

## **CADERNO DE EXERCÍCIOS – PARTE 01**

### **1 CAPÍTULO 01: INTRODUÇÃO**

#### **1.1 EXERCÍCIOS CONCEITUAIS**

- 1.1.1 Qual a diferença entre as terminologias *banco de dados* e *base de dados*?
- 1.1.2 A Fig. 1.4 ilustra a multidisciplinaridade da mineração de dados. Discuta como cada uma das áreas citadas contribui para a mineração de dados.
- 1.1.3 Discuta as principais similaridades e diferenças entre a análise de grupos e a classificação de dados.
- 1.1.4 Sabendo que a tarefa de mineração de regras de associação busca encontrar relações de coocorrência entre os atributos da base, cite cinco exemplos de problemas práticos para os quais essa tarefa é útil.
- 1.1.5 Discuta como o número de objetos e o número e tipos de atributos podem influenciar o processo de mineração de dados.
- 1.1.6 Considerando as muitas nomenclaturas já existentes na literatura, tais como Inteligência Artificial, Aprendizagem de Máquina e Inteligência Computacional, como você justificaria a necessidade de criação da nomenclatura Mineração de Dados?
- 1.1.7 Há um tipo de aprendizagem intermediário entre a supervisionada e a não supervisionada, intitulada semi-supervisionada. Explique, em linhas gerais, qual a diferença entre elas.
- 1.1.8 Explique a diferença entre, apresente e explique um exemplo de base de dados (problema) em que ambas as análises de cada alínea (a e b) abaixo podem ser feitas:
  - a) Agrupamento e classificação.
  - b) Classificação e estimação.

### **2 CAPÍTULO 02: PRÉ-PROCESSAMENTO DE DADOS**

#### **2.1 EXERCÍCIOS CONCEITUAIS**

- 2.1.1 Classifique os dados abaixo em estruturados, semiestruturados e não estruturados:
  - a) Tabela com os dados de cadastro dos funcionários de uma empresa.
  - b) Arquivos de Som.
  - c) Apresentações em PowerPoint.
  - d) Textos com palavras-chave identificadas.
  - e) Imagens com *tags*.

- 2.1.2 Os três principais tipos de problemas com os dados são incompletude, inconsistência e ruído, e as principais tarefas de pré-processamento são limpeza, integração, redução, transformação e discretização. Explique quais tarefas estão associadas a quais problemas e como se dá tal associação.
- 2.1.3 Discuta as possíveis implicações de objetos e atributos duplicados em uma base de dados no processo de mineração de dados.

## 2.2 EXERCÍCIOS NUMÉRICOS

- 2.2.1 Para a amostra da base de dados Mamo apresentada na Tabela 2.4 considere o atributo severidade como a classe alvo e faça:
- a) Impute os valores ausentes do atributo **forma** usando a moda por classe;
  - b) Impute os valores ausentes do atributo **contorno** usando o método hot-deck;
  - c) Impute os valores ausentes do atributo **densidade** usando uma constante global.
- 2.2.2 Ainda para a base de dados Mamo, faça a suavização (encaixotamento) em caixas de mesma frequência do atributo idade usando duas caixas. Faça a suavização pela média das caixas e também pelos extremos.
- 2.2.3 Normalize o atributo **dia** da base de dados da Tabela 2.3 utilizando os seguintes métodos: max-min no intervalo  $[0,1]$ ; escore-z; escalonamento decimal ( $j=2$ ); e range interquartil. Utilizando a normalização pelo escore-z ou range interquartil é possível identificar algum valor que se distancia excessivamente dos demais? Discuta.

## 2.3 DESAFIO COMPUTACIONAL

- 2.3.1 O Capítulo 2 apresenta o exemplo do processo de preparação de base de dados aplicado à base Mamo. Para este desafio, realize o mesmo processo para a base Bancos descrita na Seção 2.1.2. Realize as etapas do processo de forma similar ao exemplo do livro.

### 3 CAPÍTULO 03: ANÁLISE DESCRITIVA DE DADOS

#### 3.1 EXERCÍCIOS CONCEITUAIS

- 3.1.1 Qual é o objetivo da análise descritiva de dados?
- 3.1.2 Qual é o propósito na utilização de distribuições de frequência?
- 3.1.3 Qual a diferença entre as medidas de tendência central, as medidas de dispersão e as medidas de forma?

#### 3.2 EXERCÍCIOS NUMÉRICOS

- 3.2.1 Calcule a distribuição de frequência e desenhe o histograma simples do seguinte conjunto de valores,  $L = \{6, 7, 1, 9, 8, 2, 6, 4, 6, 4, 5, 2, 3, 1, 10, 7, 10, 2, 10, 8, 6, 5, 3, 8, 3, 1, 8, 7, 8, 7\}$ .
- 3.2.2 Para a base de dados abaixo, faça:
- Calcule o domínio, a média e a mediana de cada atributo.
  - Calcule o primeiro, segundo e terceiro quartis para todos os atributos.
  - Desenhe o diagrama em caixa para todos os atributos.

Atributo A	Atributo B	Atributo C	Atributo D
0,4190	0,1422	0,3474	0,5357
0,3908	0,0251	0,6606	0,0871
0,8161	0,4211	0,3839	0,8021
0,3174	0,1841	0,6273	0,9891
0,8145	0,7258	0,0216	0,0669
0,7891	0,3704	0,9106	0,9394
0,8523	0,8416	0,8006	0,0182
0,5056	0,7342	0,7458	0,6838
0,6357	0,5710	0,8131	0,7837
0,9509	0,1769	0,3833	0,5341
0,4440	0,9574	0,6173	0,8854
0,0600	0,2653	0,5755	0,8990
0,8667	0,9246	0,5301	0,6259
0,6312	0,2238	0,2751	0,1379
0,3551	0,3736	0,2486	0,2178
0,9970	0,0875	0,4516	0,1821
0,2242	0,6401	0,2277	0,0418
0,6525	0,1806	0,8044	0,1069
0,6050	0,0451	0,9861	0,6164
0,3872	0,7232	0,0300	0,9397

### **3.3 DESAFIO COMPUTACIONAL**

- 3.3.1 O Capítulo 3 traz o exemplo do processo de análise descritiva de dados aplicado aos atributos Mês, Dia, DMC e DC da base Fires. Para este desafio, realize o mesmo processo para os atributos X, Y, Temp e UR da base Fires. Realize as etapas do processo de forma similar ao exemplo exposto no livro.