

Universidad Nacional Abierta y a Distancia

Vicerrectoría Académica y de Investigación

Guía de aprendizaje para el desarrollo del componente práctico del curso Machine Learning Código: 203008067

1. Información general del componente práctico.

Tabla 1. Información general del componente práctico

Aspecto	Descripción
1. Estrategia metodológica	Aprendizaje basado en problemas
2. Tipología de curso	Metodológico
3. Momento de la evaluación	Intermedio
4. Puntaje de la actividad	150
5. Número de actividades del componente registradas en esta guía	1
6. Horas de trabajo independiente del estudiante	27
7. Horas de acompañamiento docente	9
8. Tipo de práctica formativa	Simulada

2. Con esta/s actividad/es de componente práctico se espera que los estudiantes logren y evidencien el siguiente resultado de aprendizaje:

Aplicar modelos de regresión y de clasificación en máquinas de aprendizaje automático para la toma de decisiones.

3. Descripción general de la(s) actividad(es) del componente práctico.

Tabla 2. *Información actividad 1*

Aspecto	Descripción
1. Escenarios de componente práctico	Simulado con apoyo de TIC
2. Tipo de actividad	Independiente
3. Número de actividad	2
4. La actividad inicia el:	martes, 2 de septiembre de 2025
5. La actividad finaliza el:	lunes, 13 de octubre de 2025

Los recursos con los que debe contar para el desarrollo de la actividad son los siguientes:

- Computador de escritorio o portátil. Conexión a internet.
- Software Python instalado en el computador o cuenta en Google Colab.
- Frank Kane. (2017). *Hands-On Data Science and Python Machine Learning : Perform Data Mining and Machine Learning Efficiently Using Python and Spark*. Packt Publishing. (pp 133-155).
- Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning - Second Edition: Vol. 2nd ed.* Packt Publishing. (pp 18-105).

La actividad consiste en:

Para el desarrollo de esta guía cada estudiante deberá seleccionar un ítem A, B, C, D o E y desarrollar los ejercicios 1, 2, 3 y 4 de acuerdo con la selección. Además, cada estudiante deberá manifestar su elección en el foro de la Fase 2 por medio de la siguiente tabla

Tabla 3. Selección ítem ejercicios

Nombre estudiante	Ítem
	A
	B
	C
	D
	E

El ejercicio 5 consiste en seleccionar un literal desarrollado por uno de sus compañeros y realizar una retroalimentación. A continuación, debes presentar en el foro las sugerencias y/o ajustes a realizar identificados en el desarrollo presentado por su compañero.

Debe presentar un solo documento con los 4 ejercicios desarrollados en el formato **Anexo 1 – Plantilla Fase 2.**

Ejercicio 1: Contextualización Dataset

De acuerdo a la selección en la **Tabla 3**, cada estudiante debe descargar los datasets descritos en el ítem correspondiente y en una celda markdown escribir con sus propias palabras la información respectiva:

- Contexto datasets.
- Tamaños.
- Variable objetivos.
- Variables asociadas.

A: Importe el dataset live-disorders (ID:8) para Regresión y credit-g (ID:31) para Clasificación.

B: Importe el dataset cholesterol (ID:204) para Regresión y blood-transfusion-service-center (ID:1464) para Clasificación.

C: Importe el dataset cloud (ID:210) para Regresión and phoneme (ID:1489) para Clasificación.

D: Importe el dataset bodyfat (ID:560) para Regresión y diabetes (ID:37) para Clasificación.

E: Importe el dataset cleveland (ID:194) para Regresión y steel-plates-fault (ID:1504) para Clsificación.

Ejercicio 2: Análisis cuantitativo y cualitativo

Según los datasets correspondientes al ítem escogido, realice un preprocesamiento de datos (eliminación valores nulos, outliers), un análisis cualitativo y cuantitativo entre las variables de la data de características y la data de la variable objetivo, para esto use:

1. Diagramas de dispersión.
2. Diagrama de barras.
3. Boxplots.
4. Tablas de contingencia.
5. Matriz de correlaciones por medio de un heatmap.

Ejercicio 3: Modelos de regresión.

Por medio del comando *train_test_split* realice una división del dataset en 75% para entrenamiento y 25% para prueba y aplique los siguientes modelos:

1. Regresión lineal.
2. Ridge.
3. Lasso.
4. Árbol de decisión.

Realice una tabla donde se comparen los resultados de las métricas MSE, MAE, RMSE y R2. Explique lo presentado en la tabla.

Ejercicio 4: Modelos de clasificación

Por medio del comando *train_test_split* realice una división del dataset en 70% para entrenamiento y 30% para prueba y aplique los siguientes modelos:

1. Regresión logística.
2. Árbol de decisión.
3. KNN.
4. Perceptrón.

Realice una tabla donde se comparen los resultados de las métricas accuracy, precisión, recall y F1-score. Explique lo presentado en la tabla. Para el mejor modelo realice una matrix de confusión y explique.

Ejercicio 5: Feedback

Adjunte un screenshot del feedback realizado a uno de sus compañeros en el foro de la **Fase 2**.

Evidencias de trabajo independiente:

Las evidencias de trabajo independiente para entregar son:

- El aporte de los 4 ejercicios publicados en el **foro** deben estar en formato **PDF** o **JPG**. En el entorno de **Evaluación** se adjuntará el archivo **.ipynb** en el formato que se encuentra en la carpeta Guía para el desarrollo del componente práctico y rúbrica de evaluación - Unidad 1 - Fase 2 - Componente práctico - Prácticas simuladas (**Anexo 1 - Plantilla Fase 2**), para aplicar modelos supervisados de machine learning.
- Entrega de un archivo en formato .pynb de los ejercicios desarrollados y nombrado *G#_(nombre_apellido)_Fase2*, por ejemplo:

G15_Rafael_Gaitan_Fase2.ipynb

Evidencias de trabajo grupal:

En esta actividad no se requieren evidencias de trabajo grupal:

4. Lineamientos generales para la elaboración de las evidencias

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

Todos los integrantes del grupo deben participar con sus aportes en el desarrollo de la actividad.

En cada grupo deben elegir un solo integrante que se encargará de entregar el producto solicitado en el entorno o escenario que haya señalado el docente.

Antes de entregar el producto solicitado deben revisar que cumpla con todos los requerimientos que se señalaron en esta guía de actividades de componente práctico.

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 Elija un elemento.

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.

5. 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 citas 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.