STATISTICS ANALYSIS DATA OF STUDENT DEPRESSION DETERMINANTS

Lizaveta Piskunova
EPAM School of Digital Competencies, Savanorių pr. 28, Vilnius

Abstract
Depression is one of the most common psychological problems faced by human society. Since students are the group with less social experience and also the group bearing lots of responsibilities, depression incidence is much higher than in the general population. To help understand this further, this article aims to discuss statistics about student depression and identify key factors which can influence depression development.
Keywords: student depression, academic pressure, correlation analysis, hypothesis testing, university mental health

1. Introduction

Since there are lots of factors in this dataset, the main aim of this project is to identify and to understand which of them affect a student's mental health, especially depression, and based on this formulate an algorithm, and as a conclusion, to provide recommendations on how to prevent development of student's depression.

2. Dataset Description and Analysis Methods

The dataset used in this research was chosen in the Kaggle educational resource. The dataset includes responses from 502 students and 11 variables: Age, Gender, Academic Pressure, Study Hours, Study Satisfaction, Financial Stress, Sleep Duration, Dietary Habits, Family History of Mental Illness, Suicidal Thoughts, and Depression status.

This research was conducted in Google Colab resource using programming language Python and libraries for data manipulation and visualization such as pandas, numpy, scipy, matplotlib, seaborn, and scikit-learn.

The implementation of the algorithm step by step included three main stages:

1. Exploratory Data Analysis (EDA):
    - Initial screening for missing values and duplicates
    - Descriptive statistics for all variables
    - Distribution visualizations via histograms and box plots
    - Categorical feature exploration using bar plots
2. Statistical Analysis:
    - One-sample and two-sample t-tests to compare means
    - Z-tests for population mean and proportions
    - ANOVA for group comparisons
    - Chi-square tests for independence of categorical variables
3. Multivariate Analysis:
    - Pearson correlation matrix (with label encoding for categorical variables)
    - PCA for dimensionality reduction and variance explanation
    - K-means clustering to detect group patterns and risk profiles

3. Analysis Results

3.1 Descriptive Statistics

This dataset has a slightly higher representation of males than females, since males are 53.2% and females are 46.8% and their difference is about 6.4%, but mostly it's balanced.

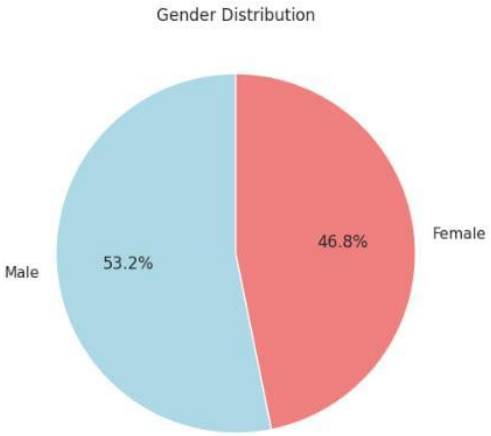

Figure 1 - Gender Distribution

So, it can be said that the sample is demographically balanced, with nearly equal gender distribution and ages ranging from 18 to 34.

Approximately 50% (252 students reported it as "yes" and 250 as "no") of the participants reported experiencing symptoms of depression. Also, about 50% of students reported they suffer from suicidal thoughts. Sleep durations were predominantly 7-8 hours, and most students rated their dietary habits as moderate or unhealthy.

Most common dietary habits, which were reported by students, are moderate (count is about 170) and unhealthy (count is about 170), so less frequent is healthy (count is about 155).

There are lots of students who do not have a family history of mental illness (count is about 255) and the number of students who have a family history of mental illness is about 235.

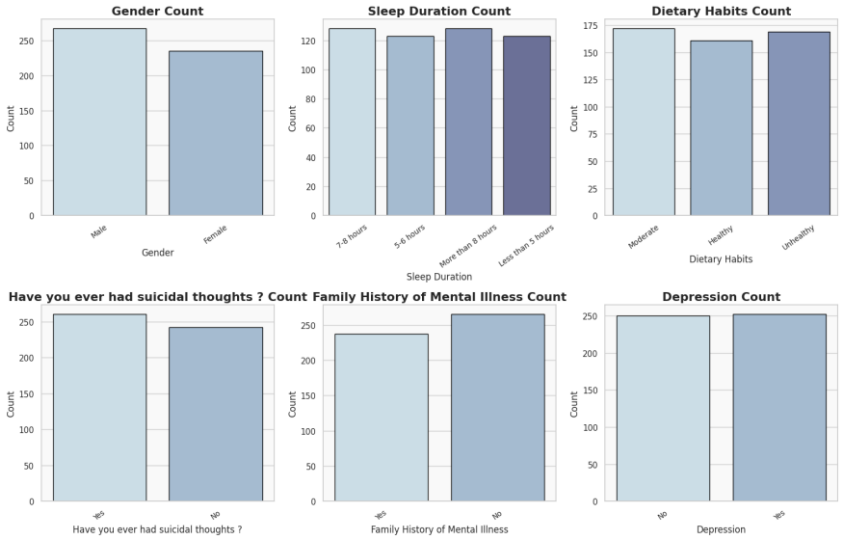

Figure 2 - Categorical Features Distributions

Academic Pressure ranges from 1 (low) to 5 (high) and the most frequent value is 3, so it means mostly students have moderate level of academic pressure. As it is for Financial Stress, where it is a balanced distribution with peaks around stress levels 1 and 2.

As for Study Satisfaction there are peaks around satisfaction levels 3 and 4, which means many students are happy with their studies. And as for Study Hours, there are more students, who study 8 and 10 hours daily, but also there are students who study a little (0-3 hours) or lots of hours (10-12).

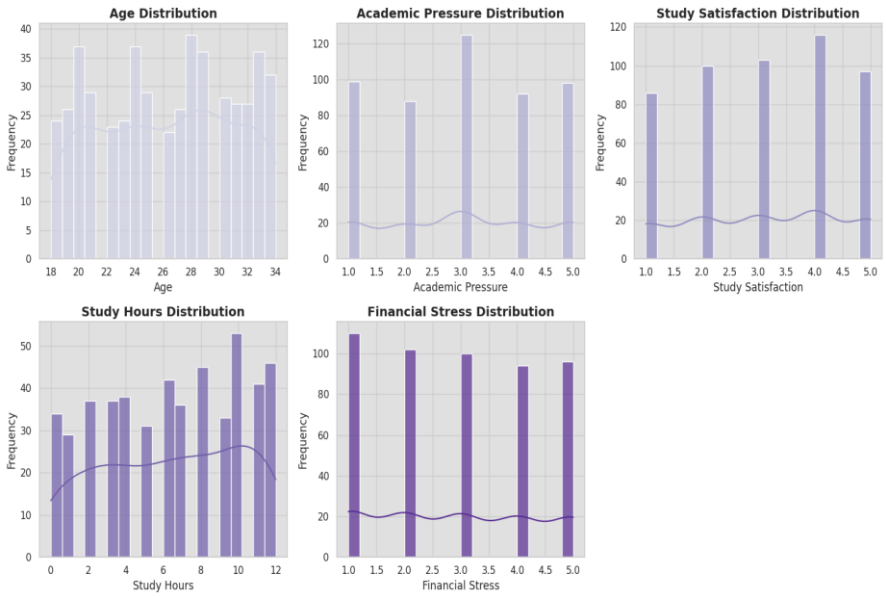

Figure 3 - Numerical features distributions

3.2 Correlation Analysis

Correlation analysis showed which factors have the strongest influence on depression development.

It provides the following results:

- Academic Pressure showed the strongest positive correlation with Depression (r = 0.48). So, higher Academic Pressure is associated with a higher possibility to report depression.
- Suicidal Thoughts were highly associated with Depression (r = 0.47). It means students who report having suicidal thoughts are likely to have depression.
- Dietary Habits and Depression have strong positive correlation, which equals 0.19. It means students with unhealthy dietary habits are likely to have depression.
- Financial Stress had a moderate positive correlation (r = 0.30).
- Study Satisfaction exhibited a negative correlation with Depression (r = -0.29). It means students, who are more satisfied with their studies, report that they don't have depression.
- Study Hours and Depression have weak positive correlation, which equals 0.20. So, more study hours are associated with a higher possibility to report depression.
- Sleep Duration and Depression have negative correlation, which equals -0.05. So, since it's a low correlation, this almost has no correlation with depression in this dataset.
- Age and Depression have weak positive correlation, which equals -0.22. It means, in this dataset older students are slightly less likely to report depression.
- Other variables such as Gender and Depression, Family History of Mental Illness and Depression have low positive correlations, which equal 0.03 and 0.06 respectively. This suggests that gender and family history do not have a statistically significant effect on depression rates in this dataset.

Figure 4 - Correlation Matrix for Numerical and Categorical Features

3.3 Hypothesis Testing

- A one-sample t-test indicated that the mean study hours (6.4) were significantly higher than the hypothesized average of 6 hours ($p = 0.015$).
- An independent two-sample t-test revealed no significant gender-based difference in study hours ($p = 0.378$).
- A paired t-test comparing Financial Stress and Academic Pressure found no significant difference in their mean values ($p = 0.394$), indicating that both are comparably impactful.
- ANOVA tests revealed no statistically significant difference in depression levels across age or gender groups.

3.4 Multidimensional Analysis

PCA reduced the dimensionality of the dataset while preserving the variance structure, identifying key contributing features to depression.

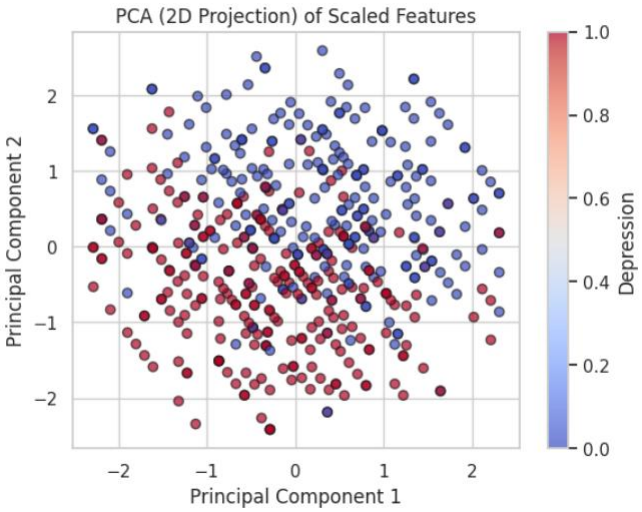

Figure 5 - Visualization of PCA on the first two principal components

K-means clustering formed two main clusters: students at high risk and low risk of depression. Some overlap between clusters suggests multidimensional complexity in diagnosing depression based on one factor.

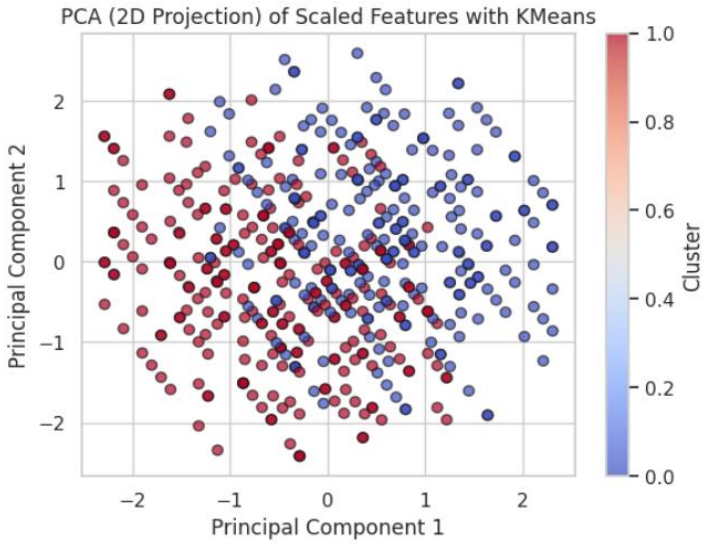

Figure 6 - Visualization of KMeans clustering

4. Analysis and Interpretation

This study provides new insights through advanced statistical analysis. Academic pressure emerged as the most critical determinant. Financial stress and suicidal ideation were also significantly linked to depressive symptoms, underscoring the compounded burden many students face.

Interestingly, family history and gender did not exhibit significant predictive power, suggesting that situational factors may outweigh inherent predispositions in this context.

When compared with the study named "Examining the Levels of Depression among University Students during the Final Assignment" by Munthe & Saragi, which focused on final-year Indonesian students experiencing depression due to final assignments, there are some similarities and differences. This research shows that academic stress is a common trigger for depression among college students and factors such as financial problems, poor time management, pressure from parents and the surrounding environment. So, both researches identify academic pressure and financial problems as key contributors to depressive symptoms. Notably, their study reported higher depression prevalence among male students, which was not reflected in this research's dataset.

Furthermore, their research provides more situational context, since the main focus is on post-pandemic time and final assignments, whereas this research's analysis captures general academic life and student well-being beyond a single academic situation. Their study complements this research's statistical insights and this research highlights how different methodologies can contribute to a more complete understanding of student mental health.

5. Recommendations

Universities should develop preventative programs focused on academic stress management and financial counselling.

Peer support and mental health services should be expanded, particularly for students reporting suicidal ideation.

Initiatives promoting healthy sleep and nutrition could play a supplementary role in mental health maintenance.

6. Conclusion

This research demonstrates that no single factor alone determines depression among students. And sum of the various factors significantly impact mental health and depression level among the student population

The combined effects of the following stressors, especially academic pressure, suicidal ideation, and financial problems, have been shown to significantly increase the likelihood of depression.

These findings align with research "50 Current Student Stress Statistics: 2025 Data, Analysis & Predictions" in research.com, which highlighted that college students face challenges such as increased academic workload, social adjustments, financial concerns and living away from home. Results of Student Depression analysis confirms the main role of academic pressure and also adds the important insight that financial stress and suicidal ideation are strong statistical indicators of depression.

The identified patterns can be used as targeted recommendations to support students' mental health through university interventions, educational programs, and professional services.

The proposed algorithm is universal and can be used for individual educational institutions, for research on global issues and student adaptation, to increase resilience and overall well-being in the academic environment. Continued research and data analysis are essential to adapting and refining these strategies.

References

1. Nurhalijah Munthe, Muhammad Putra Dinata Saragi: Examining the Levels of Depression among University Students during the Final Assignment, Journal of Educational, Health and Community Psychology Vol 12, No 1, 2023 E-ISSN 2460-8467
2. Imed Bouchrika, Phd: 50 Current Student Stress Statistics: 2025 Data, Analysis & Predictions, MAY 22, 2025: https://research.com/education/student-stress-statistics
3. John A. Rice (2007): Mathematical statistics and data analysis (3rd ed.). Duxbury Press.
4. Dimitri P. Bertsekas, John N. Tsitsiklis (2002): Introduction to probability. Athena Scientific.