

TRABAJO UT6

- CLASIFICACION -

Instrucciones:

- Se deberá elaborar un Colab y subir el fichero en el tiempo de entrega establecido
- El fichero deberá nombrarse UT6_nombre_apellido1_apellido2_dni.ipynb
- Se deberá argumentar cada uno de los puntos en el trabajo con una explicación clara
- Se deberá grabar un video de duración máxima de 10 minutos, realizando una breve explicación de cómo se ha abordado cada uno de los puntos.
- El trabajo tendrá una ponderación del 70% y el video con la defensa del 30%.

Puntos del trabajo

(1 puntos) Preprocesado de los datos

- Importar dataset
- Manejar datos missing
- Manejar datos categóricos (codificación label, ordinal, one-hot, variables dummies)
- Evitar trampa de las variables dummy
- Separar datos en conjuntos de entrenamiento y test
- Estandarizar o normalizar los datos

(1 puntos) EDA básico (Colab 5.4)

- Análisis de correlaciones, pairplot, mapas de calor, etc...
- Limpiar el modelo eliminando columnas innecesarias o correlacionadas
- Conclusiones

(1 puntos) Aplicación de algoritmos de Regresión Logística para clasificación.

(1 puntos) Aplicación de algoritmos de K-Vecinos Cercanos para clasificación.

(1 puntos) Aplicación de algoritmos de máquinas de soporte vectorial para clasificación.

- Aplicar un kernel lineal
- Aplicar un kernel no lineal
- Comparar ambos resultados y sacar unas conclusiones

(1 puntos) Aplicación de algoritmos de Naive Bayes para clasificación.

(1 puntos) Aplicación de algoritmos de Arboles de decisión para clasificación.

(1 puntos) Aplicación de algoritmos de Bosques Aleatorios para clasificación.

(1 puntos) Evaluación de los modelos

- Identificar si hay overfitting. Explicarlo
- Matriz de confusión:
 - o Accuracy
 - o Precisión
 - o Sensitivity
 - o Specifity
- Curvas ROC y AUC
- Validación cruzada

- Conclusiones

(1 puntos) Prueba automática de multiples modelos e hiperparametros (GridSearch)

- Ejecución de las pruebas
- Conclusiones