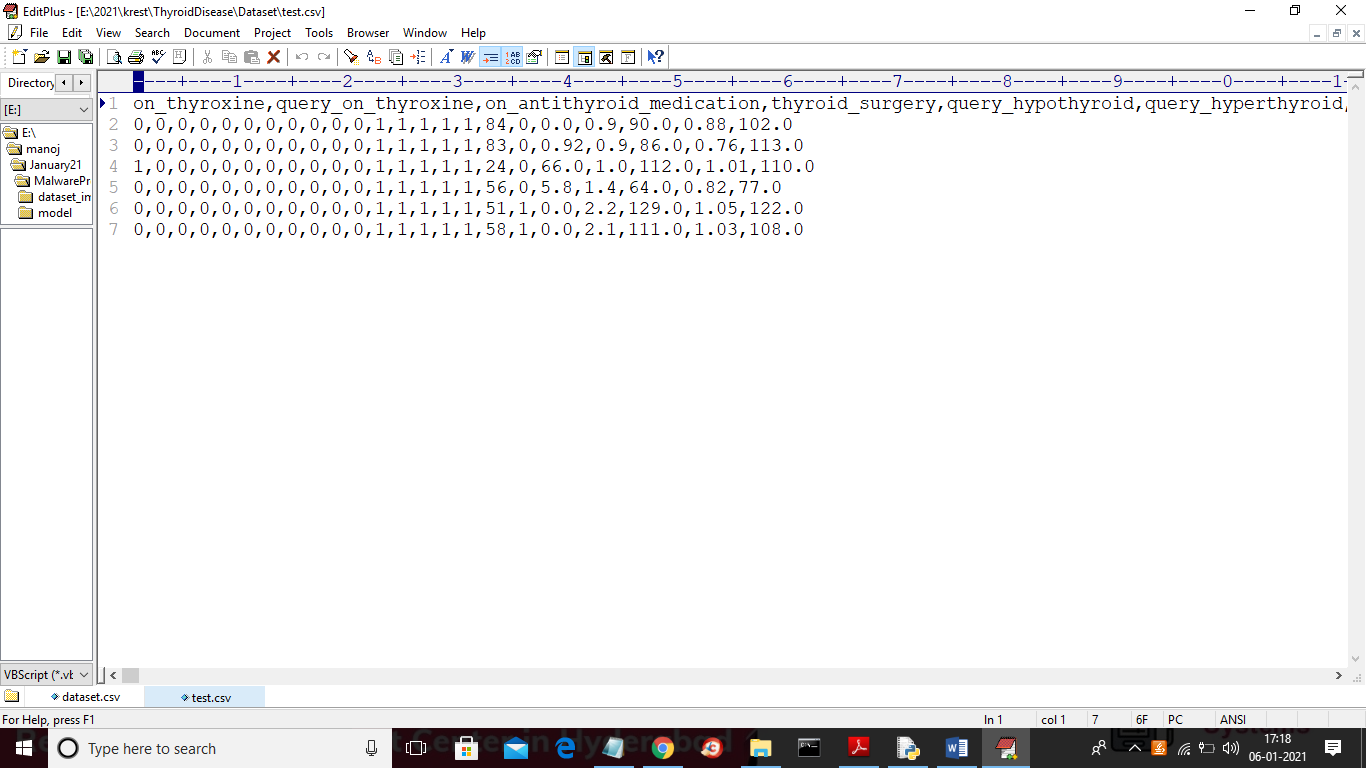
In this project we are using SVM machine learning algorithm to predict whether patient report data is normal or at risk of thyroid disease and if thyroid disease predicted from patient report then application will display proper diet and medication details. In this project we are using UCI machine learning THYROID disease dataset to train SVM algorithm and to generate prediction model. New patient test data will be applied on trained SVM model to predict whether patient is normal or at risk of thyroid disease.

Below screen shots showing dataset columns and its values



In above dataset first row contains column names and other rows contains values as 0 or 1 and if patient is under thyroid medication or surgery then its column value will be 1 else 0 and in last column contains class label as 0 or 1 where 0 means patient record is normal and 1 means patient record contains thyroid disease. In this dataset more than 3000 rows are there and 24 columns are available. For prediction all 24 columns are not available so we are applying PCA (principal component analysis) feature selection algorithm as extension concept to optimize features or to reduce columns or features which are not important for prediction. PCA will remove unnecessary columns from the dataset and use only important attributes to train SVM algorithm and due to optimize features SVM prediction accuracy can be increase.

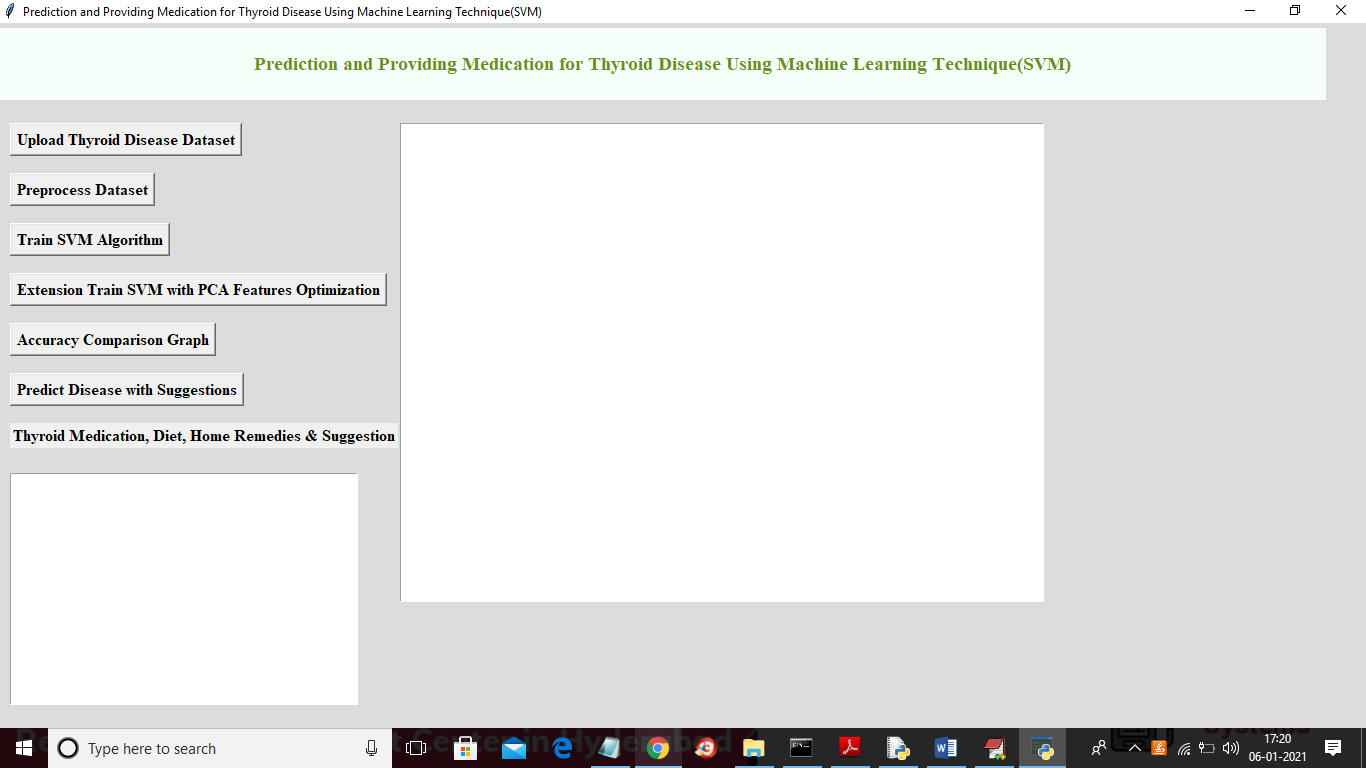
Below is the test dataset which contain no class label and SVM will predict class label for this test data



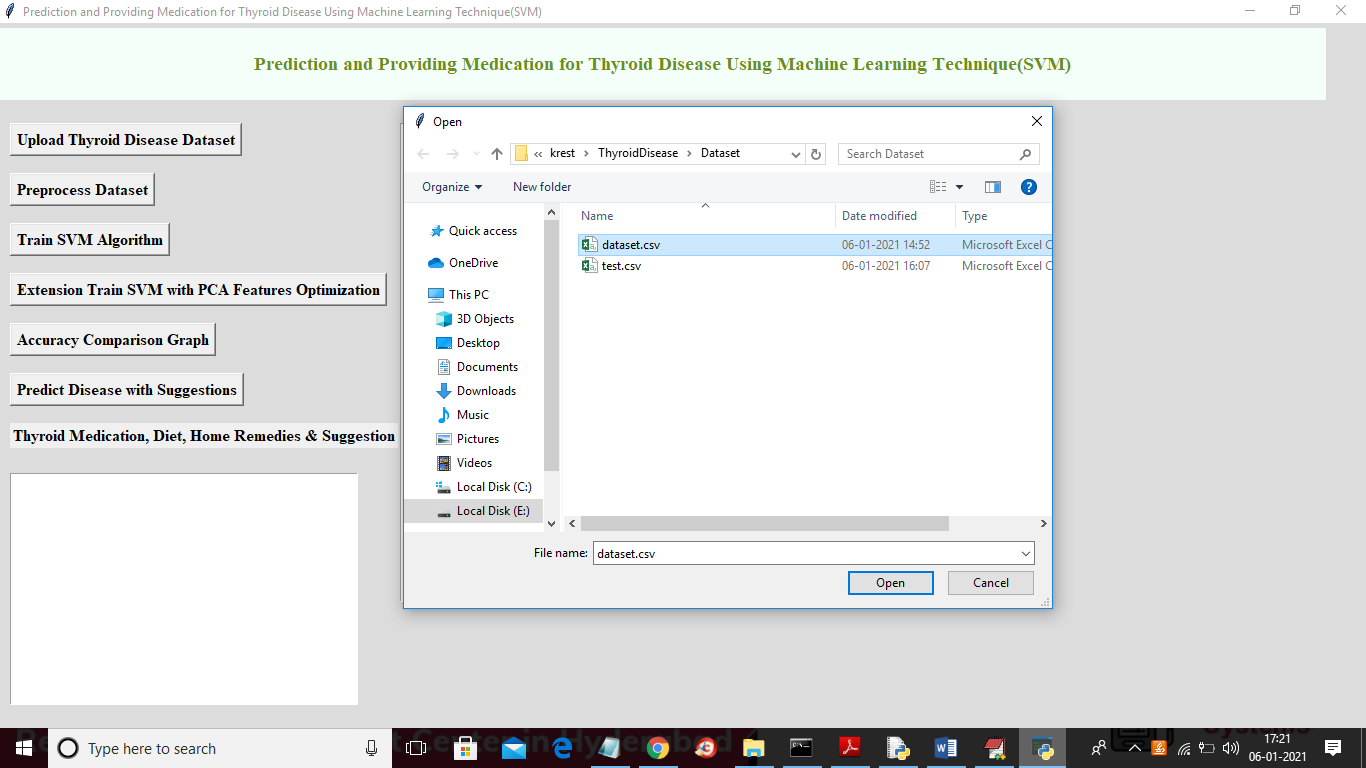
In above test dataset we can see there is no class label with value 0 or 1 and SVM will predict that value

SCREEN SHOTS

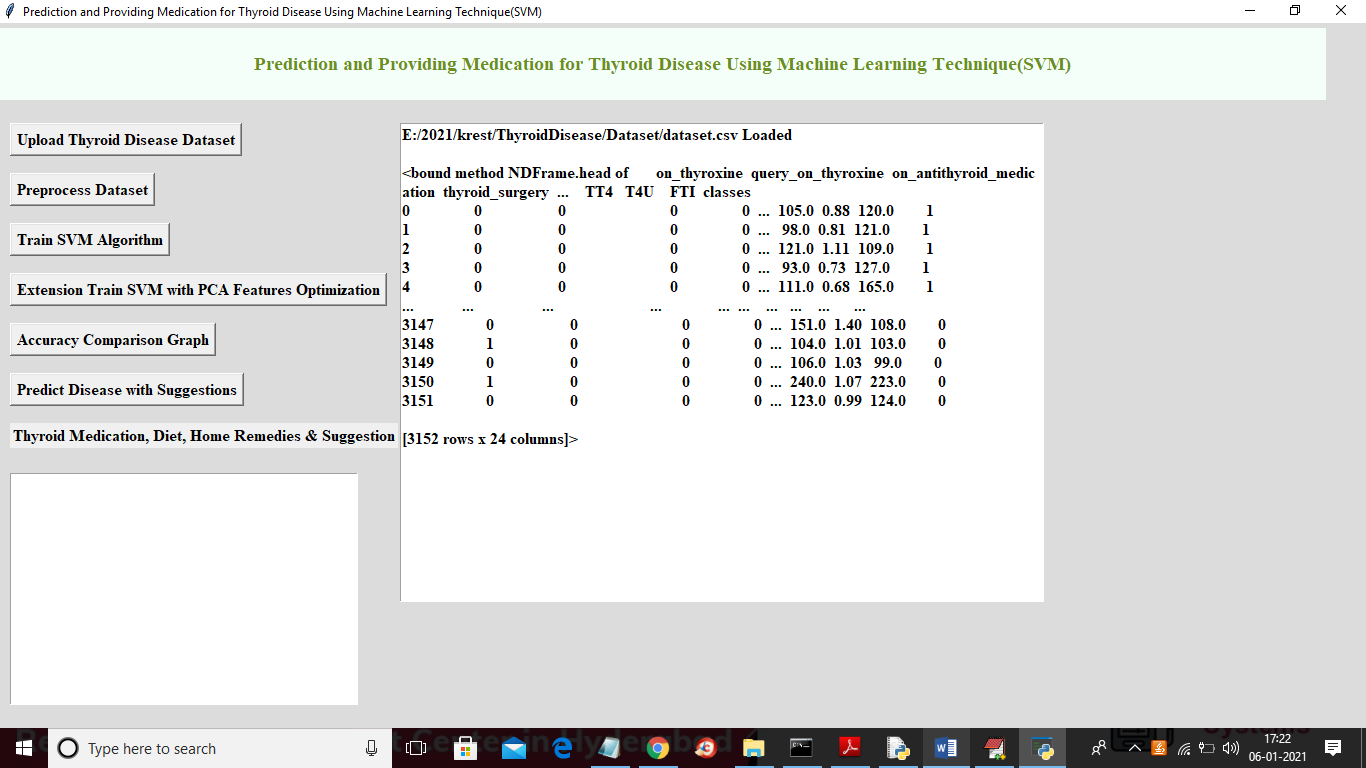
To run project double click on ‘run.bat’ file to get below screen



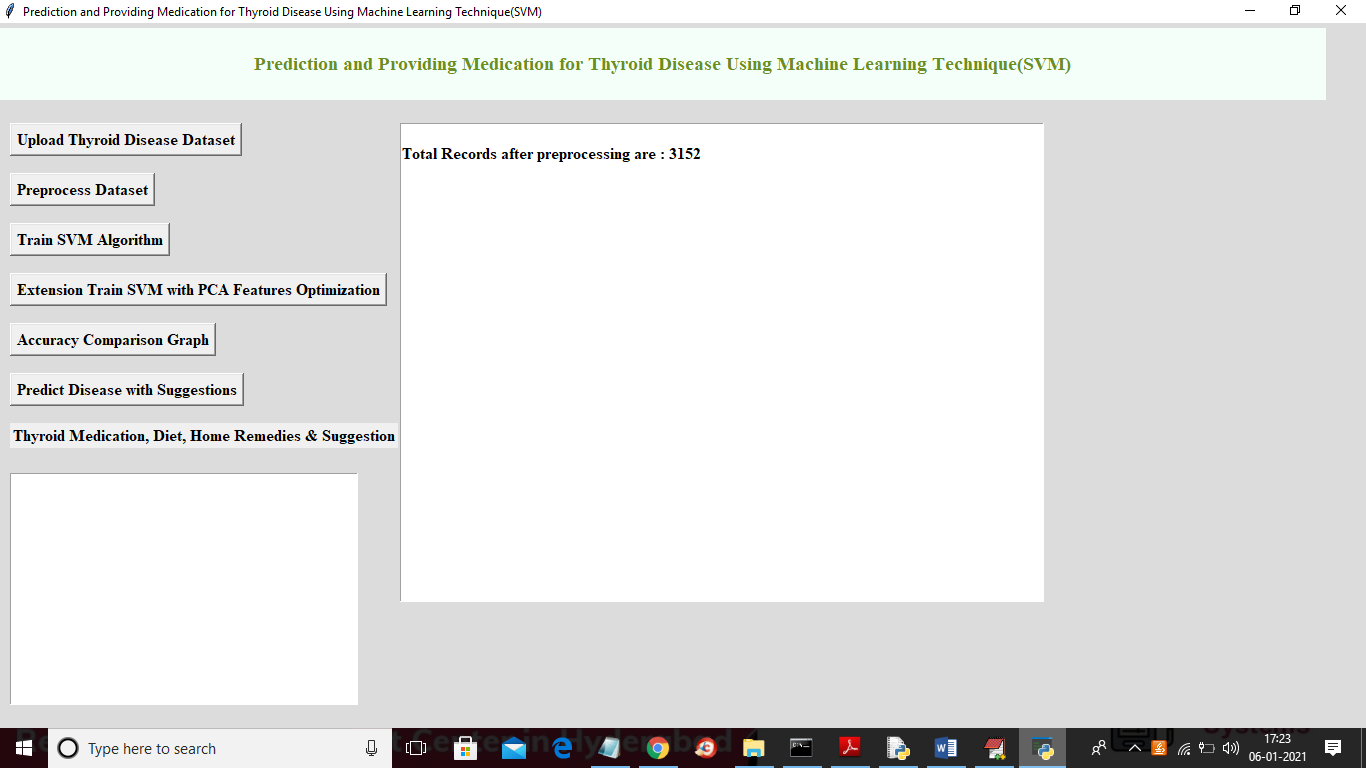
In above screen click on ‘Upload Thyroid Disease Dataset’ button to upload dataset and to get below screen



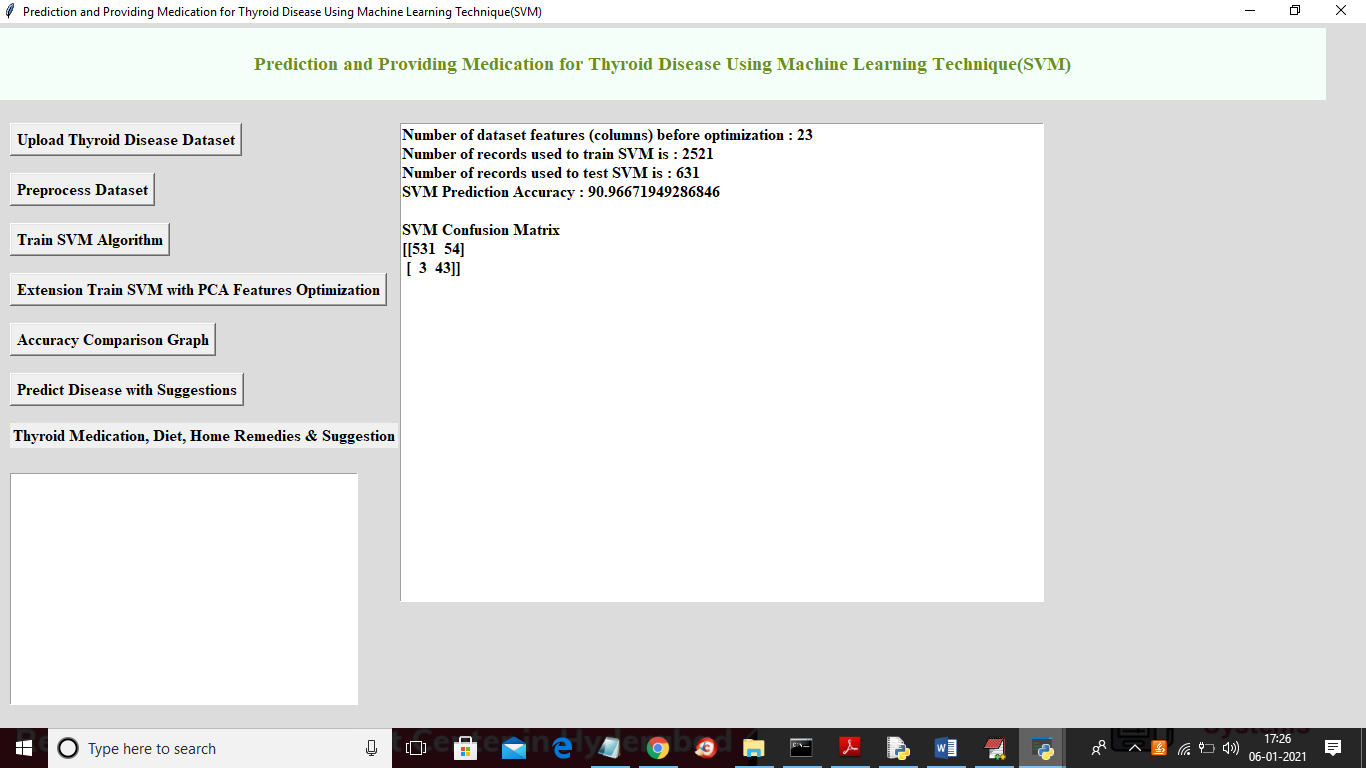
In above screen selecting and uploading ‘dataset.csv’ file and then click on ‘Open’ button to load dataset and to get below screen



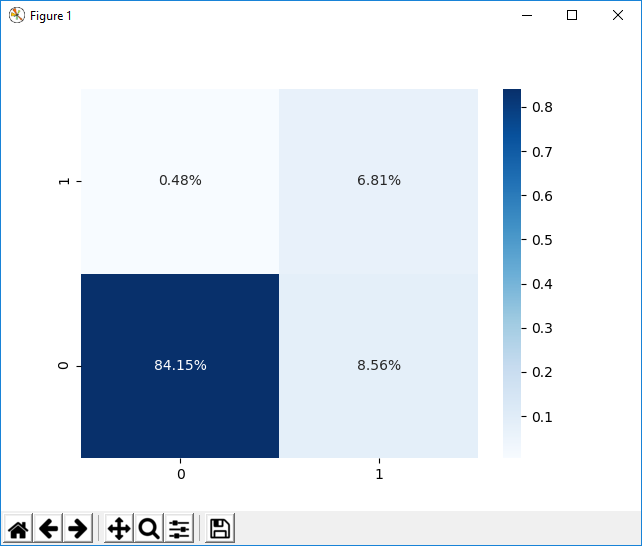
In above screen dataset loaded and displaying few records from dataset and then click on ‘Preprocess Dataset’ button to remove missing and NAN values from dataset and to separate X and Y values where X contains all dataset values and Y contains class label value.



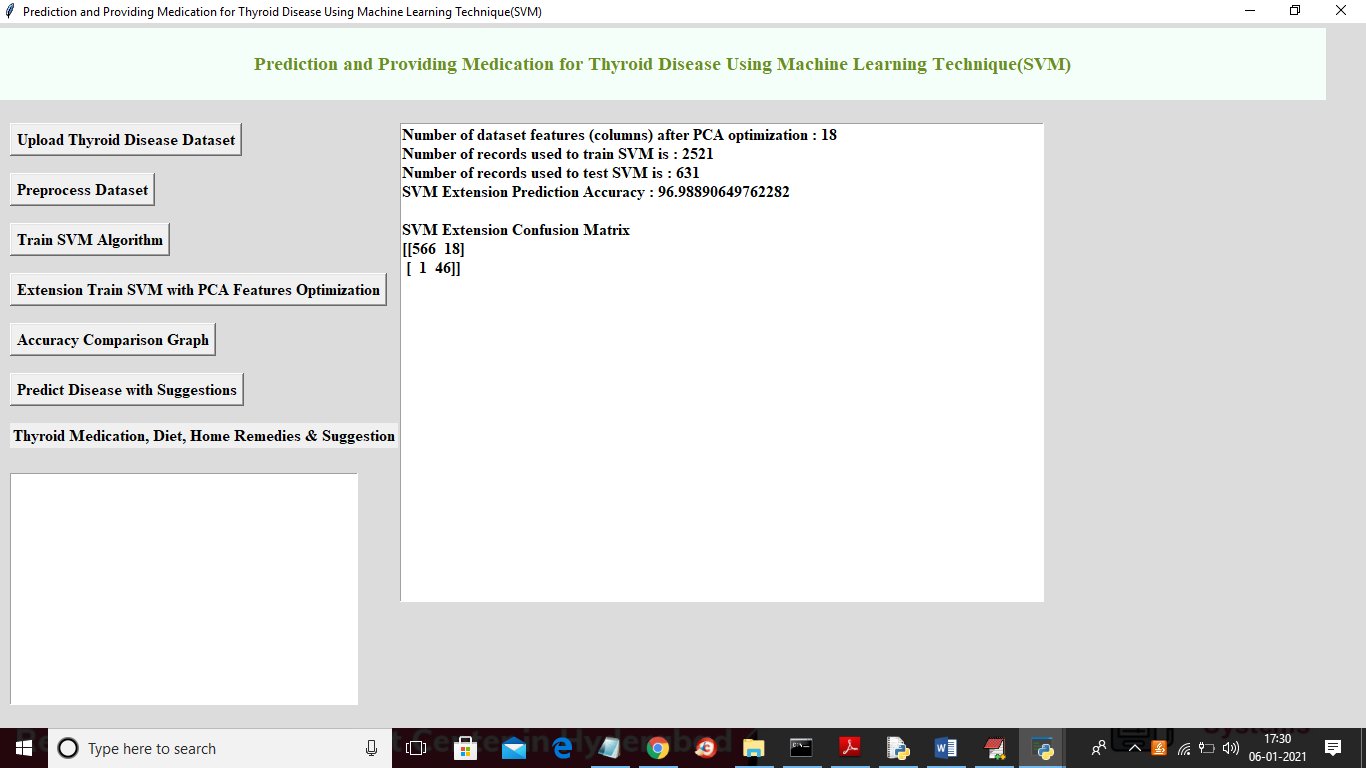
In above screen dataset showing 3152 preprocess records and now dataset is ready and now click on ‘Train SVM Algorithm’ button to split dataset into train and test and then apply SVM algorithm on train data to generate model and then model will be applied on test data to calculate prediction accuracy



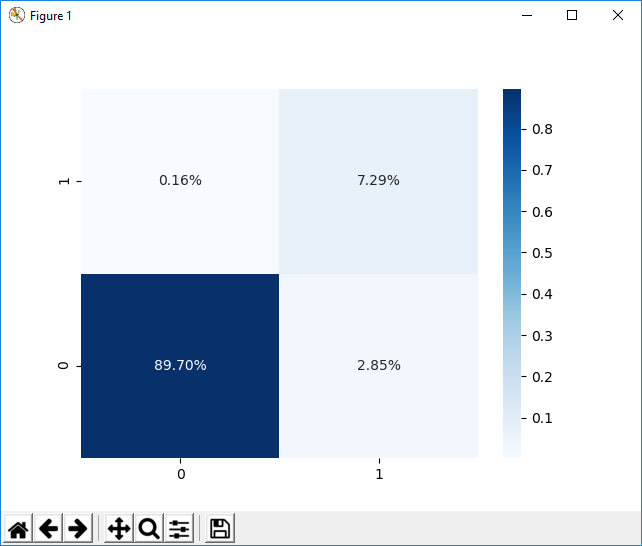
In above screen we can see dataset contains total 23 columns and using 2521 records to train SVM algorithm and using 631 test records to test SVM prediction accuracy and with normal SVM we got prediction accuracy as 90.96% and application showing confusion matrix of true and false prediction values where 531 and 3 are the true prediction and 54 and 43 are the false or incorrect prediction and below is the graph format of confusion matrix



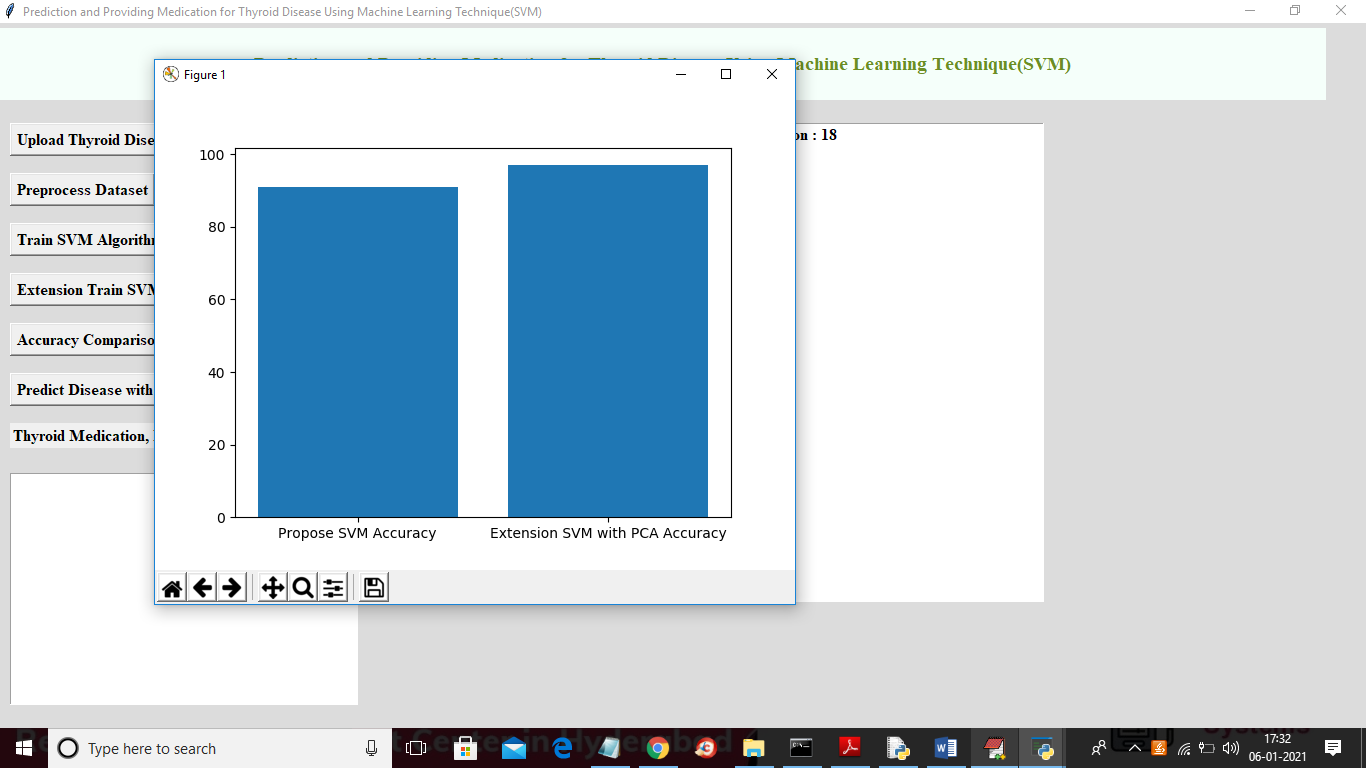
In above graph 84.15% and 6.81% is the true prediction and now click on ‘Extension Train SVM with PCA Features Optimization’ button to train SVM with PCA features optimization and to get below prediction accuracy



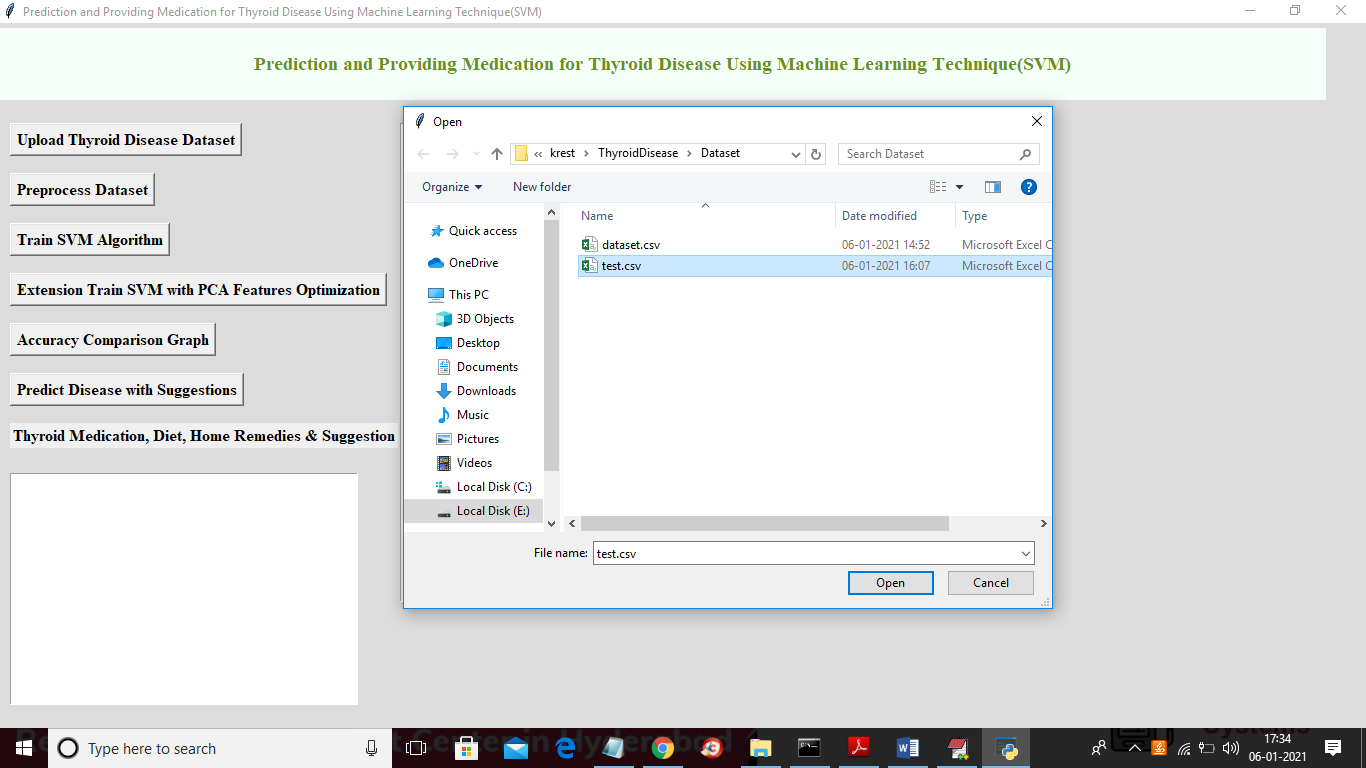
In above screen SVM with PCA extension got 96.98% prediction accuracy and confusion matrix values is also better compare to normal SVM and below is extension SVM confusion matrix graph



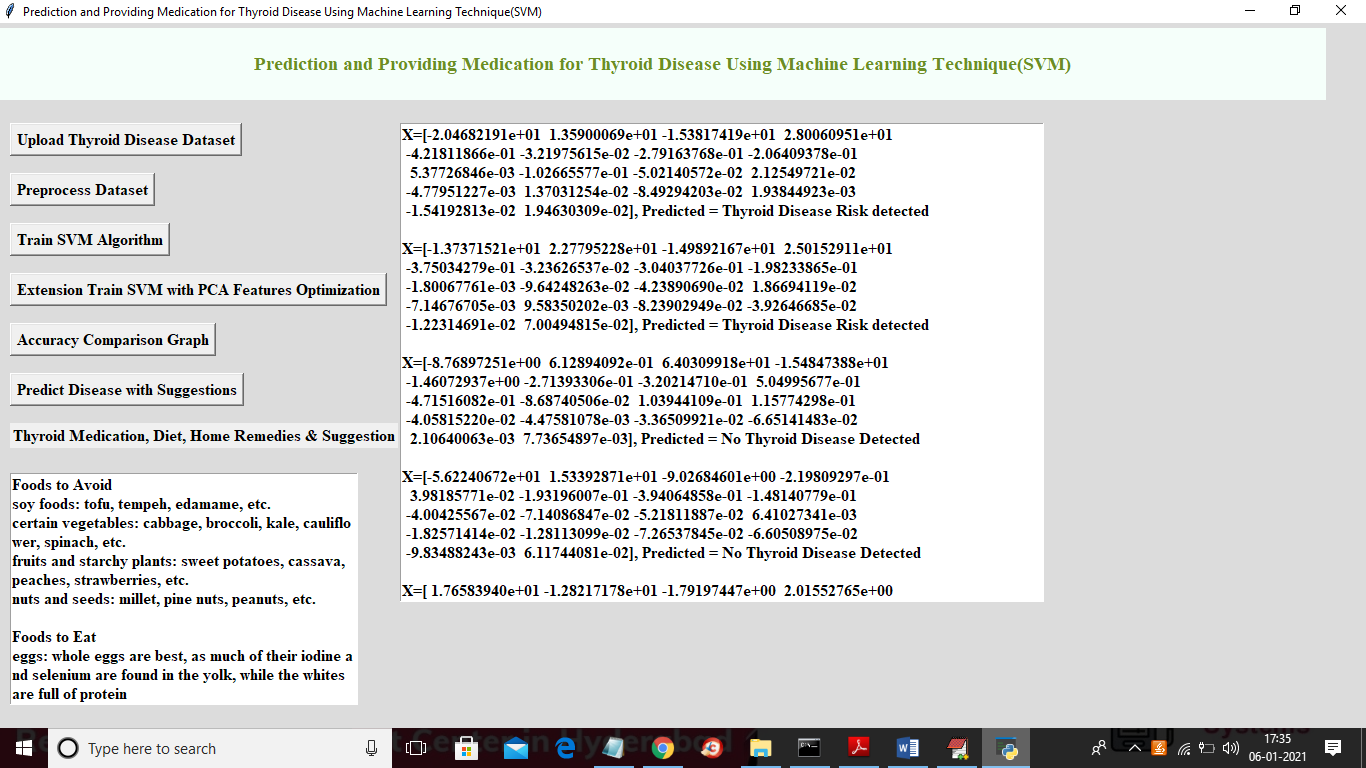
In above graph 89.70 and 7.29% is the correct prediction and other values are the false prediction. Now click on ‘Accuracy Comparison Graph’ button to get below accuracy comparison graph



In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can conclude that extension SVM with PCA is better than normal SVM and now click on ‘Predict Disease with Suggestions’ button to upload new test data and predict whether new test data contains thyroid or not



In above screen selecting and uploading ‘test.csv’ file and then click on ‘Open’ button to upload test dataset and to predict disease and to get below screen



In above screen in brackets we can see each record test value and after bracket we can see value as thyroid risk detected or not and if detected then it left box we are showing diet and medication plan as suggestion