

**Universidad Nacional Abierta y a Distancia**  
**Vicerrectoría Académica y de Investigación**  
**Curso: Análisis de datos**  
**Código: 202016908**

**Guía de actividades y rúbrica de evaluación – Tarea 4**  
Algoritmos de Aprendizaje no Supervisado

**1. Descripción de la actividad**

<b>Tipo de actividad: En grupo colaborativo</b>	
<b>Momento de la evaluación: Intermedio</b>	
<b>Puntaje máximo de la actividad: 120 puntos</b>	
<b>La actividad inicia el:</b> jueves, 18 de abril de 2024	<b>La actividad finaliza el:</b> miércoles, 15 de mayo de 2024
<b>Con esta actividad se espera conseguir los siguientes resultados de aprendizaje:</b>	
Resultado de aprendizaje 3: Aplicar algoritmos de Machine Learning no supervisado según el problema, empleando técnicas de Clustering.	
<b>La actividad consiste en:</b>	
La actividad consiste en el desarrollo de los siguientes ítems.	
<ul style="list-style-type: none"><li>• Realizar la revisión bibliográfica de la Unidad 3.</li><li>• Elaborar un cuadro sinóptico sobre los diferentes modelos de Aprendizaje no supervisado que incluya definición, casos de uso, ventajas y desventajas.</li><li>• Elaborar un listado con las siguientes definiciones: Clustering, Centroide, Dendograma, Distancia euclidean, Dispersión intra-cluster, Dispersión inter-cluster, Coeficiente de Silhouette, Índice de Calinski-Harabasz, Índice de Davies-Bouldin, Coeficiente de correlación cofenética, Inertia.</li><li>• El lenguaje a utilizar es Python, el cual se trabajará mediante Jupyter notebooks, utilizando Anaconda.</li><li>• Descargar el siguiente dataset el cual se utilizará para el desarrollo de los 2 modelos</li></ul>	

**Dataset k-means e hierarchical clustering - Mall Customer Segmentation Data:**

<https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python>

Este dataset contiene información de clientes de un mall. Las variables incluyen género, ingreso, puntaje de gasto, etc.

o descárguelo del entorno de aprendizaje junto a la guía de actividades con el nombre Anexo 5 - Dataset Mall Customer Segmentation.zip

- Con el dataset anterior diseñar los modelos de Clustering (agrupación): K-means e hierarchical clustering. Para cada algoritmo realizar los siguientes pasos:
  1. Realizar un análisis exploratorio de los datos para identificar relaciones entre variables, valores atípicos, tendencias, etc.
  2. Preprocesar los datos limpiándolos, tratando valores faltantes y transformándolos según sea necesario.
  3. Seleccionar las características más relevantes para entrenar el modelo utilizando selección de características.
  4. Entrenar el modelo configurando los diferentes hiperparámetros.
  5. Evaluar el desempeño del modelo con métricas como Coeficiente de Silhouette, Índice de Calinski-Harabasz, etc.
  6. Realizar las diferentes gráficas que permitan visualizar los resultados del modelo
  7. Interpretar, analizar y documentar los resultados obtenidos.
- Cargar en GitHub los códigos de los modelos diseñados <https://github.com/>

**Para el desarrollo de la actividad tenga en cuenta que:**

En el entorno de Información inicial debe:

- Revisar la Agenda del Curso, donde se publican las fechas de inicio y cierre de cada actividad para que se realice su entrega en las fechas allí estipuladas.

- Revisar el foro Noticias del Curso y el foro General del Curso.

En el entorno de Aprendizaje debe:

- Revisar el Syllabus del curso
- Ingresar a la Unidad 3 – Aprendizaje no supervisado - Contenidos y referentes bibliográficos y realizar las lecturas recomendadas
- Participar en el Foro de discusión - Unidad 3 - Tarea 4 - Algoritmos de Aprendizaje no Supervisado

En el entorno de Evaluación debe:

- Entregar el desarrollo de la actividad en formato PDF en el espacio Unidad 3 - Tarea 4 - Algoritmos de Aprendizaje no Supervisado - Rúbrica de evaluación y entrega de la actividad

#### **Evidencias de trabajo independiente:**

Las evidencias de trabajo independiente para entregar son:

Publicar los avances en el Foro de discusión - Unidad 3 - Tarea 4 - Algoritmos de Aprendizaje no supervisado.

#### **Evidencias de trabajo grupal:**

Las evidencias de trabajo grupal a entregar son:

En el Entorno de Evaluación **Unidad 3 - Tarea 4 - Algoritmos de Aprendizaje no supervisado**, subir un único archivo en formato PDF el cual debe nombrarse *Tarea4-GrupoNo.*, ejemplo: *Tarea4-Grupo10* y debe contener:

- Portada.
- Introducción (indicando el contenido del trabajo a entregar).
- Objetivos
- Cuadro sinóptico Aprendizaje no supervisado
- Listado de definiciones
- Enlaces de GitHub donde se puedan consultar los códigos de los modelos diseñados.
- Interpretación de los resultados de cada modelo
- Referentes bibliográficos con Normas APA.

## 2. Lineamientos generales para la elaboración de las evidencias de aprendizaje a entregar.

Para evidencias elaboradas **en grupo colaborativamente**, tenga en cuenta las siguientes orientaciones

1. Todos los integrantes del grupo deben participar con sus aportes en el desarrollo de la actividad.
2. En cada grupo deben elegir un solo integrante que se encargará de entregar el producto solicitado en el entorno que haya señalado el docente.
3. Antes de entregar el producto solicitado deben revisar que cumpla con todos los requerimientos que se señalaron en esta guía de actividades.
4. Solo se deben incluir como autores del producto entregado, a los integrantes del grupo que hayan participado con aportes durante el tiempo destinado para la actividad.

Tenga en cuenta que todos los productos escritos individuales o grupales deben cumplir con las normas de ortografía y con las condiciones de presentación que se hayan definido.

En cuanto al uso de referencias considere que el producto de esta actividad debe cumplir con las normas **APA**

En cualquier caso, cumpla con las normas de referenciación y evite el plagio académico, para ello puede apoyarse revisando sus productos escritos mediante la herramienta Turnitin que encuentra en el campus virtual.

Considere que en el acuerdo 029 del 13 de diciembre de 2013, artículo 99, se considera como faltas que atentan contra el orden académico, entre otras, las siguientes: literal e) "El plagiar, es decir, presentar como de su propia autoría la totalidad o parte de una obra, trabajo, documento o invención realizado por otra persona. Implica también el uso de citas o referencias faltas, o proponer citad donde no haya coincidencia entre ella y la referencia" y liberal f) "El reproducir, o copiar con fines de lucro, materiales educativos o resultados de productos de

investigación, que cuentan con derechos intelectuales reservados para la Universidad”

Las sanciones académicas a las que se enfrentará el estudiante son las siguientes:

- a) En los casos de fraude académico demostrado en el trabajo académico o evaluación respectiva, la calificación que se impondrá será de cero puntos sin perjuicio de la sanción disciplinaria correspondiente.
- b) En los casos relacionados con plagio demostrado en el trabajo académico cualquiera sea su naturaleza, la calificación que se impondrá será de cero puntos, sin perjuicio de la sanción disciplinaria correspondiente.

### 3. Formato de Rúbrica de evaluación

<b>Tipo de actividad: En grupo colaborativo</b>	
<b>Momento de la evaluación: Intermedio</b>	
<b>La máxima puntuación posible es de 120 puntos</b>	
<p><b>Primer criterio de evaluación:</b></p> <p>Realiza el cuadro sinóptico Aprendizaje no supervisado</p> <p><b>Este criterio representa 20 puntos del total de 120 puntos de la actividad.</b></p>	<p><b>Nivel alto:</b> El estudiante realiza de forma adecuada el cuadro sinóptico sobre los diferentes modelos de Aprendizaje no supervisado. <b>Si su trabajo se encuentra en este nivel puede obtener entre 11 puntos y 20 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante realiza parcialmente el cuadro sinóptico sobre los diferentes modelos de Aprendizaje no supervisado. <b>Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 10 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no realiza el cuadro sinóptico sobre los diferentes modelos de Aprendizaje no supervisado. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 punto</b></p>
<p><b>Segundo criterio de evaluación:</b></p> <p>Consulta las definiciones de los conceptos solicitados</p> <p><b>Este criterio representa 20 puntos del total de 120 puntos de la actividad</b></p>	<p><b>Nivel alto:</b> El estudiante consulta de forma adecuada las definiciones de los conceptos solicitados <b>Si su trabajo se encuentra en este nivel puede obtener entre 11 puntos y 20 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante consulta parcialmente las definiciones de los conceptos solicitados <b>Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 10 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no consulta las definiciones de los conceptos solicitados <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 punto</b></p>
<p><b>Tercer criterio de evaluación:</b></p> <p>Implementa los modelos de</p>	<p><b>Nivel alto:</b> El estudiante implementa de forma adecuada los modelos de K-means e Hierarchical Clustering. <b>Si su trabajo se encuentra en este nivel puede obtener entre 31 puntos y 60 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante implementa parcialmente los modelos de K-means e Hierarchical Clustering .</p>



<p>aprendizaje no supervisado</p> <p><b>Este criterio representa 60 puntos del total de 120 puntos de la actividad</b></p>	<p><b>Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 30 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no implementa los modelos de K-means e Hierarchical Clustering.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 punto</b></p>
<p><b>Cuarto criterio de evaluación:</b></p> <p>Interpreta los resultados</p> <p><b>Este criterio representa 20 puntos del total de 120 puntos de la actividad</b></p>	<p><b>Nivel alto:</b> El estudiante interpreta de forma adecuada los resultados obtenidos por los modelos.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 11 puntos y 20 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante interpreta parcialmente los resultados obtenidos por los modelos.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 10 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no interpreta los resultados obtenidos por los modelos.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 punto</b></p>