

Practical Applications of Artificial Intelligence in the Development of Predictive Models: A Case Study with ChatGPT

José Iván Calderón Carrillo, Doctor¹ 

¹Universidad Privada del Norte, Perú, jose.calderon@upn.edu.pe

Abstract—The main objective of this research is to present the role that artificial intelligence applications play in the quantitative analysis required for the development of predictive models and the support they provide in this process. The aim is to understand how these applications can influence the field of business management and continuous improvement within organizations. To this end, a practical example is presented of the use of an artificial intelligence application for the development of a predictive model, comparing it with the development of the same predictive model using a conventional method. The main result is that artificial intelligence applications facilitate and streamline the data analysis process for correct business decision-making.

Keywords— Artificial intelligence, continuous improvement, predictive models, analysis.

Aplicaciones Prácticas de Inteligencia Artificial en el Desarrollo de Modelos Predictivos: Un Caso de Estudio con ChatGPT

José Iván Calderón Carrillo, Doctor¹ 

¹Universidad Privada del Norte, Perú, jose.calderon@upn.edu.pe

Resumen– El objetivo principal de la investigación es presentar el rol que cumplen las aplicaciones de inteligencia artificial en el análisis cuantitativo que se requiere para el desarrollo de modelos predictivos y el soporte que brindan en dicho proceso. Se busca conocer como estas aplicaciones podrán influir en el campo de la gestión empresarial y mejora continua dentro de las organizaciones. Para ello, se muestra un ejemplo práctico sobre el uso de una aplicación de inteligencia artificial para el desarrollo de un modelo predictivo comparándolo con el desarrollo del mismo modelo predictivo bajo un método convencional, teniendo como principal resultado que las aplicaciones de inteligencia artificial facilitan y agilizan el proceso de análisis de datos para la correcta toma de decisiones empresariales.

Palabras clave–Inteligencia artificial, mejora continua, modelos predictivos, análisis.

I. INTRODUCCIÓN

Las organizaciones requieren estar constantemente en desarrollo debido a la agresiva competencia en las diferentes industrias. Para ello, se requiere una correcta gestión empresarial, la cual permite guiar sus operaciones, procesos y la toma de decisiones empresariales para mantener su competitividad. Es aquí donde entra el rol de la analítica empresarial.

La analítica empresarial es una tendencia que permite a las organizaciones gestionar y analizar sus datos que sirve para un correcto proceso de toma de decisiones, para mejorar sus procesos, incrementando su rentabilidad y así poder mantener su competitividad [1].

La analítica empresarial permite a una organización generar ventajas competitivas dentro de su industria, ya que al utilizar y explotar sus datos, pueden anticiparse a riesgos, innovar y mejorar su proceso de toma de decisiones. Para ello, este proceso implica la recolección, procesamiento y la correcta interpretación de datos [2].

Dentro de la analítica empresarial o analítica de datos, se encuentra el modelamiento predictivo. El modelamiento o análisis predictivos, como su propio nombre indica, tiene como objetivo predecir el comportamiento de una variable de interés en función de una o más variables predictoras. Para lograr ello, se toma como fuente y posterior análisis, los datos históricos del proceso o área de interés.

Para desarrollar un modelo predictivo, el análisis de regresión lineal es utilizada con mayor frecuencia. Este análisis examina la relación entre las variables independientes y dependientes. Este análisis es utilizado para determinar el impacto de las variables predictoras, estimar efectos y predecir la tendencia. Es ampliamente utilizada en el campo del aprendizaje automático [3].

La regresión lineal trata de predecir el valor de una variable dependiente, conociendo el valor de una o más variables predictoras o independientes. Cuando se trata de solo una variable independiente, se denomina regresión lineal simple. Y cuando se cuenta con 2 o más variables independientes, se denomina regresión lineal múltiple [4].

Como indica Molina, la ecuación general del modelo de regresión lineal múltiple es la siguiente:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Siendo Y, la variable dependiente, medida en una escala continua. b_0 , es la constante del modelo de regresión, b_i son los coeficientes para cada variable independiente y x_i son las variables independientes [4].

Hasta ahora se ha mencionado el uso del modelo de regresión lineal como herramienta para realizar el análisis predictivo como parte de la analítica empresarial, esto se aprecia en la Fig. 1.

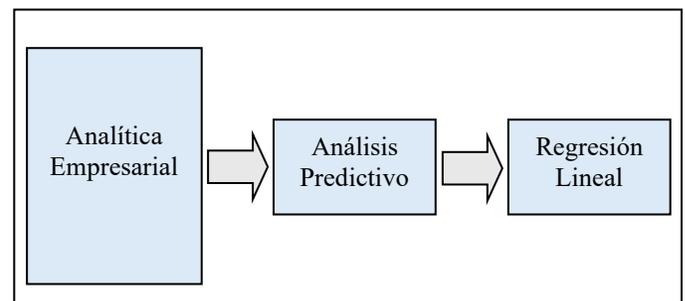


Fig. 1 Análisis predictivo, un pilar de la analítica empresarial.

Se debe mencionar que el análisis de regresión lineal no es la única técnica utilizada dentro del análisis predictivo. Existen más herramientas utilizadas como, por ejemplo, los árboles de regresión o clasificación, redes neuronales, series temporales, entre otros.

Con respecto a la analítica empresarial, no solo está soportada por el análisis predictivo. También se encuentra el análisis descriptivo y prescriptivo. El análisis descriptivo se concentra en conocer y explorar los datos, como su nombre indica, describir el comportamiento y tendencia de los datos en diferentes procesos de la organización y otros contextos. Y el análisis prescriptivo se encarga de prescribir o realizar recomendaciones en base de los hallazgos propios del uso de distintas herramientas, por ejemplo, dentro de este análisis se encuentra la simulación de Montecarlo.

Debido a que la analítica empresarial es un campo amplio, por motivos de centrar y limitar esta investigación, el desarrollo de este artículo se enfocó en el estudio del análisis predictivo, específicamente haciendo uso del modelo de regresión lineal.

El uso de la regresión lineal dentro del campo del análisis predictivo ha tenido aplicaciones en diferentes industrias y sectores organizacionales, demostrando su versatilidad y efectividad, como se muestra a continuación en los siguientes párrafos.

La investigación titulada Modelo para estimar el desempeño fiscal municipal utilizando un método de regresión lineal múltiple, presentada por Madrigal, tuvo como objetivo principal desarrollar un modelo de regresión para predecir el desempeño en las evaluaciones fiscales y poder desarrollar y ejecutar mejoras en la gestión. La investigadora tuvo como principal resultado, el desarrollo del siguiente modelo de regresión múltiple: $DFM = 0.080971 + 1.231325 * X1 \text{ capacidad fiscal} + 0.013595 * X4 \text{ autonomía tributaria} + 8.02E-05 * X5 \text{ predial} + ut$

Como indica la autora, se identificaron los factores relevantes que influyen en el desempeño fiscal municipal. Los resultados contribuyeron a la teoría de la gestión pública, sirviendo como medida de incentivo para los municipios con mayor y menor puntaje en su desempeño fiscal [5].

Para el desarrollo de este modelo de predicción, la autora indica en su investigación que utilizó el software Eviews.

En la investigación titulada Predicción de la efectividad de las pruebas rápidas realizadas a pacientes con COVID-19 mediante regresión lineal y random forest, tuvo como objetivo principal determinar la efectividad de las pruebas rápida mediante la regresión lineal de random forest haciendo uso del lenguaje de programación Python [6].

Dentro de la metodología del desarrollo de este estudio, los autores redactaron el proceso de análisis y códigos de programación en Python. Se detalla la importación de paquetes y librerías, preparación de los datos, asignación e identificación de variables, creación del modelo de regresión lineal y evaluación del modelo de regresión. Los autores obtuvieron el modelo de regresión $y = 1.14x + 7.52$. Como principales conclusiones, indican que la predicción de dicho modelo ha dado un error aceptable [6].

En la investigación titulada A Comparative Study of Linear Regression and Regression Tree, los autores tuvieron como principal objetivo realizar un modelamiento predictivo usando la técnica del modelo de regresión lineal múltiple y un

árbol de regresión, para luego comparar los resultados de dichos modelos y evaluar cuál de ellos presenta mejores predicciones. Utilizaron datos de uso libre de un repositorio de Machine Learning para poder ejecutar su estudio. Obtuvieron como resultado un MSE (error cuadrático medio) de 0.0682 y 0.0898 para el modelo de árbol de regresión y la regresión lineal múltiple respectivamente. En sus conclusiones indican que el modelo de árbol de regresión otorgó mejores resultados que el modelo de regresión lineal en función de los valores del MSE. Este hallazgo indica que el árbol de regresión tiene más precisión en la predicción [3].

Como se aprecia hasta este punto, el modelamiento predictivo presenta aplicaciones en distintas industrias y sectores, mostrando su relevancia como un análisis que debe estar siempre presente en las organizaciones.

Para poder realizar y aplicar una herramienta de análisis predictivo, como es el caso del modelo de regresión lineal, en cualquiera de sus variantes (simple, múltiple, polinomial, etc.) se requiere de ciertos componentes para poder ejecutarlo. Dichos componentes se muestran a continuación en la Fig. 2.

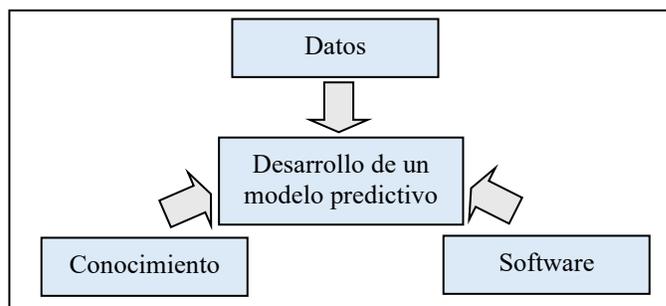


Fig. 2 Componentes para desarrollar un modelo predictivo.

La figura nos muestra que los principales componentes para el desarrollo de un modelo predictivo son: datos, uso de un software y conocimientos.

Los datos son la materia prima de todo tipo de análisis. No se puede obtener un producto final (modelo predictivo) sin materia prima. Para ello, es necesario que los datos cumplan con ciertas características, las cuales son calidad y cantidad. La calidad hace referencia a que los datos sean confiables (hayan sido correctamente registrados) y que se encuentren registrados las variables predictoras que realmente puedan estar influyendo en la variable dependiente o la variable de interés para el modelo predictivo. Para que el producto final cumpla con características de calidad, la materia prima debe cumplir con estas características de calidad.

La cantidad de los datos también influye en la efectividad del modelo de predicción. Por ello, no se recomienda trabajar este tipo de análisis con pocos datos. Poco y mucho puede ser realmente subjetivo. Una cantidad mínima adecuada de datos u observaciones puede ser en un rango de 40 a 50 registros en adelante, dependiendo del contexto.

Con respecto al software, se debe de contar con un programa, aplicación, paquete estadístico, o un entorno adecuado para poder trabajar los datos y desarrollar el modelo

de predicción. Actualmente existen muchos paquetes estadísticos para hacer diferentes tipos de análisis. Entre los más populares se pueden mencionar a Microsoft Excel, SPSS, Minitab, Jamovi, plataformas que trabajan con lenguaje de programación como Python, R, entre otros.

Todos ellos entregarán los mismos resultados si es que se ingresan y procesan los mismos datos. El inconveniente con el componente software, es que algunos de ellos requieren uso de licencias, las cuales no todos los usuarios pueden adquirir. Otro inconveniente que presenta este componente es la interfaz de usuario, hay algunos softwares que son más intuitivos y fáciles de usar que otros. Mientras que otros entornos como RStudio y Jupyter Notebook, requieren conocimiento en lenguaje de programación. Esto origina que el proceso de análisis y desarrollo del modelo de predicción sea más complicado o fácil dependiendo del software que se utilice.

Por último, tenemos el componente del conocimiento. Es el más importante de todos. Podemos tener datos de calidad, un software preparado y saber cómo utilizarlo, pero si no tenemos el conocimiento que engloba a la interpretación del modelo de predicción, será todo en vano.

El conocimiento implica todo el proceso de desarrollo, desde la extracción y tratamiento de los datos hasta la interpretación del modelo de predicción y realizar predicciones en función de este modelo. Se debe tener en cuenta cómo preparar adecuadamente los datos, descartar variables predictoras, saber interpretar adecuadamente el intercepto y coeficientes del modelo de regresión lineal y saber cómo evaluar el rendimiento del modelo de predicción.

Para tener todos estos conocimientos, los analistas o los usuarios interesados en realizar un modelo de predicción deben capacitarse y haber comprendido correctamente todos estos conceptos. Es por ello que no solo con tener los datos y un software es suficiente, si es que no se cuenta con los correctos conocimientos, este análisis no será útil.

Hasta este momento se puede apreciar que se tienen ciertas limitantes para poder desarrollar de forma correcta un modelo de predicción. Pero por motivos de extensión, este artículo solo se enfoca en las dificultades con respecto al componente del software.

¿Existirá alguna forma de facilitar el análisis con respecto al componente del software? ¿Tenemos otras alternativas u opciones en donde desarrollar dicho análisis?

En la investigación titulada *The role of ChatGPT in data science: How AI-assisted conversational interfaces are revolutionizing the Field*, los autores explican los beneficios y el apoyo que brinda ChatGPT (una aplicación de inteligencia artificial) en la ciencia de datos, es decir en el procesamiento, entrenamiento e interpretación de resultados.

En esta investigación se aprecia que los autores utilizaron ChatGPT como un apoyo en la programación con código Python, se aprecia la conversación e interacción entre el programador y ChatGPT.

Los autores mencionan que ChatGPT no siempre está en lo correcto, indican que, si el texto de entrada es ambiguo y no

se brinda el contexto adecuado, esta aplicación de inteligencia artificial puede generar respuestas incorrectas. Además de ello, aunque está entrenado en muchos campos, podría no tener suficiente conocimiento en temas específicos. Otro inconveniente es que los datos de entrenamiento de ChatGPT provienen de diferentes fuentes de internet, lo cual puede tener información incorrecta, afectando la calidad de sus respuestas. Una de sus principales conclusiones indica que ChatGPT está influyendo en la ciencia de datos, ya que es una poderosa herramienta que está ayudando a los científicos de datos en diferentes formas. Es probable que su uso se generalice aún más [7].

En la investigación titulada *Exploring the roles of AI-Assisted chatGPT in the field of data science*, aborda el potencial de ChatGPT como una herramienta de apoyo a los científicos de datos para el análisis, interpretación de los datos y como ello mejora la toma de decisiones en las industrias.

Para llevar a cabo esta investigación, los autores utilizaron una metodología de enfoque mixto. Recopilaron un conjunto de datos de diferentes dominios, aplicaron técnicas de ajuste fino mediante aprendizaje por transferencia.

Los autores indican que los modelos ajustados por ChatGPT mostraron mejoras sustanciales en tareas específicas, lo que indica un gran potencial en diferentes campos para la ciencia de datos. Además de ello, la técnica de optimización de recursos demostró la capacidad para crear modelos más eficientes sin perder rendimiento, lo que hace que ChatGPT sea una herramienta adecuada para entornos que posean recursos limitados [8].

Hasta este punto se han mencionado dos investigaciones en las cuales se ha explorado como el uso de una aplicación de inteligencia artificial (ChatGPT) contribuye al análisis, apoya con la programación y con tareas específicas dentro del campo de la ciencia de datos. Se ha mencionado que esta aplicación no siempre brinda respuestas correctas. Por otro lado, también se encuentra el dilema ético del uso de estas aplicaciones de inteligencia artificial, el cual se mencionará y desarrollará más adelante. Es por ello, que surgen las siguientes preguntas: ¿Cómo una aplicación de inteligencia artificial podrá contribuir a desarrollar un modelo de predicción? Además, ¿es práctico y confiable que una aplicación de inteligencia artificial desarrolle todo el análisis de datos?

II. MATERIALES Y MÉTODOS

Para el desarrollo de esta investigación se utilizó lo siguiente:

A. Herramientas

Se utilizó la plataforma de Google Colab, para realizar la programación en código Python y la aplicación de ChatGPT 4.0. Se desarrolló un modelo de predicción en ambos entornos donde se analizaron los mismos datos.

B. Datos

Se usó una base de datos de acceso público, llamada tips, donde se muestran variables del contexto de un restaurante.

Esta base de datos cuenta con un total de 245 registros. Se encuentra dentro de la librería de Seaborn.

C. Procedimiento

Se desarrolló un modelo de predicción, en este caso, un modelo de regresión lineal múltiple utilizando los datos de la base de datos tips, primero en el entorno de Google Colab, usando lenguaje de programación Python. Luego de haber obtenido el modelo de regresión múltiple, se realizó una serie de predicciones. Del mismo modo se utilizó ChatGPT 4.0 para desarrollar un modelo de regresión múltiple utilizando los mismos datos y se solicitó que realice las predicciones utilizando los mismos valores de entrada que se usaron en el desarrollo convencional del modelo de predicción.

Posterior a ello, se realizó una comparación de los resultados de los diferentes métodos (convencional y asistido por inteligencia artificial), validando si se obtuvieron los mismos resultados, además de poder determinar ventajas, desventajas y consideraciones para tener en cuenta al trabajar con el método asistido por inteligencia artificial.

III. RESULTADOS Y DISCUSIÓN

A continuación, en la Tabla I, se muestra los datos utilizados para el desarrollo del modelo de regresión múltiple:

TABLA I
VARIABLES DE LA BASE DE DATOS TIPS

total_bill	tip	sex	smoker	day	time	size
16.99	1.01	Female	No	Sun	Dinner	2
10.34	1.66	Male	No	Sun	Dinner	3
21.01	3.5	Male	No	Sun	Dinner	3
...
22.67	2	Male	Yes	Sat	Dinner	2
17.82	1.75	Male	No	Sat	Dinner	2
18.78	3	Female	No	Thur	Dinner	2

En la tabla se aprecia los registros sobre las propinas y el consumo de un restaurante, mostrando el detalle por género, día, cantidad de personas y si se encuentran o no en la zona de fumador. Extraído de la biblioteca Seaborn [9]

A. Desarrollo del modelo de predicción en el entorno de Google Colab

Se utilizó el siguiente código para importar las librerías correspondientes al análisis de datos y obtener la base de datos.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
data = sns.load_dataset('tips')
```

Luego de obtener la base de datos que se muestran en la Tabla I, se realizó un gráfico de dispersión entre la cantidad de

propina (variable dependiente) y el total de la factura (variable predictor 1). Para ello se escribió lo siguiente:
sns.relplot(x = 'total_bill', y = 'tip', data = data)
plt.show()

Se obtuvo el siguiente gráfico el cual se puede observar en la Fig. 3.

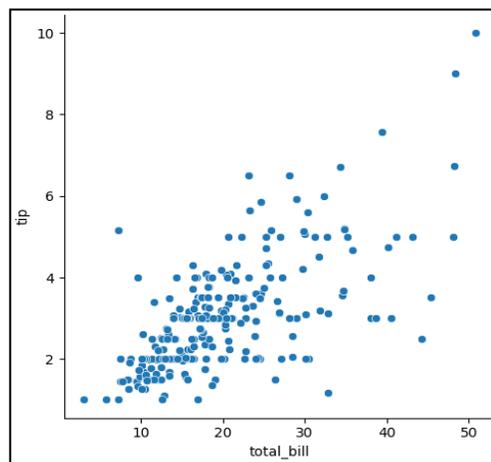


Fig. 3 Gráfico de dispersión entre el total de la factura y las propinas, realizado en Google Colab.

Posterior a ello, se realizó un gráfico de correlación entre el tamaño de la mesa y la cantidad de propina. Para obtener ello, se escribieron las siguientes líneas de código.

```
sns.relplot(x = 'size', y = 'tip', data = data)
plt.show()
```

Se obtuvo el siguiente gráfico. Ver Fig. 4.

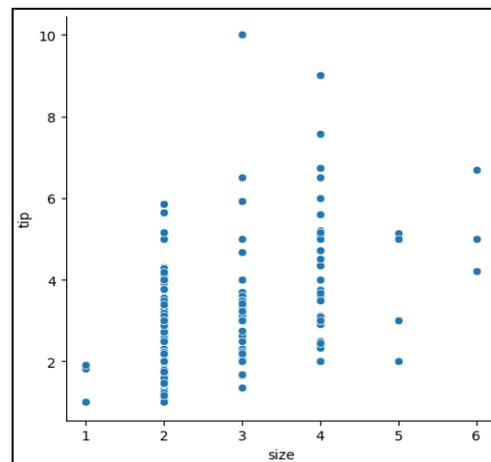


Fig. 4 Gráfico de dispersión entre el tamaño de la mesa y las propinas, realizado en Google Colab.

Como se aprecia en la Fig. 3 y Fig. 4, aparentemente existe una correlación positiva, es decir, una relación positiva entre dichas variables (el total de la factura, tamaño de la mesa y la cantidad de propina).

Es probable que las otras variables como por ejemplo el género, día, zona de fumador o no fumador, momento del día,

puedan influir en la cantidad de propinas, pero por razones prácticas y de extensión, solo se consideró como variables para desarrollar el modelo de predicción, la propina, el total de la factura y el tamaño de la mesa. El objetivo del desarrollo de este modelo de regresión múltiple será predecir la cantidad de propina en función del total de factura y el tamaño de la mesa.

Luego de realizar este breve análisis con los gráficos de dispersión, se procedió a asignar y nombrar las variables predictoras y la variable dependiente.

```
x = data[['total_bill', 'size']]
y = data['tip']
```

Posterior a ello, se separaron los datos en conjunto de entrenamiento y conjunto de prueba, en la proporción de 80% y 20% respectivamente.

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 42)
```

Luego de ello, se desarrolló el modelo de regresión lineal múltiple.

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(x_train, y_train)
```

Se creó el modelo de predicción y se mostró el error cuadrático medio de dicho modelo.

```
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
y_pred = model.predict(x_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"MSE: {mse}")
print(f"R2: {r2}")
```

El resultado de estas métricas del modelo de regresión fue lo siguiente: MSE: 0.648599619, R2: 0.481108409.

Se mostraron los valores del intercepto y de los coeficientes para desarrollar el modelo de regresión lineal múltiple

```
coeficiente = model.coef_
intercepto = model.intercept_
print(coeficiente)
print(intercepto)
```

El resultado fue lo siguiente: [0.09088, 0.26501], 0.568137092.

Es decir, el intercepto tiene un valor de 0.568. El total de factura (x1) y el tamaño de la mesa (x2), un valor de 0.091 y 0.265 respectivamente. Por lo que el modelo de regresión lineal múltiple se muestra de la siguiente manera:

$$Y = 0.568 + 0.091x_1 + 0.265x_2$$

Sí bien el modelo de regresión lineal múltiple presenta un R^2 de 0.481, un valor bajo, es decir, solo es capaz de explicar el 48.1% de la variación de las propinas en función del total de la factura y el tamaño de la mesa, el objetivo de este artículo es verificar si se tienen los mismos resultados y de forma más fácil utilizando la aplicación de inteligencia artificial.

Los valores que se utilizaron para realizar las predicciones son los siguientes:

Total de factura: 48.8, 20, 30.5, 37.5, 55.8

Tamaño de la mesa: 3, 1, 3, 2, 4

Para realizar esto en programación Python, se escribieron las siguientes líneas de código:

```
nuevos_datos = [[48.8,3], [20,1], [30.5,3], [37.5, 2], [55.8, 4]]
prediccion = model.predict(nuevos_datos)
print(f"La predicción es: {prediccion}")
```

Con ello se obtuvo los siguientes resultados predichos: 5.79819, 2.65077, 4.13506, 4.50621, 6.69937.

La interpretación de estos resultados indica lo siguiente. Para un consumo total de 48.8 dólares en una mesa con 3 personas, la cantidad de propina estimada es de 5.79 dólares. Para un consumo total de 20 dólares en una mesa con una persona, la cantidad de propina estimada es de 2.65 dólares. Para un consumo total de 30.5 dólares en una mesa con 3 personas, la cantidad estimada de propina es 4.14 dólares. Para un consumo total de 37.5 dólares en una mesa con 2 personas, la cantidad estimada de propina es 4.5 dólares. Finalmente, para un consumo total de 55.8 dólares en una mesa con 4 personas, la cantidad estimada de propina es 6.69 dólares.

B. Desarrollo del modelo de predicción asistido por ChatGPT

Para realizar este análisis dentro de este entorno, primero se adjuntaron los mismos datos en un archivo de Excel dentro de esta aplicación. Luego de ello se realizó la siguiente interacción:

Analista: Con estos datos adjuntos, realiza 2 gráficos de dispersión, teniendo en cuenta lo siguiente:

y = tips , x = total_bill

y= tips, x = size

ChatGPT: Aquí tienes los gráficos de dispersión solicitados:

Los gráficos que realizó la aplicación de inteligencia artificial se muestran a continuación. Ver Fig. 5 y Fig. 6.

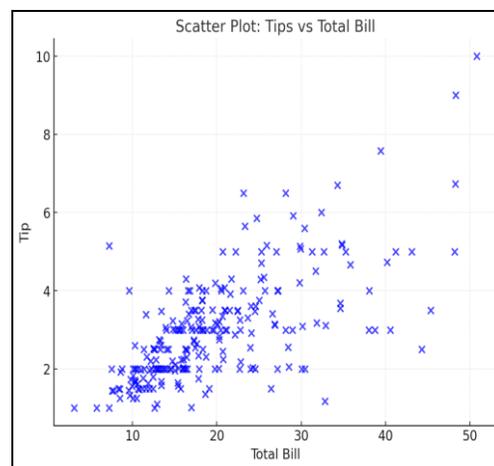


Fig. 5 Gráfico de dispersión entre total de la factura y las propinas, realizado por la inteligencia artificial.

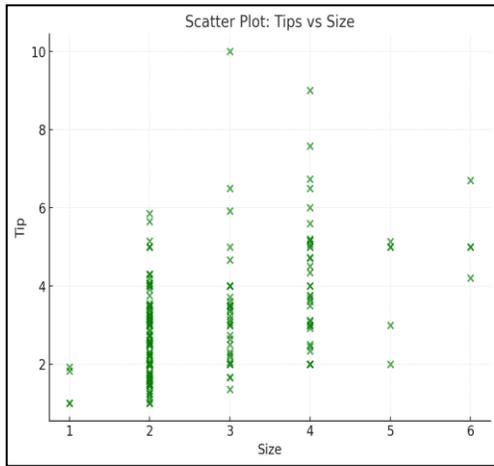


Fig. 6 Gráfico de dispersión entre el tamaño de la mesa y las propinas, realizado por la inteligencia artificial

Como se aprecia, la Fig. 5 y Fig. 6, son iguales (solo que presentan otro formato) a los gráficos de dispersión desarrollados bajo el procedimiento convencional en el entorno de Google Colab.

Luego de ello, se continuó la interacción con ChatGPT.

Analista: Con estos datos, realiza un modelo de regresión lineal múltiple teniendo en cuenta que la variable dependiente es tips y las variables independientes son total_bill y size. Muestra la ecuación. Para ello divide los datos en entrenamiento y prueba, con una proporción de 80% y 20% respectivamente, usando una semilla aleatoria de 42.

ChatGPT: El dataset contiene las siguientes columnas relevantes para el modelo:

total_bill: variable independiente.

size: variable independiente.

tip: variable dependiente.

Procederé a crear un modelo de regresión lineal múltiple utilizando las columnas relevantes y dividir los datos en conjuntos de entrenamiento (80%) y prueba (20%), usando una semilla aleatoria de 42.

El modelo de regresión lineal múltiple tiene la siguiente ecuación:

$$\text{tip} = 0.568 + 0.091 \text{ total_bill} + 0.265 \text{ size}$$

Error cuadrático medio (MSE): 0.649

Analista: calcula el R^2 y por último, realiza la predicción de las propinas si los valores del total de la factura y el tamaño de mesa presentaran estos valores:

Total de factura: 48.8, 20, 30.5, 37.5, 55.8

Tamaño de la mesa: 3, 1, 3, 2, 4

ChatGPT: Coeficiente de determinación: 0.481

Esto indica que el modelo explica el 48.1% de la variabilidad en las propinas a partir de las variables independientes.

Predicciones de las propinas:

Para los valores de factura y tamaño de la mesa proporcionados:

Total de factura: 48.8, Tamaño de la mesa: 3. Propina predicha: 5.80

Total de factura: 20.0, Tamaño de la mesa: 1. Propina predicha: 2.65

Total de factura: 30.5, Tamaño de la mesa: 3. Propina predicha: 4.14

Total de factura: 37.5, Tamaño de la mesa: 2. Propina predicha: 4.51

Total de factura: 55.8, Tamaño de la mesa: 4. Propina predicha: 6.70

Habiendo finalizado el desarrollo del modelo de predicción realizado por ambos métodos, se muestra a continuación una comparativa de los resultados obtenidos por el método convencional y el método asistido por la aplicación de inteligencia artificial, ChatGPT. Ver Tabla II.

TABLA II
EVALUACIÓN DEL DESARROLLO DEL MODELO DE PREDICCIÓN BAJO EL MÉTODO CONVENCIONAL Y ASISTIDO POR CHATGPT.

Actividad	Desarrollo Convencional	Asistido por ChatGPT
Gráficos de dispersión	Conforme	Conforme
División de los datos	Conforme	Conforme
Valor del intercepto	0.568	0.568
Valor del coeficiente x1	0.091	0.091
Valor del coeficiente x2	0.265	0.265
Valor de la predicción 1	5.798	5.8
Valor de la predicción 2	2.651	2.65
Valor de la predicción 3	4.135	4.14
Valor de la predicción 4	4.506	4.51
Valor de la predicción 5	6.699	6.7

Como se aprecia en la Tabla II, el método asistido por la aplicación de inteligencia artificial, ChatGPT, pudo realizar las mismas actividades de forma correcta, además de eso, se desarrolló el modelo de predicción de manera más rápida y fácil, coincidiendo exactamente con los resultados del método convencional (desarrollo del modelo de predicción con lenguaje de programación Python en el entorno de Google Colab). Se pudo desarrollar los gráficos de dispersión, la división de datos, mostrar la ecuación y realizar las predicciones de manera conforme.

Bajo el desarrollo mostrado anteriormente, se puede mencionar las siguientes ventajas de trabajar bajo asistencia de la aplicación de inteligencia artificial, ChatGPT:

En primer lugar, hay que destacar la facilidad al acceso del entorno, ya que es de uso gratuito y no requiere uso de licencias.

En segundo lugar, se encuentra la facilidad de solicitar o realizar actividades de análisis. No requiere conocimiento en

el manejo de un software o lenguaje programación como en otros paquetes estadísticos o entornos.

En tercer lugar, brinda interpretación de los resultados. Como se detalló párrafos anteriores, además de realizar los cálculos estadísticos, explicó que significan dichos números. Sin embargo, esto no quiere decir que todo el análisis e interpretación lo debe de realizar la aplicación de inteligencia artificial.

Hay que recordar los componentes mostrados en la Fig. 2. El componente más importante es el conocimiento. Si una persona no cuenta con conocimientos básicos para realizar cualquier tipo de análisis de datos, primero, debería capacitarse para posteriormente considerar utilizar alguna aplicación de inteligencia artificial como herramienta de apoyo o soporte.

Debemos recordar que ChatGPT (una aplicación de inteligencia artificial), es una herramienta la cual puede ser utilizada como soporte en algunos pasos del proceso de análisis de datos, como por ejemplo en el desarrollo de modelos predictivos. Y como toda herramienta, puede ser mal utilizada si es que no se cuenta con el conocimiento requerido, teniendo como consecuencia la generación de errores en el procedimiento del desarrollo y resultados.

En la investigación titulada *The Potential and Concerns of Using AI in Scientific Research*, los autores describen las distintas aplicaciones y usos de la inteligencia artificial en el ámbito de la investigación científica. Además de ello, los autores describieron algunas preocupaciones, siendo una de ellas en el aspecto ético. Indican que los investigadores deben ser transparentes en el uso de aplicaciones de inteligencia artificial en sus trabajos, indicar como se ha utilizado. Además de ello, los autores mencionan que existe la probabilidad de que la información o análisis que se haya realizado con inteligencia artificial no siempre es confiable, precisa o correcta [10].

Es por esta razón que no todo el trabajo o proceso de análisis lo debe de realizar la inteligencia artificial y menos aún si no se realiza la validación correspondiente de los resultados. Podemos utilizar a la inteligencia artificial como una herramienta para realizar parte de nuestro análisis, por ejemplo, si en un entorno o software tenemos dificultad para realizar un gráfico de dispersión, entonces podemos realizar este gráfico bajo asistencia de la aplicación de inteligencia artificial.

Teniendo esto en cuenta, se puede hacer uso de estas aplicaciones en el ámbito empresarial como una herramienta de apoyo para los analistas de datos. En la investigación titulada la importancia de la competitividad y su influencia en la rentabilidad de las organizaciones. Caso de estudio de una fábrica de calzado, se menciona que, si una empresa busca incrementar su rentabilidad, deberá primero mejorar su competitividad. Y la competitividad se ve influenciada en aspectos como la calidad de productos, productividad de procesos, costos de producción y sostenibilidad [11].

Esto representa una clara aplicación y uso de la inteligencia artificial como una herramienta de apoyo para el

desarrollo de modelos predictivos en el contexto empresarial, ya que se puede analizar factores que tienen algún tipo de asociación y predecir cual será el valor de una variable en función de otra, por ejemplo, la productividad de un proceso en función de la cantidad de trabajadores y el ritmo de trabajo de las máquinas, o quizás el grado de calidad de un producto en función de la cantidad y proporción de sus diferentes insumos, etc. Si se logra realizar este tipo de análisis de forma correcta, se podrá obtener beneficios en la organización.

Con los antecedentes y el caso práctico presentado en esta investigación, se puede concluir que las aplicaciones de inteligencia artificial, en este caso, ChatGPT, es una herramienta de apoyo muy útil para las actividades de análisis de datos, y siendo más específicos para el desarrollo de modelos predictivos. Sin embargo, hay que tener en consideración que las aplicaciones de inteligencia artificial cometen errores o no brindan el análisis o información correcta si es que no se ingresa de forma adecuada las instrucciones o no se cuenta con el conocimiento adecuado sobre el análisis que se desea realizar.

Finalmente, las aplicaciones de inteligencia artificial facilitan y agilizan el proceso de desarrollo de análisis de datos, sin embargo, debemos recordar que son solo una herramienta, y por lo tanto, también debemos hacer un uso adecuado y ético de estas.

REFERENCIAS

- [1] E. Medina, "Analítica: Tendencia para optimizar la toma de decisiones a nivel empresarial," *Dictam. Libr.*, vol. 22, no. 29, pp. 54–66, 2021.
- [2] R. Treviño, F. Rivera, and A. Garza, "La analítica de datos como ventaja competitiva en las organizaciones," *Vinculatégica EFAN*, vol. 6, no. 2, pp. 1063–1074, 2020, doi: 10.29105/vtga6.2-520.
- [3] P. Jadhav, V. Patil, and S. Gore, "A Comparative Study of Linear Regression and Regression Tree," *Int. Conf. Commun. Inf. Process.*, 2020, doi: 10.2139/ssrn.3645883.
- [4] A. Molina, "El misterio de los calcetines desaparecidos: Calidad de un modelo de regresión lineal," *Rev. Electron. AnestesiaR*, vol. 11, no. 6, pp. 1–6, 2024.
- [5] G. Madrigal, "Modelo para estimar el desempeño fiscal municipal utilizando un método de regresión lineal múltiple," *Econ. Soc. y Territ.*, vol. 24, no. 74, pp. 1–23, 2024, doi: 10.22136/est20241939.
- [6] D. Patiño, C. Munive, L. Cevallos, and M. Botto, "Predicción de la Efectividad de las Pruebas Rápidas Realizadas a Pacientes con COVID-19 mediante Regresión Lineal y Random Forest," *Ecuadorian Sci. J.*, vol. 5, no. 2, pp. 31–43, 2021, doi: 10.46480/esj.5.2.108.
- [7] H. Hassani and E. S. Silva, "The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field," *Big Data Cogn. Comput.*, vol. 7, no. 2, 2023, doi: 10.3390/bdcc7020062.
- [8] L. Valli, N. Sujatha, M. Mech, and V. Lokesh, "Exploring the roles of AI-Assisted ChatGPT in the field of data science," *E3S Web Conf.*, vol. 491, pp. 1–14, 2024, doi: 10.1051/e3sconf/202449101026.
- [9] M. Waskom, "seaborn: statistical data visualization," 2021. <https://seaborn.pydata.org/>
- [10] Z. Khlaif et al., "The Potential and Concerns of Using AI in Scientific Research: ChatGPT Performance Evaluation," *JMIR Med. Educ.*, vol. 9, no. 1, 2023, doi: 10.2196/47049.
- [11] J. Calderón, "La importancia de la competitividad y su influencia en la rentabilidad de las organizaciones. Caso de estudio de una fábrica de calzado," *Glob. Bus. Adm. J.*, vol. 8, no. 2, pp. 1–14, 2024, [Online]. Available: https://revistas.urp.edu.pe/index.php/Global_Business