CIA – 1

 # CS3802--Machine Learning Algorithms Lab

Adithya V |
BTech CSE (IoT) - A | 21011102009

Exercise 5
Use the teleco-customer-churn dataset for the following:

1. Perform the necessary pre-processings.
2. Apply all the classification algorithms (KNN, Logisitc Regression, Naive Bayes, Decision Trees, SVM) on this dataset and print the accuracies.
3. Find which algorithm gave the best accuracy.
4. Provide a justification as to why that algorithm provided the best accuracy.
5. zip the code files and the justification file and attach the zipped folder in the submission page.

Justification:

Algorithm	Justification	Accuracy
Logistic Regression	Effective in binary classification, good balance	97.54%
K-Nearest Neighbors	Effective in cluster-based problems, sensitivity	87.36%
Naive Bayes	Assumes independence, computationally efficient	97.26%
Decision Trees	Captures complex relationships, potential overfitting	100.00%
SVM	Effective in high-dimensional spaces, needs tuning	78.85%

Logistic Regression:

Justification: Logistic Regression is known for its simplicity, interpretability, and effectiveness in binary classification problems. It performed well in this case, with an accuracy of 97.54%.
Performance Comparison: Highest accuracy among all the algorithms.

K-Nearest Neighbors:

Justification: K-Nearest Neighbors is effective when there are clear clusters in the data, but it might be sensitive to irrelevant features. The accuracy is 87.36%.

Performance Comparison: Lower accuracy compared to Logistic Regression and Decision Trees.
Considering further tuning and feature selection to improve performance, but it might not be the best choice in its current state.

Naive Bayes:

Justification: Naive Bayes assumes independence between features and is computationally efficient. It performed well with an accuracy of 97.26%.
Performance Comparison: Similar accuracy to Logistic Regression.
Naive Bayes is a good choice, but its performance might be sensitive to the independence assumption.

Decision Trees:

Justification: Decision Trees can capture complex relationships and achieved 100% accuracy, indicating potential overfitting.
Performance Comparison: Highest accuracy but prone to overfitting on the training data.

SVM:

Justification: SVMs are effective in high-dimensional spaces but might require careful hyperparameter tuning. It achieved an accuracy of 78.85%.
Performance Comparison: Lower accuracy compared to other algorithms.

Conclusion:

Logistic Regression and Naive Bayes show strong performances with similar accuracies.
Decision Trees achieved 100% accuracy but might be overfitting.
K-Nearest Neighbors and SVM have lower accuracies and may benefit from further tuning.

In this comparison, Logistic Regression stands out as a strong candidate due to its high accuracy and interpretability. It's also crucial to evaluate each model on a separate test set to ensure their generalization to new, unseen data.