A screenshot of a graph

AI-generated content may be incorrect.

Eight classification models were evaluated on the Breast Cancer Wisconsin dataset to assess predictive performance. Logistic Regression achieved the highest accuracy at 97.10%, followed closely by Naïve Bayes (96.49%) and Kernel SVM (95.91%). These models demonstrated strong generalization with minimal false predictions.

Naïve Bayes, despite its assumption of feature independence, performed exceptionally well, yielding a balanced confusion matrix with only 3 false positives and 3 false negatives. Tree-based models such as Decision Tree, Random Forest, and XGBoost showed competitive results (accuracy ~94.7%), though they exhibited slightly higher false negative rates.

KNN and Linear SVM both reached 95.32% accuracy, indicating solid performance with low misclassification. Overall, Logistic Regression and Naïve Bayes emerged as top contenders, offering high accuracy and reliable classification with minimal tuning.

These results underscore the importance of evaluating multiple models, as simpler algorithms like Naïve Bayes can rival more complex ones when applied to well-preprocessed data.

Gitgub link: <https://github.com/shannaka/Breast-Cancer-Wisconsin>