BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FCC
INGENIERIA EN TECNOLOGIAS DE LA INFORMACIÓN

# INTELIGENCIA DE NEGOCIOS

## ACTIVIDAD 7
## “REGRESIÓN LOGISTICA”

EQUIPO:
DIEGO BUSTAMANTE DOMINGUEZ
ROCÍO RAMÍREZ FABIÁN

PROFESOR: Alfredo García Suárez

31/03/2025

# INTRODUCCIÓN

La regresión logística es una técnica de análisis de datos que utiliza las matemáticas para encontrar las relaciones entre dos factores de datos. Luego, utiliza esta relación para predecir el valor de uno de esos factores basándose en el otro. Normalmente, la predicción tiene un número finito de resultados, como un sí o un no.

Por ejemplo, supongamos que desea adivinar si el visitante de su sitio web va a hacer clic en el botón de pago de su carrito de compras o no. El análisis de regresión logística analiza el comportamiento de los visitantes anteriores, como el tiempo que permanecen en el sitio web y la cantidad de artículos que hay en el carrito. Determina que, si anteriormente los visitantes pasaban más de cinco minutos en el sitio y agregaban más de tres artículos al carrito, hacían clic en el botón de pago. Con esta información, la función de regresión logística puede predecir el comportamiento de un nuevo visitante en el sitio web.

La regresión logística es una técnica importante en el campo de la inteligencia artificial y el machine learning (AI/ML). Los modelos de ML son programas de software que puede entrenar para realizar tareas complejas de procesamiento de datos sin intervención humana. Los modelos de ML creados mediante regresión logística ayudan a las organizaciones a obtener información procesable a partir de sus datos empresariales. Pueden usar esta información para el análisis predictivo a fin de reducir los costos operativos, aumentar la eficiencia y escalar más rápido. Por ejemplo, las empresas pueden descubrir patrones que mejoran la retención de los empleados o conducen a un diseño de productos más rentable.

**Pasos que se siguió para realizar la regresión logística :**
**(México, Roma, Venecia)**

1. Carga de librerias en Python y carga del archivo con Pandas.

```
#Cargamos librerias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from funpymodeling.exploratory import freq_tbl
import scipy.special as special
from scipy.optimize import curve_fit
import seaborn as sns
from sklearn.metrics import r2_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
```

✓ 1.8s

```
df = pd.read_csv('Mexico.csv')
```

✓ 0.9s

2. Limpieza de datos (relleno de valores nulos).

```
df= df.fillna(method="bfill")
df= df.fillna(method="ffill")
df = df.bfill()
df= df.ffill()
```

✓ 0.6s

3. Convertimos las variables que fueron necesarias en variables.

**Transformando variables a tipo dicotómicas**

```
#bathrooms_text
df["bathroomsd"] = df["bathrooms_text"].astype(str).apply(lambda x: 1 if "2" in x or "3" in x or "4" in x else 0)
#room_type Entire home
df["E_room_typed"] = df["room_type"].apply(lambda x: 1 if x == "Entire home/apt" else 0)
#room_type Private home
df["P_room_typed"] = df["room_type"].apply(lambda x: 1 if x == "Private room" else 0)
#room_type Hotel home
df["H_room_typed"] = df["room_type"].apply(lambda x: 1 if x == "Hotel room" else 0)
#instant_bookabled
df["instant_bookabled"] = df["instant_bookable"].apply(lambda x: 1 if x == "t" else 0)
#Price
df["price"] = df["price"].replace('[\$,]', '', regex=True).astype(float)
Mprice = df["price"].median()
df["D_price"] = df["price"].apply(lambda x: 1 if x >= Mprice else 0)
#host_total_listings_count
Mlistings = df["host_total_listings_count"].median()
df["D_host_listings"] = df["host_total_listings_count"].apply(lambda x: 1 if x >= Mlistings else 0)
#host_total_listings_count
Maccuracy = df["host_total_listings_count"].median()
df["D_accuracy"] = df["review_scores_accuracy"].apply(lambda x: 1 if x >= Maccuracy else 0)
#availability_30
Mavailability = df["availability_30"].median()
df["D_availability"] = df["availability_30"].apply(lambda x: 1 if x >= Mavailability else 0)
#availability_90
Mavailability90 = df["availability_90"].median()
df["D_availability90"] = df["availability_90"].apply(lambda x: 1 if x >= Mavailability90 else 0)
```

4. Selección de variables dependientes e independientes, dividimos el conjunto de datos en la parte de entrenamiento con el 30% y prueba con el resto de los datos.

```
#Declaramos las variables dependientes e independiente para las regresion Logisitica
Vars_Indep = df[['D_accommodates', 'bathroomsd', 'E_room_typed']]
Var_Dep = df['D_price']
X = Vars_Indep
y = Var_Dep

#Dividimos el conjunto de datos en la parte de entrenamiento y prueba:
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state= None)

#Se escalan los datos
escalar = StandardScaler()
#Para realizar el escalamiento de las variables "X" tanto de entrenamiento como de prueba, utilizaremos
X_train = escalar.fit_transform(X_train)
X_test = escalar.transform(X_test)

#Definimos el algoritmo a utilizar
from sklearn.linear_model import LogisticRegression
algoritmo = LogisticRegression()

algoritmo.fit(X_train, y_train)

y_pred = algoritmo.predict(X_test)
y_pred
✓ 0.0s
```

5. Obtuvimos la matriz de confusión

```
#Verifico la matriz de Confusion
from sklearn.metrics import confusion_matrix
matriz = confusion_matrix(y_test, y_pred)
print('Matriz de Confusion:')
print(matriz)
✓ 0.0s

Matriz de Confusion:
[[1894 2024]
 [ 601 3366]]
```

6. Calculamos la precisión del modelo

```
#Caulculo de la precisuin del modelo
from sklearn.metrics import precision_score

precision = precision_score(y_test, y_pred, average='binary')
print("Precision del modelo: ")
print(precision)
✓ 0.0s

Precision del modelo:
0.6244897959183674
```

7. Calculamos la Exactitud modelo

```
from sklearn.metrics import accuracy_score

exactitud = accuracy_score(y_test, y_pred)
print("Exactitud del modelo: ")
print(exactitud)
```

8. Finalmente calculamos la sensibilidad del modelo

```python
#Calculo la sensibilidad del modelo
from sklearn.metrics import recall_score
sensibilidad = recall_score(y_test, y_pred, average="binary")
print('Sensibilidad del modelo:')
print(sensibilidad)
```

✓ 0.0s

```
Sensibilidad del modelo:
0.8485001260398286
```

## Resultados

| CASO | VARIABLE DEPENDIENTE | VARIABLES INDEPENDIENTES | PAÍS | PRECISIÓN | EXACTITUD | SENSIBILIDAD |
|---|---|---|---|---|---|---|
| 1 | price | accomodates<br>bathrooms<br>room_type | México | 0.6244898 | 0.66708941 | 0.848500126 |
| | | | Roma | 0.6591727 | 0.64623704 | 0.643546971 |
| | | | Venecia | 0.6399383 | 0.65881363 | 0.706984668 |
| 2 | availability_30 | price<br>host_listing<br>accuracy | México | 0.5240036 | 0.52092846 | 0.847619048 |
| | | | Roma | 0.6706096 | 0.60703019 | 0.591018444 |
| | | | Venecia | 0.5404438 | 0.52376946 | 0.606425703 |
| 3 | price | availability<br>number_of_reviews<br>review_scores_rating | México | 0.5394243 | 0.53805175 | 0.544604498 |
| | | | Roma | 0.6086957 | 0.59125732 | 0.51618705 |
| | | | Venecia | 0.588137 | 0.57088767 | 0.57189277 |
| 4 | availability_30 | bathrooms<br>number_of_reviews price | México | 0.541129 | 0.53640284 | 0.660595619 |
| | | | Roma | 0.681777 | 0.61514196 | 0.599203187 |
| | | | Venecia | 0.5393701 | 0.53639041 | 0.672667758 |
| 5 | price | accomodates<br>review_scores_rating<br>availability_30 | México | 0.5549537 | 0.58929477 | 0.971065104 |
| | | | Roma | 0.6463621 | 0.65660207 | 0.689530686 |
| | | | Venecia | 0.6526807 | 0.65376525 | 0.690789474 |
| 6 | price | instant_bookable<br>review_scores_rating<br>availability_90 | México | 0.5154561 | 0.51128869 | 0.51584507 |
| | | | Roma | 0.5946429 | 0.58900406 | 0.59252669 |
| | | | Venecia | 0.6526807 | 0.65376525 | 0.690789474 |
| 7 | number_of_review | bathrooms<br>availability_30<br>host_listing | México | 0.5355217 | 0.52866565 | 0.568300572 |
| | | | Roma | 0.5531915 | 0.56421812 | 0.616226071 |
| | | | Venecia | 0.5930521 | 0.58056374 | 0.586743044 |
| 8 | number_of_review | price<br>accommodates<br>review_scores_rating | México | 0.5526642 | 0.56037544 | 0.673038229 |
| | | | Roma | 0.5678571 | 0.568274 | 0.572972973 |
| | | | Venecia | 0.5281276 | 0.53260412 | 0.533955857 |
| 9 | price | bathrooms<br>review_scores_accuracy<br>review_scores_rating | México | 0.7717292 | 0.64408929 | 0.422383575 |
| | | | Roma | 0.7665953 | 0.60477693 | 0.317939609 |
| | | | Venecia | 0.6992665 | 0.63819941 | 0.482293423 |
| 10 | price | vailabilit_90<br>host_total_listings_count<br>number_of_reviews | México | 0.509901 | 0.50761035 | 0.519939425 |
| | | | Roma | 0.5838565 | 0.5750338 | 0.576106195 |
| | | | Venecia | 0.534005 | 0.53344552 | 0.534453782 |

# CONCLUSIÓN

La regresión logística es una técnica fundamentales una técnica fundamental en el análisis de datos cuando se busca modelar relaciones entre variables y predecir resultados categóricos . Su principal ventaja es la capacidad de estimar la probabilidad de que ocurra un evento. en función de diferentes factoresen función de diferentes factores, lo que la hace útil en Múltiples áreas, desde la economía hasta la ciencia de datos.

Durante su aplicación, es crucial la correcta selección y transformación de variables , especialmente cuando se trabaja con datos que requieren conversión a un formato dicotómico. Además, la evaluación del modelo mediante precisión, exactitud y sensibilidad permite determinar su rendimiento y confiabilidad.