

**STAT222**

**Advanced Statistics**

**Final Project Report**

Submitted to:   
Mr. Hafiz Arslan Ramzan

Predicting Stress And Anxiety Risks Among Teens

**Submitted By:**

Mohammad Hasnain (462247)

Mohammad Umar (461532)

Mohid Arshad (455977)

Tahir Mehmood (458593)

**Class: BSDS-1**

TABLE OF CONTENTS

[Predicting Student Stress Levels Using Statistical and Machine Learning Methods 3](#_Toc198551112)

[1. Introduction 3](#_Toc198551113)

[2. Literature Review 3](#_Toc198551114)

[3. Dataset Description & Preprocessing 3](#_Toc198551115)

[4. Exploratory Data Analysis / Visualizations 4](#_Toc198551116)

[5. Methodology & Model Pipeline 7](#_Toc198551117)

[Model Architecture Diagram: 8](#_Toc198551118)

[Input → Preprocessing → PCA & Correlation → Decision Tree → Prediction & Evaluation 8](#_Toc198551119)

[6. Advanced Statistical Techniques Used 8](#_Toc198551120)

[7. Model Evaluation & Performance Metrics 11](#_Toc198551121)

[8. Research Insights and Discussion 13](#_Toc198551122)

[9. Future Work & Limitations 13](#_Toc198551123)

[10. References 13](#_Toc198551124)

# Predicting Student Stress Levels Using Statistical and Machine Learning Methods

# 1. Introduction

Student stress is a growing concern, directly affecting mental health, academic performance, and well-being. Our *Objective* is to *predict student stress levels* *(low, medium, high)* using *features* *such as* *anxiety, self-esteem, depression, and academic pressure*. By analyzing this data statistically and through *a machine learning model*, we aim to uncover patterns that can guide interventions.

# 2. Literature Review

Recent studies have used machine learning to predict student stress:

* Filippis & Al Foysal (2024) found **psychological indicators** like self-esteem and sleep quality as **strong predictors of stress**.
* Rois et al. (2021) used **physiological features** like pulse and sleep status to reach **~89% accuracy** using **Random Forest**.
* Arya et al. (2024) applied **multiple ML algorithms**, achieving **~90% accuracy** with **Naïve Bayes**.
* Daza et al. (2023) identified **SVM** and **logistic regression** as **highly effective** in a review of **stress prediction studies**.
* Singh et al. (2024) showed **SVM** achieving **~95% accuracy** in distinguishing **stress levels**.
* Ding et al. (2023) used a **hybrid boosted model** to achieve **near-perfect prediction accuracy**.
* These works highlight the importance of psychological and academic features and demonstrate that ML models are reliable for stress prediction.

# 

# 3. Dataset Description & Preprocessing

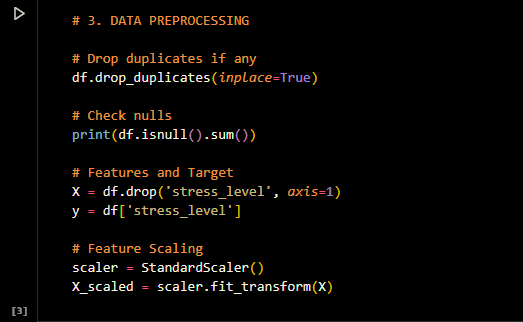
Dataset: Survey-based, containing 1,100 records across psychological, physiological, academic, and social dimensions.

Target: stress\_level (encoded as 0 = low, 1 = medium, 2 = high).

Features: Include anxiety, self-esteem, sleep quality, depression, etc.

#### Preprocessing:

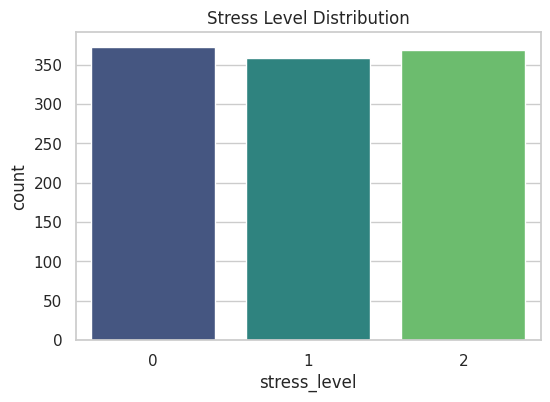
* Verified no missing values.
* Applied z-score normalization using StandardScaler.
* Removed duplicate records.
* Target variable encoded as categorical.



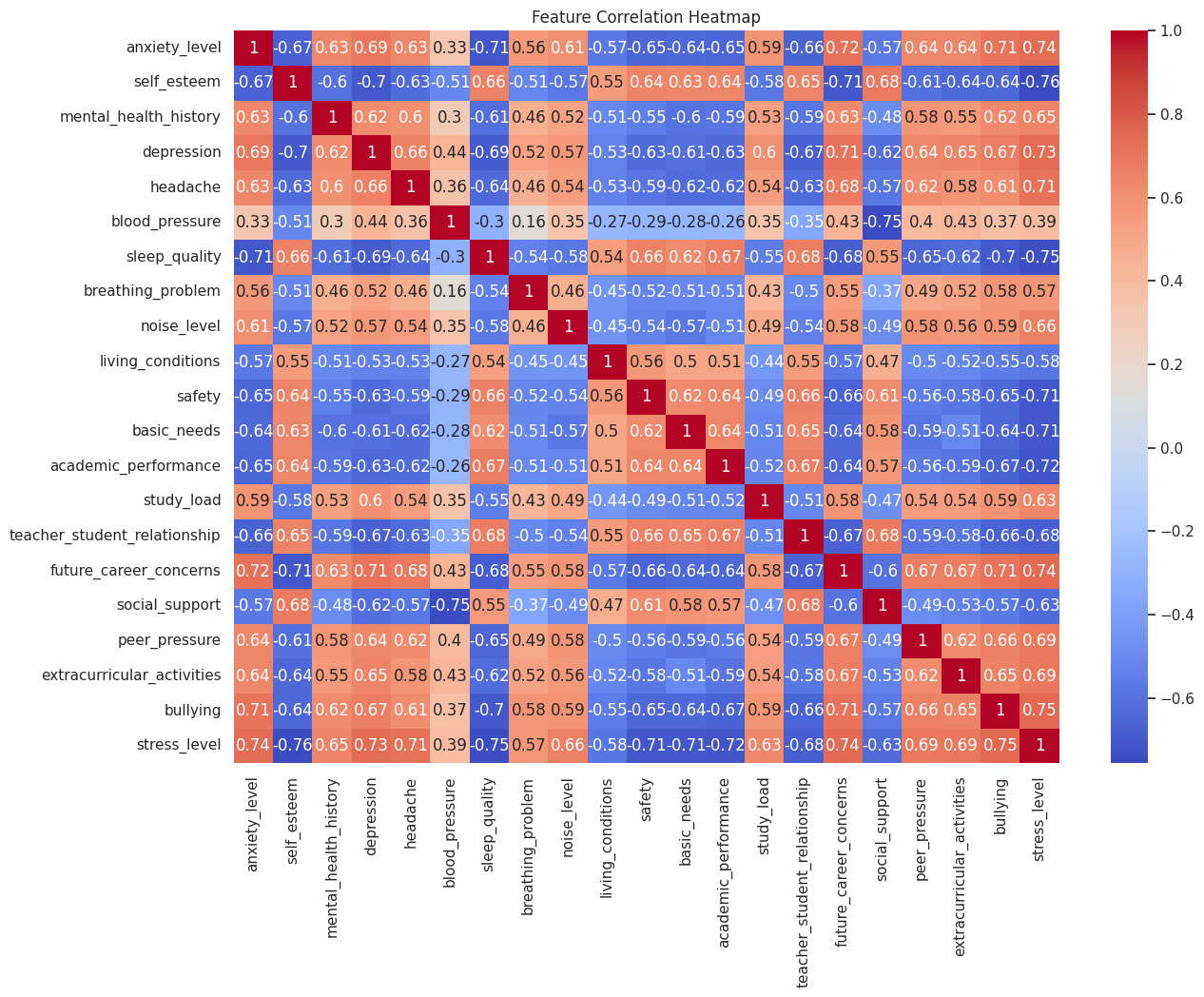
# 

# 4. Exploratory Data Analysis / Visualizations

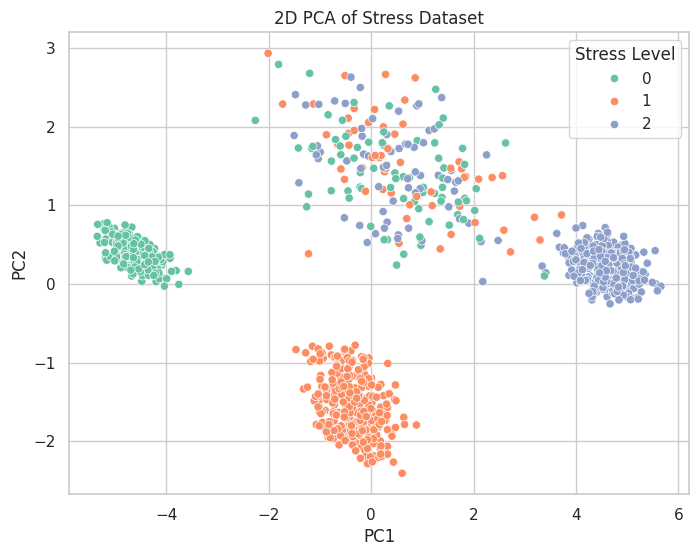
Class Distribution: Even distribution of stress levels.



Correlation Heatmap: Highlights strong links (e.g., anxiety–depression).



PCA Scatter Plot: Shows class separation in reduced dimensions.



Boxplots: Visualize trends in features like anxiety, study load, self-esteem.

A screenshot of a graph

AI-generated content may be incorrect.

# 

# 5. Methodology & Model Pipeline

We followed a structured machine learning pipeline:

Data Splitting: train\_test\_split with **82%** for training and **18%** for testing.

Model Used: DecisionTreeClassifier from scikit-learn.

#### Training:

dt = DecisionTreeClassifier(random\_state=42)

dt.fit(X\_train, y\_train)

#### Prediction:

y\_pred = dt.predict(X\_test)

Evaluation: Metrics include accuracy, precision, recall, F1-score, and confusion matrix.

### Model Architecture Diagram:

### Input → Preprocessing → PCA & Correlation → Decision Tree → Prediction & Evaluation

Diagram of a diagram with different colored squares

AI-generated content may be incorrect.

# 6. Advanced Statistical Techniques Used

#### a. Correlation Analysis

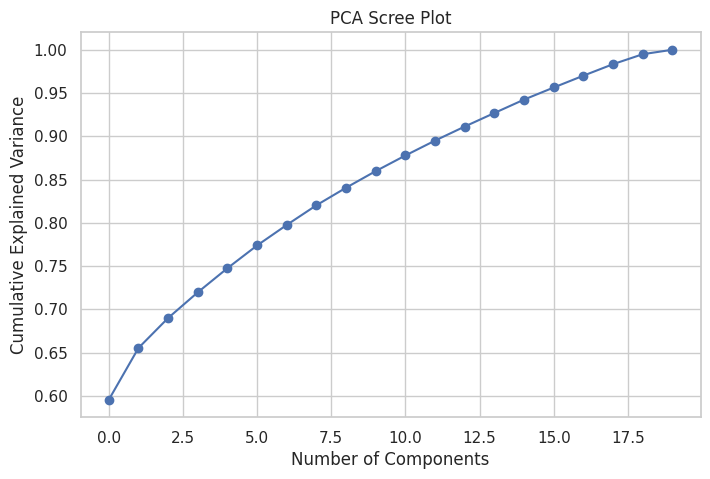
* **Pearson correlation** used to explore dependencies.
* Notable: ***Anxiety*** and ***academic performance*** showed **strong negative correlation**.

A colorful grid with numbers

AI-generated content may be incorrect.

#### b. Principal Component Analysis (PCA)

* Used to ***reduce dimensionality*** and ***visualize*** separability.
* **PCA** plot shows ***partial clustering*** of stress levels.



A graph of stress dataset

AI-generated content may be incorrect.

# 7. Model Evaluation & Performance Metrics

* Accuracy: ~90%
* F1-Scores:
* Class 0 (Low): 0.91
* Class 1 (Medium): 0.89
* Class 2 (High): 0.90
* Confusion Matrix: Most predictions lie on the diagonal (correct classifications).

#### Macro-Averaged Scores:

* Precision: 0.90
* Recall: 0.90
* F1-score: 0.90

The Decision Tree successfully identifies key features influencing stress.

*Matrix of Decision Tree Performance*

A graph of blue rectangular bars

AI-generated content may be incorrect.

*Confusion Matrix of Decision Tree*

A blue and white diagram with numbers and a blue square

AI-generated content may be incorrect.

*Comparison of Models*

A graph of different colored bars

AI-generated content may be incorrect.

# 8. Research Insights and Discussion

* **High anxiety** and **low self-esteem** *strongly* **signal high stress**.
* **The decision tree model** *reflects* **real-world stress factors**.
* **Sleep quality, academic pressure**, and **social support** *emerged* as **key influences**.
* **This predictive system** *can be used* for early **intervention** and **mental health** screening in universities.

# 9. Future Work & Limitations

#### Limitations:

* *Relies* on self-reported data (subject to bias).

#### Future Improvements:

* Try **ensemble models** (e.g., **XGBoost**).
* Incorporate **sensor data** (e.g., **wearables**).
* Apply **time-series modeling** for tracking stress over semesters.
* Validate international datasets for broader applicability.

# 10. References

* Arya, S., et al. (2024). Predicting the stress level of students using supervised machine learning. *Indian Journal of Engineering, 21*.
* Daza, A., et al. (2023). Systematic review of ML techniques to predict anxiety/stress. *Informatics in Medicine Unlocked, 43*.
* Ding, C., et al. (2023). Hybrid machine learning for stress prediction. *PeerJ Computer Science, 9*.
* Filippis, R., & Al Foysal, A. (2024). Stress factor analysis with ML. *Discover Artificial Intelligence, 4*.
* Rois, R., et al. (2021). ML stress prediction among Bangladeshi students. *JHPN, 40*.
* Singh, A., et al. (2024). Detecting stress with ML. *arXiv Preprint*.
* Dataset Source: StressLevelDataset.csv by Filippis & Al Foysal (2024).
* Tools: Python, Pandas, Scikit-learn, Seaborn, Matplotlib.