

Description of the DataBase that used in the exercice

Feature Name	Type	Role	Description
Sepal Length	Numerical (continuous)	Explanatory	Length of the sepal — the outer part of the flower that protects the petals.
Sepal Width	Numerical (continuous)	Explanatory	Width of the sepal.
Petal Length	Numerical (continuous)	Explanatory	Length of the petal — the inner colorful part of the flower.
Petal Width	Numerical (continuous)	Explanatory	Width of the petal.
Species	Categorical (nominal)	Target	Type (species) of the iris flower: <i>Iris-setosa</i> , <i>Iris-versicolor</i> , or <i>Iris-virginica</i> .

Question 1 : Load and Explore the Iris Dataset

Load the **Iris dataset** from the sklearn.datasets library into a pandas DataFrame.

```
##
```

```
from sklearn.datasets import load_iris
```

```
iris = load_iris(as_frame=True)
```

```
##
```

Display the first few rows of the dataset to understand its structure and check the unique values of the target variable.

Export the dataset to a CSV file named **Your_Name_IrisDatabase.csv** (use ; as the separator and do not include the index).

Question 2 : Describe the Data

Perform an initial data exploration to better understand the dataset.

- Display basic information (df.info() and df.describe())
- Count the number of samples for each target class
- Verify if there are any missing values
-

Question 3 : Split the Dataset

Split the dataset into training and testing subsets (80% / 20%) while maintaining class balance using the stratify parameter.

For example, if the class *Iris-setosa* represents 30% of the samples in the original dataset, this proportion should also be preserved in both the training and test subsets.

Question 4 : Build the Decision Tree

Train a **Decision Tree Classifier** on the training data using default parameters.

Then, evaluate it on the test data and display the accuracy score

Question 5 : Model Evaluation and Interpretation

Evaluate the model more thoroughly using a confusion matrix to analyze which classes are well or poorly predicted.

Question 6 : Controlling Model Complexity

Experiment with the following Decision Tree hyperparameters:

max_depth

min_samples_split

min_samples_leaf

criterion ('gini' or 'entropy')

Train different trees and observe how changing these parameters affects accuracy and overfitting.

Question 7 : Cross-Validation and Model Selection

Use **cross-validation** (`cross_val_score`) to evaluate the stability of your Decision Tree model for different parameter values.

Compare models with different `max_depth` or `ccp_alpha` values to choose the most reliable one.