

Guía de Análisis de Datos

Docente: Daniel Alexis Nieto Mora

Departamento: Sistemas

Asignatura: Análisis de datos

Carrera: Ingeniería en Ciencia de Datos

Código del grupo: 190304018-1

Actividad Evaluativa: 3

1. Indicaciones Generales

El plazo de entrega para esta Actividad Evaluativa 3 está fijado para el día **martes 14 de Octubre**. No se permitirán entregas posteriores a esta fecha.

- Los notebooks se deben subir al mismo repositorio de GitHub utilizado en la entrega 2, en una carpeta llamada **evaluación 3**.
- Se podrá realizar una presentación en clase (recomendado) o un video de máximo 25 minutos, equivalente al tiempo máximo de exposición.
- En caso de optar por el video, debe existir un formato de presentación (diapositivas) en el que se evidencie la participación de todos los integrantes y se desarrollen de forma coherente, ordenada y profundo los temas asignados.

2. Temas de Exposición por Equipos

Cada equipo debe preparar una exposición de **20 a 25 minutos** sobre el tema asignado, apoyándose en los datasets sugeridos para la parte práctica.

Eq1. Fundamentos de Machine Learning y primer caso aplicado

- Conceptos básicos: IA vs ML vs Deep Learning.
- Tipos de aprendizaje: supervisado, no supervisado y por refuerzo.
- Flujo de trabajo en proyectos de ML.

- **Aplicación práctica:** regresión lineal simple.
- Dataset 1: House Prices - Kaggle
- Dataset 2: California Housing - Scikit-learn

Eq2. Algoritmos de regresión en Machine Learning

- Regresión lineal, polinómica y logística.
- Supuestos y limitaciones.
- **Aplicación práctica:** regresión y clasificación binaria.
- Dataset 1: Student Performance - Kaggle
- Dataset 2: Breast Cancer Wisconsin - UCI

Eq3. Árboles de decisión y métodos de ensamble

- Árboles de decisión, Random Forest y Gradient Boosting.
- **Aplicación práctica:** comparación entre modelos.
- Dataset 1: Titanic - Kaggle
- Dataset 2: Bank Marketing - UCI

Eq4. Métodos de clasificación supervisada avanzados

- Support Vector Machines (SVM), K-Nearest Neighbors (KNN).
- **Aplicación práctica:** clasificación de imágenes.
- Dataset 1: MNIST Digits - Kaggle
- Dataset 2: Fashion MNIST - Zalando

Eq5. Métodos de aprendizaje no supervisado

- Clustering: K-means, DBSCAN, jerárquico.
- Reducción de dimensionalidad: PCA, t-SNE, UMAP.
- **Aplicación práctica:** segmentación y reducción de dimensionalidad.
- Dataset 1: Wholesale Customers - UCI
- Dataset 2: Mall Customers - Kaggle

Eq6. Redes neuronales artificiales y Deep Learning

- Perceptrón multicapa (MLP), funciones de activación y retropropagación.
- **Aplicación práctica:** red neuronal en Keras o PyTorch.
- Dataset 1: MNIST Original - Yann LeCun
- Dataset 2: Digits - Scikit-learn

Eq7. Introducción práctica a Deep Learning con PyTorch

- Tensores y operaciones básicas.
- Construcción y entrenamiento de un modelo sencillo.
- **Aplicación práctica:** clasificador de imágenes en PyTorch.
- Dataset 1: CIFAR-10 - Canadian Institute for AI
- Dataset 2: Dogs vs Cats - Kaggle

Eq8. Evaluación y validación de modelos de Machine Learning

- Overfitting, underfitting, validación cruzada.
- Métricas de clasificación y regresión.
- **Aplicación práctica:** evaluación de modelos.
- Dataset 1: Telco Customer Churn - Kaggle
- Dataset 2: Credit Card Default - UCI

Eq9. Interpretabilidad y explicabilidad de modelos de Machine Learning

- Importancia de la interpretabilidad (confianza, ética, regulaciones).
- Métodos globales: importancia de variables, PDP.
- Métodos locales: LIME, SHAP.
- **Aplicación práctica:** interpretación de un modelo.
- Dataset 1: Loan Prediction - Kaggle
- Dataset 2: Adult Income - UCI