



# Éxito de la aplicación de Machine Learning en la empresa

## Sistemas de Recomendación

30.08.2023

Balague Joaquín

Procesamientos de datos con Python

Ticmas academy

## Introducción

Los sistemas de recomendación son utilizados por Netflix, recomendando series o películas, Spotify sugiriendo canciones y artistas y Amazon ofreciendo productos. Todo E-Commerce exitoso utiliza esta herramienta para obtener ventajas competitivas y potenciar sus ventas.

## ¿Qué son los Sistemas ó Motores de Recomendación?

Los sistemas de recomendación, ("recommender systems") son algoritmos que intentan predecir lo que deseara adquirir un usuario en particular. Cómo por ejemplo : productos, canciones, películas, juegos, etc.

Son un tipo de algoritmo de aprendizaje automático que se utilizan tanto en problemas de clasificación como de regresión. Estos algoritmos no son supervisados ni tampoco supervisados, sino que se consideran un tipo de aprendizaje basado en memoria.

Estos intentan personalizar al máximo lo que ofrecerán a cada usuario. Dada la cantidad de información individual que podemos recabar de las personas y nos posibilita tener una mejor tasa de aciertos, mejorando la experiencia de la persona con cierta exactitud.

## Tipos de motores

Entre las estrategias más usadas encontramos:

- **Popularidad:** Aconseja por "popularidad" de los productos. Por ejemplo, "los más vendidos" globalmente, se ofrecen a todos los usuarios por igual sin personalizar. Es de fácil implementación y en algunos casos es efectiva.
- **Basado en contenido:** A partir de productos visitados por el usuario, se intenta predecir qué busca el usuario y ofrecer productos similares.
- **Colaborativa:** Utiliza la información de masas para identificar perfiles similares y aprende de los datos para recomendar productos de forma individual.

A continuación se profundizará en el motor de tipo colaborativo.

## Collaborative Filtering (Filtrado colaborativo)

Proporciona recomendaciones personalizadas a los usuarios en función de su comportamiento y preferencias pasadas. Implica el uso de similitudes entre usuarios e ítems simultáneamente para proporcionar recomendaciones. El filtrado colaborativo se puede dividir en dos clases: **basado en usuarios y basado en ítems**.

### Basado en usuarios:

- Se identifican usuarios similares
- Se recomiendan nuevos ítems a otros usuarios basado en el rating dado por otros usuarios similares (que no haya valorado este usuario)

### Basado en ítems:

- Calcula la similitud entre ítems
- Encuentra los mejores ítems similares a los que un usuario no tenga evaluados y los recomienda.

### Predicción de gustos (basado en usuarios) :

Se intentará encontrar usuarios similares, para ofrecerle ítems bien valorados para ese perfil en concreto. Hay diversas maneras de medir o calcular la similitud entre usuarios y de ello dependerá que se den buenas recomendaciones. Se debe tener en cuenta que se busca la similitud entre gustos del usuario sobre esos ítems, no se busca perfiles similares por ser del mismo sexo, edad ó nivel educativo. Sólo se utilizará los ítems que ha experimentado, valorado para agrupar usuarios parecidos.

Una de las maneras de medir esa similitud se llama **distancia por coseno de los vectores** y este crea un espacio vectorial con n dimensiones correspondientes a los n ítems y sitúa los vectores siendo su medida el valor rating de cada usuario (a ese ítem) Luego calcula el ángulo entre los vectores partiendo de la coordenada cero. A poca distancia entre ángulos, se corresponde con usuarios con mayor similitud.

Este método no es siempre perfecto, pero es bastante útil y rápido de calcular

## Cálculo de Ratings

Una vez que tenemos la matriz de similitud, emplearemos otra operación matemática para calcular las recomendaciones.

$$R_U = \left( \sum_{u=1}^n R_u * S_u \right) / \left( \sum_{u=1}^n S_u \right)$$

La fórmula para calcular los ratings faltantes: sería algo así como “Matriz de similitud PROD.VECTORIAL ratings / (sumatoria de cada fila de ratings) Transpuesta

Se realiza lo siguiente : cada rating se multiplica por el factor de similitud de usuario que dio el rating. La predicción final por usuario será igual a la suma del peso de los ratings dividido por la “suma ponderada”.

## Conclusiones

Crear un sistema de recomendación en Python con Machine Learning se centra en tener los datos correctos y un volumen alto. También es central el valor que utilizaremos como rating, siendo una valoración real de cada usuario ó un valor artificial que creemos adecuado. Tomamos a rating como ese puntaje que surge de la intersección entre usuario e ítems en un dataset. Luego se evalúa entre las opciones de motores basado en usuarios, basado en ítems y se selecciona el que menor error tenga. No se debe dejar de probar en el mundo real y ver qué porcentaje de aciertos (o feedback) generan los usuarios reales de tu aplicación.