# PSYCHIATRIC PATIENT MONITORING


## Abstract:

Especially after covid-19 pandemic depression and anxiety increased by 25% among people around the world. Mental health is inextricably linked with the person's work efficiency, physical health, social behavior, and many more. Can be seen poor mental health has very grave complications, for instance, family conflicts, a problem with various addictions, most imperative depletion in work efficiency, and suicide. Poor Mental health issue has become more common among the employees due to stress at the workplace, burnout, toxic work culture, and many more. Poor mental health ultimately leads to the overall performance of the firm.

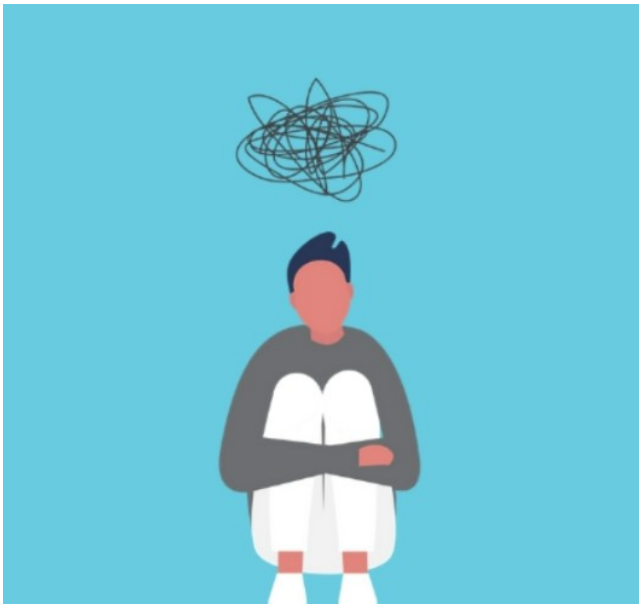
In this project, we analyze the stress among the people using a basic stress detection technique called PHQ9 questionnaire and do the analysis of the data. The main objective of this project is the determine the status of the mental health among the people. Using machine learning we detect people who are depressed could help patients and get the support they need more quickly and easily, while also reduces healthcare expenses and the stress . This analysis can bring the conclusion whether deterioration of mental health should take into consideration or not in determining the actual cause of the work efficiency and overall leverage enhancement to the firm. Furthermore, in this report, we will discuss the limitations and future aspects of the application.


## Introduction:

Mental health is the imperative factor that affects the person's behavior, emotions, reasoning, and social interactions. Bad mental health has grave consequences in society. It can be seen in the many horrific incidents, for instance, loss of health status, self-harm even suicides. It can be seen the exponential increment of the mental health-related issue around the world as the covid pandemic hit the world very hard. Especially due to burnout, disturbance in the schedule, toxic work culture mental health of the employees has become a primary concern for most corporate giants and small scale individual firms around the world.

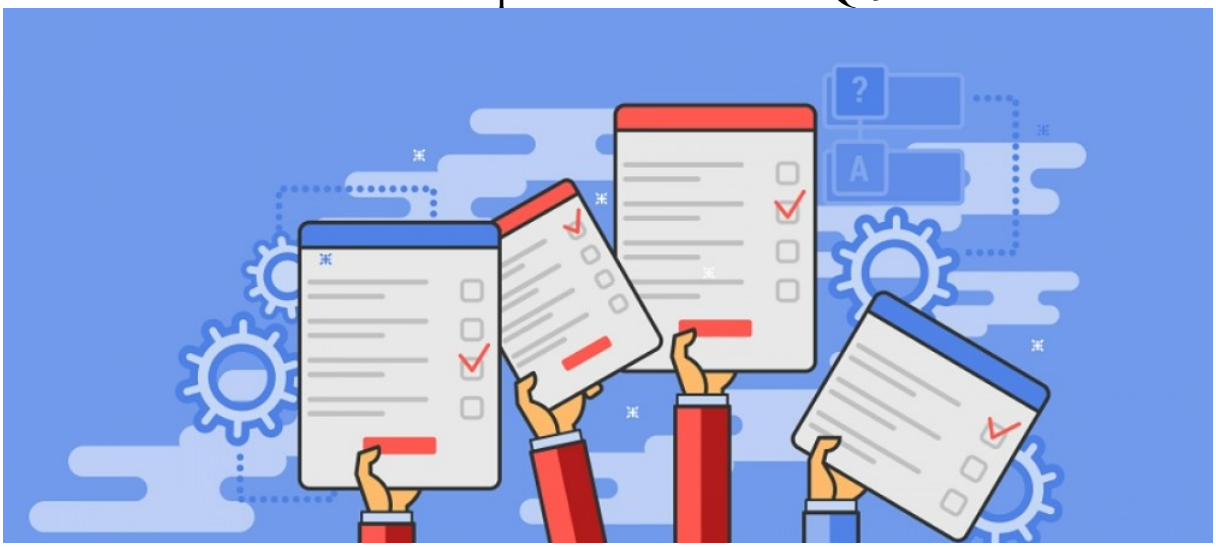
Mental health can be diagnosed using the individual responses of the persons to the basic mental health detection technique PHQ-9. Machine learning is the technique which constructs the system which improve it self by experience, using probability and statistical techniques.it allow

the research personnel to extract the useful information from the particular data. The algorithm used in the machine learning is very useful to determine and forecast the future event, analysis and many more.



## 1 Problem specification

- The main aim is to work on questionnaire of PHQ-9 .



- Classification of who are  depressed and not depressed .

- It followed the **osemin** method - " **Obtain, Scrub, Explore, Model, and interpret**

- Modeling done from simple to complex

- Use a little feature as necessary

- Accuracy

- Prediction

## 1.1 Dataset

1.CDC NHANES DATA.

2. The files for computing the depression target variable can be accessed at https://wwwn.cdc.gov/nchs/nhanes/default.aspx on the NHANES website. Select the year, choose "Questionnaire Data," and click the link for the XPT file next to "Mental Health - Depression Screener" to download the exact data for each year group.

3.Answers to a depression screening questionnaire are included in the file. It consists of nine questions used to diagnose depression, as well as a tenth question about how tough the preceding nine items were. The first nine items of the questionnaire are evaluated, and a score of 10 or higher indicates depression.

4.The questionnaire uses DSM diagnostic criteria, and a study indicated that it has 88 percent specificity and sensitivity for serious depression with a score of 10 or higher threshold.

5. Use the particular data which is consistent over the years and easily be founded in the person's medical history

## 1.2 Problem analysis

After getting answers from the people for questionaries , we will

1.Train the unsupervised algorithm
2.Observe the clusters formed by algorithm
3.Draw conclusions based on clusters formed
4.We will  this algorithm to website
5.In the website we can get final result of the mental health

**PATIENT HEALTH QUESTIONNAIRE-9 (PHQ-9)**

| **Over the last 2 weeks, how often have you been bothered by any of the following problems?** *(Use "✔" to indicate your answer)* | **Not at all** | **Several days** | **More than half the days** | **Nearly every day** |
|---|---|---|---|---|
| **1.** Little interest or pleasure in doing things | 0 | 1 | 2 | 3 |
| **2.** Feeling down, depressed, or hopeless | 0 | 1 | 2 | 3 |
| **3.** Trouble falling or staying asleep, or sleeping too much | 0 | 1 | 2 | 3 |
| **4.** Feeling tired or having little energy | 0 | 1 | 2 | 3 |
| **5.** Poor appetite or overeating | 0 | 1 | 2 | 3 |
| **6.** Feeling bad about yourself — or that you are a failure or have let yourself or your family down | 0 | 1 | 2 | 3 |
| **7.** Trouble concentrating on things, such as reading the newspaper or watching television | 0 | 1 | 2 | 3 |
| **8.** Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual | 0 | 1 | 2 | 3 |
| **9.** Thoughts that you would be better off dead or of hurting yourself in some way | 0 | 1 | 2 | 3 |

FOR OFFICE CODING __0__ + _____ + _____ + _____

**=Total Score:** _____

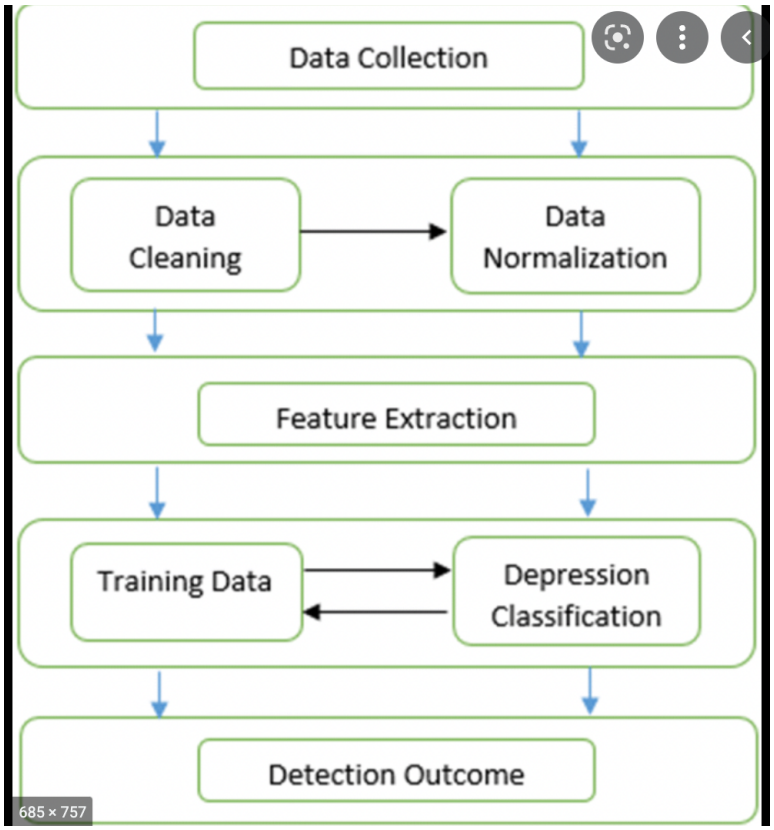
**If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?**

| **Not difficult at all** | **Somewhat difficult** | **Very difficult** | **Extremely difficult** |
|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ |

## 2 Design and Milestones

1.Data preprocessing
2.Train the model
3. Choose the algorithm
4. Accuracy
5. Design of platform with Ml algorithm
6.Website Design
7. Result

### 2.1 Proposed Method




We used the PHQ-9 cutoff score and machine learning algorithm approaches to do a diagnostic meta-analysis.
1.Random forest,
2.Support vector machine (SVM),
3.K-nearest neighbor (KNN),
4.Artificial neural network (ANN),

5.10-fold cross-validation were the machine learning algorithms used.

Random forest (RF) is a dependable classifier that selects and ranks variables with the greatest potential to distinguish across target classes using predictions produced from ensembles of decision trees.
A separating hyperplane is a discriminant classifier that can be defined as a Support Vector Machine. The concept of hyperplane includes a generalization of the maximal margin classifier.

For classification and regression, the K-Nearest Neighbors (kNN) algorithm is utilized. Pattern recognition and predictive analysis are two areas where it excels.

The neural network is a type of artificial intelligence. (ANN) is a machine learning technique that was developed to simulate the human brain. It has strong fault tolerance and parallel processing makes it quick and scalable.

## 2.2 Data processing

1. Data preprocessing

```
In [8]:
# Dropping the final column
df_target.drop(columns=['DPQ100'], inplace=True)
```

```
In [9]:
df_target.head()
```

Out[9]:

| SEQN | DPQ010 | DPQ020 | DPQ030 | DPQ040 | DPQ050 | DPQ060 | DPQ070 | DPQ080 | DPQ090 |
|---|---|---|---|---|---|---|---|---|---|
| 31130 | nan | nan | nan | nan | nan | nan | nan | nan | nan |
| 31131 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 31132 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 31134 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 31139 | 0.000 | 0.000 | 0.000 | 0.000 | 3.000 | 1.000 | 0.000 | 0.000 | 0.000 |

```
In [10]:
df_target.info()
```

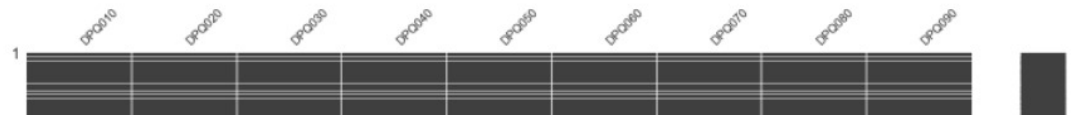
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 40496 entries, 31130 to 102956
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   DPQ010  36471 non-null  float64
 1   DPQ020  36465 non-null  float64
 2   DPQ030  36463 non-null  float64
 3   DPQ040  36458 non-null  float64
 4   DPQ050  36458 non-null  float64
 5   DPQ060  36455 non-null  float64
 6   DPQ070  36453 non-null  float64
 7   DPQ080  36452 non-null  float64
```

1.Eliminate the unwanted features
2.Delete the null values

In [11]:
```
# Checking the null values
df_target.isna().sum()
```

Out[11]:
```
DPQ010    4025
DPQ020    4031
DPQ030    4033
DPQ040    4038
DPQ050    4038
DPQ060    4041
DPQ070    4043
DPQ080    4044
DPQ090    4049
dtype: int64
```

In [12]:
```
# Looking at the pattern of null values
import missingno as ms
ms.matrix(df_target);
```




In [13]:
```
# Removing all null entries
df_target.dropna(inplace=True)
```

In [14]:
```
# Checking for null values and how many entries are left
df_target.info()
```

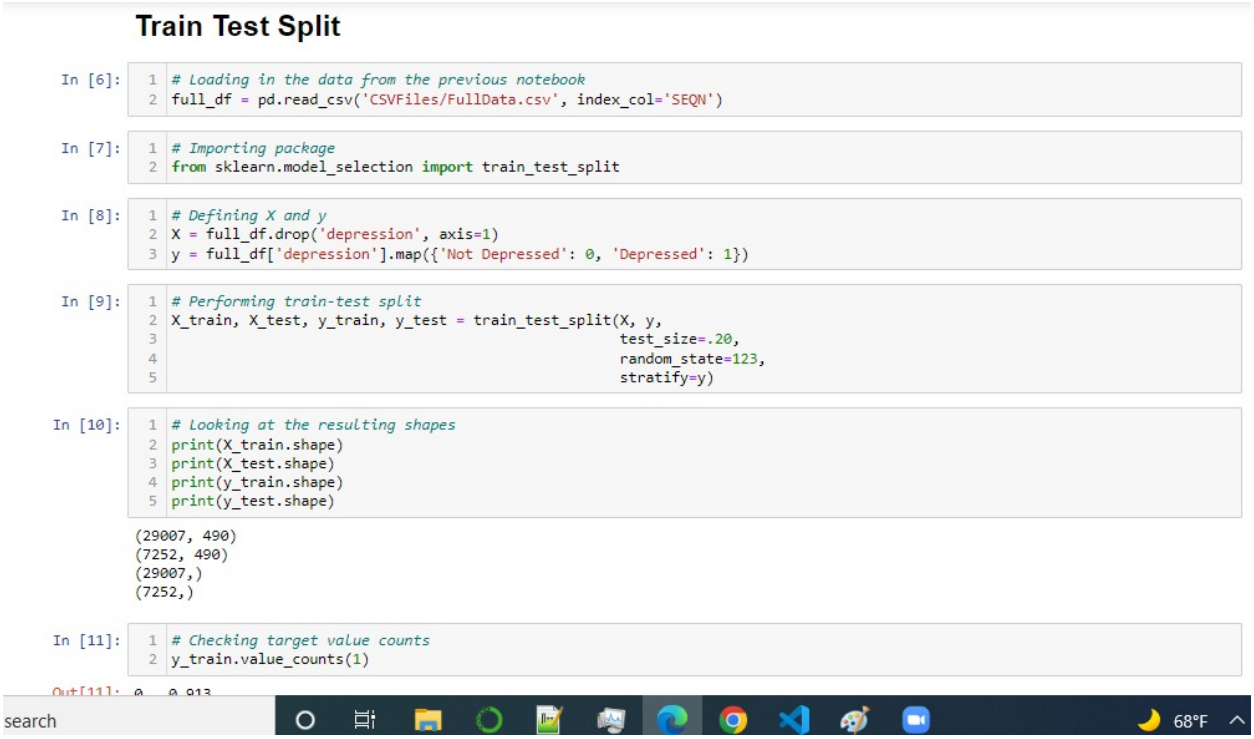
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 36447 entries, 31131 to 102956
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   DPQ010  36447 non-null  float64
 1   DPQ020  36447 non-null  float64
 2   DPQ030  36447 non-null  float64
 3   DPQ040  36447 non-null  float64
 4   DPQ050  36447 non-null  float64
 5   DPQ060  36447 non-null  float64
 6   DPQ070  36447 non-null  float64
 7   DPQ080  36447 non-null  float64
 8   DPQ090  36447 non-null  float64
dtypes: float64(9)
memory usage: 2.8 MB
```

In [15]:
```
df_target.head()
```

Out[15]:

| SEQN | DPQ010 | DPQ020 | DPQ030 | DPQ040 | DPQ050 | DPQ060 | DPQ070 | DPQ080 | DPQ090 |
|---|---|---|---|---|---|---|---|---|---|
| 31131 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 31132 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

3.Feature selection
4.Split the training and testing data



**Train Test Split**

In [6]:
```
# Loading in the data from the previous notebook
full_df = pd.read_csv('CSVFiles/FullData.csv', index_col='SEQN')
```

In [7]:
```
# Importing package
from sklearn.model_selection import train_test_split
```

In [8]:
```
# Defining X and y
X = full_df.drop('depression', axis=1)
y = full_df['depression'].map({'Not Depressed': 0, 'Depressed': 1})
```

In [9]:
```
# Performing train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=.20,
                                                    random_state=123,
                                                    stratify=y)
```

In [10]:
```
# Looking at the resulting shapes
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(29007, 490)
(7252, 490)
(29007,)
(7252,)
```

In [11]:
```
# Checking target value counts
y_train.value_counts(1)
```

Out[11]: 0 0.913

search 68°F

## 2.3 Validation methods

We are using train test split method where we are diving the data into train and test and validate the training set.

Cross-validation is a popular method used for estimating prediction error and model selection.

## Limitations

1.Inaccuracy to form subgroups.
2.Inability to determine which group is near to severe depression and mild depression.
3.Validity of the assessments score of PHQ-9.

## Future Work

We are done with preprocessing and also designed algorithm and have accuracy and we need to link our algorithm to website where we will provide data of user and predict the level of depression.

## References


1.World health Organization Depression Overview: https://www.who.int/news-room/fact-sheets/detail/depression

2. National Institute of Mental Health on Major Depression: https://www.nimh.nih.gov/health/statistics/major-depression.shtml

3. American Psychological Association Briefing Series on the Role of Psychology in Health Care: https://www.apa.org/health/briefs/primary-care.pdf

4. Association of Integrated Team-Based Care With Health Care Quality, Utilization, and Cost: https://jamanetwork.com/journals/jama/fullarticle/2545685

5. The PHQ-9: validity of a brief depression severity measure https://pubmed.ncbi.nlm.nih.gov/11556941/

6. Take the PHQ-9 depression screener online: https://www.mdcalc.com/phq-9-patient-health-questionnaire-9