

Data Science

Laboratorio 8 - Puesta en producción

Desafíos durante el proceso de selección del modelo y puesta en producción

Uno de los principales desafíos fue la selección del modelo adecuado para predecir los precios de alquiler. Se probó con varios algoritmos, y al final se optó por una red neuronal debido a la complejidad y la no linealidad de los datos. Otro gran reto fue la necesidad de normalizar tanto los datos de entrada como los de salida para asegurar que la red convergiera correctamente. También otro desafío que encontramos fue la carga del modelo entrenado y la codificación de variables categóricas ya que no estábamos muy seguros de la forma de exportación y carga correcta de los encoders.

Aprendizajes más significativos

Durante este laboratorio, aprendimos sobre la importancia del preprocesamiento de datos, especialmente en la normalización y codificación de variables categóricas. También sobre la integración de modelos de aprendizaje automático en una UI usando *streamlit*. Consideramos que esta plataforma es útil para realizar una webapp de manera fácil y entendible para los usuarios. Como actuales trabajadores en backend y frontend consideramos que este tipo de aplicaciones hacen que sea mucho más fácil enfocarse más en el desarrollo del modelo que en la integración de estos a una interfaz, consideramos que estaremos utilizando esta herramienta en el futuro ya que fue de mucha utilidad.

Evaluación crítica de las fortalezas y limitaciones del enfoque utilizado

Entre las fortalezas, consideramos que la elección de una red neuronal fue adecuado, ya que permitió modelar relaciones complejas entre las diferentes características de las propiedades y el precio de alquiler. También usar *Streamlit* fue efectivo para crear una interfaz intuitiva para ingresar y obtener resultados de una manera amigable. Por último, usar normalizadores y encoders ayudó a mantener la consistencia y calidad de los datos de entrada para el modelo.

Entre las limitaciones, consideramos que la red neuronal necesita un volumen considerable de datos para ser entrenada adecuadamente, porque en escenarios con pocos datos, el modelo puede sobreajustarse. También que cualquier error en la transformación de los datos puede tener un impacto significativo en la calidad de las predicciones.

Sugerencias para mejorar el proceso de desarrollo y despliegue de modelos de Machine Learning en el futuro

Como grupo consideramos que para mejorar el desarrollo y despliegue de modelos de ML, es importante automatizar el preprocesamiento de datos con scikit-learn. También ampliar el dataset, porque un mayor volumen de datos permite un entrenamiento más robusto y mejora la capacidad del modelo de generalizar. También creemos que Implementar visualizaciones más interactivas como gráficos 3D, pueden dar información más amplia que las visualizaciones básicas. Por último, sería recomendable supervisar el modelo en producción, con alertas por posibles desviaciones, para asegurar un buen rendimiento de este.