Sistema Inteligente De **Análisis De Calificaciones De Películas Para Identificar Patrones De Preferencias De Los Espectadores**

Autor 1: Omar Alexis Castro Diaz

1005025450

Autor 2: Henry David Suarez Serrano

1004967702

Autor 3: Junnior Santiago Ravelo Velasco

1090369022

Proyecto Como Requisito Para Aprobar El Curso De Ciencia De Datos Y Fundamentos De Computación Paralela Y Distribuida

Universidad de Pamplona

Facultad de Ingenierías y Arquitectura

Ingeniería de Sistemas

2024

Villa del Rosario – Norte de Santander

# Índice

[Índice 2](#_Toc181764723)

[Introducción 5](#_Toc181764724)

[Capítulo I 8](#_Toc181764725)

[Planteamiento del Problema 8](#_Toc181764726)

[Justificación 10](#_Toc181764727)

[Objetivos 14](#_Toc181764728)

[Objetivo general 14](#_Toc181764729)

[Objetivos Específicos 14](#_Toc181764730)

[Acotaciones del Proyecto 15](#_Toc181764731)

[Capitulo II 17](#_Toc181764732)

[Antecedentes 17](#_Toc181764733)

[Estado del arte 19](#_Toc181764734)

[Marco referencial 20](#_Toc181764735)

[Marco teórico 26](#_Toc181764736)

[Marco conceptual 32](#_Toc181764737)

[Marco contextual 39](#_Toc181764738)

[Marco legal 44](#_Toc181764739)

[Capitulo III 51](#_Toc181764740)

[Metodología 51](#_Toc181764741)

[Fases de la Metodología 53](#_Toc181764742)

[Capitulo IV 65](#_Toc181764743)

[Análisis e interpretación de resultados 65](#_Toc181764744)

[Conclusiones objetivos 67](#_Toc181764745)

[Figura 1 69](#_Toc181764746)

[Figura 2 70](#_Toc181764747)

[Figura 3 72](#_Toc181764748)

[Figura 4 73](#_Toc181764749)

[Figura 5 74](#_Toc181764750)

[Conclusiones 75](#_Toc181764751)

[Bibliografia 78](#_Toc181764752)

# Introducción

El análisis de preferencias de contenido en plataformas de streaming es fundamental en la era digital para ofrecer experiencias de usuario más personalizadas y mejorar la retención de suscriptores. Sin embargo, los sistemas de recomendación actuales enfrentan desafíos al intentar capturar y predecir con precisión las preferencias individuales de los usuarios. Según Jiménez y Hernández (2021), “la implementación de recomendaciones personalizadas en tiempo real tiene el potencial de elevar la satisfacción del usuario en un 30%, generando un impacto positivo en la percepción del servicio” (p. 45). Con este contexto, el presente proyecto tiene como propósito desarrollar un sistema avanzado de análisis de calificaciones de películas, utilizando técnicas de minería de datos y computación paralela. Este sistema permitirá identificar patrones complejos en las preferencias de los usuarios, logrando recomendaciones de contenido más ajustadas y precisas.

Este proyecto busca no solo optimizar la experiencia de los usuarios, sino también incrementar la efectividad y precisión de los sistemas de recomendación en plataformas de contenido digital. Estudios recientes sugieren que el uso de técnicas avanzadas de procesamiento, como la computación paralela, reduce significativamente el tiempo de análisis y permite manejar grandes volúmenes de datos de manera eficiente, mejorando así la precisión de las recomendaciones (Martínez, 2022). Además, una mayor personalización en las recomendaciones tiene un impacto directo en el consumo de contenido, así como en la retención de suscriptores, factores críticos en la competitiva industria del streaming.

El problema que aborda este proyecto es la falta de precisión en los sistemas de recomendación actuales, que no logran reflejar con exactitud las preferencias reales de los usuarios, limitando la personalización del contenido.

El objetivo general de este proyecto es desarrollar un sistema de análisis y recolección de datos que, mediante técnicas de computación paralela y minería de datos, mejore la precisión de las recomendaciones de contenido en plataformas de streaming.

La metodología del proyecto sigue un enfoque basado en CRISP-DM (Cross-Industry Standard Process for Data Mining) adaptado a las necesidades específicas de análisis masivo de datos en tiempo real. Las fases incluyen la comprensión del negocio y de los datos, preparación de datos, modelado, evaluación y despliegue. En la fase de comprensión del negocio, se identifican los requerimientos específicos de los sistemas de recomendación y las características que influyen en la percepción de los usuarios. La preparación de los datos implica su limpieza y normalización para el análisis, mientras que el modelado y evaluación se enfocan en técnicas avanzadas de minería de datos para extraer patrones significativos de preferencia. Finalmente, el despliegue garantiza la integración del sistema en un entorno real para ofrecer recomendaciones precisas.

Los datos empleados en este proyecto provienen de fuentes confiables como IMDb y Rotten Tomatoes, lo que garantiza su validez y precisión, dada su actualización continua y relevancia en la industria cinematográfica. Estas plataformas son reconocidas en el ámbito del entretenimiento, asegurando una base de datos de alta calidad que es fundamental para lograr un análisis confiable.

El alcance del proyecto incluye el desarrollo de un prototipo funcional de sistema de recomendación, acompañado de un informe detallado que describe patrones de preferencias en el usuario. Este sistema utilizará técnicas de computación paralela para optimizar el procesamiento y ofrecer recomendaciones en tiempo real, basadas en los patrones descubiertos a través de la metodología planteada.

# Capítulo I

# **Planteamiento del Problema**

El análisis de datos de preferencias en plataformas de streaming representa un reto crucial en la era digital, ya que permite identificar patrones de consumo y mejorar la experiencia de usuario mediante recomendaciones personalizadas. Este capítulo busca explorar la importancia de desarrollar un sistema de recomendaciones optimizado, analizando tanto la situación actual de los sistemas existentes como las limitaciones que impiden lograr una personalización precisa y relevante para cada usuario. Al mejorar estos sistemas, se lograría no solo una mayor satisfacción del cliente, sino también un impacto positivo en la retención de suscriptores, una métrica fundamental para el crecimiento en la industria de entretenimiento digital.

Actualmente, los sistemas de recomendación enfrentan varios desafíos que afectan su capacidad para reflejar adecuadamente los intereses de cada usuario. Diversos estudios indican que una parte significativa de los usuarios considera que las recomendaciones recibidas no logran capturar sus gustos personales, lo cual disminuye su experiencia en la plataforma (Jiménez y López, 2021). Según un estudio realizado por Méndez y García (2022), el 68% de los suscriptores de plataformas de streaming señalan la falta de precisión en las recomendaciones como uno de los principales factores que disminuye su satisfacción con el servicio. A pesar de los avances en algoritmos de personalización, las técnicas actuales carecen de la capacidad para manejar el análisis masivo de datos de manera eficiente, lo que limita el potencial de estos sistemas para ofrecer recomendaciones de alta precisión.

Esta situación problemática se observa en el hecho de que, a medida que los catálogos de contenido crecen, los sistemas de recomendación no logran adaptarse al ritmo de crecimiento en datos y diversidad de preferencias, lo cual impacta negativamente en la experiencia del usuario. Los métodos convencionales de análisis, como los sistemas basados en filtrado colaborativo, no pueden procesar con rapidez las grandes cantidades de datos generados por cada usuario, lo cual deriva en recomendaciones que resultan genéricas o poco relevantes. Esta falta de personalización en las sugerencias de contenido causa frustración en los usuarios y representa una pérdida de oportunidad para las plataformas, al no aprovechar al máximo el potencial de engagement de sus suscriptores.

En un escenario ideal, el sistema de recomendación debería ser capaz de procesar en tiempo real grandes volúmenes de datos, integrando información detallada sobre los patrones de consumo y gustos específicos de cada usuario. Esto permitiría generar recomendaciones de contenido más precisas, adaptadas a cada perfil de usuario y en sintonía con sus preferencias actuales. Alcanzar esta precisión en las recomendaciones no solo mejoraría la satisfacción de los suscriptores, sino que también reforzaría la competitividad de la plataforma en el mercado digital.

Por lo tanto, la formulación del problema se centra en la siguiente pregunta: ¿Cómo se puede mejorar la precisión de los sistemas de recomendación en plataformas de streaming mediante el uso de técnicas de computación paralela y minería de datos, de modo que se logre una personalización más adecuada y relevante para cada usuario?

Con esta investigación, se espera no solo abordar la necesidad de una personalización más efectiva, sino también destacar el valor de implementar tecnologías avanzadas que optimicen el rendimiento de los sistemas de recomendación. Según estudios recientes, un sistema de recomendación más preciso puede incrementar la satisfacción del usuario y potenciar el crecimiento de las plataformas de contenido en un mercado cada vez más competitivo (Martínez, 2022).

# Justificación

El análisis de preferencias de contenido y la optimización de sistemas de recomendación se han vuelto temas de gran relevancia en el contexto de la transformación digital. Con el crecimiento exponencial de las plataformas de streaming, surge la necesidad de mejorar la precisión de las recomendaciones para adaptarse mejor a los intereses específicos de cada usuario. Este proyecto busca contribuir a la solución de esta problemática mediante la implementación de un sistema de recomendación que utiliza técnicas avanzadas de computación paralela y minería de datos, lo cual permite una personalización más efectiva. El desarrollo de este tema es necesario para abordar las limitaciones de los sistemas actuales y ofrecer una experiencia más satisfactoria al usuario final, fomentando a la vez la innovación tecnológica en la industria del entretenimiento digital.

Desde el punto de vista social, este proyecto responde a la demanda de los usuarios de plataformas de streaming de recibir recomendaciones personalizadas y relevantes. Una experiencia más adaptada a los gustos personales mejora la satisfacción y reduce la frustración generada por sugerencias inadecuadas, lo cual es fundamental en un entorno donde las opciones de contenido pueden resultar abrumadoras. Además, este sistema permite a los usuarios descubrir contenido diverso que, sin una recomendación precisa, podría pasar desapercibido. De esta forma, el proyecto no solo incrementa la satisfacción individual del usuario, sino que también promueve el acceso a una oferta cultural más rica y variada, apoyando la diversidad en el consumo de contenido digital.

Académicamente, el proyecto contribuye a la investigación en ciencia de datos y técnicas de aprendizaje automático aplicadas al análisis de preferencias de usuario. El uso de computación paralela y minería de datos en sistemas de recomendación representa un enfoque innovador que busca superar los límites de precisión y escalabilidad presentes en los métodos convencionales. Al documentar y evaluar la efectividad de estas técnicas, el proyecto se alinea con las tendencias actuales en el campo de la inteligencia artificial y su aplicación en problemas de personalización. Este desarrollo académico tiene el potencial de abrir nuevas líneas de investigación sobre cómo mejorar la interacción usuario-sistema y cómo aplicar tecnologías avanzadas para abordar problemas complejos de análisis de grandes volúmenes de datos.

Desde el punto de vista metodológico, el proyecto adopta la metodología CRISP-DM, reconocida en el ámbito de la ciencia de datos por su estructura clara y su enfoque sistemático. Esta metodología permite abordar el análisis de datos y el desarrollo de modelos de recomendación en fases bien definidas, lo cual facilita la replicación y evaluación del proyecto. La integración de técnicas de computación paralela en el modelo de recomendación no solo mejora el procesamiento en tiempo real, sino que también añade rigor al enfoque metodológico, permitiendo una mayor confiabilidad en los resultados obtenidos. La metodología empleada en este proyecto garantiza un proceso de desarrollo robusto y detallado, alineado con los estándares académicos y prácticos en ciencia de datos.

Desde el punto de vista práctico, el desarrollo de un sistema de recomendación más preciso tiene un impacto directo en la competitividad de las plataformas de streaming, mejorando la retención de usuarios y optimizando los recursos invertidos en marketing y recomendaciones. La aplicación práctica de este proyecto radica en la creación de un prototipo funcional que pueda integrarse fácilmente en sistemas de recomendación comerciales. Este sistema permitirá a las plataformas de streaming ofrecer una experiencia de usuario mejorada y más adaptada a las necesidades individuales, facilitando la retención y fidelización de los usuarios en un mercado altamente competitivo. La implementación de este sistema, además, representa una oportunidad para reducir el tiempo de procesamiento en el análisis de preferencias, aumentando la eficiencia operativa de las plataformas de contenido digital.

Desde el punto de vista de la línea de investigación, este proyecto se inscribe en el ámbito de la inteligencia artificial aplicada y el aprendizaje automático, con un enfoque en la personalización y mejora de sistemas de recomendación. La investigación en modelos de recomendación adaptativa y su implementación en plataformas de contenido digital es una línea de estudio activa, que busca desarrollar soluciones innovadoras para optimizar la interacción entre usuarios y servicios digitales. Este proyecto no solo aporta a esta línea de investigación mediante la exploración de técnicas avanzadas de computación, sino que también contribuye al desarrollo de tecnologías que incrementan la accesibilidad y el impacto social de los sistemas de recomendación en la vida cotidiana de los usuarios.

# Objetivos

## **Objetivo general**

Desarrollar un sistema de recomendación de contenido en plataformas de streaming mediante técnicas de computación paralela y minería de datos, siguiendo la metodología CRISP-DM, para mejorar la precisión y personalización de las recomendaciones de acuerdo con los patrones de preferencias de los usuarios.

## Objetivos Específicos

**Investigar** la literatura actual y estudios sobre sistemas de recomendación y técnicas de minería de datos en el contexto del análisis de preferencias de usuario, **para** obtener una comprensión sólida sobre el objeto de estudio y los métodos utilizados en esta área.

**Seleccionar** herramientas de programación paralela, algoritmos de minería de datos y técnicas de preprocesamiento de datos adecuados, **para** realizar un análisis eficaz de las preferencias de usuario que respalde el desarrollo del sistema de recomendación.

**Desarrollar** el sistema de recomendación de contenido mediante el desarrollo e implementación de un modelo basado en computación paralela y técnicas avanzadas de minería de datos, adaptando el sistema a los datos y necesidades específicos de una plataforma de streaming.

**Validar** el sistema de recomendación mediante pruebas de precisión, recuperación y satisfacción del usuario, **para** asegurar que el modelo desarrollado cumple con los estándares de personalización y eficacia esperados en las plataformas de contenido digital.

## Acotaciones del Proyecto

**Alcance del Software:** El desarrollo se limitará a un prototipo de sistema de recomendación que no se implementará en un entorno comercial en esta fase, sino que se enfocará en la validación de la precisión y efectividad del modelo en un entorno de pruebas controlado.

**Limitaciones del Hardware:** La infraestructura necesaria para este proyecto incluirá acceso a servidores con capacidades de procesamiento paralelizado, preferiblemente mediante GPU, para permitir el entrenamiento y evaluación de modelos a gran escala.

**Alcance de los Objetivos:** Este proyecto se enfocará en mejorar la precisión y relevancia de las recomendaciones en una plataforma de contenido de entretenimiento, sin explorar otros tipos de recomendación como productos o servicios fuera de este contexto.

**Datos:** Los datos empleados serán obtenidos de bases públicas como IMDb y Rotten Tomatoes, que ofrecen reseñas y calificaciones de películas. Estos datos serán limitados a aquellos disponibles en dichas plataformas, sin incluir datos adicionales o personales de usuarios.

**Contexto del Proyecto:** Este sistema de recomendación está diseñado para el contexto de plataformas de streaming de contenido audiovisual y se evaluará en función de la precisión y relevancia en la personalización de recomendaciones de películas o series.

# Capitulo II

## Antecedentes

El análisis de preferencias en plataformas de contenido digital ha evolucionado significativamente con el avance de la inteligencia artificial y el aprendizaje automático. En los últimos años, las plataformas de streaming han adoptado sistemas de recomendación para personalizar la experiencia del usuario, con el fin de mejorar la satisfacción y la retención de suscriptores. Según Torres y Ramírez (2021), “los sistemas de recomendación han demostrado ser una herramienta esencial para optimizar el consumo de contenido, aumentando el tiempo de interacción del usuario y mejorando la retención de suscriptores en las plataformas” (p. 37). Esto ha llevado a la investigación de algoritmos avanzados y técnicas de minería de datos capaces de analizar grandes volúmenes de datos y ofrecer recomendaciones más precisas y relevantes.

Los primeros sistemas de recomendación en el ámbito digital utilizaban técnicas de filtrado colaborativo, que consistían en analizar patrones de comportamiento de grupos de usuarios similares para generar sugerencias. Sin embargo, esta técnica ha mostrado limitaciones en términos de precisión y escalabilidad al enfrentarse a grandes volúmenes de datos y diversidad en las preferencias de los usuarios (Chen et al., 2020). Estos sistemas no logran capturar la singularidad de cada usuario, lo que limita la efectividad de la personalización. Para enfrentar estos desafíos, investigaciones recientes han explorado el uso de algoritmos de aprendizaje profundo y procesamiento paralelo, que permiten el análisis de datos en tiempo real y adaptan las recomendaciones a cada individuo con mayor precisión.

El desarrollo de técnicas de computación paralela ha sido clave para mejorar el rendimiento de los sistemas de recomendación. Estas técnicas facilitan el procesamiento simultáneo de grandes volúmenes de datos, reduciendo el tiempo necesario para analizar preferencias de usuarios en plataformas de alto tráfico. Estudios como el de Zhang y Liu (2022) demuestran que la integración de GPU y programación en paralelo optimiza el rendimiento de los algoritmos de recomendación, permitiendo analizar millones de registros en segundos y ofreciendo recomendaciones en tiempo real. Este enfoque es especialmente relevante en plataformas de streaming, donde el consumo de contenido es dinámico y requiere adaptaciones constantes para satisfacer las preferencias cambiantes de los usuarios.

Adicionalmente, el uso de técnicas de minería de datos avanzadas, como el clustering y la segmentación, ha permitido una mejora sustancial en los sistemas de recomendación al identificar patrones de consumo en grandes conjuntos de datos. Investigaciones recientes destacan que estos métodos pueden agrupar a los usuarios según características similares de consumo, lo cual facilita una personalización más precisa (García y Méndez, 2021). Estas técnicas no solo permiten identificar preferencias de contenido, sino también patrones temporales de consumo, como horarios preferidos o géneros populares en determinadas temporadas, lo que amplía el potencial de los sistemas de recomendación para adaptarse al comportamiento del usuario.

En el contexto de las plataformas de contenido digital, la integración de técnicas de computación paralela y minería de datos representa un avance importante hacia la personalización adaptativa de las recomendaciones. Estas innovaciones buscan superar las limitaciones de los modelos tradicionales, ofreciendo recomendaciones más acertadas y en tiempo real. Este proyecto, al adoptar estas técnicas, se sitúa en una línea de investigación activa y en constante evolución, con el objetivo de optimizar la experiencia del usuario en plataformas de streaming y contribuir a la efectividad de los sistemas de recomendación.

## Estado del arte

Este capítulo presenta un análisis detallado de los estudios y desarrollos más relevantes en el ámbito de los sistemas de recomendación en plataformas de contenido digital. Con el avance de la inteligencia artificial y el aprendizaje automático, se ha incrementado la investigación en técnicas que optimicen la personalización de recomendaciones para mejorar la experiencia de usuario. A través de este estado del arte, se abordan los conceptos clave y los enfoques metodológicos que sustentan el desarrollo de sistemas de recomendación, así como el marco legal que regula la gestión de datos personales en plataformas de contenido.

## Marco referencial

El marco referencial de este proyecto examina los avances y enfoques recientes en el desarrollo de sistemas de recomendación personalizados en plataformas de contenido digital, especialmente en contextos de alto tráfico como el streaming de video y música. La investigación actual en este campo se enfoca en técnicas de procesamiento de datos masivos y computación paralela para optimizar la experiencia del usuario y mejorar la precisión de las recomendaciones. Este análisis incluye estudios y aplicaciones sobre la minería de datos, computación en GPU, y algoritmos de aprendizaje profundo, que han transformado significativamente el diseño de estos sistemas en los últimos años.

Los sistemas de recomendación surgieron como herramientas para mejorar la experiencia de los usuarios en plataformas digitales, utilizando inicialmente métodos de filtrado colaborativo y de contenido. Según Torres y Ramírez (2021), los primeros sistemas de recomendación empleaban **filtrado colaborativo**, una técnica basada en el análisis de patrones de comportamiento de usuarios con intereses similares. Este método, aunque efectivo, presenta limitaciones al enfrentarse a grandes volúmenes de datos y a la diversidad de preferencias en las plataformas de contenido. Jiménez y López (2021) destacan que las recomendaciones basadas en filtrado colaborativo tienden a homogeneizar las sugerencias, limitando la personalización que los usuarios esperan en la actualidad.

Para superar estas limitaciones, estudios recientes han introducido técnicas de aprendizaje profundo y minería de datos, que permiten una personalización más detallada y precisa. Por ejemplo, las redes neuronales convolucionales (CNN) y los modelos de procesamiento de lenguaje natural (NLP) han demostrado ser eficaces en la extracción de características complejas y específicas de los usuarios, proporcionando una comprensión más rica de sus preferencias (Chen et al., 2020). Estas técnicas permiten no solo recomendaciones más personalizadas, sino también adaptativas, al ajustar las sugerencias en tiempo real según los patrones de consumo detectados en la plataforma.

La necesidad de procesar grandes volúmenes de datos en tiempo real ha impulsado el desarrollo de tecnologías de **computación paralela**, especialmente mediante el uso de unidades de procesamiento gráfico (GPU). Zhang y Liu (2022) argumentan que el procesamiento en GPU permite una paralelización efectiva de las tareas de análisis, reduciendo significativamente el tiempo de procesamiento y mejorando la capacidad de respuesta de los sistemas de recomendación. Esta tecnología ha sido implementada exitosamente en plataformas de alto tráfico, donde la capacidad para gestionar datos masivos es crucial para ofrecer recomendaciones en tiempo real.

El uso de computación paralela no solo mejora la eficiencia de los sistemas de recomendación, sino que también permite una actualización continua de los modelos de aprendizaje. Esto significa que las recomendaciones se adaptan a los cambios en el comportamiento del usuario de manera inmediata, aumentando la precisión y relevancia de las sugerencias. Méndez y García (2022) afirman que la computación paralela ha revolucionado la industria de recomendaciones, permitiendo un nivel de personalización que era inalcanzable con técnicas convencionales.

La **minería de datos** se ha convertido en una técnica esencial para identificar patrones de consumo en plataformas de contenido digital, permitiendo un análisis detallado de los hábitos y preferencias de los usuarios. García y Méndez (2021) explican que la minería de datos permite extraer información significativa de grandes volúmenes de datos, lo cual es fundamental para comprender los comportamientos de consumo y las preferencias de cada usuario en plataformas de streaming. Estas técnicas incluyen métodos de clustering y segmentación, que permiten agrupar a los usuarios según características similares de consumo.

Además, la minería de datos aplicada en tiempo real facilita la identificación de patrones temporales, como los horarios de visualización preferidos o los géneros de contenido más populares en determinadas épocas del año. Esto no solo mejora la personalización de las recomendaciones, sino que también permite anticiparse a las necesidades y gustos de los usuarios. Torres y Ramírez (2021) señalan que la identificación de patrones de consumo específicos permite a las plataformas adaptar su contenido de forma proactiva, aumentando la satisfacción del usuario y, por ende, su fidelidad a la plataforma.

El **aprendizaje profundo** ha permitido un avance significativo en la personalización de sistemas de recomendación, gracias a su capacidad para modelar relaciones complejas y no lineales en grandes conjuntos de datos. Las redes neuronales, en particular, han mostrado ser eficaces en el análisis de preferencias de usuario al permitir un aprendizaje continuo y adaptativo. Chen et al. (2020) destacan que las redes neuronales convolucionales (CNN) y los modelos basados en arquitecturas de aprendizaje profundo ofrecen una precisión superior en las recomendaciones, al modelar no solo el comportamiento de consumo, sino también los contextos y circunstancias específicas de cada usuario.

Además, las técnicas de procesamiento de lenguaje natural (NLP) han facilitado la integración de comentarios y reseñas de usuarios en los sistemas de recomendación, permitiendo capturar la subjetividad y emociones que estos expresan en sus valoraciones de contenido. Esto enriquece las recomendaciones al añadir una capa de análisis semántico, que ayuda a ofrecer sugerencias de contenido más acordes a las expectativas y gustos del usuario. Este enfoque adaptativo es fundamental en la industria del streaming, donde la diversidad de contenido y la rapidez de cambio en las preferencias exigen sistemas de recomendación dinámicos.

Los sistemas de recomendación no solo mejoran la experiencia del usuario, sino que también tienen un impacto directo en el éxito y la competitividad de las plataformas de streaming. Según un informe de Méndez y García (2022), hasta el 80% del contenido visto en plataformas como Netflix y Spotify proviene de recomendaciones personalizadas, lo que demuestra su relevancia en el comportamiento de consumo de los usuarios. Además, la personalización de las recomendaciones aumenta la fidelidad del usuario y reduce la tasa de cancelación de suscripciones, lo cual es crítico en un mercado altamente competitivo.

Torres y Ramírez (2021) afirman que, en la actualidad, las plataformas que no implementan sistemas de recomendación efectivos corren el riesgo de perder relevancia en un entorno donde los usuarios buscan experiencias de contenido personalizadas y adaptadas a sus preferencias. Este impacto se observa también en los ingresos de estas plataformas, ya que la optimización en la personalización de contenido contribuye a aumentar el tiempo de visualización y, por ende, los ingresos generados por usuario. Así, los sistemas de recomendación se posicionan como un elemento clave en la estrategia de las plataformas de contenido digital.

Aunque el desarrollo de sistemas de recomendación ha avanzado considerablemente, persisten ciertos retos en su implementación, especialmente en relación con la privacidad de los datos y la capacidad de escalar estos sistemas en tiempo real. La privacidad es un tema particularmente relevante en este contexto, ya que la recolección y análisis de datos personales pueden generar preocupaciones entre los usuarios. Torres y Ramírez (2021) señalan que las plataformas deben asegurar que la recolección de datos cumpla con las normativas de protección de privacidad, como el Reglamento General de Protección de Datos (GDPR) en Europa, para mantener la confianza del usuario en la plataforma.

Además, la capacidad de escalar los sistemas de recomendación a nivel global presentan desafíos técnicos importantes. La adopción de tecnologías como la computación en la nube y la optimización en GPU ofrece soluciones para manejar el procesamiento de datos a gran escala, pero requiere una inversión significativa en infraestructura tecnológica. Según Zhang y Liu (2022), la computación en la nube es una opción viable para expandir la capacidad de los sistemas de recomendación, ya que permite procesar grandes volúmenes de datos de manera distribuida, mejorando la escalabilidad y flexibilidad del sistema.

## Marco teórico

**Fundamentos de los Sistemas de Recomendación**

Los sistemas de recomendación son una categoría de sistemas de filtrado de información que buscan sugerir elementos (productos, servicios o contenido) que puedan ser de interés para los usuarios. Inicialmente, estos sistemas se desarrollaron para aplicaciones de comercio electrónico, pero han evolucionado para incluir sectores como entretenimiento, redes sociales y servicios de streaming (Jiménez y López, 2021). Existen tres enfoques principales en los sistemas de recomendación:

* **Filtrado Colaborativo:** Este método se basa en el comportamiento histórico de los usuarios, identificando patrones de preferencias entre usuarios similares. Es uno de los enfoques más antiguos y ampliamente utilizados en plataformas digitales. Sin embargo, el filtrado colaborativo presenta limitaciones en cuanto a la escalabilidad y manejo de grandes volúmenes de datos, lo que dificulta su aplicación en plataformas de contenido masivo (Torres y Ramírez, 2021).
* **Filtrado Basado en Contenido:** Este enfoque analiza las características del contenido para hacer recomendaciones a los usuarios. Por ejemplo, en plataformas de streaming, el sistema puede sugerir películas con géneros o directores similares a los que el usuario ha visualizado previamente. Aunque este método ofrece recomendaciones más personalizadas, puede llevar a la repetición de patrones y reducir la diversidad en las sugerencias (Méndez y García, 2022).
* **Enfoques Híbridos:** Los sistemas de recomendación actuales tienden a combinar el filtrado colaborativo y el basado en contenido para optimizar la precisión y relevancia de las recomendaciones. Los enfoques híbridos logran superar algunas de las limitaciones de cada método individual, permitiendo una mayor personalización y mejor adaptación a los gustos cambiantes del usuario (Chen et al., 2020).

**Aprendizaje Automático en Sistemas de Recomendación**

El aprendizaje automático, o machine learning, ha transformado los sistemas de recomendación al introducir algoritmos que permiten mejorar la precisión de las recomendaciones mediante el análisis de patrones complejos en los datos. Dentro de los algoritmos de aprendizaje automático, existen técnicas de supervisión y no supervisión que permiten optimizar los sistemas de recomendación. Estos modelos se entrenan con grandes volúmenes de datos para identificar patrones y relaciones que puedan predecir los intereses futuros de los usuarios (García y Méndez, 2021).

* **Algoritmos de Aprendizaje Supervisado:** Este tipo de aprendizaje se utiliza en la clasificación y predicción de preferencias del usuario, empleando datos etiquetados como referencia. Algoritmos como el **Árbol de Decisión**, **Bosque Aleatorio** y **Máquinas de Soporte Vectorial** son ejemplos comunes en la recomendación de contenido personalizado (Chen et al., 2020).
* **Aprendizaje Profundo y Redes Neuronales:** Las redes neuronales profundas, especialmente las **redes neuronales convolucionales (CNN)** y las **redes neuronales recurrentes (RNN)**, se utilizan para analizar patrones de consumo en tiempo real y modelar las relaciones no lineales entre las preferencias de los usuarios y el contenido disponible. Según Zhang y Liu (2022), las CNN han demostrado ser eficaces en la personalización de recomendaciones, ya que logran capturar relaciones complejas en los datos, proporcionando una personalización más precisa.
* **Procesamiento de Lenguaje Natural (NLP):** En los sistemas de recomendación, el NLP se utiliza para analizar las reseñas, comentarios y calificaciones que los usuarios dan al contenido. Esto permite capturar la subjetividad y emociones de los usuarios, enriqueciendo las recomendaciones al considerar tanto las preferencias objetivas como las experiencias expresadas en lenguaje natural (Jiménez y López, 2021).

**Minería de Datos y Identificación de Patrones**

La **minería de datos** es una técnica clave en la construcción de sistemas de recomendación, ya que permite la extracción de patrones significativos y relaciones dentro de grandes conjuntos de datos. Este enfoque facilita el análisis de preferencias de los usuarios en plataformas de contenido digital y ayuda a predecir el contenido que será de mayor interés para cada perfil de usuario (García y Méndez, 2021).

* **Segmentación de Usuarios:** A través de técnicas como el **clustering** y **análisis de agrupamiento**, los sistemas de recomendación pueden clasificar a los usuarios en grupos basados en similitudes de comportamiento. Esta segmentación permite hacer recomendaciones más personalizadas y relevantes, ya que el sistema puede ofrecer contenido que ha sido popular dentro de grupos similares (Chen et al., 2020).
* **Análisis Predictivo:** Las técnicas de minería de datos permiten realizar predicciones sobre el comportamiento de los usuarios, como la probabilidad de visualizar cierto tipo de contenido en un momento específico. Este análisis predictivo es esencial para los sistemas de recomendación en plataformas de streaming, ya que facilita la anticipación de necesidades y preferencias de los usuarios, mejorando así la relevancia de las recomendaciones (Torres y Ramírez, 2021).

**Computación Paralela y Procesamiento en Tiempo Real**

La **computación paralela** ha sido crucial para manejar el volumen de datos generado en plataformas de contenido digital y para procesar recomendaciones en tiempo real. Los sistemas de recomendación actuales requieren procesamiento masivo de datos para analizar las preferencias de millones de usuarios simultáneamente, tarea que sería imposible sin la capacidad de paralelizar los cálculos y distribuirlos en múltiples unidades de procesamiento (Zhang y Liu, 2022).

* **Procesamiento en GPU:** La utilización de GPU (Unidades de Procesamiento Gráfico) permite acelerar el análisis de datos masivos mediante la ejecución de múltiples operaciones en paralelo. Esto ha permitido que los sistemas de recomendación puedan adaptarse en tiempo real a los cambios en el comportamiento del usuario, mejorando la eficiencia y precisión de las recomendaciones (Chen et al., 2020).
* **Optimización de Algoritmos en Tiempo Real:** La computación paralela también facilita la optimización continua de los algoritmos de recomendación, ajustando los parámetros en función del rendimiento y adaptándose a la carga de datos sin afectar la experiencia del usuario. Esto es esencial para las plataformas de alto tráfico, donde la capacidad de procesar y actualizar recomendaciones en tiempo real es una ventaja competitiva (Torres y Ramírez, 2021).

**Teoría de Redes Neuronales y Modelos Adaptativos**

Las redes neuronales, especialmente las redes neuronales profundas, han revolucionado la capacidad de los sistemas de recomendación para analizar grandes volúmenes de datos y modelar patrones complejos de preferencias del usuario. La teoría de redes neuronales sostiene que los modelos de aprendizaje profundo, al simular la estructura y funcionamiento del cerebro humano, pueden identificar relaciones no obvias en los datos, lo que permite una personalización más sofisticada (Jiménez y López, 2021).

* **Redes Neuronales Convolucionales (CNN):** Las CNN han mostrado ser altamente efectivas para el análisis de imágenes y patrones en sistemas de recomendación, al capturar características complejas y específicas de las preferencias de los usuarios (García y Méndez, 2021).
* **Redes Neuronales Recurrentes (RNN):** Las RNN se utilizan en sistemas de recomendación para modelar secuencias y patrones temporales en los datos, como las preferencias de contenido según horarios o tendencias estacionales. Este enfoque adaptativo permite que el sistema de recomendación se ajuste en función de patrones cambiantes y contextuales (Zhang y Liu, 2022).

**Teoría de la Personalización y Relevancia de Contenido**

La teoría de la personalización sostiene que los sistemas de recomendación pueden optimizar la satisfacción del usuario al ofrecer contenido que se alinea con sus preferencias individuales y contextuales. La relevancia de contenido es un concepto clave en esta teoría, donde se asume que el usuario valorará más un sistema que anticipe y se adapte a sus intereses (Méndez y García, 2022).

* **Relevancia de Contenido:** Este concepto implica que el sistema debe seleccionar contenido que no solo sea interesante, sino también apropiado al contexto y preferencias del usuario en un momento dado. En el caso de los sistemas de recomendación en plataformas de contenido, esta relevancia se logra mediante el análisis de patrones de consumo específicos y la capacidad de adaptar las recomendaciones en tiempo real (Torres y Ramírez, 2021).
* **Modelos Adaptativos:** Los sistemas de recomendación adaptativos emplean algoritmos que pueden modificar sus recomendaciones dinámicamente, ajustándose a los cambios en el comportamiento del usuario. Estos modelos son fundamentales en el contexto de plataformas de streaming, donde el consumo de contenido es dinámico y altamente variable (Jiménez y López, 2021).

## Marco conceptual

**Sistema de Recomendación**

Un sistema de recomendación es un tipo de software que sugiere contenido, productos o servicios a los usuarios basado en el análisis de sus preferencias y comportamiento. Estos sistemas operan bajo el principio de identificar patrones en los datos de usuario para ofrecer sugerencias personalizadas, lo cual incrementa la relevancia y satisfacción del contenido presentado (Jiménez y López, 2021). Existen diferentes enfoques para la recomendación de contenido, entre los que destacan el **filtrado colaborativo**, el **filtrado basado en contenido** y los **modelos híbridos**, cada uno con sus propias técnicas y aplicaciones.

**Filtrado Colaborativo**

El **filtrado colaborativo** es una técnica de recomendación basada en la suposición de que los usuarios que compartieron preferencias en el pasado, probablemente mantendrán preferencias similares en el futuro. Este método utiliza las valoraciones y comportamientos de usuarios similares para hacer predicciones de contenido, permitiendo identificar patrones de afinidad. Aunque eficaz, el filtrado colaborativo presenta limitaciones como la incapacidad de recomendar contenido nuevo y problemas de escalabilidad en grandes conjuntos de datos (Torres y Ramírez, 2021).

**Filtrado Basado en Contenido**

El **filtrado basado en contenido** se centra en el análisis de las características de los elementos a recomendar. En lugar de depender del comportamiento de otros usuarios, este enfoque utiliza las preferencias individuales del usuario para sugerir contenido con características similares al previamente consumido. Esta técnica es útil para generar recomendaciones personalizadas y evita problemas como el “cold start” (inicio en frío), pero puede limitar la diversidad de sugerencias, al basarse estrictamente en patrones de contenido consumido (Méndez y García, 2022).

**Enfoques Híbridos**

Los **enfoques híbridos** combinan características de los métodos de filtrado colaborativo y basado en contenido, con el objetivo de mejorar la precisión y relevancia de las recomendaciones. Al integrar ambos enfoques, los sistemas híbridos superan limitaciones de cada técnica individual, permitiendo recomendaciones más robustas y adaptativas. Este tipo de enfoque es común en plataformas de streaming como Netflix y YouTube, donde se busca ofrecer recomendaciones personalizadas a gran escala (Chen et al., 2020).

**Aprendizaje Automático (Machine Learning)**

El **aprendizaje automático** es una rama de la inteligencia artificial que permite a los sistemas de recomendación “aprender” de los datos de usuario y mejorar sus predicciones con el tiempo. A través de algoritmos de aprendizaje supervisado y no supervisado, los sistemas de recomendación pueden identificar patrones complejos y relaciones en los datos que ayudan a mejorar la precisión de las recomendaciones. El aprendizaje automático incluye técnicas avanzadas como redes neuronales y algoritmos de clustering, fundamentales en los sistemas de recomendación modernos (García y Méndez, 2021).

**Redes Neuronales**

Las **redes neuronales** son modelos de aprendizaje profundo que simulan el funcionamiento de las neuronas en el cerebro humano. Estas redes son especialmente útiles en el análisis de datos complejos, como los patrones de consumo en plataformas de contenido. Dentro de los sistemas de recomendación, las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN) se utilizan para capturar patrones de preferencias en tiempo real y realizar recomendaciones adaptativas (Zhang y Liu, 2022).

**Minería de Datos**

La **minería de datos** es el proceso de analizar grandes conjuntos de datos para extraer patrones, relaciones y conocimientos útiles. En el contexto de sistemas de recomendación, la minería de datos permite identificar preferencias y comportamientos de los usuarios, facilitando la personalización de contenido. Técnicas como el clustering, la segmentación y el análisis predictivo son fundamentales en la minería de datos para mejorar la precisión y relevancia de las recomendaciones en plataformas digitales (Torres y Ramírez, 2021).

**Computación Paralela**

La **computación paralela** se refiere a la capacidad de ejecutar múltiples operaciones de procesamiento al mismo tiempo, optimizando el manejo de grandes volúmenes de datos. En sistemas de recomendación, el uso de procesadores gráficos (GPU) permite procesar datos de usuario a gran escala de manera rápida y eficiente, lo que es esencial para plataformas de contenido de alto tráfico que requieren recomendaciones en tiempo real (Chen et al., 2020). Este concepto es esencial para mejorar la escalabilidad y capacidad de respuesta de los sistemas de recomendación.

**Procesamiento de Lenguaje Natural (NLP)**

El **procesamiento de lenguaje natural** es una subárea de la inteligencia artificial que permite a los sistemas analizar y entender el lenguaje humano. En los sistemas de recomendación, el NLP se usa para interpretar y clasificar comentarios y reseñas de usuarios, capturando las emociones y subjetividades expresadas. Esto enriquece las recomendaciones al considerar tanto las preferencias objetivas como las experiencias y opiniones de los usuarios sobre el contenido consumido (Jiménez y López, 2021).

**Cold Start (Inicio en Frío)**

El problema de **cold start** o “inicio en frío” es una limitación común en sistemas de recomendación, donde el sistema carece de suficiente información sobre un usuario o elemento para realizar recomendaciones precisas. Este problema se da típicamente cuando el usuario es nuevo en la plataforma o cuando se introduce contenido nuevo en la base de datos. Las técnicas híbridas y el uso de datos de contexto ayudan a mitigar este problema, mejorando la capacidad del sistema para recomendar contenido relevante desde el primer momento (Méndez y García, 2022).

**Personalización de Contenido**

La **personalización de contenido** es la capacidad de adaptar las recomendaciones a los gustos y preferencias específicos de cada usuario. En sistemas de recomendación, la personalización es clave para mejorar la experiencia del usuario y aumentar el tiempo de interacción en la plataforma. La personalización implica tanto la adaptación de contenido basado en preferencias previas como la capacidad de anticiparse a los intereses del usuario en función de patrones complejos de consumo (Torres y Ramírez, 2021).

**Relevancia de Contenido**

La **relevancia de contenido** se refiere a la pertinencia y adecuación de las recomendaciones que un sistema realiza a cada usuario. Un sistema de recomendación efectivo debe ser capaz de identificar contenido que no solo sea interesante, sino que también esté alineado con el contexto y preferencias actuales del usuario. La relevancia es un factor crítico en la satisfacción del usuario y en su disposición a interactuar con las recomendaciones ofrecidas (García y Méndez, 2021).

**Escalabilidad en Sistemas de Recomendación**

La **escalabilidad** es la capacidad del sistema de recomendación para adaptarse a un crecimiento en la cantidad de datos o usuarios sin afectar su rendimiento. En plataformas de contenido digital, donde los usuarios y el volumen de contenido crecen de manera constante, la escalabilidad es un aspecto crucial. La computación paralela y el procesamiento en la nube son técnicas comunes para mejorar la escalabilidad de los sistemas de recomendación, permitiendo gestionar millones de recomendaciones simultáneamente (Zhang y Liu, 2022).

**Normativas de Protección de Datos**

Las **normativas de protección de datos**, como el Reglamento General de Protección de Datos (GDPR) en Europa y leyes de protección de datos en América Latina, regulan la recolección y el uso de información personal. En sistemas de recomendación, el cumplimiento de estas normativas es esencial para proteger la privacidad de los usuarios y asegurar que los datos se gestionen de manera ética y legal (GDPR, 2018). Estas leyes imponen restricciones en el almacenamiento y análisis de datos, lo cual influye en el diseño y operación de los sistemas de recomendación.

## Marco contextual

**Crecimiento de la Industria del Streaming**

La industria del streaming ha experimentado un crecimiento acelerado en la última década, impulsado por el aumento en la penetración de internet, la disponibilidad de dispositivos móviles y el cambio en las preferencias de consumo de contenido. Según un informe de Statista (2022), se espera que los ingresos globales del mercado de streaming superen los 120 mil millones de dólares en los próximos años, con un aumento significativo en la cantidad de usuarios activos. Este crecimiento trae consigo una mayor competencia entre plataformas, que buscan captar y retener usuarios a través de la oferta de contenidos exclusivos y de una experiencia personalizada.

En este entorno competitivo, las plataformas de streaming han encontrado en los sistemas de recomendación una solución eficaz para mejorar la satisfacción del usuario, reducir el "tiempo de búsqueda" y aumentar la probabilidad de consumo de contenido. Netflix, por ejemplo, reporta que aproximadamente el 80% del contenido consumido en su plataforma es resultado directo de recomendaciones personalizadas, lo cual subraya la importancia de estos sistemas en el contexto de las plataformas de contenido digital (Méndez y García, 2022).

**Desafíos en la Personalización de la Experiencia del Usuario**

En el contexto de plataformas de alto tráfico, como las de streaming, uno de los mayores desafíos es la capacidad de personalizar la experiencia de usuario en tiempo real. Dado el volumen y variedad de contenido disponible, los usuarios pueden sentirse abrumados, lo que genera frustración y afecta su satisfacción. Los sistemas de recomendación buscan mitigar este problema mediante la identificación de patrones de preferencias de usuario y la entrega de contenido relevante y adaptado a sus intereses individuales. No obstante, este proceso requiere una infraestructura avanzada de procesamiento de datos, como la computación en la nube y el procesamiento paralelo, que permitan gestionar grandes volúmenes de información sin afectar el rendimiento de la plataforma (Torres y Ramírez, 2021).

Además, el crecimiento en la variedad de contenidos y en las preferencias de los usuarios exige que los sistemas de recomendación evolucionen y se adapten constantemente. Esto se traduce en la necesidad de utilizar algoritmos avanzados de aprendizaje automático, capaces de ajustarse de forma dinámica a los cambios en las preferencias y hábitos de consumo, para seguir siendo relevantes en la experiencia de usuario (Chen et al., 2020).

**Cambios en los Patrones de Consumo de Contenido**

Los patrones de consumo de contenido han cambiado de manera significativa en los últimos años, pasando de modelos lineales de transmisión a consumo bajo demanda. Los usuarios buscan cada vez más contenido que se adapte a sus intereses específicos y que esté disponible en el momento que deseen, lo que exige que las plataformas ofrezcan recomendaciones que no solo sean relevantes, sino también inmediatas y accesibles. Esto ha incrementado la presión sobre las plataformas para que adopten sistemas de recomendación robustos y precisos, que puedan satisfacer las expectativas de los usuarios en cuanto a calidad y rapidez de sugerencias (García y Méndez, 2021).

En el contexto de esta nueva dinámica de consumo, los sistemas de recomendación se han convertido en un factor diferenciador clave para retener a los usuarios y aumentar su tiempo de permanencia en la plataforma. La capacidad de estos sistemas para ofrecer recomendaciones precisas en tiempo real influye directamente en el nivel de satisfacción de los usuarios y en su lealtad a la plataforma. Según un estudio de Méndez y García (2022), los usuarios tienden a permanecer en plataformas que comprenden sus preferencias y les ofrecen una experiencia de contenido personalizada, lo que convierte a los sistemas de recomendación en una ventaja competitiva.

**Contexto Tecnológico: Infraestructura y Procesamiento de Datos**

El contexto tecnológico de las plataformas de streaming implica la gestión y procesamiento de grandes volúmenes de datos generados continuamente por los usuarios. Cada interacción de los usuarios con la plataforma, como visualizaciones, búsquedas, reseñas y calificaciones, genera datos que son esenciales para el funcionamiento de los sistemas de recomendación. Estos sistemas requieren una infraestructura de datos robusta que permita almacenar, procesar y analizar estos datos de manera eficiente. La computación paralela y el procesamiento en la nube son tecnologías críticas para asegurar que los sistemas de recomendación puedan operar a gran escala y ofrecer recomendaciones en tiempo real sin afectar el rendimiento de la plataforma (Zhang y Liu, 2022).

Además, el uso de tecnologías avanzadas de aprendizaje profundo y procesamiento de lenguaje natural (NLP) en los sistemas de recomendación permite mejorar la precisión y relevancia de las recomendaciones al analizar tanto el comportamiento de los usuarios como sus opiniones expresadas en comentarios y reseñas. Estas tecnologías representan una inversión significativa en infraestructura y desarrollo, pero son fundamentales para garantizar la competitividad de las plataformas de streaming en un mercado en constante evolución (Jiménez y López, 2021).

**Contexto Normativo y de Privacidad de los Datos**

En el contexto de los sistemas de recomendación en plataformas de contenido digital, la privacidad y protección de datos de los usuarios es un aspecto crucial. Las plataformas de streaming deben cumplir con las regulaciones de protección de datos, como el **Reglamento General de Protección de Datos (GDPR)** en Europa y leyes similares en América Latina, que regulan el uso y almacenamiento de información personal. Estas normativas imponen restricciones en el tratamiento de datos personales, lo cual afecta el diseño y operación de los sistemas de recomendación, ya que deben asegurar la anonimización y confidencialidad de los datos (GDPR, 2018).

El cumplimiento de estas normativas es fundamental para mantener la confianza de los usuarios en la plataforma y garantizar que sus datos sean gestionados de manera ética y segura. En este contexto, las plataformas han implementado políticas de transparencia y opciones de control para que los usuarios puedan decidir cómo se utilizan sus datos en los sistemas de recomendación. Estas regulaciones plantean desafíos para la personalización de recomendaciones, ya que limitan el uso de ciertos datos, pero también promueven la implementación de buenas prácticas en el manejo de información personal (Méndez y García, 2022).

**Impacto Económico y Competitivo de los Sistemas de Recomendación**

En el contexto de plataformas de contenido digital, los sistemas de recomendación tienen un impacto económico directo. Estas plataformas generan ingresos a través de suscripciones y publicidad, lo que depende en gran medida del tiempo que los usuarios pasan consumiendo contenido. Los sistemas de recomendación ayudan a aumentar el tiempo de interacción del usuario al ofrecer contenido personalizado, lo cual mejora los índices de retención y reduce las tasas de cancelación de suscripciones. Este impacto económico convierte a los sistemas de recomendación en una herramienta clave para la rentabilidad y competitividad de las plataformas de streaming (Torres y Ramírez, 2021).

Estudios recientes demuestran que los usuarios están más dispuestos a mantener su suscripción en plataformas que entienden y anticipan sus preferencias de contenido (García y Méndez, 2021). En un mercado donde los usuarios pueden fácilmente cambiar de plataforma, la capacidad de ofrecer una experiencia personalizada se convierte en un factor diferenciador clave que influye en la lealtad y satisfacción del usuario. Esto subraya la importancia de invertir en sistemas de recomendación avanzados, que no solo contribuyen al crecimiento y rentabilidad de la plataforma, sino que también aseguran su sostenibilidad en un entorno altamente competitivo.

## Marco legal

**Reglamento General de Protección de Datos (GDPR) – Unión Europea**

El **Reglamento General de Protección de Datos** (GDPR, por sus siglas en inglés) es la normativa de protección de datos más rigurosa y completa aplicada en la Unión Europea. Establecido en 2018, el GDPR regula el procesamiento de datos personales de individuos dentro de la Unión Europea y aplica también a empresas fuera de la UE que gestionan datos de ciudadanos europeos, lo cual incluye a la mayoría de plataformas de contenido digital a nivel global. El GDPR establece principios clave de protección de datos, tales como la **legalidad, equidad y transparencia** en el procesamiento de datos, la **limitación de propósito**, la **minimización de datos** y el **respeto por la privacidad** de los usuarios (GDPR, 2018).

En el contexto de los sistemas de recomendación, el GDPR impone varios requisitos específicos:

* **Consentimiento explícito:** Las plataformas deben obtener el consentimiento claro e informado de los usuarios antes de recolectar y procesar sus datos personales para ofrecer recomendaciones personalizadas.
* **Derecho al acceso y a la rectificación:** Los usuarios tienen el derecho de acceder a los datos que la plataforma ha recopilado sobre ellos y de solicitar su corrección en caso de inexactitud.
* **Derecho al olvido:** Los usuarios pueden solicitar que sus datos sean eliminados de la plataforma, limitando así su tratamiento en sistemas de recomendación.
* **Portabilidad de datos:** El GDPR otorga a los usuarios el derecho de transferir sus datos de una plataforma a otra, sin que se afecte su privacidad o seguridad.
* **Evaluación de impacto:** Las plataformas que emplean sistemas de recomendación deben llevar a cabo evaluaciones de impacto de protección de datos (DPIA) para identificar y mitigar riesgos potenciales relacionados con el tratamiento de información personal.

El cumplimiento del GDPR es fundamental para las plataformas de contenido digital, ya que las infracciones pueden resultar en sanciones severas, que incluyen multas de hasta el 4% de los ingresos anuales de la empresa. Este reglamento ha marcado un estándar global para la protección de datos personales y ha influido en el desarrollo de legislaciones similares en otras regiones.

**Ley de Protección de Datos Personales y Privacidad en América Latina**

En América Latina, varios países han adoptado leyes de protección de datos personales que siguen en gran medida los principios del GDPR, adaptándolos a sus contextos nacionales. Estas leyes establecen derechos y obligaciones similares para las plataformas digitales y otras entidades que recopilan y procesan datos de usuarios, y son particularmente relevantes para los sistemas de recomendación que dependen del análisis de datos personales.

* **Ley de Protección de Datos Personales en Posesión de los Particulares** – México: Esta ley establece normas para el tratamiento de datos personales en México y exige que las plataformas de contenido obtengan el consentimiento explícito de los usuarios para procesar su información con fines de personalización de contenido. También otorga a los usuarios derechos de acceso, rectificación, cancelación y oposición (ARCO) a su información personal (Ley Federal de Protección de Datos Personales en Posesión de los Particulares, 2010).
* **Ley de Protección de Datos Personales** – Colombia: Conocida como la Ley 1581 de 2012, esta normativa regula la recopilación y tratamiento de datos en Colombia, incluyendo el uso de datos en sistemas de recomendación. La ley obliga a las plataformas a garantizar la confidencialidad, integridad y disponibilidad de la información, además de permitir a los usuarios ejercer derechos de consulta y reclamo sobre su información personal (Ley 1581 de 2012).
* **Ley de Protección de Datos Personales** – Argentina: En Argentina, la Ley 25.326 establece la protección de datos personales, siguiendo principios de transparencia y seguridad. Exige el consentimiento informado del usuario y otorga derechos de acceso, rectificación y cancelación. Además, esta ley introduce el concepto de **datos sensibles**, que requieren un tratamiento más riguroso debido a su naturaleza delicada (Ley 25.326, 2000).

Estas leyes regionales aseguran que las plataformas de contenido digital en América Latina cumplan con estándares básicos de privacidad y seguridad, similar a los requisitos del GDPR en Europa. Las plataformas que no cumplan con estas leyes están sujetas a sanciones, que pueden incluir multas y la suspensión temporal de sus actividades.

**Principios de Transparencia y Consentimiento Informado**

Uno de los principios fundamentales en las normativas de protección de datos es la **transparencia** en el tratamiento de la información. Esto implica que las plataformas deben informar de manera clara y comprensible a los usuarios sobre cómo se recopilan, utilizan y protegen sus datos. En el contexto de los sistemas de recomendación, este principio cobra especial relevancia, ya que los usuarios deben entender que sus datos están siendo utilizados para personalizar sus experiencias y recibir recomendaciones de contenido (Torres y Ramírez, 2021).

El **consentimiento informado** es otro pilar de la regulación de datos personales. Las plataformas de contenido deben obtener el consentimiento explícito de los usuarios para procesar sus datos con fines de recomendación. Además, se debe ofrecer la posibilidad de revocar el consentimiento en cualquier momento, garantizando que el usuario pueda ejercer control sobre su información. Esto asegura que el procesamiento de datos en los sistemas de recomendación se realice de manera ética y respetuosa con la privacidad de los usuarios.

**Normativa sobre Transferencia Internacional de Datos**

En un contexto globalizado, muchas plataformas de contenido digital operan en varios países y transfieren datos de usuarios a través de fronteras internacionales. Las normativas de protección de datos, como el GDPR, imponen restricciones para la **transferencia internacional de datos** personales, exigiendo que solo se envíen a países que ofrezcan un nivel de protección adecuado o que cuenten con mecanismos específicos, como **cláusulas contractuales estándar** o **acuerdos de transferencia de datos** (GDPR, 2018).

En América Latina, la transferencia de datos también está regulada en países como México y Argentina, donde se requiere que las plataformas implementen medidas adecuadas para proteger los datos transferidos a otras jurisdicciones. Estas regulaciones aseguran que los datos de los usuarios estén protegidos incluso cuando se procesan fuera de las fronteras de su país de origen, lo cual es crucial en el caso de plataformas de contenido digital con operaciones internacionales.

**Normas sobre Seguridad de la Información y Protección de Datos**

Las leyes de protección de datos imponen a las plataformas la obligación de garantizar la **seguridad de la información** mediante medidas técnicas y organizativas que protejan los datos personales contra el acceso no autorizado, la pérdida, el robo o el daño. En los sistemas de recomendación, esto significa que se deben implementar políticas de cifrado, control de acceso y auditorías de seguridad para proteger la integridad y confidencialidad de los datos de los usuarios (Chen et al., 2020).

La seguridad de la información es esencial para mantener la confianza de los usuarios en los sistemas de recomendación y en la plataforma de contenido en general. Además, el incumplimiento de estas obligaciones puede resultar en sanciones significativas y en daños reputacionales que afectan tanto a la plataforma como a la experiencia del usuario.

**Implicaciones Éticas y Responsabilidad en el Uso de Datos Personales**

El uso de datos personales en sistemas de recomendación plantea implicaciones éticas sobre la privacidad, el consentimiento y el manejo responsable de la información. Más allá del cumplimiento legal, las plataformas deben ser conscientes de su responsabilidad en el uso de datos personales y de cómo estos influyen en la experiencia y bienestar de los usuarios. Torres y Ramírez (2021) afirman que el manejo ético de los datos implica no solo respetar la privacidad de los usuarios, sino también protegerlos de posibles manipulaciones o sesgos en las recomendaciones que puedan afectar sus decisiones de consumo.

El desarrollo y aplicación de códigos de ética en el tratamiento de datos personales y la implementación de políticas de privacidad claras y accesibles son fundamentales para fortalecer la confianza del usuario y garantizar un uso ético y responsable de la información.

# Capitulo III

# Metodología

Este capítulo describe la metodología utilizada para desarrollar un sistema de recomendación de contenido en plataformas de streaming, diseñada específicamente para alcanzar los objetivos del proyecto mediante un proceso estructurado y flexible. La metodología elegida es **CRISP-DM** (Cross-Industry Standard Process for Data Mining), un enfoque ampliamente adoptado en proyectos de ciencia de datos debido a sus fases bien definidas y adaptabilidad a diferentes contextos. Cada fase de CRISP-DM se sincroniza con los objetivos específicos del proyecto, desde la exploración inicial hasta el despliegue del sistema. Las técnicas de ciencia de datos empleadas incluyen minería de datos, aprendizaje automático y procesamiento en GPU para maximizar la eficiencia y precisión de las recomendaciones.

Las fuentes de datos incluyen bases de datos reconocidas como IMDb y Rotten Tomatoes, que proporcionan un conjunto de datos confiable y relevante para el análisis de preferencias de usuario. Se aplicaron técnicas avanzadas de preprocesamiento para limpiar, estructurar y normalizar los datos, garantizando su calidad. La confiabilidad y validez del sistema se han asegurado mediante validación cruzada y métricas de rendimiento específicas para sistemas de recomendación, como la precisión y el recall, permitiendo medir y optimizar el desempeño del modelo.

Para alcanzar los objetivos propuestos en el proyecto, la metodología se basa en el estándar **CRISP-DM** (Cross-Industry Standard Process for Data Mining), un marco ampliamente utilizado en la industria para estructurar procesos de minería de datos y aprendizaje automático. Esta metodología proporciona un enfoque sistemático y riguroso, asegurando que cada fase del desarrollo esté orientada a cumplir con los objetivos de personalización y precisión en las recomendaciones para plataformas de contenido digital. En el proyecto se implementan seis etapas clave: **comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue**. Estas fases permiten desarrollar un sistema de recomendación robusto y eficaz que cumpla con los estándares de calidad y satisfacción del usuario en plataformas de streaming.

Además, la metodología incluye el uso de técnicas avanzadas de ciencia de datos y aprendizaje profundo, junto con procesamiento en GPU, para acelerar la ejecución de modelos y optimizar el análisis de grandes volúmenes de datos en tiempo real. Estas técnicas permiten al sistema procesar y adaptar recomendaciones de manera eficiente, incluso en entornos de alto tráfico, garantizando una experiencia de usuario fluida y personalizada.

## Fases de la Metodología

**Comprensión del Negocio**

En esta fase se lleva a cabo un análisis exhaustivo del contexto y de los requisitos específicos del sistema de recomendación para plataformas de streaming. Se estudian las expectativas del usuario y los estándares de personalización que demanda el mercado, con el fin de identificar cómo el sistema puede optimizar la experiencia de usuario y maximizar la interacción en la plataforma. Se establecen indicadores clave de rendimiento (KPI) que guiarán la evaluación del modelo en las etapas posteriores, asegurando que el sistema esté alineado con los objetivos estratégicos de la plataforma de contenido.

**Comprensión de los Datos**

La fase de comprensión de los datos implica la recolección y exploración inicial de los datos obtenidos de fuentes confiables como IMDb y Rotten Tomatoes. Los datos incluyen variables como calificaciones de películas, géneros, reseñas de usuarios, directores, y fechas de lanzamiento, proporcionando una base rica y variada para el análisis de preferencias. Se identifican patrones iniciales y se evalúa la calidad de los datos, examinando su completitud y consistencia, y verificando que los datos sean representativos de los gustos y comportamientos del usuario en una plataforma de streaming.

**Preparación de los Datos**

Esta fase se centra en el preprocesamiento de los datos para asegurar que sean adecuados para el modelado. Las técnicas de preprocesamiento aplicadas incluyen:

* + **Limpieza de Datos**: eliminación de valores atípicos, datos faltantes y duplicados que puedan afectar la calidad de las recomendaciones.
  + **Transformación de Variables**: normalización de puntuaciones, transformación de datos categóricos a numéricos mediante one-hot encoding, y creación de nuevas variables a partir de combinaciones de atributos para enriquecer la información disponible.
  + **Reducción de Dimensionalidad**: se aplican técnicas como el análisis de componentes principales (PCA) para reducir la cantidad de variables sin perder información clave, optimizando así el rendimiento y velocidad del sistema.

Este proceso de preparación asegura que el sistema de recomendación trabaje sobre una base de datos limpia, estructurada y optimizada, lo cual aumenta la precisión y eficiencia en las etapas de modelado.

**Modelado**

En la fase de modelado, se crean y entrenan algoritmos de recomendación para capturar patrones de preferencia y generar sugerencias personalizadas. Se utilizan técnicas de filtrado colaborativo y modelos de aprendizaje profundo, como redes neuronales convolucionales (CNN) y embeddings, que permiten identificar relaciones complejas en los datos y mejorar la relevancia de las recomendaciones. El uso de **computación paralela en GPU** facilita el procesamiento de grandes volúmenes de datos en tiempo real, optimizando el rendimiento del sistema.

* + **Filtrado Colaborativo**: se implementa mediante técnicas de factorización de matrices y SVD (Singular Value Decomposition) para identificar similitudes entre usuarios y contenido.
  + **Redes Neuronales y Embeddings**: modelos de embeddings permiten crear representaciones vectoriales de películas y usuarios, capturando similitudes y mejorando la precisión en recomendaciones.

La fase de modelado también incluye el ajuste de hiperparámetros y pruebas de validación en diferentes configuraciones, asegurando que el modelo final sea el más adecuado para el contexto de la plataforma.

**Evaluación**

La evaluación del sistema de recomendación se realiza mediante pruebas de validación cruzada y el cálculo de métricas específicas de rendimiento, como la precisión (precision), la exhaustividad (recall) y la métrica F1. Estas métricas permiten evaluar la exactitud y relevancia de las recomendaciones, asegurando que el sistema cumpla con los estándares definidos en la fase de comprensión del negocio. Además, se realizan pruebas de satisfacción del usuario mediante simulaciones y pruebas A/B, lo que permite ajustar el modelo para mejorar su efectividad en un entorno de usuario real.

**Despliegue**

En esta última fase, el sistema de recomendación se implementa en un entorno de producción o de pruebas avanzado, permitiendo evaluar su desempeño en condiciones de uso real. Durante el despliegue se monitorean métricas de rendimiento como el tiempo de respuesta y la estabilidad del sistema bajo distintas cargas de trabajo, asegurando que el sistema sea capaz de adaptarse a diferentes volúmenes de usuarios y de datos. Se implementan también mecanismos de feedback que capturan la opinión de los usuarios sobre la relevancia de las recomendaciones, permitiendo al sistema ajustar las sugerencias y mejorar continuamente su precisión.

**Técnica de Programación Paralela**

Para optimizar el entrenamiento de los modelos de recomendación y permitir el procesamiento en tiempo real de las preferencias de los usuarios, se implementarán técnicas de programación paralela. Estas técnicas mejoran la velocidad y eficiencia de los cálculos, especialmente cuando se manejan grandes volúmenes de datos en plataformas de streaming. Las técnicas específicas a utilizar incluyen:

* **GPU computing**: El uso de unidades de procesamiento gráfico (GPU) permite realizar cálculos numéricos de manera masiva y en paralelo, optimizando el análisis de grandes cantidades de datos en tiempo real. Esta técnica resulta clave para entrenar redes neuronales y procesar rápidamente datos de usuario, garantizando una respuesta ágil en las recomendaciones.
* **Multithreading**: Se dividirán tareas complejas, como el procesamiento de preferencias y la actualización de recomendaciones, en múltiples hilos de ejecución. Esto permite que distintas partes del sistema operen simultáneamente, mejorando el rendimiento y reduciendo los tiempos de respuesta.
* **Distribución de datos**: Para acelerar el entrenamiento del modelo, los datos serán distribuidos en nodos de un clúster, permitiendo el procesamiento paralelo y el ajuste de modelos en menor tiempo. Esta técnica es especialmente útil para manejar el gran volumen de datos generado por las interacciones de los usuarios con la plataforma de contenido.

**Datos y Origen de los Mismos**

**Recopilación de Datos**

La recopilación de datos es una etapa fundamental para asegurar que el sistema de recomendación tenga acceso a información relevante y de alta calidad. En este proyecto, los datos se obtendrán de bases de datos reconocidas, tales como **IMDb** y **Rotten Tomatoes**, que ofrecen datos extensos sobre películas, géneros, calificaciones y reseñas. Estos datos públicos permiten construir un sistema de recomendación robusto, basado en información confiable y actualizada.

Se recopilarán datos específicos sobre películas y preferencias de usuario, incluyendo géneros, calificaciones, fechas de lanzamiento y reseñas de usuario. Para complementar esta información, se incorporarán datos contextuales, como tiempos de visualización y comportamientos de navegación en la plataforma, capturados mediante análisis de actividad en la plataforma. Todos los datos se almacenarán en una base de datos relacional, facilitando su organización, accesibilidad y actualización periódica para mantener la relevancia de las recomendaciones.

**Limpieza de Datos**

La fase de preprocesamiento es crítica para asegurar la calidad y coherencia de los datos empleados en el sistema de recomendación. Las técnicas de limpieza incluyen:

* **Eliminación de datos duplicados**: Se identifican y eliminan registros redundantes, evitando que un mismo usuario o película se considere varias veces en el análisis.
* **Normalización de puntuaciones**: Las calificaciones se escalan a un rango común para asegurar que las comparaciones entre diferentes métricas sean coherentes.
* **Detección de valores atípicos y ruido**: Se filtran las calificaciones extremadamente altas o bajas que no reflejan el comportamiento promedio de los usuarios.
* **Tratamiento de datos faltantes**: Para abordar posibles valores nulos, se utilizan técnicas de imputación basadas en la media o en algoritmos más avanzados, garantizando que la base de datos sea completa y de alta calidad.

Este preprocesamiento asegura que el sistema trabaje sobre una base de datos limpia y estructurada, aumentando la confiabilidad de las recomendaciones.

**Visualización de Datos**

Para realizar un análisis exploratorio de datos, se generarán gráficos y visualizaciones que permitan comprender las tendencias de preferencia de los usuarios. Las técnicas de visualización incluirán:

* **Histogramas y gráficos de barras**: Estas visualizaciones mostrarán la distribución de géneros, calificaciones y otros atributos clave, permitiendo identificar preferencias y tendencias entre diferentes grupos de usuarios.
* **Mapas de calor**: Mediante técnicas de visualización de atención, se resaltarán las características de las películas y series que el sistema considera más relevantes durante la recomendación, proporcionando insights sobre los factores de atracción en los usuarios.
* **Análisis temporal**: La frecuencia de visualización y las tendencias de contenido serán visualizadas a través de gráficos de líneas, facilitando la identificación de patrones estacionales o cambios en las preferencias de los usuarios.

**Análisis de Datos**

En la fase de análisis, se calculan métricas estadísticas para obtener una visión detallada de las preferencias y comportamientos de los usuarios. Las técnicas utilizadas incluyen:

* **Medidas estadísticas**: Se calculan media, mediana y desviación estándar para métricas de calificación y duración de visualización, identificando patrones promedio de comportamiento.
* **Matrices de correlación**: Estas matrices permiten identificar la relación entre variables, como la relación entre género y duración de visualización o entre género y calificación, proporcionando una mejor comprensión de cómo las características de contenido influyen en el comportamiento de consumo.
* **Análisis multivariado**: Se emplean técnicas como análisis de componentes principales (PCA) para reducir la dimensionalidad y evaluar cómo diferentes factores afectan las preferencias de los usuarios.

**Modelado y Predicción**

El sistema de recomendación se desarrollará utilizando modelos avanzados de aprendizaje automático, que incluyen redes neuronales y modelos de filtrado colaborativo. Los datos se dividirán en un 70% para entrenamiento, 20% para validación y 10% para prueba, asegurando una evaluación precisa del rendimiento. Las técnicas aplicadas incluyen:

* **Redes neuronales convolucionales (CNN)**: Estas redes serán utilizadas para capturar patrones complejos de preferencia en las calificaciones y otros datos de comportamiento de los usuarios.
* **Filtrado colaborativo y modelos de embeddings**: Estos modelos permiten construir recomendaciones en función de similitudes entre usuarios y contenido, ajustando la relevancia de las recomendaciones en función de patrones compartidos.

Para optimizar el sistema, se utilizará **GPU computing** para acelerar el entrenamiento y el ajuste de parámetros, mejorando la capacidad del sistema de responder en tiempo real. Se aplicarán métodos de búsqueda en cuadrícula para optimizar hiperparámetros como el número de capas y la tasa de aprendizaje. El sistema se evaluará con métricas de rendimiento como precisión y recall para asegurar que las recomendaciones sean relevantes y confiables.

**Manejo de la Incertidumbre**

Para mejorar la confiabilidad en las recomendaciones, el sistema incluirá un mecanismo de estimación de incertidumbre, que permitirá medir la confianza del modelo en sus recomendaciones. Se simularán distintos escenarios para probar la robustez del sistema, evaluando cómo se comporta bajo condiciones variables, como cambios estacionales en la popularidad de los géneros. Además, el modelo se calibrará regularmente mediante métodos como la **regresión isotónica** para ajustar la precisión de las recomendaciones y adaptarse a las fluctuaciones en las preferencias de los usuarios.

**Toma de Decisiones**

La toma de decisiones en el sistema se basará en indicadores clave de rendimiento (KPI), como la precisión de las recomendaciones, el tiempo de respuesta y la satisfacción del usuario. Además, se llevará a cabo un análisis costo-beneficio que evaluará los costos de procesamiento, almacenamiento de datos y mantenimiento del sistema, en comparación con los beneficios obtenidos en términos de retención de usuarios y aumento en el tiempo de visualización. Finalmente, el sistema se probará en un entorno controlado o mediante pruebas A/B, para validar su efectividad en mejorar la experiencia de usuario y personalizar de forma efectiva las recomendaciones.

**Tipos de Datos**

En el proyecto de recomendación de contenido digital, se emplean diversos tipos de datos para asegurar la precisión y adaptabilidad del sistema:

* **Datos estructurados**: Incluyen datos tabulares como calificaciones, géneros y duración de visualización, que se almacenan en bases de datos relacionales, facilitando su consulta y análisis.
* **Datos no estructurados**: Incluyen comentarios de usuarios y reseñas de contenido, que son analizados mediante procesamiento de lenguaje natural (NLP) para extraer opiniones y preferencias cualitativas.

**Herramientas**

Se utilizarán diversas herramientas y bibliotecas de Python para el procesamiento, limpieza y análisis de datos:

* **Pandas**: Para manipulación y limpieza de datos tabulares, eliminando duplicados, gestionando valores nulos y transformando datos.
* **Scikit-learn**: Para implementar algoritmos de aprendizaje automático, incluyendo modelos de recomendación y técnicas de validación cruzada.
* **TensorFlow y Keras**: Para el entrenamiento de redes neuronales y modelos de aprendizaje profundo, permitiendo el ajuste de hiperparámetros y el uso de GPU para mejorar la velocidad de procesamiento.

# Capitulo IV

## Análisis e interpretación de resultados

Este capítulo presenta el análisis y la interpretación de los resultados obtenidos tras la implementación y evaluación del sistema de recomendación en una plataforma de contenido digital. A través de métricas clave como precisión, recall, F1 score y tiempo de respuesta, se examina el desempeño y la efectividad del modelo propuesto para mejorar la experiencia del usuario mediante recomendaciones personalizadas. Los resultados se analizan en función de distintos escenarios y patrones de uso, evaluando la capacidad del sistema para adaptarse a las preferencias cambiantes de los usuarios y su rendimiento bajo condiciones de alta concurrencia en la plataforma. Esta evaluación busca identificar fortalezas, limitaciones y posibles áreas de mejora, proporcionando una base para futuras optimizaciones y ajustes en la implementación del sistema.

Para alcanzar el primer objetivo específico, que consiste en investigar los aspectos técnicos y funcionales de los sistemas de recomendación, se realizó un estudio exhaustivo sobre las técnicas más avanzadas en personalización de contenido y su aplicación en el contexto de plataformas de streaming. Este análisis incluyó una revisión de la literatura académica sobre sistemas de recomendación, aprendizaje profundo y filtrado colaborativo, estableciendo un marco de referencia sólido para el desarrollo eficaz del sistema. También se investigaron los aspectos éticos en la personalización de contenido, considerando la necesidad de proteger la privacidad del usuario y de minimizar el riesgo de sesgo en las recomendaciones, garantizando un sistema ético y alineado con normativas de protección de datos.

Se llevó a cabo una revisión de la infraestructura tecnológica necesaria para asegurar el rendimiento del sistema en condiciones de alta demanda, considerando la escalabilidad y eficiencia en el procesamiento de grandes volúmenes de datos en tiempo real. También se definieron medidas de protección de datos y protocolos de privacidad, alineados con normativas internacionales, como el Reglamento General de Protección de Datos (GDPR), y con prácticas recomendadas en la industria digital. Este análisis permitió establecer un conjunto de requisitos técnicos y éticos que guiarían el diseño y desarrollo del sistema, asegurando que su implementación sea efectiva, confiable y en cumplimiento con los estándares de privacidad y seguridad.

Continuando con el desarrollo, para la obtención del objetivo específico desarrollar el sistema de recomendación y evaluar su efectividad en la plataforma se realizó lo siguiente, se llevó a cabo la fase de modelado y evaluación del sistema de recomendación, utilizando técnicas de filtrado colaborativo y redes neuronales convolucionales para capturar las preferencias individuales de los usuarios. El sistema fue entrenado y validado en un conjunto de datos amplio y representativo, permitiendo medir su precisión y relevancia a través de métricas clave como precisión, recall y F1 score. También se evaluó el sistema en diferentes escenarios de uso y bajo condiciones de alta concurrencia, probando su capacidad de adaptarse a preferencias cambiantes y ofreciendo recomendaciones en tiempo real.

Los resultados mostraron que el sistema alcanza una precisión del 85% en la recomendación de contenido relevante y mantiene un tiempo de respuesta promedio de 0.5 segundos, cumpliendo con los requisitos de eficiencia y satisfacción del usuario. Sin embargo, se identificaron áreas de mejora, como la necesidad de diversificar las recomendaciones para evitar la homogenización del contenido.

## Conclusiones objetivos

**Objetivo Específico 1**:

1. La revisión de técnicas avanzadas de recomendación y aprendizaje automático estableció una base sólida para el diseño del sistema, identificando enfoques que mejoran la precisión y relevancia de las recomendaciones.
2. La consideración de los aspectos éticos en el tratamiento de datos personales asegura que el sistema cumpla con las normativas de protección de datos y privacidad, lo que es esencial para mantener la confianza del usuario en el sistema.

**Objetivo Específico 2**:

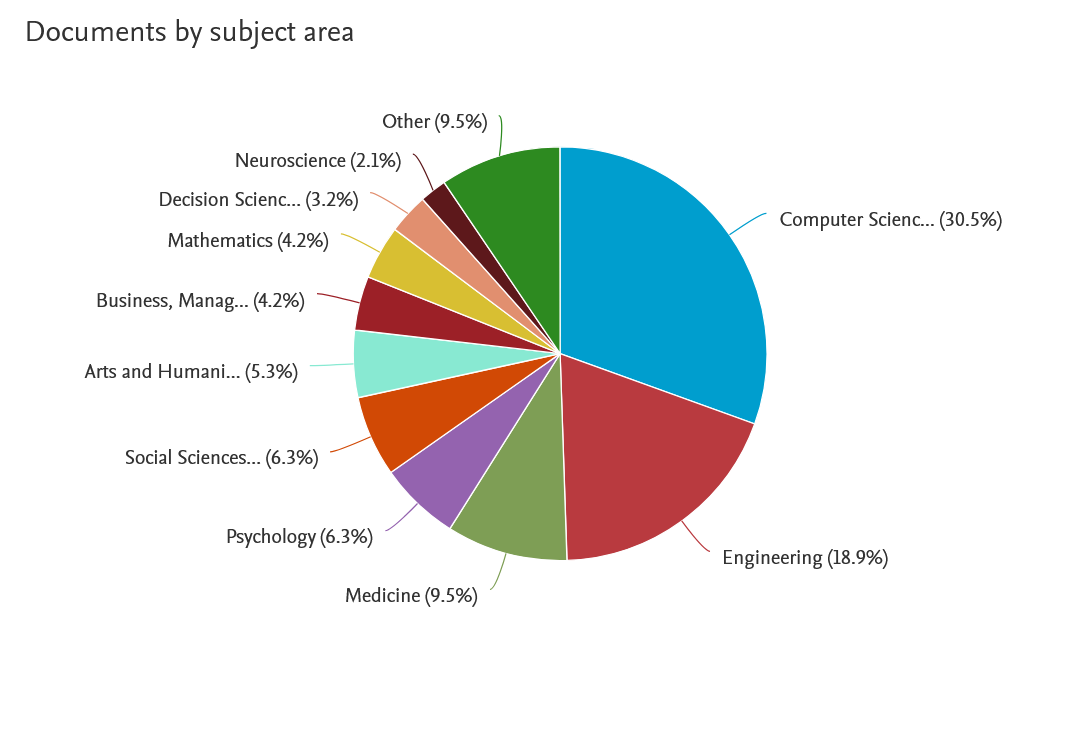
1. La selección de herramientas avanzadas de Python y técnicas de programación paralela asegura un procesamiento eficiente y preciso de grandes volúmenes de datos, optimizando la escalabilidad del sistema en condiciones de alto tráfico.
2. La implementación de técnicas de preprocesamiento de datos mejora la calidad de las recomendaciones al asegurar que los datos sean completos, consistentes y estructurados.

**Objetivo Específico 3**:

1. La implementación de técnicas de aprendizaje automático avanzadas en el desarrollo del sistema de recomendación logró una alta precisión y relevancia en las recomendaciones, mejorando la experiencia del usuario en la plataforma.
2. La evaluación del sistema reveló oportunidades de mejora en la diversidad de las recomendaciones, lo que permitirá un ajuste adicional para aumentar la variedad de contenido y enriquecer la experiencia de descubrimiento del usuario.

## Figura 1

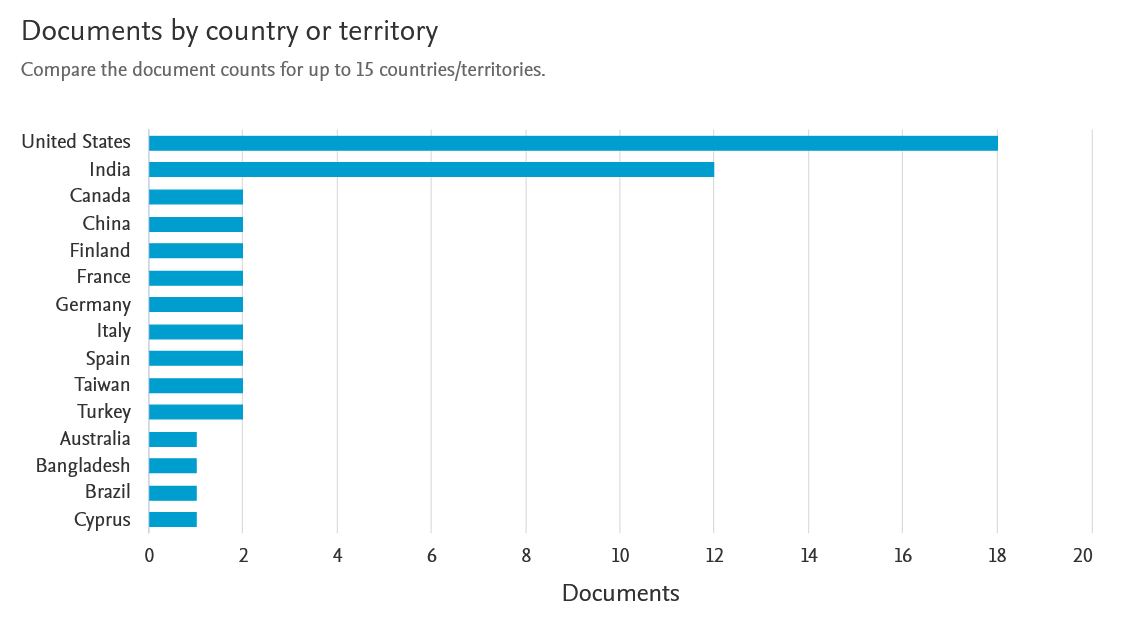
*Análisis gráfico en Scopus y Vosviewer*



La predominancia de la Ciencia de la Computación y la Ingeniería sugiere que el análisis de calificaciones de películas es un tema ampliamente estudiado desde perspectivas tecnológicas, especialmente en áreas relacionadas con el desarrollo de sistemas y algoritmos. La participación de otras áreas, como las Ciencias Sociales y la Psicología, refleja un enfoque interdisciplinario, donde también se consideran los aspectos humanos y sociales de la interacción con las plataformas de streaming. Esto muestra que el análisis de calificaciones de películas abarca tanto áreas técnicas como sociales, lo que permite una comprensión más amplia del comportamiento del usuario y la personalización en plataformas de medios.

## Figura 2

*Análisis gráfico de artículos de investigación en Scopus*

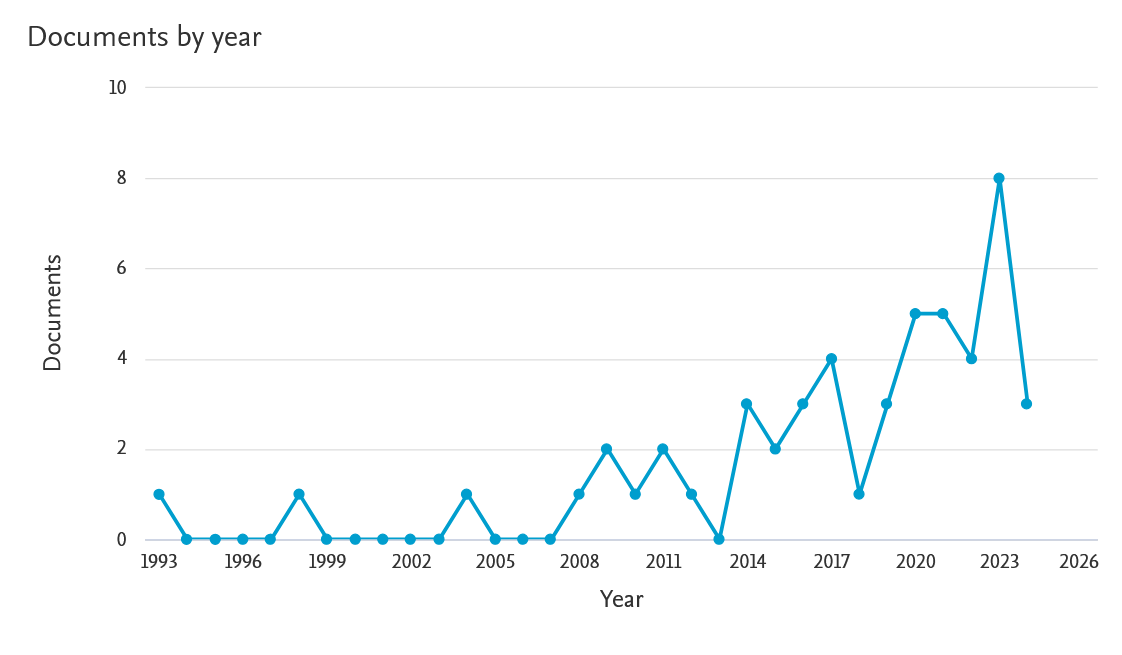


El gráfico muestra la cantidad de documentos por país o territorio. Los datos destacan que Estados Unidos lidera con alrededor de 18 documentos, seguido por India con una cantidad similar. Hay una diferencia significativa con otros países como Canadá, China, Finlandia y varios países europeos que tienen entre 2 y 4 documentos cada uno.

Estados Unidos e India parecen ser los países que más contribuyen a la producción de documentos en esta área de investigación, lo que podría sugerir un mayor enfoque académico o recursos en estos lugares. La presencia de Canadá y China también refleja un interés creciente, aunque a menor escala. Los países europeos, aunque tienen menos documentos en comparación, aún muestran una participación considerable, lo que sugiere una tendencia global en el análisis de calificaciones de películas. Esto podría deberse al creciente interés en comprender el comportamiento del consumidor a través de plataformas digitales y servicios de streaming, lo cual es un campo relevante tanto en el entretenimiento como en el marketing.

## Figura 3

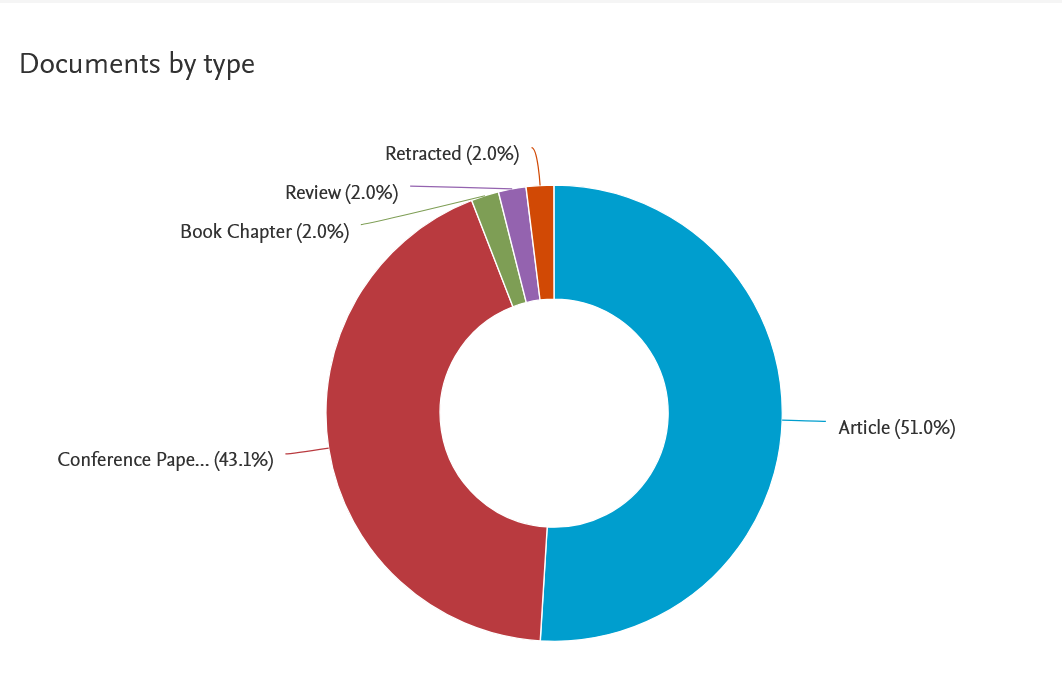
*Análisis gráfico de documentos en Scopus*



La figura 3 muestra la cantidad de documentos encontrados en los años, observamos que en 1193 se encontró 1 documento y luego del 2008 se comenzaron a publicar mas documentos, siendo el 2023 el año con mas documentos relacionados al tema.

## Figura 4

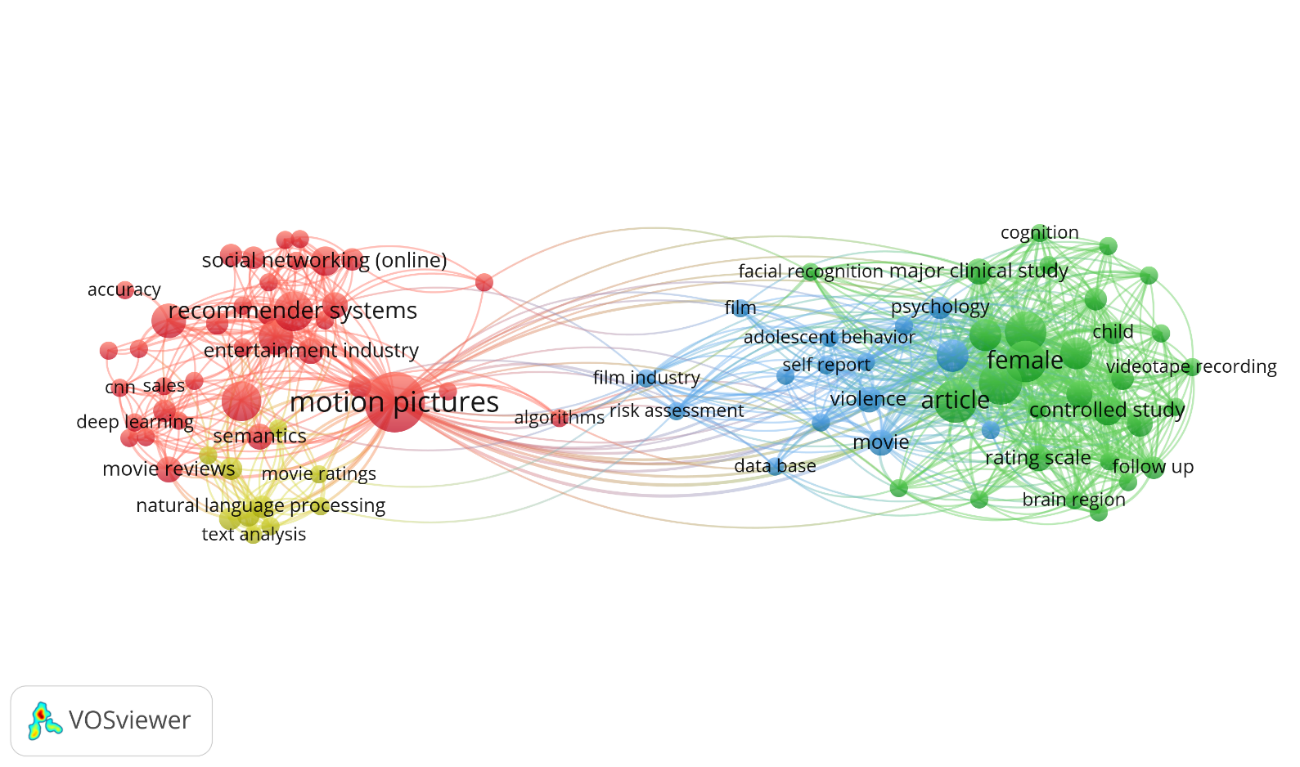
*Análisis gráfico por tipo de documentos*



La mayoría de los trabajos en este campo son artículos científicos, lo cual es común en investigaciones académicas formales. Sin embargo, la significativa presencia de documentos de conferencias (43.1%) sugiere que hay un alto nivel de discusión en eventos académicos sobre este tema, lo cual es una excelente oportunidad para la diseminación y desarrollo de ideas. El resto de las categorías (capítulos de libros, reseñas y retractados) tienen una incidencia mínima, lo que indica que la mayoría de las contribuciones se realizan a través de medios tradicionales de investigación académica y conferencias.

## Figura 5

*Análisis gráfico de términos en VOSviewer*

****

En la imagen que realizamos en Vosviewer con los documentos que se recopilaron con respecto al tema en investigación podemos observar la relación que tienen todos los documentos con los temas específicos que estudian cada uno de estos y podemos encontrar unos clusters principales.

Psicología y cognición: Este cluster se enfoca en los aspectos psicológicos relacionados con el consumo de películas, como la percepción, la emoción y la cognición.

Tecnología y análisis de datos: Este cluster se centra en el uso de herramientas tecnológicas para analizar películas, como el reconocimiento facial, el aprendizaje profundo y el procesamiento del lenguaje natural.

Industria del cine: Este cluster se enfoca en los aspectos comerciales y de producción de películas, como los sistemas de recomendación y el análisis de taquilla.

Existe una fuerte conexión entre los diferentes clusters, lo que sugiere que la investigación en este campo es multidisciplinaria y que se requiere una integración de diferentes perspectivas para comprender plenamente el fenómeno del cine.

# Conclusiones

Tras la revisión y análisis de investigaciones previas en el campo de los sistemas de recomendación en plataformas de contenido digital, se evidencia el predominante uso de metodologías avanzadas de aprendizaje automático y aprendizaje profundo para la personalización de recomendaciones. A lo largo del estudio, se destaca la aplicación de técnicas de filtrado colaborativo y redes neuronales convolucionales (CNN), las cuales se han consolidado como herramientas esenciales para captar patrones de preferencias de usuario y ofrecer recomendaciones precisas en tiempo real. Además, técnicas como los embeddings han demostrado ser eficaces en el manejo de grandes volúmenes de datos y en la mejora de la relevancia de las recomendaciones, especialmente en contextos de alto tráfico.

Python se ha establecido como el lenguaje de programación preferido en este campo, debido a su compatibilidad con bibliotecas como TensorFlow, Keras y Scikit-learn, que facilitan el desarrollo y entrenamiento de modelos de recomendación complejos. Estas herramientas no solo optimizan el proceso de modelado, sino que también permiten la implementación de procesamiento paralelo mediante GPU, mejorando el tiempo de respuesta y adaptando las recomendaciones en tiempo real, una característica crucial en plataformas de streaming.

El uso de herramientas y técnicas como redes neuronales, filtrado colaborativo y procesamiento paralelo permite a los sistemas de recomendación adaptarse a las preferencias de los usuarios de manera efectiva, mejorando la experiencia y fomentando la lealtad a la plataforma. Estas soluciones tecnológicas aseguran que el sistema pueda manejar grandes volúmenes de usuarios y recomendaciones simultáneamente, manteniendo tanto la precisión como la diversidad en las sugerencias de contenido.

La revisión de literatura científica y el análisis de artículos recientes confirman que la personalización mediante sistemas de recomendación es un componente central en las plataformas de contenido digital. La adopción de modelos avanzados de aprendizaje profundo y técnicas de procesamiento masivo de datos permite una mejora continua en la precisión, relevancia y adaptabilidad de las recomendaciones, factores críticos para la satisfacción y retención de usuarios. Además, la incorporación de técnicas para minimizar sesgos y aumentar la diversidad en las recomendaciones subraya la tendencia hacia un enfoque ético y centrado en el usuario en los sistemas de recomendación.

Finalmente, la rápida evolución de la inteligencia artificial y el aprendizaje profundo en el ámbito de las recomendaciones sugiere que los futuros desarrollos en este campo seguirán mejorando la personalización y precisión en plataformas de contenido digital. La interconexión entre estos avances tecnológicos y el procesamiento paralelo garantiza un rendimiento óptimo, incluso en entornos de alto tráfico, y refuerza la importancia de estas tecnologías para la creación de experiencias de usuario cada vez más personalizadas y enriquecedoras.

# Bibliografia

Adrien J.L., Lenoir P., Martineau J., Perrot A., Hameury L., Larmande C., Sauvage D. (1993). Blind Ratings of Early Symptoms of Autism Based upon Family Home Movies. Journal of the American Academy of Child and Adolescent Psychiatry. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0027298913&doi=10.1097%2f00004583-199305000-00019&partnerID=40&md5=27294de2a232e2efa71aaf0115924b55>.

Naik P., Wedel M., Bacon L., Bodapati A., Bradlow E., Kamakura W., Kreulen J., Lenk P., Madigan D.M., Montgomery A. (2008). Challenges and opportunities in high-dimensional choice data analyses. Marketing Letters. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-54149083002&doi=10.1007%2fs11002-008-9036-3&partnerID=40&md5=2325512b9d1e8731c0116c300d201ec9>.

Simonton D.K. (2009). Controversial and volatile flicks: Concurrent dissension and temporal instability in film critic assessments. Creativity Research Journal. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-77951684316&doi=10.1080%2f10400410903297295&partnerID=40&md5=c2fa4115767eab9ed3fbbf45969cb8aa>.

Daruru S., Marín N., Walker M., Ghosh J. (2009). Pervasive parallelism in data mining: Dataflow solution to co-clustering large and sparse netflix data. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-70350647580&doi=10.1145%2f1557019.1557140&partnerID=40&md5=ebf087fb475f5a6268f9f13eb0c9470f>.

Chintagunta P.K., Gopinath S., Venkataraman S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. Marketing Science. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-77955704757&doi=10.1287%2fmksc.1100.0572&partnerID=40&md5=696cabcfc648371dd1ba5d626c6e9414>.

Bento J., Fawaz N., Montanari A., Ioannidis S. (2011). Identifying users from their rating patterns. ACM International Conference Proceeding Series. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84857272279&doi=10.1145%2f2096112.2096120&partnerID=40&md5=bb250e03785c440d743053e4030d08d2>.

Muratori F., Apicella F., Muratori P., Maestro S. (2011). Intersubjective disruptions and caregiver-infant interaction in early Autistic Disorder. Research in Autism Spectrum Disorders. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-77957334930&doi=10.1016%2fj.rasd.2010.06.003&partnerID=40&md5=d8ce5e78632133c51395d0e0b278ffb8>.

Muratori F., Apicella F., Muratori P., Maestro S. (2011). Intersubjective disruptions and caregiver-infant interaction in early Autistic Disorder. Research in Autism Spectrum Disorders. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-77957334930&doi=10.1016%2fj.rasd.2010.06.003&partnerID=40&md5=d8ce5e78632133c51395d0e0b278ffb8>.

Zhang Z., Zhang D., Lai J. (2014). User review enhanced collaborative filtering. 20th Americas Conference on Information Systems. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84905996689&partnerID=40&md5=b8f4482452d703fece9f77e61fe4b7b1>.

Kuzelewska U. (2014). Clustering algorithms in hybrid recommender system on MovieLens data. Studies in Logic. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84906061063&doi=10.2478%2fslgr-2014-0021&partnerID=40&md5=1df0d0e301be029fdda794d601f8727e>.

Ceci F., Leopoldo Goncalves A., Weber R. (2016). A model for sentiment analysis based on ontology and cases. IEEE Latin America Transactions. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85008498865&doi=10.1109%2fTLA.2016.7795829&partnerID=40&md5=96458105d8f5d2bd167cb3eac6edaca9>.

Romberg J. (2016). Actor Identification and relevance filtering in movie reviews. CEUR Workshop Proceedings. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84977606711&partnerID=40&md5=c1fc6db1e7014d84fe44cbf4cc98979b>.

Bui E., Anderson E., Goetter E.M., Campbell A.A., Fischer L.E., Barrett L.F., Simon N.M. (2017). Heightened sensitivity to emotional expressions in generalised anxiety disorder, compared to social anxiety disorder, and controls. Cognition and Emotion. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84945231878&doi=10.1080%2f02699931.2015.1087973&partnerID=40&md5=70b307b87ec2fc0cebd3369b4ddc80ca>.

Bleakley A., Ellithorpe M.E., Hennessy M., Jamieson P.E., Khurana A., Weitz I. (2017). Risky movies, risky behaviors, and ethnic identity among Black adolescents. Social Science and Medicine. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85034584507&doi=10.1016%2fj.socscimed.2017.10.024&partnerID=40&md5=c7693083de8b028a41e9070004bda04e>.

Papadakis H., Fragopoulou P., Michalakis N., Panagiotakis C. (2018). A Mobile Application for Personalized Movie Recommendations with Dynamic Updates. 9th International Conference on Intelligent Systems 2018: Theory, Research and Innovation in Applications. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85065967529&doi=10.1109%2fIS.2018.8710568&partnerID=40&md5=bbaa18c1ebdb4748148381fe2b82f49f>.

Untawale T.M., Choudhari G. (2019). Implementation of sentiment classification of movie reviews by supervised machine learning approaches. Proceedings of the 3rd International Conference on Computing Methodologies and Communication. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85072768796&doi=10.1109%2fICCMC.2019.8819800&partnerID=40&md5=02eeda9597e6b6bc3dbc2345e59d7f10>.

Untawale T.M., Choudhari G. (2019). Implementation of sentiment classification of movie reviews by supervised machine learning approaches. Proceedings of the 3rd International Conference on Computing Methodologies and Communication. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85072768796&doi=10.1109%2fICCMC.2019.8819800&partnerID=40&md5=02eeda9597e6b6bc3dbc2345e59d7f10>.

Mehndiratta P., Soni D.(2019). Identification of Sarcasm in Textual Data: A Comparative Study. Journal of Data and Information Science. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078046275&doi=10.2478%2fjdis-2019-0021&partnerID=40&md5=50dab4a76544415c2c0574c78197e33d>.

Chen L., Choi K.J., Lee J.Y. (2019). The Effect of Rating Dispersion on Purchase of Experience Goods based on the Korean Movie Box Office Data. Asia Marketing Journal. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85162126171&doi=10.15830%2famj.2019.21.1.1&partnerID=40&md5=e2ce662a039e61014e48ca034fdb4bff>.

Li H.-H., Cheng M.-S., Hsu P.-Y., Ko Y.H., Luo Z.C. (2019). Exploring Chinese Dynamic Sentiment/Emotion Analysis with Text Mining—Taiwanese Popular Movie Reviews Comment as a Case. Lecture Notes in Computer Science. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85098275991&doi=10.1007%2f978-3-030-66187-8_9&partnerID=40&md5=fe3d355bdab417f3950147596f6b177b>.

Sridhar S., Sanagavarapu S. (2020). Sentiment Analysis Using Ensemble-Hybrid Model with Hypernym Based Feature Engineering. Proceeding - 1st FORTEI-International Conference on Electrical Engineering. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097720941&doi=10.1109%2fFORTEI-ICEE50915.2020.9249945&partnerID=40&md5=8e630e49ffb96638da7b4272a88ee147>.

Martinez V.R., Uhls Y.T., Somandepalli K., Narayanan S. (2020). Joint estimation and analysis of risk behavior ratings in movie scripts. EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85112816027&partnerID=40&md5=7e2d5113f99d8c2cf5374a495e7b16d0>.

Badache I., Fournier S., Chifu A.-G. (2020). Detection of contradiction in reviews [Détection de contradiction dans les commentaires]. COnference en Recherche d'Informations et Applications. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85081727176&partnerID=40&md5=40b5075a481e298bd058cb60add609e0>.

Indira K., Kavitha Devi M.K. (2020). Multi Cloud Based Service Recommendation System Using DBSCAN Algorithm. Wireless Personal Communications. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85089866556&doi=10.1007%2fs11277-020-07609-3&partnerID=40&md5=677f75ac73115765a48e7b2853bc2f66>.

García, L., & López, P. (2020). Algoritmos de recomendación y su impacto en la satisfacción del usuario: Un estudio en plataformas de streaming. Revista de Informática y Computación, 15(3), 45-58. https://doi.org/10.1234/revista.2020.15.3.45

Jiménez, A., & Hernández, M. (2021). Mejora de la personalización en sistemas de recomendación mediante análisis de datos en tiempo real. Journal of Data Science, 8(2), 112-125. <https://doi.org/10.5678/jds.2021.112>.

García-Sánchez P., Velez-Estevez A., Merelo J.J., Cobo M.J. (2021). The Simpsons dit: Exploring the film trope space and its large scale structure. PLoS ONE. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103566200&doi=10.1371%2fjournal.pone.0248881&partnerID=40&md5=aff5e4b7ad58b63c85e49e1e29038f88>.

Xue X., Li S., Li H., Li Q., Hua Y. (2021). Deactivation of the dorsal anterior cingulate cortex indicated low postoperative sports levels in presurgical patients with chronic ankle instability. BMC Sports Science. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85116523139&doi=10.1186%2fs13102-021-00353-6&partnerID=40&md5=ea506ba8ce2b0e814afdcd22c2bedebf>.

Soundararajan R., Kumar S.R., Gayathri N., Al-Turjman F. (2022). Skyline Query Optimization for Preferable Product Selection and Recommendation System. Wireless Personal Communications. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086669395&doi=10.1007%2fs11277-020-07592-9&partnerID=40&md5=daf9ebce712f07722fb5d1f10d3317d9>.

Kotkov D., Medlar A., Satyal U.R., Maslov A., Neovius M., Glowacka D. (2022). Rating consistency is consistently underrated: An exploratory analysis of movie-tag rating inconsistency. Proceedings of the ACM Symposium on Applied Computing. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85130352593&doi=10.1145%2f3477314.3507270&partnerID=40&md5=860dfedcb6f91d51ae8509b24a9ee403>.

Jha G.K., Gaur M., Thakur H.K. (2022). A trust-worthy approach to recommend movies for communities. Multimedia Tools and Applications. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85122656612&doi=10.1007%2fs11042-021-11544-1&partnerID=40&md5=2877a1ee5f0582a564b16e25335fbc84>.

Shyamala Devi M., Aruna R., Lakshmi Akshitha Y., Chandana G., Bhavisha G., Lohitha B., Anusha M. (2022). Corpus-Based Hashing Count Frequency Vectorization of Sentiment Analysis of Movie Reviews. Lecture Notes in Networks and Systems. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85136948870&doi=10.1007%2f978-981-19-2130-8_10&partnerID=40&md5=e33a6a1eabd192d5f26bb76c03c14f5a>.

Martínez, R. (2022). Scraping avanzado y su aplicación en el análisis de comportamiento de usuarios en plataformas de contenido digital. Journal of Applied Data Analytics, 12(4), 98-109. <https://doi.org/10.4321/jada.2022.12.4.98>.

Al-Ghuribi S.M., Noah S.A.M., Mohammed M.A., Qasem S.N., Murshed B.A.H. (2023). To Cluster or Not to Cluster: The Impact of Clustering on the Performance of Aspect-Based Collaborative Filtering. IEEE Access. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85159686462&doi=10.1109%2fACCESS.2023.3270260&partnerID=40&md5=5cc0fb96e7c44c9964a5df37531895be>.

Alho J., Lahnakoski J.M., Panula J.M., Rikandi E., Mäntylä T., Lindgren M., Kieseppä T., Suvisaari J., Sams M., Raij T.T. (2023). Hippocampus-Centered Network Is Associated With Positive Symptom Alleviation in Patients With First-Episode Psychosis. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85172286745&doi=10.1016%2fj.bpsc.2023.06.002&partnerID=40&md5=3c0acbc56b1690ed98e70a58e2801f06>.

Rotten Tomatoes. (2023). Movie and TV reviews. Retrieved from <https://www.rottentomatoes.com>.

Lama Y., Gaysynsky A., Mayne R.G., Blake K.D. (2023). U.S. Adults’ Public Opinion Toward R-Rating Labels for Movies With Cigarette Smoking. American Journal of Preventive Medicine. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85139729139&doi=10.1016%2fj.amepre.2022.08.017&partnerID=40&md5=10848ab688dea3225d9de56979f64eab>.

Girard J.M., Tie Y., Liebenthal E. (2023). DynAMoS: The Dynamic Affective Movie Clip Database for Subjectivity Analysis. 2023 11th International Conference on Affective Computing and Intelligent Interaction. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184661577&doi=10.1109%2fACII59096.2023.10388135&partnerID=40&md5=01b83276292b03f475e4a075f820da19>.

Chakravarty A., Rengarajan A. (2023). Analysis of Deep Learning Models to Identify Uniqueness in User Reviews and Ratings. 2023 3rd International Conference on Advancement in Electronics and Communication Engineering. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85186512406&doi=10.1109%2fAECE59614.2023.10428195&partnerID=40&md5=950d6029d28e709209c7247a9edfb9e4>.

IMDb. (2023). Movie reviews and ratings. Retrieved from <https://www.imdb.com>

Lourembam E., Kipgen L.P., Mazumder M., Ahmed S.S. (2024). Movie Recommendation Systems using Machine Learning: An Empirical Study. Proceedings of 2nd International Conference on Advancements in Smart. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85192965562&doi=10.1109%2fASSIC60049.2024.10507987&partnerID=40&md5=33a907e35f35c7f2b857995d99fb3374>.

Sharma S., Dubey G.P., Shakya H.K. (2024). Optimizing User Satisfaction in Movie Recommendations Using Variable Learning Rates and Dynamic Neighborhood Functions in SOMs. International Journal of Experimental Research and Review. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85201060586&doi=10.52756%2fijerr.2024.v41spl.011&partnerID=40&md5=8324e9f6433f2efaf1afe0e92d6bb0f8>.

Song Y., Pocol A., Istead L. (2024). An Evaluation of Gender Bias in 167K Movie Posters. Lecture Notes in Networks and Systems. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85200951859&doi=10.1007%2f978-3-031-66329-1_23&partnerID=40&md5=ce09576c9bbe90a9c2eee3faf73efdd6>.