Project Title

Iris Flower Classification

Name: NIRANJAN KUMAR

Roll no- 2312res418

Date: 21 August 2025

Problem Statement

- The iris flower, scientifically known as Iris, is a distinctive genus of flowering plants. Within this genus, there are three primary species: Iris setosa, Iris versicolor, and Iris virginica. These species exhibit variations in their physical characteristics, particularly in the measurements of their sepal length, sepal width, petal length, and petal width.

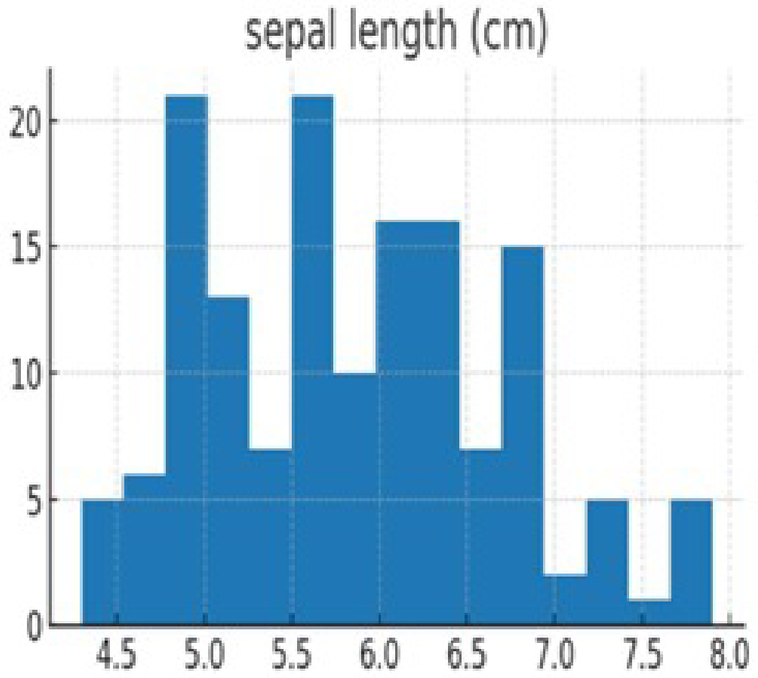
Objective

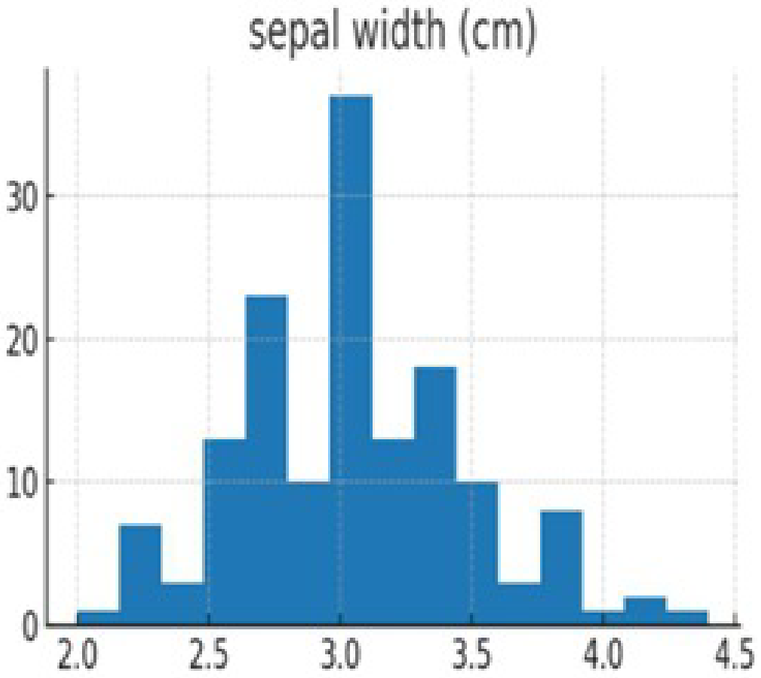
- The objective of this project is to develop a machine learning model capable of learning from the measurements of iris flowers and accurately classifying them into their respective species. The model's primary goal is to automate the classification process based on the distinct characteristics of each iris species.

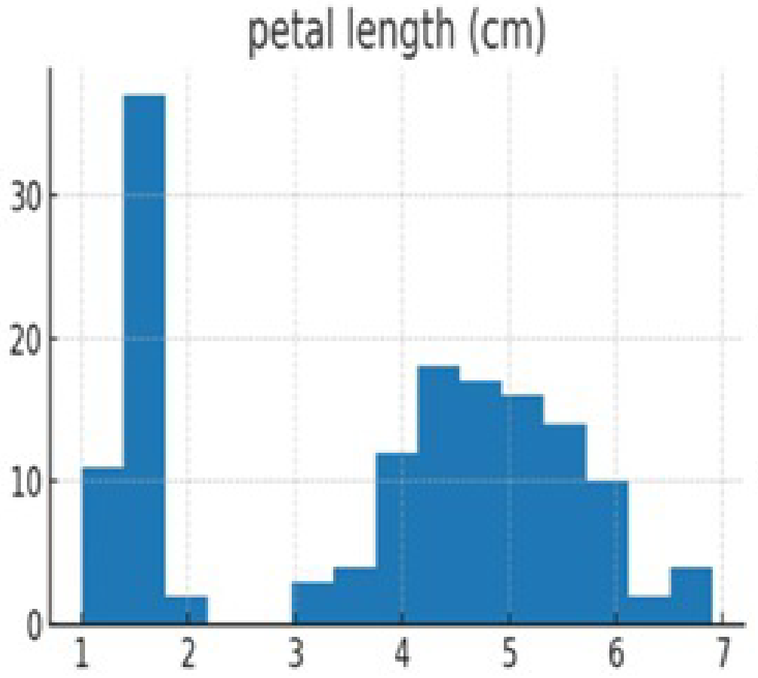
Project Details

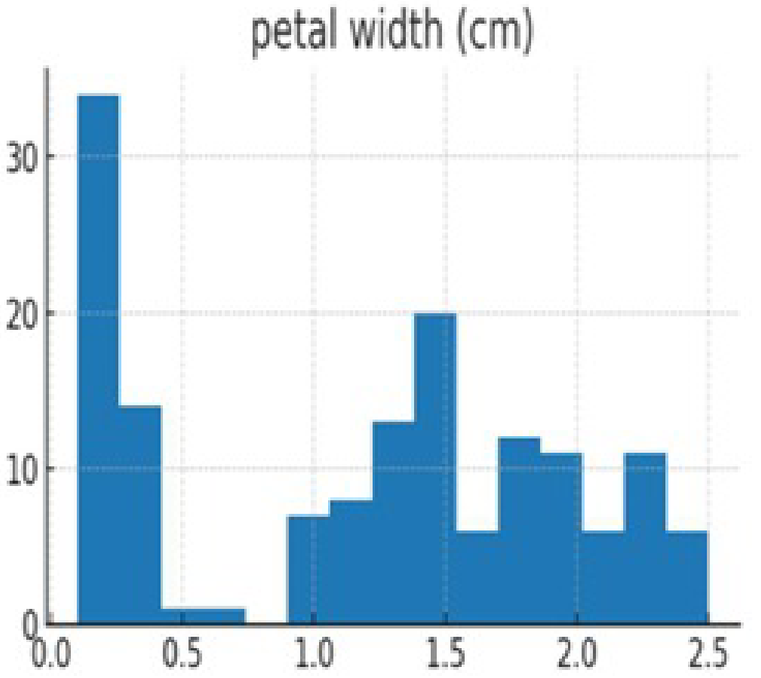
- **Iris Species:** The dataset consists of iris flowers, specifically from the species setosa, versicolor, and virginica.
- **Key Measurements:** The essential characteristics used for classification include sepal length, sepal width, petal length, and petal width
- **Machine Learning Model:** The project involves the creation and training of a machine learning model to accurately classify iris flowers based on their measurements

Feature Distributions



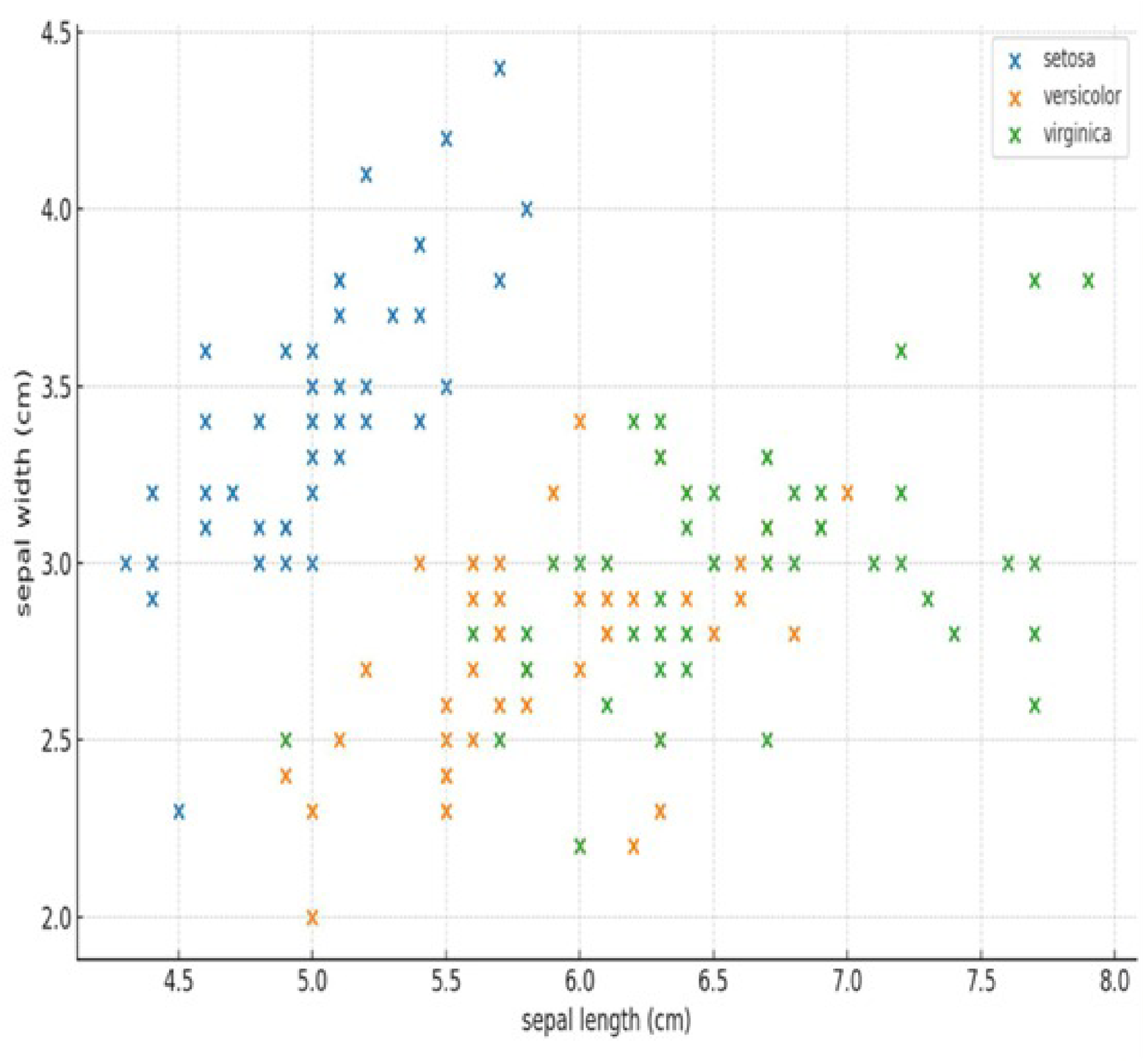






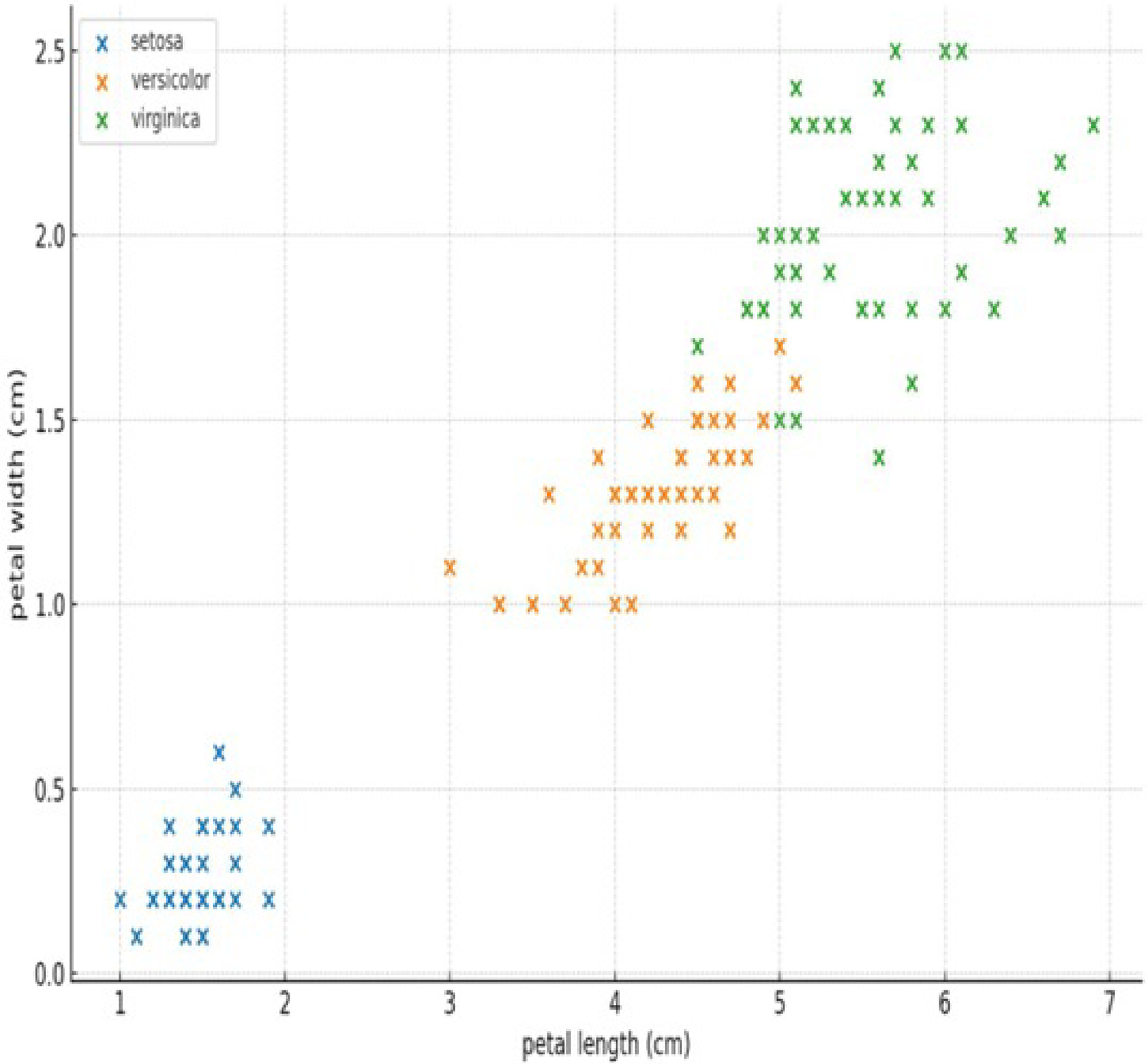
Sepal length (cm)	Sepal width (cm)	Petal length (cm)	Petal width (cm)
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2

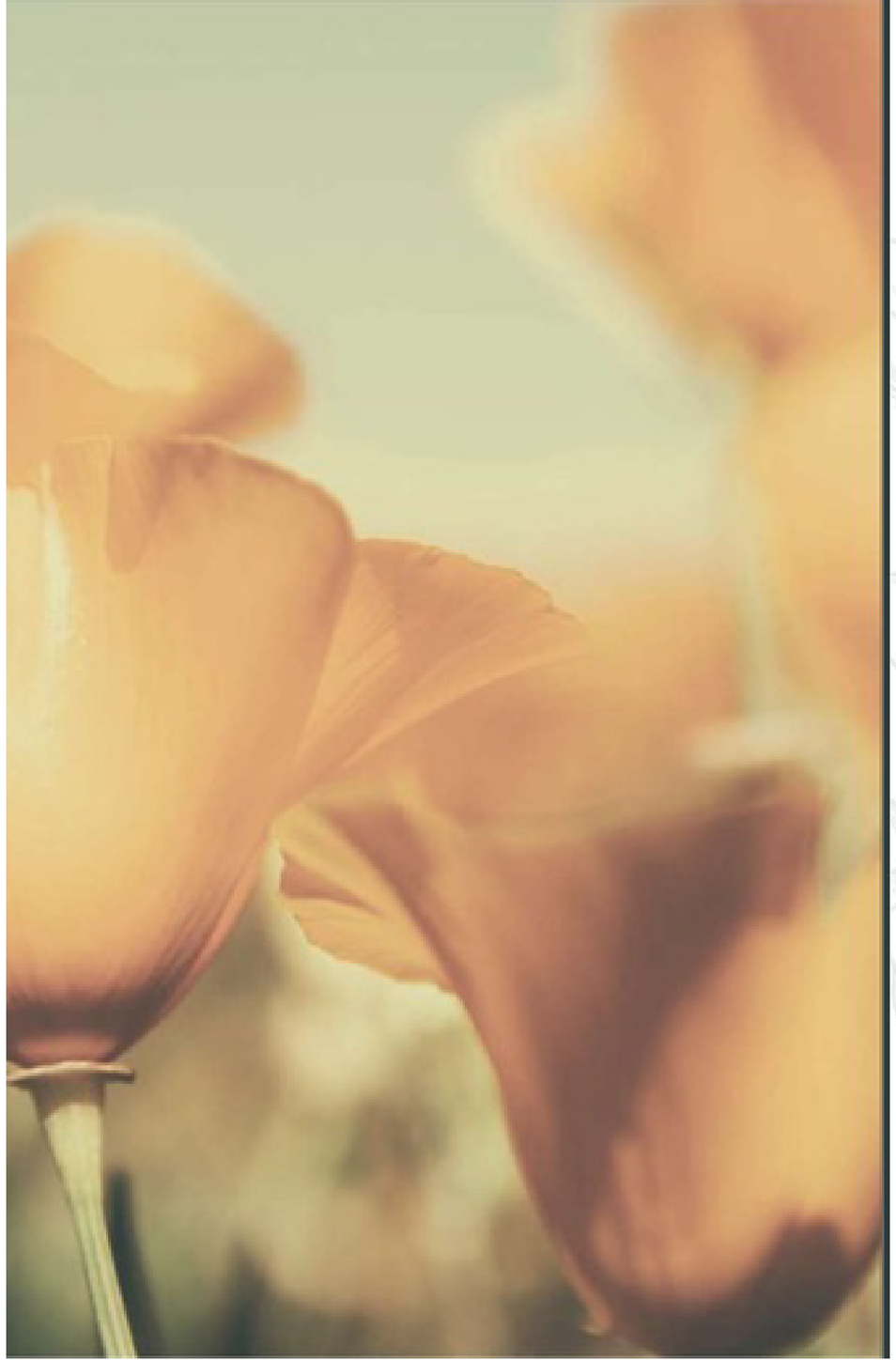
Dataset Snapshot (first 5 rows)



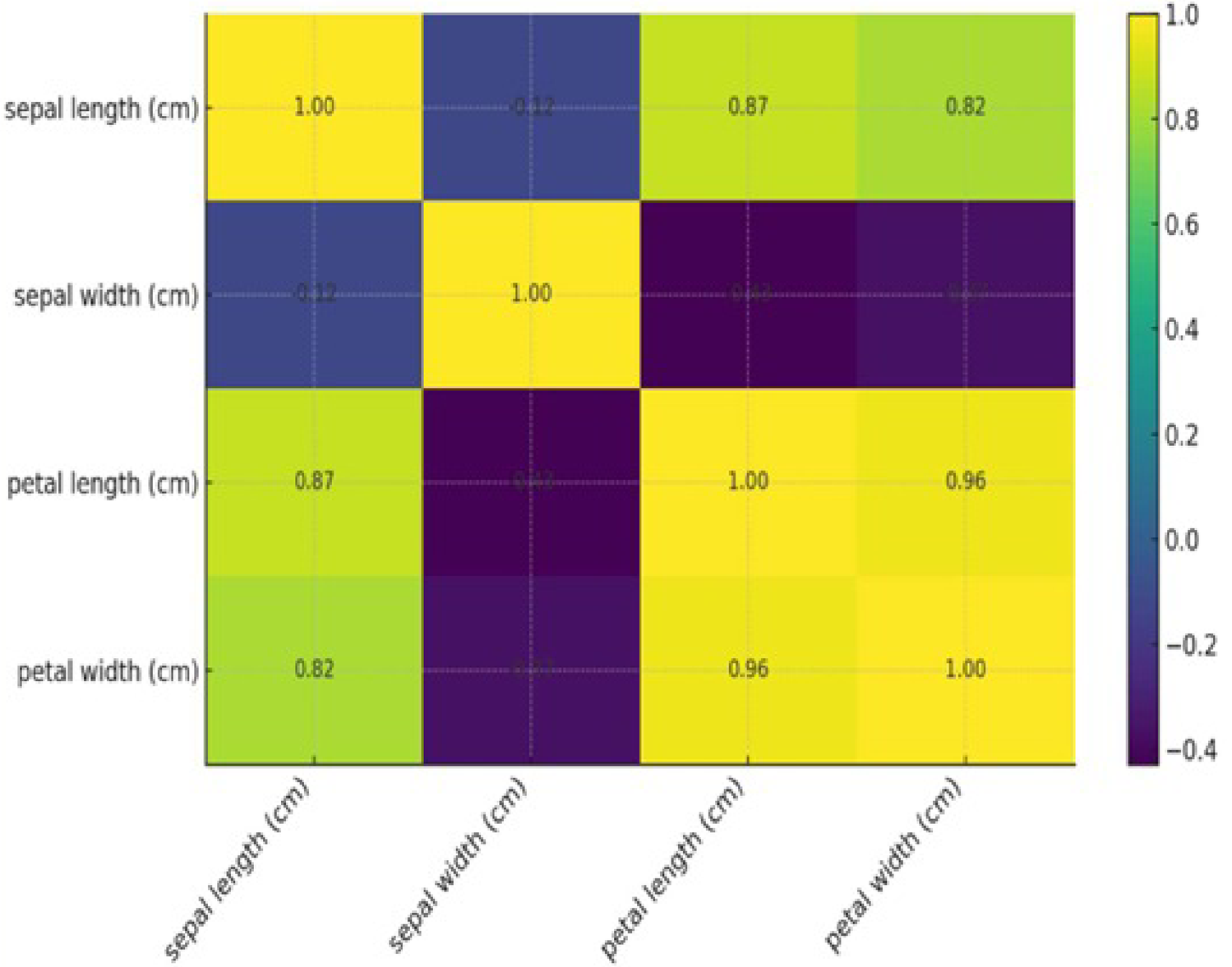
Sepal Analysis (Length vs Width)

Petal Analysis (Length vs Width)





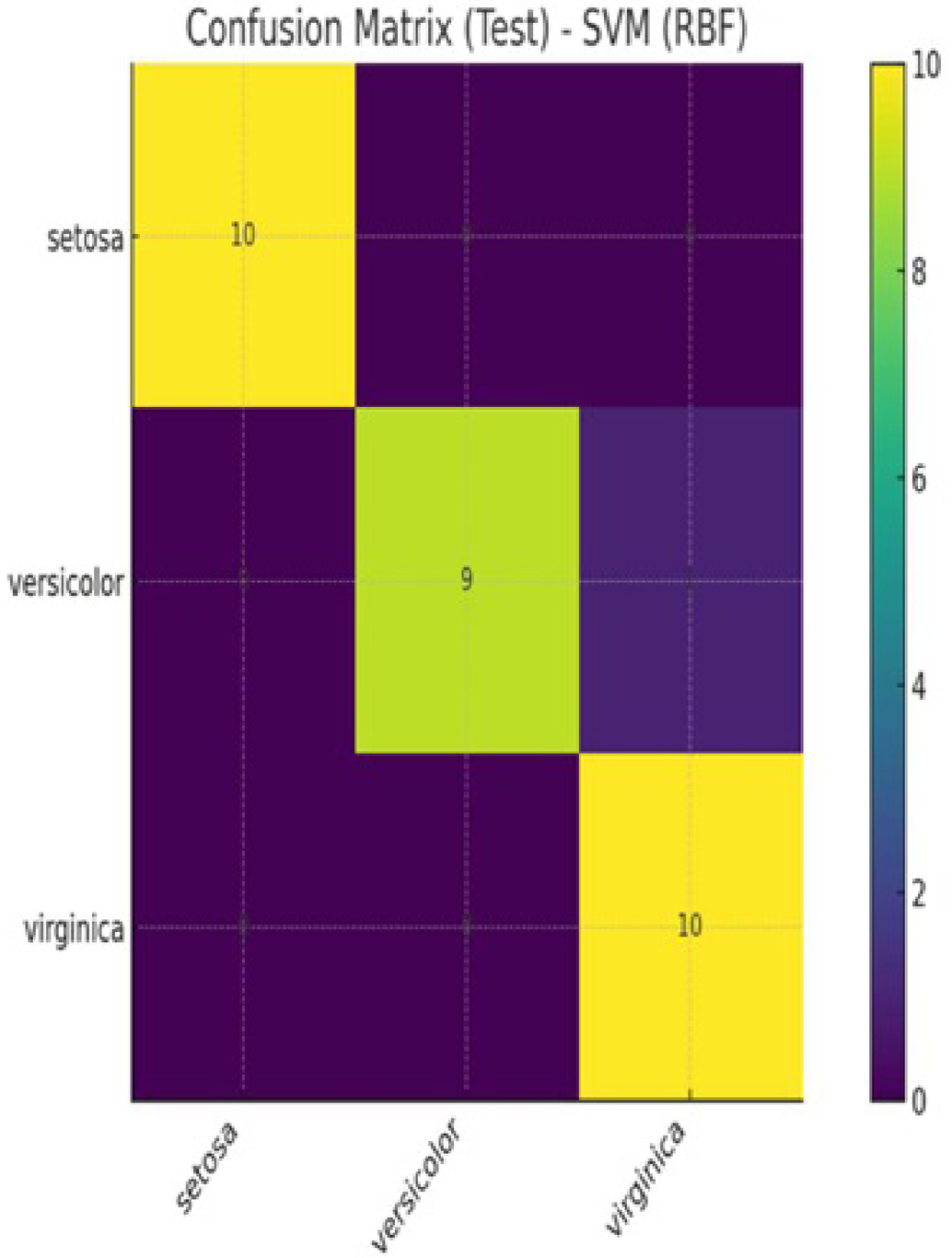
Correlation Heatmap



Model Building Process

	Logistic regression	Logistic regression tuned	Decision Tree	Decision Tree tuned	Random Forest	Random Forest tuned	SVM	SVM tuned	XGB	XGB tuned	Naive Bayes	Naive Bayes tuned	Neural Network	Neural Network tuned
Precision Train	0.980952	0.990741	1.000000	0.954548	1.000000	0.971693	0.980952	0.980952	1.000000	1.000000	0.942857	0.942857	0.980952	0.990741
Precision Test	0.979167	0.979167	0.979167	0.960784	0.979167	0.979167	0.979167	0.979167	0.979167	0.979167	0.979365	0.979365	0.960784	0.979167
Recall Train	0.980952	0.990476	1.000000	0.952381	1.000000	0.971429	0.980952	0.980952	1.000000	1.000000	0.942857	0.942857	0.980952	0.990476
Recall Test	0.977778	0.977778	0.977778	0.955556	0.977778	0.977778	0.977778	0.977778	0.977778	0.977778	0.977778	0.977778	0.955556	0.977778
Accuracy Train	0.980952	0.990476	1.000000	0.952381	1.000000	0.971429	0.980952	0.980952	1.000000	1.000000	0.942857	0.942857	0.980952	0.990476
Accuracy Test	0.977778	0.977778	0.977778	0.955556	0.977778	0.977778	0.977778	0.977778	0.977778	0.977778	0.977778	0.977778	0.955556	0.977778
F1 macro Train	0.980952	0.990478	1.000000	0.952353	1.000000	0.971434	0.980952	0.980952	1.000000	1.000000	0.942857	0.942857	0.980952	0.990478
F1 macro Test	0.977692	0.977692	0.977692	0.955093	0.977692	0.977692	0.977692	0.977692	0.977692	0.977692	0.977806	0.977806	0.955093	0.977692

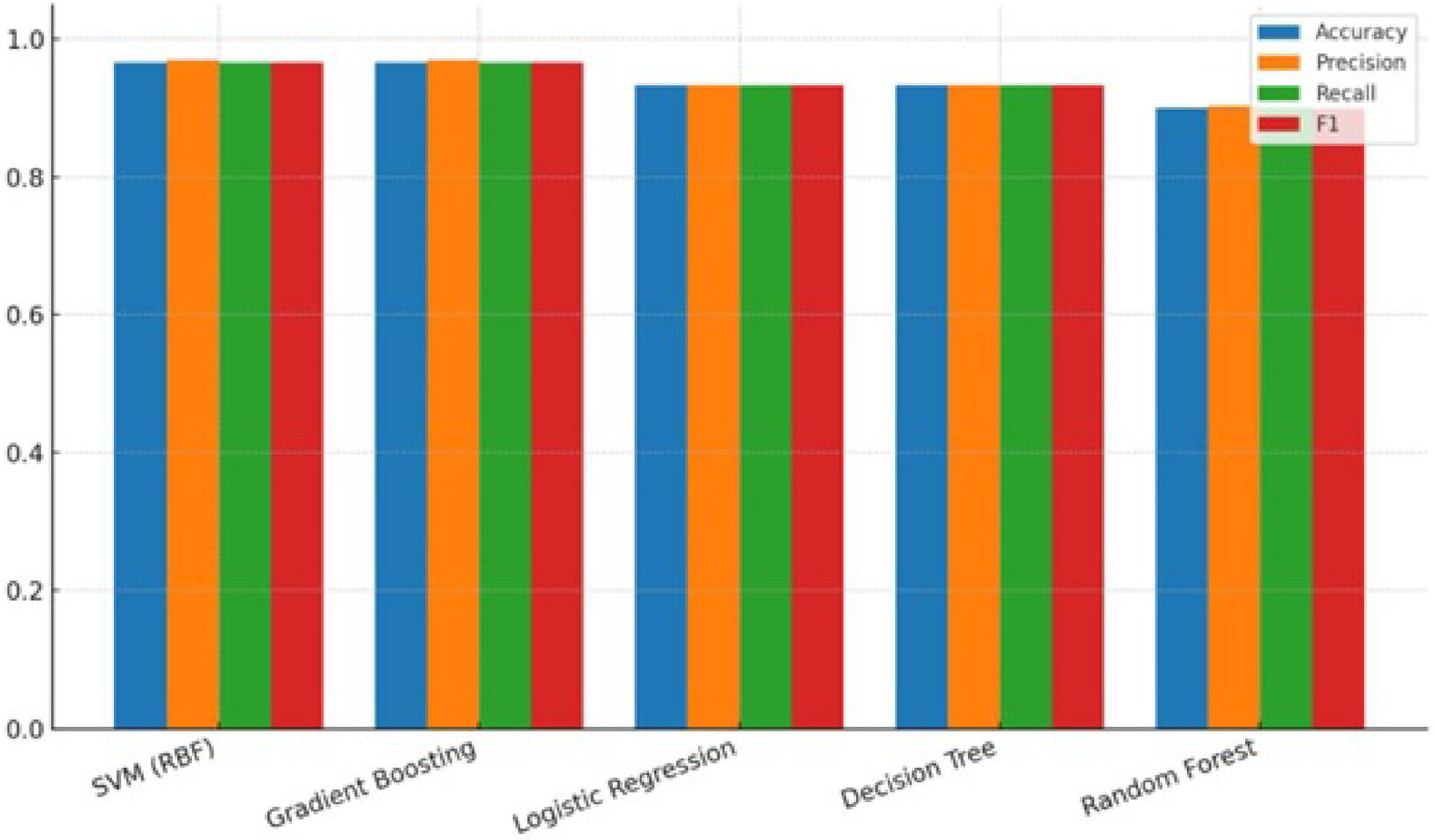
Confusion Matrix (Test) - SVM(RBF)



Insights & Results

- Best Model: SVM (RBF) (Accuracy: 96.67%, F1: 96.66%)
- Petal features show strongest separability; Setosa is linearly separable.
- Tree ensembles perform robustly without scaling; LR/SVM benefit from scaling.

Model Comparison: Accuracy, Precision, Recall, F1



Future Scope

- Add more flower species and real-world field data.
- Hyperparameter tuning and cross-validation.
- Export model and deploy a simple web app for demo.

Conclusion

- ML can classify Iris species with high accuracy. Comparative evaluation helps select a reliable model. Visualizations confirm strong feature patterns, especially in petal measurements.

Thank you