

# DATA SCIENCE

---

Cristiane Neri Nobre

# Data Science

- **Data Science** é a disciplina de processamento e análise de dados com o objetivo de extrair conhecimento valioso.
- O termo "Ciência dos Dados" foi inventado na década de 1960. No entanto, realmente tomou forma apenas recentemente, quando a tecnologia se tornou suficientemente madura.

# Descoberta de conhecimento

De acordo com Wives [27], **descobrir conhecimento** significa **identificar, receber informações relevantes**, e poder **processá-las e agregá-las** ao conhecimento prévio de seu usuário, mudando o estado de seu conhecimento atual, a fim de que determinada situação ou problema possa ser resolvido.

# Mineração de Dados

É um termo utilizado para descrever o processo de descoberta de padrões interessantes e desconhecidos em grandes bancos de dados

A mineração de dados é uma parte do processo geral de descoberta de conhecimento em bancos de dados (KDD - Knowledge Discovery in Databases)

# Mineração de Dados

O processo KDD foi definido por [Fayyad et al. (1996)] como "o processo não trivial de identificar padrões válidos, novos, potencialmente úteis e, finalmente, compreensíveis em dados".

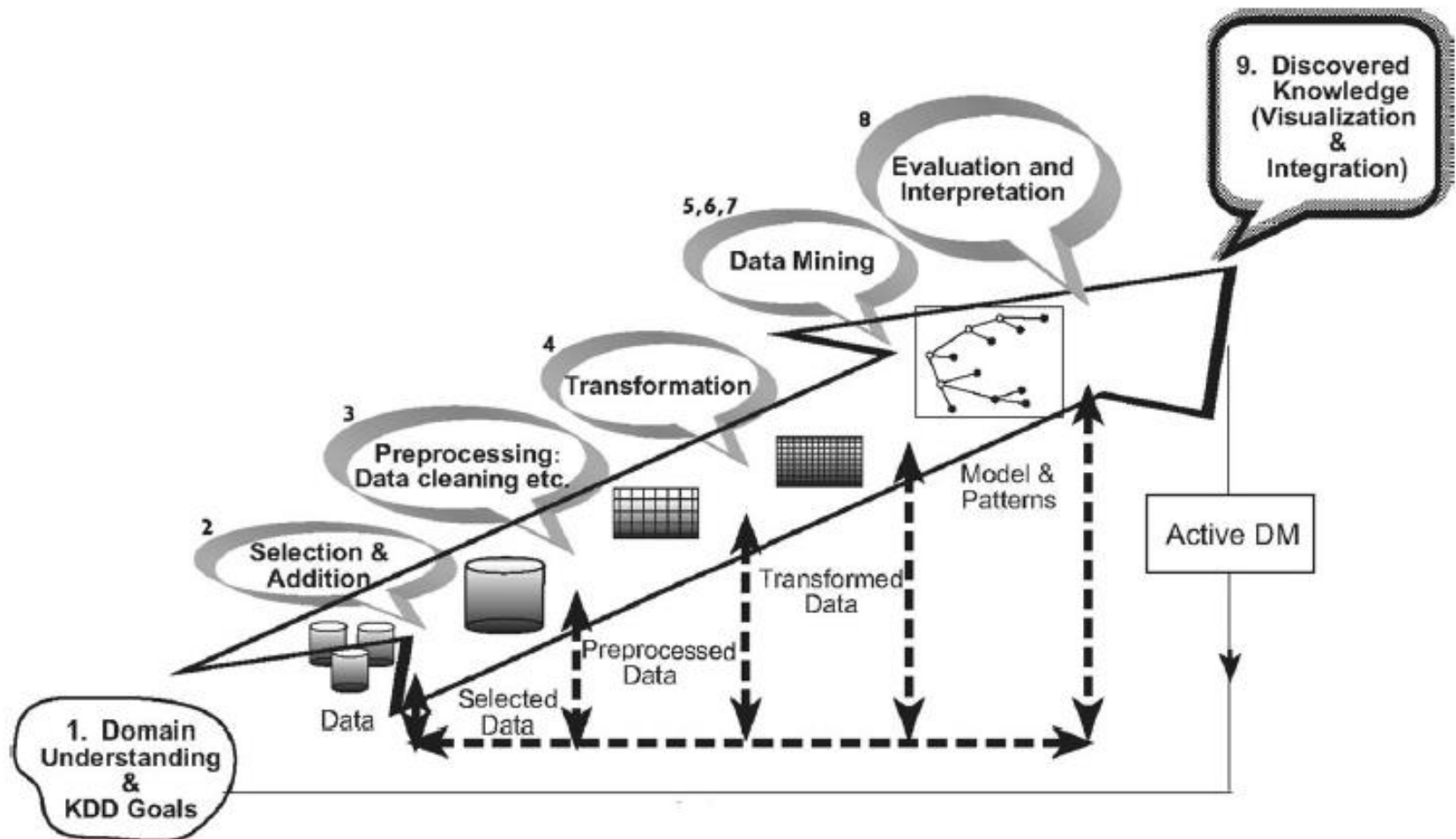
Friedman (1997a) considera o processo KDD como uma análise automática de dados exploratórios de grandes bancos de dados.

# Mineração de Dados

Alguns pesquisadores e profissionais usam o termo "mineração de dados" como sinônimo do processo completo de KDD.

Vários pesquisadores, como [Brachman e Anand (1994)], [Fayyad et al. (1996)] e [Reinartz (2002)] propuseram diferentes formas de dividir o processo KDD em fases.

# O processo de KDD



Extraído de Rokach, Lior. Data mining with decision trees: theory and applications. 2nd edition. 2015  
DM = Data Mining