



EXTRACCIÓN DE CONOCIMIEN TOS EN BASES DE DATOS

ING. LUIS ENRIQUE MASCOTE CANO



INVESTIGACIÓN DE
TÉCNICAS DE VISUALIZACIÓN
Lic. Ricardo
Hernández Martínez
Fecha de Entrega:
30/11/25

Introducción

La visualización y representación de información es un componente esencial dentro del proceso de extracción de conocimiento, ya que permite transformar datos complejos en representaciones comprensibles que facilitan la interpretación, el análisis y la toma de decisiones. Estas técnicas son fundamentales en el campo de la ciencia de datos, minería de datos y analítica, pues ayudan a revelar patrones, tendencias, relaciones y anomalías que no serían evidentes mediante tablas o números únicamente.

Este reporte presenta una investigación sobre las principales técnicas de visualización de información y su relación directa con el proceso de extracción del conocimiento.

2. Importancia de la Visualización en la Extracción de Conocimiento

En el proceso de KDD (Knowledge Discovery in Databases), la visualización cumple múltiples funciones:

- **Exploración de datos (EDA):** permite al analista comprender la estructura, distribución y comportamiento de los datos.
- **Detección de patrones:** facilita la identificación de correlaciones, grupos, tendencias o desviaciones.
- **Validación de modelos:** ayuda a evaluar el rendimiento de técnicas de clasificación, regresión y clustering.
- **Comunicación de resultados:** convierte hallazgos complejos en conocimientos accesibles para tomadores de decisiones.

La visualización, por tanto, no solo muestra información: **permite generar conocimiento a partir de los datos.**

3. Técnicas de Visualización de Datos

3.1. Gráficos Univariados

a) Histogramas

Representan la distribución de una variable numérica.

- **Uso:** ver dispersión, asimetría, valores extremos.
- **Aporte al conocimiento:** ayudan a determinar si se necesita normalización o transformación.

b) Gráficos de barras

Comparan categorías discretas.

- **Uso:** análisis de frecuencias.
- **Aporte:** permiten ver jerarquías y niveles.

3.2. Gráficos Bivariados

a) Diagramas de dispersión (Scatter Plots)

Representan la relación entre dos variables numéricas.

- **Aporte:** visualizan correlaciones, agrupamientos y outliers.

b) Mapas de calor (Heatmaps)

Usan colores para indicar intensidad, especialmente en matrices de correlación.

- **Aporte:** facilitan identificar relaciones fuertes entre variables.

3.3. Técnicas Avanzadas

a) Gráficos 3D

Permiten visualizar relaciones entre tres variables.

- **Aporte:** útiles en clustering o datos multivariados.

b) Diagramas de caja (Boxplots)

Muestran distribución, mediana y valores atípicos.

c) Visualización de clustering

Incluye:

- Gráficos de dispersión coloreados por cluster.
- Dendrogramas para clustering jerárquico.
- Mapas de densidad.

Aportan información clave sobre separabilidad, cohesión y estructura de los grupos.

d) Visualización tras reducción de dimensionalidad

Métodos como PCA y t-SNE permiten proyectar datos complejos a 2D o 3D.

- **Apporte:** facilitan identificar patrones que no son visibles en el espacio original.

4. Técnicas de Representación de Información

Además de gráficos, la representación incluye formas estructuradas y textuales que apoyan la extracción de conocimiento.

4.1. Tablas Resumen

Muestran estadísticas descriptivas como media, desviación estándar, mínimos y máximos.

- **Apporte:** dan contexto antes de aplicar modelos o visualizaciones.

4.2. Mapas conceptuales y diagramas

Representan relaciones entre conceptos.

- **Aporte:** sintetizan conocimiento derivado del análisis.

4.3. Dashboards

Integrar múltiples visualizaciones en un sistema interactivo.

- **Aporte:** permiten el análisis dinámico y en tiempo real.

4.4. Informes Narrativos

Combina texto, gráficos y tablas para comunicar hallazgos.

- **Aporte:** hacen accesible el conocimiento a personal no técnico.

5. Relación entre Visualización y Extracción de Conocimiento

La visualización está presente en todas las etapas del proceso KDD:

- **Preprocesamiento:** identificación de datos faltantes, outliers y comportamientos anómalos.
- **Selección y transformación:** análisis visual de correlaciones y variabilidad.
- **Minería de datos:** validación gráfica de modelos.
- **Interpretación y evaluación:** uso de gráficos para comunicar los resultados.

La visualización es, por tanto, una herramienta clave para **convertir datos en decisiones informadas**.

6. Conclusiones

La visualización y representación de información son componentes esenciales en la extracción de conocimiento porque permiten comprender datos complejos de manera intuitiva, identificar patrones relevantes y comunicar hallazgos de forma efectiva. Sin estas técnicas, el análisis sería más lento, menos preciso y menos accesible para los distintos perfiles dentro de una organización.

El uso adecuado de gráficos, diagramas, dashboards y representaciones textuales favorece un proceso de análisis más profundo y con mayor valor agregado.

7. Referencias

- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press.
- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
- Ware, C. (2021). *Information Visualization: Perception for Design*. Morgan Kaufmann.