



**INSTITUTO POLITÉCNICO NACIONAL**

ESCUELA SUPERIOR DE CÓMPUTO

INGENIERÍA EN SISTEMAS COMPUTACIONALES



## **Inteligencia Artificial**

### **Laboratorio 8: Datasets**

Marco Isaí Flores Vicencio

Jesús Pérez González

Sebastián Reyes Núñez

**Grupo:**

6CV3

## Introducción

En este laboratorio, se explora el uso y manejo de datasets en el análisis de datos mediante el conjunto de datos "Iris Plant". Este dataset contiene características morfológicas de tres especies de flores de iris: Iris setosa, Iris versicolor, e Iris virginica. A través de este laboratorio, se busca comprender el proceso de lectura y almacenamiento de un archivo de datos en una estructura adecuada, como una matriz o un dataframe, utilizando herramientas de programación.

## Marco Teórico

Un conjunto de datos es una colección organizada de información que se utiliza en la ciencia de datos para realizar análisis y extraer conclusiones. En el contexto de este laboratorio, el conjunto de datos "Iris Plant" permite aplicar conceptos de clasificación y análisis exploratorio, siendo un recurso ampliamente utilizado en la enseñanza y práctica del aprendizaje automático y el análisis estadístico.

Existen diferentes estructuras de datos que facilitan la manipulación y almacenamiento de información de datasets, entre las que destacan las matrices y los dataframes. Una matriz es una estructura de datos de tipo array multidimensional que se emplea para almacenar datos en filas y columnas, permitiendo realizar operaciones algebraicas de manera eficiente. En el contexto de la programación científica y análisis de datos, las matrices son especialmente útiles para almacenar valores numéricos que requieren operaciones rápidas y eficientes. Por otro lado, un 'dataframe' es una estructura de datos tabular que permite almacenar datos de diferentes tipos en columnas, como números, textos y valores booleanos.

## Desarrollo

Empezamos importando la librería 'pandas', la cuál se encarga de la manipulación y análisis de los datos.

```
import pandas as pd
```

Definimos el nombre del archivo que contiene el dataset. En este caso el archivo se llama 'bezdekIris.data'

```
file_name = 'bezdekIris.data'
```

Después definimos una cadena con los nombres de las columnas del dataset, en este caso se definen con las medidas de las flores y la clase a la que pertenece.

Una vez definida, lee el archivo usando 'pd.read\_csv'. Las opciones de 'header=None' y 'names=column\_names' indican que el archivo no tiene una línea de encabezado y asigna los nombres de las columnas definidas.

```
column_names = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'class']
iris_df = pd.read_csv(file_name, header=None, names=column_names)
```

Finalmente, agrupa el dataset por la columna clase y muestra las primeras 5 filas de cada clase. Así podemos visualizar algunos casos de cada clase dentro del dataset.

```
print(iris_df.groupby('class').head(5))
```

## Results

	sepal_length	sepal_width	petal_length	petal_width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
50	7.0	3.2	4.7	1.4	Iris-versicolor
51	6.4	3.2	4.5	1.5	Iris-versicolor
52	6.9	3.1	4.9	1.5	Iris-versicolor
53	5.5	2.3	4.0	1.3	Iris-versicolor
54	6.5	2.8	4.6	1.5	Iris-versicolor
100	6.3	3.3	6.0	2.5	Iris-virginica
101	5.8	2.7	5.1	1.9	Iris-virginica
102	7.1	3.0	5.9	2.1	Iris-virginica
103	6.3	2.9	5.6	1.8	Iris-virginica
104	6.5	3.0	5.8	2.2	Iris-virginica

PS C:\Users\Windows PC\OneDrive\Inteligencia Artificial>

	sepal_length	sepal_width	petal_length	petal_width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
50	7.0	3.2	4.7	1.4	Iris-versicolor
51	6.4	3.2	4.5	1.5	Iris-versicolor
100	6.3	3.3	6.0	2.5	Iris-virginica
101	5.8	2.7	5.1	1.9	Iris-virginica

PS C:\Users\Windows PC\OneDrive\Inteligencia Artificial>

## Conclusiones

En este laboratorio, se logró comprender y aplicar técnicas básicas de manipulación de datos mediante el uso del conjunto de datos "Iris Plant". La lectura y almacenamiento del archivo de datos en estructuras como matrices y dataframes permitieron explorar de manera eficiente las características y clases de las especies de iris. A través del uso de herramientas como pandas, se facilitaron las operaciones de agrupamiento y visualización, lo cual es fundamental en la preparación de datos para análisis posteriores o implementación de modelos de aprendizaje automático.

## Referencias

Del Río, F. M. (2024, 29 enero). *Tema 8 Tipos de datos: data frames | Programación con R.*

<https://www4.ujaen.es/~fmartin/R/tipos-de-datos-data-frames.html>

Navarro, S. (2024, 17 abril). ¿Qué son los datasets? [4 sitios donde encontrarlos]. *KeepCoding*

*Bootcamps*. <https://keepcoding.io/blog/que-son-datasets/>

Solis, D. C. (2023, 1 mayo). Datasets: Qué son y cómo acceder a ellos. *OpenWebinars.net*.

<https://openwebinars.net/blog/datasets-que-son-y-como-acceder-a-ellos/>