



Tecnológico de Monterrey

Campus Puebla

José Antonio Villanueva Arenas A01737550

Analítica de datos y herramientas de inteligencia artificial II

Actividad 3.1 regresión no lineal

11 Abril 2025

Tabla de coeficientes y correlaciones

Tabla de correlaciones

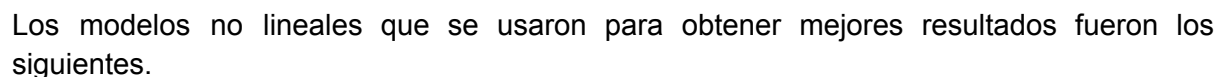
Variable dependiente	Variable independiente	Modelo	Correlación lineal	Correlación con no lineal	mejora
calculated_host_listings_count	host_listings_count	func_polinomial_inversa	0.69	0.79	+0.10
calculated_host_listings_count	host_listings_count	func_cuadratica	0.69	0.81	+0.12
review_scores_accuracy	review_scores_rating	func_cuadratica	0.63	0.67	+0.04
review_scores_accuracy	review_scores_rating	func_cociente_polinomio	0.63	0.63	+0.0
host_total_listings_count	host_listings_count	función cuadrática	0.65	0.65	+0.0
host_total_listings_count	host_listings_count	func_exponencia	0.65	0.66	+0.1
accomodates	beds	función cuadrática	0.44	0.44	+0.0
accomodates	beds	func_valor_absoluto	0.44	0.44	0.0
reviews_per_month	id	función cuadrática	0.33	0.36	+0.03
reviews_per_month	id	func_logaritmica	0.33	0.34	+0.01
price	availability_365	función cuadrática	0.31	0.36	+0.05
price	availability_365	func_valor_absoluto	0.31	0.31	+0.0

Tabla de determinación

Variable dependiente	Variable independiente	Modelo	Determinación con no lineal
calculated_host_listings_count	host_listings_count	func_polinomial_inversa	0.63
calculated_host_listings_count	host_listings_count	func_cuadratica	0.65
review_scores_accuracy	review_scores_rating	func_cuadratica	0.45
review_scores_accuracy	review_scores_rating	func_cociente_polinomio	0.39
host_total_listings_count	host_listings_count	función cuadrática	0.43
host_total_listings_count	host_listings_count	func_exponencia	0.43
accommodates	beds	función cuadrática	0.19
accommodates	beds	func_valor_absoluto	0.19
reviews_per_month	id	función cuadrática	0.13
reviews_per_month	id	func_logaritmica	0.11
price	availability_365	función cuadrática	0.13
price	availability_365	func_valor_absoluto	0.09

Gracias a las tablas se puede ver que mientras mayor es la determinación, suele ser más alta la correlación aunque en algunos modelos encontramos determinaciones muy bajas, lo que explica el porcentaje de los datos que se encuentran en la línea de regresión.

En primer lugar se generó un Heat Map para obtener las correlaciones de un modelo simple y se determinó el objetivo de superarlas con dos modelos de regresión no lineal.



- $y = ax^2 + bx + c$ ("Función cuadrática")
- $y = a \cdot \exp(-bx) + c$ ("Función exponencial")
- $y = 1/a \cdot x$ ("Función inversa")
- $y = a \cdot \sin(x) + b$ ("Función senoidal")
- $y = a \cdot \tan(x) + b$ ("Función tangencial")
- $y = a \cdot \text{abs}(x) + b \cdot x + c$ (Función Valor absoluto)
- $y = (a \cdot x^2 + b) / c \cdot x$ (Función cociente entre polinomios)
- $y = a \cdot \log(x) + b$ (Función logarítmica)

- $y = a*x + b*x + c*x$ (Función lineal con producto de coeficientes)
- $y = 1/a*x^{**2}$ (Función cuadrática inversa)
- $y = a/b*x^{**2} + c*x$ (Función polinomial inversa)

Después de aplicar los modelos para cada variable se obtuvieron estos resultados:

Variable dependiente	Variable independiente	Modelo	Correlación lineal	Correlación con no lineal	mejora
calculated_host_listings_count	host_listings_count	func_polinomial_inversa	0.69	0.79	+0.10
calculated_host_listings_count	host_listings_count	func_cuadratica	0.69	0.81	+0.12
review_scores_accuracy	review_scores_rating	func_cuadratica	0.63	0.67	+0.04
review_scores_accuracy	review_scores_rating	func_cociente_polinomio	0.63	0.63	+0.0
host_total_listings_count	host_listings_count	función cuadrática	0.65	0.65	+0.0
host_total_listings_count	host_listings_count	func_exponencia	0.65	0.66	+0.1
accommodates	beds	función cuadrática	0.44	0.44	+0.0
accommodates	beds	func_valor_absoluto	0.44	0.44	0.0
reviews_per_month	id	función cuadrática	0.33	0.36	+0.03
reviews_per_month	id	func_logaritmica	0.33	0.34	+0.01
price	availability_365	función cuadrática	0.31	0.36	+0.05

price	availability_365	func_valor_absoluto	0.31	0.31	+0.0
-------	------------------	---------------------	------	------	------

En general, se pudo ver que:

Con una mayor correlación, suele haber una mayor determinación, lo que sugiere que el modelo explica mejor la variabilidad de la variable dependiente. Sin embargo, esto no sucede en todos los modelos con alta correlación, algunos muestran una variabilidad baja. Por otro lado, en la mayoría de los casos, el modelo no lineal mejora respecto al lineal, aunque en algunos las mejoras son mínimas. Por último, las variables price y reviews_per_month presentan baja determinación, por lo que se puede inferir que no son muy certeros estos modelos.