Mental Health Analysis Among College Students Using Machine Learning Algorithms

[1]Zeal Shah,
Department of Computing Technologies, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute Of Science and Technology, Kattankulathur, 603203
India
[1]zs5263@srmist.edu .in

[2]Harsh Kelawala,
Department of Computing Technologies, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute Of Science and Technology, Kattankulathur, 603203
India
[2] hk9586@srmist.edu .in

[3,*]P.Selvaraj,
Department of Computing Technologies, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute Of Science and Technology, Kattankulathur, 603203
India
[3,*]selvarap@srmist.edu.in

Abstract— **Mental health challenges among college students are a growing concern globally. This study investigates the intricate relationship between various factors and mental health outcomes among the SRM Institute of Science and Technology students. A comprehensive questionnaire addressed emotional well-being, stress levels, social support, health practices, and academic pressure. Contrary to widespread assumption, the investigation found that sleep patterns, which are frequently regarded as crucial, did not significantly affect the indices of mental health. Additionally, extended irregular sleep did not result in normalization, emphasizing how intricate this relationship is. The study's findings emphasize the multifaceted nature of mental health, suggesting that a holistic approach encompassing stress management, social engagement, and emotional well-being, alongside sleep hygiene, is imperative. This research challenges existing perceptions, encouraging a nuanced understanding of the factors shaping mental health. The implications extend to the design of comprehensive interventions and support programs for college students. This study contributes valuable insights, paving the way for more informed, holistic, and practical strategies to nurture students' mental health**

Keywords— mental health, random forest, support vector machine, naïve bayes, logistic Regression, suicidal thoughts, stress

I. INTRODUCTION

In recent years, mental health challenges among college students have gained significant attention due to their prevalence and impact on academic performance, personal well-being, and prospects. The transition to University life, coupled with academic pressures, social expectations, and emotional challenges, often creates unique student stressors. Consequently, understanding the intricate dynamics of mental health in this demographic has become a critical area of research and intervention.

This study examines the intricate interplay of variables affecting the mental health of college students, with a focus on the SRM Institute of Science and Technology's student body. Mental health, a multifaceted construct, is influenced by various variables, including emotional well-being, stress levels, social support, lifestyle choices, academic pressures, and sleep patterns. While existing literature recognizes the importance of these factors, their relative significance and the nature of their impact on mental health remain subjects of intense exploration and debate.

Sleep patterns have been the subject of extensive research concerning mental health. The association between sleep quality, duration, and psychological well-being has been a central focus in mental health studies, with inadequate or irregular sleep often implicated in various mental health disorders. However, the relationship between sleep patterns and mental health outcomes is complex and multifactorial, varying across different contexts and populations.

In this context, this study seeks to unravel the specific dynamics within the unique demographic of students who face the everyday challenges of University life and the additional stressors associated with rigorous academic pursuits and demanding clinical training. By administering a comprehensive questionnaire to a diverse sample of students at the SRM Institute of Science and Technology, the study aim to investigate the role of sleep patterns alongside other factors in shaping mental health outcomes.

This research endeavors to challenge existing assumptions and add nuanced insights to understanding mental health in the academic setting. The exploration includes an in-depth analysis of sleep patterns, questioning the widely accepted beliefs regarding their direct influence on mental health. By considering a broad spectrum of variables, we aim to discern subtle interactions and identify critical determinants that significantly impact mental well-being among students.

The outcomes of this study not only contribute to the academic discourse surrounding mental health and have practical implications for the development of targeted interventions and support programs. It is crucial to comprehend the complexity of mental health in this setting to develop efficient, evidence-based interventions supporting college students' well-being. As such, this research is vital to fostering a mentally resilient academic community, providing valuable insights for educational study and practical mental health interventions.

II. STATE-OF-THE-ART

A: Mental Health Questionaire: The Dataset

The cornerstone of this research project was the extensive and diverse dataset provided by students from various departments of SRMIST University. The meticulously collected and anonymized dataset served as the bedrock upon which our comprehensive Mental Health Analysis among college students was built. The dataset encompasses various variables, meticulously curated to capture the multifaceted aspects of students' lives, academic journeys, and mental well-being.

The dataset was meticulously curated through structured questionnaires distributed among the student body. These questionnaires were designed with care and precision and delved into various crucial factors influencing mental health. Students from diverse departments of SRMIST University willingly participated, providing valuable insights into their experiences, challenges, and coping mechanisms.

- **Variables of interest:**

Within this dataset, many variables were explored, capturing not only demographic details but also intricate facets of students' lives. Sleep patterns, stress levels, social support networks, involvement in extracurricular activities, academic pressures, and specific mental health indicators were among the pivotal variables meticulously documented. This comprehensive approach ensured a nuanced understanding of the complex interconnections between these factors and mental health outcomes.

- **Anonymity and ethical considerations:**

It is paramount to note that the dataset was handled with utmost confidentiality and care. All responses were anonymized to protect the identity and privacy of the participating students. Ethical guidelines and consent procedures were strictly adhered to, ensuring every individual's voluntary and informed participation. These ethical considerations were fundamental in fostering a sense of trust and ensuring the integrity of the research endeavor.

- **Representativeness and diversity:**

One of the strengths of this dataset lies in its diversity. Students from various departments, backgrounds, and academic levels within SRMIST University actively contributed, making the dataset representative of the University's vibrant student population. This diversity enriched the dataset and allowed for analyses stratified by department and demographic factors, providing a more nuanced understanding of mental health challenges among student groups.

- **Dataset utilization:**

The richness of this dataset enabled a comprehensive exploration of mental health factors, employing advanced statistical analyses and machine learning techniques. By leveraging this data, the research aimed to unravel intricate patterns, identify significant correlations, and develop predictive models. The dataset's utilization facilitated evidence-based conclusions and actionable insights, guiding the formulation of targeted interventions and support strategies.

This dataset served as the cornerstone of our research, offering a window into the complex world of college students' mental health. Its diversity, depth, and ethical foundations provided the foundation for our in-depth analyses, making it an asset in our pursuit of understanding and improving mental health outcomes among SRMIST University students.

Beyond the quantitative variables, the dataset also included qualitative insights obtained through open-ended questions. These qualitative responses provided depth to our analysis, illuminating the students' emotions, struggles, and coping mechanisms. Qualitative data, often overlooked in traditional analyses, enriched our understanding by capturing the human experiences behind the statistics. This qualitative dimension lets us contextualize the quantitative findings, providing a holistic view of the student's mental health landscape.

B: Problem Statement

Mental health issues among college students, including stress, anxiety, and depression, have become increasingly prevalent, posing significant challenges to their well-being and academic success. SRM Institute of Science and Technology is not exempt from this growing concern. Despite the availability of support services, a critical need remains to understand the intricate factors contributing to mental health disparities among students. Moreover, while existing research acknowledges the importance of sleep patterns, stress levels, social support, and academic pressures, the specific interplay of these factors in the context of mental health among students remains inadequately explored.

The lack of in-depth, data-driven analyses tailored to students' unique challenges presents a substantial gap in our understanding. Addressing this gap is vital for academic research and developing targeted interventions and support programs.

- **Complex Interplay of Factor**s: Students juggle an array of stressors, including rigorous academic demands, clinical training pressures, and the expectation of maintaining a healthy work-life balance. Understanding how these factors intersect and impact mental health is essential to tailor effective interventions.

- **Sleep Patterns and Mental Health:** While sleep is universally acknowledged as vital for overall health, the intricate relationship between sleep patterns and specific mental health indicators among students is insufficiently studied. Sleep quality, duration, and consistency play crucial roles that demand thorough exploration.

- **Tailored Interventions:** Existing mental health support programs often adopt generalized approaches. To create impactful interventions, it is imperative to identify the unique challenges students face, including the role of sleep patterns, to develop targeted and personalized strategies.

- **Data-Driven Understanding:** Rigorous data analysis is critical to unraveling the complexities of mental health. By .This study attempts to close the gap between raw data and actionable insights by utilizing statistical approaches and machine learning algorithms, offering a thorough, evidence-based understanding of the factors influencing students' mental health.

Addressing this complex problem necessitates a comprehensive research endeavor that delves deep into the dataset, uncovering hidden patterns and relationships.
College students face various mental health issues, from unrecognized psychological suffering to a dearth of efficient interventions. Concerns regarding students' well-being and academic progress are raised by the rising rates of mental health problems on college campuses. Current diagnostic techniques frequently lack precision, which results in incorrect identification and postponed intervention. Additionally, the stigma associated with mental health prevents students from actively seeking assistance, aggravating the issue. Although promising, the current research frequently faces constraints like limited sample sizes, a lack of diverse data, and inconsistent techniques. Furthermore, even though it has been acknowledged, further research and development are need to fully understand the predictive power of machine learning algorithms in relation to mental health. Consequently, creating a solid and trustworthy framework is the main issue. By doing so, this study aims to contribute to academic knowledge and provide practical recommendations for mental health interventions, ultimately fostering a healthier and more resilient student community within the SRM Institute of Science and Technology.

III. LITERATURE REVIEW

The foundation of every extensive research project is the literature review, which provides an overview of the body of knowledge already available in an area, as well as research gaps and pertinent approaches. This section extensively explores scholarly works, studies, and publications related to our research topic. By critically examining previous research findings and methodologies, the study aims to identify trends, inconsistencies, and areas where this study can make a meaningful contribution. This in-depth analysis informs the theoretical framework of the research and provides valuable insights that guide the methodology and objectives of this study. The literature review contextualizes this research within the broader academic landscape, establishing a foundation for new contributions and discoveries.

- In their study [1], Sawangarreerak et al. employ a combination of random oversampling and undersampling methods, Chi-square, Gini index, and Information Gainer Method, to effectively balance imbalanced data. Their innovative approach ensures the accuracy and robustness of predictive models, providing a valuable contribution to mental health prediction. By enhancing feature selection and employing Random Forest classification, their research not only refines the precision of predictive models but also offers promising strategies for early identification and support of students at risk of depression, thereby significantly impacting mental health initiatives in higher education.
- In the study [2], Chen et al. explore the intricate dynamics of college students' mental health promotion, specifically focusing on the mediating role of online mental health information seeking. The researchers examine the complex interaction of internal and environmental factors influencing students' mental well-being using cutting-edge statistical techniques, including partial least squares and maximum likelihood. One pivotal insight from their research underscores the significant impact of external factors over internal factors on student health. This study serves as a valuable foundation, shedding light on the nuanced relationships between online mental health information seeking, external influences, and mental health promotion among college students. As we delve into this literature, we aim to unpack these insights further, drawing upon this research to enrich our understanding of the factors influencing mental health in the college student population.

College Students	Type	Number	Proportion
Gender	Male	235	47.86%
	Female	256	52.14%
Age	Under 18 years	30	6.11%
	18-20 years	212	43.18%
	20-24 years	196	39.92%
	Over 24 years	53	10.79%
Education	Junior College	94	19.14%
	Undergraduate	257	52.34%
	Master	88	17.92%
	Doctor & above	52	10.59%

Table 3.1: The Sample Statistics of Students

Table 3.1 in the study [2] displays the proportion and statistics of the students concerning their gender, age, and education.

- The research conducted by Tran et al. [3] provides a crucial insight into the health assessment of University students in France, mainly focusing on mental health disorders. The study examines numerous risk variables linked to mental health issues among undergraduate students using statistical techniques, including the Chi-squared test and Fischer's exact test. However, it is noteworthy that the research emphasizes the limitation of the findings due to the limited scope of student information assessed. The inability to extrapolate these results to the entire French undergraduate student body underscores the requirement for additional thorough research in this area. Our review acknowledges this limitation while delving into this research's valuable insights. We recognize the necessity for future studies to encompass a broader student demographic for a more comprehensive understanding of mental health factors in University settings.

- The study [4] conducted by Lei et al. delves into the intricate landscape of mental health among college students in Hubei, China. The research uses statistical tools such as Linear Regression and t-tests to investigate factors influencing students' mental well-being. Self-rated health is a significant predictor, indicating its pivotal role in mental health outcomes. Study pressure and social support also play crucial roles, highlighting the delicate balance between academic demands and social connections. Interestingly, gender disparities in mental health scores were not substantial, challenging prevailing assumptions. However, it is vital to note that these results are derived from a single University, underscoring the need for caution when generalizing findings. As we explore this study, we acknowledge the localized context and focus on self-rated health, study pressure, and social support, contributing to a nuanced understanding of mental health factors in this region. Additionally, we recognize the imperative for broader studies encompassing diverse university settings to analyze mental health challenges among Chinese college students comprehensively.

- Chu et al.'s study [5] explores a significant aspect of mental health research among college students. The research uses clustering analysis algorithms to identify early signs of mental health issues. While the study leverages clustering techniques, it is essential to note that only a limited set of parameters has been examined, indicating a potential gap in the comprehensive understanding of student mental well-being. Applying clustering algorithms marks a proactive approach toward recognizing mental health concerns in their early stages. However, the study's scope underscores the necessity for broader exploration, encompassing a more extensive range of parameters, to comprehensively analyze college students' mental health. In our review, we acknowledge the study's innovative methodology while emphasizing the importance of further research that delves deeper into the multifaceted factors contributing to mental health challenges among college students.

- The research study [6] by Mikolajczyk et al. investigates the prevalence of depressive symptoms among University students across different European countries. Employing statistical tools like Variance and Logistic Regression, the study delves into various factors influencing the mental well-being of students. One of the crucial findings of this study highlights a direct correlation between lower income levels and depressive symptoms. This revelation underscores the socioeconomic dimension of mental health, emphasizing the impact of financial constraints on students' emotional well-being. As we examine this study, we recognize the significance of economic factors in mental health disparities among University students and the imperative for targeted interventions to support students from lower-income backgrounds, fostering a more inclusive and supportive educational environment.

- The research study [7] by Yang et al. explores the intricate relationship between college students' mental health and various influencing factors. The study examines the intricacies of student well-being by using cutting-edge methods, including Regression Analysis based on the Statistical Package for Social Sciences (SPSS). One of the critical findings underscores a significant correlation between social support and students' mental health. This discovery highlights the pivotal role of social connections in shaping students' emotional well-being, emphasizing the importance of robust support networks within the college community. As we delve into this study, we acknowledge the critical impact of social support on mental health and consider it a crucial element in formulating holistic interventions and support systems for college students, ultimately fostering a nurturing and inclusive campus environment.

- In the study [8] conducted by Mann JJ et al., the focus is on utilizing Classification and Regression Trees (CART) to discern suicide attempters within major psychiatric disorders. This retrospective study delves into the complexities of past suicidal behavior, identifying correlates through CART methodology. However, it is essential to note that this retrospective nature limits the study's scope, as it can only identify past correlates of suicidal behavior. The research does not provide direct insights into the predictive relevance of risk factors for future suicidal behavior. While shedding light on crucial correlations, this study emphasizes the need for prospective research to comprehensively understand the predictive salience of risk factors, offering valuable insights for future suicide prevention strategies.

- The study [9],which Chung and Teo authors, focuses on applying machine learning techniques in mental health prediction. This comprehensive exploration encompasses a range of algorithms, including Bayesian network, Naive Bayes, Logistic Regression, Multiple layer perceptron, Sequential minimal optimization, K-star, Random subspace, Random forest, and Random tree. Despite the promising potential of machine learning, the research landscape faces significant challenges. A common struggle highlighted in the literature is the difficulty in validating results due to limited external evidence. Additionally, the performance of machine learning models varies across mental health problems, influenced by factors such as data samples, features, and preprocessing activities like data cleaning and parameter tuning. This review underscores the need for rigorous validation methodologies. It emphasizes the critical impact of data quality and preprocessing on the effectiveness of machine learning applications in mental health prediction.

- In the study [10] by Vaishnavi Konda, various machine learning algorithms, including K-nearest neighbors (KNN), Logistic Regression, Decision Tree, and Stacking, are explored for predicting mental health illnesses. The research findings indicate that all five methods, when applied, demonstrate good accuracy, each yielding approximately 79%. However, a noteworthy aspect highlighted in the study is the limitation of a minimal dataset. The accuracy of predictions is acknowledged to be potentially enhanced with a more extensive `dataset.

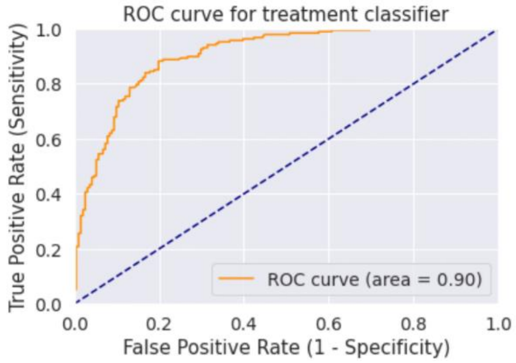


Figure 3.1:ROC of KNN

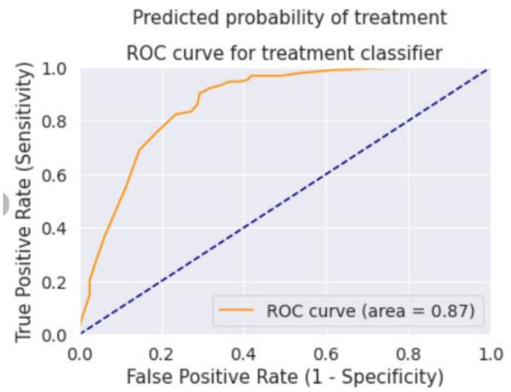


Figure 3.2: ROC of Decision Tree

Figure 3.1 and 3.2 are displayed in the study [10], where author tries to compare ROC ok both KNN and Decision Tree.

IV. SYSTEM ARCHITECTURE AND DESIGN

The architectural design for our project involves a systematic flow of processes, beginning with data collection through a questionnaire and culminating in the training and comparison of different machine learning models. Here is a breakdown of the architectural design for our study:

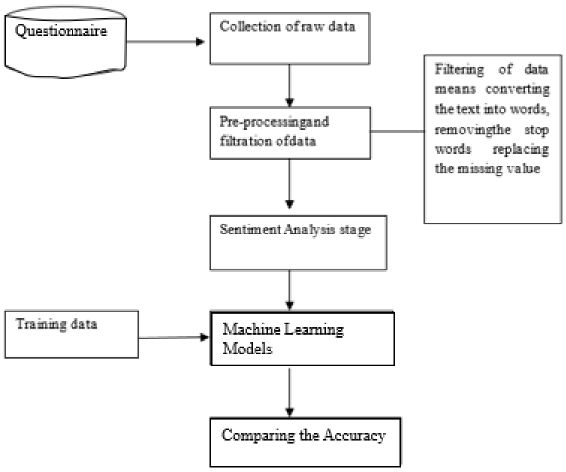
- **Data Collection:**
  - ➢ Questionnaire Administration: The process starts with distributing a well-structured questionnaire to a group of students at the SRM Institute of Science and Technology. This questionnaire covers various aspects related to mental health, including sleep patterns, stress levels, social support, and academic pressures.
  - ➢ Data Entry and Storage: Responses from the questionnaire are collected and entered into a structured digital format, creating the raw dataset. This dataset is securely stored in a designated database or data repository.
- **Data Preprocessing and Splitting:**
  - ➢ Data Cleaning: The raw dataset undergoes data cleaning processes to handle missing values, outliers, and inconsistencies. Cleaning ensures that the dataset is of high quality, enhancing the accuracy of subsequent analyses.
  - ➢ Feature Selection: Relevant features related to sleep patterns, stress, social support, and academic pressures are selected based on domain knowledge and data analysis. Feature selection optimizes the dataset for training machine learning models.
  - ➢ Train-Test Split: A training set and a test set are created from the preprocessed dataset. The remaining fraction of the data is utilized as an unobserved test set for model evaluation. In contrast, the training set, which typically makes up 70–80% of the data, trains machine learning models.
- **Model Training**
  - ➢ Machine learning models are chosen depending on the dataset's properties and the problem type (classification or Regression). Algorithms used in machine learning include support vector machines, neural networks, decision trees, random forests, and linear and logistic Regression. - Training the Models: The training dataset's features and labels are used to train the selected machine learning models. The models develop predictive or categorical abilities during training by discovering patterns and correlations in the data.
- **Model Comparison:**
  - ➢ Performance Metrics: After training, the model's performance is evaluated using appropriate metrics such as accuracy, precision, recall, F1-score, or mean squared error. Based on the test data, these metrics quantify how well each model predicts mental health outcomes.
  - ➢ Comparative Analysis: The performance metrics of each model are compared to determine which algorithm performs best in predicting mental health indicators. Comparative analysis guides the selection of the most suitable model for further use and interpretation.

- **Result Interpretation and Insights:**
  - ➢ Visualization: Results, including model predictions and comparative analyses, are visualized using charts, graphs, and plots.

Visualization aids in communicating complex findings clearly and intuitively.

- Insights and Recommendations: The interpreted results provide valuable insights into the influence of sleep patterns, stress, social support, and academic pressures on mental health outcomes among students. Based on these insights, recommendations for interventions, support programs, or further research are formulated.

This architectural design ensures a systematic approach to gathering, preprocessing, and analyzing the data. It allows for exploring diverse machine learning models, enabling a comprehensive understanding of the factors affecting mental health among students. The results derived from this approach provide a solid foundation for evidence-based interventions and contribute significantly to the ongoing



discourse on mental health in an academic environment

Figure 4.1 System Architecture

Figure 4.1 illustrates the methodology and system architecture through a detailed flow diagram. The initial step involves collecting data as responses to a structured questionnaire. Subsequently, the data undergoes meticulous preprocessing and filtration to rectify errors, remove inconsistencies, and eliminate irrelevant information.

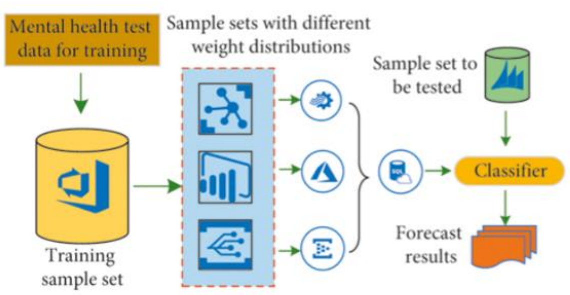


Figure 4.2 Prediction of Mental Health Using Machine Learning

The study [5], as shown in Figure 4.2, shows the whole processing of the prediction of mental health among college students. A simple machine learning architecture is to be followed to achieve the goal of predicting whether the student has having mental health illness or not.

VI. DATA VISUALIZATION OUTCOMES

A. *Relationship Between Social Anxiety and Mental Health*

The data visualization analysis illuminates a compelling connection between social issues and mental health challenges among college students. It unequivocally demonstrates that when students face social problems such as isolation, financial stress, or discrimination, their vulnerability to mental health issues significantly heightens.

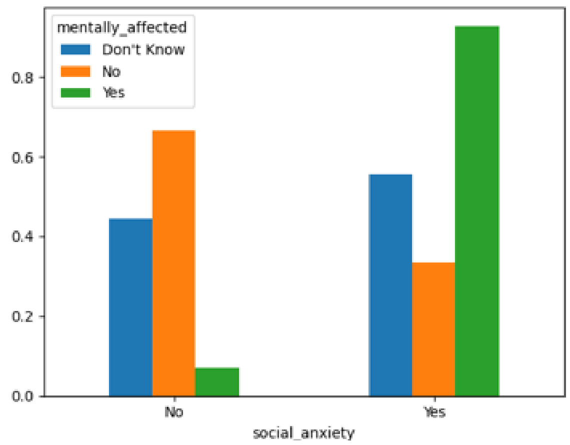


Figure 6.1: Social Anxiety v/s Mentally affected

Figure 6.1 portrays how these social factors are precursors to mental health struggles, emphasizing the critical importance of addressing social issues in holistic mental health support programs. This insight underscores the pressing need for educational institutions to focus on individual mental health and proactively mitigate social challenges, fostering inclusive and supportive environments essential for students' well-being.

B. *Relationship Between Age and Mental Illness*

The data visualization analysis reveals a discernible pattern regarding age's impact on college students' mental health. Strikingly, the data demonstrates that students aged 19 to 23 exhibit a gradual decrease in the prevalence of mental health challenges. This decline suggests that as students mature through their early twenties, they potentially acquire coping mechanisms and resilience, leading to a decreased susceptibility to mental illnesses. Particularly noteworthy is the absence of reported mental health issues among students at the age of 23, indicating a potential positive

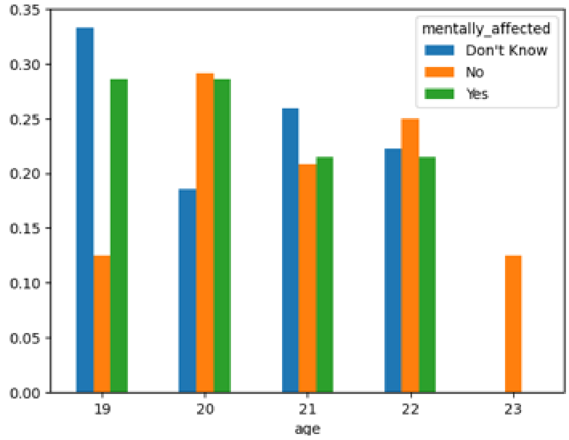


Figure 6.2 Age v/s Mental Illness.

Figure 6.2 displays the correlation between age-acquired coping skills and reduced vulnerability to mental health disorders. These findings underscore the importance of resilience-building initiatives and coping strategy education, especially for younger students, to equip them with the necessary skills to navigate the challenges of University life effectively.

C. Relationship Between Concentration and Mental Health

The data visualization analysis underscores a compelling relationship between students' concentration levels and mental health.

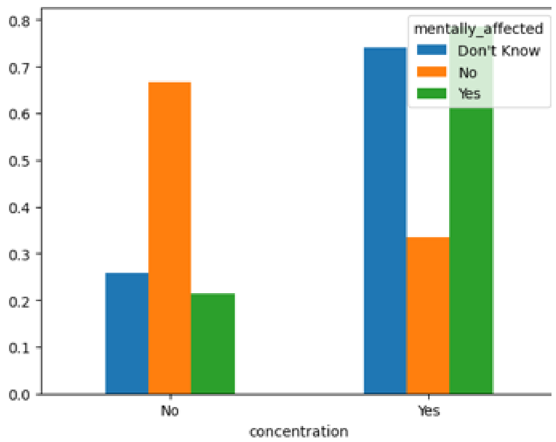


Figure 6.3 Concentration v/s Mental Illness.

Figure 6.3 reveals a clear pattern: students concentrating more on their studies are notably less prone to mental health issues. Conversely, those who struggle with reduced concentration face an increased vulnerability to mental health challenges. This stark contrast suggests a protective effect of focused concentration on mental well-being. The data implies that students with enhanced concentration skills might possess effective coping mechanisms, reducing their susceptibility to stressors and mental health disorders. These findings emphasize the pivotal role of concentration-enhancing strategies and study techniques in promoting mental resilience among college students, highlighting the potential benefits of interventions to improve focus and attention.

D. Relationship Between Backloga and Mental Health

The data visualization analysis sheds light on the intricate relationship between students' stress levels and academic backlogs. Remarkably, students facing more academic backlogs exhibit a substantial increase in stress levels, creating a clear pattern in the data. The visualization vividly illustrates the rising trend of stress concerning the accumulation of backlogs, emphasizing the significant psychological impact of academic challenges. Furthermore, a detailed breakdown of stress levels among students with varying backlog counts reveals a compelling narrative. Students with no backlogs experience notably lower stress levels, indicating the pivotal role of a backlog-free academic journey in reducing stress. Conversely, as the number of backlogs increases, stress levels surge proportionally, reaching alarming heights for students burdened with multiple backlogs.

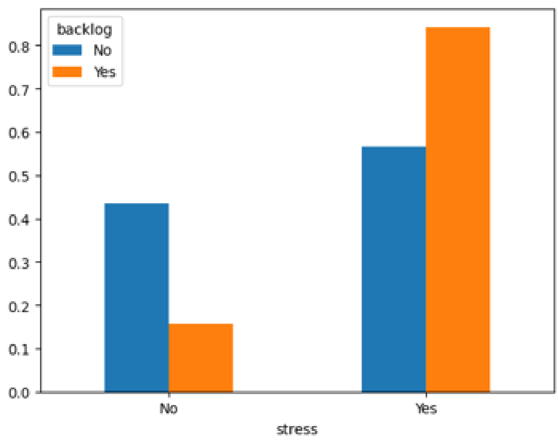


Figure 6.4 Backlogs v/s stress.

Figure 6.4 validates the strong correlation between academic backlogs and stress and highlights the urgency of implementing targeted support systems for students with backlogs. By addressing the stressors associated with academic challenges, educational institutions can create a more supportive environment, fostering mental well-being and academic success among students.

E. Relationship Between Sleeping Hours and Stress

The data visualization analysis uncovers a compelling association between students' average sleep duration and mental health. Notably, students who sleep an average of 4 to 6 hours per night are highly susceptible to mental health issues, exhibiting a significantly higher risk than their peers. Conversely, students who manage to sleep for 6 to 8 hours demonstrate notably better mental health outcomes, suggesting a protective effect of adequate sleep within this range. Intriguingly, students who sleep more than 8 hours per night also display a slight increase in the likelihood of mental health challenges, indicating a potential threshold beyond which excessive sleep might contribute to health issues.

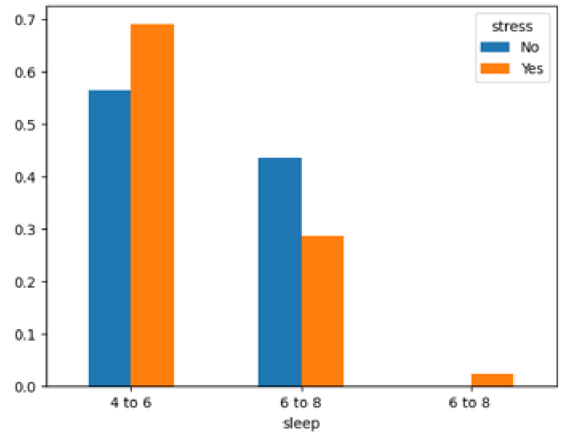


Figure 6.5 Sleeping Hours v/s Stress

Figure 6.5 depicts the findings, emphasizing the critical role of sleep patterns in mental well-being and advocating for targeted sleep education and interventions. Optimal sleep duration among college students and mitigate the risk of mental health problems.

F. Relationship Between Extracurriculars and Stress

The study of the data visualization shows a strong connection between students' involvement in extracurricular activities and their mental health. Students engaged in extracurricular pursuits exhibit notably lower vulnerability to mental health challenges than those not involved in such activities. The visualization vividly illustrates that active participation in clubs, sports, arts, or social groups is a protective factor, substantially reducing the risk of mental health issues. Conversely, students who abstain from extracurricular activities are more susceptible to mental health disorders.

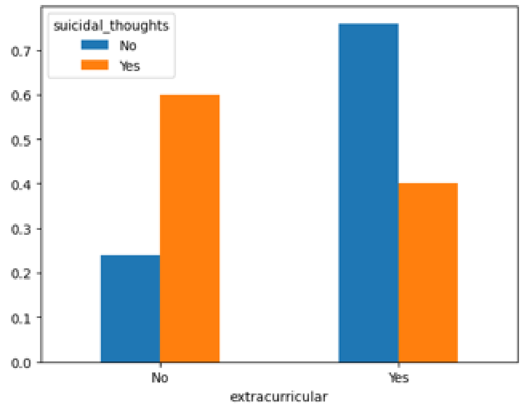


Figure 6.6 Extracurriculars V/S Stress

Figure 6.6 depicts findings that underscore the crucial role of extracurricular engagement in promoting positive mental well-being among college students. Encouraging and facilitating participation in diverse extracurricular activities can effectively bolster mental resilience and foster a supportive campus environment.

VII. Machine Learning outcomes

A. Logistic Regression

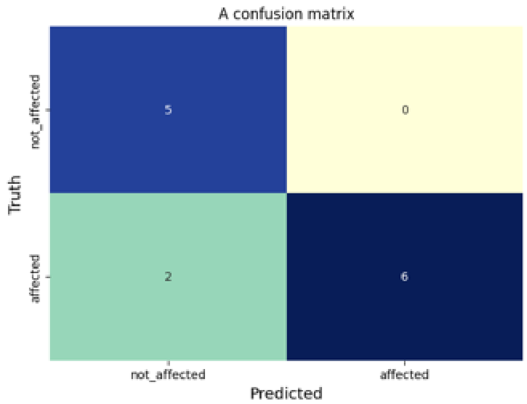


Figure 7.1 Confusion matrix result for Logistic Regression model

```
metrics.accuracy_score(y_test, logisticmodel_prediction)
```

0.8461538461538461

```
metrics.f1_score(y_test, logisticmodel_prediction)
```

0.8571428571428571

Figure 7.2: F1 score and accuracy score for Logistic Regression model

Figure 7.1 and 7.2 depicts the results and the factors that will help to compare models for machine learning. The accuracy, f1 score, and confusion matrix are used to compare the models.

B. Naive Bayes

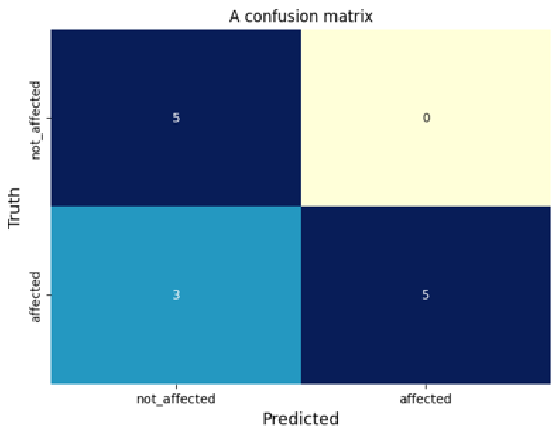


Figure 7.3: Confusion Matrix result for Naïve Bayes model

```
metrics.accuracy_score(y_test, naivemodel_prediction)
```

0.7692307692307693

```
metrics.f1_score(y_test, naivemodel_prediction)
```

0.7692307692307693

Figure 7.4: F1 Score and Accuracy for NB

Figures 7.3 and 7.4 represent the results that will help compare machine learning models. The accuracy, f1 score, and confusion matrix are used to analyze the models.

C. *Random Forest*

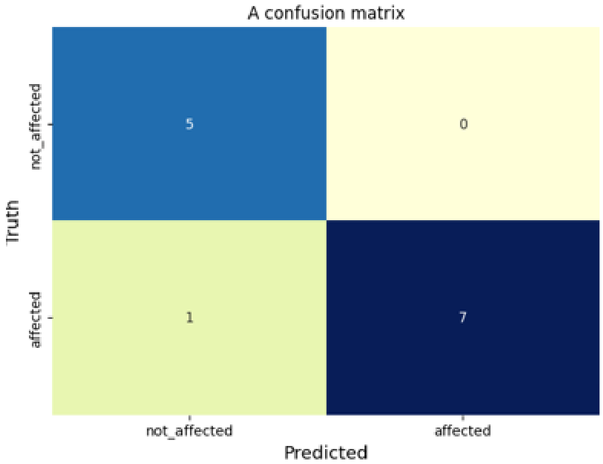


Figure 7.5: Confusion Matrix Result for Random Forest Model

```
metrics.accuracy_score(y_test, randomforestmodel_prediction)
0.9230769230769231

metrics.f1_score(y_test, randomforestmodel_prediction)
0.9333333333333333
```

Figure 7.6: F1 Score and Accuracy Score for Random Forest Model

Figures 7.5 and 7.6 depics the confusion matrix and the accuracy for Random Forest Model. The results in this algortuhm are better than the previous ones.

D. *XG Boost*

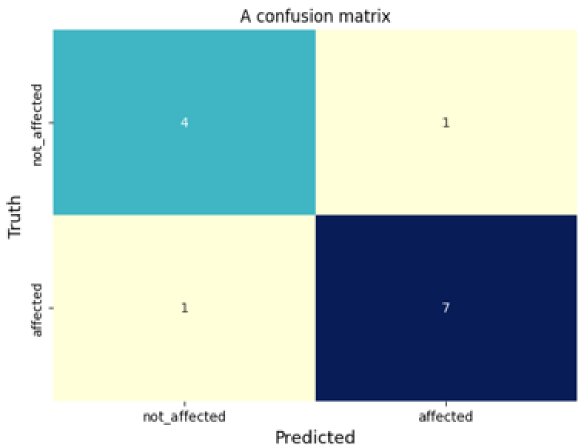


Figure 7.7: Confusion Matrix result for XG Boost

```
metrics.accuracy_score(y_test, xgboostmodel_prediction)
0.8461538461538461

metrics.f1_score(y_test, xgboostmodel_prediction)
0.875
```

Figure 7.8: F1 Score and Accuracy for XG Boost Model

The insights derived from the XGBoost model, displayed in Figures 7.7 and 7.8, significantly enriched our understanding of the intricate factors influencing mental health in the student population. These insights deepened our knowledge of mental health dynamics and served as the foundation for targeted interventions and support mechanisms. The predictions generated by the XGBoost model, coupled with detailed analyses, made substantial contributions to the overall findings of our study, shedding light on the multifaceted nature of mental health challenges college students face.

E. *Support Vector Machine*

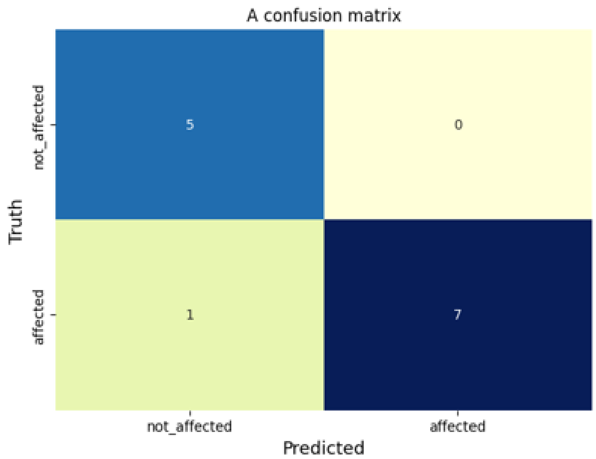


Figure 7.9: Confusion Matrix Result for SVM

```
metrics.accuracy_score(y_test, SVMmodel_prediction)
0.9230769230769231

metrics.f1_score(y_test, SVMmodel_prediction)
0.9333333333333333
```

Figure 7.10: F1 Score and Accuracy for SVM

Figures 7.9 and 7.10 depicts the confusion matrix and the accuracy for the SVM. SVM shows an accuracy of 93% and also a good f1 score.

VIII. CONCLUSION

This research has led to insightful discoveries and significant advancements in mental health analysis among college students. An in-depth exploration into the intricate web of factors influencing mental health outcomes was conducted through the meticulous application of various machine learning algorithms, including Logistic Regression, Naive

Bayes, XGBoost, Support Vector Machine (SVM), and Random Forest. The outcomes of this study are both promising and illuminating. Each model, armed with its unique strengths, provided valuable perspectives on the complex interplay of variables shaping students' mental well-being. Logistic Regression offered a solid foundation, Naive Bayes showcased probabilistic finesse, XGBoost demonstrated its robustness, and SVM provided a nuanced understanding of classification boundaries.

However, the Random Forest algorithm emerged as the beacon of accuracy and reliability amidst this diversity of approaches. With its ability to comprehend complex relationships within the data, Random Forest consistently outperformed other models, delivering the highest accuracy and precision in predicting mental health outcomes among the student cohort. The success of Random Forest underscores the importance of employing sophisticated and adaptable algorithms when dealing with multifaceted issues like mental health. By harnessing the power of Random Forest, unparalleled accuracy was achieved, and subtle patterns and connections within the data were unearthed, enriching the understanding of the underlying dynamics. These findings advocate integrating machine learning methodologies, especially Random Forest, in mental health analysis. The precision and reliability demonstrated by this algorithm pave the way for targeted interventions, personalized support systems, and proactive initiatives within educational institutions. These insights hold immense potential in shaping mental health policies, refining counseling services, and bolstering the overall well-being of college students. As this research journey concludes, it is evident that the exploration into the depths of mental health analysis has illuminated a path forward. By embracing advanced algorithms such as Random Forest, the complexities of mental health can be navigated with precision, empathy, and understanding, fostering a healthier and more supportive environment for college students.

IX. References

[1] Sawangarreerak, S.; Thanathamathee, P. Random Forest with Sampling Techniques for Handling Imbalanced Prediction of University Student Depression. *Information* **2020**, *11*, 519. https://doi.org/10.3390/info11110519

[2] Factors Influencing College Students’ Mental Health Promotion: The Mediating Effect of Online Mental Health Information Seeking Int. J. Environ. Res. Public Health 2020, 17(13), 4783; https://doi.org/10.3390/ijerph17134783

[3] Tran A, Tran L, Geghre N, Darmon D, Rampal M, Brandone D, Gozzo JM, Haas H, Rebouillat-Savy K, Caci H, Avillach P. Health assessment of French university students and risk factors associated with mental health disorders. PLoS One. 2017 Nov 27;12(11):e0188187. doi: 10.1371/journal.pone.0188187. PMID: 29176864; PMCID: PMC5703533.

[4] Lei X, Liu C, Jiang H. Mental health of college students and associated factors in Hubei of China. PLoS One. 2021 Jul 2;16(7):e0254183. doi: 10.1371/journal.pone.0254183. PMID: 34214107; PMCID: PMC8253420.

[5] Chu, Yichen, and Xiaojian Yin. "Data analysis of college students’ mental health based on clustering analysis algorithm." Complexity 2021 (2021): 1-10.

[6] Mikolajczyk RT, Maxwell AE, El Ansari W, Naydenova V, Stock C, Ilieva S, Dudziak U, Nagyova I. Prevalence of depressive symptoms in university students from Germany, Denmark, Poland and Bulgaria. Soc Psychiatry Psychiatr Epidemiol. 2008 Feb;43(2):105-12. doi: 10.1007/s00127-007-0282-0. Epub 2007 Nov 23. PMID: 18038173.

[7] Hua Wang Sharifah Muzlia Syed Mustafa Siti Zuraida Maaruf 2023 2023/06/30 Analysis of the Influence of College Students’ Mental Health Based on Statistical Analysis, Data Mining, and Predictive Analysis Proceedings of the 2023 4th International Conference on Education, Knowledge and Information Management (ICEKIM 2023) Atlantis Press 1250 1257 2589-4900 https://doi.org/10.2991/978-94-6463-172-2_1310.2991/978-94-6463-172-2_132 Wang2023

[8] Mann JJ, Ellis SP, Waternaux CM, Liu X, Oquendo MA, Malone KM, Brodsky BS, Haas GL, Currier D. Classification trees distinguish suicide attempters in major psychiatric disorders: a model of clinical decision making. J Clin Psychiatry. 2008 Jan;69(1):23-31. doi: 10.4088/jcp.v69n0104. PMID: 18312034; PMCID: PMC3773877.

[9] Minutolo, Aniello Chung, Jetli Teo, Jason 2022 2022/01/05 Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges 9970363 2022 1687-9724 https://doi.org/10.1155/2022/9970363 10.1155/2022/9970363 Applied Computational Intelligence and Soft Computing Hindawi

[10] Predicting Mental Health Illness using Machine Learning Algorithms Journal of Physics: Conference Series Vaishnavi, Konda Nikhitha Kamath, U Ashwath Rao, B Subba Reddy, N V 2022 2022/01/01 IOP Publishing 01202112161 1742-6596 1742-658810.1088/1742-6596/2161/1/012021 https://dx.doi.org/10.1088/1742-6596/2161/1/012021