FINAL PROYECT
IRIS SPECIES CLASSIFICATION

Learning objectives:

At the end of this activity the student will be able to:

a. *Integrate and demonstrate the knowledge acquired during the course through an end-to-end data mining project.*
b. *Apply independently the complete data mining pipeline: understanding, cleaning, modeling, and evaluation.*
c. *Select and justify the techniques and algorithms used in a classification problem.*
d. *Communicate the workflow and results through a dashboard.*

The Iris dataset is one of the most recognized benchmarks in machine learning. It contains 150 flower samples, each belonging to one of three species: *Iris setosa*, *Iris versicolor*, and *Iris virginica*.

Each sample includes four numerical features:

- Sepal length
- Sepal width
- Petal length
- Petal width

The goal of this project is to train a classification model capable of predicting the species of a flower based on these four measurements.

Instructions

1. <u>Design your own workflow:</u> Based on the knowledge acquired throughout the semester, you must design your own workflow that includes the steps you consider necessary to achieve a correct classification of the flowers (e.g., data understanding, preprocessing, modeling, validation, etc.).

   You may select a classification model such as Random Forest or any other algorithm you consider suitable for this type of problem.

2. <u>Develop an interactive dashboard:</u> Using Streamlit, create an interactive dashboard that visually communicates your project's results. The dashboard must include:

✓ Model metrics (Accuracy, Precision, Recall, F1).
✓ A panel where the user can enter the sepal and petal length/width to obtain a predicted species and visualize the result in a 3D scatter plot that shows the position of the new sample relative to the dataset.
✓ Additional visualizations (such as histograms or scatter matrices) that help explain the data.

   The dashboard should be well-organized, interactive, and user-friendly.

3. <u>Video presentation:</u> Record a short presentation video in wich you explain:

✓ The methodology (workflow) you designed for the classification model.
✓ The reasoning and justification behind your methodological choices.
✓ A demonstration of your dashboard, explaining how it works and how it visualizes the predictions.

Deliverables

➢ Link to the video presentation.
➢ A GitHub repository containing:
  - Proyect.py (Streamlit app or main script)
  - requirements.txt
  - README.md describing purpose and instructions.
  - All team members' names must appear in the GitHub repository and the deployment.

The activity must be completed in **groups of at least 2 and at most 4 participants**.

Best of luck!

"The three chief virtues of a programmer are: Laziness, Impatience and Hubris."
Larry Wall