SUICIDAL Tendency Detection

Peoples are taking lots of stress due to today's competitive environment in almost all fields like working area and educational area diverting peoples into stress which sometime causes them to take suicide steps. To overcome from this problem author of this paper using PATIENTS dataset to train machine learning algorithms and then this trained model can take patient current status values as input and then predict whether that patient is having any suicidal thoughts. If suicidal thoughts detected then doctors will take necessary steps to help patients in recovering from those thoughts.

To implement this project author has used KAGGLE suicidal attempt and stress dataset and then train this dataset with deep learning algorithm called CNN. This dataset contains NON-NUMERIC characters which will not understand by MACHINE LEARNING algorithms so author applying Neural Machine Translation algorithm called Natural Language Processing (NLP) to translate all non-numeric characters to numeric characters which will understand by machine learning algorithms.

In this paper we have trained existing Random Forest and propose CNN algorithm to train dataset and then evaluate both algorithms performance in terms of accuracy, precision, recall and FSCORE.

Note: u r asking to record voice and facial expression to detect depression or suicidal tendency but we don't any senor or devices to record so things so we are using depression and suicidal dataset from KAGGLE which contains columns to detect depression. From facial expression suicide depression can be detected as patient can be depress for some other reason not for suicide so we need to used accurate suicide or depression related dataset to detect such tendency so we choose below dataset

Below is the dataset screen used in this project

In above dataset first row contains dataset column names and remaining rows contains dataset values and in above dataset we can see some values are numeric and some are non-numeric and this non-numeric characters will be translate to numeric format by using NLP technique. NLP will assign numeric ID to each unique non-numeric characters and this ID's will be used to train ML algorithms. In below screen we can see dataset contains 'suicidal_attempt' and 'stressed' column.
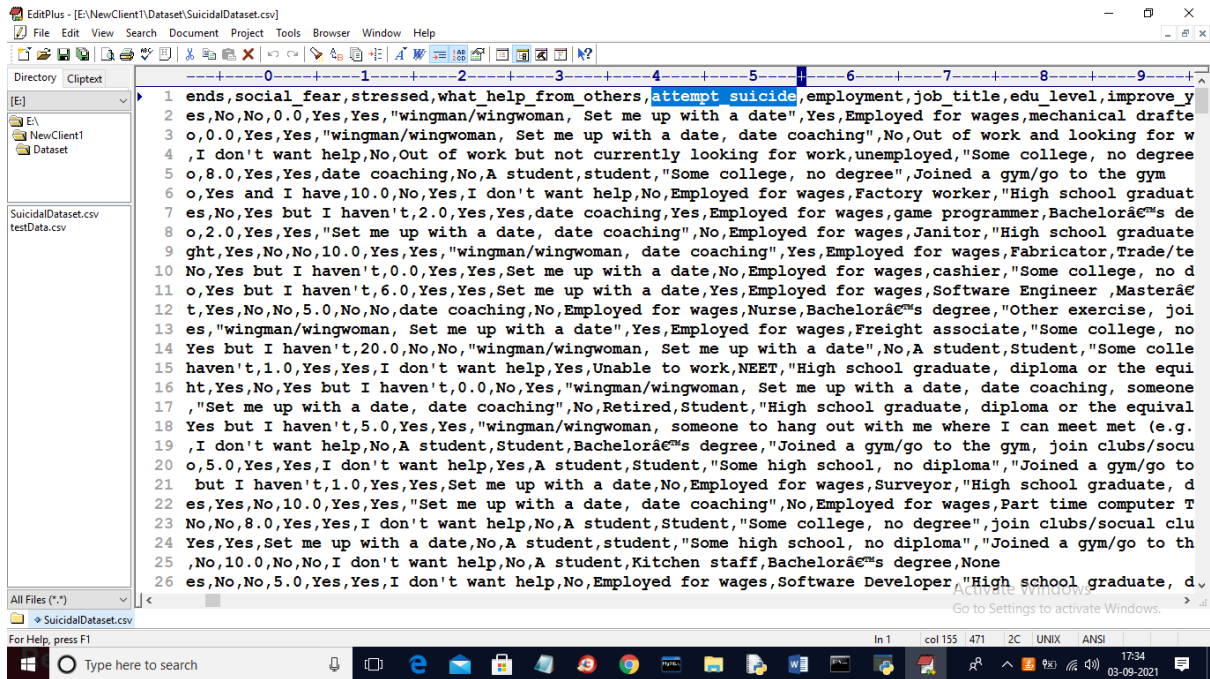
In above dataset screen you can see dataset contains 'attempt_suicide' and 'stressed' column. We will used above dataset to train ML algorithms.

SCREEN SHOTS

To run project double click on 'run.bat' file to get below screen
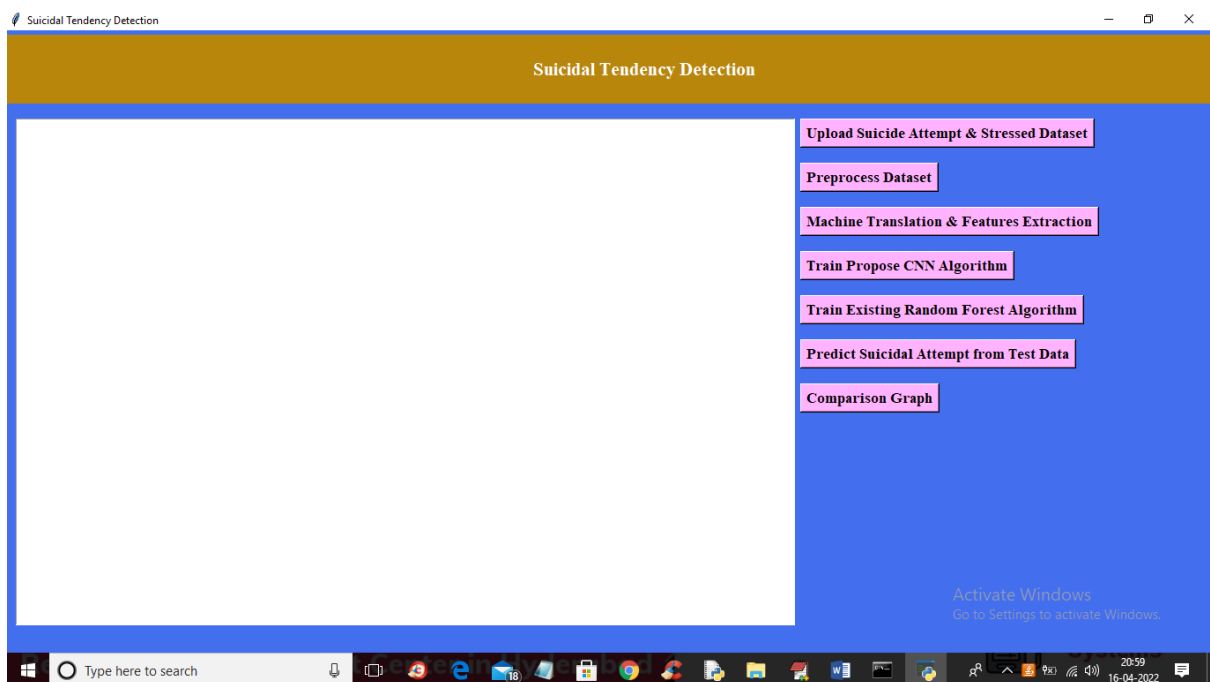
In above screen click on 'Upload Suicide Attempt & Stressed Dataset' button to upload dataset and to get below screen
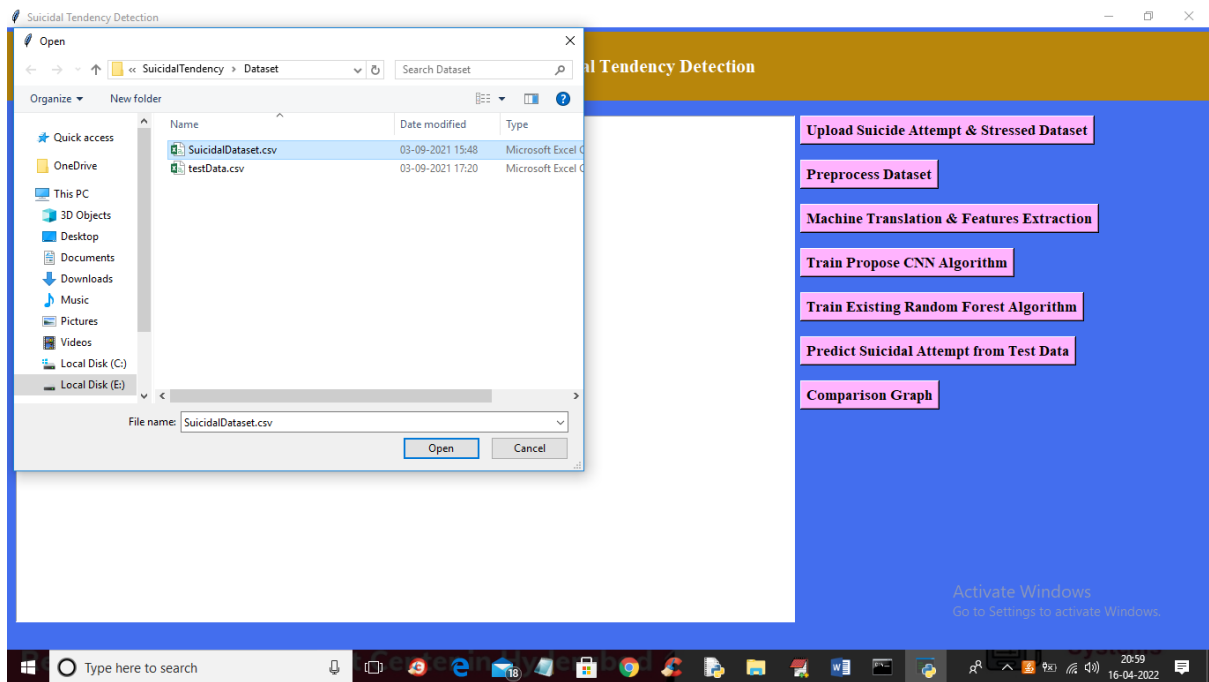
In above screen selecting and uploading 'Suicidal' dataset and then click on 'Open' button to load dataset and to get below screen
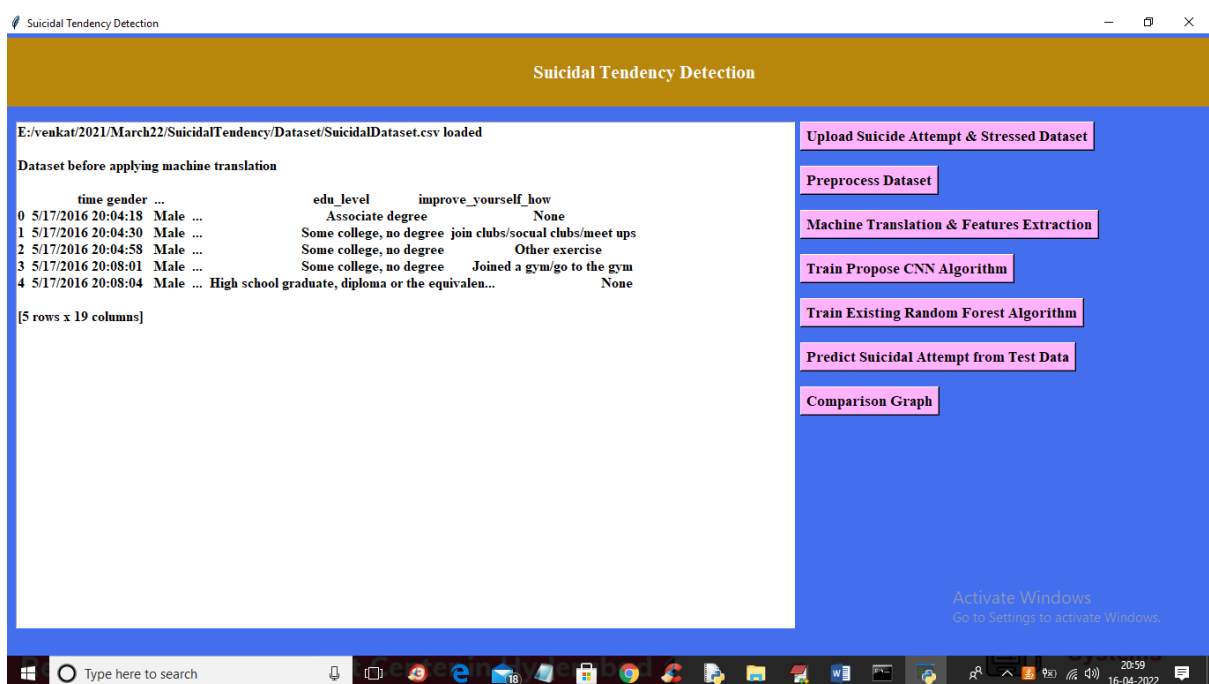
In above screen dataset loaded and we can see some records from dataset and dataset contains some non-numeric characters and to translate them first click on 'Preprocess Dataset' button to remove missing values and then replace with 0
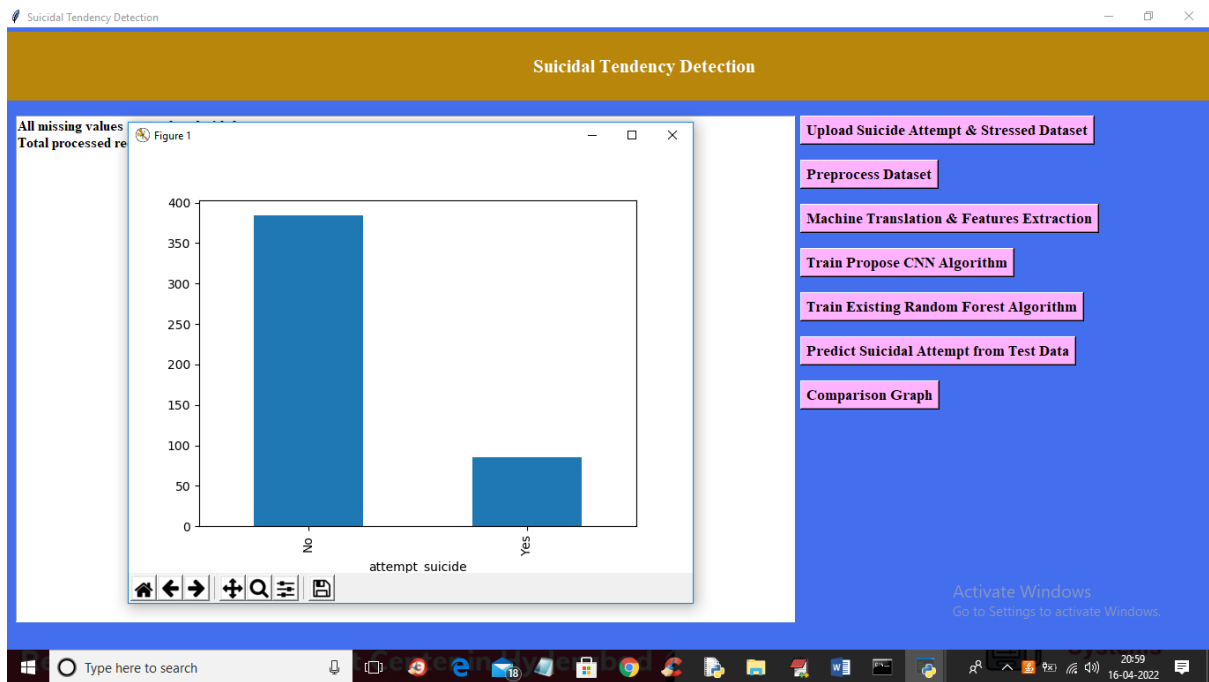
In above screen we can see all missing data is replaced with 0 and we can see dataset contains total 469 records. In graph we can see total patients with and without suicidal thought. In above graph X-axis represents YES and NO values and y-axis represents total counts of YES and NO patients. YES means patients has suicidal thoughts and NO means patients has no suicidal thoughts. Now close above graph and then click on 'Machine Translation & Features Extraction' button to translate all dataset NON-NUMERIC features to NUMERIC features.

In above screen we can see complete dataset is translated to numeric data and in below two lines we can see dataset using 614 records to train CNN algorithms and using 154 records for testing CNN performance. Now train and test data is ready and now click on 'Train Propose CNN Algorithm' button to train CNN with above dataset and to get below output
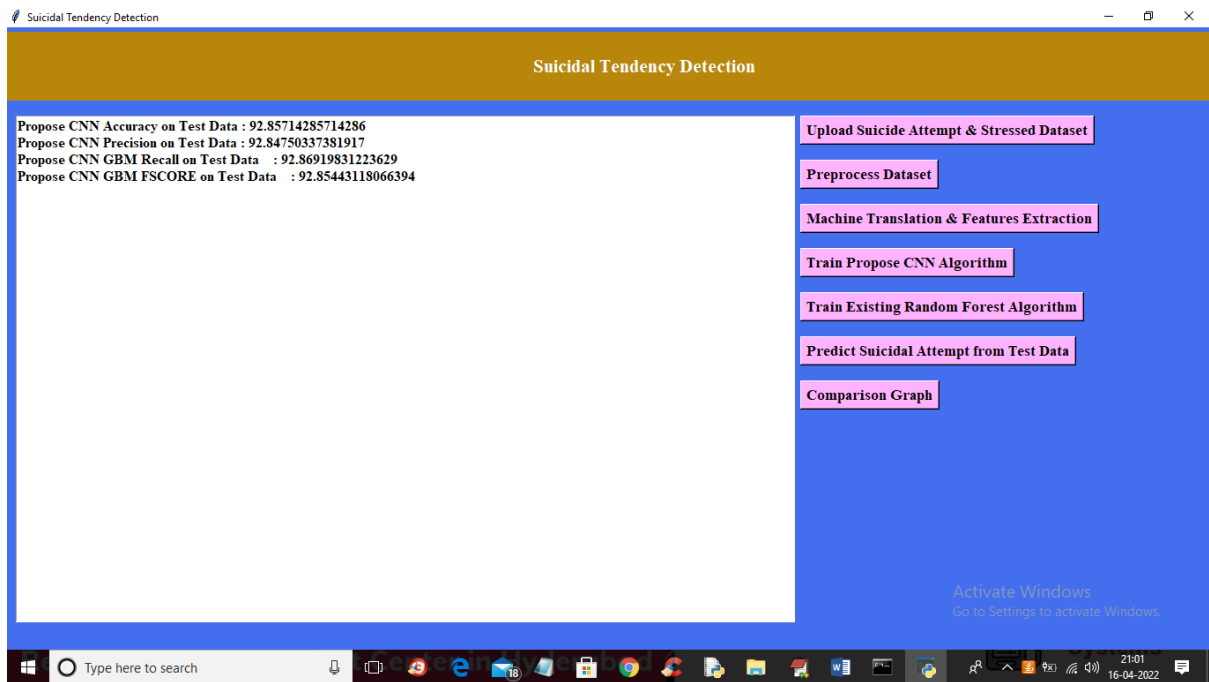
In above screen we can see with CNN we got 92% accuracy and now click on 'Train Existing Random Forest Algorithm' button to train existing Random Forest algorithm on same data and calculate accuracy
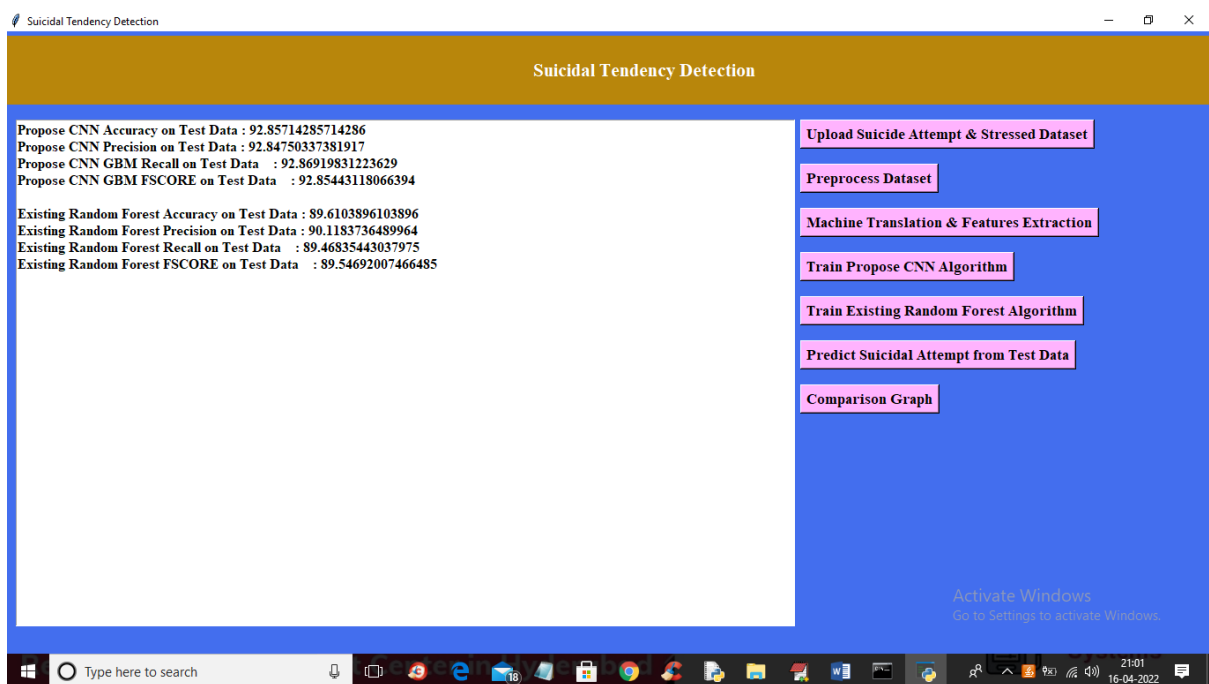
In above screen with existing random forest algorithm we got 89% accuracy and now click on 'Predict Suicidal Attempt from Test Data' button to upload test data and then CNN will predict whether test patient records has any suicidal and NO suicidal thoughts.

In above screen we are 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 brackets we can see patient test records values and after arrow symbol '====>' we can see prediction result from CNN as whether patient has suicidal thoughts or NOT. You can scroll down above screen to get all predicted records like below screen

Now click on 'Comparison Graph' button to get below graph

In above graph x-axis represents algorithm names and y-axis represents accuracy, precision, recall and FSCORE in different colour bars. In above graph we can see CNN is performing well compare to existing Random Forest algorithm.