A picture containing text

Description automatically generated

Análisis de grandes volúmenes de datos

Dr. Néstor Velasco Bermeo

Actividad 4.3 Avance de proyecto 1

Sistema de Recomendación

Mayo 17 2024

**EQUIPO 8**

Joel Orlando Hernández Ramos

A00759664

María de los Ángeles Rabelero Campos

A01793031

Daniel Gómez Álvarez

A01794498

# Descripción General del Proyecto

Anteriormente, se había mencionado que los Sistemas de Recomendación son herramientas de filtrado de información que proporcionan a los usuarios el contenido adecuado y personalizado, reduciendo significativamente el esfuerzo y el tiempo necesario para buscar información relevante en internet. Estos sistemas utilizan complejos algoritmos para predecir las necesidades y deseos de los usuarios, mejorando así continuamente la precisión de sus recomendaciones a través del análisis de grandes volúmenes de datos.

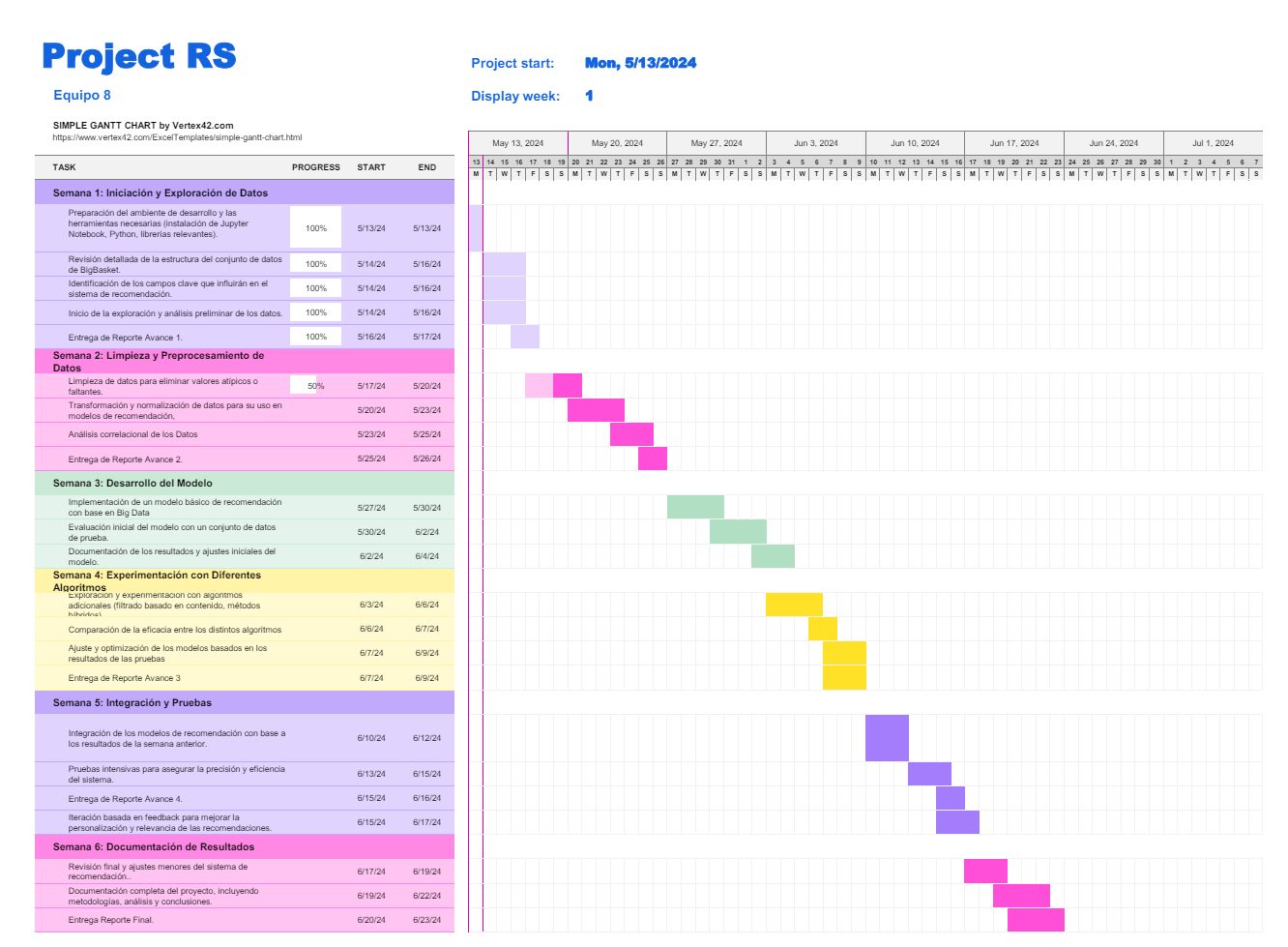
Aunado a esto, los sistemas de recomendación son una importante herramienta en el mundo empresarial, pues optimiza indicadores relevantes comerciales como las ventas, la participación de los usuarios y las tasas de retención de clientes. La implementación estratégica de sistemas de recomendación puede llevar a resultados empresariales mejorados, desde tasas de clics más altas y ventas, hasta una mayor lealtad del cliente y diferenciación competitiva.

En este sentido, el Equipo 8 del curso Análisis de grandes volúmenes de dato creará un sistema de recomendación adaptado para el sector del comercio electrónico, utilizando el conjunto de datos públicos de BigBasket, que abarca una lista completa de productos. Este sistema está diseñado para mejorar la experiencia de compra ofreciendo recomendaciones de productos personalizadas que se alinean con las preferencias y comportamientos de compra de cada cliente.

El núcleo de este proyecto implica analizar factores como el historial de compras, las valoraciones de los productos y las demografías de los usuarios para predecir y recomendar con precisión los productos que probablemente serán comprados. Mediante el uso de tecnologías de grandes datos, el sistema manejará grandes volúmenes de datos, permitiendo la identificación de patrones y tendencias significativos que contribuyen a realizar recomendaciones precisas.

Además, el proyecto evaluará diversos algoritmos y técnicas empleados en sistemas de recomendación, incluyendo el filtrado colaborativo, el filtrado basado en contenido y métodos híbridos. Esta exploración ayudará a identificar el enfoque más eficiente para personalizar las sugerencias de productos de manera efectiva, aumentando potencialmente tanto la satisfacción del cliente como la rentabilidad del negocio.

En general, este proyecto no solo busca desarrollar un sistema de recomendación funcional, sino también pretende profundizar en la comprensión del comportamiento de los consumidores en línea y la dinámica del mercado de comercio electrónico, proporcionando al estudiante una valiosa experiencia práctica en la aplicación de análisis de grandes datos en contextos empresariales reales.



# Descripción de la Base de Datos

Big Basket, como una de las tiendas de comestibles en línea más grandes de la India. En este sentido, la empresa cuenta con una base de datos pública en formato **.csv** que incluye una variedad de información, como listas de productos, rating de los productos, precio comercial y precio de venta, etc. En este sentido, la base de datos elegida tiene una estructura general de 27,555 productos, cada uno con 10 dimensiones enlistadas a continuación:

1. **index**: Un índice entero único para cada entrada de producto.
2. **product**: El nombre del producto.
3. **category**: La categoría amplia bajo la cual se clasifica el producto (11 categorías únicas).
4. **sub\_category**: Una clasificación más específica dentro de la categoría principal (90 subcategorías únicas).
5. **brand**: El nombre de la marca asociada con el producto. El conjunto de datos incluye 2,313 marcas únicas, destacando la diversidad en la oferta de productos.
6. **sale\_price**: El precio de venta del producto en el conjunto de datos.
7. **market\_price**: El precio de mercado o minorista del producto antes de cualquier descuento.
8. **type**: Proporciona una clasificación aún más detallada del producto, especificando tipos o características particulares (por ejemplo, "Aceite y sérum para cabello" bajo "Cuidado del cabello").
9. **rating**: Una calificación numérica dada por los clientes, aunque no todos los productos tienen una calificación.
10. **description**: Una descripción textual que proporciona detalles sobre el producto; algunas entradas pueden faltar.

En este sentido, es importante destacar que esta base de datos cuenta con características que la hacen una opción robusta para el desarrollo de este proyecto.

* **Categorización Detallada:** La estructura del conjunto de datos incluye categorías y subcategorías, lo que permite al sistema de recomendación segmentar productos de manera más eficiente y sugerir artículos que sean más relevantes según las preferencias específicas del usuario. Por ejemplo, un cliente interesado en productos de cuidado personal se beneficiaría de recomendaciones que se ajusten precisamente a su historial de búsqueda y compra en esa subcategoría.
* **Información sobre Precios:** Incluye tanto precios de venta como de mercado, lo cual podría ser útil para comparaciones de precios, análisis de descuentos y comprensión de las decisiones de compra de los consumidores.
* **Retroalimentación del Consumidor**: La disponibilidad de calificaciones permite el análisis de las preferencias de los clientes y el rendimiento de los productos.
* **Información sobre Marcas:** El conocimiento de las marcas de los productos permite al sistema ofrecer recomendaciones que no solo se ajustan a las preferencias de tipo de producto, sino también a lealtades o aversiones hacia marcas específicas, lo cual es un factor decisivo en las decisiones de compra de muchos consumidores.

Finalmente, con base en estas características, se espera que este conjunto de datos proporciona una fuente rica de información para analizar patrones de compra, popularidad de productos y tendencias de mercado. El sistema de recomendación se espera que cuente con las siguientes características.

* **Optimización de la Experiencia de Usuario:** Al implementar un sistema de recomendación que utilice este conjunto de datos, los usuarios podrán descubrir productos que quizás no encuentren por sí mismos, mejorando la navegación y la satisfacción del cliente, lo que a su vez puede aumentar la fidelidad y las tasas de conversión.
* **Relevancia Comercial y Escalabilidad:** Al tratarse de un dataset derivado de un contexto de comercio electrónico real, las recomendaciones generadas son más aplicables y prácticas en escenarios comerciales reales. Además, la diversidad y el volumen de los datos facilitan la escalabilidad del sistema a medida que se expande el inventario o se diversifica la base de clientes.

# Exploración Inicial y Análisis del Conjunto de Datos

En este primer avance se realizó un primer análisis y limpieza de los datos. Los pasos de dicha tarea están descritas del paso 3 al paso 9 en el Jupyter notebook que puede ser visualizado con detalle en el siguiente repositorio privado de GitHub:

<https://github.com/DanielGAMNA/TC4034_EQUIPO8/blob/main/Proyecto_Avance_1_8.ipynb>

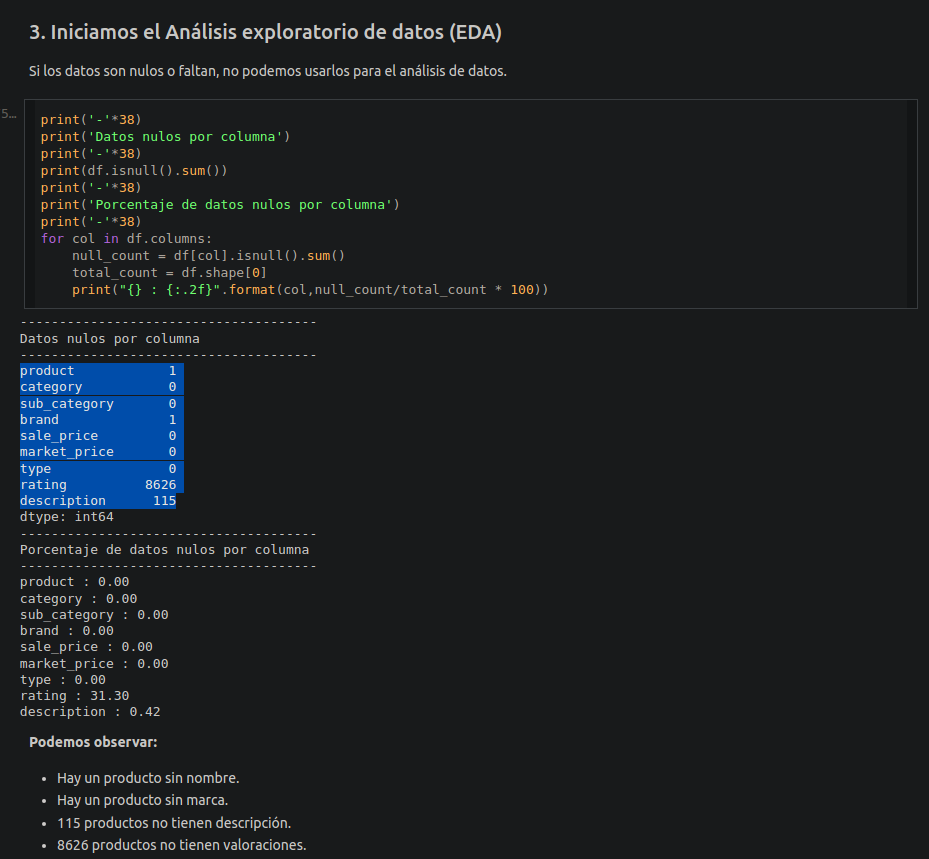


Figura 1. Análisis Exploratorio de los datos

# Algoritmo de Recomendación Básico

Adicionalmente, si implemento un algoritmo básico de recomendación como prueba. Dicha tarea está descrita del paso 10 al paso 12 en el Jupyter notebook que puede ser visualizado con detalle en el siguiente repositorio privado de GitHub:

<https://github.com/DanielGAMNA/TC4034_EQUIPO8/blob/main/Proyecto_Avance_1_8.ipynb>



Figura 2. Algoritmo de Recomendación

# Conclusiones

En conclusión, el proyecto del Equipo 8 ha logrado establecer las bases fundamentales para el desarrollo de un sistema de recomendación orientado al sector del comercio electrónico. A través del análisis del conjunto de datos de BigBasket, hemos identificado características importantes que facilitarán la generación de recomendaciones personalizadas. Este sistema promete no solo mejorar la experiencia de compra del usuario, sino también incrementar la eficiencia comercial y la rentabilidad de las empresas que implementen esta tecnología.

Los desafíos encontrados, especialmente en la limpieza y tratamiento de los datos, han sido superados, demostrando que la selección de la base de datos parece ser la correcta. La implementación inicial de un algoritmo de recomendación básico ha indicado un potencial significativo, y se espera que su refinamiento mediante técnicas más avanzadas mejore notablemente su precisión y aplicabilidad. Se espera poder aplicar técnicas de Big Data que se vayan enseñando durante el curso.

Las próximas etapas del proyecto incluirán una evaluación más exhaustiva de diversos algoritmos y técnicas de recomendación. Este análisis permitirá optimizar el sistema para que responda de manera eficaz a las necesidades cambiantes de los consumidores y del mercado.