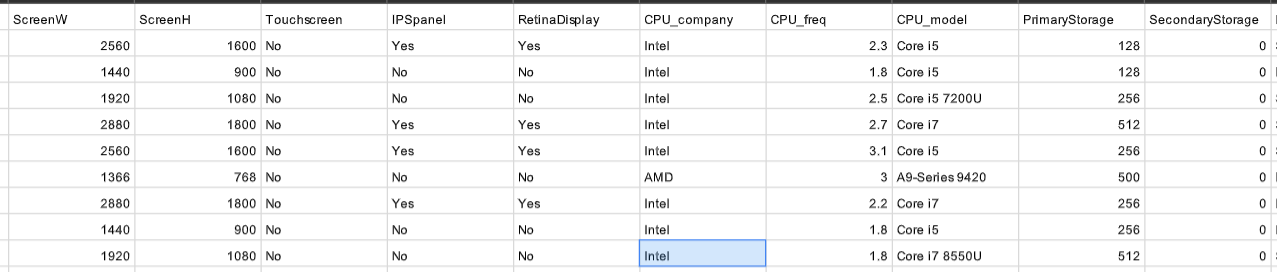
**Jorge Luis Chuqimia Parra**

**Del dataset trabajado en los exámenes, describa el mismo a detalle**

**El dataset elegido, denominado Laptop Prices, contiene información relacionada con las especificaciones técnicas y el precio en euros de laptops de diferentes marcas. Es un dataset limpio y preprocesado, optimizado para su uso en algoritmos de aprendizaje automático. A continuación, se describen los atributos principales:**

1. **Company: Nombre del fabricante de la laptop.**
2. **Product: Marca y modelo del dispositivo.**
3. **TypeName: Tipo de laptop (Notebook, Ultrabook, Gaming, etc.).**
4. **Inches: Tamaño de la pantalla en pulgadas.**
5. **Ram: Cantidad de memoria RAM (en GB).**
6. **OS: Sistema operativo preinstalado.**
7. **Weight: Peso de la laptop (en kilogramos).**
8. **Price\_euros: Precio de la laptop en euros (atributo objetivo o "target").**
9. **Screen: Resolución de la pantalla (Standard, Full HD, 4K, etc.).**
10. **ScreenW: Ancho de la pantalla (en píxeles).**
11. **ScreenH: Altura de la pantalla (en píxeles).**
12. **Touchscreen: Si la laptop tiene o no pantalla táctil.**
13. **IPSpanel: Si la pantalla incluye un panel IPS.**
14. **RetinaDisplay: Si la pantalla incluye una pantalla Retina.**
15. **CPU\_company: Compañía fabricante del procesador.**
16. **CPU\_freq: Frecuencia del procesador (en Hz).**
17. **CPU\_model: Modelo del procesador.**
18. **PrimaryStorage: Capacidad del almacenamiento principal (en GB).**
19. **PrimaryStorageType: Tipo de almacenamiento principal (HDD, SSD, etc.).**
20. **SecondaryStorage: Capacidad de almacenamiento secundario (si aplica).**
21. **SecondaryStorageType: Tipo de almacenamiento secundario (si aplica).**
22. **GPU\_company: Compañía fabricante del procesador gráfico.**
23. **GPU\_model: Modelo del procesador gráfico.**

**El dataset incluye laptops de diversas marcas como Apple, HP, Acer, Dell, Asus, entre otras, lo que lo convierte en un conjunto de datos representativo para realizar análisis sobre las características que influyen en el precio de las laptops.**

****

**Describa claramente del objetivo de investigación a partir del dataset elegido**

**El objetivo principal de la investigación es identificar y analizar las características técnicas de las laptops que tienen mayor impacto en su precio.**

**Selección del clasificador (acorde a los datos supervisado, no supervisado). El clasificador puede, pero no necesariamente depender del preprocesamiento.**

**Justificar el clasificador (máximo 2 planas con fuente ISBN, DOI)**

El problema que se plantea en este estudio es la predicción del precio de las laptops en euros, basándose en un conjunto de atributos técnicos de las laptops. Dado que la variable objetivo es continuamente escalada (**precio en euros**), estamos ante un problema de regresión supervisada. En este tipo de problemas, la tarea consiste en predecir un valor numérico basado en un conjunto de características de entrada. Por lo tanto, la elección de un modelo de regresión supervisada es fundamental para abordar este tipo de predicción.

**1. Modelos de Regresión Supervisada**

Los modelos de regresión supervisada se basan en el uso de datos etiquetados, donde tanto las características de entrada como el valor de la variable objetivo están disponibles durante el entrenamiento del modelo. En este caso, las características de entrada son los atributos de las laptops (como RAM, tamaño de pantalla, tipo de almacenamiento, etc.), y el valor objetivo es el precio de la laptop. Los modelos de regresión permiten establecer una relación matemática entre estas características y el valor de salida, generando una función que puede predecir el precio de las laptops cuando se le proporcionan nuevos datos de entrada.

**2. Elección del Algoritmo de Regresión**

Existen diversos algoritmos de regresión que podrían aplicarse a este problema, y la elección del algoritmo depende de las características del conjunto de datos y de los objetivos del modelo. A continuación, se presentan algunos de los algoritmos más relevantes y su justificación:

**2.1. Regresión Lineal**

La regresión lineal es uno de los algoritmos más simples y ampliamente utilizados para problemas de regresión. Este modelo se basa en la suposición de que existe una relación lineal entre las características de entrada y el precio de la laptop. Aunque este modelo es fácil de interpretar y rápido de entrenar, su principal limitación radica en su incapacidad para capturar relaciones no lineales entre las características.

* Es un modelo simple, eficiente y fácil de implementar.
* No puede modelar relaciones complejas o no lineales entre las variables, lo que podría ser una limitación en este caso.

**2.2. Árboles de Decisión para Regresión**

Los árboles de decisión son modelos que dividen los datos en diferentes "ramas" basadas en condiciones sobre las características de entrada. Cada "hoja" del árbol corresponde a un valor de predicción. Este modelo es útil para capturar relaciones no lineales entre las características y el precio.

* No requiere suposiciones sobre la relación entre las variables, y puede manejar tanto características numéricas como categóricas.
* Los árboles de decisión pueden ser propensos al sobreajuste, especialmente cuando el árbol es muy profundo.

**2.3. Random Forest para Regresión**

El Random Forest es un conjunto de árboles de decisión, donde varios árboles se entrenan en subconjuntos aleatorios de los datos. Luego, las predicciones de los árboles se combinan para obtener una predicción final. Este enfoque mejora la precisión y reduce el riesgo de sobreajuste.

* Es muy robusto frente al sobreajuste y puede manejar datos no lineales de manera eficaz.
* Requiere más tiempo de entrenamiento y no es tan interpretable como un solo árbol de decisión.

**3. Comparación y Selección Final del Modelo**

La elección final del algoritmo debe basarse en un equilibrio entre precisión, complejidad y capacidad de interpretación. La regresión lineal es adecuada para establecer una línea base

**Fuente**

* *"An Introduction to Statistical Learning"* de Gareth James, Daniela Witten, Trevor Hastie y Robert Tibshirani, ISBN 978-1-4614-7138-7.
* Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" de Aurélien Géron
* Revistas como Journal of Machine Learning Research, Machine Learning, y IEEE Transactions on Pattern Analysis and Machine Intelligence