

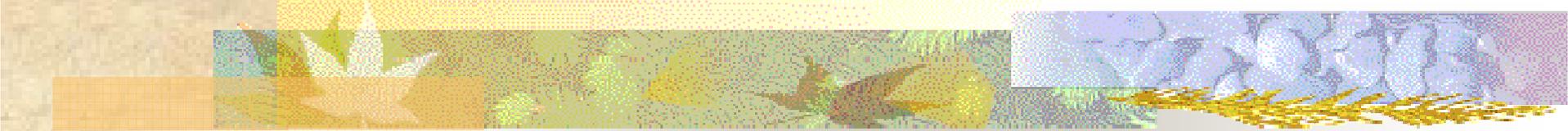
# Mineração de Dados

- Data cleansing
- Data integration
- Data selection
- Data transformation
- Data mining
- Pattern evaluation
- Knowledge representation



# Data cleansing

- Missing values
  - Ignorar a tupla
  - Preencher manualmente
  - Substituir todos os valores ausentes por uma cte
  - Usar a média
  - Usar a média de acordo com a classe
  - Preencher com o valor mais provável
    - Regressão, Bayes, árvores de decisão



# Data cleansing

- Noisy data
  - Binning
  - clustering
  - Inspeção humana e através de programa de computador
  - Regressão
  - inconsistências



# Data Transformation

- Smoothing
- Aggregation
- Generalização
- Normalização
- Construção de atributo (feature construction)



# Data Reduction

- Aggregation
- Redução de dimensão
- Compressão
  - Wavelet transform, PCA...
- Numerosity reduction
  - Modelos paramétricos, histogramas, amostragem...
- Discretização e construção de hierarquias