

**DATA MINING KDD.** Actualmente se generan grandes volúmenes de información en redes sociales, bancos, tiendas departamentales, cines, etc.

**¿Qué es la minería de datos? (Data Mining).** La minería de datos en realidad es el núcleo de todo un proceso llamado Descubrimiento de Conocimiento en Base de Datos, es un proceso metodológico para encontrar un “modelo” válido, útil y entendible que describa patrones de acuerdo a la información proporcionada. El KDD es un proceso metodológico y secuencial que se sigue para encontrar conocimiento en un conjunto de datos en bruto. Este conjunto de datos es una colección de información ya sea cuantitativa o cualitativa y está compuesto por variables o atributos (columnas) y caos (filas).

- **Fases del KDD**

**Abstracción del escenario.** Se trata de entender la problemática a la que nos vamos a enfrentar y tener contexto para proponer soluciones viables y reales.

**Selección de datos.** Debemos de elegir datos disponibles para realizar el estudio e integrarlos en uno solo que puedan favorecer a llegar a alcanzar los objetivos.

**Limpieza y preprocesamiento.** Esto implica eliminar variables o atributos con datos faltantes o eliminar información no útil para este tipo de tareas.

**Transformación de los datos.** Se mejora la calidad de los datos con transformaciones que involucran ya sea reducción de dimensiones o transformaciones.

**Selección de la apropiada tarea de minería de datos.** Se refiere a elegir el paradigma apropiado de Minería de Datos, ya sea la clasificación, regresión o agrupación, según los objetivos que se haya planteado para la investigación (como puede ser predicción o descripción).

**Elección del algoritmo de minería de datos.** Seleccionar la técnica o algoritmo, o incluso más de uno, para la búsqueda del patrón y obtener conocimiento.

**Aplicación del algoritmo.** Es posible que se ejecuten varias veces los algoritmos intentando ajustar los parámetros que optimicen los resultados.

**Evaluación.** Una vez aplicados los algoritmos al conjunto de datos, se evalúa los patrones que se generaron y el rendimiento que se obtuvo, esto para verificar que cumpla con las metas planteadas en las primeras fases.

**Aplicación.** Si los resultados de la evaluación se satisfacen, la última etapa es simplemente aplicar el conocimiento encontrado al contexto y comenzar a resolver sus problemáticas.