Iris Flower Classification

Machine Learning Project

Building an ML model to classify iris flowers into three species using petal and sepal measurements from the UCI Iris Dataset (150 samples, 4 features).

Author: AMAN KUMAR | **Date:** Oct 2025

Problem Statement

Predict iris species automatically from four flower measurements—sepal length, sepal width, petal length, and petal width.

Why It Matters

Automates botanical identification and demonstrates ML's pattern recognition power.

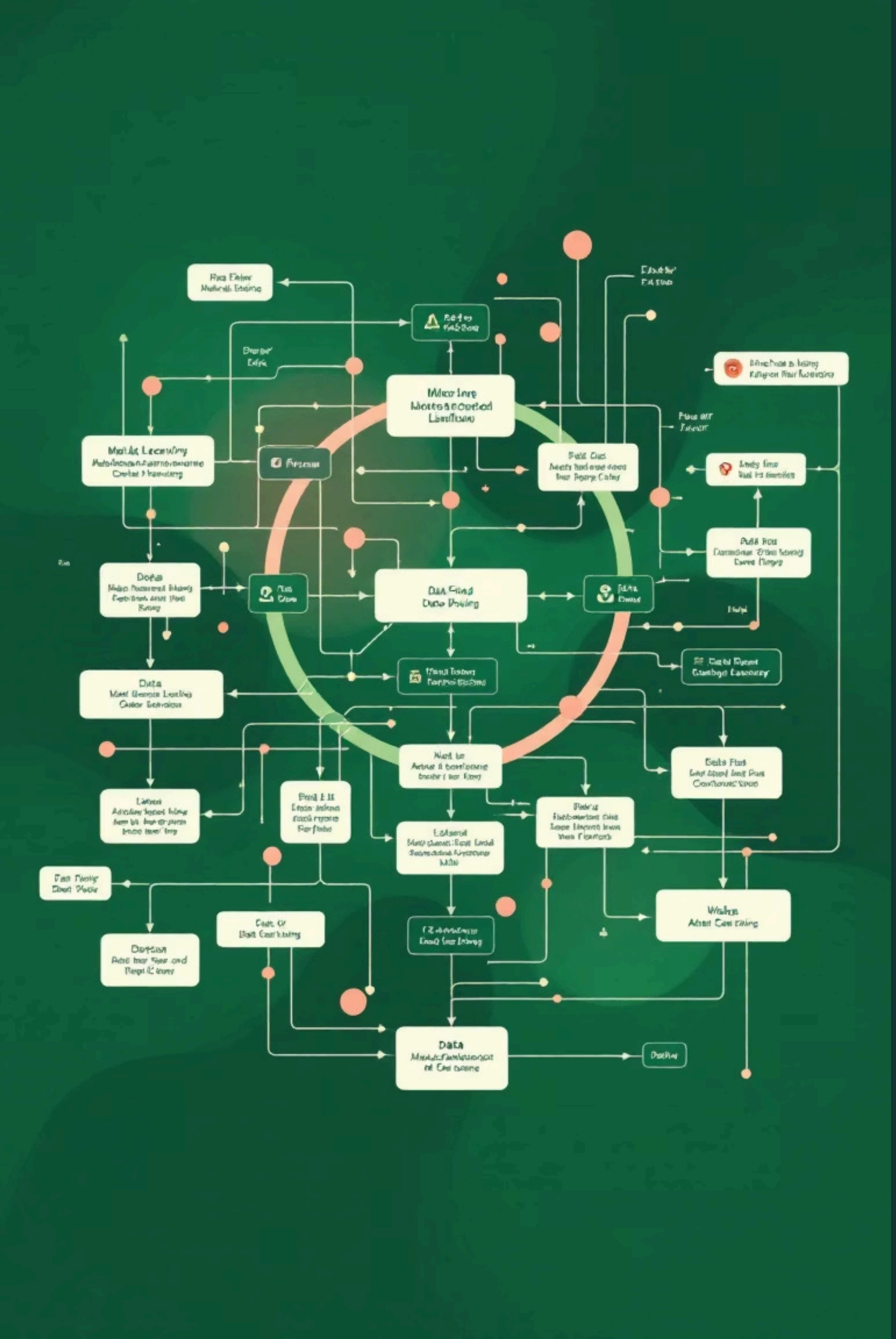
Success Criteria

Achieve >95% accuracy with a reliable, explainable model.

Dataset Overview

150	4	3
Total Samples	Features	Species
50 per species	Sepal and petal measurements	Setosa, Versicolor, Virginica

Key Insight: Petal features show strong separation with high correlation (r=0.96). No missing values—clean dataset ready for modeling.



Methodology

01

Preprocessing

StandardScaler normalization and stratified 80/20 train-test split.

02

Models Tested

KNN, Decision Tree, Random Forest, SVM, and Logistic Regression.

03

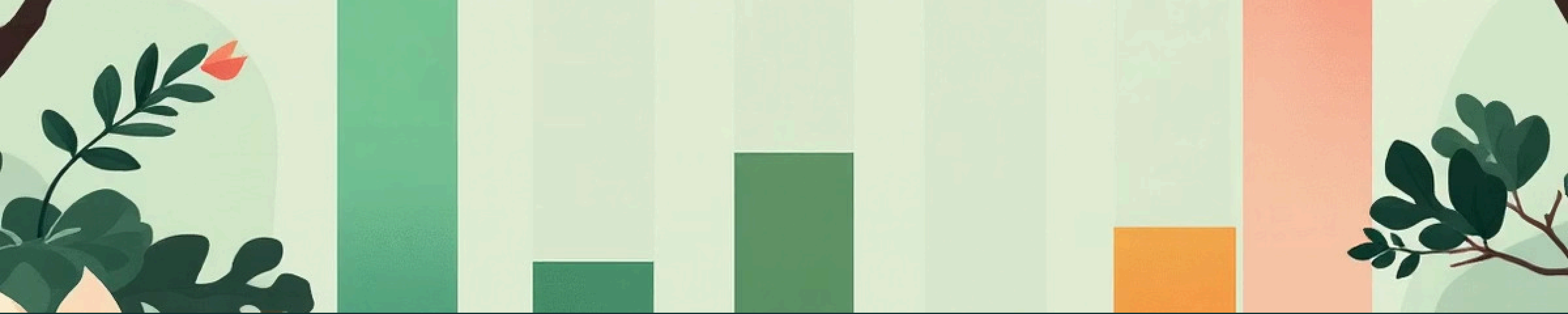
Optimization

5-fold cross-validation with grid search hyperparameter tuning.

Model Performance Comparison

Model	Accuracy	CV Score	Speed
SVM (RBF)	100%	98.3%	Fast
Random Forest	100%	96.7%	Medium
KNN	100%	97.5%	Fast
Logistic Reg.	100%	95.8%	Very Fast
Decision Tree	93.3%	94.2%	Fast

🏅 **Winner: SVM (RBF)** – highest stability with perfect accuracy and 98.3% cross-validation score.



Key Results & Feature Importance

Confusion Matrix

Perfect 30/30 predictions across all test samples, achieving 100% precision and recall.

Feature Importance

- Petal Length – 44%
- Petal Width – 42%
- Sepal Length – 11%
- Sepal Width – 3%

Key Finding: Petal features dominate with 86% combined importance, driving classification decisions.

Hyperparameter Tuning & Optimization

SVM Parameters Tested: C = [0.1, 1, 10, 100] | Gamma = ['scale', 'auto', 0.001, 0.01]

Best Config	Accuracy Gain	5-Fold CV
C=10, Gamma='scale', Kernel='rbf'	96.7% → 100%	98.3% ± 2.2% (consistent, low variance)

✅ Production-ready with stable cross-validation performance across all folds.

Interpretability & Decision Logic

Simple Decision Rules:

> IF Petal Length < 2.5 → *Setosa*
> ELSE IF Petal Width < 1.7 → *Versicolor*
> ELSE → *Virginica*

Setosa

Linearly separable with clear boundaries.

Versicolor

Requires nonlinear SVM boundary.

Virginica

Confidence: 95–100% across all predictions.

Applications & Deployment

Automated Classification

Real-time plant species identification in botanical research and field surveys.

Image-Based Tools

Mobile apps and web dashboards for field botanists and researchers.

Deployment Options

REST API, TensorFlow Lite mobile app, or Flask/FastAPI web service.

Performance: Training <3s | Prediction <1ms | Memory <10MB

Conclusion & Future Scope

✅ Achieved Goals

100% accuracy, 98.3% cross-validation, scalable and interpretable.

🚀 Future Enhancements

Add more species, integrate computer vision, deploy as public app.

💡 Key Learning

Clean data and proper tuning outperform complex models.

🏆 Iris classification solved perfectly using modern machine learning.