**Aplicación que predice el precio de los coches V0**

**Contexto:**

Predecir el precio de un coche a veces es un dolor de cabeza. Hay compañías que te dan un presupuesto, pero algunas personas no confían si su coche esta siendo valorado correctamente o no, e incluso se sienten estafados. Los datos obtenidos son de diferentes coches con variedad de marcas y precios.

**Problema:**

* Muchas agencias de coches = diferentes presupuestos de coches
* Las agencias de coches son 100% confiables?
* Tienes que pagar un extra si la agencia contratada tase el coche.
* Necesitamos una APP que te de un presupuesto mas acorde al coche que quieres valorar, abierto a todo el mundo.
* Diferentes modelos y características: Hay una gran variedad de modelos de automóviles y características diferentes que pueden afectar el precio de un coche.
* Condiciones del coche: La edad, el kilometraje y la condición del coche pueden afectar significativamente su precio.

**Análisis:**

1. Definición de los datos

|  |  |
| --- | --- |
| **Fuente** | <https://archive.ics.uci.edu/ml/datasets.php> |
| **Formato** | Datos de 205 coches y hay 26 atributos |
| **Atributos** | 1. symboling  2. normalized-losses  3. make  4. fuel-type  5. aspiration  6. num-of-doors  7. body-  8. drive-wheels  9. engine-location  10. wheel-base  11. length  12. width  13. height  14. curb-weight  15. engine-type  16. num-of-cylinders  17. engine-size  18. fuel-system  19. bore  20. stroke  21. compression-ratio  22. horsepower  23. peak-rpm  24. city-mpg  25. highway-mpg  26. price |

1. Tenemos que preparar los datos: para ello debemos quitar las columnas que no aportan nada o estén vacías, eliminar los datos duplicados, seleccionar la categoría que queremos predecir, en nuestro caso es el “precio”, seguidamente tenemos que hacer una búsqueda exhaustiva de las posibles correlaciones. Finalmente, ya tendremos preparados para poder analizar los algoritmos de Machine learning.
2. visualización de los datos

Gráfico, Histograma

Descripción generada automáticamente

Vemos que hay una gran cantidad de coches con precios en torno a 10K y 20 K, por tanto, se puede ver que hay mas muestras en este entorno debido a que hay más facilidad de pago para las familias de clase media.

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamente

En el histograma de coches anterior vemos que la marca Toyota es la que mas destaca.

Gráfico, Histograma

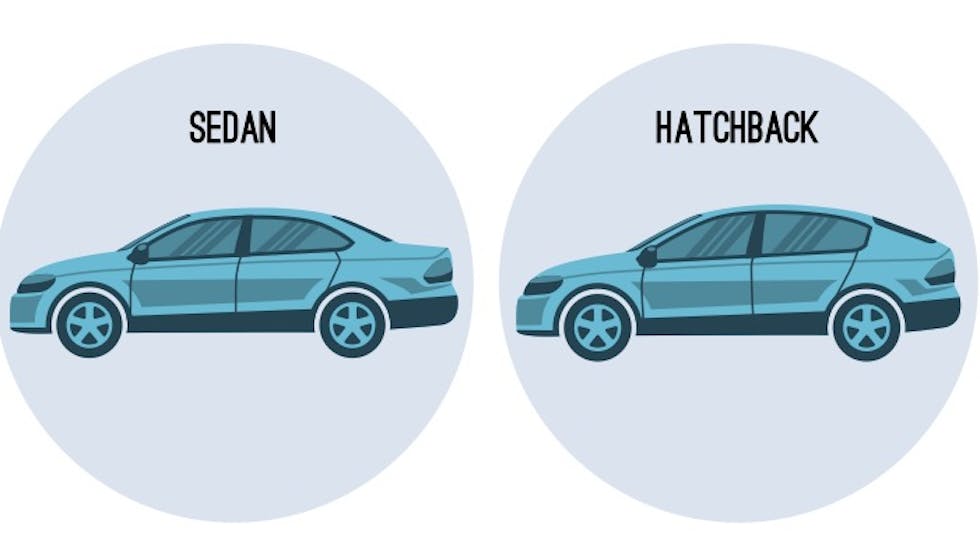
Descripción generada automáticamente

(En este histograma podemos ver que el dataset es un tanto antigua ya que hoy en día la mayoría de los coches es a gasolina, entiendo que significa gasolina) Vemos que hay mas coches a gas que diesel. Esto puede ser a que los coches a diesel son mas caros, y por supuesto el diesel no es nada barato hoy en día.

Gráfico, Histograma

Descripción generada automáticamente

El tipo de coche que predomina es el “sedan” ya que es el que tiene mas demanda, seguida del “hatchback”.



1. Gráfico, Gráfico de rectángulos

   Descripción generada automáticamenteCorrelaciones:

Si dos atributos tienen una correlación alta, significa que están fuertemente relacionados entre sí. Por lo tanto, la información que proporcionan puede ser redundante, y uno de los atributos podría eliminarse sin perder demasiada información. En algunos casos, la alta correlación entre atributos puede ser útil, como en el caso de la edad y la experiencia laboral en el análisis de la capacidad para desempeñar un trabajo.

Por otro lado, si dos atributos tienen una correlación baja, significa que están débilmente relacionados entre sí. En este caso, es probable que la información que proporcionan sea complementaria, y ambos atributos pueden ser útiles para predecir un resultado o tomar una decisión.

Es importante tener en cuenta que la correlación no siempre indica causalidad. Solo porque dos atributos están correlacionados no significa que uno cause el otro. Es posible que haya otros factores que estén causando la correlación. Por lo tanto, es importante realizar un análisis cuidadoso de los datos para determinar si la correlación entre los atributos es significativa y si es útil para los propósitos de la tarea de predicción o toma de decisiones en cuestión.

Como nos interesa el precio, tenemos que buscar que atributos tienen más correlación con esta: engine-size, horsepower, etc.

1. Selección de los mejores atributos

Gráfico, Gráfico de barras

Descripción generada automáticamenteGráfico, Gráfico de barras

Descripción generada automáticamente

Gráfico, Gráfico de barras

Descripción generada automáticamenteGráfico, Gráfico de barras

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

**Algoritmos**

**Acciones**

**Conclusiones**