🌸 Iris Flower Classification – Pro Version

Project Report

🔍 1. Introduction

The Iris dataset is one of the most famous datasets in machine learning. It contains **150 samples** of iris flowers, each with four features: sepal length, sepal width, petal length, and petal width.
 The goal is to classify flowers into three species: **Setosa, Versicolor, and Virginica**.

This project demonstrates a **complete ML workflow** including data exploration, visualization, preprocessing, model training, cross-validation, hyperparameter tuning, and evaluation.

🛠️ 2. Project Workflow

Pipeline:

Dataset → Exploration → Visualization → Preprocessing → Model Training
→ Cross-Validation → Hyperparameter Tuning → Evaluation → Results

Steps:

- **Dataset:** Iris dataset (150 samples, 3 classes, 4 features)

- **Exploration & Visualization:** Pairplots, distribution plots

- **Preprocessing:** Feature scaling (important for LR, KNN, SVM)

- **Model Training & Evaluation:**

    - Logistic Regression

    - K-Nearest Neighbors (KNN, tuned with GridSearchCV)

    - Decision Tree (tuned with GridSearchCV)

○ Support Vector Machine (SVM, best model)

● **Validation:** 5-fold cross-validation, confusion matrix, decision boundary plots

---

🤖 3. Models Used

Model	Description
Logistic Regression	Linear model, interpretable
KNN	Instance-based, requires scaling
Decision Tree	Rule-based, interpretable, tuned depth
SVM	High-performance, robust to non-linear data

📊 4. Results

✅ Cross-Validation Accuracy

Model	5-Fold CV Accuracy
Logistic Regression	0.9600
KNN	0.9600 (best k=5)
Decision Tree	0.9533 (best max_depth=4)
SVM	**0.9667 (Best Model)**

👉 **SVM achieved the highest CV accuracy (96.7%)** and test accuracy of ~97%.
👉 Logistic Regression and KNN performed similarly well (~96%).
👉 Decision Tree was slightly lower (~95%) but more interpretable.

📉 Key Insights

- **SVM** was the **best performing model**.

- **Scaling** was essential for Logistic Regression, KNN, and SVM.

- **Confusion matrix** confirmed balanced classification across all species.

- Visualization of decision boundaries provided clear model interpretability.

📂 5. Figures Generated

- Pairplot of Iris features – `figures/pairplot.png`

- Feature distributions – `figures/feature_distributions.png`

- Confusion Matrix (SVM) – `figures/confusion_matrix.png`

- Decision Boundary (SVM, first 2 features) – `figures/decision_boundary.png`

🚀 6. Conclusion

The **Iris Flower Classification – Pro Version** showcases a **professional end-to-end ML workflow**:

- Data exploration & visualization

- Preprocessing (scaling)

- Multiple models compared

- Cross-validation & hyperparameter tuning

- Evaluation with confusion matrix & decision boundaries