

**Entregable:** Actividad publicada en el repositorio de GitHub. Debe enviarme el link del repositorio por medio del formulario: <https://forms.gle/amcV43qc2HMyX61D8> **antes del 8 noviembre miércoles a las 12:00 media noche**

Pueden nombrar el repositorio como guste y utilizar los notebook que considere. Puede copiar los códigos de clase y hasta las soluciones de internet, pero la interpretación pedida a las decisiones e interpretación deben ser originales y de su autoría, acordes con lo implementado en el código.

### PROBLEMA 1

**Modelo a implementar:** Kmeans

**Archivo fuente:** usuarios\_twitter.csv

**Objetivo:** agrupar usuarios por sus rasgos de personalidad

**Variable de entrada:** crea tu modelo seleccionando solo 3 características

- usuario: nombre en Twitter
- “op”: Openness to experience – grado de apertura mental a nuevas experiencias, curiosidad, arte
- “co”: Conscientiousness – grado de orden, prolijidad, organización
- “ex”: Extraversion – grado de timidez, solitario o participación ante el grupo social
- “ag”: Agreeableness – grado de empatía con los demás, temperamento
- “ne”: Neuroticism, – grado de neuroticismo, nervioso, irritabilidad, seguridad en sí mismo.
- Wordcoun: Cantidad promedio de palabras usadas en sus tweets

**Variable de salida:** Categoría – Actividad laboral del usuario (1. Actor/actriz, 2. Cantante, 3. Modelo, 4. Tv, series, 5. Radio, 6. Tecnología, 7. Deportes, 8. Política y 9. Escritor)

**Debes responder a:**

- ¿Cuál es el número adecuado de cluster?
- ¿Qué interpretación le dan a los grupos que se han formado?

### PROBLEMA 2

**Modelo a implementar:** RNA. Deben realizar una búsqueda de hiperparámetros mediante GridSearchCV or RandomizedSearchCV. Utilicen 5 K-Fold validación cruzada.

**Datos:** Se simulan observaciones en dos dimensiones, pertenecientes a tres grupos, cuya separación no es perfecta.

```
from sklearn.datasets import make_blobs
X, y = make_blobs(
    n_samples = 500,
    n_features = 2,
    centers = 3,
    cluster_std = 1.2,
```

```
shuffle = True,  
random_state = 0  
)
```

**Objetivo:** clasificar a qué grupo pertenece cada muestra

**Debes responder a:**

- ¿Cuál es la configuración de la RNA que te generó mejor accuracy? considerando la métrica accuracy para decidir, complejidad del modelo y gasto computacional.

**Recomendaciones:**

- La mejor solución no necesariamente es la de Gridsearch
- Solo selecciona máximo dos hiperparámetros: número de neuronas y número de capas.