

OASIS INFOBYTE INTERNSHIP (Data Science)

Task 1: Iris Flower Classification

Name: Palak Parmar

College: Bansal Group Of Institute

Submission Date: 04/11/2025

Introduction: The Iris dataset contains measurements of iris flowers from three species (setosa, versicolor, virginica). This simple project shows how to use Python to build a model that learns from measurements and predicts the species. We keep the code minimal and suitable for beginners.

Libraries to install (run once in terminal): pip install numpy pip install scikit-learn pip install matplotlib

Code (copy and run in Python):

```
# Beginner-friendly Iris classification (no dataframe columns printed)

from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report
import matplotlib.pyplot as plt

# Load data (features and labels only)
iris = load_iris()
X = iris.data # features as a plain array
y = iris.target # labels as numbers 0,1,2

# Split into train and test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Scale features (recommended for many models)
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Create and train model
model = LogisticRegression(max_iter=200)
model.fit(X_train, y_train)
# Make predictions and show results
```

```
y_pred = model.predict(X_test)
print("Accuracy:", round(accuracy_score(y_test, y_pred), 4))
print("\nClassification Report:")
print(classification_report(y_test, y_pred))

# Simple plot: first two features (optional)
plt.scatter(X_test[:,0], X_test[:,1], c=y_pred, marker='o')
plt.xlabel('Feature 1 (sepal length)')
plt.ylabel('Feature 2 (sepal width)')
plt.title('Test data predictions (colored by predicted class)')
plt.show()
```

Expected Output (example): - A printed Accuracy value such as: Accuracy: 0.9667 - Classification Report showing precision, recall and f1-score for each class - A simple scatter plot (optional) showing predicted classes (colors) Note: - This code does not print dataframe columns or any extra boxes. - It uses plain arrays (X, y) for simplicity so beginners can focus on steps: load -> split -> scale -> train -> predict.

Conclusion: This beginner-friendly implementation demonstrates a complete pipeline for classification using the Iris dataset. You can edit your personal details on the first page and submit the PDF. No extra decorative boxes are included.