

PROPUESTA INICIAL PROYECTO FINAL ANS – GRUPO 13.

**Resumen**

El mercado de servicios de streaming ha experimentado un crecimiento exponencial en los últimos años, estableciéndose como una revolución del entretenimiento “personalizado” y ofreciendo a los consumidores una amplia variedad de opciones. Sin embargo, esta proliferación de plataformas presenta un desafío significativo para los usuarios: la "sobrecarga de opciones". Este fenómeno puede llevar a la parálisis en la toma de decisiones o a la elección de servicios que no se ajustan a las necesidades individuales. Este problema queda perfectamente esbozado en palabras del nobel de economía Richard Thaler *"Demasiadas opciones pueden paralizar a los consumidores, llevarlos a tomar decisiones ineficientes, o incluso hacer que eviten tomar decisiones por completo."*

Con este contexto, surge la necesidad de una herramienta que ayude a los consumidores a navegar por este complejo panorama de opciones de streaming. El objetivo principal de nuestro trabajo es desarrollar una guía integral y accesible que permita a los usuarios evaluar de manera precisa y fundamentada las principales plataformas de streaming en función de factores clave como el costo de suscripción, la calidad del contenido y la variedad de títulos disponibles. Para lograr esto, se recopilarán y analizarán datos de reseñas de usuarios, puntuaciones de películas y popularidad en sitios como IMDB.

Esta guía no solo ayudará a los consumidores a tomar decisiones informadas, sino que también proporcionará información valiosa a los proveedores de servicios de streaming, destacando áreas de mejora y oportunidades de crecimiento en un mercado altamente competitivo.

En resumen, nuestro trabajo pretende ser una herramienta esencial para maximizar el valor de la inversión de las personas en servicios de streaming, proporcionando una evaluación equilibrada y rigurosa de las opciones disponibles en el mercado.

**Introducción**

¿cómo determinar cuál servicio ofrece la mejor relación calidad-precio en función de las necesidades y preferencias individuales? La vasta oferta de contenido, combinada con distintas estructuras de precios y diversidad de títulos, puede resultar abrumadora para los usuarios al momento de tomar una decisión informada y es nuestro deseo poder ayudar a los millones de usuarios de estas plataformas en la toma de una decisión que los satisfaga.

El cliente potencial de esta propuesta son los millones de usuarios de plataformas de streaming que desean maximizar el valor de su suscripción, obteniendo el mejor balance entre costo, calidad y variedad de contenido. A nivel organizacional, las empresas que operan en este espacio también pueden beneficiarse de este análisis, ya que proporcionará información valiosa sobre cómo se perciben sus servicios en comparación con la competencia y qué áreas podrían mejorar para captar más usuarios.

El problema que buscamos resolver pertenece al área de análisis no supervisado (ANS) dentro del aprendizaje automático. Específicamente, utilizaremos técnicas de clustering y reducción de dimensionalidad para agrupar y analizar las plataformas de streaming según factores críticos. Estas técnicas esperamos nos permitan descubrir patrones ocultos y relaciones subyacentes entre las distintas plataformas, ofreciendo una evaluación objetiva y profunda que guiará tanto a consumidores como a proveedores en la toma de decisiones más informadas y estratégicas.

**Revisión preliminar de la literatura:**

La literatura existente sobre la evaluación de servicios de streaming se ha centrado en diversos aspectos, como la calidad del contenido, la experiencia del usuario y el costo de suscripción. Estudios nacionales e internacionales han utilizado métodos similares para abordar preguntas relacionadas con la relación calidad-precio de estos servicios.

Por ejemplo, investigaciones previas han empleado análisis de clústeres para agrupar servicios de streaming según características como la resolución de video, la variedad de contenido y la frecuencia de actualización de la biblioteca. Otros estudios han utilizado encuestas de satisfacción del usuario para evaluar la percepción de valor. Así mismo, para las plataformas el agrupado de características a contenidos también es fundamental en su sistema interno de recomendaciones, como la plataforma Netflix que en sus inicios utilizó un algoritmo de tipo SVD para la agrupación y recomendación de contenido.

Nuestro enfoque se diferencia al combinar tanto métricas objetivas como subjetivas en un análisis integral. Una medida subjetiva podría ser la popularidad de las películas. Mientras que algunos estudios se han centrado exclusivamente en aspectos técnicos o en la percepción del usuario, nuestra propuesta busca integrar ambos enfoques para proporcionar una evaluación más completa y equilibrada.

**Descripción de los datos:**

Los datos se obtienen de la base de título de la pagína IMDB y están estructurados de la siguiente manera:

* title: Título de la película o serie. Categórica nominal.
* platform: Plataforma digital donde se puede encontrar la película (por ejemplo, Netflix, Amazon Prime, etc.). Categórica nominal.
* type: Tipo de contenido (IMAGE, SHOW). Categórica nominal.
* imdbVotes: Número de votos que la película ha recibido en IMDB. Numérica discreta.
* imdbScore: Puntuación media de la película en IMDB. Numérica continua.
* tmdbPopularity: Puntuación de popularidad de la película en TMDB. Numérica continua.
* tmdbScore: Puntuación media de la película en TMDB. Numérica continua.
* runtime: Duración de la película en minutos. Numérica discreta.
* genres: Géneros a los que pertenece la película (puede haber más de uno). Categórica nominal
* director: Director de la película. Categórica nominal.
* year\_range: Rango de años en el que fue lanzada la película. Categórica ordinal.
* url: URL de la página de la película en su respectiva plataforma. Categórica nominal

Una descripción más amplia de estos se encuentra en el repositorio de github del proyecto en el documento EDA.ipynb les pedimos por favor remitirse a este dado lo extenso de este proceso.

**Propuesta metodológica:**

Para abordar la pregunta de interés, se planea utilizar el algoritmo de K-means clustering debido a su simplicidad y eficacia en la identificación de grupos homogéneos dentro de los datos. Este algoritmo es adecuado para agrupar los servicios de streaming en función de sus características y evaluar su relación calidad-precio. Otros algoritmos candidatos incluyen DBSCAN, que también podrían ser útiles para identificar patrones en los datos, sin embargo, a medida que avance el curso podríamos seguir probando modelos más sofisticados.

**Referencias:**

* [IMDB](https://www.imdb.com)
* [TMDB](https://www.themoviedb.org)
* [Netflix tech blog](https://netflixtechblog.com/)
* [Estudio Percepcion Consumidor Servicio Streaming](https://repositorio.usm.cl/server/api/core/bitstreams/d67c307d-1a49-4c90-b54d-335457c3938d/content)
* [Comparativa Streaming](https://www.xataka.com/basics/comparativa-disney-netflix-hbo-movistar-prime-video-apple-filmin-catalogo-funciones-precios)
* [Recomendaciones de Netflix](https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429)