Q3:

Model Performance Comparison on Breast Cancer Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-Score |
| SVM | 0.97 | 0.98 | 0.96 | 0.97 |
| k-NN (k=3) | 0.956 | 1.0 | 0.891 | 0.943 |
| Perceptron | 0.947 | 0.955 | 0.913 | 0.933 |
| Logistic Regression | 0.956 | 0.977 | 0.913 | 0.944 |

Based on the results from the **Breast Cancer dataset**, **SVM** stands out as the best performing model with an accuracy of 97%, precision of 98%, recall of 96%, and an F1-score of 97%. This high performance is due to SVM's ability to maximize the margin between classes, making it highly effective in distinguishing between malignant and benign tumors. In contrast, **k-NN** (with k=3) achieved an accuracy of 95.6% and perfect precision (100%), but its recall was lower at 89.1%, meaning it missed a few malignant cases. The F1-score of 94.3% reflects this trade-off between precision and recall. **Perceptron** performed well with a test accuracy of 94.7%, precision of 95.5%, and recall of 91.3%, but it showed signs of overfitting, as indicated by a perfect training accuracy of 100%. **Logistic Regression**, with an accuracy of 95.6%, precision of 97.7%, recall of 91.3%, and an F1-score of 94.4%, also provided a strong balance between precision and recall. While both Logistic Regression and k-NN performed similarly in terms of accuracy, SVM consistently outperformed them in terms of overall metrics. Therefore, **SVM** is the best model for this dataset, offering the most reliable performance in classifying breast cancer cases.