



PROCESO DE GESTIÓN DE FORMACIÓN PROFESIONAL INTEGRAL
TALLER DE CIENCIA DE DATOS
(Machine Learning – Regresión Logística)

IDENTIFICACIÓN DEL TALLER

- **Denominación del Programa de Formación:** Tecnólogo en Análisis y Desarrollo de Software
- **Código del Programa de Formación:** 228118
- **Nombre del Proyecto:** 2417781 - Desarrollo de software como recurso para la gestión de información de las organizaciones en el contexto de la innovación, transformación digital y en el marco de políticas y requisitos de la empresa.
- **Fase del Proyecto:** Evaluación
- **Actividad de Proyecto:** Desarrollar las tareas de configuración y puesta en marcha del software
- **Competencia:** Construcción del software
- **Resultados de Aprendizaje Alcanzar:** Codificar la solución que cumpla con el diseño establecido.
- **Duración del Taller:** 04 horas

2. DESCRIPCION DEL TALLER:

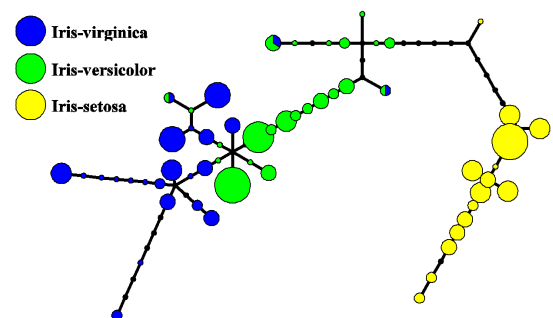
2.1 Contexto:

El conjunto de datos flor Iris o conjunto de datos iris de Fisher es un conjunto de datos multivariante introducido por Ronald Fisher en 1936, como un ejemplo de análisis discriminante lineal, según los siguientes factores:

- Cuantificación la variación morfológica de la flor Iris de tres especies relacionadas.
- Las especies se coleccionaron en la Península de la Gaspesia, bajo las mismas condiciones.

El conjunto de datos contiene 150 muestras de cada una de tres especies de Iris (Iris setosa, Iris virginica e Iris versicolor). Se midió cuatro rasgos de cada muestra: el largo y ancho del sépalo y pétalo, en centímetros. Basado en la combinación de estos cuatro rasgos, Fisher desarrolló un modelo discriminante lineal para distinguir entre una especie y otra.

Basado en el modelo análisis discriminante de Fisher, este conjunto de datos se convirtió en un caso de prueba típico por muchas técnicas clasificaciones estadísticas en aprendizaje automático como en máquinas de soporte vectorial.





2.2 Información de las etiquetas:

- **Id:** identificar único de cada flor de iris.
- **SepalLengthCm:** largo del sépalo de la flor de iris en centímetros.
- **SepalWidthCm:** ancho del sépalo de la flor de iris en centímetros.
- **PetalLengthCm:** largo del petalo de la flor de iris en centímetros.
- **PetalWidthCm:** ancho del petalo de la flor de iris en centímetros.
- **Species:** tipo de especie de la flor de irirs (Iris setosa, Iris virginica e Iris versicolor).

2.3 Procedimiento sugerido:

1. Preparación de los datos.
2. División de los datos
3. Construcción y aprendizaje del modelo de regresión logística
4. Evaluación de modelo de machine learning
5. Predicción futura

2.4 Requerimientos

- La fundación agropecuaria de Antioquia; cuenta con 5mil hectáreas de plantación de Flores de Iris en el municipio de Barbosa con la novedad que las semillas fueron mezcladas en su plantación y después de 6 meses de su proceso de cultivo no es posible clasificar cada flor a la vista humana para ser enviadas en exportación al continente Asiático.
Por lo anterior se requiere predecir qué tipo de Flor de Iris corresponde según las siguientes características:
 - SepalLengthCm: 4.8
 - SepalWidthCm: 3.7
 - PetalLengthCm: 1.5
 - PetalWidthCm: 0.2
- La solución presentada debe ser una aplicación web según estándares de programación (validación y seguridad).

2.5 Restricciones

- Para solución del ejercicio solo se podrá utilizar el método de regresión logística

2.6 Dataset

- https://github.com/majash29/CienciaDatosSENAV02/raw/main/04-ModeloRegresion/07_TallerRegresionLogistica/Dataset/CultivoAntioquia.xlsx

3. ENTREGABLES DEL TALLER

- En el LMS se debe entregar la URL del repositorio de GitHub con la siguiente información:
 - Notebook de Python con los scripts utilizados.



- Código fuente utilizado para la implementación del IA como una aplicación Web de predicción.
- URL del despliegue realizado de la aplicación.

4. CONTROL DEL DOCUMENTO

	Nombre	Cargo	Dependencia	Fecha
Autor (es)	Alvaro Pérez Niño	Instructor	ADSO	30 agosto de 2023