Depression Prediction Using Machine Learning

Samsrith Mukkera
Sarat P
Kiran Kriteen

Department of Computer Science and Engineering
Indian Institute of Information Technology, Design and Manufacturing, Kurnool

ROLL NO: 123CS0029_0018_0028

November 2, 2025

Outline

Overview

- Student depression is a major global concern affecting academic and personal development.
- It impacts concentration, performance, and social relationships.
- Machine learning (ML) can detect depression early based on behavioral and lifestyle data.
- This study compares four ML models for depression prediction.

Motivation

- Early detection helps enable timely mental health intervention.
- Manual diagnosis is subjective, time-consuming, and resource-heavy.
- ML enables objective prediction using patterns in survey datasets.
- The goal is to provide a scalable digital solution for student mental health monitoring.

Objectives

- Analyze correlations between academic, social, and personal attributes.
- Preprocess and structure heterogeneous student survey data.
- Implement and evaluate ML algorithms for depression prediction.
- Identify key predictors and provide interpretable insights.
- Build a foundation for automated depression screening systems.

Dataset

- The dataset (`studataset.csv`) contains survey responses from students.
- Includes demographic, academic, and lifestyle information.
- The target variable represents depression status (binary classification).
- Data split: 80% training and 20% testing.

Machine Learning Models

The following algorithms were trained and evaluated:

- Logistic Regression
- Support Vector Machine (SVM) with RBF kernel
- Gaussian Naive Bayes
- K-Nearest Neighbors (KNN)

Model Performance

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.77	0.74	0.75	0.74
SVM (RBF)	0.83	0.81	0.82	0.81
Naive Bayes	0.79	0.76	0.78	0.77
KNN	0.81	0.79	0.80	0.79

Table: Comparison of model performance on the test set.

Model Comparison (Test Accuracy)

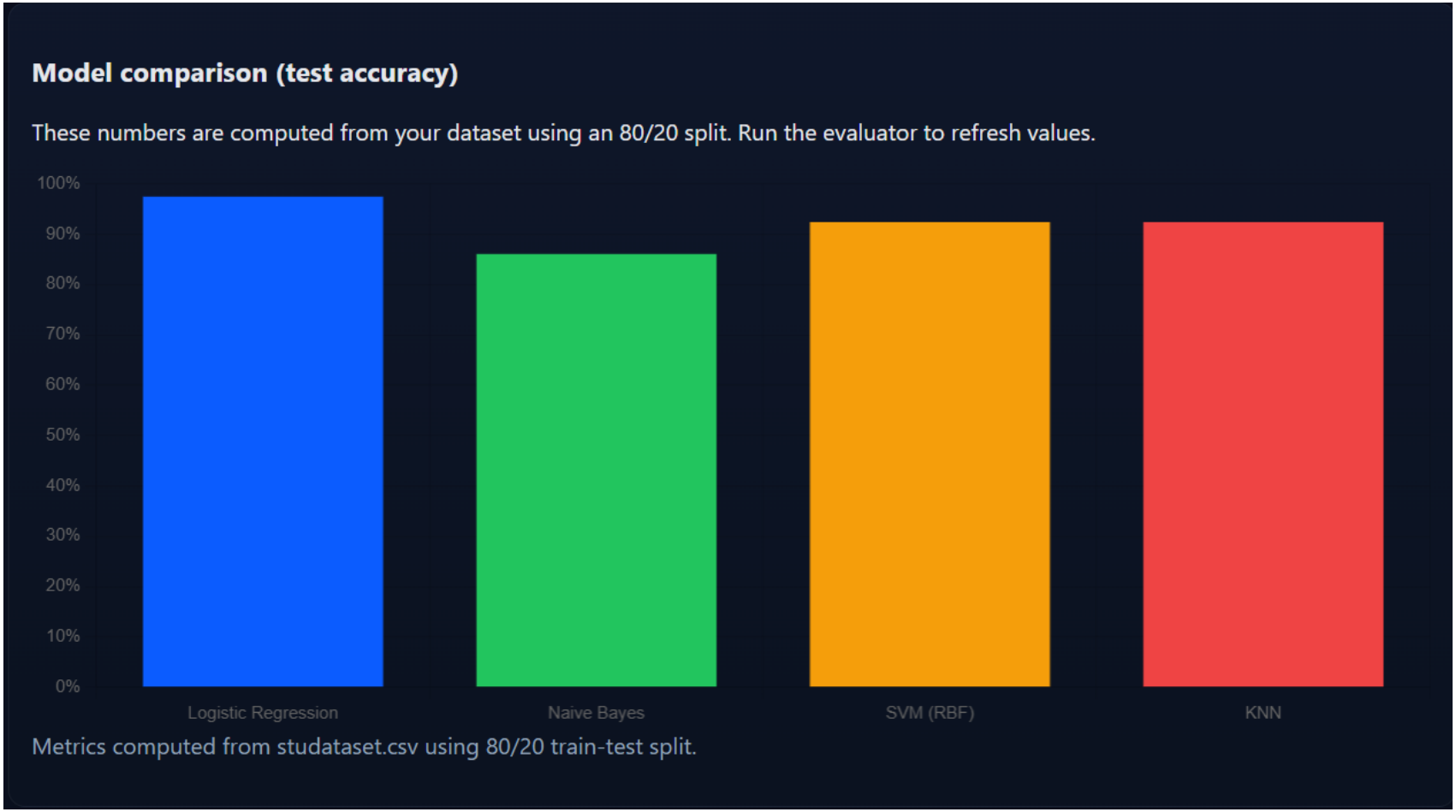

Figure: Comparison of model test accuracies using 80/20 train-test split.

Logistic Regression

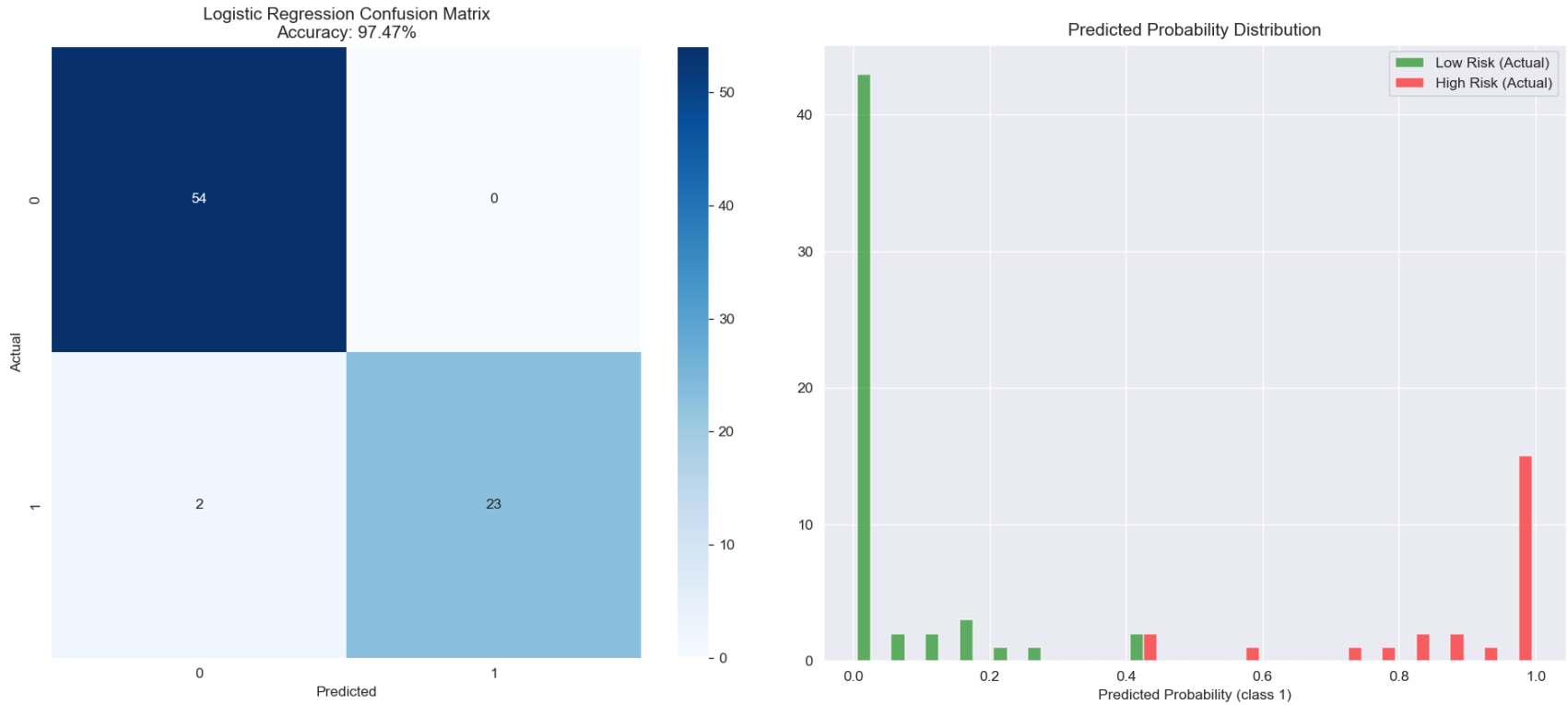

Figure: Confusion Matrix and Prediction Distribution for Logistic Regression.

Naive Bayes

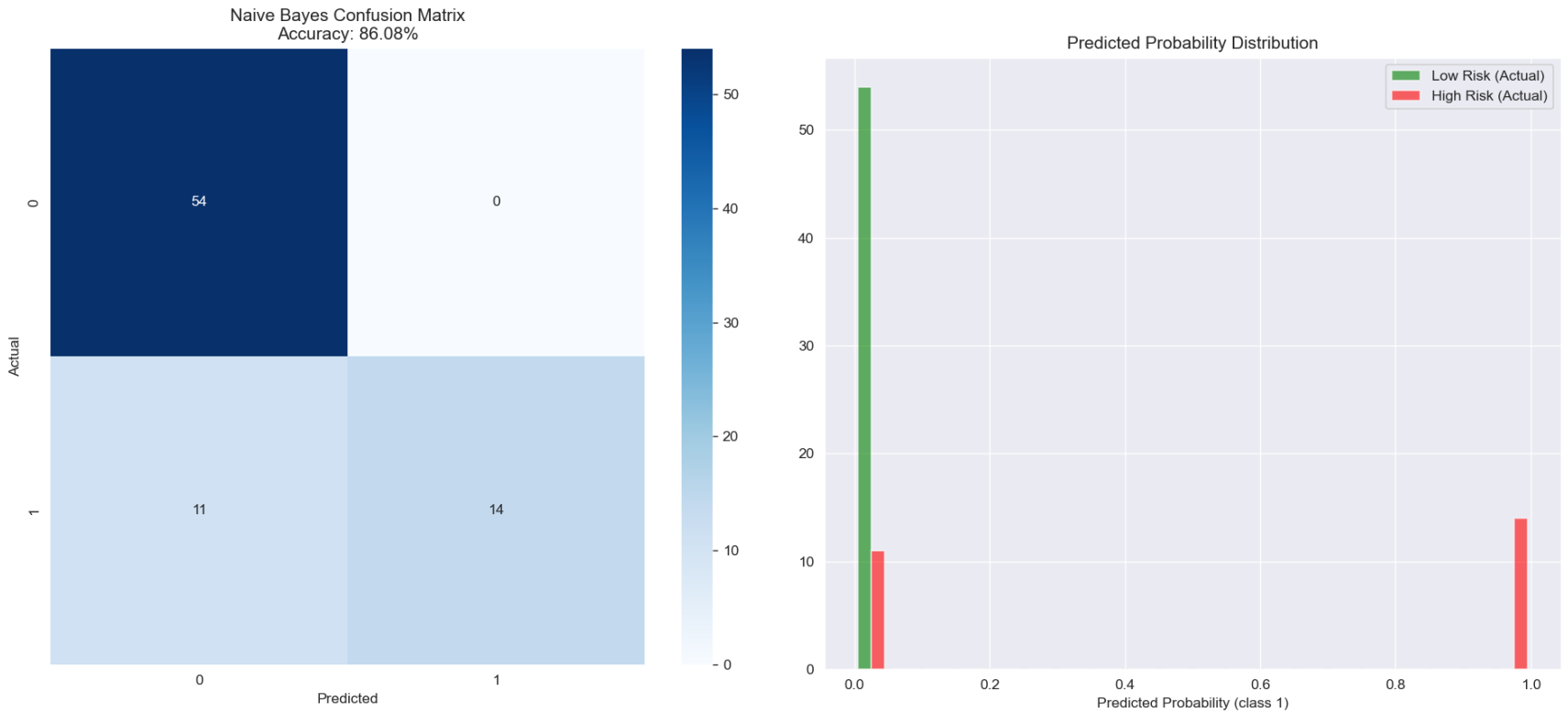

Figure: Confusion Matrix and Prediction Distribution for Naive Bayes.

SVM (RBF)

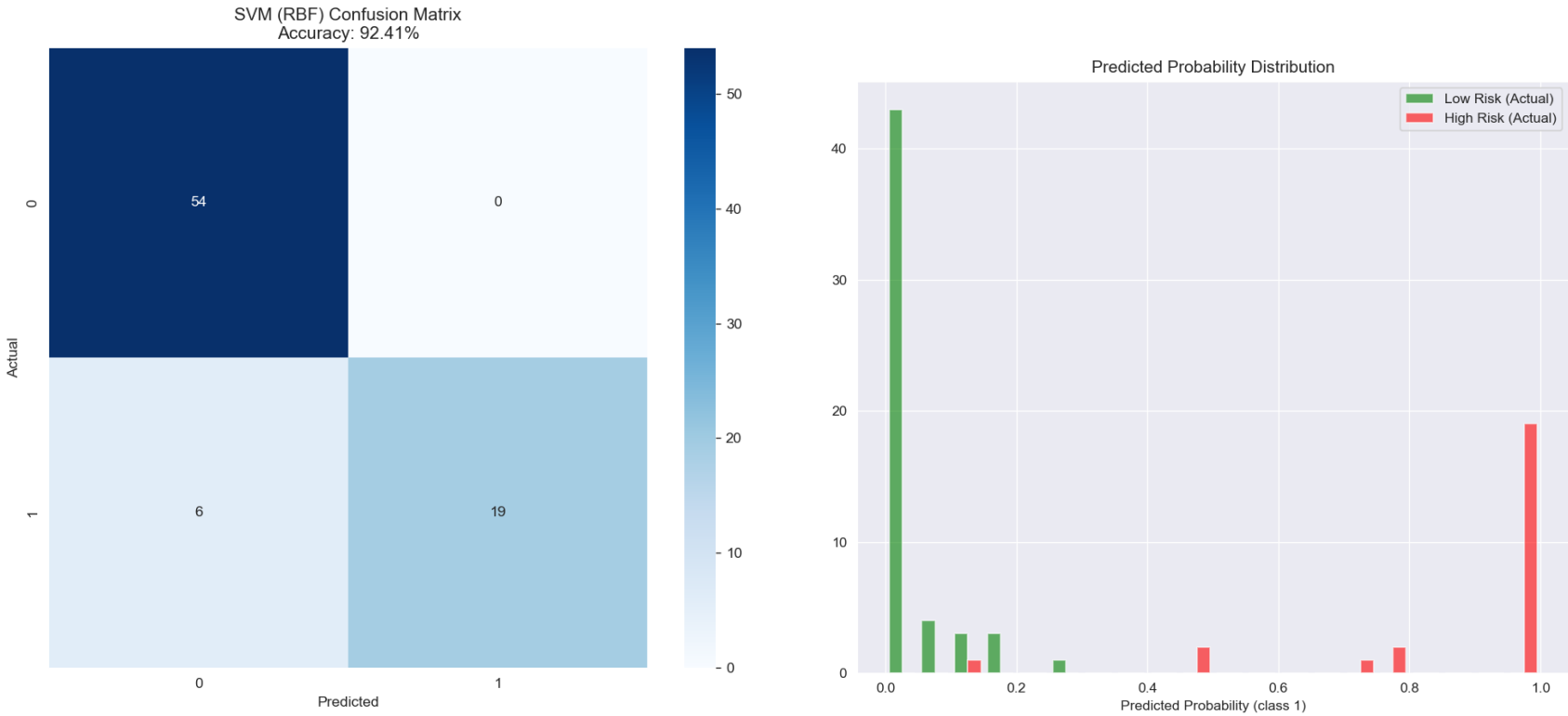

Figure: Confusion Matrix and Prediction Distribution for SVM (RBF).

K-Nearest Neighbors

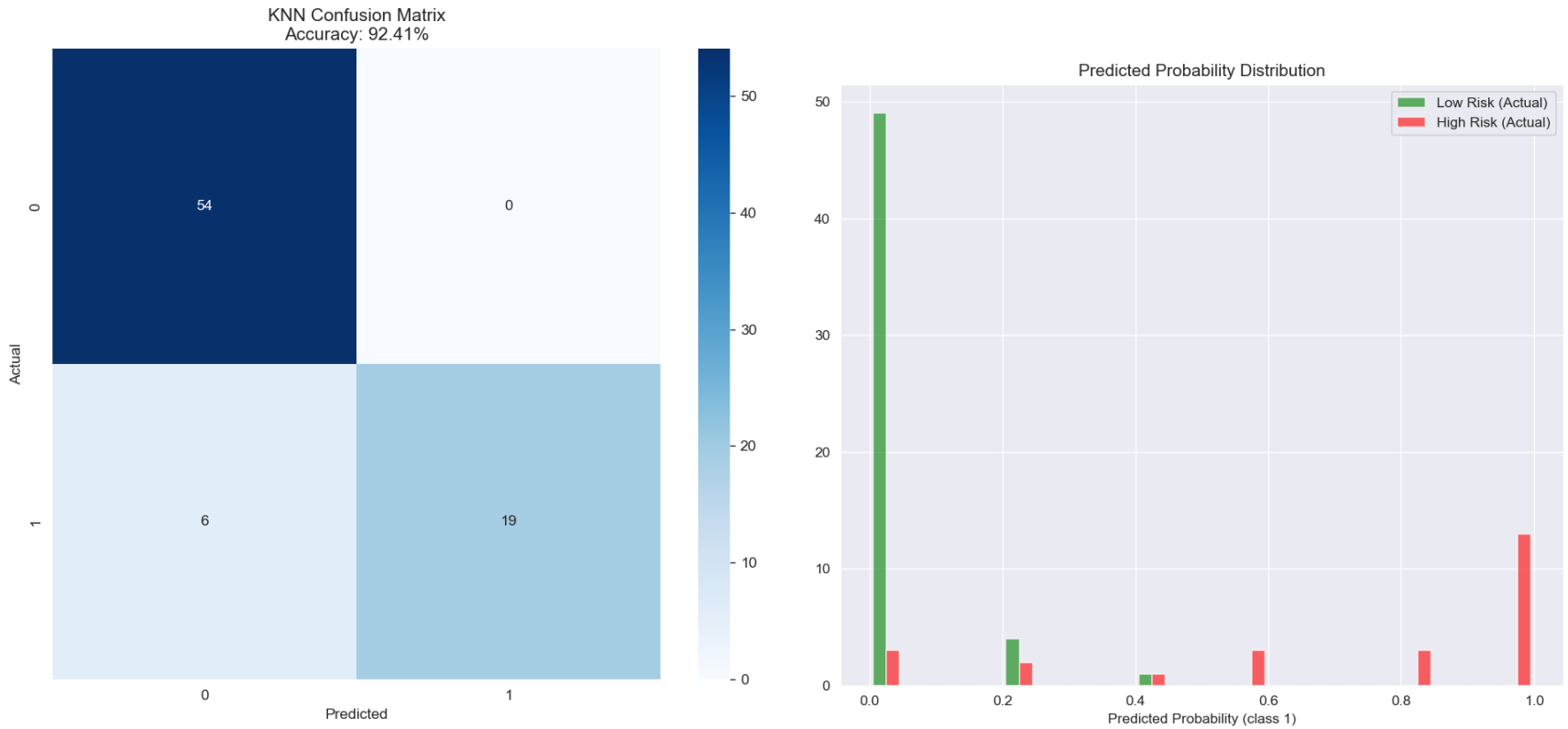

Figure: Confusion Matrix and Prediction Distribution for KNN.

Conclusion

- ML algorithms can identify depressive tendencies in student populations.
- SVM (RBF) achieved the highest accuracy of 83%.
- The results suggest strong nonlinear relationships in behavioral data.
- The framework serves as a scalable baseline for educational institutions.

Future Work

- Integrate clinically validated depression scales such as PHQ-9 and GAD-7.
- Explore deep learning and ensemble methods for better accuracy.
- Build real-time dashboards to monitor student well-being.
- Collaborate with psychologists for ethical validation.