Predicting Student Depression Using Machine Learning: A Comparative Study of Classification Models

Fahim Shahriar Rafi (24141188)
Department of Computer Science
BRAC University
Dhaka, Bangladesh
fahim.shahriar.rafi@g.bracu.ac.bd

Tasnia Mahjabin Maliha (22101644)
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
tasnia.mahjabin.maliha@g.bracu.ac.bd

Razin Sufian (24141183)
Department of Computer Science
BRAC University
Dhaka, Bangladesh
razin.sufian@g.bracu.ac.bd

Md. Mahmudul Hasan (22201198)
Department of Computer Science
BRAC University
Dhaka, Bangladesh
md.mahmudul.hasan8@g.bracu.ac.bd

This project investigates the application of various Machine Learning algorithms to predict whether a student experiences depression, based on behavioral, academic, and psychological indicators. Using the Student Depression dataset, we implemented and compared multiple classifiers including Logistic Regression, Random Forest, XGBoost, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, and Naive Bayes. Each model was evaluated both with and without Principal Component Analysis (PCA) to assess the impact of dimensionality reduction on performance. The dataset was preprocessed through column filtering, handling missing values via imputation, label encoding, and feature scaling. Model performance was assessed using accuracy, confusion matrix, and classification reports. Among all models, SVM and Logistic Regression achieved the highest accuracy, indicating a predominantly linear relationship between the input features and the target variable.

Keywords—Student depression prediction, machine learning, classification models, mental health, support vector machine (SVM), logistic regression, dimensionality reduction, principal component analysis (PCA), feature engineering, supervised learning.

I. Introduction

In today's fast-paced academic environment, students are increasingly subjected to overwhelming study pressure, often leading to undiagnosed cases of depression. This mental health burden not only affects their emotional well-being but also impacts their physical health, academic performance, and in severe cases, can lead to suicidal tendencies. Given the seriousness of the issue, early detection of depression among students is crucial. The objective of this project is to develop a classification-based machine learning model capable of predicting whether a student is suffering from depression. By identifying at-risk individuals, timely psychological or medical intervention can be initiated. To achieve this, we analyze a wide range of behavioral, academic, and psychological attributes of students. The project employs both traditional and ensemble machine learning models, and evaluates them using classification accuracy and other performance metrics to determine the most effective approach.

II. Dataset Description

The dataset used in this project is the *Student Depression Dataset*, obtained from Kaggle[1], which contains responses from 27,901 students. Each record includes a diverse range of features spanning demographic, academic, psychological, and lifestyle-related factors. The dataset consists of 18 columns in total, with the target variable being **Depression**, which is labeled as 0 (not depressed) or 1 (depressed).

To understand the distribution of depression cases in the dataset, we visualized the target variable. As shown in **Figure 1**, the dataset presents a moderate class imbalance, with a higher number of students labeled as depressed. This imbalance is important to consider when interpreting the evaluation results of the classification models.

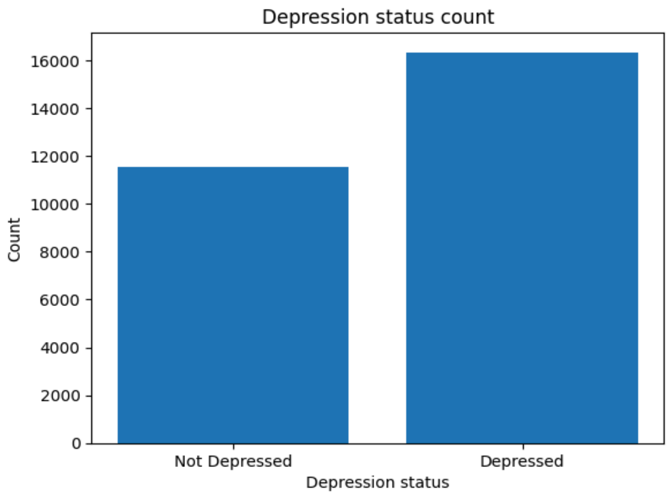

Figure 1

The dataset includes both categorical and numerical features. Categorical attributes such as **Gender**, **City**, **Profession**, **Dietary Habits**, and **Degree** provide demographic context. Numerical features include **Age**, **Academic Pressure**, **Work Pressure**, **CGPA**, **Study Satisfaction**, and **Work/Study Hours**, among others. Additionally, features like **Suicidal Thoughts** and **Family History of Mental Illness** serve as potentially strong indicators of mental health conditions.

Several features were found to contain missing values, particularly **CGPA**, **Job Satisfaction**, and **Financial Stress**.

These were addressed during the data preprocessing phase using column removal or imputation techniques depending on the importance of the feature.

Overall, the dataset offers a rich set of attributes that can be used to train machine learning models to predict depression in students based on multiple behavioral and contextual factors.

III. Methodology

This section outlines the complete pipeline used in building and evaluating the student depression prediction models. It includes data preprocessing, feature transformation, dimensionality reduction, model selection, and evaluation.

A. *Data preprocessing*

The initial phase involved cleaning and preparing the raw dataset. Several preprocessing steps were carried out to ensure model readiness:

- **Column Removal**: Irrelevant or redundant columns such as IDs were dropped.
- **Handling Missing Values**:
    - Non-essential rows with missing values were removed.
    - Important features with missing data (e.g., CGPA, Financial Stress) were imputed using statistical methods such as mean or mode imputation.
- **Encoding Categorical Variables**: Label encoding was applied to convert categorical features such as Gender, City, Degree, and Dietary Habits into numerical representations.
- **Feature Scaling**: Continuous features were normalized using **Min-Max Scaling** to bring them into a uniform range, which is essential for distance-based models like KNN and SVM.

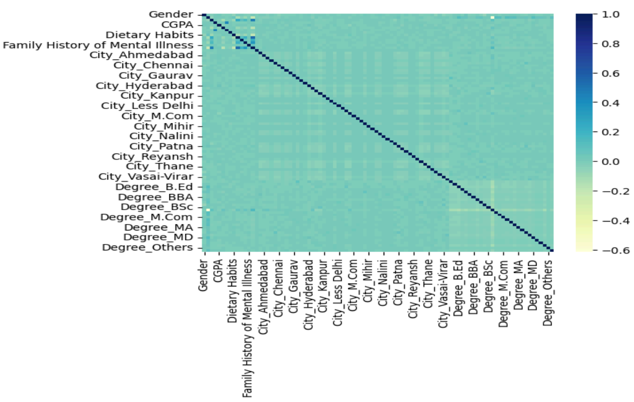

Figure 2

B. *Dimensionality reduction with PCA*

To analyze whether dimensionality reduction could improve performance, we applied **Principal Component Analysis (PCA)** on the scaled dataset. PCA transforms the original feature space into a lower-dimensional space while retaining most of the data's variance. We retained enough principal components to preserve a high percentage of the variance.

Models were trained and evaluated both **with** and **without PCA**, allowing a comparative study of their effectiveness in reduced dimensions.

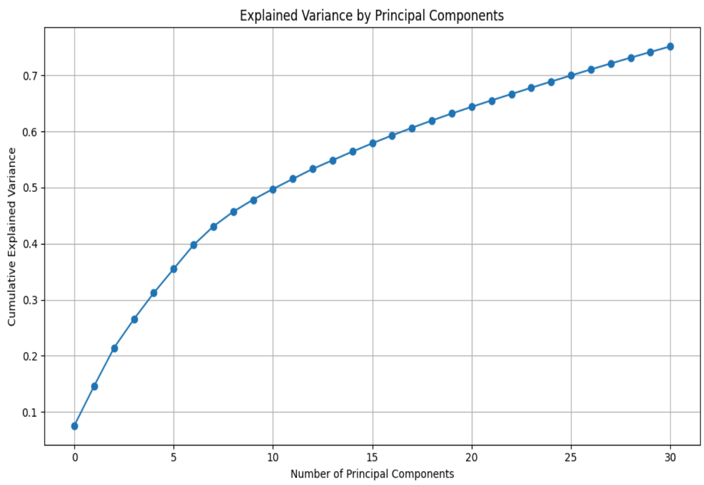

Figure 3

C. *Model Selection*

We trained and tested the following classification models:

- Logistic Regression
- Decision Tree
- Random Forest
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)
- Gaussian Naive Bayes
- XGBoost

Each model was trained on the same preprocessed dataset using an **80-20 train-test split**. Hyperparameters were kept at default values to ensure fairness and reproducibility across models. PCA-transformed data was also used to retrain the models for comparative analysis.

D. Evaluation Metrics

Model performance was measured using the following evaluation tools:

- **Accuracy**: Proportion of correctly classified instances.
- **Confusion Matrix**: Visual representation of prediction results (true vs. false labels).
- **Classification Report**: Includes precision, recall, and F1-score for each class.

IV. Results and Evaluation

To evaluate the effectiveness of each classification model, we used an 80-20 train-test split and assessed performance based on accuracy, confusion matrix, and classification reports. Each model was trained and tested both with and without Principal Component Analysis (PCA) to investigate the impact of dimensionality reduction on performance.

A. Accuracy Comparison

The classification accuracy of all models is shown in **Figure 4**. Models trained without PCA generally performed better than those with dimensionality reduction. Among the models tested, **Support Vector Machine (SVM)** and **Logistic Regression** achieved the highest accuracy scores, indicating strong predictive capabilities in a linearly separable feature space. Ensemble methods like Random Forest and XGBoost also performed well, although they slightly lagged behind the linear models in this dataset.

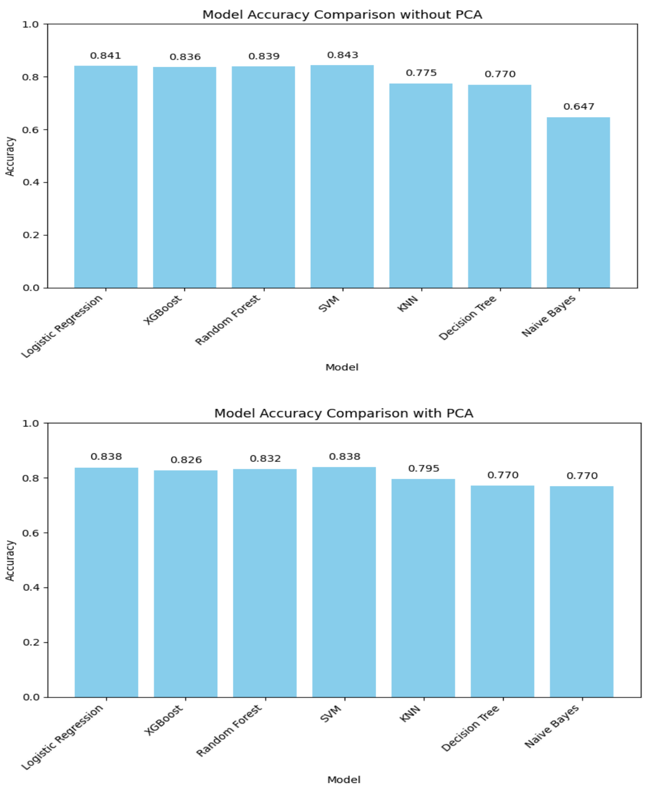

Figure 4

B. Confusion Matrices

Confusion matrices were plotted for the top-performing models to further analyze their classification behavior.

- **Figure 5** shows the confusion matrix for SVM, which had the best accuracy. The model correctly classified the majority of both depressed and non-depressed students, though some false positives were present.
- **Figure 6** presents the confusion matrix for Logistic Regression, which exhibited a similar pattern of strong performance with slightly more misclassifications than SVM.

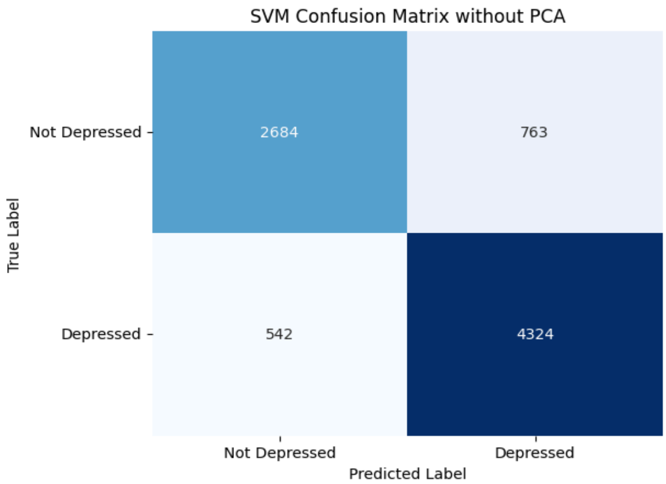

Figure 5

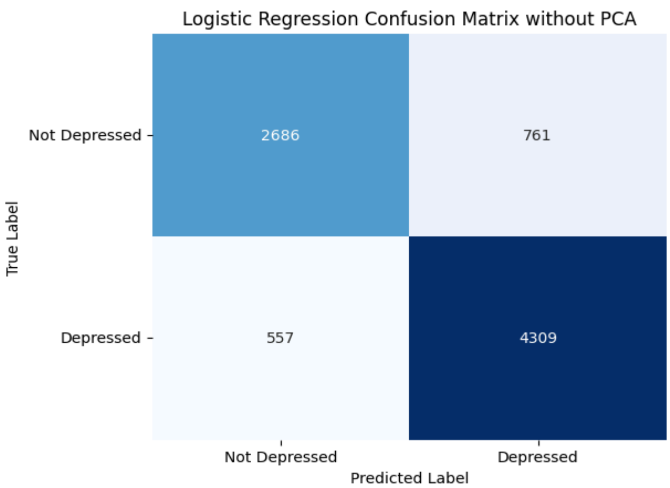

Figure 6

C. Classification Reports

Classification reports provided deeper insight into the balance of precision and recall for each model. SVM achieved both high precision and recall, resulting in a strong F1-score for both classes. Logistic Regression demonstrated similar behavior, reinforcing its suitability for this binary classification task.

Table 1 presents the classification report for SVM, which achieved the highest accuracy without dimensionality reduction. The model shows high recall for the depressed class, making it effective for identifying at-risk students.

Class Label	Precision	Recall	F1-Score	Support
0 (Not Depressed)	0.83	0.78	0.80	3447
1 (Depressed)	0.85	0.89	0.87	4866
Accuracy			0.84	8313
Macro Avg	0.84	0.83	0.84	8313
Weighted Avg	0.84	0.84	0.84	8313

Table 1

V. Discussion

The results of our experiments provide several key insights into the performance of machine learning models for predicting student depression.

First, the models trained on the original, full-dimensional dataset consistently outperformed their PCA-reduced counterparts. This suggests that most features carried useful information, and dimensionality reduction in this case may have led to a loss of critical variance. While PCA is often effective in eliminating redundancy, its use here did not improve classification outcomes, highlighting the importance of evaluating dimensionality reduction on a case-by-case basis.

Among the models tested, **Support Vector Machine (SVM)** and **Logistic Regression** emerged as the top performers, both achieving high accuracy and strong F1-scores across both classes. Their effectiveness suggests that the decision boundary between depressed and non-depressed students may be largely linear in nature. This aligns with our observation that complex ensemble methods such as Random Forest and XGBoost, while still strong, did not outperform simpler linear models.

Another important observation is the strong recall achieved by SVM for the depressed class. This is especially valuable in a mental health context, where failing to identify a student in need of help can have serious consequences. High recall ensures that most students with depressive mentality are successfully detected by the model, even at the cost of a few false positives.

Overall, the analysis emphasizes the potential of simple yet powerful classification algorithms for mental health prediction tasks. It also underscores the importance of balanced evaluation metrics like F1-score and confusion matrix insights when dealing with real-world imbalanced datasets.

VI. Conclusion and Future Work

This study explored the effectiveness of various machine learning algorithms in predicting student depression based on a diverse set of academic, behavioral, and psychological features. By applying models such as Logistic Regression, SVM, KNN, Decision Tree, Random Forest, XGBoost, and Naive Bayes, we conducted a thorough comparative analysis of classification performance. The results highlighted that **Support Vector Machine (SVM)** outperformed all other models in terms of accuracy and recall, particularly for identifying students at risk of depression.

The findings emphasize the importance of proper data preprocessing, including handling missing values, encoding categorical variables, and feature scaling, all of which played a critical role in model effectiveness. Although dimensionality reduction using PCA was explored, it did not significantly enhance model performance, suggesting that the original feature space contained rich and non-redundant information.

Looking ahead, future work can focus on incorporating **more diverse datasets** including behavioral logs, social media activity, or survey-based emotional assessments to improve model robustness. Additionally, integrating **ensemble learning techniques with advanced tuning** or **neural networks** could further optimize classification outcomes. Addressing **class imbalance** using resampling techniques or cost-sensitive learning is also a promising direction. Ultimately, such predictive tools can be embedded into early intervention systems to support student mental health at scale.

References

[1] A. Shamim, "Student Depression Dataset," Kaggle, 2022. [Online]. Available: https://www.kaggle.com/datasets/adilshamim8/student-depression-dataset