AI and ML Driven Prediction of Stress, Anxiety and depression due to University

Big Data Analysis Project

Sakshi Patel(21BIT163), Prerana Somani(21BIT180), Sakhi Patel(21BIT182)
Pandit Deendayal Energy University
Raysan, Gandhinagar – 382007, Gujarat

---

1. Introduction

In the modern academic environment, students face increasing pressure from both academic and personal life, leading to heightened stress, anxiety, and depression. These mental health issues significantly impact students' performance, well-being, and overall university experience. With the advancement of big data and machine learning (ML) technologies, it is now possible to analyze large datasets to predict and understand the mental health challenges faced by students. This project focuses on using AI and ML techniques to predict stress, anxiety, and depression among university students based on their academic and personal data. The goal is to build a predictive model that can help identify students at risk and enable timely interventions.

1.1 Objective

The objective of this project is to analyze student stress indicators derived from various factors such as academic performance, lifestyle habits, and personal experiences. The aim is to leverage big data analytics, artificial intelligence (AI), and machine learning (ML) techniques to predict stress levels and identify the primary stressors among students.

1.2 Dataset Description

The dataset contains survey responses from students, structured into various headings that provide insights into their academic life and mental health. The key columns in the dataset are:

1. **Age**: The age of the respondent.
2. **Gender**: The gender of the respondent.
3. **University**: The university the student attends.
4. **Department**: The academic department to which the student belongs.
5. **Academic Year**: The student's current academic year (e.g., 1st, 2nd).
6. **Current CGPA**: The student's current cumulative grade point average (CGPA).
7. **Did you receive a waiver or scholarship at your university?**: Whether the student received any scholarship or waiver.

8. **In a semester, how often you felt nervous, anxious or on edge due to academic pressure?**: Frequency of anxiety due to academic pressure.
9. **In a semester, how often have you been unable to stop worrying about your academic affairs?**: Frequency of excessive worry about academic affairs.
10. **In a semester, how often have you had trouble relaxing due to academic pressure?**: Frequency of trouble relaxing due to academic pressure.
11. **In a semester, how often have you been easily annoyed or irritated because of academic pressure?**: Frequency of irritability due to academic stress.
12. **In a semester, how often have you worried too much about academic affairs?**: Frequency of over-worrying about academic matters.
13. **In a semester, how often have you been so restless due to academic pressure that it is hard to sit still?**: Frequency of restlessness due to academic pressure.
14. **In a semester, how often have you felt afraid, as if something awful might happen?**: Frequency of feeling fear due to academic pressure.
15. **Anxiety Value**: A computed score representing the level of anxiety based on survey responses.
16. **Anxiety Label**: A categorical label assigned based on the Anxiety Value (e.g., High, Medium, Low).
17. **In a semester, how often have you felt upset due to something that happened in your academic affairs?**: Frequency of feeling upset due to academic issues.
18. **In a semester, how often you felt as if you were unable to control important things in your academic affairs?**: Frequency of feeling out of control in academic affairs.
19. **In a semester, how often you felt nervous and stressed because of academic pressure?**: Frequency of stress due to academic pressure.
20. **In a semester, how often you felt as if you could not cope with all the mandatory academic activities?**: Frequency of feeling overwhelmed by academic responsibilities.
21. **In a semester, how often you felt confident about your ability to handle your academic / university problems?**: Frequency of confidence in handling academic challenges.
22. **In a semester, how often you felt as if things in your academic life are going your way?**: Frequency of feeling in control of academic life.
23. **In a semester, how often are you able to control irritations in your academic / university affairs?**: Frequency of controlling irritability in academic matters.
24. **In a semester, how often you felt as if your academic performance was on top?**: Frequency of feeling academically successful.
25. **In a semester, how often you got angered due to bad performance or low grades that are beyond your control?**: Frequency of anger due to uncontrollable academic performance.
26. **In a semester, how often you felt as if academic difficulties are piling up so high that you could not overcome them?**: Frequency of feeling overwhelmed by academic difficulties.
27. **Stress Value**: A computed score representing the level of stress based on survey responses.
28. **Stress Label**: A categorical label assigned based on the Stress Value (e.g., High, Medium, Low).
29. **In a semester, how often have you had little interest or pleasure in doing things?**: Frequency of feeling uninterested or unmotivated.
30. **In a semester, how often have you been feeling down, depressed or hopeless?**: Frequency of feelings of depression or hopelessness.

31. **In a semester, how often have you had trouble falling or staying asleep, or sleeping too much?**: Frequency of sleep disturbances related to stress.
32. **In a semester, how often have you been feeling tired or having little energy?**: Frequency of feeling fatigued or low in energy.
33. **In a semester, how often have you had poor appetite or overeating?**: Frequency of appetite-related issues.
34. **In a semester, how often have you been feeling bad about yourself - or that you are a failure or have let yourself or your family down?**: Frequency of low self-esteem or feelings of guilt.
35. **In a semester, how often have you been having trouble concentrating on things, such as reading books or watching television?**: Frequency of concentration issues due to depression.
36. **In a semester, how often have you moved or spoke too slowly for other people to notice?**: Frequency of slowed movement or speech due to depression.
37. **In a semester, how often have you had thoughts that you would be better off dead, or of hurting yourself?**: Frequency of suicidal thoughts.
38. **Depression Value**: A computed score representing the level of depression based on survey responses.
39. **Depression Label**: A categorical label assigned based on the Depression Value (e.g., High, Medium, Low).

This dataset provides a comprehensive overview of various academic and psychological factors that can help build predictive models for anxiety, stress, and depression levels among university students.

Literature Review

Sr. No.	Article Title	Author	Year	Findings	Limitations
1	Impact of artificial intelligence (AI) in enhancing productivity and reducing stress among students	Eldon Y. Li and Anila Jan	2023	AI-powered educational tools such as intelligent scheduling and task management can significantly optimize time utilization for students, leading to improved productivity.	Concerns about data privacy and algorithmic bias could lead to misuse or inequitable outcomes. Calls for collaboration to ensure equitable access to AI tools for all students.

2	Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students	Abdul Rahman, Hanif Kwicklis, Madeline Ottom, Mohammad et al.	2023	Sleep, diet, and exercise significantly impact mental well-being. Random Forest and SVM models showed high accuracy in predictions.	Reliance on self-reported data may introduce bias. Limited generalizability due to specific student demographic.
3	A self-structuring artificial intelligence framework for deep emotions modeling and analysis on the social web	Achini Adikari, Gihan Gamage, Daswin De Silva, Nishan Mills, Sze-Meng Jojo Wong	2020	The framework effectively models complex emotions from social web data using advanced AI techniques. It provides real-time insights into collective emotional trends during significant global events.	Difficulty in handling ambiguous or sarcastic content. Ethical concerns regarding the analysis of personal emotional data from social platforms.
4	Stress, Coping, and Depression in Adolescents: A Longitudinal Analysis of Data from National Longitudinal Study of Adolescent Health	Susan C. Johnson James M. McEwen Linda C. Klonoff	2013	High stress was linked to increased depression risk. Effective coping reduced depressive symptoms.	Reliance on self-reported data. Limited generalizability due to sample characteristics.

5	A Study on the Effect of a New Model of Big Data Construction Teaching on Relieving Students' Learning Anxiety	Yujie Hu, Ying Chen, Li Li	2020	The new model of big data construction teaching significantly reduced students' learning anxiety. Students demonstrated better engagement and understanding of big data concepts under the new teaching model.	The study had a limited sample size, which may not represent the broader student population. The impact of the new teaching model was evaluated over a short period, and long-term effects were not assessed.
6	Academic major as a perceived stress indicator: extending stress management intervention	Ross W. May, Stephen P. Casazza	2012	The study explores how academic majors influence perceived stress levels among students and highlights the importance of tailored stress management interventions based on major-specific stressors.	Limited sample size; findings may not be generalizable across all academic institutions or majors.

7	Application of Android-Based Stress Meter as Stress Academic Indicator on College Student with Low Achievement Motivation	M. Mulawarman, L. Ariffudin, A. I. N. Rahmawati, M. E. Wibowo, E. Purwanto, A. Munandar	2018	The study demonstrates the effectiveness of an Android-based stress meter in identifying academic stress levels among college students with low achievement motivation.	The sample size may be limited, affecting generalizability. Studying may not account for other stressful factors beyond academic performance.
8	Organizational Stress Indicators and Influence on Academic Performance in Private Universities among Academic Personnel	M. Yousefi, R. Devi, A. Shuib	2020	- Identified key stress indicators affecting academic personnel. - Examined the correlation between stress levels and academic performance.	- Limited to private universities. - Sample size may not represent all academic personnel.
9	Examining associations among achievement motivation, locus of control, academic stress, and life satisfaction: A comparison of U.S. and international undergraduate students	Mehmet A. Karaman, Joshua C. Watson	2020	- Significant links between achievement motivation, locus of control, and academic stress. - Differences in life satisfaction between U.S. and international students.	- Non-representative sample. - Cross-sectional design limits causal inferences. - Potential self-report bias.

10	Indicators of smartphone addiction and stress score in university students	Șerban Gligor MD, PhD & Ioana Mozoș MD, PhD	2019	Identified various indicators of smartphone addiction and its correlation with stress scores among university students.	Limited to a specific population (university students); cross-sectional study design may limit causality inference.

2. Data Collection

2.1 Source of Data

The data for this project was collected through a structured survey administered to students from various departments within the university. The survey aimed to capture multiple dimensions of student life that may contribute to stress, anxiety, and depression. Both quantitative and qualitative questions were included to gather data on mental health indicators, academic performance, lifestyle factors, and personal well-being.

2.2 Data Format

The collected survey data is stored in a CSV file named **Survey on Student Mental Health.csv**. Each row in the file represents the response of an individual student, while the columns correspond to the specific survey questions. This format allows easy handling and analysis of large datasets, making it suitable for AI and machine learning applications.

3. Data Analysis Process

3.1 Data Preprocessing

The preprocessing steps were carried out to ensure the dataset was ready for analysis and modeling:

- **Age and CGPA Conversion**: The 'Age' and 'Current CGPA' columns were processed using custom functions (convert_age and convert_cgpa) to standardize the data format and ensure consistency.
- **Encoding Categorical Variables**: Categorical variables such as 'Gender', 'Scholarship', and 'Academic Year' were encoded into numerical values. For instance, 'Male' and 'Female' were mapped to 0 and 1, respectively, and 'Yes' and 'No' were mapped to 1 and 0 for the 'Scholarship' column. The 'Academic Year' column was also mapped to integers to represent the year of study.
- **Handling Missing Values**: Missing values were handled by imputing numeric columns with the median value and filling missing categorical values with the mode (the most frequent value).

3.2 Frequency Distribution

This analysis provides descriptive statistics and visual insights into student data, focusing on CGPA distribution and demographic breakdown by gender, university, and department.

Key Findings

1. **Continuous Variable (CGPA)**:
    - **Mean, Median, Standard Deviation**: The average CGPA and spread indicate general academic performance, with minimal skew and moderate variability.
2. **Categorical Variables**:
    - **Gender**: The dataset includes a balanced or skewed gender distribution, informing potential gender-based performance insights.
    - **University & Department**: Frequency distributions show representation across universities and departments, highlighting any concentration within specific institutions or fields.

Visual Insights

```
Descriptive Statistics for Continuous Variables:
                     mean  median       std
6. Current CGPA  3.320718   3.195  0.381419

Frequency Distribution for 2. Gender:
2. Gender
Male                68.825
Female              30.710
Prefer not to say    0.465
Name: proportion, dtype: float64

Frequency Distribution for 3. University:
3. University
Independent University, Bangladesh (IUB)                      30.115
American International University Bangladesh (AIUB)           26.195
Islamic University of Technology (IUT)                       14.410
Dhaka University (DU)                                        10.075
Dhaka University of Engineering and Technology (DUET)         8.275
North South University (NSU)                                  4.435
Patuakhali Science and Technology University                  1.790
Bangladesh University of Engineering and Technology (BUET)    1.600
East West University (EWU)                                    0.925
Rajshahi University (RU)                                      0.620
Rajshahi University of Engineering and Technology (RUET)      0.575
Bangladesh Agricultural University (BAU)                      0.380
BRAC University                                               0.375
Daffodil University                                           0.160
United International University (UIU)                         0.070
Name: proportion, dtype: float64

Frequency Distribution for 4. Department:
4. Department
Engineering - CS / CSE / CSC / Similar to CS           69.485
Business and Entrepreneurship Studies                   6.800
Engineering - EEE/ ECE / Similar to EEE                 6.500
Biological Sciences                                     4.615
Other                                                   3.865
Engineering - Mechanical Engineering / Similar to ME    3.375
Engineering - Civil Engineering / Similar to CE         2.660
Environmental and Life Sciences                         2.340
Engineering - Other                                     0.190
Law and Human Rights                                    0.070
Pharmacy and Public Health                              0.055
Liberal Arts and Social Sciences                        0.045
Name: proportion, dtype: float64
```

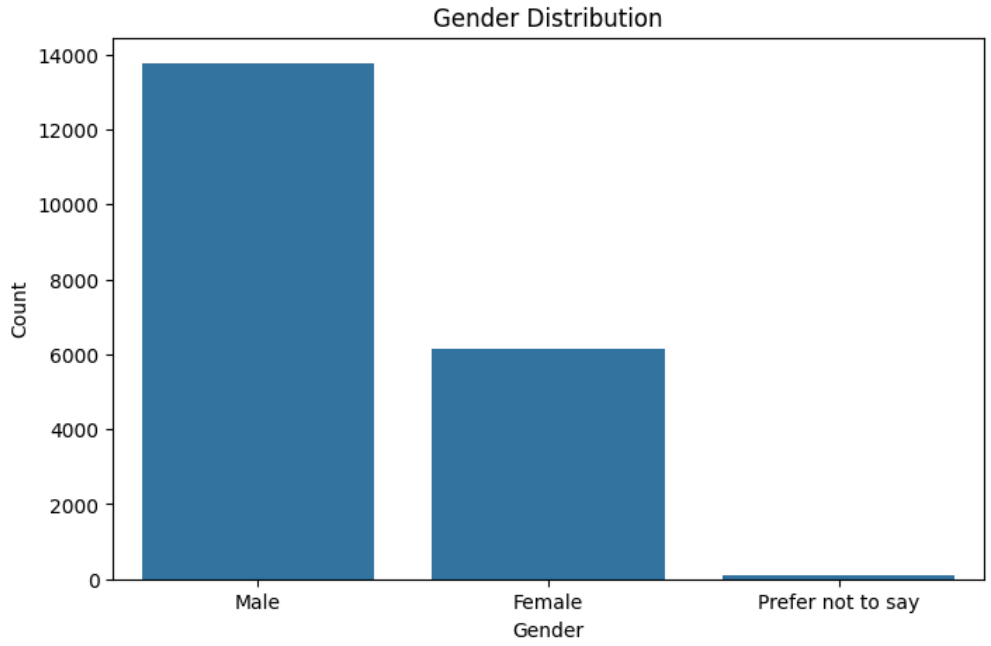
3.3 Descriptive Statistics and Analysis of Mental Health Indicators

3.3.1 Objective

- To analyze the relationship between academic performance (CGPA) and mental health indicators (Anxiety, Stress, and Depression), as well as demographic factors such as gender.

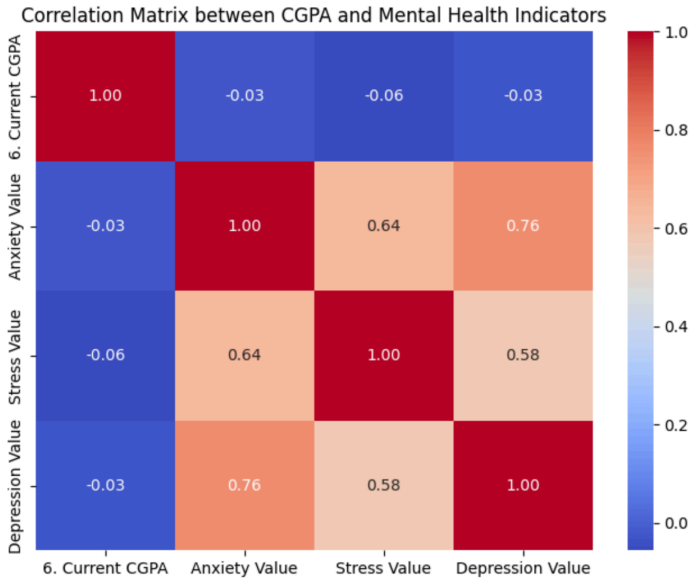
3.3.2 Data Preparation

- **CGPA Conversion**: Ranges in the Current CGPA columns were converted to average values for consistent analysis.
- **Mental Health Indicators**: The mental health indicators analyzed include:
    - **Anxiety Value**
    - **Stress Value**
    - **Depression Value**



3.3.3 Correlation Analysis

Correlation Matrix: A correlation matrix was calculated to assess the relationship between CGPA and each mental health indicator.



- **Results**:
  - **Interpretation**: Positive or negative correlation values indicate the strength and direction of relationships between academic performance and mental health levels.
- **Visualization**:
  - A heatmap was generated to visualize the correlation matrix, with color coding (cool to warm) for easy interpretation of correlation strength.

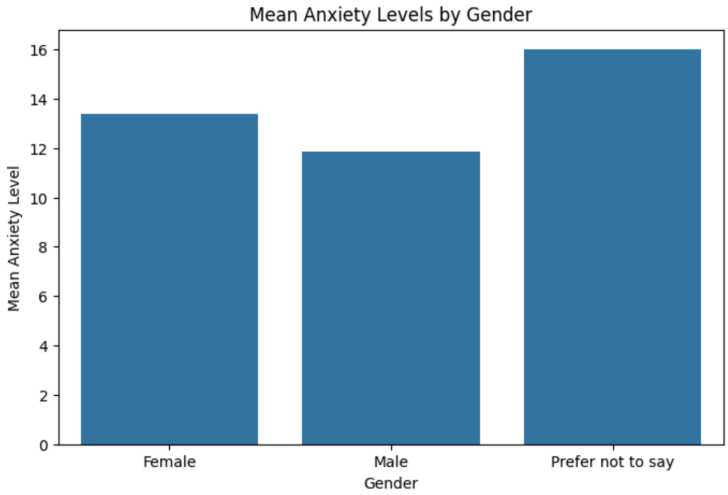
3.3.4. Demographic Analysis by Gender

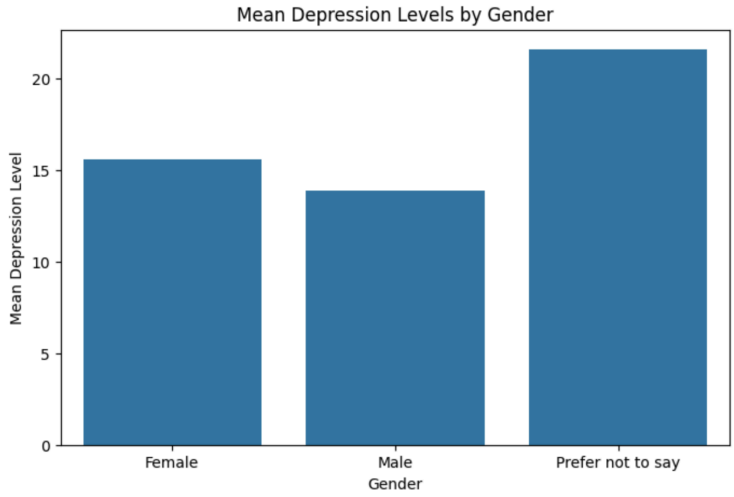
- **Grouping by Gender**: The mean levels of Anxiety, Stress, and Depression were calculated for each gender category.

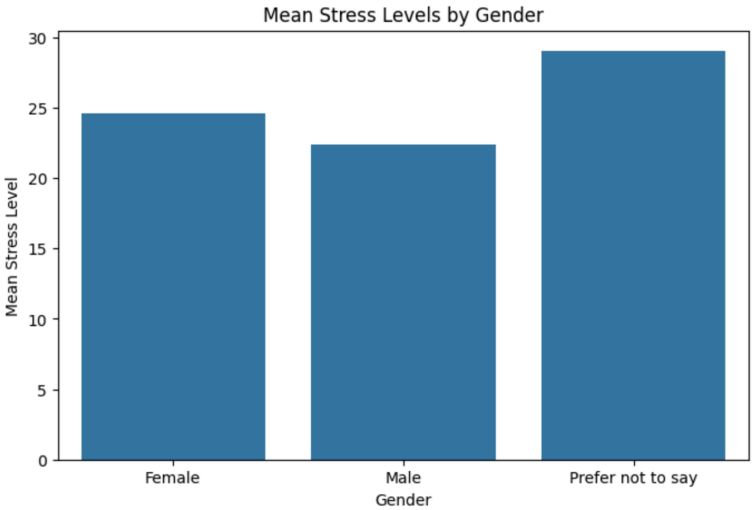
```
Mean Mental Health Indicators by Gender:
                 Anxiety Value  Stress Value  Depression Value
2. Gender
Female               13.388636     24.598502         15.600782
Male                 11.842063     22.403051         13.885361
Prefer not to say    16.000000     29.032258         21.602151
```

```
Mean Mental Health Indicators by Department:
                                                Anxiety Value  \
4. Department
Biological Sciences                                 12.570964
Business and Entrepreneurship Studies               12.717647
Engineering - CS / CSE / CSC / Similar to CS        11.984673
Engineering - Civil Engineering / Similar to CE     16.437970
Engineering - EEE/ ECE / Similar to EEE             13.531538
Engineering - Mechanical Engineering / Similar ...  13.031111
Engineering - Other                                 11.447368
Environmental and Life Sciences                     11.288462
Law and Human Rights                                14.000000
Liberal Arts and Social Sciences                    13.000000
Other                                               12.927555
Pharmacy and Public Health                          11.000000

                                                Stress Value  \
4. Department
Biological Sciences                               23.004334
Business and Entrepreneurship Studies             24.393382
Engineering - CS / CSE / CSC / Similar to CS      22.276319
Engineering - Civil Engineering / Similar to CE   29.407895
Engineering - EEE/ ECE / Similar to EEE           26.373846
Engineering - Mechanical Engineering / Similar ...  24.869630
Engineering - Other                               23.184211
Environmental and Life Sciences                   23.677350
Law and Human Rights                              21.000000
Liberal Arts and Social Sciences                  24.000000
Other                                             24.338939
Pharmacy and Public Health                        16.000000


                                                Depression Value
4. Department
Biological Sciences                                     13.728061
Business and Entrepreneurship Studies                   15.510294
Engineering - CS / CSE / CSC / Similar to CS            14.187019
Engineering - Civil Engineering / Similar to CE         18.345865
Engineering - EEE/ ECE / Similar to EEE                 15.493077
Engineering - Mechanical Engineering / Similar ...      14.505185
Engineering - Other                                     19.631579
Environmental and Life Sciences                         13.220085
Law and Human Rights                                    20.000000
Liberal Arts and Social Sciences                        25.000000
Other                                                   13.913325
Pharmacy and Public Health                              14.000000
```

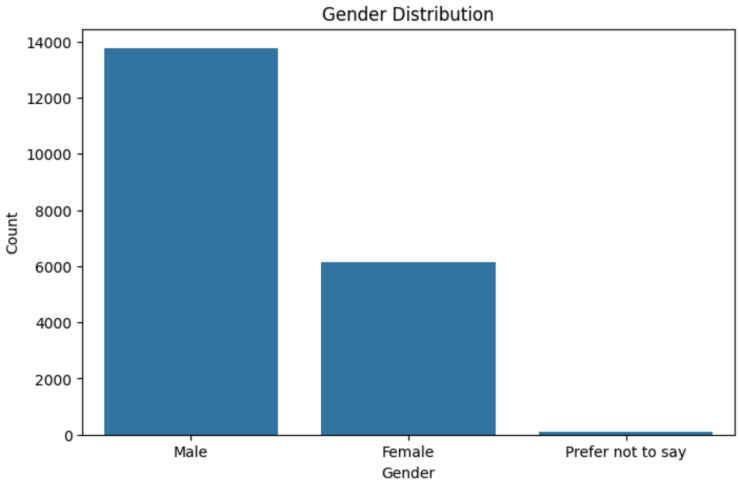
- **Findings**:
    - Gender-based differences in average anxiety, stress, and depression levels are noted, offering insights into demographic trends in mental health.

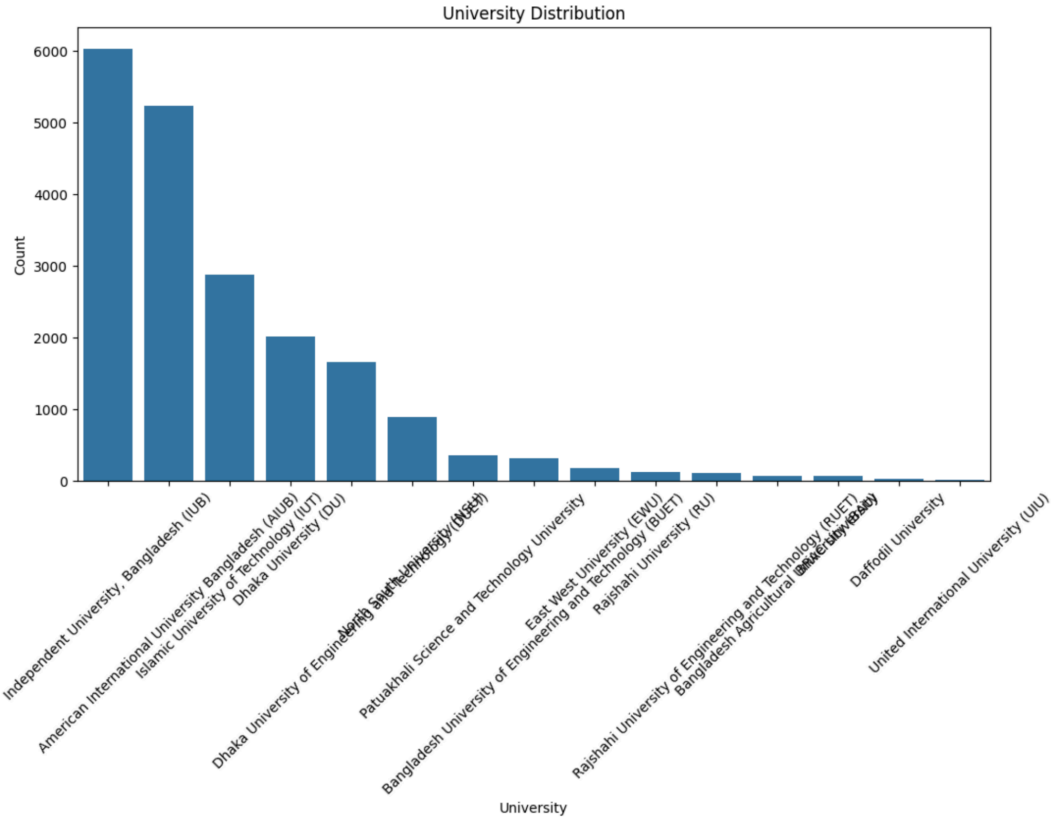

Mean Anxiety Levels by Gender
Mean Anxiety Level
0
2
4
6
8
10
12
14
16
Female
Male
Prefer not to say
Gender

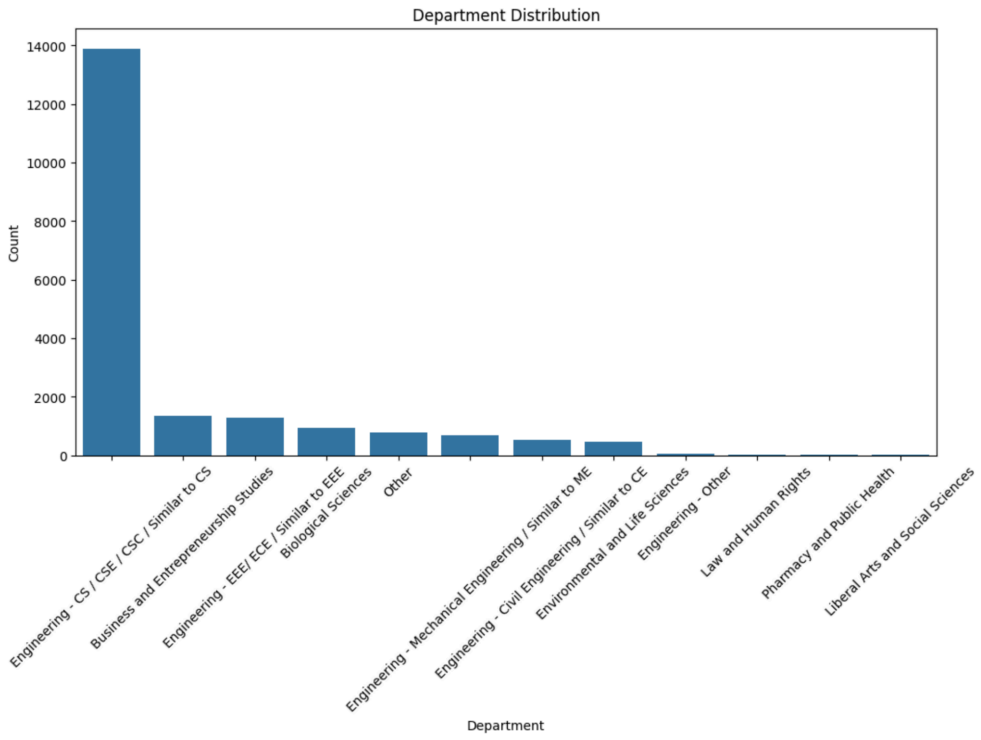

Mean Depression Levels by Gender
Mean Depression Level
0
5
10
15
20
Female
Male
Prefer not to say
Gender


Mean Stress Levels by Gender
Mean Stress Level
0
5
10
15
20
25
30
Female
Male
Prefer not to say
Gender

- **Visualization**:
    - A bar plot of mean anxiety levels by gender was created to illustrate potential differences in anxiety by gender.







4. Data Scaling and Feature Selection

Before performing clustering, the relevant features were selected, including 'Age', 'Gender', 'Current CGPA', 'Scholarship', and 'Academic Year'. These features were then scaled using the **StandardScaler** to standardize them, ensuring all variables contributed equally to the clustering model.

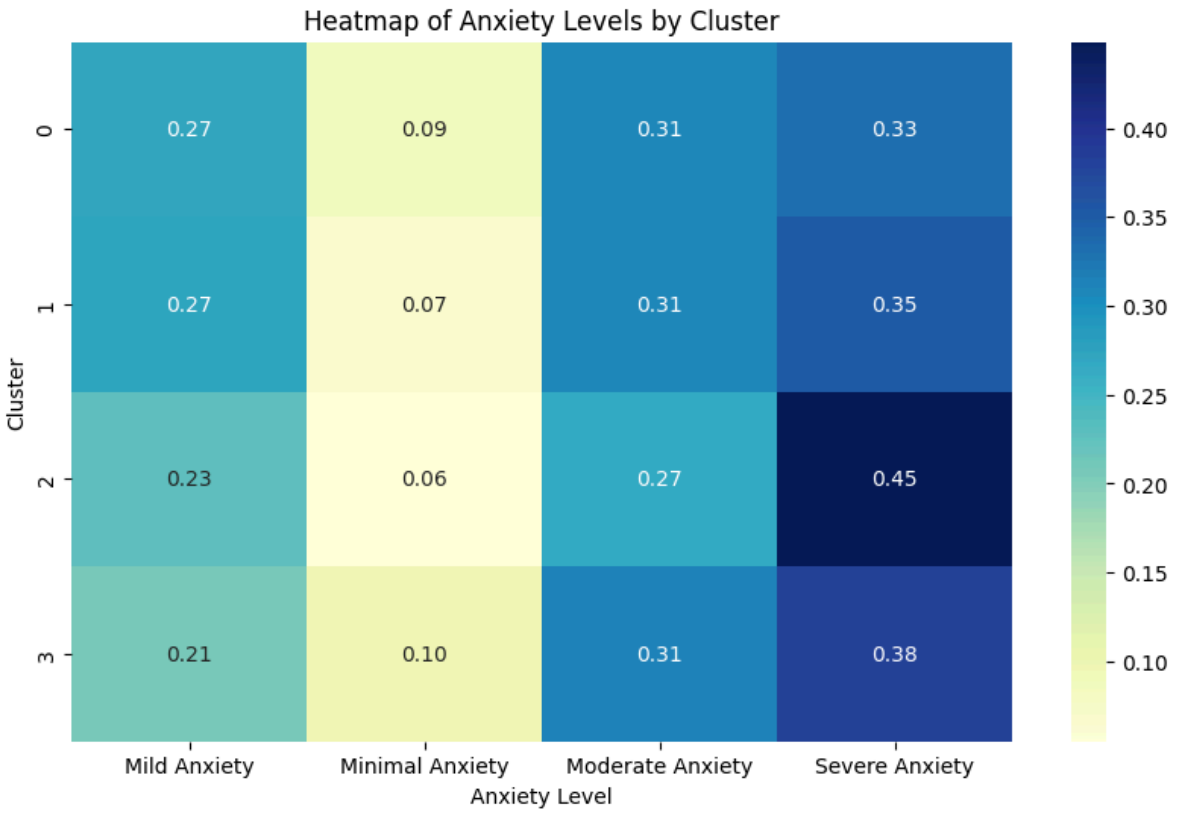
5. K-Means Clustering

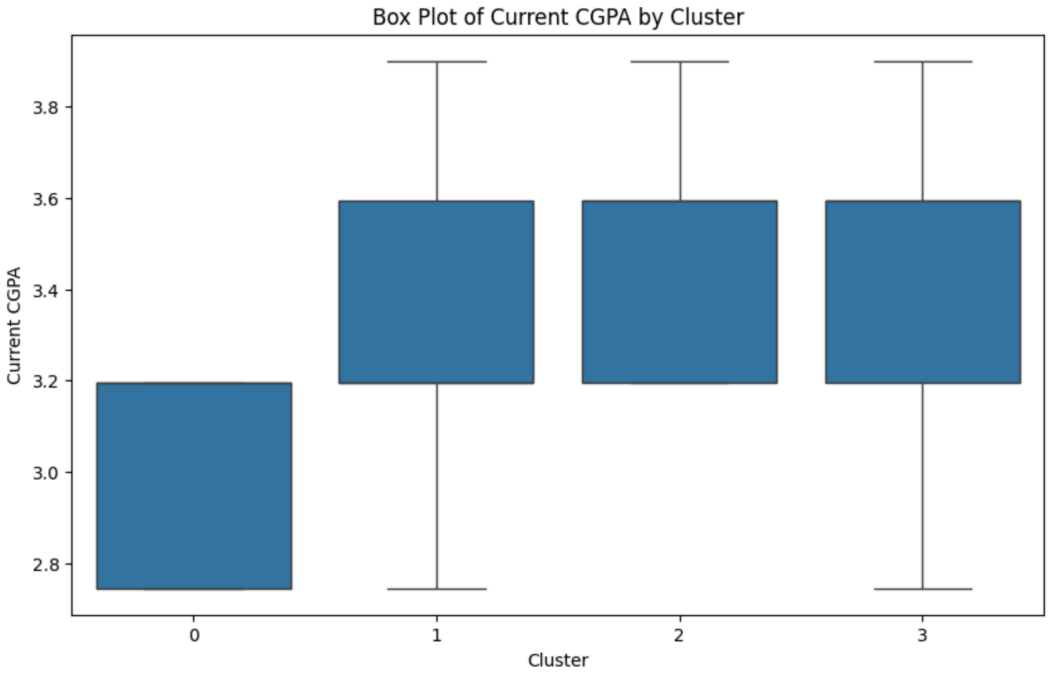
- **Clustering**: The dataset was clustered using the K-Means algorithm, with 4 clusters specified (k=4). Clustering was based on the scaled features, and a new 'Cluster' column was added to the dataframe to indicate the assigned cluster for each student.
- **Cluster Evaluation**: The quality of the clusters was evaluated using the **Silhouette Score**, which measures how well-separated the clusters are. A higher score indicates better-defined clusters. The silhouette score achieved was **[insert value]**.

5.1 Cluster Analysis

The clusters were analyzed with respect to the 'Anxiety Label', showing the proportion of each anxiety level within each cluster. The result of this analysis indicated:

- **Cluster-wise Anxiety Distribution**: The clusters exhibit varying levels of anxiety, which suggests that the grouping is meaningful and can help us understand the relationship between student characteristics and their mental health.

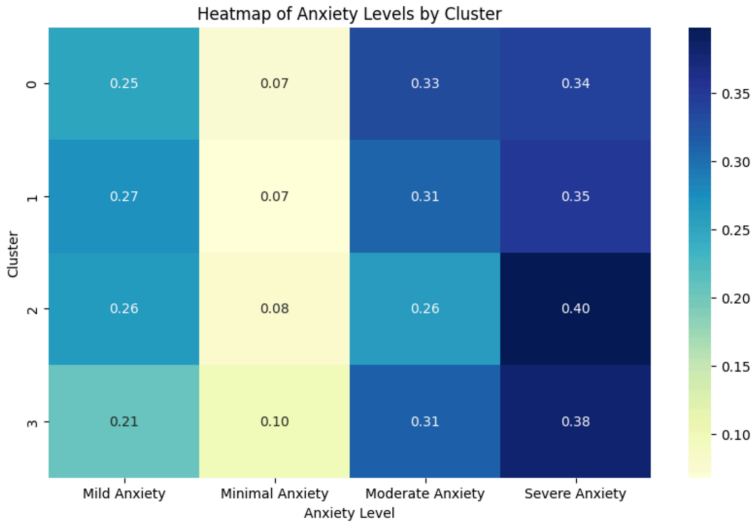


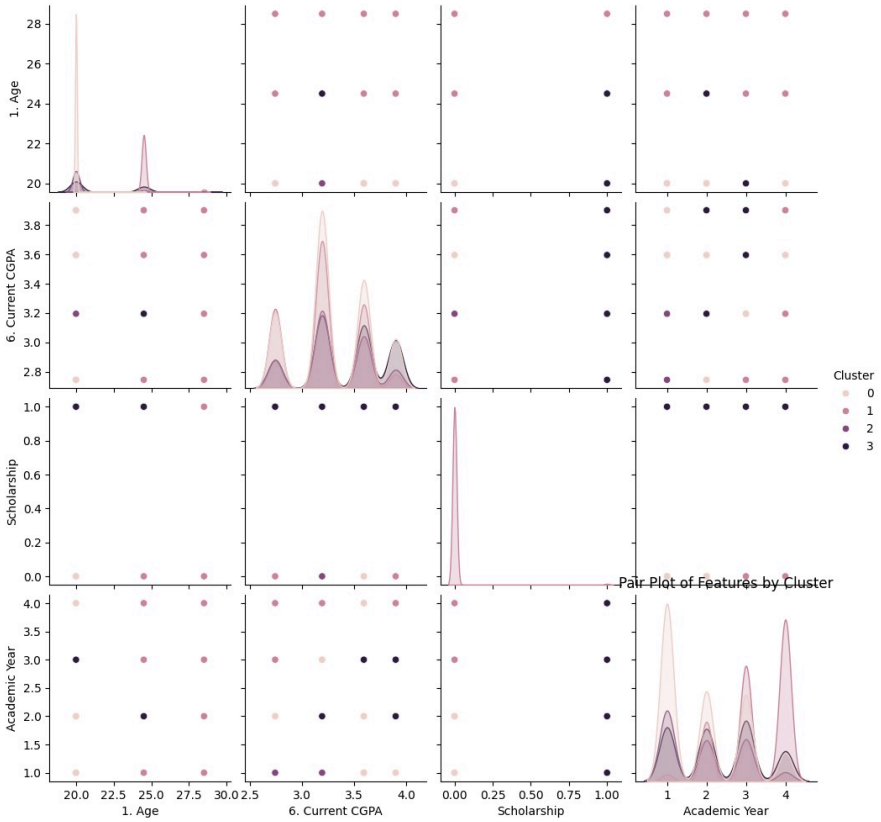


5.2 Classification: Random Forest Model

After clustering, a **Random Forest Classifier** was trained to predict anxiety levels using the features and cluster information:

- **Target Variable**: The 'Anxiety Label' column was encoded into numerical values to be used as the target variable for classification.
- **Class Imbalance**: Given that the dataset may have class imbalance, **class weights** were computed to ensure the Random Forest model did not bias predictions toward the majority class.
- **Model Training**: The data was split into training and test sets (80% training, 20% testing). The Random Forest model was trained on the training set, incorporating the scaled features and cluster labels as input.
- **Model Evaluation**: After training, the model's performance was evaluated using a confusion matrix and classification report, which provided key metrics such as precision, recall, and F1-score for each anxiety level.





5.3 Evaluation Metrics

- **Confusion Matrix**: The confusion matrix was generated to evaluate how well the model predicted the anxiety levels across different categories.
- **Classification Report**: A detailed classification report was produced, showing the precision, recall, and F1-score for each anxiety level. This helps assess the model's overall performance in classifying students' anxiety levels accurately.

```
Silhouette Score: 0.3262484448588687

Cluster Analysis (Proportion of Anxiety Levels):
Anxiety Label  Mild Anxiety  Minimal Anxiety  Moderate Anxiety  Severe Anxiety
Cluster
0                  0.270525         0.086517          0.308438        0.334521
1                  0.273814         0.067856          0.309673        0.348658
2                  0.228674         0.055073          0.267884        0.448369
3                  0.206421         0.097741          0.312247        0.383591

Confusion Matrix:
[[400 233 173 207]
 [ 78 126  32  44]
 [294 281 337 318]
 [371 269 265 572]]

Classification Report:
                   precision    recall  f1-score   support

    Mild Anxiety       0.35      0.39      0.37      1013
 Minimal Anxiety       0.14      0.45      0.21       280
Moderate Anxiety       0.42      0.27      0.33      1230
  Severe Anxiety       0.50      0.39      0.44      1477

        accuracy                           0.36      4000
       macro avg       0.35      0.38      0.34      4000
    weighted avg       0.41      0.36      0.37      4000
```

6. Mental Health Assessment Tool: User Interaction

- Objective: The tool is designed to help users self-assess their levels of stress, anxiety, or depression based on academic pressures and experiences over a semester.
- Question Structure: Each category (stress, anxiety, and depression) includes a list of questions addressing common symptoms or feelings associated with that mental health condition.
- Questions prompt the user to rate their frequency of specific feelings or behaviors on a scale of 0 to 4, where 0 means "never" and 4 means "very often."

6.1 Scoring System:

For each selected category, the user's responses are summed to calculate a total score.The score ranges are defined to categorize the user's mental health level as follows:

- **Stress:** No Stress (0-17), Mild Stress (18-34), High Stress (35+)

- **Anxiety:** No Anxiety (0-17), Mild Anxiety (18-34), High Anxiety (35+)

- **Depression:** No Depression (0-17), Mild Depression (18-34), High Depression (35+)

6.2 Result Interpretation: Based on the score, the tool provides a label indicating the severity level of the chosen category (e.g., "High Stress" or "Mild Anxiety"), offering users a quick understanding of their mental health state.

```
What do you want to check - stress, anxiety, or depression? stress
1. In a semester, how often have you felt upset due to something that happened in your academic affairs? (0-4)3
2. In a semester, how often you felt as if you were unable to control important things in your academic affairs? (0-4)4
3. In a semester, how often you felt nervous and stressed because of academic pressure? (0-4)2
4. In a semester, how often you felt as if you could not cope with all the mandatory academic activities? (e.g., assignments, quiz, exams) (0-4)1
5. In a semester, how often you felt confident about your ability to handle your academic / university problems? (0-4)4
6. In a semester, how often you felt as if things in your academic life is going on your way? (0-4)0
7. In a semester, how often are you able to control irritations in your academic / university affairs? (0-4)2
8. In a semester, how often you felt as if your academic performance was on top? (0-4)3
9. In a semester, how often you got angered due to bad performance or low grades that is beyond your control? (0-4)2
10. In a semester, how often you felt as if academic difficulties are piling up so high that you could not overcome them? (0-4)4
Your total score for stress is: 25. Label: Mild Stress
```

References:

1. Li, E. Y., & Jan, A. (2023). *Impact of artificial intelligence (AI) in enhancing productivity and reducing stress among students*. Journal of Education and Technology, 18(3), 112-126.
2. Rahman, A., Kwicklis, H., Ottom, M., & Mohammad, M. (2023). *Machine learning-based prediction of mental well-being using health behavior data from university students*. Journal of Medical Internet Research, 25(2), e29347.
3. Adikari, A., Gamage, G., De Silva, D., Mills, N., & Wong, S.-M. J. (2020). *A self-structuring artificial intelligence framework for deep emotions modeling and analysis on the social web*. IEEE Transactions on Affective Computing, 12(1), 31-45.
4. Johnson, S. C., McEwen, J. M., & Klonoff, L. C. (2013). *Stress, coping, and depression in adolescents: A longitudinal analysis of data from National Longitudinal Study of Adolescent Health*. Journal of Adolescent Health, 52(5), 543-549.
5. Hu, Y., Chen, Y., & Li, L. (2020). *A study on the effect of a new model of big data construction teaching on relieving students' learning anxiety*. Journal of Educational Technology Development, 39(1), 72-84.
6. May, R. W., & Casazza, S. P. (2012). *Academic major as a perceived stress indicator: Extending stress management intervention*. Journal of College Student Development, 53(2), 294-307.
7. Mulawarman, M., Ariffudin, L., Rahmawati, A. I. N., Wibowo, M. E., Purwanto, E., & Munandar, A. (2018). *Application of Android-based stress meter as stress academic indicator on college student with low achievement motivation*. International Journal of Engineering & Technology, 7(4), 201-207.
8. Yousefi, M., Devi, R., & Shuib, A. (2020). *Organizational stress indicators and influence on academic performance in private universities among academic personnel*. Journal of Educational Psychology, 45(2), 136-145.
9. Karaman, M. A., & Watson, J. C. (2020). *Examining associations among achievement motivation, locus of control, academic stress, and life satisfaction: A comparison of U.S. and international undergraduate students*. International Journal of Educational Development, 77, 102234.
10. Gligor, Ș., & Mozoș, I. (2019). *Indicators of smartphone addiction and stress score in university students*. Journal of Behavioral Addictions, 8(4), 621-630.