

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.

ScreenW	ScreenH	Touchscreen	IPSPanel	RetinaDisplay	CPU_company	CPU_freq	CPU_model	PrimaryStorage	SecondaryStorage
2560	1600	No	Yes	Yes	Intel	2.3	Core i5	128	0
1440	900	No	No	No	Intel	1.8	Core i5	128	0
1920	1080	No	No	No	Intel	2.5	Core i5 7200U	256	0
2880	1800	No	Yes	Yes	Intel	2.7	Core i7	512	0
2560	1600	No	Yes	Yes	Intel	3.1	Core i5	256	0
1366	768	No	No	No	AMD	3	A9-Series 9420	500	0
2880	1800	No	Yes	Yes	Intel	2.2	Core i7	256	0
1440	900	No	No	No	Intel	1.8	Core i5	256	0
1920	1080	No	No	No	Intel	1.8	Core i7 8550U	512	0

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. Random Forest Regressor

Es un modelo basado en árboles que funciona muy bien con datos complejos y no lineales, como los de este conjunto de datos, que tienen muchas características (tamaño de pantalla, marca, tipo de almacenamiento, etc.). El Random Forest es muy robusto contra el sobreajuste, especialmente cuando se tiene un conjunto de datos grande con muchas variables.

- Maneja bien relaciones no lineales entre las variables.
- Es fácil de usar y no requiere mucho preprocesamiento.
- Es resistente al sobreajuste.

2. Support Vector Machines (SVM) para Regresión (SVR)

Las SVM pueden ser muy eficaces para problemas de regresión, especialmente cuando los datos tienen una estructura no lineal. SVR puede encontrar una función que mapea las características de entrada a un valor continuo, en este caso, el precio de la laptop.

- Muy bueno para datos complejos y no lineales.
- Puede ser más preciso que otros modelos en ciertas condiciones, pero requiere un buen ajuste de los parámetros.

3. Redes Neuronales (ANN para Regresión)

Las redes neuronales son potentes para modelar relaciones no lineales complejas, lo que las hace adecuadas para predecir precios basados en un conjunto de datos con muchas características. A medida que la cantidad de datos aumenta, las redes neuronales pueden mejorar su rendimiento.

- Capaz de aprender representaciones complejas de los datos.
- A medida que la cantidad de datos aumenta, las redes neuronales tienden a mejorar.

3. Selección Final del Modelo

Random Forest Regressor: Dada que trabaja muy bien con operaciones complejas

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