

**Universidad Nacional Abierta y a Distancia**  
**Vicerrectoría Académica y de Investigación**  
**Guía por actividades para el desarrollo del componente práctico del curso**  
**Machine Learning - 203008067**

**1. Información general del componente práctico.**

<b>Estrategia de aprendizaje:</b> Aprendizaje basado en problemas
<b>Tipo de curso:</b> Metodológico
<b>Momento de la evaluación:</b> Intermedio
<b>Puntaje máximo del componente:</b> 150 puntos
<b>Número de actividades del componente registradas en esta guía:</b> 1
<b>Con este componente se espera conseguir los siguientes resultados de aprendizaje:</b> <ul style="list-style-type: none"> <li>• Adaptar máquinas de soporte vectorial para la toma de decisiones.</li> <li>• Aplicar algoritmos de agrupamiento para estructurar datos de tal manera que puedan ser clasificados y usados en la extracción de información.</li> </ul>

**2. Descripción general actividad del componente práctico.**

<b>Escenarios de componente práctico: Con Apoyo TIC</b>	
<b>Tipo de actividad:</b> Independiente	
<b>Número de actividad:</b> 1	
<b>Puntaje máximo de la actividad:</b> 150 puntos	
<b>La actividad inicia el:</b> lunes, 10 de junio de 2024	<b>La actividad finaliza el:</b> domingo, 14 de julio de 2024
<b>Los recursos con los que debe contar para el desarrollo de la actividad son los siguientes:</b> <ul style="list-style-type: none"> <li>• Computador de escritorio o portátil.</li> <li>• Conexión a internet.</li> <li>• Software Python instalado en el computador o cuenta en Google Colab.</li> </ul>	
<b>La actividad consiste en:</b> Usar la base de datos "Anexo 2 - productividad_confeccion.csv", que se encuentra en la carpeta <a href="#">Guía para el desarrollo del componente práctico y rúbrica de evaluación - Unidad 1 - Fase 3 - Componente práctico - Prácticas simuladas</a> , para aplicar máquinas de soporte vectorial. Posteriormente, divide el conjunto de datos en <i>clusters</i> , para asignar un cluster a cada dato.	

Tenga en cuenta que debe realizar todas las actividades en Python.

### **Acción 1: Support Vector Machine.**

- Seleccionar una base de entrenamiento (80%) y una base de prueba (20%).
- Con la base de entrenamiento aplicar un modelo al nivel de productividad real (alto y bajo) de un equipo en función de las siguientes variables: el día de la semana, el cuarto del mes, departamento, número de trabajadores, tiempo estándar asignado (SMV), trabajo en progreso (WIP) y tiempo extra.
- Realizar las predicciones sobre la base de prueba. Calcule la precisión (accuracy) sobre los datos de prueba.

### **Acción 2: K-Means Clustering.**

Se quiere agrupar los equipos de trabajo, usando el método *K-Means Clustering*, basados en los datos del "Anexo 2 - productividad\_confeccion.csv", lo cual permitirá identificar patrones y similitudes entre equipos para tomar decisiones informadas sobre la gestión de la producción y la asignación de recursos. Para esto, se necesita dar respuesta a las siguientes preguntas:

- ¿Cuántos clusters sería adecuado seleccionar para representar eficientemente los diferentes tipos de equipos de trabajo?
- ¿Cuál es el perfil característico de cada cluster en términos de las variables proporcionadas?
- ¿Cómo podrían utilizarse estos clusters para mejorar la gestión de la producción y la asignación de recursos en la fábrica?

### **Acción 3: Agglomerative Clustering.**

Use el método *Agglomerative Clustering* sobre los equipos de trabajo para entender los patrones de producción compartidos. Seleccione mínimo seis variables que usted estime más relevantes para construir los clusters y explique su escogencia. Con lo anterior dé solución a lo siguiente:

- Encuentre el número óptimo de *clusters* utilizando dendrogramas.

- b. Entrene el modelo de agrupamiento jerárquico.
- c. Visualice los *clusters*. Cuál es el perfil característico de cada clúster?

#### **Acción 4: Conclusiones.**

Redacte unas conclusiones generales (en una celda markdown) de los resultados de las acciones 1 a 3.

Para el desarrollo de esta actividad, es necesario que el estudiante revise en el entorno de Aprendizaje (Unidad 2), las siguientes referencias:

Giuseppe Bonaccorso. (2018). *Machine Learning Algorithms: Popular Algorithms for Data Science and Machine Learning, 2nd Edition: Vol. 2nd ed.* Packt Publishing. (pp. 340-355).

Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning - Second Edition: Vol. 2nd ed.* Packt Publishing. (pp. 363-372).

Condiciones para entregar aporte:

Debe compartir un archivo de extensión **.ipynb** en el foro colaborativo nombrado **Fase4\_nombre\_apellido\_Grupo#.ipynb** en el formato suministrado en la carpeta [Guía para el desarrollo del componente práctico y rúbrica de evaluación - Unidad 2 - Fase 4 - Componente práctico - Prácticas simuladas](#).

Importante: el archivo .ipynb debe ejecutarse sin errores.

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

En el entorno de Información inicial debe:

- Revisar la agenda del curso.
- Revisar las subsecciones de Acompañamiento docente.

En el entorno de Aprendizaje debe:

- Interactuar, en el foro de la Fase 4, con su tutor y sus compañeros de grupo sobre las temáticas abordadas, realimentar dos o más respuestas.
- Presentar en el foro de la Fase 4 los avances de cada una de las actividades. No se aceptan enlaces a Google Colab o archivos como Word o PDF.
- Consultar el material bibliográfico de la Unidad 2.

En el entorno de Evaluación debe: Entregar el archivo en formato .ipynb, de las actividades desarrolladas y nombrado G#\_(nombre del estudiante)\_Fase\_4, por ejemplo: G1\_Pepito Perez\_Fase\_4.ipynb

**Evidencias de trabajo independiente:** Las evidencias de trabajo independiente para entregar son:

- El desarrollo de las tres actividades publicadas en el foro en debe estar con extensión .ipynb, como se solicita en las condiciones de entrega.

**Evidencias de trabajo grupal:** En esta actividad no se requieren evidencias de trabajo grupal son:

### 3. Lineamientos generales para la elaboración de las evidencias

Para evidencias elaboradas **de forma Independiente**, tenga en cuenta las siguientes orientaciones

1. Identifique los recursos y referentes de la unidad a la que corresponde la fase.
2. Intervenga en el foro de discusión aplicando las normas de netiqueta virtual, evidenciando siempre respeto por las ideas de sus compañeros y del cuerpo docente.
3. Antes de entregar el producto solicitado deben revisar que cumpla con todos los requerimientos que se señalaron en esta guía de actividades, rúbrica de evaluación y por parte del tutor en el foro de discusión.
4. No cometa fraudes, ni plagios ni actos que atenten contra el normal desarrollo académico de las actividades.

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.

#### 4. Formato de Rúbrica de evaluación

<b>Tipo de actividad: Independiente</b>	
<b>Momento de la evaluación: Intermedio</b>	
<b>La máxima puntuación posible es de 150 puntos</b>	
<b>Primer criterio de evaluación:</b>  Aplica máquinas de soporte vectorial a modelos de aprendizaje supervisado.  <b>Este criterio representa 50 puntos del total</b>	<b>Nivel alto:</b> El estudiante aplica máquinas de soporte vectorial a cada uno de los modelos propuestos. Comparte de manera oportuna sus avances en el foro colaborativo. <b>Si su trabajo se encuentra en este nivel puede obtener entre 43 puntos y 50 puntos</b>  <b>Nivel Medio:</b> El estudiante aplica máquinas de soporte vectorial, sin embargo, no lo hace para la totalidad de los modelos propuestos o no comparte de manera oportuna sus avances en el foro colaborativo. <b>Si su trabajo se encuentra en este nivel puede obtener entre 30 puntos y 42 puntos</b>



<p><b>de 150 puntos de la actividad.</b></p>	<p><b>Nivel bajo:</b> El estudiante no aplica máquinas de soporte vectorial a los modelos de regresión y clasificación y no participa de manera oportuna en el foro colaborativo <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 29 puntos</b></p>
<p><b>Segundo criterio de evaluación:</b></p> <p>Aplica K-Means para encontrar <i>cluster centers</i> para los datos</p> <p><b>Este criterio representa 50 puntos del total de 150 puntos de la actividad</b></p>	<p><b>Nivel alto:</b> El estudiante aplica el algoritmo <i>k-means clustering</i> a los datos correctamente. Comparte de manera oportuna sus avances en el foro colaborativo <b>Si su trabajo se encuentra en este nivel puede obtener entre 43 puntos y 50 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante aplica el algoritmo <i>k-means clustering</i> a los datos, sin embargo, no lo logra encontrar un <i>cluster centers</i> representativos para regiones de datos o no comparte de manera oportuna sus avances en el foro colaborativo. <b>Si su trabajo se encuentra en este nivel puede obtener entre 30 puntos y 42 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no aplica el algoritmo <i>k-means clustering</i> a los datos y no participa de manera oportuna en el foro colaborativo. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 29 puntos</b></p>
<p><b>Tercer criterio de evaluación:</b></p> <p>Adapta algoritmos de agrupamiento a los datos para obtener dendrogramas.</p> <p><b>Este criterio representa 40</b></p>	<p><b>Nivel alto:</b> El estudiante adapta correctamente el algoritmo de agrupamiento y obtiene dendrogramas que permiten interpretar los resultados. Comparte de manera oportuna sus avances en el foro colaborativo <b>Si su trabajo se encuentra en este nivel puede obtener entre 34 puntos y 40 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante adapta el algoritmo agrupamiento a los datos, sin embargo, no lo logra trazar dendrogramas o no comparte de manera oportuna sus avances en el foro colaborativo. <b>Si su trabajo se encuentra en este nivel puede obtener entre 24 puntos y 33 puntos</b></p>

<p><b>puntos del total de 150 puntos de la actividad</b></p>	<p><b>Nivel bajo:</b> El estudiante no adapta el algoritmo de agrupamiento a los datos y no participa de manera oportuna en el foro colaborativo  <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 23 puntos</b></p>
<p><b>Cuarto criterio de evaluación:</b></p> <p>Documento solicitado en el entorno de Evaluación.</p> <p><b>Este criterio representa 10 puntos del total de 150 puntos de la actividad</b></p>	<p><b>Nivel alto:</b> El estudiante entrega, en carpeta comprimida ZIP en el entorno de evaluación, todos los elementos solicitados.  <b>Si su trabajo se encuentra en este nivel puede obtener entre 9 puntos y 10 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante entrega el documento, en carpeta comprimida ZIP en el entorno de evaluación, pero no cumple con todos los elementos solicitados.  <b>Si su trabajo se encuentra en este nivel puede obtener entre 6 puntos y 8 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no entrega, en carpeta comprimida ZIP en el entorno de Evaluación, todos los elementos solicitados .  <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 5 puntos</b></p>