**A logo of a company

Description automatically generatedA logo of a university

AI-generated content may be incorrect.UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN**

**Facultad de Ingeniería Mecánica y Eléctrica**

**Inteligencia Artificial y Redes Neuronales**

**Docente:** Daniel Isaías López Páez

**Grupo:** 013

**Act 05:** Procesamiento de conjuntos de datos en Python

**Integrantes:**

|  |  |
| --- | --- |
| Nombre | Matricula |
| Abraham Giovanni Aguirre Bravo | 2014222 |
| Karla Judith Corona Castro | 1923390 |
| Ramiro Emiliano Martínez De La Cruz | 2014197 |
| María Fernanda Garza Barbosa | 1945003 |

Ciudad Universitaria, San Nicolás de los Garza, a 01 de abril del 2025

***PREGUNTAS***:

1. ¿Cómo afecta la limpieza de datos al análisis del dataset?

La limpieza de datos es fundamental para asegurar la calidad y confiabilidad del análisis del dataset. En Python, se suele hacer con bibliotecas como **pandas** y **numpy**, eliminando valores nulos, duplicados y corrigiendo inconsistencias en los datos.

* **Datos limpios**: Generan análisis precisos y modelos de Machine Learning más confiables.
* **Datos sucios**: Introducen ruido, errores y sesgos que pueden llevar a conclusiones incorrectas.

1. ¿Por qué es importante la normalización en Machine Learning?

La normalización es crucial cuando las variables tienen diferentes escalas, ya que evita que los valores más grandes dominen el entrenamiento del modelo. Se usa comúnmente en algoritmos como redes neuronales y KNN.

1. ¿Cómo interpretarías los datos visualizados?

La interpretación de datos visualizados depende del tipo de gráfico utilizado. Al interpretar los datos, consideramos:

* **Distribución**: Si es normal, sesgada o multimodal.
* **Tendencias**: Aumento o disminución en series temporales.
* **Outliers**: Valores atípicos que podrían influir en los resultados.
* **Correlaciones**: Relaciones entre variables en gráficos de dispersión.

**CONCLUSIÓN GRUPAL:**

La limpieza de datos garantiza que el análisis sea confiable al eliminar valores erróneos, nulos o duplicados, lo que mejora la precisión de los resultados. La normalización es fundamental en Machine Learning, ya que equilibra las escalas de las variables y evita que algunas dominen el modelo, mejorando su desempeño. Finalmente, la visualización de datos facilita la identificación de patrones, tendencias y outliers, lo que ayuda a extraer conclusiones significativas. Un buen preprocesamiento de datos es esencial para obtener modelos más precisos y análisis más efectivos.

Link :