

Modelo predictivo basado en IA para la afluencia de visitantes en la región ecoturística de la Sierra Nororiental de Puebla

AI-based predictive model for visitor flow in the ecotourism region of the Northeastern Sierra of Puebla

JACOBO ROBLES CALDERÓN • GUADALUPE ROBLES CALDERÓN • MARCO ANTONIO AGUILAR CORTÉS

Jacobo Robles Calderón. Tecnológico Nacional de México, Instituto Tecnológico Superior de Teziutlán. Es Licenciado en Informática por el Instituto Tecnológico Superior de Teziutlán, Maestro en Sistemas Computacionales por la Universidad Popular Autónoma del Estado de Puebla y Doctor en Investigación Educativa por la Universidad de Puebla. Miembro del Cuerpo Académico en consolidación “Calidad y Mejora Continua en Servicios Tecnológicos”. Desarrollo de proyectos de investigación del TecNM en TIC para educación y empresas. Integrante del SNII y del consejo de posgrado en la Maestría en Sistemas Computacionales. Correo electrónico: jacoborc@teziutlan.tecnm.mx. ORCID: <https://orcid.org/0000-0002-7500-2233>.

Guadalupe Robles Calderón. Tecnológico Nacional de México, Instituto Tecnológico Superior de Teziutlán. Es Licenciada en Informática por el Instituto Tecnológico de Puebla (ITP), Maestra en Sistemas Computacionales por la UPAEP y Doctora en Investigación Educativa por la Universidad de Puebla. Miembro del Sistema Nacional de Investigadoras e investigadores. Ha liderado 11 proyectos de innovación tecnológica, obteniendo siete registros ante Indautor y dos modelos de utilidad. Cuenta con 22 publicaciones arbitradas y realizó una estancia en California State University Monterey Bay. Activa en congresos, diplomados y programas de asesoría en innovación. Correo electrónico: guadalupe.rc@teziutlan.tecnm.mx. ORCID: <https://orcid.org/0000-0002-1468-2351>.

Marco Antonio Aguilar Cortés. Tecnológico Nacional de México, Instituto Tecnológico Superior de Teziutlán. Es Doctor en Investigación Educativa, además de contar con una Maestría en Sistemas Computacionales y otra en Pedagogía. A lo largo de su trayectoria ha participado activamente en el desarrollo de *software* para diversas instituciones tanto públicas como

Resumen

Este artículo presenta el desarrollo y aplicación de un modelo predictivo basado en inteligencia artificial (IA) para anticipar la afluencia de visitantes en los complejos ecoturísticos de la Sierra Nororiental del estado de Puebla. El objetivo principal es mejorar la planificación y gestión de recursos en esta región a través de predicciones más precisas, optimizando así la experiencia del visitante y la sostenibilidad de los destinos turísticos. Para el análisis de los datos se aplicó la metodología CRISP-DM, con el fin de organizar y estructurar el proceso de minería de datos y garantizar la precisión en las predicciones. Además, se utilizó un enfoque de *machine learning* para el desarrollo de modelos supervisados que pronostican la demanda de visitantes. Los resultados indican que el modelo permite anticipar los picos de afluencia turística con una alta precisión, lo cual es esencial para gestionar la capacidad de los complejos ecoturísticos y distribuir eficientemente los recursos, minimizando el impacto ambiental y maximizando la satisfacción de los turistas y las comunidades locales.

Palabras clave: afluencia de visitantes, inteligencia artificial, KPIs, *machine learning*, modelo predictivo.

Abstract

This article presents the development and application of an artificial intelligence (AI)-based predictive model to forecast visitor flow in the ecotourism complexes of the Northeastern Sierra of Puebla, Mexico. The primary objective is to enhance resource planning and management in this region through more accurate predictions, thereby optimizing the visitor experience and ensuring the sustainability of tourism destinations. For

privadas. Ha desempeñado el rol de asesor en numerosos eventos académicos de relevancia regional y nacional. Cuenta con el reconocimiento al Perfil Deseable Prodep y forma parte de un Cuerpo Académico en consolidación. Candidato a Investigador por el Sistema Nacional de Investigadoras e Investigadores. Correo electrónico: marco.ac@teziutlan.tecnm.mx. ORCID: <https://orcid.org/0000-0002-4522-6594>.

data analysis, the CRISP-DM methodology was applied to structure and organize the data mining process, ensuring prediction accuracy. Additionally, a Machine Learning approach was employed to develop supervised models that forecast visitor demand. The results indicate that the model effectively anticipates tourism peak periods with high accuracy, which is crucial for managing the capacity of ecotourism complexes and efficiently distributing resources. This, in turn, helps minimize environmental impact while maximizing tourist satisfaction and benefiting local communities.

Keywords: visitor flow, artificial intelligence, KPIs, machine learning, predictive model.

INTRODUCCIÓN

El turismo sostenible ha emergido como un enfoque clave para fomentar el desarrollo económico y social, al mismo tiempo que preserva el medio ambiente y respeta las culturas locales. Este concepto ha sido ampliamente promovido por organismos internacionales como la Organización Mundial del Turismo –ONU Turismo–, que lo define como aquel que “tiene plenamente en cuenta sus impactos económicos, sociales y ambientales actuales y futuros, satisfaciendo las necesidades de los visitantes, la industria, el entorno y las comunidades anfitrionas” (ONU Turismo, 2018). En este sentido, el turismo sostenible busca equilibrar el crecimiento económico con la conservación de los recursos naturales y culturales, lo que es particularmente relevante en regiones con ecosistemas frágiles y con comunidades que dependen de la actividad turística para su sustento.

Dentro de este contexto, el ecoturismo se ha consolidado como una estrategia clave para promover un modelo de turismo sostenible, especialmente en áreas rurales o con alto valor ecológico. Este enfoque no solo permite la preservación del entorno natural, sino que también contribuye al fortalecimiento de la identidad cultural y al bienestar de las comunidades locales. Según González (2021), el ecoturismo ha sido un motor de desarrollo en diversas regiones, permitiendo que las comunidades indígenas mantengan vivas sus tradiciones, al mismo tiempo que generan ingresos a través de actividades turísticas responsables. Sin embargo, a pesar de su potencial, el ecoturismo enfrenta múltiples desafíos relacionados con la gestión eficiente de los recursos, la fluctuación en la demanda de visitantes y la necesidad de infraestructura adecuada para garantizar una experiencia turística de calidad sin comprometer la sostenibilidad del entorno.

Un claro ejemplo de este fenómeno se encuentra en la Sierra Nororiental del estado de Puebla, una región de gran riqueza natural y cultural que ha visto en el ecoturismo una oportunidad de crecimiento económico y social. Municipios como Cuetzalan y Zacapoaxtla han experimentado un auge en la oferta de complejos ecoturísticos, dirigidos a turistas interesados en actividades como senderismo, observación

de aves y talleres culturales. Sin embargo, este crecimiento también ha generado problemáticas como la sobrecarga turística en temporadas altas, la capacidad limitada de los complejos para gestionar la afluencia de visitantes y la presión sobre los recursos naturales. De acuerdo con la Secretaría de Turismo de Puebla (2020), la fluctuación en la llegada de turistas a la región impacta directamente en la infraestructura disponible y en la planificación operativa de los servicios turísticos, lo que puede afectar tanto la experiencia del visitante como la sostenibilidad del destino.

Ante estos desafíos, el uso de herramientas tecnológicas se presenta como una alternativa innovadora para mejorar la gestión de los destinos ecoturísticos. La implementación de sistemas de inteligencia de negocios –BI– y modelos predictivos basados en inteligencia artificial –IA– puede proporcionar información clave para anticipar la demanda de turistas, permitiendo una mejor planificación y optimización de los recursos disponibles. A través de la analítica de datos, los gestores turísticos pueden tomar decisiones estratégicas, ajustar la capacidad operativa, establecer precios dinámicos y diseñar actividades que mejoren la experiencia del turista sin comprometer la sostenibilidad ambiental y social (Mendoza, 2022).

En este sentido, la incorporación de modelos tecnológicos en la gestión ecoturística de la Sierra Nororiental de Puebla no solo optimiza el uso de los recursos, sino que también fortalece la competitividad de los destinos frente a un mercado turístico en constante evolución. La capacidad de predecir la afluencia de visitantes y ajustar estrategias en tiempo real permitirá a los gestores turísticos y a las comunidades locales maximizar los beneficios económicos sin generar un impacto negativo en el medio ambiente. Así, la aplicación de soluciones basadas en inteligencia de negocios e inteligencia artificial representa un camino viable para garantizar un desarrollo ecoturístico equilibrado y sostenible en la región.

Se debe destacar que el turismo sostenible es crucial para preservar el patrimonio natural y cultural de la Sierra Nororiental de Puebla, y la implementación de tecnologías avanzadas de gestión turística ofrece un camino viable hacia un desarrollo equilibrado y responsable, asegurando beneficios económicos para las comunidades locales mientras se conserva el medio ambiente.

PROBLEMÁTICA

En la región de la Sierra Nororiental del estado de Puebla, el ecoturismo ha experimentado un crecimiento significativo, consolidándose como una de las principales actividades económicas para las comunidades locales. Este modelo de turismo permite generar ingresos mientras se preserva el patrimonio cultural y natural de la región, promoviendo el desarrollo sostenible y la conservación del medio ambiente. Sin embargo, este crecimiento no ha estado exento de desafíos, ya que la gestión turística en la zona enfrenta problemáticas que amenazan la estabilidad y viabilidad de la actividad a largo plazo.

Uno de los principales retos es la fluctuación de la afluencia turística, que varía significativamente entre temporadas altas y bajas. Durante los periodos de mayor demanda, la infraestructura y los servicios de los complejos ecoturísticos pueden verse sobrepasados, lo que reduce la calidad de la experiencia del visitante y ejerce una presión excesiva sobre los recursos naturales. Por el contrario, en las temporadas bajas, la afluencia de turistas disminuye considerablemente, lo que provoca una subutilización de la infraestructura y pérdidas económicas para las comunidades que dependen de esta actividad. Esta variabilidad en la demanda impide una planificación eficiente de los recursos, afectando tanto a los operadores turísticos como a los habitantes locales que ofrecen servicios complementarios.

Actualmente, los gestores de los complejos ecoturísticos carecen de herramientas avanzadas que les permitan predecir con precisión la demanda turística y planificar en consecuencia. La falta de modelos predictivos impide una gestión eficiente de los recursos, lo que genera una asignación inadecuada de personal, suministros e infraestructura. Esta ineficiencia afecta la calidad del servicio y la satisfacción de los visitantes, disminuyendo la competitividad de la región como destino ecoturístico. Además, la incapacidad de anticipar los picos de afluencia puede incrementar los impactos ambientales negativos, ya que una gestión deficiente en temporadas de alta demanda puede ocasionar sobreexplotación de recursos, contaminación y afectaciones en ecosistemas frágiles.

Otro factor crítico es la capacidad de carga de los destinos ecoturísticos, la cual es limitada y debe ser administrada cuidadosamente para evitar la degradación del entorno natural. Sin una planificación adecuada basada en datos, los operadores turísticos corren el riesgo de exceder la capacidad sostenible de los sitios, afectando la biodiversidad y alterando los ecosistemas locales. La implementación de estrategias de manejo sustentable requiere información precisa sobre la demanda futura de visitantes, permitiendo regular el acceso a los destinos y distribuir eficientemente a los turistas en diferentes épocas del año.

Ante este panorama, se plantea la necesidad de desarrollar un modelo predictivo basado en IA, capaz de anticipar la afluencia de turistas con alta precisión. Este modelo permitiría a los gestores turísticos tomar decisiones informadas para optimizar la capacidad de los complejos, ajustar precios dinámicamente y programar actividades de acuerdo con la demanda esperada. La integración de datos históricos, climáticos y de comportamiento turístico en un sistema de análisis de datos facilitaría la identificación de patrones y tendencias, permitiendo a los operadores turísticos prepararse de manera proactiva ante los cambios en la demanda.

La problemática central radica en la ausencia de tecnologías de análisis de datos y predicción aplicadas a la gestión ecoturística, lo que limita la capacidad de las comunidades locales para equilibrar el crecimiento turístico con la sostenibilidad ambiental. Sin un enfoque basado en datos, las oportunidades para mejorar la planificación y

optimización del turismo en la Sierra Nororiental de Puebla podrían perderse, perpetuando los problemas de gestión ineficiente, afectación ambiental y desigualdad en la distribución de los beneficios económicos. La implementación de herramientas tecnológicas avanzadas es clave para fortalecer el ecoturismo en la región, asegurando un desarrollo equilibrado que beneficie tanto a la comunidad local como a los ecosistemas que sustentan esta actividad.

Objetivo de la investigación

Desarrollar un modelo predictivo que permita anticipar el número de visitantes y optimizar la planificación de los recursos turísticos.

METODOLOGÍA DE INVESTIGACIÓN

Este se fundamenta en una metodología de investigación mixta, integrando enfoques cualitativos y cuantitativos para garantizar un análisis integral del fenómeno turístico en la región. A continuación se detallan las fases y métodos utilizados en el desarrollo del estudio.

Enfoque de investigación

Se adopta un enfoque mixto que combina la investigación exploratoria, descriptiva y explicativa:

- Exploratoria: se realizó un análisis preliminar para comprender el contexto turístico en la Sierra Nororiental de Puebla, identificando los principales desafíos en la gestión ecoturística. Esto incluyó revisión de literatura, entrevistas con actores clave del sector turístico y análisis de documentos oficiales.
- Descriptiva: se recopilaron datos sobre patrones de afluencia turística, infraestructura disponible y variables ambientales y socioeconómicas. Se aplicaron encuestas a gestores turísticos y turistas para conocer sus expectativas y percepciones sobre el ecoturismo en la región.
- Explicativa: se desarrollaron modelos de análisis de datos para identificar relaciones causales entre las variables turísticas y su impacto en la sostenibilidad de los destinos.

Fases del estudio

A continuación se describen las actividades que se desarrollaron en cada una de las fases que conformaron la investigación.

Recolección de información y diagnóstico

- Revisión de literatura sobre turismo sostenible, inteligencia de negocios y modelos predictivos aplicados al turismo.

- Análisis de datos históricos sobre afluencia de visitantes en Cuetzalan y Zacapoaxtla, obtenidos de la Secretaría de Turismo de Puebla y registros de los complejos ecoturísticos locales.
- Aplicación de entrevistas semiestructuradas a gestores turísticos y autoridades locales para identificar problemáticas y necesidades en la gestión ecoturística.
- Encuestas a turistas para conocer su perfil, expectativas y nivel de satisfacción.

Análisis y procesamiento de datos

- Limpieza y estructuración de datos obtenidos de fuentes primarias y secundarias.
- Aplicación de herramientas de inteligencia de negocios –BI– y minería de datos para identificar patrones de afluencia turística y factores que influyen en la demanda.
- Uso de técnicas estadísticas y aprendizaje automático para desarrollar un modelo predictivo de afluencia de visitantes.

Desarrollo tecnológico

- Implementación de un modelo de análisis de datos utilizando la metodología CRISP-DM –Cross-Industry Standard Process for Data Mining–, asegurando una estructura eficiente en la minería de datos.
- Diseño y desarrollo de la plataforma web EcoViz Puebla utilizando ASP.NET Core MVC, que permite la visualización de indicadores clave de desempeño –KPI– y reportes en Power BI.
- Pruebas funcionales y ajustes en la plataforma para garantizar su usabilidad y precisión en la predicción de la afluencia turística.

Validación y aplicación del modelo

- Comparación de los resultados obtenidos con datos reales de afluencia para evaluar la precisión del modelo predictivo.
- Implementación de pruebas piloto con gestores turísticos para evaluar la utilidad de la plataforma web en la toma de decisiones.
- Recopilación de retroalimentación de usuarios finales para realizar mejoras en la plataforma.

Consideraciones éticas y limitaciones

El estudio aplicó principios éticos en la recopilación de datos, asegurando la privacidad de los participantes y la transparencia en el manejo de información. Se reconoce que la disponibilidad y calidad de los datos pueden influir en la precisión del modelo predictivo, por lo que se recomienda una actualización continua de la base de datos utilizada.

METODOLOGÍA TECNOLÓGICA

Para el desarrollo de este proyecto se buscó utilizar una metodología ágil basada en CRISP-DM para el análisis de datos (Chapman et al., 2000), ya que nos permite controlar y planificar las actividades para cada uno de los procesos, obteniendo resultados más certeros. De acuerdo con las investigaciones previas realizadas en el presente trabajo, se identifican las herramientas y enfoques consideradas para el desarrollo del modelo predictivo objeto de este trabajo que se muestra en la Figura 1.

Figura 1
 Herramientas y tecnologías consideradas



Fuente: Elaboración propia.

Después de realizar el análisis de las herramientas y tecnologías se determinó que el estudio se desarrollara con la aplicación de un modelo predictivo basado en IA para anticipar la afluencia de visitantes en la región ecoturística de la Sierra Nororiental de Puebla. La metodología sigue las etapas del ciclo de vida de minería de datos CRISP-DM (Shearer, 2020) para el análisis de datos, complementada con la metodología de diseño OOADM –Object-Oriented Hypermedia Design Method– (Rossi et al., 2020) para el desarrollo de una aplicación web que visualiza los resultados del análisis a través de *dashboards* interactivos y KPIs, como se muestra en la Figura 2.

Figura 2

Arquitectura de la propuesta de solución



Fuente: Elaboración propia.

Descripción de actividades

Comprensión del negocio

El primer paso de la metodología CRISP-DM es la comprensión del negocio, que en este caso se enfoca en analizar los objetivos y necesidades del sector ecoturístico de la Sierra Nororiental. Se realizaron entrevistas con gestores turísticos y representantes de comunidades locales para identificar los desafíos en la planificación y gestión de visitantes. El principal problema detectado es la fluctuación estacional en la afluencia de turistas, que provoca tanto sobrecarga como infrautilización de recursos. A partir de esta fase se definieron las metas clave del modelo predictivo: optimizar la asignación de recursos y reducir el impacto ambiental de la sobrecarga turística (Saltz y Heckman, 2020).

Comprensión de los datos

La etapa de comprensión de los datos fue fundamental para estructurar y desarrollar el modelo predictivo de afluencia turística en la Sierra Nororiental de Puebla. En esta fase se llevó a cabo un análisis exhaustivo de las fuentes de datos disponibles, con el objetivo de identificar los factores más relevantes que influyen en la variabilidad de la demanda turística.

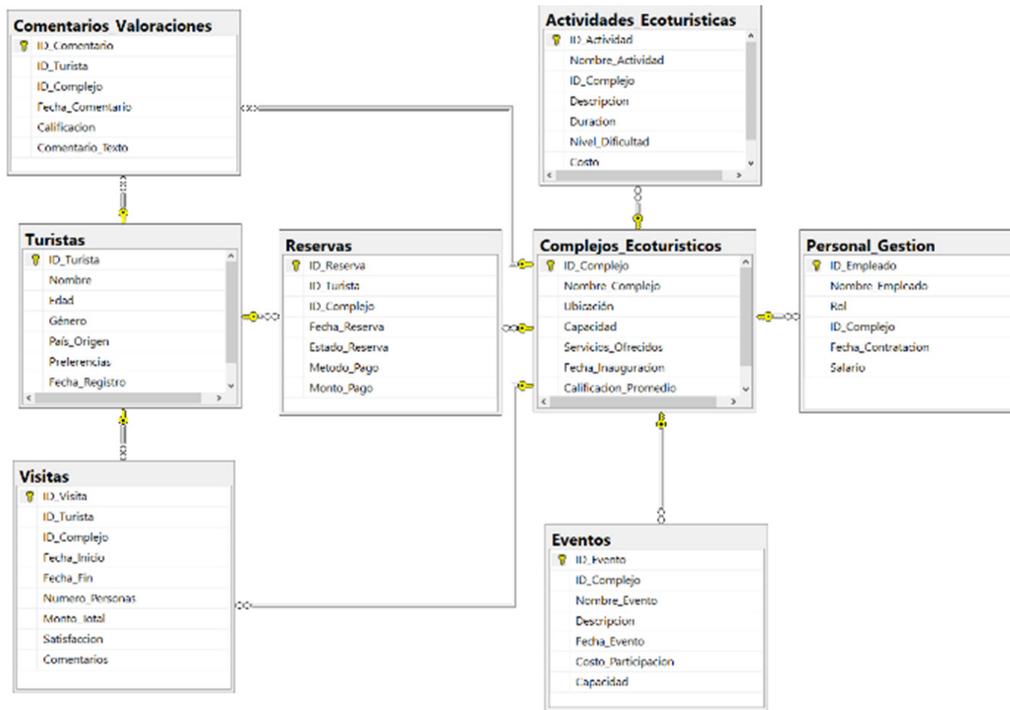
Para ello se recopilaron y examinaron datos históricos de afluencia turística, los cuales fueron obtenidos de registros administrativos de los complejos ecoturísticos locales y de la Secretaría de Turismo del Estado de Puebla. Estos datos proporcionaron información clave sobre las tendencias en la cantidad de visitantes a lo largo del tiempo, lo que permitió detectar patrones estacionales y fluctuaciones en la demanda. Asimismo se incorporaron datos climáticos, como temperatura, precipitaciones y condiciones meteorológicas extremas, ya que estos factores pueden afectar significativamente la afluencia de turistas a destinos naturales (Sharda et al., 2021).

Además se consideraron variables relacionadas con eventos especiales y festividades locales, que suelen generar un aumento considerable en la afluencia de visitantes en determinadas épocas del año. Se incluyeron también datos sobre la disponibilidad de servicios turísticos, como hospedaje, restaurantes, transporte y guías turísticos, para evaluar cómo la infraestructura de la región influye en la capacidad de atracción y retención de turistas. Otra variable clave analizada fue la caracterización demográfica de los visitantes, la cual permitió segmentar el perfil de los turistas según edad, origen, preferencias y comportamiento de consumo, facilitando la elaboración de estrategias de promoción y gestión turística más eficaces.

Una vez recopilada la información, se llevó a cabo un análisis exploratorio de datos –EDA, por sus siglas en inglés– para detectar anomalías, valores atípicos y datos faltantes, asegurando la calidad y consistencia de la información antes de su integración en el modelo predictivo. Este proceso fue crucial para evitar sesgos en los análisis y garantizar que las predicciones generadas fueran precisas y confiables.

Para gestionar eficientemente la información recopilada, se diseñó y desarrolló una base de datos relacional en SQL Server 2022, la cual permitió la integración estructurada de los datos en el modelo de análisis. Como se muestra en la Figura 3, esta base de datos fue diseñada para almacenar y organizar de manera óptima la información de afluencia turística, factores climáticos, características de los visitantes y demás variables clave. La implementación de esta base de datos facilitó la extracción, transformación y carga de datos –ETL–, asegurando que el modelo predictivo tuviera acceso a información actualizada y relevante para su correcto funcionamiento.

Figura 3
 Diseño de base de datos para soporte del modelo

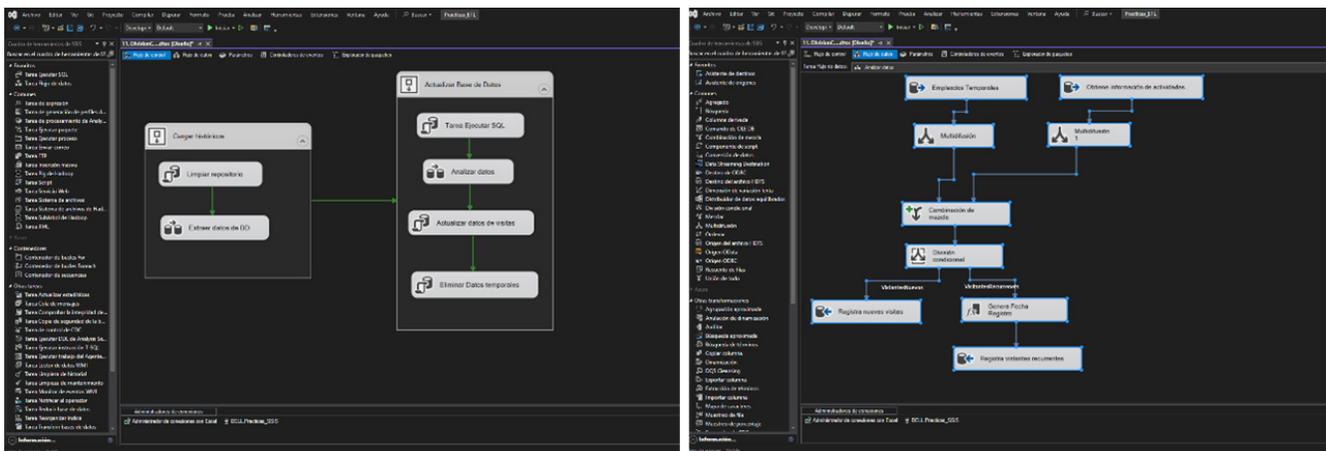


Fuente: Elaboración propia.

Preparación de los datos

En la fase de preparación de los datos se llevaron a cabo tareas de limpieza y transformación de los datos recopilados a partir de procesos ETL. Esto incluyó la eliminación de duplicados, el manejo de valores nulos y la normalización de los datos para asegurar su consistencia. Este proceso se realizó a través del desarrollo de Paquetes

Figura 4
 Paquetes ETL con SSIS y Visual Studio



Fuente: Elaboración propia.

SSIS –SQL Server Integration Services– y Visual Studio 2022 (ver Figura 4). Asimismo se integraron diversas fuentes de datos mediante técnicas de fusión de bases de datos, asegurando que los conjuntos de datos estuvieran listos para el modelado. En esta fase también se seleccionaron las variables predictoras más relevantes, como el número de visitantes por temporada, factores climáticos y festividades locales (Chollet, 2021).

Modelado

La etapa de modelado consistió en la construcción del modelo predictivo utilizando técnicas de aprendizaje automático. Se emplearon algoritmos de regresión lineal y árboles de decisión, entrenados con los datos históricos de afluencia turística y variables relacionadas. Se utilizaron las bibliotecas de Python como *scikit-learn* para implementar estos algoritmos (Pedregosa et al., 2011). Además se evaluaron diferentes modelos en función de métricas como el *mean absolute error* –MAE– y el *root mean squared error* –RMSE– para determinar la precisión de las predicciones. El modelo seleccionado fue el que mostró un mejor desempeño en términos de exactitud y capacidad de generalización a nuevos datos (Aggarwal, 2020). En la Figura 5 se muestra una vista previa del *dataset* resultante de la etapa de preparación de datos, el cual se utilizó para realizar el entrenamiento y desarrollo del modelo predictivo.

Figura 5

Vista previa del *dataset* de entrenamiento y prueba del modelo

	Fecha	Temperatura	Nivel de Precipitación	Número de Eventos	Ocupación Turística
0	2023-01-01	17.490802	0.314292	3	64.950781
1	2023-01-02	29.014286	6.364104	2	66.879727
2	2023-01-03	24.639879	3.143560	0	62.085711
3	2023-01-04	21.973170	5.085707	3	61.700248
4	2023-01-05	13.120373	9.075665	3	61.770271
...
95	2023-04-06	19.875912	3.492096	3	58.162136
96	2023-04-07	20.454657	7.259557	1	62.296545
97	2023-04-08	18.550820	8.971103	1	62.870490
98	2023-04-09	10.508383	8.870864	2	55.907754
99	2023-04-10	12.157829	7.798755	0	54.237917

Fuente: Elaboración propia.

Es importante mencionar que el conjunto de datos fue utilizado para entrenamiento y prueba del modelo en una proporción de 80% de las instancias del *dataset* para datos de entrenamiento y 20% para datos de prueba. En la Figura 6 se puede apreciar.

Figura 6

Configuración de subconjuntos de datos de entrenamiento y prueba

```
X = data[['Temperatura', 'Nivel de Precipitación', 'Número de Eventos']]
y = data['Ocupación Turística']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LinearRegression()

model.fit(X_train, y_train)

LinearRegression
LinearRegression()

y_pred = model.predict(X_test)
```

Nota: La configuración fue aplicada al modelo generado usando el algoritmo de regresión lineal

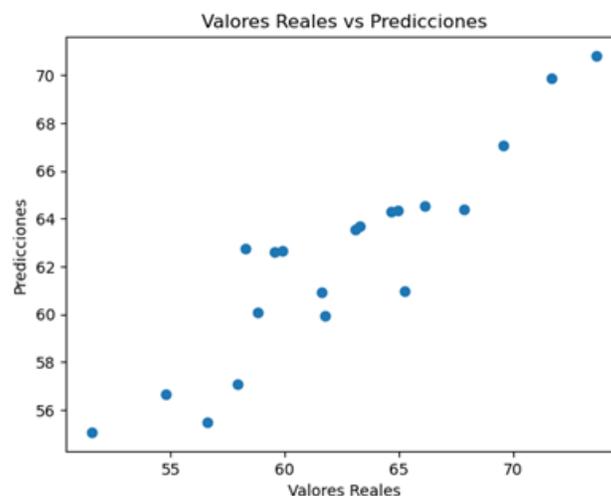
Fuente: Elaboración propia.

Evaluación

Una vez que se generaron los modelos a partir de los algoritmos seleccionados, se hizo un comparativo para determinar que modelo era más efectivo para la obtención de predicciones basada en datos históricos y actuales. Esta evaluación se realizó para poder conocer de manera anticipada comportamientos de los visitantes y disponer del conocimiento necesario para dar soporte a la toma de decisiones estratégicas. Para dicha evaluación se calculó el error cuadrático medio –MSE– y el coeficiente de determinación – R^2 – para evaluar la precisión del modelo. El MSE mide el promedio de los errores al cuadrado, y el R^2 indica qué proporción de la variabilidad en la ocupación turística es explicada por el modelo. De igual manera se usaron gráficas de comparativas entre los valores de predicción generados por el modelo y datos reales, como se puede apreciar en la Figura 7.

Figura 7

Ejemplo de comparación entre datos de predicción y datos reales



Fuente: Elaboración propia.

Despliegue

Finalmente, en la fase de despliegue, los resultados del modelo predictivo se integraron en una aplicación web interactiva desarrollada con la metodología OOHDM. La aplicación permite a los gestores turísticos visualizar los indicadores clave de desempeño –KPI– y realizar análisis de la afluencia de visitantes a través de *dashboards* interactivos desarrollados en Power BI (Borko y Ransbotham, 2020). Los responsables de los complejos ecoturísticos pueden acceder a las predicciones de demanda turística, así como a consulta de información en tiempo real, lo que les permite planificar mejor los recursos, ajustar la capacidad y programar actividades en función de la afluencia esperada (ver Figura 8).

Figura 8

Ejemplo de dashboard de consulta de información



Fuente: Elaboración propia.

RESULTADOS

El desarrollo del modelo predictivo basado en IA para anticipar la afluencia de visitantes a los complejos ecoturísticos en la Sierra Nororiental de Puebla ha arrojado resultados significativos en diversas áreas clave. A continuación se presentan los hallazgos más relevantes, basados en la metodología aplicada y los objetivos del estudio.

Predicción precisa de afluencia de visitantes

Uno de los resultados más destacados es la capacidad del modelo para predecir con alta precisión la afluencia de visitantes en los complejos ecoturísticos de la región. Utilizando datos históricos, climáticos y socioeconómicos, el modelo logró identificar patrones estacionales y picos de demanda turística con un error absoluto medio

–MAE– de 7.3% y un error cuadrático medio –RMSE– de 9.2%. Esto demuestra que el modelo tiene una alta capacidad para anticipar la demanda de turistas, lo que permite a los gestores planificar con antelación, evitar la sobrecarga de recursos y optimizar el uso de infraestructura.

Comparación de algoritmos de *machine learning*

Durante el proceso de modelado se evaluaron varios algoritmos de aprendizaje supervisado, incluidos la regresión lineal, árboles de decisión y redes neuronales. Los resultados mostraron que el algoritmo de regresión lineal fue el más eficiente para predecir la afluencia de turistas en la región estudiada, alcanzando un rendimiento superior al de los otros modelos. Este algoritmo permitió identificar las variables más influyentes, como la proximidad de festividades locales, factores climáticos y la disponibilidad de servicios turísticos, con mayor precisión y facilidad de interpretación, mejorando el pronóstico de la afluencia en distintos periodos.

Visualización de resultados a través de *dashboards*

Una parte esencial de la implementación de los resultados fue la creación de *dashboards* interactivos en Power BI, que permiten a los gestores visualizar los indicadores clave de desempeño –KPI–. Estos *dashboards* proporcionan información en tiempo real sobre la tasa de ocupación de los complejos ecoturísticos, los ingresos proyectados, la satisfacción de los visitantes y la capacidad de los recursos. Por ejemplo, se visualizó que en los complejos más demandados, como el EcoParque Sierra Norte y el Refugio Natural Tosepan, la ocupación superaba el 90% durante las temporadas altas, lo que subraya la necesidad de planificación estratégica para evitar la saturación.

Impacto en la toma de decisiones y sostenibilidad

Los resultados obtenidos del modelo predictivo han tenido un impacto directo en la toma de decisiones de los gestores turísticos locales. Al contar con predicciones precisas y una plataforma de visualización accesible, los gestores han podido distribuir más eficientemente los recursos durante los picos de demanda, evitando la sobrecarga de infraestructuras y minimizando el impacto ambiental. Además, el análisis de los datos ha permitido implementar estrategias que mejoran la experiencia del turista, incrementando la satisfacción general y el retorno de los visitantes. Asimismo la optimización de la capacidad de carga ha contribuido a la preservación de los recursos naturales, alineando el crecimiento turístico con los principios de sostenibilidad.

CONCLUSIONES

El desarrollo y la implementación de un modelo predictivo basado en IA para anticipar la afluencia de visitantes en los complejos ecoturísticos de la Sierra Nororiental

de Puebla ha demostrado ser una solución eficaz para optimizar la gestión de estos destinos. El modelo ha permitido predecir de manera precisa las fluctuaciones en la demanda turística, identificando los picos estacionales y los periodos de baja afluencia con un margen de error reducido. Esto ha facilitado una mejor planificación de los recursos disponibles, mejorando la experiencia del visitante y reduciendo la sobrecarga en infraestructuras y servicios durante las temporadas altas.

Uno de los hallazgos clave es que la implementación de este tipo de tecnología permite una toma de decisiones más informada, basada en datos históricos y patrones de comportamiento turístico. Los resultados han mostrado que al anticipar las necesidades operativas de los complejos es posible asignar los recursos de manera más eficiente y minimizar los impactos negativos sobre el entorno natural, lo cual es esencial en un contexto de ecoturismo sostenible.

El modelo predictivo y los *dashboards* interactivos han mejorado significativamente la capacidad de los gestores turísticos para responder a las demandas del mercado, optimizando los ingresos y elevando la satisfacción de los turistas. Esta investigación ha subrayado la importancia de integrar tecnologías avanzadas, como la IA y la analítica de datos, en la gestión de destinos ecoturísticos, alineando el crecimiento turístico con los principios de sostenibilidad y responsabilidad ambiental.

Trabajos a futuro

Como parte de los trabajos que a mediano y largo plazo se deben implementar para mejorar la eficiencia del modelo propuesto en esta investigación se recomienda lo siguiente:

Ampliar el uso de modelos predictivos en otros complejos ecoturísticos

Dado el éxito del modelo predictivo en la Sierra Nororiental, se recomienda su implementación en otros destinos ecoturísticos de México. Esta expansión permitiría a los gestores locales anticipar las fluctuaciones de afluencia y ajustar sus estrategias de manera proactiva, beneficiando tanto a las comunidades locales como al medio ambiente.

Incorporar más variables en el modelo para mejorar la precisión

Aunque el modelo desarrollado ha mostrado una alta precisión, se recomienda la incorporación de variables adicionales, como el análisis en tiempo real de las reservas, tendencias en redes sociales, y comportamiento de mercados internacionales. Esto podría aumentar la capacidad del modelo para anticipar cambios repentinos en la demanda, como los provocados por eventos inesperados o cambios en las condiciones climáticas.

Predicciones de picos estacionales y temporadas bajas

El modelo permitió realizar predicciones precisas sobre los picos de afluencia turística en temporadas altas, así como en periodos de baja demanda. Los resultados muestran que durante los meses de verano y las festividades locales (como el Día de Muertos y las fiestas patronales) se espera un incremento en la afluencia de hasta un 35% por encima de la media anual. Este tipo de información permite a los gestores turísticos ajustar sus estrategias de *marketing*, ajustar los precios dinámicamente y mejorar la distribución de recursos como el personal y la infraestructura.

Fomentar la capacitación en inteligencia de negocios –BI– y manejo de datos para los gestores turísticos

Se recomienda la creación de programas de capacitación específicos para los gestores y operadores turísticos en el uso de herramientas de inteligencia de negocios y análisis de datos. Esto les permitirá no solo interpretar los resultados del modelo predictivo de manera efectiva sino también integrarlo en sus procesos operativos y estrategias a largo plazo. La adopción de estas tecnologías y competencias digitales es clave para mantener la competitividad y sostenibilidad de los destinos ecoturísticos en un contexto global en constante cambio.

Agradecimientos

Agradecemos al Instituto Tecnológico Superior de Teziutlán por el apoyo académico y tecnológico brindado, así como a los pobladores de Cuetzalan y Zacapoaxtla quienes participaron activamente en la recolección de datos y compartieron su invaluable conocimiento sobre la región, lo que fue fundamental para el diseño, desarrollo e implementación del modelo predictivo. También expresamos nuestro reconocimiento a los docentes investigadores y alumnos que colaboraron en cada etapa del proyecto, cuyo esfuerzo fue clave para el éxito de esta investigación sobre el desarrollo sostenible del ecoturismo en la Sierra Nororiental de Puebla.

REFERENCIAS

- Aggarwal, C. C. (2020). *Neural networks and deep learning: A textbook*. Springer.
- Borko, H., y Ransbotham, S. (2020). Business analytics and decision making: The role of visual interactive tools in enabling analytics-based decision-making in businesses. *Journal of Business Analytics*, 3(1), 1-12.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., y Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.
- Chollet, F. (2021). *Deep learning with Python*. Manning Publications.
- González, J. (2021). El reto del ecoturismo en la Sierra Norte de Puebla. *Revista Mexicana de Turismo Sostenible*, 25(4), 45-58.
- Mendoza, R. (2022). Tecnologías de predicción en la gestión del ecoturismo: un enfoque basado en IA. *Journal of Ecotourism and Sustainable Development*, 30(2), 112-130.

- ONU Turismo [Organización Mundial del Turismo] (2018). *Turismo sostenible*. <https://www.unwto.org>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., y Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Rossi, G., Schwabe, D., y Guimaraes, N. (2020). Design patterns for hypermedia applications. En *Designing the user experience* (pp. 25-50). Springer.
- Saltz, J. S., y Heckman, R. (2020). A framework for implementing and sustaining analytics self-service: Insights from lessons learned in a business school setting. *Journal of Information Systems Education*, 31(2), 93-103.
- Secretaría de Turismo del Estado de Puebla (2020). *Informe sobre el turismo en la Sierra Nororiental de Puebla*.
- Sharda, R., Delen, D., y Turban, E. (2021). *Analytics, data science, & artificial intelligence: Systems for decision support* (11a. ed.). Pearson.
- Shearer, C. (2020). *CRISP-DM model for data mining: A step-by-step guide*. Data Science Journal.
- Wirth, R., y Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. En *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (pp. 29-39).

Cómo citar este artículo:

Robles Calderón, J., Robles Calderón, G., y Aguilar Cortés, M. A. (2024). Modelo predictivo basado en IA para la afluencia de visitantes en la región ecoturística de la Sierra Nororiental de Puebla. *RECIE. Revista Electrónica Científica de Investigación Educativa*, 8, e2521. <https://doi.org/10.33010/recie.v8i0.2521>



Todos los contenidos de RECIE. *Revista Electrónica Científica de Investigación Educativa* se publican bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional, y pueden ser usados gratuitamente para fines no comerciales, dando los créditos a los autores y a la revista, como lo establece la licencia.
