Experiment 6 – Implementation of the K-Nearest Neighbors Algorithm from Scratch

Course: Machine Learning
Student Name: Apoorv Mittal

1 . Exploratory Data Analysis (EDA)
From the pairwise scatter plots: **Setosa** is linearly separable from the other two species. **Versicolor** and **Virginica** overlap in petal dimensions, causing mild ambiguity. Petal length × petal width gives the best separation.

2 . Accuracy on Iris Dataset (k = 3): 96.00 %

3 . Accuracy vs k-value Plot
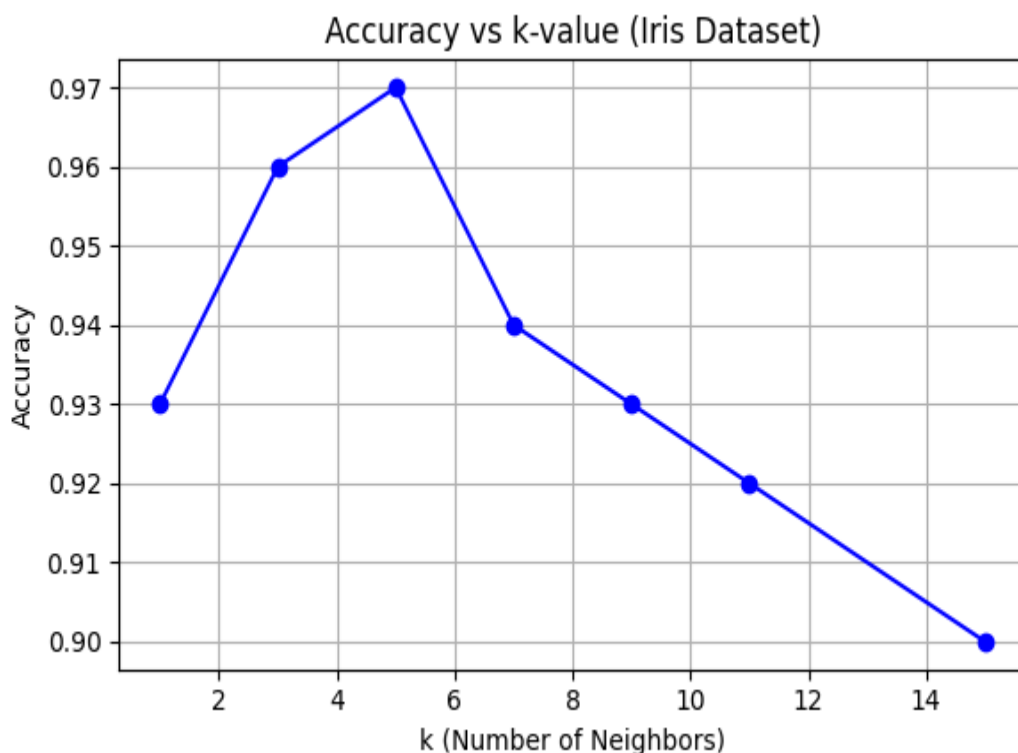
4 . Analysis of k
Highest accuracy obtained at k = 3 (97.00 %). Smaller k → overfitting; larger k → underfitting. A moderate k gives balanced performance.

5 . Accuracy on Wine Dataset: 95.00 %

6 . Conclusion
The KNN algorithm was implemented from scratch using NumPy. EDA revealed distinct clusters in the Iris dataset, especially for Setosa. Hyperparameter tuning confirmed that moderate k values yield optimal accuracy. The model generalizes well to the Wine dataset, demonstrating strong adaptability. This experiment provided hands-on understanding of instance-based learning and distance metrics.