



Regional Antioquia
Centro de Servicios y Gestión Empresarial

SENASoft

Machine Learning - Ingeniería de Datos

INSTRUCCIONES GENERALES

Querido aprendiz: a continuación, se describe el caso estudio médico, cual debe desarrollar aplicando las fases de la metodología CRIPS-MD para los procesos de ciencia de datos. Esta prueba busca identificar los aprendices que representaran al centro de formación en las categorías de Machine Learning e Ingeniería de Datos en la competencia de SENASoft:

2. DESARROLLO DE LA ACTIVIDAD

ENTENDIMIENTO DEL NEGOCIO:

Imaginemos que eres un estudiante de ciencias de la salud y estás interesado en desarrollar una herramienta de diagnóstico preliminar para enfermedades comunes. Deseas crear un programa que tome en cuenta los síntomas presentados por un paciente y los utilice para determinar la posible enfermedad que podría estar afectándolo. Esto podría resultar especialmente útil en entornos donde no se cuenta con acceso inmediato a profesionales médicos o en situaciones donde se requiere una evaluación inicial rápida.

El desafío radica en desarrollar un script utilizando un árbol de clasificación que sea capaz de tomar una serie de síntomas como entrada y proporcionar una estimación precisa de la enfermedad probable.

Etiquetas de información

- *Fiebre*
- *Tos*
- *Dolor de Garganta*
- *Congestión Nasal*
- *Dificultad Respiratoria*
- *Enfermedad (Gripe, Resfriado Común o Sinusitis)*

Nota: *cero (0) no presenta síntomas – uno (1) presenta síntomas.*

Entregables:

1. Script de Python con el proceso completo desarrollado para la generación del modelo de Machine learning



Regional Antioquia
Centro de Servicios y Gestión Empresarial

3. CRITERIOS DE EVALUACION:

- Se demuestra una comprensión clara del problema y los objetivos del proyecto.
- Se muestra evidencia de una exploración adecuada del dominio y los datos disponibles.
- Se lleva a cabo una limpieza de datos efectiva para abordar problemas como valores faltantes, duplicados, errores, atípicos entre otros.
- Se realiza una transformación adecuada de los datos para prepararlos para el análisis.
- Se selecciona y aplica correctamente el modelo o algoritmo apropiado para abordar las preguntas del proyecto.
- Se realiza una división adecuada de los datos en conjuntos de entrenamiento, validación y prueba.
- Se ajustan los hiperparámetros del modelo de manera efectiva y se realiza una validación cruzada si es necesario.
- Se documentan y justifican las decisiones de modelado tomadas.
- Se evalúa el rendimiento del modelo utilizando métricas adecuadas según el tipo de problema
- Se analizan los resultados obtenidos y se interpretan de manera clara y precisa.

4. CONTROL DEL DOCUMENTO

	Nombre	Cargo	Dependencia	Fecha
Autor (es)	Alvaro Pérez Niño	Instructor	ADSO	05 de julio de 2023
	Daniel David Benavides	Instructor	ADSO	05 de julio de 2023