



PROCESO DE GESTIÓN DE FORMACIÓN PROFESIONAL INTEGRAL
TALLER DE CIENCIA DE DATOS
(Machine Learning – Árboles Clasificación)

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:

La agricultura de precisión es una de las técnicas del agro que se encuentra en auge en el valle de aburra; debido a que ayuda a los agricultores a tomar decisiones informadas sobre la estrategia agrícola. A continuación, se socializa un conjunto de datos que permitiría a los agricultores construir un modelo predictivo para recomendar los cultivos más adecuados para crecer en una granja en particular en función de varios parámetros, tales como: nitrógeno, potasio, fósforo, temperatura, humedad entre otros. El conjunto de datos contiene 2201 muestras diferentes tipos de cultivos como arroz, maíz, garbanzo, frijol, lentejas, granada, banana, entre otros.

2.2 Información de las etiquetas:

- **N-** relación del contenido de nitrógeno en el suelo
- **P-** relación de contenido de fósforo en el suelo
- **K-** relación del contenido de potasio en el suelo
- **temperatura-** temperatura en grados Celsius
- **humedad** - humedad relativa en %
- **ph-** valor de ph del suelo
- **precipitacion** - precipitación en mm



2.3 Procedimiento sugerido:

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

2.4 Requerimientos

- El sistema debe estar en capacidad de predecir cual es tipo de cultivo óptimo bajo las siguientes condiciones:
 - $N = 13$
 - $P = 5$
 - $K = 8$
 - Temperatura = 2.385.340.379
 - Humedad = 9.010.522.549
 - $Ph = 7.474.710.503$
 - Precipitación = 103.923.226
- 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 Árboles.

2.6 Dataset

- https://github.com/majash29/CienciaDatosSENAV02/raw/main/05-ModeloArboles/08_TallerArbolesClasificacion/Dataset/agricultura.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