
EVALUACIÓN 3 - MODELOS DE SEGMENTACIÓN **INTELIGENCIA ARTIFICIAL**

Docente Jazna Meza Hidalgo
Noviembre 2024

DESCRIPCIÓN

Analizar cómo podrían formarse grupos de individuos con características comunes mediante el uso del modelo KMeans. Al finalizar, cada grupo debería interpretar sus características y su posible relación con aspectos como estabilidad matrimonial, nivel educativo, ocupación y otros factores personales o familiares. Deberán justificar su elección del número de clusters y evaluar la interpretación de estos grupos. Además, es clave que exploren y expliquen las posibles implicaciones de estos segmentos en términos de tendencias de comportamiento o satisfacción en las relaciones personales, si fuera el caso.

Esta actividad permitirá practicar la selección de parámetros en KMeans y la interpretación de datos segmentados dentro de un contexto social y personal relevante.

Se deberá investigar acerca del parámetro init y realizar variantes de todos los valores posibles de ese parámetro.

El conjunto de datos proporciona información sobre:

1. rate_marriage: How rate marriage
 - 1 = very poor,
 - 2 = poor,
 - 3 = fair,
 - 4 = good,
 - 5 = very good
2. age : Age
3. yrs_married: Number of years married. Interval approximations.
4. children: Number of children
5. religious: How religious,
 - 1 = not,
 - 2 = mildly,
 - 3 = fairly,
 - 4 = strongly
6. educ: Level of education,
 - 9 = grade school,

- 12 = high school,
- 14 = some college,
- 16 = college graduate,
- 17 = some graduate school,
- 20 = advanced degree

7. occupation:

- 1 = student,
- 2 = farming, agriculture; semi-skilled, or unskilled worker;
- 3 = white-collor;
- 4 = teacher counselor social worker, nurse; artist, writers; technician, skilled worker,
- 5 = managerial, administrative, business,
- 6 = professional with advanced degree

8. occupation_husb : Husband's occupation. Same as occupation.

9. affairs: measure of time spent in extramarital affairs

REQUERIMIENTOS DE LA EVALUACIÓN

1. Construcción de modelos de segmentación
2. Analizar el comportamiento de los modelos usando técnicas de escalado y el uso de un hiper parámetro
3. Analizar los grupos propuestos por uno de los modelo de segmentación

FORMATO DE LA ENTREGA

Debe entregar el notebook de base entregado completando cada una de las secciones contenidas en él. Todos los comentarios y las tablas de resultados que se requieren deben estar incluidas en el notebook.

PLAZOS DE ENTREGA

En la plataforma ADECCA, **JUEVES 21 DE NOVIEMBRE** hasta las 18:00. Se aceptan entregas posteriores de acuerdo con los siguientes descuentos de la calificación final:

FECHA DE ENTREGA	DESCUENTO A APLICAR
21 noviembre a las 18:01 horas hasta el 21 noviembre a las 18:59 horas	Descuento = 2 puntos
21 noviembre a las 19:00 horas hasta el 21 noviembre a las 19:59 horas	Descuento = 3 puntos
Después de las 20:00 del 21 noviembre	NO CONVIENE ENTREGAR

LISTA DE COTEJO

1. Generales
 - Respeta nombre del notebook entregado (620454-ModelosSegmentacion-Equipo-X.ipynb donde X corresponde a su número de equipo)
 - Indica información de los integrantes del equipo
 - Todas las interpretaciones y justificaciones aparecen como texto (evitando que aparezcan como comentarios del código)
2. Obtener el número óptimo de clusters usando 2 de las 3 técnicas de escalado revisadas en clases y de acuerdo con las siguientes combinaciones:
 - Usando inercias para calcular el óptimo:
 - Usando los datos escalados con la técnica 1 y valor por defecto de *init*
 - Usando los datos escalados con la técnica 1 y valor asignado a *init*
 - Usando los datos escalados con la técnica 2 y valor por defecto de *init*
 - Usando los datos escalados con la técnica 2 y valor asignado a *init*
 - Usando el índice de silhouette para calcular el óptimo:
 - Usando los datos escalados con la técnica 1 y valor por defecto de *init*
 - Usando los datos escalados con la técnica 1 y valor asignado a *init*
 - Usando los datos escalados con la técnica 2 y valor por defecto de *init*
 - Usando los datos escalados con la técnica 2 y valor asignado a *init*
3. Construir los modelos de segmentación considerando todos los óptimos obtenidos (al menos uno de estos modelos debe tener un índice de silhouette superior a 0.8)
4. Generar una tabla con todos los resultados anteriores indicando el número de cluster óptimo que calcula cada variante.
5. Elija el modelo cuyo índice de silhouette sea el más alto y seleccione dos de los grupos del modelo de segmentación generado considerando un análisis de:
 - Correlación de 2 ó 3 variables seleccionadas
 - Medidas estadísticas de tendencia central y de dispersión
 - Realizar 2 comentarios acerca de lo obtenido en los puntos anteriores

NOTA Cada ítem de la lista de cotejo se evalúa con presencia o ausencia.

CÁLCULO CALIFICACIÓN FINAL

La lista de cotejo define la calificación del proyecto. La calificación individual de cada estudiante va a depender de la respuesta a una pregunta asociada al desarrollo del proyecto:

1. Responde correctamente: calificación = calificación proyecto
2. Responde de forma parcial: calificación = 50% calificación proyecto
3. Responde de forma incorrecta: calificación = 40% calificación proyecto