

Universidad Nacional Abierta y a Distancia
Vicerrectoría Académica y de Investigación

Curso: Análisis de datos

Código: 202016908

Guía de aprendizaje– Tarea 3 Algoritmos de Aprendizaje Supervisado

1. Datos de la/el Tarea

Tabla 1. Tabla de descripción

Aspecto	Descripción
1. Tipo de actividad	Colaborativa
2. Momento de la evaluación	Intermedio
3. Unidad gestora	Escuela de Ciencias Básicas Tecnología e Ingeniería ECBTI
4. Puntaje de la/el Tarea	120 puntos
5. La actividad inicia el:	lunes, 17 de marzo de 2025
6. La actividad finaliza el:	domingo, 13 de abril de 2025
7. Horas de trabajo independiente del estudiante	36 horas

2. Descripción detallada de la actividad de aprendizaje

Con el desarrollo de esta actividad se espera que se alcance el siguiente resultado 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:

- Realizar la revisión bibliográfica de la Tarea 3
- 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/chenngs/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

- 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 esta actividad se requieren los siguientes materiales y recursos:

- Carlos Véliz. (2020). [Aprendizaje automático. Introducción al aprendizaje profundo](#). El Fondo Editorial de la Pontificia Universidad Católica del Perú.
https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2600876&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_I Cap 3, 4, 5 y 6
- David Julian. (2016). [Designing Machine Learning Systems with Python](#). Packt Publishing.
https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1218065&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover. Cap 2
- Giuseppe Bonaccorso. (2018). [Machine Learning Algorithms](#) : Popular Algorithms for Data Science and Machine Learning, 2nd Edition: Vol. 2nd ed. Packt Publishing.
<https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1881497&lang>

=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover Cap 3, 4, 5 y 8

- Minguillón, J. Casas, J. y Minguillón, J. (2017). [Minería de datos: modelos y algoritmos](https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/58656). Editorial UOC. <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/58656>. Cap 4, 5, 13
- Pratap Dangeti. (2017). [Statistics for Machine Learning](https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1560931&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover) : Build Supervised, Unsupervised, and Reinforcement Learning Models Using Both Python and R. Packt Publishing. https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1560931&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover Cap 2, 3 y 4
- Romero Villafranca, R. y Zúnica Ramajo, L. (2020). [Métodos estadísticos para ingenieros](https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/129644). Editorial de la Universidad Politécnica de Valencia. <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/129644>. Cap 12
- Ramírez Cortés, B. S. . (2022). [Big Data Analytics aplicado al estudio del desempeño académico en un curso universitario](https://doi.org/10.22490/25394088.6493). Publicaciones E Investigación, 16(4). <https://doi.org/10.22490/25394088.6493>

Para el desarrollo de esta actividad debe seguir los siguientes pasos:

Paso 1: Revisar el Syllabus del curso

Paso 2: 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.

Paso 3: Revisar el foro Noticias y foro General del Curso

Paso 4: Revisar los contenidos y referentes bibliográficos de la Tarea 3

Paso 5: Participar en el Foro de discusión - Tarea 3 - Algoritmos de Aprendizaje Supervisado

Paso 6: Entregar el desarrollo de la actividad en formato PDF en el espacio Tarea 3 - Algoritmos de Aprendizaje Supervisado - Rúbrica de evaluación y entrega de la actividad

3. Indicaciones para el desarrollo y entrega de las evidencias de aprendizaje.

Las evidencias de aprendizaje son las acciones, productos o procesos observables que se realizan y/o entregan para manifestar las capacidades, habilidades, aptitudes y actitudes adquiridas, y que, a su vez, servirán al docente para verificar y evaluar su desempeño.

Las evidencias a desarrollar independientemente son:

- Publicar los avances en el Foro de discusión - Tarea 3 - Algoritmos de Aprendizaje Supervisado.

Las evidencias a desarrollar colaborativamente son:

En el Entorno de Aprendizaje y Evaluación **Tarea 3 - Algoritmos de Aprendizaje Supervisado – Rúbrica de evaluación y entrega de la actividad**, subir un único archivo en formato PDF el cual debe nombrarse *Tarea3-GrupoNo.*, ejemplo: *Tarea3-Grupo10* y debe contener:

- Portada.
- Introducción
- 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.

Para su desarrollo y entrega 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 independientes 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.

4. Situaciones de orden académico

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.