Iris Classification using K-Nearest Neighbors (KNN)
Internship Task 6

1. Abstract
This project implements the K-Nearest Neighbors (KNN) classification algorithm on the Iris dataset to understand distance-based learning. The workflow includes data normalization, hyperparameter tuning, model evaluation using confusion matrix and accuracy metrics, and visualization of decision boundaries.

2. Problem Statement
To build an efficient and accurate multi-class classification model using the KNN algorithm for predicting iris flower species based on their morphological measurements.

3. Dataset Description
The Iris dataset consists of 150 samples belonging to three species: Setosa, Versicolor, and Virginica. Each sample contains four features: sepal length, sepal width, petal length, and petal width.

4. Methodology
- Data Loading and Exploration
- Feature Normalization
- Train-Test Split
- KNN Model Training
- Hyperparameter Tuning
- Model Evaluation
- Decision Boundary Visualization

5. Model Training
The KNN model was trained using different values of K to find the optimal number of neighbors. Feature normalization was applied to prevent scale bias in distance calculations.

6. Model Evaluation
Model performance was evaluated using accuracy score, confusion matrix, and classification report. The optimal K value resulted in high classification accuracy across all three iris species.

7. Decision Boundary Visualization

Decision boundaries were visualized using two selected features to illustrate class separation. This provides intuitive understanding of how the KNN algorithm classifies different regions in feature space.

8. Results and Discussion

The KNN classifier achieved excellent performance with minimal misclassifications. Hyperparameter tuning significantly improved generalization and stability of the model.

9. Conclusion

This project demonstrates a complete machine learning workflow for multi-class classification using KNN. The inclusion of normalization, hyperparameter tuning, and visualization techniques ensures high accuracy and interpretability, making the model suitable for real-world applications.