Milestone

Project Title: How can machine learning help identify and address the key factors impacting college students' mental health?

Team members (Group N):

- Linn Juge
- Alexandra Kelber
- Lora Maillard

Summary

I. Introduction (why)

A. Motivation

Mental health challenges among college students have reached alarming proportions, representing a complex and multifaceted issue that significantly impacts not only academic performance but also the overall well-being of individuals within higher education institutions. The modern academic environment, characterized by increased competition, high expectations, and the diverse psychosocial challenges faced by students, underscores the critical need for comprehensive and proactive mental health support systems. Our project aims to leverage machine learning techniques to identify and address key factors affecting college students' mental health. With the increasing availability of digital data, including academic records and online communication, we believe data-driven approaches can provide valuable insights and contribute to proactive mental health support.

B. State of the Art

This study underscores the pressing need for addressing students' mental health amid societal pressures. It proposes a mental health evaluation system using a joint optimization algorithm with an improved decision tree and artificial neural network. The research employs data mining and joint learning for analysis, demonstrating improved accuracy and efficiency. The article explores machine learning applications in predicting psychological issues, citing successful examples. The proposed system aims to effectively tackle contemporary psychological challenges through data-driven approaches, emphasizing the importance of mental health guidance.

Another article highlights the increasing global burden of mental health issues, particularly among college students, emphasizing the need for urgent attention. It explores the emerging problems of anxiety, depression, stress, and suicidal thoughts within this demographic. The paper focuses on the application of Artificial Intelligence (AI) and Machine Learning (ML) in addressing mental health challenges among students. Utilizing bibliometric analysis, the study identifies a critical need to align efforts with the United Nations Sustainable Development Goal 3, emphasizing awareness at various levels.

C. Available Datasets

We have come across diverse datasets encompassing academic records, survey responses, and social media interactions. They allow us to capture both quantitative and qualitative aspects of students' experiences, forming the basis for our analysis.

II. Hypotheses and Specific Objectives (What)

A. Hypotheses

We made a list of specific topics that can impact students' mental health, so that we can later focus on the most important questions after analysing the data available on different sources.

1. **Academic Stress:**
    - How does academic stress vary among students across different academic levels (freshmen, sophomores, etc.) and majors?
2. **Impact of Extracurricular Activities/ Physical Activity:**
    - How does participation in extracurricular activities relate to the mental health of students?
    - Are certain types of extracurricular activities more beneficial for mental well-being than others?
3. **Access to Mental Health Resources:**
    - How knowledgeable are students about mental health issues, and does higher mental health literacy correlate with better mental health outcomes?
4. **Relationship between Sleep Patterns and Mental Health:**
    - How do sleep patterns and sleep quality impact the mental health of students?
    - Are there differences in sleep habits between students with varying levels of stress or anxiety?
5. **Social Support on Campus:**
    - How do social support networks influence the mental health of students?
6. **Technology Use and Mental Health:**
    - What is the relationship between the use of technology (social media, screen time) and mental health outcomes among students?
    - Are there patterns of technology use that are associated with higher stress levels?
7. **Academic Performance and Mental Health:**
    - How does mental health status correlate with academic performance among students?
    - Are there specific mental health factors that are more strongly associated with academic struggles?
8. **Financial Stress:**
    - To what extent does financial stress contribute to mental health issues among students?
9. **Transition Periods and Mental Health:**
    - How do students experience mental health during transitional periods (e.g., transitioning from high school to college), and are there specific challenges faced during this period?

- Are there specific interventions that can support students during these transitional phases?

10. **Impact of Remote Learning:**
    - How has the shift to remote learning during the COVID-19 pandemic affected the mental health of students, and are there differences in experiences between various academic levels?
11. **Relationship between Peer Support and Mental Health:**
    - What role does peer support play in the mental health of students, and how do different types of social relationships influence well-being?
12. **Coping Mechanisms:**
    - How do students cope with stress, and are there effective coping mechanisms that contribute to better mental health outcomes?
13. **Eating Habits and Mental Health:**
    - Is there a correlation between dietary habits, such as nutrition and meal regularity, and the mental health of students?
14. **Time Management and Stress:**
    - How does time management relate to stress levels among students, and are there time management strategies associated with better mental health?
15. **Impact of Campus Environment:**
    - What role does the physical environment of the campus play in the mental health of students, and how do factors like green spaces or noise levels contribute?
16. **Impact of Cultural and Racial Identity:**
    - How does cultural and racial identity impact the mental health experiences of students, and are there unique stressors associated with specific cultural backgrounds?
17. **Impact of Part-Time Employment:**
    - How does part-time employment during the school year affect the mental health of students, and are there differences based on the nature of the work?

B. Specific Objectives

Our objectives include :

1. *Developing Predictive Models:* Build ML models to predict mental health outcomes based on academic and sentiment features.
2. *Feature Engineering:* Refine feature extraction methods to enhance the informativeness of academic and textual data.

3. *Evaluating Model Performance:* Assess the accuracy and interpretability of the developed models using rigorous evaluation metrics.

III. Methods (How)

A. Data collection

The first step is to collect diverse datasets from various sources, including academic records, survey responses... The goal is to create one or several comprehensive datasets that capture both quantitative and qualitative aspects of students' experiences. We will try to explore relevant platforms to access the necessary data.

Then the collected data has to be cleaned and preprocessed. This involves addressing missing or inconsistent values and encoding categorical variables. To facilitate quantitative analysis, we will convert categorical responses to numeric values.

B. Exploratory Data Analysis (EDA)

Our initial exploration will involve generating descriptive statistics for relevant variables. We aim to gain insights into key metrics related to mental health indicators and other pertinent features. We will employ visualizations like histograms and scatter plots to identify trends and patterns within the dataset. Insights from EDA will guide subsequent analyses.

C. Feature Engineering

Feature engineering is a critical step in refining our dataset. We will apply transformations, including scaling and normalization, to enhance the effectiveness of our features in predicting mental health outcomes.

D. Machine Learning Models

For predictive analysis, we plan to select machine learning models suited for classification tasks related to mental health outcomes. The dataset will be divided into training and testing sets, with careful consideration given to randomization. We will perform hyperparameter tuning to optimize model performance. We may also use neural networks to predict whether a person have mental health problems based on different data we have on them.

Evaluation metrics, such as accuracy, precision, and recall, will be chosen to assess the models' predictive capabilities. The selection of these metrics aligns with the specific objectives of our study, emphasizing the importance of accurately predicting mental health outcomes.

IV. Preliminary experiments

A. Data Collection

We gathered data from diverse sources, including academic records, surveys, and questionnaires, gathered from students to capture subjective experiences.

We found several datasets on different websites:

Student Mental health (kaggle.com)

Student Stress Factors: A Comprehensive Analysis (kaggle.com)

Student Stress Survey Jan2020 OPENDATA.xlsx (figshare.com)

Relationship between academic stress, emotional intelligence and eating behavior in university students (zenodo.org)

Student anxiety dataset (kaggle.com)

Relationship Between Smartphone Usage and Sleep Quality & Mental Health of College Students During the COVID-19 Pandemic (openicpsr.org)

Awareness of insomnia on physical and mental health of medical student: An online cross-sectional survey - Mendeley Data

StudentLife (kaggle.com)

Students Mental Health Assessments (kaggle.com)

These datasets contain to some extent the data needed to respond to the many hypotheses we had about what is impacting students' mental health. Especially, the third one contains many data concerning most of our questions, so it will be one that we will use for our project.

B. Data Preprocessing and EDA

We decided to focus on 3 of these datasets for the moment : Student Mental health (kaggle.com) , Student Stress Survey Jan2020 OPENDATA.xlsx (figshare.com), Students Mental Health Assessments (kaggle.com) , but the second one is the one that we worked on the most, because it contains many questions and aspects of a student life that we can analyse to get the more accurate results.

First, we had to preprocess the data, which means do the data cleaning and integration.

This process involved combining data from different sources, which included datasets and spreadsheets, renaming columns, converting data types, addressing missing values, deleting useless rows and columns, scaling numeric values, encoding categorical variables.

On this data, we then made histograms in order to understand the features more and choose which are the best machine learning methods we can apply.

C. Baseline Model:

The baseline model for the first dataset utilizes Random Forest Classifiers, leveraging features like gender, age, course, study year, CGPA, and marital status, to predict anxiety and depression. Trained on split datasets, they provide initial insights into mental health prediction among students. The models are evaluated using accuracy metrics, confusion matrices, and classification reports. Additionally, Random Forest Regressors analyze feature importance, highlighting key contributors to predictions. These foundational models serve as a starting point for further refinement and feature engineering to enhance predictive accuracy in mental health.

D. Next Steps

Building on our preliminary results, our next steps include:

1. We plan to explore more ML models, such as Neural networks, Classification, PCA to improve predictive accuracy.
2. We have to continuously refine feature engineering to better capture nuances in finding the factors impacting students' mental health.

To check the work that has been done for the moment, follow this link:

GitHub Repository: loramaillard/Data_Science_Final_project (github.com)

Conclusion

In conclusion, our project, focusing on college students' mental health, leverages machine learning to gain insights from diverse datasets. This milestone marks our progress in developing a comprehensive solution for understanding and addressing college students' mental health using ML.