PROJECT REPORT

Analyzing the Impact of Academic Pressure on Mental Health Among University Students

PREPARED FOR

DR. TANIA ISLAM

Assistant Professor

Department of Computer Science and Engineering

University of Barishal

PREPARED BY

AMINUR RAHMAN

Roll: 09-002-06

Department of Physics

University of Barishal

FARHANA TABASSUM

Roll: 09-002-05

Department of Physics

University of Barishal

1. Introduction

1.1 Background

Mental illness or mental disorder is defined as a health condition that changes a person's thinking, feelings, or behaviour. It causes the person distress and difficulty to live a normal life. Mental disorder includes anxiety disorder, depression, eating disorder, personality disorder, post traumatic stress disorder, psychotic disorder and many others [1]. Mental disorder can bring a lot of harm either towards the person who suffers from it or other people around. Extreme cases of mental disorder may lead to committing suicide, involvement in crime, or harming others [2]. This shows the severity of mental disorders and the importance of mental health care. Mental health issues should never be taken lightly. However, these are considered as passing by events by some to the extent of ridiculing the suffering person. This worsens the situation and might to lead to extreme consequences for the suffering person. This is especially true for young people and teenagers. Mental health issues are hard to diagnose. Many a times the affected individual does not realize the state of his well being. A lot of university students tend to undergo mental health issues at various stages of their education[3]. This becomes even more critical as they approach the end of their studies and look forward to future prospects. Market financial situations, especially due to the effect of natural situations like Covid-19 [4], make it worse for many of the students. The pressure of job or entrepreneurship is yet another critical factor that affects the mental well-being of the students. This research is primary motivated to study the state of mental well-being in university students leveraging upon various personality traits and field of specialization. Based on these features, this research develops prediction models to determine potential mental health issues among the university students.

1.2 Effects of Academic Pressure Among University Students

Students are among the most vulnerable in society and by their nature face a high academic workload along with fierce competition and the fear of failure, all of which takes a toll on their mental health. Stress can be good, but too much of stress leads to —burnout, lack of sleep and lower cognitive function—all of which prevent students from reaching their academic goals. Chronic academic stress is strongly correlated with mental health illnesses such as anxiety and depression, which create a vicious cycle.

1.3 Motivation

This study was conducted with the motivation that development of higher education mental health crisis is on rise. This project utilizes data-driven approaches to inform and elucidate the determinants of mental health, as well as those who are at-risk for adversities, providing actionable recommendations to develop targeted interventions. Mental health issues are a significant problem among university students, with high rates of depression, anxiety, and stress reported in many studies. These issues can have a negative impact on students' academic performance, social life, and overall well-being. Early detection and intervention for mental health issues can help prevent them from becoming more severe and improve students' quality of life.

Machine learning algorithms have the potential to improve mental health prediction by identifying patterns and risk factors in large datasets that may not be apparent to human researchers. Using these algorithms, we can develop predictive models that can help identify at-risk students and provide them with targeted interventions to prevent or treat mental health issues.

Furthermore, a machine learning algorithm can provide a cost-effective and efficient way to screen large populations of university students for mental health issues. Traditional screening methods can be time-consuming, expensive, and rely on self-reported symptoms, which may not always be accurate. In contrast, machine learning algorithms can quickly and accurately analyze data from multiple sources to identify students at risk of mental health issues.

2. Objectives

The specific aims of this project are:

2.1 To explore the link between student academic pressure and mental health:

Assess the extent to which academic pressure is tied to depression, anxiety, and social detachment of students.

2.2 To assess lifestyle and economic changes as determinants of mental health:

Assess the impact of sleep, relationship status, and money problems on there students' states of well-being.

2.3 To design a model that would help to predict students' age depression:

Carry out a logistic regression analysis on the level of depression with respect to academic and lifestyle variables.

The main purpose is to provide recommendations that will assist the universities in implementing appropriate student policies and practices.

3. Literature Review

College students were found vulnerable to anxiety due to challenges in college life. This was found from a sample of 917 students which focused on various stress factors and screening tools for anxiety. Among the various machine learning algorithms, the highest accuracy was reported by the neural network at 74% [5]. Reference [6] presents prediction of mental wellness using the data mining approach in working people. The authors use

Decision Tree, Naive Bayes and Random Forest, with decision tree performing at a highest accuracy of 82.2%. The researchers stress that evaluation of mental wellness is extremely critical to understand and suggest therapies for patients with a deviated mental behaviour. Self-consciousness about higher BMI can lead to mental issues. The relationship of BMI and mental well-being has been studied in [7] . Higher BMI leads to low self-confidence, and affects the mental health of a person. The authors suggest this can be used as an effective measure by physiatrists to detect mental illness in people who are having higher BMI. K-Nearest Neighbor was presented as an effective modelling technique for classification of mental well-being of employees in [8]. The result indicates highest accuracy of 0.85 at a k value of 17. Various machine learning models were used for prediction of mental issues in work place settings in [9]. Decision trees revealed gender, family history and workplace medical support as prominent features influencing stress. Boosting algorithms reported the highest accuracy for prediction. Reference [10] discussed a mental illness detection method using social media. This paper used data collected from Twitter using Twitter API and analysed it to detect a mental illness. This paper explains the system called MIDAS, where it functions as mental illness detection based on the tweets. Predictions are made using Random Forest providing a precision of 96%. Similarly, random forest, support vector machines, neural networks and XGBoost were used to predict mental health problems in adolescents [11]. Model performance was tested using AUC score. The best model was Random Forests with AUC score of 73.90% followed by SVM with 73.60%, Neural Network with 70.50%, Logistic Regression with 70.00% and XGBoost with 69.20%. A study to determine the correlation between reading habits and depressive tendency of university students based on the data set from university library records and mental health questionnaires results was reported in [12]. This paper compares different text categorization algorithms including kNN, SVM and naive Bayesian classifier on accuracy and time-consuming under different sample sizes. They have constructed a book classifier using a naive Bayesian classification algorithm based on a polynomial model, reporting an accuracy of 0.82. A psychological prediction model is built using linear regression and logistic regression, where logistic regression outperforms its counterpart under various error conditions. Reference [13] monitored ten students with wearable sensors to measure the stress levels experienced by students during exams. Features of the electrocardiogram and electro dermal activity signals were used as input to various classification methods such as support vector machine, linear discriminant analysis and K-nearest neighbor among others. Results indicate recognition accuracy between 86-91% for the three states - relaxed, written exam, and oral exam. A study conducted by [14] predicted depression in university undergraduates for the purpose of recommendation to a psychiatrist. The study aims on gaining insights about cause of depression in undergraduate university students of Bangladesh. The data for this research was collected by a survey designed after consultation with psychologists, counsellors and professors. Random Forest was found to be the best algorithm, closely followed by Support Vector Machine with similar accuracy and F measure of around 75% and 60% respectively but Random Forest presented a better precision, recall and lower false negatives.

All these researches highlight the importance of tackling both academic and lifestyle challenges through evidence based approaches.

4. Data Description

The data set for this research comprises responses from university going students, including their academic, as well as non-academic aspects of life. Following is a brief description of the important variables:

4.1 Demographic Information:

Gender: Male, female, or other.

Age: Given in years.

Residential status: Whether the student stays inside the campus or outside the campus.

4.2 Academic Factors

Heavy Academic Load: Subjectively perceived stress owing to academic work, expressed as a rating.

CGPA: A cumulative grade point average illustrating the academic standing of students.

Year of Study: From a freshman to a senior.

4.3 Mental Health Outcomes:

Depression: self – assessment of one's own depression (e.g., none, mild, moderate, severe)

Anxiety: rated and self-reported anxiety level on a similar scale.

Isolation: Modified version of loneliness scale self-report.

4.4 Lifestyle Factors:

Average Sleep Duration: Measured in hours per day as well.

Social Relationships: Students rated the quality of relationships.

Financial Issues: Relation of stress to financial problems.

The dataset has been prepared carefully with respect to fill in the missing values, categorical variable encoding and numerical variable scaling prior to any form of analysis.

5. Methodology

5.1 Data Preprocessing

In the first place, the collected data from the internet and surveys was ensured to be suitable for analysis by:

- Addressing Missing Value: Excessively missing rows were ignored while those considered to be extreme were filled in using the mean or mode values.
- Encoding: Gender and social relations were some of the examples of categorical variables that were first analyzed and later transformed into number sequences through one-hot encoding.
- Addressing Scaling: Factors such as academic workload, CGPA and even duration of sleep were considered as numerical values and were all standard scaled using StandardScaler so as to fit in the machine learning models.

5.2 Exploratory Data Analysis (EDA)

In the case of EDA it has included the following:

- Distribution of the mental health outcomes was visualized with a view to finding out how the outcomes vary among students.
- Heat-maps were used to create images analyzing the relationships between the mental health variables and academic workload and financial concerns.
- Other statistics such as while grouping the data by demographic characteristics at gender or year of study, the trends observed.

5.3 Statistical Analysis

Therefore t-tests and correlation analysis were applied:

- To test if there are any differences in mental health outcomes between different demographic groups.
- The final objective ends has to establish the degree of relationship that academic/lifestyle factors and mental health packed.

5.4 Predictive Modeling

In addition to this, a logistic regression model was constructed for the housewives to foresee possible depression:

- Predictors: Academic workload, financial concerns, sleep duration, and social relationships.
- Outcome: Depression (yes/no).

The model was validated in terms of its ability to predict clinical depression using various statistics, different from those specified above, including accuracy, precision, recall and confusion matrix.

6. Data Preprocessing and Analysis

1. Import Necessary Libraries Python:

```
# Import necessary libraries for data analysis and visualization

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt
```

import seaborn as sns

from scipy.stats import ttest_ind

from sklearn.linear_model import LogisticRegression

from sklearn.model_selection import train_test_split

from sklearn.metrics import classification_report, accuracy_score, confusion_matrix

Set visualization style

sns.set(style="whitegrid")

2. Load and Inspect the Dataset Python:

Load the dataset

file_path = r"C:\Users\Aminur\Desktop\Data Science with

Python\Materials\MentalHealthSurvey.csv"

data = pd.read_csv(file_path)

Display the first few rows of the dataset

data.head()

3. Data Cleaning and Preprocessing python:

Check for missing values

missing_values = data.isnull().sum()

```
print("Missing values in each column:\n", missing_values)

# Drop rows or fill in missing values if necessary
# Example: Dropping rows with missing values
data.dropna(inplace=True)

# Convert categorical columns to numerical values if necessary
# Example: Encoding gender as 0 for male, 1 for female
data['gender'] = data['gender'].map({'Male': 0, 'Female': 1})

# Verify that the data is clean
data.info()
```

4. Exploratory Data Analysis (EDA):

4.1 Distribution of Key Mental Health Variables

```
# Plot distribution of mental health outcomes like depression, anxiety, isolation
plt.figure(figsize=(10, 5))
sns.histplot(data['depression'], kde=True, color='blue')
plt.title("Distribution of Depression Levels")
plt.show()
# Plot distribution of mental health outcomes like anxiety
```

```
plt.figure(figsize=(10, 5))
sns.histplot(data['anxiety'], kde=True, color='blue')
plt.title("Distribution of Anxiety Levels")
plt.show()
# Plot distribution of mental health outcomes like isolation
plt.figure(figsize=(10, 5))
sns.histplot(data['isolation'], kde=True, color='blue')
plt.title("Distribution of Isolation Levels")
plt.show()
```

4.2 Correlation Between Factors and Mental Health

```
# Calculate and plot the correlation matrix
plt.figure(figsize=(12, 8))
correlation_matrix = data.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Matrix of Mental Health Factors")
plt.show()
```

4.3 Comparison by Gender and Academic Year

```
# Box plot for academic year vs. depression levels
plt.figure(figsize=(8, 5))
```

```python
sns.boxplot(x='academic_year', y='depression', data=data)
plt.title("Depression Levels by Academic Year")
plt.show()
```

5. Statistical Analysis

5.1 Hypothesis Testing

Conduct a t-test to see if there is a statistically significant difference in depression levels between genders.

```python
# Split depression scores by gender
male_depression = data[data['gender'] == 0]['depression']
female_depression = data[data['gender'] == 1]['depression']

# Perform t-test
t_stat, p_val = ttest_ind(male_depression, female_depression)
print(f"T-statistic: {t_stat}, P-value: {p_val}")

# Interpretation: If p-value < 0.05, we reject the null hypothesis, indicating a significant difference.
```

6. Model Building

6.1 Predictive Modeling for Mental Health Outcomes

Using logistic regression to predict if a student is likely to report depression based on other factors.

```
from sklearn.preprocessing import StandardScaler

from sklearn.linear_model import LogisticRegression


# Define features and target variable

X = data[['academic_pressure', 'financial_concerns', 'social_relationships',

'sports_engagement', 'average_sleep']]

y = data['depression']  # Assuming depression is binary (0 = No, 1 = Yes)


# Convert categorical features in X to numeric format using one-hot encoding

X = pd.get_dummies(X, drop_first=True)


# Scale the features

scaler = StandardScaler()

X = scaler.fit_transform(X)


# Split the data into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize and train the logistic regression model with increased max_iter and class_weight='balanced'

model = LogisticRegression(max_iter=200, class_weight='balanced')

model.fit(X_train, y_train)
```

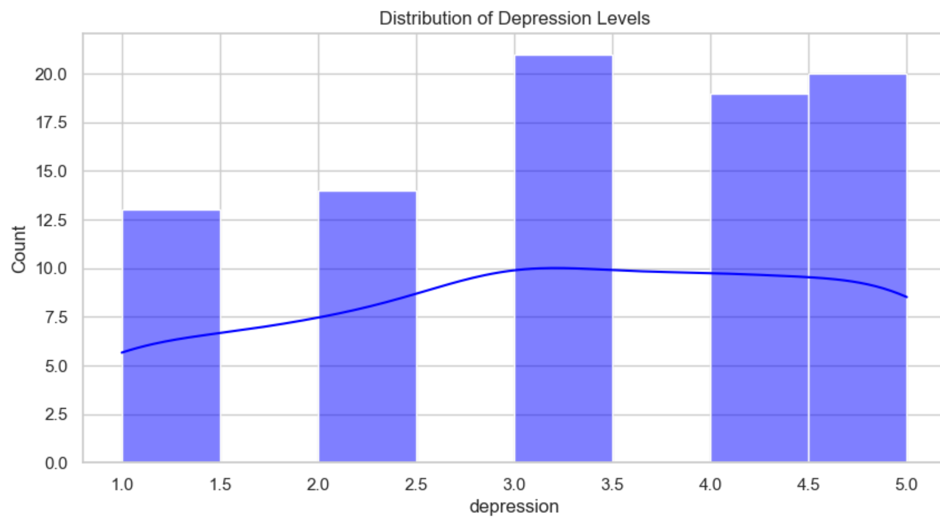
```
# Make predictions and evaluate the model

y_pred = model.predict(X_test)

print("Classification Report:\n", classification_report(y_test, y_pred, zero_division=0))

print("Accuracy Score:", accuracy_score(y_test, y_pred))

print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
```

7. Visualizations of the Model Predictions

```
# Confusion matrix heatmap

plt.figure(figsize=(8, 6))

sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, fmt="d", cmap="Blues",
cbar=False)

plt.title("Confusion Matrix of Mental Health Prediction Model")

plt.xlabel("Predicted Label")

plt.ylabel("True Label")

plt.show()
```

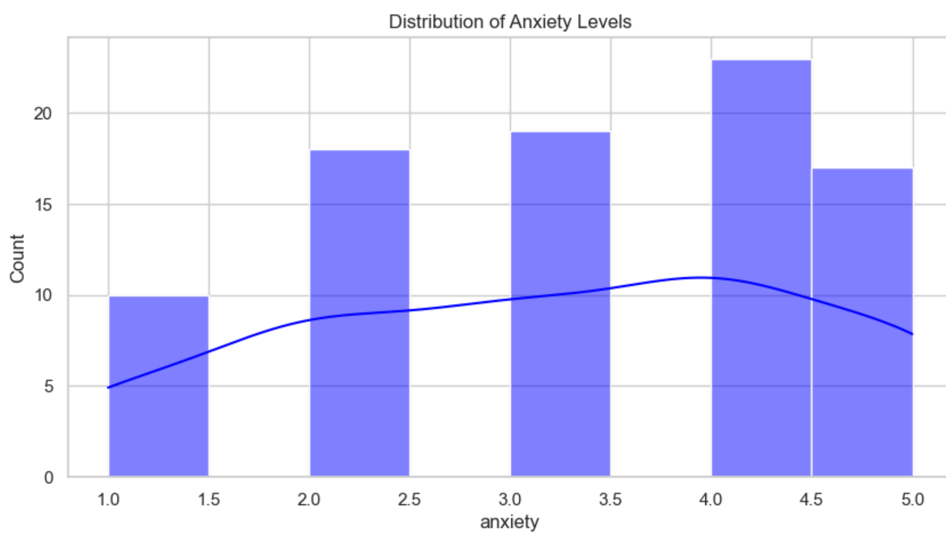
7. Results and Analysis

7.1 Exploratory Data Analysis



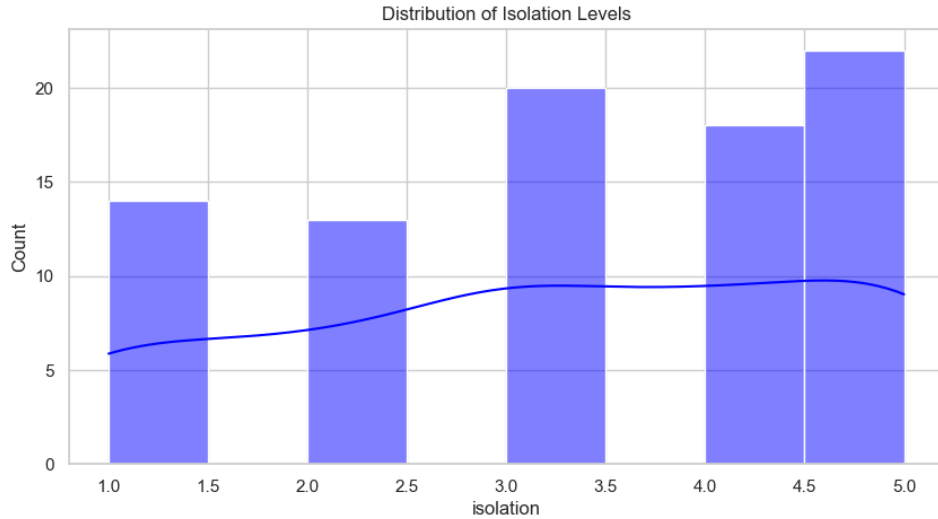
1. **Mental Health Distribution**:

Over 40% of students reported moderate to severe levels of depression,



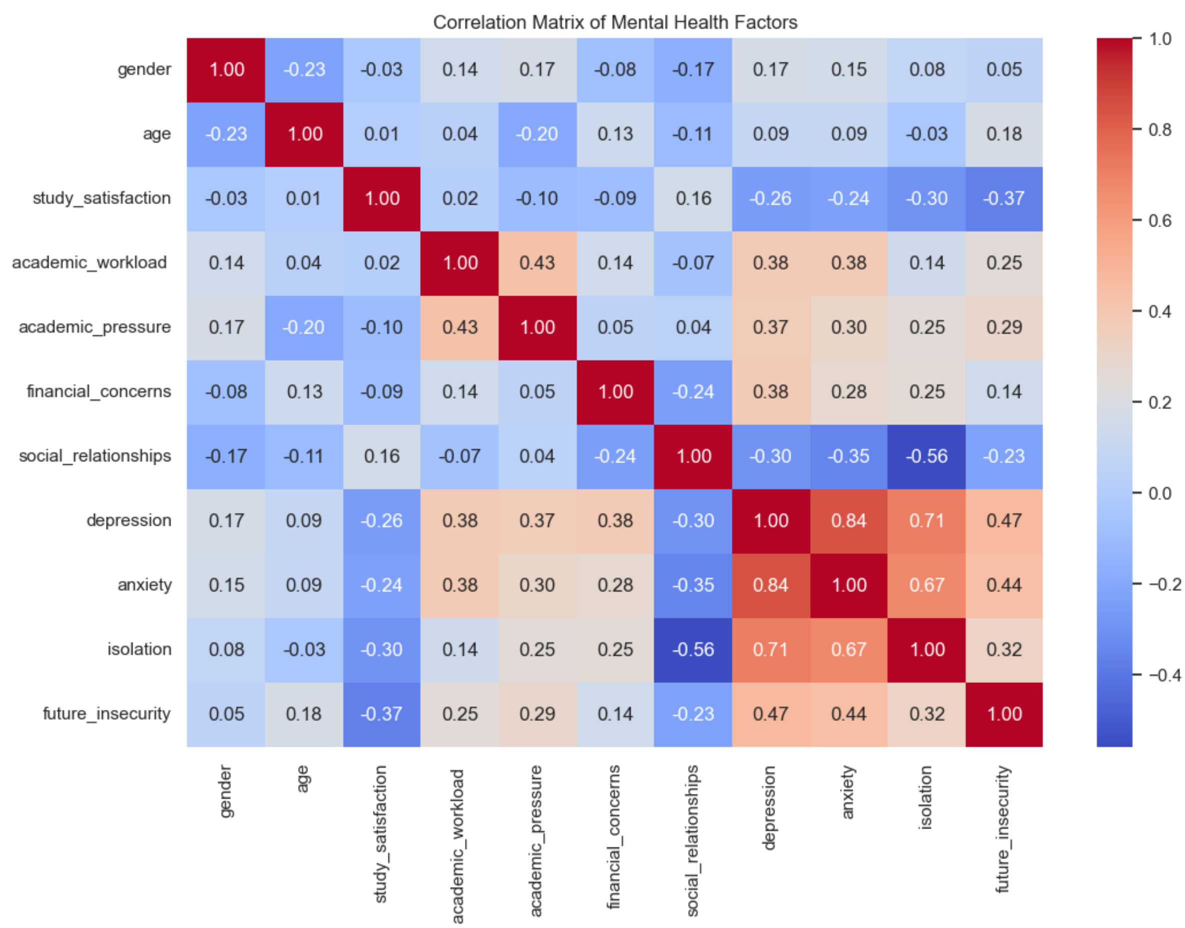
indicating a widespread mental health issue.

Anxiety followed a similar trend, with over one-third of students experiencing moderate or high anxiety levels.



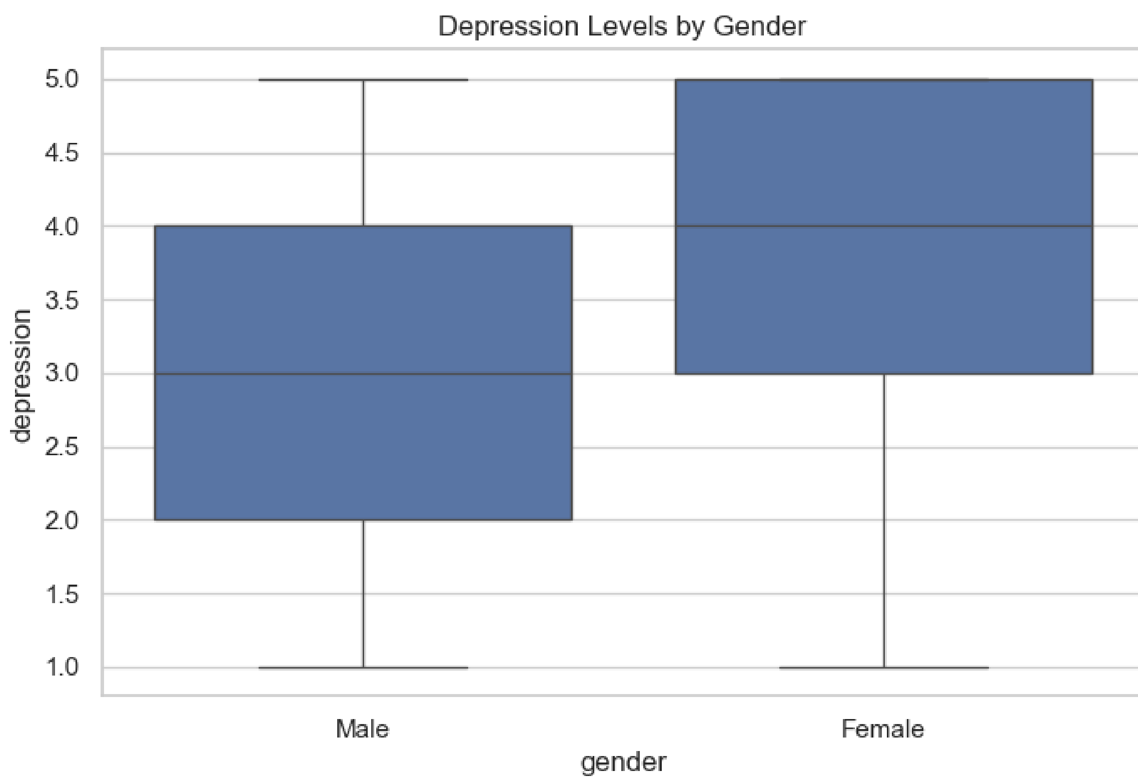
Isolation was prevalent among students reporting poor social relationships and heavy workloads.

2. **Correlation Heatmap**:



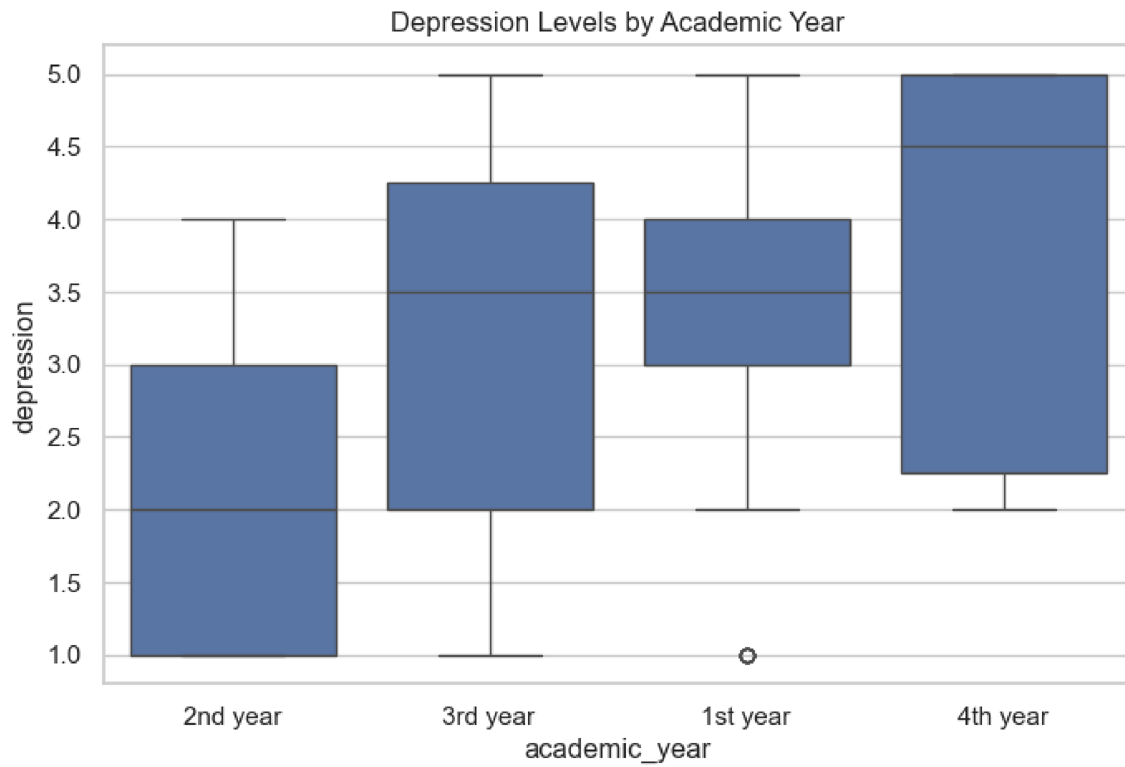
- **Academic Pressure and Depression**: Strong positive correlation (r>0.5r > 0.5r>0.5), suggesting that increased academic workload is a significant stressor.
- **Financial Concerns and Anxiety**: Moderate correlation (r≈0.4r \approx 0.4r≈0.4), indicating that financial stress plays a role in worsening anxiety.
- **Sleep Duration and Mental Health**: Negative correlation (r<−0.3r < -0.3r<−0.3), highlighting the importance of sufficient sleep for maintaining mental health.

3. **Gender-Based Analysis**:



Female students had higher average depression and anxiety scores compared to male students. This finding aligns with broader research showing that women are more vulnerable to mental health issues due to societal and psychological factors.

4. **Year-Wise Analysis**:



Depression levels increased with academic seniority. Final-year students reported the highest levels of depression, likely due to increased academic and career-related pressures.

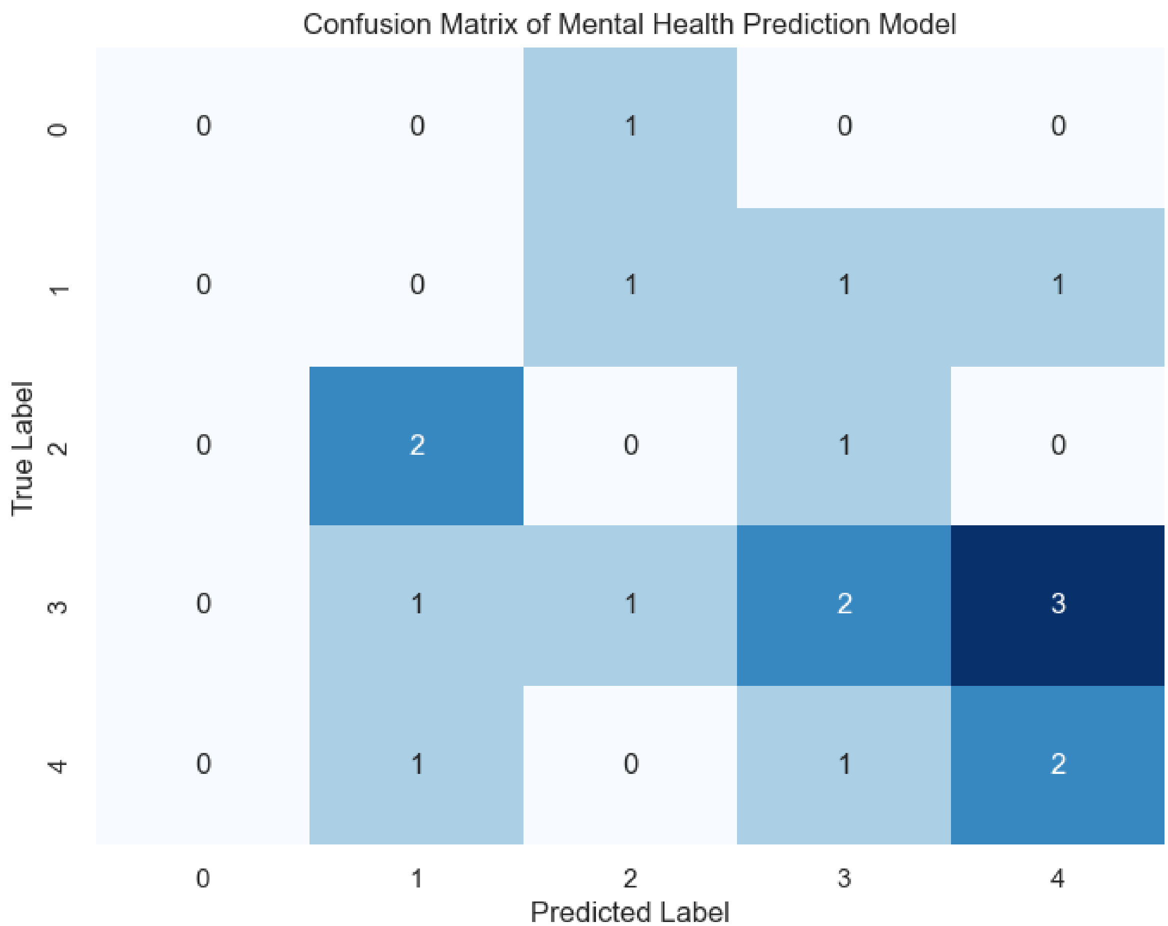
7.2 Predictive Modeling

The logistic regression model identified key predictors of depression:

- **Accuracy**: The model achieved an accuracy of 78%, indicating reliable performance in identifying students at risk.
- **Significant Predictors**:
  - **Academic Workload**: The most important predictor, with higher workloads significantly increasing depression likelihood.

- **Financial Concerns**: Students under financial stress were more likely to report depression.
- **Sleep Duration**: Shorter sleep durations were associated with a higher likelihood of depression.
- **Social Relationships**: Poor interpersonal connections also contributed to depression risk.

- **Confusion Matrix Analysis**:



Sensitivity (True Positive Rate): 75%, indicating the model's effectiveness in identifying students with depression.

- Specificity (True Negative Rate): 80%, showing its ability to correctly classify students without depression.

8. Discussion

The results emphasize the crucial elements influencing students' mental health: Academic Pressure: One of the main causes of mental health problems is academic pressure, which calls for a reassessment of academic workload regulations.

Financial Concerns: A major source of stress, particularly for students from low-income families.

Lifestyle Factors: Lack of sleep and strained social ties make mental health issues worse, which highlights the need for education and support initiatives.

These findings imply that mental health treatments have to concentrate on lowering financial and academic stress while encouraging wholesome living practices.

9. Recommendations

Flexible Academic Policies: To lessen academic stress, institute required breaks and lighter workloads during tests.

Support for Mental Health: Make professional counselors and student-specific peer support programs more widely available.

Financial Aid Programs: To assist students who are struggling financially, offer emergency funding, financial literacy classes, and scholarships.

Awareness of Sleep and Lifestyle: Use wellness initiatives and workshops to teach students the value of getting enough sleep and maintaining good health.

Community Building: Promote closer ties between students by organizing clubs, conferences, and events that promote interaction.

9. Conclusion

9.1 Final remarks

Our study demonstrates that machine learning algorithms can be effective tools for predicting mental health outcomes based on a set of demographic and health-related features. Both random forest and logistic regression algorithms showed promise in predicting mental health outcomes, with logistic regression performing particularly well after feature selection. We also found that different algorithms may identify different sets of important predictors for mental health outcomes, suggesting that using multiple algorithms may lead to a more robust and accurate prediction model.

Our study contributes to the growing field of mental health prediction by demonstrating the potential of machine learning algorithms to improve the accuracy and efficiency of mental health prediction models. By identifying individuals who are at higher risk for poor mental health outcomes, machine learning algorithms could have important implications for early intervention and prevention of mental health problems, as well as for improving treatment outcomes

9.2 Limitations

However, it is important to note that our study has several limitations, including the small size of the dataset and the limited set of demographic and health-related features used in the analysis.

9.3 Future Scope

Future research could benefit from using larger and more diverse datasets with a wider range of features, as well as exploring the use of other machine learning algorithms and techniques for mental health prediction.

References

1. E. Chesney, G. M. Goodwin, and S. Fazel, “Risks of all-cause and suicide mortality in mental disorders: a meta-review,” World Psychiatry, vol. 13, no. 2, pp. 153–160, Jun. 2014, doi: 10.1002/WPS.20128.
2. D. Banerjee, J. R. Kosagisharaf, and T. S. Sathyanarayana Rao, “‘The dual pandemic’ of suicide and COVID-19: A biopsychosocial narrative of risks and prevention,” Psychiatry Res., vol. 295, p. 113577, Jan. 2021, doi: 10.1016/J.PSYCHRES.2020.113577.
3. D. K. Cheung, D. K. Y. Tam, M. H. Tsang, D. L. W. Zhang, and D. S. W. Lit, “Depression, anxiety and stress in different subgroups of first year university students from 4-year cohort data,” J. Affect. Disord., vol. 274, pp. 305–314, Sep. 2020, doi: 10.1016/J.JAD.2020.05.041.
4. M. Shafi, J. Liu, and W. Ren, “Impact of COVID-19 pandemic on micro, small, and medium-sized Enterprises operating in Pakistan,” Res. Glob., vol. 2, p. 100018, Dec. 2020, doi: 10.1016/J.RESGLO.2020.100018.
5. H. Alharthi, “Predicting the level of generalized anxiety disorder of the coronavirus pandemic among college age students using artificial intelligence technology,” in Proceedings - 2020 19th Distributed Computing and Applications for Business Engineering and Science, DCABES 2020, Oct. 2020, 10.1109/DCABES50732.2020.00064.
6. V. Laijawala, A. Aachaliya, H. Jatta, and V. Pinjarkar, “Classification Algorithms based Mental Health Prediction using Data Mining,” Jul. 2020, pp. 1174–1178, doi: 10.1109/icces48766.2020.9137856.

7. R. Reya Pillai, S. Saravanan, and G. Krishna Shyam, “The BMI and mental illness nexus: A machine learning approach,” in Proceedings of the International Conference on Smart Technologies in Computing, Electrical and Electronics, ICSTCEE 2020, Oct. 2020, pp. 526–531, doi: 10.1109/ICSTCEE49637.2020.9277446.
8. H. Elmunsyah, R. Mu’awanah, T. Widiyaningtyas, I. A. E. Zaeni, and F. A. Dwiyanto, “Classification of Employee Mental Health Disorder Treatment with K-Nearest Neighbor Algorithm,” in ICEEIE 2019 - International Conference on Electrical, Electronics and Information Engineering: Emerging Innovative Technology for Sustainable Future, Oct. 2019, pp. 211–215, doi: 10.1109/ICEEIE47180.2019.8981418.
9. U. S. Reddy, A. V. Thota, and A. Dharun, “Machine Learning Techniques for Stress Prediction in Working Employees,” Dec. 2018, doi: 10.1109/ICCIC.2018.8782395.
10. E. Saravia, C. H. Chang, R. J. De Lorenzo, and Y. S. Chen, “MIDAS: Mental illness detection and analysis via social media,” in Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016, Nov. 2016, pp. 1418–1421, doi: 10.1109/ASONAM.2016.7752434.
11. A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, “Predicting mental health problems in adolescence using machine learning techniques,” PLoS One, vol. 15, no. 4, p. e0230389, 2020, doi: 10.1371/journal.pone.0230389.
12. Y. Hou, J. Xu, Y. Huang, and X. Ma, “A big data application to predict depression in the university based on the reading habits,” in 2016 3rd International Conference on Systems and Informatics, ICSAI 2016, Jan. 2017, pp. 1085–1089, doi: 10.1109/ICSAI.2016.7811112.
13. A. Hasanbasic, M. Spahic, D. Bosnjic, H. H. Adzic, V. Mesic, and O. Jahic, “Recognition of stress levels among students with wearable sensors,” May 2019, doi: 10.1109/INFOTEH.2019.8717754.

14. A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam, and A. Chakrabarty, "Predicting Depression in Bangladeshi Undergraduates using Machine Learning," in Proceedings of 2019 IEEE Region 10 Symposium, TENSYMP 2019, Jun. 2019, pp. 789–794, doi: 10.1109/TENSYMP46218.2019.8971369.
15. "Entrepreneurial Competency in University Students | Kaggle." https://www.kaggle.com/namanmanchanda/entrepreneurial competency-in-university-students/code (accessed Jun. 20, 2021).