

IA y Datos: El Combustible de la Transformación Digital

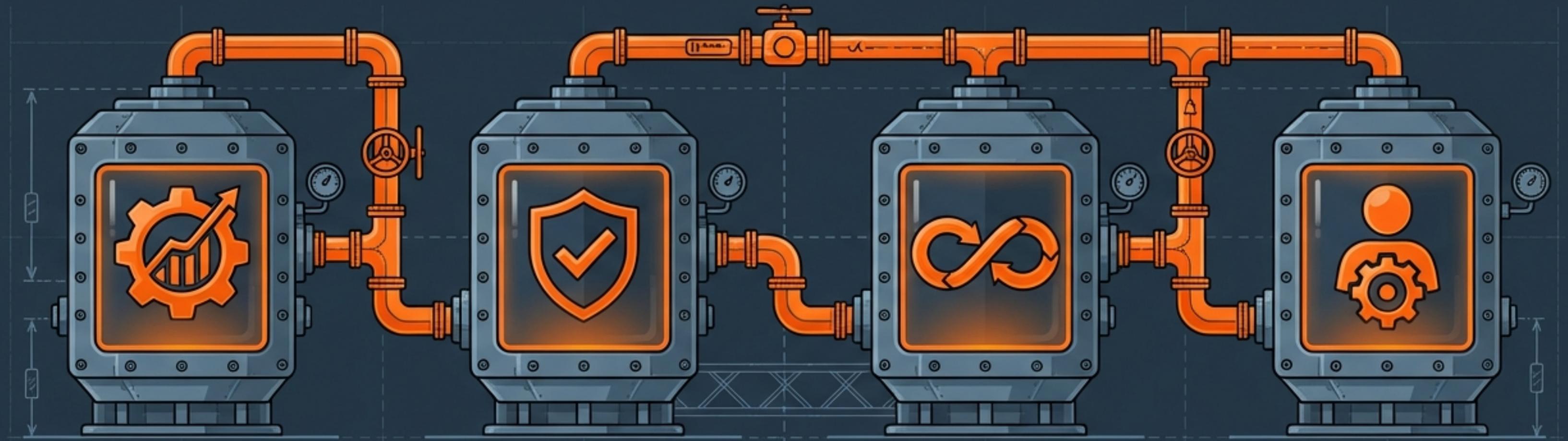
Digitalización aplicada a los sectores productivos - Tema 9

UNIDAD 9: INTELIGENCIA ARTIFICIAL Y TRATAMIENTO DE DATOS



Sin Datos de Calidad, No Hay Inteligencia

La inteligencia artificial (IA) ha transformado los sectores productivos, pero su eficacia depende absolutamente de la “materia prima” que consume. Los modelos requieren volúmenes masivos de información para operar con precisión.



Entrenamiento

Alimentar algoritmos (ML/DL) para aprender tareas específicas.

Validación y Prueba

Asegurar que el modelo funcione con datos nuevos y desconocidos.

Mejora Continua

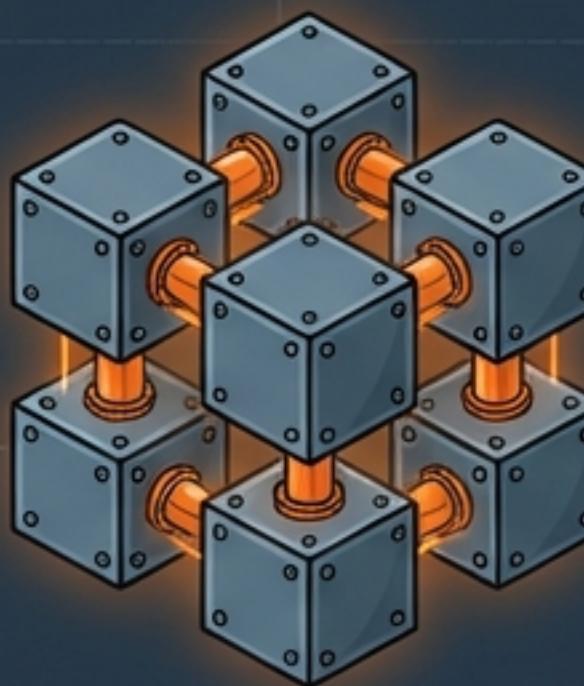
Adaptación constante a nuevas tendencias mediante datos recientes.

Personalización

Adaptar respuestas y recomendaciones al comportamiento del usuario.

Clasificación de la Materia Prima: Tipos de Datos

Estructurados



Formatos fijos
(filas/columnas).
Ej: Bases de datos SQL.

Semiestructurados



Etiquetas sin esquema rígido.
Ej: XML y JSON.

XML

```
<empinfo>
  <employees>
    <employee>
      <employse>
        <name>James Kirk</name>
        <ages>40</age>
      </employee>
    ...
  </employees>
</empinfo>
```

JSON

```
{
  "empinfo": {
    "employees": [
      {
        "name": "James Kirk",
        "age": 40,
      },
      ...
    ]
  }
}
```

No Estructurados



Sin formato predefinido.
Ej: Textos, imágenes,
vídeo, audio.

Otros Fuentes

- Datos de Sensores: IoT y electrónica.
- Datos de Red y Sociales: Registros de navegación e interacciones.

LOGÍSTICA DEL DATO: RECOLECCIÓN Y ALMACENAMIENTO

Fuentes



Base de Datos



IoT



Redes Sociales

Métodos
(Scraping, APIs)

Almacenamiento

Log Files

Analytics
Data



Videos &
Pictures

Application
Data



Backup &
Archival



Artificial
Intelligence (AI)

Advanced
Analytics

Machine
Learning (ML)

Analyze Data

Amazon S3

Almacenamiento de objetos para
cualquier cantidad de datos

Tecnologías de Almacenamiento

- **SQL (Relacional):** Tablas estructuradas, esquema rígido.
- **NoSQL (Flexible):** Documentos o clave-valor (MongoDB).
- **La Nube:** Soluciones escalables (AWS S3, Google Cloud).

El Escudo de Seguridad: Privacidad y Protección

La recolección de datos conlleva la responsabilidad de proteger la información sensible y cumplir con la ley.

- 1. Regulaciones:** Cumplimiento estricto del GDPR para derechos del usuario.
- 2. Anonimización vs. Pseudonimización:**
 - **Anonimización:** Eliminar irreversiblemente la info identificativa.
 - **Pseudonimización:** Reemplazar datos con pseudónimos (reversible).
- 3. Cifrado:** Protección criptográfica en tránsito y reposo.
- 4. Control de Acceso:** Políticas estrictas para personal autorizado.
- 5. Consentimiento:** Transparencia total sobre el uso de los datos.



LA REFINERÍA: LIMPIEZA Y PREPARACIÓN

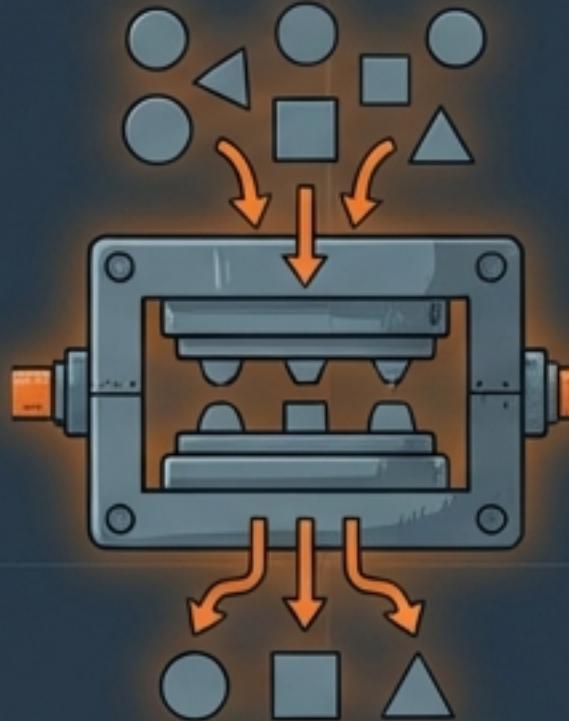
EL 90% DEL TIEMPO EN PROYECTOS DE IA
SE GASTA EN LA SELECCIÓN, LIMPIEZA,
ENRIQUECIMIENTO Y TRANSFORMACIÓN DE DATOS.

EL PROCESO DE LIMPIEZA



- Limpieza**
- Imputación:**
Rellenar valores faltantes.
- Eliminación:**
Borrar duplicados.
- Corrección:**
Arreglar errores tipográficos.

TÉCNICAS DE TRANSFORMACIÓN



- Transformación**
- Normalización:**
Ajustar a escala común.
- Estandarización:**
Media 0, desviación 1.
- Codificación (One-hot):**
Convertir texto categórico a numérico.

Optimización: Reducción y Transformación

Simplificar la complejidad del conjunto de datos para mejorar la velocidad y precisión.

Reducción de Dimensionalidad:

Disminuir el número de variables sin perder información clave.

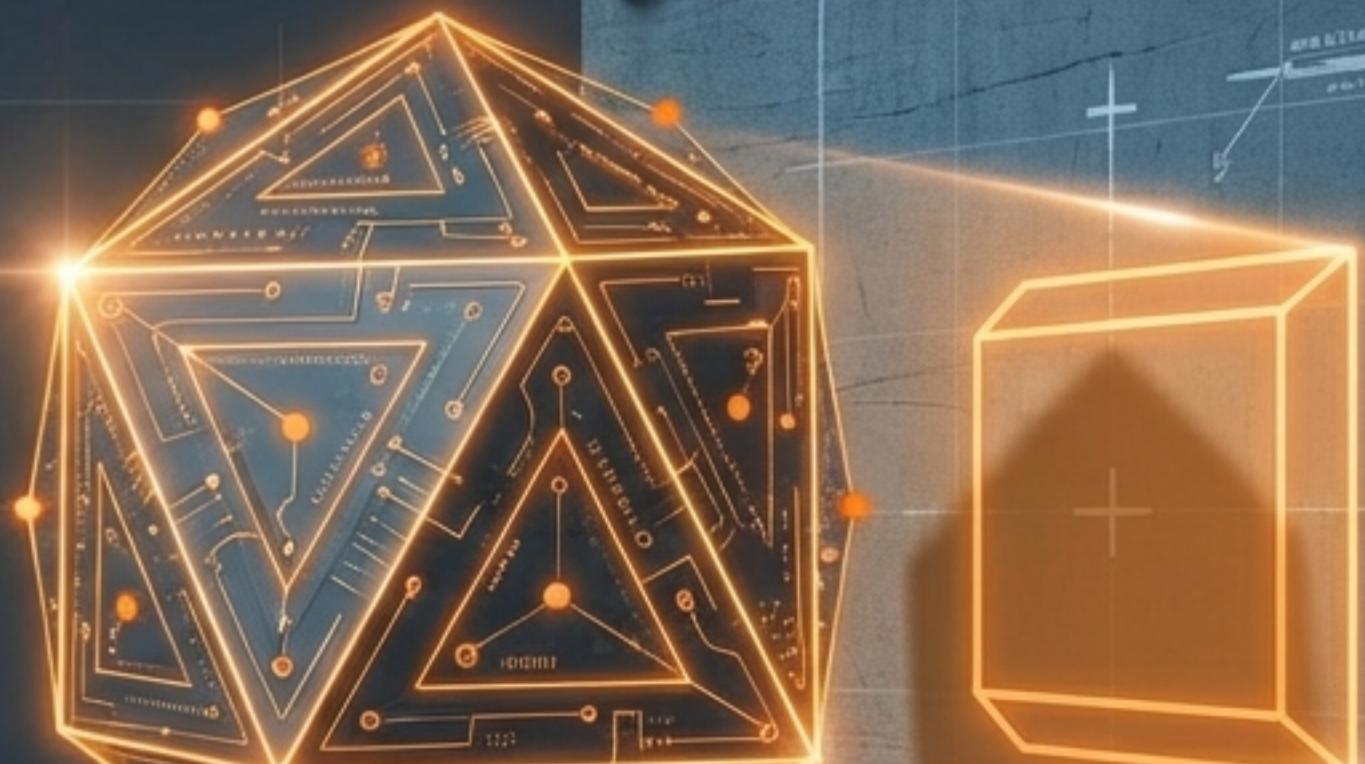
Técnicas Clave:

- **PCA (Análisis de Componentes Principales):**

Transformar datos a un espacio de menor dimensión capturando la variabilidad.

- **Análisis de Correlación:** Identificar y eliminar variables redundantes.

- **Agregación:** Resumir datos mediante medias, sumas o conteos.

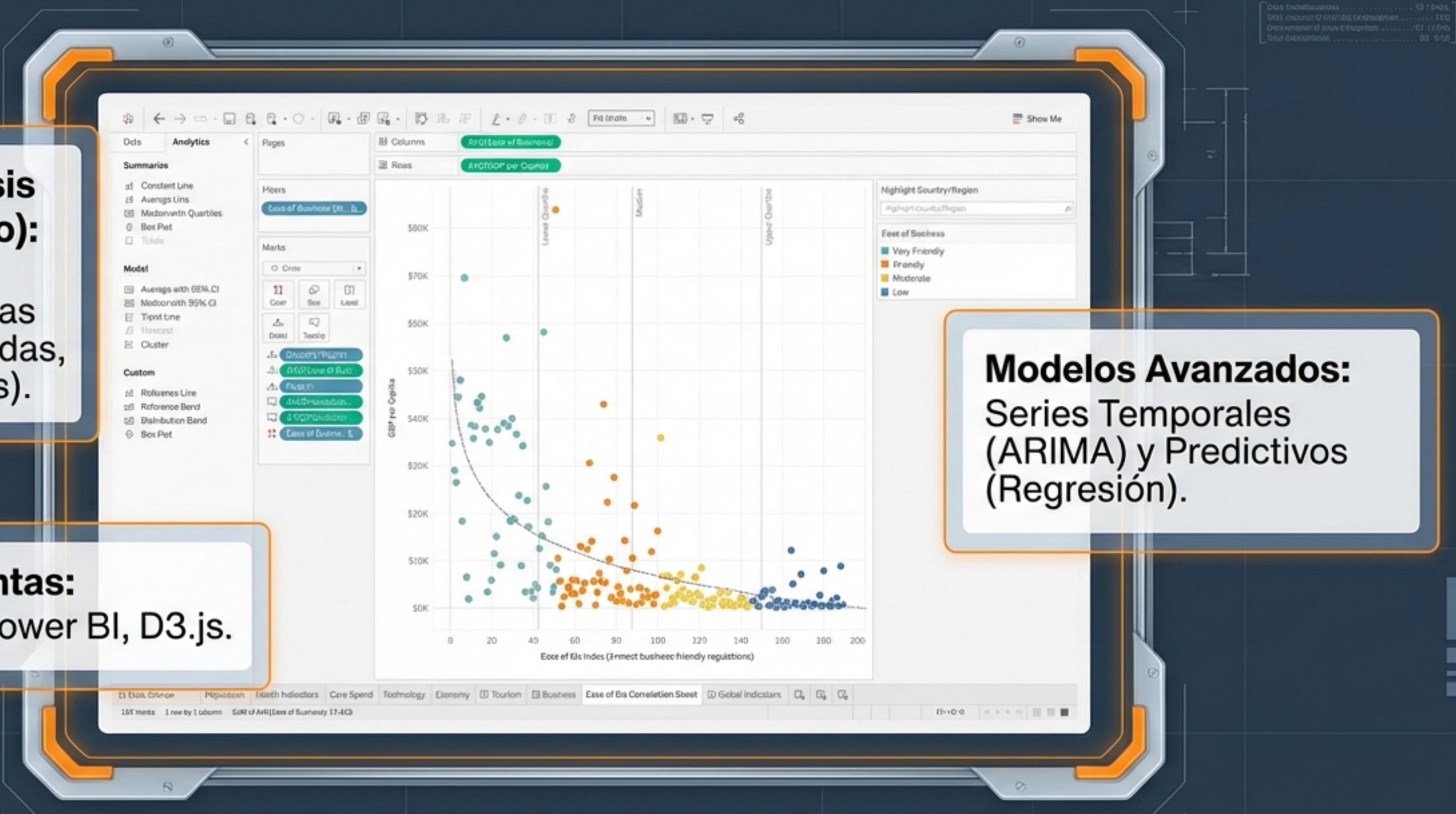


El Descubrimiento: Análisis y Visualización (EDA)

EDA (Análisis Exploratorio):
Entender características (medias, modas, desviaciones).

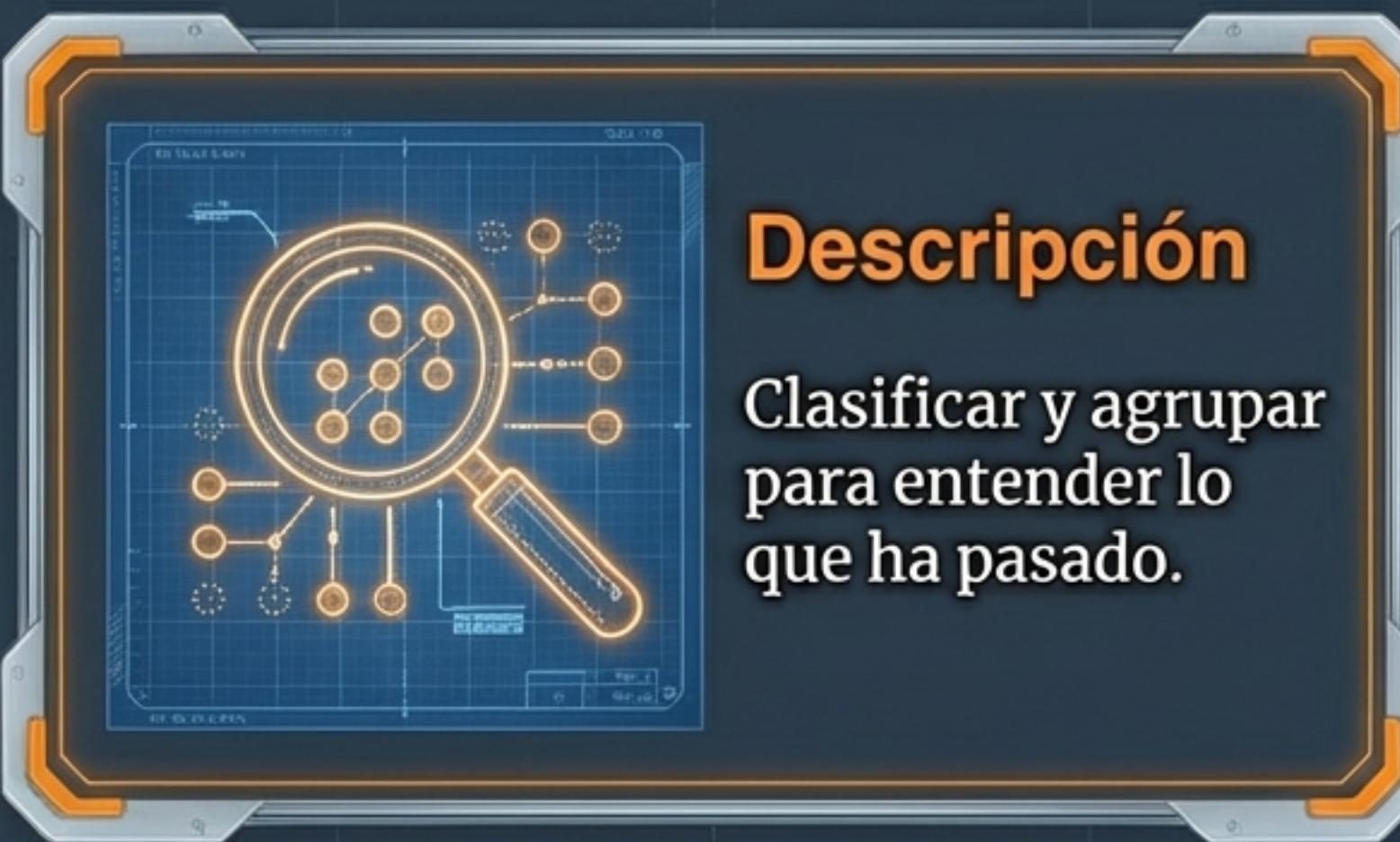
Herramientas:
Tableau, Power BI, D3.js.

Modelos Avanzados:
Series Temporales (ARIMA) y Predictivos (Regresión).



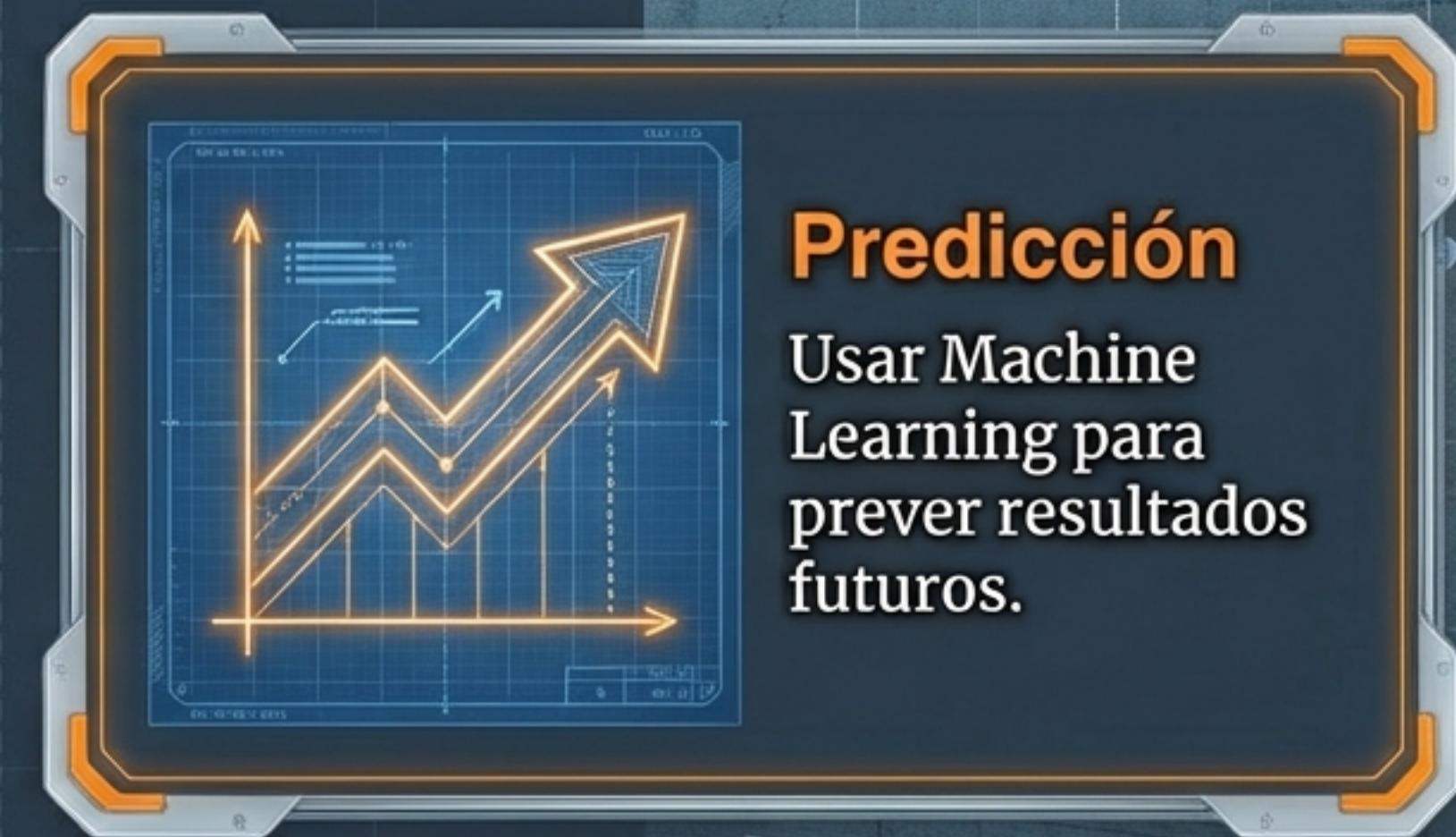
Minería de Datos (KDD): Encontrando Patrones

Knowledge Discovery in Databases (KDD): El proceso de descubrir patrones, correlaciones y tendencias en grandes volúmenes de datos.



Descripción

Clasificar y agrupar para entender lo que ha pasado.



Predictión

Usar Machine Learning para prever resultados futuros.

Patrones: Regularidades. | **Features:** Atributos analizados. | **Instancias:** Casos individuales.

El Ciclo de Vida del Conocimiento



Datos Brutos:
Fuente original.

Selección:
Subconjunto relevante.

Preproceso: Transformación:
Limpieza.
Formato adecuado.

Minería:
Extracción de patrones.

Interpretación:
Conocimiento accionable.

SCALE: 1:50

Caso Práctico: Alimentando al E-commerce

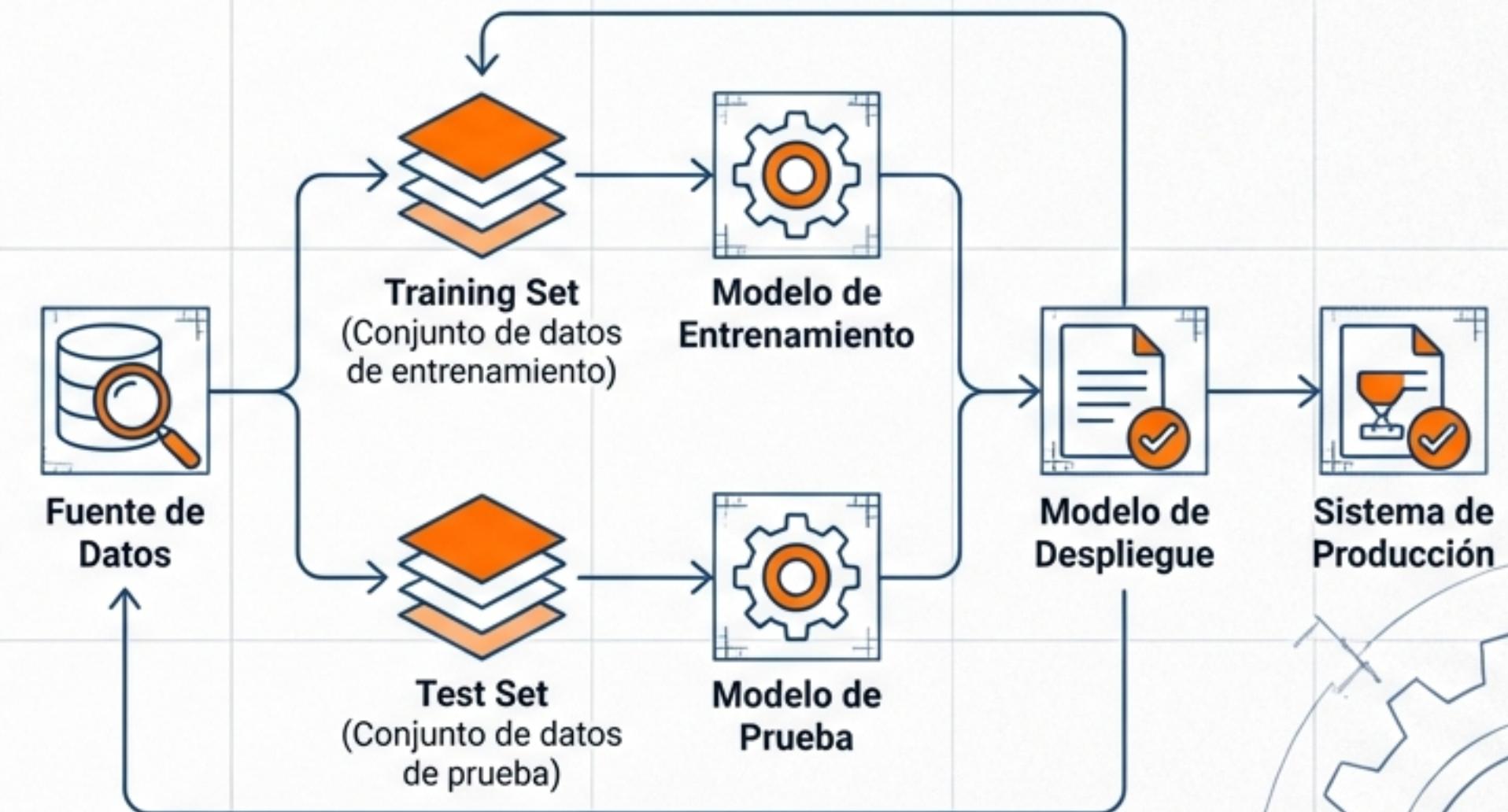
Reto: Personalización de la experiencia de usuario.

El Proceso:

- **Recopilación:** Historial, compras, reseñas.
- **Limpieza:** Eliminación de atípicos.
- **Entrenamiento:** Aprendizaje de relaciones ('quien compra X, mira Y').

Resultado:

- Recomendaciones precisas.
- +Ventas y Fidelidad.



Caso Práctico: Inteligencia en el Supermercado

Reto: Aumentar lealtad analizando el comportamiento físico.



Resolución Crítica: La IA No es Infalible

Sesgos en Entrenamiento:
Si los datos están sesgados, la respuesta también lo estará.

Falta de Contexto:
La IA procesa estadística, no 'comprende' el mundo.

Datos Desactualizados:
Los modelos tienen fechas de corte de conocimiento.

Alucinaciones:
Generación de respuestas lógicas pero falsas.



CONCLUSIÓN: La IA es una herramienta de apoyo. El juicio humano es el filtro final necesario.

Resumen Ejecutivo: Mapa del Ecosistema de Datos



Conclusión y Referencias

Los datos son el activo más valioso de la era moderna. El tratamiento de datos transforma información bruta en activos estratégicos. La calidad del dato define el techo de inteligencia de cualquier organización.

**Referencias Bibliográficas:

- Arntz, M., Gregory, T. y Zierahn, U. (2016). *The Risk of Automation...*
- Benjamins, R. (2022). *A data-driven company.*
- Nilsson, N. J. (2010). *The Quest for Artificial Intelligence.*
- Russell, S. y Norvig, P. (2021). *Artificial Intelligence: A Modern Approach.*
- OECD Social, Employment and Migration Working Papers.