

Desafío - Clasificación de calidad de ramen

En este desafío tendrás la oportunidad de poner a prueba los conceptos aprendidos durante la sesión. Los ejercicios están diseñados para reforzar practicar lo explicado en clases y poder implementar un caso real.

Lee todo el documento antes de comenzar el desarrollo individual, para asegurarte de tener el máximo de puntaje y enfocar bien los esfuerzos. Asegúrate de seguir las instrucciones específicas en cada ejercicio y de completar los requerimientos adicionales, si los hubiera.

Tiempo asociado: 6 horas cronológicas

Descripción

En este proyecto de Machine Learning, se te ha encomendado la tarea de construir un modelo de clasificación que determine si un paquete de ramen es "bueno" o "malo" basándose en sus características específicas. Para ello, se proporciona un conjunto de datos que contiene información sobre diferentes marcas de ramen y sus atributos, así como una etiqueta que indica si el ramen fue considerado como "bueno" o "malo" por expertos o usuarios.

Para cumplir este objetivo necesitarás:

1. Realizar un análisis exhaustivo de calidad de datos para identificar y tratar posibles valores faltantes, valores atípicos y errores en los datos. Garantizar que el conjunto de datos esté limpio y preparado para su posterior procesamiento y entrenamiento de los modelos.
2. Realizar el preprocesamiento de los datos, incluyendo la normalización o estandarización de las características, codificación de variables categóricas y división del conjunto de datos en conjuntos de entrenamiento y prueba.
3. Dividir los datos en conjuntos de entrenamiento y prueba y entrenar dos modelos de clasificación: uno utilizando Regresión Logística y otro utilizando Support Vector Machine (SVM). Estos modelos servirán para determinar si un paquete de ramen es "bueno" o "malo" en función de las características proporcionadas y realiza una búsqueda de hiperparámetros.
4. Evaluar el rendimiento de ambos modelos utilizando métricas relevantes para problemas de clasificación, como precisión, recall, F1-score y el área bajo la curva ROC (AUC-ROC). A partir de esto, debes comparar y analizar los resultados obtenidos

por cada modelo para determinar cuál es más adecuado para este problema específico.

Requerimientos

1. Analiza, prepara y codifica adecuadamente los datos, y realiza su división para entrenamiento y prueba. **(3 puntos)**
2. Aplica modelos de clasificación previamente entrenados y probados. **(4 puntos)**
3. Evalúa el rendimiento de algoritmos de clasificación, utilizando las métricas adecuadas. **(3 puntos)**



¡Mucho éxito!

Consideraciones y recomendaciones

- Debes entregar la solución en un archivo de Jupyter Notebook, con el código y las explicaciones necesarias