

Resumen del Proyecto Final

Tiempo estimado: 5 minutos

En este proyecto, utilizarás un conjunto de datos del mundo real y practicarás inteligencia artificial generativa para crear códigos en Python que puedan realizar preparación de datos, análisis y desarrollo de modelos.

Escenario del Proyecto

Ha sido contratado como Científico de Datos por una firma de consultoría. La firma tiene un cliente que es un concesionario de automóviles de reventa. Tienen una característica especial en los coches Ford y quieren que tu firma diseñe un modelo que pueda predecir el precio de cotización óptimo para los coches en su lote. Te proporcionamos datos de ventas de los últimos años. El conjunto de datos contiene diferentes características de los coches y el precio al que fueron vendidos.

Las tareas asignadas a ti son las siguientes:

- 1. Puede haber algunas entradas duplicadas y algunos valores faltantes en el conjunto de datos. La limpieza de datos será parte de la tarea.
- 2. Debes realizar un análisis exploratorio de datos para obtener información detallada sobre los datos y determinar el efecto de diferentes características en el precio. Algunas solicitudes específicas del cliente incluyen:
 - a. Identificar el número de ventas para cada tipo de combustible
 - b. Identificar qué tipo de transmisión tiene más valores atípicos de precio
- 3. Comparar los modelos con regresiones lineales, polinómicas y de ridge en variables simples y múltiples para encontrar el modelo de mejor rendimiento.
- 4. Realice una Búsqueda en Rejilla en el modelo de regresión de Ridge para identificar el hiperparámetro óptimo del modelo para un mejor rendimiento.

Acerca del conjunto de datos

Este conjunto de datos contiene precios de venta de reventa para automóviles Ford. Este es un conjunto de datos públicos disponible en el sitio web de [Kaggle](#) como [Conjunto de Datos de Precios de Automóviles Ford](#) bajo la licencia [CC0: Dominio Público](#) . El conjunto de datos ha sido ligeramente modificado para el propósito de este proyecto.

A continuación se explican los atributos de este conjunto de datos.

Variable	Descripción
modelo	Nombre del modelo del automóvil
año	Año de fabricación del automóvil.
transmisión	Tipo de transmisión (Automática, Manual o Semiautomática)
kilometraje	Número de millas recorridas
tipoDeCombustible	El tipo de combustible que usa el automóvil (Gasolina, Diésel, Híbrido, Eléctrico, Otro)
impuesto	Impuesto anual a pagar en USD
millas por galón	Millas por Galón que recorre el automóvil
Tamaño del motor	Tamaño del motor del automóvil
precio	Precio del automóvil en USD

Tareas en el proyecto

Las tareas del proyecto son la preparación de datos, el análisis y el desarrollo de modelos. Basado en el conjunto de datos, debes escribir instrucciones para generar los códigos en Python para realizar tareas específicas. Puedes acceder a un entorno de pruebas basado en JupyterLite para probar los códigos generados utilizando las instrucciones del aula de IA Generativa.

Autor(es)

[Abhishek Gagneja](#)

