

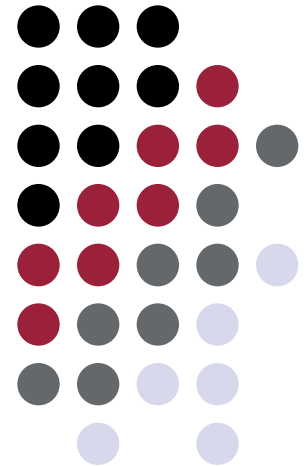
# Mineração de Dados



Universidade Federal  
de Ouro Preto

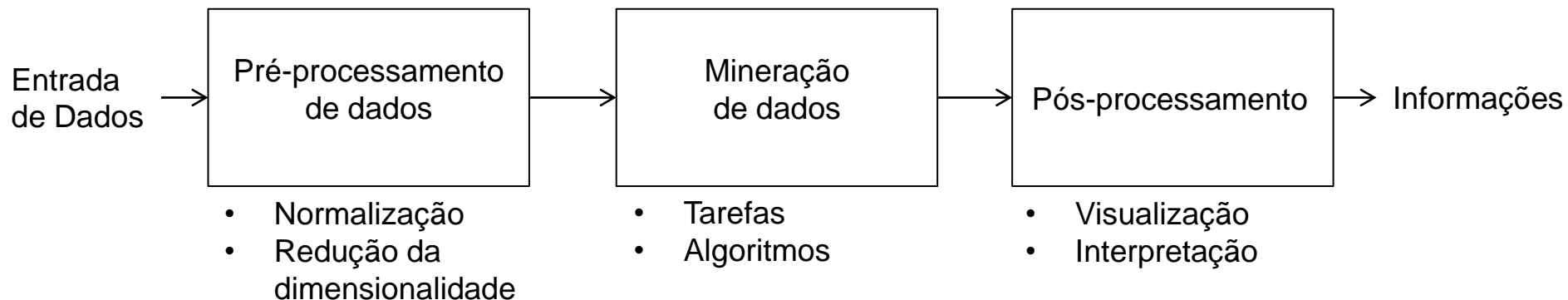
## CEA462 – Sistemas de Apoio à Decisão

Prof. MSc. George H. G. Fonseca  
Universidade Federal de Ouro Preto





- Mineração de dados
  - Processo de descoberta automática de informações úteis em grandes repositórios de dados
- Faz parte do processo de descoberta de conhecimento em bancos de dados (KDD)
  - Processo de conversão de dados brutos em informações úteis





- Negócios

- Detecção de perfis de clientes
- Vendas direcionadas
- Detecção de fraudes
- Prever o valor de ações

- Ciência e engenharia

- Detectar câncer em pacientes
- Identificar doenças de coração
- Identificar digitais
- Processamento de linguagem natural



- Classificação
  - Identificar potenciais inadimplentes
- Agrupamento
  - Agrupar tipos de doenças semelhantes
- Associação
  - Padrões de comportamento de clientes
- Detecção de anomalias
  - Detecção de mensagens spam



- Amostragem
  - Consiste em selecionar um subconjunto dos dados de entrada para derivar o modelo
    - Aleatoriamente
  - Há a perda de precisão, uma vez que não se usa o conjunto completo de dados
  - O conjunto de dados pode ser demasiadamente grande e a perda de precisão insignificante



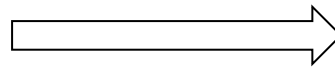
- Normalização
  - Consiste em colocar os atributos em uma mesma escala, de modo que possuam o mesmo 'peso' na derivação do modelo
    - Menor valor assumira 0 e o maior assumirá 1
    - Os demais figurarão proporcionalmente no intervalo  $[0, 1]$
  - A transformação de atributos discretos em contínuos e vice-versa também faz parte da normalização



- Normalização

- Exemplo
- Renda mensal:  $900 \equiv 0 / 6000 \equiv 1$
- Casa própria: Não  $\equiv 0 /$  Sim  $\equiv 1$
- Dependentes:  $0 \equiv 0 / 4 \equiv 1$

ID	Renda mensal	Casa própria	Dependentes
1	1000	Não	1
2	6000	Sim	2
3	1300	Não	4
4	2000	Não	1
5	1100	Não	3
6	1800	Sim	0
7	4300	Sim	2
8	0900	Não	1
9	2700	Sim	1



ID	Renda mensal	Casa própria	Dependentes
1	0,02	0,00	0,25
2	1,00	1,00	0,50
3	0,08	0,00	1,00
4	0,22	0,00	0,25
5	0,04	0,00	0,75
6	0,18	1,00	0,00
7	0,67	1,00	0,50
8	0,00	0,00	0,25
9	0,35	1,00	0,25

# Bibliografia



- *Introdução ao Data Mining*. Steinbach, Michael; Kumar, Vipin; Tan, Pang-ning, Rio de Janeiro: Ed. Ciência Moderna, 2009. Capítulos 1 e 2.

