



Tecnológico de Monterrey

Actividad 4.1 (Regresión Logística)

Pilar Méndez Briones A01736843

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

Fecha

27/04/2025

1. **host_identity_verified**

Variable dependiente	Objetivo	Variables independientes	Precisión (t)	Precisión (f)	Exactitud	Sensibilidad (t)	Sensibilidad (f)	Matriz de confusión
host_identity_verified	Las variables objetivo son t y f	'availability_365','host_acceptance_rate','host_total_listings_count'	0.870	0.00	0.870	1.0	0.0	[[0 383] [0 2585]]
Observaciones	El modelo predijo correctamente todos los casos de hosts verificados ('t'), pero no pudo identificar ni un solo caso de hosts no verificados ('f'). El modelo tiene buena exactitud, pero porque los datos estaban desequilibrados: ser verificado es "la norma" en Airbnb y por eso hay muy pocos en 'f'.							

2. **host_response_time**

Variable dependiente	Objetivo	Variables independientes	Precisión (a day)	Precisión (a few days more')	Exactitud	Sensibilidad (a day)	Sensibilidad (a few days more')	Matriz de confusión
host_response_time	Las variables objetivo son 'a day' y 'a few days more'	'host_response_rate','host_acceptance_rate','host_listings_count'	0.954	0.00	0.954	1.0	0.0	[[2833 0] [135 0]]

Observaciones	El modelo predice perfectamente los casos de respuesta en un día ('a day'), pero no logra detectar ningún caso de respuesta más lenta. Como la mayoría de los anfitriones cumplen con responder en un día, el modelo aprende rápidamente este patrón dominante.
---------------	---

3. host_is_superhost

Variable dependiente	Objetivo	Variables independientes	Precisión (t)	Precisión (f)	Exactitud	Sensibilidad (t)	Sensibilidad (f)	Matriz de confusión
host_is_superhost	Las variables objetivo son t y f	'availability_365','number_of_reviews','host_listings_count'	0.373	0.814	0.796	0.082	0.967	[[2316 79] [526 47]]
Observaciones	El modelo es muy bueno prediciendo quién no es superhost, pero muy malo prediciendo quién sí es. Es fácil para el modelo suponer que alguien no es superhost, porque es lo más común en la plataforma, y por eso es que la precisión y sensibilidad de la variable 't' son muy bajas a comparación de 'f'.							

4. room_type

Variable dependiente	Objetivo	Variables independientes	Precisión (Entire home/apt)	Precisión (Public or shared room)	Exactitud	Sensibilidad (Entire home/apt)	Sensibilidad (Public or shared room)	Matriz de confusión
----------------------	----------	--------------------------	-----------------------------	-----------------------------------	-----------	--------------------------------	--------------------------------------	---------------------

room_type	Las variables objetivo son Entire home/apt y Public or shared room	'review_scores_rating', 'accommodates', 'price'	0.810	0.865	0.813	0.990	0.201	[[2278 21] [534 135]]
Observaciones	El modelo predice muy bien los alojamientos completos, pero tiene dificultad para los espacios compartidos. Los hospedajes completos tienen características más consistentes en el precio, tamaño y puntuación, lo cual facilita al modelo reconocerlos.							

5. host_listings_count

Variable dependiente	Objetivo	Variables independientes	Precisión (Una propiedad)	Precisión (Mas de una)	Exactitud	Sensibilidad (Una propiedad)	Sensibilidad (f)	Matriz de confusión
host_listings_count	Las variables objetivo son Una propiedad y Más de una	'maximum_nights', 'accommodates', 'price'	0.952	0.0	0.952	1.0	0.0	[[0 142] [0 2826]]
Observaciones	El modelo predice perfectamente a los anfitriones que tienen una propiedad, pero no identifica a los que tienen más de una. Por eso es que la mayoría de los anfitriones en Airbnb solo manejan una propiedad, probablemente porque son dueños particulares, no empresas de renta masiva.							

6. bedrooms

[illegible]

7. beds

[illegible]

8. number_of_reviews_130d

Variable dependiente	Objetivo	Variables independientes	Precisión (No tiene)	Precisión (Últimos 30 días)	Exactitud	Sensibilidad (No tiene)	Sensibilidad (Últimos 30 días)	Matriz de confusión
number_of_reviews_130d	Las variables objetivo son No tiene y Últimos 30 días	'availability_365','host_acceptance_rate','price'	0.815	0.276	0.806	0.985	0.230	[[0 383] [0 2585]]
Observaciones	Altísima sensibilidad en 'no tiene': el modelo casi nunca falla al detectar alojamientos sin reseñas recientes. Pero hay una baja sensibilidad en 'últimos 30 días': ignora completamente alojamientos activos. Esto se debe a que la mayoría tiene reseñas fuera de los 30 días y por eso le cuesta reconocer a los que apenas tienen reseñas.							

9. availability_30

Variable dependiente	Objetivo	Variables independientes	Precisión (Hasta 14 días)	Precisión (De 14 a 30 días)	Exactitud	Sensibilidad (Hasta 14 días)	Sensibilidad (De 14 a 30 días)	Matriz de confusión
Availability_30	Las variables objetivo son Hasta 14 días y De 14 a 30 días	'number_of_reviews','host_acceptance_rate','price'	0.683	0.523	0.680	0.985	0.034	[[33 918] [30 1987]]

