KLE Technological University
Creating Value, Leveraging Knowledge
Dr. M. S. Sheshgiri Campus, Belagavi

# Department of Electronics and Communication Engineering

Machine Learning
Course Project
on

# Anxiety Prediction Of Student Behavior Based On Day To Day Life


**By:**

| **Name** | **SRN** |
|---|---|
| **Sanskruti Mirajkar** | 02FE22BEC084 |
| **Shraman Kanthi** | 02FE22BEC092 |
| **Swati Patil** | 02FE22BEC112 |
| **Darshan Modekar** | 02FE22BEC119 |


**Semester: V, 2024-2025**

**Under the Guidance of**

**Prof. Neelam Somannavar**

KLE Technological University,
Dr. M. S. Sheshgiri College of Engineering and Technology
BELAGAVI-590 008

**2023-2024**

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

# CERTIFICATE

This is to certify that project entitled **"Anxiety Prediction Of Student Behavior Based On Day To Day Life"** is a bonafide work carried out by the student team of **" Sanskruti (02FE22BEC084) , Shraman (02FE22BEC092),Swati (02FE22BEC112), Darshan(02FE21BEC119)"**. The project report has been approved as it satisfies the requirements with respect to the Machine Learnin work prescribed by the university curriculum for B.E. (VI Semester) in Department of Electronics and Communication Engineering of KLE Technological University Dr. M. S. Sheshgiri CET Belagavi campus for the academic year 2023-2024.

| | | |
|---|---|---|
| **Prof. Neelam S** | **Dr. Dattaprasad A. Torse** | **Dr. S. F. Patil** |
| **Guide** | **Head of Department** | **Principal** |

**External Viva:**

**Name of Examiners** **Signature with date**

1.

2.

# ACKNOWLEDGMENT

## Acknowledgment

We extend our heartfelt gratitude to our guide, Prof. Neelam Somannavar, whose invaluable guidance and unwavering support have been instrumental throughout the entirety of this project. Her expertise in Machine Learning has provided us with deep insights and clarity, enabling us to overcome challenges and achieve our project goals. Prof. Neelam Somannavar's comprehensive knowledge of algorithm selection, feature engineering, and model evaluation techniques has significantly contributed to the successful development of our system. Her suggestions on techniques such as cross-validation, hyperparameter tuning, and performance metrics have been crucial in ensuring the accuracy and robustness of our solution.

Moreover, her emphasis on the importance of data preprocessing, such as normalization and handling missing values, along with her guidance on effective visualization methods for model results, have not only enhanced the technical quality of our work but also improved the overall presentation of our findings. Prof. Somannavar's mentorship has fostered a spirit of learning and innovation, encouraging us to explore advanced concepts and refine our approach at every stage.

Throughout this journey, her dedication, constructive feedback, and unwavering encouragement have been a source of motivation, inspiring us to push the boundaries of our understanding of Machine Learning principles. Her contributions have been vital in navigating the complexities of the project and ensuring its success.

We sincerely thank Prof. Neelam Somannavar for her invaluable guidance, support, and encouragement, which have been the cornerstone of our efforts and achievements in this project.

-The project team

# Contents

# List of Figures

## List of Abbreviations

| Abbreviation | Description |
|---|---|
| K-NN | K Nearest Neighbours |
| CGPA | Cumulative Grade Point Average |
| HMM | Hidden Markov Model |
| SVM | Support Vector Mechanism |

## ABSTRACT

The prevalence of anxiety among students has significantly increased in recent years, emphasizing the need for timely detection and intervention. This study explores the prediction of anxiety levels in students using various machine learning algorithms, including K-Nearest Neighbors (K-NN), Naïve Bayes, and Logistic Regression. The dataset, collected through targeted questionnaires, encompasses factors such as age, gender, university, department, year, CGPA, scholarship status, and indicators like nervousness, worrying, relaxation, irritation, fear, anxiety values, anxiety labels, and stress labels, alongside other demographic variables and symptoms of anxiety. By transforming these attributes into quantifiable parameters, the models are trained and evaluated for their effectiveness in predicting anxiety. Among the algorithms, the one demonstrating the highest accuracy, lowest error, and best performance is selected.

# 1 INTRODUCTION

In today's fast faced world, the mental health of individuals, especially students, is increasing at risk. Academic pressure, career ambitions and personal relationship issues are significantly contributors to Anxiety., Depression and stress for student, these stressors are often magnified due to challenges or balancing daily academic responsibilities, future career planning and social expectations. According to the World Health Organization, depression is among the most prevalent mental health disorders, affecting over 300 million people globally, and is often accompanied by Anxiety and Stress. These mental health conditions can hinder a student's ability to focus, learn and maintain a healthy lifestyle, impacting overall academic performance and well-being Numerous predicting models have been developed to address mental health disorders, particularly focusing on common symptoms such as feeling of guilt, worthlessness, helplessness, restlessness and even societal thoughts generalize anxiety disorders. Gad, for instance, manifest through symptoms like nervousness, sleeplessness, gastrointestinal issues and lack of concentration stress Similarly, is characterized by symptoms such as inability to relax irritability, fatigue, chronic headaches and frequent infection Differentiating these disorders accurately pose a challenge for machines as many symptoms overlap among anxiety, depression and stress. To address this, we propose a machine learning based approach to credit anxiety level in students by analyzing behavioral patterns in daily life, utilizing the depression, anxiety. We aim to classify the severity of these mental health conditions into five categories. In the study, we apply machine learning algorithms such as Naïve Bayes, K Nearest Neighbors and Logistic Regression to predict and classify anxiety. The goal is to enhance accuracy in diagnosing these mental health issues among students and provide a foundation for effective intervention.

# 2 LITERATURE SURVEY

The rise in mental health issues, particularly anxiety, among students highlights the need for predictive models that facilitate early identification and intervention. Research demonstrates that machine learning models can effectively predict mental health conditions by analyzing behavioral and demographic data.

Reece et al. [6] used a Hidden Markov Model (HMM) to assess depression and PTSD among Twitter users, revealing the potential of social media data in identifying mental health risks. Braithwaite et al. [7] employed decision trees to evaluate suicide risk from tweets, achieving 92 percentage accuracy, showing how text-based data can reveal mental health risks.

Sau et al. (2017) analyzed mental health among older adults using various models, the random forest model was found to be the most accurate, achieving 91 percentage accuracy, particularly in dealing with complex datasets .In a study by Priya, Anu. (2020 ) the model was used to predict anxiety, stress, and depression. with algorithms such as KNN and random forest, selecting random forest for its effective handling of imbalanced data. Hou et al. [8] used reading habits to predict depression, with Naïve Bayes performing best, highlighting how specific behaviors can indicate mental health issues. This study builds on these approaches by targeting student anxiety prediction through daily life factors, like CGPA and demographic variables. By assessing KNN, support vector machine (SVM), Decision tree, Naïve Bayes, Random Forest, and Logistic Regression models, the study aims to identify the most effective model to support early interventions, promoting mental well-being and academic success. After truly reviewing the available literature, we referred to multiple research papers that provided valuable insight into the domain of our project. Among these we identified 1 research that stood out due to its relevance depth, analysis and alignment with our objectives.

# 3 METHODOLOGY

## 3.1 Dataset Description And Preprocessing

In this paper, we used the dataset provided on Kaggle. This dataset is about university students' anxiety. The dataset contains 1977 data points (rows) and 39 features (columns).

### TABLE 1: Attribute Descriptions

| **Column Name** | **Column Description** |
|---|---|
| Age | Participant's age. |
| Gender | Participant's gender. |
| University | The name of the university where the individual is enrolled or graduated from. |
| Year | The current year of study or the academic year the individual is in. |
| CGPA | Cumulative Grade Point Average. |
| Scholarship | Indicates whether the individual is receiving a scholarship, typically a binary value (yes/no). |
| Nervous | A measure of the individual's feelings of nervousness, potentially on a scale. |
| Worrying | Indicates the frequency or intensity of worrying thoughts experienced by the individual. |
| Relaxing | A measure of how often the individual engages in relaxing activities or feels relaxed. |
| Irritated | A measure of how often the individual feels irritated or annoyed. |
| Overthinking | Indicates the tendency of the individual to engage in overthinking or excessive rumination. |
| Restless | A measure of how often the individual feels restless or unable to relax. |
| Afraid | Indicates the frequency or intensity of feelings of fear experienced by the individual. |
| Anxiety Value | A numerical score quantifying the individual's level of anxiety. |
| Anxiety Label | A categorical label representing the individual's anxiety level (e.g., low, moderate, high). |
| Upset | A measure of how often the individual feels upset or emotionally disturbed. |

| | |
|---|---|
| Control | Indicates the individual's perceived level of control over their emotions or situation. |
| Stressed | A measure of how often the individual feels stressed or overwhelmed. |
| Cope | Indicates the individual's ability to cope with stress or difficult situations. |
| Confident | A measure of the individual's self-confidence or belief in their abilities. |
| Flow | Indicates the experience of being in a state of flow, characterized by deep engagement in activities. |
| Irritation Control | A measure of how well the individual can control feelings of irritation. |
| Performance | A measure of the individual's performance in academic or other relevant contexts. |
| Anger | Indicates the frequency or intensity of feelings of anger experienced by the individual. |
| Difficulties | A measure of the difficulties the individual faces in their academic or personal life. |
| Stress Value | A numerical score quantifying the individual's level of stress. |
| Stress Label | A categorical label representing the individual's stress level (e.g., low, moderate, high). |
| Interest | A measure of the individual's interest in their studies or activities. |
| Down | Indicates how often the individual feels down or low in mood. |
| Sleep | A measure of the quality or quantity of sleep the individual gets. |
| Energy | Indicates the individual's energy levels, possibly affecting their daily activities. |
| Appetite | A measure of the individual's appetite or eating patterns. |
| Self-Worth | Indicates the individual's sense of self-worth or self-esteem. |
| Concentration | A measure of the individual's ability to concentrate on tasks or studies. |
| Movement | Indicates the individual's level of physical activity or restlessness. |
| Suicidal Thoughts | A measure of whether the individual experiences thoughts of self-harm or suicide. |
| Depression Value | A numerical score quantifying the individual's level of depression. |
| Depression Label | A categorical label representing the individual's depression level (e.g., low, moderate, high). |

| | Age | Gender | University | Department | Year | CGPA | Scholarship | Nervous | Worrying | Relaxing | Irritated | Overthinking | Restless | Afraid | AnxietyValue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 18-22 | Female | Independent University, Bangladesh (IUB) | Engineering - CS / CSE / CSC / Similar to CS | Fourth Year or Equivalent | 2.50 - 2.99 | No | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 10 |
| **1** | 18-22 | Male | Independent University, Bangladesh (IUB) | Engineering - CS / CSE / CSC / Similar to CS | First Year or Equivalent | 3.80 - 4.00 | No | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 9 |
| **2** | 18-22 | Male | Independent University, Bangladesh (IUB) | Engineering - CS / CSE / CSC / Similar to CS | First Year or Equivalent | 3.00 - 3.39 | No | 2 | 1 | 1 | 0 | 2 | 2 | 2 | 10 |
| **3** | 18-22 | Male | Independent University, Bangladesh (IUB) | Engineering - CS / CSE / CSC / Similar to CS | First Year or Equivalent | 3.40 - 3.79 | No | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| **4** | 18-22 | Male | Independent University, Bangladesh (IUB) | Engineering - CS / CSE / CSC / Similar to CS | First Year or Equivalent | 3.40 - 3.79 | No | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **1972** | 23-26 | Male | Bangladesh Agricultural University | Biological Sciences | Fourth Year or | 3.40 - 3.79 | No | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 11 |

Figure 1: Dataset description

## 3.2 Identifying duplicate values

Identifying duplicate values in a dataset is a crucial step in data preprocessing, especially in projects involving data analysis, machine learning, or statistical modeling. Duplicate values can arise due to data entry errors,system glitches,or merging datasets from multipe sources.These duplicates can introduce biases, inflate the significance of certain observations, and lead to misleading results in analysis. In the context of our project, which involves analyzing academic performance and mental health attributes, duplicate entries could misrepresent patterns, such as the prevalence of anxiety or depression, or skew the relationship between variables like stress and academic performance. By identifying and removing duplicates, you ensure the dataset accurately represents unique individuals, leading to more reliable and valid insights. This step is particularly important when the goal is to draw meaningful conclusions about the population being studied or to build predictive models based on clean and unbiased data. Removing duplicate values is essential to ensure the dataset accurately represents unique observations, avoids bias, and maintains the integrity of the analysis or model. Therefore, we must delete duplicate values for reliable and valid results.

```
 Age Gender                                        University  \
1304  18-22   Male          Independent University, Bangladesh (IUB)
1342  18-22   Male          Independent University, Bangladesh (IUB)
1363  18-22   Male  American International University Bangladesh (...

                                  Department                      Year  \
1304  Engineering - CS / CSE / CSC / Similar to CS  First Year or Equivalent
1342  Engineering - CS / CSE / CSC / Similar to CS  First Year or Equivalent
1363  Engineering - CS / CSE / CSC / Similar to CS  Third Year or Equivalent

             CGPA Scholarship  Nervous  Worrying  Relaxing  Irritated  \
1304  3.40 - 3.79          No        1         1         1          1
1342  3.40 - 3.79         Yes        1         1         1          1
1363  3.00 - 3.39          No        1         1         1          1

      Overthinking  Restless  Afraid  Anxiety Value  AnxietyLabel  Upset  \
1304             1         1       1              7  Mild Anxiety      2
1342             1         1       1              7  Mild Anxiety      2
1363             1         1       1              7  Mild Anxiety      2

      Control  Stressed  Cope  Confident  Flow  IrritationControl  \
1304        2         2     2          2     2                  2
1342        2         2     2          2     2                  2
1363        2         2     2          2     2                  2

      Performance  Anger  Difficulties  StressValue      StressLabel  \
1304            2      2             2           20  Moderate Stress
1342            2      2             2           20  Moderate Stress
1363            2      2             2           20  Moderate Stress

      Interest  Down  Sleep  Energy  Appetite  SelfWorth  Concentration  \
1304         1     1      1       1         1          1              1
1342         1     1      1       1         1          1              1
1363         1     1      1       1         1          1              1

      Movement  SuicidalThoughts  DepressionValue  DepressionLabel
1304         1                 1                9  Mild Depression
1342         1                 1                9  Mild Depression
1363         1                 1                9  Mild Depression
Total duplicate rows: 3
```

Figure 2: Duplicate Values

## 3.3 Identifying unique values

Identifying unique values in a dataset is vital for gaining meaningful insights and ensuring accurate analysis. In the context of our project, which focuses on academic performance and mental health attributes, unique values help us understand the diversity and range of responses, such as variations in anxiety, stress, or depression levels among individuals. This process allows us to identify patterns, correlations, and outliers while ensuring each data point is distinct and contributes to the overall findings. Recognizing unique values also prevents redundancy, supports better decision-making, and enhances the reliability of conclusions drawn from the data. Hence it gives us target column

### TABLE 2: Identifying Duplicate Values

| **Attributes** | **Unique Values** |
|---|---|
| Age | 5 |
| Gender | 3 |
| University | 15 |
| Department | 12 |
| Year | 5 |
| CGPA | 6 |

| | |
|---|---|
| Scholarship | 2 |
| Nervous | 4 |
| Worrying | 4 |
| Relaxing | 4 |
| Irritated | 4 |
| Overthinking | 4 |
| Restless | 4 |
| Afraid | 4 |
| Anxiety Value | 2 |
| Anxiety Label | 4 |
| Upset | 5 |
| Control | 5 |
| Stressed | 5 |
| Cope | 5 |
| Confident | 5 |
| Flow | 5 |
| Irritation Control | 5 |
| Performance | 5 |
| Anger | 5 |
| Difficulties | 5 |
| Stress Value | 38 |
| Stress Label | 3 |
| Interest | 4 |
| Down | 4 |
| Sleep | 4 |
| Energy | 4 |
| Appetite | 4 |
| Self-Worth | 4 |
| Concentration | 4 |
| Movement | 4 |
| Suicidal Thoughts | 4 |
| Depression Value | 28 |
| Depression Label | 6 |

## 3.4 Handling Outliers

Handling outliers is a critical step in data preprocessing to ensure the accuracy and reliability of analysis. Outliers are extreme values that deviate significantly from the rest of the data and can arise due to measurement errors, data entry mistakes, or genuine variability. In our project, which examines academic performance and mental health attributes, outliers could distort statistical measures like mean and standard deviation, leading to misleading results. Proper handling, such as removing or transforming outliers based on their impact, ensures that the dataset accurately reflects underlying trends and relationships, improving the quality of insights and the performance of predictive models.

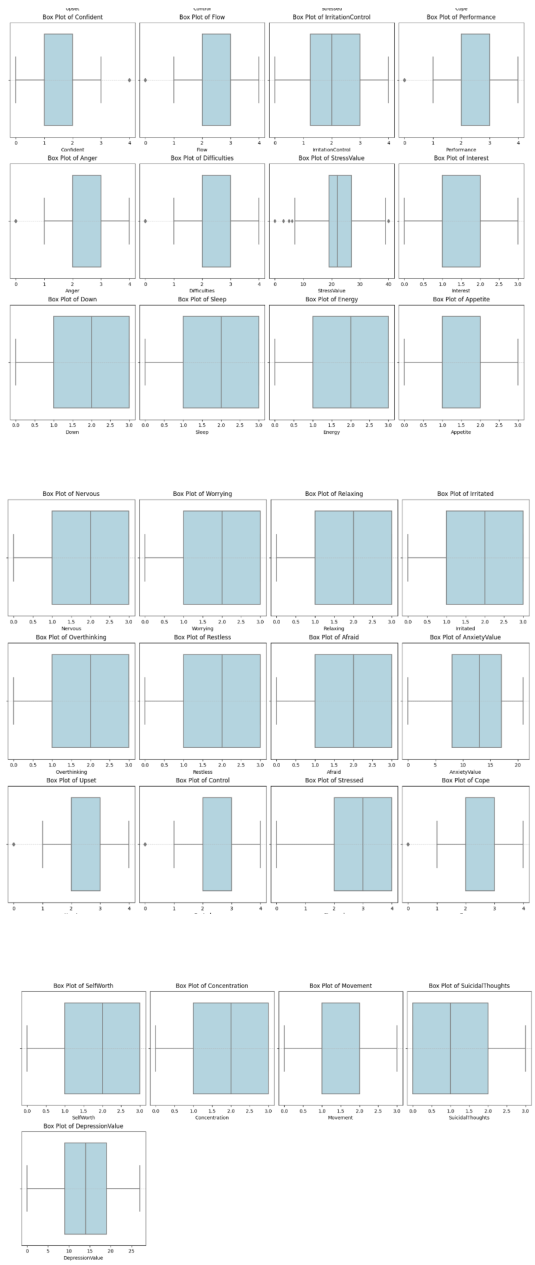


Figure 3: Box plot of outliers

Dropping outliers is essential to maintain the integrity of our dataset and enhance the accuracy of our analysis. Outliers can skew results, distort statistical calculations, and lead to misleading interpretations, particularly in our project focused on academic performance and mental health attributes. By removing these extreme values, we ensure that the remaining data more accurately represents the true patterns and relationships within the population, ultimately leading to more reliable insights and improved predictive modeling.

## 3.5 Data Cleaning

Data cleaning is an essential step in data analysis focused on identifying and resolving errors or inconsistencies within a dataset to enhance its accuracy and dependability. This process often involves addressing missing values, correcting typos, removing duplicates, and addressing outliers that may distort analysis results. In our project, which examines academic performance and mental health attributes, effective data cleaning ensures that the dataset accurately reflects the true characteristics of the population being studied. By standardizing formats, removing irrelevant information, and ensuring consistency across entries, we can enhance the dataset's usability, leading to more valid insights and better-informed decision-making. Ultimately, thorough data cleaning lays the groundwork for successful analysis, predictive modeling, and effective interpretation of results.

## 3.6 Correlation Matrix

A correlation matrix is a table that illustrates the relationship between multiple variables in a dataset by displaying correlation coefficient .Each cell in the matrix represents the correlation between two variables, with value ranging from -1 to +1.A value of +1 signifies a perfect positive correlation meaning that when one variable increases the other also increases .A value of -1 indicates a perfect negative correlation, whereas increases in one variable leads to decrease in other .A correlation of 0 suggests no relationship between the two variables . In our project, a correlation matrix is invaluable for understanding the relationships among various academic performance and mental health attributes, such as anxiety, stress, and depression levels. By visualizing these correlations, we can identify patterns and potential dependencies, helping to uncover insights about how these factors interact. This information is crucial for further analysis, hypothesis testing, and building predictive models, as it allows us to focus on the most relevant variables and better understand the dynamics at play in our data. Overall, a correlation matrix serves as a powerful tool for exploratory data analysis, guiding our research direction and informing subsequent steps in our project.

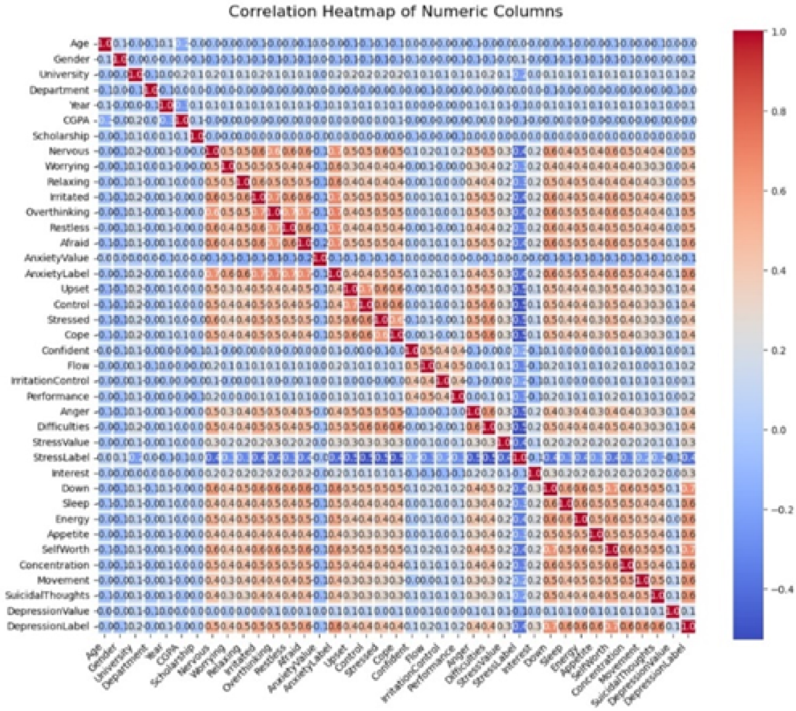


Figure 4: Heat Map

From the above graph we came to a conclusion that there is a high correlation between the below attributes Nervous, Worrying, Relaxing, Irritated, Overthinking, Restless, Afraid and Anxiety label, Upset, Control, Stressed, Cope, Confident, Down, Sleep, Energy, Appetite, Self-worth, Concentration, Movement, Depression Level. So out of all these attributes, I can retain single attributes so that it will be having summary of other attributes so that it can be correlated

## 3.7 Models with Accuracy

### Logistic Regression

Logistic regression is a powerful tool for predicting anxiety levels in students based on various data sets containing features such as Nervous, Worrying, Relaxing, Irritated, Overthinking, Restless, Afraid and Anxiety label, Upset, Control, Stressed, Cope, Confident, Down, Sleep, Energy, Appetite, Self-worth, Concentration, Movement, Depression Level patterns and survey response by modeling the probability of students experiencing anxiety, logistic regression uses these features to establish relationships with the target variable. Its probabilistic framework allows it to estimate the likelihood of anxiety, enabling precise classification based on chosen thresholds. Logistic regression is interpretable, meaning the influence of individual factor on anxiety can be quantified through feature coefficient. This makes it particularly valuable in education and psychological studies, as it not only predicts outcome, but also provides actionable insight for intervention strategies. By effectively handling binary classification tasks, Logistic Regression is a robust choice of anxiety prediction across diverse student datasets.

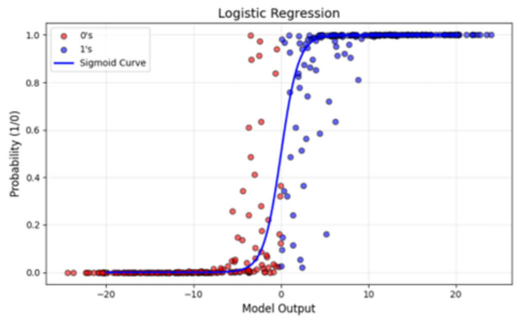


Figure 5: Logistic Regression

After applying Logistic Regression on the trained dataset, we have calculated different parameter i.e. Precision, Recall, F1-Score, Support and Accuracy of anxiety as shown in the Table 2. We have got an Accuracy of 68.10 percentage for Logistic Regression.

```
Accuracy: 0.6810
Accuracy: 68.10%
Confusion Matrix:
[[  0   0   0   0   0  24]
 [  0   0   0   0   0  30]
 [  0   0   0   0   0  28]
 [  0   0   0  40   0   1]
 [  0   0   0   0  11  21]
 [  1   1   5   4  11 218]]
```

Figure 6: Confusion Matrix

## TABLE 2: Identifying Duplicate Values

| SL. No | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 24 |
| 1 | 0.00 | 0.00 | 0.00 | 30 |
| 2 | 0.00 | 0.00 | 0.00 | 28 |
| 3 | 0.91 | 0.98 | 0.94 | 41 |
| 4 | 0.50 | 0.34 | 0.41 | 32 |
| 5 | 0.68 | 0.91 | 0.78 | 240 |

### K-Nearest Neighbors (k-NN)

The K nearest neighbor KN algorithm is a simple and flexible machine learning method that can be applied to predict student anxiety using behavioral academic and personal data. This algorithm evaluates the similarity between a student's dataset such as academic scores, sleep patterns, survey results, and social activities—and that of other students in the dataset. It determines the kk nearest neighbors by calculating distances (e.g., using Euclidean distance) to find students with similar characteristics. Anxiety is then predicted based on the predominant class (e.g., "anxious" or "not anxious") among these nearest neighbors. KNN is particularly effective for datasets with complex, non-linear relationships, as it does not rely on any assumptions about data distribution. By selecting an appropriate kk value, the model can balance noise reduction and accuracy, making it a valuable tool for identifying anxiety in diverse student datasets.

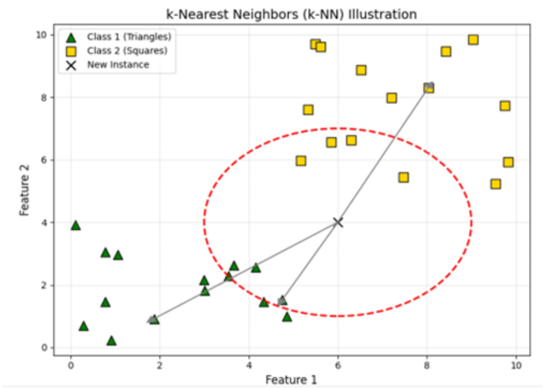


Figure 7: K-NN

After applying K-Nearest Neighbors(KNN) on the trained dataset, we have calculated different parameter i.e. Precision, Recall, F1-Score, Support and Accuracy of anxiety as shown in Table 3. We have got an Accuracy of 53.92 percentage for K-Nearest Neighbors (KNN).

```
Accuracy: 0.5392
Accuracy in percentage: 53.92%
Confusion Matrix:
[[  3   1  10   0   0  10]
 [  2   2  11   0   1  14]
 [  0   3   9   0   0  16]
 [  0   0   1  25   0  15]
 [  0   0   0   0  23   9]
 [  9   6  27  13  34 151]]
```

Figure 8: Confusion Matrix

## TABLE 2: Identifying Duplicate Values

| SL. No | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.22 | 0.46 | 0.30 | 24 |
| 1 | 0.22 | 0.27 | 0.24 | 30 |
| 2 | 0.48 | 0.57 | 0.52 | 28 |
| 3 | 1.00 | 1.00 | 1.00 | 41 |
| 4 | 0.36 | 0.97 | 0.53 | 32 |
| 5 | 0.99 | 0.61 | 0.76 | 240 |

### Naive Bayes

Naive Bayes is a chance-based machine learning algorithm that can be effectively used to predict anxiety in students by analyzing various features, such as academic performance, lifestyle habits, and survey responses. The model applies Bayes' theorem, assuming that all features are conditionally independent, to calculate the probability of a student being ”anxious” or ”not anxious” based on the input data. Despite its simplicity, Naive Bayes performs well for classification tasks, especially with categorical data. By learning the likelihood of anxiety based on individual features and combining them to make predictions, it provides a fast and efficient way to analyze student data. This algorithm is particularly useful for datasets with clear class distinctions and works well even with smaller datasets, making it an excellent choice for predicting anxiety among students in a structured and interpretable manner.

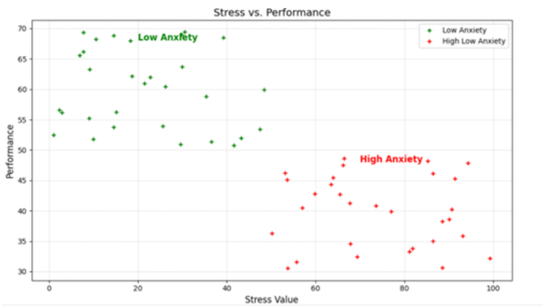


Figure 9: Naive Bayes

After applying Naive Bayes on the trained dataset, we have calculated different parameter i.e. Precision, Recall, F1-Score, Support and Accuracy of anxiety as shown in Table 4.We have got an Accuracy of 64.30 percentage for Naive Bayes.

```
Accuracy: 0.6430
Accuracy in percentage: 64.30%
Confusion Matrix:
[[ 11   9   4   0   0   0]
 [ 11   8  11   0   0   0]
 [  1  10  16   0   0   1]
 [  0   0   0  41   0   0]
 [  0   0   0   0  31   1]
 [ 26  10   2   0  55 147]]
```

Figure 10: Confussion Matrix

## TABLE 2: Identifying Duplicate Values

| SL. No | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.22 | 0.46 | 0.30 | 24 |
| 1 | 0.22 | 0.27 | 0.24 | 30 |
| 2 | 0.48 | 0.57 | 0.52 | 28 |
| 3 | 1.00 | 1.00 | 1.00 | 41 |
| 4 | 0.36 | 0.97 | 0.53 | 32 |
| 5 | 0.99 | 0.61 | 0.76 | 240 |

# 4 Result and Discussion

## 4.1 Visualization

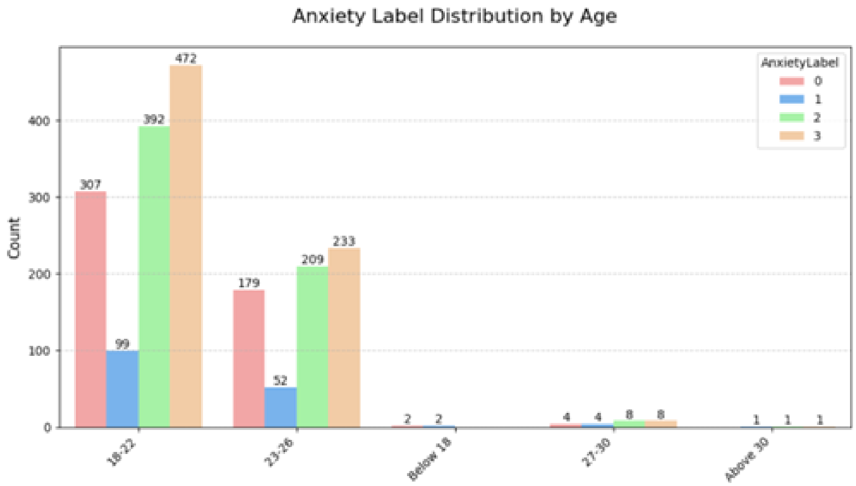


Figure 11: Anxiety label Distribution vs Age

The above Fig.11 shows the anxiety Vs Age distribution. Here we can identify that the age group of 18-22 have more students suffering from anxiety

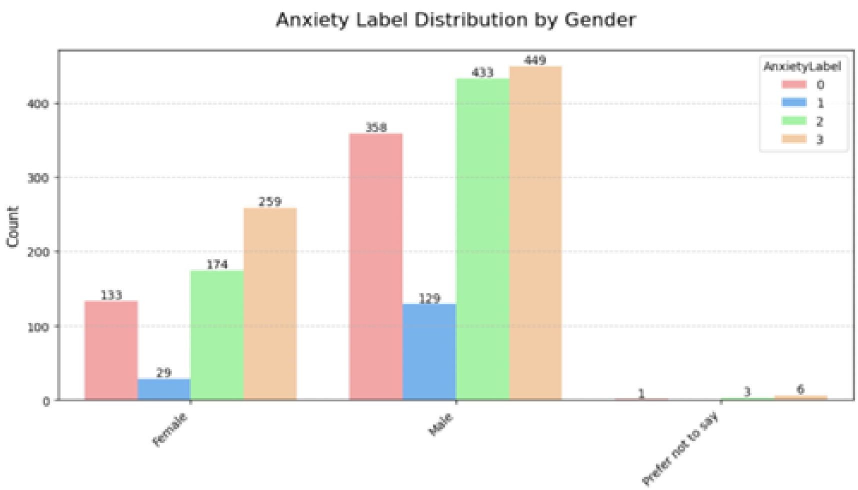


Figure 12: Anxiety label Distribution vs Gender

The above Fig.12 shows the anxiety Vs Gender distribution. Here we can identify that Male students suffer more from anxiety.

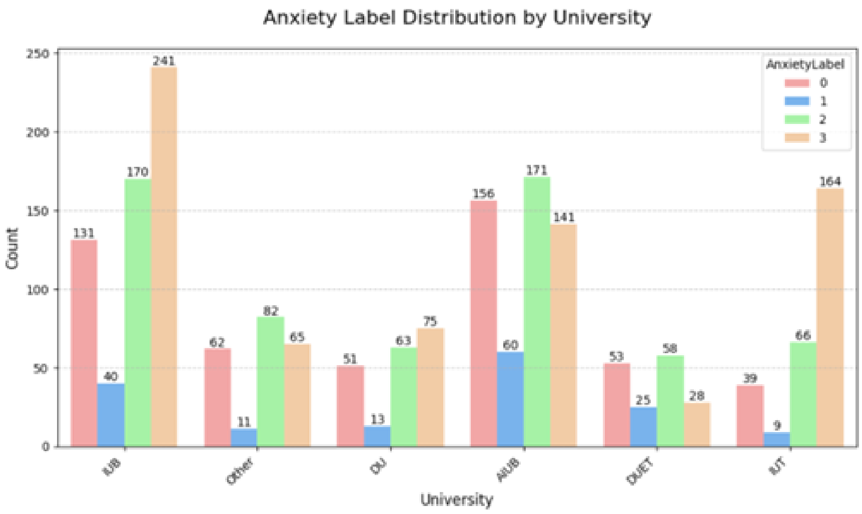


Figure 13: Anxiety label Distribution vs University

The above Fig.13 show Anxiety label Vs University. Here we can identify that University of IUB is suffering more from Anxiety

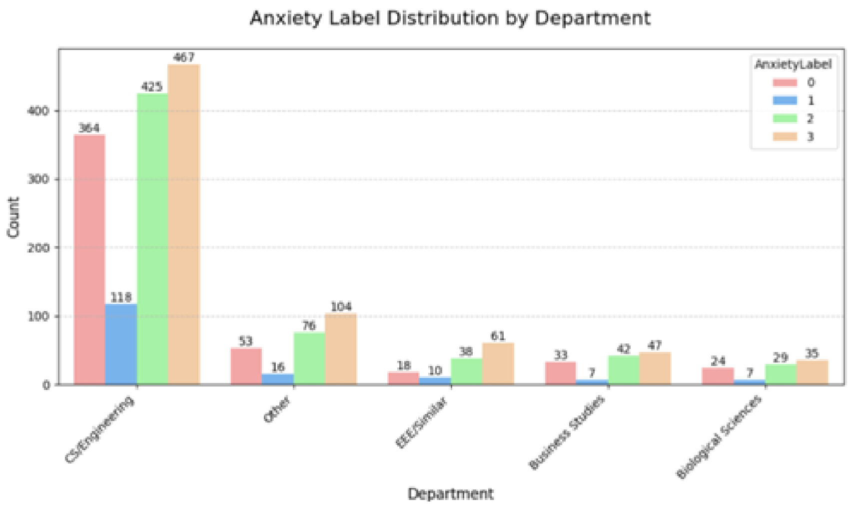


Figure 14: Anxiety label Distribution vs Department

The above Fig.14. shows Anxiety label Vs Department. Here we can anxiety that CSE Engineering is suffering more from Anxiety

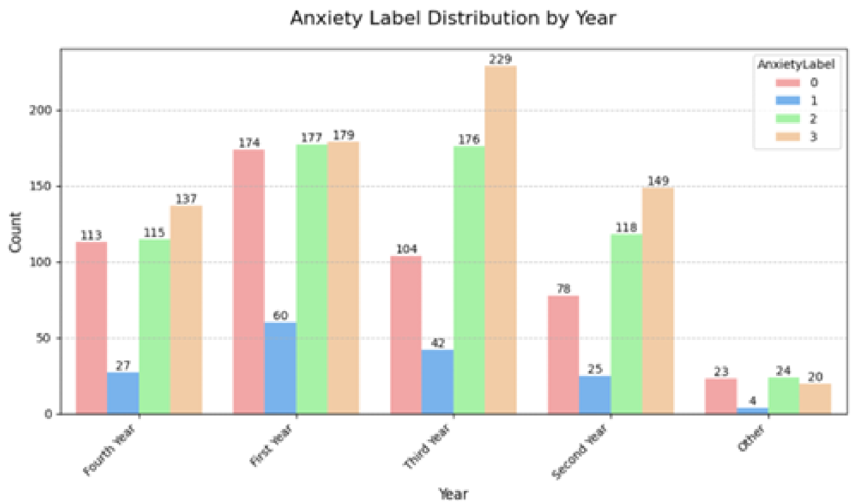


Figure 15: Anxiety label Distribution vs Year

The above Fig.15. shows Anxiety label Vs Year. Here we can identify that Third-Year students is suffering more from Anxiety

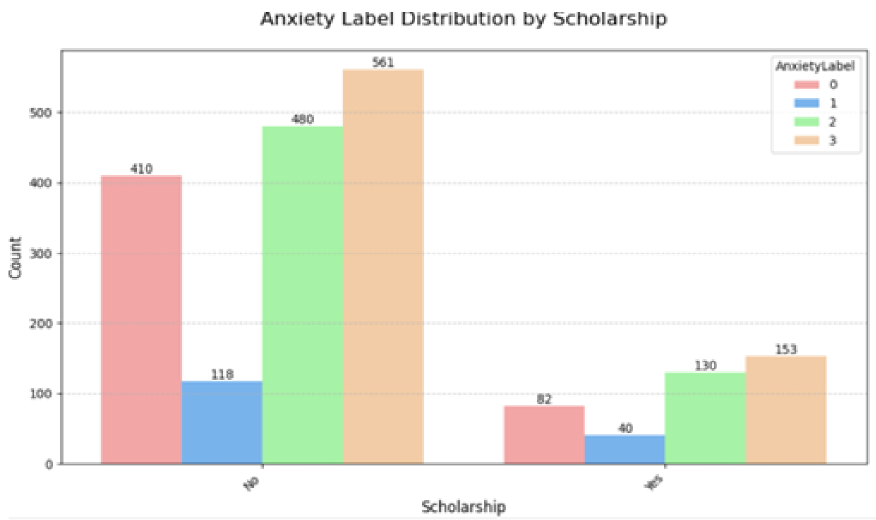


Figure 16: Anxiety label Distribution vs Scholarship

The above Fig.16 shows Anxiety label Vs scholarship. Here we can identify that students who have not received Scholarship has suffered more from Anxiety.

## 4.2 Results

### TABLE 3: Model Evaluation Metrics

| SL. No | Model | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| 0 | Logistic Regression | 0.55087 | 0.683544 | 0.60466 | 68.35 |
| 1 | K-Nearest Neighbors | 0.58557 | 0.52911 | 0.53990 | 52.911 |
| 2 | Naïve Bayes | 0.79680 | 0.64300 | 0.67920 | 64.303 |

We have showcased the Accuracy of anxiety of students of different dataset in the form of table and bar graph as shown in Fig.17.

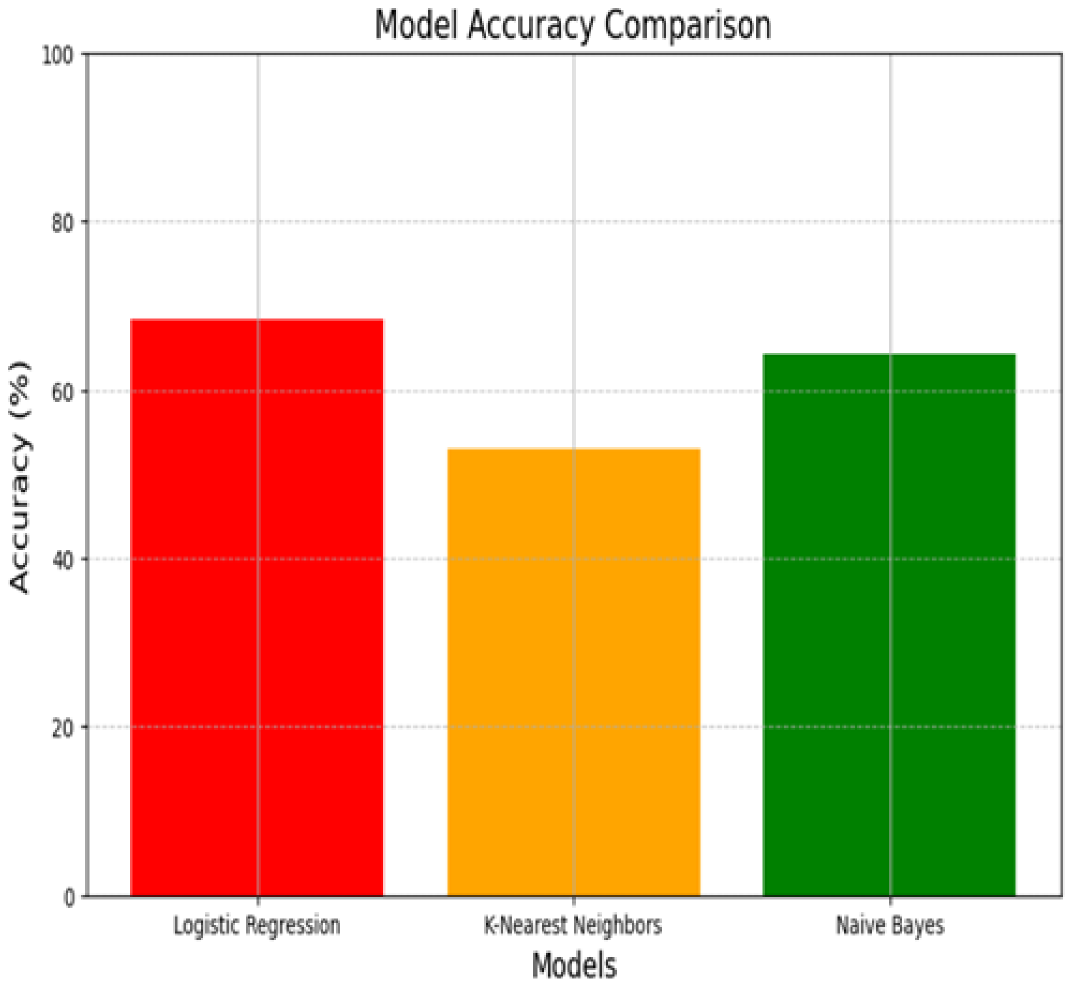


Figure 17: Accuracy of Models

# 5 Conclusion

In our research we investigated how machine learning techniques can detect individuals who might be prone to anxiety by analyzing observable behavioral patterns. The model was designed to analyze various input features such as academic performance, social interaction, sleep patterns and lifestyle habits. Logistic Regression proved to be an effective and interpretable method for predicting the likelihood of Anxiety, offering valuable insight into the contributing factors. The results demonstrated that such a tool can provide early warnings, enabling timely intervention to address mental health concerns among students.

# 6 Applications

The prediction of anxiety levels among students using machine learning algorithms has several practical applications:

- **Early Detection:** The models can facilitate the early detection of anxiety in students, enabling timely interventions and support from mental health professionals.
- **Tailored Interventions:** Understanding the factors contributing to anxiety can help in designing personalized mental health programs and strategies to alleviate stressors in students' lives.
- **Institutional Support:** Educational institutions can utilize the findings to implement support systems, workshops, and counseling services targeted at high-risk groups.
- **Research Development:** This study can serve as a foundation for further research into the mental health challenges faced by students, guiding future studies and intervention programs.

# 7 Future Scope

The future scope of this research encompasses various avenues for improvement and further exploration:

- **Integration of Additional Data:** Future studies can include more diverse data sources, such as physiological indicators, academic performance metrics, and lifestyle habits to enhance prediction accuracy.
- **Development of Real-Time Monitoring Tools:** Creating applications or platforms for real-time monitoring of students' mental health could provide continuous support and timely interventions.
- **Longitudinal Studies:** Conducting longitudinal studies to monitor the changes in anxiety levels over time and the impact of interventions will provide deeper insights into effective coping strategies.
- **Exploration of Advanced Algorithms:** Investigating advanced machine learning techniques such as deep learning or ensemble methods could further improve prediction models and their applicability in diverse contexts.

# Bibliography


[1] M. Tubio-Fungeririno, E. Cernades, Of Goncalves, et al., "Viability Study of Machine Learning-Based Prediction of COVID-19 Pandemic Impact in Obsessive-Compulsive Disorder Patients," *Frontiers in Neurorobotics*, vol. 16, pp. 807584, 2022. doi:10.3389/fninf.2022.807584.

[2] E. H. Jang, et al., "Automated detection of panic disorder based on multi-modal physiological signals using machine learning," *ETRI Journal*, pp. 1-14, 2022. https://doi.org/10.4218/etrij.2021-0299.

[3] C. H. Tsai et al., "Attack Prediction Using Wearable Devices and Machine Learning: Development and Cohort Study," *JMIR Medical Informatics*, vol. 10, no. 2, e33063, 2022. doi:10.2196/33063. PMID:35166679; PMCID:PMC8889475.

[4] M. D. Nemesure, M. V. Heinz, R. Huang et al., "Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence," *Scientific Reports*, vol. 11, 1980, 2021. https://doi.org/10.1038/s41598-021-81368-4.

[5] K. Na, S. Cho, S. Cho, "Machine learning-based discrimination of panic disorder from other anxiety disorders," *Journal of Affective Disorders*, vol. 278, pp. 1-4, 2021.

[6] M. Xing, F. M. Jacklynn, H. Klump, "Classification of Social Anxiety Disorder with Support Vector Machine Analysis Using Neural Correlates of Social Signals of Threat," *Frontiers in Psychiatry*, vol. 11, 2020.

[7] X. Chen, I. H., Zheng X, Huang J., "Effects of music therapy on COVID-19 patients' anxiety, depression, and life quality: a protocol for systematic review and meta-analysis," *Medicine*, vol. 100, no. 26, pp. 26419, 2021.

[8] Priya, Anu, Shruti Garg, and Neha Prerna Tigga, "Predicting anxiety, depression and stress in modern life using machine learning algorithms," *Procedia Computer Science*, vol. 167, pp. 1258-1267, 2020.