

CASO PRÁCTICO 2

Ciencia y analítica de datos

Enunciado:

Utilice el dataset de cáncer de mama (Diagnostic Wisconsin Breast Cancer Database) donde tenemos los datos de 569 mujeres y 30 características relevantes, el cual se trabajó en la unidad 1 realizando un preprocesamiento de datos y un análisis exploratorio de datos. Se pretende aprender el tipo de cáncer de mama (maligno=1, benigno=0) utilizando los algoritmos de clasificación revisados en la unidad 2, así como realizar la evaluación del rendimiento de los algoritmos usando las métricas y gráficas respectivas.

Se pide:

- Descargue el dataset:
<http://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>
- Cargue el conjunto de datos en Python
- Realizar el preprocesamiento de datos y análisis exploratorio de datos necesarios.
- Implementar, entrenar y afinar todos los algoritmos de clasificación estudiados utilizando TensorFlow y Keras.
- Evaluar los algoritmos usando las métricas y gráficas de rendimiento.
- Realice un informe conciso en Word comparando el rendimiento de todos los algoritmos.
- Suba el script de Python y el informe al LMS (Moodle).

SOLUCIÓN:

Conforme a la rúbrica de conocimientos y destrezas a ser adquiridos en la Unidad (valoración porcentual 100%), se busca que el maestrante:

- Realice un preprocesamiento y análisis exploratorio de datos en Python (20%).
- Entrene y afine los diferentes algoritmos de clasificación (20%).
- Evalúe los algoritmos usando las métricas y gráficas de rendimiento (20%).
- Redacte el informe de manera concisa, comparando el rendimiento de los algoritmos (20%).
- Presentación del caso práctico de manera concisa (20%).