

Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias de la Computación

Ingeniería en Tecnologías de la Información

Primavera 2025



Inteligencia de Negocios

Regresión Logística Reporte

Docente:

Alfredo García Juárez

Alumno:

Antonio Morales Badillo

Introducción

La regresión logística es un modelo de aprendizaje automático supervisado que se usa para predecir resultados de tipo categórico, es decir, cuando la variable que quieres predecir tiene dos o más posibles clases como: Si/No, 1/0, Es spam/No es spam. Su objetivo principal es predecir probabilidades de pertenecer a una categoría específica. Transformando los valores numéricos en probabilidades entre 0 y 1, utiliza la función sigmoide.

Casos de regresión logística

Los puntos negros representaran las variables dependientes, mientras que los puntos blancos sus variables independientes.

- **host_is_superhost:** Indica si el anfitrión es un superanfitrión.
 - **host_acceptance_rate:** Tasa de aceptación de reservas.
 - **host_listings_count:** Número de propiedades que tiene el anfitrión.
 - **review_scores_rating:** Puntaje promedio de reseñas.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.54787234	0.63386176	0.48938866
Los ángeles	t (true) / f (false)	0.41185834	0.6054034	0.21174304
Ottawa	t (true) / f (false)	0.57681159	0.60282021	0.50895141

Los resultados obtenidos de los modelos en los 3 países parecen ser apenas aceptable, siendo México y Ottawa los que obtuvieron mejores resultados. Mientras que Los ángeles fue quien obtuvo el peor resultado, indicando que el modelo tubo problemas para identificar los casos positivos. Por lo que se interpreta que la variables independientes no tienen un impacto muy significativo.

- **host_identity_verified:** Indica si la identidad del anfitrión está verificada.
 - **host_has_profile_pic:** Si el anfitrión tiene foto de perfil.
 - **host_listings_count:** Número de propiedades que tiene el anfitrión.
 - **host_verifications:** Verificaciones realizadas por el anfitrión.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.95396322	0.95396322	1
Los ángeles	t (true) / f (false)	0.88541821	0.88541821	1
Ottawa	t (true) / f (false)	0.91421857	0.91421857	1

Los resultados obtenidos de los 3 modelos en los 3 países son muy buenos. Mostrando que las variables independientes tienen un gran impacto en la predicción. Con esto se da entender de que los usuarios que están verificados tiene X cantidad de propiedades, tiene foto de perfil y como vimos en la regresión lineal, mayormente 2 verificaciones.

- **has_availability:** Indica si la propiedad tiene disponibilidad.
 - **availability_30:** Disponibilidad de la propiedad en los próximos 30 días.
 - **availability_90:** Disponibilidad de la propiedad en los próximos 90 días.
 - **availability_365:** Disponibilidad de la propiedad durante todo el año.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.97311351	0.97311351	1
Los ángeles	t (true) / f (false)	0.96905996	0.96905996	1
Ottawa	t (true) / f (false)	0.96827262	0.96827262	1

Los resultados obtenidos de los 3 modelos en los 3 países son muy buenos. Mostrando que las variables independientes tienen un gran impacto en la predicción. Las propiedades tienen disponibilidad en estos intervalos, es muy probable que la propiedad este disponible.

- **instant_bookable:** Indica si la propiedad es instantáneamente reservable.
 - **number_of_reviews:** Número total de reseñas.
 - **review_scores_accuracy:** Puntaje de precisión de las reseñas.
 - **review_scores_cleanliness:** Puntaje de limpieza de las reseñas.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.45033113	0.60279011	0.08854167
Los ángeles	t (true) / f (false)	0.58333333	0.75011103	0.00207223
Ottawa	t (true) / f (false)	0	0.7720329	0

Los resultados obtenidos de los 3 modelos en los 3 países no son muy buenos. En caso de Ottawa se debe que el dataset tiene muy pocos datos y el modelo no puede aprender correctamente. Mientras que para México y los ángeles, las variables independientes no tiene un gran impacto para determinar si la propiedad se puede reservar de forma instantánea.

- **Bathrooms:** Número de baños en la propiedad. Se categorizo si tiene o no tiene baños.
 - **bathrooms_text:** Texto que describe la cantidad de baños.
 - **property_type:** Tipo de propiedad.
 - **review_scores_cleanliness:** Puntaje de limpieza de las reseñas.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.99556119	0.99556119	1
Los ángeles	t (true) / f (false)	0.99629904	0.99629904	1
Ottawa	t (true) / f (false)	0.99529965	0.99529965	1

Los resultados obtenidos de los 3 modelos en los 3 países son casi perfectos. Esto tiene mucho sentido ya que según el tipo de propiedad deben tener casi una moda en cuanto a cantidad de baños, además que bathrooms_texto es casi exactamente lo mismo que la cantidad de baños. Por último al haber mas baños los anfitriones deben prestar más atención al aseo del hogar.

- **Bedrooms:** Número de dormitorios en la propiedad. Se categorizo si tiene más de 3 camas.
 - **beds:** Número total de camas.
 - **review_scores_value:** Puntaje de valor de las reseñas.

- **availability_60**: Disponibilidad en los próximos 60 días.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.76576577	0.97564997	0.33864542
Los ángeles	t (true) / f (false)	0.66926677	0.90991858	0.29916318
Ottawa	t (true) / f (false)	0.83333333	0.94594595	0.51282051

Los resultados obtenidos de los 3 modelos en los 3 países son buenos. El único problema es la sensibilidad de los modelos, lo que sugiere que el modelo necesita refinamiento para mejorar su capacidad de identificar propiedades con más de 3 camas.

- **host_response_time**: Tiempo de respuesta del anfitrión. Se categorizo el anfitrión responde en menos de una hora.
 - **host_response_rate**: Tasa de respuesta del anfitrión.
 - **number_of_reviews_l30d**: Número de reseñas en los últimos 30 días.
 - **review_scores_communication**: Puntaje de comunicación en las reseñas.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.87282501	0.87596703	0.98735798
Los ángeles	t (true) / f (false)	0.81818913	0.819245	0.98231713
Ottawa	t (true) / f (false)	0.8346056	0.82843713	0.97619048

Los resultados obtenidos de los 3 modelos en los 3 países son bastante buenos. Es muy probable a que seba a que se están utilizando 2 variables independientes relacionadas con la evaluación de la comunicación del anfitrión. Muy probablemente lo que baje el rendimiento del modelo sea la cantidad de reseñas, pero no sería mucho.

- **room_type**: Tipo de habitación ofrecida. Se categorizo si la habitación no es privada.
 - **beds**: Número total de camas.
 - **review_scores_location**: Puntaje de ubicación de las reseñas.
 - **review_scores_rating**: Puntuación promedio de todas las reseñas

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.63380282	0.74148061	0.80838323
Los ángeles	t (true) / f (false)	0.73751128	0.73042191	0.98463238
Ottawa	t (true) / f (false)	0.85990338	0.75323149	0.7007874

Los resultados obtenidos de los 3 modelos en los 3 países son buenos. Siendo el peor el de México, seguido de los ángeles, muy probablemente los modelos necesitan refinamientos para mejorar la identificación de habitaciones no privadas.

- **Accommodates**: Número de personas que la propiedad puede alojar. Se categorizo si la propiedad tiene capacidad X personas diferente 4.
 - **minimum_nights**: Número mínimo de noches requeridas.
 - **maximum_nights**: Número máximo de noches permitidas.
 - **calculated_host_listings_count**: Número de propiedades que tiene el anfitrión.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.78135701	0.78135701	1

Los ángeles	t (true) / f (false)	0.82028127	0.82028127	1
Ottawa	t (true) / f (false)	0.80376028	0.80376028	1

Los resultados obtenidos de los 3 modelos en los 3 países son buenos. Tiene sentido que los usuarios que buscan opciones de alojamiento para grupos.

- **Price:** Precio de la propiedad. Se categoriza si es precio es no mayor a la media.
 - **review_scores_value:** Puntaje de valor de las reseñas.
 - **review_scores_location:** Puntaje de ubicación de las reseñas.
 - **review_scores_cleanliness:** Puntaje de limpieza de las reseñas.

País	Categorización	Precisión	Exactitud	Sensibilidad
México	t (true) / f (false)	0.75358275	0.75358275	1
Los ángeles	t (true) / f (false)	0.6827262	0.6827262	1
Ottawa	t (true) / f (false)	0.78356773	0.78356773	1

Los resultados obtenidos de los 3 modelos en los 3 países son buenos. Las variables independientes al ser puntajes de reseña sirven como indicadores de calidad. Además que los anfitriones pueden ajustar sus precios ante la competencia. Seguramente los valores máximos y mínimos afectan drásticamente al modelo.

Conclusiones

Los resultados indican que la mayoría de modelo es eficaz en la identificación de la relación entre las variables dependientes e independientes, con sensibilidad perfecta en algunos casos, mientras que en otros presenta sensibilidad baja, sugiriendo áreas de mejora. Pero la baja precisión en ciertos casos resalta la necesidad de hacer ajustes en los modelos, ya sea incorporando más variables o utilizar técnicas de modelado avanzadas. En general, la alta sensibilidad en la identificación de propiedades adecuadas para grupos y precios accesibles refleja la demanda del mercado, lo que sugiere que los anfitriones podrían estar ajustando sus ofertas en respuesta a las necesidades de los usuarios.