

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 3
 Algoritmos de Aprendizaje Supervisado

1. Descripción de la actividad

Tipo de actividad: En grupo colaborativo	
Momento de la evaluación: Intermedio	
Puntaje máximo de la actividad: 120 puntos	
La actividad inicia el: jueves, 21 de marzo de 2024	La actividad finaliza el: miércoles, 17 de abril de 2024
Con esta actividad se espera conseguir los siguientes resultados de aprendizaje: Resultado de aprendizaje 2: Aplicar algoritmos de Machine Learning supervisado según el problema, empleando métodos de modelado predictivo como regresión y clasificación	
La actividad consiste en: La actividad consiste en el desarrollo de los siguientes ítems. <ul style="list-style-type: none"> • Realizar la revisión bibliográfica de la Unidad 2. • Elaborar un cuadro sinóptico sobre los diferentes modelos de Aprendizaje Supervisado que incluya definición, casos de uso, ventajas y desventajas. • Elaborar un listado con las siguientes definiciones: Datos de Train, Datos de Validation y Test, GridSearchCV, One Hot Encoding, Matriz de confusión, Precision, Accuracy, Specificity, Recall, F1 Score, curva ROC, R cuadrado. • El lenguaje a utilizar es Python, el cual se trabajará mediante Jupyter notebooks, para esto es necesario instalar Anaconda, que es una distribución libre y abierta de los lenguajes Python y R, utilizada en ciencia de datos, y aprendizaje automático. 	

Descargue e instale Anaconda desde:

<https://www.anaconda.com/products/distribution>

- ✓ Dentro de Anaconda, instalar y ejecutar (Launch) Jupyter notebooks
- Descargar los datasets: Realizar el registro en la plataforma Kaggle desde <https://www.kaggle.com/> botón "Register"

1. **Dataset Regresión Lineal** – Vehicle dataset:
<https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho>

Contiene información técnica y de mercado de automóviles. El objetivo es predecir el precio del automóvil en función de sus atributos

2. **Dataset Regresión Logística** – Heart Disease Cleveland UCI:
<https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci>

Contiene información de pacientes como edad, sexo, presión sanguínea, colesterol, etc. la variable objetivo es la presencia de enfermedad cardíaca.

3. **Dataset Árboles de Decisión** – Red Wine Quality:
<https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009>

Contiene información de 10 clases de calidad del vino

o descárguelos del entorno de aprendizaje junto a la guía de actividades con los nombres Anexo 2 - Dataset Vehicle.zip, Anexo 3 - Dataset Heart Disease Cleveland UCI.zip y Anexo 4 - Dataset Red Wine Quality.zip

- Con los datasets anteriores diseñar los modelos predictivos de Regresión Lineal, Regresión Logística y Árboles de decisión. 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. Dividir el dataset en Train y Test para evaluar correctamente el modelo.
 5. Entrenar el modelo configurando los diferentes hiperparámetros.
 6. Evaluar el desempeño del modelo en el conjunto de Test con métricas como precisión, recall, F1-score, etc.
 7. Realizar las diferentes gráficas que permitan visualizar los resultados del modelo.
 8. Interpretar, analizar y documentar los resultados obtenidos.
- Crear una cuenta en GitHub y cargar 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 2 – Aprendizaje supervisado - Contenidos y referentes bibliográficos y realizar las lecturas recomendadas
- Participar en el Foro de discusión - Unidad 2 - Tarea 3 - Algoritmos de Aprendizaje Supervisado

En el entorno de Evaluación debe:

- Entregar el desarrollo de la actividad en formato PDF en el espacio Unidad 2 - Tarea 3 - Algoritmos de Aprendizaje 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 2 - Tarea 3 - Algoritmos de Aprendizaje Supervisado.

Evidencias de trabajo grupal:

Las evidencias de trabajo grupal a entregar son:

En el Entorno de Evaluación **Unidad 2 - Tarea 3 - Algoritmos de Aprendizaje Supervisado**, subir un único archivo en formato PDF el cual debe nombrarse *Tarea3-GrupoNo.*, ejemplo: *Tarea3-Grupo10* y debe contener:

- Portada.
- Introducción (indicando el contenido del trabajo a entregar).
- Objetivos
- Cuadro sinóptico Aprendizaje 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

Tipo de actividad: En grupo colaborativo	
Momento de la evaluación: Intermedio	
La máxima puntuación posible es de 120 puntos	
<p>Primer criterio de evaluación:</p> <p>Realiza el cuadro sinóptico Aprendizaje supervisado</p> <p>Este criterio representa 15 puntos del total de 120 puntos de la actividad.</p>	<p>Nivel alto: El estudiante realiza de forma adecuada el cuadro sinóptico sobre los diferentes modelos de Aprendizaje supervisado. Si su trabajo se encuentra en este nivel puede obtener entre 8 puntos y 15 puntos</p> <p>Nivel Medio: El estudiante realiza parcialmente el cuadro sinóptico sobre los diferentes modelos de Aprendizaje supervisado. Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 7 puntos</p> <p>Nivel bajo: El estudiante no realiza el cuadro sinóptico sobre los diferentes modelos de Aprendizaje supervisado. Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 punto</p>
<p>Segundo criterio de evaluación:</p> <p>Consulta las definiciones de los conceptos solicitados</p> <p>Este criterio representa 15 puntos del total de 120 puntos de la actividad</p>	<p>Nivel alto: El estudiante consulta de forma adecuada las definiciones de los conceptos solicitados Si su trabajo se encuentra en este nivel puede obtener entre 8 puntos y 15 puntos</p> <p>Nivel Medio: El estudiante consulta parcialmente las definiciones de los conceptos solicitados Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 7 puntos</p> <p>Nivel bajo: El estudiante no consulta las definiciones de los conceptos solicitados Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 punto</p>
<p>Tercer criterio de evaluación:</p> <p>Implementa los modelos de</p>	<p>Nivel alto: El estudiante implementa de forma adecuada los modelos de regresión lineal, regresión logística y árboles de decisión. Si su trabajo se encuentra en este nivel puede obtener entre 38 puntos y 75 puntos</p> <p>Nivel Medio: El estudiante implementa parcialmente los modelos de regresión lineal, regresión logística y árboles de decisión.</p>

<p>aprendizaje supervisado</p> <p>Este criterio representa 75 puntos del total de 120 puntos de la actividad</p>	<p>Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 37 puntos</p> <p>Nivel bajo: El estudiante no implementa los modelos de regresión lineal, regresión logística y árboles de decisión.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 punto</p>
<p>Cuarto criterio de evaluación:</p> <p>Interpreta los resultados</p> <p>Este criterio representa 15 puntos del total de 120 puntos de la actividad</p>	<p>Nivel alto: El estudiante interpreta de forma adecuada los resultados obtenidos por los modelos.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 8 puntos y 15 puntos</p> <p>Nivel Medio: El estudiante interpreta parcialmente los resultados obtenidos por los modelos.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 2 puntos y 7 puntos</p> <p>Nivel bajo: El estudiante no interpreta los resultados obtenidos por los modelos.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 1 punto</p>