**ASSIGNMENT-1**

**Detailed Overview and Summary of the Using KNN Algorithm on Iris and Simulated Data**

INTRODUCTION:

This project applies the K-Nearest Neighbors (KNN) algorithm to both the Iris dataset and a custom synthetic dataset. It begins with an introduction to KNN and the Iris data, then explores key parameters of the KNeighborsClassifier, trains models, evaluates their predictions, and visualizes the resulting decision boundaries. The analysis incorporates train-test splits, confusion matrices, and accuracy scores, providing a clear picture of model performance across both datasets.

Essential Steps Performed With Explanations and Output Results::

* The iris dataset is loaded, split with an 80-20 train-test division, and a KNN classifier is fit without specifying hyperparameters.
* Accuracy on training data is measured at 97.5%97.5%.
* After parameter specification (n\_neighbors=5, weights='uniform', etc.), the accuracy on the iris test set is 96.7%96.7%.
* The help (KNeighborsClassifier) output provides detailed guidance on the model’s configurable parameters.
* Synthetic two-dimensional data are created around three cluster centers for visual demonstration.
* The synthetic data is also split by an 80-20 ratio. KNN achieves perfect test accuracy (100%100%) on this dataset owing to its clear cluster separation.
* Predictions and accuracy metrics are printed directly for both training and simulated datasets.
* Confusion matrices are calculated and displayed for the simulated dataset, visually confirming the model’s perfect classification.

Data Generation:

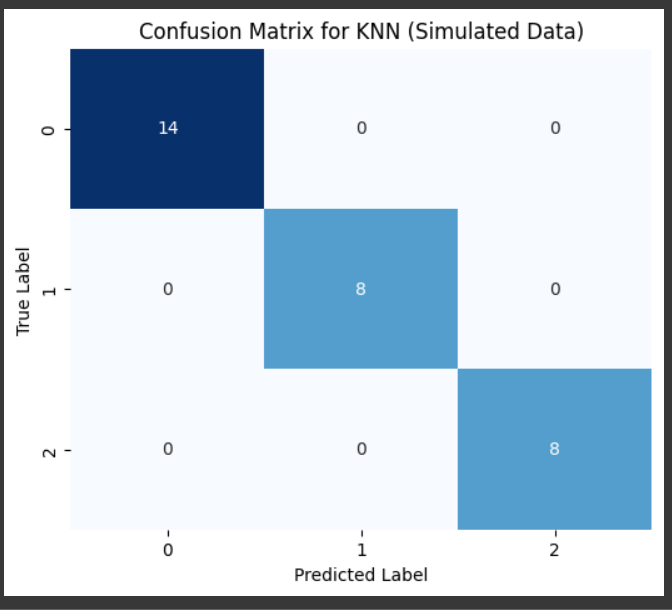
* Synthetic data was created using make\_blobs around three cluster centers, forming three artificial classes for classification tasks.
* This allowed evaluation of KNN on well-separated, controlled clusters.

Model Training and Testing

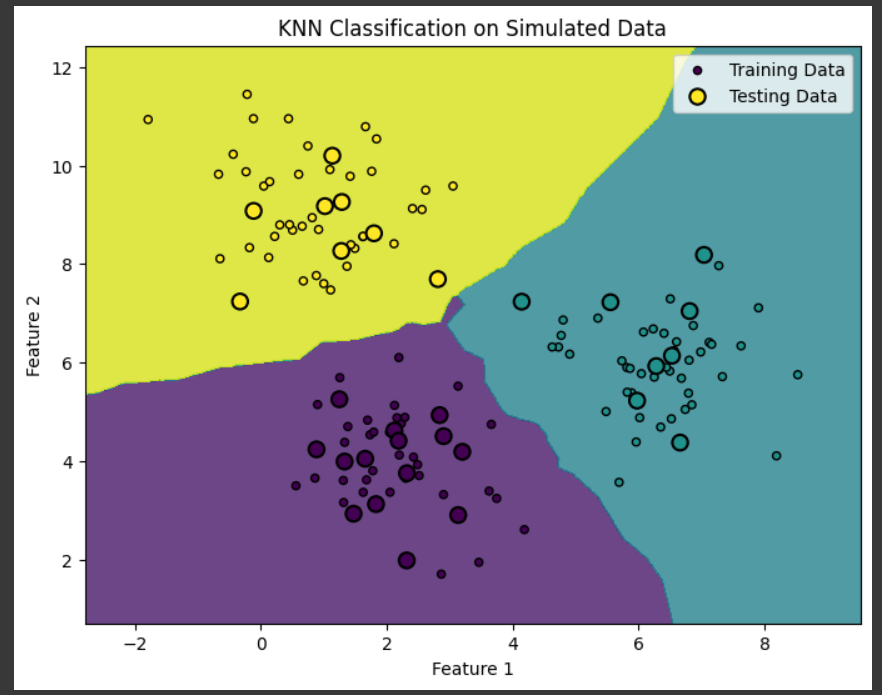
* The synthetic dataset was similarly split into 80% training and 20% testing.
* The KNN classifier, trained as before, achieved 100% accuracy on test data, reflecting perfectly separable simulated clusters.

Confusion Matrix and Visualization:

* A confusion matrix heatmap confirmed perfect classification for simulated test data: every predicted class matched the true label.



VISUAL PLOT:

The final plot in the notebook provides a visual representation of KNN’s classification boundaries on synthetic data. The colored regions illustrate the decision surfaces for each class, while the smaller circles represent the training samples and the larger circles represent the test samples. The clear separation of points within their respective colored regions highlights the strong performance of the model.

SUMMARY:

This assignment clearly shows how parameter tuning and evaluation techniques can improve KNN. By working with both classic and custom datasets, it highlights why choosing the right distance metrics and parameter settings is so important. The evaluation through accuracy scores and confusion matrices makes the results easy to interpret. Visualizing the decision boundaries also helps us see how well KNN works when clusters are distinct—leading to perfect accuracy and no confusion in the matrix. Altogether, this demonstrates that KNN is a strong choice for data with clear clustering patterns, while also reminding us that careful parameter tuning and visualization are key for truly understanding and trusting the model’s performance.