

Limpeza de Dados

Agente Educacional

Sérgio M. Dias

Agenda

Dados incompletos
Dados com ruídos
Dados inconsistentes



Dados podem estar (**quase sempre estarão!**)

Incompletos

- Atributos ausentes
- Valores de atributos ausentes
- Dados agregados (*frequente em DWs*)

Com ruídos

- Erros (*ex: idade = -50 erro!*)
- Outliers (*valores discrepantes*)

Inconsistentes

- *Ex: idade = 43 e data de nascimento = 03/07/1997*
- *Atributos com valores "1,2,3" passa a assumir "a,b,c"*

Dados Incompletos

Causas mais comuns para dados ausentes

- Mau funcionamento de equipamento
- Dado não foi originalmente considerado importante
- Dado inconsistente, removido da base

Como tratar?

- Ignorar registro ou atributo
- Preencher valores manualmente
- Preencher valores automaticamente

Constante, mediana, moda, média dos vizinhos, média global, média de registros da mesma classe, métodos de aprendizado de máquina

[Exercício Limpeza de Dados - KNIME]





Obrigado!

Agente Educacional

Sérgio M. Dias

sergio.dias@serpro.gov.br | #31 6539

Demais agentes educacionais sobre o assunto

Marcelo Pita | marcelo.pita@serpro.gov.br | #81 8794

Gustavo Torres | gustavo.gamatorres@serpro.gov.br | #31 6950