

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

EdiPlus - [E:\NewClient1\Dataset\SuicidalDataset.csv]

File Edit View Search Document Project Tools Browser Window Help

Directory Cliptext

[E:]

E:\NewClient1Dataset

SuicidalDataset.csvtestData.csv

All Files (*.*)

SuicidalDataset.csv

For Help, Press F1

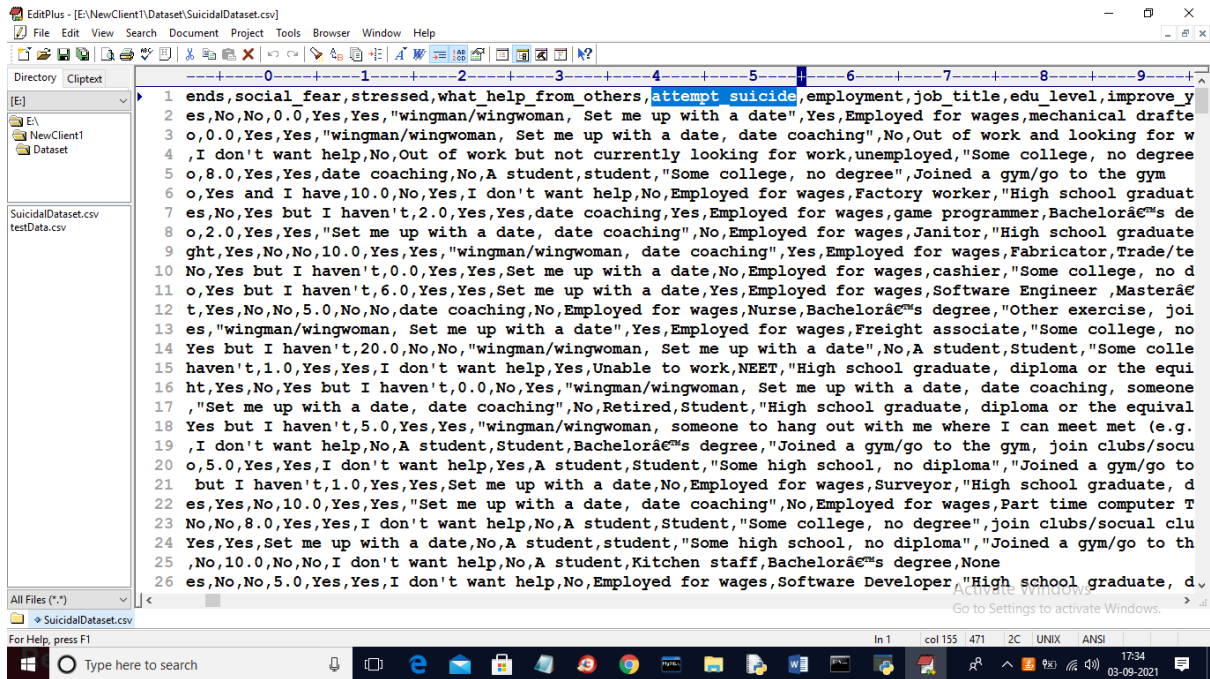
	1	2	3	4	5	6	7	8	9	10					
1	time	gender	sexuality	age	income	race	bodyweight	virgin	prostitution	legal	pay	sex	friends	social	f
2	5/17/2016	20:04:18	Male	Straight	35	"\$30,000 to \$39,999"	White non-Hispanic	Normal weight	Yes	No	No	0.0	Yes	Yes	
3	5/17/2016	20:04:30	Male	Bisexual	21	"\$1 to \$10,000"	White non-Hispanic	Underweight	Yes	No	No	0.0	Yes	Yes	
4	5/17/2016	20:04:58	Male	Straight	22	\$0	White non-Hispanic	Overweight	Yes	No	No	10.0	Yes	Yes	I don't want
5	5/17/2016	20:08:01	Male	Straight	19	"\$1 to \$10,000"	White non-Hispanic	Overweight	Yes	Yes	No	8.0	Yes	Yes	
6	5/17/2016	20:08:04	Male	Straight	23	"\$30,000 to \$39,999"	White non-Hispanic	Overweight	No	No	Yes	and I h			
7	5/17/2016	20:09:09	Male	Straight	24	"\$50,000 to \$74,999"	White non-Hispanic	Normal weight	Yes	No	Yes	but			
8	5/17/2016	20:10:56	Male	Straight	22	"\$1 to \$10,000"	White non-Hispanic	Underweight	Yes	No	No	2.0	Yes	Yes	
9	5/17/2016	20:11:13	Female	Gay/Lesbian	24	"\$20,000 to \$29,999"	White non-Hispanic	Normal weight	Yes	No	No				
10	5/17/2016	20:11:52	Male	Straight	20	"\$10,000 to \$19,999"	White non-Hispanic	Overweight	Yes	No	Yes	but I			
11	5/17/2016	20:13:37	Male	Straight	33	"\$50,000 to \$74,999"	White non-Hispanic	Overweight	No	No	Yes	but I h			
12	5/17/2016	20:17:40	Female	Straight	32	"\$50,000 to \$74,999"	Hispanic (of any race)	Overweight	Yes	No	No	5			
13	5/17/2016	20:18:37	Male	Straight	25	"\$1 to \$10,000"	Asian	Normal weight	Yes	No	No	0.0	Yes	Yes	wingman/w
14	5/17/2016	20:26:57	Male	Straight	24	"\$1 to \$10,000"	White non-Hispanic	Normal weight	No	No	Yes	but I hav			
15	5/17/2016	20:30:21	Female	Straight	23	\$0	White non-Hispanic	Normal weight	Yes	No	Yes	but I haven't	1.0	Y	
16	5/17/2016	20:30:26	Male	Straight	22	"\$30,000 to \$39,999"	Hispanic (of any race)	Normal weight	Yes	No	Yes				
17	5/17/2016	20:30:49	Male	Straight	23	\$0	White non-Hispanic	Underweight	Yes	Yes	No	6.0	No	Yes	"Set me up w
18	5/17/2016	20:32:26	Female	Straight	29	"\$1 to \$10,000"	white and asian	Normal weight	Yes	No	Yes	but I hav			
19	5/17/2016	20:34:35	Male	Straight	21	"\$1 to \$10,000"	Asian	Normal weight	No	No	No	12.0	No	No	I don't want
20	5/17/2016	20:39:14	Male	Straight	17	"\$1 to \$10,000"	White non-Hispanic	Overweight	Yes	Yes	No	5.0	Yes	Yes	
21	5/17/2016	20:40:22	Male	Straight	29	"\$75,000 to \$99,999"	caucasian	Normal weight	Yes	No	Yes	but I haven'			
22	5/17/2016	20:41:29	Male	Straight	32	"\$10,000 to \$19,999"	White non-Hispanic	Normal weight	Yes	Yes	No	10.0			
23	5/17/2016	20:42:09	Female	Straight	20	"\$1 to \$10,000"	White non-Hispanic	Normal weight	Yes	No	No	8.0	Yes	Yes	Set m
24	5/17/2016	20:42:10	Male	Straight	24	\$0	Hispanic (of any race)	Normal weight	Yes	No	No	40.0	Yes	Yes	
25	5/17/2016	20:44:36	Male	Straight	20	"\$1 to \$10,000"	White non-Hispanic	Normal weight	Yes	No	No	10.0	No	N	
26	5/17/2016	20:46:31	Male	Straight	21	"\$20,000 to \$29,999"	White non-Hispanic	Normal weight	Yes	No	No	5.0			

Go to Settings to activate Windows.

In 1 col 1 471 74 UNIX ANSI

17:30 03-09-2021

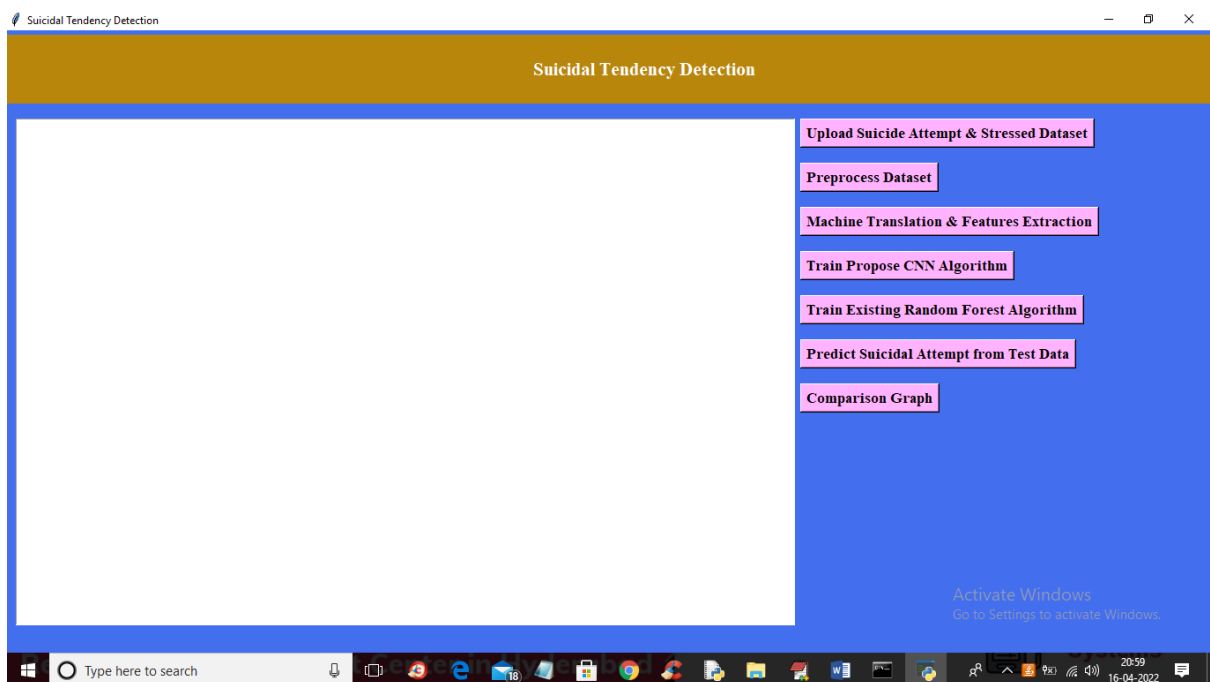
In above dataset first row contains dataset column names and remaining rows contains dataset values and in above dataset we can see some values are numeric and some are non-numeric and this non-numeric characters will be translate to numeric format by using NLP technique. NLP will assign numeric ID to 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.



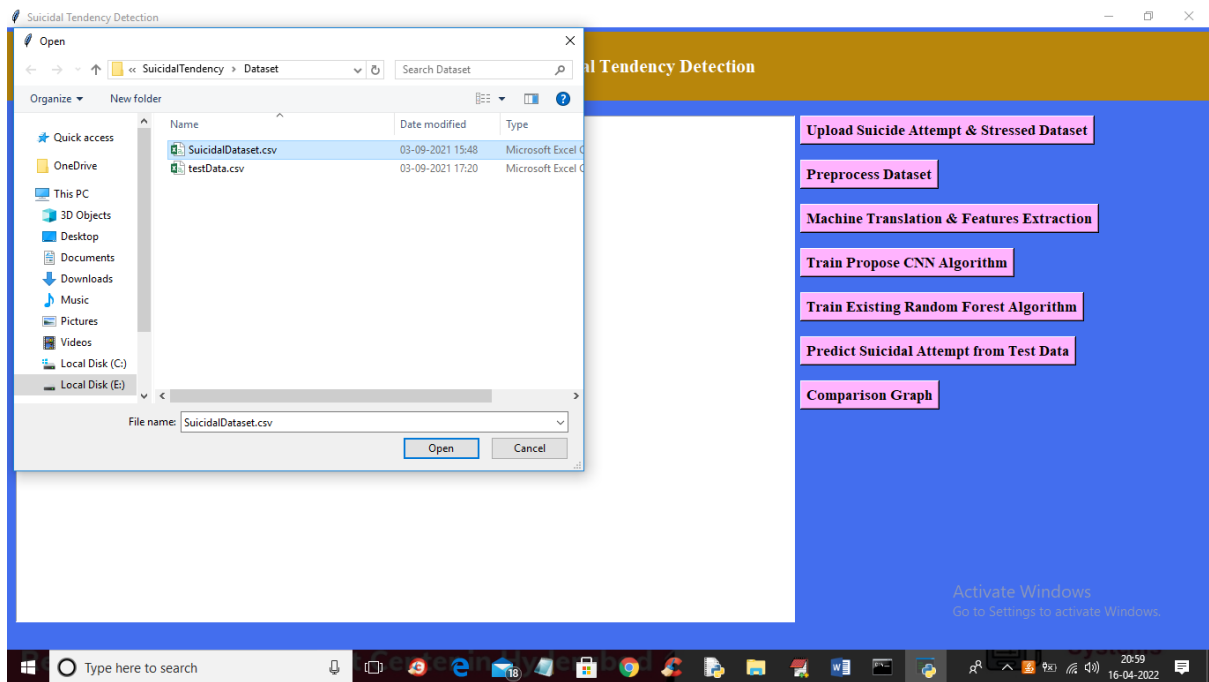
In above dataset screen you can see dataset contains 'attempt_suicide' and 'stressed' column. We will used above dataset to train ML algorithms.

SCREEN SHOTS

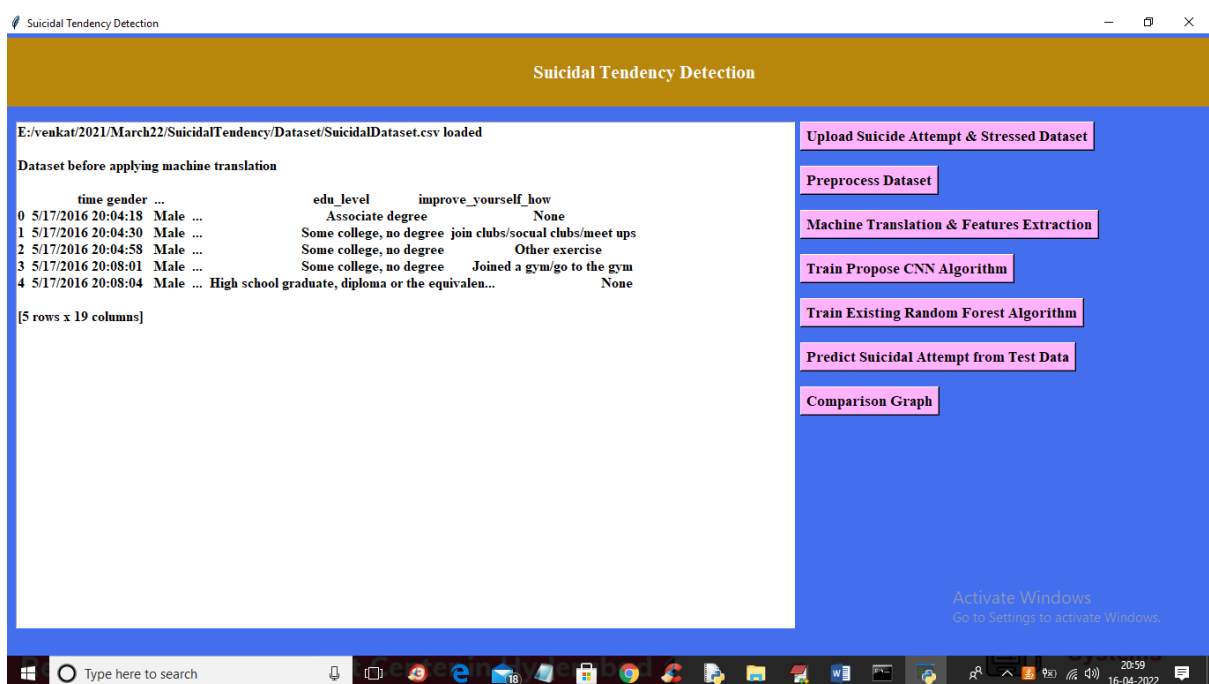
To run project double click on 'run.bat' file to get below screen



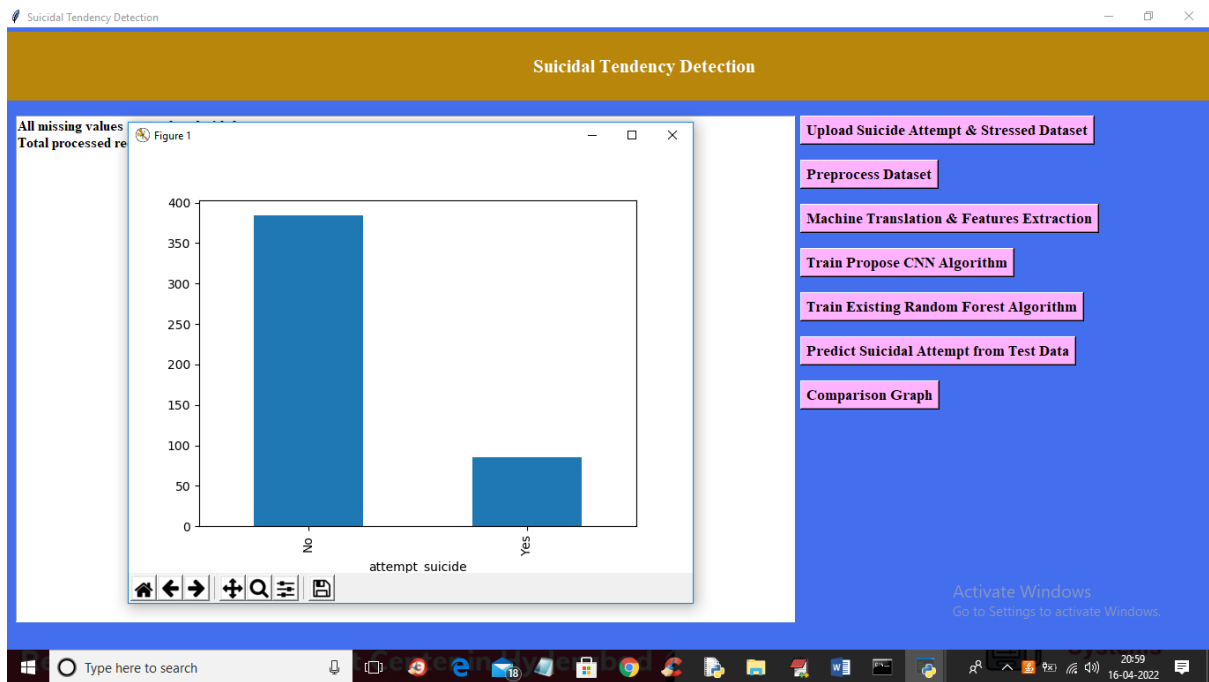
In above screen click on 'Upload Suicide Attempt & Stressed Dataset' button to upload dataset and to get below screen



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.

Suicidal Tendency Detection

Dataset after applying machine translation

	gender	sexuality	age	race	bodyweight	...	attempt_suicide	employment	job_title	edu_level	improve_yourself_how
0	1	2	35	18	0	...	1	2	243	0	33
1	1	0	21	18	3	...	0	4	4	6	58
2	1	2	22	18	2	...	0	5	272	6	36
3	1	2	19	18	2	...	0	1	263	6	15
4	1	2	23	18	2	...	0	2	62	3	33
...
464	0	2	26	18	2	...	0	2	43	4	15
465	1	2	31	18	0	...	0	2	219	3	43
466	0	2	17	0	0	...	1	2	255	7	20
467	1	2	18	0	1	...	0	5	172	3	21
468	1	2	28	18	0	...	0	2	156	4	48

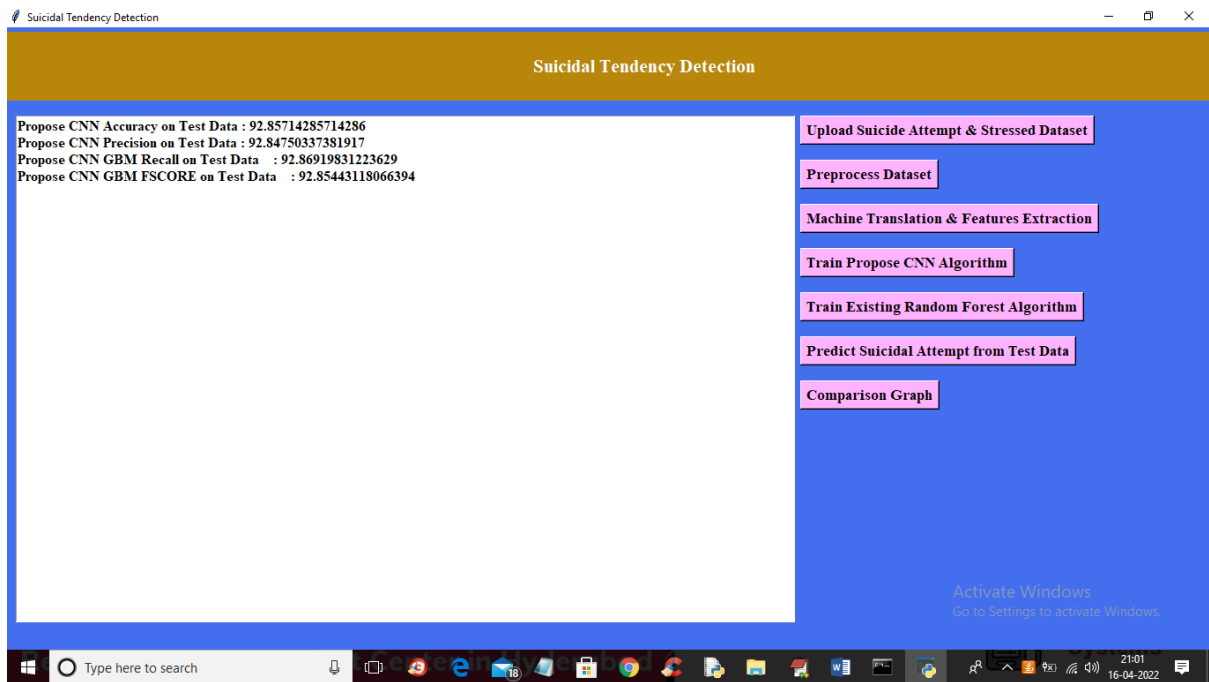
[469 rows x 17 columns]

Total records used to train machine learning LITE GBM Algorithm is : 614
Total records used to test machine learning LITE GBM Algorithm is : 154

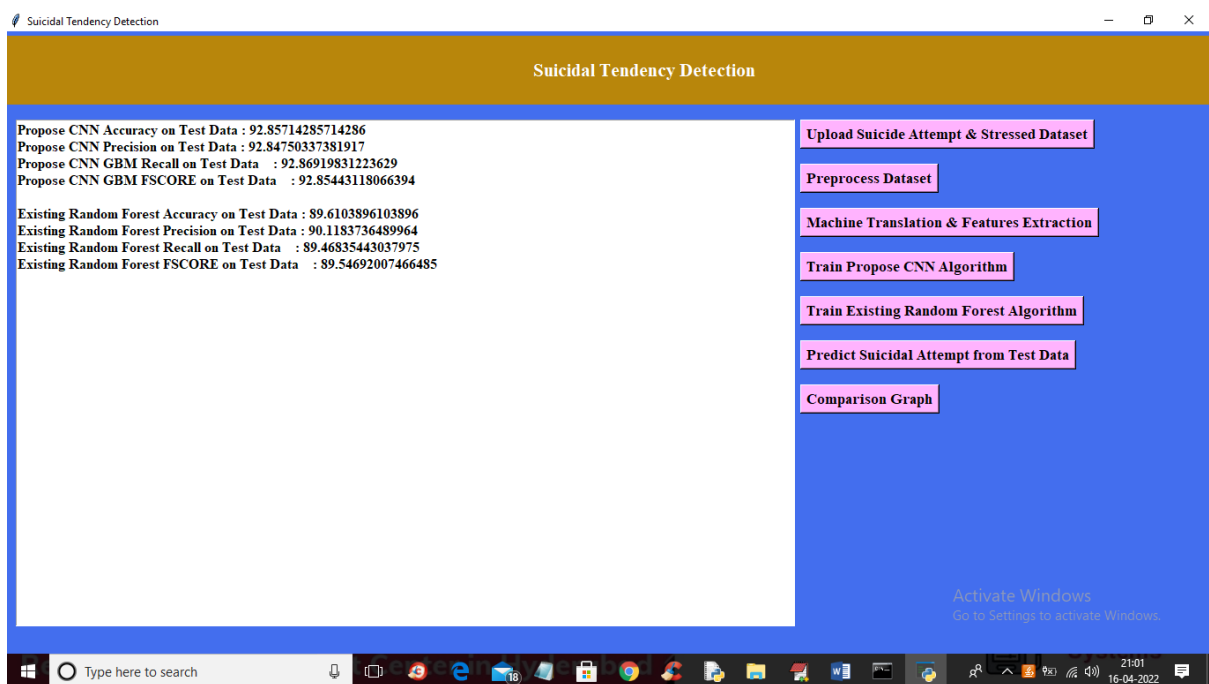
Upload Suicide Attempt & Stressed Dataset
Preprocess Dataset
Machine Translation & Features Extraction
Train Propose CNN Algorithm
Train Existing Random Forest Algorithm
Predict Suicidal Attempt from Test Data
Comparison Graph

Activate Windows
Go to Settings to activate Windows.

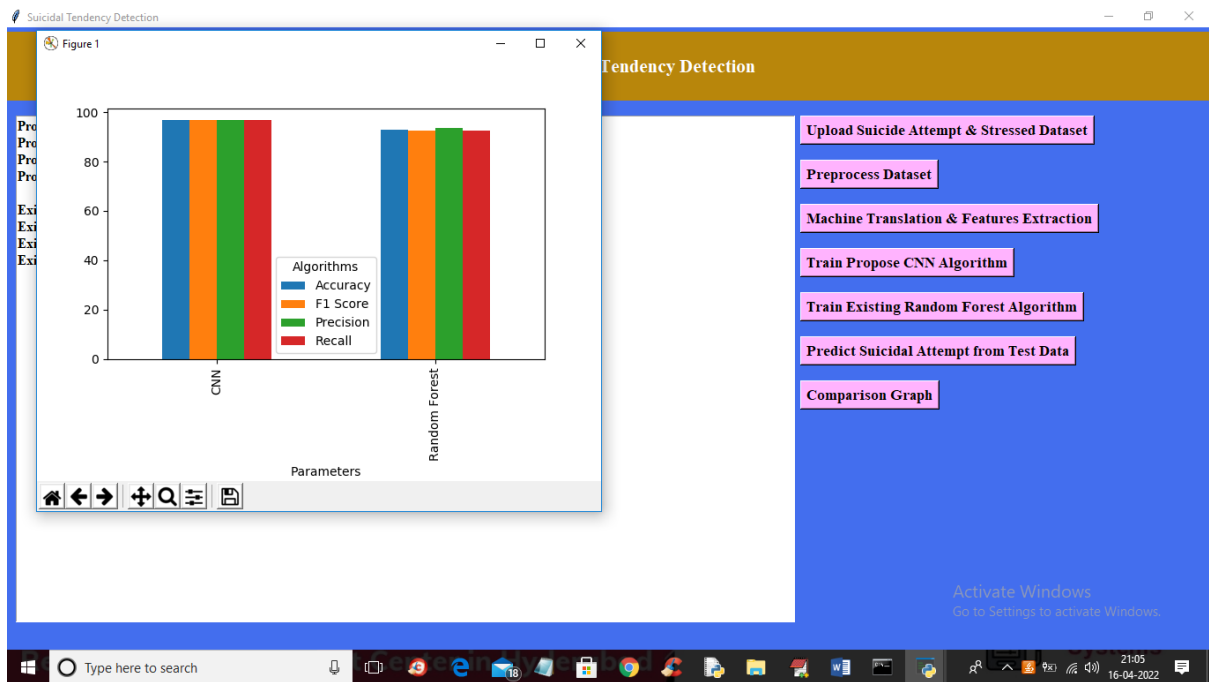
In above screen we can see complete dataset is translated to numeric data and in below two lines we can see dataset using 614 records to train CNN algorithms and using 154 records for testing CNN performance. Now train and test data is ready and now click on 'Train Propose CNN Algorithm' button to train 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 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.