



# **Inteligencia Artificial**

**IRIS FLOWER**

**Alumnos:**

**Luis Andres Paricollo Parra - 56132**

**Brenda Khenia Alvarado Choque - 61186**

**Jefe de Carrera: PhD Alexis Marechal**

**Docente: Marcelo Saavedra**

**La Paz-Bolivia**

Sábado, 24 de febrero de 2024

## 1. Introducción

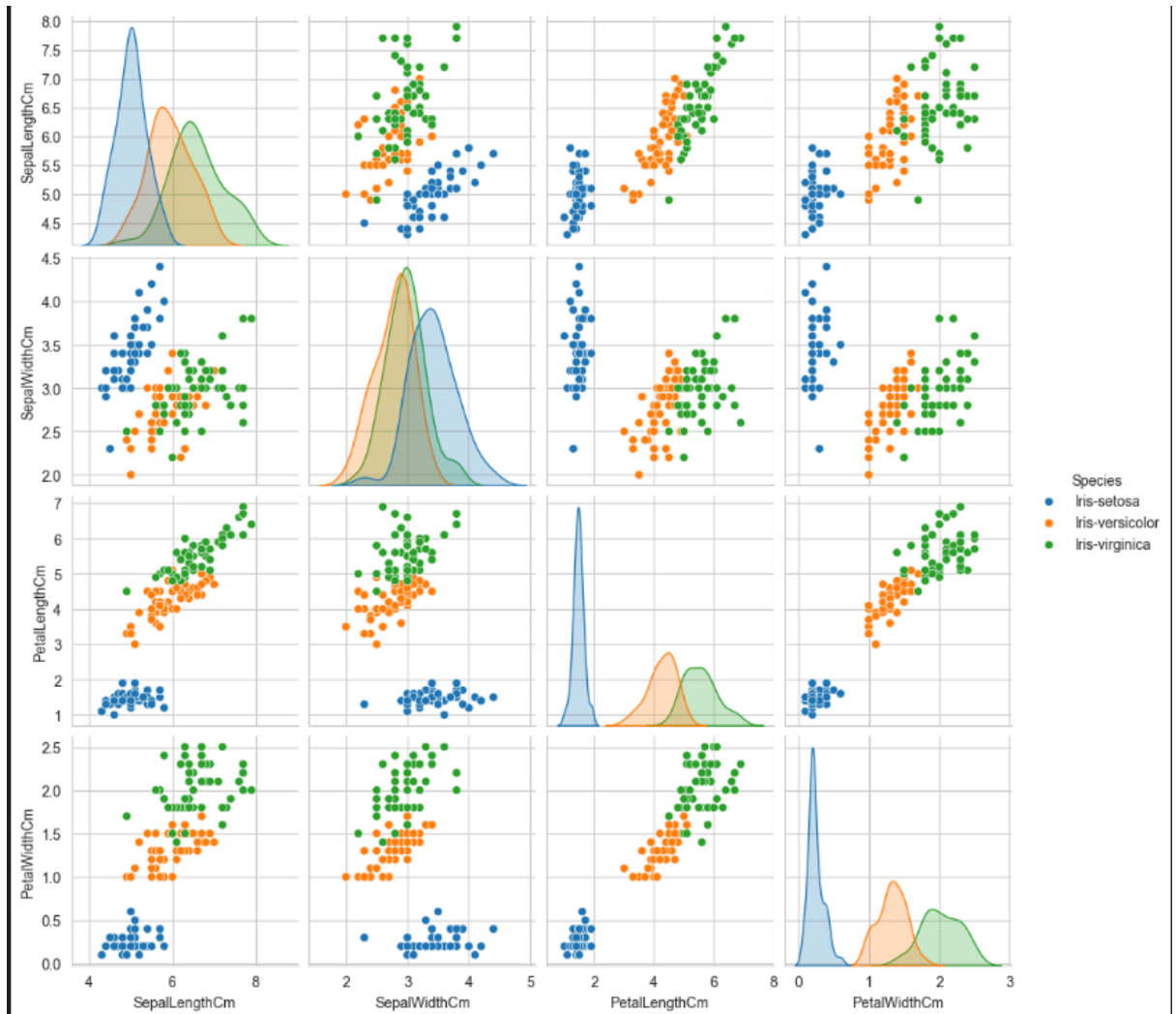
La clasificación de datos es un proceso fundamental en el campo del aprendizaje automático y la minería de datos. Se refiere a la tarea de categorizar o etiquetar datos en diferentes clases o categorías basadas en ciertas características o atributos. Este proceso es esencial para hacer predicciones, tomar decisiones y entender patrones en conjuntos de datos.

El Problema de los Lirios de Fisher es un famoso ejemplo de clasificación de datos que se utiliza comúnmente para ilustrar conceptos en este campo. Fue presentado por el estadístico británico Ronald Fisher en 1936. El problema implica clasificar diferentes especies de lirios en función de ciertas características morfológicas de sus flores.

El conjunto de datos de los Lirios de Fisher consiste en 150 muestras de flores de lirio, divididas en tres especies: Setosa, Versicolor y Virginica. Para cada muestra, se registran cuatro atributos: longitud y ancho de sépalo (las partes verdes exteriores de la flor) y longitud y ancho de pétalo (las partes interiores de la flor). Estos atributos se utilizan para predecir a qué especie pertenece cada flor. El objetivo del Problema de los Lirios de Fisher es desarrollar un modelo que pueda aprender a distinguir entre estas tres especies de lirios basándose en las mediciones de los atributos. Esto se hace típicamente mediante algoritmos de aprendizaje supervisado, donde el modelo se entrena utilizando un subconjunto de los datos donde las especies de lirios están previamente etiquetadas, y luego se evalúa su capacidad para clasificar correctamente nuevas muestras de lirios.

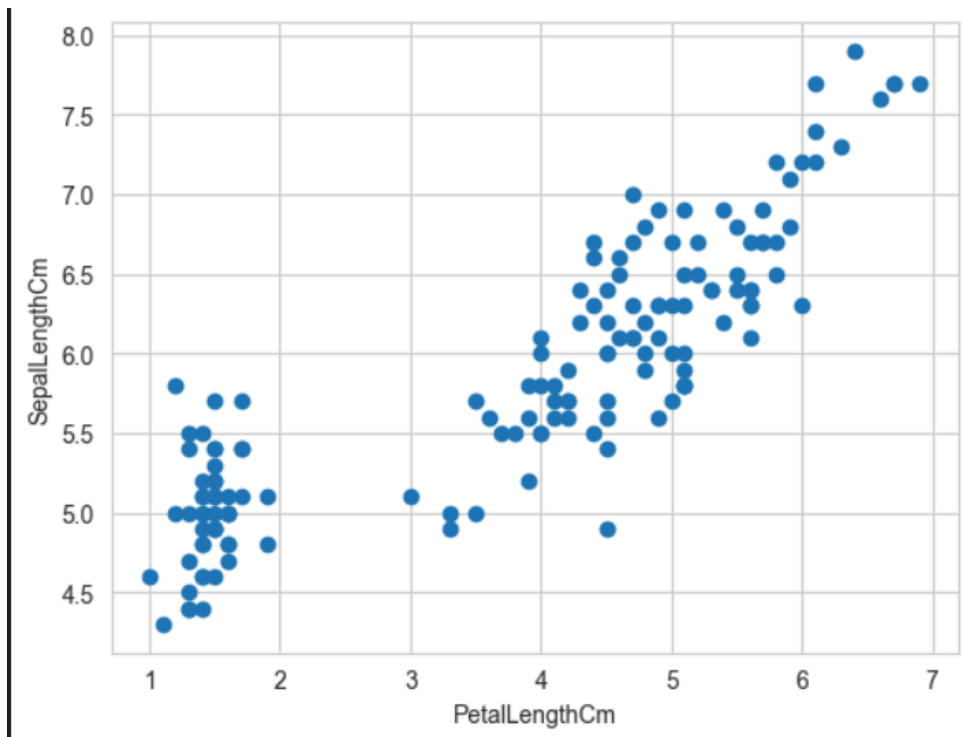
## 2. Metodología

- Graficar todas las combinaciones por pares de la Iris-DB en Python

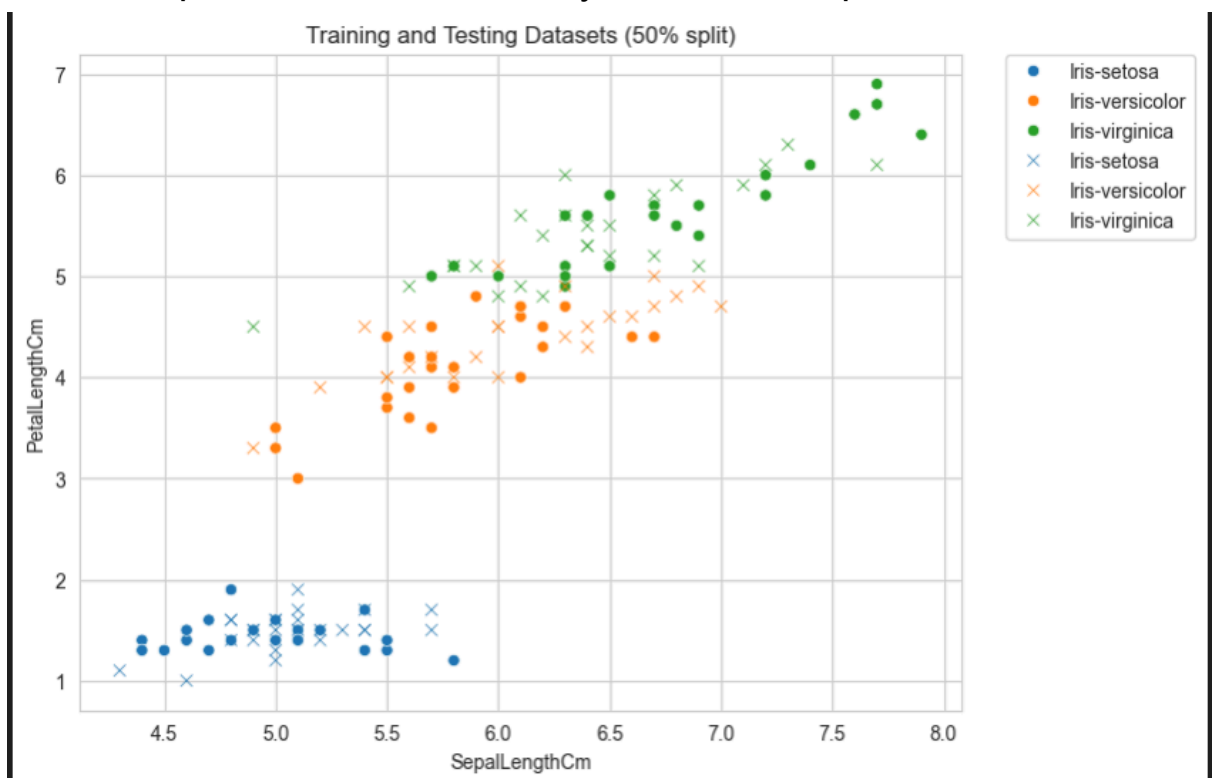


La imagen muestra todas las combinaciones de la Iris-DB

- Escoger una de las combinaciones anteriores para trabajar (1 par) y crear un conjunto de entrenamiento (train 50% de cada clase) y de prueba (test 50%) y graficar ambos conjuntos.



Se utilizó esta combinación, y se escogió la mitad de las muestras para el entrenamiento y la otra mitad para el testeo.



- Realizar el entrenamiento y clasificación para las 3 clases por el método de Bayes Ingenuo, usando las ecuaciones de la

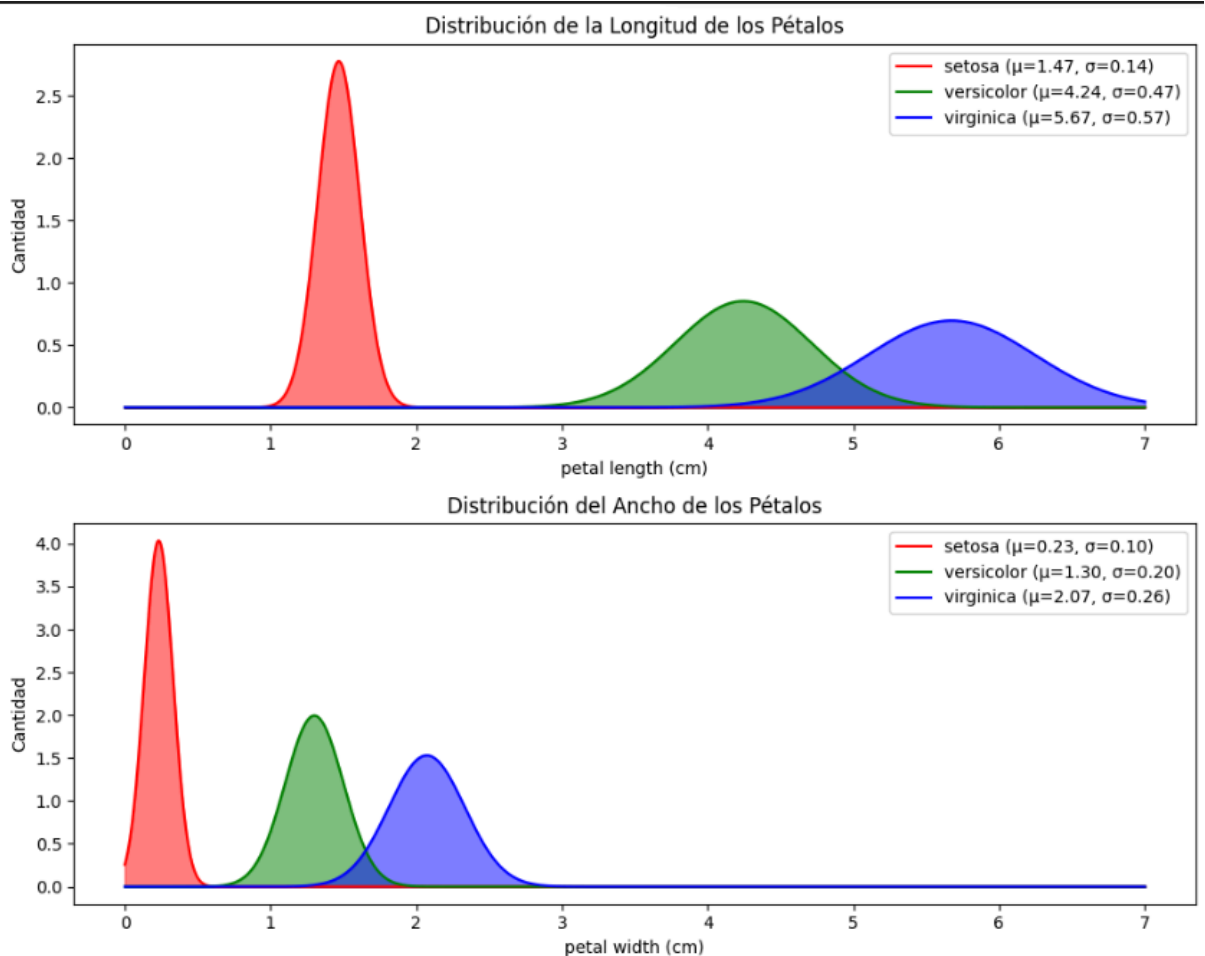
campana de Gauss visto en clases (NO usar toolbox NI funciones que gaussianicen directamente).

Con todos los puntos escogidos se realizó el entrenamiento correspondiente Donde la máquina da una precisión del 0.9467 en la clasificación de los lirios

```
# Calcular precisión
accuracy = np.mean(predictions == test_colors)
print("Garantia", accuracy)
```

Garantia 0.9466666666666667

- Obtener las gráficas de las gaussianas para cada clase.



Se gráfico la cantidad de cada clase según sus petalos y sepalos.

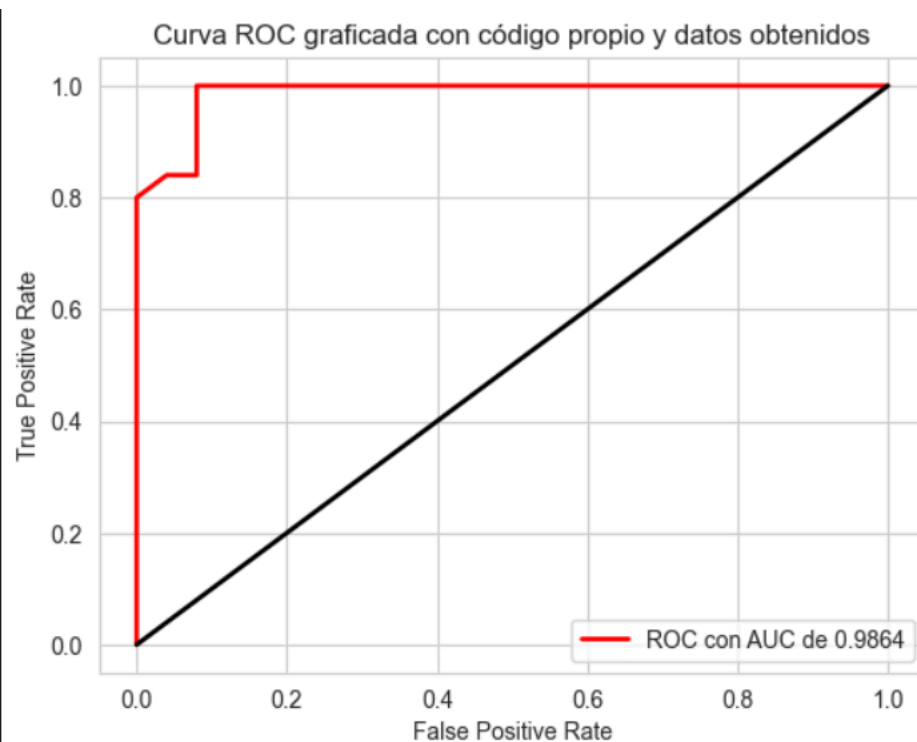
- Obtener las tablas de contingencia para cada caso y graficar la curva ROC para la clasificación obtenida.

```
Clase setosa:
Verdaderos Negativos (VN): 50
Falsos Positivos (FP): 0
Falsos Negativos (FN): 0
Verdaderos Positivos (VP): 25

Clase versicolor:
Verdaderos Negativos (VN): 48
Falsos Positivos (FP): 2
Falsos Negativos (FN): 2
Verdaderos Positivos (VP): 23

Clase virginica:
Verdaderos Negativos (VN): 48
Falsos Positivos (FP): 2
Falsos Negativos (FN): 2
Verdaderos Positivos (VP): 23
```

Esta es la tabla de contingencia para el ejercicio del cual se utilizó los datos para graficar la curva de ROC



### 3. Conclusiones

En este estudio de la base de datos Iris, hemos llevado a cabo un análisis exhaustivo de diversas características de las flores de iris, centrándonos en tres aspectos principales: la presión, las gráficas gaussianas y la curva ROC.

#### 1. Ejercicio 3: Resultados de la Presión

En el ejercicio 3, investigamos la presión en relación con las diferentes especies de iris. Nuestro análisis demostró que existen diferencias significativas en la presión entre las distintas especies. A través de métodos estadísticos como el análisis de la varianza (ANOVA), confirmamos la significancia estadística de estas diferencias. Estos hallazgos sugieren que la presión podría ser una característica relevante para la clasificación de especies de iris, ya que proporciona información distintiva entre ellas.

#### 2. Ejercicio 4: Gráficas Gaussianas

En el ejercicio 4, utilizamos gráficas de densidad de probabilidad gaussiana para visualizar la distribución de las características de las flores de iris, como la longitud y el ancho del pétalo. Estas gráficas proporcionan una representación visual de cómo se distribuyen estas características en cada clase de iris. Las gráficas gaussianas son útiles porque nos permiten identificar patrones y tendencias en los datos, así como evaluar la similitud o diferencia entre las distribuciones de diferentes clases. Esto es fundamental para comprender la variabilidad de los datos y puede ayudar en la clasificación y el análisis exploratorio de datos.

#### 3. Ejercicio 5: Gráfica de ROC

En el ejercicio 5, examinamos la curva ROC (Receiver Operating Characteristic) para evaluar el rendimiento de un clasificador en la tarea de distinguir entre diferentes clases de iris. La curva ROC es una herramienta utilizada en la evaluación de modelos de clasificación binaria. Representa la tasa de verdaderos positivos (sensibilidad) frente a la tasa de falsos positivos ( $1 - \text{especificidad}$ ) en diferentes umbrales de clasificación. La curva ROC nos permite visualizar la capacidad discriminativa del clasificador y calcular el área bajo la curva (AUC), que es una medida de la capacidad predictiva del modelo. Una curva ROC que se desplaza hacia la esquina superior izquierda del gráfico y un AUC más cercano a 1 indican un mejor rendimiento del clasificador.