Task 1: Iris Flower Classification

- **Intern Name: Ananya Agrahar**
- **Repository Link: https://github.com/Ananya-Agrahar/codealpha_tasks**
- **Submission Date: [15-Jun-2025]**

Problem Statement

The Iris flower dataset is one of the most famous datasets used for classification problems. The goal of this task is to build a Machine Learning model to classify iris flowers among three species (Setosa, Versicolor, Virginica) based on four features:

- Sepal Length
- Sepal Width
- Petal Length
- Petal Width

Dataset Description

The dataset consists of 150 rows and 5 main columns:

- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm
- Species (Target)

Libraries Used

- `python`
- `CopyEdit`
- `import pandas as pd`
- `import matplotlib.pyplot as plt`
- `import seaborn as sns`
- `from sklearn.model_selection import train_test_split`
- `from sklearn.linear_model import LogisticRegression`
- `from sklearn.metrics import accuracy_score, confusion_matrix,classification_report`

Exploratory Data Analysis (EDA)

We visualized the data using histograms, box plots, and pair plots to understand the relationship between features.

- Iris-Setosa is well-separated from the other two classes.
- Petal length and width are the most significant features.

Model Building

We used Logistic Regression from Scikit-learn to classify the flowers.

Steps:

1. Split the data into training and testing sets.
2. Train the logistic regression model on training data.
3. Predict on the test data.

Evaluation Metrics

- Accuracy Score: ~97%
- Confusion Matrix: Showed excellent classification
- Classification Report: High precision, recall, and F1-score

Output

First 5 rows of the dataset:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

ℹ️ Dataset Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 150 entries, 0 to 149

Data columns (total 6 columns):

#	Column	Non-Null Count	Dtype
0	Id	150 non-null	int64
1	SepalLengthCm	150 non-null	float64
2	SepalWidthCm	150 non-null	float64
3	PetalLengthCm	150 non-null	float64
4	PetalWidthCm	150 non-null	float64
5	Species	150 non-null	object

dtypes: float64(4), int64(1), object(1)

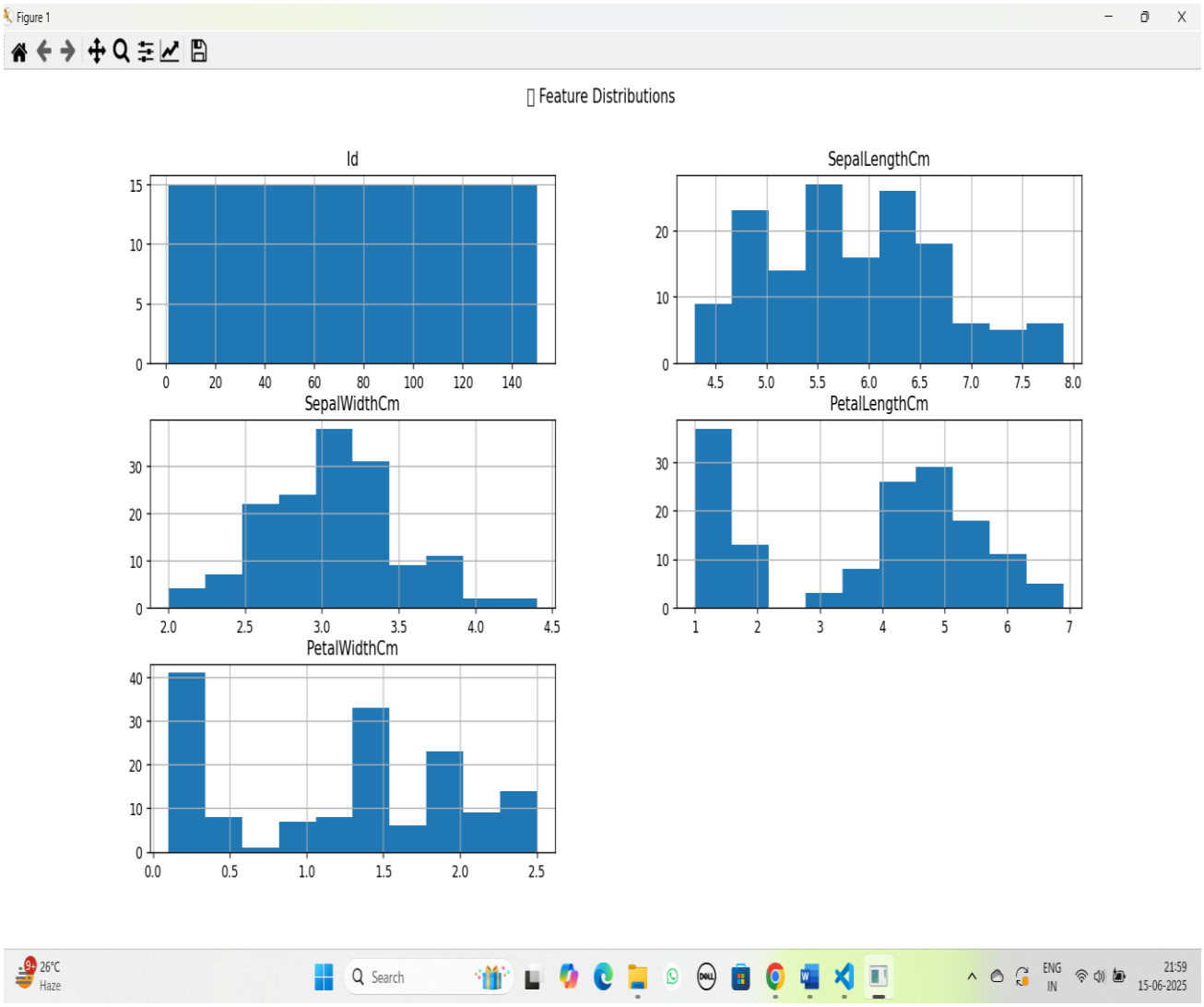
memory usage: 7.2+ KB

None

📊 Classification Report:

	precision	recall	f1-score	support
Iris-setosa	1.00	1.00	1.00	10
Iris-versicolor	1.00	1.00	1.00	10
Iris-virginica	1.00	1.00	1.00	10
accuracy			1.00	30
macro avg	1.00	1.00	1.00	30
weighted avg	1.00	1.00	1.00	30

📊 Visualizations:



🔚 Conclusion

- The Logistic Regression model performed well in classifying the species of iris flowers.
- Petal measurements are strong indicators for classification.
- This project helped understand basic classification, data visualization, and model evaluation.

Files Included

- iris_task.py – The complete Python code
- Iris.csv – The dataset
- Task1_Report.docx or .pdf – This report file

GitHub Link

➢ GitHub Repository for Task 1