**Breast Cancer Classification Model Results**

**Introduction**

This report presents the performance evaluation of various machine learning models trained on the **Breast Cancer Wisconsin Dataset**. The dataset consists of **10 numerical features** describing cell characteristics and a **binary class label**, where:

* **Benign tumor** → Label = **2** (converted to **0**)
* **Malignant tumor** → Label = **4** (converted to **1**)

The objective of this classification task is to predict whether a tumor is benign or malignant based on the given features. The dataset was split into **75% training data** and **25% test data** for model evaluation.

Each model's performance is evaluated based on:

1. **Accuracy** – The percentage of correctly classified samples.
2. **Confusion Matrix** – A summary of correct and incorrect predictions.
3. **Precision, Recall, and F1-score** – Additional evaluation metrics provided in the classification report.

**Model Performance Results**

The table below summarizes the accuracy and confusion matrix results for each classification model:

| **Model** | **Accuracy** | **Confusion Matrix** |
| --- | --- | --- |
| Logistic Regression | 0.959 | [[106 5] [ 2 58]] |
| KNN (k=5) | 0.953 | [[106 5] [ 3 57]] |
| Linear SVM | 0.959 | [[106 5] [ 2 58]] |
| Kernel SVM (RBF) | 0.965 | [[106 5] [ 1 59]] |
| Naïve Bayes | 0.959 | [[106 5] [ 2 58]] |
| Decision Tree | 0.942 | [[105 6] [ 4 56]] |
| Random Forest (10 estimators) | 0.953 | [[106 5] [ 3 57]] |
| XGBoost | 0.971 | [[106 5] [ 0 60]] |

**Key Observations**

**Best Performing Models**

* **XGBoost (97.1% Accuracy)** performed the best, correctly classifying **all malignant tumors** with a **recall of 100%**.
* **Kernel SVM (96.5%)** followed closely, misclassifying only **one malignant tumor** as benign.
* **Logistic Regression (95.9%)** and **Linear SVM (95.9%)** also performed well.

**Moderate Performance Models**

* **KNN (95.3%) and Random Forest (95.3%)** had similar performance, showing **slightly lower recall**.
* **Naïve Bayes (95.9%)** performed well, indicating that the dataset may conform to Naïve Bayes' assumption of **feature independence**.

**Weakest Performing Model**

* **Decision Tree (94.2%)** had the **lowest accuracy** and misclassified more tumors compared to other models.
* Decision Trees tend to **overfit** training data, making them **less generalizable** to new data.

**Conclusion**

This study shows that **XGBoost is the most effective model for breast cancer classification**, achieving the **highest accuracy (97.1%) and recall (100%)** for malignant tumors. It is followed by **Kernel SVM, Logistic Regression, and Linear SVM**.