# Depression Symptoms and Severity Levels in a Clinical Sample: A Comprehensive Analysis

# 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 so that 0 for 'no depression,' 1 for 'mild depression,' 2 for 'moderate depression,' and 3 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. In this project, we used many methodologies to build a reliable model for classifying the depression state. 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.

**Data Analytics Life Cycle:**

# Business Case Evaluation

# The objective is to create a machine learning model that uses 16 features from a dataset to categorize a person's depression status. By precisely estimating depression severity, the approach seeks to improve mental health monitoring and assistance.

1. **Data Identification**

Determine the information needed to accomplish the goal. In this instance, the "Depression.csv" dataset contains a number of characteristics pertaining to people's physical and mental well-being as well as their level of depression.

# Data Acquisition and Filtering

# The dataset was obtained for this project from Kaggle, an open-source website with large data sets. Filtering: Make sure the dataset only contains the target variables and pertinent features that are required for the study.

1. **Data extraction**

Take the data out of the received file and put it in a format that can be used for analysis.  
 Process: To put the data into a DataFrame for additional processing, read the "Depression.csv" file using Python libraries like pandas.

# Data validation and cleansing

# Verify the accuracy and consistency of the data. Procedure: Address any missing values and make any necessary corrections. As an illustration, you may eliminate duplicate records, impute missing values, and make sure all features are formatted correctly.

1. **Data aggregation and representation**

Arrange the information so that it may be analyzed.  
 Procedure: If needed, combine relevant aspects to aggregate data and show it in an organized manner. To make sure each feature contributes equally to the model prediction, scale them using StandardScaler.

# Data Analysis

# Utilize machine learning techniques for data analysis and model construction. Convert the desired variable (the state of depression) into numerical values. Train\_test\_split was used to divide the data into training and testing sets. Training data is used to train a Support Vector Classifier (SVC). Utilizing the testing data, assess how well the model performs.

1. **Data Visualization**

To comprehend the model's performance and spot any trends or insights, visualize the outcomes.  
 To demonstrate the model's accuracy and pinpoint the areas where it makes accurate and inaccurate predictions, use visualization tools (such as confusion matrices).

# Utilization of analysis results

Use the insights and forecasts from the model to accomplish the business goal.  
 Utilize the trained model to forecast a new person's likelihood of developing depression. The model's predictions can aid in the early identification of depression and offer prompt interventions, improving the results for mental health. Mental health specialists may also receive access to the results in order to assist them in making decisions.

# Future work:

# 1. Mobile and Web Applications (Internet of Everything)

* + **Mental Health applications:** Create or collaborate with current mental health applications that provide consumers/users with depression risk assessments, personalized advice, and resources.
  + **Telehealth Services:** Integrate the model with telehealth platforms such as IoT (Internet of Things) so that clinicians can assess patients' mental health remotely.