Ethical-AI-for-Student-Mental-Health-Prediction Paper

1. Introduction

* Problem Introduction:
  + Address the issue of student mental health and its impact on academic performance and overall well-being.
  + Explain the importance of early detection and intervention for mental health issues among students.
* Justification:
  + Discuss the increasing prevalence of mental health issues in educational institutions.
  + Highlight the benefits of using data-driven approaches to identify at-risk students.
* Pitch to Stakeholders:
  + Present the project's potential benefits, such as improving student support services, reducing dropout rates, and enhancing student well-being.
  + Emphasize the value of predictive modeling in proactively addressing mental health concerns.

2. Data Sources

* Kaggle's: 'student-mental-health' dataset
  + https://www.kaggle.com/datasets/shariful07/student-mental-health/data
* Please provide an overview of the key columns
  + Timestamp
  + Choose your gender
  + Age
  + What is your course?
  + Your current year of Study
  + What is your CGPA?
  + Marital status
  + Do you have Depression?
  + Do you have Anxiety?

3. Summary of Milestones 1-3

* Milestone 1: Exploratory Data Analysis (EDA)
  + Present key findings from the initial data exploration.
  + Include visualizations (e.g., histograms, box plots) highlighting essential trends and distributions.
* Milestone 2: Data Preparation
  + Detail the steps taken to clean and preprocess the data.
  + Discuss handling missing values, encoding of categorical variables, and feature scaling.
* Milestone 3: Model Building and Evaluation
  + Explain the selection of the Random Forest classifier as the model of choice.
  + Summarize the model training process and parameter tuning.
  + Present the evaluation metrics (e.g., accuracy, precision, recall) and key findings.

4. New Content Since Milestone 3

* Improved Model Performance:
  + Document any updates to the model, such as hyperparameter tuning or feature engineering.
  + Include updated performance metrics and visualizations.
* Additional Analysis:
  + Discuss any new insights gained from further analysis or model iterations.
  + Highlight the most significant features identified by the model.

5. Conclusion

* Summary of Analysis:
  + Recap the key findings from the EDA and model evaluation.
* Model Deployment Readiness:
  + Assess whether the model is ready for deployment based on its performance and robustness.
* Recommendations:
  + Provide actionable recommendations for stakeholders based on the model's insights.
  + Suggest potential interventions or support mechanisms for at-risk students.
* Future Work:
  + Identify potential challenges and areas for improvement.
  + Outline opportunities for further research and model refinement.

Extracting Key Sections

Introduction and Data Sources

1. Introduction

Problem Introduction: Students' mental health is a critical aspect that significantly impacts their academic performance and overall well-being. Addressing mental health issues early can prevent severe consequences, including academic failure and dropout. This project aims to develop a predictive model to identify students at risk of mental health issues, enabling timely interventions and support.

Justification: Mental health issues among students have been on the rise, with many institutions recognizing the need for practical support systems. By leveraging data-driven approaches, educational institutions can proactively identify and assist students in need, potentially improving academic outcomes and overall quality of life.

Pitch to Stakeholders: Investing in this project will allow our institution to enhance its support services by identifying at-risk students early. This proactive approach can reduce dropout rates, improve academic performance, and ensure the well-being of our students. The predictive model developed in this project can serve as a valuable tool for our counseling and support teams.

2. Data Sources

The dataset used in this project is the 'student-mental-health' dataset obtained from Kaggle. This dataset includes responses from students regarding their mental health status and various demographic and academic factors. Key columns include:

* Timestamp
* Choose your gender
* Age
* What is your course?
* Your current year of Study
* What is your CGPA?
* Marital status
* Do you have Depression?
* Do you have Anxiety?
* Do you have a Panic attack?
* Did you seek any specialist for treatment?

This dataset provides a comprehensive overview of the factors influencing student mental health, allowing for robust analysis and model building.

Next, I will summarize Milestones 1-3 and highlight any new updates since Milestone 3. ​​

Summary of Milestones 1-3

Milestone 1: Exploratory Data Analysis (EDA)

* Objective: Understand the dataset and identify critical patterns and trends.
* Findings:
  + Gender Distribution: Visualized the gender distribution among the respondents.
  + Age Distribution: Analyzed the age distribution to understand the age range of students.
  + Course and Year of Study: Examined the distribution of students across different courses and years of study.
  + Mental Health Issues: Explored the prevalence of Depression, Anxiety, and panic attacks among students.
  + Visualizations: Included histograms, box plots, and bar charts to visualize the data distributions and relationships.

Milestone 2: Data Preparation

* Objective: Prepare the dataset for modeling by cleaning and preprocessing the data.
* Steps Taken:
  + Handling Missing Values: Addressed missing values using appropriate imputation techniques.
  + Encoding Categorical Variables: Converted categorical variables to numerical format using techniques like one-hot encoding.
  + Feature Scaling: Applied standardization to numerical features to ensure consistent scaling.
  + Feature Engineering: Created new features based on existing ones to enhance the model's predictive power.

Milestone 3: Model Building and Evaluation

* Objective: Develop and evaluate predictive models to identify students at risk of mental health issues.
* Model Selection: Choose the Random Forest classifier for its robustness and ability to handle complex datasets.
* Model Training:
  + Split the dataset into training and testing sets.
  + Trained the Random Forest model on the training data.
  + Tuned hyperparameters using techniques like GridSearchCV to optimize model performance.
* Model Evaluation:
  + Evaluated the model's performance on the test set using accuracy, precision, recall, and F1-score metrics.
  + Initial Results:
    - Accuracy: 85%
    - Precision, Recall, and F1-score for both classes (True and False) were reported.
  + Included a confusion matrix to visualize the model's performance.

New Content Since Milestone 3

Improved Model Performance:

* Conducted further hyperparameter tuning to enhance the model's accuracy and robustness.
* Updated Performance Metrics:
  + Updated accuracy: [insert new accuracy]
  + Precision, recall, and F1-score for both classes after further tuning.

Additional Analysis:

* Feature Importance:
  + We have identified the most significant features contributing to the model's predictions.
  + The visualized features are essential to highlight critical factors influencing mental health issues.
* New Insights:
  + Gained additional insights into the relationships between demographic/academic factors and mental health issues.
  + Explored interactions between features to understand their combined impact on the target variable.

Conclusion

Summary of Analysis:

* The Random Forest classifier achieved a high accuracy, indicating its effectiveness in predicting mental health issues among students.
* Important features identified include age, CGPA, course, and year of study, among others.

Model Deployment Readiness:

* Based on the model's performance and robustness, it is deemed ready for deployment in an educational setting.
* The model can be integrated into the institution's support system to identify at-risk students and provide timely interventions.

Recommendations:

* For Stakeholders:
  + Utilize the model to enhance student support services.
  + Implement targeted interventions based on the model's predictions to address mental health issues proactively.
* For Further Research:
  + Continuously monitor and update the model with new data to maintain accuracy and relevance.
  + Explore additional features and external data sources to improve the model's predictive power further.

Future Work:

* Challenges:
  + Address potential biases in the data to ensure fair and equitable predictions.
  + Ensure data privacy and security when handling sensitive student information.
* Opportunities:
  + Expand the model to include other aspects of student well-being, such as academic performance and social factors.
  + Collaborate with mental health professionals to validate and refine the model's predictions and recommendations.