



PREDICCIÓN DE PRECIOS DE VEHÍCULOS

MACHINE LEARNING MODELS

MAYCO CASTELLANOS

INDICE

- ✓ LIMPIEZA Y PREPARACIÓN DE DATOS.
- ✓ DESARROLLO DE MODELOS
- ✓ COMPARAR RANDOM FOREST VS. XGBOOST.
- ✓ OPTIMIZAR MODELOS.
- ✓ RESULTADOS Y HALLAZGOS



LIMPIEZA



NA values:

Se imputaron valores nulos a través de la media agrupada por marca.



Valores Duplicados:

Se eliminaron los duplicados.



Outliers:

Se realizó eliminación de un registro que contenía outlier en la variable año (2060)

× × × ×
× × × ×
× × × ×
× × × ×

PROCESAMIENTO

Codificación: Se Aplicó One Hot Encoding a variables categóricas

Escalado: Se utilizó Min-Max Scaling para normalizar las variables numéricas.

División: El dataset se dividió en un 70% para entrenamiento y un 30% para prueba.

No se toma en cuenta la variable categorica model por su alta cardinalidad para los modelos

MAKE

FUELTYPE

TRANSMISION

YEAR

MILEAGE

ENGINE
SIZE

TAX

MPG

~~MODEL~~

× × × ×
× × × ×
× × × ×
× × × ×

MODELOS



Fase preliminar (baseline)

Resultados MAE



REGRESION LINEAL



2962.55



RANDOM FOREST



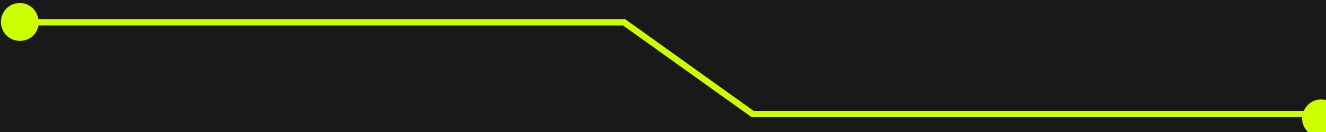
1440.67



XGBOOST



1496.08



MODELOS



Fase preliminar (baseline)

Resultados MAE



REGRESION LINEAL



~~2962.55~~



RANDOM FOREST



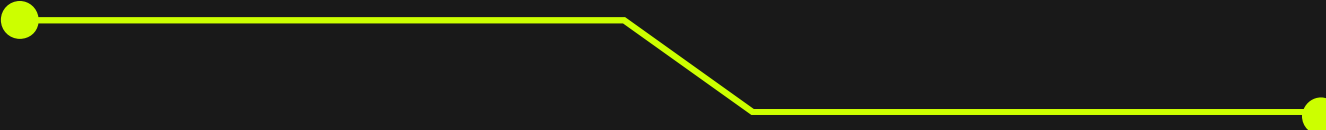
1440.67



XGBOOST



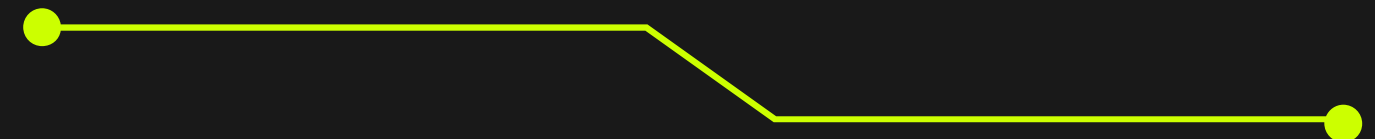
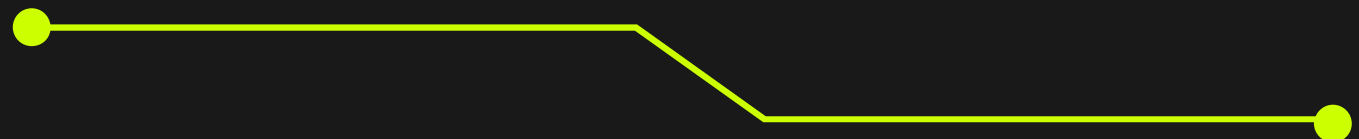
1496.08



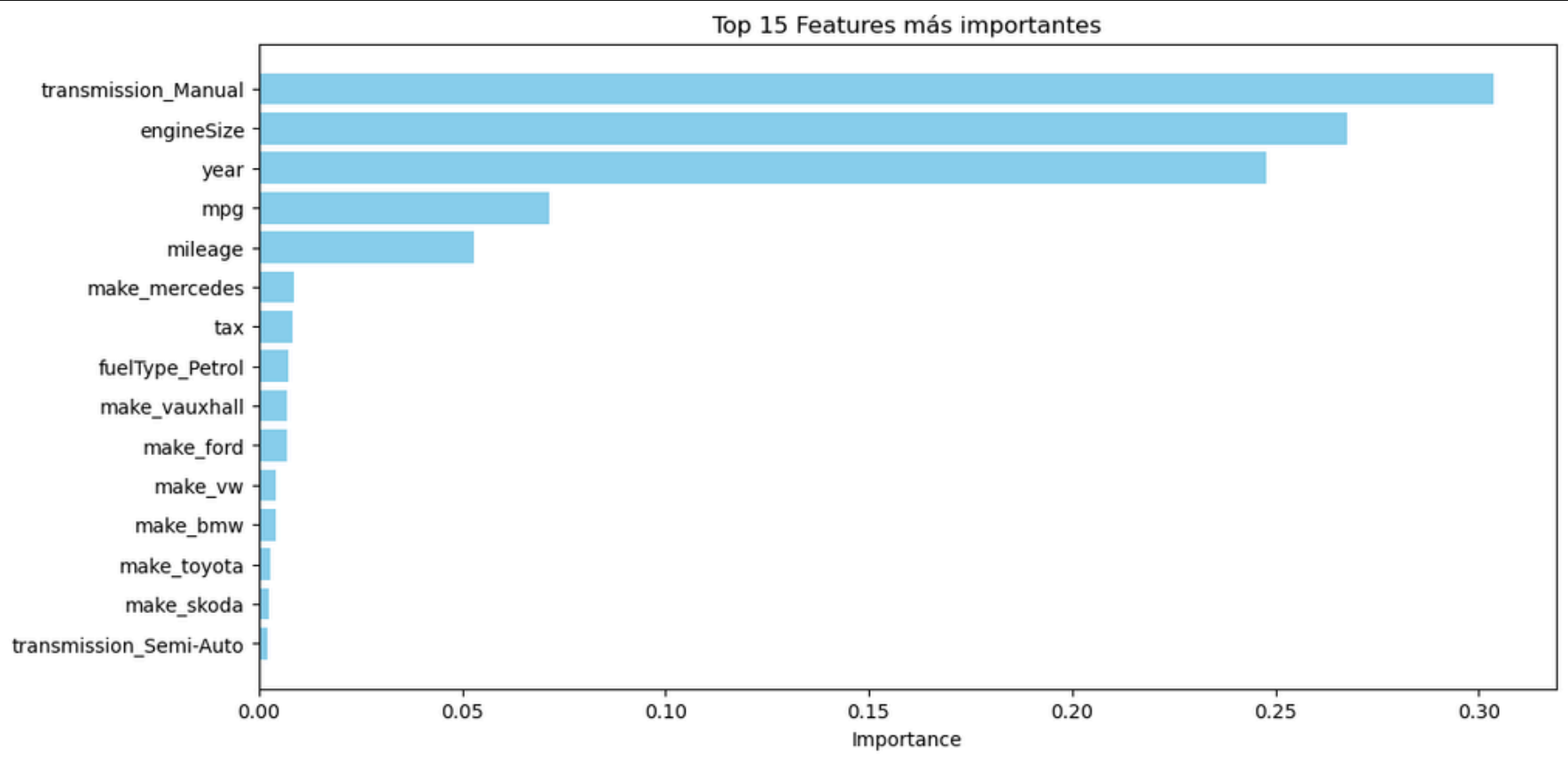
OPTIMIZACION



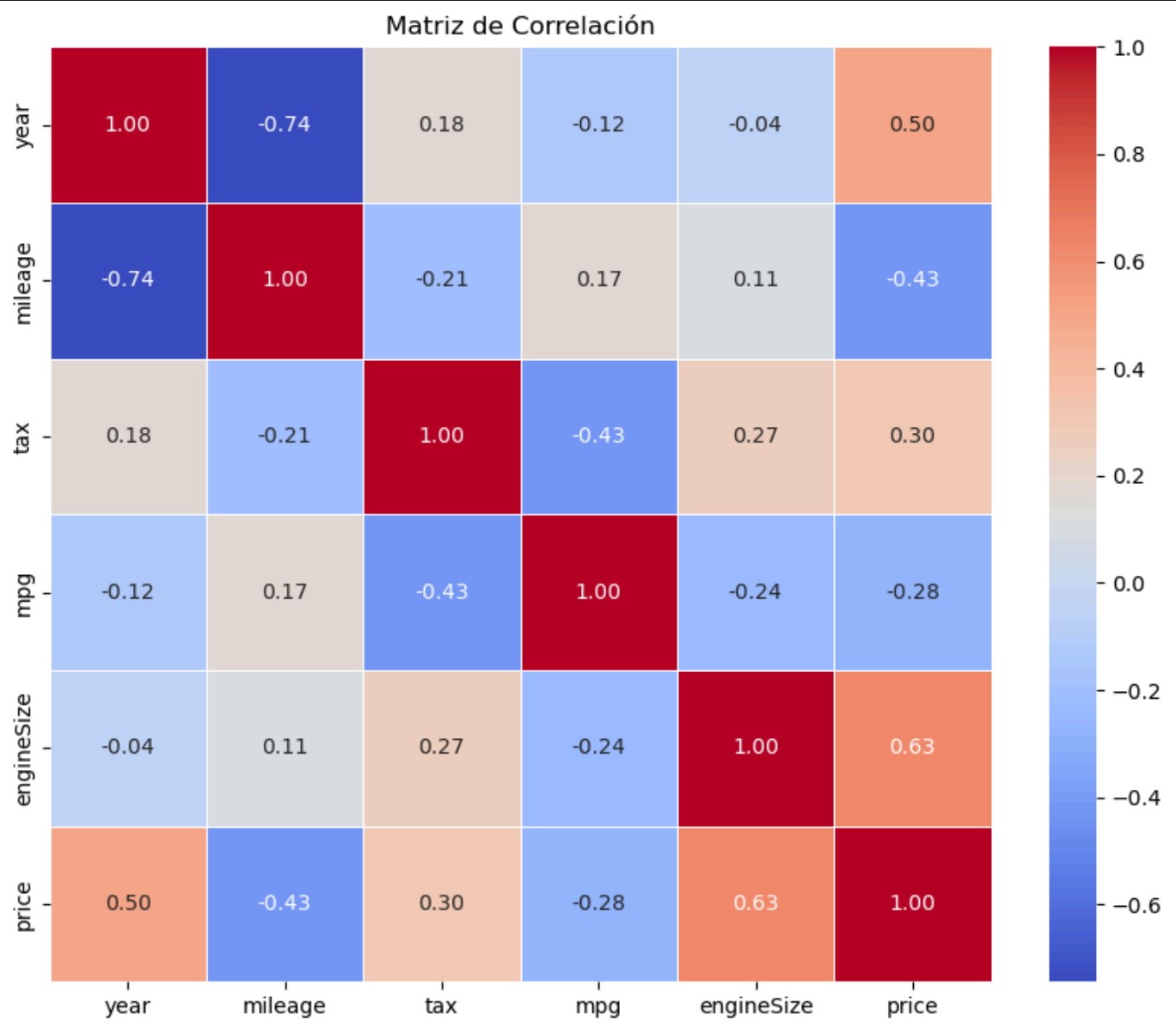
MODELADO CON MEJORES HIPERPARAMETROS
GRID SEARCH
SELECCIÓN DE MEJORES FEATURES



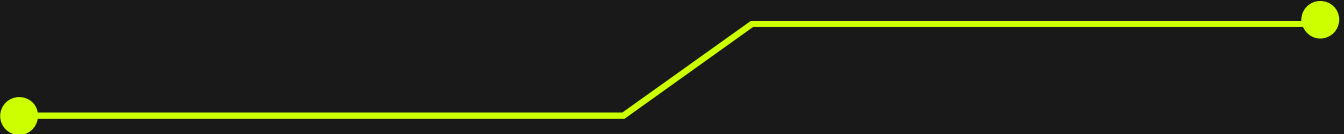
SELECCION DE MEJORES FEATURES



FEATURE_IMPORTANCE_



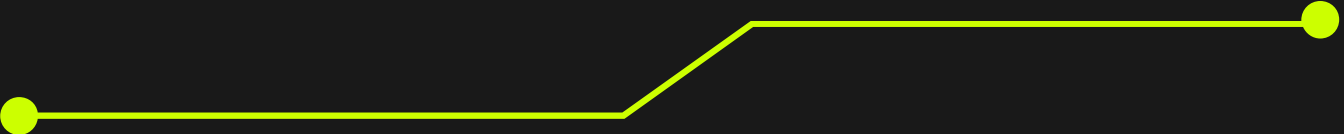
RESULTADOS



	Ajuste de Hiperparametros	Grid Search	Mejores Features
Random Forest	1410.67	1496.78	1553.23
XGBoost	1386.53	1358.74	1562.91



RESULTADOS



	Ajuste de Hiperparametros	Grid Search	Mejores Features
Random Forest	1410.67	1496.78	1553.23
XGBoost	1386.53	1358.74	1562.91



CONCLUSIONES

Eficiencia computacional: XGBoost demostró ser más eficiente en tiempo de entrenamiento, ya que utiliza técnicas como el boosting secuencial y la poda de árboles, lo que lo hace más ágil que Random Forest, que entrena múltiples árboles de forma paralela.

Complejidad del modelo: Random Forest es más sencillo de interpretar y menos sensible a los hiperparámetros, lo que puede hacerlo una mejor opción para aplicaciones rápidas o con menos necesidad de ajuste fino.

XGBoost, aunque más complejo, ofrece mayor control y optimización, lo que permite obtener un modelo más ajustado si se configura correctamente..

THANK YOU

