

Important Points

- The stated problem can be solved individually or in a group of maximum five.
- This project will follow viva and the performance in viva is key for good evaluation.
- The project must be submitted on or before the deadline. Late submission will have ZERO grade.
- Bring your laptop at the time of viva.
- Viva will be on 07-Feb-2025 at 02:00 to 04:00 PM

Analysis and Prediction of Mental Health Indicators Using Readability, Sentiment, and Behavioral Features

Objective: The objective of this project is to explore and analyze a dataset containing linguistic, psychological, and behavioral attributes, with a particular focus on mental health-related features. Using various Exploratory Data Analysis (EDA) techniques and predictive modeling approaches, we aim to:

1. Understand the relationships between the variables.
2. Perform predictive modeling on key mental health-related outcomes.
3. Evaluate the performance of Decision Tree (DT), k-Nearest Neighbors (KNN), and Linear Regression models for selected target variables.

Dataset Overview: The dataset contains features related to:

- **Psychological indicators** (e.g., economic stress, isolation, substance use, domestic stress).
- **Behavioral attributes** (e.g., Days_Indoors, Growing_Stress, Changes_Habits, Mood_Swings, Coping_Struggles, Work_Interest, Social_Weakness).
- **Categorical variables** such as Gender, Country, Occupation, self_employed, family_history, mental_health_interview, and care_options.

The dataset is suitable for analyzing psychological data in the context of mental health, with potential applications in understanding patterns and predicting outcomes. **Link of Dataset:**

<https://www.kaggle.com/datasets/bhavikjikadara/mental-health-dataset/data>

Task 1:

Explore the dataset, by using several types of Exploratory Data Analysis (EDA) techniques. Make the graphs and plots of following:

1. **Explore individual columns to understand their distributions and summary statistics.**
 - **Categorical Features** (e.g., Gender, Country, Occupation, self_employed, family_history, mental_health_interview, care_options):

- Frequency counts.
- Bar plots or pie charts to visualize category distributions.
- **Numerical Features** (e.g., Days_Indoors, Growing_Stress, Changes_Habits, Mood_Swings, Coping_Struggles, Work_Interest, Social_Weakness):
 - Descriptive statistics (mean, median, variance, etc.).
 - Histograms, boxplots, or density plots.
- 2. Examine relationships between two columns.**
 - **Numerical vs. Numerical:**
 - Scatter plots or correlation heatmaps (e.g., Days_Indoors vs. Growing_Stress).
 - **Numerical vs. Categorical:**
 - Boxplots or violin plots (e.g., Days_Indoors by Gender).
 - **Categorical vs. Categorical:**
 - Cross-tabulations and heatmaps for correlation (e.g., self_employed vs. family_history).
- 3. Analyze relationships among multiple variables.**
 - Pairplots or scatterplot matrices to study relationships across numerical columns.
 - Grouped bar plots (e.g., Gender, Occupation, and mental_health_interview).
- 4. Focus on features related to mental health:**
 - Analyze relationships between family_history, treatment, and mental_health_interview.
 - Study the impact of Days_Indoors on Work_Interest or Social_Weakness.
 - Assess how care_options availability varies by Country or Gender.
- 5. Understand how features are related:**
 - Correlation matrices to identify highly related features.
- 6. Detect unusual values in numerical columns:**
 - Boxplots and z-scores.
 - Influence of outliers on mental health features like Mood_Swings or Growing_Stress.

Task 2:

The target variable (or things to be predicted) should ideally be chosen based on the context of the dataset and the problem you're trying to solve. Given the columns provided, potential predictions could be related to mental health indicators or behavior patterns. Here are some possibilities:

1. Predicting Mental Health Treatment

- **Target Variable:** treatment
- **Objective:** Predict whether an individual is likely to seek or require mental health treatment based on factors like family history, stress levels, mood swings, and social weakness.

2. Predicting Days Spent Indoors

- **Target Variable:** Days_Indoors
- **Objective:** Estimate how many days a person spends indoors based on stress levels, coping struggles, and other behavioral traits.

3. Predicting Stress Levels

- **Target Variable:** Growing_Stress
- **Objective:** Use features such as Days_Indoors, Mood_Swings, and Coping_Struggles to predict an individual's stress level.

4. Predicting Changes in Habits

- **Target Variable:** Changes_Habits
- **Objective:** Predict the likelihood or extent of changes in habits based on mental health history, mood swings, and stress factors.

5. Predicting Work Interest

- **Target Variable:** Work_Interest
- **Objective:** Determine if stress, social weakness, or mental health factors impact a person's interest in work.

6. Predicting Likelihood of Mental Health Issues

- **Target Variable:** Mental_Health_History
- **Objective:** Predict whether a person has a history of mental health issues based on current stressors, family history, and coping mechanisms.

Key Considerations for Selection

- Evaluate model performance using metrics:
 - Classification targets: Accuracy, precision, recall, F1-score.
 - Regression targets: Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared (R^2).
- Comparative Analysis:
 - Compare models based on metrics.
- Assess trade-offs between simplicity, interpretability, and performance.

Task 3:

Comprehensive Report:

1. Introduction:

- Overview of the project and dataset.
- Importance of mental health analysis using behavioral features.

2. EDA Findings:

- Key patterns and distributions in categorical and numerical variables.
- Relationships and correlations between features.
- Insights into mental health-related attributes.

3. Modeling Results:

- Description of the models and target variables.
- Performance metrics for DT, KNN, and Linear Regression.
- Comparison of models and key observations.

4. Discussion:

- Implications of findings for mental health analysis.
- Potential applications of predictive models in real-world scenarios.

5. Conclusion and Recommendations:

- Summary of insights and model performance.
- Suggestions for improving the dataset and extending the analysis.