

PROGRAMACIÓN 2024-2

TRABAJO DE APLICACIÓN No. 3

En este tercer trabajo desarrollarás un proyecto de analítica de datos, usando técnicas de aprendizaje supervisado usando la base de datos seleccionada en el trabajo de aplicación No. 2.

Fecha de presentación: 22 de noviembre de 2024.

Objetivo

El objetivo de este proyecto es desarrollar dos modelos de aprendizaje en Python, con las bibliotecas Numpy, Pandas y Scikit Learn.

Instrucciones

1. **Base de Datos:** Se usará la misma base de datos seleccionada en [UCI Machine Learning Repository](#) para el trabajo de aplicación No. 2.
2. **Modelos de aprendizaje:** Selecciona al menos dos algoritmos de aprendizaje automático supervisado adecuados para tu problema. Utiliza técnicas de búsqueda de hiperparámetros como GridSearchCV o RandomizedSearchCV para encontrar los mejores hiperparámetros para cada algoritmo.
3. **Realizar las curvas de aprendizaje:** Utiliza las curvas de aprendizaje para evaluar el desempeño de los modelos en función del tamaño del conjunto de entrenamiento. Esto te ayudará a identificar problemas como overfitting o underfitting.
4. **Realizar una evaluación diagnóstica:** Para cada modelo, realiza un diagnóstico de posibles problemas como overfitting o bias. Proporciona una recomendación justificada sobre qué pasos seguir para mejorar el desempeño, como recolectar más datos, ajustar los hiperparámetros o probar otros algoritmos.
5. **Compara el desempeño de los dos modelos:** Una vez tengas los modelos definitivos, crea una estrategia de comparación de los dos modelos. Puedes revisar pruebas de nivel de confiabilidad entre evaluadores ([Inter-rater reliability](#))

6. **Conclusiones:** Con base en los resultados, redacta un resumen con las principales conclusiones sobre el comportamiento de los modelos analizados.

Entregables

Para la entrega, se deberá crear un repositorio en GitHub en el que se depositen todos los archivos del trabajo, que deben seguir la siguiente estructura:

```
|
+- README.md
+- 01 - modelo 1.ipynb
+- 02 - modelo 2.ipynb
+- 03 - comparación.ipynb
+- 04 - dataset.csv
+- 05 - informe.pdf
```

Descripción de los archivos del repositorio

README.md Archivo Markdown con la descripción breve del proyecto.

01 - modelo 1.ipynb y **02 - modelo 2.ipynb** muestran, respectivamente, las dos soluciones planteadas y la evaluación de los modelos (curvas de aprendizaje, matrices de confusión, etc.).

03 - comparación.ipynb muestra la estrategia de comparación de los dos modelos implementados, usando análisis de concordancia.

04 - dataset.csv archivo con los datos usados.

05 - informe.pdf archivo que resume toda la estrategia de solución planteada, los resultados obtenidos y las conclusiones a las que se logró llegar.