Advance Machine Learning

Assignment -1

The aim of this assignment is to learn how to run a basic machine learning project using Python.

Initially, the **Iris dataset** and trained a model called **K-Nearest Neighbors (KNN)** utilized to check how well it can classify flowers.

After that, a **new simulated dataset with three groups (classes)** were created and repeated the same steps with KNN.

This helps to practice splitting data into training and testing parts, training a model, and checking its accuracy.

The goal was to understand how KNN works and to see how its performance changes when the data is simple (well-separated) or a bit more challenging.

Method followed:

- Split both datasets into 80% training and 20% testing using train_test_split.

- Trained a KNeighborsClassifier (KNN) model on the training data.

- Evaluated the model using accuracy score on both training and testing sets.

- Also plotted a confusion matrix and scatter plots to visualize the results.

Results (Regarding Iris Dataset) -

- **Train accuracy:** 0.975

- **Test accuracy:** 0.967

- The KNN model performed very well on the Iris dataset.
- It reached about **97.5% accuracy on training data** and **96.7% on testing data**.
- The results show that the model got the patterns correctly and also generalized well to new data.

```
···   Predictions from the classifier:
      [0 1 2 0 2 0 1 1 0 1 1 0 0 0 0 0 0 0 2 0 2 1 1 1 0 2 1 1 2 0 2 0 2 1 2 2 1
       1 1 2 2 0 2 2 0 1 0 2 2 0 1 1 0 0 1 1 1 1 2 1 2 0 0 1 1 2 0 2 1 0 2 2 1 2
       2 0 0 2 1 1 2 0 1 1 0 1 1 2 2 1 0 2 0 2 0 0 1 2 2 1 2 2 0 1 1 0 2 2 2 1 2
       2 2 0 0 1 0 2 2 1]
      Target values:
      [0 1 2 0 2 0 1 1 0 1 1 0 0 0 0 0 0 0 2 0 2 1 1 1 0 2 1 1 2 0 2 0 2 2 2 2 1
       1 1 1 2 0 2 2 0 1 0 2 2 0 1 1 0 0 1 1 1 1 2 1 2 0 0 1 1 1 0 2 1 0 2 2 1 2
       2 0 0 2 1 1 2 0 1 1 0 1 1 2 2 1 0 2 0 2 0 0 1 2 2 1 2 2 0 1 1 0 2 2 2 1 2
       2 2 0 0 1 0 2 2 1]
      0.975

···   0.9666666666666667
```

Simulated Dataset

- **Train accuracy:** 1.000

- **Test accuracy:** 1.000

- **Observation:** Accuracy is perfect because the blobs are well separated, so KNN can easily classify.
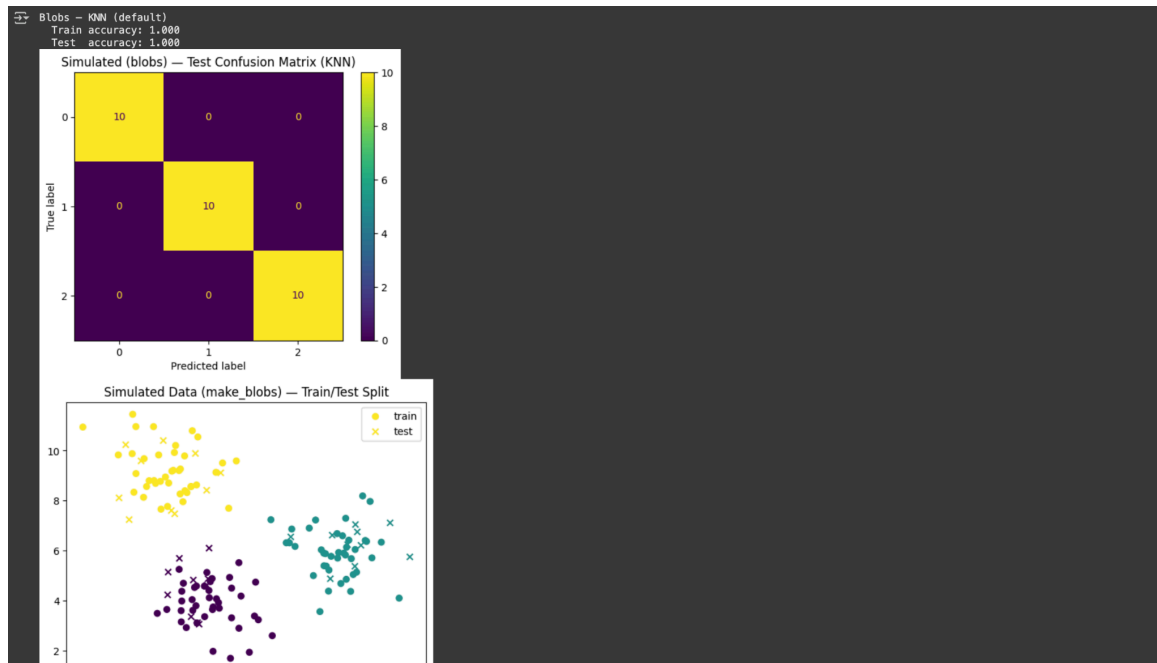
Interpretation:

The simulated dataset had three clear groups of points.

When training the KNN model on this data, it reached **100% accuracy** on both training and testing sets.

This might have happened because the clusters were separated, so there was no confusion between classes.

It shows that KNN works well when the data is clean and the classes are far apart.

Conclusion

In this assignment, I learned how to:

- Train and test a KNN classifier using Python and scikit-learn.

- Measure accuracy for both training and testing sets.

- Use confusion matrices and scatter plots to visualize results.

The results show how the structure of data strongly affects machine learning performance.