Universal Analyst Report

Comprehensive Analysis of iris Dataset

Report Date: 2025-08-21 12:39:39
Dataset Name: iris
Analysis Type: Full EDA & Modeling
Report Version: 1.0

Universal Analyst Model Report

Date of Analysis: 2025-08-21 12:39:37

Dataset: iris

Step 1: Dataset Overview

- info: Metadata from preprocessing step

Step 2: Exploratory Data Analysis (EDA)

Exploratory Data Analysis Report

Dataset Overview

- Number of rows: 150
- Number of columns: 5

Summary Statistics

Numerical Features

	count	mean	median	std	min	max	skew	kurtosis
sepallength	150	5.84333	5.8	0.828066	4.3	7.9	0.311753	-0.573568
sepalwidth	150	3.054	3	0.433594	2	4.4	0.330703	0.241443
petallength	150	3.75867	4.35	1.76442	1	6.9	-0.271712	-1.39536
petalwidth	150	1.19867	1.3	0.763161	0.1	2.5	-0.103944	-1.33525

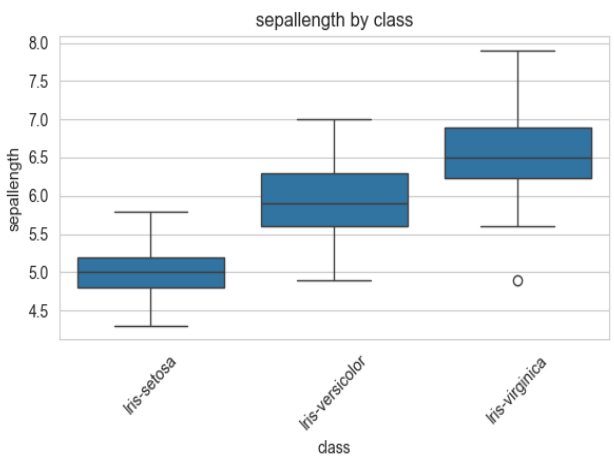
Categorical Features

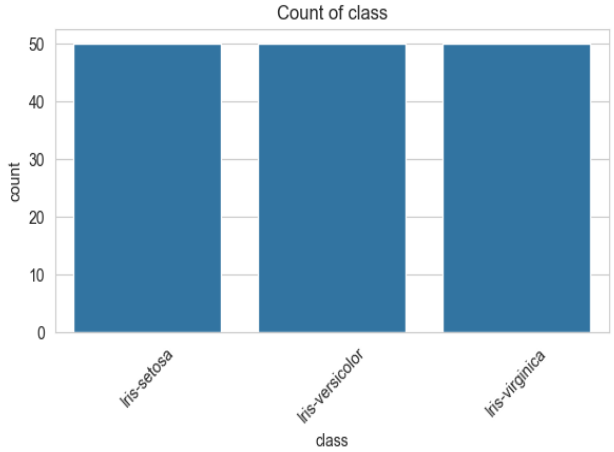
	unique_count	mode_freq	missing
class	3	50	0

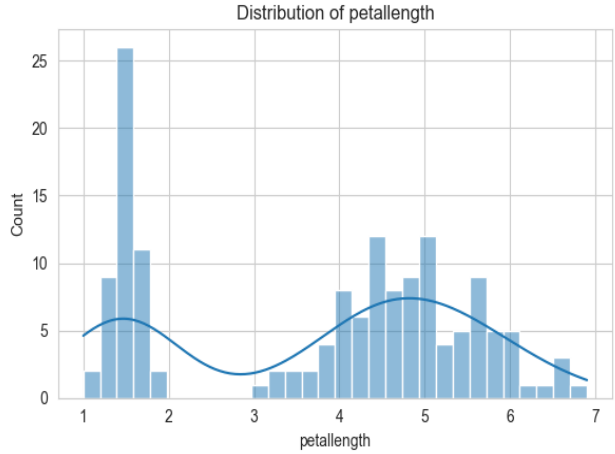
Key Insights

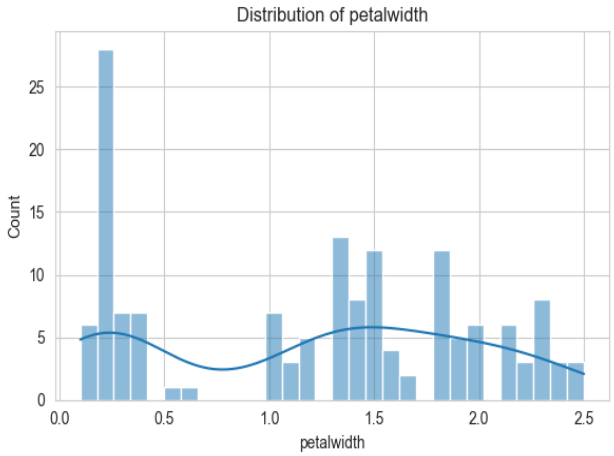
- Feature 'sepalwidth' has 4 potential outliers.
- Features 'sepallength' and 'petallength' have strong correlation: 0.87
- Features 'sepallength' and 'petalwidth' have strong correlation: 0.82
- Features 'petallength' and 'petalwidth' have strong correlation: 0.96
- Features with high variance: petallength

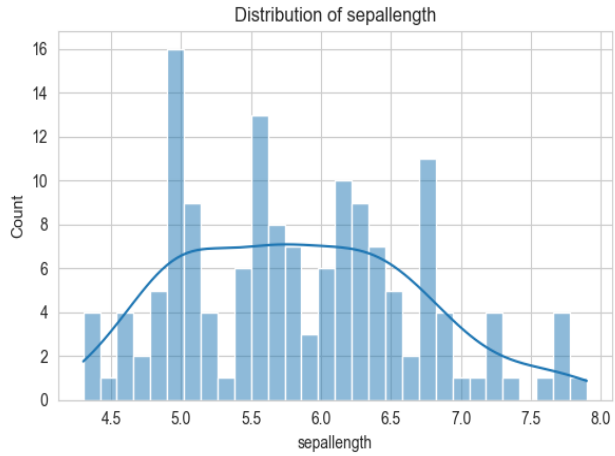
Visualizations

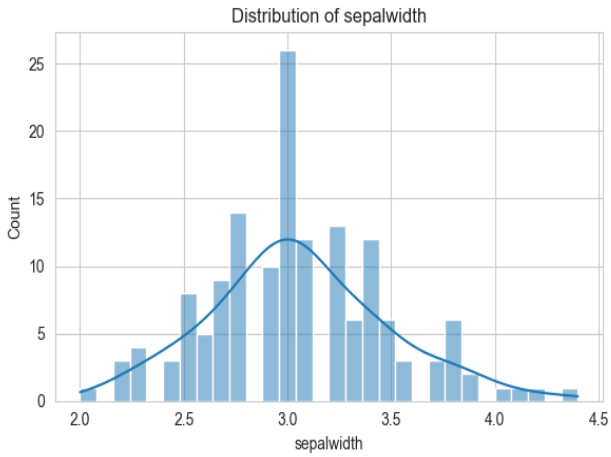


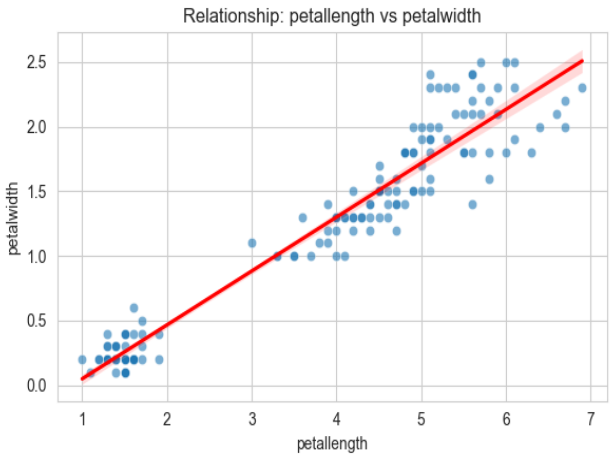












Step 3: Insight Extraction

Data Insight Report

Dataset Summary

- Number of rows: 150
- Number of columns: 5
- Target column: class
- Problem type: classification

Top Influential Features

- petallength: Mutual Information Score = 0.9926
- petalwidth: Mutual Information Score = 0.9856
- sepallength: Mutual Information Score = 0.5114
- sepalwidth: Mutual Information Score = 0.2898

Summary Statistics of Top Features

- petallength: Mean = 3.7587, Median = 4.3500, Std = 1.7644
- petalwidth: Mean = 1.1987, Median = 1.3000, Std = 0.7632
- sepallength: Mean = 5.8433, Median = 5.8000, Std = 0.8281
- sepalwidth: Mean = 3.0540, Median = 3.0000, Std = 0.4336

Outlier Counts per Numeric Feature

- sepallength: 0 outliers detected
- sepalwidth: 4 outliers detected
- petallength: 0 outliers detected
- petalwidth: 0 outliers detected

Next Steps

- Consider building predictive models using the identified influential features.

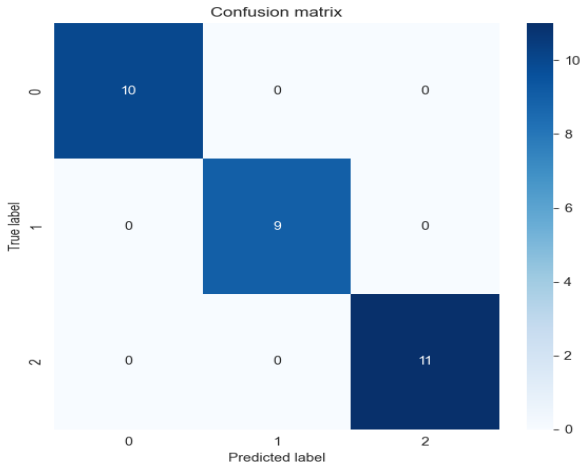
Step 4: Modeling and Prediction

Model Evaluation Report

Problem type: classification

LogisticRegression

- accuracy: 1.0000
- precision: 1.0000
- recall: 1.0000

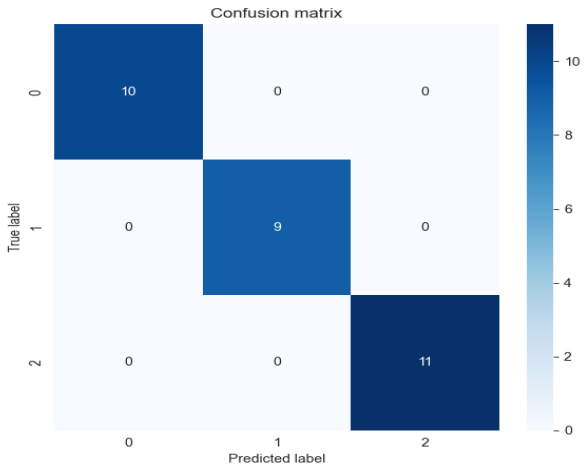


- f1_score: 1.0000

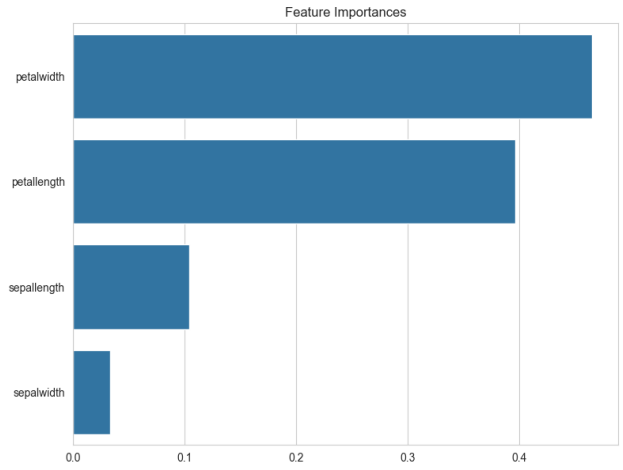
Feature Importance

RandomForestClassifier

- accuracy: 1.0000
- precision: 1.0000
- recall: 1.0000

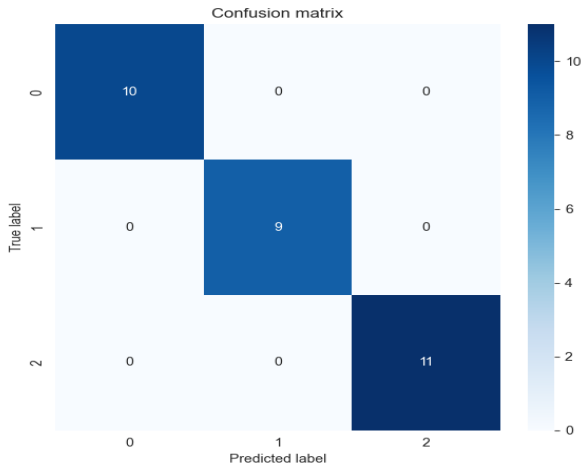


- f1_score: 1.0000



SVC

- accuracy: 1.0000
- precision: 1.0000
- recall: 1.0000



- f1_score: 1.0000

Feature Importance

Conclusion

This report summarizes the data ingestion, preprocessing, exploratory analysis, insights, and modeling results. Further analysis and model tuning may be required based on business needs.