

Trabajo Final

Curso: Complejidad Algorítmica

Universidad Peruana de Ciencias Aplicadas

Sección: CC41

Docente: Abraham Sopla Maslucán

Integrantes:

Ruiz Torres, Erick Hérnan (u202118946)

Alcántara Cruz Rodrigo Alonso(u202216698)

Conza Hualpa, Alexia Evelyn (u202118038)

2024 - I

Contents

1	Descripción del problema	3
2	Descripción del conjunto de datos (dataset)	3
2.1	Características y origen de datos móviles de análisis	3
2.1.1	Origen de los datos	4
3	Propuesta	6
4	Diseño del aplicativo	7
4.1	Diagrama de clase y diagrama de base de datos	7
4.2	Módulo para la adecuación de datos fuente a estructura de datos tipo grafo	7
4.3	Módulo para la representación de datos en el grafo	7
4.4	Interfaz tentativa (visual) para usuarios	7
5	Validación de resultados y pruebas	7
6	Conclusiones	7
7	Referencias bibliográficas	7

1 Descripción del problema

El problema que veremos a cabo en este proyecto se enfoca en las recomendaciones de autos eléctricos, el objetivo es brindar a todos los usuarios sugerencias personalizadas de autos en base a características como precio, año, modelo, etc. La recomendación de estos vehículos es fundamental en la venta online ya que ayuda a los usuarios a obtener su auto adecuado en base a sus necesidades. Como menciona Pigna (2023):

“Los automóviles son ampliamente utilizados como un medio de transporte personal, permitiendo a las personas llegar a sus destinos de manera cómoda y rápida. Esto es especialmente valioso en áreas donde el transporte público es limitado o poco eficiente.”

Para estar en contexto con este problema, podremos usar un sistema en grafos. Estos son estructuras de datos que permiten las relaciones entre distintos elementos de forma eficiente, ahora con esta información, podemos representar los autos con las características mencionadas anteriormente como los nodos del grafo y las relaciones se pueden representar con aristas. Esto permitiría modelar las similitudes de los vehículos o crear una comparativa de estos para que el usuario pueda elegir acorde a su presupuesto y necesidad. Imaginemos el siguiente caso: El usuario busca un automóvil de marca tesla del año 2022, el sistema puede usar algoritmos de búsqueda en el grafo establecido para poder encontrar mediante un filtro todos los automóviles disponibles que cumplan las especificaciones del usuario.

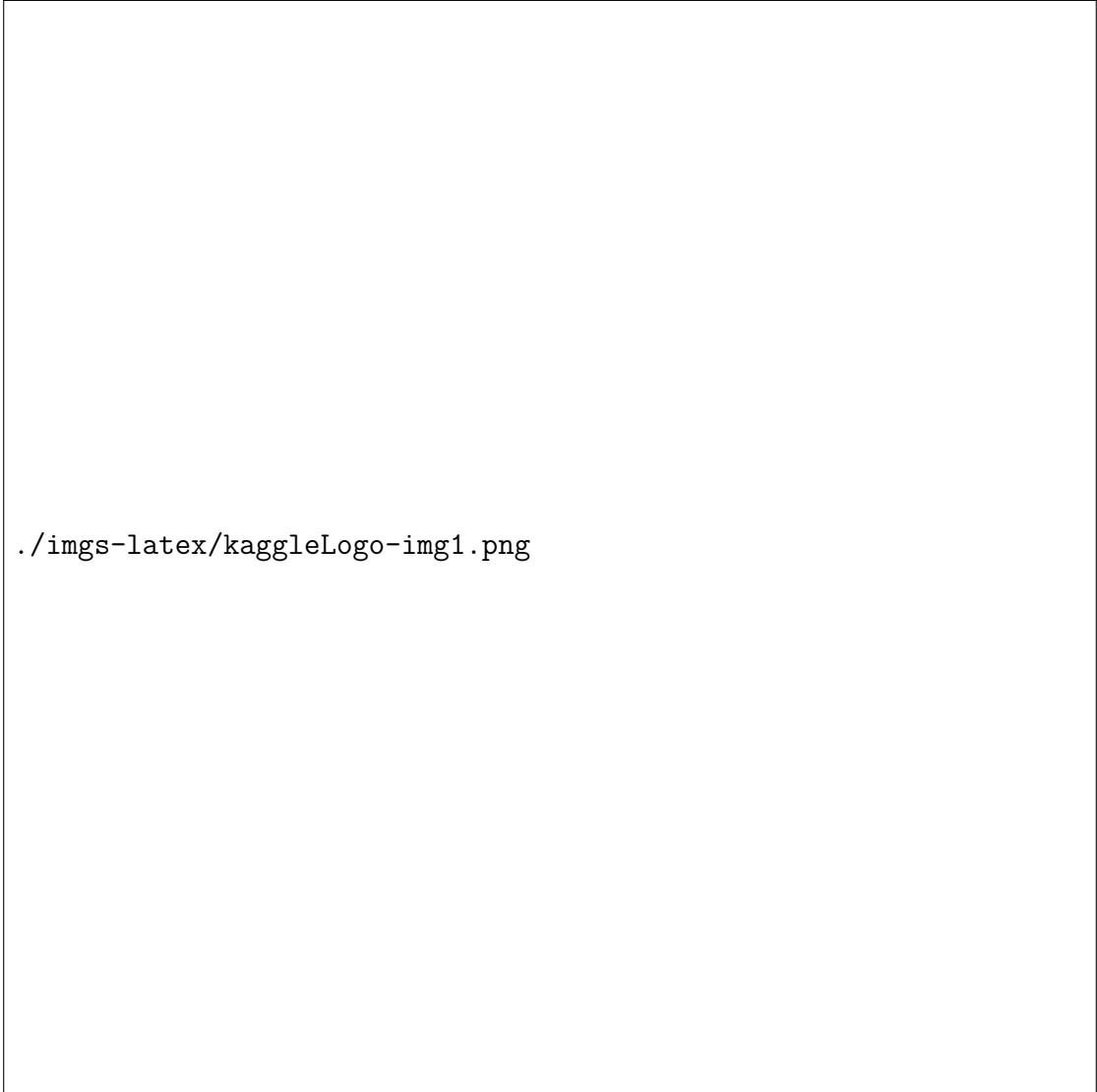
2 Descripción del conjunto de datos (dataset)

2.1 Características y origen de datos móviles de análisis

Vamos a abordar el problema mencionado anteriormente teniendo en cuenta ciertos criterios específicos, que incluyen el uso de un conjunto de datos creado para analizar y evaluar diferentes estrategias y soluciones. Por lo tanto, hemos desarrollado un conjunto de datos para recomendaciones de autos eléctricos, que contiene información detallada sobre autos eléctricos y sus características. En esta sección, se explicarán las características y el origen de los datos, además de presentar gráficos visuales que muestran la complejidad de las relaciones entre los autos eléctricos. El objetivo es proporcionar a los usuarios sugerencias personalizadas según sus preferencias y necesidades específicas, como la modelo, año del modelos, ubicación , etc. Este tipo de recomendación es crucial, ya que facilita a los clientes la búsqueda de vehículos eléctricos que se ajusten a sus necesidades individuales.

2.1.1 Origen de los datos

Los datos que se emplearán para el análisis de este proyecto principalmente han sido sacados de la plataforma en línea Kaggle. El nombre del dataset en línea es *Electric Vehicle Population Dataset* o, en su versión en español, *Conjunto de datos de población de vehículos eléctricos*. A su vez, se ha hecho generación de datos en algunas columnas de la base de datos final a utilizar en WattzFinder, el procedimiento detallado se explicará a continuación:



`./imgs-latex/kaggleLogo-img1.png`

Al principio, se recopila información importante sobre las recomendaciones de autos eléctricos durante un período determinado.

`./imgs-latex/dataExplorerer-csv-screenshot-img1.png`

Luego se traduce los nombres de las columnas para tener una base de datos más organizada, debido a que nuestro proyecto Wattzfinder trabajará como un Sistema de recomendación basado en grafos se necesitará datos cuantitativos para poder utilizar diferentes métodos algorítmicos de grafos por datos cuantitativos, por ello se utilizará la biblioteca pandas para poder cargar la base de datos en un archivo csv en un DataFrame.

Uno de los principales problemas de nuestra Dataset sacada de Kaggle es que en la columna de Precios la mayoría valores son = 0, lo cual no tendría sentido para nuestro sistema de recomendación de carros electrónicos, por lo que en el siguiente código se extraerán y reemplazarán los precios que son = 0 con los datos originales de Kaggle que != 0, se utilizarán los datos previos y se distribuirán en las filas necesarias de manera aleatoria, a su vez se utilizará percentiles para mantener la distribución general de precios de la dataset original y completar la mayoría de datos vacíos en la columna Precios.

Luego se generarán 2 columnas adicionales por el mismo código “Descuentos” y “Precio Final”, los descuentos se agregarán desde un 8% hasta un 40%, este descuento solo se aplicará a las filas con datos generados en Precios, es decir que los datos sacados en la dataset de Kaggle tendrán un descuento de 0%, esto se hará con el motivo que luego puedan ser identificables y observar si hay una correlación imprevista cuando hagamos el análisis final de nuestro proyecto y corregir si existe un error.

Aparte de utilizar la bibliotecas ‘pandas’ para los procesos relacionados con el archivo csv, se utilizará ‘numpy’ para el procesamiento de números en las 3 columnas modificadas en cuanto a la aleatoriedad y los percentiles.

En cuánto al código en general se utilizará las bibliotecas ‘sys’ y ‘datetime’ para traspasar el archivo con el script python como argumento y modificar su nombre con la fecha respectiva del procesamiento de datos.

El código python es el siguiente:

3 Propuesta

El objetivo de nuestro proyecto WatzzFinder es desarrollar un sistema de recomendación avanzado basado en algoritmos de búsqueda en grafos. Este sistema representará y almacenará datos sobre vehículos eléctricos y sus interrelaciones de manera visual, permitiendo a los usuarios seleccionar el vehículo más adecuado según sus preferencias personales.

4 Diseño del aplicativo

4.1 Diagrama de clase y diagrama de base de datos

4.2 Módulo para la adecuación de datos fuente a estructura de datos tipo grafo

4.3 Módulo para la representación de datos en el grafo

4.4 Interfaz tentativa (visual) para usuarios

5 Validación de resultados y pruebas

Validamos los resultados a través de...

6 Conclusiones

La recomendación de vehículos es crucial en las ventas online, ya que ayuda a los usuarios a encontrar el auto adecuado según sus necesidades individuales. El uso de un sistema en grafos permite modelar las relaciones entre diferentes autos de manera eficiente, representando las similitudes entre ellos y permitiendo una comparación fácil para que los usuarios elijan según su presupuesto y necesidad.

7 Referencias bibliográficas

- Pigna, A. (2023, 26 septiembre). *¿Para qué sirve el auto? Utilidad, función y ventajas*. Kavak. Available at: <https://www.kavak.com/mx/blog/para-que-sirve-el-auto-utilidad-funcion-y-ventajas>
- Carazo, F., & Amat, J. (2023, Abril). *Detección de comunidades en grafos y redes con python*. Available at: <https://cienciadedatos.net/documentos/pygml02-detecion-comunidades-grafos-redes-python>
- Javi, G. (2020, Agosto 28). *Spectral Clustering*. Available at: https://javi897.github.io/Spectral_clustering/
- Lubos, D., Fritscher, E., & Kriz, J. (2013). *Movie Recommendation Based on Graph Traversal Algorithms*. Available at: <http://www2.fiiit.stuba.sk/~bielik/publ/abstracts/2013/televido-dexa2013.pdf>