

**INSTITUTO TECNOLÓGICO DE BUENOS AIRES – ITBA**  
**MAESTRIA EN CIENCIA DE DATOS**

## **Proyecto de Tesis**

# **Sistema de Recomendación de Videojuegos Híbrido Integrando Comportamiento de los Usuarios, Características de los Juegos y Valoraciones de la Comunidad**

**ALUMNO:** Flores, Matias Gabriel (Leg. N° 105811)

**DIRECTOR:** Abalde, Roberto

**TALLER DE TESIS II**

**PROFESORES:** Mon, Alicia

**BUENOS AIRES**  
**PRIMER CUATRIMESTRE, 2025**

## Índice

1. Introducción	2
2. Planteamiento del Problema	2
2.1. Problema	2
2.2. Propuesta	3
2.3. Objetivo General	3
2.4. Objetivos Específicos	4
2.5. Preguntas de Investigación	4
3. Estado de la Cuestión	4
4. Hipótesis y Variables	6
4.1. Hipótesis	6
4.2. Variables	6
5. Metodología	7
5.1. Entendimiento del Negocio	7
5.2. Entendimiento de los Datos	7
5.3. Preparación de los Datos	7
5.4. Modelado	8
5.5. Evaluación de los Modelos, Resultados y Conclusiones:	8
6. Cronograma de Actividades	8
7. Referencias - Bibliografía	10

## Palabras Clave:

Sistemas de Recomendación – Ciencia de Datos – Filtrado Colaborativo – Basado en Contenidos – Recomendaciones Híbridas – Videojuegos – Análisis de Sentimiento

## **1. Introducción**

Durante los últimos años, el crecimiento de plataformas como YouTube, Netflix, Spotify y muchas otras, ha traído consigo una integración cada vez más profunda de sistemas de recomendación en la vida diaria de los usuarios. Estos algoritmos son responsables de sugerir productos o contenidos que podrían resultar de interés para cada persona en función de su historial de interacciones y preferencias pasadas. Su uso ha crecido exponencialmente y hoy en día se encuentran en aplicaciones de diversos tipos, desde redes sociales como Facebook e Instagram, hasta plataformas de streaming como Netflix y HBO Max, así como en sitios de comercio electrónico como Amazon y MercadoLibre.

La implementación de estos sistemas responde al problema de la sobrecarga de opciones. A medida que aumenta la cantidad de información digital y el número de usuarios en Internet, también lo hace el número de opciones entre las cuales los consumidores deben elegir. Este exceso de posibilidades puede dificultar las decisiones de los usuarios y sobrecargar su experiencia. Un sistema de recomendación permite, entonces, filtrar esta gran cantidad de opciones, destacando únicamente las más relevantes para cada usuario.

En el contexto de los videojuegos, esta problemática es igualmente significativa. Plataformas de videojuegos, como Steam, ofrecen miles de títulos en diversas categorías y géneros, lo que hace que los jugadores puedan encontrarse abrumados al intentar descubrir juegos nuevos que se relacionen con sus gustos y preferencias.

Los sistemas de recomendación actuales en estas plataformas, aunque son útiles, se basan en filtrado colaborativo, un enfoque que utiliza el comportamiento de otros usuarios con intereses similares para hacer recomendaciones. Sin embargo, este tipo de recomendación deja de lado otros aspectos valiosos, como los atributos específicos de los videojuegos (género, desarrollador, fecha de lanzamiento) y las opiniones de la comunidad expresadas en las reseñas. Por ello, surge la necesidad de explorar un sistema de recomendación híbrido que integre múltiples fuentes de información: el comportamiento histórico de los usuarios, los atributos de los videojuegos y el análisis de las opiniones en las reseñas. Este enfoque permitirá obtener recomendaciones más precisas, relevantes y ajustadas a las expectativas de los jugadores.

Además, la inclusión de un modelo de análisis de sentimiento en las reseñas representa una innovación que agrega una capa emocional a las recomendaciones, permitiendo a los usuarios descubrir juegos bien valorados por la comunidad y, en última instancia, mejorar su experiencia de usuario en la plataforma. Por estas razones, este estudio no solo busca mejorar la efectividad de las recomendaciones en plataformas de videojuegos, sino también contribuir al sector del entretenimiento, integrando técnicas avanzadas de procesamiento de lenguaje natural y aprendizaje automático.

## **2. Planteamiento del Problema**

### **2.1. Problema**

Las plataformas de videojuegos se han convertido en puntos clave de interacción para millones de jugadores en todo el mundo. Para mejorar la experiencia del usuario y aumentar la retención, estas

plataformas recurren a sistemas de recomendación que sugieren juegos en función de las preferencias individuales de cada usuario. Sin embargo, los sistemas de recomendación actuales en plataformas como Steam dependen principalmente del filtrado colaborativo, una técnica que basa sus sugerencias en las interacciones pasadas de los usuarios. Este enfoque identifica patrones de comportamiento entre usuarios similares, permitiendo predecir qué juegos pueden interesarles. Aunque el filtrado colaborativo ha demostrado ser efectivo en varios contextos, tiene limitaciones clave que afectan su efectividad en la recomendación de videojuegos.

Entre estas limitaciones se destaca la dependencia del historial de interacciones, que dificulta la recomendación para nuevos usuarios (problema de arranque en frío) o de juegos que recién se lanzan y carecen de datos de interacción suficientes. Además, el filtrado colaborativo no explota otras fuentes de información valiosas que podrían enriquecer la experiencia del usuario, tales como los atributos específicos de los juegos y las reseñas y opiniones de la comunidad. Estas reseñas reflejan no solo las preferencias de los jugadores, sino también sus emociones y opiniones sobre la calidad de los juegos. Ignorar estos elementos deja fuera un contexto relevante que podría hacer las recomendaciones más precisas y adaptadas a los gustos y expectativas de cada usuario.

## 2.2. Propuesta

Para superar estas limitaciones y mejorar la precisión de las recomendaciones, se propone el desarrollo de un sistema de recomendación híbrido que combine tres enfoques:

- **Filtrado colaborativo:** Este modelo se basará en el comportamiento histórico de los usuarios, aprovechando las interacciones pasadas para detectar similitudes en las preferencias entre usuarios. Al incorporar técnicas de filtrado colaborativo, el sistema podrá ofrecer recomendaciones personalizadas basadas en patrones de comportamiento compartidos.
- **Modelo basado en contenido:** Para complementar al filtrado colaborativo, se integrará un modelo que utilice los atributos específicos de los videojuegos, tales como el género, los desarrolladores y las etiquetas asociadas al juego. Al considerar las características de cada juego, el sistema podrá recomendar títulos que se asemejen a los que el usuario ha disfrutado previamente, incluso si no hay datos de interacción directa.
- **Análisis de sentimiento:** Finalmente, el sistema incluirá un componente de análisis de sentimiento, que utilizará técnicas de procesamiento del lenguaje natural (NLP) para analizar las reseñas de la comunidad. Este análisis capturará el tono emocional de las reseñas (positivo, negativo o neutro), ajustando las recomendaciones en función de la percepción general de la comunidad. De esta manera, se priorizarán los juegos con mejor valoración emocional, lo cual puede influir significativamente en la decisión de compra del usuario.

## 2.3. Objetivo General

Desarrollar un sistema de recomendación híbrido para la plataforma de videojuegos que combine filtrado colaborativo, análisis de atributos de contenido y análisis de sentimiento en las reseñas de los usuarios,

con el fin de mejorar la precisión y relevancia de las recomendaciones, optimizando así la experiencia del usuario.

## **2.4. Objetivos Específicos**

1. Implementar un modelo de filtrado colaborativo que permita analizar patrones de interacción entre usuarios y videojuegos, identificando los gustos en común entre ellos y generando recomendaciones en función del comportamiento histórico de los usuarios.
2. Desarrollar un modelo basado en contenido que utilice los atributos específicos de cada videojuego, tales como género, desarrollador y características, para recomendar títulos similares a los que el usuario consumió previamente.
3. Incorporar un módulo de análisis de sentimiento que utilice técnicas de procesamiento de lenguaje natural (NLP) para analizar las reseñas de los usuarios, capturando el tono emocional (positivo, negativo o neutro) y ajustando las recomendaciones en función de la percepción general de la comunidad.
4. Implementar un modelo híbrido que combine todos los modelos mencionados anteriormente.
5. Comparar la efectividad del sistema híbrido frente a un sistema de recomendación basado exclusivamente en filtrado colaborativo, evaluando su rendimiento mediante métricas de exactitud, precisión, recall y F1-score.

## **2.5. Preguntas de Investigación**

1. ¿Cómo influye la integración de modelos basados en contenido y análisis de sentimiento en la mejora de las recomendaciones, en comparación con un enfoque centrado únicamente en filtrado colaborativo?
2. ¿Qué contribución específica realiza cada módulo del sistema híbrido al desempeño global del recomendador?
3. ¿Qué desafíos y limitaciones surgen al aplicar técnicas de procesamiento de lenguaje natural para analizar reseñas de usuarios en videojuegos?
4. ¿Qué grado de alcance logra el sistema propuesto sobre el catálogo de videojuegos disponible, y cómo se vincula este resultado con la diversidad de las recomendaciones generadas?
5. ¿Hasta qué punto la combinación de distintos enfoques mejora la precisión del modelo frente a estrategias tradicionales?

## **3. Estado de la Cuestión**

Los sistemas de recomendación han experimentado un notable crecimiento en los últimos años, siendo

implementados en una amplia variedad de sectores, incluyendo plataformas de contenido, comercio electrónico y videojuegos. Estos sistemas buscan proporcionar recomendaciones personalizadas al usuario mediante el análisis de sus preferencias y comportamiento histórico. Sin embargo, debido a la complejidad de las interacciones y al volumen de datos que manejan, los algoritmos de recomendación han enfrentado limitaciones en términos de escalabilidad, precisión y relevancia.

Por un lado, el filtrado colaborativo es uno de los enfoques más comunes en sistemas de recomendación. Según Aggarwal (2016) [1], este método se basa en identificar patrones de comportamiento entre usuarios con gustos similares, generando recomendaciones en función de interacciones previas. Sin embargo, el filtrado colaborativo presenta limitaciones importantes, como la dependencia de datos históricos, la susceptibilidad al problema del "arranque en frío" y la dificultad para escalar en plataformas con grandes volúmenes de usuarios e ítems (Cacheda et al., 2011) [2].

Koren, Bell y Volinsky (2009) [6] destacan las técnicas de factorización de matrices como una mejora en el filtrado colaborativo, ya que permiten descomponer la matriz de interacciones usuario-ítem en vectores latentes que capturan relaciones complejas. Esta técnica ha demostrado ser efectiva en la generación de recomendaciones, particularmente en el caso de grandes volúmenes de datos. Sin embargo, aunque la factorización de matrices optimiza la precisión, sigue siendo limitada en términos de contexto y atributos específicos de los ítems.

Por otro lado, la recomendación basada en contenido es una técnica que considera los atributos de los ítems para sugerir aquellos que se alineen con las preferencias del usuario. Este enfoque se ha vuelto especialmente relevante en contextos como el de los videojuegos, donde características como el género y desarrollador juegan un papel importante en las decisiones de los usuarios (Kane, 2018) [7]. Este tipo de recomendación complementa al filtrado colaborativo al permitir recomendaciones en escenarios de "arranque en frío" o cuando se tienen pocos datos de interacción de los usuarios.

A pesar de esto, la recomendación basada en contenido también tiene sus desafíos, especialmente en la identificación de atributos que realmente sean significativos para la experiencia del usuario. Según Çano y Morisio (2017) [5], la combinación de filtrado colaborativo y modelos basados en contenido en un sistema híbrido puede mejorar la precisión y la personalización de las recomendaciones, aprovechando tanto las preferencias de los usuarios como las características de los ítems.

El análisis de las opiniones de los usuarios, como las reseñas en línea, es otra dimensión que ha cobrado relevancia en sistemas de recomendación. Cox y Keinman (2015) [3] señalan que, en la industria de los videojuegos, tanto las críticas profesionales como las reseñas de los usuarios influyen notablemente en las decisiones de compra. Las reseñas positivas y la percepción general del juego pueden hacer que un título sea más atractivo, mientras que las reseñas negativas pueden reducir su popularidad.

El análisis de sentimiento, que clasifica las opiniones de los usuarios como positivas, negativas o neutrales, permite capturar el contexto emocional detrás de las interacciones con los ítems (Dang et al., 2021) [7]. Este enfoque, cuando se integra en un modelo de recomendación, puede aumentar la relevancia de las recomendaciones al priorizar productos que generan experiencias positivas entre los usuarios. El análisis de sentimiento también puede mejorar la experiencia del usuario al recomendar ítems que han recibido una buena valoración emocional en la comunidad, ayudando a los usuarios a evitar productos que otros consideran de baja calidad o poco satisfactorios.

La evaluación de sistemas de recomendación es esencial para validar su efectividad y relevancia. Herlocker et al. (2004) [4] proponen métricas como la precisión, el recall y el F1-score para evaluar los modelos de recomendación en términos de su capacidad para predecir correctamente las preferencias de los usuarios. Chen y Liu (2017) [8] destacan que las métricas de precisión y recall son especialmente útiles en modelos híbridos, ya que miden la efectividad del sistema en identificar y recomendar los ítems que son relevantes para el usuario.

La combinación de estos enfoques en un sistema de recomendación híbrido representa una solución innovadora que responde a las necesidades actuales de personalización en plataformas como Steam. Aprovechar la intersección entre el filtrado colaborativo, el modelo basado en contenido y el análisis de sentimiento permite desarrollar un sistema de recomendación más robusto y adaptativo, capaz de capturar no solo las preferencias explícitas e implícitas de los usuarios, sino también el contexto emocional que puede influir en su decisión de compra. Esta evolución en los sistemas de recomendación promete optimizar la experiencia del usuario y, en última instancia, contribuir al crecimiento de la industria de los videojuegos.

## **4. Hipótesis y Variables**

### **4.1. Hipótesis**

La integración de un sistema híbrido que combine filtrado colaborativo, los atributos de los juegos y análisis de sentimiento en las reseñas mejora significativamente la precisión y relevancia de las recomendaciones en comparación con sistemas que utilizan solo filtrado colaborativo.

### **4.2. Variables**

- **Calificación de Usuarios sobre Videojuegos** (Dependiente): Si recomienda o no el Videojuego.
- **Reviews de Usuarios sobre Videojuegos** (Independiente).
- **Atributos del Videojuego** (Independiente):
  - **Géneros.**
  - **Tags** (Palabras clave que representan características del juego).
  - **Precio.**
  - **Fecha de Lanzamiento.**
  - **Modalidad de juego.**
- **Cantidad de Videojuegos Consumidos** (Independiente).

## 5. Metodología

### 5.1. Entendimiento del Negocio

La implementación de un sistema de recomendación efectivo es fundamental para plataformas de videojuegos que desean optimizar la experiencia del usuario, mejorar la precisión de las recomendaciones y aumentar la retención de sus clientes. Este estudio, centrado en un sistema de recomendación híbrido, busca los siguientes objetivos de negocio:

- Aumentar la relevancia y precisión de las recomendaciones al proporcionar sugerencias adaptadas a los intereses y emociones de los usuarios.
- Incrementar la satisfacción y retención de los usuarios, mejorando su experiencia de navegación y descubrimiento de nuevos videojuegos en la plataforma.
- Fomentar la adquisición de nuevos videojuegos a través de recomendaciones más personalizadas que incrementen la intención de compra.

### 5.2. Entendimiento de los Datos

En esta fase, se exploran y comprenden los datos disponibles en el dataset de usuarios y videojuegos. Las tareas clave incluyen:

- **Identificación de fuentes de datos:** Datasets de interacciones entre usuarios y videojuegos, atributos de videojuegos y reseñas.
- **Análisis exploratorio:** Comprensión de la estructura de los datos, evaluación de la distribución de atributos y detección de problemas de calidad (valores faltantes, duplicados o inconsistencias).
- **Revisión de los atributos:** Familiarización con los datos clave que alimentarán el sistema híbrido, tales como géneros, desarrolladores, fechas de lanzamiento y precios.

### 5.3. Preparación de los Datos

Una vez comprendidos los datos, se realizará una preparación para optimizar su uso en el modelo de recomendación. Esta fase incluye:

- **Limpieza de datos:** Eliminación de valores nulos o duplicados y tratamiento de inconsistencias.
- **Preprocesamiento de reseñas:** Tokenización, eliminación de stop words y transformación de las reseñas en representaciones numéricas para su análisis de sentimiento.
- **Generación de vectores de atributos:** Uso de técnicas como TF-IDF para representar los atributos del contenido y facilitar el cálculo de similitudes en el modelo basado en contenido.



#### 5.4. Modelado

- **Implementación de Modelo de Filtrado Colaborativo:** Primero se realizará la implementación de modelos de filtrado colaborativo, para intentar replicar los que se utilizan actualmente en la industria. Para el desarrollo de estos modelos, se creará una matriz de usuarios y videojuegos para determinar las interacciones entre los mismos, y luego se aplicarán técnicas de Knn y de factorización de matrices para realizar la predicción del sistema, y finalmente realizar las recomendaciones.
- **Implementación de Modelo Basado en Contenido:** Para la implementación de este algoritmo, se utilizarán los atributos de los videojuegos, y mediante técnicas como TF-IDF se determinará la relevancia de los atributos para cada videojuego. Luego, se aplicarán técnicas de similitud para detectar cuales juegos son más parecidos al juego objetivo, y luego realizar las recomendaciones.
- **Implementación de Análisis de Sentimiento:** Para realizar el análisis de sentimiento sobre las reseñas de los usuarios y mejorar la precisión de las recomendaciones, se emplearán técnicas de procesamiento del lenguaje natural (NLP) que permitirán identificar el estado emocional (positivo, negativo o neutro) de cada review. Este análisis permitirá ajustar las recomendaciones para priorizar juegos que generen experiencias positivas entre los usuarios.
- **Implementación de Modelo Híbrido:** El modelo de recomendación híbrido combinará filtrado colaborativo, modelo basado en contenido, y análisis de sentimiento para ofrecer recomendaciones personalizadas que tengan en cuenta tanto las preferencias del usuario como la percepción de la comunidad sobre cada videojuego.

#### 5.5. Evaluación de los Modelos, Resultados y Conclusiones:

La evaluación del sistema de recomendación se realizará a través de métricas de exactitud, precisión, recall y F1-score. Los resultados mostrarán si la combinación de métodos en el modelo híbrido mejora la relevancia de las recomendaciones. Finalmente, se extraerán conclusiones sobre la efectividad de cada modelo y se determinará cual es el que proporciona mejores resultados.

### 6. Cronograma de Actividades

**Etapas 1:** Investigación y Análisis Exploratorio de los Datos (2 meses)

**Etapas 2:** Implementación y Evaluación de Modelo de Filtrado Colaborativo (1 mes)

**Etapas 3:** Implementación y Evaluación de Modelo Basado en Contenidos (1 mes)

**Etapas 4:** Implementación y Evaluación de Modelo de Análisis de Sentimientos (1 mes)

**Etapas 5:** Implementación y Evaluación de Modelo Híbrido (1 mes)

**Etapas 6:** Comparación entre los modelos y modificaciones adicionales a los modelos (1 mes)

**Etapas:** 7 etapas (3 meses)

**Duración total estimada:** 9 meses

	Mes								
	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb
Etapas									
Etapas 1									
Etapas 2									
Etapas 3									
Etapas 4									
Etapas 5									
Etapas 6									
Etapas 7									

Tabla 1. Cronograma de Actividades.

## 7. Referencias - Bibliografía

- [1] Aggarwal, C. C. (2016). An Introduction to Recommender Systems. In Recommender Systems (pp. 1–28). Springer International Publishing. [https://doi.org/10.1007/978-3-319-29659-3\\_1](https://doi.org/10.1007/978-3-319-29659-3_1)
- [2] Cacheda, F., Carneiro, V., Fernández, D., & Formoso, V. (2011). Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web*, 5(1). <https://doi.org/10.1145/1921591.1921593>
- [3] Cox, J., & Kaimann, D. (2015). How do reviews from professional critics interact with other signals of product quality? Evidence from the video game industry. *Journal of Consumer Behaviour*, 14(6), 366–377. <https://doi.org/10.1002/cb.1553>
- [4] Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004, January). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*. <https://doi.org/10.1145/963770.963772>.
- [5] Çano, E., & Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*. IOS Press. <https://doi.org/10.3233/IDA-163209>
- [6] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37. <https://doi.org/10.1109/MC.2009.263>
- [7] Dang, C. N., Moreno-García, M. N., & De la Prieta, F. (2021). An approach to integrating sentiment analysis into recommender systems. *Sensors*, 21(16). <https://doi.org/10.3390/s21165666>
- [8] Chen, M., & Liu, P. (2017). Performance evaluation of recommender systems. *International Journal of Performability Engineering*, 13(8), 1246–1256. <https://doi.org/10.23940/ijpe.17.08.p7.12461256>
- [9] Kane, F. (2018). Building Recommender Systems with Machine Learning and AI. Independently published. ISBN 1718120125, 9781718120129.