



Tecnológico de Monterrey

Campus Puebla

José Antonio Villanueva Arenas A01737550

Analítica de datos y herramientas de inteligencia artificial II

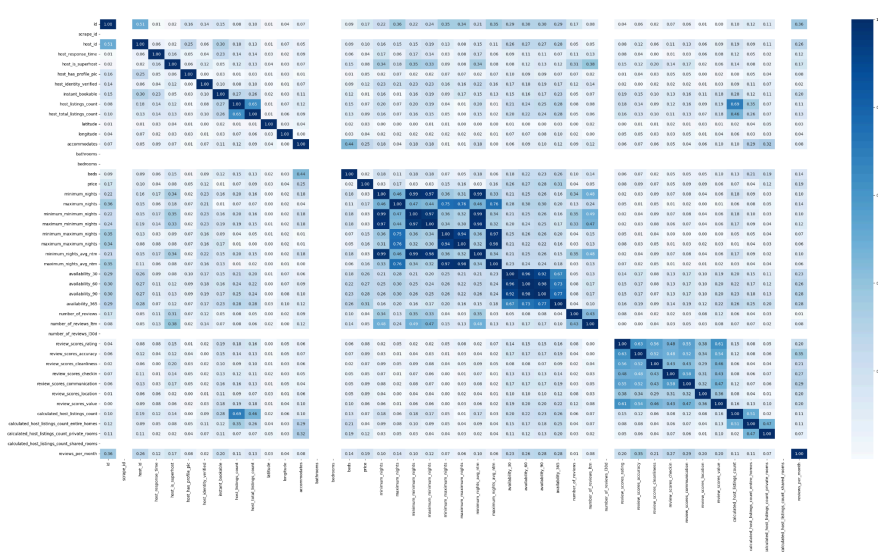
Actividad 4.1 regresión logística

25 Abril 2025

Reporte

Se realizaron modelos de regresión logística para identificar precisión, exactitud y sensibilidad de las variables objetivo: `host_is_superhost`, `host_has_profile_pic`, `host_identity_verified`, `instant_bookable`, `host_response_time`, `price`, `reviews_per_month`, `accommodates`, `minimum_nights`, `minimum_nights`. En este caso se analizaron los datos de Airbnb Berlín.

En primer lugar, se convirtieron algunas variables a numéricas y se generó un Heat Map para obtener las correlaciones más altas y usar las tres correlaciones más altas como variables independientes.



El primer modelo realizado tuvo los siguientes resultados:

Variable dependiente: `host_is_superhost`

Variables independientes: `number_of_reviews_ltm`, `minimum_nights`, `minimum_nights`.

Precisión	Exactitud	Sensibilidad
0.52, 0.80	0.78	0.21, 0.94

Este modelo tiene un cierto desbalance de clases, por lo que se puede decir que no detecta de forma correcta a los verdaderos superhosts.

El segundo modelo realizado tuvo los siguientes resultados:

Variable dependiente: `host_has_profile_pic`

Variables independientes: `host_id`, `id`, `availability_30`

Precisión	Exactitud	Sensibilidad
0.976, 0	0.976	1.0, 0

Este modelo tiene un claro desbalance de clases.

El tercer modelo realizado tuvo los siguientes resultados:

Variable dependiente: host_identity_verified

Variables independientes: maximum_minimum_nights, minimum_minimum_nights, minimum_nights

Precisión	Exactitud	Sensibilidad
0.91, 0	0.90	0.99, 0

Este modelo tiene un claro desbalance de clases.

El cuarto modelo realizado tuvo los siguientes resultados:

Variable dependiente: instant_bookable

Variables independientes: host_id, calculated_host_listings_count, host_listings_count

Precisión	Exactitud	Sensibilidad
0.59, 0.77	0.74	0.28, 0.92

Aunque el modelo tiene buena exactitud, se ve un desbalance de clase significativo.

El quinto modelo realizado tuvo los siguientes resultados:

Variable dependiente: host_response_time

Variables independientes: instant_bookable, minimum_minimum_nights, minimum_nights_avg_ntm

Precisión	Exactitud	Sensibilidad
0.64, 0-46	0.59	0.73, 0.36

El modelo no es malo, aunque la exactitud y la precisión podrían estar un poco más altas, lo que lo haría un modelo confiable para predecir, si el tiempo de respuesta es menor a una hora o mayor.

El sexto modelo realizado tuvo los siguientes resultados:

Variable dependiente: price

Variables independientes: availability_365, availability_90, availability_60

Precisión	Exactitud	Sensibilidad
------------------	------------------	---------------------

0.70, 0.67	0.69	0.73, 0.64
------------	------	------------

El modelo tiene buen equilibrio y métricas aceptables, lo que significa que la disponibilidad sí influye en tener un precio mayor y se puede hacer una predicción relativamente acertada.

El séptimo modelo realizado tuvo los siguientes resultados:

Variable dependiente: reviews_per_month

Variables independientes: id, review_scores_accuracy, review_scores_communication

Precisión	Exactitud	Sensibilidad
0.0, 0.94	0.94	0.0, 1

Este modelo tiene un claro desbalance de clases.

El octavo modelo realizado tuvo los siguientes resultados:

Variable dependiente: accommodates

Variables independientes: beds, calculated_host_listings_count_private_rooms, calculated_host_listings_count_entire_homes

Precisión	Exactitud	Sensibilidad
0.63, 0.80	0.77	0.44, 0.90

El modelo tiene resultados no muy altos ni muy bajos, sin embargo, por la sensibilidad, no se recomienda usarlo para hacer una predicción.

El noveno modelo realizado tuvo los siguientes resultados:

Variable dependiente: minimum_nights

Variables independientes: minimum_nights_avg_ntm, minimum_minimum_nights, maximum_minimum_nights

Precisión	Exactitud	Sensibilidad
0.95, 0.99	0.99	0.75, 0.99

El modelo permite hacer buenas predicciones debido a sus resultados.

El décimo modelo realizado tuvo los siguientes resultados:

Variable dependiente: maximum_nights

Variables independientes: maximum_maximum_nights, maximum_nights_avg_ntm, minimum_maximum_nights

Precisión	Exactitud	Sensibilidad
0.77, 0.95	0.86	0.94, 0.81

Este modelo tiene un muy buen desempeño y sin problemas se puede usar para hacer una predicción.

Comparativa

En primer lugar, los modelos de las variables: `maximum_nights` y `minimum_nights` tienen muy buenos resultados en las métricas de precisión, exactitud y sensibilidad. Por lo tanto, estos son los mejores modelos para realizar predicciones. Por otro lado, los peores modelos fueron `host_is_superhost`, `instante_bookable` y `reviews_per_month`, los primeros dos mostraron una sensibilidad muy baja y el de `reviews_per_month` fue un modelo fallido. Por lo tanto, se recomendaría probar con otra variable o analizar el porqué de los resultados. Y en los otros modelos se ve claramente un desbalance de clases.

Conclusiones

Es muy interesante cómo en el mismo conjunto de datos se pueden encontrar algunos modelos muy buenos y otros con resultados poco satisfactorios. Con este análisis se puede ver la importancia de usar el conjunto de métricas disponible, ya que analizar solo una de las tres puede sesgar los resultados y hacerte creer que el modelo es mejor de lo que realmente es.