Student Performance Prediction Report

K Nearest Neighbors Model (KNN)

Reasoning: Implemented KNN for its simplicity and effectiveness in classification tasks. KNN relies on the principle that similar data points tend to belong to the same class, making it suitable for predicting student performance based on similar instances from the training data.

Evaluation Metrics:

K-Value	F1 Score	Accuracy
3	0.33730158730158727	0.35443037974683544
5	0.37081550566479177	0.379746835443038
7	0.3743645606390704	0.3924050632911392
9	0.3773991838507968	0.379746835443038
11	0.4018456966832536	0.4177215189873418

Support Vector Machine Model (SVM)

Reasoning: SVM was chosen due to its ability to handle high-dimensional data and its effectiveness in capturing complex relationships between features. It works well for classification tasks with clear margins of separation between classes, which can be beneficial for predicting student performance.

Evaluation Metrics:

C-Value	F1 Score	Accuracy
0.1	0.23846022824050417	0.35443037974683544
1	0.405396934602301	0.4936708860759494
10	0.42238741650506356	0.4936708860759494
100	0.4197012138188609	0.4810126582278481

Gaussian Naive Bayes Model

Reasoning: Gaussian Naive Bayes was selected for its simplicity and speed, especially for small to medium-sized datasets. Despite its "naive" assumption of feature independence, it often performs surprisingly well in practice and can provide a baseline performance for comparison with other models.

Findings

Model	Accuracy
KNN	0.4177215189873418
SVM	0.6234177215189873
GaussianNB	0.5569620253164557

Conclusions

In this report, we explored the application of different machine learning models for predicting student performance. Through experimentation with KNN, SVM, and Gaussian Naive Bayes, we observed varying levels of performance across different evaluation metrics. The SVM model achieved the highest accuracy among the tested models, followed by KNN and Gaussian Naive Bayes. Also found that when adjusting K-values for KNN and C values for SVM, the accuracy varied. Also various kernel types affected the SVM accuracy score.

Limitations

The limitations of this study include potential challenges in generalizing the findings beyond the specific dataset used. Additionally, the accuracy of the models heavily relies on the selected features, which may overlook important predictors of student performance. Moreover, SVM, particularly with high-dimensional data, can pose computational challenges and require significant resources for training and inference.

Future Works

Future research endeavors could focus on exploring advanced feature engineering techniques to identify more informative features or combinations of features. Additionally, optimizing SVM through fine-tuning hyperparameters and exploring different kernel functions could enhance its performance and scalability, especially for larger datasets.