

**Engenharia e Administração de Sistemas de Banco de Dados –  
Mineração de Dados**  
Prova

Questão	Valor	Nota
1	2,0	
2	2,0	
3	2,0	
4	2,0	
5	2,0	
Total	10,0	

Nome: \_\_\_\_\_

**Instruções:** *Somente serão consideradas respostas nos espaços marcados. Caso absolutamente necessário, continue no verso e indique claramente que a resposta está no verso.*

1. Uma conhecida cadeia de lojas eletrônicas pretende instituir um sistema de promoções para clientes a partir da análise do histórico de compras por categoria de produtos. As seguintes transações serão consideradas na análise:

Tabela 1: Transações

	celular	TV	eletrônicos	informática	livro	esporte-lazer	casa-decoração
1	não	sim	não	sim	sim	não	não
2	sim	não	sim	sim	sim	não	não
3	não	sim	não	sim	sim	não	não
4	sim	sim	não	sim	sim	não	não
5	não	não	sim	não	não	não	não
6	não	não	não	não	sim	não	não
7	não	não	não	sim	não	não	não
8	não	não	não	não	não	não	sim
9	não	não	não	não	não	sim	sim
10	não	não	não	não	não	sim	não

(a) Calcule o suporte para conjuntos de itens  $\{celular\}$ ,  $\{TV, celular\}$  e  $\{celular, informática, livro\}$ , tratando cada ID de transação como uma cesta de compras.

(b) Use os resultados da parte (a) para calcular a confiança para as regras  $\{celular\} \rightarrow \{TV\}$  e  $\{celular\} \rightarrow \{informática, livro\}$ .

(c) Indique uma regra envolvendo o item *celular* que tenha confiança maior do que 0,9.

2. Considere a existência de duas classes (*classe 1* e *classe 2*) composta por 5 instâncias cada. Considere ainda a existência de 3 diferentes vetores de características (*descriptor 1*, *descriptor 2*, e *descriptor 3*), que definem pontos do espaço  $\mathbb{R}^2$ . As Figuras 1(a), 1(b), e 1(c) mostram as classes 1 e 2 no plano cartesiano para os descritores 1, 2 e 3, respectivamente.

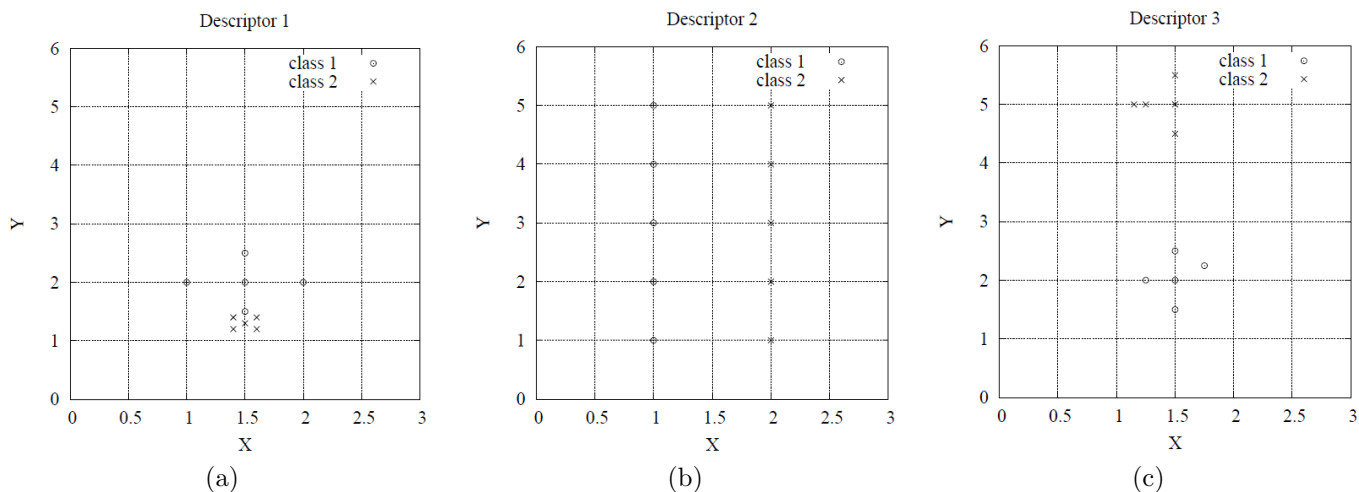


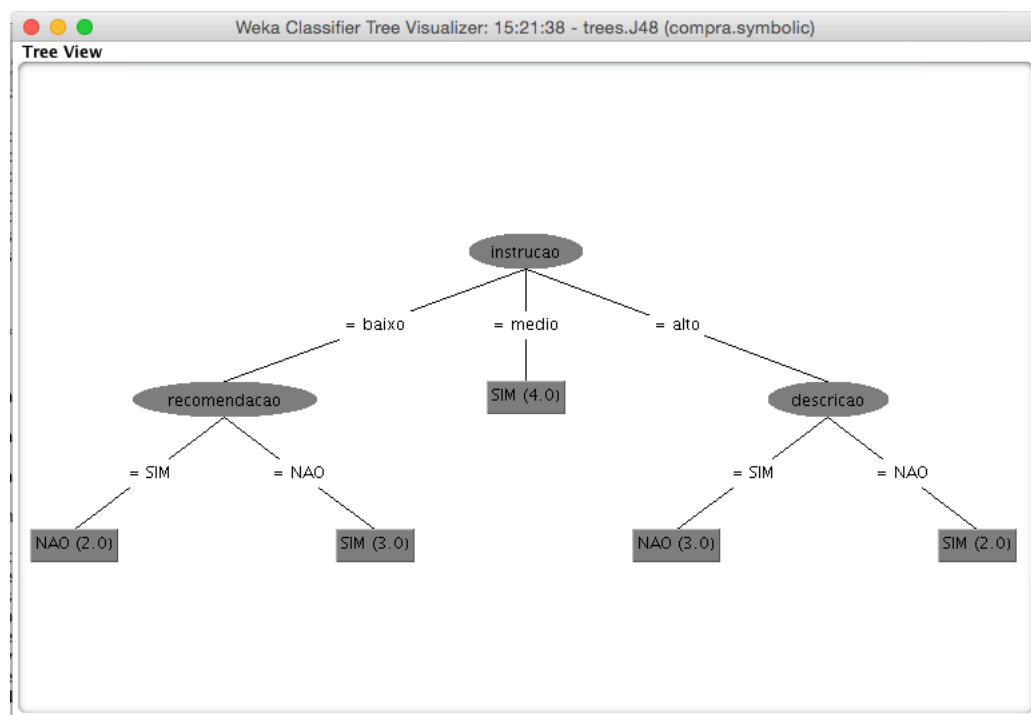
Figura 1: (a) Descritor 1. (b) Descritor 2. (c) Descritor 3.

Suponha que você deseja determinar a classe da nova instância definida pelo ponto **(2,0; 4,0)**, usando o algoritmo KNN, com  $K=1$ . Esta nova instância seria classificada em qual classe, segundo os descritores? Sua resposta seria diferente se fosse usado  $K=3$ ? Justifique suas respostas.

3. Considere a árvore de decisão abaixo. Ela refere-se a um modelo que pretende prever se compras serão efetivadas em uma *site* de comércio eletrônico. Quatro variáveis são consideradas no modelo:

- **instrucao**: refere-se ao grau de instrução do cliente que pode ser *baixo*, *médio*, ou *alto*;
- **recomendacao**: registra se o cliente acessou a página do produto por um *link* recomendado (valor *SIM*) ou não (valor *NÃO*);
- **descricao**: registra se o cliente buscou informações adicionais sobre o produto, clicando em algum *link* de descrição (há dois valores possíveis: *SIM* ou *NÃO*);
- **duracao**: refere-se ao tempo dispensado pelo cliente no *site*, que pode ser *baixo*, *médio*, ou *alto*.

As classes do modelo indicam a efetivação ou não de uma compra (classe *SIM* ou classe *NÃO*).



(a) Indique uma instância de teste que seria classificada como pertencendo à classe “NÃO”.

(b) Como seriam classificadas as instâncias abaixo?

- duracao = alto, descricao = NAO, recomendacao = SIM, instrucao = baixo
- duracao = alto, descricao = SIM, recomendacao = NAO, instrucao = alto
- duracao = alto, descricao = SIM, recomendacao = NAO, instrucao = medio



4. Considere o seguinte conjunto de pontos:  $p_1 = (2, 10)$ ,  $p_2 = (2, 5)$ ,  $p_3 = (8, 4)$ ,  $p_4 = (5, 8)$ ,  $p_5 = (7, 5)$ ,  $p_6 = (6, 4)$ ,  $p_7 = (1, 2)$  e  $p_8 = (4, 9)$ . Suponha que os pontos  $p_1$ ,  $p_4$  e  $p_7$  sejam usados como centróides iniciais do algoritmo de agrupamento K-means ( $k = 3$ ). Usando a distância de Manhattan (i.e., dados pontos  $(x_1, y_1)$  e  $(x_2, y_2)$ ,  $distancia_{Manhattan} = |x_1 - x_2| + |y_1 - y_2|$ ), calcule os três clusters para cada uma das primeiras iterações do algoritmo.

5. Considere a tabela abaixo que contém registros de compras realizadas *on-line*.

Cód. Produto	Descrição	Cód. Item	U.F.	Qtd. Pedida	Qtd. Aprovada	Qtd. Atendida	Valor Total	Situação
3095	SABONETE LIQUIDO, PRONTO USO, 5 LT	3095	U.F. GALAO COM 5 LITROS	4.000		4	23.32	ATENDIDO
971925	REATOR ELETRONICO 2X32WX220V	954799	UNIDADE	25.000		25	344.50	ATENDIDO
711	DETERGENTE NEUTRO CONCENTRADO	711	FRASCO COM 500 MILILITROS	5.000		5	4.37	ATENDIDO
51204	ENVELOPE DE PAPELARIA KRAFT 240X340 MM	51204	UNIDADE	150.000		150	29.85	ATENDIDO
18	ACUCAR REFINADO	18	PACOTE COM 5 QUILOGRAMAS	6.000		6	62.50	ATENDIDO
46239	MEXEDOR P/ CAFE, PLASTICO, 9CM, TIPO REMO	46239	AMAS PACOTE COM 500 UNIDADE	4.000		4	10.64	ATENDIDO
46238	ADOCANTE DIETETICO, SACARINA/CLICAMATO	46238	UNIDADE	3.000		3	4.12	ATENDIDO
6655	CAIXA ARQUIVO COR BRANCA COM LOGOTIPO	6655	UNIDADE	5.000		5	8.05	ATENDIDO
79439	PANO DE LIMPEZA, ALVEJADO, MED 74X45CM	79439	UNIDADE	2.000		2	2.92	ATENDIDO
63556	LAMPADA FLUOR. TUBULAR 32W	63556	UNIDADE	40.000		40	109.60	ATENDIDO
59	BARBANTE DE ALGODAO, 8 FIOS 184 MTS.	59	UNIDADE	1.000		1	3.68	ATENDIDO
17749	CORRETIVO LIQUIDO A BASE D'AGUA	17749	UNIDADE	1.000		1	0.83	ATENDIDO
406	PAPEL ALMACO C/PAUTA, SULFITE(200X275)MM	406	DEZENA	62.000		62	31.27	ATENDIDO
50146	SACO LIXO, USO DOMESTICO, PRETO, 100L	50146	PACOTE COM 100 UNIDADE	4.000		4	116.00	ATENDIDO
71558	CAPA DE PROCESSO, EM PVC INCOLOR, 48 X	71558	UNIDADE	20.000		20	32.04	ATENDIDO
33654	COPO DESCARTAVEL, POLIESTIRENO, CAP. 200ML	33654	CENTO	150.000		150	319.72	ATENDIDO
73159	BORRACHA DE PAPELARIA EM PVC	73159	UNIDADE	2.000		2	1.30	ATENDIDO
17954	BORRACHA PARA DESENHO, MACIA	17954	UNIDADE	2.000		2	1.13	ATENDIDO
182	COLCHETE LATONADO N.10, CX C/72 A 100 PC	182	CAIXA COM 72 UNIDADE	1.000		1	1.85	ATENDIDO
48967	PAPEL TAMANHO CARTA, 216 X 279MM	48967	S PACOTE COM 500 FOLHAS	20.000		20	164.12	ATENDIDO
199307	PAPEL SULFITE TAMANHO A4, 210X297MM	199307	PACOTE COM 500 FOLHAS	70.000		70	610.70	ATENDIDO
110247	FORM.CAPA DE PROCESSO	110247	UNIDADE	30.000		30	21.55	ATENDIDO
745	ESPONJA PARA LIMPEZA	745	UNIDADE	1.000		1	0.39	ATENDIDO
101514	PAPEL HIGIENICO P/ DISPENSER ROLO 300 M	101514	UNIDADE	80.000		80	205.00	ATENDIDO
492637	TOALHA DE PAPEL-SIMPLES-CLASSE I-BRANCO	492637	PACOTE COM 1250 UNIDADE	85.000		85	732.04	ATENDIDO
971348	MARCADORES DE PAGINA, SETINHA, POLIESTER	31246	CARTELA COM 1 CARTELA	2.000		2	5.80	ATENDIDO
82006	RECADADO AUTO	82006	PACOTE COM 4 BLOCOS	4.000		4	13.36	ATENDIDO
517292	ADESIVO(38X50)MM,S/PAUTA COLA PAPEL - BASTAO	517292	COM 50 FOLHAS	4.000		4	15.40	ATENDIDO
59466	SACO LIXO, COLETA SELET., VERMELHO, 100L	59466	UNIDADE	1.000	1.000	1	11.37	ATENDIDO
59465	SACO DE LIXO, COLETA SELET., AZUL, 100L	59465	PACOTE COM 20 UNIDADE	1.000	1.000	1	7.95	ATENDIDO

a) Identifique e caracterize um problema de mineração envolvendo estes dados. Qual técnica de mineração de dados seria utilizada? Por quê?

b) Como seriam organizados os dados de entrada da técnica escolhida, caso o software Weka fosse utilizado? Que tipo de pré-processamento seria necessário?

