



Pontificia Universidad  
**JAVERIANA**  
Cali

**Recomendación personalizada de paquetes turísticos mediante aprendizaje automático con integración de datos demográficos, históricos y de preferencias**

Programa de Maestría en Ingeniería

Presentado por:

Juan Sebastián Alvear Arteaga

Dirigido por:

Juan Camilo Paz Roa (Ph.D)

Pontificia Universidad Javeriana Cali  
Facultad de Ingeniería y Ciencias  
06 de 2025

# Recomendación personalizada de paquetes turísticos mediante aprendizaje automático con integración de datos demográficos, históricos y de preferencias

Juan Sebastián Alvear<sup>1</sup>, Juan Camilo Paz<sup>1</sup>

<sup>1</sup> *Facultad de Ingeniería y Ciencias, Universidad Javeriana Cali, Cali, Colombia*

## Resumen

Actualmente la personalización de servicios es un elemento de gran diferencia competitiva. Servicios de transmisión de videos, música y redes sociales logran dicha personalización mediante modelos de recomendación que dejan al alcance del usuario productos de su mayor preferencia. Este exitoso esquema también es susceptible de ser aplicado en la industria turística, donde el uso de modelos de recomendación es escaso. Es así como los modelos de aprendizaje automático e inteligencia artificial inherentes a los modelos de recomendación se han convertido en una oportunidad estratégica para agencias de viajes que buscan aumentar la satisfacción del cliente y las ventas de sus planes ofrecidos. Este estudio propone un modelo de recomendación de paquetes turísticos que toma como entrada 17 variables con información del cliente con características demográficas, históricas y cruces de preferencias entre el cliente y los paquetes turísticos. Este primer prototipo es basado en un modelo de random forest que predice la calificación esperada de viaje a partir de estas variables. La metodología para entrenar el modelo incluyó la comparación de modelos de regresión lineal, Lasso, árbol de decisión y random forest, evaluados con métricas de error cuadrático medio (MSE), raíz del error cuadrático medio (RMSE) y error porcentual absoluto medio (MAPE). Los resultados muestran que el modelo Random Forest obtuvo el menor error (RMSE = 0.3361, MSE = 0.1130 y MAPE = 6.01), siendo seleccionado gracias a su balance entre precisión y robustez. Este trabajo confirma la viabilidad del uso de técnicas de aprendizaje automático para lograr la personalización de recomendaciones turísticas.

Palabras clave: Turismo, Recomendación, Aprendizaje Automático, Machine Learning, Inteligencia artificial.

## 1. Introducción

El sector turístico en Colombia ha sido fuertemente impactado por la tecnología en los últimos años, cambiando la forma en que es percibido por las nuevas generaciones y la promoción de paquetes turísticos desde las agencias de viajes tradicionales.

Según datos del ministerio de comercio, industria y turismo, el turismo representó aproximadamente el 2,3% del Producto interno bruto de Colombia, cifra que le permite ascender al pasar de la undécima a la novena posición, en el listado de sectores que más aportan al valor agregado nacional [1].

En cuanto a gasto de turismo receptor, la cifra brindada por el DANE se aproxima a los 50,8 Billones de pesos, representando un crecimiento del 13,8% respecto al año anterior del gasto turístico en Colombia en 2023, el 74% provino del turismo receptivo mientras que el 26% restante del doméstico. Gran parte de las utilidades de empresas turísticas locales se presentan en productos de turismo emisor (entiéndase como los viajes realizados por colombianos al exterior) [2].

La nueva tendencia de compras en el sector turístico se ha encaminado a la anticipación y satisfacción de necesidades y deseos de turistas basados en el uso de herramientas de BDA (Big Data Analytics), IoT y herramientas de minería de datos [3]. Grandes compañías turísticas como Expedia Group y Booking están usando la IA para mejorar la planificación de viajes y la experiencia de compra. Entre sus

avances destacan la planificación de viajes conversacional con IA generativa, que permite a los usuarios planificar de manera intuitiva y personalizada usando las aplicaciones móviles. Además, han lanzado una función de compra inteligente de hoteles que sugiere opciones basadas en las preferencias del usuario y las características de los destinos.

El presente trabajo busca desarrollar un modelo de recomendación de paquetes turísticos que integre analítica de datos a la gestión de operaciones de la empresa Mayorista de turismo Alta Calidad, la cual está situada en la ciudad de Santiago de Cali, con una trayectoria de 25 años en el negocio de turismo emisor. Con esta herramienta se espera mejorar el nivel de servicio, así como la interacción con clientes y tiempos de respuesta.

En este trabajo se desarrolló un modelo predictivo de afinidad entre clientes y paquetes turísticos, con base en técnicas de aprendizaje automático, aplicado a un conjunto de datos históricos recopilados en la empresa Mayorista de turismo Alta Calidad.

Para alcanzar este objetivo, se construyó una metodología basada en aprendizaje automático utilizando la biblioteca Scikit-learn. El proceso incluyó la transformación de variables categóricas mediante codificación OneHotEncoder, la evaluación de cuatro algoritmos de regresión (Regresión Lineal, Lasso, Árbol de Decisión y Random Forest), y la validación de su desempeño utilizando métricas como Error Cuadrático Medio

(MSE), Raíz del Error Cuadrático Medio (RMSE) y Error Porcentual Absoluto Medio (MAPE). La variable objetivo fue la calificación media histórica del viaje, mientras que las variables predictoras incluyeron características del cliente, tipo de plan adquirido, duración del viaje y preferencias.

Los resultados mostraron que el modelo Random Forest presentó el mejor rendimiento, con un RMSE de 0.3361 y un MAPE de 6.01%, superando a los demás modelos tanto en precisión como en robustez. El modelo fue posteriormente entrenado con el 100% de los datos disponibles y exportado como archivo pkl, permitiendo su integración con nuevas observaciones en hojas de cálculo. Este trabajo evidencia la viabilidad de aplicar modelos de recomendación basados en aprendizaje automático en el sector turístico, permitiendo generar sugerencias personalizadas, mejorar la experiencia del cliente y optimizar los ingresos de las agencias.

## 2. Revisión de literatura

Uno de los principales componentes de la inteligencia artificial dentro del marco de turismo inteligente son los sistemas de recomendación. Estos sistemas son métodos creados para proveer recomendaciones relevantes de productos, basados en gustos y comportamientos, teniendo un amplio rango de acción dentro de lo que destaca destinos, acomodaciones, actividades y atracciones locales (J. Doe et al, 2019) [3] Existen varios tipos de sistemas de recomendación usados en turismo

inteligente, los cuales los de mayor uso son: filtrado colaborativo, filtrado basado en contenido y modelos híbridos [5].

**Filtrado colaborativo:** este modelo se basa en la recolección de preferencias y comportamientos de un grupo de usuarios con el fin de generar recomendaciones asociadas. Este acercamiento supone que las preferencias pasadas podrían continuar con elecciones futuras.

Este método ha sido efectivo en la generación de recomendaciones relevantes al analizar patrones en una amplia base de usuarios. Las limitaciones con este tipo de método se basan en el arranque inicial, donde pueda existir información insuficiente de nuevos usuarios o nuevos destinos, estos factores limitan la eficacia de este sistema [6].

**Filtrado basado en contenido:** al contrario del modelo anterior, este filtrado se concentra en las características de los productos por sí mismas y sugiere productos que son comparables a clientes que ya los han encontrado atractivos previamente.

Este modelo fue desarrollado para evitar las limitaciones descritas previamente del modelo colaborativo. Este método sobresale en la entrega de sugerencias relevantes, incluso para nuevos usuarios o destinos, sin embargo, podrían llegar a ser demasiado específicos, lo que limita la diversidad de recomendaciones [7].

**Modelos híbridos:** Estos modelos combinan aspectos del filtrado colaborativo y filtrado basado en contenido con el fin de producir recomendaciones más precisas.

Modelos híbridos han emergido para contrarrestar las limitaciones de los modelos anteriores, usando las mejores características de ambos métodos para proveer recomendaciones más específicas. Estos sistemas pueden analizar interacciones de información de usuarios y atributos de destinos, brindando un conjunto diversificado de sugerencias. Modelos híbridos han demostrado una mejora en asertividad consistente y recomendaciones relevantes, enlazando satisfacción de usuario en el sector [8].

Aparte de los sistemas de recomendación previamente explicados, el alcance de la inteligencia artificial en el “Turismo inteligente” va más allá, teniendo opciones como ChatBots, asistentes virtuales, proyecciones futuras de comportamientos y preferencias y asignación de recursos. Aparte de estos desarrollos, la IA puede predecir picos de períodos de viajes, permitiendo a negocios optimizar su staff y niveles de inventarios mesurados, además que análisis predictivos pueden ayudar a identificar tendencias emergentes de viajes, permitiendo a negocios adaptar su oferta en relación a la evolución a la demanda de clientes [9].

La aplicación de la IA en el turismo inteligente no está exenta de desafíos. Seguridad de la información y privacidad

de datos son preocupaciones significativas, por lo que la IA requiere de información personal extensa para funcionar efectivamente. Asegurar el uso ético y la protección de esta información es fundamental para mantener la confianza del usuario. También existe el desafío del sesgo algorítmico, donde los sistemas de IA pueden perpetuar inadvertidamente sesgos existentes en los datos, dando lugar a recomendaciones injustas o discriminatorias [9]. Para poder cubrir estos desafíos se requiere un gran esfuerzo colaborativo entre desarrolladores de inteligencia artificial, empresarios del turismo y grupos de interés que aseguren que las aplicaciones de inteligencia artificial se encuentran alineadas con las necesidades y expectativas de usuarios.

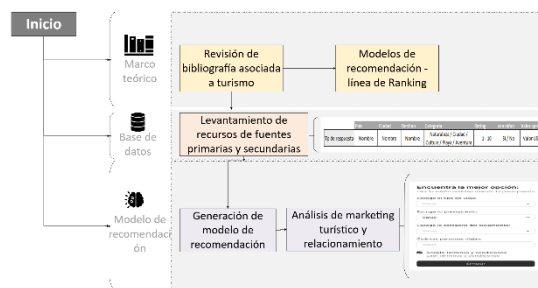
Los sistemas de recomendación turísticos basados en inteligencia artificial han ganado creciente aceptación en contextos digitales, especialmente en plataformas globales como Booking, Airbnb o TripAdvisor, que operan con grandes volúmenes de datos y arquitecturas de recomendación en tiempo real. Sin embargo, en mercados regionales como el de Cali, Colombia, muchas agencias de viajes tradicionales aún carecen de herramientas que les permitan sistematizar la asignación de planes turísticos según las características de sus clientes. Este trabajo se diferencia precisamente en que adapta técnicas de aprendizaje automático a un entorno con recursos limitados, demostrando que es posible implementar un modelo robusto, interpretable y funcional sin necesidad de

una infraestructura tecnológica avanzada. A continuación, se listan algunos elementos diferenciadores en relación al uso de este tipo de metodologías en el sector:

- Validación cuantitativa con métricas tradicionales que permiten comparar objetivamente el rendimiento de modelos.
- Portabilidad del modelo hacia entornos empresariales reales sin necesidad de plataformas sofisticadas.
- Enfoque en la fidelización del cliente como variable estratégica de personalización turística.

### 3. Metodología

La metodología implementada en esta investigación combina técnicas de procesamiento de datos, modelamiento estadístico y aprendizaje automático, orientadas a predecir la afinidad entre clientes y paquetes turísticos. Gracias al apoyo del sector privado, se pudo generar una base de datos histórica la cual se compone en 17 variables predictoras que incluyen datos sociodemográficos, preferencias de viaje, perfil de cliente, presupuesto, último destino y calificación otorgada. El vector de salida corresponde a la calificación esperada del viaje.



**Figura 1.** Diagrama general de actividades a ejecutar en la investigación

### Modelo de sugerencias:

#### Descripción del Dataset

Campo	Descripción
Nombre	Nombre del cliente que compra el paquete turístico
Preferencias	Tipo de viaje que prefiere el cliente (Playa, Ciudad, Cultura)
Edad	Edad de cliente
Género	Femenino/Masculino
Estado civil	Soltero/ Casado, Divorciado/Unión Libre
Estrato	Definir estrato socioeconómico de cliente
hijos	Cantida de hijos
Visa USA	Posee Visa a USA
Tolerancia horas de viaje	Tolerancia horas de viaje
Restricciones	Tiene restricciones alimentarias, desazamiento, propiocepi
Presupuesto	Perfilamiento de cliente según compras históricas
Perfil	Tipo de perfil del cliente
Viaja con	Acompañamiento de cliente en viajes previos
Geografía	País visitado
Paquete comercial	Paquete ofrecido
Público objetivo	Tipo de viajero que podría interesar paquetes turísticos
Preferencia paquete Tur	Preferencia de paquete turístico
Calificación media Histórica	Calificación del viaje según preferencias de viaje

**Tabla 1.** Datos sociodemográficos relacionados con paquetes turísticos con calificación

#### Preprocesamiento

La fase de preprocesamiento sirve para identificar y transformar variables categóricas, mediante código OneHotEncoder, con el fin de poderlas incluir en modelos matemáticos sin introducir sesgo [12].

Para asegurar que el flujo del proceso sea replicable y eficiente, se estructura el

modelo con base en Scikit-learn, el cual integra el preprocesamiento y modelo de regresión en un único objeto serializable (Modelo random forest). Este fue entrenado con un conjunto de entrenamiento (80%) y validado sobre un conjunto de prueba (20%) utilizando las métricas estándar: MSE (Error cuadrático medio), RMSE (Raíz del error cuadrático medio) y MAPE (Error porcentual absoluto medio), siguiendo prácticas aceptadas en evaluación de modelos supervisados [13].

Se compararon cuatro algoritmos de regresión:

- Regresión lineal clásica (baseline);
- Lasso con  $\alpha=0.01$  (modelo regularizado para selección de variables);
- Árbol de decisión con  $\text{max\_depth}=5$ ;
- Random Forest con configuración por defecto ( $n\_estimators=100$ ).

Las herramientas usadas en la elaboración del código fueron: Lenguaje Python, librerías tipo Scikit-learn usadas para transformar variables, ejecutar pruebas de eficiencia y determinar mejor modelo. Por último, las bases de datos y los resultados generados por el modelo se alojan en archivos de Excel guardados localmente en computadora personal.

#### 4. Resultados

Con el fin de determinar el modelo más adecuado para predecir la afinidad entre clientes y paquetes turísticos, se entrenaron y evaluaron cuatro algoritmos de regresión: Regresión Lineal, Lasso,

Árbol de Decisión y Random Forest. Los modelos fueron evaluados sobre un conjunto de prueba (20% de los datos), utilizando las métricas de Error Cuadrático Medio (MSE), Raíz del Error Cuadrático Medio (RMSE) y Error Porcentual Absoluto Medio (MAPE), a partir de un total de 64 registros aportados por la empresa.

Modelo	MSE	RMSE	MAPE (%)
Regresión lineal	15.742	12.547	19.62
Lasso	0.1809	0.4254	8.09
Árbol de decisión	0.2700	0.5197	9.70
Random Forest	0.1130	0.3361	6.01

**Tabla 2.** Comparativa de desempeño cuantitativo según modelo

Los resultados muestran que el modelo Random Forest obtuvo el mejor rendimiento general con un RMSE de 0.3361 y un MAPE de 6.01%, llevando a ser escogido gracias a su balance entre precisión y robustez.

Posteriormente, el modelo fue entrenado con la totalidad de los datos y guardado como archivo.pkl utilizando la librería joblib. Este archivo fue utilizado para realizar predicciones sobre nuevos registros de clientes contenidos en un archivo Excel (datos\_nuevos.xlsx), al cual se le añadió una nueva columna denominada "Calificación viaje". El archivo resultante (predicciones\_randomForest.xlsx) contiene las observaciones originales más la predicción generada, y está preparado para ser usado por áreas comerciales o de mercadeo para asignación de paquetes turísticos personalizados.

La discusión de estos resultados cobra mayor relevancia al considerar que muchas agencias de viajes tradicionales no cuentan con mecanismos automatizados para sugerir paquetes turísticos a sus clientes, lo cual limita su capacidad para personalizar ofertas y fidelizar usuarios. Tal como lo destacan diversos autores, entre ellos J. Doe y Smith and Jhonson, la integración de sistemas de recomendación en el sector turístico no solo mejora la experiencia del viajero, sino que también incrementa la eficiencia operativa y reduce la dependencia de criterios subjetivos por parte de los asesores comerciales [3] [4].

Además, diversos autores como Şeker y Ali Louati sostienen que los sistemas híbridos de recomendación — aquellos que combinan filtrado colaborativo, análisis de contenido y aprendizaje supervisado— permiten una mejora sustancial en la precisión de la predicción y en la capacidad de segmentar audiencias, especialmente en entornos con limitaciones de recursos tecnológicos [10] [13].

La implementación de este tipo de modelos en agencias caleñas podría representar una solución concreta al bajo nivel de automatización, permitiéndoles competir en mejores condiciones con plataformas digitales globales.

Finalmente, los resultados también concuerdan con estudios como los de Rodríguez Hernández, que resaltan la utilidad de la inteligencia artificial para analizar datos históricos de comportamiento y transformar esos

aprendizajes en propuestas de valor diferenciadas para cada cliente. En consecuencia, este proyecto no solo valida la aplicabilidad de modelos predictivos en el sector turístico local, sino que también propone una ruta práctica de adopción tecnológica alineada con las capacidades reales de las empresas.

## 5. Conclusiones

Este enfoque permitió validar un modelo funcional, eficiente y replicable, capaz de predecir la afinidad entre clientes y paquetes con base en datos históricos, y con potencial de integración en herramientas como Excel, sistemas de gestión de relaciones con el cliente (CRM) o PyXLL para entornos empresariales.

La implementación de sistemas de recomendación inteligentes en agencias de viajes tradicionales en Cali puede representar una ventaja competitiva crucial en un mercado caracterizado por la homogeneidad en la oferta. A diferencia de los métodos convencionales, los modelos predictivos basados en inteligencia artificial permiten identificar patrones en las preferencias de los usuarios, lo que facilita la generación automatizada de paquetes personalizados, como se evidencia en propuestas híbridas exitosas que combinan filtrado colaborativo y basado en contenido.

La ausencia de herramientas tecnológicas en estas agencias limita su capacidad para captar y fidelizar clientes en un entorno altamente digitalizado. Los

resultados obtenidos con el modelo Random Forest implementado muestran una mejora en la precisión de las recomendaciones, lo que confirma la utilidad de estas técnicas para incrementar la satisfacción del cliente y el aumento en transacciones comerciales buscando la fidelización de clientes propios.

La adopción de modelos de recomendación, además de mejorar la calidad del servicio ofrecido, representa una estrategia viable para aumentar la rentabilidad de las agencias de viajes mediante la optimización del marketing relacional. Al automatizar la selección de

actividades turísticas según intereses individuales, se reduce la carga operativa y se amplía la capacidad de respuesta a la demanda del mercado, replicando resultados positivos obtenidos por otros autores que han evidenciado un aumento en la retención de clientes gracias a la personalización de servicios [12]

## 6. Referencias:

- [1] Página web del **ministerio de comercio industria y turismo**, noticias de turismo: <https://www.mincit.gov.co/prensa/noticias/turismo/turismo-ocupa-lugar-esencial-economia-colombiana#:~:text=La%20participaci%C3%B3n%20del%20turismo%20en,aportan%20al%20valor%20agregado%20nacional.>), 2024.
- [2] Página web del departamento administrativo nacional de estadística – **DANE**, cuenta satélite de turismo. <https://www.dane.gov.co/index.php/comunicados-y-boletines/cuentas-y-sintesis-nacionales/turismo>, 2023.
- [3] J. Doe et al., "Artificial Intelligence-Assisted Tourism Service Recommendation Algorithms: A Comprehensive Review," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 3, pp. 123-135, 2019.
- [4] Smith and Johnson, "Enhancing Smart Tourism with AI-Assisted Recommendation Systems," IEEE Internet of Things. Journal, vol. 8, no. 4, pp. 5678-5689, 2020.
- [5] Kim and Park, "User Preference Analysis for Hotel Recommendation Using Machine Learning Techniques," IEEE. International Conference on Data Mining, pp. 56-63, 2018.
- [6] Li and Wang, "Context-Aware Recommendation System for Cultural Heritage Tourism," IEEE Transactions on multimedia, vol. 18, no. 5, pp. 678-689, 2018.
- [7] Zhang et al., "Enhancing Tourism Experience with Reinforcement Learning-Based Recommendation Systems," IEEE. Transactions on Services Computing, vol. 25, no. 4, pp. 345-357, 2019.
- [8] Gupta et al., "Hybrid Collaborative and Content-Based Filtering for City Exploration," IEEE Transactions on Intelligent transportation Systems, vol. 28, no. 2, pp. 456-467, 2021.
- [9] X. Sun, "Smart Tourism: Design and Application of Artificial Intelligence-Assisted Tourism Service Recommendation Algorithms" IEEE Electrical Systems vol. 20-9 pp.728-735, 2024.

- [10] Ali Louati 1, Hassen Louati 2, Meshal Alharbi 3 , Elham Kariri 1 , Turki Khawaji 1, Yasser Almubaddil 1 and Sultan Aldwsary, Machine Learning and Artificial Intelligence for a Sustainable Tourism: A Case Study on Saudi Arabia, Information 2024, 15,516, 2024.
- [11] Figueira et al., "Multiple criteria decisión análisis" state of the arte surveys, Springer, 2005.
- [12] Liu et al., "A Hybrid Recommendation System for Tourist Attractions Based on Collaborative Filtering and Content-Based Filtering," IEEE International Conference on Tourism and Technology, pp. 234-241, 2020.
- [13] F. Şeker, "Combining the power of artificial intelligence and mathematical modelling: A hybrid technique for enhanced forecast of tourism receipts", IEEE. European Journal of tourism research, vol. 7658, pp. 814-817, 2024
- [14] Y. M. Rodríguez Hernández, "Sistema de recomendación de actividades turísticas mejorando la experiencia del viajero con inteligencia artificial," Revista Científica Dominio de las Ciencias, vol. 9, no. 1, pp. 963–979, 2023.