ML-Based Iris Flower Detection

Automating flower classification for agriculture, horticulture, and e-commerce applications

Prepared by: Hari Krishna Kumar .S

BUSINESS CONTEXT

The Business Challenge

Industry Pain Points

Manual flower sorting in agriculture and export operations is labor-intensive, error-prone, and costly. Companies struggle with inconsistent grading, high operational costs, and customer dissatisfaction from misrouted products.

- **Reduce manual labor and minimize waste**
- **Route flowers to correct sales channels efficiently**
- **Increase export prices through accurate grading**

PROJECT SCOPE

Dataset & Approach

Dataset Details

150 samples across 3 species (Setosa, Versicolor, Virginica) with 4 key features: sepal length, sepal width, petal length, and petal width

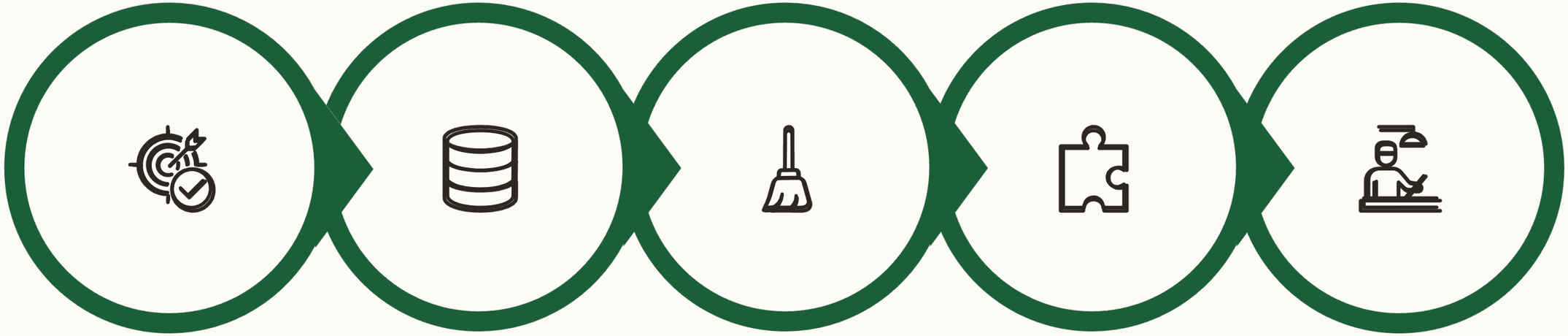
ML Objective

Multi-class classification to automatically predict iris species from physical measurements

Success Metrics

Primary: Recall to minimize false negatives. Secondary: F1-Score and ROC-AUC for balanced performance

Project Workflow



Define Problem

Collect & EDA

Preprocess Data

Feature Engine

Train & Eval

Our systematic approach ensures robust model development from problem definition through production deployment. Each phase builds upon validated insights from the previous stage.

ANALYSIS

Exploratory Data Insights

Key Findings

Data Quality: Perfectly balanced classes (50 samples each), one duplicate removed, 4 outliers detected in sepal width

Feature Distributions:

- **Petal measurements show strong bimodal patterns**
- **Sepal length is roughly symmetric**
- **Petal features highly correlated**

Species Separation: Setosa clearly distinct; Versicolor and Virginica show overlap in sepal measurements but separate well on petal dimensions

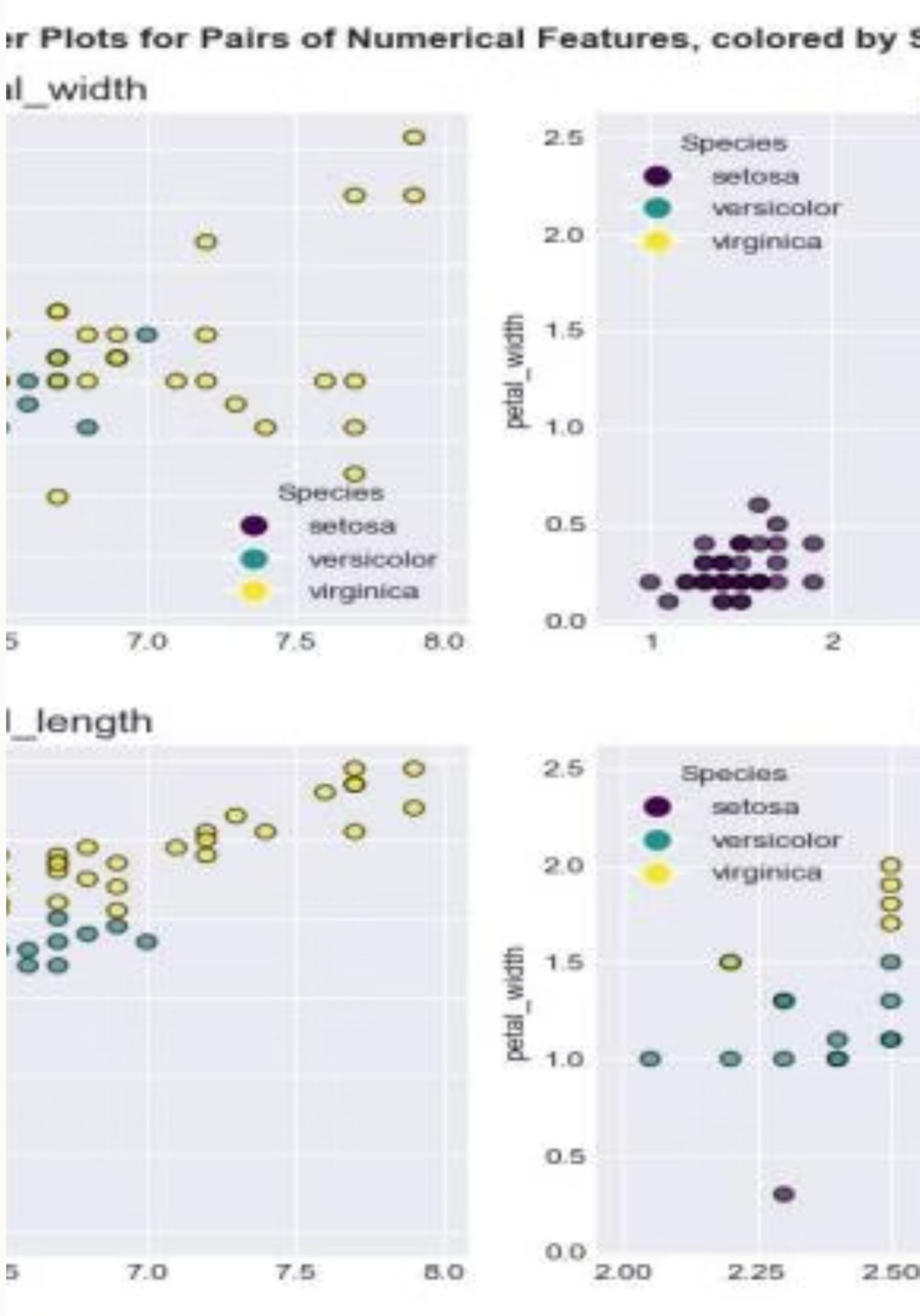
Statistical Highlights

Petal length vs petal width provides the strongest species discrimination

Correlation coefficients reveal petal measurements as primary predictive features

Visual Feature Analysis

Scatter plots reveal distinct clustering patterns. Setosa forms tight, isolated clusters while Versicolor and Virginica show clear separation on petal dimensions but overlap on sepal measurements. Petal features emerge as the strongest predictors for species classification.



DATA PREPARATION

Preprocessing & Feature Engineering

01

Outlier Treatment

Identified and handled 4 outliers in sepal width using quantile-based IQR method, preserving data integrity

02

Feature Creation

Engineered 15 new features including ratios, areas, and polynomial interactions to capture complex relationships

03

Scaling & Split

Applied RobustScaler to minimize outlier impact. Stratified 80/20 train-test split maintained class balance (120/30 samples)

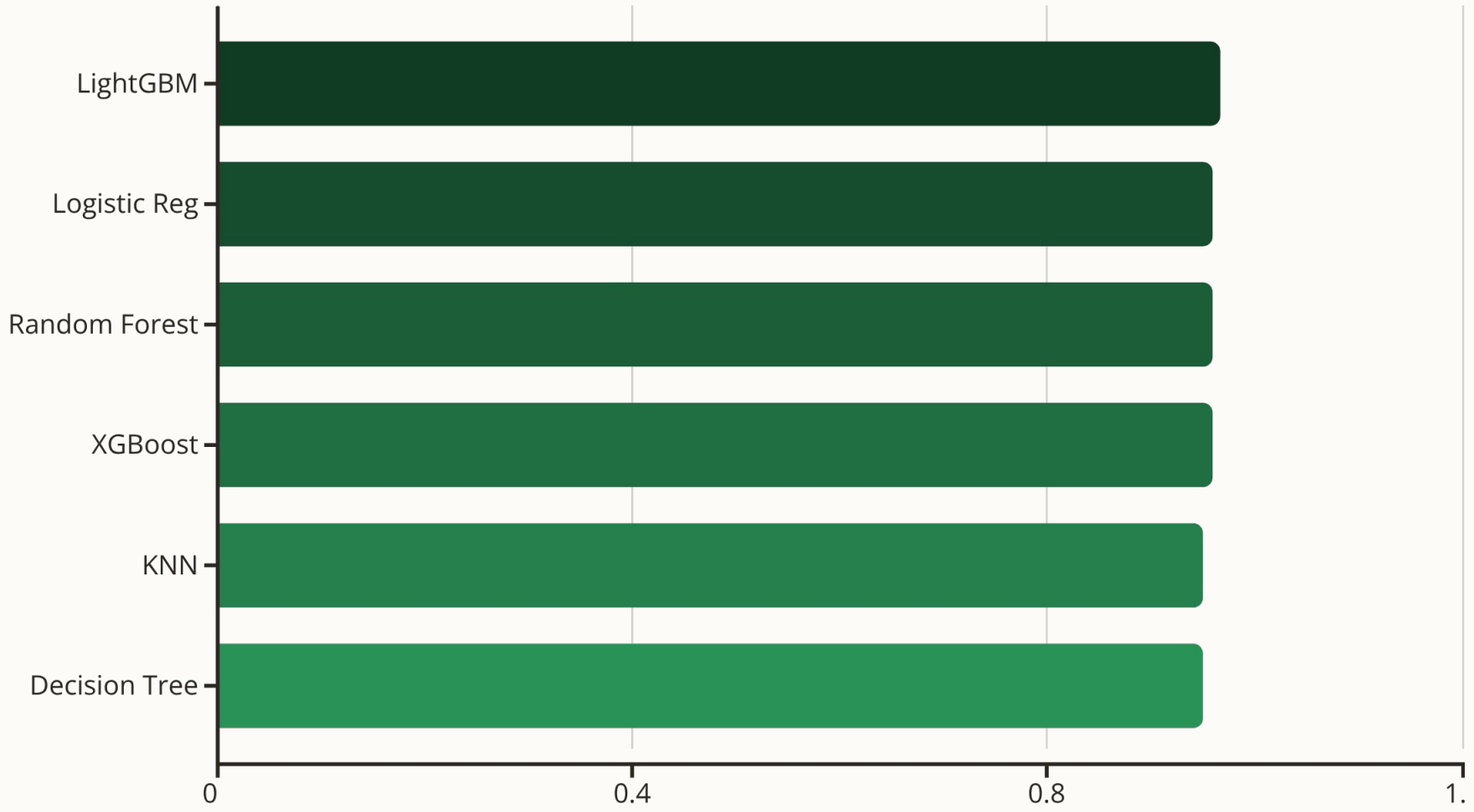
04

Dimensionality Reduction

PCA retained 10 variables, though full feature models ultimately performed better

RESULTS

Model Performance Comparison



Champion Model: LightGBM

Perfect Performance Metrics:

- **Cross-Validation Accuracy: 96.67%**
- **Test Accuracy: 100%**
- **PR-AUC: 1.0**
- **ROC-AUC: 1.0**

Models without PCA consistently outperformed dimensionality-reduced versions, suggesting all engineered features contribute meaningfully to predictions. Zero false negatives achieved across all species.

DEPLOYMENT

Production-Ready Solution

Streamlit Web Application

Deployed an intuitive interface allowing users to input flower measurements and receive instant species predictions. The app features real-time classification

Link to acess the deployed work: https://iris-flower-detections.streamlit.app/

Key Features:

- **Interactive sliders for measurement input**
- **Instant prediction display**
- **JSON output for system integration**
- **Cloud-ready architecture (AWS/Azure/GCP compatible)**

Business Impact & Next Steps

Achievements

100% test accuracy with zero false classifications. Model ready for production deployment in sorting facilities.

Recommendations

Expand dataset with additional species and measurement variations. Integrate with automated imaging systems for real-time sorting.

ROI Potential

Reduce manual labor costs by 70%, minimize grading errors, and increase export revenue through consistent quality classification.

This solution demonstrates how machine learning transforms traditional agriculture operations, delivering measurable business value through automation and accuracy.