

Plan Formativo: Ciencia de Datos	Nivel de Dificultad
Módulo: Aprendizaje de Máquina	Medio
Tema: Regresión Logística	
Intención del aprendizaje o aprendizaje esperado:	
<ul style="list-style-type: none">• Elaborar un modelo predictivo aplicando el algoritmo de Regresión Logística para resolver un problema de clasificación utilizando el lenguaje Python.	
Ejercicios planteados	
<p>El archivo rrhh.csv contiene información de la situación laboral de los trabajadores de la empresa ABAC. Las variables disponibles son:</p> <ul style="list-style-type: none">• Estado: Estado actual del trabajador en la empresa (1: desvinculado, 0: vinculado).• Edad: Edad del trabajador en años.• Ratio.Pago: Medida de pago por hora (numerico)• Salario: Salario mensual en dólares que tiene o tenía el trabajador• Dias.trabajados: Días que lleva o llevaba trabajando en la empresa• Ausencias: Días que ha faltado a trabajar• Sexo: Sexo del trabajador (Female , Male)	

- **Estado.Civil:** Estado civil del trabajador (1: divorciado, 2: casado,3: separado, 4: soltero, 5: viuda)
 - **Departamento:** Lugar de trabajo en la empresa (Admin Offices,..)
 - **Posicion:** Cargo del trabajador/empleado (Accountant I ,....)
 - **Desempeño:** Clasificación del desempeño del trabajador.
1. Cargue los datos y revise los formatos de cada variable, recuerde codificar las variables como numéricas o factores según corresponda.
 2. Realice un análisis descriptivo de sus datos. Determinar si existen observaciones faltantes, en el caso de existir tome la decisión de omitirlas del estudio u omitir la variable. Evalúe si existen posibles incongruencias en la fuente de datos (ej: edades negativas). Y finalmente analice la presencia de valores atípicos en las variables. Comente.
 3. Realice análisis de cómo se relacionan las variables **continuas** con la variable de interés. Acompañe con gráficos y estadísticas. ¿Qué variables pudieran resultar significativas a la hora de modelar la probabilidad de que el trabajador sea desvinculado a la empresa?
 4. Realice análisis de cómo se relacionan las variables **categorías** con la variable de interés. Acompañe con gráficos y estadísticas. ¿Qué variables pudieran resultar significativas a la hora de modelar la probabilidad de que el trabajador sea desvinculado a la empresa?

Caso

APRENDIZAJE DE MÁQUINA SUPERVISADO

Preguntas guía

-

Recursos Bibliográficos:

[1] Cassis, A. (2015). Inteligencia artificial 101.

<https://inteligenciaartificial101.wordpress.com/2015/10/20/aprendizajesupervisado/>

[2] Métodos de Clasificación

<https://bookdown.org/content/2274/metodos-de-clasificacion.html>

[3] Regresión Logística

<https://conceptosclaros.com/que-es-regresion-logistica/>

[4] Función Sigmoidal

<https://ichi.pro/es/que-es-la-funcion-sigmoidal-como-se-implementa-en-regresion-logistica-77981969140323>