**Predicción de precios de viviendas - House Prices: Advanced Regression Techniques**

Este proyecto incluye un notebook donde se documentan todas las notas, conclusiones y pasos para llegar a ellas que solo requiere importarlo a Kaggle, cambiar la ruta de los archivos CSV de Kaggle y ejecutar las celdas con "Run All".

**Índice**

1. Introducción

2. Exploración y análisis de datos

3. Preparación de datos

4. Selección de características

5. Modelado

6. Evaluación de modelos

7. Conclusiones

8. Anexos

**Predicción de precios de viviendas - House Prices: Advanced Regression Techniques**

**1. Introducción**

El objetivo de este proyecto es predecir precios de viviendas utilizando datos del reto 'House Prices: Advanced Regression Techniques'

en Kaggle. Este análisis busca construir un modelo robusto que pueda predecir el precio de una vivienda basándose en sus características.

**2. Exploración y análisis de datos**

Se realizó un análisis inicial de los datos, destacando:

- Variables con valores nulos y su impacto.

- Distribuciones de variables clave como `SalePrice` y `LotArea`.

- Análisis de outliers para identificar valores atípicos que influencian el modelo.

**3. Preparación de datos**

Los pasos clave del preprocesamiento incluyeron:

- Imputación de valores nulos utilizando técnicas como el promedio o moda según el caso.

- Transformaciones logarítmicas en variables como `SalePrice` para estabilizar su distribución.

- Escalado de variables para normalizar los datos.

**Predicción de precios de viviendas - House Prices: Advanced Regression Techniques**

**4. Selección de características**

Se utilizaron técnicas de análisis de importancia de características y PCA para identificar las variables más relevantes: - `['GrLivArea', ' OverallQual', 'GarageArea', 'KitchenQual', 'TotalBaños', 'GarageCars'].

**5. Modelado**

Se entrenaron varios modelos, incluyendo:

- Regresión lineal, Ridge, Lasso.

- Algoritmos de árboles como Random Forest y Gradient Boosting.

Cada modelo fue evaluado utilizando métricas como RMSE y R².

**6. Evaluación de modelos**

El modelo Random Forest mostró el mejor desempeño con un RMSE más bajo y un R² más alto. Gráficos de predicción vs valores reales ayudaron a validar la calidad del modelo.

**7. Conclusiones**

- Las características más importantes fueron `OverallQual` y `GrLivArea`.

- La transformación logarítmica mejoró el rendimiento del modelo.

- Se recomienda explorar modelos avanzados o agregar datos externos para mejorar el rendimiento.

**8. Anexos**

**Predicción de precios de viviendas - House Prices: Advanced Regression Techniques**

Se incluyen:

- Código relevante.

- Tablas detalladas de métricas.

- Documentación de transformaciones.