Supervised ML (Assignment3)

Assignment Project

Problem Statement

A botanical research centre is studying different species of Iris flowers to automate the plant identification process. Currently, botanists manually identify flower species by measuring sepal and petal dimensions, which is:

- Time-consuming
- Error-prone
- Not scalable

The research centre wants to build a **Machine Learning based classification system** that can automatically predict the **species of an Iris flower** based on its physical measurements.

You are hired as a **ML Engineer** to build classification models using:

- K-Nearest Neighbours (KNN)
- Logistic Regression
- Naive Bayes

Your task is to train all three models, compare their performances, and identify the best algorithm.

Dataset Description

Feature	Description
SepalLengthCm	Length of sepal (cm)
SepalWidthCm	Width of sepal (cm)
PetalLengthCm	Length of petal (cm)
PetalWidthCm	Width of petal (cm)
Species (Target)	Iris-setosa, Iris-versicolor, Iris-virginica

Note - Iris dataset is a small, well-balanced & clean dataset with only 150 samples split across 3 categories (50 each). So the models are going to perform extremely good for this dataset but that should not be a realistic expectation from real-life data that we deal with. Because real-life data is not always balanced, has noise & needs a lot of pre-processing.

To make the predictions a little bit worse - only train with 50% of the data & test on 100%.