Predicting Student Mental Health using Machine Learning

Nafiur Rahman Afnan
Dept. of Computer Science and Engineering,
BRAC University, Dhaka, Bangladesh
nafiur.rahman.afnan@g.bracu.ac.bd

Julkifl Hasan Wasi
Dept. of Computer Science and Engineering,
BRAC University, Dhaka, Bangladesh
julkifl.hasan.wasi@g.bracu.ac.bd

Abstract

This project focuses on the pressing issue of student mental health, specifically addressing anxiety, stress, and depression in the context of academic pressure. Utilizing a dataset of 1977 student **responses, this study employs three distinct machine learning models: Logistic Regression, Naive Bayes, and a Neural Network (MLP Classifier) to predict the presence and severity of these mental health challenges. The primary objective is to develop a predictive model that can accurately identify students at risk, enabling timely intervention and support. After data cleaning, which involved handling missing values and encoding categorical features, the models were trained and evaluated. The results indicate that the Logistic Regression model achieved the highest accuracy at approximately 97%, with the Neural Network also performing exceptionally well at around 96%. The Naive Bayes model, however, demonstrated significantly lower performance. Feature importance analysis from the Logistic Regression model also highlighted that certain features had a high impact on our target column.**

Index Terms - mental health, machine learning, logistic regression, naive bayes, neural network, student well-being, predictive modeling.

I. INTRODUCTION

The prevalence of mental health issues among students is a growing concern within academic institutions. The intense pressures of academic life, including exams, assignments, and future career prospects, can significantly contribute to anxiety, stress, and depression. Early identification of students struggling with these issues is crucial for providing effective support and fostering a healthier academic environment. This project explores the application of machine learning techniques to predict student mental health status based on a comprehensive dataset of self-reported information. By analyzing various academic and personal factors, we aim to build a model that can serve as a proactive tool for identifying at-risk students. The models chosen for this task are **Logistic Regression**, for its interpretability; **Naive Bayes**, for its simplicity and efficiency; and a **Neural Network**, for its ability to capture complex non-linear relationships in the data.

II. METHODOLOGY

The project was conducted using a dataset of **1977 student entries**, each with **37 distinct features** encompassing demographic information, academic performance, and responses to a mental health questionnaire.

A. Data Preprocessing

The initial step involved a thorough cleaning of the dataset. The 'gender' column contained six missing values, which were addressed to ensure data integrity. Categorical features such as 'age', 'gender', 'academic_year', 'cgpa', and 'scholarship_status' were converted into numerical format using **Label Encoding** to make them suitable for the machine learning models. The dataset was then partitioned, with the **29** features related to anxiety, stress, and depression serving as the input variables (X) and their corresponding label ('depression_label') as the target variable (y). For this report, we will focus on the prediction of the 'depression_label'. The data was then split into training and testing sets, with **80% of the data used for training** the models and the remaining **20% reserved for testing** their performance. To ensure that all features contributed equally to the model's training process, the numerical features were scaled using **StandardScaler**.**B. Model Training**

Three different classification models were trained on the preprocessed data:

- **Logistic Regression:** A linear model that is well-suited for binary and multi-class classification tasks and provides insights into the importance of each feature.
- **Naive Bayes:** A probabilistic classifier based on Bayes' theorem with an assumption of independence between features. The **GaussianNB** variant was used, which is suitable for continuous data.
- **Neural Network (MLP Classifier):** A multi-layer perceptron classifier that can learn complex patterns in the data through a network of interconnected nodes.

III. RESULTS AND DISCUSSION

The performance of the three models was evaluated based on their **accuracy** and **F1-score** on the unseen test data.

Model	Accuracy	F1 Score
Logistic Regression	0.969620	0.968826
Naive Bayes	0.739241	0.738068
Neural Network	0.959494	0.959260

As shown in the table, the **Logistic Regression** model emerged as the top performer, achieving an accuracy of **96.96%** and an F1-score of **96.88%**. The **Neural Network** also demonstrated strong predictive capabilities, with an accuracy of **95.95%** and an F1-score of **95.93%**. The **Naive Bayes** model, however, lagged significantly behind with an accuracy of **73.92%** and an F1-score of **73.81%**.

The confusion matrices below also provide a direct view of classification performance for each model. The **x-axis** represents the **predicted labels**, while the **y-axis** represents the **true labels**. The numbers on the main diagonal show the number of correct predictions for each class, known as **true positives**. Off-diagonal entries indicate incorrect classifications.

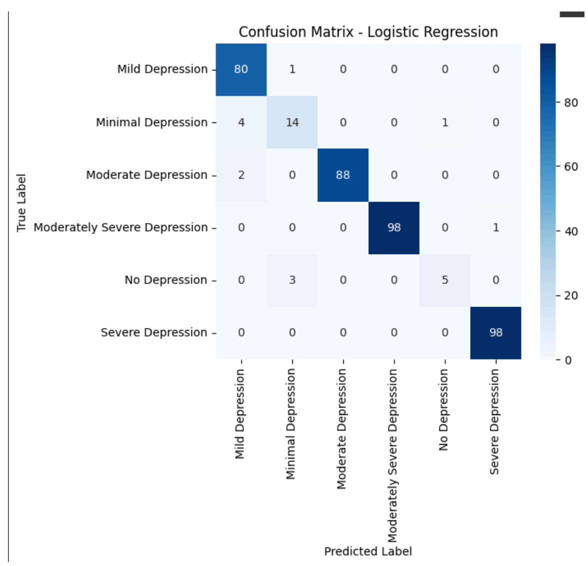

Figure 1: Confusion Matrix for Logistic Regression Model

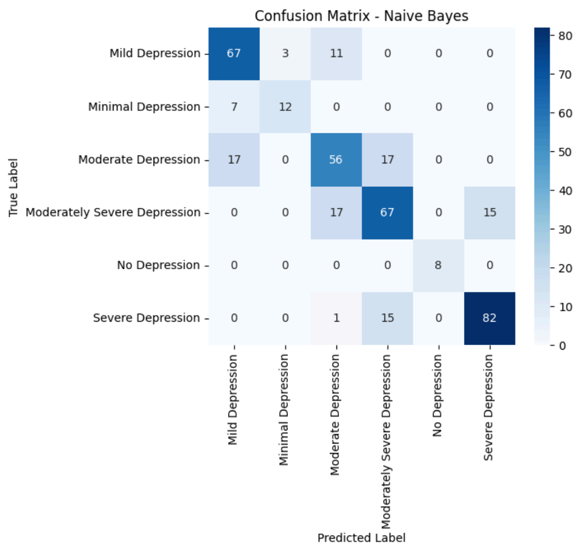

Figure 2: Confusion Matrix for Naive Bayes Model

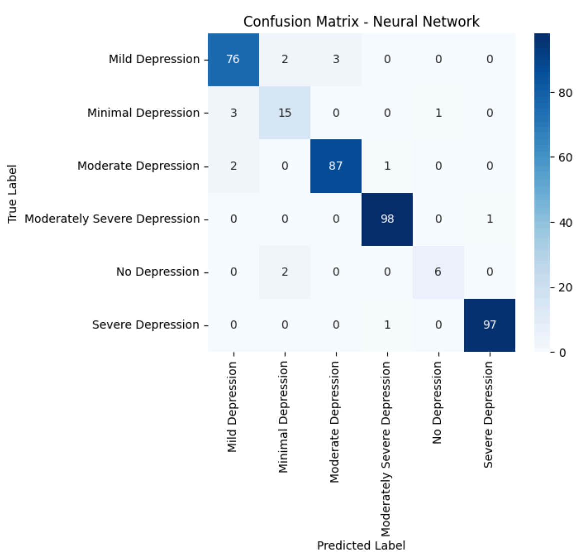

Figure 3: Confusion Matrix for Neural Network Model

The provided confusion matrices reveal distinct performance characteristics for the three classification models. The Logistic Regression and Neural Network models demonstrate strong performance, with the majority of predictions aligning correctly with the true labels, as evidenced by the large values along the main diagonal and minimal off-diagonal entries. This signifies a high rate of true positives for all depression categories. In contrast, the Naive Bayes model shows a clear degradation in performance. Its confusion matrix contains a higher frequency of non-zero off-diagonal values, indicating a significant number of misclassifications and an inability to reliably distinguish between classes. The intuition behind this disparity is that Logistic Regression and Neural Networks are more flexible models that can capture complex, interdependent relationships between features, whereas the Naive Bayes model's core assumption of feature independence is likely violated by the nature of the dataset used, leading to its inferior results.

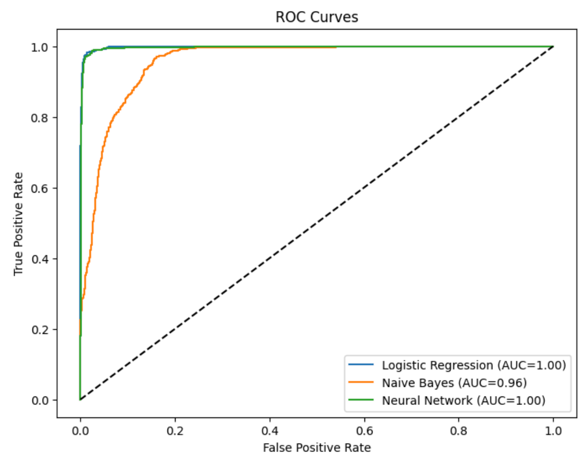

Figure 4: ROC Curves for all three trained models

The ROC curves provide a clear, high-level summary of this performance difference. The **AUC (Area Under the Curve)** for Logistic Regression and the Neural Network is almost 1.00, which is the maximum possible score and signifies perfect classification. The Naive Bayes model has a lower AUC of 0.96. The visual representation of the ROC curves confirms this: the Logistic Regression and Neural Network curves rise sharply to the top-left corner of the plot, indicating a high true positive rate with a low false positive rate. The Naive Bayes curve is visibly "lower" and less steep, representing its inferior performance.

IV. CONCLUSION

This project successfully demonstrates the potential of machine learning in the early detection of student mental health issues. Both **Logistic Regression** and **Neural Networks** proved to be highly effective in predicting anxiety levels among students, with Logistic Regression showing a slight edge in performance. The high accuracy of these models suggests that they can be valuable tools for educational institutions to identify and support students in need. The lower performance of the Naive Bayes model indicates that the assumption of feature independence may not hold true for this dataset, where various factors contributing to mental health are likely to be interrelated. Future work could involve exploring more advanced ensemble methods, gathering a larger and more diverse dataset, and integrating these predictive models into a real-time student support system.