



Tecnológico de Monterrey

Actividad 4.1 (Regresión Logística)

Lucero Jannete López García A01736938

**Analítica de datos y herramientas de inteligencia artificial
I (Gpo 101)**

Fecha
22/04/2025

Variable dependiente	Variables independientes	Precisión con variable objetivo 'f'	Precisión con variable objetivo 't'	Sensibilidad con variable objetivo 'f'	Sensibilidad con variable objetivo 't'	Exactitud
host_is_superuserhost	minimum_nights', 'minimum_nights', 'minimum_nights_avg_ntm'	0.5985	0.47	0.9778	0.03	0.5957
Observaciones	Dan mejores resultados con la variable objetivo donde el el host no es super_host					

Variable dependiente	Variables independientes	Precisión con variable objetivo 'Less than a day'	Precisión con variable objetivo 'More than a day'	Sensibilidad con variable objetivo 'Less than a day'	Sensibilidad con variable objetivo 'More than a day'	Exactitud
host_response_time	accommodates', 'review_scores_rating', 'bedrooms'	0.9606	0	1	0	0.9606
Observaciones	Dan mejores resultados dónde la variable objetivo del tiempo de respuesta es menos de un día, esto ya sea porque que tiene más datos.					

Variable dependiente	Variables independientes	Precisión con variable objetivo 'City scrape'	Precisión con variable objetivo 'More than a day'	Sensibilidad con variable objetivo 'City scraped'	Sensibilidad con variable objetivo 'More than a day'	Exactitud
Source	minimum_nights_avg_ntm', 'maximum_minimum_nights', 'host_listing'	0.9367	0	1	0	0.9367

	s_count'					
Observaciones	Dan mejores resultados dónde la variable objetivo del tipo de fuente es derivado de datos recientes.					

Variable dependiente	Variables independientes	Precisión con variable objetivo 'Good score'	Precisión con variable objetivo 'Bad score'	Sensibilidad con variable objetivo 'Good score'	Sensibilidad con variable objetivo 'Bad score'	Exactitud
review_scores_rating	minimum_nights_avg_ntm', 'maximum_minimum_nights', 'host_listings_count'	0.7436	0	1	0	0.7436
Observaciones	Dan mejores resultados dónde la variable objetivo de la puntuación es buena					

Variable dependiente	Variables independientes	Precisión con variable objetivo 'Alta demanda'	Precisión con variable objetivo 'Baja demanda'	Sensibilidad con variable objetivo 'Alta demanda'	Sensibilidad con variable objetivo 'Baja demanda'	Exactitud
availability_90	minimum_nights_avg_ntm', 'maximum_minimum_nights', 'host_listings_count'	0.8472	0	1	0	0.8472
Observaciones	Dan mejores resultados dónde la variable objetivo de la disponibilidad es alta, al parecer es constante y hay más datos					

Variable dependiente	Variables independientes	Precisión con variable objetivo 'Muchas reseñas'	Precisión con variable objetivo 'Pocas reseñas'	Sensibilidad con variable objetivo 'Muchas reseñas'	Sensibilidad con variable objetivo 'Pocas reseñas'	Exactitud
review_scores_cleanliness	reviews_per_month', 'review_scores_checkin', 'number_of_reviews'	0.7799	0.51	0.9695	0.09	0.7659
Observaciones	Dan mejores resultados dónde la variable objetivo de la puntuación de limpieza es con muchas reseñas, esto se debe a que hay más datos y están conforme con el servicio de limpieza.					

Variable dependiente	Variables independientes	Precisión con variable objetivo 'Precios bajos'	Precisión con variable objetivo 'Precios altos'	Sensibilidad con variable objetivo 'Precios bajos'	Sensibilidad con variable objetivo 'Precios altos'	Exactitud
Price	availability_30', 'maximum_minimum_nights', 'host_listings_count'	0.847	0	1	0	0.847
Observaciones	Hay mejores resultados cuando los precios son bajos.					

Variable dependiente	Variables independientes	Precisión con variable objetivo '1 cama'	Precisión con variable objetivo 'más de 2 camas'	Sensibilidad con variable objetivo '1 cama'	Sensibilidad con variable objetivo 'más de dos camas'	Exactitud
----------------------	--------------------------	--	--	---	---	-----------

beds	availability_30', 'maximum_minimum_nights', 'host_listings_count'	0.8716	0	1	0	0.8716
Observaciones	Hay mejores resultados cuando solo hay una cama por lugar					

Variable dependiente	Variables independientes	Precisión con variable objetivo 'Alta aceptación'	Precisión con variable objetivo 'Media alta aceptación'	Sensibilidad con variable objetivo 'Alta aceptación'	Sensibilidad con variable objetivo 'Media alta aceptación'	Exactitud
host_acceptance_rate	availability_30', 'maximum_minimum_nights', 'host_listings_count'	0.9189	0	1	0	0.9189
Observaciones	Hay mejores resultados cuando hay altas aceptaciones por parte del anfitrión.					

Variable dependiente	Variables independientes	Precisión con variable objetivo 'Pocas propiedades'	Precisión con variable objetivo 'Varias propiedades'	Sensibilidad con variable objetivo 'Pocas propiedades'	Sensibilidad con variable objetivo 'Varias propiedades'	Exactitud
calculated_host_listings_count	availability_30', 'maximum_minimum_nights', 'host_listings_count'	0.91	0.35	0.95	0.223	0.89

Observaciones	Hay mejores resultados cuando hay pocas propiedades por anfitrión debido a que no todo cuentan con varias.
---------------	--

Aunque en varios de los modelos la sensibilidad aparece como 1.0, esto no significa necesariamente que estén funcionando de manera perfecta. De hecho, al contrastar esta métrica con la precisión y la exactitud, se puede notar un comportamiento que merece revisión ya que, la sensibilidad de 1.0 en tantos modelos puede indicar dos cosas:

- Que el modelo podría estar prediciendo todo como clase positiva, lo cual es riesgoso porque no está aprendiendo a distinguir correctamente entre clases.
- Que puede existir un desequilibrio de clases en los datos, es decir, que hay muchas más observaciones de una clase que de la otra, y eso puede sesgar el aprendizaje del modelo.

Incluso, en algunos casos que revisé, la matriz de confusión indicaba que el modelo no estaba prediciendo ninguna clase negativa (por ejemplo: $\begin{bmatrix} 4049 & 0 \\ 344 & 0 \end{bmatrix}$).

De igual forma matrices que dan precisión cero y sensibilidad cero, esto debido al desbalance de los datos, que ocasiona que a los datos con poca información los ponga como cero.

En conclusión, es importante no dejarnos llevar solo porque la sensibilidad es perfecta. sino que también analice la precisión, la exactitud y la matriz de confusión es necesaria para entender el panorama completo. Solo así puedo estar seguro de que el modelo realmente está funcionando bien y no está dando resultados desbalanceados.