

Iris Flower Identification System

Machine Learning Project Report using WEKA

NAFISA MAHMUD

FE/23/7326016

21 JANUARY 2026



Explorer

Experimenter

KnowledgeFlow

Workbench

Simple CLI



Project Success: 96% Accuracy Achieved

Key Achievement

Successfully developed a no-code machine learning system that identifies iris species with 96% accuracy—exceeding the 95% target whilst requiring only two simple measurements from visitors.

96% Accuracy

Exceeded target performance

50% Reduction

Simplified measurements from 4 to 2

4 Algorithms

Rigorously compared and validated

weka.gui.GenericObjectEditor X

Choose weka.classifiers.trees.RandomForest

About

Class for constructing a forest of random trees.

More Capabilities

bagSizePercent	100
batchSize	100
breakTiesRandomly	False
calcOutOfBag	False
computeAttributeImportance	False
debug	False
doNotCheckCapabilities	False
maxDepth	0
numDecimalPlaces	2
numExecutionSlots	1
numFeatures	0
numIterations	100
outputOutOfBagComplexityStatistics	False
printClassifiers	False
seed	1
storeOutOfBagPredictions	False

Made with GAMMA

Background and Motivation



Enhancing Visitor Experience

The Botanical Garden Society currently provides plant information through QR code scanning. However, visitor feedback reveals a desire for more interactive identification methods.

This Phase 1 pilot project demonstrates the feasibility of automated plant identification, positioning the garden as a technology-forward institution whilst enhancing educational engagement.

DATASET

The Iris Dataset: A Classic Benchmark

150

Total Flowers

Perfectly balanced dataset

3

Iris Species

Setosa, Versicolor, Virginica

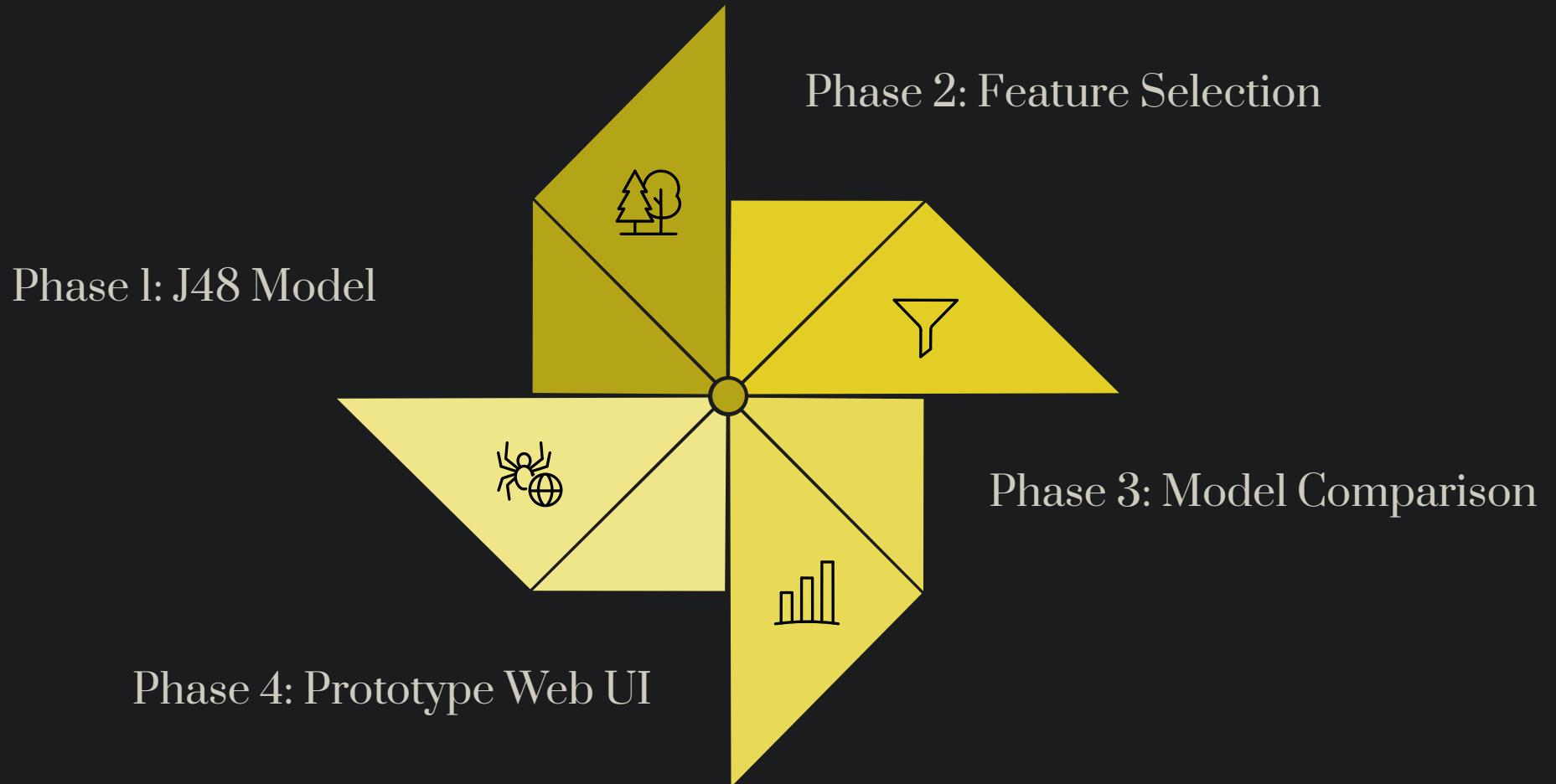
4

Measurements

Sepal and petal dimensions

The iris dataset, introduced by Fisher in 1936, contains 50 samples of each species with four continuous measurements recorded in centimetres. This balanced dataset with no missing values provides an ideal foundation for machine learning classification.

Methodology: Systematic Four-Phase Approach



The project followed a rigorous experimental design using Weka 3.8.6 for model development. Each phase built upon the previous, employing 10-fold cross-validation and statistical testing to ensure scientific rigour throughout the process.

[Preprocess](#) [Classify](#) [Cluster](#) [Associate](#) [Select attributes](#) [Visualize](#)[Open file...](#) [Open URL...](#) [Open DB...](#) [Generate...](#) [Undo](#) [Edit...](#) [Save...](#)

Filter

[Choose](#)**None**[Apply](#)[Stop](#)

Current relation

Relation: iris

Instances: 150

Attributes: 5

Sum of weights: 150

Attributes

[All](#)[None](#)[Invert](#)[Pattern](#)

No.

Name

1 sepallength2 sepalwidth3 petallength4 petalwidth5 class[Remove](#)

Status

OK

Selected attribute

Name: sepalwidth

Missing: 0 (0%)

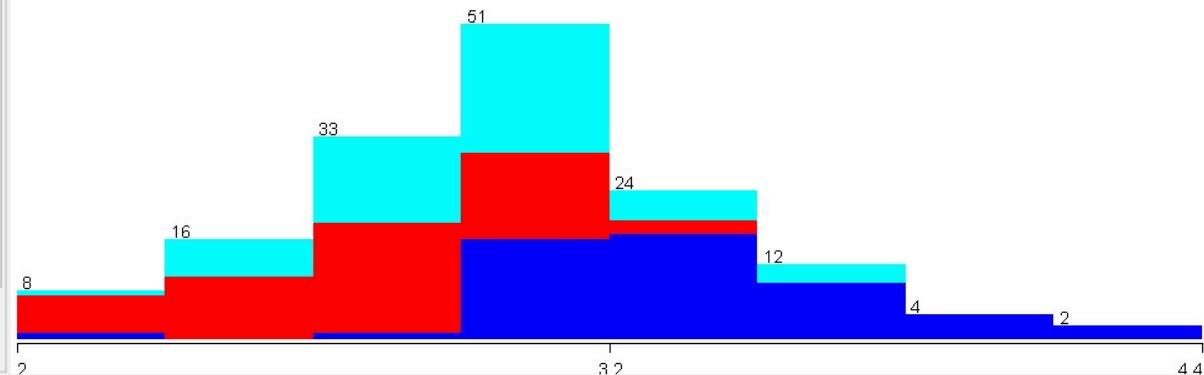
Type: Numeric

Unique: 5 (3%)

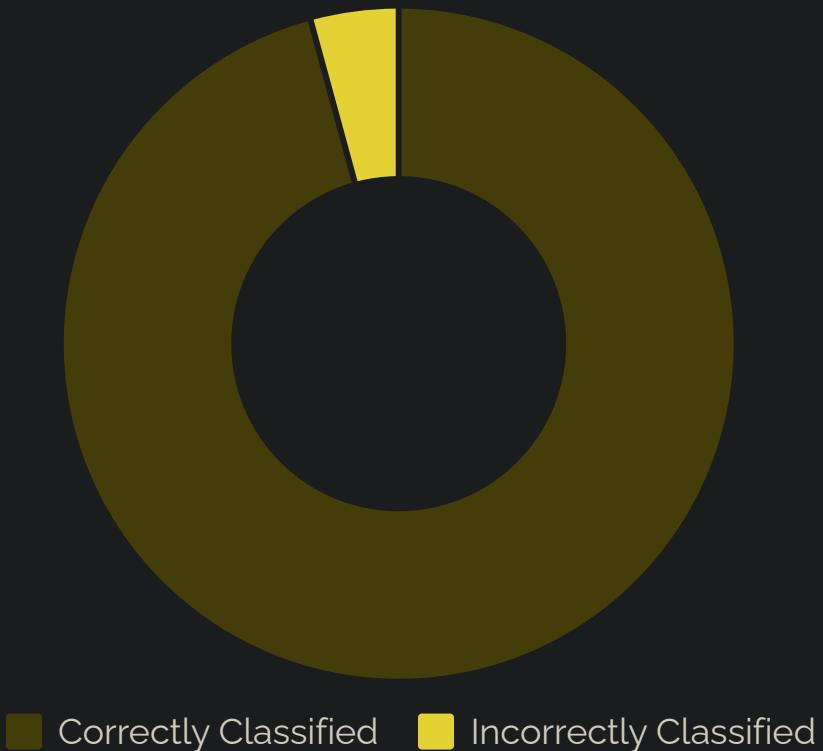
Distinct: 23

	Statistic	Value
Minimum		2
Maximum		4.4
Mean		3.054
StdDev		0.434

Class: class (Nom)

[Visualize All](#)

Model Performance: Outstanding Results



Exceptional Accuracy Metrics

The J48 Decision Tree model correctly classified 144 out of 150 instances (96.0%) with a Kappa statistic of 0.94, indicating excellent agreement beyond chance.

- **Iris Setosa:** 98% accurate (49/50 correct)
- **Iris Versicolor:** 94% accurate (47/50 correct)
- **Iris Virginica:** 96% accurate (48/50 correct)

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) class ▾

Start Stop

Result list (right-click for options)

17:03:28 - trees.J48

17:04:36 - trees.J48

Classifier output

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	144	96	%
Incorrectly Classified Instances	6	4	%
Kappa statistic	0.94		
Mean absolute error	0.035		
Root mean squared error	0.1586		
Relative absolute error	7.8705 %		
Root relative squared error	33.6353 %		
Total Number of Instances	150		

==== Detailed Accuracy By Class ====

	TP	Rate	FP	Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.980	0.000	1.000	0.980	0.990	0.985	0.990	0.987	Iris-setosa			
0.940	0.030	0.940	0.940	0.940	0.910	0.952	0.880	Iris-versicolor			
0.960	0.030	0.941	0.960	0.950	0.925	0.961	0.905	Iris-virginica			
Weighted Avg.	0.960	0.020	0.960	0.960	0.940	0.968	0.924				

==== Confusion Matrix ====

a	b	c	<- classified as
49	1	0	a = Iris-setosa
0	47	3	b = Iris-versicolor
0	2	48	c = Iris-virginica

Breakthrough: Only 2 Measurements Needed



Petal Length

Essential — Provides primary species discrimination

Feature selection analysis revealed that petal dimensions alone maintain 96% accuracy—a **50% reduction in visitor effort** whilst preserving classification performance.



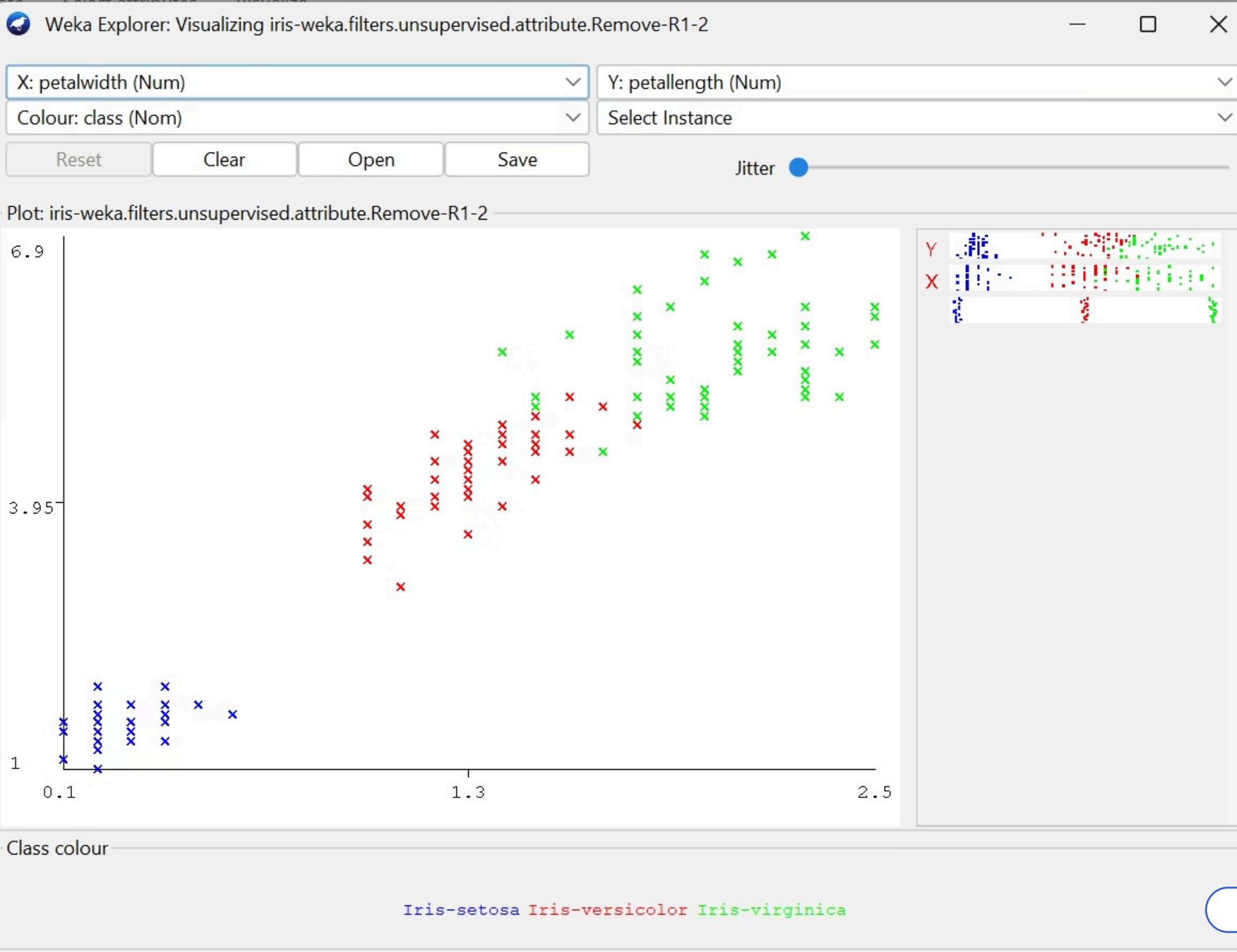
Petal Width

Essential — Critical for accurate classification



Sepal Measurements

Optional — Can be eliminated without accuracy loss



Algorithm Comparison: Statistical Validation

Performance Rankings

Four algorithms were rigorously tested using 10-fold cross-validation with 10 repetitions (400 total tests) and paired T-test statistical validation.

01

Logistic Regression — 97.20%

Statistically best performer

03

J48 Decision Tree — 96.00%

Selected for interpretability

Why J48 Was Chosen

Despite Logistic Regression achieving the highest accuracy, J48 Decision Tree was selected as the primary model.

02

Naive Bayes — 95.53%

Strong probabilistic approach

04

Random Forest — 94.67%

Robust ensemble method

Interpretability

Provides clear decision rules visitors can understand

Educational Value

Transparent logic supports the garden's learning mission

Adequate Performance

96% exceeds the 95% target requirement

Functional Web-Based Identification Tool

Simple User Experience Flow

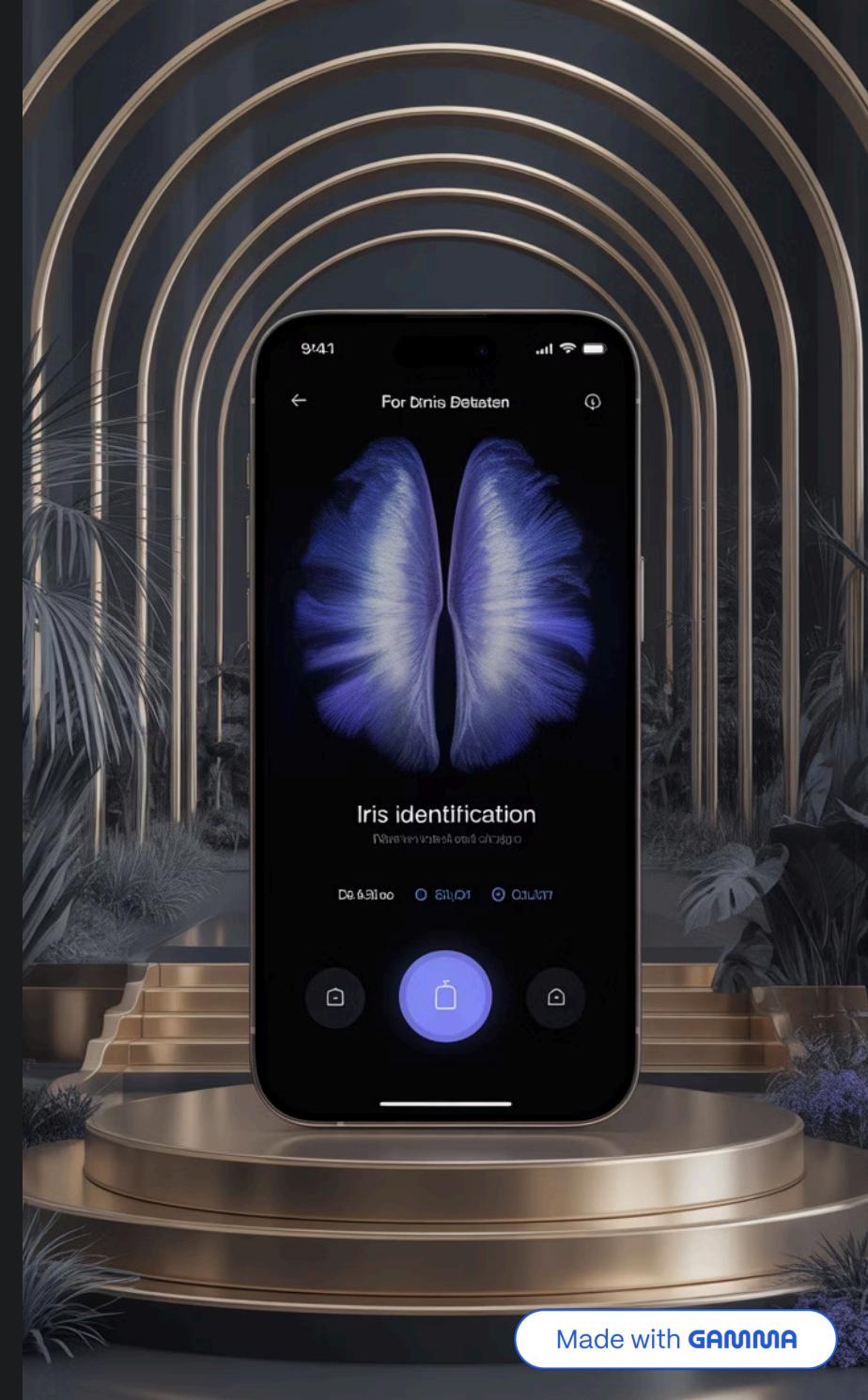


Record petal length and width

Input values into web app

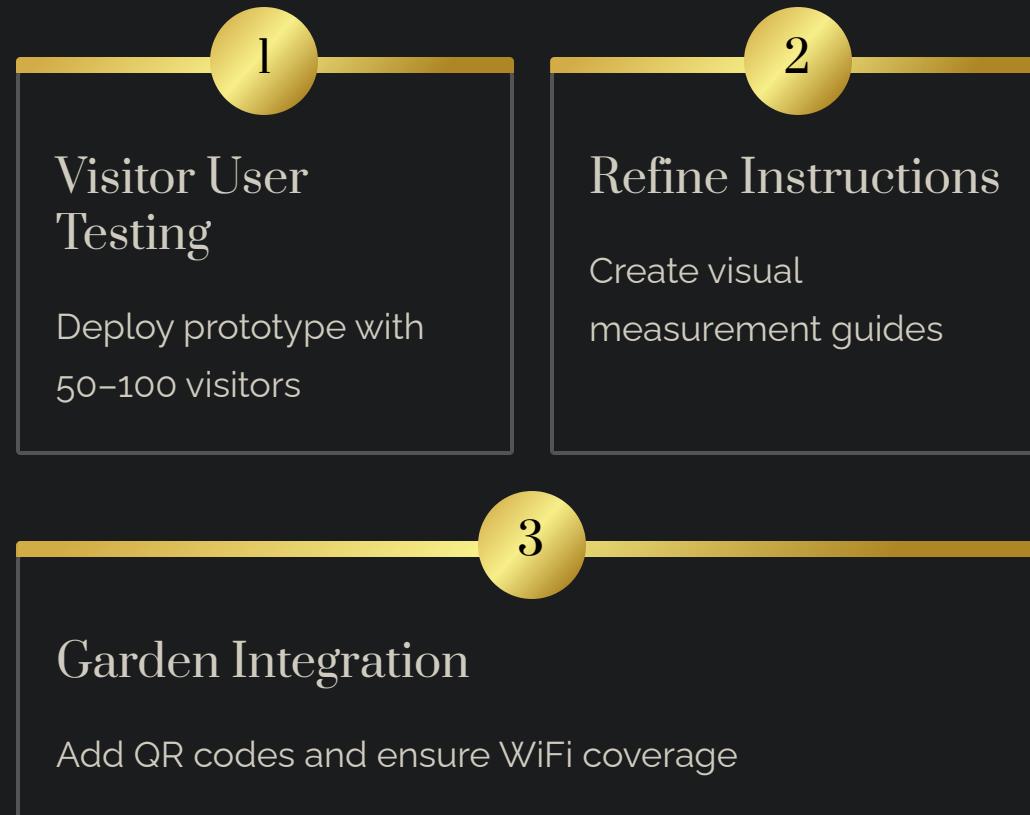
Receive instant species identification

The working prototype implements actual J48 decision rules, delivering identification results in under one second. Average completion time: approximately 30 seconds per flower.



Next Steps: Proceed to Visitor Testing

Immediate Actions



Project Validated

The combination of **high accuracy (96%)**, simplified user experience (only 2 measurements), interpretable decision rules, and low implementation cost makes this an excellent candidate for full deployment.

Estimated Budget: £5,000–10,000 for pilot testing phase, with expansion to 20+ species feasible within 6 months.



Iris Flower Identifier

Botanical Garden Society - Visitor App Prototype

Petal Length (cm)

e.g., 4.5



Measure from base to tip of petal

Petal Width (cm)

e.g., 1.5

Measure at widest point of petal



Identify Flower

Iris Flower Identifier

Botanical Garden Society - Visitor App Prototype

Petal Length (cm)

5

Measure from base to tip of petal

Petal Width (cm)

4

Measure at widest point of petal



Identify Flower



Iris Virginica

Large, showy petals. Found in wetland areas.

link to the project document:

[https://drive.google.com/drive/folders/lJq7enQT
Vxf3pZVQ_imGH_XTYKIwoCgMS?usp=sharing](https://drive.google.com/drive/folders/lJq7enQT
Vxf3pZVQ_imGH_XTYKIwoCgMS?usp=sharing)

Thank you

