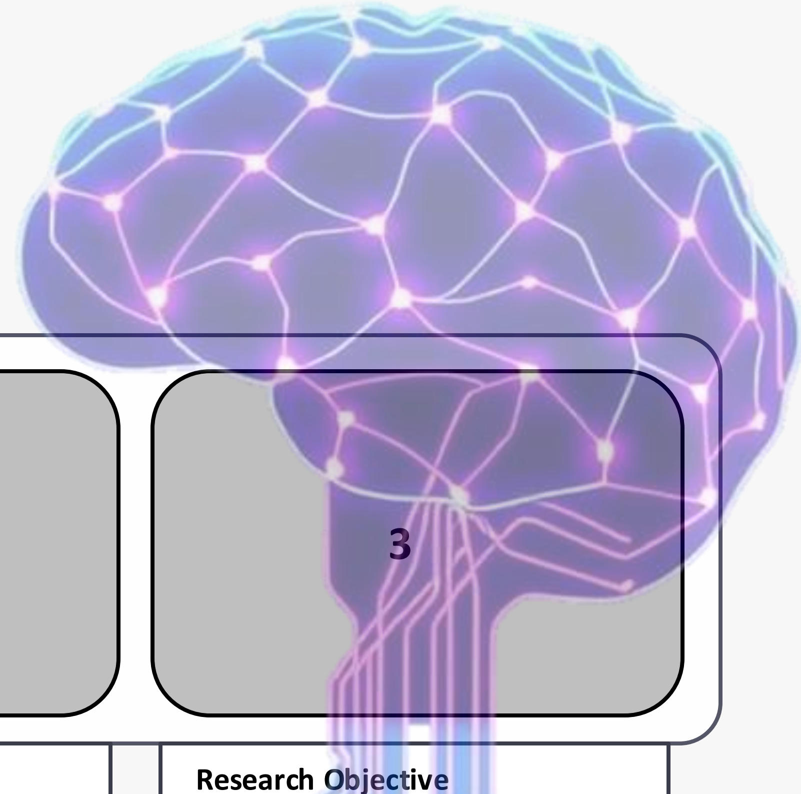
AI-Powered Mental Health Prediction Using Machine Learning

Presented at 4th International Conference on Advanced Computing Technologies & Applications (ICACTA 2025)

Iqra Fatima Umang, Aafreen
Department of Computer Applications
Invertis University, Bareilly
Emails: syedaaiqraa1308@gmail.com, afreenkhan20202@gmail.com

Mentored by: Mr. Nazmul Arefin
Department of Computer Applications, Invertis University

Introduction & Motivation



1

2

3

Rising Need for Mental Health Solutions

The increasing prevalence of depression among students necessitates innovative, data-driven approaches to mental health assessment. Limited access to professionals highlights the need for scalable alternatives.

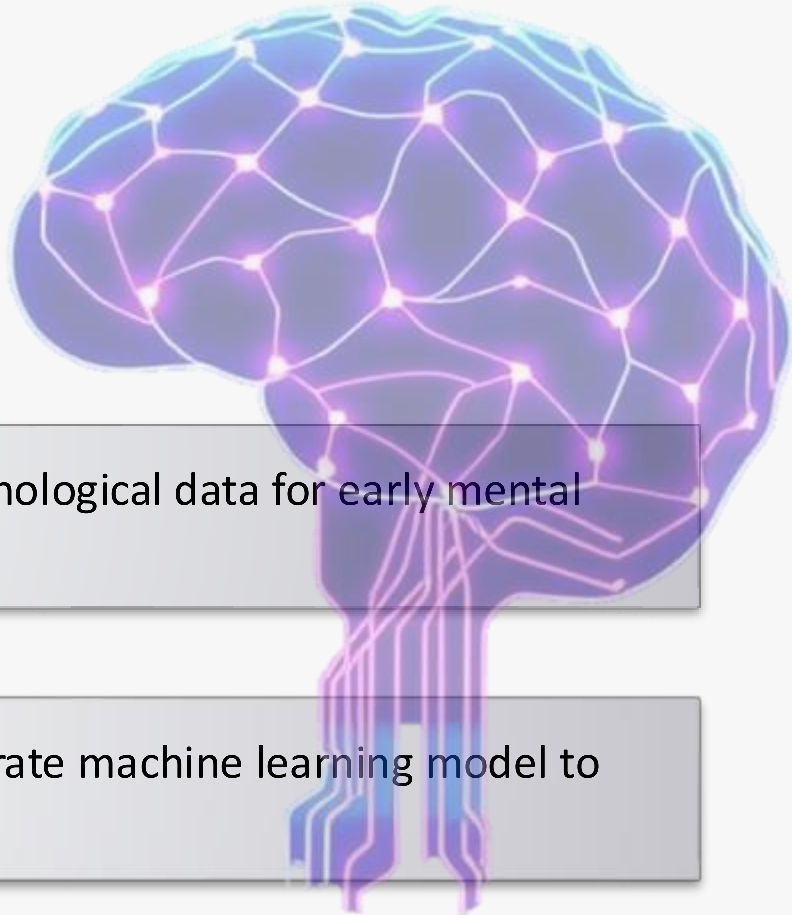
AI and Machine Learning Applications

Recent advancements enable the use of machine learning to detect early mental distress by analyzing behavioral and psychological data, offering new potential for intervention in academic settings.

Research Objective

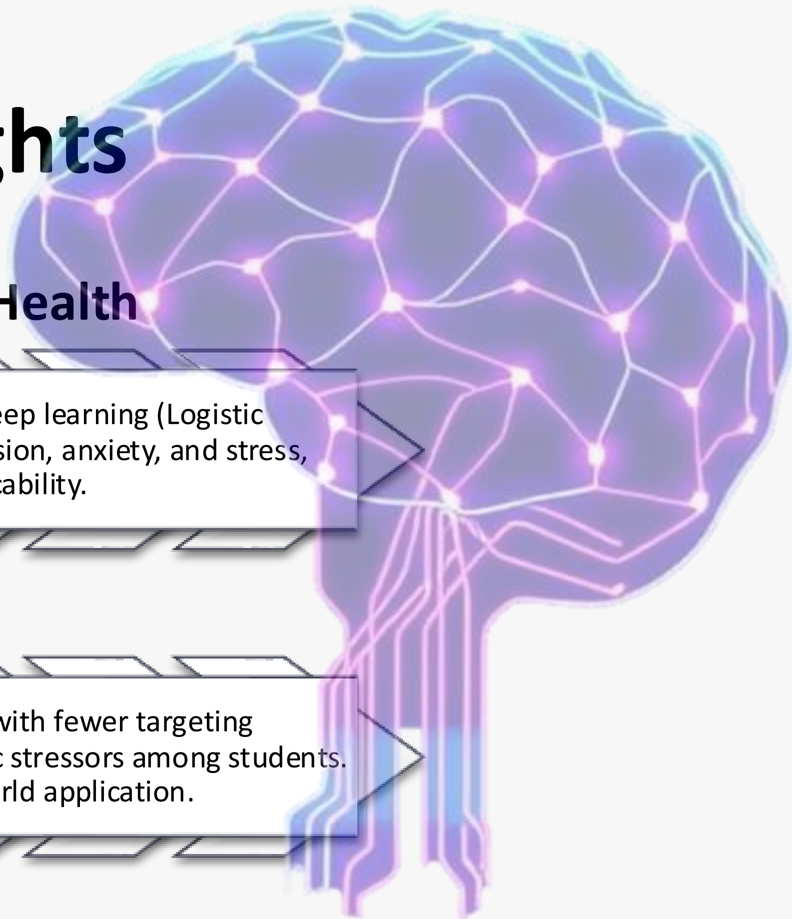
This study aims to develop an interpretable and accurate predictive model for student depression using machine learning, supporting early warning systems in educational institutions.

Objective



- Utilize behavioral, academic, and psychological data for early mental health risk detection.
- Develop an interpretable and accurate machine learning model to predict depression in students.
- Enable real-world usability through a simple, user-friendly interface for non-technical users.

Literature Review: Key Insights



[1] ML Approaches to Mental Health

Prior research demonstrates effective use of ML and deep learning (Logistic Regression, SVM, Random Forest) for detecting depression, anxiety, and stress, with high reported accuracy and practical clinical applicability.

[2] Academic Contexts & Gaps

Many existing studies focus on general mental health, with fewer targeting depression caused by academic and socio-demographic stressors among students. GUI deployment remains underrepresented for real-world application.

[3] Current Study's Innovation

This research narrows the gap by focusing on student-specific depression, utilizing public datasets, comprehensive feature analysis, and GUI deployment to enhance usability.

Proposed Methodology: System Overview

Stepwise Approach

The methodology follows a pipeline: data collection, preprocessing, exploratory data analysis, model training (Logistic Regression, SVM, Random Forest), hyperparameter tuning, model comparison, evaluation, and GUI deployment.

GUI for Practical Application

A Tkinter-based GUI is built to facilitate user-friendly depression prediction, bridging technical models and real-world usability for students.

Data Collection and Preprocessing

Dataset Details

Dataset from Kaggle has 502 student survey responses with 11 features: demographic (gender, age), behavioral, academic (study satisfaction, pressure), and psychological measures (financial stress, suicidal thoughts).

Preprocessing Steps

Data cleaning involved handling missing values, encoding categorical, standardizing categorical data and ordinal variables, and feature scaling to improve consistency and ensure compatibility with ML algorithms.

Exploratory Data Analysis & Target Analysis

Correlation and Key Variables

EDA revealed strong positive correlations between depression and academic pressure, financial stress, and suicidal thoughts. Study satisfaction showed a negative correlation, suggesting a protective effect.

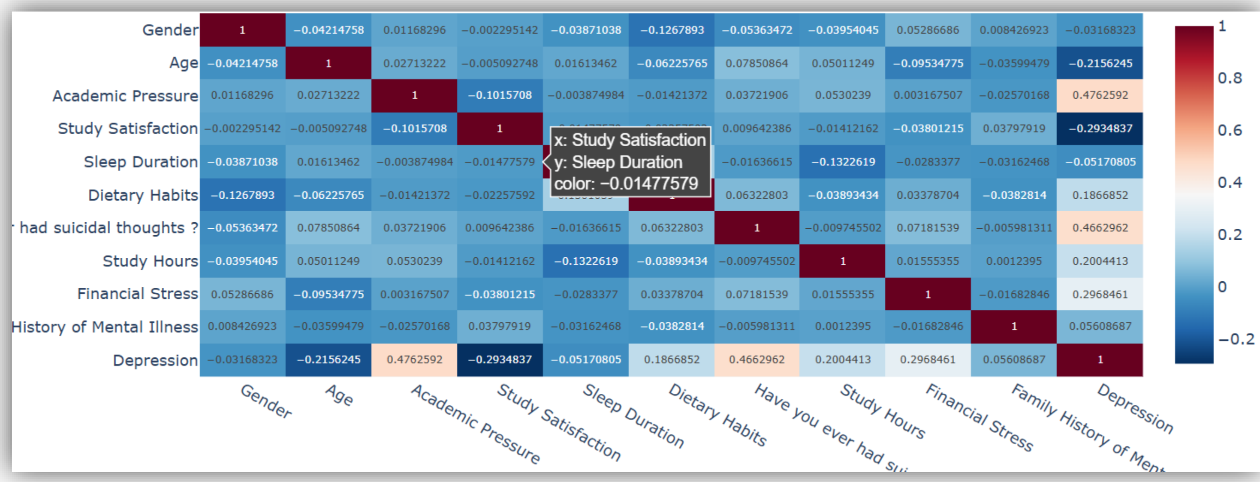


Figure 1: Correlation Heatmap of Key Features
This heatmap visualizes the pairwise correlation between features in the dataset. Strong correlations with depression are observed for academic pressure, financial stress, and suicidal thoughts.

Class Distribution & Task Framing

The target variable (depression) is binary and well-balanced. Features were selected based on relevance and correlation, framing the problem as a supervised binary classification task.

Machine Learning Models: Overview & Algorithms

Logistic Regression
Achieved highest accuracy (98.02%), perfect recall for detecting depression, and maintained strong interpretability via feature coefficients analysis.

Support Vector Machine
Strong performance (97.03% accuracy), with robust recall. Utilized kernel-based classification and decision function for optimal separation.

Random Forest Classifier
Offered reasonable results (92.08% accuracy), able to model nonlinear relationships, but had higher false positives.

Model Evaluation & Comparison

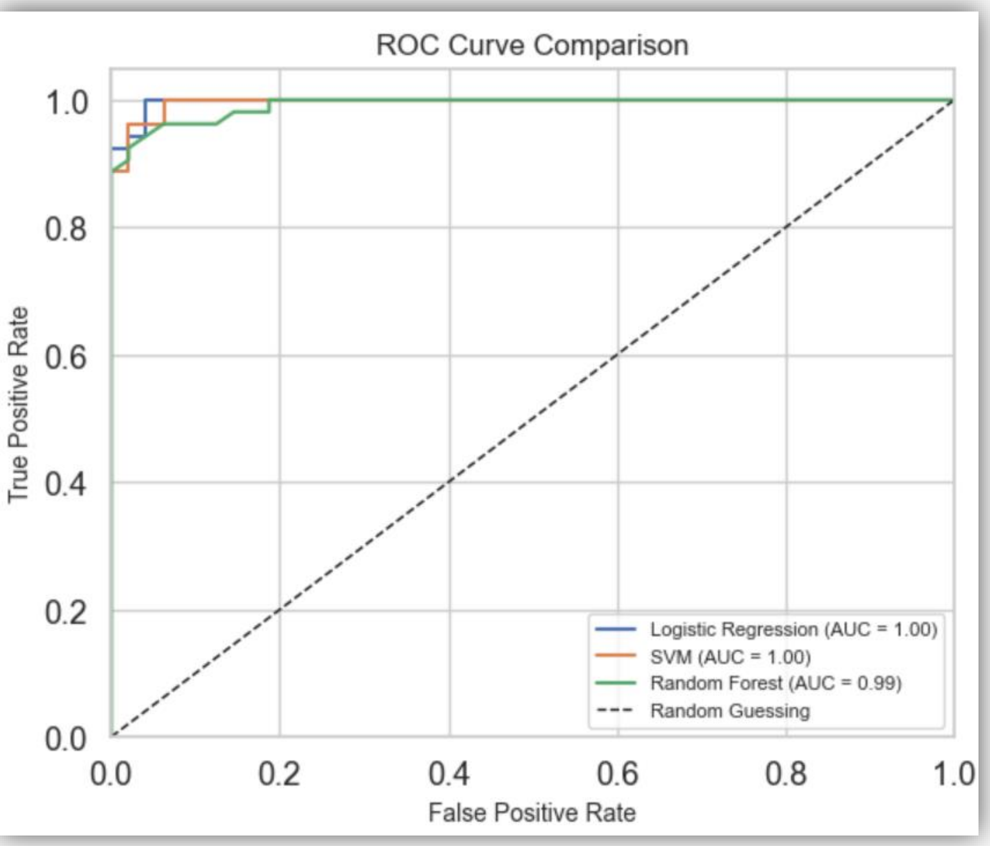


Figure 2: Receiver Operating Characteristic (ROC) Curve Comparison

Metric-Based Assessment

Evaluated all models using accuracy, precision, recall, F1-score, and confusion matrices. Logistic Regression excelled in all metrics, demonstrating reliability.

ROC Curve Analysis

ROC curves and AUC confirmed strong discriminative power for all models, with Logistic Regression and SVM leading. Visual comparisons further supported these findings.

Feature Importance & Model Interpretation

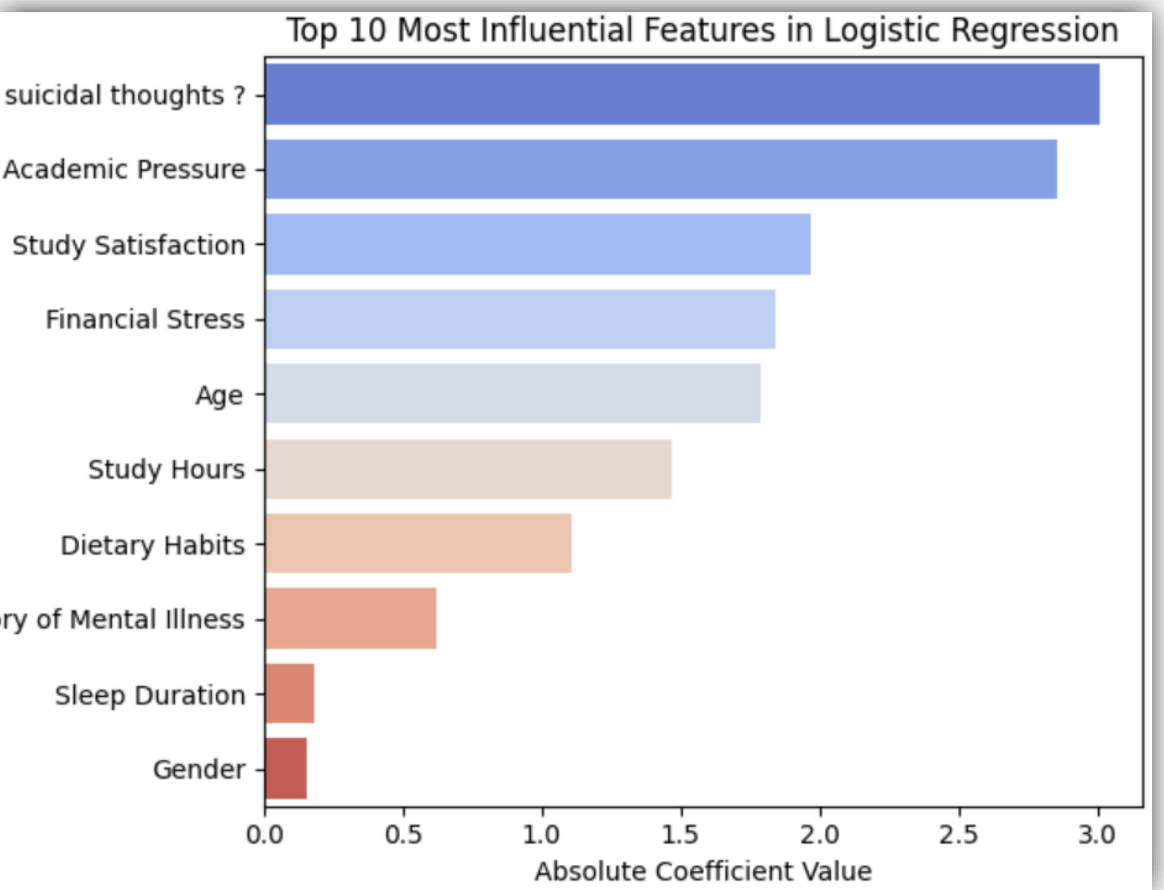


Figure 3: Top 10 Most Influential Features in the Logistic Regression Model Based on Absolute Coefficients

Logistic Regression Coefficient Analysis

Feature importance derived from logistic regression coefficients revealed suicidal thoughts, academic pressure, study satisfaction, and financial stress as major indicators.

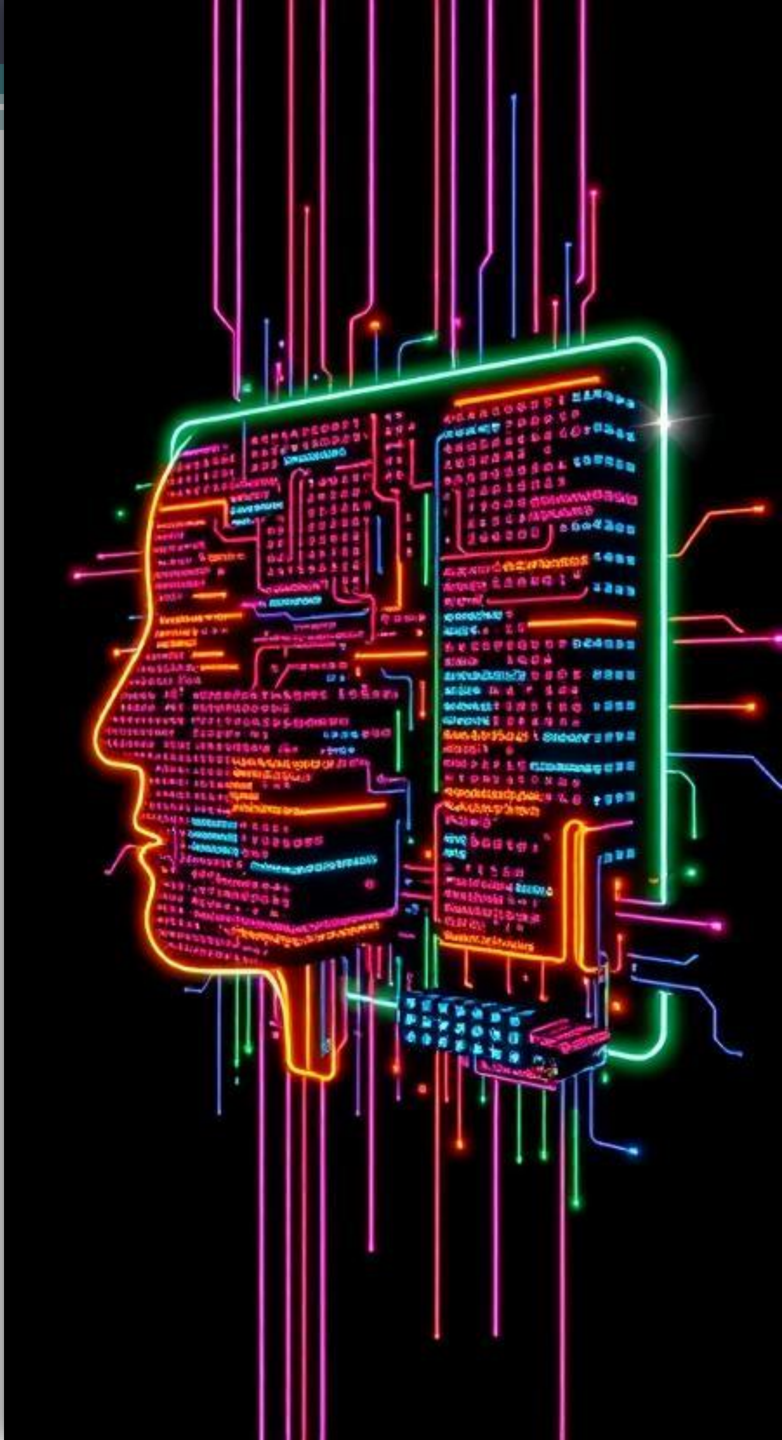
Implications for Interventions

Interpretable models provide actionable insights, allowing educators and counselors to target interventions to high-impact factors affecting student mental health

Hyperparameter Tuning & Regularization

Tuning and Optimization
Grid search with cross-validation identified optimal hyperparameters for Logistic Regression (C=10, L2 penalty). L1 regularization (Lasso) was also explored for model sparsity and interpretability.

Results Consistency
All model variants achieved identical top accuracy (98.02%), indicating robust dataset and effective preprocessing, with regularization enhancing interpretability.



Deployment: Model Saving & GUI Integration

Model Saving for Real-World Use

The finalized Logistic Regression model and scaler were saved using joblib for consistent future predictions, enabling transferability across environments.

GUI Development with Tkinter

A user-friendly graphical interface allows non-technical users to input relevant features and receive instant depression risk prediction, supporting proactive screening.

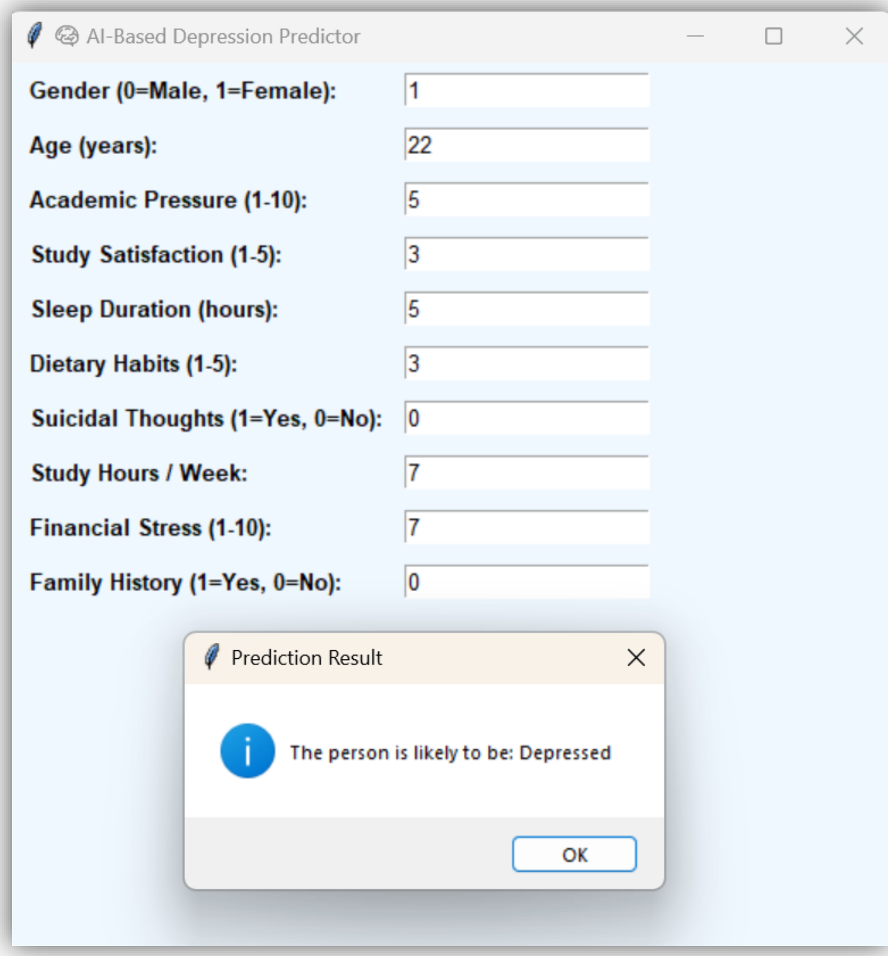


Figure 4: Screenshot of the Depression Predictor GUI built using Tkinter, showcasing user input fields and prediction output.

Results & Discussion

Performance Overview

Logistic Regression outperformed other classifiers with the fewest false positives and consistently high accuracy, precision, recall, and F1-score across cross-validation and test sets.

Usability Evaluation

The GUI was tested with various inputs, reliably delivering accurate predictions and confirming retention of model capabilities post-deployment.

Conclusion & Future Scope

Conclusion

- Developed a reliable and user-friendly AI model to predict depression risk in students.
- Logistic Regression achieved the highest accuracy (98.01%), proving effective for classification.
- The model highlights academic pressure, financial stress, and low study satisfaction as key indicators.
- A simple GUI was created to enable easy, real-time mental health screening.

Future Scope

- Expand testing with larger, diverse, and clinically validated datasets.
- Integrate the system into mobile and web platforms for wider accessibility.
- Enhance accuracy using advanced models like ensemble methods or deep learning.
- Address ethical concerns with robust data privacy, transparency, and responsible AI practices.

Thank you for your time and attention.