

Análisis de datos - Trabajo práctico integrador.

Parte 2.

1. Introducción y motivación

Para este trabajo final deberá realizar la preparación del set de datos utilizado en el TP 1 para el entrenamiento de un modelo de IA.

2. Consignas

El análisis debe abordar los siguientes aspectos:

- Análisis de balance/desbalance de clases ante un posible problema de clasificación.
 - Plantear/describir el problema y analizar la variable target.
 - Proponer y aplicar mecanismos de balance en caso de ser necesario y justificar la selección.
- Análisis y selección de features:
 - Implementar tests estadísticos para evaluar relaciones entre features.
 - Realizar una selección de variables basándose en su relevancia y significancia.
 - Justificar la elección de las variables seleccionadas.
 - Crear nuevos features en caso de ser necesario. Justificar.
- Transformación de variables:
 - Aplicar técnicas de conversión de variables, por ejemplo: codificación, discretización.
 - Escalar y / o normalizar los features.
- Reducción de la dimensionalidad:
 - Implementar dos técnicas de reducción de la dimensionalidad. Comparar.
 - Comparar el dataset original con el dataset reducido, evaluando las ventajas y desventajas de la reducción.
 - Analizar cómo la reducción de la dimensionalidad afecta la interpretación y el rendimiento de futuros modelos predictivos.

3. Entrega

Consideraciones:

- Se deberá hacer una entrega por grupo a través del campus.
- Deberán informar que el repositorio GitHub (el mismo del TP 1) está actualizado con la notebook correspondiente a esta segunda parte, correctamente documentada y organizada.
- La entrega estará habilitada desde el día jueves **12/06/2025 a las 00:00** hasta el día lunes **16/06/2025 a las 23:59** (hora Argentina).
- Pasada la ventana de tiempo, el campus cierra la posibilidad de entrega de forma automática. Ante cualquier eventualidad, contactar a las docentes.
- Al finalizar, asegurarse que la actividad figure “Entregada”.

4. Evaluación

Se evaluará:

- El planteo de supuestos y limitaciones en el análisis estadístico.
- La comprensión de los temas vistos:
 - Conversión de variables
 - Selección de features
 - Reducción de dimensionalidad
 - etc.
- La calidad de las explicaciones, conclusiones y justificaciones de los pasos realizados.