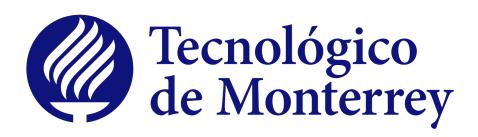
Instituto Tecnológico y de Estudios Superiores de Monterrey



Analítica de Datos y Herramientas de Inteligencia Artificial II

Actividad 4.1: Reporte de regresión logística

Alumno:

Diego Sánchez Márquez

A01734778

Grupo 501

Modelos con mejores resultados

Los modelos con mayor precisión, exactitud y sensibilidad fueron los que tienen como variable objetivo las variables de: has_availability and host_has_profile_pic. El modelo donde se utilizaron las variables independientes de beds, bedrooms y accommodates tiene coeficientes de precisión y exactitud de 0.99, y sensibilidad de 1.00, lo que indica una gran capacidad de predicción. Otro modelo que arrojó muy buenos resultados, fue realizado con la variable objetivo has_availability con variables independientes como availability_365, host_total_listings_count y estimated_occupancy_1365d, logrando los mismos niveles de precisión, exactitud y sensibilidad.

También, el modelo con variable objetivo host_has_profile_pic, explicado por host_response_time, host identity verified y host_acceptance_rate, obtuvo resultados de 0.98 en precisión y exactitud, y de 1.00 en sensibilidad. Este modelo también se considera confiable.

Modelos con resultados moderados

El modelo que intenta predecir host_is_superhost utilizando beds, bedrooms y accommodates tiene una precisión de 0.66, exactitud de 0.65 y una sensibilidad de 0.99.

Los dos modelos que utilizan instant_bookable como variable objetivo también presentan un rendimiento moderado. El primero utiliza las variables independientes de price, property_type y host_identity_verified y el segundo price, host_identity_verified y availability_30. Ambos modelos muestran coeficientes similares, con precisión entre 0.68 y 0.69 y sensibilidad alrededor del 0.94.

Modelos con resultados bajos

Los modelos que intentan predecir estimated_occupancy_1365d y number_of_reviews presentan resultados muy bajos en todos los coeficientes. En el caso de estimated_occupancy_1365d, la precisión y la sensibilidad son de 0.02 y 0.04. Para number of reviews, los valores son todavía más pequeños: 0.01 en precisión y sensibilidad.

Conversión de variables

Con base en el desempeño de los modelos, se convirtieron las siguientes variables a dicotómicas:

- host_response_time: Se dividieron los intervalos obtenidos en fast response time y slow response time
- estimated_occupancy_1365d: Se dividieron los intervalos en high occupancy y low occupancy.
- number_of_reviews: Se dividieron los intervalos en high number of reviews y low number of reviews

Conclusión

Los modelos con mejores resultados fueron aquellos que predicen has_availability y host_has_profile_pic, debido a su alta precisión, exactitud y sensibilidad. Por el contrario, los modelos donde se intentó explicar las variables de estimated_occupancy_l365d y number of reviews presentaron un bajo desempeño.

A partir de estos resultados, se convirtieron a formato dicotómico las variables host_response_time, estimated_occupancy_l365d y number_of_reviews, dividiéndolas en categorías como fast/slow response time, high/low occupancy y high/low number of reviews, con el objetivo de facilitar su análisis y mejorar el desempeño de los modelos de regresión logística.