## Prediction of level of depression of individuals

-by Anil, Jignesh & Rathan

### Abstract

This project aims to develop a machine-learning model that can classify an individual's depression state using other variables. We are using a dataset called Depression.csv, which has 16 features. The target feature of our project is the depression state of the individual, and we aim to accurately predict the level of depression (no depression, mild depression, moderate depression, severe depression) based on other features. Our approach involves preprocessing the data to handle missing values, scaling features for optimal model performance, and training and validating the model using robust machine-learning techniques. The successful implementation of this model could enhance mental health monitoring and support, contributing to better mental health outcomes.

### Introduction

Nowadays, Depression is one of the significant problems for many people, mainly Adults. It affects a person's mental health, resulting in anxiety, panic attacks, and damage to physical health. Detection of Depression is essential, as well as providing the proper treatment. Machine Learning (ML) is one of the emerging technologies that can help find hidden patterns in data using various techniques. This project aims to apply one of those machine learning (ML) techniques to develop a model that can classify the state of the Depression based on its severity.

### Methods

### In this project, we used many methodologies to build a reliable model for classifying the depression state. Firstly, we mapped the target variable string values to the numeric values: 0 for 'no depression,' 0 for 'mild depression,' 1 for 'moderate depression,' and 1 for 'severe depression.' Next, we used StandardScaler to scale the features so that each column contributes equally to the model prediction. Scaling is vital in classification, as it helps arrange the data within the same range. As our model mainly focuses on classification, we utilized a Support Vector Classifier (SVC), a classification model in Python from the sci-kit learn library. This library provides the implementation of this model simply and efficiently. We also used train\_test\_split from sklearn, which divides the data into train and test data. Finally, we used a confusion matrix to measure the model's accuracy, making detecting errors easy. The rows of this matrix represent the actual outcomes, whereas the columns represent the predictions of our model.

### Data Procurement:

Kaggle is one of the open-source platforms for extensive collections of data sets. We downloaded our data set from Kaggle, which provided relevant and adequate information for training our model.

# Big Data Analytics Life Cycle

### Stage1: Business Case Evaluation

The goal of our project is to identify the depression state of an individual based on other column values.

### Stage2: Data Identification

We collected our dataset from Kaggle website which is an internal source.

### Stage3: Data Acquisition and Filtering

import pandas as pd  
import numpy as np  
import pandas as pd  
from sklearn import svm  
from sklearn.model\_selection import train\_test\_split, cross\_val\_score  
from sklearn.preprocessing import StandardScaler  
from sklearn.metrics import classification\_report, confusion\_matrix

df=pd.read\_csv("/Users/rathanrajdasari/Downloads/Big Data/csvs/Deepression.csv")

df.columns

Index(['Number ', 'Sleep', 'Appetite', 'Interest', 'Fatigue', 'Worthlessness',  
 'Concentration', 'Agitation', 'Suicidal Ideation', 'Sleep Disturbance',  
 'Aggression', 'Panic Attacks', 'Hopelessness', 'Restlessness',  
 'Low Energy', 'Depression State'],  
 dtype='object')

### Stage4: Data Extraction

As our target variable is in object type we mapped those values to numbers to make it more compatible

maps={"Mild":0,"Moderate":1,"Severe":1,"No depression":0,"\tModerate":1,"\tNo depression":0,"\tSevere":1,"2\tNo depression":0,"5\tNo depression":0}

### Imputation

df.fillna(method='ffill',inplace=True)

### Stage5: Data Validation and Cleansing

As the 'Number' column is the id of the data frame and as it does not contribute anything to the target variable, we are deleting it

### Stage6: Data Aggregation and Representation

As we are dealing with only one dataset, this stage is skipped

### Stage7: Data Analysis

It includes finding correlation between the variables, target and numerical variables exploration, normalization, splitting the data, building a model, and finding performance of the model.

col=['Sleep', 'Appetite', 'Interest',  
 'Concentration', 'Agitation', 'Sleep Disturbance', 'Panic Attacks', 'Hopelessness',  
 'Low Energy']  
X=df[col]  
y=df['Depression State']

### Splitting the data

from sklearn.model\_selection import train\_test\_split  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X,y, test\_size=0.2, random\_state=42)

### Scaling the data using StandardScaler

from sklearn.preprocessing import StandardScaler  
sca = StandardScaler()  
X\_train = sca.fit\_transform(X\_train)  
X\_test = sca.transform(X\_test)

### Results

from sklearn import svm  
clf = svm.SVC(kernel='rbf')#, decision\_function\_shape='ovr') # 'ovr' stands for one-vs-rest  
  
# Train the SVM model  
clf.fit(X\_train, y\_train)  
  
Confusion Matrix:  
 [[37 26]  
 [13 32]]

### Efficiency of the model

efficency=1-((26+13)/(37+26+13+32))  
print("The efficiency of the model is:”, round(efficency,2))

The efficency of the model is: 0.64

### Stage8: Data Visualization

Visualizing the data

### Stage9: Utilization of Analysis results

The results can be used to find out the depression level of the individual based on the other variables.

### Conclusion

This project aims to develop a reliable model for detecting depression. We applied machine learning to the classification of depression levels of individuals, which involved data pre-processing, cleaning, validation, extraction, and analysis. By using predictive analytics, we perform effective classification to determine depression levels. It also involved StandardScaler for feature scaling, mapping the target variable, and classification using SVC. Finally, we used a confusion matrix to find the evaluation metrics of the model.

### Future work

This model can be integrated with the healthcare systems in assisting treatment and observing depression. Also, it can be combined with the existing mobile mental health applications in the assessment of depression and providing precautions.