



Ingeniería en Ciencias de Datos y Matemáticas

Escuela de Ingeniería y Ciencias, Campus Monterrey

Inteligencia artificial avanzada para la ciencia de datos I (TC3006C.102)

Reto Análisis del contexto y la normatividad Momento de Retroalimentación

Presenta:

Karla Andrea Palma Villanueva (A01754270)

Docentes:

Jesús Adrian Rodríguez Rocha

Alfredo Esquivel Jaramillo

Mauricio Gonzalez Soto

Frumencio Olivas Alvarez

Antonio Carlos Bento

Hugo Terashima Marín

Monterrey, Nuevo León, México. 11 de septiembre de 2024

1. Reflexión American Housing

Ahora bien, de igual forma se manejó un proyecto en donde se pudieran poner en práctica la aplicación de las diversas métricas y técnicas ejercidas en clase. Este análisis se centró en el conjunto de datos de viviendas de California de Scikit-Learn, implementando regresión logística para clasificar las casas como superiores o inferiores a \$300,000 en valor. El conjunto de datos contenía 9 variables, incluyendo ingresos, antigüedad de la vivienda, número de habitaciones, población y datos de ubicación. Se creó una clasificación binaria basada en el valor medio de la vivienda.

El modelo alcanzó una precisión del 85 %, con un recall del 80 % para las casas más caras, lo que demuestra un buen desempeño en la identificación de propiedades de alto valor. La curva ROC mostró un AUC de 0.90, lo que indica una excelente discriminación entre las clases. A pesar de que se utilizaron técnicas de regularización para mitigar el sobreajuste el modelo mostró cierto sesgo hacia la clase mayoritaria, con un rendimiento inferior en la clase minoritaria. teniendo un nivel medio de varianza, lo que podría indicar cierto sobreajuste a los patrones de entrenamiento. En general, el modelo logró un buen rendimiento, pero tiene margen de mejora para generalizar a patrones más complejos.

Este análisis es importante para aplicaciones en el sector inmobiliario, ayudando a identificar patrones de precios de viviendas e informar decisiones de inversión. El sólido desempeño en propiedades caras es particularmente valioso para el análisis del mercado de alto nivel. Además, la capacidad de predecir con precisión el valor de las viviendas puede ser una herramienta crucial para los agentes inmobiliarios y los inversores, permitiéndoles tomar decisiones más informadas y estratégicas.

Asimismo, los desarrolladores de viviendas y los urbanistas pueden beneficiarse de este tipo de análisis al planificar nuevos proyectos. Al comprender mejor los factores que influyen en el valor de las propiedades, pueden diseñar comunidades y desarrollos que maximicen el valor para los futuros residentes. Esto no solo mejora la rentabilidad de los proyectos, sino que también contribuye a la creación de entornos de vida más atractivos y sostenibles.

2. Reflexión Titanic

El análisis de datos realizado se centró en el conjunto de datos del Titanic, obtenido a través de Kaggle, una plataforma reconocida para compartir conjuntos de datos con fines educativos y de investigación. En el respectivo archivo se almacena información histórica de pasajeros de dominio público, lo cual sirve como base para manejar los datos de manera responsable y a fines de investigación y modelación.

Los datos se utilizaron para desarrollar un modelo de aprendizaje automático que pudiera predecir la supervivencia de los pasajeros. En el proceso de desarrollo se abordaron aspectos como la imputación de valores nulos, transformación de variables categóricas y estandarización; además de desarrollar un modelo predictivo en el que se pusieron en práctica, el *cross validation* y el ajuste de hiperparámetros, para garantizar resultados confiables. El modelo final, basado en una Máquina de Vectores de Soporte (SVM), fue seleccionado en función de su rendimiento y su potencial para mejorar.

El modelo escogido para cumplir con el propósito fue el de Máquina de Vectores de Soporte (SVM), en función de su rendimiento y su potencial para mejorar. Se obtuvo un rendimiento sólido, con una precisión y un F1-score de 0.88, sin mostrar signos de sobreajuste. De igual forma se utilizaron métricas de evaluación adecuadas, como el F1-score, dada la descompensación de clases en el conjunto de datos.