

Supervised Learning Model Evaluation: Type 2 Diabetes (T2D) Detection

The table below summarizes the performance of six baseline Machine Learning (ML) models applied to the UCI diabetes dataset containing features of the Early-Stage Diabetes symptoms. Models were evaluated on the same test set using standard classification metrics.

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Logistic Regression	0.923	0.920	0.920	0.920	0.977
K-Nearest Neighbors	0.894	0.870	0.900	0.880	0.977
Decision Tree	0.951	0.960	0.950	0.950	0.965
Random Forest	0.990	0.990	0.990	0.990	1.000
Naive Bayes	0.913	0.910	0.910	0.910	0.961
Support Vector Machine	0.990	0.990	0.990	0.990	0.996

Observations

- Random Forest and SVM achieved the best performance, with near-perfect scores across all metrics.
- Logistic Regression and Naive Bayes showed solid balance between precision and recall.
- KNN lagged slightly in F1 Score due to trade-offs between sensitivity and specificity.