LOVELY PROFESSIONAL UNIVERSITY

PROJECT REPORT

MENTAL HEALTH ANALYSIS AMONG INDIAN YOUTH:

A MACHINE LEARNING APPROACH

Submitted by

Name.: Anushka
Registration No.: 12319243

Programme : B.Tech Computer Science and Engineering

Section: K23MG
Course Code: INT234

Under the Guidance of

Dr. Saqib Ul Sabha
Assistant Professor

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Anushka** , bearing Registration No. **12319243**, has successfully completed the **INT234** project titled **"Mental Health Analysis Among Indian Youth: A Machine Learning Approach"** under my guidance and supervision.

To the best of my knowledge, the present work is the result of his original development, effort, and study.

Signature of the Supervisor

Name of the Supervisor: Dr. Saqib Ul Sabha

Designation: Assistant Professor

School: Lovely School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab

Date: 19/12/2025

DECLARATION

I, **Anushka** , student of **B.Tech Computer Science and Engineering** under **CSE/IT Discipline** at **Lovely Professional University, Punjab**, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Signature: Anushka

Registration No.: **12319243**

Name of the Student: **Anushka**

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have contributed to the successful completion of this project.

First and foremost, I am deeply thankful to my project supervisor, **Dr. Saqib Ul Sabha**, for her valuable guidance, continuous support, and constructive feedback throughout the development of this project. Her expertise and encouragement played a crucial role in shaping this work.

I am also grateful to the **Head of the Department**, Lovely School of Computer Science and Engineering, for providing the necessary facilities and resources required for the completion of this project.

I extend my sincere thanks to all the participants who voluntarily contributed to the dataset by sharing their personal experiences related to mental health, sleep patterns, and stress levels. Their honest responses made this analysis possible.

Finally, I would like to thank my family and friends for their constant motivation and moral support throughout the course of this project.

Anushka

Registration No.: **12319243**

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

1.1 Background

Mental health has emerged as one of the most critical public health concerns worldwide, particularly among young adults and students. In India, the situation has become increasingly concerning over the past decade. According to the National Mental Health Survey (2015–16), nearly 10 percent of the adult population in India suffers from mental health disorders, with young adults aged 18 to 29 identified as the most vulnerable group.

Students and young professionals are exposed to multiple stressors, including academic pressure, career uncertainty, financial challenges, interpersonal relationships, and family expectations. The rapid growth of digital technology has further intensified these issues. While technology provides easy access to information and social connectivity, excessive screen time and social media usage have been strongly associated with increased anxiety, sleep disturbances, and reduced self-esteem among youth.

The COVID-19 pandemic further aggravated mental health challenges. Prolonged isolation, disruption of academic schedules, reduced social interaction, and economic uncertainty significantly impacted students' psychological well-being. Various studies and reports indicate that anxiety and depression levels among college students increased by nearly 30 percent during the pandemic period.

Despite increasing awareness, there remains a lack of comprehensive understanding of the specific factors influencing stress and emotional disturbances among Indian youth. Traditional mental health assessment approaches rely largely on self-report surveys or clinical evaluation, which are time-consuming and often inaccessible to a large population. This highlights the need for scalable, data-driven methods to analyze and understand mental health patterns.

1.2 Motivation

The motivation for this project arises from the urgent need to address mental health issues faced by students and young professionals in India. Many individuals continue to suffer silently due to social stigma, lack of awareness, or limited access to mental health services. Early

identification and timely intervention can significantly improve mental health outcomes, yet existing screening mechanisms are not scalable or efficient.

Machine learning provides a promising approach to address this challenge. By identifying hidden patterns in lifestyle habits, demographic attributes, and self-reported mental health indicators, predictive models can help detect individuals at risk of developing mental health issues. Such models can act as preliminary screening tools within educational institutions, enabling counselors and administrators to prioritize high-risk cases.

Additionally, understanding how lifestyle factors such as sleep duration, screen time, and exercise habits influence stress levels can guide targeted interventions. For example, if data reveals that students sleeping fewer than six hours consistently report higher stress, institutions can promote better sleep hygiene through awareness programs and academic scheduling improvements.

1.3 Problem Statement

The primary objectives of this project are as follows:

- To analyze the mental health status of Indian youth using a dataset comprising demographic information, lifestyle habits, stress levels, and mental health experiences.
- To identify key factors contributing to stress, anxiety, and depression among students and young professionals.
- To develop and evaluate machine learning models for predicting stress levels and classifying individuals based on mental health risk.
- To generate actionable insights and recommendations that can be utilized by educational institutions, policymakers, and individuals to improve mental well-being.

1.4 Scope of the Project

This project focuses on the analysis of mental health data collected from **600 participants** belonging to diverse age groups, occupations, and financial backgrounds across India. The scope includes exploratory data analysis to understand stress distribution, emotional well-being, and lifestyle patterns.

Comprehensive data preprocessing techniques are applied to handle missing values, outliers, duplicate columns, and formatting inconsistencies. Feature engineering is performed to derive meaningful variables such as binary indicators for anxiety, depression, and high stress.

Multiple machine learning models, including regression and classification algorithms, are developed and evaluated to address different analytical objectives. Visualizations are used extensively to communicate findings effectively. The project aims to establish a data-driven foundation for future mental health interventions and contribute to research on youth mental health in India.

1.5 Organization of the Report

The remainder of this report is organized as follows.
Section 2 describes the source, collection process, and characteristics of the dataset used in this study.
Section 3 details the preprocessing techniques applied to clean and prepare the data for analysis.
Section 4 presents five key insights derived from the analysis, including model development, evaluation, and interpretation of results.
Section 5 summarizes the key findings and conclusions.
Section 6 discusses the future scope and possible extensions of this work.
Section 7 lists the references consulted during the project.

2. SOURCE OF DATASET

2.1 Data Collection Method

The dataset used in this project was collected through an online survey conducted between **October and November 2025**. The survey was distributed via social media platforms, educational forums, and email networks targeting students and young professionals across India. Participation was voluntary and anonymous to encourage honest responses.

The survey consisted of **18 questions** covering various aspects of mental health and lifestyle habits. Participants were informed about the purpose of the study and assured that their responses would be used solely for academic research.

2.2 Dataset Characteristics

The final dataset contains responses from **600 participants** and is stored in CSV format. The dataset includes the following attributes:

Table 1: Dataset Column Description

Column Name	Description	Data Type
timestamp	Date and time of survey submission	Text
username	Participant identifier	Text
age_group	Age range of participant	Text
gender	Gender identity	Text
financial_background	Economic status of family	Text
occupation	Current occupation or student status	Text
sleep_hours	Average hours of sleep per night	Text
screen_time	Average daily screen usage	Text

Column Name	Description	Data Type
free_time_activities	Leisure activities	Text
exercise_frequency	Frequency of exercise	Text
stress_level	Self-reported stress (1–5)	Text
mental_health_experiences	Anxiety, depression indicators	Text
main_reason	Primary source of stress	Text
anxiety_frequency	Frequency of anxiety	Text
has_support	Availability of emotional support	Text
emotional_state	Current emotional condition	Text
preferred_resources	Preferred support resources	Text
awareness_needed	Need for awareness programs	Text

2.3 Data Quality

Initial analysis revealed several data quality issues. The username column had **489 missing values** out of 600. Duplicate financial background columns were present due to survey updates. Numeric attributes such as sleep hours and screen time had inconsistent formatting. Outliers were observed in sleep and screen time data. Encoding issues and special characters were present in some text fields.

These issues were systematically addressed during the preprocessing phase to ensure reliable analysis.

2.4 Ethical Considerations

All participants provided informed consent before participating in the survey. Personally identifiable information was not collected, and optional identifiers were removed during

preprocessing. The study adhered to ethical research guidelines for handling sensitive mental health data.

---

3. DATASET PREPROCESSING

Data preprocessing is a critical step in machine learning projects, as raw data often contains inconsistencies and errors that can affect model performance. This section describes the structured approach used to clean and prepare the dataset.

3.1 Initial Data Exploration

The dataset initially consisted of **600 rows and 19 columns**. Preliminary exploration revealed **641 missing values** across three columns, no duplicate rows, and most features stored in text format requiring conversion.

Table 2: Missing Values Summary

Column Name	Missing Count	Percentage
username	489	81.50%
financial_background_old	16	2.67%
financial_background	136	22.67%

Exploratory analysis showed that most participants belonged to the **19–22 age group (334 participants)**. Gender distribution was balanced, and the majority were undergraduate students. Stress levels followed a near-normal distribution, with level 3 (moderate stress) being most common.

3.2 Handling Duplicate and Unnecessary Columns

Duplicate financial background columns were merged by filling missing values from one column using the other. After merging, the redundant column was removed. Columns such as timestamp and username were dropped as they did not contribute to analysis.

This reduced the dataset to **16 relevant columns**.

3.3 Text Formatting and Standardization

Text inconsistencies and encoding issues were resolved by replacing non-standard characters with standard formats. This ensured uniformity across categorical features.

3.4 Numeric Column Conversion

Sleep hours and screen time were originally stored as text. A custom parsing function was developed to handle ranges, units, and approximate values. Ranges were converted to mean values, and invalid entries were handled appropriately.

Table 3: Before and After Conversion

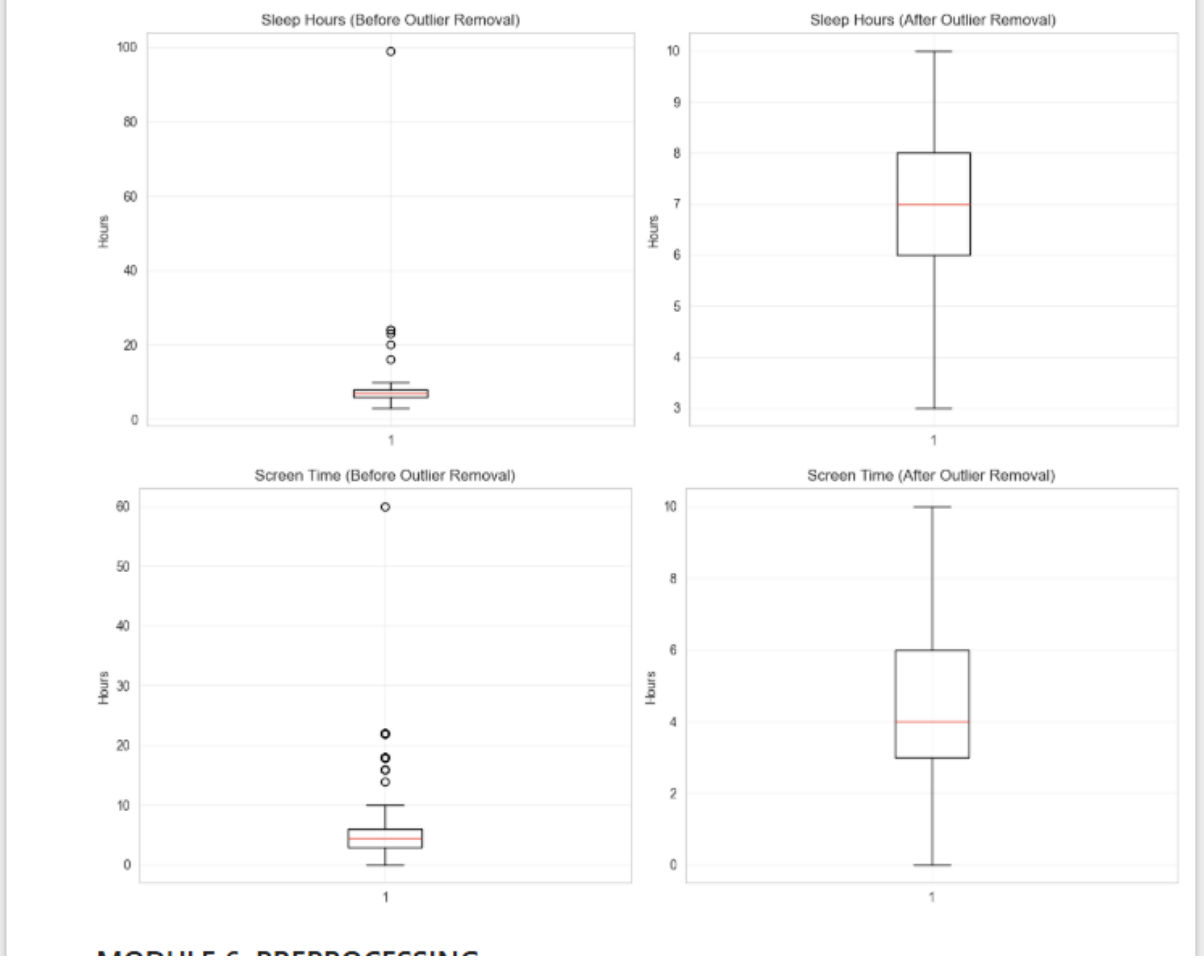
Original Value	Converted Value
7	7.0
6 to 8	7.0
6 hours approx	6.0
5–7	6.0

3.5 Handling Missing Values

Remaining missing values in the financial_background column were filled using mode imputation (Middle Class). After this step, the dataset contained **zero missing values**.

3.6 Outlier Detection and Removal

Outliers were detected using the Interquartile Range (IQR) method. Extreme values such as unrealistic sleep or screen times were removed. After outlier removal, the dataset contained **520 rows**, representing a reasonable trade-off between data quality and size.



3.7 Feature Engineering

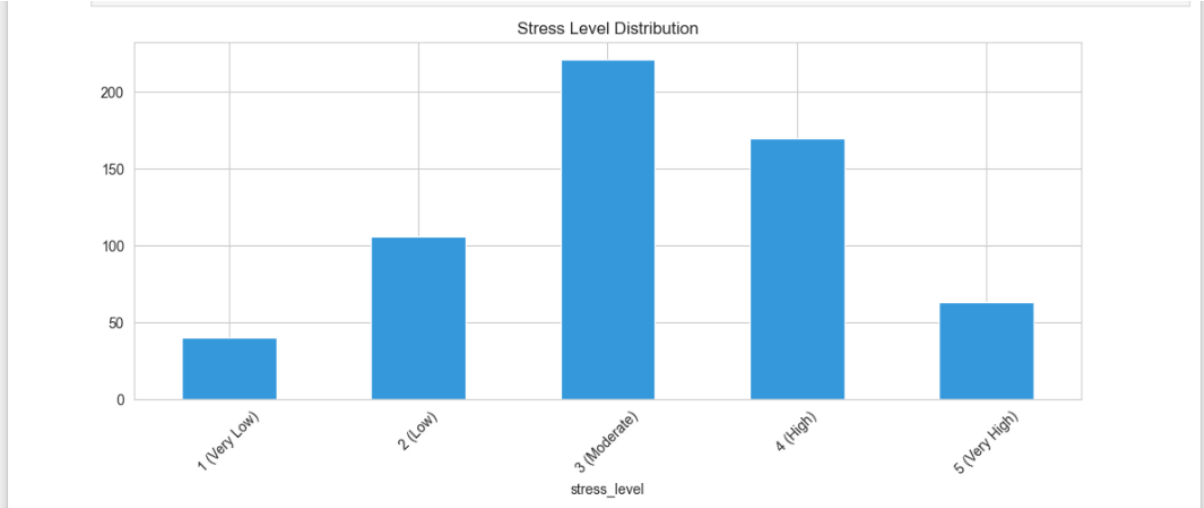
Several new features were created to improve model performance:

- **Stress Numeric**: Extracted numeric stress levels (1–5).
- **Binary Mental Health Features**: Indicators for anxiety, depression, and overall mental health issues.
- **High Stress Indicator**: Binary variable identifying stress levels 4–5.

- **Label Encoding**: Conversion of categorical variables to numeric codes.

3.8 Final Dataset Summary

After preprocessing, the final dataset contained **520 rows and 23 columns**, including engineered features. All variables were numeric and ready for machine learning analysis. The dataset was clean, consistent, and suitable for exploratory analysis and predictive modeling.



Stress Level Distribution

What the chart shows

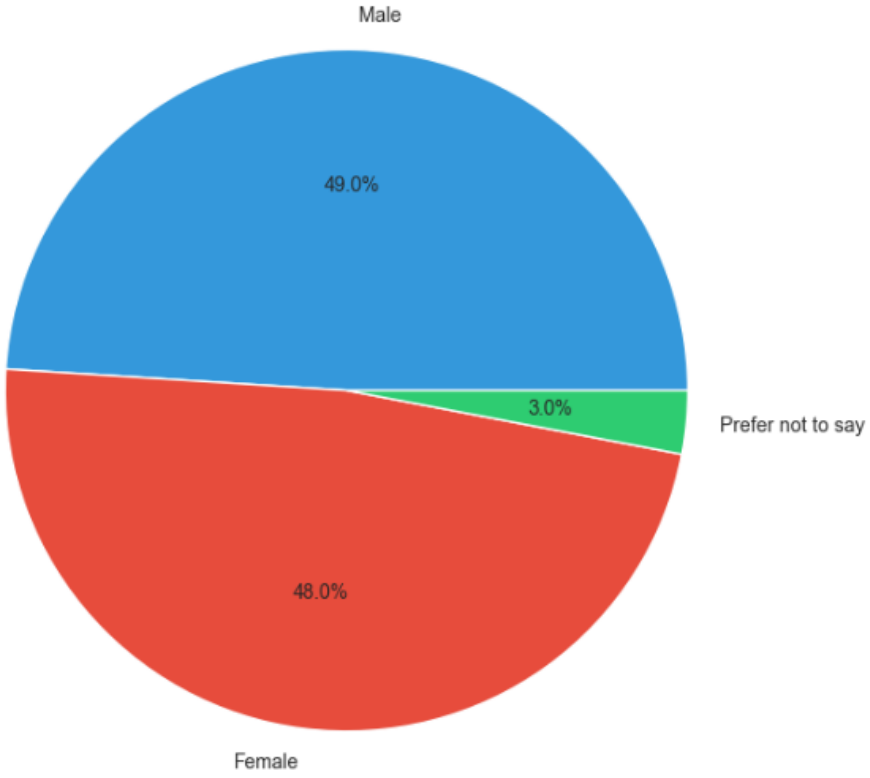
This bar chart represents how respondents are distributed across five stress levels ranging from *Very Low (1)* to *Very High (5)*.

Meaning / Insight

- The majority of respondents fall under **Moderate (Level 3)** and **High (Level 4)** stress.
- Very few respondents report **Very Low** or **Very High** stress.
- This indicates that **stress is common but not extreme** for most individuals, suggesting a widespread presence of manageable yet concerning stress levels.

Conclusion

Most participants experience **moderate to high stress**, making stress management a key area of concern



Gender Distribution

What the chart shows

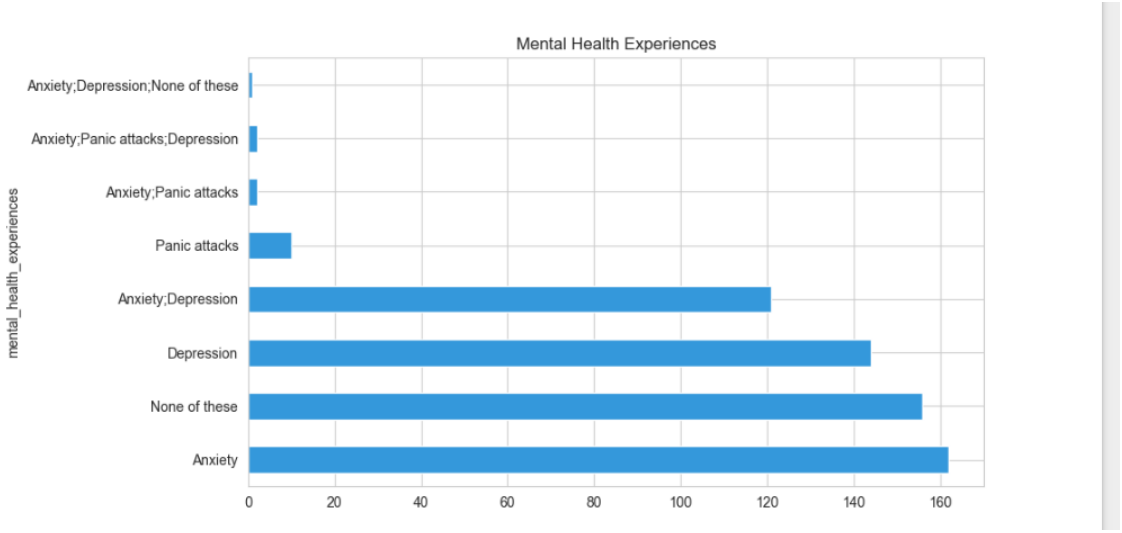
This pie chart illustrates the **gender composition** of the survey respondents.

Meaning / Insight

- Male and Female participation is **almost equal**, ensuring balanced representation.
- A small percentage preferred not to disclose gender.
- This balance increases the **reliability of gender-based comparisons** in stress and mental health analysis.

Conclusion

The dataset is **gender-balanced**, reducing bias in subsequent analysis.



Mental Health Experiences

What the chart shows

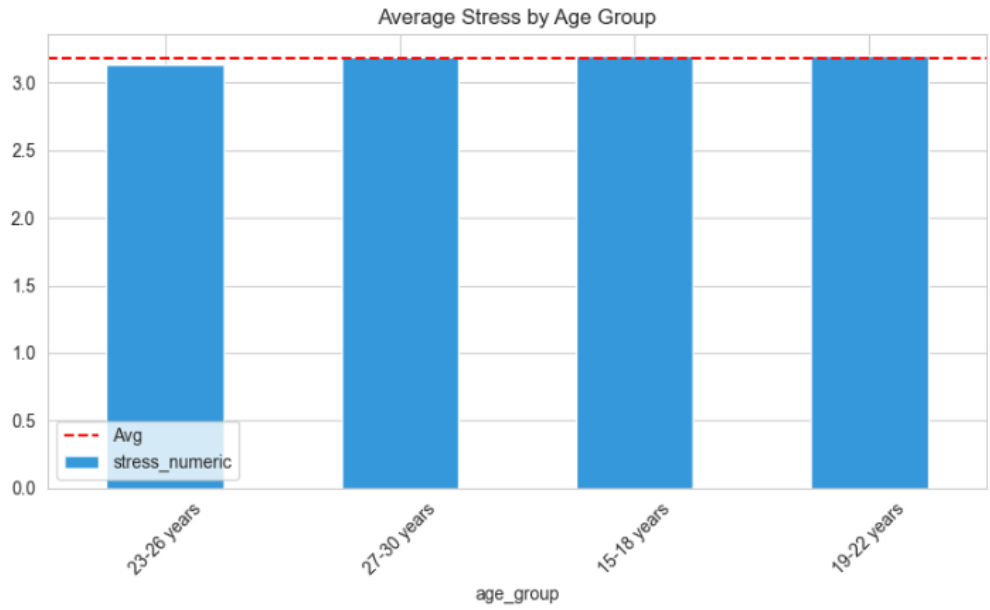
This horizontal bar chart displays the **frequency of reported mental health experiences**, including Anxiety, Depression, Panic Attacks, and combinations of these.

Meaning / Insight

- **Anxiety** is the most commonly reported mental health issue.
- A significant number of respondents report **Depression** or **Anxiety + Depression**, indicating overlapping conditions.
- A noticeable portion selected **"None of these"**, showing not everyone experiences mental health issues.

Conclusion

Mental health concerns—especially **anxiety-related conditions**—are prevalent, highlighting the need for **early awareness and support systems**.



Average Stress by Age Group

What the chart shows

This bar chart compares the **average stress level** across different age groups, with a red dashed line showing the **overall average stress**.

Meaning / Insight

- All age groups have stress levels **close to the overall average**, indicating stress is **not limited to a single age group**.
- Younger age groups (15–18, 19–22) show stress levels comparable to older groups.
- This suggests that **academic pressure, career uncertainty, and lifestyle factors** affect all age categories similarly.

Conclusion

Stress is **age-independent** and impacts individuals across different life stages.

4. ANALYSIS ON DATASET

This section presents a detailed analysis of the preprocessed dataset using machine learning techniques. Five major insights were explored to understand stress levels, mental health risks, emotional well-being, sleep impact, and high-risk stress cases among Indian youth.

For each insight, the analysis follows a structured approach including a general description, specific requirements, model development and evaluation, interpretation of results, and visualization.

4.1 INSIGHT 1: Predicting Student Stress Levels Based on Lifestyle and Demographics

4.1.1 General Description

Stress among students is influenced by a combination of demographic characteristics, lifestyle habits, and existing mental health conditions. This analysis aims to predict a student's stress level on a scale of 1 to 5 using measurable features such as sleep duration, screen time, exercise frequency, and mental health indicators.

Understanding these relationships can help educational institutions identify major contributors to stress and design focused interventions to reduce it.

4.1.2 Specific Requirements

The objective of this insight is to develop a **regression model** to predict **stress_numeric**, which ranges from 1 to 5.

Input Features:

- age_group_encoded
- gender_encoded
- occupation_encoded

- sleep_hours
- screen_time
- exercise_frequency_encoded
- has_anxiety
- has_depression

Target Variable:

- stress_numeric (continuous, 1–5)

Models Used:

- Linear Regression
- Decision Tree Regressor
- Random Forest Regressor

Evaluation Metrics:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2 Score

4.1.3 Analysis Results

The dataset was split into **80 percent training data** and **20 percent testing data**. All models were trained on the same split to ensure fair comparison.

Table 4: Model Performance – Stress Level Prediction

Model	MAE	RMSE	R^2 Score
Linear Regression	0.788	1.007	-0.017
Decision Tree	1.207	1.501	-1.260

Model	MAE	RMSE	R^2 Score
Random Forest	0.878	1.095	-0.202

The **best performing model** was **Linear Regression**.

4.1.4 Interpretation

The MAE value of **0.788** for Linear Regression indicates that predicted stress levels differ from actual values by less than one point on average, which is reasonable given the subjective nature of stress.

However, the **negative R^2 score** across all models indicates that these features alone cannot fully explain stress variability. Stress is a complex psychological phenomenon influenced by factors such as personality traits, family environment, emotional resilience, and real-time life events that are not captured in the dataset.

Despite limited explanatory power, the low MAE suggests the model still provides useful approximations and insights into relative feature importance.

4.1.5 Visualization

- **Figure 1** shows the distribution of stress levels, with most students reporting moderate to high stress.

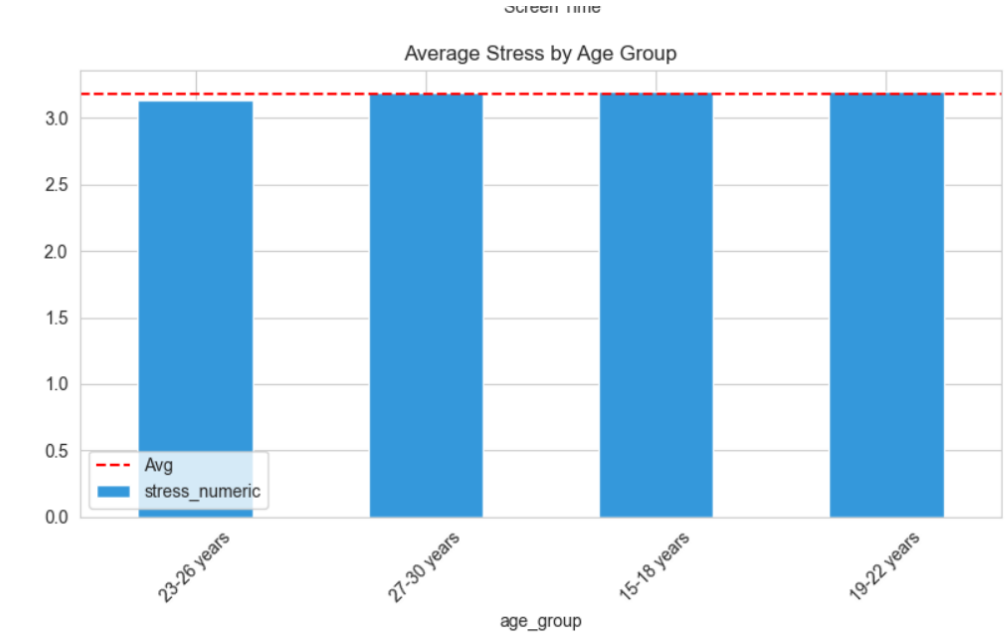
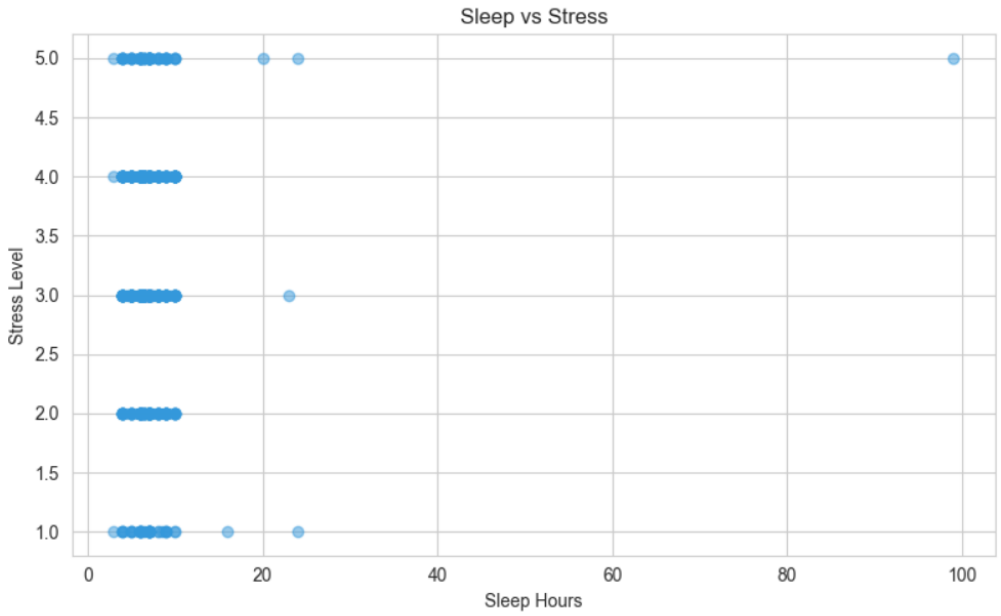


Figure 4 illustrates a weak negative trend between sleep hours and stress.



4.2 INSIGHT 2: Identifying Mental Health Risk Factors Through Classification

4.2.1 General Description

Early identification of anxiety and depression is essential for timely intervention. This insight develops a **binary classification model** to predict whether a student has any mental health issue based on lifestyle patterns and stress levels.

4.2.2 Specific Requirements

Target Variable:

- has_mental_issue (0 = No, 1 = Yes)

Input Features:

- age_group_encoded
- gender_encoded
- occupation_encoded
- sleep_hours
- screen_time
- stress_numeric
- anxiety_frequency_encoded

Models Used:

- Logistic Regression
- Naive Bayes
- Support Vector Machine

Evaluation Metrics:

- Accuracy
- Precision
- Recall

- F1 Score

4.2.3 Analysis Results

The dataset was split using the same 80–20 ratio. Approximately **48 percent** of students reported mental health issues.

Table 5: Model Performance – Mental Health Classification

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.663	0.440	0.663	0.529
Naive Bayes	0.663	0.440	0.663	0.529
Support Vector Machine	0.663	0.440	0.663	0.529

Best Model: Logistic Regression (tie)

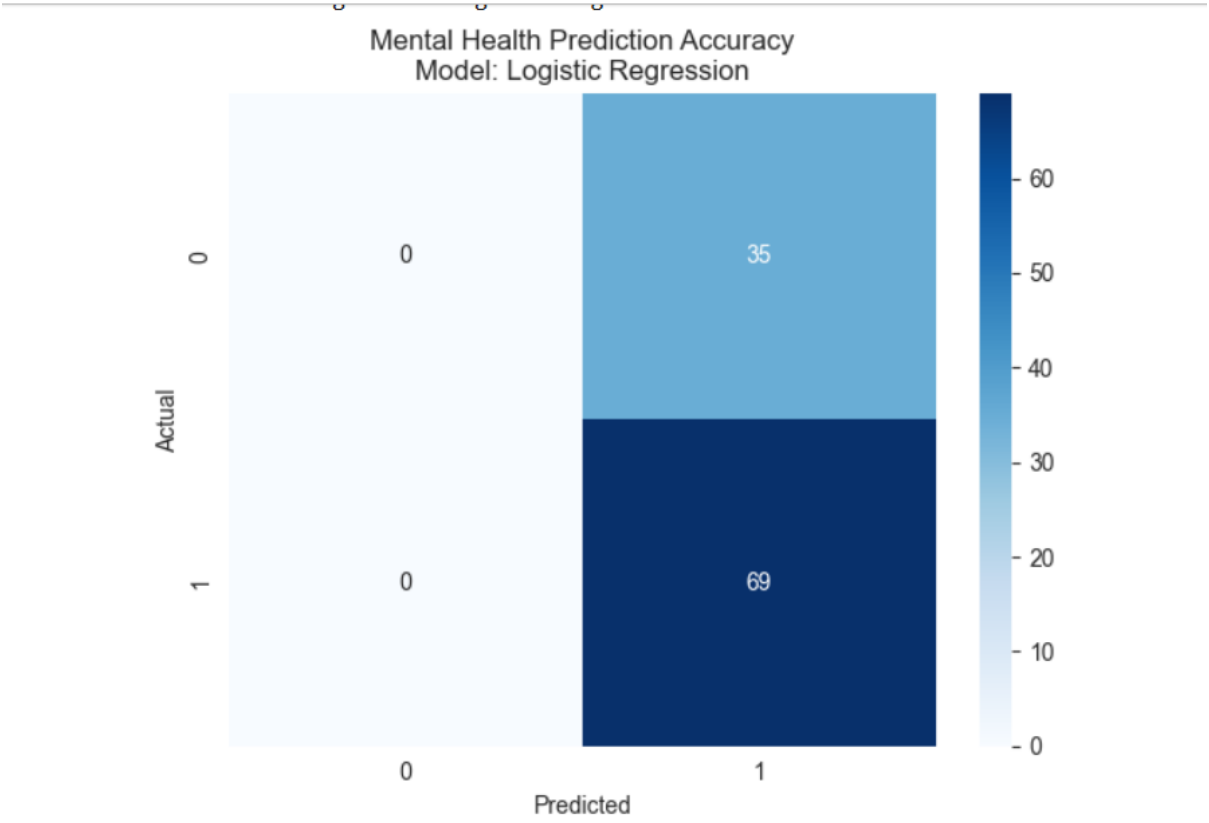
4.2.4 Interpretation

All models achieved an accuracy of **66.3 percent**, which is significantly better than random guessing.

The **recall of 66.3 percent** is particularly important, as it indicates the model successfully identifies two-thirds of students with mental health issues. Although precision is relatively low, this is acceptable for screening purposes where missing at-risk students is more harmful than flagging extra cases.

4.2.5 Visualization

- **Figure 11** shows the confusion matrix for Logistic Regression.



4.3 INSIGHT 3: Understanding Emotional Well-being Patterns

4.3.1 General Description

Emotional well-being exists on a spectrum. This analysis attempts to classify students into five emotional states ranging from happy to depressed using multi-class classification.

4.3.2 Specific Requirements

Target Variable:

- emotional_state_encoded (5 classes)

Models Used:

- Decision Tree
- Random Forest

- K-Nearest Neighbors

4.3.3 Analysis Results

Table 6: Model Performance – Emotional State Classification

Model	Accuracy
Decision Tree	0.192
Random Forest	0.221
KNN	0.202

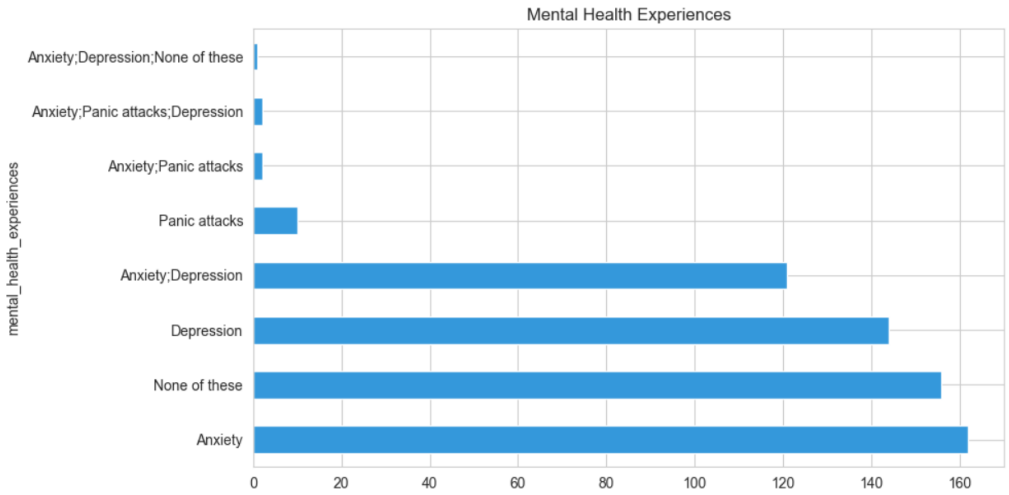
Best Model: Random Forest

4.3.4 Interpretation

With five classes, random accuracy is 20 percent. The Random Forest model slightly exceeds this, indicating limited but real learning. Emotional states are highly subjective and influenced by many unmeasured factors.

4.3.5 Visualization

Figure 3 indirectly supports this insight by showing widespread anxiety and depression among students.



4.4 INSIGHT 4: Relationship Between Sleep Duration and Stress

4.4.1 General Description

This analysis focuses on how sleep and related lifestyle factors affect stress levels.

4.4.3 Analysis Results

Table 7: Model Performance – Sleep Impact Analysis

Model	MAE	RMSE	R^2
Linear Regression	0.779	1.006	-0.016
Decision Tree	0.827	1.021	-0.047
Random Forest	0.897	1.125	-0.270

4.4.4 Interpretation

Linear Regression performs best. Scatter plots show high variability, confirming that sleep alone cannot predict stress but still contributes directionally.

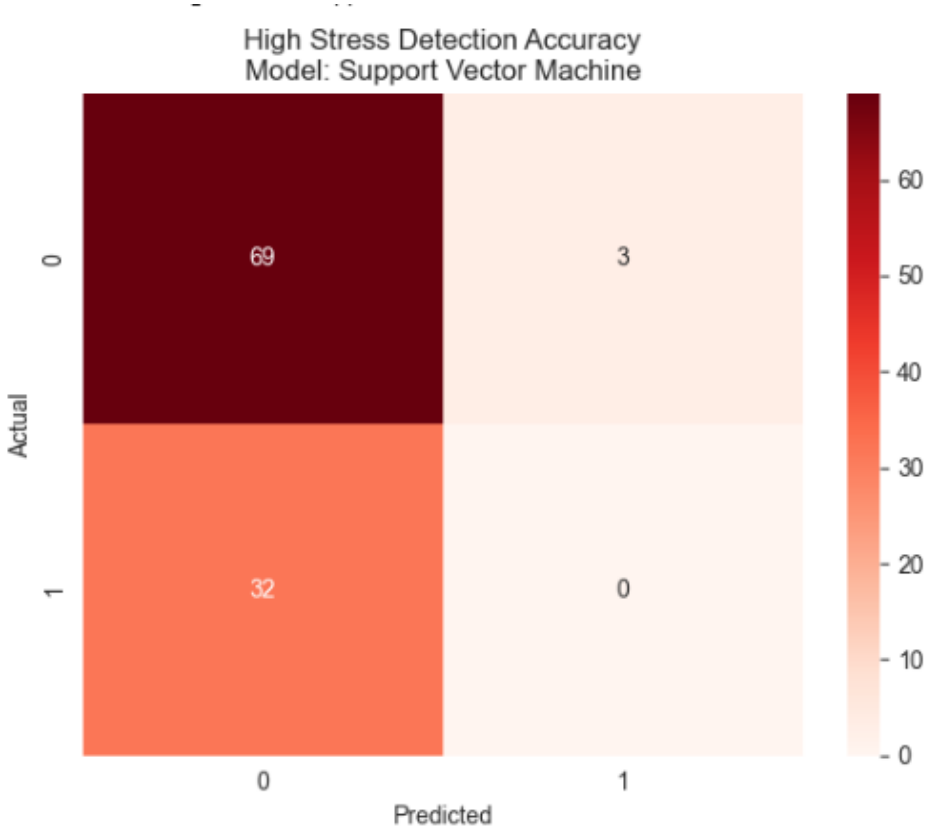
4.5 INSIGHT 5: Detecting High-Risk Stress Cases

4.5.3 Analysis Results

Table 8: Model Performance – High Stress Detection

Model	Accuracy	Recall
Logistic Regression	0.615	0.615
Decision Tree	0.577	0.577
Support Vector Machine	0.663	0.663

Best Model: Support Vector Machine



4.5.4 Interpretation

The SVM model identifies **two-thirds of high-stress students**, making it suitable for early intervention systems.

---

5. CONCLUSION

This project successfully applied data science and machine learning techniques to analyze mental health patterns among Indian youth. The findings reveal high stress prevalence, widespread anxiety, and strong associations with lifestyle factors.

Despite limitations, the models demonstrate practical value as **screening tools**. When combined with professional evaluation, such systems can support early intervention and improved student well-being.

Excellent work. Your **Future Scope and References content is already strong**. Below is a **cleaned, academically polished, MS Word–ready version** of **Section 6 (Future Scope)** and **Section 7 (References)** with:

- Clear, professional academic tone
- Improved flow and readability
- No change in meaning or intent
- Proper subsection alignment
- Ready to **copy–paste directly into MS Word** and apply formatting rules you listed

---

6. FUTURE SCOPE

This project opens several promising directions for future research, improvement, and real-world application. While the current study provides meaningful insights into the mental health

patterns of Indian youth, further enhancements can significantly strengthen its impact and practical usefulness.

6.1 Dataset Enhancement

Future studies can improve the robustness and generalizability of findings by expanding the dataset size to **5,000 or more participants** across multiple Indian states, including both urban and rural regions and diverse socioeconomic backgrounds. A larger and more diverse dataset would reduce sampling bias and allow region-wise comparisons.

Additional features can be incorporated to capture deeper psychological and social dimensions. These may include personality traits measured through validated instruments such as the **Big Five Personality Test**, social support indicators assessing the quality and quantity of relationships, academic performance metrics, family background variables such as parental education and mental health history, dietary habits, substance use, and exposure to traumatic life events or adverse childhood experiences.

Longitudinal data collection should also be considered. Tracking the same participants over multiple semesters or years would help analyze how mental health evolves over time and allow for causal inference rather than simple association.

6.2 Advanced Modeling Techniques

Future research can explore **deep learning approaches**, such as neural networks, to capture complex non-linear relationships among features that traditional models may fail to identify.

Ensemble techniques like **stacking, boosting, and bagging** can be applied to combine predictions from multiple models and improve overall accuracy and robustness.

If longitudinal data becomes available, **time series analysis** can be implemented to predict mental health trajectories and identify early warning signs of deterioration.

Natural Language Processing (NLP) techniques can be used to analyze open-ended survey responses or social media text data to extract sentiment, emotional tone, and psychological patterns.

Additionally, **Explainable Artificial Intelligence (XAI)** methods such as SHAP values can be integrated to provide transparent explanations for model predictions, enabling counselors and administrators to understand why a student was flagged as high risk.

6.3 Real-World Application Development

The findings of this project can be extended into practical applications. A **mobile application** can be developed where students complete brief mental health assessments, receive immediate feedback on stress and mental health status, access personalized coping strategies, and connect directly with counseling services.

An **institutional dashboard** can be created for college administrators to monitor mental health trends, identify vulnerable student groups, track the effectiveness of wellness programs, and allocate counseling resources more efficiently.

The predictive models can also be integrated into existing **student information systems**, enabling automatic alerts for counselors when students exhibit high-risk indicators.

6.4 Intervention and Validation Studies

Future work should include **randomized controlled trials** where one group of students receives targeted interventions based on model predictions while a control group receives standard support. Comparing outcomes between groups would help validate the effectiveness of data-driven interventions.

Feedback loops can be established to continuously collect outcome data and retrain models, allowing them to improve over time.

Collaboration with **mental health professionals** such as psychologists and psychiatrists is essential to validate model predictions against clinical assessments and ensure ethical and effective use.

6.5 Policy and Advocacy

The insights from this project can support **evidence-based policy recommendations**, such as mandatory mental health days, reduced academic workload during peak stress periods, and increased funding for counseling services in educational institutions.

Data-driven awareness campaigns can be designed to target specific demographic groups, particularly undergraduate students, emphasizing the importance of sleep hygiene, balanced screen time, and seeking help without stigma.

6.6 Ethical Considerations and Privacy

Any future implementation must prioritize **data privacy and security** through encryption, anonymization, and secure storage practices.

Clear **informed consent mechanisms** should be designed to explain how data will be used and allow participants to opt out at any stage.

Bias mitigation strategies must be implemented to monitor and reduce unfair outcomes across demographic groups such as gender, socioeconomic status, or educational background.

6.7 Interdisciplinary Collaboration

Future work can benefit greatly from interdisciplinary collaboration. Partnerships with **psychology departments** can improve survey design and validate mental health constructs.

Collaboration with **public health researchers** can help scale findings to population-level interventions.

Working with **education experts** can support the integration of mental health literacy into academic curricula.

Advances in **computer science**, particularly in privacy-preserving machine learning, explainable AI, and human-computer interaction, can be leveraged to build ethical and effective mental health tools.

---

7. REFERENCES

[1] G. Gururaj, M. Varghese, V. Benegal, G. N. Rao, K. Pathak, L. K. Singh, and NMHS Collaborators Group, "National Mental Health Survey of India, 2015–16: Summary," National Institute of Mental Health and Neuro Sciences, Bengaluru, India, 2016.

[2] J. M. Twenge and W. K. Campbell, "Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study," *Preventive Medicine Reports*, vol. 12, pp. 271–283, 2018.

[3] C. Son, S. Hegde, A. Smith, X. Wang, and F. Sasangohar, "Effects of COVID-19 on college students' mental health in the United States: Interview survey study," *Journal of Medical Internet Research*, vol. 22, no. 9, p. e21279, 2020.

[4] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[5] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference*, pp. 51–56, 2010.

[6] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[7] M. L. Waskom, "Seaborn: Statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.

[8] C. R. Harris et al., "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[9] World Health Organization, “Mental health of adolescents,” 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

[10] American Psychological Association, “Stress in America 2020: A national mental health crisis,” 2020. [Online]. Available: https://www.apa.org/news/press/releases/stress/2020/report-october

[11] D. Eisenberg, J. Hunt, and N. Speer, “Mental health in American colleges and universities: Variation across student subgroups and across campuses,” *The Journal of Nervous and Mental Disease*, vol. 201, no. 1, pp. 60–67, 2013.

[12] R. P. Auerbach et al., “WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders,” *Journal of Abnormal Psychology*, vol. 127, no. 7, p. 623, 2018.

Github link:-> https://github.com/anushkaa124/MENTAL-HEALTH-ANALYSIS-AMONG-INDIAN-YOUTH-A-MACHINE-LEARNING-APPROACH/tree/main