

PROYECTO FINAL CIENCIA DE DATOS II

Entrenamiento y optimización de modelos de Machine
Learning



DATASET PRECIOS DE AUTOMOVILES

"Este proyecto se centra en el análisis de un conjunto de datos de precios de automóviles, cuyo objetivo principal es predecir el precio de un vehículo en función de atributos clave como la marca, el modelo, el año, el tamaño del motor, el tipo de combustible, la transmisión, el kilometraje, el número de puertas, el número de propietarios anteriores y más. El precio de los automóviles es una variable de gran interés tanto para compradores como para vendedores, así como para aplicaciones de tasación automática. Este análisis puede contribuir a una mejor comprensión de los factores que influyen en los precios de los vehículos usados."

"Este análisis podría ser beneficioso para varias audiencias, incluyendo concesionarios de automóviles, plataformas de compra-venta de autos, empresas de seguros y financiamiento, y consumidores interesados en conocer el valor de mercado de vehículos usados. Además, la industria automotriz podría usar este análisis para realizar predicciones de precios y optimizar estrategias de ventas." Problema : *¿Cómo podemos predecir el precio de un automóvil basado en sus características (marca, modelo, año, tamaño del motor, tipo de combustible, transmisión, kilometraje, número de puertas, y cantidad de propietarios anteriores)

PROCESO

1-Preprocesamiento y Exploración de Datos: Comenzamos con la carga de los datos y la limpieza de los mismos, donde nos aseguramos de manejar valores faltantes y variables categóricas. Utilizamos One-Hot Encoding para convertir También real

2-Modelado inicial: Probamos varios modelos de regresión, como Regresión Lineal , Árboles de Decisión ,Bosque aleatorio , utilizaMSE (Error CuadráR² (coeficiente Inicialmente, todos los modelos ofrecieron buenos resultados, pero algunas limitaciones. En particular, el Árbol de Decisión y el Random Forest resultaron ser más adecuados

3-Optimización del Modelo: A continuación, aplicamos optimización de hiperparámetros en el modelo de *RanBosque aleatorio usando *RandomiRandomizedSearchCV para encontrar la mejor combinación de parámetrosn_estimators), lo profundomax_depth),min_samples_splity `min_muestras_hojamin_samples_leaf). Con la optimización, conseguimos mejorar significativamente el rendimiento del modelo, reduciendo el MSE y aumentando el R²..

4-Evaluación y Comparación de Modelos: Evaluamos el rendimiento de cada modelo utilizando el conjunto de prueba y validación cruzada . El modelo de **RaSe optimizó Random ForestMSE de 266.832y un R² de 0.97 , lo que indica que el modelo es capaz de Aunque la Regresión Lineal también se muestraBosque aleatorio .

5-Modelo Final y Selección: Después de una evaluación exhaustiva, el modelo final seleccionado fue el Random Forest optimizado , debido a su capacidad superior.

CONCLUSION FINAL

Hemos logrado desarrollar un modelo predictivo confiable que estima el precio de los vehículos con alta precisión. El modelo de Random Forest optimizado es el más publicitario