A Two-Stage Model to Predict Surgical Patients Lengths of Stay From an Electronic Patient Database

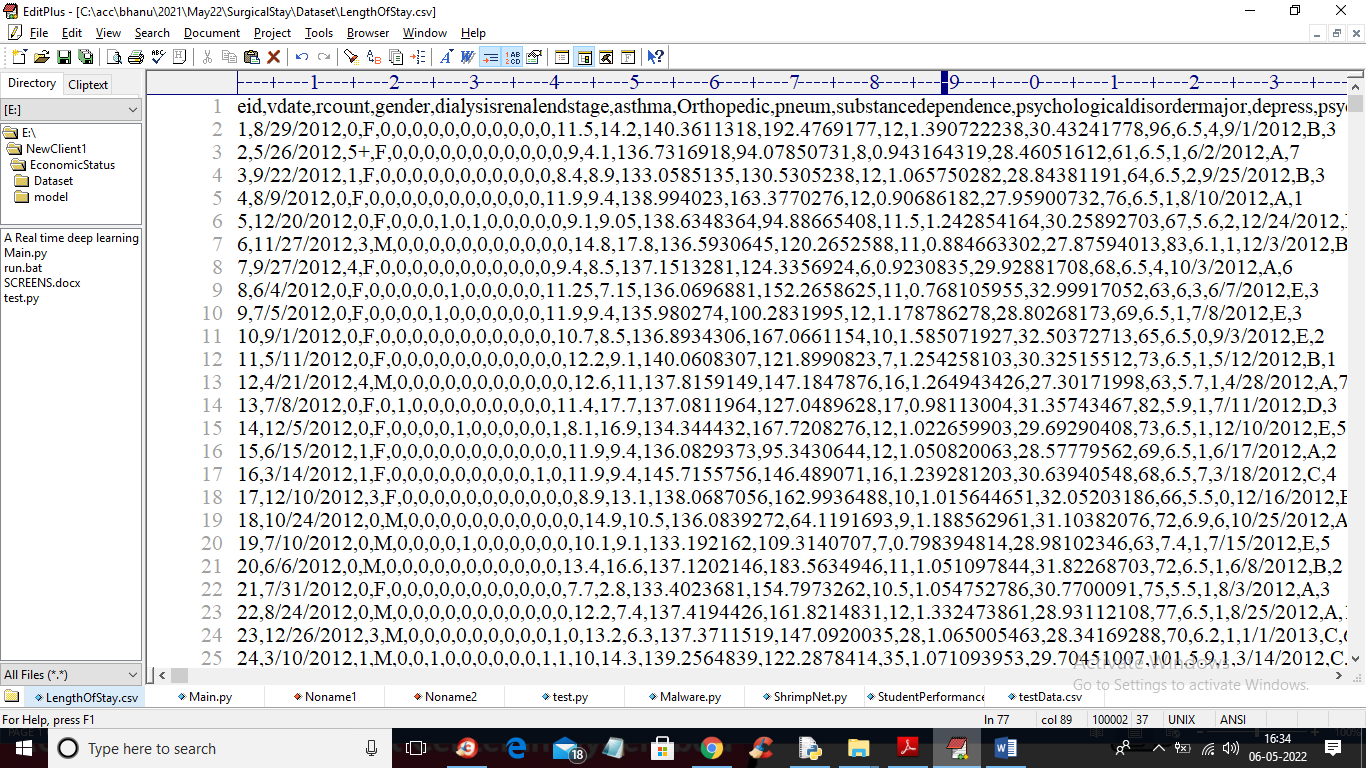
In this paper author is introducing two-stage algorithm called CART (classification and regression tree) to predict patient hospital stay after surgery. Now-a-days it’s become difficult for patients and hospital to maintain inflows of the patient as if patient stay for longer duration due to miscalculation after surgery then it will raise bill amount for the patient and hospital resources like rooms and other articles will be wasted and cannot be used for other patients.

To overcome from this problem author introducing two stage algorithm which will predict hospital stay by using classification in STAGE 1 and then further enhance this classification output by adding STAGE 2 technique called clustering and both this stages will run by using algorithm called CART. CART will split or partition data into cluster which help in getting accurate classification result as similar data will be in same cluster.

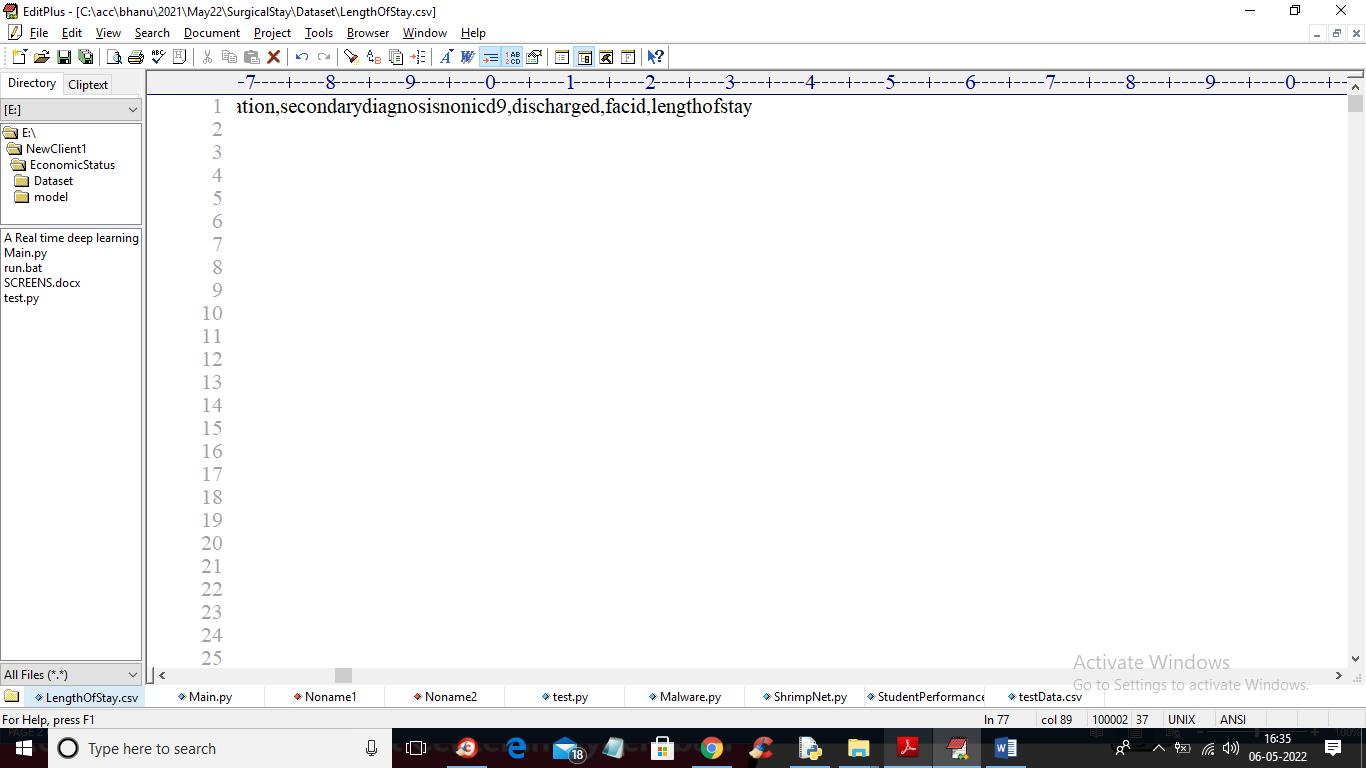
To implement this project author has collected real patient and hospital dataset and not publish this dataset on internet so we have patient LENGTH OF STAY (LOS) dataset from internet.

Propose algorithm will apply strategic technique (such as classification, regression and clustering by partition data into tree cluster) to predict patient LOS so its prediction output will be better. Existing algorithms such as Random Forest will not use any strategic technique so its prediction will not better and its error will be low compare to propose CART algorithm as its not applying any strategic technique.

Below screen showing dataset details



In above screen first row contains dataset column names and remaining rows are the dataset values of patient records and in last column we can see LOS values



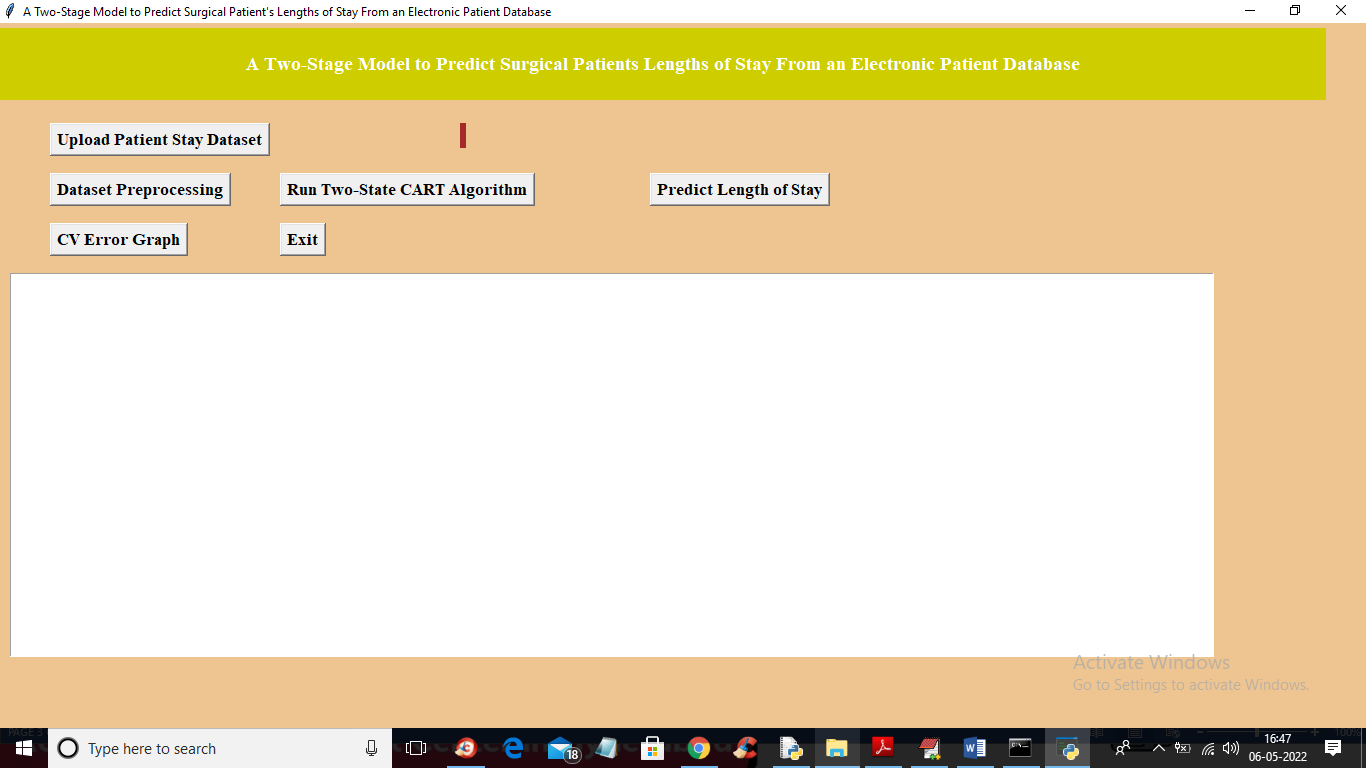
In above screen last column name LENGTH OF STAY will be used as class label. In paper at page number 4 you can read the advantages of CART algorithm

To implement this project we have designed following modules

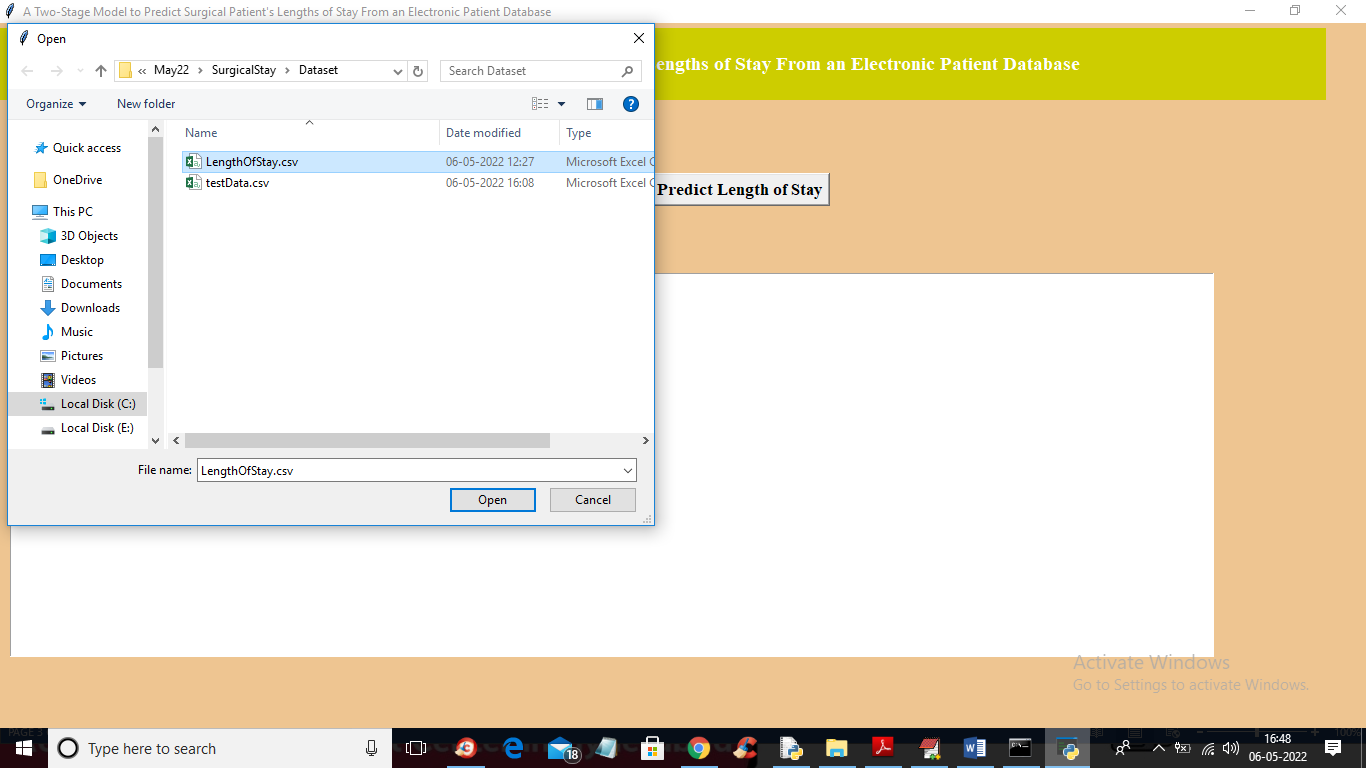
1. Upload Patient Stay Dataset: using this module we will upload dataset to application
2. Dataset Preprocessing: using this module we will read dataset and then replace missing values and then encode all non-numeric data to numeric values and then split dataset into train and test
3. Run Two-State CART Algorithm: processed train dataset will be train with Random Forest, KNN and CART algorithm by using classification in STAGE 1 and then train all 3 algorithm in STAGE 2 using clustering and regression algorithm. All this algorithm get trained using cross validation algorithm which train each algorithm multiple time and then find average error rate.
4. Predict Length of Stay: using this module we will upload patient test data and then CART algorithm will analyse patient test data and predict length of stay
5. CV Error Graph: using this module we will plot misclassification error rate for each algorithm

SCREEN SHOTS

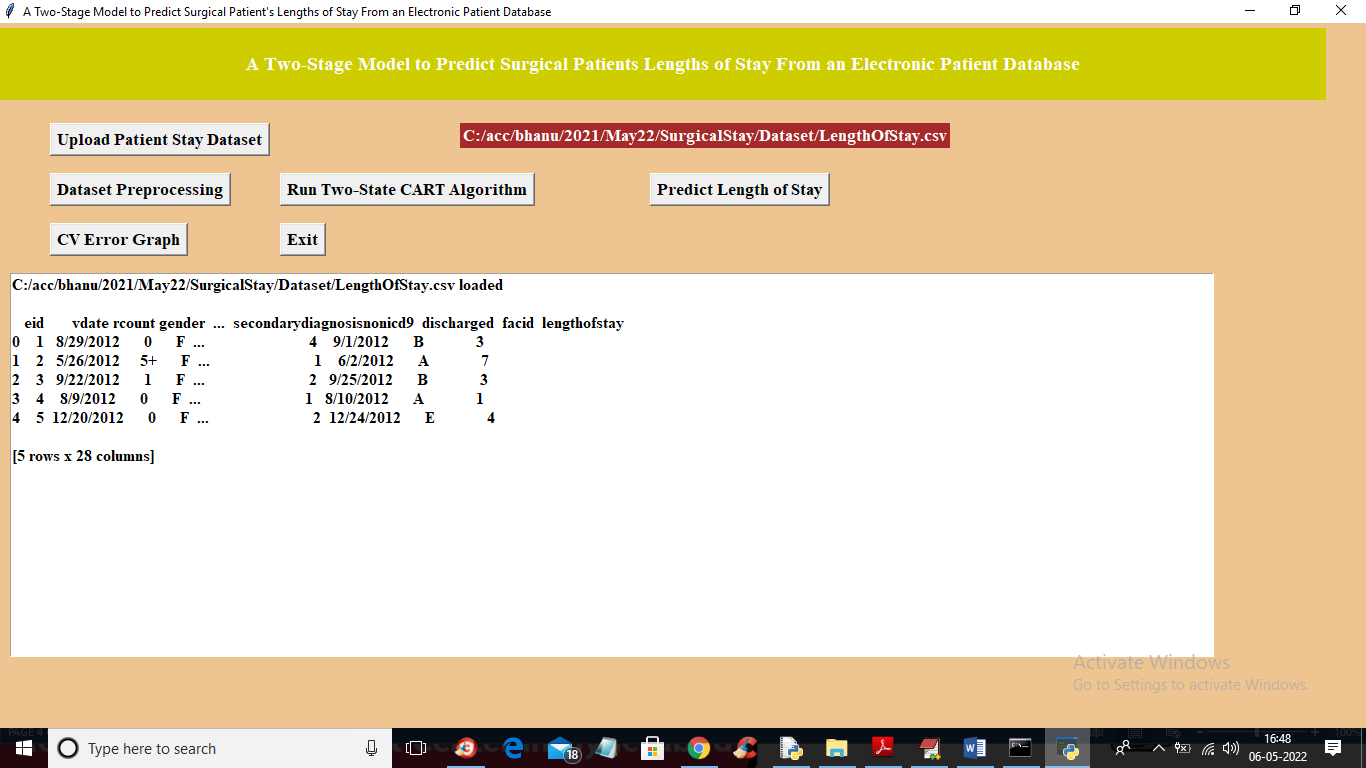
To run project double click on ‘run.bat’ file to get below output



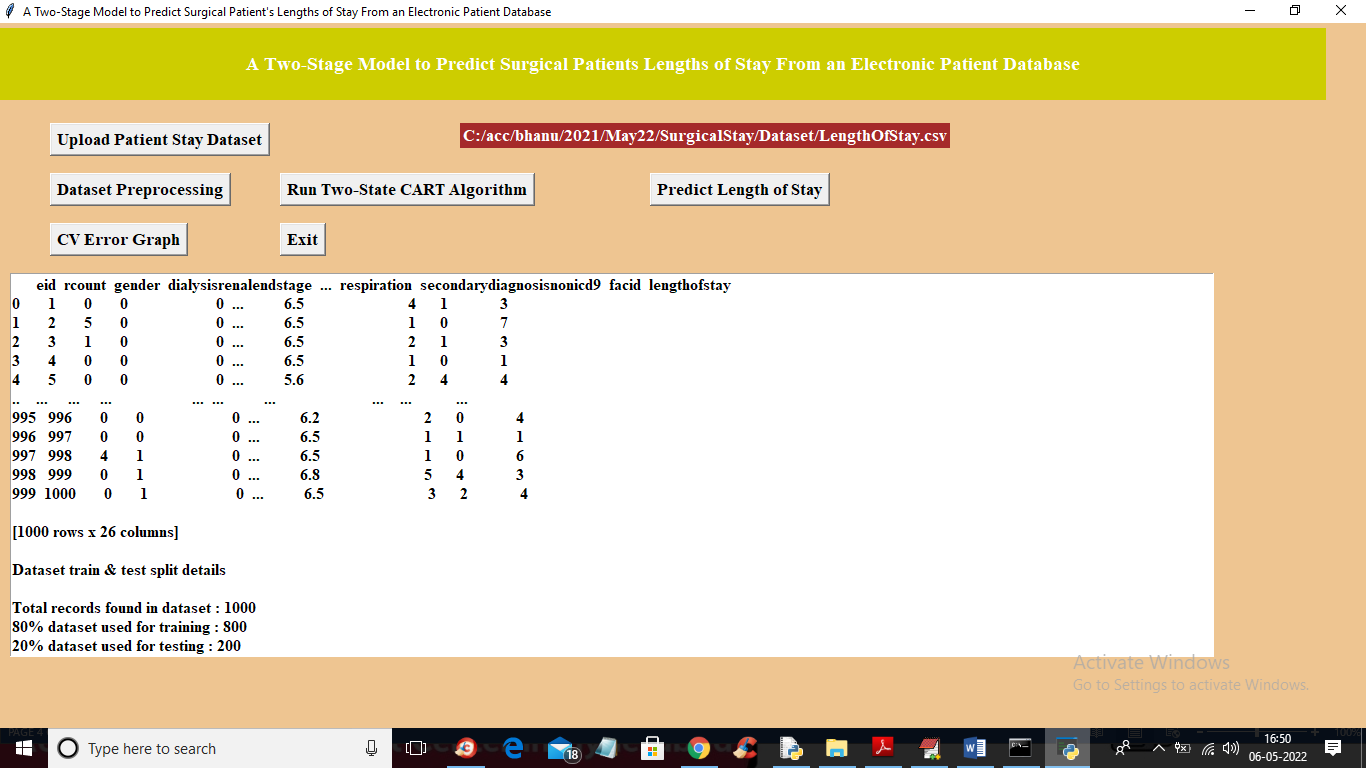
In above screen click on ‘Upload Patient Stay Dataset’ button to upload dataset



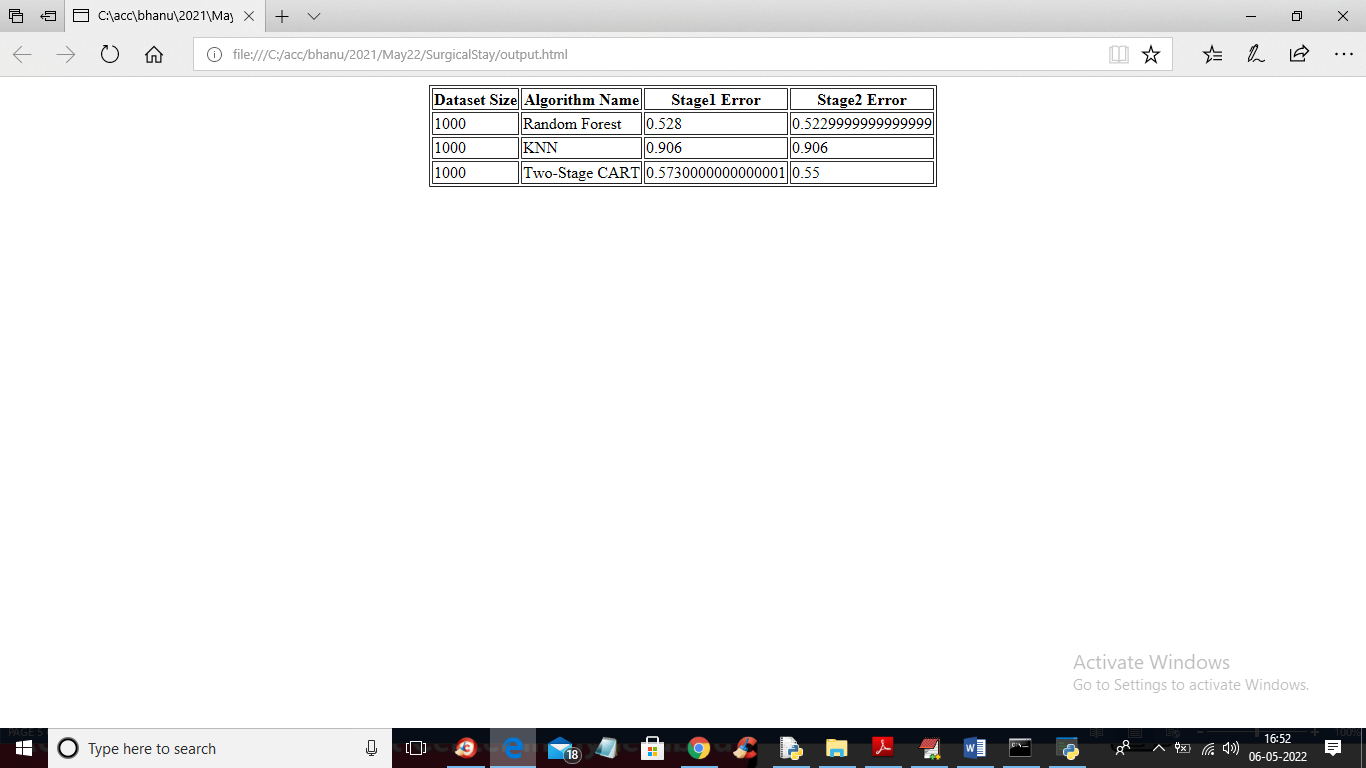
In above screen selecting and uploading dataset and then click on ‘Open’ button to get below output



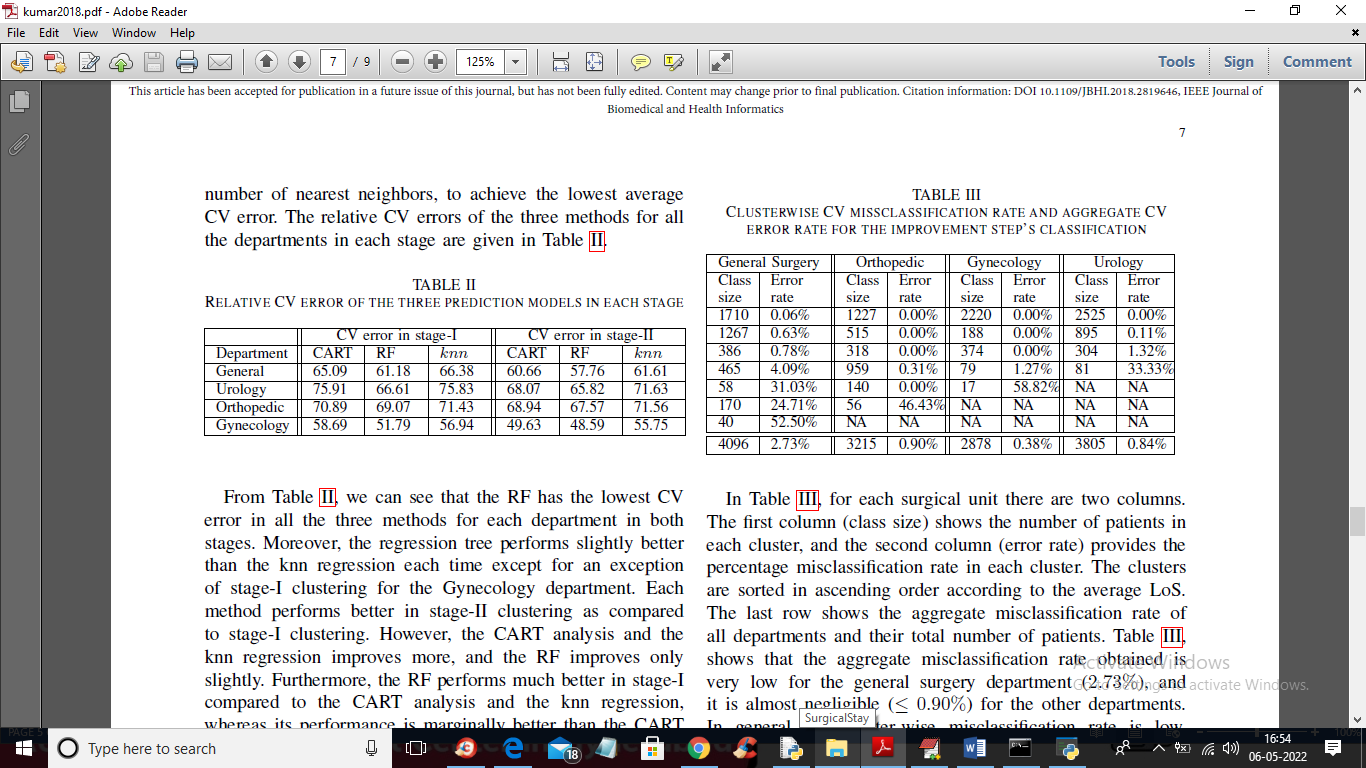
In above screen dataset loaded and we can see dataset contains some non-numeric values so we need to process dataset to convert or encode non-numeric to numeric values. Now click on ‘Dataset Preprocessing’ button to clean and convert data to numeric



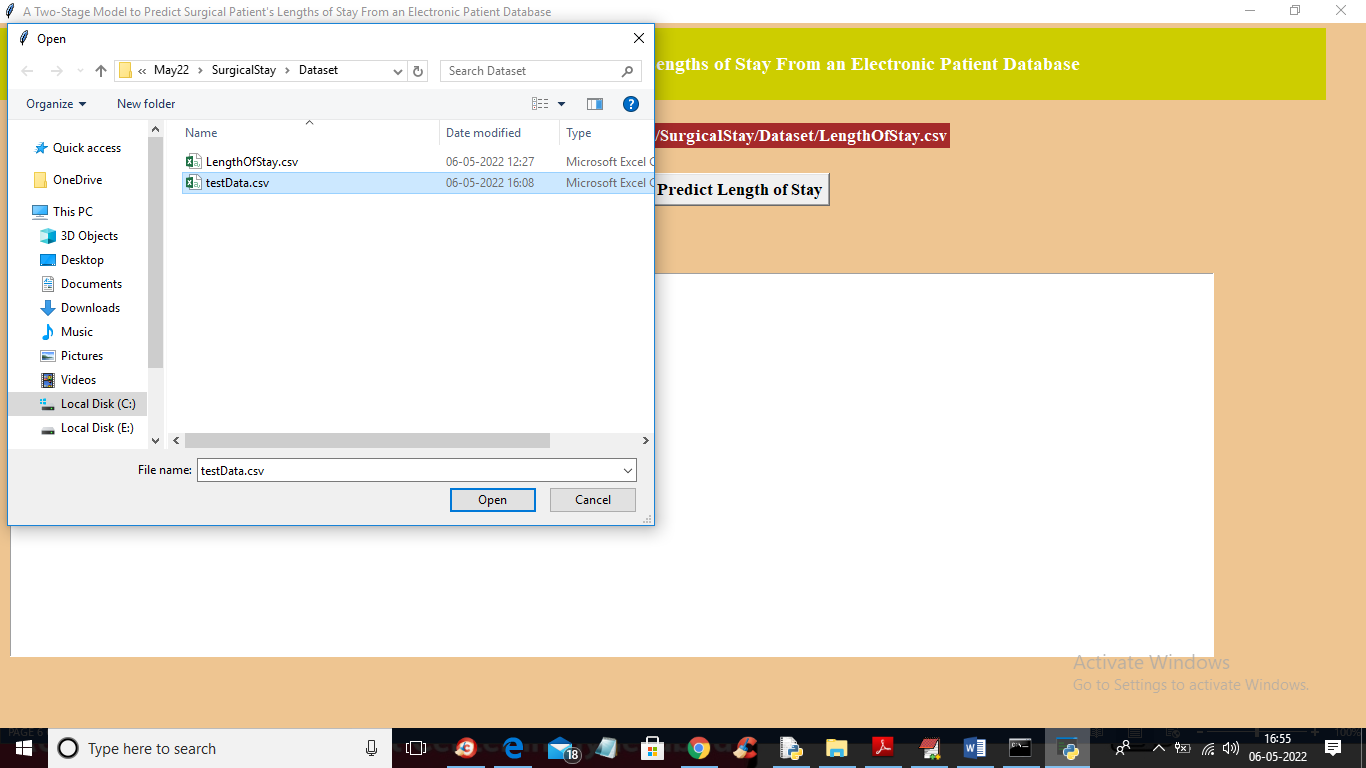
In above screen we can see all dataset values converted to numeric and we can see dataset contains 1000 records and 800 using for training and 200 for testing. Now dataset is ready and now click on ‘Run Two-State CART Algorithm’ button to train all algorithm and get below output



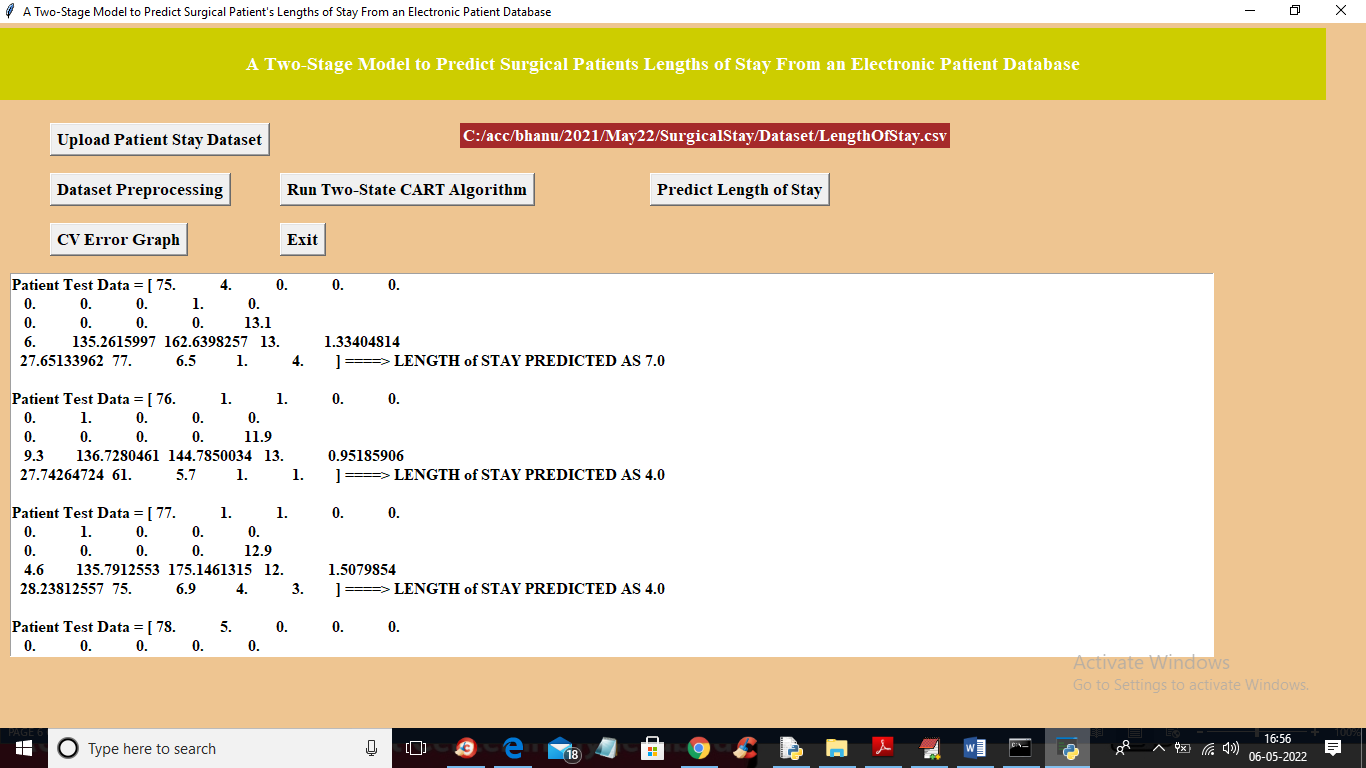
In above screen we can see misclassification error rate for each algorithm in stage 1 and 2 and in above screen we can see Random Forest error rate is less compare to propose CART but CART using strategic techniques and Random Forest not applying and in below PAPER output also we can see Random Forest error is less compare to CART



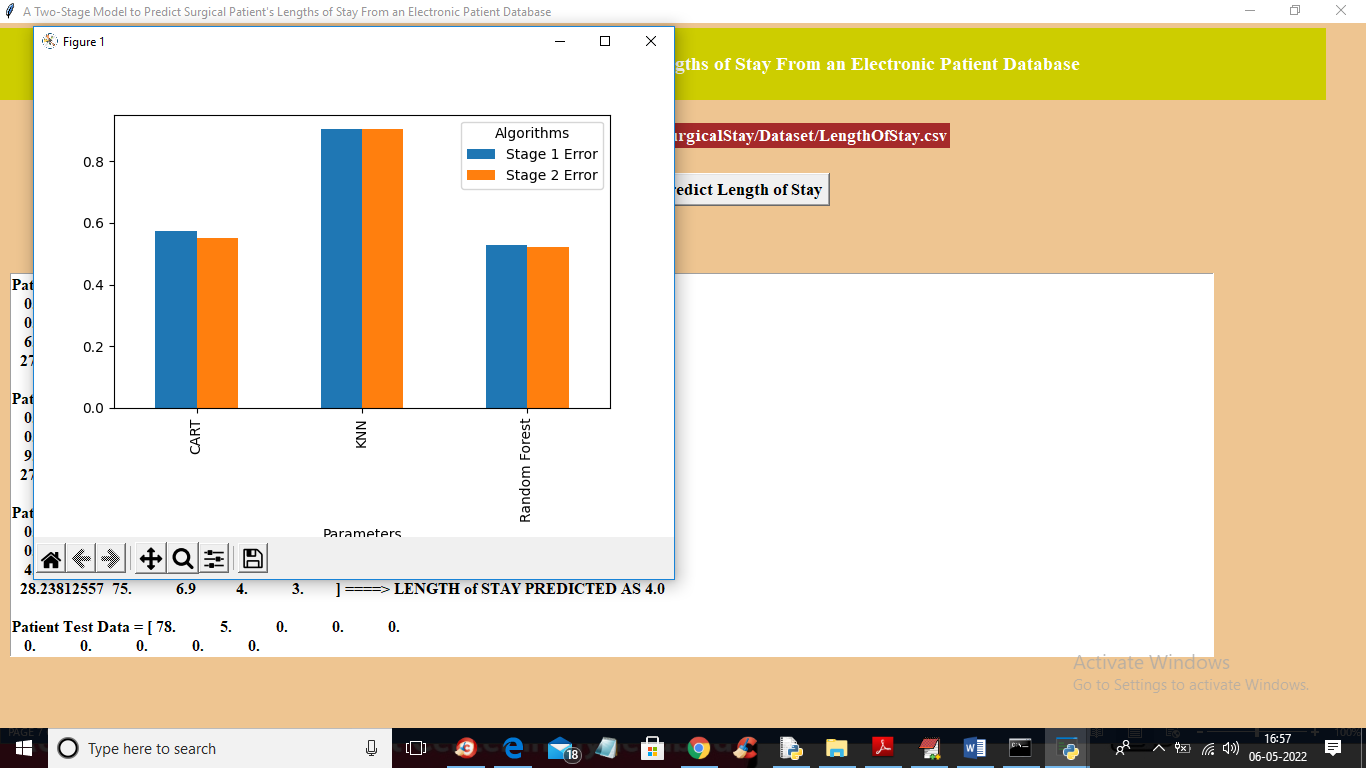
In above screen in first table we can see random forest error rate is less compare to propose CART. Now click on ‘Predict Length of Stay’ button to upload test data like below screen



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



In above screen in square bracket you can see the patient test data and after arrow symbol you can see length of stay as 7 days and similarly for each test record you can see predicted length of days. Now click on ‘CV Error Graph’ button to get cross validation error of each algorithm



In above graph x-axis represents algorithm name and y-axis represents error rate for each stage 1 and 2 and in all 3 algorithms random forest error rate is less.