



Universidad
Nacional Autónoma
de México



INSTITUTO DE
INVESTIGACIONES
EN MATEMÁTICAS
APLICADAS Y
EN SISTEMAS

Licenciatura en Ciencia de Datos

Minería de Datos

Tarea 3

Resumen BD en KDD

Integrante:

- Villalón Pineda Luis Enrique

¿Qué es y para qué sirve una BD de Conocimientos en KDD?

Una Base de Conocimientos es como una biblioteca digital donde guardamos todo lo valioso que descubrimos al analizar grandes cantidades de datos.

Imagínate que tienes millones de registros de ventas, comportamiento de clientes o datos médicos. Después de procesarlos y encontrar patrones importantes - como "los clientes que compran café también tienden a comprar galletas" o "ciertos síntomas aparecen juntos frecuentemente" - necesitas un lugar donde almacenar estos hallazgos.

Aquí es donde entra la Base de Conocimientos: guarda estos descubrimientos de forma organizada para que puedas:

- Reutilizar lo aprendido sin tener que hacer todo el análisis otra vez
- Entender fácilmente los patrones encontrados
- Tomar mejores decisiones basadas en información real y procesada

Es básicamente tu "memoria institucional" de todo el conocimiento extraído de los datos, listo para usar cuando lo necesites en nuevos proyectos o decisiones importantes.

¿Cómo está estructurada?

Una Base de Conocimientos en KDD está organizada en cinco secciones principales:

1. La "ficha técnica" (Metadatos): Aquí se guarda la información básica: de dónde vinieron los datos, cuándo se procesaron, qué tan confiables son, etc. Es como tener la "historia clínica" de tus datos.
2. Los "tesoros encontrados" (Resultados): Todo lo valioso que descubriste: patrones de comportamiento, reglas tipo "si pasa A, entonces probablemente pase B", modelos que pueden predecir resultados, grupos de clientes similares, etc.
3. Las "calificaciones" (Medidas de evaluación): Números que te dicen qué tan buenos son tus descubrimientos. Como cuando un profesor te pone una nota: "este patrón es 85% confiable" o "este modelo acierta 9 de cada 10 veces".
4. El "manual de uso" (Reglas y heurísticas): Instrucciones sobre cómo interpretar y aplicar lo que encontraste. Te explica qué significa cada resultado y cómo usarlo en la práctica.
5. El sistema de "actualización" (Módulos de actualización): Permite agregar nuevos descubrimientos sin borrar los anteriores, manteniendo todo organizado y actualizado.

Algo mas representativo o forma de verlo en un esquema es de la siguiente manera:

1. Metadatos

- Fuente de los datos
- Fecha de recolección
- Calidad y formato

2. Resultados del análisis

- Reglas de asociación
- Patrones frecuentes
- Modelos predictivos o descriptivos

3. Medidas de evaluación

- Precisión
- Confianza
- Soporte
- Error

4. Reglas y heurísticas

- Condiciones para aplicar el conocimiento
- Casos de uso prácticos

5. Módulos de actualización

- Nuevos descubrimientos
- Versionado de modelos o reglas

En pocas palabras es como tener un asistente que no solo guarda tus hallazgos, sino que también te dice qué tan buenos son y cómo usarlos correctamente.

Ejemplo

Supongamos que estamos en una empresa que analiza las compras de un supermercado, por lo que tenemos lo siguiente:

- Metadatos:
 - Fuente: Base de datos de ventas 2024
 - Formato: CSV

- Calidad: 98% de registros completos
- Resultados:
 - Regla descubierta: *“Si un cliente compra pan y mantequilla, hay un 70% de probabilidad de que también compre mermelada”*.
- Medidas de evaluación:
 - Soporte: 15%
 - Confianza: 70%
 - Lift: 1.5
- Reglas/heurísticas:
 - Usar la regla para promociones conjuntas.
- Actualización:
 - Cada mes se incorporan las nuevas ventas para refinar las reglas.