**CASO**

**PREDICCION DEL PRECIO AUTOS DE SEGUNDA MANO**

Se desea desarrollar un sistema que integre un modelo predictivo de aprendizaje automático para evaluar el precio de un auto de segunda mano que considere el número de años, el tipo de combustible, cuántos kilómetros ha recorrido, etc.

El cliente requiere un modelo predictivo de aprendizaje automático que pronostique el precio de un auto basándose en sus especificaciones.

El grupo de análisis de requerimientos piensa que se debe utilizar   
Puede utilizar este flujo como plantilla para resolver cualquier problema de Regresión ML supervisada.

Luego de entrevistas de elicitación con el cliente, se redactó en un documento de trabajo el flujo de requerimientos funcionales y no funcionales como sigue:

* La lectura de los datos se debe realizar en Python
* Se debe realizar claramente la identificación de la variable objetivo
* Se debe mostrar la observación de la distribución de la variable objetivo
* Se tiene que realizar una exploración básica de datos
* Las columnas inútiles se deben rechazar
* El Análisis exploratorio visual de la distribución de datos tiene que mostrar un histograma y gráficos de barras.

El cliente proporcionara un dataset con datos de características de los autos para ser tomada como base para el sistema a desarrollar.

El dataset "PrecioAutoSegunda.csv" contiene 1435 datos de precios de autos y está compuesto de los siguientes campos y su descripción.

**Precio**: El precio del auto en dólares

**Tiempo**: La edad del auto en meses

**Recorrido**: recorrido de Cuántos KMS se utilizó el auto

**Combustible**: Tipo de combustible Gasolina/Diesel/GNC del auto

**Potencia**: Potencia del auto en HP

**MetColor**: Si el auto tiene color metálico o no

**Automático**: Si el auto tiene transmisión automática o no

**CC**: El tamaño del motor del auto en cilindros cúbicos

**Puertas**: El número de puertas del auto

**Peso**: El peso del auto

Los analistas de diseño y modelamiento advierten que los desarrolladores siempre les advierten a los analistas que el éxito del proyecto está en el tratamiento de los datos,   
y que es uno de los pasos más importantes en el aprendizaje automático, por lo que se debe entender bien los datos y el dominio antes de intentar aplicar cualquier algoritmo de aprendizaje automático, por eso recomiendan que:

* Se tiene que realizar la selección de características basada en la distribución de datos
* Debe haber un Tratamiento de valores atípicos
* Debe haber un Tratamiento de valores perdidos
* Debe haber un Análisis de correlación visual
* Debe haber un Análisis estadístico de correlación (selección de características)
* Se debe realizar una conversión de datos a numéricos para ML
* Se debe realizar un muestreo y validación cruzada K-fold
* Se debe probar múltiples algoritmos de regresión
* Observar la información resumida de los datos
* Tipos de datos, valores perdidos basados en el número de valores no nulos Vs filas totales, etc.
* Eliminar las variables de los datos que tienen demasiados valores perdidos (Valores perdidos > 30%).
* Eliminar las variables cualitativas que no pueden ser utilizadas en ML.
* Aplicar la estadística descriptiva para visualizar los parámetros
* Encontrar valores únicos para cada columna para entender qué columna es categórica y cuál es continua, típicamente si el número de valores únicos es < 20 entonces la variable es probable que sea una categoría de lo contrario continua.

Actividades de modelamiento

1. Identifique y organice los requerimientos funcionales y no funcionales según lo colectado en el documento de trabajo que se obtuvo de las entrevistas de elicitación con el cliente, además agregue requerimientos que se obtuvieron de las técnicas de observación, etnografía y encuesta. (4 pts.)

**Solo como referencia.**

<https://www.mercawise.com/estudios-de-mercado-en-mexico/estudio-de-mercado-sobre-autos>

<https://noticias.coches.com/consejos/diez-preguntas-imprescindibles-si-compras-un-coche-usado/45010>

1. Describir o plantear el problema graficar un árbol de problemas con los problemas secundarios. Redactar el Objetivo con los objetivos específicos
2. Realizar el catalogo de requisitos considerando los motivos y las restricciones asociadas.
3. Modele los casos de uso para el sistema predicción del precio de autos de segunda mano. (4 pts.)
4. Genere historias de usuario de los casos de uso para describir los procedimientos. (4 pts.)

Desarrollo

1. Del caso anterior, utilizando la información proporcionada, desarrolle un modelo en Python en Colab como prototipo del sistema a desarrollar antes de ser desplegado a producción del sistema (8 pts.)