

# Proceso KDD

Integrantes: Benjamin Urbina, Josue Bustos, Cristhian Montenegro, Brandon Muñoz



**UNIVERSIDAD**  
**DE LA SERENA**  
CHILE

## ¿Qué es el proceso KDD?

El proceso KDD (Knowledge Discovery in Databases) es una metodología utilizada para extraer conocimiento útil y valioso a partir de grandes volúmenes de datos. Este proceso es clave en áreas como la minería de datos, ciencia de datos y análisis de grandes bases de datos. Este cuenta con una serie de pasos los cuales hablaremos a continuación.

Este proceso surgió como una respuesta a la creciente necesidad de extraer información valiosa de las enormes cantidades de datos que comenzaron a acumularse en las bases de datos a partir de la década de 1980. Con el avance de las tecnologías de la información y la creación de sistemas más complejos de almacenamiento de datos, se hizo evidente que los métodos tradicionales para analizar la información no eran suficientes para encontrar patrones útiles en estos grandes volúmenes de datos.

## Pasos del proceso KDD

**Recolección de datos:** Este es el primer paso, donde se recopilan los datos que provienen de diferentes fuentes, como bases de datos, archivos o incluso sensores. Aquí se asegura tener toda la información necesaria para el análisis.

**Procesamiento:** Una vez que los datos están reunidos, es necesario limpiarlos y prepararlos. Esto significa corregir errores, eliminar duplicados, manejar datos faltantes y normalizar la información. El objetivo es dejar los datos en un estado óptimo para su posterior análisis, evitando resultados incorrectos o sesgados.

**Transformación de datos:** En esta fase, los datos se modifican o transforman para que puedan ser analizados más fácilmente. A menudo, esto implica reducir el tamaño del conjunto de datos o convertirlos en un formato más adecuado para las técnicas de minería de datos. Por ejemplo, puedes agrupar datos, crear nuevas variables o seleccionar sólo los atributos más relevantes.

**Minería de datos:** Este es el corazón del proceso KDD. Aquí se aplican algoritmos y técnicas de análisis avanzadas para descubrir patrones, relaciones y tendencias ocultas en los datos. Dependiendo del objetivo, puedes usar técnicas como clasificación, agrupación (clustering), reglas de asociación o regresión. El tipo de técnica depende de la naturaleza del problema que se esté resolviendo.

**Interpretación y evaluación:** Finalmente, los resultados obtenidos de la minería de datos se interpretan y se evalúan. Es crucial comprender si los patrones descubiertos son útiles, significativos y aplicables en la práctica. Este paso incluye presentar los hallazgos de manera clara y sencilla, ayudando a la toma de decisiones o a mejorar procesos.