

Trabajo Práctico 3

Enunciado

En este trabajo práctico deben elegir con qué dataset trabajar para resolver un problema de clasificación con algunos de los modelos presentados en el curso .

Hay datasets disponibles en:

https://github.com/Digital-House-DATA/ds_blend_2021_img/wiki/Public-datasets

O pueden elegir libremente otro dataset que tenga al menos 1000 registros.

Objetivos:

- Seleccionar los datos a modelar. Justificar esa selección.
- Entrenar al menos tres modelos de clasificación distintos sobre estos datos.
- Evaluar la performance de los modelos y comparar ventajas y desventajas de cada uno.
- Determinar la importancia de las features incluidas en la predicción. Interpretar.

Requisitos

Los materiales deberán ser entregados en una Jupyter notebook que satisfaga los requerimientos del proyecto. La notebook deberá estar debidamente comentada. Además, los grupos deberán crear un repositorio para el proyecto (anonimizado) en Github. Para la presentación deben armar algunos slides y grabar un video que exponga los resultados obtenidos en no más de 10 minutos.

Material a entregar

Presentar un informe estructurado (PPT o Google Slides). Éste debe constar en una introducción (planteo del problema, la pregunta, la justificación del dataset utilizado, etc.), un desarrollo de los análisis realizados (modelos propuestos, ventajas y desventajas, comparación de performance, visualizaciones) y una exposición de los principales resultados y conclusiones.

Fecha de entrega

- El material deberá entregarse en la **clase 34** del curso.

Evaluación

- Seleccionar los datos a modelar. Justificar esa selección.
- Entrenar al menos tres modelos de clasificación distintos sobre estos datos.
- Evaluar la performance de los modelos y comparar ventajas y desventajas de cada uno.
- Determinar la importancia de las features incluidas en la predicción. Interpretar.

/10	Discusión del dataset y presentación del problema: Selección del dataset, definición y alcance del problema
/10	Discusión del método: Entrenamiento modelo 1 de clasificación y evaluación de performance, determinación de features importantes.
/10	Discusión del método: Entrenamiento modelo 2 de clasificación y evaluación de performance, determinación de features importantes.
/10	Discusión del método: Entrenamiento modelo 3 de clasificación y evaluación de performance, determinación de features importantes.
/10	Discusión de los resultados: Comparación de los modelos, ventajas y desventajas. Comparación de importancia de features en cada modelo.
/10	Adecuada profundidad del trabajo
/10	Adecuada presentación de conclusiones
/10	Formato del informe (introducción/cuerpo/conclusión, gráficos)
/10	Presentación del informe (estructura del texto, storytelling, figuras)
/10	Presentación del video (tiempo adecuado, claridad en la exposición, información relevante para la audiencia definida)
/100	Nota final