Group 5 | Abhinav Koukuntla, Logan Buchanan, Nisna Tanna, Ishita Rout, Joe Posillico and Sravanth Tumma
ITCS 3162 - Data Mining
11/25/2024

Dataset: https://www.kaggle.com/datasets/ikynahidwin/depression-student-dataset

Introduction

In the current educational environment, students often face significant academic pressure, which can negatively affect their mental health and overall well-being. Understanding the factors contributing to academic stress and mental health issues is essential for developing effective support systems within educational institutions. Our dataset explores the relationship between various stressors, including financial stress, and students' mental health and academic satisfaction**.** By analyzing this data, we aim to identify key predictors of student well-being and provide insights that can help improve the educational experience for students.

Data

The dataset we have chosen to work on for our project is called “Depression Student Dataset” which was published by the authors Nurrizky Arum Jatmiko, Daru Okta, and Adinda Salsabila on the website Kaggle. The dataset itself contains 502 different total entries with 11 different features. Five of these features are of numerical data types: Age, Academic Pressure, Study Satisfaction, Study Hours, and Financial Stress. The remaining 6 features are of the categorical data type and are as follows Gender, Sleep Duration, Dietary Habits, Suicidal Thoughts, Family History, and Depression. We will use these eleven features above to create models that can be used to answer the posed research question.

Methodology Plan

For our analysis, we plan to use Linear Regression and K-means Clustering to evaluate relationships and patterns within the dataset.

1. Linear Regression: We will use linear regression to examine how factors like financial stress, academic workload, and study hours directly affect each other when examining students' study satisfaction and mental health indicators. This will help us understand the impact of each variable on overall student well-being.
2. K-Means Clustering: Clustering will allow us to group students with similar characteristics, such as similar levels of stress or mental health status. This clustering will make it easier to identify distinct groups within the student population and tailor interventions accordingly.

Additionally, we will use correlation analysis to identify the strength of relationships between different variables and use visualization techniques to present our findings. This approach will enable us to look at key information and make data-driven recommendations for improving student support systems.

The variables that we will be trying to predict will be depression which is a boolean value of either yes or no. The predictor variables will be most of the remaining filters: Gender, Age, Academic Pressure, Study Satisfaction, Sleep Duration, Dietary Habits, Study Hours, Financial Stress, Family History of Mental Illness, and Have you ever had suicidal thoughts?

Evaluation Plan
Our primary goal is to determine which factors most significantly influence students' study satisfaction and mental health. To achieve this, we will:

1. Develop linear regression models to quantify the impact of financial stress, academic workload, and other variables on study satisfaction and mental health indicators.
2. Apply K-Means Clustering to segment students into groups based on their stress levels and mental health status, allowing us to identify patterns and common characteristics within each cluster.
3. Validate our models using metrics such as Inertia and Silhouette Score for the K-means model and metrics such as mean absolute error, mean squared error, root mean squared error, and R^2 to evaluate the linear regression model. These metrics will aid us in making conclusions about our models.

We will also conduct surveys to gather additional feedback from students, ensuring that our data-driven insights are aligned with their real-world experiences. Secondary questions include identifying which demographic groups are most affected by academic stress and exploring the role of social support in mitigating mental health issues.

Initial Tasks:
- Introduction - Nisna Tanna [12/01]
- Pre-processing - Ishita Rout [12/02]
- Linear Regression Model - Joe Posillico [12/03]
- Clustering - Logan Buchanan [12/03]
- Impact-Sravanth [12/04]
- Conclusion - Abhinav Koukuntla Due: [12/04]
- Code upload to Github - Everyone

Github:
https://github.com/joepo95/ITCS-3162-Group-5-Project

Group Expectations
We will set due dates for each of our above tasks and work together to ensure they are all completed. If there is a lack of communication or issues meeting the task deadline, we will first give a warning. If the warning does not work, we will contact either Professor Benedict or a TA to try to resolve the issues. Finally, if this is unsuccessful, a letter grade reduction will be issued.