

# Autonomous DS Crew - cleaned Analysis

Generated by Autonomous Data Science Crew | 2026-02-22 11:49:19

Dataset: **cleaned** | Task: **classification**

## Terminology Guide

Read these terms first. They are used throughout this report.

Term	Meaning	Why It Matters
<b>Target Column</b>	The column the model is trying to predict.	Defines the learning objective and evaluation logic.
<b>Feature</b>	An input variable used by the model for prediction.	Feature quality strongly drives model performance.
<b>Cross Validation (CV)</b>	Repeated train/validation splits used to estimate stability.	Helps detect models that perform well only on one split.
<b>accuracy</b>	Primary score used to compare candidate models.	Higher value indicates better predictive performance for this task.
<b>Overfitting Gap</b>	Difference between train score and test score.	Large gap means the model memorizes training data and may generalize poorly.

## Dataset Overview



## Feature Explanations

Feature	Type	Null %	Unique	Role	What it means
sepal length (cm)	float64	0.0	35	feature	Numeric feature used as a quantitative signal.
sepal width (cm)	float64	0.0	23	feature	Numeric feature used as a quantitative signal.
petal length (cm)	float64	0.0	43	feature	Numeric feature used as a quantitative signal.
petal width (cm)	float64	0.0	22	feature	Numeric feature used as a quantitative signal.
target	int64	0.0	3	target	Prediction target column used by the model.

## Exploratory Data Analysis

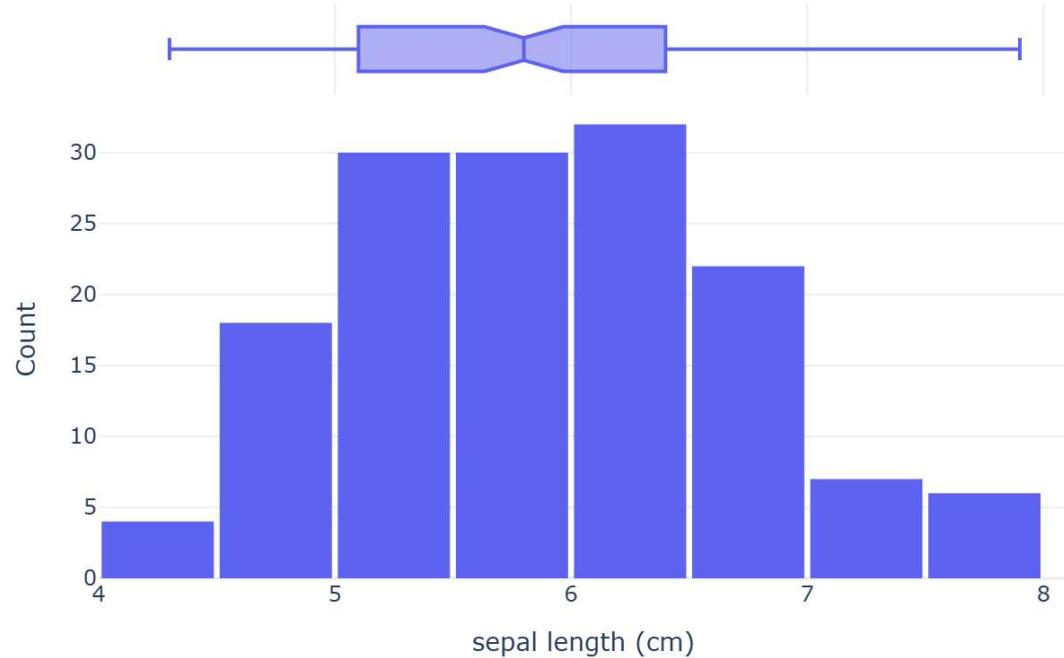
EDA found 5 numeric and 0 categorical features. Top correlation pair: petal length (cm) vs petal width (cm) (0.9628).

### Chart Findings

#### Distribution: sepal length (cm)

sepal length (cm): mean=5.844, median=5.8, std=0.831. Use this chart to inspect spread and skew.

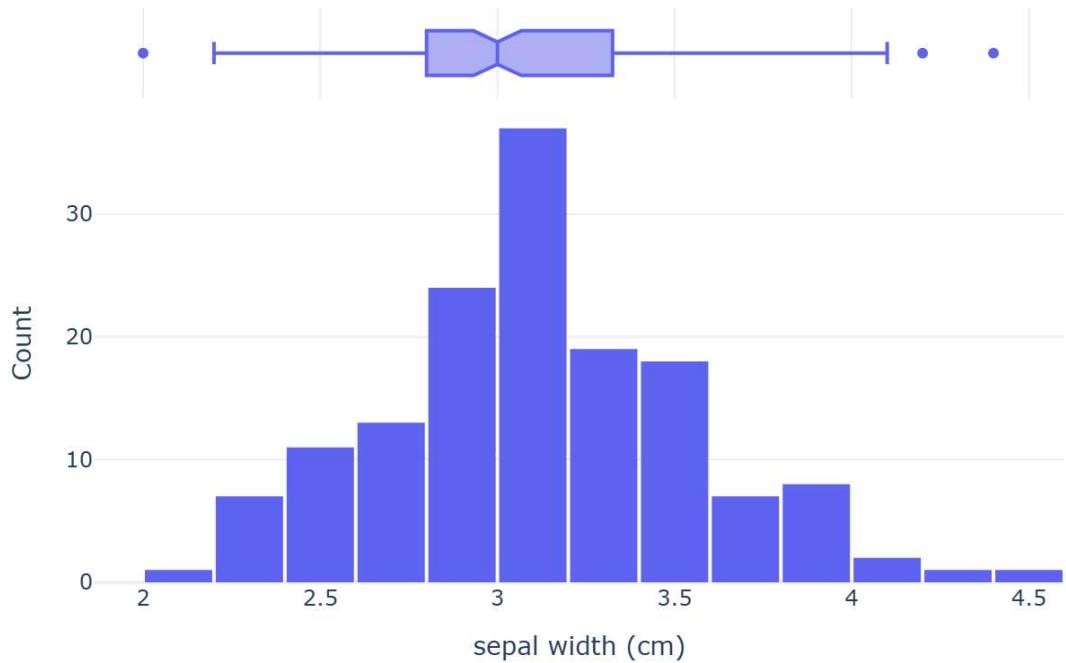
Distribution of sepal length (cm)



#### Distribution: sepal width (cm)

sepal width (cm): mean=3.06, median=3.0, std=0.436. Use this chart to inspect spread and skew.

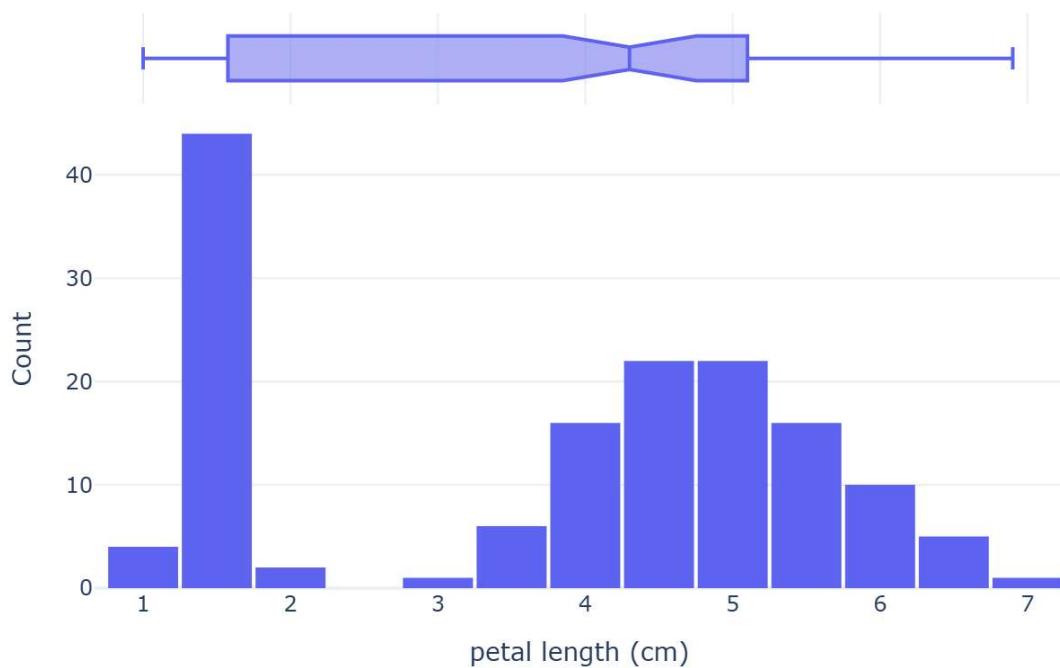
Distribution of sepal width (cm)



Distribution: petal length (cm)

petal length (cm): mean=3.749, median=4.3, std=1.768. Use this chart to inspect spread and skew.

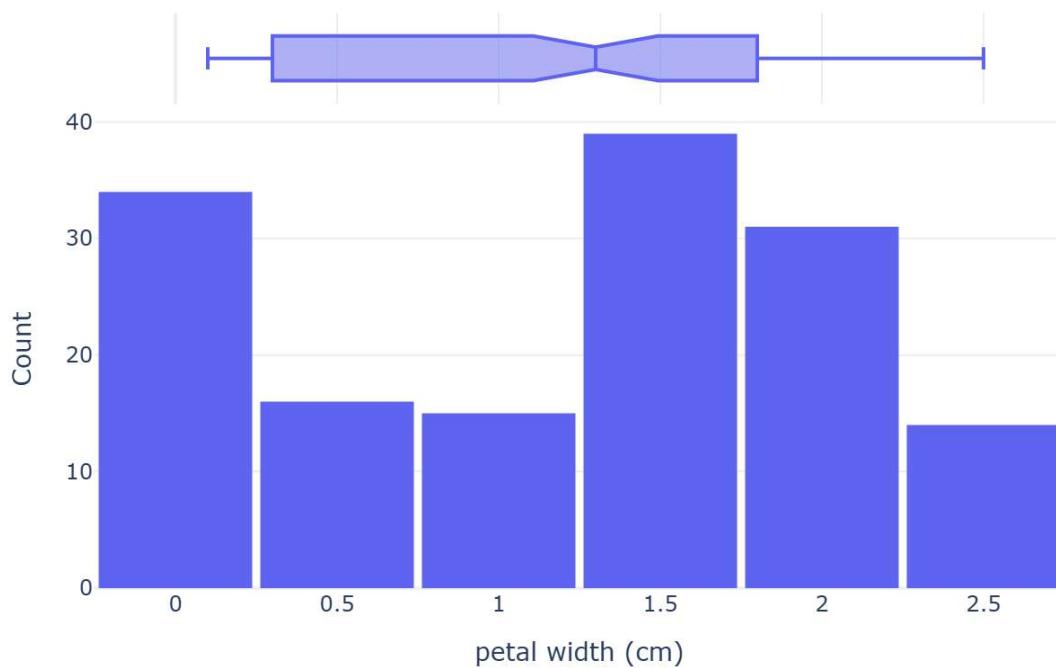
### Distribution of petal length (cm)



### Distribution: petal width (cm)

petal width (cm): mean=1.195, median=1.3, std=0.763. Use this chart to inspect spread and skew.

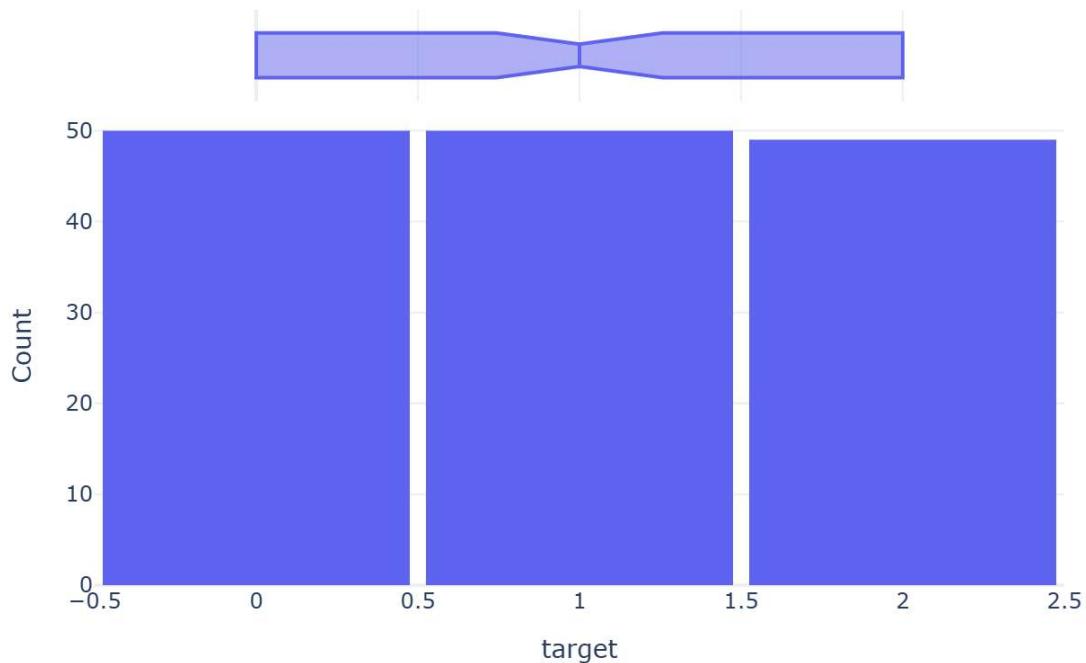
Distribution of petal width (cm)



### Distribution: target

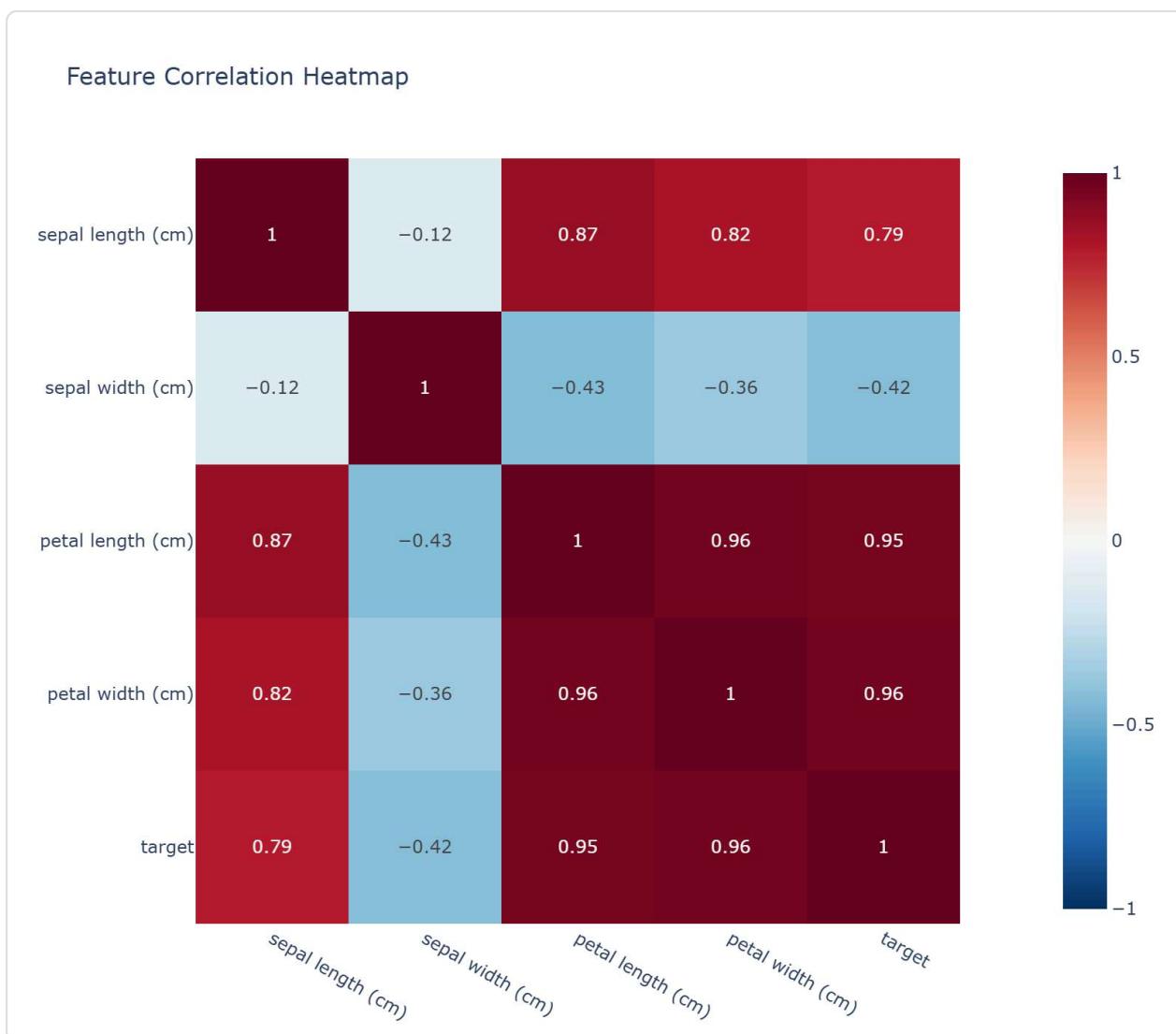
target: mean=0.993, median=1.0, std=0.818. Use this chart to inspect spread and skew.

### Distribution of target



### Correlation Heatmap

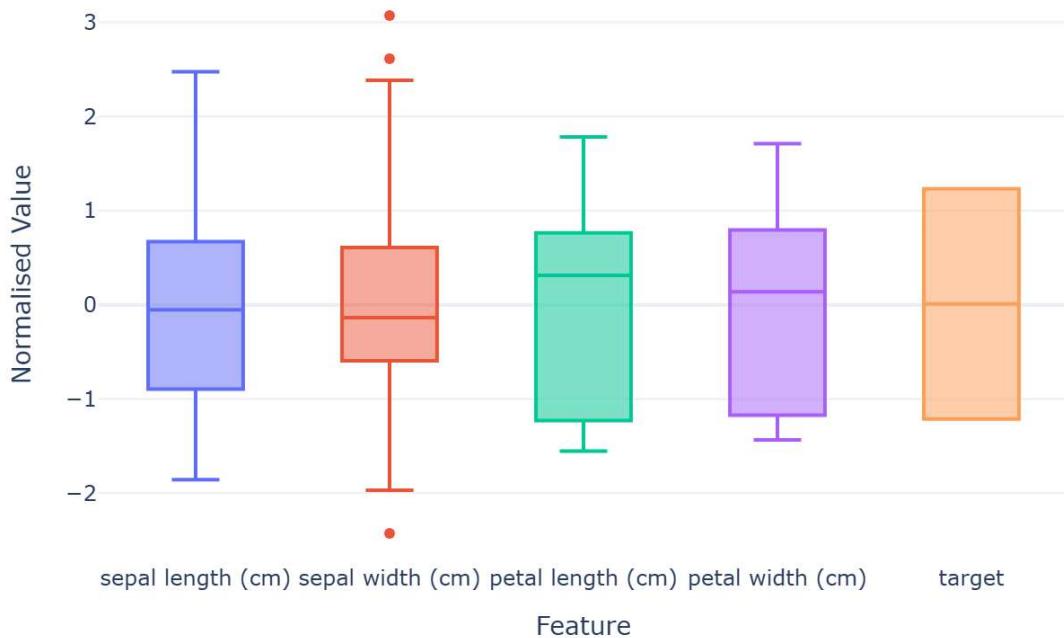
Top correlation pair: petal length (cm) vs petal width (cm) (0.9628).



## Outlier Box Plot

Outlier box plot highlights extreme values. Review long tails and unusual ranges.

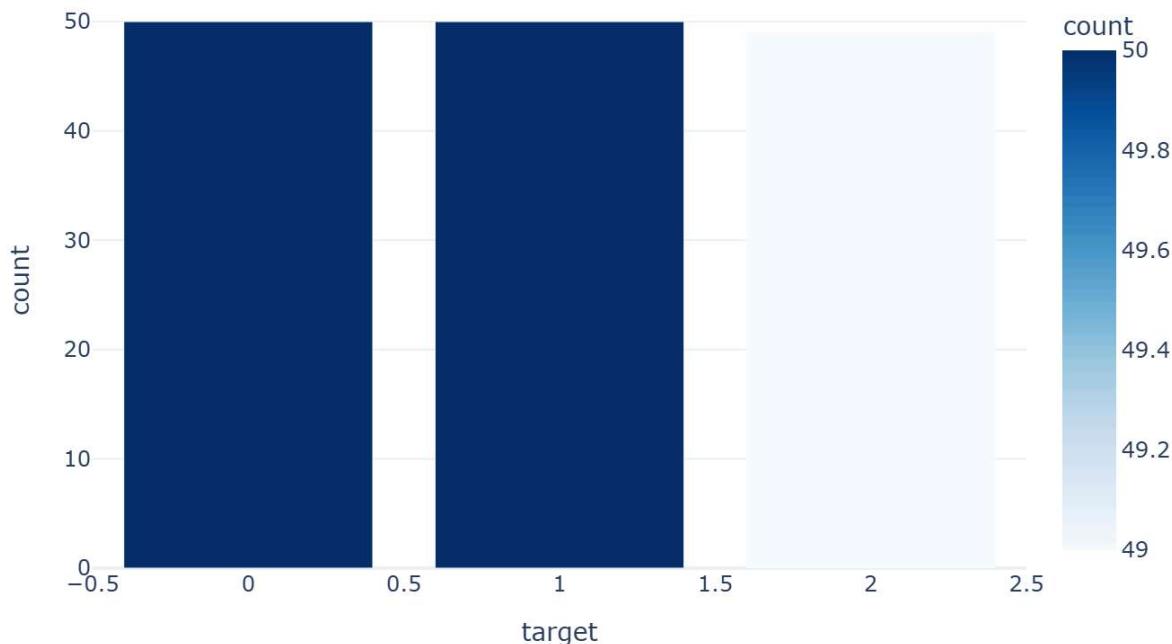
## Outlier Detection — Normalised Box Plots



## Target Distribution

Target distribution shows class/value balance: 0: 33.56%, 1: 33.56%, 2: 32.89%.

### Target Distribution — target



## Model Performance

### Best Model: GradientBoosting



### Model Leaderboard

Rank	Model	CV Score	Test Score	Status
1	<b>GradientBoosting</b>	0.9319	0.9667	<b>Best</b>

Rank	Model	CV Score	Test Score	Status
2	<b>SVM</b>	0.9652	0.9667	Evaluated
3	<b>LogisticRegression</b>	0.9656	0.9333	Evaluated
4	<b>RandomForest</b>	0.9319	0.9333	Evaluated
5	<b>ExtraTrees</b>	0.9486	0.9333	Evaluated

AutoML selected GradientBoosting as best by accuracy. Top feature signals:  
num\_sepal length (cm), num\_sepal width (cm), num\_petal length (cm), num\_petal width (cm).

## Overfitting Assessment

**0.9916**

TRAIN SCORE

**1.0**

TEST SCORE

**-0.85%**

GAP

NONE

SEVERITY

Model generalises well. No overfitting detected.

## Key Insights and Recommendations

Finding 1: Dataset has 149 rows and 5 columns. Finding 2: Target 'target' unique values: 3. Distribution: 0: 33.56%, 1: 33.56%, 2: 32.89%. Finding 3: Best model: GradientBoosting with accuracy=0.9667. Recommended actions: Remove or combine highly correlated features to reduce instability..

### Key Findings

- Dataset has 149 rows and 5 columns.
- Target 'target' unique values: 3. Distribution: 0: 33.56%, 1: 33.56%, 2: 32.89%.
- Best model: GradientBoosting with accuracy=0.9667.
- Model quality is strong; overfitting severity is none.
- Top correlation pair: petal length (cm) vs petal width (cm) (0.9628).
- Top predictive features: num\_sepallength (cm), num\_sepawidth (cm), num\_petallength (cm), num\_petalwidth (cm).

### Recommended Actions

- Remove or combine highly correlated features to reduce instability.

## What DS Crew Did

- Loaded and validated dataset 'cleaned.csv'.
- Cleaned data and produced analysis dataset with 149 rows.
- Ran EDA and generated visual chart artifacts.
- Trained candidate models and selected GradientBoosting.
- Evaluated model quality with accuracy=0.9667 and overfitting severity=none.
- Compiled findings and produced one HTML report with embedded charts and explanations.

## Top Predictive Features

Rank	Feature	Importance
1	num_sepallength (cm)	0.0032
2	num_sepawidth (cm)	0.0158
3	num_petal length (cm)	0.3491
4	num_petal width (cm)	0.6319

Autonomous Data Science Crew | Powered by CrewAI + Groq + MLflow

Report generated: 2026-02-22 11:49:19