

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.

Estrategia de aprendizaje: Aprendizaje basado en problemas
Tipo de curso: Metodológico
Momento de la evaluación: Intermedio
Puntaje máximo del componente: 150 puntos
Número de actividades del componente registradas en esta guía: 1
Con este componente se espera conseguir los siguientes resultados de aprendizaje: Aplicar modelos de clasificación y tareas de regresión en máquinas de aprendizaje automático para la toma de decisiones.

2. Descripción general actividad(es) del componente práctico.

Escenarios de componente práctico: Con Apoyo TIC	
Tipo de actividad: Independiente	
Número de actividad: 1	
Puntaje máximo de la actividad: 150 puntos	
La actividad inicia el: viernes, 22 de marzo de 2024	La actividad finaliza el: jueves, 18 de abril de 2024
Los recursos con los que debe contar para el desarrollo de la actividad son los siguientes: <ul style="list-style-type: none"> • Computador de escritorio o portátil. • Conexión a internet. • Software Python instalado en el computador o cuenta en Google Colab. • Kane, F. (2017). <i>Hands-On Data Science and Python Machine Learning</i>. Packt Publishing. • Raschka, S., & Mirjalili, V. (2017). <i>Python Machine Learning - Second Edition: Vol. 2nd ed.</i> Packt Publishing. 	
La actividad consiste en: Usar la base de datos "Anexo 2 - productividad_confeccion.csv", que se encuentra en la carpeta Guía para el desarrollo del componente práctico y rúbrica de evaluación - Unidad 1 - Fase 3 - Componente práctico - Prácticas simuladas , para aplicar modelos supervisados de machine learning.	

Tenga en cuenta que debe realizar toda la actividad en Python.

Contexto de la base de datos: La industria de la confección es uno de los ejemplos clave de la globalización industrial en esta era moderna. Es una industria altamente intensiva en mano de obra con muchos procesos manuales. Satisfacer la enorme demanda global de productos de confección depende en gran medida del desempeño de producción y entrega de los empleados en las empresas de fabricación de prendas. Por lo tanto, es altamente deseable entre los tomadores de decisiones en la industria de la confección rastrear, analizar y predecir el rendimiento de productividad de los equipos de trabajo en sus fábricas.

La data contiene información de las variables:

- a. *date*: Fecha en MM-DD-YYYY
- b. *day*: Día de la semana.
- c. *quarter*: Cuarto del mes. Un mes se dividió en cuatro partes.
- d. *department*: Departamento asociado a la instancia
- e. *team_no*: Número de equipo asociado con la instancia
- f. *no_of_workers*: Número de trabajadores en cada equipo.
- g. *no_of_style_change*: Número de cambios en el estilo de un producto.
- h. *targeted_productivity*: Productividad objetivo establecida por la Autoridad para cada equipo en cada día.
- i. *smv*: Valor Minuto Estándar, es el tiempo asignado para una tarea.
- j. *wip*: Trabajo en progreso. Incluye la cantidad de elementos sin terminar por productos.
- k. *over_time*: Representa la cantidad de tiempo extra de cada equipo en minutos.
- l. *incentive*: Representa la cantidad de incentivo financiero (en BDT) que permite o motiva un curso de acción particular.
- m. *idle_time*: La cantidad de tiempo que la producción estuvo interrumpida por varias razones.
- n. *idle_men*: El número de trabajadores que estuvieron inactivos debido a la interrupción de la producción.
- o. *actual_productivity*: Porcentaje real de productividad que entregaron los trabajadores.

Acción 1: Aplicación modelos lineales.

- Seleccionar una base de entrenamiento (80%) y una base de prueba (20%).
- Con la base de entrenamiento aplicar un modelo de regresión lineal para predecir la productividad real a partir del número de trabajadores de cada equipo y el tiempo extra en minutos.
- Realizar las predicciones sobre la base de prueba.
- Con los resultados del ítem anterior, realizar un diagrama de dispersión, en el eje X ubicar los valores predichos y en el eje Y ubicar los valores de productividad real.

Acción 2: Aplicación modelo logístico.

- Construir la variable 'nivel_productividad' con niveles alto y bajo. Considere nivel alto a productividad real > 0.7 y nivel bajo a productividad real ≤ 0.7 .
- Con la base de entrenamiento aplicar un modelo de regresión logística para predecir el 'nivel_productividad' basado en el número de trabajadores de cada equipo y el tiempo extra en minutos.
- Realizar las predicciones sobre la base de prueba. Calcule la precisión (accuracy) sobre los datos de prueba.

Acción 3: Aplicación K-NN.

- Con la base de entrenamiento, aplicar un modelo KNN de aprendizaje automático que pueda aprender del número de trabajadores de cada equipo y el tiempo extra en minutos, para predecir "nivel_productividad".
- Realizar las predicciones sobre la base de prueba. Calcule la precisión (accuracy) sobre los datos de prueba.

Acción 4: Aplicación Árboles de decisión (CART).

- Con la base de entrenamiento, aplicar un modelo de árboles de decisión para distinguir "nivel_productividad" bajo y alto usando todas las características.
- Usando la función `graphviz()` visualizar el diagrama de árbol.
- Realizar las predicciones sobre la base de prueba. Calcule la precisión (accuracy) sobre los datos de prueba.

Acción 5: Conclusiones.

Redacte unas conclusiones generales (en una celda markdown) de los resultados de las acciones 1 a 4.

Para el desarrollo de estas acciones, es necesario que el estudiante revise en el entorno de Aprendizaje (Unidad 1), las siguientes referencias:

Kane, F. (2017). *Hands-On Data Science and Python Machine Learning*. Packt Publishing. (pp. 183-195).

Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning - Second Edition: Vol. 2nd ed*. Packt Publishing. (pp. 90-103).

Condiciones para entregar aporte:

Debe compartir un archivo de extensión **.ipynb** en el foro colaborativo nombrado **Fase3_nombre_apellido_Grupo#.ipynb** en el formato suministrado en la carpeta [Guía para el desarrollo del componente práctico y rúbrica de evaluación - Unidad 1 - Fase 3 - 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 3, con su tutor y sus compañeros de grupo sobre las temáticas abordadas, realimentar dos o más respuestas.
- Presentar en el foro de la Fase 3 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 1.

En el entorno de Evaluación debe: Entregar el archivo en formato .ipynb, de las actividades desarrolladas y nombrado G#_(nombre del estudiante)_Fase_3, por ejemplo: G1_Pepito Perez_Fase_3.ipynb

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

- El desarrollo de las cinco 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.

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 debe 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

finés 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

Tipo de actividad: Independiente	
Momento de la evaluación: Intermedio	
La máxima puntuación posible es de 150 puntos	
<p>Primer criterio de evaluación:</p> <p>Aplica el algoritmo k-NN y presenta conclusiones.</p> <p>Este criterio representa 30 puntos del total de 150 puntos de la actividad.</p>	<p>Nivel alto: El estudiante aplica correctamente el algoritmo k-NN. Comparte de manera oportuna sus avances en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 26 puntos y 30 puntos</p> <p>Nivel Medio: El estudiante no aplica correctamente el algoritmo k-NN y/o no realiza los ajustes sugeridos por el tutor y/o no comparte de manera oportuna sus avances en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 18 puntos y 25 puntos</p> <p>Nivel bajo: El estudiante no aplica el algoritmo k-NN y no participa de manera oportuna en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 17 puntos</p>
<p>Segundo criterio de evaluación:</p> <p>Construye Modelos lineales usando regresión lineal y logística y presenta conclusiones.</p> <p>Este criterio representa 60 puntos del total</p>	<p>Nivel alto: El estudiante construye todos los diferentes modelos lineales solicitados. Comparte de manera oportuna sus avances en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 51 puntos y 60 puntos</p> <p>Nivel Medio: El estudiante construye modelos lineales, sin embargo, no lo hace para la totalidad de modelos solicitados y/o no realiza los ajustes sugeridos por el tutor y/o no comparte de manera oportuna sus avances en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 36 puntos y 50 puntos</p>

de 150 puntos de la actividad	Nivel bajo: El estudiante no construye modelos lineales y no participa de manera oportuna en el foro colaborativo. Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 35 puntos
Tercer criterio de evaluación: Aplica árboles de decisión a modelos de regresión y clasificación y presenta conclusiones. Este criterio representa 50 puntos del total de 150 puntos de la actividad	Nivel alto: El estudiante aplica árboles de decisión a cada uno de los modelos propuestos. Comparte de manera oportuna sus avances en el foro colaborativo. Si su trabajo se encuentra en este nivel puede obtener entre 43 puntos y 50 puntos Nivel Medio: El estudiante aplica árboles de decisión, sin embargo, no lo hace para la totalidad de los modelos propuestos y/o no realiza los ajustes sugeridos por el tutor y/o no comparte de manera oportuna sus avances en el foro colaborativo. Si su trabajo se encuentra en este nivel puede obtener entre 30 puntos y 42 puntos Nivel bajo: El estudiante no aplica árboles de decisión a los modelos de regresión y clasificación y no participa de manera oportuna en el foro colaborativo. Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 29 puntos
Cuarto criterio de evaluación: Documento solicitado en el entorno de Evaluación. Este criterio representa 10 puntos del total de 150 puntos de la actividad	Nivel alto: El estudiante entrega, en carpeta comprimida ZIP en el entorno de evaluación el archivo solicitado. Si su trabajo se encuentra en este nivel puede obtener entre 9 puntos y 10 puntos Nivel Medio: El estudiante entrega el documento, en carpeta comprimida ZIP en el entorno de evaluación, pero no cumple con todos los elementos solicitados. Si su trabajo se encuentra en este nivel puede obtener entre 6 puntos y 8 puntos

	<p>Nivel bajo: El estudiante no entrega, en carpeta comprimida ZIP en el entorno de Evaluación, con todos los elementos solicitados</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 5 puntos</p>
--	---