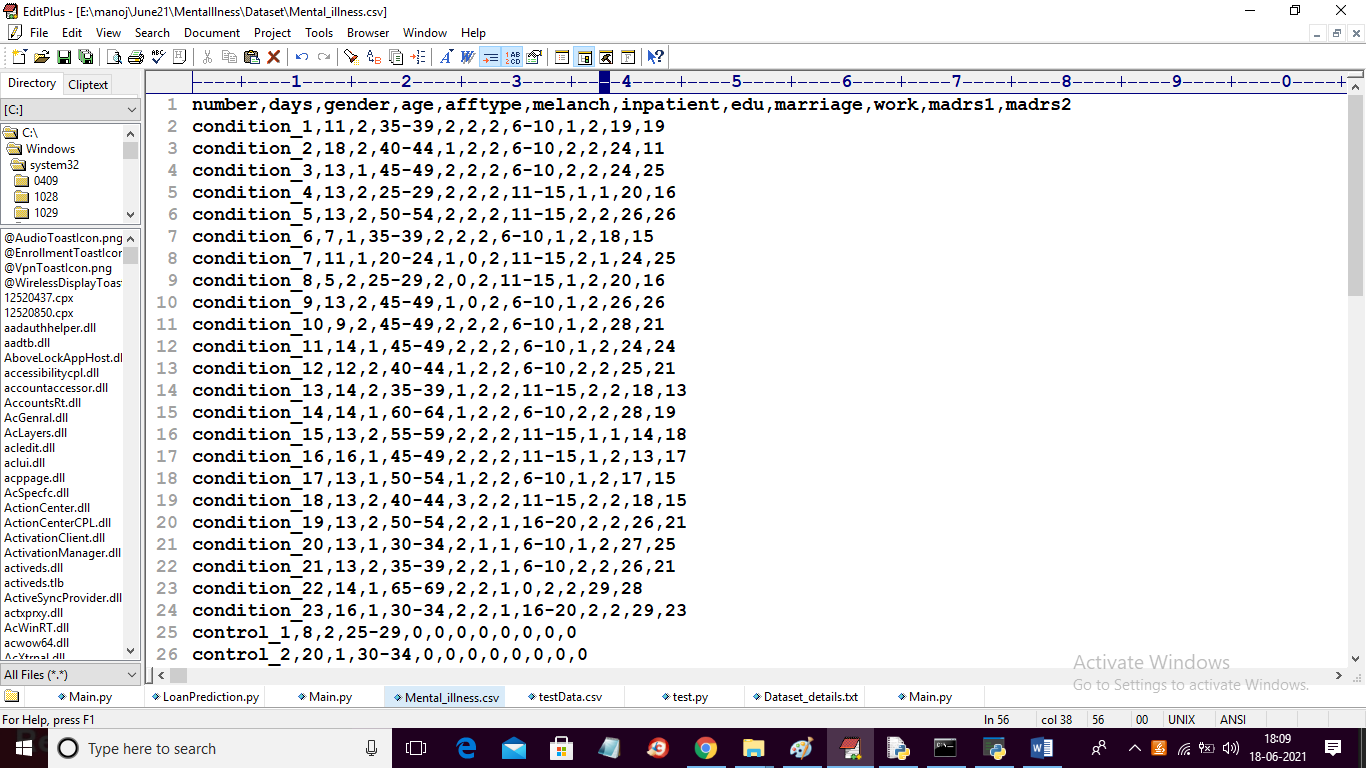
Data Analytics in Mental Healthcare

This paper describing concept to detect cause of mental condition such as personality disorder, bipolar disorder or depression order by applying Machine Learning Algorithms. In this project we are using Random Forest machine learning algorithm to predict mental disorder and this algorithm will be trained with previous patient history dataset and then build a trained model and this trained model will be applied on new patient data to predict their mental disorder. Random Forest trained model will used Artificial Intelligence technique to detect mental disorder and this technique works similar to human brain as human brain will take decision based on current environment and similarly random forest trained model will predict mental disorder by analysing his medical history data.

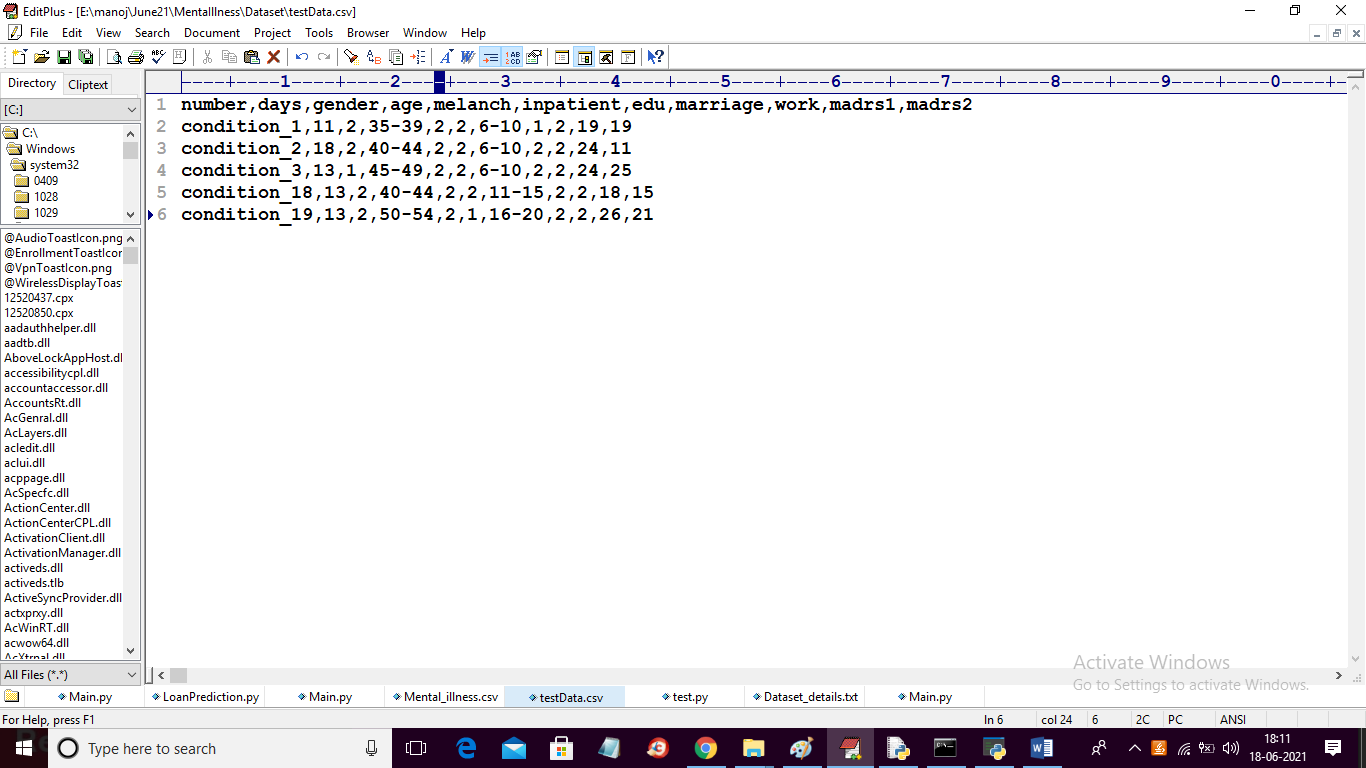
To train machine learning model we are using below dataset



About dataset details you can read below para

Dataset contains the following columns; number (patient identifier), days (number of days of measurements), gender (1 or 2 for female or male), age (age in age groups), afftype (1: bipolar II, 2: unipolar depressive, 3: bipolar I), melanch (1: melancholia, 2: no melancholia), inpatient (1: inpatient, 2: outpatient), edu (education grouped in years), marriage (1: married or cohabiting, 2: single), work (1: working or studying, 2: unemployed/sick leave/pension), madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped).

In above dataset ‘afftype’ column represents mental disorder and ML will get trained on above dataset and below is the new Patient Test Data



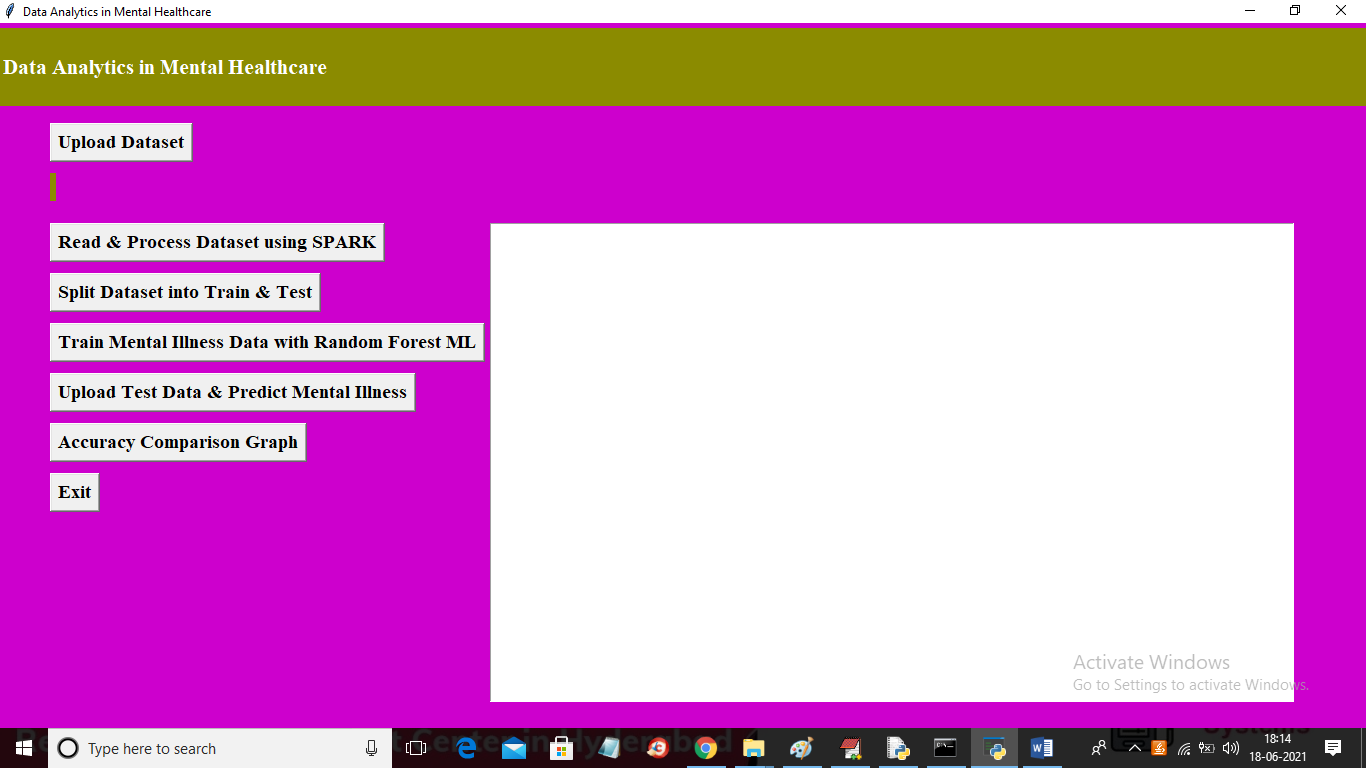
In above test data we have patient medical record but we don’t have ‘afftype’ mental disorder column and when we applied ML model on above test data then ML will predict mental disorder ‘afftype’ column.

To implement this project we have used BIG DATA SPARK technology which will read and process dataset faster. This project consists of following modules

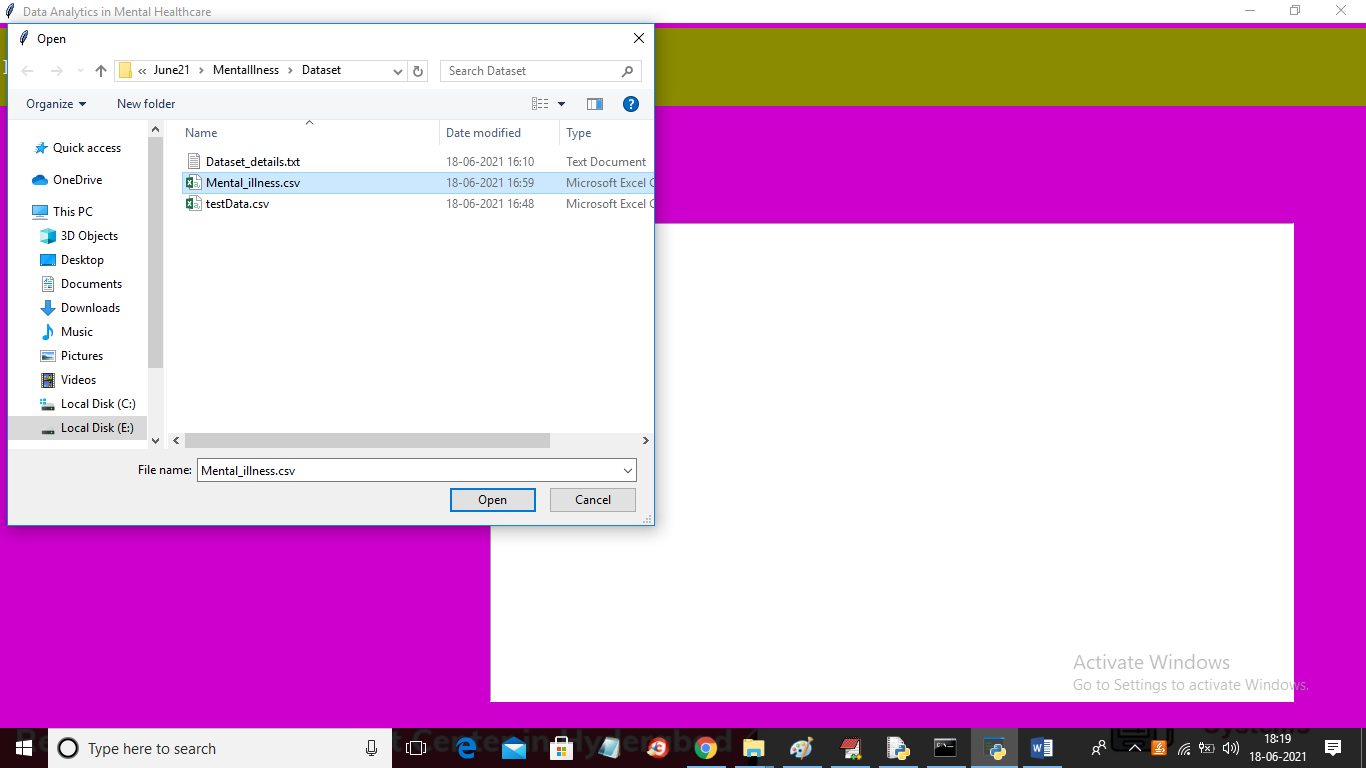
1. Upload Dataset: Using this module we will upload dataset to application
2. Read & Process Dataset using SPARK: using this module spark will read and process dataset and then convert non-numeric age and education data to numeric format
3. Split Dataset into Train & Test: using this module we will split dataset into train and test where application used 80% dataset for training and 20% dataset for testing
4. Train Mental Illness Data with Random Forest ML: Using this module we will train above dataset with Random Forest Machine Learning algorithm and then calculate its prediction accuracy and error rate on test data
5. Upload Test Data & Predict Mental Illness: using this module we will upload test data and then trained ML model will predict mental disorder for that test data
6. Accuracy Comparison Graph: using this module we will plot accuracy and error rate of random forest ML algorithm

SCREEN SHOTS

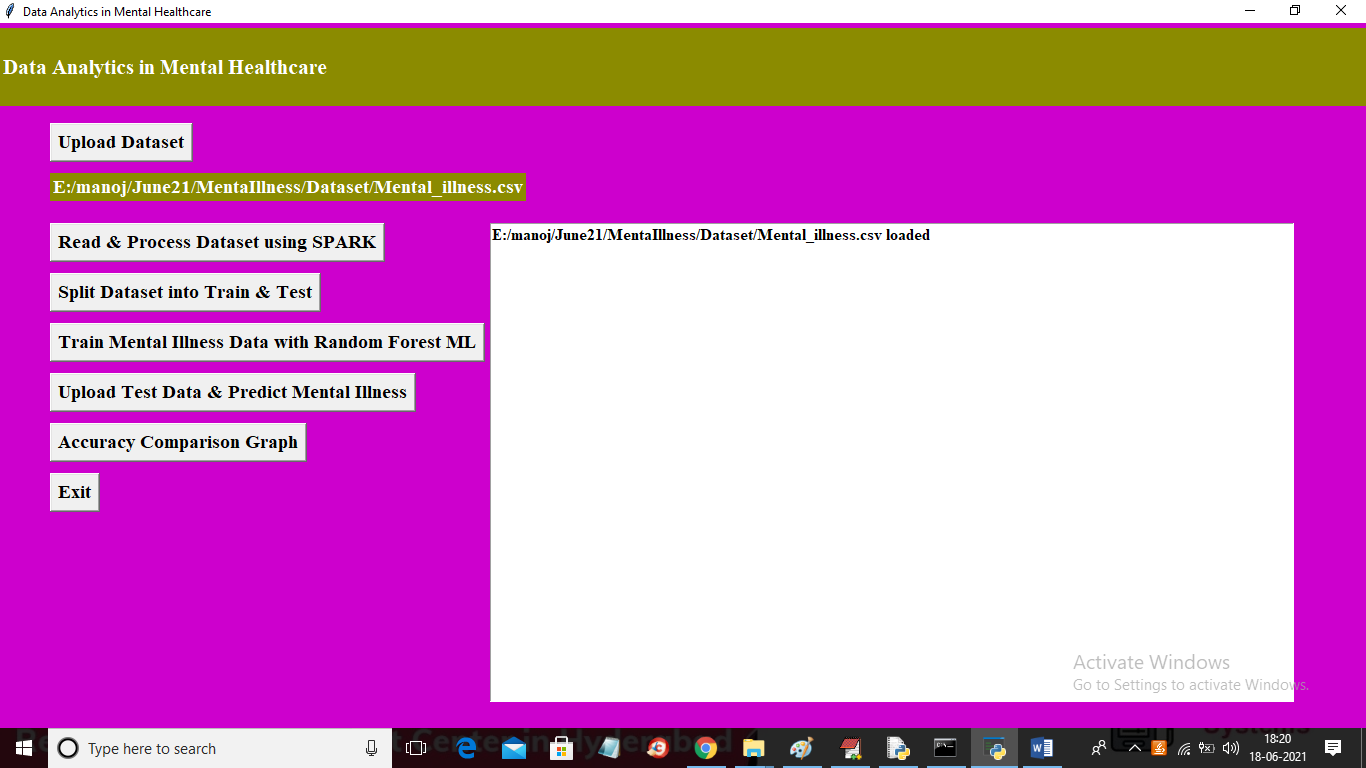
To run project double click on ‘run.bat’ file to get below screen



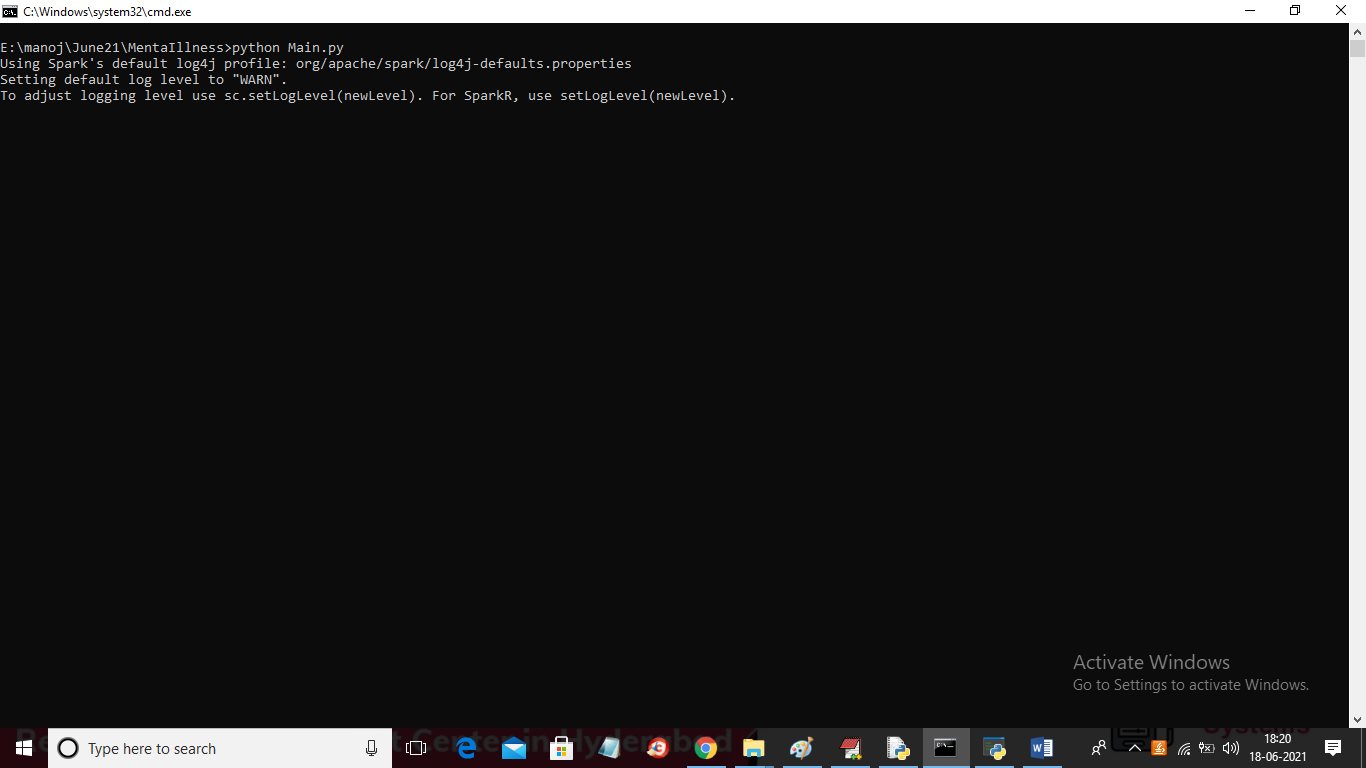
In above screen click on ‘Upload Dataset’ button to upload dataset and to get below screen



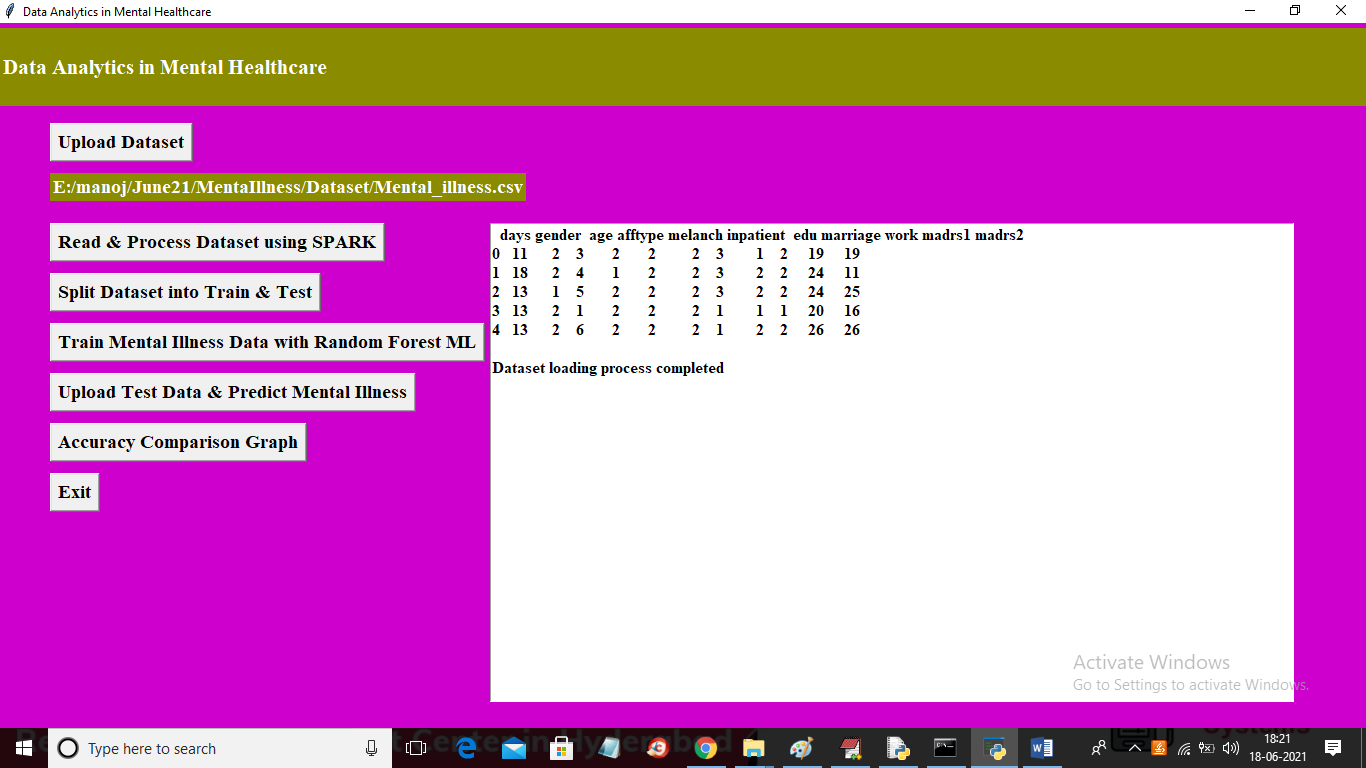
In above screen selecting and uploading ‘Mental\_illness.csv’ dataset file and then click on ‘Open’ button to load dataset and to get below screen



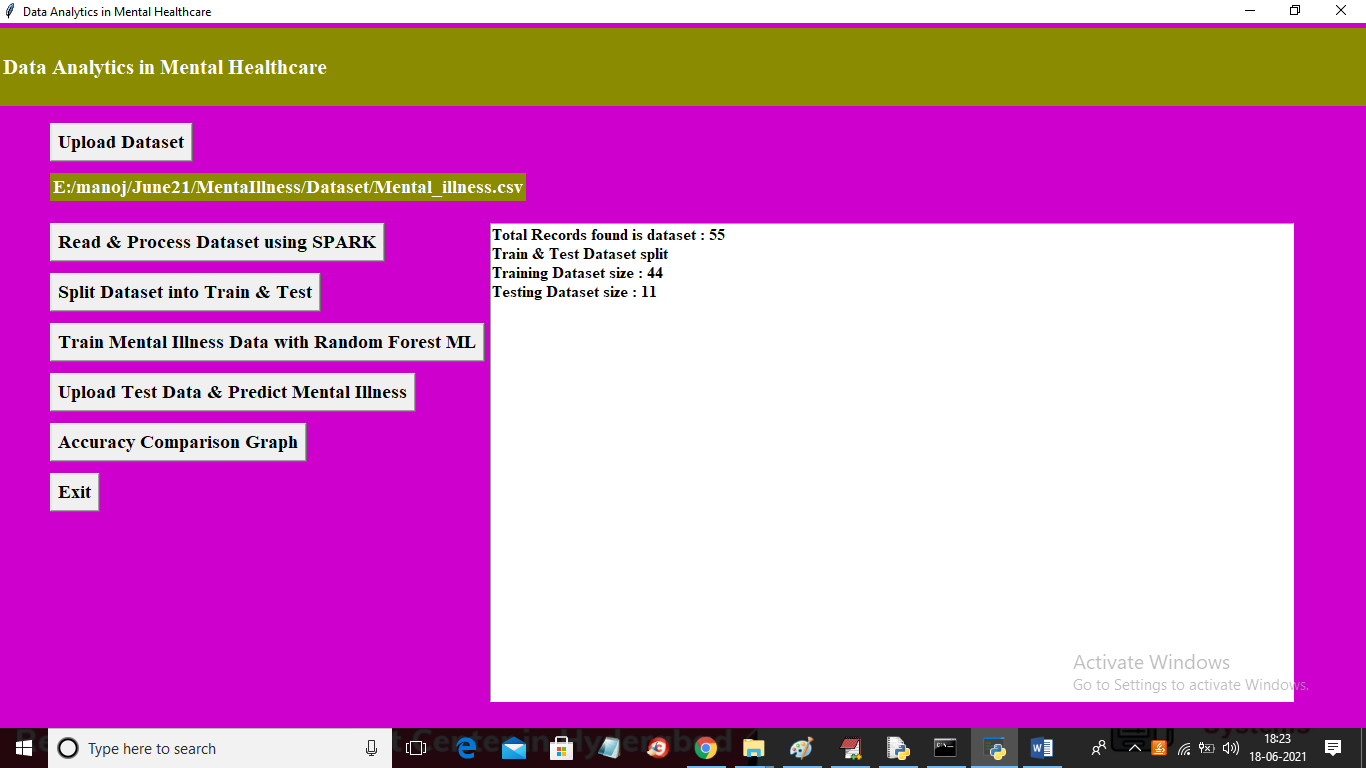
In above screen dataset loaded and now click on ‘Read & Process Dataset using SPARK’ button to read dataset using SPARK python API



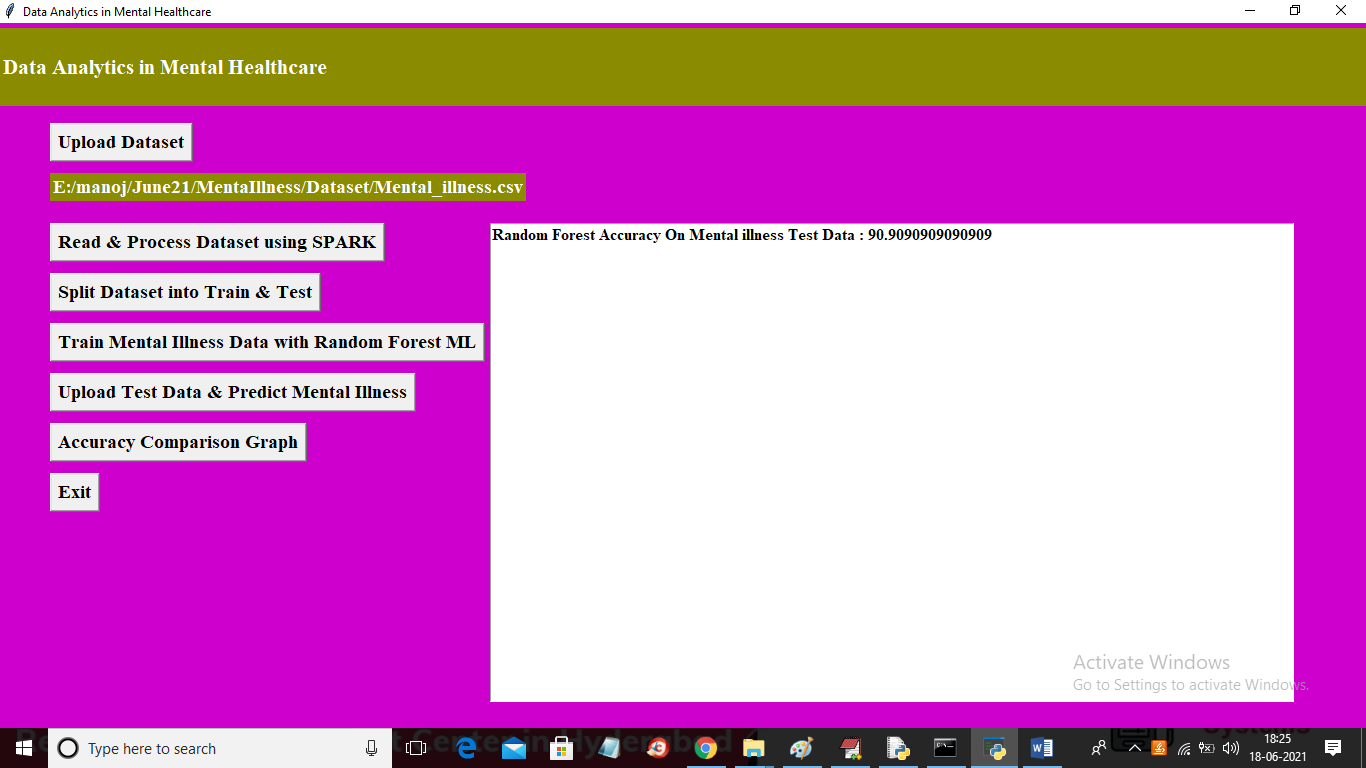
In above screen we can see python using SPARK API to read dataset and after reading dataset will get below screen



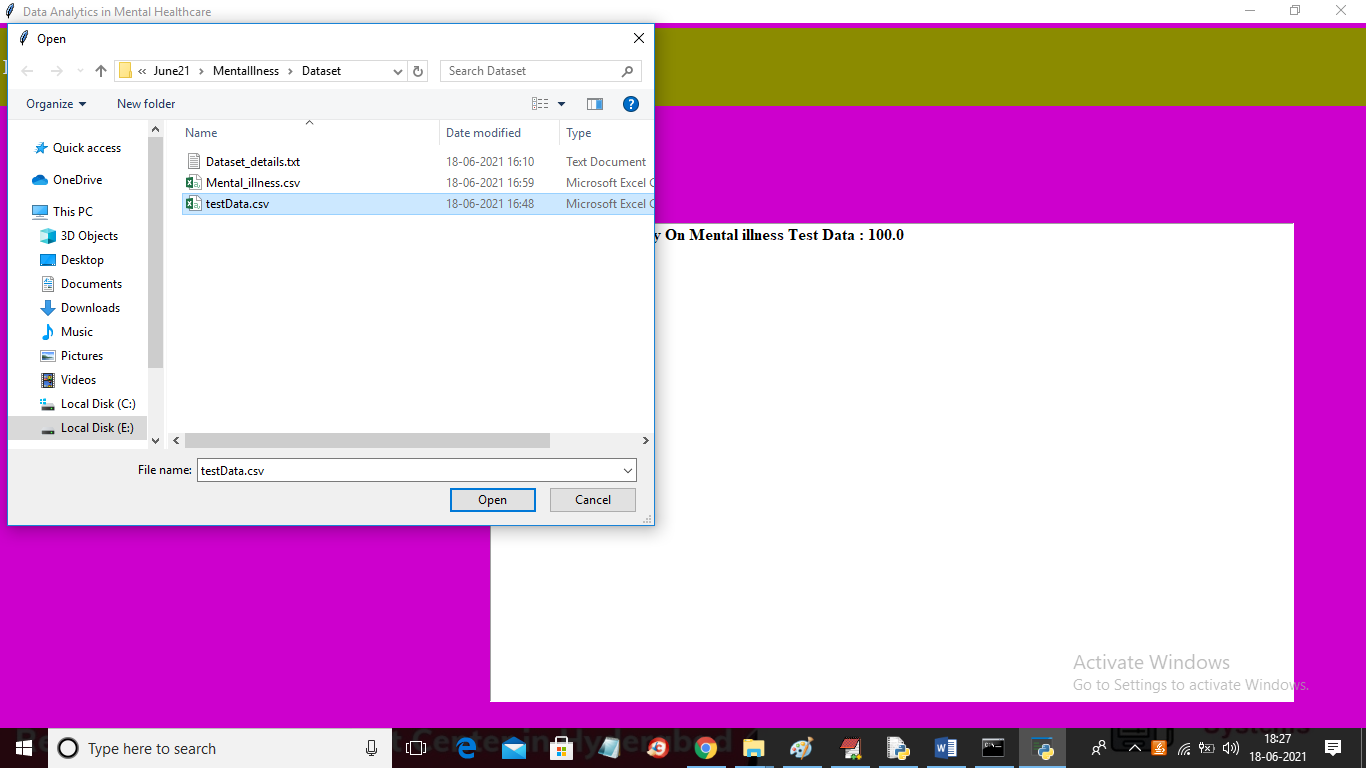
In above screen SPARK read entire dataset and we can see AGE and EDUCATION column is converted to numeric as in dataset age values is in group (35-40) but SPARK has assigned using code to age group to convert that value to numeric as machine learning algorithms accept only numeric values. Now in above screen dataset is ready and now click on ‘Split Dataset into Train & Test’ button to divide dataset into train and test part



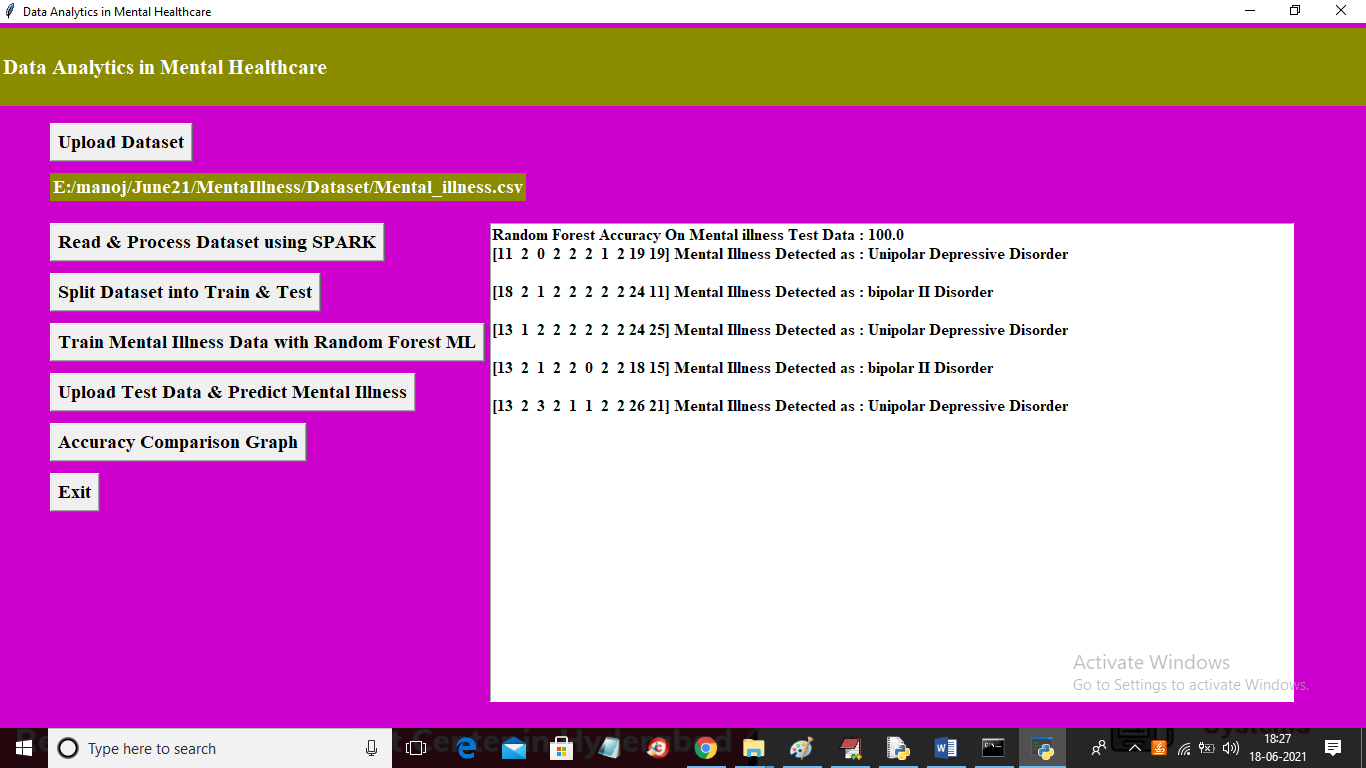
In above screen we can see dataset divide into train and test part where for training application using 44 records and for testing using 11 records and now train and test data is ready and now click on ‘Train Mental Illness Data with Random Forest ML’ button to train random forest with above dataset



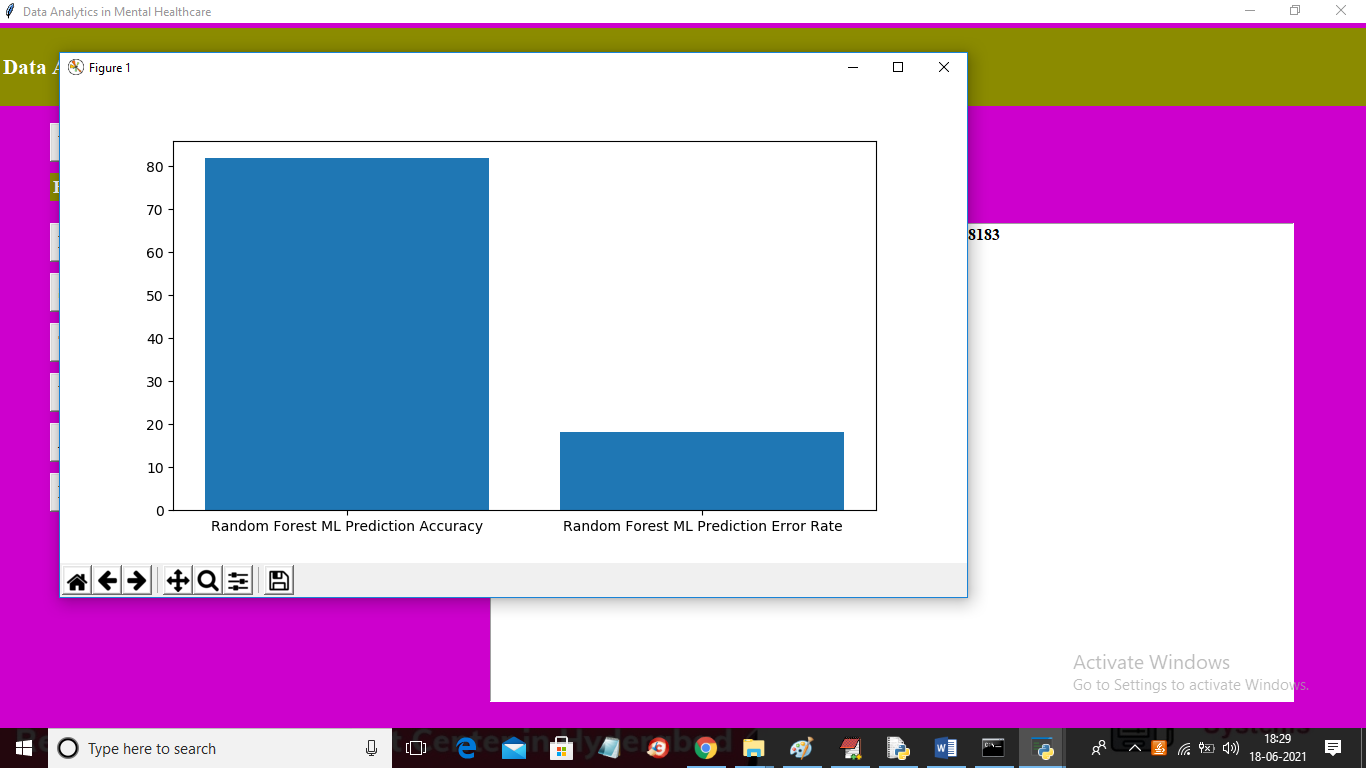
In above screen with random forest we got 90% accuracy and now click on ‘Upload Test Data & Predict Mental Illness’ button to upload test data



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



In above screen in square bracket we can see patient medical data and based on medical dataset random forest ML has predicted illness as BIPOLAR, UNIPOLAR depression etc. Now click on ‘Accuracy Comparison Graph’ button to get below graph



In above graph x-axis represents random forest accuracy and error rate and y-axis represents its values and in above graph we can see correctly prediction accuracy is higher than error rate so ML model is accurate in prediction