



Prova Objetiva de conhecimentos Específicos - Residência em TI - Business Intelligence e Analytics

Professor(a): Comissão de seleção UFRN

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



Nome:**Matrícula:****Assinatura:** _____**Questão 1***Valor da questão: 0,33*

Considerando a fase de Implementação no CRISP-DM, qual dos seguintes aspectos é crucial para garantir a adoção bem-sucedida dos resultados da mineração de dados?

- a) Re-treinar o modelo periodicamente usando o conjunto de dados original para garantir sua estabilidade.
- b) Desenvolver uma interface de usuário complexa que permita aos usuários finais ajustar os parâmetros do modelo.
- c) Garantir que os resultados da mineração de dados sejam apresentados de maneira compreensível para os stakeholders, incluindo a tradução de insights técnicos em ações de negócios.
- d) Focar exclusivamente na precisão do modelo, assegurando que os resultados da mineração de dados sejam sempre os mais precisos possíveis.

Questão 2*Valor da questão: 0,33*

Considere o esquema relacional abaixo, no qual 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) $\sigma_{placa}(\sigma_{ano = 2023; valor \leq 90000})$
- c) $\pi_{placa}(\sigma_{ano = 2023 \text{ AND } valor \leq 90000}(VEICULO))$
- d) $\sigma_{placa}(\pi_{valor \leq 90000 \text{ AND } ano = 2023}(VEICULO))$

Questão 3*Valor da questão: 0,33*

Considere o seguinte código Python:

```
df = pd.DataFrame({'A': [1, 2, np.nan],  
                  'B': [5, np.nan, np.nan],  
                  'C': [5, 10, 15]})
```

	A	B	C
0	1.0	5.0	5
1	2.0	NaN	10
2	NaN	NaN	15

Qual das linhas de código abaixo é responsável por excluir todos os dados ausentes presentes no DataFrame?

- a) `df.dropna()`
- b) `df.dropna(axis=2)`
- c) `df.dropna(thresh=2)`
- d) `df.dropna(axis=1)`

Questão 4*Valor da questão: 0,33*

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

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

Questão 5

Valor da questão: 0,33

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) malignant, malignant, benign
- c) benign, benign, malignant
- d) benign, malignant, benign

Questão 6

Valor da questão: 0,33

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

Questão 7

Valor da questão: 0,33

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

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

Questão 8

Valor da questão: 0,33

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

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

Questão 9

Valor da questão: 0,33

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

Questão 10

Valor da questão: 0,33

Foi implementado, nas versões mais recentes do Python (3.10 e posteriores), uma estrutura de seleção múltipla, qual das estruturas abaixo representa essa implementação?

- a) SWITCH-CASE
- b) CASE-IN
- c) CASE-OF
- d) MATCH-CASE

Questão 11

Valor da questão: 0,33

Qual é o objetivo principal do Metabase?

- a) Criar websites dinâmicos.
- b) Armazenar grandes volumes de dados.
- c) Visualizar e analisar dados de forma simples e intuitiva.
- d) Gerenciar servidores de banco de dados.

Questão 12

Valor da questão: 0,33

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
- b) K-means
- c) KNN
- d) Naive Bayes

Questão 13

Valor da questão: 0,33

Seja a seguinte sequência de operações de Álgebra Relacional:

$$\pi A3, \pi A4 (\sigma A1 = A2 \text{ (Produto } x \text{ Estoque)})$$

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

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

Questão 14

Valor da questão: 0,33

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

```
x = [5, 4, 3, 2, 1]
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) [3,1,5,4,2]
- b) [3,4,1,2,5]
- c) [4,3,1,5,2]
- d) [4,3,2,1,5]

Questão 15

Valor da questão: 0,33

Analise as seguintes afirmativas:

- I. K-means 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 é um método de Clustering. 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), um modelo completo para a realização de um projeto de DM.

Podemos afirmar corretamente que:

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

Questão 16

Valor da questão: 0,33

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. Sobre este conceito, é correto afirmar que:

- 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.

Sobre as afirmativas acima, é correto dizer que:

- a) Estão corretas as afirmativas II e IV.
- b) Estão corretas as afirmativas I e IV.
- c) Estão corretas as alternativas III e IV.
- d) Estão corretas as alternativas I, III e IV.

Questão 17

Valor da questão: 0,33

Sobre Banco de Dados Relacional, é correto afirmar que:

- I. Que a chave primária é usada para estabelecer relacionamentos entre tabelas.
- II. Que a chave primária é utilizada para identificar unicamente cada registro em uma tabela.
- III. Que a normalização é um processo de remoção de redundância dos dados para melhorar a integridade.

Sobre as afirmativas acima, é correto dizer que:

- a) Estão corretas as afirmativas I e II.
- b) Estão corretas as afirmativas I e III.
- c) Estão corretas as alternativas I, II e III.
- d) Estão corretas as alternativas II e III.

Questão 18

Valor da questão: 0,33

Qual cláusula SQL é usada para agrupar dados em uma consulta?

- a) GROUP BY
- b) MERGE
- c) COMBINE
- d) SORT BY

Questão 19

Valor da questão: 0,33

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

- a) Clustering
- b) Classificação
- c) Associação
- d) Regressão Linear

Questão 20

Valor da questão: 0,33

Qual é a finalidade dos "cards" no Metabase?

- a) São mensagens de erro do sistema.
- b) São cartões de crédito virtuais usados para pagamentos online.
- c) São peças de um quebra-cabeça virtual.
- d) São mini-relatórios ou visualizações individuais de dados.

Questão 21

Valor da questão: 0,33

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

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

Questão 22

Valor da questão: 0,33

Sobre o processamento avançado de linguagem natural e aprendizado profundo:

- I. O aprendizado não supervisionado é inútil para tarefas de PLN.
- II. Redes neurais convolucionais são exclusivamente utilizadas para processamento de imagens e não aplicáveis em PLN.
- III. O BERT (Bidirectional Encoder Representations from Transformers) permite a compreensão contextual de palavras em textos.
- IV. Modelos generativos, como GPT (Generative Pre-trained Transformer), não conseguem gerar textos coerentes e contextuais.

Escolha a(s) afirmação(ões) correta(s)

- a) Estão corretas II e III.
- b) Está correta apenas III.
- c) Estão corretas III e IV.
- d) Estão corretas I e II.

Questão 23

Valor da questão: 0,33

Considerando as técnicas de processamento de linguagem natural para compreensão de texto.

- I. A desambiguação de sentidos de palavras é uma tarefa simples que não requer conhecimento contextual.
- II. Redes neurais recorrentes (RNNs) não podem ser usadas para processar sequências de texto devido à sua natureza linear.
- III. Sistemas do tipo "Question Answering" (QA) baseados em PLN podem extrair respostas de um corpus de texto utilizando compreensão de leitura profunda e inferência.
- IV. A análise de sentimentos é exclusivamente baseada em contagem de palavras positivas e negativas, sem necessidade de modelos complexos.

Escolha a(s) afirmação(ões) correta(s)

- a) Está correta apenas III.
- b) Estão corretas I e II.
- c) Estão corretas III e IV.
- d) Estão corretas II e III.

Questão 24

Valor da questão: 0,33

A respeito das técnicas de otimização e treinamento para arquiteturas de PLN baseadas em aprendizado profundo, qual(is) das seguintes afirmações é/são verdadeira(s)?

- I. Técnicas de otimização como o Gradiente Descendente Estocástico (SGD) não são aplicáveis a modelos baseados em Transformer devido à sua complexidade computacional.
- II. A inicialização dos pesos em arquiteturas profundas de PLN, como os Transformers, não influencia significativamente o desempenho do modelo.
- III. O uso de técnicas de aprendizado profundo tem mostrado ser eficaz na redução da necessidade de recursos computacionais para treinamento de modelos de PLN.
- IV. Regularizações como dropout e normas L2 são cruciais para evitar o overfitting em modelos de PLN altamente parametrizados.

- a) Estão corretas II e III.
- b) Está correta apenas IV.
- c) Estão corretas I e II.
- d) Estão corretas III e IV.

Questão 25

Valor da questão: 0,33

A respeito dos desafios e limitações dos embeddings de palavras, qual(is) das seguintes afirmações é/são correta(s)?

- I. Embeddings de palavras estáticos, como Word2Vec, são capazes de distinguir entre diferentes sentidos de uma palavra com base no contexto de uso.
- II. Uma das principais limitações dos embeddings de palavras é sua incapacidade de capturar relações polissêmicas efetivamente, dado que cada palavra é representada por um único vetor.
- III. A geração de embeddings para palavras fora do vocabulário (OOV) não é trivialmente resolvida por modelos como GloVe e Word2Vec sem a necessidade de técnicas adicionais.
- IV. A incorporação de conhecimento externo, como ontologias ou bases de conhecimento semântico, é comumente realizada em todas as técnicas de embeddings de palavras para enriquecer suas representações.

- a) Estão corretas II e III.
- b) Estão corretas I e II.
- c) Está correta apenas II.
- d) Estão corretas III e IV.

Questão 26

Valor da questão: 0,33

Considere o uso de classes abstratas em Python. Qual das seguintes afirmações é verdadeira ao implementar uma classe abstrata usando o módulo abc?

- a) Uma classe abstrata deve ter pelo menos um método abstrato e não pode conter qualquer implementação de método.
- b) Uma classe abstrata pode ser instanciada diretamente, desde que todos os seus métodos abstratos sejam sobrescritos.
- c) Métodos abstratos em uma classe abstrata podem conter uma implementação completa, que pode ser chamada usando o método `super()` em classes derivadas.
- d) Para definir um método como abstrato, ele deve ser decorado com `@abstractmethod` e a classe deve herdar de `ABC`.

Questão 27

Valor da questão: 0,33

Na duck typing (tipagem do pato, em tradução literal) do Python, o módulo `typing` introduziu o conceito de Protocol para definir uma "interface" que classes podem implementar. O que diferencia o Protocol das Abstract Base Classes (ABCs) em Python?

- a) Protocol é usado exclusivamente para tipagem estática, não afetando o comportamento em tempo de execução, enquanto ABCs impõem restrições de implementação em tempo de execução.
- b) ABCs suportam verificação de tipo em tempo de execução, enquanto Protocol só é verificado em tempo de compilação.
- c) Protocol permite a herança múltipla, enquanto as ABCs não.
- d) Classes podem satisfazer um Protocol simplesmente implementando os métodos e propriedades necessários, sem precisar de herança explícita.

Questão 28

Valor da questão: 0,33

Considerando os Bancos NoSQL e o conceito de consistência eventual, qual das seguintes afirmações é verdadeira?

- a) A consistência eventual aplica-se apenas a bancos de dados de chave-valor e não é relevante para outros tipos de Bancos NoSQL.
- b) A consistência eventual significa que, se nenhuma nova atualização for feita a um dado específico, eventualmente todas as cópias desse dado no sistema irão convergir para o mesmo valor.
- c) A consistência eventual garante que, após uma transação ser concluída, todas as leituras retornarão imediatamente o mesmo valor escrito.
- d) A consistência eventual é um modelo de consistência forte onde as atualizações são instantaneamente visíveis a todos os nós.

Questão 29

Valor da questão: 0,33

No contexto de Bancos NoSQL, o que significa o teorema CAP?

- a) Que um sistema de banco de dados pode fornecer Consistência, Acessibilidade e Particionamento, tudo ao mesmo tempo.
- b) Que um sistema de banco de dados pode fornecer Consistência, Disponibilidade e Tolerância a Partições, mas apenas duas dessas podem ser garantidas simultaneamente.
- c) Que um sistema de banco de dados é inerentemente incapaz de escalar além de um certo ponto devido a restrições computacionais.
- d) Que um sistema de banco de dados deve priorizar CPU, Armazenamento e Performance acima de tudo.

Questão 30

Valor da questão: 0,33

No contexto de aprendizado de máquina, por que o monitoramento contínuo de modelos em produção é crítico?

- a) Para identificar a deriva dos dados (data drift) e a degradação do modelo, garantindo que o modelo continue com desempenho conforme esperado ao longo do tempo.
- b) Para detectar e corrigir automaticamente bugs no código de ML sem intervenção humana.
- c) Para garantir que os modelos continuem operando dentro dos limites de uso de recursos de computação e armazenamento.
- d) Exclusivamente para monitorar a latência de inferência e otimizar o tempo de resposta do modelo.