

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Desarrollo de software



Extracción de Conocimiento en Bases de Datos

V.1. Reporte de investigación de técnicas de visualización (50%)

IDGS91N

PRESENTA:

Juan Carlos Medina Sánchez

DOCENTE:

Enrique Mazcote

Chihuahua, Chih., 29 de noviembre de 2025

1. Introducción

La visualización de información es una etapa fundamental dentro del proceso de extracción del conocimiento, ya que permite transformar grandes volúmenes de datos en representaciones gráficas fáciles de interpretar. A través de técnicas adecuadas, se facilita el descubrimiento de patrones, tendencias, comportamientos y anomalías que no serían evidentes mediante tablas o datos sin procesar.

En el campo de la analítica, ciencia de datos y minería de información, la visualización actúa como un puente entre el análisis técnico y la comprensión humana. Este reporte presenta las principales técnicas de visualización y representación de información, su utilidad y su relación directa con la extracción de conocimiento.

2. ¿Qué es la visualización de información?

La visualización de información es el proceso de representar datos de manera gráfica con el fin de facilitar su interpretación. Permite que los usuarios comprendan relaciones, identifiquen patrones, detecten problemas y tomen decisiones basadas en datos.

Algunas características clave:

- Hace visibles patrones complejos.
- Facilita análisis exploratorio de datos (EDA).
- Permite comunicar resultados de manera clara.
- Es fundamental para usuarios no técnicos.

3. Relación entre visualización y extracción del conocimiento

La extracción del conocimiento implica varias etapas: recopilación, limpieza, transformación, modelado y representación de resultados. La visualización participa en dos momentos esenciales:

3.1. Visualización para análisis exploratorio (EDA)

Se utiliza para entender cómo están distribuidos los datos antes del modelado:

- Distribución de variables
- Correlaciones
- Valores atípicos
- Patrones temporales

Favorece decisiones sobre:

- Selección de variables
- Selección de algoritmo
- Preprocesamiento

3.2. Visualización para la comunicación de resultados

Después del modelado, la visualización permite:

- Interpretar métricas
- Presentar predicciones
- Mostrar clústeres
- Comparar modelos
- Explicar decisiones basadas en datos

Por ello, la visualización es esencial tanto para la comprensión del problema como para la toma de decisiones basadas en conocimiento.

4. Técnicas de visualización y representación de información

4.1.1. Histogramas

Representan la distribución de una variable continua.

Útiles para:

- Detectar sesgo
- Identificar multimodalidad
- Analizar dispersión

4.1.2. Diagramas de barras

Representan valores categóricos.

Útiles para:

- Comparar grupos
- Identificar frecuencias
- Analizar conteos

4.1.3. Diagramas de líneas

Permiten analizar series de tiempo.

Aplicaciones:

- Tendencias
- Estacionalidad
- Cambios abruptos

4.2. Visualización multivariable

4.2.1. Mapas de calor (Heatmaps)

Representan matrices de datos mediante una escala de colores.

Aplicaciones:

- Matrices de correlación
- Intensidad de actividad
- Señales biométricas

4.2.2. Gráficos de dispersión (Scatter plots)

Permiten observar relaciones entre dos variables.

Son útiles para:

- Detectar correlación
- Identificar outliers
- Visualizar clústeres

4.2.3. Gráficos de burbujas

Extensión del scatter plot usando una tercera variable como tamaño.

4.3. Visualización avanzada y análisis de patrones

4.3.1. Mapas de clústeres (Cluster Plots)

Usados en modelado no supervisado (K-means, DBSCAN).

Permiten ver cómo se agrupan los datos.

4.3.2. PCA Biplot

Representa los primeros componentes de PCA.

Sirve para:

- Visualizar alta dimensionalidad en 2D
- Identificar relaciones e influencias de variables

4.3.3. t-SNE

Técnica avanzada para visualizar datos complejos en 2D o 3D.
Especialmente útil para:

- Imágenes
- Texto
- Datos con patrones no lineales

4.4. Técnicas de representación para reportes ejecutivos

4.4.1. Dashboards

Plataformas como Power BI o Tableau permiten:

- Indicadores clave (KPIs)
- Gráficas interactivas
- Filtros dinámicos
- Información para toma de decisiones

4.4.2. Infografías

Combinan imágenes, texto y datos.

Se utilizan para transmitir conclusiones de forma rápida y visualmente atractiva.

5. Importancia de la visualización en la generación de conocimiento

La visualización contribuye directamente al proceso de Knowledge Discovery (KDD) porque:

- **Reduce la complejidad** de los datos.
- **Facilita la interpretación humana.**
- **Permite validar hipótesis** formadas durante el análisis.
- **Aumenta la precisión del modelo** al revelar problemas o patrones ocultos.
- **Comunica resultados** a personal no técnico de manera clara.

Sin estos elementos, sería difícil transformar datos crudos en conocimiento útil.

6. Conclusiones

La visualización y representación de la información son elementos esenciales dentro del análisis de datos y el proceso de extracción del conocimiento. Permiten comprender, explorar y comunicar información que de otra manera pasaría

desapercibida. Las técnicas de visualización van desde gráficos básicos hasta métodos avanzados como PCA y t-SNE, cada una con aplicaciones particulares dentro del análisis de datos.

Comprender estas técnicas no solo mejora el análisis exploratorio y la calidad del modelado, sino que fortalece la capacidad de presentar resultados de manera profesional, efectiva y orientada a la toma de decisiones.

7. Referencias (formato APA)

- Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.
- Knafllic, C. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.
- Ware, C. (2013). *Information Visualization: Perception for Design*. Morgan Kaufmann.