



Tecnológico de Monterrey

Actividad 3 (Regresión Logística)

Enrique Rosales Mijangos

15/10/2025

Gestión de proyectos de plataformas tecnológicas

Introducción

Para realizar predicciones de variables dicotómicas, se necesita de la regresión logística, la cuál usa herramientas como la matriz de confusión. Para la práctica, se analizarán 10 variables dicotómicas, para las cuales se analizarán métricas como la precisión, exactitud, sensibilidad y el puntaje f1.

Transformación de variables

Para el ejercicio, se consideraron las siguientes variables y se especifica la transformación que requirió:

'host_is_superhost': No requirió de transformación

'host_has_profile_pic': No requirió de transformación

'host_identity_verified': No requirió de transformación

'instant_bookable': No requirió de transformación

'host_response_time': Se agruparon las respuestas de antes de un día y después de 1 día.

Observaciones de “Sin respuesta” se transformaron a “a few days or more”

'host_response_rate': Se dividieron las calificaciones entre más de 90 y menos de 90.

'host_acceptance_rate': Se dividieron las calificaciones entre más de 80 y menos de 80.

'property_type': Se agruparon en otros aquellas observaciones con menos repeticiones.

'room_type': Se agruparon en otros aquellas observaciones con menos repeticiones.

'license': Se dividió en host sin licencia y con licencia.

Regresiones logísticas

Host is superhost

```
Matriz de Confusión:  
[[3990  909]  
 [2916 1548]]  
Precisión del modelo:  
0.577758470894874  
0.63003663003663  
Exactitud del modelo:  
0.5914770906760654  
Sensibilidad del modelo:  
0.8144519289650949  
0.3467741935483871  
Puntaje F1 del modelo:  
0.44733420026007803
```

Se presenta una precisión muy similar tanto para T como F, por lo cual el modelo puede predecir casi de igual manera ambas clases. Se presenta una buena exactitud (0.59). En cuanto a la sensibilidad, se muestra una mayor brecha, por lo que la proporción de positivos detectados no es tan buena para una clase. Al analizar el balance (F1), el puntaje de 0.44 no es tan bueno. El modelo predice menos del 50%.

Host has profile pic

```
Matriz de Confusión:  
[[3932  902]  
 [2901 1628]]  
Precisión del modelo:  
0.5754427045221718  
0.6434782608695652  
Exactitud del modelo:  
0.5938267649257717  
Sensibilidad del modelo:  
0.8134050475796442  
0.3594612497240009  
Puntaje F1 del modelo:  
0.46125513528828443
```

En este caso la precisión del modelo tiene similitud entre ambas clases, así mismo la exactitud fue de 0.59, siendo aceptable. Pero al considerar la sensibilidad, las dos clases difieren más, por esta razón el puntaje F1 se ve afectado y termina en 0.46. Por lo que igualmente no se considera un modelo bueno, ya que deja más de la mitad de las observaciones predichas de manera random.

Host identity verified

```
Matriz de Confusión:  
[[ 0 1276]  
 [ 0 8087]]  
Precisión del modelo:  
0.0  
0.8637188935170351  
Exactitud del modelo:  
0.8637188935170351  
Sensibilidad del modelo:  
0.0  
1.0  
Puntaje F1 del modelo:  
0.9268767908309455
```

El modelo inicial claramente muestra un problema de balanceo, podría predecir sólo de una clase, más no de la otra. Se requiere un balanceo.

Oversampling

```
Matriz de Confusión:  
[[ 394  882]  
 [2271 5816]]  
Precisión del modelo:  
0.14784240150093808  
0.8683189011645267  
Exactitud del modelo:  
0.6632489586670939  
Sensibilidad del modelo:  
0.30877742946708464  
0.7191789291455423  
Puntaje F1 del modelo:  
0.7867433209333784
```

Al realizar el balanceo los parámetros mejoran en gran medida, pero igualmente la precisión es muy distinta para las dos clases, igualmente con la sensibilidad. Por otro lado, la exactitud puede considerarse aceptable. Al final el puntaje F1 termina en 0.78, lo cuál sugiere que el modelo es bueno.

Instant bookable

```
Matriz de Confusión:  
[[ 0 4361]  
 [ 0 5002]]  
Precisión del modelo:  
0.0  
0.5342304816832212  
Exactitud del modelo:  
0.5342304816832212  
Sensibilidad del modelo:  
0.0  
1.0  
Puntaje F1 del modelo:  
0.6964148973198747
```

Mal modelo inicial, no predice igualmente para las dos clases.

Reponderación de clases

```
Matriz de Confusión:  
[[3198 1163]  
 [3546 1456]]  
Precisión del modelo:  
0.47419928825622776  
0.5559373806796487  
Exactitud del modelo:  
0.4970629071878671  
Sensibilidad del modelo:  
0.7333180463196515  
0.29108356657337064  
Puntaje F1 del modelo:  
0.3821020863403753
```

La precisión es similar para las dos clases, y aunque todos los parámetros mejoraron, aún el modelo podría mejorar. Esto se podría hacer eligiendo otra variable independiente explicativa.

host response time

```
Matriz de Confusión:  
[[4022  903]  
 [2840 1598]]  
Precisión del modelo:  
0.5861264937336054  
0.6389444222311076  
Exactitud del modelo:  
0.6002349674249706  
Sensibilidad del modelo:  
0.8166497461928934  
0.3600721045515998  
Puntaje F1 del modelo:  
0.46058509871739445
```

Tiene buena exactitud, buena precisión similar para las dos clases, pero la sensibilidad sigue un poco desigual. El puntaje final fue de 0.46, puede considerarse un modelo que predice de manera random.

host response rate

```
Matriz de Confusión:  
[[8255   0]  
 [1108   0]]  
Precisión del modelo:  
0.8816618605147922  
0.0  
Exactitud del modelo:  
0.8816618605147922  
Sensibilidad del modelo:  
1.0  
0.0  
Puntaje F1 del modelo:  
0.9371097740946759
```

Requiere de balanceo.

Reponderación de clases

```
Matriz de Confusión:  
[[3683 4572]  
 [ 245  863]]  
Precisión del modelo:  
0.9376272912423625  
0.15878564857405703  
Exactitud del modelo:  
0.485528142689309  
Sensibilidad del modelo:  
0.4461538461538462  
0.7788808664259927  
Puntaje F1 del modelo:  
0.6046129853073956
```

En este modelo la sensibilidad es mucho mejor, aunque con ligera inclinación de mejor predicción hacia una clase, aún se logra un buen puntaje F1 (0.6), este es un modelo aceptable para predecir la variable host response rate.

host acceptance rate

```
Matriz de Confusión:  
[[7095    0]  
 [2268    0]]  
Precisión del modelo:  
0.7577699455302788  
0.0  
Exactitud del modelo:  
0.7577699455302788  
Sensibilidad del modelo:  
1.0  
0.0  
Puntaje F1 del modelo:  
0.8621946773605541
```

Requiere de balanceo.

Reponderación de clases

```
Matriz de Confusión:  
[[3179 3916]  
 [ 301 1967]]  
Precisión del modelo:  
0.9135057471264367  
0.33435322114567395  
Exactitud del modelo:  
0.5496101676812988  
Sensibilidad del modelo:  
0.448062015503876  
0.8672839506172839
```

Tanto la precisión como exactitud y sensibilidad muestran buena capacidad de predicción, con valores similares, el modelo logra predecir de manera no random.

Property type

```
Matriz de Confusión:  
[[ 542 3979]  
 [1027 3815]]  
Precisión del modelo:  
0.3454429572976418  
0.489479086476777  
Exactitud del modelo:  
0.4653423048168322  
Sensibilidad del modelo:  
0.11988498119884981  
0.7878975629904998  
Puntaje F1 del modelo:  
0.1779967159277504
```

El modelo muestra mucha brecha en la sensibilidad, lo cual afecta mucho en el puntaje F1, el modelo predice mejor para una clase que para otra.

Room type

```
Matriz de Confusión:  
[[8226  33]  
 [1102   2]]  
Precisión del modelo:  
0.8818610634648371  
0.05714285714285714  
Exactitud del modelo:  
0.8787781693901527  
Sensibilidad del modelo:  
0.9960043588812205  
0.0018115942028985507  
Puntaje F1 del modelo:  
0.003511852502194908
```

Requiere balanceo.

Sobremuestreo

```
Matriz de Confusión:  
[[6372 1887]  
 [ 595  509]]  
Precisión del modelo:  
0.9145973876847998  
0.2124373956594324  
Exactitud del modelo:  
0.7349140232831357  
Sensibilidad del modelo:  
0.7715219760261532  
0.46105072463768115
```

Aunque la precisión para una clase es poca y no cercana a la otra clase, la sensibilidad es aceptable y se tiene buena exactitud, podría haber un ligero desbalance, pero con ajustes el modelo podría ser muy bueno.

License

```
Matriz de Confusión:  
[[7699    0]  
 [1664    0]]  
Precisión del modelo:  
0.8222791840222151  
0.0  
Exactitud del modelo:  
0.8222791840222151  
Sensibilidad del modelo:  
1.0  
0.0  
Puntaje F1 del modelo:  
0.9024733325518697
```

Requiere balanceo

Reponderación

```
Matriz de Confusión:  
[[2554 5145]  
 [ 354 1310]]  
Precisión del modelo:  
0.8782668500687758  
0.2029434546862897  
Exactitud del modelo:  
0.41268824094841394  
Sensibilidad del modelo:  
0.3317313936874919  
0.7872596153846154  
Puntaje F1 del modelo:  
0.48156877533704157
```

En este caso se muestra igualmente un ligero desbalance, y aunque se pueda predecir en mayor medida una clase, el modelo es mucho mejor que el inicial, se logra corregir en gran medida el desbalance y se convierte en un modelo aplicable.