Sistema de recomendación para la industria de la ropa.

*Jorge Eduardo Enciso, Cristian Cristancho, Nikolas Sebastian Rodriguez*

*Pontificia Universidad Javeriana*

# INTRODUCCION

Una de las principales industrias que dejan expresar la opinión de las personas es la industria de la moda, mediante esta las personas pueden comunicar información de ellos mismos como por ejemplo la personalidad o profesión de uno, por esta razón las personas buscan ropa que los identifique y este proceso puede ser demorado ya que cada persona es única y encontrar ropa con la que se sientan cómodos e identificados puede ser un proceso muy complicado. Usando la tecnología como medio la industria de la moda o ropa puede facilitar el proceso de obtención de las prendas y puede brindar un servicio más personalizado a la hora de comprar. Las compras de los usuarios dependen de muchos factores tales como variables demográficas, posición geográfica y las relaciones interpersonales que pueden tener. La combinación de estas variables hace que el proceso de recomendación por parte de las industrias sea cada vez más difícil.

Uno de los grandes retos de la industria es como identificar los intereses de los usuarios y de esta forma recomendarles productos que tengan mayor probabilidad a ser comprados. Gracias a que la mayoría de los negocios de moda han migrado a internet, el número de transacciones en línea también ha aumentado drásticamente. El e-commerce es fundamental para la mayoría de los negocios de ropa porque le permite al distribuidor estar en contacto directo con el cliente. Por esta razón, es primordial que estos negocios cuenten con un sistema de recomendación que ayude al desarrollo del proyecto.

# Datos

La base de datos es provista para resolver el desafío de Recysis 2022, este es un desafío que se presenta todos los años el cual tiene como objetivo desarrollar un sistema de recomendación usando inteligencia artificial. Este año la empresa que fue la fuente de información de los datos fue Dressipi, la cual es una compañía de tecnología el cual da recomendaciones a retailers de talla mundial, generando experiencias personalizadas a los compradores durante todo el proceso de compra de productos.

El objetivo del desafío de este año es realizar una predicción usando el histórico y las características de las prendas compradas por las sesiones de los usuarios. La predicción estará a nivel de sesión y se tendrá que predecir las próximas prendas que el usuario o sesión va a comprar según su historial de compras.

La base de datos esta constituida por tres archivos el cual tiene los siguientes datos.

1. Sesiones: Los productos vistos por la sesión durante un día.
2. Compras: Las compras que se realizaron durante el día. Solo habrá una compra por sesión.
3. Características de los productos: Características de los productos codificadas.

La métrica de desempeño que se usara para la medir la efectividad del modelo es el promedio ranking reciproco el cual es una medida que comprueba que tan correcta es la secuencia. Ya que es Esta métrica es muy usada en problemas secuenciales y en este caso lo que va a medir es el orden de compra de los siguientes productos comprados. El rango de esta métrica es de 0-1 donde tiene el objetivo de maximizar esta misma.

Actualmente la tabla de resultados esta constituida por los 50 equipos que mejor promedio reciproco tengan, el valor para ocupar un puesto entre estos equipos esta entre 0.18 y 0.21.

# Exploración de datos

Para la exploración de datos se hará una descripción inicial de los datos iniciales de entrenamiento. En primer lugar, la base de datos esta constituida de 4,743,820 sesiones, las cual son 1,000,000 sesiones únicas y contienen 23,496 productos únicos.

A picture containing shape

Description automatically generated

Imagen 1 Cantidad de vistas por sesión

En promedio la cada sesión contiene revisa 4 productos, sin embargo, tiene una desviación estándar de 6 productos, lo cual quiere decir que existe una dispersión muy grande entre los clientes, y estos miran entre 0-10 productos en general, En la imagen 1 se puede observar que la cantidad de productos vistos tiene una distribución similar a la exponencial y que los percentiles son cuantil 1 es de 1 producto y la mediana es de 3.

Por otro lado, también están las compras realizadas por las sesiones en el set de entrenamiento. La base de compras esta constituida por 1,000,000 registros los cuales han comprado 18,907 productos diferentes.

Además de estas compras se tiene las características de los productos los cuales muestran que existen 23,691 productos únicos con 73 características diferentes entre estos. En la imagen 2, se puede evidenciar la cantidad de características únicas que tienen los productos en general. Donde en promedio cada producto tiene entre 14 y 24 características diferentes, donde la mayoría de los productos tiene entre 24 y 25 características.

Chart, histogram

Description automatically generated

Imagen 2 Cantidad de características por producto

Asimismo, se hizo un conteo de las características más repetidas para tener conocimiento de posibles sesgos que puede tener la base de entrenamiento a la hora de predecir productos. A partir de esto como se presenta en la imagen 3, la característica que se repite alrededor de 4,366 es la característica 30 seguida de la característica 4 con un total de 272 apariciones. Este análisis ayuda a entender cuales son esas características que casi no aparecen en la base de datos y en el caso de que se quiera disminuir dimensionalidad se podrá eliminar los productos que tengan poca ocurrencia y que estén repetidas.

A picture containing chart

Description automatically generated

Imagen 3 Conteo de las características repetidas.

Por último, en la exploración de datos se encuentra la cantidad de valores por categoría única. Como se puede apreciar en la imagen 4, las características únicas por categoría. En total hay 73 categorías únicas las cuales en promedio cada categoría tiene 12 características únicas con una desviación estándar de 13, es decir la mayoría de los datos van a estar entre 1 y 25 características.

Chart, bar chart

Description automatically generated

Imagen 3 Histogramas de las características por categoría.

# Metodologías Empleadas

1. Autoencoder
2. RNN
3. Item2Vec

# Protocolo experimental

# Conclusiones

Reconocimiento

Este trabajo no es financiado por ningún tipo de institución.

Referencias

1. Chakraborty, S., Hoque, M. S., Rahman Jeem, N., Biswas, M. C., Bardhan, D., & Lobaton, E. (2021). Fashion Recommendation Systems, Models and Methods: A Review. Informatics, 8(3), 49. doi:10.3390/informatics8030049