

Prueba de progreso 1: Predicción de la Calidad del Vino. Clasificación vs regresión

2025

Usando un conjunto de datos con diversas características de vinos, predice la calidad de estos usando un algoritmo de regresión y otro de clasificación y compara el desempeño de ambos. Por lo tanto, para cada tarea (regresión y clasificación) tendrás que seleccionar y usar un algoritmo, analizar su desempeño y extraer conclusiones de los resultados. Puedes y es recomendable probar diferentes algoritmos y parámetros antes de seleccionar el que genere mejores resultados en cada tarea. Todas las pruebas deberán verse reflejadas en el cuaderno de Jupyter a entregar.

Datos

Debe usarse este conjunto de datos sobre la calidad del vino, que contiene información sobre varias propiedades químicas de cada vino, así como una puntuación de calidad.

Para cargar los datos, puede hacerse a través de una librería de Python desarrollada para ese fin que puede instalarse a través del comando pip:

```
pip install ucimlrepo
```

Y para cargarlos desde Python, puede usarse el siguiente código:

```
from ucimlrepo import fetch_ucirepo

# Descarga del dataset
wine_quality = fetch_ucirepo(id=186)

# Datos brutos en formato Dataframe de pandas
X = wine_quality.data.features
y = wine_quality.data.targets
```



```
# Metadatos
print(wine_quality.metadata)

# Información de las variables
print(wine_quality.variables)
```

A continuación tienes una tabla describiendo cada campo del conjunto de datos, pero es recomendable que profundices en ellos usando la información que tienen disponible en su página web o a través del artículo científico que publicaron las personas que crearon los datos en el que utilizan técnicas basadas en regresión.

Nombre de Variable	Función	Tipo	Faltan valores?
fixed_acidity	Característica	Continua	No
volatile_acidity	Característica	Continua	No
citric_acid	Característica	Continua	No
residual_sugar	Característica	Continua	No
chlorides	Característica	Continua	No
free_sulfur_dioxide	Característica	Continua	No
$total_sulfur_dioxide$	Característica	Continua	No
density	Característica	Continua	No
pН	Característica	Continua	No
sulphates	Característica	Continua	No
alcohol	Característica	Continua	No
quality	Objetivo	Categórica	No

Normas y estructura

- Esta tarea se hará en **grupos** previamente establecidos
- El código deberá estar alojado en un **repositorio git**, ya sea GitHub, GitLab, Codeberg o cualquier otro.
 - Antes de empezar la tarea cada grupo deberá compartir acceso al repositorio donde esté alojado el código al profesor, para que este pueda hacer seguimiento de los cambios y avances en el trabajo.
- El **envío** consistirá en uno (preferible) o varios cuadernos de **Jupyer Notebook bien documentados**, valorando particularmente el uso combinado de código Python y Markdown.



Procedimiento

En esta sección se explicitan los pasos a dar necesarios para llevar a cabo la prueba satisfactoriamente. Esto no significa que no puedan darse otros pasos similares en sustitución o complementandolos si crees que son adecuados.

• Elección de algoritmo

- Selecciona un algoritmo de regresión para predecir la calidad como una variable continua.
- Selecciona un algoritmo de clasificación para predecir la calidad como una variable categórica.

La elección del algoritmo debe ser justificada basada en su rendimiento con el conjunto de datos.

• Evaluación de los modelos

- Para regresión, usa métricas como Mean Absolute Error (MAE), Mean Square Error (MSE), o \mathbb{R}^2 .
- Para regresión, usa al menos una de las siguientes técnicas: confusion matrix, ROC o F1-score.



La elección de técnicas de evaluación de modelos apropiadas será valorada.

Comparación

Después de entrenar y evaluar los dos modelos deberás comparar sus desempeños. Para ello, incluye una sección en el cuaderno de Jupyter en la que expliques cuál de las aproximaciones funciona mejor y reflexiones acerca de por qué.

Evaluación

La valoración final de la prueba está formada por: - Un 7,5% perteneciente a la **primera** prueba de progreso. - Un 8% perteneciente a las **prácticas** de laboratorio.

En total hace un 15,5% del total de la asignatura que podrá ser recuperado si hace falta.



Envío

Deberá entregarse en el campus virtual el **cuaderno de Jupyter** (fichero .ipynb) o un zip incluyendo varios cuadernos a través de la tarea que será habilitada para dicho fin.

La fecha límite para el envío será el 30 de marzo de 2025.