Predicting Student Depression: A Machine Learning Approach Using Survey Data

Abstract
Student mental health, particularly depression, is a growing concern. This research investigates the efficacy of machine learning models in predicting depression among students using a comprehensive survey dataset. After rigorous preprocessing, including handling missing values, feature engineering, and one-hot encoding, several classification models—Logistic Regression, Random Forest, XGBoost, and a Stacking ensemble—were trained and evaluated. The Stacking ensemble model achieved the highest performance with a ROC-AUC of 0.925 on the test set. Feature importance analysis revealed 'Suicidal Thoughts (Binary)', 'Academic Pressure', and 'Financial Stress' as key predictors. These findings underscore the potential of data-driven approaches for early identification of students at risk, enabling timely intervention.

1. Introduction

The prevalence of depression among students has significant implications for academic performance, social integration, and long-term well-being. Early detection is crucial for providing support and mitigating adverse outcomes. Traditional screening methods can be resource-intensive. This study explores the application of machine learning (ML) techniques to develop a predictive model for student depression based on self-reported survey data, aiming to identify key contributing factors and build an accurate classification system.

2. Methodology

- **Dataset:** The study utilized the "Student Depression Dataset," initially comprising 27,901 entries and 18 features. After cleaning (e.g., removing rows with '?' in 'Financial Stress', and 'Others' categories in 'Dietary Habits' and 'Degree'), the final dataset for modeling contained 27,851 student records. Features encompassed demographics (Age, Gender, City), academic factors (Profession, Academic Pressure, CGPA, Degree), lifestyle choices (Sleep Duration, Dietary Habits), and mental health indicators (Work/Study Hours, Financial Stress, Family History of Mental Illness, Suicidal Thoughts). The target variable was binary 'Depression' (Yes/No).
- **Data Preprocessing:**
    1. **Cleaning:** Rows with ambiguous ('?') or highly infrequent/vague ('Others') categorical entries were removed. The 'id' column was dropped. 'City' (high cardinality) and 'Profession' (low variance, dominated by 'Student') were removed after exploratory data analysis (EDA).
    2. **Feature Engineering:** Textual 'Sleep Duration' was parsed into numerical hours. 'Have you ever had suicidal thoughts ?' and 'Family History of Mental Illness' were converted to binary (0/1) representations.
    3. **Encoding:** Remaining categorical features ('Gender', 'Dietary Habits', 'Degree') were one-hot encoded.
    4. **Scaling:** Numerical features (9 in total, e.g., 'Age', 'CGPA', 'Sleep Duration') were standardized using StandardScaler.

- **Modeling:** The dataset was split into 70% training and 30% testing sets, stratified by the target variable. Four models were implemented:
    1. Logistic Regression (LR)
    2. Random Forest (RF)
    3. XGBoost (XGB)
    4. Stacking Ensemble (LR, RF, XGB as base learners; Logistic Regression as meta-learner)
       Hyperparameters for LR, RF, and XGB were tuned using GridSearchCV with 5-fold cross-validation, optimizing for ROC-AUC.
-
- **Evaluation:** Model performance was primarily assessed using ROC-AUC and Accuracy on the test set. Classification reports, confusion matrices, ROC curves, and calibration curves were also analyzed.

3. Results and Discussion

All models demonstrated strong predictive capabilities. The performance metrics on the test set are summarized below:

Model	ROC-AUC	Accuracy
Logistic Regression	0.924	0.8504
Random Forest	0.921	0.8451
XGBoost	0.925	0.8490
Stacking Ensemble	**0.925**	**0.8490**

The Stacking ensemble and XGBoost models yielded the best ROC-AUC scores (0.9249 and 0.9248 respectively, rounded to 0.925). The confusion matrix for the best performing model (XGBoost, due to slightly simpler interpretation than stacking for feature importance and being a component of the best stacker) showed a good balance in predicting both depressed and non-depressed students, though with some misclassifications inherent in complex real-world data.

Random Forest feature importance analysis highlighted 'Suicidal Thoughts Binary' as the most influential predictor, followed by 'Academic Pressure', 'Financial Stress', 'Age', and 'Work/Study Hours'. This aligns with existing literature on risk factors for depression. Calibration curves indicated that the models were reasonably well-calibrated, with XGBoost showing strong reliability.

4. Conclusion and Future Work

This study successfully developed machine learning models capable of predicting student depression with high accuracy and ROC-AUC scores exceeding 0.92. The Stacking ensemble demonstrated the most robust performance. Key predictors identified, such as suicidal ideation and academic stress, offer actionable insights for campus mental health initiatives.

Future work could involve:

- Incorporating more diverse data sources (e.g., academic records, social media activity with consent).
- Exploring advanced deep learning architectures.
- Developing a real-time deployment strategy for proactive student support.

This research demonstrates the significant potential of machine learning to contribute to student well-being by enabling early identification and intervention for depression.