



PROJECT BREAK II
MACHINE LEARNING
AMAZON PRICE TOOL

NORBERTO HUGO FERNÁNDEZ LLAUGER

OBJETIVO DEL PROYECTO

El proyecto pretende predecir precios de productos de Amazon, con el objetivo de ayudarme a elegir cuales puede ser una buena opción vender y cuales no.

Dado el problema que es, el enfoque elegido fue el de regresión supervisada.

COMPRENSIÓN Y LIMPIEZA DE DATOS

-El dataset inicial tiene una dimensionalidad de (25874, 30)

Problemas encontrados:

- Bastantes columnas con información poco útil
- Outliers muy dispares
- Tratamiento de categóricas

Soluciones aplicadas:

- Eliminación de columnas innecesarias
- Creación de columnas para mejor categorización
- Limitar outliers
- Tratamiento de target con Boxcox



DIVISIÓN DEL DATASET Y FEATURE ENGINEERING

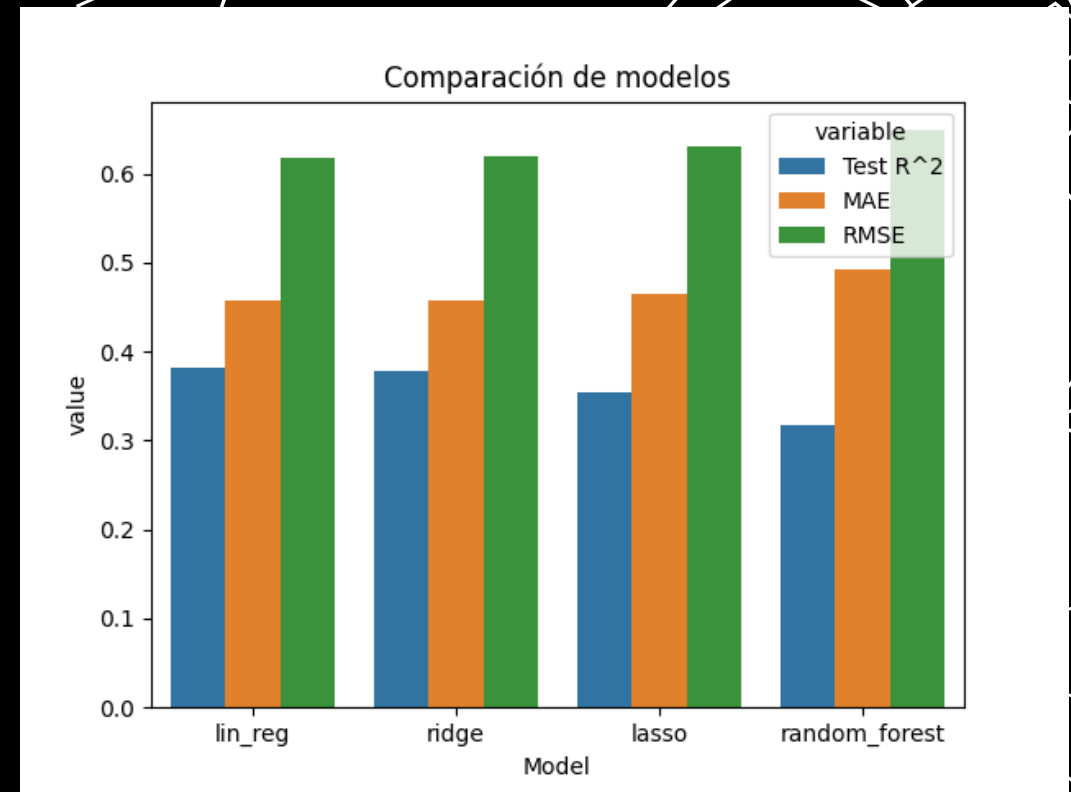
División de Train/Test al principio del análisis

Modificaciones a columnas categóricas para
mejor funcionamiento del modelo

MODELADO

Modelos probados:

- Linear Regression
- Ridge Regression
- Lasso Regression
- Random Forest Regressor
- Gradient Boosting Regressor



RESULTADOS Y CONCLUSIONES

Habría que mejorar la calidad e interpretación de los datos, parece que para precios muy bajos no generaliza mal.

| | Model | Train R ² | Test R ² | MAE | MSE | RMSE |
|---|---------------|----------------------|---------------------|----------|----------|----------|
| 0 | lin_reg | 0.426933 | 0.381650 | 0.449299 | 0.371436 | 0.609455 |
| 1 | ridge | 0.422712 | 0.376461 | 0.451799 | 0.374552 | 0.612007 |
| 2 | lasso | 0.392731 | 0.349950 | 0.461129 | 0.390477 | 0.624882 |
| 3 | random_forest | 0.899754 | 0.297240 | 0.492625 | 0.422140 | 0.649723 |