

Inteligencia Artificial - Laboratorio 8 -

Instrucciones:

- Deben unirse a uno de los grupos de Canvas de nombre "Laboratorio 8 #", donde N es un número entre 1 y 23. Los grupos pueden ser de 2 a 4 personas.
- Esta actividad debe realizarse en grupos.
- Sólo es necesario que una persona del grupo suba el trabajo a Canvas.
- No se permitirá ni se aceptará cualquier indicio de copia. De presentarse, se procederá según el reglamento correspondiente.
- Tendrán hasta la fecha indicada en Canvas para entregar este laboratorio.
 - No se confíen, aprovechen el tiempo en clase para entender todos los ejercicios y avanzar lo más posible.

Task 1 - Mixture Models y Segmentación de Clientes de Banco

Al igual que el laboratorio anterior, deberá construir un modelo basado en aprendizaje no supervisado para el dataset dado. Lo que se busca con este modelo es poder segmentar a clientes de una entidad bancaria para poder así tomar decisiones a nivel comercial y potenciar las operaciones de la empresa. Recuerden que:

- Deben hacer una breve exploración con los datos. Esto implica, pero no está limitado a:
 - Hacer encoding de las variables que se necesiten
 - Revisar si el dataset está balanceado, caso no estarlo, aplicar alguna técnica para balancearlo lo más y mejor posible
 - Escalar las variables si considera necesario
 - Selección de variables
- Recuerde definir de forma clara y razonada (es decir, diga el por qué de su elección) de una métrica de desempeño principal

Task 1.1 -Mixture Models

Recordando lo que se mencionó en el laboratorio anterior para recalcar la importancia de este tipo de análisis.

En el mercado tan competitivo de la actualidad, el tener un mejor entendimiento de los clientes ya no es algo opcional para las empresas, sino un *must* estratégico y competitivo. Esto debido a que los *insights* que se extraen de estos deberían guiar las operaciones. La segmentación de clientes es un proceso en el cual se divide a las personas que usan los servicios de una empresa dada, basándose en características compartidas entre los grupos. Las diferentes empresas pueden contar con distintos tipos de datos de sus clientes desde edad, tipo de vida, ingresos y mucho más.

Usando el dataset de este [enlace](#), implemente un modelo de K-Means para segmentar a los clientes en la cantidad de grupos que ustedes decidan. Para ello considere lo siguiente

- Si considera necesario haga split del dataset como en los laboratorios anteriores
 - Recuerde que es un problema de aprendizaje no supervisado
- Seleccione el número de clusters que mejor se adapte al dataset. Justifique su elección
- ¿Podría PCA ayudarle a mejorar sus clusters? ¿Por qué?
- Muestre al menos una gráfica que muestre cómo se dividen los clusters seleccionados
 - Seleccione dos features, justifique su elección
 - Puede considerar investigar sobre Principal Component Analysis (PCA)
 - Recuerde mostrar la covarianza como una elipse para cada uno de los clusters
- Provea una métrica de desempeño, justificando su elección
- Compare sus resultados con los que obtuvo en el laboratorio 7
 - ¿Cuál es mejor? ¿Por qué?
- Considerando sus resultados del laboratorio 7, ¿cuándo cree que es mejor usar Mixture Models y K-Means?

Inteligencia Artificial - Laboratorio 8 -

Para este task **no usen librerías**, sino implementen el algoritmo por ustedes mismos. Pueden usar librerías para la lectura del archivo, métricas de desempeño, división de dataset, etc. Pero el algoritmo principal deben hacerlo ustedes.

Luego:

- Repita los pasos para entrenar su modelo, pero ahora usando librerías y compare los resultados. ¿Cuál implementación fue mejor? ¿Por qué? (Responda como parte del readme de su repositorio)

Entregas en Canvas

1. Link al repositorio de los integrantes del grupo.
 - a. Deberán subir el código también a Canvas por temas de Acreditación

Evaluación

1. [4 pts.] Task 1.1 (Incluyendo respuestas en comentarios)
2. [1 pts.] Respuestas relacionadas con el laboratorio 7

Recuerden que si no se hace un buen preprocesamiento de datos, los modelos no serán adecuados. Ergo, si obvian este paso pueden llegar a perder hasta 2 puntos por task.

Total 5 pts.