



Processo Seletivo – Programa de Residência em Tecnologia da Informação – Tribunal de Justiça do Estado do Acre

Professor(a): Comissão de seleção

Os cartões de resposta devem ser lidos
usando o aplicativo Multiprova Corretor.



iOS



ANDROID

1 prova

Nome:

Matrícula:

Assinatura: _____

Questão 1

Valor da questão: 1,00

Considere o seguinte código Python:

```
import pandas as pd

data = {'Empresa': ['GOOG', 'GOOG', 'MSFT', 'MSFT', 'FB', 'FB'],
        'Nome': ['Sam', 'Charlie', 'Amy', 'Vanessa', 'Carl', 'Sarah'],
        'Venda': [200, 120, 340, 124, 243, 350]}

df = pd.DataFrame(data)

por_companhia = df.groupby("Empresa")
```

Qual das linhas de código abaixo é responsável por permitir a seguinte saída (output)?

	Nome	Venda
Empresa		
FB	Sarah	350
GOOG	Sam	200
MSFT	Vanessa	340

- a) `por_companhia.std()`
- b) `por_companhia.mean()`
- c) `por_companhia.min()`
- d) `por_companhia.max()`

Questão 2

Valor da questão: 1,00

Dada a seguinte base de dados:

#	X1	X2	X3	X4	X5	X6	X7	X8	X9	Class
1	5	2	3	4	2	7	3	6	1	malignant
2	5	4	4	5	7	10	3	2	1	benign
3	3	1	1	1	2	2	3	1	1	benign
4	6	8	8	1	3	4	3	7	1	benign
5	4	1	1	3	2	1	3	1	1	benign
6	8	10	10	8	7	10	9	7	1	malignant
7	1	1	1	1	2	10	3	1	1	benign
8	2	1	2	1	2	1	3	1	1	benign
9	8	7	5	10	7	9	5	5	4	malignant
10	7	4	6	4	6	1	4	3	1	malignant

Utilizando o classificador k-NN (*k-nearest neighbors* ou “k-vizinhos mais próximos”) com os valores 1, 2 e 3 para k, e a distância Euclidiana como métrica de distância, classifique esta instância de teste: (10, 10, 10, 4, 8, 1, 8, 10, 1). Quais serão os resultados para os diferentes valores de k?

- a) benign, malignant, malignant
- b) benign, benign, malignant
- c) malignant, malignant, benign
- d) benign, malignant, benign

Questão 3

Valor da questão: 1,00

Com base nos sistemas de banco de dados NoSQL, assinale a alternativa que correlaciona corretamente os SGBD's no NoSQL e seus modelos estruturais.

- a) Cassandra: Modelo Orientado a Colunas – Neo4J: Modelo Baseado em Grafos – Redis: Modelo Orientado a Documentos – MongoDB: Modelo Chave-Valor.
- b) MongoDB: Modelo Orientado a Colunas – Cassandra: Modelo Baseado em Grafos – Neo4J: Modelo Orientado a Documentos – Redis: Modelo Chave-Valor.
- c) Cassandra: Modelo Orientado a Colunas – Neo4J: Modelo Baseado em Grafos – MongoDB: Modelo Orientado a Documentos – Redis: Modelo Chave-Valor.
- d) Redis: Modelo Orientado a Colunas – Cassandra: Modelo Baseado em Grafos – MongoDB: Modelo Orientado a Documentos – Neo4J: Modelo Chave-Valor.

Questão 4

Valor da questão: 1,00

Qual é o resultado da seguinte consulta SQL, que usa a função RANK () ?

```
SELECT nome, salario,  
RANK() OVER (ORDER BY salario DESC) AS ranking  
FROM funcionarios;
```

- a) Retorna o nome e o salário de cada funcionário, mas não inclui informações de classificação.
- b) Retorna o nome e o salário de cada funcionário, em ordem alfabética decrescente.
- c) Retorna um erro, pois as funções de janela não são suportadas em banco de dados relacionais.
- d) Retorna o nome, o salário e a classificação de cada funcionário com base no salário, em ordem decrescente.

Questão 5

Valor da questão: 1,00

Qual é o principal objetivo do Apache Airflow?

- a) Gerenciar contêineres Docker.
- b) Automatizar a execução de fluxos de trabalho.
- c) Servir como um servidor web.
- d) Armazenar grandes volumes de dados.

Questão 6

Valor da questão: 1,00

Qual é o resultado da seguinte consulta SQL, que envolve junção de tabelas?

```
SELECT clientes.nome, COUNT(pedidos.id)  
FROM clientes  
LEFT JOIN pedidos ON clientes.id = pedidos.cliente_id  
GROUP BY clientes.nome;
```

- a) Retorna o nome de cada cliente e o número total de pedidos em toda a tabela de pedidos.
- b) Retorna o nome de cada cliente e a média de pedidos feitos por cliente.
- c) Retorna o nome de cada cliente e o número de pedidos feitos por aquele cliente.
- d) Retorna um erro, pois não é possível fazer uma junção entre essas tabelas.

Questão 7

Valor da questão: 1,00

Qual é o resultado da seguinte consulta SQL, que usa uma subconsulta correlacionada?

```
SELECT nome, salario  
FROM funcionarios AS f  
WHERE salario > (SELECT AVG(salario) FROM funcionarios WHERE departamento =  
f.departamento);
```

- a) Retorna os nomes e salários dos funcionários cujos salários são menores do que a média de seus colegas no mesmo departamento.
- b) Retorna os nomes e salários dos funcionários cujos salários são maiores do que a média de seus colegas no mesmo departamento.
- c) Retorna os nomes e salários de todos os funcionários.
- d) Retorna um erro, pois subconsultas correlacionadas não são suportadas.

Questão 8

Valor da questão: 1,00

Em Ciência de Dados, Python é uma das linguagens de programação mais utilizadas. A esse respeito, é correto afirmar que a linguagem de programação Python:

- a) Possui recursos para controle de fluxo, tais como: *if-else*, *switch-case*, *while* e *for* em todas as suas versões.
- b) Foi desenvolvida com o intuito de substituir a linguagem de programação C por causa de sua altíssima performance.
- c) Pode ser utilizada como uma linguagem de programação funcional.
- d) Mostra-se ideal para desenvolvimento rápido e criação de scripts em razão de sua natureza compilada.

Questão 9

Valor da questão: 1,00

Você deseja encontrar a quantidade de produtos em cada categoria na tabela "produtos". Qual código SQL atinge esse objetivo?

- a) `SELECT categoria, COUNT(*) FROM produtos GROUP BY categoria;`
- b) `SELECT categoria, MAX(quantidade) FROM produtos GROUP BY categoria;`
- c) `SELECT categoria, SUM(quantidade) FROM produtos GROUP BY categoria;`
- d) `SELECT categoria, AVG(preco) FROM produtos GROUP BY categoria;`

Questão 10

Valor da questão: 1,00

Uma *black box* é como uma caixa opaca onde você não consegue observar o seu interior. No contexto de Aprendizado de Máquina, na *black box* você pode visualizar entradas e saídas, mas não o seu funcionamento. Qual dos algoritmos abaixo é considerado uma *black box*?

- a) MLP (Perceptron de Múltiplas Camadas)
- b) k-NN (k-vizinhos mais próximos)
- c) k-means (k-médias)
- d) Naive Bayes (Bayes ingênuo)

Questão 11

Valor da questão: 1,00

Analise as seguintes afirmativas:

- I. *k-means* (k-médias) é um algoritmo de aprendizado não supervisionado, em que se calcula a distância entre os objetos da base e cada um dos centroides; em que se atribui cada objeto ao centroide mais próximo;
- II. k-NN (k-vizinhos mais próximos) é um método de *Clustering* (agrupamento). Uma vez que os agrupamentos e os seus centroides são identificados, é fácil atribuir novos objetivos para um cluster baseado na distância do objeto do centroide mais próximo.
- III. O CRISP-DM é uma metodologia abrangente de mineração de dados e um modelo de processo que fornece, para os usuários de *data mining* (DM, mineração de dados), um modelo completo para a realização de um projeto de DM.

Podemos afirmar corretamente que:

- a) Apenas as afirmativas I e III estão corretas.
- b) Apenas as afirmativas I e II estão corretas.
- c) Apenas as afirmativas II e III estão corretas.
- d) Todas as afirmativas estão corretas.

Questão 12

Valor da questão: 1,00

O que é normalização em bancos de dados relacionais?

- a) Um processo de criptografia de dados sensíveis.
- b) Um processo de adição de redundância aos dados para melhorar o desempenho.
- c) Um processo de conversão de dados em formato JSON.
- d) Um processo de remoção de redundância dos dados para melhorar a integridade.

Questão 13

Valor da questão: 1,00

Qual é a principal finalidade do *Data Lake* no contexto de análise de dados?

- a) Facilitar a geração de relatórios em tempo real.
- b) Servir como um banco de dados transacional.
- c) Realizar análises estatísticas avançadas.
- d) Armazenar grandes volumes de dados não estruturados.

Questão 14

Valor da questão: 1,00

Você deseja criar uma nova tabela chamada "estoque_backup" que seja uma cópia exata da tabela "estoque". Qual código SQL realiza essa tarefa?

- a) COPY estoque TO estoque_backup;
- b) BACKUP TABLE estoque TO estoque_backup;
- c) CREATE TABLE estoque_backup AS SELECT * FROM estoque;
- d) INSERT INTO estoque_backup SELECT * FROM estoque;

Questão 15

Valor da questão: 1,00

O que é uma chave primária em um banco de dados relacional?

- a) Uma chave usada para criptografar dados sensíveis.
- b) Uma chave usada para estabelecer relacionamentos entre tabelas.
- c) Uma chave usada para identificar unicamente cada registro em uma tabela.
- d) Uma chave usada para acessar o sistema de gerenciamento de banco de dados.

Questão 16

Valor da questão: 1,00

_____ é uma classe de métodos ensemble que utiliza classificadores de árvore de decisão. É uma combinação de preditores de árvores tal que cada árvore depende dos valores de um vetor aleatório amostrado independentemente e com a mesma distribuição para todas as árvores.

Assinale a alternativa que preenche corretamente a lacuna do trecho acima:

- a) Random Forest (RF, floresta aleatória)
- b) k-Nearest Neighbors (k-NN, k-vizinhos mais próximos)
- c) Support Vector Machine (SVM, máquina de vetor de suporte)
- d) Regressão logística

Questão 17

Valor da questão: 1,00

Considere o esquema relacional abaixo, no qual a placa é a chave primária:
VEICULO(idVeiculo, placa, cor, modelo, marca, ano, valor)

Qual é a expressão em álgebra relacional a ser aplicada sobre esse esquema, de forma a obter as placas dos VEICULOS com ano igual a 2023 e valor menor ou igual a R\$ 90.000,00?

- a) $\pi_{placa}(\pi_{ano = 2023; \pi_{valor > 90000}}(VEICULO))$
- b) $\pi_{placa}(\sigma_{ano = 2023 \text{ AND } valor \leq 90000}(VEICULO))$
- c) $\sigma_{placa}(\pi_{valor \leq 90000 \text{ AND } ano = 2023}(VEICULO))$
- d) $\sigma_{placa}(\sigma_{ano = 2023; valor \leq 90000})$

Questão 18

Valor da questão: 1,00

A seguir é apresentado um trecho de código escrito na linguagem Python.

```
x = [2, 3, 1, 5, 4]

for i in range(len(x)-1):
    x[i], x[i+1] = x[i+1], x[i]

print(x)
```

O valor da variável x após a execução desse trecho de código será:

- a) [4,2,5,3,1]
- b) [2,3,1,5,4]
- c) [3,1,5,4,2]
- d) [3,4,1,2,5]

Questão 19

Valor da questão: 1,00

Suponha que para resolver um determinado problema, você aplicou quatro técnicas de classificação, sendo: Árvore de Decisão, k-NN (k-vizinhos mais próximos), Naive Bayes (Bayes ingênuo) e Perceptron de Múltiplas Camadas. Ao comparar o resultado dos classificadores, você percebeu que o Naive Bayes teve uma taxa de acerto bem menor que as taxas dos outros classificadores. O que é possível afirmar sobre o conjunto de treinamento:

- a) Possui os atributos independentes e/ou os atributos numéricos não obedecem a uma distribuição normal.
- b) Possui os atributos independentes e/ou os atributos numéricos obedecem a uma distribuição normal.
- c) Possui os atributos dependentes e/ou os atributos numéricos obedecem a uma distribuição normal.
- d) Possui os atributos dependentes e/ou os atributos numéricos não obedecem a uma distribuição normal.

Questão 20

Valor da questão: 1,00

Seja a seguinte sequência de operações da Álgebra Relacional:

$$\pi_{A1,A2}(\sigma_{A1=5}(A \bowtie_{A1=B3} B))$$

Considerando essa sequência da esquerda para a direita, que operações foram empregadas?

- a) Projeção, seleção e junção
- b) Seleção, junção e projeção
- c) Seleção, projeção e junção
- d) Junção, seleção e projeção

Questão 21

Valor da questão: 1,00

Sobre o Esquema Estrela (*Star Schema*) de um *Data Warehouse*, é **INCORRETO** afirmar que:

- a) a tabela fato é ligada sempre a duas ou mais dimensões.
- b) o Esquema Estrela é composto por tabelas do tipo fato e dimensão.
- c) a granularidade, também chamada de detalhe, é o menor nível da hierarquia da dimensão.
- d) a dimensão descreve os dados que estão nos fatos, ou seja, as chaves.

Questão 22

Valor da questão: 1,00

Big Data surgiu a partir da necessidade de manipular um grande volume de dados e, com isso, novos conceitos foram introduzidos, como o *Data Lake*. Considere as afirmações abaixo e assinale a alternativa correta.

- I. É o resultado de sucessivas operações de mineração de dados, sendo um ambiente no qual é possível ter relatórios e dashboards de maneira amigável para os analistas de negócio.
 - II. Serve como um banco de dados transacional.
 - III. É projetado para armazenar dados de diversas fontes e formatos, não havendo a necessidade da definição de um esquema de dados para inserir novos itens.
 - IV. Tem como finalidade armazenar grandes volumes de dados não estruturados.
- a) Estão corretas as afirmações I, III e IV.
 - b) Estão corretas as afirmações III e IV.
 - c) Estão corretas as afirmações I e IV.
 - d) Estão corretas as afirmações II e IV.

Questão 23

Valor da questão: 1,00

Qual técnica é usada para segmentar clientes em grupos com base em características semelhantes?

- a) Associação.
- b) Classificação.
- c) Regressão linear.
- d) Clustering (agrupamento).

Questão 24

Valor da questão: 1,00

O que é o teorema CAP (Consistência, Disponibilidade e Tolerância a Partições) em relação a bancos de dados NoSQL?

- a) É um teorema que descreve as limitações na capacidade de um sistema distribuído em garantir simultaneamente consistência, disponibilidade e tolerância a partições.
- b) É um teorema que define os tipos de dados que podem ser armazenados em bancos de dados NoSQL.
- c) É um teorema que descreve a estrutura de tabelas em bancos de dados NoSQL.
- d) É um teorema que descreve a escalabilidade de bancos de dados NoSQL.

Questão 25

Valor da questão: 1,00

Qual é a diferença entre as listas e as tuplas em Python?

- a) Listas podem conter elementos de tipos diferentes, mas as tuplas não.
- b) Listas são mutáveis, enquanto tuplas são imutáveis.
- c) Não há diferença, os termos são usados indistintamente.
- d) Tuplas são usadas para iterar sobre sequências, enquanto listas são usadas para armazenar elementos.

Questão 26

Valor da questão: 1,00

Considere o fluxo de transformação do *Pentaho Data Integration* apresentado a seguir:



Os passos S1, S2 e S3 realizam, respectivamente, as operações de:

- a) Leitura de Arquivo CSV, Split de Colunas e Calculadora de Valores.
- b) Leitura de Arquivo de Propriedades, Seleção de Linhas Únicas e Calculadora de Valores.
- c) Leitura de Banco de Dados, Join de Colunas e Log.
- d) Leitura de Arquivo JSON, Split de Colunas e Saída em Formato de Tabela.

Questão 27

Valor da questão: 1,00

A modelagem de *Data Warehouses* pode ser feita seguindo diferentes esquemas. Sobre esse tópico, analise as afirmativas:

- I. No esquema estrela, os dados são organizados em uma tabela dimensão e muitas tabelas fatos;
- II. O esquema floco de neve é uma variação do esquema estrela, onde algumas tabelas fatos são normalizadas, dividindo, assim, os dados em tabelas adicionais;
- III. Quando várias tabelas fatos compartilham as mesmas tabelas dimensões, temos nesse caso, uma constelação de fatos;
- IV. O esquema floco de neve é uma variação do esquema estrela, onde algumas tabelas dimensão são normalizadas, dividindo, assim, os dados em tabelas adicionais.

Podemos afirmar corretamente que:

- a) somente a afirmativa II está correta.
- b) somente as afirmativas II, III e IV estão corretas.
- c) somente as afirmativas I, II e IV estão corretas.
- d) somente as afirmativas III e IV estão corretas.

Questão 28

Valor da questão: 1,00

Em uma consulta SQL, qual cláusula é utilizada para limitar o número de registros (tuplas) retornados?

- a) LIMIT
- b) RESTRICT
- c) COUNT
- d) TOP

Questão 29

Valor da questão: 1,00

Em relação ao processamento de linguagem natural (PLN), analise as afirmativas a seguir.

- I. O PLN envolve a compreensão e a geração de linguagem natural humana.
- II. A tarefa principal do PLN é traduzir textos de uma língua para outra.
- III. O PLN não é utilizado para tarefas de processamento de voz.
- IV. O PLN é aplicado em sistemas de recuperação de informações e assistentes virtuais.

Estão corretas as afirmativas:

- a) I e IV, apenas.
- b) II e III, apenas.
- c) I e III, apenas.
- d) I, II e III.

Questão 30

Valor da questão: 1,00

Dois tabelas de página são mantidas durante a vida de uma transação: a tabela de página atual e a tabela de página cópia. Quando a transação inicia, as duas tabelas são idênticas. A tabela de página cópia nunca é alterada durante a execução da transação. A tabela de página atual é alterada quando a transação processa uma operação de escrita. Quando a transação é parcialmente efetivada, a tabela de página cópia é descartada e a tabela de página atual torna-se a nova tabela de página. Se a transação for abortada, a tabela de página atual é descartada. Qual é a técnica de recuperação do banco de dados em caso da falha descrita acima?

- a) Paginação de sombra (*shadow*).
- b) Recuperação imediata.
- c) Recuperação baseada em Log.
- d) Recuperação adiada.

Questão 31

Valor da questão: 1,00

Dentro do contexto de Mineração de Dados, podemos afirmar:

- I. A limpeza dos dados é uma das etapas do processo.
- II. O treinamento é uma parte fundamental para efetuar a limpeza dos dados.
- III. A busca por padrões consiste em um dos possíveis objetivos da Mineração de Dados.
- IV. O escalonamento dos valores é uma etapa possível na preparação dos dados.

Sobre as afirmativas anteriores, é correto dizer que:

- a) Somente a alternativa I está correta.
- b) Estão corretas somente as afirmativas III e IV.
- c) Somente a alternativa II está incorreta.
- d) Estão incorretas as afirmativas II e IV.

Questão 32

Valor da questão: 1,00

O objetivo é obter a lista dos 3 funcionários com os maiores salários na tabela "funcionarios". Qual código SQL atinge esse objetivo?

- a) `SELECT nome, salario FROM funcionarios ORDER BY salario DESC LIMIT 3;`
- b) `SELECT nome, salario FROM funcionarios ORDER BY salario ASC FETCH FIRST 3 ROWS ONLY;`
- c) `SELECT nome, salario FROM funcionarios ORDER BY salario ASC LIMIT 3;`
- d) `SELECT nome, salario FROM funcionarios ORDER BY salario DESC OFFSET 3 FETCH NEXT 3 ROWS ONLY;`

Questão 33

Valor da questão: 1,00

A modelagem de *Data Warehouses* pode ser feita seguindo diferentes esquemas. Sobre esse tópico, analise as afirmativas:

- I. No esquema estrela, os dados são organizados em uma tabela dimensão e muitas tabelas fatos;
- II. O esquema floco de neve é uma variação do esquema estrela, onde algumas tabelas fatos são normalizadas, dividindo, assim, os dados em tabelas adicionais;
- III. Quando várias tabelas fatos compartilham as mesmas tabelas dimensões, temos nesse caso, uma constelação de fatos.

Podemos afirmar corretamente que:

- a) somente a afirmativa III está correta.
- b) somente as afirmativas I e II estão corretas.
- c) somente a afirmativa II está correta.
- d) somente as afirmativas II e III estão corretas.

Questão 34

Valor da questão: 1,00

Considere o seguinte código Python que utiliza compreensão de lista:

```
numeros = [1, 2, 3, 4, 5]
resultado = [x for x in numeros if x % 2 != 0]
```

Qual é o valor da variável resultado após a execução deste código?

- a) [2, 4, 6]
- b) [2, 4]
- c) Erro de sintaxe.
- d) [1, 3, 5]

Questão 35

Valor da questão: 1,00

Qual dos códigos Python executa o seguinte procedimento: Dada uma lista de palavras, criar uma nova lista que contenha apenas as palavras com mais de 5 caracteres.

- a) palavras = ["python", "programação", "dados", "ciência"]
palavras_longas = [p for p in palavras if p.count() > 5]
- b) palavras = ["python", "programação", "dados", "ciência"]
palavras_longas = [p for p in palavras if len(p) > 5]
- c) palavras = ["python", "programação", "dados", "ciência"]
palavras_longas = [p for p in palavras if p.length() > 5]
- d) palavras = ["python", "programação", "dados", "ciência"]
palavras_longas = [p for p in palavras if p.size() > 5]

Questão 36

Valor da questão: 1,00

Qual das seguintes afirmações é verdadeira sobre bancos de dados relacionais?

- a) São altamente escaláveis horizontalmente.
- b) Eles são especialmente adequados para armazenar dados semi-estruturados.
- c) Não requerem linguagem de consulta.
- d) Usam tabelas para armazenar dados e relacionamentos entre eles.

Questão 37

Valor da questão: 1,00

São exemplos de bancos de dados relacionais e bancos No-SQL, respectivamente:

- a) SQLite, Cassandra, PostgreSQL e Neo4j.
- b) Redis, DB2, Couchbase e MongoDB.
- c) MongoDB, MS Access, MySQL e DataStax.
- d) MySQL, Firebird, MongoDB e HBase.

Questão 38

Valor da questão: 1,00

Qual é a diferença entre a cláusula WHERE e a cláusula HAVING em uma consulta SQL?

- a) WHERE é usada para agrupar registros com base em critérios especificados, enquanto HAVING é usada para filtrar registros com base em critérios especificados.
- b) WHERE e HAVING são usadas de forma intercambiável em consultas SQL.
- c) WHERE é usada para agrupar registros com base em critérios especificados, enquanto HAVING é usada para ordenar registros com base em critérios especificados.
- d) WHERE é usada para filtrar registros com base em critérios especificados, enquanto HAVING é usada para filtrar grupos agregados com base em critérios especificados.

Questão 39

Valor da questão: 1,00

O que é um DAG (*Directed Acyclic Graph*, grafo acíclico direcionado) no contexto do Apache Airflow?

- a) Um tipo de banco de dados NoSQL.
- b) Uma estrutura de dados para armazenar registros de log.
- c) Um formato de arquivo usado para armazenar configurações.
- d) Uma representação visual de um fluxo de trabalho com tarefas e suas dependências.

Questão 40

Valor da questão: 1,00

Qual dos códigos Python listados executa o seguinte procedimento: Dada uma string, verificar se ela contém a palavra "Python"

- a) texto = "Python é uma linguagem de programação"
contem_python = "Python" in texto
- b) texto = "Linguagem de programação"
contem_python = texto.contains("Python")
- c) texto = "Linguagem de programação Python"
contem_python = texto.index("Python")
- d) texto = "Python é uma linguagem de programação"
contem_python = texto.indexOf("Python") != -1