1. **Selección de datos.**En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar. Es la etapa donde los datos relevantes para el análisis son extraídos desde la o las fuentes de datos.
2. **Preprocesamiento.**Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos en una forma manejable, necesaria para las fases posteriores. En esta etapa se utilizan diversas estrategias para manejar datos faltantes o en blanco, datos inconsistentes o que están fuera de rango, obteniéndose al final una estructura de datos adecuada para su posterior transformación.
3. **Transformación.**Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada.  Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente.
4. **Data Mining.** Es la fase de modelamiento propiamente tal, en donde métodos inteligentes son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que están contenidos u “ocultos” en los datos.
5. **Interpretación y Evaluación.** Se identifican los patrones obtenidos y que son realmente interesantes, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos.

Elementos de la minería de datos

La **minería de datos** comprende cinco elementos principales:

**1. Procesos ETL:** extraer, transformar y cargar datos transaccionales en los sistemas de almacenamiento de datos.

**2. Almacén de datos:** recopilar y gestionar los datos en un sistema de base de datos multidimensional.

**3. Acceso a los datos:** proporcionar acceso a los datos a los analistas de negocio, profesionales de tecnología deinformación y usuarios autorizados.

**4. Análisis de datos:** analizar los datos de todas las aplicaciones.

**5. Presentación**: de los datos en un formato útil, como un gráfico o una tabla, que facilite una buena interpretación de su contenido.

El análisis es una parte sustancial de la **minería de datos** pero no se lleva a cabo siempre de la misma forma. De hecho, existen diferentes niveles de análisis disponibles entre los que destacan las **visualizaciones de datos**, que se basan en modelos gráficos; las **redes neuronales artificiales**, que aprenden a través de la formación; los**árboles de decisión,** que generan reglas para la clasificación de conjuntos de datos; o los **algoritmos genéticos,** que no son otra cosa que técnicas de optimización que utilizan procesos como combinación genética, mutación y la selección natural en un diseño basado en los conceptos de evolución natural.