

# Netflix

November 4, 2022

## 1 PRÁCTICA 1: Desgranando el Sistema de Recomendación de Netflix

Asignatura: Aprendizaje automático

Marta Venegas Pardo

### 1.1 Breve historia introductoria

El objetivo de la práctica es describir el algoritmo de recomendación de Netflix, pero previamente, haremos una breve descripción de los comienzos de esta plataforma, que se ha convertido en la plataforma de streaming más valorada en la actualidad.

**Netflix** es una plataforma de streaming Estadounidense fundada en agosto de 1997 por Reed Hastings y Marc Randolph que en Julio de este año contaba con más de 200 millones de suscriptores.

Sin embargo, esta empresa no ha sido siempre una plataforma de streaming, ya que en sus comienzos alquilaba DVD por correo postal en Estados Unidos. Una característica diferenciadora era que la empresa no penalizaba a los clientes por devolver los videos fuera de la fecha establecida, y hasta que el título no fuera devuelto, el cliente no podía alquilar otro. No es hasta el año 1999, un año después de su fundación, que la compañía comenzó a ofrecer una suscripción de pago para apoder alquilar DVD's ilimitados. En el año 2007 fué introducido el servicio en streaming a la carta en dispositivos personales.

Netflix, comienza en el año 2011 a producir contenido original, por lo que no solo ofrecía a sus suscriptores películas, series y documentales existentes, sino que comienza a lanzar su propio contenido como series o documentales exclusivas de la plataforma. Sin embargo, no fué hasta el año 2010 cuando empezó su expansión a nivel internacional, comenzando por Canadá (2010), luego Latinoamérica y el Caribe (2011), Holanda, Alemania, Austria, Bélgica, Francia, Luxemburgo y Suiza (2014). En el año 2020 ya está disponible en más de 190 países. Netflix gana en el año 2018 un Óscar por el documental Ícaro. Otra cifra destacable es el exponencial crecimiento en cuanto a número de suscriptores en el primer trimestre del año 2020, coincidiendo con la pandemia por el COVID-19, con un aumento de 16 millones de nuevos usuarios.

### 1.2 Ideas clave de la lectura

Después de la lectura del artículo propuesto y de alguna que otra búsqueda por internet podemos extraer las siguientes ideas claves iniciales:

- En el año 2020 Netflix gana más de 37 millones de nuevos suscriptores

- Recogida de gran cantidad de datos (interacción con la web, correlación entre el gusto y las preferencias del usuario, contenido de netflix, hora del día, dispositivo, tiempo de visualización) toda esta información analizada genera la recomendación de contenido personalizada para cada usuario.
- Nuevo algoritmo de recomendaciones basadas en análisis de gráficos: para captar la atención de los usuarios (estudios muestran que los usuarios perderán el interés en el uso de la plataforma si no encuentran nada interesante en los primeros 90 segundos)
- AI para la recomendación de contenido: inicialmente pide gustos para comenzar a recomendar, pero cuanto más se use, mas acertada será la recomendación
- Es muy importante la experiencia del usuario y la personalización para cada usuario.
- Empezó alquilando DVD por correo postal en EEUU (1998) y no fué hasta 2007 que lanzó el servicio de reproducción de videos a la carta
- Usa la nube de AWS desde Junio de 2016 (también Open Connect), empieza en 2008
- Distintos sistemas recomendadores (transformación): puntuación -> ránking -> generación de páginas -> maximizar la experiencia del usuario (videos personalizados, top-N, contenido interesante, películas populares actualmente)

### 1.3 Descripción del algoritmo

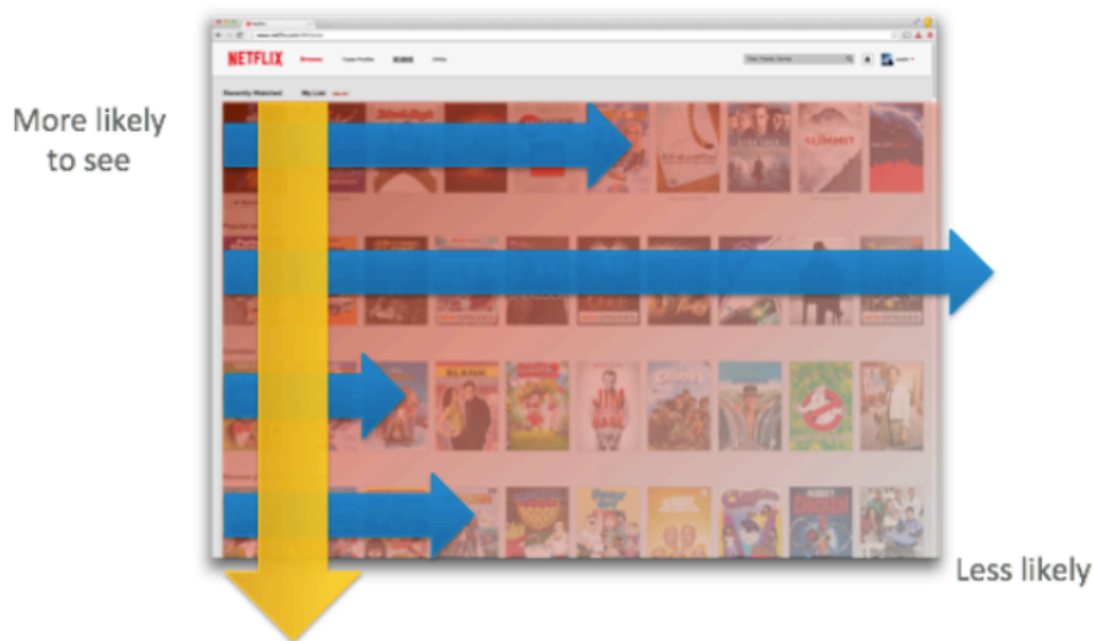
Los sistemas de recomendación ayudan a los diferentes usuarios a elegir películas o series de entre un catálogo de más de 15000 títulos diferentes. Es importante captar la atención de los usuarios, ya que los clientes perderán el interés en la plataforma si durante los primeros dos minutos no ha encontrado algo para ver. Es por eso que estos sistemas de recomendación tienen que ofrecer la mayor personalización posible para el usuario, maximizando así la experiencia de éste.

Cuando un nuevo usuario se registra en la plataforma, se le pide que elija entre diferentes títulos para así poder iniciar la recomendación de diferentes series o películas que podrían gustarle. Conforme vamos consumiendo contenido de la plataforma, el sistema de recomendación comienza a comprender los gustos de preferencias de cada uno de sus usuarios, ofreciendo así un contenido personalizado en función de los datos recopilados. El contenido es ofrecido de la siguiente manera: **encuentra similitudes entre perfiles diferentes, y por tanto hará recomendaciones basadas en perfiles similares.**

Netflix va recopilando todo tipo de información, como el dispositivo utilizado, la hora del día, el tiempo que pasamos en la plataforma, ubicación, título, género y también si hemos terminado de ver lo que iniciamos o no. Toda esta información es almacenada y utilizada para la creación de los diferentes algoritmos de recomendación.

Cuando el usuario accede a la página principal de su perfil, ya está accediendo a contenido personalizado, atendiendo a diferentes categorías:

- **Seleccionadas para el usuario:** se trata de una recomendación de títulos exclusiva para cada usuario específico en función del contenido consumido.
- **Tendencias:** Se trata de contenido popular en la plataforma
- **Contenido similar a:** Son títulos cuyo contenido es similar a otra serie o película que el usuario haya visto recientemente o el tiempo que haya pasado viéndola
- **Noticias:** Se trata de nuevos títulos que podrían ser interesantes para ciertos usuarios



A continuación se muestra una imagen que nos puede ayudar a comprender mejor como los diferentes algoritmos de recomendación hacen que los títulos se ordenen en función de las preferencias de cada usuario.

El orden de recomendación es el siguiente: *de izquierda a derecha y de arriba a abajo*, es decir, el título que el algoritmo ha considerado que el usuario podría elegir es el que se encuentra en la primera posición de la primera fila, después el segundo y así hasta acabar la primera fila y pasar a la siguiente.

Encontramos dos modelos principales que utiliza la plataforma para filtrar los datos:

- Filtro basado en contenido: compara la información de cada título a nivel individual, como reseñas o comentarios
- Filtro colaborativo: este algoritmo trata de encontrar la relación existente entre las preferencias y el comportamiento de cada usuario

A continuación, netflix va a calificar cada título en una categoría (Cada una de las filas de la página de inicio)

Netflix utiliza diferentes algoritmos de recomendación:

- **Clasificación de video recomendada (PVR):** se trata de un algoritmo que divide los títulos en función de sus características (películas románticas, documentales sobre animales,...) y ofrece contenido a cada usuario en función de sus preferencias personales.
- **Clasificación del Top-N:** clasifica los títulos en función de las preferencias top de cada usuario, determinando así los títulos más demandados de cada categoría.
- **Clasificación de contenido interesante para su posterior visualización:** algoritmo que comprueba si un usuario ha terminado el título que empezó o no, y analiza datos como por ejemplo cuánto tiempo ha pasado desde la última vez que lo vió, dónde lo dejó de ver o en qué dispositivo, calculando así la probabilidad de que el usuario quiera continuar viéndolo.

- **Clasificación de películas populares:** este algoritmo analiza que tipo de contenido es más popular en momento actual atendiendo a información que puede llevar a conocer la tendencia de visualización de los usuarios, por ejemplo, teniendo en cuenta eventos mundiales (el covid hizo que se dispararan las visualizaciones de la película contagio, he de decir que yo la ví durante la pandemia) o festivos como Halloween, Navidad o San Valentín, donde las visualizaciones de títulos de estos temas se disparan.
- **Recomendaciones basadas en la portada de las películas:** este es un nuevo algoritmo de recomendación que utiliza imágenes para captar el interés de los usuarios, de tal manera que crea diferentes miniatras para cada una de sus series o películas. Así ofrece a cada uno de sus clientes aquella imagen que pueda resultar más interesante basándose en la información almacenada sobre éste, es decir, el algoritmo personaliza las imágenes del encabezado de las películas para tratar de llegar a más usuarios. Este último algoritmo es un ejemplo clave de mejora de la experiencia de usuario, personalizando cada perfil en función de los datos de éste.

Se estima que más del 80% del contenido consumido de Netflix proviene de los diferentes sistemas de recomendación, es decir, la gran mayoría de visualizaciones provienen de los algoritmos de recomendación.

Ahora nos preguntamos ¿qué datos son los que Netflix utiliza realmente para determinar la probabilidad de que un usuario vaya a ver una serie o película?

Se recogen datos como interacción del usuario con el sitio web (títulos buscados, guardados, calificaciones,...), hora del día, tipo de dispositivo, tiempo de visualización, correlaciones entre gustos de individuos dentro de un mismo grupo de preferencias o datos sobre el contenido de la plataforma (género, idioma, actores,...).

El sistema de recomendación también se basa en lo que se conoce como *A/B testing*, en el que se prueban distintas opciones relacionadas con sugerencias de películas, miniaturas o diferentes organizaciones de títulos para tratar de averiguar que despertará mayor interés en el usuario, es decir, tratar de averiguar que título será mas probable que un usuario determinado seleccione. Estos test suelen realizarse antes de introducir cambios en la plataforma, ya que se miden las métricas antes y después de realizar el cambio únicamente para ciertos usuarios obteniendo así datos que refuten las decisiones tomadas.

Un ejemplo sencillo de un posible test A/B:

Dividiremos a los usuarios de netflix que ven con frecuencia títulos románticos en dos grupos A y B. Al grupo A, que se trata del 50% de usuarios que ven con frecuencia películas románticas le mostraremos una miniatura de una escena romántica en diferentes títulos y al otro 50% no. Es importante que el resto de condiciones se mantengan constantes en ambos grupos. A continuación, se recogerán datos acerca del número de visualizaciones que han tenido estas películas y se medirá en qué grupo se han visto más películas. Si sale que en el grupo A se han visto más veces estos títulos ofrecidos con la miniatura romántica, nos indicaría que el mostrar miniaturas relacionadas con los gustos del usuario hace que éstos vayan a ver esos títulos con mayor probabilidad. Esto se traduciría en que si a un usuario le gustan las comedias románticas, el sistema de recomendación hará que lo que vea sea una miniatura que presenta una escena romántica de la película que el sistema le esté recomendando.

Video interesante: <https://www.youtube.com/watch?v=f8OK1HBEgn0>

### 1.3.1 Fun Fact

Para terminar, expondremos un dato interesante sobre Netflix. Este es que esta plataforma utiliza AWS, un servicio en la nube (también Open Connect). El motivo de migrar todo a la nube se debe a un fallo que hubo en uno de los data centers de Netflix durante tres días en agosto de 2008, días en los que el servicio de envíos de DVDs estuvo parado debido a que una base de datos estaba dañada. La empresa tardó más de ocho años en trasladarse por completo a la nube, pero ésta le ha ofrecido mayor eficacia.

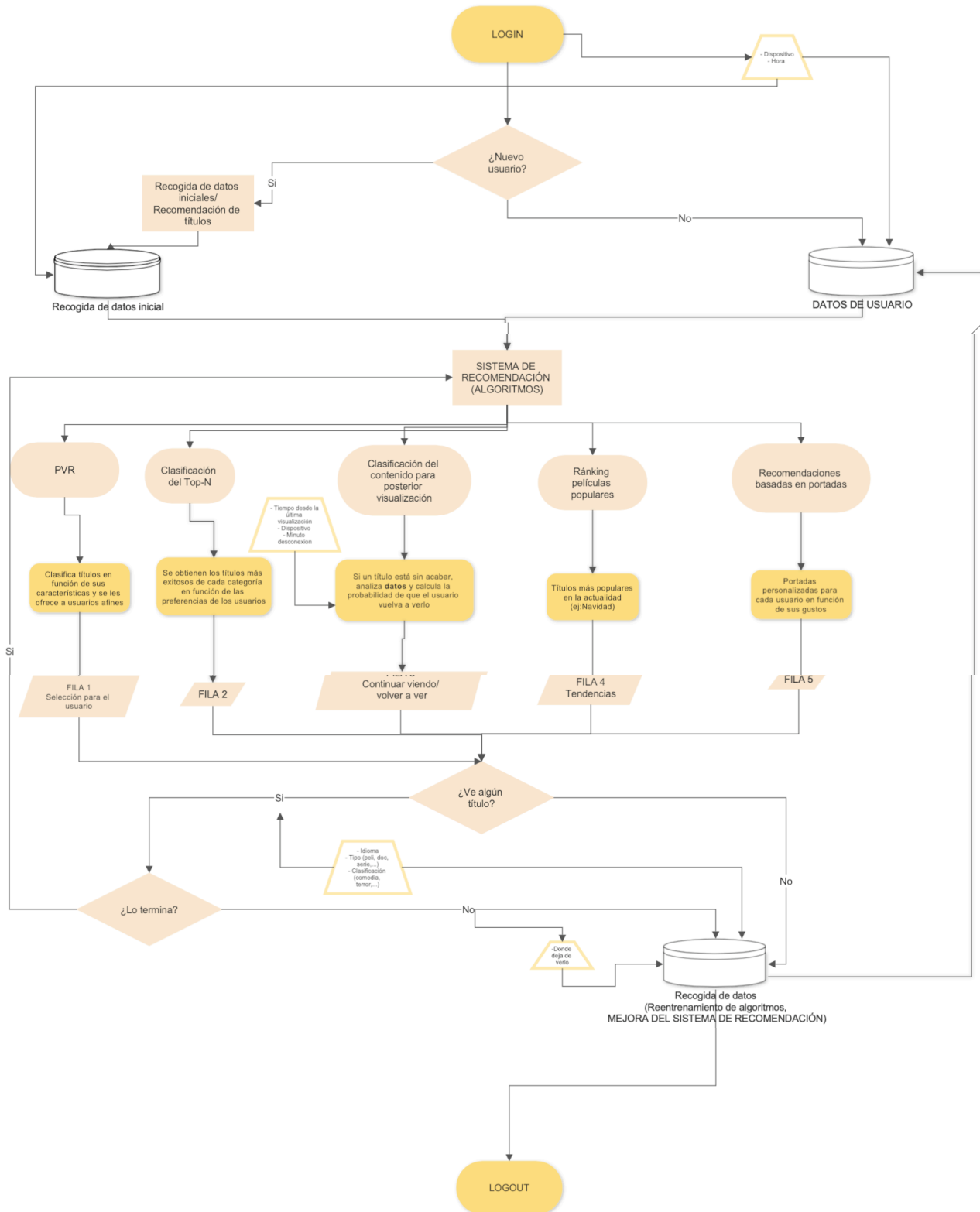
Todo lo que ocurre antes de pulsar el botón de play ocurre en el servidor backend y se ejecuta en el AWS y lo que sucede después de presionar play es gestionado por Open Connect, almacena los videos de Netflix en distintas partes del mundo y las reproduce en cada dispositivo.

La eficacia que se ha comentado anteriormente se consigue de la siguiente manera: Netflix usa AWS en tres regiones: Irlanda y dos en EEUU, obteniendo así una cobertura mundial. Dentro de cada región, Netflix opera en tres zonas de disponibilidad diferentes. La principal ventaja de tener las tres regiones es que si una de ellas falla, el resto se podrían ocupar perfectamente de cubrir a los usuarios de la región que se ha caído, es decir, si nosotros estamos haciendo uso de la plataforma en España, lo más probable es que estemos conectados a la región de Irlanda, pero si se cae la región de Irlanda, se nos redirigiría a una de las dos regiones de EEUU y la caída pasaría desapercibida para nosotros. Esto se conoce como *evacuar* una región y se puede hacer en un tiempo de 6 minutos.

Bibliografía:

- <https://recostream.com/blog/recommendation-system-netflix>
- <https://histografias.com/infografia-historia-netflix.html>
- <https://netflixtechblog.com/what-is-an-a-b-test-b08cc1b57962>
- <https://www.xataka.com/streaming/la-compleja-infraestructura-detras-de-netflix-que-pasa-cuando-le-das-al-play>
- <https://towardsdatascience.com/deep-dive-into-netflixs-recommender-system-341806ae3b48>

# ALGORITMO NETFLIX



MARTA VENEGAS PARDO