Summary Report

Data Generation and Splitting:

The dataset was generated with three cluster centers at [2,4], [6,6], and [1,9].

The data set consists of 150 samples, which were split into:

80% Training Data (120 samples)

20% Testing Data (30 samples)

A random seed of 7 was used to ensure reproducibility.

Model Training and Evaluation:

KNN classifiers are trained using the training data.

Predictions are made for both the training and testing data.

The accuracy scores were:

Training Accuracy: 1.0 (100%)

Testing Accuracy: 1.0 (100%)

Data Visualization

The training dataset plot shows three well-separated clusters, confirming the dataset's suitability for classification. The confusion matrix heatmap visually confirms that all predictions are correct. The testing dataset plot also shows a distribution similar to the training dataset, reinforcing the model's effectiveness.

In Conclusion, The results of this analysis demonstrate that the KNN classifier performed exceptionally well, achieving 100% accuracy on both the training and testing datasets. The perfect classification suggests that the dataset was well-structured with minimal overlap between classes, making it an ideal case for KNN. While this result is highly promising, real-world datasets often have more complexity, requiring techniques such as cross-validation, hyperparameter tuning, and feature scaling to ensure robustness. In future analyses, exploring different values for k, adding noise to the dataset, or testing the model on real-world data could provide additional insights into its effectiveness in more challenging scenarios.