

-- CONHECIMENTOS ESPECÍFICOS --

Julgue os próximos itens, relativos a aprendizado de máquina.

- 76** O algoritmo *k-means* é uma técnica de agrupamento por partição que visa encontrar *k* grupos disjuntos de observações, na qual a escolha inicial dos *k* centroides é feita aleatoriamente; o algoritmo itera ajustando esses centroides com base na similaridade das observações, até convergir para uma solução final, que define as partições.
- 77** O algoritmo *k-NN* é um método de aprendizado baseado em instâncias que classifica novos exemplares por meio de um modelo previamente treinado, o que dispensa a necessidade de armazenar os dados de treinamento para futuras comparações.
- 78** No algoritmo *random forest*, cada árvore de decisão é treinada com um subconjunto diferente de variáveis, mas todas utilizam exatamente o mesmo conjunto de observações, para garantir que a variância do modelo seja reduzida sem comprometer o viés.
- 79** O algoritmo de *backpropagation*, utilizado para treinar redes neurais, calcula o erro na saída da rede e ajusta os pesos da rede aleatoriamente, para minimizar o erro na próxima iteração.
- 80** Considere que, em um modelo de regressão linear ajustado para prever a produção de soja (*Y*, em sacas por hectare) com base na quantidade de fertilizante aplicado (*X*, em kg/ha), a equação estimada foi $Y = 30 + 0,8X$. Nesse caso, o coeficiente angular dessa equação indica que, ao aplicar 10 kg/ha a mais de fertilizante, espera-se um aumento de 8 sacas por hectare na produção de soja, e, sem aplicação de fertilizante, a produção esperada de soja será de 30 sacas por hectare.

No que se refere à linguagem Python, julgue os itens a seguir.

- 81** Se, no código a seguir, a variável `producao` armazena a produção, em toneladas, de diferentes produtos agrícolas, então essa variável é do tipo dicionário e, após a execução do código, a soma da produção de milho e de soja será de 2300 toneladas, e a chave `trigo` não existirá mais em `producao`.

```
producao = {
    "milho":1200,
    "soja":800,
    "trigo":500,
    "algodao":300}

producao["milho"] += producao.get("trigo",0)
producao["soja"] -= producao.get("algodao",0)
producao["batata"] = producao.pop("trigo")*2

print(producao)
```

- 82** Considere o seguinte código, que armazena dados de produção de café de certa região.

```
fazendas = ["Fazenda A", "Fazenda B",
            "Fazenda A", "Fazenda B"]
safras = [2023, 2023, 2024, 2024]
quantidades_kg = [1500, 2300, 1800, 2100]
dados = [
    ["Fazenda", "Safr", "Quantidade_ kg"],
    [fazendas[0], safras[0],
     quantidades_kg[0]],
    [fazendas[1], safras[1],
     quantidades_kg[1]],
    [fazendas[2], safras[2],
     quantidades_kg[2]],
    [fazendas[3], safras[3],
     quantidades_kg[3]],
    ]
```

Nessas condições, é correto afirmar que a execução do comando `print(dados[3][2])` retornará o valor 2300.

- 83** O código apresentado a seguir utiliza a estrutura de dados do tipo conjuntos e, após a sua execução, a variável `culturas` conterá {"Café", "Soja"} e a variável `total_culturas` terá 2 como valor.

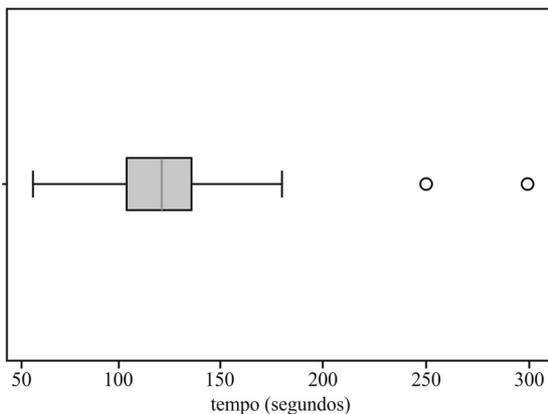
```
def culturas_comuns(fazenda1, fazenda2):
    return fazenda1.intersection(fazenda2)
fazenda_a = {"Milho", "Soja", "Café",
            "Trigo"}
fazenda_b = {"Café", "Arroz", "Soja",
            "Feijão"}
culturas = culturas_comuns(fazenda_a,
                          fazenda_b)
total_culturas = len(culturas)
print("Culturas comuns:", culturas)
print("Total de culturas comuns:",
      total_culturas)
```

Acerca de governança de dados, julgue os itens que se seguem.

- 84** De acordo com o DAMA-DMBOK (2017), a governança de dados é responsável por garantir a execução direta das atividades de gerenciamento de dados, inclusive a aplicação de métricas, políticas e ferramentas para assegurar sua qualidade e conformidade.
- 85** A linhagem de dados (*data lineage*) refere-se ao rastreamento completo do ciclo de vida dos dados, com entradas, saídas, regras aplicadas e transformações realizadas, e é útil para a depuração de problemas em fluxos de processamento.
- 86** O CDO (*chief data officer*) é responsável por definir e executar estratégias para a governança, a qualidade e o uso dos dados em diversas áreas, experimentos científicos, gestão operacional e tomada de decisão estratégica, o que garante que as informações sejam organizadas e utilizadas de forma eficiente para gerar inovação e valor para a instituição.
- 87** A distinção entre os mecanismos de ausência de dados é essencial para a análise da qualidade de dados; no caso de uma ausência classificada como MAR (*missing at random*), a probabilidade de um valor estar ausente depende exclusivamente da própria variável ausente, mas não de outras variáveis do conjunto de dados.
- 88** No mecanismo MCAR (*missing completely at random*), a ausência de um valor ocorre de maneira completamente aleatória, ou seja, a probabilidade de ausência não depende da própria variável nem de qualquer outro atributo do conjunto de dados.

- 89 Se, nos dados de criação de suínos, um suíno foi registrado com 100 kg de peso e 2 meses de idade, então, como os valores registrados são incomuns para a espécie, tem-se, nesse caso, um exemplo de valores inconsistentes; essa discrepância pode indicar erro de preenchimento, como entrada incorreta de dados, ou a manipulação intencional para alterar informações e deve ser corrigida com técnicas como a remoção do registro inconsistente ou a substituição do peso por uma média calculada a partir de outros suínos da mesma idade.
- 90 No método do *Z-score*, um valor é considerado um *outlier* se seu *Z-score* for superior a 3 ou inferior a -3 .
- 91 Considere que, na aplicação de técnicas de anonimização em um estudo acerca da influência de fatores na produtividade de lavouras de soja, os nomes e os números de inscrição dos agricultores foram substituídos por códigos numéricos aleatórios, em atendimento ao princípio de anonimização de identificadores; e que as coordenadas geográficas foram alteradas para microrregiões, e informações, como tipo de solo e variedade de soja, foram agrupadas em categorias amplas, respeitado o princípio de anonimização de atributos. Nessas condições, é correto afirmar que foram adotadas medidas que protegem a privacidade dos agricultores, enquanto permitem a análise dos dados.
- 92 A transformação de valores numéricos de um atributo quantitativo para outro valor numérico pode ser realizada por meio de funções matemáticas simples, como logarítmica, inverso e módulo, porém, essa transformação só pode ser feita por meio da normalização escalar, isto é, a padronização (*Z-score*) é inadequada para esse fim.

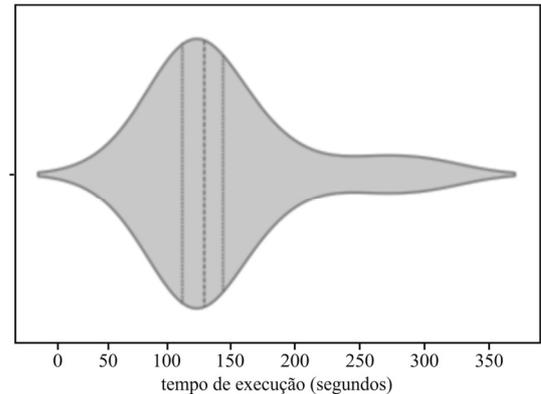
Para o treinamento de um modelo de *machine learning*, foram realizados 16 experimentos, cujos tempos de execução (em segundos) registrados foram os seguintes: 100, 111, 123, 100, 178, 101, 102, 117, 300, 121, 123, 124, 165, 250, 55, 104. A partir desse *dataset*, foi plotado o gráfico a seguir.



Com base nessas informações, julgue os próximos itens.

- 93 O diagrama *box plot* apresentado indica que 50% dos tempos de execução estão abaixo de 119 s, o que corresponde à mediana.
- 94 O gráfico permite inferir que não há *outliers*; logo, todos os valores estão dentro da média esperada para os dados.

- 95 Considerando o gráfico de violino para o *dataset* em questão, mostrado a seguir, é correto afirmar que esse gráfico pode ser substituído por um diagrama de *box plot*, sem perda de informações importantes sobre a distribuição dos dados.



- 96 As informações fornecidas indicam que a presença de *outliers* no *dataset* influencia mais a mediana que a média.
- 97 As informações fornecidas indicam que o primeiro quartil corresponde a 101,25 s e o terceiro quartil, a 154,75 s.

No que se refere a inteligência artificial, matemática computacional e computação aplicada, julgue os itens subsequentes.

- 98 *Heaps* são estruturas de dados que permitem operações eficientes de inserção, remoção e acesso; seu uso pode melhorar significativamente a eficiência de algoritmos de ordenação.
- 99 O modelo de redes neurais convolucionais (CNNs) elimina totalmente a necessidade de pré-processamento dos dados de imagem, o que é uma das vantagens para a sua utilização na agropecuária.
- 100 A notação Big O é crucial para avaliar a eficiência de técnicas de processamento de dados, pois estabelece tanto o limite superior quanto o limite inferior para o crescimento do tempo de execução ou consumo de recursos, independentemente do tamanho da entrada.
- 101 As CNNs podem ser aplicadas na agricultura digital para a análise de imagens obtidas por meio de veículos aéreos não tripulados (VANTs), o que permite a detecção automática de pragas e doenças.

Julgue os itens que se seguem, acerca de modelagem de dados e de planejamento estatístico amostral e experimental.

- 102 A extração de características transforma os dados brutos em informações otimizadas para a construção de modelos preditivos.
- 103 Embora haja o risco de eliminar variáveis que contenham informações relevantes, a seleção de atributos contribui para a redução da complexidade dos modelos e pode, em alguns casos, melhorar a generalização ao reduzir o *over fitting* e eliminar totalmente os vieses.
- 104 A seleção de atributos em um modelo preditivo pode reduzir o *over fitting* e, ao mesmo tempo, garantir menor variância sem impactar o viés, desde que os atributos removidos não sejam altamente correlacionados com a variável resposta.
- 105 O método de reamostragem *bootstrap* melhora a estimativa do desempenho de um modelo ao criar múltiplas amostras de treinamento a partir dos dados originais, o que reduz a variância das estimativas e proporciona intervalos de confiança mais precisos para métricas de avaliação.
- 106 O princípio da aleatorização em experimentos estatísticos é essencial para reduzir viés e garantir que diferenças observadas entre grupos possam ser atribuídas ao tratamento aplicado, em vez de a fatores externos não controlados.

107 Em um experimento de comparação entre múltiplos tratamentos, o uso do teste de hipóteses para comparar médias entre grupos tratados deve ser acompanhado de ajustes para múltiplas comparações, a fim de controlar o risco de falsos positivos.

A respeito de sistemas para rastreabilidade e protocolos para certificação e diferenciação de processos de produção e produtos de origem animal, julgue os itens a seguir.

108 A adoção de sistemas digitais para rastreabilidade e protocolos de certificação contribui para a diferenciação competitiva dos produtos de origem animal no mercado.

109 A incorporação da rastreabilidade digital como um processo de rotina na cadeia produtiva de produtos de origem animal enfrenta desafios significativos, incluindo a necessidade de maior engajamento dos setores produtivos, bem como a conscientização dos consumidores.

110 A integração de diversos sistemas de rastreabilidade em uma cadeia produtiva unificada elimina os desafios de interoperabilidade e a necessidade de adaptações regionais.

Espaço livre