

INTRODUCTION

Depression (major depressive disorder) is a common and serious medical illness that negatively affects how you feel, the way you think and how you act. Fortunately, it is also treatable.

Depression causes feelings of sadness and/or a loss of interest in activities you once enjoyed. It can lead to a variety of emotional and physical problems and can decrease your ability to function at work and at home.

Machine learning is a branch of artificial intelligence in computer science which addresses the use of data algorithms to emulate human learning gradually bettering its accuracy efficiency. Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans.

This project is built with this in mind. Machine learning algorithms take into account various factors from a person's life and use it to determine if they are under the effects of depression and compare the accuracies of different algorithms.

PROBLEM DEFINITION

- To predict if a person has depression or not based on the person's previous mental health disorders . By testing the data on various ML classification algorithms and checking the accuracy.
- 6 algorithms are tested on 3 datasets namely, DEPRESSION.csv, DEPRESSIONv4.csv and Depression Dataset.csv

CHAPTER 4 - PROJECT DESCRIPTION

The project is to help people understand if they have depression and to help mental care workers such as doctors, psychologists and therapists in predicting whether their patients have depression, to determine if the people are under the effects of depression and compare the accuracies of different algorithms. Three datasets and one real time data is used.

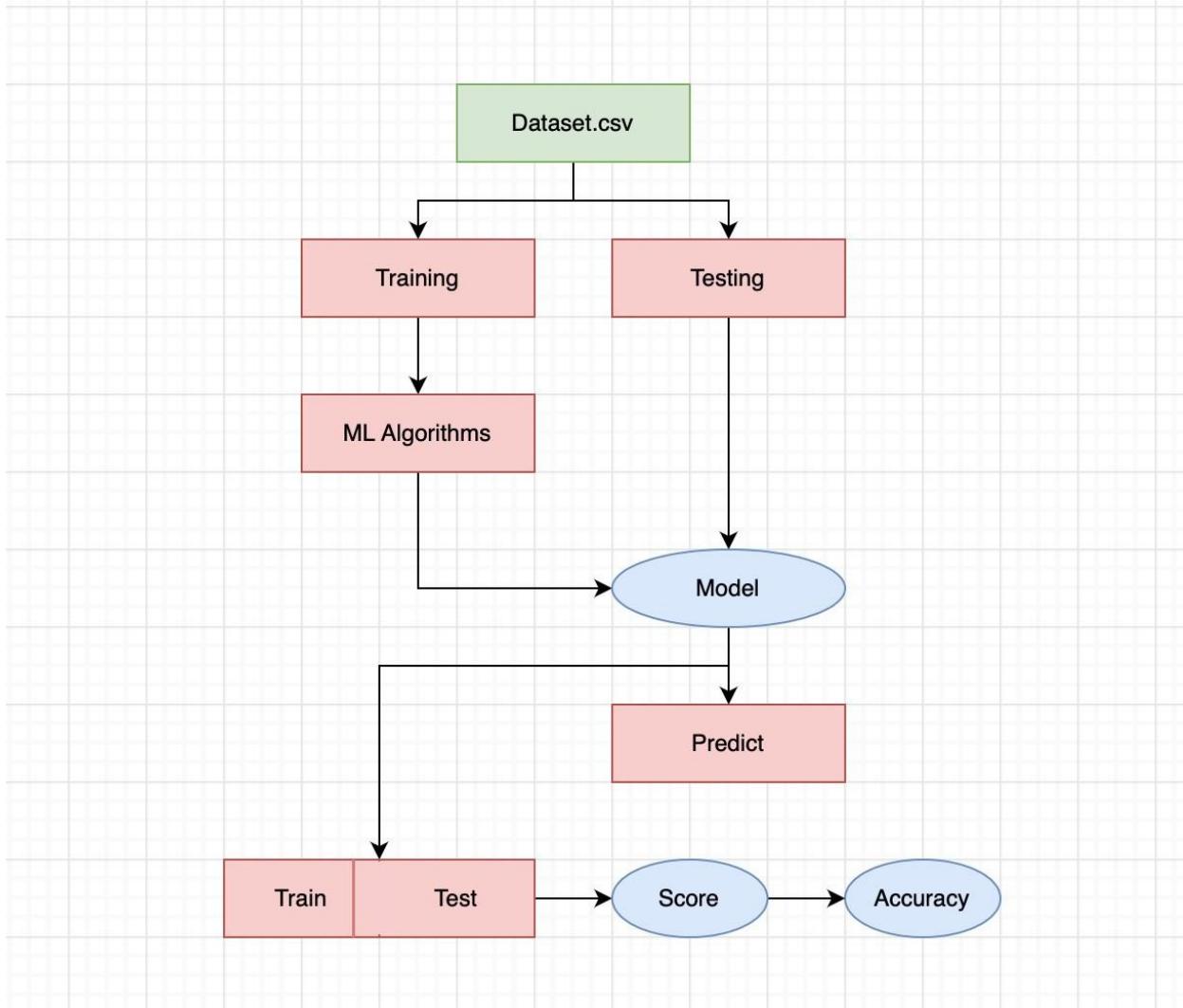
- 1) DEPRESSION.csv
- 2) DEPRESSIONv4.csv (label encoded)
- 3) Depression Dataset.csv
- 4) Real Time Depression Dataset

There are 6 classification ML algorithms :

Gaussian Naive Bayes Algorithm, Random Forest Classifier Algorithm, K-Nearest Neighbour Classifier Algorithm, Decision Tree Classifier Algorithm, Support Vector Machine Algorithm and Logistic Regression Algorithm.

METHODOLOGY

1. DESIGN



DATASETS

We have experimented on two different versions of a dataset namely DEPRESSION.csv which is experimented on without label encoding and DEPRESSIONv4.csv with label encoding.

Original dataset before pre-processing

I. DEPRESSION.csv

The screenshot shows a Jupyter Notebook cell with the following code:

```
[1]: import numpy as np
import pandas as pd
df = pd.read_csv('Mental health Depression disorder Data.csv')
df.head()
```

Below the code, the dataset is displayed as a Pandas DataFrame. The columns are:

	Entity	Code	Year	Schizophrenia (%)	Bipolar disorder (%)	Eating disorders (%)	Anxiety disorders (%)	Drug use disorders (%)	Depression (%)	Alcohol use disorders (%)
0	Afghanistan	AFG	1990	0.160560	0.697779	0.101855	4.828830	1.677082	4.071831	0.672404
1	Afghanistan	AFG	1991	0.160312	0.697961	0.099313	4.829740	1.684746	4.079531	0.671768
2	Afghanistan	AFG	1992	0.160135	0.698107	0.096692	4.831108	1.694334	4.088358	0.670644
3	Afghanistan	AFG	1993	0.160037	0.698257	0.094336	4.830864	1.705320	4.096190	0.669738
4	Afghanistan	AFG	1994	0.160022	0.698469	0.092439	4.829423	1.716069	4.099582	0.669260

We manually changed the values of depression column by applying a condition i.e if value ≥ 3 then the person is depressed(represented by 1) and if value < 3 then the person is not depressed(represented by 0)

After changing the values of the depression column, we get

The screenshot shows a Jupyter Notebook cell with the command `df.head()`. Below the command, the dataset is displayed as a Pandas DataFrame. The columns are:

	Schizophrenia (%)	Bipolar disorder (%)	Eating disorders (%)	Anxiety disorders (%)	Drug use disorders (%)	Alcohol use disorders (%)	Depression
0	0.160560	0.697779	0.101855	4.828830	1.677082	0.672404	1
1	0.160312	0.697961	0.099313	4.829740	1.684746	0.671768	1
2	0.160135	0.698107	0.096692	4.831108	1.694334	0.670644	1
3	0.160037	0.698257	0.094336	4.830864	1.705320	0.669738	1
4	0.160022	0.698469	0.092439	4.829423	1.716069	0.669260	1

Counting the number of rows and columns in the dataset

```
[2]: (6468, 7)
```

OUR DATASET HAS 6468 ROWS AND 7 COLUMNS OF DATA

Count of all the columns having (NaN,na) values

```
Schizophrenia (%)      0
Bipolar disorder (%)   0
Eating disorders (%)   0
Anxiety disorders (%)  0
Drug use disorders (%) 0
Alcohol use disorders (%) 0
Depression             0
dtype: int64
```

FROM THE ABOVE RESULTS, WE CAN SEE THAT NONE OF THE COLUMNS HAVE ANY NaN OR na VALUES.

Counting the number of people having depression(indicated as 1) or not(indicated as 0) based on the values of the 6 features

```
1    4857
0    1611
Name: Depression, dtype: int64
```

FROM THE ABOVE RESULTS, WE CAN SEE THAT 4857 ARE DEPRESSED AND 1611 ARE NOT DEPRESSED.

Count of all the columns having (NaN,na) values

```
Schizophrenia (%)      0
Bipolar disorder (%)   0
Eating disorders (%)   0
Anxiety disorders (%)  0
Drug use disorders (%) 0
Alcohol use disorders (%) 0
Depression             0
dtype: int64
```

FROM THE ABOVE RESULTS, WE CAN SEE THAT NONE OF THE COLUMNS HAVE ANY NaN OR na VALUES.

Counting the number of people having depression(indicated as 1) or not(indicated as 0) based on the values of the 6 features

```
1    4857
0    1611
Name: Depression, dtype: int64
```

FROM THE ABOVE RESULTS, WE CAN SEE THAT 4857 ARE DEPRESSED AND 1611 ARE NOT DEPRESSED.

DATA VISUALIZATION

Plotting our data using sea born(Visualizing the counts by creating a count plot)

DEPRESSION.csv

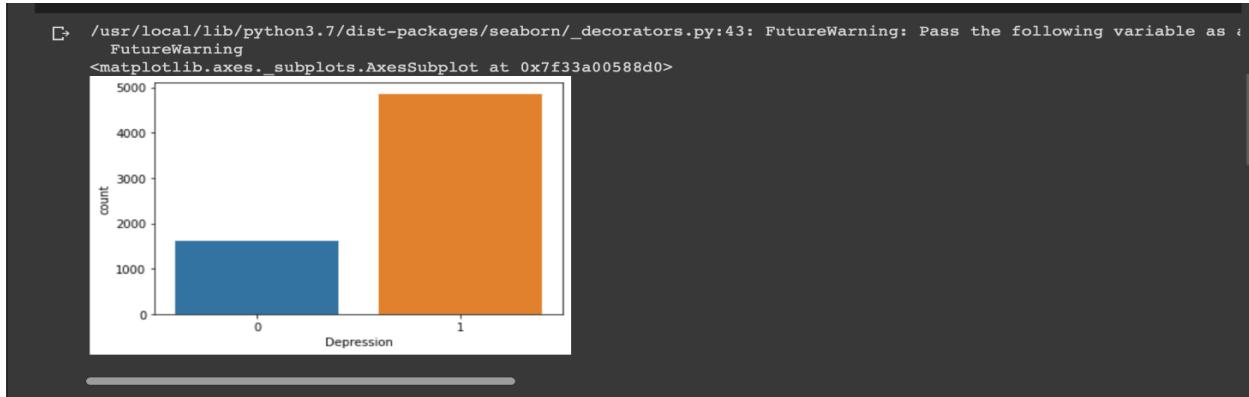
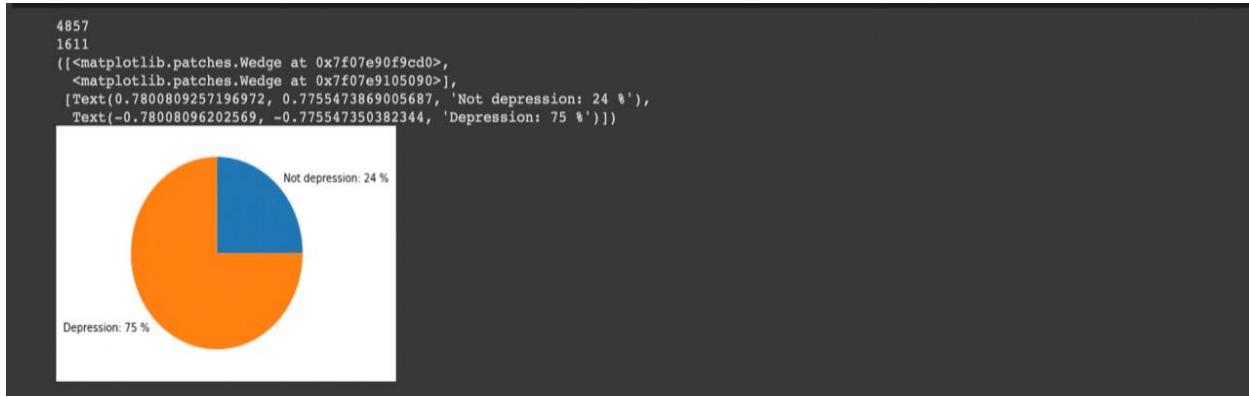


CHART DISPLAYING DEPRESSED (1) AND NOT DEPRESSED (0) .

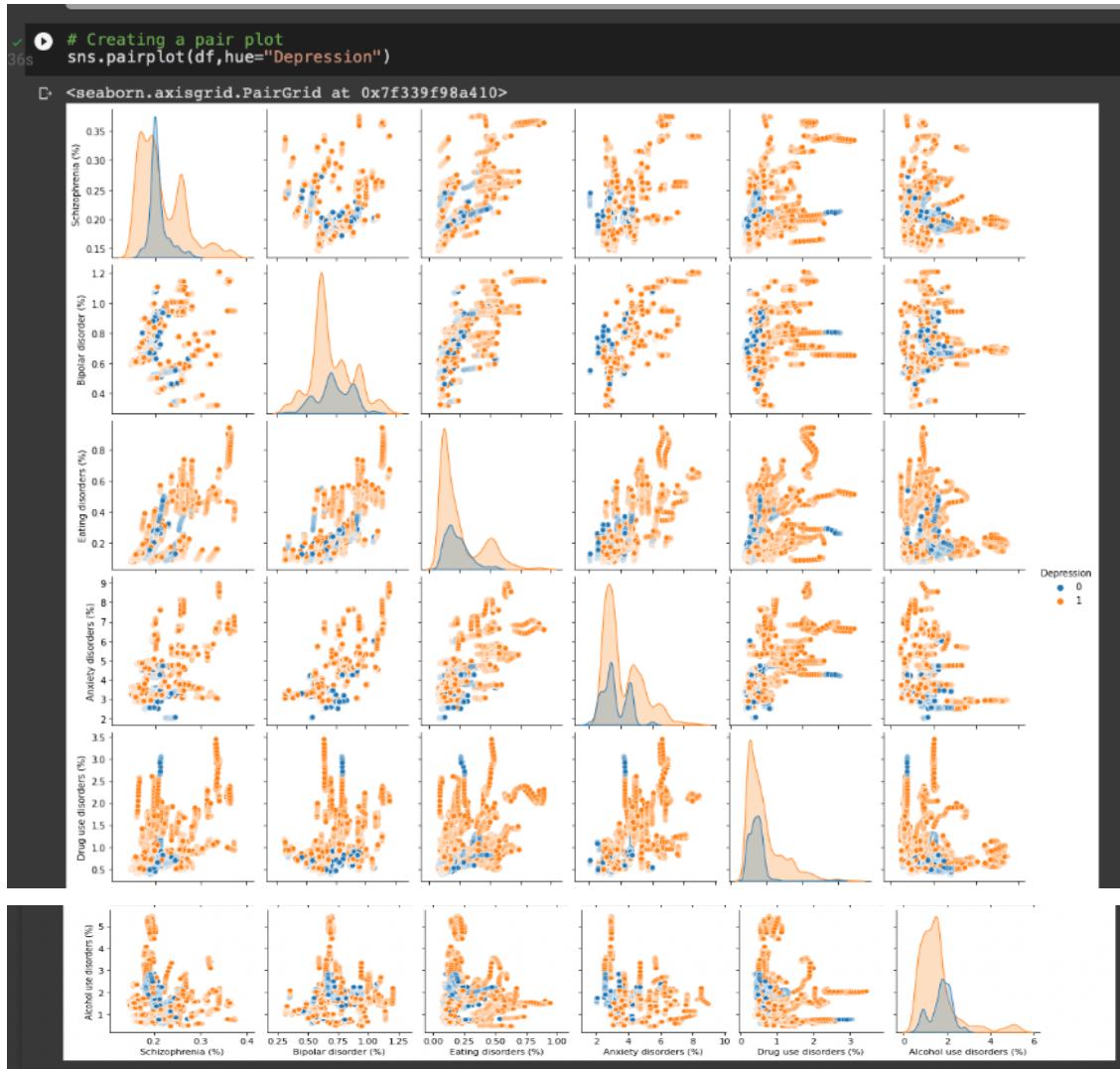
Creating a pie chart

A Pie Chart is a circular statistical plot that can display only one series of data. The area of the chart is the total percentage of the given data. The area of slices of the pie represents the percentage of the parts of the data. The slices of pie are called wedges.



Creating a pair plot

A 'PAIR PLOT' IS ALSO KNOWN AS A SCATTER PLOT IN WHICH ONE VARIABLE IN THE SAME DATA ROW IS MATCHED WITH ANOTHER VARIABLE'S VALUE.

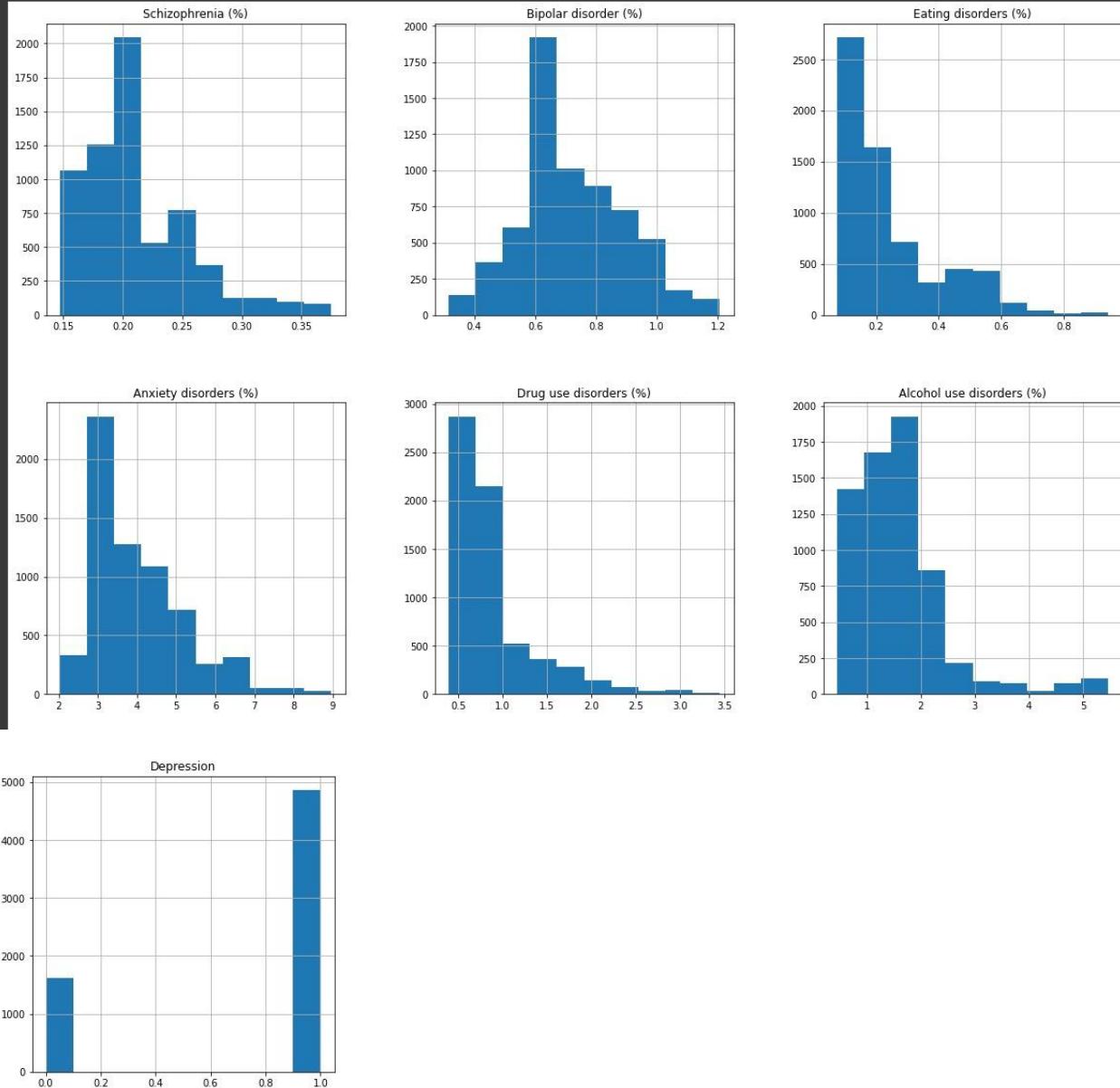


THIS SHOWS THE PAIR PLOT OF ALL THE COLUMNS HIGHLIGHTING DEPRESSED (ORANGE) AND NOT DEPRESSED (BLUE) .

Creating a histogram

A histogram is basically used to represent data provided in a form of some groups. It is an accurate method for the graphical representation of numerical data distribution. It is a type of bar plot where X-axis represents the bin ranges while Y-axis gives information about frequency.

```
hist = df.hist(figsize = (20,20))
```



RUNNING THE DATASET ON DIFFERENT ML ALGORITHMS

We apply classification i.e decision making algorithms to the dataset considered. We test the accuracy for each algorithm. Totally, we run it on 6 algorithms i.e Gaussian Naive Bayes Algorithm, Random Forest Classifier Algorithm, K-Nearest Neighbour Classifier Algorithm, Decision Tree Classifier Algorithm, Support Vector Machine Algorithm and Logistic Regression Algorithm.

Gaussian Naive Bayes Algorithm Output

```
      Schizophrenia (%) Bipolar disorder (%) Eating disorders (%) \
0       0.160560      0.697779      0.101855 \
1       0.160312      0.697961      0.099313 \
2       0.160135      0.698107      0.099692 \
3       0.160037      0.698257      0.094336 \
4       0.160022      0.698469      0.092439 \
...        ...        ...        ...
6463    0.155670      0.607993      0.117248 \
6464    0.155993      0.608610      0.118973 \
6465    0.156465      0.609363      0.119470 \
6466    0.157111      0.610234      0.121456 \
6467    0.157963      0.611242      0.124446 \
          Anxiety disorders (%) Drug use disorders (%) Alcohol use disorders (%) \
0       4.828830      1.677082      0.672404 \
1       4.829740      1.684746      0.671768 \
2       4.821108      1.694334      0.670644 \
3       4.830864      1.705320      0.669738 \
4       4.829423      1.716069      0.669260 \
...        ...        ...        ...
6463    3.090168      0.766280      1.515641 \
6464    3.093964      0.766914      1.515470 \
6465    3.098687      0.771802      1.514751 \
6466    3.104294      0.772275      1.513269 \
6467    3.110926      0.772648      1.510943 \
[6468 rows x 6 columns]
0      1
1      1
2      1
3      1
4      1
...
6463   1
6464   1
6465   1
6466   1
6467   1
Name: Depression, Length: 6468, dtype: int64
0.7594310451453309
[0 0 1 ... 1 1 0]
```

ACCURACY FOR THE ABOVE ALGORITHM IS 0.759 .

Random Forest Classifier Algorithm Output

```
    Schizophrenia (%) Bipolar disorder (%) Eating disorders (%) \
0      0.160560      0.697779      0.101855 \
1      0.160312      0.697961      0.099313
2      0.160135      0.698107      0.096692
3      0.160037      0.698257      0.094336
4      0.160022      0.698469      0.092439
...
6463     ...          ...
6464     0.155670      0.607993      0.117248
6465     0.155993      0.608610      0.118073
6466     0.156465      0.609363      0.119470
6467     0.157111      0.610234      0.121456
6468     0.157963      0.611242      0.124443

    Anxiety disorders (%) Drug use disorders (%) Alcohol use disorders (%)
0      4.828830      1.677082      0.672404
1      4.829740      1.684746      0.671768
2      4.831108      1.694334      0.670644
3      4.830864      1.705320      0.669738
4      4.829423      1.716069      0.669260
...
6463     ...          ...
6464     3.090168      0.766280      1.515641
6465     3.093964      0.768914      1.515470
6466     3.098687      0.771802      1.514751
6467     3.104294      0.772275      1.513269
6468     3.110926      0.772648      1.510943

[6468 rows x 6 columns]
0      1
1      1
2      1
3      1
4      1
...
6463     1
6464     1
6465     1
6466     1
6467     1
Name: Depression, Length: 6468, dtype: int64
0.9938157081014224
[1 1 1 ... 1 1 0]
```

ACCURACY FOR THE ABOVE ALGORITHM IS 0.993 .

K-Nearest Neighbour Classifier Algorithm Output

```
    Schizophrenia (%) Bipolar disorder (%) Eating disorders (%) \
0      0.160560      0.697779      0.101855 \
1      0.160312      0.697961      0.099313
2      0.160135      0.698107      0.096692
3      0.160037      0.698257      0.094336
4      0.160022      0.698469      0.092439
...
6463     ...          ...
6464     0.155670      0.607993      0.117248
6465     0.155993      0.608610      0.118073
6466     0.156465      0.609363      0.119470
6467     0.157111      0.610234      0.121456
6468     0.157963      0.611242      0.124443

    Anxiety disorders (%) Drug use disorders (%) Alcohol use disorders (%)
0      4.828830      1.677082      0.672404
1      4.829740      1.684746      0.671768
2      4.831108      1.694334      0.670644
3      4.830864      1.705320      0.669738
4      4.829423      1.716069      0.669260
...
6463     ...          ...
6464     3.090168      0.766280      1.515641
6465     3.093964      0.768914      1.515470
6466     3.098687      0.771802      1.514751
6467     3.104294      0.772275      1.513269
6468     3.110926      0.772648      1.510943

[6468 rows x 6 columns]
0      1
1      1
2      1
3      1
4      1
...
6463     1
6464     1
6465     1
6466     1
6467     1
Name: Depression, Length: 6468, dtype: int64
0.9857761286332715
[1 1 1 ... 1 1 0]
```

ACCURACY FOR THE ABOVE ALGORITHM IS 0.985.

Decision Tree Classifier Algorithm

Output

```
  Schizophrenia (%)  Bipolar disorder (%)  Eating disorders (%) \
0   0.160560        0.697779        0.101855 \
1   0.160312        0.697961        0.099313 \
2   0.160135        0.698107        0.099692 \
3   0.160037        0.698257        0.094336 \
4   0.160022        0.698469        0.092439 \
...
6463  0.155670        0.607993        0.117248 \
6464  0.155993        0.608610        0.118073 \
6465  0.156465        0.609363        0.119470 \
6466  0.157111        0.610234        0.121456 \
6467  0.157963        0.611242        0.124443 \
...
Anxiety disorders (%)  Drug use disorders (%)  Alcohol use disorders (%)
0    4.828830        1.677082        0.672404 \
1    4.829740        1.684746        0.671768 \
2    4.831108        1.694334        0.670644 \
3    4.830864        1.705320        0.669738 \
4    4.829423        1.716069        0.669268 \
...
6463  3.090168        0.766280        1.515641 \
6464  3.093964        0.768914        1.515470 \
6465  3.098687        0.771802        1.514751 \
6466  3.104294        0.772275        1.513269 \
6467  3.110926        0.772648        1.510943 \
[6468 rows x 6 columns]
0   1
1   1
2   1
3   1
4   1
...
6463  1
6464  1
6465  1
6466  1
6467  1
Name: Depression, Length: 6468, dtype: int64
0.9839208410636983
[1 1 1 ... 1 1 0]
```

ACCURACY FOR THE ABOVE ALGORITHM IS 0.983.

Support Vector Machine Algorithm

Output

```
  Schizophrenia (%)  Bipolar disorder (%)  Eating disorders (%) \
0   0.160560        0.697779        0.101855 \
1   0.160312        0.697961        0.099313 \
2   0.160135        0.698107        0.099692 \
3   0.160037        0.698257        0.094336 \
4   0.160022        0.698469        0.092439 \
...
6463  0.155670        0.607993        0.117248 \
6464  0.155993        0.608610        0.118073 \
6465  0.156465        0.609363        0.119470 \
6466  0.157111        0.610234        0.121456 \
6467  0.157963        0.611242        0.124443 \
...
Anxiety disorders (%)  Drug use disorders (%)  Alcohol use disorders (%)
0    4.828830        1.677082        0.672404 \
1    4.829740        1.684746        0.671768 \
2    4.831108        1.694334        0.670644 \
3    4.830864        1.705320        0.669738 \
4    4.829423        1.716069        0.669268 \
...
6463  3.090168        0.766280        1.515641 \
6464  3.093964        0.768914        1.515470 \
6465  3.098687        0.771802        1.514751 \
6466  3.104294        0.772275        1.513269 \
6467  3.110926        0.772648        1.510943 \
[6468 rows x 6 columns]
0   1
1   1
2   1
3   1
4   1
...
6463  1
6464  1
6465  1
6466  1
6467  1
Name: Depression, Length: 6468, dtype: int64
0.821273964131107
[1 1 1 ... 1 1 0]
```

ACCURACY FOR THE ABOVE ALGORITHM IS 0.821.

Logistic Regression Algorithm Output

```

Schizophrenia (%) Bipolar disorder (%) Eating disorders (%) \
0 0.160560 0.597779 0.101855
1 0.160312 0.697961 0.099313
2 0.160135 0.698187 0.096692
3 0.160037 0.698257 0.094336
4 0.160022 0.698469 0.092439
...
6463 0.155670 0.607993 0.117248
6464 0.155650 0.608110 0.117252
6465 0.156465 0.609363 0.119470
6466 0.157111 0.610234 0.121456
6467 0.157963 0.611242 0.124443

Anxiety disorders (%) Drug use disorders (%) Alcohol use disorders (%)
0 4.828830 1.677082 0.672404
1 4.827440 1.674746 0.671768
2 4.831168 1.691534 0.679444
3 4.830864 1.705320 0.669738
4 4.829423 1.716069 0.669260
...
6463 3.090168 0.766280 1.515641
6464 3.093964 0.768914 1.515470
6465 3.098687 0.771802 1.514751
6466 3.104294 0.772275 1.513269
6467 3.110926 0.772648 1.510943

[6468 rows x 6 columns]
0 1
1 1
2 1
3 1
4 1
...
6463 1
6464 1
6465 1
6466 1
6467 1
Name: Depression, Length: 6468, dtype: int64
0 7884972170686456
[1 1 ... 1 1]

```

ACCURACY FOR THE ABOVE ALGORITHM IS 0.788.

II. DEPRESSIONv4.csv(dataset after label encoding)

	Schizophrenia	Bipolar disorder	Eating disorders	Anxiety disorders	Drug use disorders	Alcohol use disorders	Depression
0	1	2	0	2	1	0	1
1	1	2	0	2	1	0	1
2	1	2	0	2	1	0	1
3	1	2	0	2	1	0	1
4	1	2	0	2	1	0	1

We have label encoded all the columns by determining the max value of every column and splitting them into 4 quartiles.

RUNNING THE DATASET ON DIFFERENT ML ALGORITHMS

Gaussian Naive Bayes algorithm

Output

```
Schizophrenia Bipolar disorder Eating disorders Anxiety disorders \
0           1            2            0            2
1           1            2            0            2
2           1            2            0            2
3           1            2            0            2
4           1            2            0            2
...
6463        1            2            0            1
6464        1            2            0            1
6465        1            2            0            1
6466        1            2            0            1
6467        1            2            0            1

Drug use disorders Alcohol use disorders
0           1            0
1           1            0
2           1            0
3           1            0
4           1            0
...
6463        0            1
6464        0            1
6465        0            1
6466        0            1
6467        0            1

[6468 rows x 6 columns]
0           1
1           1
2           1
3           1
4           1
...
6463        1
6464        1
6465        1
6466        1
6467        1
Name: Depression, Length: 6468, dtype: int64
0.8286951144094001
[1 1 1 ... 1 1 1]
```

ACCURACY OF THE ABOVE ALGORITHM IS 0.8286951144094001

Random Forest Classifier algorithm

Output

```
Schizophrenia  Bipolar disorder  Eating disorders  Anxiety disorders  \
0              1                  2                  0                  2
1              1                  2                  0                  2
2              1                  2                  0                  2
3              1                  2                  0                  2
4              1                  2                  0                  2
...
6463           1                  2                  0                  1
6464           1                  2                  0                  1
6465           1                  2                  0                  1
6466           1                  2                  0                  1
6467           1                  2                  0                  1

Drug use disorders  Alcohol use disorders
0                  1                  0
1                  1                  0
2                  1                  0
3                  1                  0
4                  1                  0
...
6463           0                  1
6464           0                  1
6465           0                  1
6466           0                  1
6467           0                  1

[6468 rows x 6 columns]
0      1
1      1
2      1
3      1
4      1
...
6463   1
6464   1
6465   1
6466   1
6467   1
Name: Depression, Length: 6468, dtype: int64
0.8769325912183055
[1 1 1 ... 1 1 1]
```

ACCURACY OF THE ABOVE ALGORITHM IS 0.8769325912183055

K-Nearest Neighbour Classifier algorithm

Output

```
      Schizophrenia  Bipolar disorder  Eating disorders  Anxiety disorders \
0                  1                  2                  0                  2
1                  1                  2                  0                  2
2                  1                  2                  0                  2
3                  1                  2                  0                  2
4                  1                  2                  0                  2
...
6463                 1                  2                  0                  1
6464                 1                  2                  0                  1
6465                 1                  2                  0                  1
6466                 1                  2                  0                  1
6467                 1                  2                  0                  1

      Drug use disorders  Alcohol use disorders
0                      1                      0
1                      1                      0
2                      1                      0
3                      1                      0
4                      1                      0
...
6463                     0                      1
6464                     0                      1
6465                     0                      1
6466                     0                      1
6467                     0                      1

[6468 rows x 6 columns]
0      1
1      1
2      1
3      1
4      1
...
6463    1
6464    1
6465    1
6466    1
6467    1
Name: Depression, Length: 6468, dtype: int64
0.865182436611008
[1 1 1 ... 1 1 1]
```

ACCURACY OF THE ABOVE ALGORITHM IS 0.865182436611008

Decision Tree Classifier algorithm

Output

```
    Schizophrenia  Bipolar disorder  Eating disorders  Anxiety disorders \
0                  1                  2                  0                  2
1                  1                  2                  0                  2
2                  1                  2                  0                  2
3                  1                  2                  0                  2
4                  1                  2                  0                  2
...
...                ...                ...
6463                 1                  2                  0                  1
6464                 1                  2                  0                  1
6465                 1                  2                  0                  1
6466                 1                  2                  0                  1
6467                 1                  2                  0                  1

    Drug use disorders  Alcohol use disorders
0                      1                      0
1                      1                      0
2                      1                      0
3                      1                      0
4                      1                      0
...
...                ...
6463                 0                      1
6464                 0                      1
6465                 0                      1
6466                 0                      1
6467                 0                      1

[6468 rows x 6 columns]
0      1
1      1
2      1
3      1
4      1
...
6463    1
6464    1
6465    1
6466    1
6467    1
Name: Depression, Length: 6468, dtype: int64
0.8769325912183055
[1 1 1 ... 1 1 1]
```

ACCURACY OF THE ABOVE ALGORITHM IS 0.8769325912183055

Support Vector Machine algorithm

Output

```
      Schizophrenia  Bipolar disorder  Eating disorders  Anxiety disorders  \
0                  1                  2                  0                  2
1                  1                  2                  0                  2
2                  1                  2                  0                  2
3                  1                  2                  0                  2
4                  1                  2                  0                  2
...
...               ...
6463                 1                  2                  0                  1
6464                 1                  2                  0                  1
6465                 1                  2                  0                  1
6466                 1                  2                  0                  1
6467                 1                  2                  0                  1

      Drug use disorders  Alcohol use disorders
0                      1                      0
1                      1                      0
2                      1                      0
3                      1                      0
4                      1                      0
...
...               ...
6463                 0                      1
6464                 0                      1
6465                 0                      1
6466                 0                      1
6467                 0                      1

[6468 rows x 6 columns]
0      1
1      1
2      1
3      1
4      1
...
6463    1
6464    1
6465    1
6466    1
6467    1
Name: Depression, Length: 6468, dtype: int64
0.8701298701298701
[1 1 1 ... 1 1 1]
```

ACCURACY OF THE ABOVE ALGORITHM IS 0.8701298701298701

Logistic Regression algorithm

Output

```
      Schizophrenia  Bipolar disorder  Eating disorders  Anxiety disorders  \
0                  1                  2                  0                  2
1                  1                  2                  0                  2
2                  1                  2                  0                  2
3                  1                  2                  0                  2
4                  1                  2                  0                  2
...
6463                 1                  2                  0                  1
6464                 1                  2                  0                  1
6465                 1                  2                  0                  1
6466                 1                  2                  0                  1
6467                 1                  2                  0                  1

      Drug use disorders  Alcohol use disorders
0                      1                      0
1                      1                      0
2                      1                      0
3                      1                      0
4                      1                      0
...
6463                     0                     1
6464                     0                     1
6465                     0                     1
6466                     0                     1
6467                     0                     1

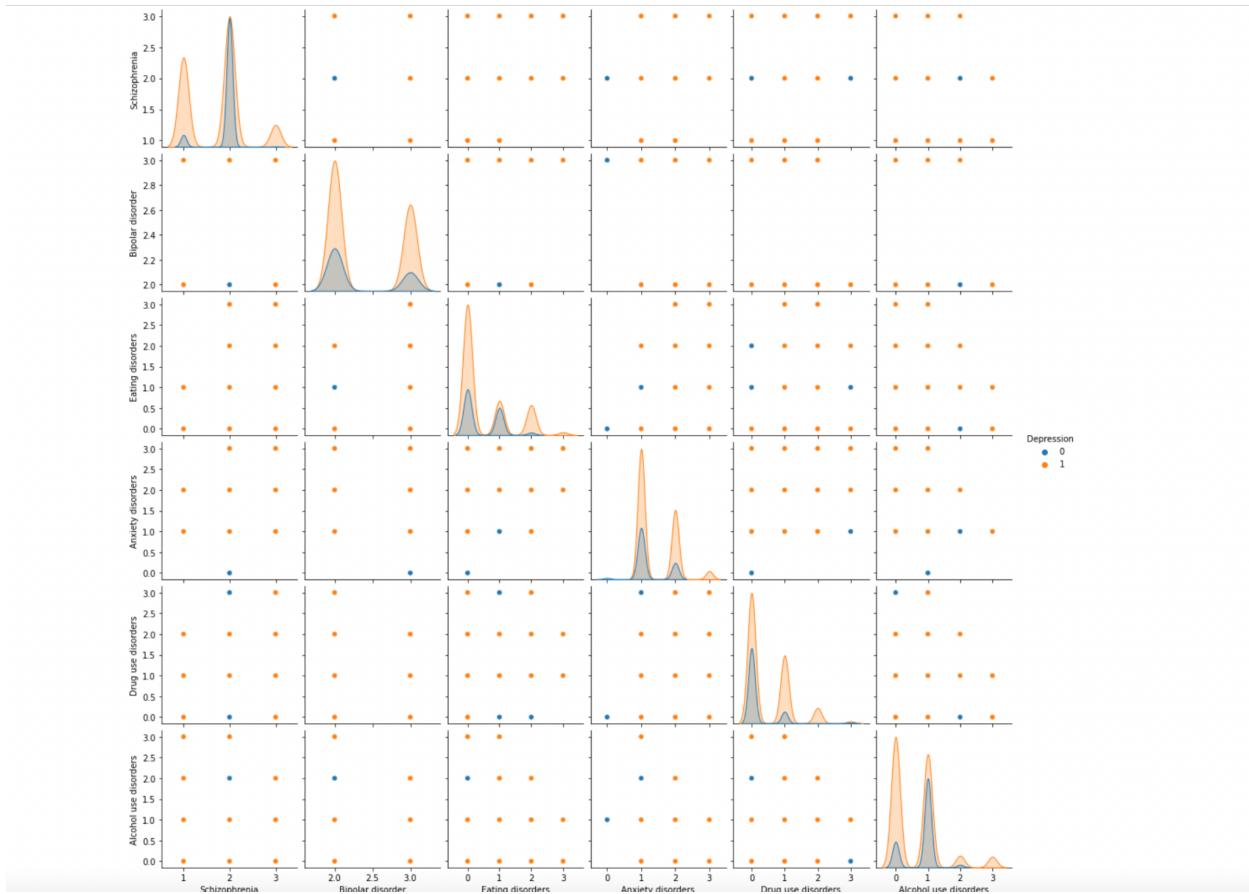
[6468 rows x 6 columns]
0      1
1      1
2      1
3      1
4      1
...
6463    1
6464    1
6465    1
6466    1
6467    1
Name: Depression, Length: 6468, dtype: int64
0.8070500927643784
[1 1 1 ... 1 1 1]
```

ACCURACY OF THE ABOVE ALGORITHM IS 0.8070500927643784

DEPRESSIONv4.csv

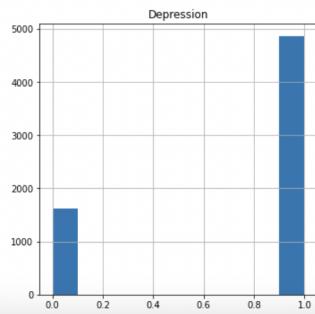
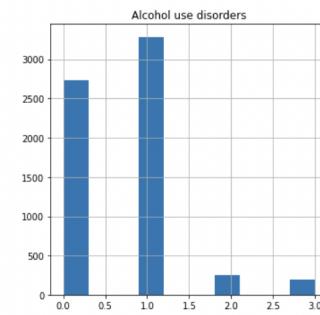
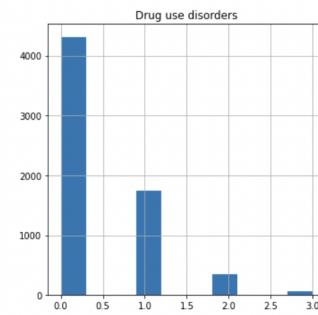
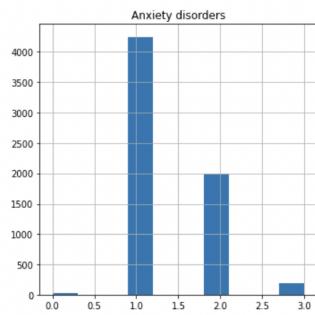
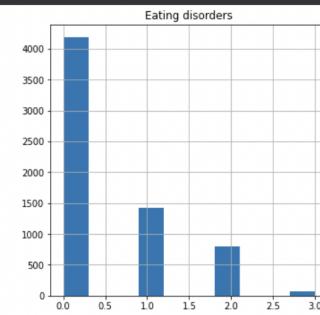
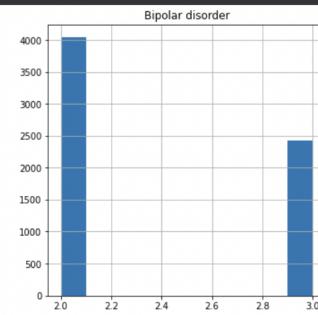
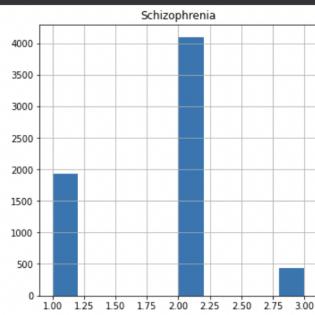
Creating a pair plot

A 'PAIR PLOT' IS ALSO KNOWN AS A SCATTER PLOT IN WHICH ONE VARIABLE IN THE SAME DATA ROW IS MATCHED WITH ANOTHER VARIABLE'S VALUE.



THIS SHOWS THE PAIR PLOT OF ALL THE COLUMNS HIGHLIGHTING DEPRESSED(ORANGE) AND NOT DEPRESSED(BLUE).

Creating a histogram



III. Depression Dataset.csv

	AGERNG	GENDER	EDU	PROF	MARSTS	RESDPL	LIVWTH	ENVSAT	POSSAT	FINSTR	...	ANXI	DEPRI	ABUSED	CHEAT	THREAT	SUICIDE	INFER	CONFL	DE
0	26-30	Female	Post Graduate	Unemployed	Unmarried	Town	With Family	Yes	Yes	No	...	Yes	No	No	No	No	No	No	No	\\
1	26-30	Male	Post Graduate	Service holder (Private)	Unmarried	City	With Family	Yes	No	Yes	...	Yes	Yes	Yes	No	No	No	No	Yes	\\
2	21-25	Male	HSC	Student	Unmarried	City	With Family	Yes	Yes	No	...	Yes	Yes	No	No	No	No	No	No	\\
3	16-20	Male	HSC	Student	Unmarried	City	With Family	No	Yes	No	...	Yes	Yes	No	Yes	No	No	No	No	\\
4	21-25	Male	Graduate	Student	Unmarried	Town	With Family	No	Yes	Yes	...	Yes	Yes	No	No	No	No	Yes	\\	

5 rows x 31 columns

```
[ ] df.columns
Index(['AGERNG', 'GENDER', 'EDU', 'PROF', 'MARSTS', 'RESDPL', 'LIVWTH',
       'ENVSAT', 'POSSAT', 'FINSTR', 'DEBT', 'PHYEX', 'SMOKE', 'DRINK',
       'ILLNESS', 'PREMED', 'EATDIS', 'AVGSLP', 'INSOM', 'TSSN', 'WRKPRE',
       'ANXI', 'DEPRI', 'ABUSED', 'CHEAT', 'THREAT', 'SUICIDE', 'INFER',
       'CONFLICT', 'LOST', 'DEPRESSED'],
      dtype='object')
```

Depression Dataset.csv after label encoding

	AGERNG	GENDER	EDU	PROF	MARSTS	RESDPL	LIVWTH	ENVSAT	POSSAT	FINSTR	...	ANXI	DEPRI	ABUSED	CHEAT	THREAT	SUICIDE	INFER	CONFLICT	LOST	DE
0	2	0	2	5	2	1	0	1	1	0	...	1	0	0	0	0	0	0	1	0	\\
1	2	1	2	3	2	0	0	1	0	1	...	1	1	1	0	0	0	0	1	0	0
2	1	1	1	4	2	0	0	1	1	0	...	1	1	0	0	0	0	0	0	0	0
3	0	1	1	4	2	0	0	0	1	0	...	1	1	0	1	0	0	0	0	0	0
4	1	1	0	4	2	1	0	0	1	1	...	1	1	0	0	0	0	0	1	1	0

5 rows x 31 columns

RUNNING THE DATASET ON DIFFERENT ML ALGORITHMS

Gaussian Naive Bayes algorithm

Output

```
      ENVSAT POSSAT FINSTR INSOM ANXI DEPRI ABUSED CHEAT THREAT \
0         1     1      0     0     1     0      0     0     0
1         1     0      1     1     1     1      1     0     0
2         1     1      0     0     1     1      1     0     0
3         0     1      0     0     1     1      1     0     1
4         0     1      1     1     1     1      1     0     0
...
599        0     1      1     1     0     1      0     0     0
600        0     0      1     0     1     1      1     0     0
601        0     0      1     0     1     0      0     0     0
602        1     1      0     0     1     0      0     0     0
603        1     1      1     0     0     0      0     0     0

      SUICIDE INFER CONFLICT LOST
0         0     0      1     0
1         0     1      0     0
2         0     0      0     0
3         0     0      0     0
4         0     1      1     0
...
599        0     0      0     1
600        0     1      0     0
601        0     0      0     0
602        0     1      0     0
603        0     0      0     1

[604 rows x 13 columns]
0         0
1         1
2         0
3         1
4         1
...
599        1
600        1
601        1
602        0
603        0
Name: DEPRESSED, Length: 604, dtype: int64
0.7814569536423841
```

Random Forest Classifier algorithm

Output

```
    ENVSAT POSSAT FINSTR INSOM ANXI DEPRI ABUSED CHEAT THREAT \
0      1      1      0      0      1      0      0      0      0
1      1      0      1      1      1      1      1      0      0
2      1      1      0      0      1      1      0      0      0
3      0      1      0      0      1      1      0      1      0
4      0      1      1      1      1      1      0      0      0
...
599     ...     ...     ...     ...     ...     ...     ...
600     0      1      1      1      0      1      0      0      0
601     0      0      1      0      1      1      1      0      0
602     1      1      0      0      1      0      0      0      0
603     1      1      1      0      0      0      0      0      0

    SUICIDE INFER CONFLICT LOST
0      0      0      1      0
1      0      1      0      0
2      0      0      0      0
3      0      0      0      0
4      0      1      1      0
...
599     ...     ...     ...
600     0      0      0      1
601     0      1      0      0
602     0      1      0      0
603     0      0      0      1

[604 rows x 13 columns]
0      0
1      1
2      0
3      1
4      1
...
599     1
600     1
601     1
602     0
603     0
Name: DEPRESSED, Length: 604, dtype: int64
0.8675496688741722
```

K-Nearest Neighbour Classifier algorithm

Output

```
    ENVSAT POSSAT FINSTR INSOM ANXI DEPRI ABUSED CHEAT THREAT \
0      1      1      0      0      1      0      0      0      0
1      1      0      1      1      1      1      1      0      0
2      1      1      0      0      1      1      0      0      0
3      0      1      0      0      1      1      0      1      0
4      0      1      1      1      1      1      0      0      0
...
599     0      1      1      1      0      1      0      0      0
600     0      0      1      0      1      1      1      0      0
601     0      0      1      0      1      0      0      0      0
602     1      1      0      0      1      0      0      0      0
603     1      1      1      0      0      0      0      0      0

    SUICIDE INFER CONFLICT LOST
0        0      0      1      0
1        0      1      0      0
2        0      0      0      0
3        0      0      0      0
4        0      1      1      0
...
599     0      0      0      1
600     0      1      0      0
601     0      0      0      0
602     0      1      0      0
603     0      0      0      1

[604 rows x 13 columns]
0      0
1      1
2      0
3      1
4      1
...
599     1
600     1
601     1
602     0
603     0
Name: DEPRESSED, Length: 604, dtype: int64
0.8609271523178808
```

Decision Tree Classifier algorithm

Output

```
    ENVSAT POSSAT FINSTR INSOM ANXI DEPRI ABUSED CHEAT THREAT \
0      1      1      0      0      1      0      0      0      0
1      1      0      1      1      1      1      1      0      0
2      1      1      0      0      1      1      1      0      0
3      0      1      0      0      1      1      1      0      1
4      0      1      1      1      1      1      0      0      0
...
599     0      1      1      1      0      1      0      0      0
600     0      0      1      0      1      1      1      0      0
601     0      0      1      0      1      0      0      0      0
602     1      1      0      0      1      0      0      0      0
603     1      1      1      0      0      0      0      0      0

    SUICIDE INFER CONFLICT LOST
0      0      0      1      0
1      0      1      0      0
2      0      0      0      0
3      0      0      0      0
4      0      1      1      0
...
599     0      0      0      1
600     0      1      0      0
601     0      0      0      0
602     0      1      0      0
603     0      0      0      1

[604 rows x 13 columns]
0      0
1      1
2      0
3      1
4      1
...
599     1
600     1
601     1
602     0
603     0
Name: DEPRESSED, Length: 604, dtype: int64
0.8211920529801324
```

Support Vector Machine algorithm

Output

```
      ENVSAT POSSAT FINSTR INSQM ANXI DEPRI ABUSED CHEAT THREAT \
0          1     1     0     0     1     0     0     0     0
1          1     0     1     1     1     1     1     0     0
2          1     1     0     0     0     1     1     0     0
3          0     1     0     0     1     1     1     0     1
4          0     1     1     1     1     1     0     0     0
..      ...
599        0     1     1     1     0     1     0     0     0
600        0     0     1     0     1     1     1     0     0
601        0     0     1     0     1     0     0     0     0
602        1     1     0     0     1     0     0     0     0
603        1     1     1     0     0     0     0     0     0

      SUICIDE INFER CONFLICT LOST
0          0     0     1     0
1          0     1     0     0
2          0     0     0     0
3          0     0     0     0
4          0     1     1     0
..      ...
599        0     0     0     1
600        0     1     0     0
601        0     0     0     0
602        0     1     0     0
603        0     0     0     1

[604 rows x 13 columns]
0      0
1      1
2      0
3      1
4      1
..
599    1
600    1
601    1
602    0
603    0
Name: DEPRESSED, Length: 604, dtype: int64
0.9072847682119205
```

Logistic Regression algorithm

Output

```
    ENVSAT POSSAT FINSTR INSOM ANXI DEPRI ABUSED CHEAT THREAT \
0      1      1      0      0      1      0      0      0      0
1      1      0      1      1      1      1      1      0      0
2      1      1      0      0      1      1      0      0      0
3      0      1      0      0      1      1      0      1      0
4      0      1      1      1      1      1      0      0      0
...
599     0      1      1      1      0      1      0      0      0
600     0      0      1      0      1      1      1      0      0
601     0      0      1      0      1      0      0      0      0
602     1      1      0      0      1      0      0      0      0
603     1      1      1      0      0      0      0      0      0

    SUICIDE INFER CONFLICT LOST
0      0      0      1      0
1      0      1      0      0
2      0      0      0      0
3      0      0      0      0
4      0      1      1      0
...
599     0      0      0      1
600     0      1      0      0
601     0      0      0      0
602     0      1      0      0
603     0      0      0      1

[604 rows x 13 columns]
0      0
1      1
2      0
3      1
4      1
...
599     1
600     1
601     1
602     0
603     0
Name: DEPRESSED, Length: 604, dtype: int64
0.9072847682119205
```

IV. REAL TIME DATASET

Data collected from DSU students

20 Questions taken from Beck's algorithm

Mental Health Survey

Fill this form with respect to your recent mental state. Data obtained is intended for project use only.
Survey conducted only for undergrad students.
[There's cake at the end]

 suchitradas931@gmail.com (not shared) [Switch account](#) 

* Required

1. Sadness *

I do not feel sad.
 I feel sad.
 I am sad all the time and I can't snap out of it.
 I am so sad and unhappy that I can't stand it.

2. Future *

I am not particularly discouraged about the future.
 I feel discouraged about the future.
 I feel I have nothing to look forward to.
 I feel the future is hopeless and that things cannot improve.

100 responses

Accepting responses

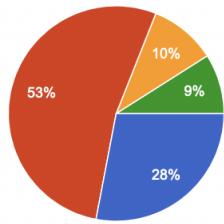
Summary

Question

Individual

1. Sadness

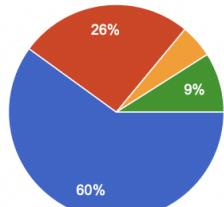
100 responses

 Copy

- I do not feel sad.
- I feel sad.
- I am sad all the time and I can't snap out of it.
- I am so sad and unhappy that I can't stand it.

2. Future

100 responses

 Copy

- I am not particularly discouraged about the future.
- I feel discouraged about the future.
- I feel I have nothing to look forward to.
- I feel the future is hopeless and that things cannot improve.

GUI

After feature selection using Boruta, we get 13 most appropriate features

ENVSAT, POSSAT, FINSTR, INSOM, ANXI, DEPRI, ABUSED, CHEAT, THREAT, SUICIDE, INFER, CONFLICT, LOST

We have built a GUI using python displaying these 13 features. Prediction is done using the Decision Tree Algorithm.

Screenshots

The screenshot shows a window titled "Depression Detection System". It contains 13 pairs of questions and radio button options. The questions are: Environmental satisfaction, Achievement satisfaction, Financial stress, Insomnia, Anxiety, Deprivation, Abused, Cheated, Threat, Suicidal, Inferiority complex, Recent conflict, and Recent loss. Each question has two radio button options: "Yes" and "No". The "No" option is selected for all questions except "Inferiority complex" where "Yes" is selected. To the right of the "Recent loss" question, the text "Probability of depression is low" is displayed in green. At the bottom right is a "Submit" button.

Question	Yes	No
Environmental satisfaction	<input type="radio"/>	<input checked="" type="radio"/>
Achievement satisfaction	<input type="radio"/>	<input checked="" type="radio"/>
Financial stress	<input type="radio"/>	<input checked="" type="radio"/>
Insomnia	<input type="radio"/>	<input checked="" type="radio"/>
Anxiety	<input type="radio"/>	<input checked="" type="radio"/>
Deprivation	<input type="radio"/>	<input checked="" type="radio"/>
Abused	<input type="radio"/>	<input checked="" type="radio"/>
Cheated	<input type="radio"/>	<input checked="" type="radio"/>
Threat	<input type="radio"/>	<input checked="" type="radio"/>
Suicidal	<input type="radio"/>	<input checked="" type="radio"/>
Inferiority complex	<input checked="" type="radio"/>	<input type="radio"/>
Recent conflict	<input type="radio"/>	<input checked="" type="radio"/>
Recent loss	<input type="radio"/>	<input checked="" type="radio"/>



Depression Detection System

Environmental satisfaction

Yes No

Achievement satisfaction

Yes No

Financial stress

Yes No

Insomnia

Yes No

Anxiety

Yes No

Deprivation

Yes No

Abused

Yes No

Cheated

Yes No

Threat

Yes No

Suicidal

Yes No

Inferiority complex

Yes No

Recent conflict

Yes No

Recent loss

Yes No

Probability of depression is high

EXPERIMENTATION

We used different datasets from various sources like, github, kaggle, and even collected real time data using google forms.

We used :

- 1) DEPRESSION.csv
- 2) DEPRESSIONv4.csv (label encoded)
- 3) Depression Dataset
- 4) Real Time Depression Dataset (collected on google forms, of only undergraduate students of Dayananda Sagar University)

We use 3 algorithms for feature selection :

SelectKBest

mRMR

Boruta

CHAPTER 8 - TESTING AND RESULTS

We used the following datasets :

- 1) DEPRESSION.csv
- 2) DEPRESSIONv4.csv (label encoded)
- 3) Depression Dataset

We ran the above datasets in the 6 classification ML algorithms :

- Logistic regression
- Decision tree
- Random forest
- Support vector machine
- K nearest neighbour
- Naive bayes

The following are the accuracy rate we got

1) DEPRESSION.csv

Gaussian Naive's Bayes algorithm - 0.7594310451453309

Random Forest Classifier algorithm - 0.9925788497217068

K-Nearest Neighbour Classifier algorithm - 0.9857761286332715

Decision Tree Classifier algorithm - 0.9820655534941249

Support Vector Machine algorithm - 0.821273964131107

Logistic Regression algorithm - 0.7884972170686456

2) DEPRESSIONv4.csv (label encoded)

Gaussian Naive's Bayes algorithm - 0.8286951144094001

Random Forest Classifier algorithm - 0.8769325912183055

K-Nearest Neighbour Classifier algorithm - 0.865182436611008

Decision Tree Classifier algorithm - 0.8769325912183055

Support Vector Machine algorithm - 0.8701298701298701

Logistic Regression algorithm - 0.8070500927643784

3) Depression Dataset

Gaussian Naive's Bayes algorithm -

Random Forest Classifier algorithm - 0.8675496688741722

K-Nearest Neighbour Classifier algorithm - 0.8609271523178808

Decision Tree Classifier algorithm - 0.8211920529801324

Support Vector Machine algorithm - 0.9072847682119205

Logistic Regression algorithm - 0.9072847682119205

We have used 6 classification algorithms(Gaussian Naive Bayes Algorithm, Random Forest Classifier Algorithm, K-Nearest Neighbor Classifier Algorithm, Decision Tree Classifier Algorithm, Support Vector Machine Algorithm and Logistic Regression Algorithm.) to predict if an individual is depressed or not based on the values of the 6 features (Schizophrenia, Bipolar disorder, Eating disorders, Anxiety disorders, Drug use disorders, Alcohol use disorders) for DEPRESSION.csv and DEPRESSIONv4.csv and 13 features(ENVSAT, POSSAT, FINSTR, INSOM, ANXI, DEPRI, ABUSED, CHEAT, THREAT, SUICIDE, INFER, CONFLICT, LOST) for Depression Dataset.csv

We have built a GUI using python displaying these 13 features. Prediction is done using the Decision Tree Algorithm.