Iris Flower Classification — Pro Version

This project classifies iris flowers into Setosa, Versicolor, and Virginica using multiple machine learning models. It demonstrates a complete ML workflow: data exploration, visualization, preprocessing, model training, cross-validation, hyperparameter tuning, evaluation, and visualization of results.

Models Implemented: Logistic Regression (linear model) K-Nearest Neighbors (instance-based, tuned with GridSearchCV) Decision Tree (rule-based, tuned with GridSearchCV) Support Vector Machine (SVM, best performing model)

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)

Key Insights: SVM achieved the highest accuracy (96.7% CV, 97% test set). KNN and Logistic Regression performed similarly well (~96%). Decision Tree slightly lower (~95%), but interpretable. Scaling was essential for Logistic Regression, KNN, and SVM. Confusion matrix confirmed balanced performance across all classes.

Figures Generated:

Pairplot of Iris features (`figures/pairplot.png`) Feature distributions (`figures/feature_distributions.png`) Confusion matrix for best model (SVM) Decision boundary (SVM, first 2 features)

Conclusion:
This project demonstrates a complete ML workflow, from data exploration to advanced evaluation. The Support Vector Machine was the best performing model, achieving 97% accuracy. By including visualization, cross-validation, hyperparameter tuning, and multiple models, this project reflects a professional approach to ML problem-solving suitable for internship evaluation.