**Metodología KDD (Knoledge Discovery in Databases).**

<http://sedici.unlp.edu.ar/bitstream/handle/10915/21220/Documento_completo.pdf?sequence=1&isAllowed=y>

<https://www.researchgate.net/profile/Vasudha_Bhatnagar/publication/2403357_A_Proposal_for_Data_Mining_Management_System/links/570a826f08aea660813722e3.pdf>

La KDDo descubrimiento de conocimiento en bases de datos, es un proceso tecnológico de múltiples pasos, para el descubrimiento de patrones objetivo de una base de datos. “El descubrimiento de conocimiento en bases de datos es un campo de la inteligencia artificial de rápido crecimiento, que combina técnicas del reconocimiento de patrones, estadística, bases de datos, y visualización para automáticamente extraer conocimiento (o información), de un nivel bajo de datos (bases de datos)” [FAYY97][[1]](#footnote-1).

**Los Pasos del KDD**

Se utilizan una serie de pasos, en los que se incluyen decisiones tomadas por el usuario. De acuerdo a las necesidades de los agentes involucrados se utilizan más o menos pasos.

**1)** **Definición de un objetivo:** Entender el dominio de aplicación, cuál es el problema a resolver, y cuáles son los objetivos.

**2)** **Definición de un subconjunto de datos:** del conjunto de datos originales se define un subconjunto apropiado, en respuesta al problema que se desea resolver. En este paso también se depuran variables irrelevantes.

**3)** **Etapa de limpieza**: se identifican valores faltantes, atípicos, erróneos, se normalizan los valores de las variables. Las grandes cantidades de información recolectadas por medio de métodos automáticos pueden contener valores fuera de rango (edades negativas), combinaciones incorrectas de datos (Sexo: masculino; embarazada: si), entre otros. Ingresar datos erróneos podría arrojar resultados alejados del comportamiento real.

**4)** **Reducción de dimensiones:** Reducción de columnas que se repiten (fecha de nacimiento y edad).

**5)** **Minería de datos**: se elige la herramienta de minería de datos adecuada al problema a resolver, (predecir, explicar, clasificar, agrupar, etc) asi como los parámetros de las redes utilizadas (arquitectura de la red, datos de entrenamiento, de validación y de testeo, etc). Posteriormente se procede a la identificación de patrones y relaciones entre datos.

**6)** **Interpretación de los datos:** Presentación de los datos por medio de gráficos y tablas para la explicación de los hallazgos.

**7)** **Consolidar el conocimiento** En esta etapa se debería consolidar el conocimiento ganado, prueba de los modelos creados aplicándolos al mundo real.

1. [FAYY97] Usama Fayyad y Evangelos Simoudis. (1997). Data Mining and Knowledge Discovery in Databases. [↑](#footnote-ref-1)