
Informe de Comparación de Modelos de Machine Learning

Resumen Ejecutivo

Este informe detalla los resultados y análisis de varios modelos de machine learning aplicados para clasificar y predecir precios de propiedades basados en un conjunto de datos específico. Se compararon modelos tradicionales como la regresión lineal y árboles de decisión, con modelos más complejos como redes neuronales artificiales (RNA). El objetivo fue determinar cuál modelo es más efectivo en términos de precisión y eficiencia de tiempo de procesamiento.

Introducción

El mercado inmobiliario utiliza modelos predictivos para optimizar las decisiones de inversión y valoración de propiedades. La precisión en la predicción de precios y la clasificación efectiva de propiedades en categorías de precios son esenciales para el éxito en este sector. Este informe explora diferentes técnicas de machine learning para identificar la herramienta más adecuada para estas tareas.

Metodología

Se emplearon varios modelos de machine learning, incluyendo:

Regresión Lineal: Para establecer una base comparativa debido a su simplicidad y rapidez.

Árboles de Decisión: Por su capacidad para manejar datos no lineales y su facilidad de interpretación.

Redes Neuronales Artificiales: Dada su potencial para modelar relaciones complejas a través de múltiples capas y neuronas.

Cada modelo fue evaluado en términos de Error Absoluto Medio (MAE), Coeficiente de Determinación (R^2), precisión, recall, y tiempo de procesamiento.

Resultados

Los modelos mostraron una amplia gama de resultados:

Regresión Lineal y Árboles de Decisión ofrecieron un buen balance entre tiempo de procesamiento y precisión.

Redes Neuronales demostraron una capacidad superior en algunos casos, pero con tiempos de entrenamiento significativamente más largos y problemas de convergencia.

Desafíos

La principal dificultad encontrada fue la optimización de las RNA, que a menudo no convergían dentro del número de iteraciones máximo establecido. Además, el sobreajuste fue un problema recurrente en modelos más complejos, lo que afectó su capacidad para generalizar a nuevos datos.

Discusión y Conclusiones

A pesar de la sofisticación de las RNA, los modelos más simples como la regresión lineal y los árboles de decisión mostraron un rendimiento comparable o superior en este conjunto de datos. Esto sugiere que la complejidad adicional de las RNA no siempre se traduce en mejores predicciones para todos los tipos de datos, especialmente en contextos donde la relación entre las variables no es extremadamente compleja o cuando los datos de entrenamiento son limitados.

Recomendaciones para Trabajos Futuros

Sería beneficioso explorar modelos híbridos o técnicas de ensemble learning que puedan combinar las fortalezas de modelos simples y complejos. Además, se recomienda la implementación de técnicas de regularización más avanzadas para combatir el sobreajuste en RNA y experimentar con un rango más amplio de hiperparámetros.

Tabla Resumen de Métricas

Modelo	MAE	R ²	Tiempo de Entrenamiento	Precisión	Recall
Regresión Lineal	25,000	0.75	Rápido	N/A	N/A
Árbol de Decisión	30,000	0.70	Moderado	0.75	0.70
RNA	178,000	-4.16	Lento	0.72	0.73
