

Aprendizaje de máquina

Modelos de clasificación

1. Para este ejercicio se utilizarán los resultados de un análisis químico realizado a distintos vinos producidos en una misma región de Italia, pero provenientes de distintos cultivos [2]. El análisis encontró la presencia de 13 componentes en 3 diferentes clases de vino. Su objetivo es construir un modelo de clasificación que permita predecir el tipo de vino a partir de sus componentes.
2. Para este ejercicio se utilizará un conjunto de datos que evalúa el nivel riesgo financiero que tiene una persona a la hora de solicitar un crédito en un banco. Los datos se encuentran disponibles en [2]. Su objetivo es construir un modelo de clasificación que permita predecir el riesgo financiero, y ayude al banco a tomar la decisión sobre qué tan riesgoso puede ser aprobar un crédito a un cliente.

En ambos casos, como parte del proceso usted deberá:

- A. Análisis descriptivo
- B. Análisis exploratorio (identifique características más relevantes)
- C. Preparación de los datos (limpieza y pipelines de preprocesamiento)
- D. Generar modelos con diferentes combinaciones de características
- E. Evaluación de desempeño en datos de entrenamiento-pruebas.
 - A. Generar matriz de confusión
 - B. Reporte de clasificación
 - C. Curva ROC para comparar los modelos
 - D. Identifique posibles casos de subajuste, sobreajuste.
- F. A partir de los resultados obtenidos, seleccione el modelo que a su modo de ver ofrece la mejor solución. Justifique su respuesta.

Para tener en cuenta

- Recuerde utilizar pipelines siempre que sea posible
- Esta actividad puede realizarse en grupos de máximo 4 personas
- La actividad tiene una duración estimada de 2 horas
- La solución al ejercicio debe entregarse a través de Google Colab
- Para la calificación se dará mayor peso a los análisis y conclusiones obtenidas

Recursos

[1] https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html

[2] https://raw.githubusercontent.com/jpospinalo/MachineLearning/main/Logistic%20Regression/german_credit_data.csv