



# ***“TIPOS Y CONJUNTOS DE DATOS EN MINERIA DE DATOS”***

UNIVERSIDAD TECNOLÓGICA DE SAN LUIS RIO COLORADO

JULIA GARCIA

EXTRACCION DE CONOCIMIENTO EN BASES DE DATOS

FIDEL OROZCO CASTRO

IDGS9-1

9 DE OCTUBRE DEL 2024

## MODELO

Representación matemática o computacional de un proceso. Se ajusta para aprender patrones y relaciones entre un sistema o conjunto de datos. Para que pueda realizar predicciones, toma de decisiones o tareas en específico.

### TIPOS DE DATOS:

- **Datos estructurados:** Organizados mediante filas y columnas que representan atributo o característica específica.
- **Datos no estructurados:** No contienen una estructura fija y suelen incluir texto, imágenes, audio y video.
- **Datos semiestructurados:** Contienen cierta estructura, pero no se ajustan completamente a un formato tabular (JSON o XML)
- **Datos temporales:** Incluyen información registrada en función del tiempo (incluyen datos climáticos, datos financieros, registros de eventos en sistemas).
- **Datos transaccionales:** Representan transacciones individuales y se utilizan comúnmente en análisis de cestas de mercado y análisis de asociación (ventas, compras, transacciones bancarias).
- **Datos objetivo:** Se conocen como etiquetas o valores de respuesta, son la variable que se busca predecir o moldear en un problema de aprendizaje automático.
- **Datos de entrenamiento:** Son los utilizados para entrenar a los modelos de aprendizaje automático.
- **Datos de prueba:** Se utilizan para evaluar el rendimiento del modelo después de su entrenamiento.

## CONJUNTO DE DATOS

- **Conjunto de datos desequilibrados:** Una o varias clases o categorías tienen muchas más instancias que otras, creando un desequilibrio.
- **Conjunto de datos etiquetados y no etiquetados:** Tiene etiquetas o categorías asignadas previamente, lo que los hace adecuados para el aprendizaje supervisado. Los no etiquetados se utilizan en técnicas de aprendizaje no supervisado.
- **Conjunto de datos grandes:** Son extremadamente grandes y complejos que superan la capacidad de procesamiento y análisis de herramientas (software y hardware).