

Evaluation Summary

In this study, Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest classifiers were used to predict student exam pass/fail outcomes based on hours studied, attendance, and previous scores.

Logistic Regression provided a strong baseline performance with good accuracy and F1-score while remaining highly interpretable. Its linear nature makes it easy to understand how each feature influences the probability of passing the exam.

KNN demonstrated the ability to capture non-linear patterns in the data but was sensitive to feature scaling and the choice of the number of neighbors. Although its performance was competitive, the model lacked interpretability since predictions depend on distances to neighboring data points.

Random Forest achieved the highest overall accuracy and F1-score by combining multiple decision trees, reducing variance and improving generalization. However, its ensemble structure makes it less interpretable compared to simpler models.

Overall, the results show that increasing model complexity improves predictive performance but reduces interpretability. Logistic Regression is preferable when transparency is required, while Random Forest is suitable when predictive accuracy is the primary goal.