



Misión 1

**INTELIGENCIA
ARTIFICIAL**



Explorador



Tema 5: Visualización de datos



Campista, llegó el momento de retar tus conocimientos y que los pongas a prueba a través de los diferentes recursos que encontraras en este espacio como son: conceptos, ejemplos, herramientas, actividades prácticas y retos, los cuales te ayudaran alcanzar los objetivos trazados en el nivel explorador.

Visualización de Datos en el Ciclo de Vida de una Aplicación de Aprendizaje de Máquina

La visualización de datos es una herramienta esencial para entender, analizar y comunicar información compleja de manera efectiva. A través de gráficos y representaciones visuales, se pueden identificar patrones, tendencias y anomalías en los datos que podrían no ser evidentes de otra manera. A continuación, se describen algunos tipos comunes de gráficos utilizados en la visualización de datos, junto con sus aplicaciones y ejemplos.

Tipos de Gráficos

- **Histogramas:** Los histogramas son gráficos que muestran la distribución de un conjunto de datos continuos. Dividen los datos en intervalos o "bins" y muestran la frecuencia de los datos en cada intervalo. Los histogramas son útiles para entender la distribución y la forma de los datos, así como para identificar patrones como la normalidad o la presencia de sesgos.
 - **Aplicaciones**
 - Análisis de la distribución de consumos energéticos en diferentes regiones.
 - Evaluación de la distribución de rendimientos de cultivos en distintas temporadas.

Ejemplo:

Un histograma puede ser utilizado para mostrar la distribución de la cantidad de energía solar generada diariamente en una planta solar. Esto puede ayudar a identificar los días con producción anómala o evaluar la variabilidad de la producción a lo largo del tiempo.

<https://youtu.be/qBigTkBLU6g>

- **Gráficos de cajas (Box plots):** Los gráficos de cajas, o box plots, son utilizados para mostrar la distribución de un conjunto de datos a través de sus cuartiles. Un gráfico de cajas muestra la mediana, los cuartiles, y los valores atípicos de un conjunto de datos. Son útiles para identificar la dispersión, la simetría y la presencia de valores atípicos.

- **Aplicaciones**

- Comparación de la producción de energía entre diferentes tipos de paneles solares.
- Evaluación de la variabilidad en los rendimientos de cultivos bajo diferentes prácticas agrícolas.

Ejemplo:

Un gráfico de cajas puede mostrar la distribución de la eficiencia energética de diferentes tipos de turbinas eólicas. Esto permite comparar la mediana de eficiencia, la dispersión de los datos y detectar turbinas que tienen un rendimiento significativamente diferente del resto.

<https://youtu.be/lefGOMFbI90>

- **Gráficos de Densidad:** Los gráficos de densidad, o gráficos de densidad de kernel (KDE), son utilizados para estimar la densidad de probabilidad de una variable continua. A diferencia de los histogramas, los gráficos de densidad suavizan los datos y proporcionan una visualización más fluida de la distribución de los datos.

- **Aplicaciones**

- Análisis de la densidad de consumo energético en diferentes horarios del día.
- Evaluación de la distribución de valores de humedad del suelo en diferentes regiones agrícolas.

Ejemplo:

Un gráfico de densidad puede mostrar la distribución de la demanda de energía en una comunidad durante un día típico. Esto ayuda a visualizar los picos y valles en el consumo de energía y puede ser útil para planificar la generación y distribución de energía.

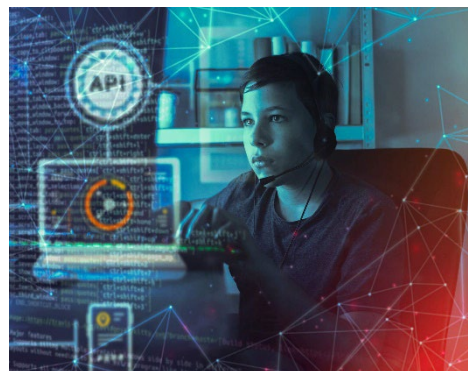
<https://youtu.be/88lpvX0YEso>

Aprendizajes Prácticos

Estos ejercicios permitirán a los estudiantes aplicar los conceptos aprendidos en situaciones prácticas y específicas.

- Práctica: Análisis Gráfico de Variables y su Impacto en el Modelo de Machine Learning.

Objetivo: El estudiante generará e interpretará histogramas, gráficos de cajas y gráficos de densidad para analizar las variables del conjunto de datos previamente elegido. El objetivo es extraer insights clave que informen y mejoren el desarrollo del modelo de machine learning, entendiendo cómo las características de los datos pueden influir en la solución del problema identificado.



TAREAS

1. GENERACIÓN DE GRÁFICOS

- **Objetivo:** Visualizar la distribución de las variables numéricas del conjunto de datos para detectar patrones, sesgos y outliers.
- **Tareas:**
 - **Generar un histograma para cada variable numérica.**
 - **Ejemplo:** En un conjunto de datos sobre la adopción de energía solar en comunidades rurales, se genera un histograma para la variable "Costo de

instalación por kW". El histograma muestra una distribución sesgada hacia la derecha, indicando que la mayoría de las instalaciones tienen un costo relativamente bajo, pero hay algunas con costos significativamente altos, probablemente en áreas remotas.

- **Crear un gráfico de cajas (boxplot) para cada variable numérica.**
 - *Ejemplo:* Se crea un gráfico de cajas para la variable "Consumo energético anual" en un estudio sobre comunidades que utilizan energía eólica. El gráfico revela la presencia de varios outliers, que corresponden a hogares con un consumo energético extremadamente alto, posiblemente debido a la falta de eficiencia energética en sus instalaciones.
- **Desarrollar un gráfico de densidad para cada variable numérica.**
 - *Ejemplo:* Se desarrolla un gráfico de densidad para la variable "Proporción de energía renovable utilizada" en un conjunto de datos sobre la transición energética en zonas urbanas. El gráfico muestra que la mayoría de las zonas tienen una alta proporción de energía renovable, pero con una segunda densidad menor en áreas donde la proporción es significativamente baja, lo que sugiere desigualdades en la transición energética.

2. SELECCIÓN E INTERPRETACIÓN DE GRÁFICOS

- *Objetivo:* Identificar y analizar las características clave de al menos cinco variables a partir de sus gráficos.
- *Tareas:*
 - **Seleccionar al menos cinco gráficos para su análisis (pueden ser histogramas, gráficos de cajas o gráficos de densidad).**
 - *Ejemplo:* Se seleccionan gráficos de densidad para las variables "Proporción de energía renovable utilizada" y "Emisiones de CO2 por

hogar" en un conjunto de datos sobre comunidades energéticas. Se observa que las áreas con baja proporción de energía renovable tienden a tener emisiones de CO2 más altas, lo que podría sugerir una relación directa entre la adopción de energías limpias y la reducción de emisiones.

- **Interpretar cada gráfico seleccionado, abordando aspectos:**

- *Ejemplo:* Al interpretar el gráfico de cajas para "Costo de instalación por kW" en la adopción de energía solar, se observa que los outliers de alto costo podrían ser áreas rurales con menos acceso a proveedores, lo que aumenta los costos de instalación. Este hallazgo es relevante para el modelo, ya que estos outliers podrían influir en las decisiones de subsidios o apoyos financieros para fomentar la adopción.
- *Ejemplo:* La densidad de la variable "Consumo energético anual" muestra que los hogares con un alto consumo podrían estar utilizando tecnologías menos eficientes. Este insight sugiere que el modelo podría incluir variables que evalúen la eficiencia energética para mejorar las predicciones sobre el impacto de la adopción de energías renovables.

3. CONCLUSIONES Y RELEVANCIA PARA EL MODELO

- *Objetivo:* Derivar conclusiones de los gráficos interpretados y evaluar su impacto en el desarrollo del modelo de machine learning.
- *Tareas:*
 - **Extraer conclusiones sobre la naturaleza de los datos con base en los gráficos analizados.**
 - *Ejemplo:* Se concluye que la variabilidad en los costos de instalación de energía solar puede ser un factor clave para modelar la adopción de tecnologías limpias en comunidades rurales, lo que sugiere la necesidad de políticas diferenciadas para diferentes regiones.

- **Argumentar cómo los insights obtenidos son relevantes para el desarrollo del modelo.**
 - *Ejemplo:* La identificación de áreas con baja proporción de energía renovable y altas emisiones de CO2 es crucial para desarrollar un modelo que pueda predecir qué comunidades necesitan mayor apoyo en la transición hacia energías limpias con un consumo energético extremadamente alto, posiblemente debido a la falta de eficiencia energética en sus instalaciones.
- **Explicar cómo estos insights podrían influir en la solución del problema, considerando su impacto potencial en el rendimiento del modelo de machine learning.**
 - *Ejemplo:* Los insights obtenidos sobre el alto costo de instalación en áreas rurales sugieren que el modelo podría beneficiarse de incluir variables geográficas para mejorar la precisión en la predicción de la adopción de energía solar, lo que a su vez podría guiar la asignación de recursos y subsidios en la transición energética.

4. DOCUMENTACIÓN

- **Objetivo:** Presentar de manera clara y estructurada los resultados y conclusiones del análisis gráfico.
- **Tareas:**
 - **Redactar un informe que incluya:**
 - Los gráficos seleccionados
 - La interpretación de cada gráfico
 - Las conclusiones y su relevancia para el modelo de Machine Learning
 - **Asegurarse de que el informe sea conciso, claro y visualmente efectivo, utilizando gráficos y explicaciones que respalden los puntos principales.**

Material Complementario

Campista, en este espacio encontraras ayudas en diferentes formatos que pueden potenciar tu proceso de aprendizaje.

- **StatQuest: Histograms, Clearly Explained:**
<https://www.youtube.com/watch?v=qBigTkBLU6g>
- **¿Qué es y cómo funciona la INTELIGENCIA ARTIFICIAL?**
<https://www.youtube.com/watch?v=24Uz1mBksL4&t=25s>
- **Curvas de densidad | Khan Academy en Español**
<https://www.youtube.com/watch?v=Nx9NwPoTKL8>