



Profesor(a): Jorge Bozo
NRC: 76783



“Diseño e implementación de un sistema híbrido de recomendación turística basado en datos objetivos”

Lunes 20 de octubre de 2025



Integrantes:

Valentina Astudillo

Cesar Mella

Jean Lucas Peñaloza

Índice:

1. Introducción	3
2. Descripción del problema	4
2.1 La crisis de confianza: La erosión de la fiabilidad a través de reseñas fraudulentas.	4
2.2 El sesgo de popularidad: La falla estructural de los algoritmos	4
3. Propuesta de solución	6
3.1 Combatiendo la crisis de confianza con datos objetivos	6
3.2 Rompiendo el sesgo de popularidad con diversidad estadística	6
3.3 Ofreciendo una personalización auténtica y contextualizada	7
4. Diseño de la solución	8
4.1 Filtrado colaborativo	8
4.2. Filtrado basado en contenido	8
4.3 Filtrado híbrido	8
4.4 Modelamiento de sistema híbrido	9
Bibliografía:	11

1. Introducción

La planificación de un viaje es un proceso profundamente personal, su principal característica es que se basa en los gustos y preferencias de los viajeros, quienes a su vez sustentan la selección de destinos en recomendaciones y calificaciones de otros usuarios. La promesa subyacente es clara: la veracidad y calidad de la opinión de un par puede ser la diferencia entre una experiencia de viaje memorable y una decepcionante. Sin embargo, el ecosistema digital en el que se apoya este modelo presenta fracturas estructurales que ponen en jaque su fiabilidad.

Mediante una investigación se reveló una crisis de confianza sistémica. Plataformas clave en la toma de decisiones de los viajeros se ven inundadas por millones de reseñas falsas, diseñadas para manipular la percepción y los algoritmos con fines económicos. Esta erosión de la confianza se agrava por un desafío técnico inherente a los sistemas de recomendación: el "sesgo de popularidad". Los algoritmos, por diseño, tienden a recomendar excesivamente los destinos que ya son populares, creando un ciclo de retroalimentación que concentra a las multitudes en los mismos lugares y homogeneiza la oferta turística. El resultado es una paradoja: una herramienta pensada para personalizar la elección termina por estandarizarla, ocultando destinos menos conocidos y contribuyendo directamente al sobreturismo.

Frente a este panorama, donde la autenticidad es socavada por la desinformación y la personalización es limitada por el sesgo algorítmico, surge la necesidad de un nuevo paradigma. Este proyecto propone una alternativa radical: un sistema de recomendación que no se base en la subjetividad de las opiniones, sino en la objetividad de los datos. El objetivo es desarrollar una plataforma que utilice datos históricos y estadísticos de viajes por países para generar sus sugerencias. Al analizar patrones de viaje reales, flujos turísticos a lo largo del tiempo, y métricas cuantificables, buscamos ofrecer una herramienta más transparente, diversa y fiable, capaz de descubrir joyas ocultas y de alinear verdaderamente las recomendaciones con los intereses genuinos del viajero, más allá de las distorsiones de la popularidad digital.

2. Descripción del problema

El modelo actual de planificación de viajes digitales se enfrenta a una doble falla sistémica que compromete su fiabilidad y sostenibilidad. Por un lado, sufre una crisis de confianza derivada de la manipulación generalizada del contenido generado por el usuario; por otro, está limitado por una arquitectura algorítmica que, por diseño, fomenta la concentración y la homogeneización. Estos dos problemas se retroalimentan, creando un ecosistema que a menudo desinforma al viajero y agrava las presiones sobre los destinos turísticos.

2.1 La crisis de confianza: La erosión de la fiabilidad a través de reseñas fraudulentas.

El pilar de las plataformas de recomendación turística es la confianza en la autenticidad de las opiniones de otros viajeros. Sin embargo, este pilar está sistemáticamente erosionado por la proliferación a gran escala de reseñas falsas. La motivación es principalmente económica: una calificación positiva puede aumentar drásticamente los ingresos de un negocio, creando un incentivo poderoso para manipular las percepciones. Plataformas como Tripadvisor eliminan millones de reseñas fraudulentas anualmente, incluyendo contenido generado por los propios negocios para autoevaluarse positivamente (boosting), así como reseñas pagadas a "granjas de bots" especializadas.

Esta "armamentización" de las reseñas convierte la mayor fortaleza del sistema —la comunidad— en su mayor vulnerabilidad. Para el viajero, esto introduce una capa fundamental de incertidumbre, obligándolo a cuestionar la veracidad de la información en la que basa sus decisiones y devaluando la promesa de una recomendación auténtica y fiable.

2.2 El sesgo de popularidad: La falla estructural de los algoritmos

Más allá de la manipulación deliberada, los sistemas de recomendación actuales padecen un problema estructural conocido como "sesgo de popularidad" (popularity bias). Los algoritmos están diseñados para recomendar elementos que ya son populares, creando un ciclo de retroalimentación que se autopropaganda: un destino popular recibe más visibilidad, lo que genera más visitas y reseñas, lo que a su vez refuerza su popularidad en el sistema.

Este ciclo tiene dos consecuencias negativas directas:

2.2.1 Agravamiento del sobreturismo: Al concentrar la atención en un número reducido de "puntos calientes", los algoritmos contribuyen directamente a la



masificación y a los problemas asociados de degradación ambiental, gentrificación y presión sobre los servicios públicos.

2.2.2 Homogeneización de la experiencia: El sesgo de popularidad reduce drásticamente la diversidad de las opciones presentadas al usuario. Destinos, alojamientos y actividades menos conocidos, que podrían ofrecer experiencias más auténticas y distribuir los beneficios del turismo de manera más equitativa, permanecen invisibles, atrapados en la "larga cola" del catálogo digital.

En resumen, el problema que se busca resolver es que el actual ecosistema de recomendación turística está construido sobre una base defectuosa. Depende de opiniones subjetivas que son fácilmente manipulables y de algoritmos que, por su naturaleza, limitan la elección en lugar de expandirla. Esto crea una experiencia de planificación poco fiable para el usuario y fomenta un modelo de turismo insostenible para los destinos.

3. Propuesta de solución

La solución propuesta consiste en abandonar la dependencia de las opiniones subjetivas y a menudo manipuladas de los usuarios, para construir un sistema sobre una base de datos históricos, objetivos y estadísticamente relevantes. En lugar de preguntar "¿Qué dicen los usuarios que les gusta?", el sistema se centrará en responder "¿Cuáles son los patrones de viaje reales y demostrables a lo largo del tiempo?".

3.1 Combatiendo la crisis de confianza con datos objetivos

El sistema propuesto neutraliza el problema de las reseñas falsas al eliminar por completo su influencia en el proceso de recomendación. En lugar de basarse en calificaciones y comentarios, que son susceptibles de fraude, la plataforma analizará métricas agregadas y verificables, tales como:

1. Flujos turísticos históricos: Cantidad de visitantes por país, región o ciudad a lo largo de diferentes años y temporadas.
2. Origen de los viajeros: Patrones que conectan las preferencias de viajeros de un país con ciertos tipos de destinos.

3.2 Rompiendo el sesgo de popularidad con diversidad estadística

El modelo actual sufre de un ciclo de retroalimentación donde la popularidad genera más popularidad, concentrando a los turistas y ocultando alternativas. La solución propuesta rompe este ciclo al analizar el conjunto completo de datos estadísticos, no solo los "éxitos". Esto permite:

1. Identificar "joyas ocultas": El sistema puede detectar destinos que, aunque no tengan el volumen masivo de los "puntos calientes", muestran un crecimiento constante, altas tasas de satisfacción (medidas por la duración de la estancia o el gasto diversificado) o una fuerte popularidad entre nichos de viajeros específicos.
2. Descubrir patrones emergentes: Al analizar datos a lo largo del tiempo, el sistema puede identificar tendencias y destinos emergentes antes de que se saturen, ofreciendo al usuario la oportunidad de ser un pionero.
3. Promover una distribución más equitativa del turismo: Al dar visibilidad a una gama más amplia de destinos basados en su mérito estadístico, la plataforma



puede ayudar a desviar los flujos turísticos de las zonas sobresaturadas, contribuyendo a un modelo de turismo más sostenible.

3.3 Ofreciendo una personalización auténtica y contextualizada

Mientras que los sistemas actuales personalizan basándose en las opiniones de "usuarios similares", este modelo lo hará basándose en la correlación entre perfiles de viajeros y patrones de viaje reales. Un usuario podría definir su perfil no solo por gustos ("playa", "historia"), sino por comportamientos de viaje deseados ("viajes largos", "temporada baja", "presupuesto medio").

4. Diseño de la solución

Para el diseño de la solución, el enfoque más adecuado será un sistema híbrido, con un fuerte componente de filtrado basado en contenido.

4.1 Filtrado colaborativo

Este método se basa en la idea de que si a una persona A le gusta lo mismo que a una persona B, a la persona A también le gustará lo que le gusta a la persona B pero que aún no ha descubierto. En el turismo, esto se traduce en recomendar destinos basándose en las valoraciones o comportamientos de usuarios con gustos similares.

4.2. Filtrado basado en contenido

Este enfoque recomienda elementos basándose en sus características o atributos. En este caso, se recomendarían destinos que tengan atributos similares a las preferencias que un usuario ha definido.

- La solución se centra en datos objetivos y atributos verificables de los destinos, no en opiniones. El "contenido" de estos ítems (los destinos) sería:
 - Flujos turísticos históricos.
 - Origen de los viajeros.
 - Tasas de crecimiento.
 - Duración media de la estancia.
- Del mismo modo, el perfil del usuario se define por su "contenido", es decir, sus preferencias explícitas sobre el comportamiento de viaje: "viajes largos", "temporada baja", "presupuesto medio".
- El sistema funcionará cruzando los atributos del perfil del usuario con los atributos de los destinos, lo cual es la definición de un sistema de filtrado basado en contenido.

4.3 Filtrado híbrido

Un sistema híbrido combina dos o más técnicas de recomendación para potenciar las fortalezas y mitigar las debilidades de cada una. Esta es la opción que mejor se adapta a la sofisticación de la propuesta.

- Combina lo mejor de dos mundos: se puede usar el filtrado basado en contenido como el motor principal para hacer coincidir las preferencias del usuario (presupuesto, temporada) con los datos objetivos de los destinos.
- Reinventa el filtrado colaborativo: aunque se descartan las opiniones subjetivas, la solución busca "patrones de viaje reales". Un modelo híbrido podría implementar una forma de filtrado colaborativo basado en datos objetivos. Por ejemplo, podría agrupar usuarios no por sus *calificaciones*, sino por sus *patrones*

de comportamiento objetivos (ej: "viajeros de Chile que realizan viajes de bajo presupuesto en primavera"). Luego, el sistema podría recomendar destinos que son populares dentro de ese grupo específico y objetivo, revelando así las "joyas ocultas".

- Supera el "arranque en frío": un problema común en los sistemas colaborativos es que no pueden recomendar nada a usuarios nuevos por falta de datos . Este modelo, al tener una base de filtrado por contenido, puede ofrecer recomendaciones desde el primer momento, ya que solo necesita que el usuario defina sus preferencias de viaje.

4.4 Modelamiento de sistema híbrido

Para este sistema híbrido se plantean 2 etapas secuenciales que permitan generar una respuesta completa.:.

4.4.1 Etapa 1 (basada en contenido): El usuario introduce sus preferencias (tipo de viaje, presupuesto, temporada). El sistema filtra y presenta una lista inicial de destinos cuyos atributos objetivos (datos históricos, estacionales, etc.) coinciden con estas preferencias.

4.4.2 Etapa 2 (colaborativo-objetivo): Sobre esa lista filtrada, el sistema puede refinar las sugerencias analizando los patrones de viaje de grupos de viajeros con perfiles de comportamiento similares (definidos por datos, no por opiniones).

4.5 Arquitectura tecnológica y prototipo de validación

La arquitectura se compone de tres capas principales:

1. Capa de datos (fuente de contenido):

Se utilizará el archivo "world_tourism_economy_data.xlsx". Este archivo es la implementación directa del "Perfil de atributos objetivos". Proporciona los datos económicos y de flujo (TSA) que alimentan el motor de filtrado basado en contenido. Es la fuente de "verdad objetiva" que reemplaza a las opiniones subjetivas.

2. Capa lógica (motor de recomendación):

Uso de Python 3, utilizando las bibliotecas Pandas (para cargar, limpiar, organizar y analizar los datos) y Scikit-learn (para hacer el trabajo inteligente, agrupar usuarios, normalizar datos, etc). Esto sería el "cerebro" del sistema. Todo el diseño híbrido se ejecuta aquí.

En síntesis se utilizarían para:

- Pandas: para ingestar, limpiar y procesar los datos del archivo .xlsx, creando los atributos clave que el motor de contenido necesita.



- Python: para calcular las puntuaciones del filtro de contenido y del filtro colaborativo (reinventado).
- Scikit-learn: para la normalización de los atributos (asegurando una comparación justa) y el cálculo de similitud.

3. Capa de presentación (interfaz de validación):

Esta capa es crucial para la validación del diseño. Se contempla utilizar Streamlit ya que actúa como una interfaz de usuario interactiva que permite probar los casos de uso en tiempo real. Permite al usuario definir su "perfil de usuario" a través de *sliders* y *selectores*, enviando estas preferencias a la Capa lógica (Python) y mostrando los resultados. Su uso es fundamental para validar la solución al problema de "arranque en frío".

Bibliografía:

Streamlit • A faster way to build and share data apps. (s. f.). <https://streamlit.io/>

pandas - Python Data Analysis Library. (s. f.). <https://pandas.pydata.org/>

scikit-learn: machine learning in Python — scikit-learn 1.7.2 documentation. (s. f.).

<https://scikit-learn.org/stable/>

Murel, J., & Kavlakoglu, E. (2025, February 13). ¿Qué es el filtrado colaborativo? *Ibm.com.*

<https://www.ibm.com/es-es/think/topics/collaborative-filtering>