Lab 7 Assignment

Lab: k-NN vs Naïve Bayes (Clean + Noisy)

Objectives:

1. Implement the k-Nearest Neighbors (k-NN) classifier.

2. Implement the Naïve Bayes classifier.

3. Compare the performance of the two algorithms.

Tools: Scikit-learn

Dataset: Wine Dataset (from sklearn.datasets import load_wine)

Supporting Content:

• k-NN: An instance-based learning algorithm that classifies a data point based on how its neighbors are classified.

• Naïve Bayes: A probabilistic classifier based on applying Bayes' theorem with strong (naïve) independence assumptions between the features.

• Feature Scaling: k-NN is distance-based, so scaling is crucial. Naïve Bayes is not affected by scaling.

Lab Tasks:

1. Load the Wine dataset and split it into training and testing sets.

2. Standardize the features (important for k-NN!).

3. Train a k-NN classifier (KNeighborsClassifier) with n_neighbors=5. Calculate its accuracy.

4. Train a Naïve Bayes classifier (GaussianNB). Calculate its accuracy.

5. Compare the accuracy of the two models.

Exercise:

1. For k-NN, experiment with different values of k (1, 5, 15). Plot k against accuracy. What is the optimal k and why does performance change?

2. Why is feature scaling critical for k-NN but not for Naïve Bayes? Test this by running k-NN on the unscaled data and reporting the accuracy.

3. For the Naïve Bayes model, use predict_proba to examine the predicted probabilities for the first test instance.

Dataset Link:

🔗 Wine Dataset (UCI Machine Learning Repository)
https://archive.ics.uci.edu/ml/datasets/wine