

Tarea N° 2
Inteligencia Artificial

VR-JF

Fecha: Lunes 20 de Octubre de 2025

Plazo: 3 semanas

En esta tarea, deberá comparar distintos algoritmos de aprendizaje supervisado y no supervisado sobre un mismo conjunto de datos de su elección. Su dataset deberá tener al menos 10.000 filas y 7 columnas utilizables. Además, seleccione una de las columnas como su etiqueta **Y**, la cual debe ser discreta y con más de 2 clases.

1. **(30 puntos)** Deberá ejecutar tres algoritmos de clustering: K-Means, K-Means++ y MeanShift, utilizando al menos cuatro configuraciones distintas para cada técnica (deberá justificar la elección de parámetros). Para el entrenamiento, use únicamente el 80% de los datos, omitiendo la etiqueta **Y**. A continuación, evalúe las doce configuraciones obtenidas mediante una métrica de su elección (por ejemplo, *Silhouette Score*) y seleccione las tres de mejor desempeño. Luego, aplique cada una de estas configuraciones al 20% restante de los datos, asignando a cada muestra el cluster correspondiente (obtenidos desde el entrenamiento). Finalmente, compare la etiqueta **Y** real de cada muestra con la etiqueta dominante dentro del cluster al que pertenece y analice si este procedimiento resulta razonable para asignar etiquetas faltantes.
2. **(30 puntos)** Utilizando el mismo conjunto de datos previamente seleccionado, diseñe e implemente una técnica que permita entrenar en paralelo múltiples instancias (al menos tres por técnica) de Regresión Logística y SVM, variando sus hiperparámetros (por ejemplo: batch size, tasa de aprendizaje, etc, según sea el caso). Los parámetros de cada configuración deberán definirse en un archivo de configuración externo, y el entrenamiento deberá realizarse utilizando el 80% de los datos. Cada modelo se evaluará periódicamente (solo con datos de entrenamiento), y por cada cinco épocas deberá descartarse la configuración con peor desempeño entre todas las configuraciones restantes. Para las dos mejores configuraciones, presente métricas de evaluación utilizando el conjunto de testing, es decir, el 20% de los datos. Analice los resultados obtenidos en función de los hiperparámetros seleccionados.

Importante

En caso de utilizar herramientas generativas, se deberá declarar específicamente el propósito para el que fueron empleadas. Está estrictamente prohibido su uso para la elaboración de análisis de resultados.

Asimismo, algunos trabajos, seleccionados de forma aleatoria, podrán ser objeto de una interrogación oral. En tales casos, la calificación podrá ser modificada en función del

desempeño en dicha instancia. Si el trabajo no es seleccionado, la nota original se mantendrá.

Condiciones de entrega

Se indican las siguientes condiciones para la entrega de la tarea:

- La tarea se desarrolla **individual o en parejas** a través de Jupyter Notebook. Además del código, será necesario presentar un video explicativo con cada uno de los puntos de la tarea. El video deberá tener una duración máxima de 15 minutos y deberá estar apoyado con slides.
- La tarea se entregará vía el Canvas del curso. La fecha y hora límite para la entrega es el día Domingo 9 de Noviembre a las 23.59. Cualquier entrega atrasada no se considerará y será calificada con la nota mínima de inmediato. **No habrá extensión de plazo.**