Classifying iris flower data using Naive Bayes, Random Forest, and KNN (k-nearest neighbors) algorithms.

Sengthong Xayavong

Abstract: This study evaluates the performance of three machine learning algorithms—Naive Bayes, Random Forest, and k-Nearest Neighbors (k-NN)—for the classification of Iris flower species using the classic Iris dataset. The dataset comprises 150 samples with four numerical features: sepal length, sepal width, petal length, and petal width. Each algorithm was implemented and tested using a train-test split methodology, with accuracy, precision, recall, and F1-score as evaluation metrics. Experimental results demonstrate that Random Forest achieved the highest classification accuracy at 96.7%, outperforming both Naive Bayes (93.3%) and k-NN (95.6%). The findings indicate that ensemble methods like Random Forest are particularly effective for this multiclass classification problem due to their robustness against overfitting and ability to handle feature interactions. However, k-NN also showed competitive performance, especially with optimal k-value selection. This comparative analysis provides practical insights into algorithm selection for similar low-dimensional classification tasks and underscores the importance of model evaluation beyond simple accuracy.

Keywords: Iris classification, Naive Bayes, Random Forest, K-Nearest Neighbors