

Trabajo Práctico 1 - Entrega Final

Reservas de Hotel

[75.06/95.58] Organización de Datos
Primer cuatrimestre de 2023

ALUMNO	PADRON	CORREO
BARTOCCI, Camila	105781	cbartocci@fi.uba.ar
LOURENGO, Lucía	104880	llourenco@fi.uba.ar
PATÍÑO, Franco	105126	fpatino@fi.uba.ar

Síntesis del trabajo realizado

Durante el transcurso del trabajo, se llevaron a cabo diversas etapas que involucraron el análisis y el preprocesamiento de los datos, la selección de características relevantes y la implementación de modelos de aprendizaje automático para la predicción de la cancelación de reservas de hotel. Se consideraron diferentes enfoques y se evaluaron múltiples algoritmos de clasificación, con el objetivo de obtener un modelo preciso y confiable.

Conclusiones

A continuación, se presentan las conclusiones más relevantes obtenidas a partir del trabajo realizado:

1. Importancia de las características y relevancia del preprocesamiento

Durante el proceso de ingeniería de características, se analizó la importancia de variables, detectando aquellas que más influyen en la cancelación de reservas de hotel. Aspectos como el tipo de depósito, el hecho de si se asignó o no el tipo de cuarto reservado, la antelación de la reserva, y la cantidad de cancelaciones previas mostraron una correlación significativa con las cancelaciones. Estos hallazgos resaltan la importancia de considerar las características adecuadas al desarrollar modelos predictivos en este dominio.

Tras el trabajo realizado, es claro que el preprocesamiento de datos desempeña un papel fundamental en el desarrollo de modelos de aprendizaje automático precisos y confiables. Al manipular y preparar los datos antes de alimentar a los algoritmos de entrenamiento se logra reducir la complejidad del análisis y mejorar el entendimiento del problema; el preprocesamiento tiene el potencial de influir significativamente en la calidad de las predicciones, ayudando a evitar sesgos y garantizando resultados más justos y equitativos.

2. Comparación de modelos

Se exploraron varios algoritmos de aprendizaje automático, incluyendo árboles de decisión, modelos KNN, SVM con diferentes kernels, ensambles homogéneos como XGBoost y Random Forest, ensambles híbridos como Voting y Stacking, y modelos de clasificación basados en redes neuronales, siendo el que mejor funcionó con el problema el ensamble Random Forest. Obtuvimos buenos resultados también con KNN, XGBoost y Redes Neuronales. Sin embargo, no logramos buenos modelos de ensamble híbridos ni de SVM, principalmente por la complejidad computacional que el entrenamiento de éstos amerita.

3. Evaluación de modelos

Se llevó a cabo una evaluación exhaustiva de los modelos utilizando métricas de desempeño relevantes como precisión, recall, accuracy y F1-score, en particular optimizando ésta última. También analizamos el área bajo la curva ROC, procurando que sea lo más cercana a 1 posible. Como resultado, se identificaron los modelos más efectivos para la predicción de cancelaciones de reservas de hotel, lo que proporciona una base sólida para futuras implementaciones y mejoras.

Opciones no exploradas

Aunque se lograron resultados satisfactorios en este proyecto, existen algunas opciones que quedaron fuera del alcance del trabajo actual. La exploración de las cancelaciones de reservas de hotel podría complementarse mediante el análisis de los comentarios y reseñas dejados por los clientes. La minería de texto y el procesamiento del lenguaje natural podrían ayudar a comprender mejor las razones detrás de las cancelaciones y mejorar la precisión de los modelos. Y aunque se implementaron modelos de aprendizaje automático clásicos, los modelos de aprendizaje profundo, como las redes neuronales recurrentes o las redes neuronales convolucionales, podrían capturar patrones más complejos en los datos y mejorar aún más la precisión de la predicción.