Advanced Machine Learning Models for Data Analytics

Assignment **1**
Professor: **Dr. Ruchika Arora**
Submitted By: **Rohit kumar** (Mtech_AI-**25901334**)

Student Depression Prediction using Stacking Ensemble Models

Question

Implement a **stacking ensemble model** to predict **student depression** using a dataset containing demographic, academic, lifestyle, and psychological features. Use a **Decision Tree** and **Logistic Regression** as base models, with **Logistic Regression** as the final estimator. Evaluate the model performance and interpret the results.

Objective

The goal is to implement a **stacking ensemble model** to predict **student depression** based on various personal, academic, and lifestyle factors. The ensemble combines a **Decision Tree** and **Logistic Regression** as base models, with **Logistic Regression** as the final estimator. The model is trained, evaluated, and interpreted to understand key factors contributing to student depression.

Dataset Description

The **Student Depression Dataset** contains the following fields:

- **Demographics:** ID, Gender, Age, City, Profession, Degree
- **Academic & Work Pressure:** Academic Pressure, Work Pressure, CGPA, Study Satisfaction, Job Satisfaction
- **Lifestyle & Health:** Sleep Duration, Dietary Habits, Work/Study Hours, Financial Stress

- **Psychological Factors:** Suicidal Thoughts, Family History of Mental Illness
- **Target Variable:** Depression (0 = No, 1 = Yes)

The dataset contains both **numeric and categorical features** and some missing values, which were handled during preprocessing.

Methodology

Step 1: Data Preprocessing

- Checked for missing values and inconsistencies.
- Categorical features (Gender, City, Degree, etc.) were encoded into numeric values.
- Numeric features were standardized to bring them to a comparable scale.

Step 2: Model Building

- **Base Models:** Decision Tree and Logistic Regression.
- **Final Estimator:** Logistic Regression in a stacking ensemble.
- The dataset was split into **training (80%)** and **testing (20%)** sets.

Step 3: Model Evaluation

- **Metrics Used:** Accuracy, Precision, Recall, F1-score.
- **Confusion Matrix** visualized correct and incorrect predictions.
- **ROC Curve & AUC** analyzed the model's discrimination ability.

Step 4: Feature Importance

- Decision Tree base model used to extract **feature importance**.
- Top predictive features identified for depression risk.

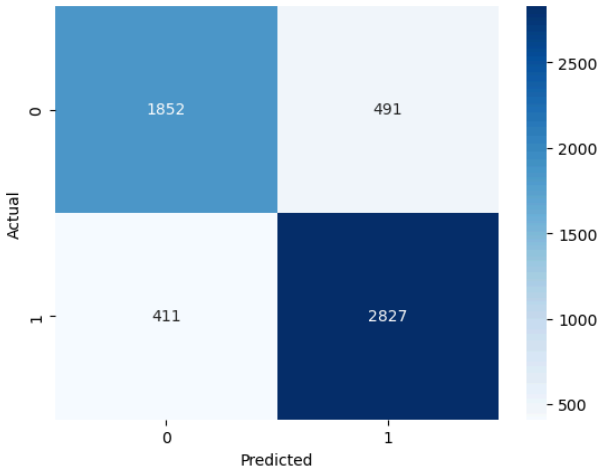
Results and Analysis

Model Performance

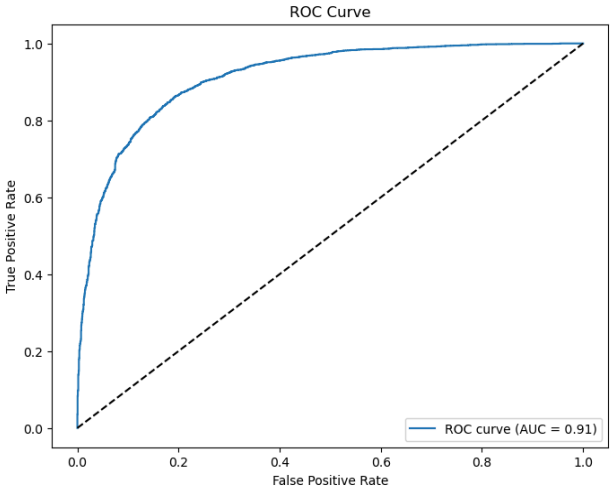
- **Accuracy:** 83.8%
- **Classification Report:**
    - Class 0 (No Depression): Precision = 0.82, Recall = 0.79, F1-score = 0.80
    - Class 1 (Depression): Precision = 0.85, Recall = 0.87, F1-score = 0.86
- **ROC Curve:** AUC = 0.90 (excellent discrimination)

Feature Importance (Top Contributors)

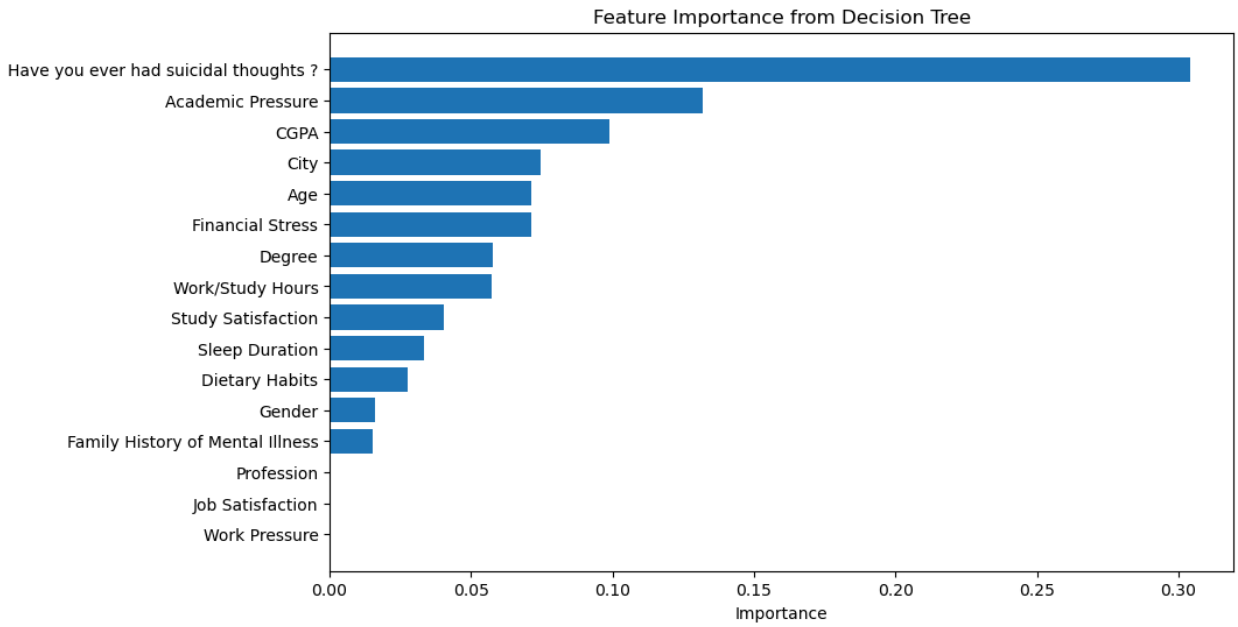
- Suicidal Thoughts: 30% – the most significant predictor of student depression.
- Academic Pressure: 13% – high academic stress strongly impacts depression risk.
- CGPA: 10% – lower or fluctuating academic performance contributes to depression.
- City: 7% – location has a moderate influence on depression likelihood.
- Age: 7% – age plays a moderate role in depression risk.
- Financial Stress: 7% – financial pressure is a notable factor.
- Degree: 6% – type of degree has some influence on mental health.
- Work/Study Hours: 6% – excessive workload moderately affects depression probability.



Confusion Matrix



Roc Curve



Feature Importance from Decision Tree

Observation:

- Suicidal thoughts and academic pressure are the **strongest indicators** of depression.
- Lifestyle and demographic factors like sleep, diet, and city also contribute but are less influential.

Insights and Recommendations

- **Early Intervention:** Students with high academic pressure or suicidal thoughts should be prioritized for counseling.
- **Monitoring CGPA & Workload:** Tracking students' CGPA and work/study hours can help identify at-risk students early.
- **Awareness Programs:** Organize mental health awareness and stress management workshops.
- **Targeted Support:** Personalized interventions for high-risk students based on top contributing features.

Conclusion

The stacking ensemble model effectively predicts student depression with **high accuracy**. Key features such as **suicidal thoughts, academic pressure, and CGPA** are strong predictors. This analysis provides actionable insights for universities and counselors to **support student mental health**.