



SIMULADO ESPECIAL

PETROBRAS

Ciência de Dados - Pré-Edital

Caderno de Prova

Nome: _____

INFORMAÇÕES SOBRE O SIMULADO

- 1 - Este simulado conta com questões focadas no concurso **Petrobras**, cargo de **Ciência de Dados**;
- 2 - A prova contém itens que abordam conhecimentos cobrados no edital do concurso;
- 3 - As questões são inéditas e foram elaboradas pelos nossos professores com base no perfil da banca organizadora;
- 4 - A participação no ranking classificatório só será permitida até o horário de liberação do gabarito;
- 5 - O link para preencher o formulário com seu gabarito está localizado logo após estas instruções;

PREENCHA SEU GABARITO

Clique no link, ou copie e cole no seu navegador, para preencher seu gabarito.

<https://forms.gle/e8WEmhGe5nseo1vZ6>

- | | | | | |
|------------|------------|------------|------------|------------|
| 01 - ABCDE | 15 - ABCDE | 29 - ABCDE | 43 - ABCDE | 57 - ABCDE |
| 02 - ABCDE | 16 - ABCDE | 30 - ABCDE | 44 - ABCDE | 58 - ABCDE |
| 03 - ABCDE | 17 - ABCDE | 31 - ABCDE | 45 - ABCDE | 59 - ABCDE |
| 04 - ABCDE | 18 - ABCDE | 32 - ABCDE | 46 - ABCDE | 60 - ABCDE |
| 05 - ABCDE | 19 - ABCDE | 33 - ABCDE | 47 - ABCDE | 61 - ABCDE |
| 06 - ABCDE | 20 - ABCDE | 34 - ABCDE | 48 - ABCDE | 62 - ABCDE |
| 07 - ABCDE | 21 - ABCDE | 35 - ABCDE | 49 - ABCDE | 63 - ABCDE |
| 08 - ABCDE | 22 - ABCDE | 36 - ABCDE | 50 - ABCDE | 64 - ABCDE |
| 09 - ABCDE | 23 - ABCDE | 37 - ABCDE | 51 - ABCDE | 65 - ABCDE |
| 10 - ABCDE | 24 - ABCDE | 38 - ABCDE | 52 - ABCDE | 66 - ABCDE |
| 11 - ABCDE | 25 - ABCDE | 39 - ABCDE | 53 - ABCDE | 67 - ABCDE |
| 12 - ABCDE | 26 - ABCDE | 40 - ABCDE | 54 - ABCDE | 68 - ABCDE |
| 13 - ABCDE | 27 - ABCDE | 41 - ABCDE | 55 - ABCDE | 69 - ABCDE |
| 14 - ABCDE | 28 - ABCDE | 42 - ABCDE | 56 - ABCDE | 70 - ABCDE |

SIMULADO NO SISTEMA DE QUESTÕES

Clique no link, ou copie e cole no seu navegador, para fazer este simulado também no SQ!

<https://bit.ly/4uVfuqL>

CONHECIMENTOS GERAIS**LÍNGUA PORTUGUESA***Fabício Dutra*

Era um comício na praça central da cidade. O prefeito, de peito estufado e voz inflamada, discursava diante de uma multidão cansada, mas ainda acostumada a ouvir promessas. Exaltava, orgulhoso, a retomada do crescimento, os novos projetos urbanos, as cores da bandeira, o hino municipal e o escambau. Atrás dele, tremulavam faixas patrióticas; ao redor, vendedores ambulantes disputavam espaço com fotógrafos e curiosos. Alguns aplaudiam sem convicção. Outros apenas observavam em silêncio, como quem já perdera a capacidade de acreditar.

De repente, o chão começou a tremer.

Primeiro, um leve estremecimento. Depois, um abalo violento que interrompeu o discurso e espalhou pânico pela praça. Os postes balançaram. As vidraças vibraram. Uma enorme fenda abriu-se no chão, rasgando o asfalto diante do palanque como se a própria cidade estivesse sendo partida ao meio. Mulheres gritaram. Crianças correram. Muitos pensaram que era a sétima trombeta do Apocalipse e se afastaram em desespero, atropelando bancos, bandeiras e caixas de som.

A guarda municipal tentou conter o caos. Isolou a área, afastou os curiosos e acalmou os ansiosos com ordens firmes e promessas vagas. O prefeito, agora pálido e sem a imponência de minutos antes, recolheu-se atrás de seus assessores. Enquanto isso, os sábios da cidade — homens antigos, conhecidos por aconselharem governos em momentos extremos — reuniram-se rapidamente sob a marquise da prefeitura. Conversaram em voz baixa, trocaram olhares graves e, após longos minutos, aproximaram-se do prefeito.

A conclusão parecia absurda, mas foi anunciada com solenidade: era necessário que um munícipe corajoso saltasse voluntariamente para o abismo. Somente assim, disseram eles, a fenda fechar-se-ia e a cidade seria poupada de uma tragédia ainda maior.

O silêncio espalhou-se pela praça.

Ninguém se moveu.

Os aplausos haviam desaparecido. As palavras patrióticas também. Restavam apenas o medo, o vento que saía das profundezas da fenda e o desconforto de

perceber que, diante do sacrifício verdadeiro, todo discurso se tornava pequeno.

Foi então que um homem se levantou no meio da multidão.

Chamava-se Marcus Lucius. Não era político, militar nem celebridade. Era apenas um cidadão comum. Caminhou lentamente até a borda do abismo, enquanto dezenas de olhos o acompanhavam em absoluto silêncio. Não demonstrava coragem teatral nem desejo de glória. Havia apenas firmeza em seu rosto.

Antes do salto, voltou-se para a multidão. Observou o prefeito, os guardas, os sábios e o povo que lotava a praça. Então disse, com voz serena:

— Façam por merecer.

E simplesmente saltou.

Seu corpo desapareceu na escuridão profunda da fenda. Durante alguns segundos, ninguém respirou. Em seguida, o chão voltou a tremer. As pedras moveram-se lentamente, como placas gigantescas empurradas por forças invisíveis. A enorme rachadura começou a fechar-se até desaparecer completamente, deixando apenas uma cicatriz irregular no asfalto da praça.

Marcus Lucius nunca mais foi visto.

Durante muitos dias, a cidade permaneceu em silêncio. Os discursos diminuíram. As bandeiras perderam parte do brilho. E muitos passaram a evitar aquela praça, talvez porque ela lembrasse a todos uma verdade incômoda: cidades não são sustentadas por palavras grandiosas, mas pelos poucos que ainda estão dispostos a sacrificar algo por elas.

1. Com base na conclusão do conto, qual é a crítica central direcionada à postura das autoridades e da sociedade em momentos de crise?
 - a) O egoísmo intrínseco da multidão, que prefere a destruição da cidade a permitir que um cidadão comum se torne um herói.
 - b) A incapacidade técnica dos sábios em prever fenômenos geológicos, recorrendo a soluções místicas por falta de conhecimento científico.
 - c) A ineficácia da retórica patriótica e institucional diante da necessidade de ações concretas que exigem renúncia pessoal.
 - d) A fragilidade das instituições democráticas, que colapsam imediatamente quando confrontadas por eventos de ordem sobrenatural.
 - e) A população da cidade demonstrou plena confiança nas instituições públicas, mantendo-se organizada e cooperativa durante toda a crise, sem demonstrar medo ou hesitação diante do sacrifício exigido.
2. No que diz respeito à progressão textual, como o autor constrói a transição entre o início do comício e o desfecho da narrativa?
 - a) Através de uma gradação descendente, em que o fervor inicial da multidão se transforma em apatia total e esquecimento histórico.
 - b) Mediante o uso de flashbacks que explicam a origem da fenda, conectando o passado da cidade ao comportamento do prefeito.
 - c) Por meio de uma ruptura abrupta que desloca o foco da grandiloquência política para a crueza de uma realidade que exige um preço tangível.
 - d) Utilizando uma estrutura cíclica, na qual a cidade retorna ao estado de normalidade absoluta, ignorando o evento ocorrido.
 - e) Por meio de uma narrativa predominantemente descritiva, que privilegia a ambientação da praça em detrimento do conflito central.
3. Quanto às regras de pontuação, assinale a alternativa que esteja incorreta:
 - a) Enquanto o prefeito exaltava o progresso da cidade diante da multidão reunida na praça, muitos habitantes permaneciam em silêncio, porque já não acreditavam, completamente nas promessas repetidas em todos os comícios.
 - b) Assim que a enorme fenda se abriu no centro da praça e o chão começou a tremer violentamente, os discursos patrióticos perderam a força, pois a população percebeu que a crise exigia atitudes concretas.
 - c) Marcus Lucius caminhou lentamente até a borda do abismo e encarou a população com serenidade, porque compreendia que palavras grandiosas não sustentam uma sociedade quando surgem momentos extremos.
 - d) Embora os sábios afirmassem que o sacrifício de um cidadão salvaria toda a cidade, ninguém se ofereceu imediatamente, de modo que o silêncio da multidão revelou o medo coletivo diante do perigo.
 - e) Depois que a fenda se fechou e a praça voltou ao aparente estado de normalidade, muitos moradores passaram a evitar aquele lugar, já que a lembrança do sacrifício expunha a fragilidade moral da própria cidade.

4. Quanto às normas de concordância, assinale a alternativa correta:
- a) Apesar de os discursos das autoridades parecer convincente durante o comício na praça central, ainda surgiam entre os moradores dúvidas e críticas que revelavam a fragilidade daquela aparente confiança coletiva.
 - b) Quando se ouviu os primeiros estrondos vindos do subsolo da cidade e a multidão começou a correr desesperadamente, tornou-se evidente a ausência de líderes preparados para enfrentar uma situação extrema.
 - c) Nem o prefeito nem os sábios da cidade conseguiu tranquilizar completamente a população assustada, porque já circulavam entre os habitantes rumores de que a tragédia ainda poderia aumentar.
 - d) Após o desaparecimento de Marcus Lucius no interior da fenda que se abriu no chão da praça, permaneceram nos rostos dos presentes expressões de culpa e espanto diante daquele sacrifício inesperado.
 - e) Embora houvessem na cidade pessoas influentes e discursos repletos de patriotismo exagerado, faltavam às autoridades coragem e disposição para assumir responsabilidades em momentos decisivos.
5. No fragmento “Enquanto isso, os sábios da cidade... reuniram-se rapidamente”, o conectivo em destaque estabelece uma relação de:
- a) Simultaneidade, situando a ação dos sábios no mesmo tempo cronológico em que a guarda municipal tentava conter o caos.
 - b) Consequência, apresentando a deliberação dos sábios como o resultado direto da palidez do prefeito.
 - c) Concessão, indicando que a reunião dos sábios ocorreu apesar do caos e do pânico generalizado.
 - d) Finalidade, expressando o objetivo dos sábios de aconselhar o governo em momentos de extrema urgência.
 - e) Explicação, introduzindo um esclarecimento acerca das razões que levaram a população a abandonar a praça.
6. Com relação às normas de concordância nominal, assinale a alternativa correta.
- a) A multidão permaneceu meio assustadas depois que enormes rachaduras surgiram no centro da praça.
 - b) Os guardas municipais mantiveram fechada as entradas da praça durante o momento de maior tensão.
 - c) Seguem anexa ao relatório as declarações dos sábios da cidade sobre o ocorrido na praça.
 - d) Os discursos patrióticos e as promessas vazias pareciam bastante inadequados diante da gravidade da situação.
 - e) Era necessário atitudes rápidas e decisões firmes para impedir o avanço do pânico coletivo.
7. No fragmento “...pelos poucos que ainda estão dispostos a sacrificar algo por elas”, a palavra “que” é classificada como pronome relativo. Qual a função sintática que esse termo desempenha na oração subordinada?
- a) Objeto direto
 - b) Sujeito
 - c) Adjunto adnominal
 - d) Complemento nominal
 - e) Predicativo do sujeito
8. Assinale a alternativa em que a lacuna deve ser preenchida com ‘à’.
- a) A guarda municipal agiu ___ pressas para conter os curiosos da praça.
 - b) Marcus Lucius permaneceu frente ___ todos antes de saltar no abismo.
 - c) O prefeito dirigiu-se ___ multidão com discursos inflamados durante o comício.
 - d) Os sábios fizeram referência ___ essa antiga profecia da cidade.
 - e) Muitos habitantes começaram ___ desconfiar das promessas das autoridades.

9. Com relação aos mecanismos de coesão utilizados no texto, assinale a alternativa correta.
- a) O pronome “eles”, em “façam por merecer”, retoma exclusivamente os sábios da cidade, responsáveis pela decisão do sacrifício.
 - b) A expressão “Enquanto isso” estabelece uma relação de oposição entre a ação da guarda municipal e a reunião dos sábios.
 - c) O pronome “que”, em “a fenda fechar-se-ia”, introduz uma oração subordinada adjetiva explicativa referente ao termo “abismo”.
 - d) O termo “assim”, em “Somente assim, disseram eles, a fenda fechar-se-ia”, funciona como elemento coesivo que retoma a ideia do sacrifício voluntário mencionado anteriormente.
 - e) A repetição da palavra “cidade” ao longo do texto provoca prejuízo coesivo, tornando a narrativa semanticamente incoerente.
10. Em relação à colocação dos pronomes oblíquos átonos, assinale a alternativa correta.
- a) O povo tinha calado-se após o desaparecimento de Marcus Lucius.
 - b) Os guardas não continham-se diante do avanço do caos.
 - c) Todos haviam impressionado-se com a coragem do cidadão.
 - d) A multidão se afastou em desespero após a abertura da fenda.
 - e) Os sábios tinham reunido-se rapidamente sob a marquise.

LÍNGUA INGLESA

Adolfo Sá

Lula backs responsible oil exploration in Brazil's Equatorial Margin

President Luiz Inácio Lula da Silva advocated on Monday (May 18) for oil exploration in the Amazon River Mouth Basin, part of the Equatorial Margin. He stressed that the activity must be carried out responsibly to avoid environmental damage.

“No one cares more about the Amazon than we do,” Lula said, referring to his government.

The president made the remarks during a visit to Paulínia (Replan), Petrobras’ largest refinery, in the state of São Paulo. **He argued that exploration is also important for national sovereignty and to prevent other countries from encroaching on the area.**

“Before long, Trump will think it’s his and go there. He thought Canada was his, he thought Greenland was his. He thought the Gulf of Mexico was his. Who’s to say he won’t think the Equatorial Margin is his too? So we’re going to occupy the area and explore for oil with the utmost responsibility, ensuring that this money can be reinvested to secure the country’s future,” Lula noted.

Last year, Petrobras obtained authorization from the Brazilian Institute of the Environment and Renewable Natural Resources (Ibama) to begin exploratory drilling operations in the Equatorial Margin. The region, in northern Brazil, is considered a new pre-salt frontier because of its oil potential.

Lula argued that Petrobras should be regarded as a national asset and must not be privatized. According to him, Brazilians would feel the financial impact of the war in the Middle East even more acutely if the company were already privately owned.

“Petrobras is earning more from oil exports, and prices have risen because of the war in Iran. So we are collecting part of this additional revenue through taxes on oil exports to subsidize diesel and gasoline prices, ensuring that the burden does not fall on Brazilians, truck drivers, or motorists. We are using government funds to prevent this cost from being passed on to the Brazilian people, who are not to blame for the war in Iran. The war in Iran is Trump’s fault,” the president stated.

<https://agenciabrasil.ebc.com.br/en/economia/noticia/2026-05/lula-backs-responsible-oil-exploration-brazils-equatorial-margin>

11. The main purpose of the text is to

- a) criticize the environmental policies of the Brazilian government.
- b) present Lula's defense of responsible oil exploration in the Equatorial Margin.
- c) compare Petrobras' performance to that of private oil companies.
- d) explain how the war in Iran affects fuel prices worldwide.
- e) announce that Ibama authorized exploratory drilling in the region.

12. In the sentence *"He argued that exploration is also important for national sovereignty and to prevent other countries from encroaching on the area"*, the word *encroaching* is closest in meaning to

- a) trespassing.
- b) retreating.
- c) financing.
- d) ignoring.
- e) protecting.

13. In *"So we're going to occupy the area and explore for oil with the utmost responsibility"*, the word *utmost* could be replaced, without change in meaning, by

- a) minimal.
- b) maximum.
- c) partial.
- d) questionable.
- e) gradual.

14. According to the text, Lula's argument for oil exploration includes the idea that

- a) Petrobras should be privatized to increase efficiency.
- b) the Amazon region has no environmental significance.

c) exploration is a matter of national sovereignty.

d) oil revenues will not benefit the Brazilian population.

e) the war in Iran has lowered oil prices globally.

15. The word *"acutely"* in *"Brazilians would feel the financial impact of the war in the Middle East even more acutely if the company were already privately owned"* means

- a) mildly.
- b) slowly.
- c) intensely.
- d) rarely.
- e) vaguely.

16. In the sentence *"So we are collecting part of this additional revenue through taxes on oil exports to subsidize diesel and gasoline prices, ensuring that the burden does not fall on Brazilians"*, the word *"so"* expresses

- a) contrast.
- b) time.
- c) condition.
- d) conclusion.
- e) addition.

17. According to the final paragraph, the Brazilian government is

- a) allowing oil prices to rise freely without intervention.
- b) taxing oil exports to prevent fuel price increases from reaching consumers.
- c) blaming the war in Iran for its own economic mismanagement.
- d) reducing Petrobras' profits by imposing strict environmental regulations.
- e) privatizing Petrobras to attract foreign investment.

18. In the sentence *"He stressed that the activity must be carried out responsibly to avoid environmental damage"*, the verb stressed could NOT be replaced by:

- a) Emphasized
- b) Underlined
- c) Highlighted
- d) Reiterated
- e) Suggested

19. In the sentence *"No one cares more about the Amazon than we do"*, the verb DO is used as:

- a) an auxiliary verb for negation.
- b) an auxiliary verb for question formation.
- c) a substitute verb avoiding repetition of "care".
- d) an emphatic auxiliary to intensify the statement.
- e) a main verb meaning "to perform".

20. In the sentence *"Who's to say he won't think the Equatorial Margin is his too? So we're going to occupy the area"*, the word SO expresses:

- a) Contrast.
- b) Addition.
- c) Purpose.
- d) Conclusion.
- e) Condition.

CONHECIMENTOS ESPECÍFICOS

BLOCO I

Lucas Ianni

21. Uma equipe de ciência de dados de uma prefeitura precisa segmentar os munícipes com perfis socioeconômicos semelhantes para direcionar políticas públicas, sem possuir qualquer rotulação prévia dos grupos. Para essa tarefa, a abordagem adequada é o aprendizado

- a) supervisionado por regressão, voltado à previsão de valores numéricos.
- b) supervisionado por classificação binária, restrito à atribuição de duas categorias.
- c) por reforço, com recompensas e punições atribuídas pelo ambiente.
- d) por transferência transdutiva, com adaptação entre domínios distintos.
- e) não supervisionado por clusterização, voltado à formação de grupos com elementos similares.

22. Em uma análise pós-treinamento, um cientista de dados verifica que seu modelo apresenta taxa de erro extremamente baixa no conjunto de treinamento e taxa de erro significativamente mais alta no conjunto de teste. Esse comportamento indica que o modelo está sofrendo de

- a) underfitting, caracterizado por alto viés e baixa capacidade de aprender padrões nos próprios dados de treino.
- b) overfitting, caracterizado por alta variância, em que o modelo memorizou ruídos e peculiaridades dos dados de treino e não generaliza para novos dados.
- c) vazamento de dados, problema em que dados de produção contaminam o conjunto de treino, gerando avaliação pessimista.
- d) subajuste de hiperparâmetros, em que a função de aquisição falhou em selecionar parâmetros adequados.
- e) baixa cardinalidade da amostra, problema exclusivo de modelos paramétricos com função fixa.

23. Sobre os métodos de avaliação e validação de modelos em aprendizado supervisionado, analise as afirmativas a seguir.

- I. Na resubstituição, o mesmo conjunto de dados é utilizado para treinar e avaliar o modelo, o que tende a superestimar o desempenho real, pois o modelo pode simplesmente memorizar as observações em vez de aprender padrões generalizáveis.
- II. Na validação cruzada k-fold, o conjunto de dados é dividido em k partições; em cada iteração, k-1 folds são utilizados para treino e o fold restante para teste, repetindo-se o processo k vezes e calculando-se a média das métricas.
- III. A validação cruzada leave-one-out (LOOCV) é um caso particular da k-fold em que k é igual ao número total de observações, sendo computacionalmente barata e indicada para conjuntos de dados muito grandes.
- IV. No bootstrap, geram-se amostras aleatórias com reposição a partir do conjunto original; as observações não selecionadas para o conjunto de treino (out-of-bag) podem ser utilizadas no conjunto de teste.

Estão corretas apenas as afirmativas

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

24. Em um projeto com elevado custo computacional para treinamento de cada combinação de hiperparâmetros, a equipe deseja adotar uma estratégia de busca que utilize um modelo probabilístico das relações entre conjuntos de hiperparâmetros e desempenho do modelo, aproveitando os resultados das iterações anteriores para decidir, de forma orientada, qual o próximo conjunto a ser avaliado. A técnica adequada para esse cenário é

- a) busca em grade (grid search), por testar exaustivamente todas as combinações possíveis dos hiperparâmetros.
- b) busca aleatória (random search), por amostrar combinações de forma totalmente independente entre as iterações.
- c) otimização bayesiana, que constrói progressivamente um modelo probabilístico da função desempenho e utiliza uma função de aquisição para selecionar o próximo conjunto de hiperparâmetros a testar.
- d) validação cruzada estratificada, que garante proporção de classes nos folds durante a avaliação do modelo.
- e) bootstrap, técnica de reamostragem com reposição que estima a variabilidade do desempenho em diferentes amostras.

25. Um analista deseja classificar novos dados com base em sua similaridade com exemplares já rotulados, atribuindo a cada novo dado a categoria predominante entre seus vizinhos mais próximos no espaço de características.

O algoritmo de aprendizado de máquina descrito é:

- a) Árvore de Decisão, pois divide os dados em ramos com base no ganho de informação dos atributos.
- b) K-Means, pois agrupa os dados em K clusters de forma não supervisionada.
- c) SVM, pois maximiza a margem entre as classes por meio de um hiperplano de separação.
- d) K-Nearest Neighbors, pois utiliza a proximidade aos exemplos rotulados para definir a classe da nova amostra.
- e) Naive Bayes, pois calcula a probabilidade condicional dos atributos dado cada classe possível.

26. Uma equipe de ciência de dados precisa construir um modelo de árvore de decisão capaz de lidar diretamente com atributos numéricos contínuos, sem necessidade de discretização manual prévia, e que aplique poda após a construção da árvore para reduzir o sobreajuste aos dados de treinamento.

O algoritmo de construção de árvore que atende simultaneamente a essas duas necessidades é:

- a) C4.5, pois introduz pontos de corte automáticos em variáveis contínuas e realiza poda da árvore após sua construção.
- b) ID3, pois utiliza ganho de informação como critério e aceita nativamente qualquer tipo de atributo numérico contínuo.
- c) Random Forest, pois é o único algoritmo de árvore que realiza poda nos seus ramos após o treinamento.
- d) KNN, pois trata cada nó da árvore como um vizinho próximo no cálculo da distância euclidiana.
- e) DBSCAN, pois agrupa atributos contínuos em regiões densas antes da construção dos nós internos.

27. Em um problema de classificação binária, um cientista de dados verificou que seu conjunto de treinamento não é linearmente separável no espaço original de características. Para continuar utilizando Máquinas de Vetores de Suporte, ele aplicou uma técnica que permite obter uma fronteira de decisão não linear, computando produtos internos em um espaço de maior dimensionalidade sem precisar calcular explicitamente as novas coordenadas dos dados.

A técnica descrita é conhecida como:

- a) hard margin, abordagem que exige separação perfeita das classes no espaço original.
- b) soft margin, mecanismo que permite a violação da margem por algumas observações dos dados.
- c) método dos mínimos quadrados, utilizado para minimizar a soma dos resíduos quadrados do modelo.

d) discretização de atributos, etapa que converte variáveis numéricas em categorias antes do treinamento.

e) kernel trick, que calcula implicitamente o produto interno no espaço de alta dimensão sem mapear as coordenadas.

28. Analise as afirmativas a seguir sobre o algoritmo Naive Bayes aplicado a problemas de classificação:

I. A suposição de independência condicional entre os atributos, dado o valor da classe, permite simplificar o cálculo da probabilidade conjunta em um produto de probabilidades individuais.

II. O algoritmo dispensa o uso do Teorema de Bayes em sua formulação, baseando-se exclusivamente nas frequências marginais dos atributos no conjunto de treinamento.

III. Apesar de a suposição de independência raramente se verificar na prática, o Naive Bayes apresenta bom desempenho em diversos cenários reais, sendo aplicação clássica na filtragem de spam.

IV. Por ser um classificador discriminativo, o Naive Bayes modela diretamente a fronteira de decisão entre as classes, sem necessidade de estimar probabilidades a priori das categorias.

Estão corretas apenas:

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

29. Um analista realizou uma regressão linear simples relacionando o número de horas semanais de exercício físico, variável independente x , ao percentual de redução de pressão arterial após três meses, variável dependente y . A equação ajustada obtida foi:

$$\hat{y} = 1,2 + 0,8x$$

Com base nesse modelo, é correto afirmar que:

- a) Para cada hora adicional de exercício semanal, espera-se uma redução de 1,2 pontos percentuais na pressão arterial, conforme indicado pelo coeficiente angular do modelo.
- b) O valor 0,8 representa o intercepto da reta, ou seja, o valor previsto de y quando a variável x é igual a zero.
- c) A cada hora adicional de exercício semanal, espera-se um aumento de 0,8 ponto percentual na redução prevista da pressão arterial.
- d) O coeficiente 1,2 representa a tangente do ângulo formado pela reta de regressão com o eixo das ordenadas.
- e) O modelo evidencia relação inversa entre as variáveis, uma vez que ambos os coeficientes apresentam sinal positivo.

30. Em uma análise exploratória de dados de clientes, uma equipe não possui rótulos prévios indicando perfis de consumo, categorias ou classes conhecidas. Ainda assim, deseja identificar grupos naturais de clientes com comportamentos semelhantes entre si e diferentes dos demais grupos. A técnica adequada para esse cenário é

- a) classificação supervisionada, pois utiliza classes previamente conhecidas para treinar o modelo.
- b) agrupamento, pois busca identificar estruturas ou subgrupos naturais em dados não rotulados.
- c) regressão linear, pois estima uma variável numérica contínua a partir de variáveis explicativas.
- d) validação cruzada, pois separa os dados em folds para medir o desempenho preditivo do modelo.

e) análise de séries temporais, pois identifica padrões sequenciais dependentes do tempo.

31. Em uma base de dados multivariada usada para agrupamento, há variáveis medidas em escalas muito diferentes, como renda anual, idade e quantidade de acessos a um sistema. Além disso, algumas variáveis apresentam forte correlação entre si. Para reduzir distorções na medição de proximidade entre observações, a equipe deseja empregar uma métrica que considere variâncias e covariâncias dos atributos. A métrica mais adequada é

- a) distância de Manhattan, pois soma as diferenças absolutas entre as coordenadas dos pontos.
- b) distância euclidiana, pois mede o comprimento da reta entre dois pontos no espaço de atributos.
- c) similaridade de cosseno, pois mede apenas o ângulo entre os vetores, ignorando completamente sua magnitude.
- d) distância de Chebyshev, pois considera apenas a maior diferença absoluta entre coordenadas.
- e) distância de Mahalanobis, pois ajusta os efeitos de escala e correlação entre variáveis por meio da matriz de covariâncias.

32. Uma equipe pretende aplicar um algoritmo de agrupamento particional em uma base de dados numérica. O método escolhido exige a definição prévia do número de grupos, inicializa centros de clusters, atribui cada observação ao centro mais próximo, recalcula esses centros pela média dos pontos atribuídos e repete o processo até a estabilização. O algoritmo descrito é

- a) k-means.
- b) k-modes.
- c) k-medoids.
- d) DBSCAN.
- e) AGNES.

33. Durante a aplicação de agrupamento hierárquico aglomerativo, um analista precisa definir como será calculada a distância entre dois clusters. Ele deseja usar um critério que una, a cada etapa, os dois grupos cuja fusão produza o menor aumento da variância interna total, favorecendo clusters compactos. O método de ligação descrito é

- a) ligação mínima, pois considera a menor distância entre pares de pontos dos dois clusters.
- b) ligação completa, pois considera a maior distância entre pares de pontos dos dois clusters.
- c) ligação média, pois utiliza a média de todas as distâncias entre pontos de clusters distintos.
- d) método de Ward, pois seleciona a fusão que minimiza o aumento da variância intracluster.
- e) ligação por cosseno, pois calcula a distância hierárquica a partir do ângulo entre centroides.

34. Em uma rede neural profunda utilizada para classificação de imagens, a equipe técnica percebeu que o uso de funções lineares em todas as camadas impedia o modelo de capturar padrões complexos, limitando a rede a combinações lineares das variáveis de entrada. Para permitir que a rede aprenda relações não lineares, como bordas, texturas e formas em imagens, deve-se utilizar

- a) funções de perda, responsáveis por substituir os pesos das conexões entre os neurônios.
- b) funções de ativação não lineares, responsáveis por introduzir não linearidade no processamento da rede.
- c) funções de custo lineares, responsáveis por impedir o ajuste dos parâmetros durante o treinamento.
- d) funções de pooling, responsáveis por classificar diretamente a imagem em uma das classes possíveis.
- e) funções de inicialização, responsáveis por eliminar a necessidade de treinamento da rede neural.

35. Uma equipe está desenvolvendo um sistema de visão computacional para identificar defeitos em imagens industriais. Como os pixels próximos tendem a formar padrões locais relevantes, como bordas, texturas e pequenas irregularidades, a equipe deseja utilizar uma arquitetura capaz de explorar a localidade espacial por meio de filtros que percorrem regiões da imagem. A arquitetura mais adequada é

- a) rede neural convolucional, pois utiliza filtros ou kernels para extrair padrões locais e pode empregar pooling para reduzir a dimensionalidade dos mapas de características.
- b) rede neural recorrente, pois processa os pixels apenas com base na dependência temporal entre palavras de uma sequência.
- c) perceptron simples, pois possui uma única camada e é suficiente para aprender qualquer padrão não linear em imagens complexas.
- d) rede de Hopfield, pois gera novas imagens sintéticas a partir de ruído aleatório de entrada.
- e) regressão logística, pois preserva automaticamente a relação espacial entre pixels vizinhos em imagens.

36. Durante o treinamento de uma rede neural multicamada, a saída prevista pelo modelo é comparada ao valor real, gerando um erro. Para ajustar pesos e vieses das camadas anteriores, o algoritmo calcula como cada parâmetro contribuiu para esse erro e propaga essa informação da camada de saída em direção às camadas internas. Esse procedimento é conhecido como

- a) max pooling, pois seleciona o maior valor de cada região da imagem durante a etapa de convolução.
- b) softmax, pois transforma todas as saídas intermediárias da rede em valores binários.
- c) forward propagation, pois propaga o erro da saída para as camadas anteriores para corrigir os pesos.
- d) backpropagation, pois retropropaga o erro pela rede para calcular gradientes e permitir a atualização dos parâmetros.
- e) dropout, pois remove definitivamente os neurônios que apresentarem maior erro durante o treinamento.

37. Durante o treinamento de uma rede neural profunda para classificação de imagens, a equipe percebeu que o erro no conjunto de treinamento continuava diminuindo, enquanto o erro no conjunto de validação passava a aumentar após determinado número de épocas. Para reduzir esse comportamento, decidiu-se aplicar uma técnica que, durante o treinamento, desativa aleatoriamente parte dos neurônios, impedindo que a rede dependa excessivamente de combinações específicas de unidades internas.

A técnica descrita é:

- a) validação cruzada, por dividir os dados em múltiplos folds de treino e validação.
- b) gradiente descendente estocástico, por atualizar os pesos usando pequenos lotes de dados.
- c) minimização da entropia cruzada, por otimizar uma função de perda adequada à classificação.
- d) dropout, por desligar aleatoriamente neurônios durante o treinamento para reduzir o sobreajuste.

e) sobreamostragem, por aumentar artificialmente a frequência das classes minoritárias.

38. Uma equipe deseja substituir uma única árvore de decisão, que apresentou alta variância e tendência a overfitting, por um método mais robusto. A estratégia escolhida treina vários modelos ligeiramente diferentes em amostras bootstrap diferentes e, adicionalmente, seleciona subconjuntos aleatórios de atributos durante a sua construção. A predição final é obtida por votação majoritária em problemas de classificação ou pela média das previsões em problemas de regressão.

O método descrito é:

- a) boosting, pois treina modelos sequencialmente para corrigir os erros dos modelos anteriores.
- b) stacking, pois utiliza as previsões de modelos base como entrada para um meta-modelo.
- c) regressão logística, pois combina linearmente atributos para estimar probabilidades de classe.
- d) k-nearest neighbors, pois classifica uma observação com base na distância em relação aos vizinhos mais próximos.
- e) random forest, pois combina múltiplas árvores treinadas com amostras e atributos aleatórios para reduzir variância e melhorar a generalização.

39. Em um projeto de Deep Learning, uma equipe deseja construir redes neurais de forma produtiva, utilizando uma biblioteca que ofereça abstrações de alto nível, como camadas, modelos, otimizadores, funções de perda e métricas. A equipe também quer evitar a implementação direta de operações matemáticas de baixo nível, como manipulação explícita de tensores e cálculo manual de gradientes, delegando essas tarefas a bibliotecas especializadas de backend.

A biblioteca adequada para esse cenário é:

- a) NumPy, por ser uma biblioteca voltada exclusivamente à criação de redes neurais profundas com camadas reutilizáveis.
- b) SciPy, por fornecer diretamente APIs de alto nível para treinamento de modelos convolucionais e recorrentes.
- c) Pandas, por permitir a definição de modelos neurais sequenciais e funcionais a partir de DataFrames.
- d) Matplotlib, por automatizar o treinamento de redes neurais por meio de funções gráficas.
- e) Keras, por fornecer uma interface de alto nível, modular e simplificada para construção de modelos de Deep Learning.

40. Em uma análise de séries temporais, uma equipe avalia uma sequência mensal de indicadores econômicos e observa que, ao longo dos anos, a média da série se desloca progressivamente, enquanto a variabilidade também se altera em determinados períodos. Antes de aplicar modelos preditivos, os analistas discutem se a série pode ser considerada estacionária.

Nesse contexto, a característica que melhor diferencia uma série temporal estacionária de uma não estacionária é:

- a) uma série estacionária sempre apresenta tendência crescente, enquanto uma série não estacionária apresenta apenas oscilações aleatórias.

- b) uma série estacionária é composta apenas por observações anuais, enquanto uma série não estacionária é composta por observações contínuas.

- c) uma série estacionária possui propriedades estatísticas, como média e variância, aproximadamente constantes ao longo do tempo, enquanto uma série não estacionária apresenta mudanças nessas propriedades.

- d) uma série estacionária não pode ser usada para previsão, enquanto uma série não estacionária é a única adequada para modelos de séries temporais.

- e) uma série estacionária é aquela que não possui defasagens, enquanto uma série não estacionária depende exclusivamente de valores passados.

BLOCO II

Lucas Ianni

41. A equipe de BI de uma rede varejista projeta um data warehouse e decide normalizar as tabelas de dimensão, separando atributos de baixa cardinalidade em tabelas auxiliares, como UF segregada de Município e Categoria separada de Produto. Essa decisão de modelagem caracteriza o esquema

- a) star schema, em que toda a hierarquia é mantida em uma única tabela de dimensão por área de análise.

- b) fact constellation, no qual múltiplas tabelas fato compartilham as mesmas dimensões agregadas.

- c) snowflake schema, em que as dimensões são normalizadas em hierarquias formando ramificações secundárias.

- d) galaxy schema, que dispensa tabela fato central e armazena tudo em dimensões interligadas.

- e) cube schema, no qual dimensões e fatos são fundidos em uma única estrutura multidimensional.

42. Um analista deseja extrair, de uma tabela CLIENTES, todos os registros cuja idade esteja compreendida entre 25 e 40 anos, incluídos os limites. A cláusula SQL apropriada para essa extração por intervalo é

- a) WHERE idade BETWEEN 25 AND 40.
- b) WHERE idade RANGES 25 TO 40.
- c) WHERE idade IN (25, 40).
- d) WHERE idade LIKE '25..40'.
- e) WHERE idade FIND IN (25, 40).

43. Um arquiteto de dados precisa revisar a conformidade de um novo data warehouse corporativo com as quatro características propostas por William Inmon. Analise as afirmativas a seguir.

- I. A orientação a assunto significa que o data warehouse organiza dados em torno dos principais temas do negócio, como vendas, clientes e produtos, em vez de organizá-los em torno das aplicações operacionais que os geram.
- II. A integração exige que dados provenientes de diferentes fontes sejam padronizados quanto a unidades de medida, nomenclatura e codificação antes de serem carregados no repositório analítico.
- III. A não volatilidade implica que, uma vez carregados, os dados do data warehouse podem ser livremente atualizados pelos usuários finais para refletir a operação corrente da empresa.
- IV. A variância no tempo refere-se ao fato de que o data warehouse armazena dados históricos com referência temporal, permitindo análises de tendências e comparações entre períodos.

Estão corretas apenas as afirmativas

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

44. Em um projeto de engenharia de dados, a equipe técnica precisa carregar grandes volumes de dados estruturados e semiestruturados em um data lake corporativo, aproveitando o poder computacional do próprio repositório de destino para realizar as transformações somente após a carga dos dados brutos. A abordagem adequada para esse cenário é

- a) ETL, pois transforma os dados em trânsito antes de carregá-los no destino final.
- b) ELT, pois carrega os dados brutos diretamente no destino e realiza as transformações posteriormente, no próprio repositório.
- c) CDC (Change Data Capture), por ser a única abordagem aplicável a repositórios de dados não estruturados.
- d) OLTP, técnica que automatiza a transformação de dados durante operações transacionais correntes.
- e) Stage Area, abordagem que dispensa qualquer transformação dos dados originais.

45. Em uma solução analítica baseada em OLAP, diferentes operações permitem ao usuário navegar pelos cubos multidimensionais. Associe as operações da Coluna I com suas respectivas descrições na Coluna II.

Coluna I

1. Drill-down
2. Drill-up (roll-up)
3. Slice
4. Dice

Coluna II

- () Seleciona uma fatia do cubo a partir do filtro aplicado em uma única dimensão.
- () Aumenta o nível de detalhamento dos dados, descendo na hierarquia de uma dimensão.
- () Seleciona um subcubo a partir de filtros aplicados em duas ou mais dimensões simultaneamente.
- () Reduz o nível de detalhamento dos dados, agregando-os ao subir na hierarquia da dimensão.

A sequência correta, de cima para baixo, é

- a) 1 – 2 – 3 – 4.
- b) 2 – 1 – 4 – 3.
- c) 3 – 1 – 4 – 2.
- d) 4 – 3 – 1 – 2.
- e) 3 – 4 – 1 – 2.

46. Durante a modelagem conceitual de um sistema de recursos humanos, a equipe identifica uma entidade DEPENDENTE cujos registros somente podem ser identificados unicamente por meio da combinação entre a chave primária da entidade FUNCIONARIO e um identificador próprio da entidade dependente, como o nome do dependente. Essa entidade é classificada como

- a) entidade associativa, pois materializa um relacionamento entre duas entidades fortes.
- b) entidade primária composta, em razão de sua chave possuir mais de um atributo.
- c) entidade subordinada hierárquica, por estar abaixo da entidade FUNCIONARIO na árvore conceitual.
- d) entidade fraca, pois sua chave primária depende da chave primária de outra entidade para garantir unicidade.
- e) entidade especializada, pois herda atributos da entidade FUNCIONARIO em relação de generalização.

47. Sobre os operadores e cláusulas da linguagem SQL, assinale a alternativa INCORRETA:

- a) O operador LIKE permite buscar padrões em colunas textuais, sendo o caractere % usado para representar zero, um ou múltiplos caracteres, e o caractere _ usado para representar exatamente um único caractere.
- b) O operador IN funciona como abreviação para múltiplas condições encadeadas com AND, retornando apenas registros que atendam simultaneamente a todos os valores especificados na lista de comparação.
- c) A cláusula DISTINCT, posicionada logo após o SELECT, elimina linhas duplicadas no conjunto de resultados retornados pela consulta.
- d) A palavra-chave RIGHT JOIN retorna todos os registros da tabela à direita da junção e os correspondentes da tabela à esquerda, exibindo valores NULL nas colunas da tabela esquerda quando não houver correspondência.
- e) A cláusula BETWEEN permite filtrar valores que se encontram dentro de um intervalo determinado, incluindo os limites informados na expressão.

48. Em um cluster HDFS implantado por uma empresa de mídia para processar grandes volumes de logs de acesso, a equipe identifica falhas intermitentes em um dos nós e precisa esclarecer qual componente concentra o gerenciamento do espaço de nomes do sistema de arquivos, controla os metadados e direciona o acesso aos blocos distribuídos entre os demais servidores. Esse componente é o

- a) DataNode, responsável por manter os metadados e a árvore de diretórios do cluster.
- b) JobTracker, encarregado de armazenar fisicamente os blocos de dados replicados.
- c) Resource Manager, dedicado exclusivamente ao gerenciamento do namespace.
- d) Secondary NameNode, que substitui automaticamente o nó principal em caso de falha sem interrupção do serviço.
- e) NameNode, encarregado de gerenciar o namespace, os metadados do sistema e controlar o acesso e a distribuição dos arquivos.

49. Um desenvolvedor inicia o uso do MongoDB e deseja entender corretamente sua organização interna. Nesse SGBD, a unidade básica de dados, que se assemelha aproximadamente a um registro do modelo relacional, é denominada

- a) tupla.
- b) shard.
- c) documento.
- d) coleção.
- e) índice.

50. Uma empresa de streaming precisa armazenar centenas de petabytes de vídeos, imagens e backups, distribuindo os dados em escala global, com acesso direto por identificadores únicos, sem percorrer hierarquias de diretórios. A tecnologia adequada a esse cenário é

- a) o armazenamento orientado a objetos, que organiza os dados em unidades independentes compostas por conteúdo, metadados e identificador único, acessados diretamente em uma arquitetura plana.
- b) o sistema de arquivos hierárquico tradicional, que dispõe os dados em estrutura de diretórios e subdiretórios, exigindo navegação para acesso aos arquivos.
- c) o armazenamento em blocos, baseado em divisão dos dados em blocos de tamanho fixo acessados por endereços lógicos, otimizado para transações de banco de dados.
- d) o armazenamento NAS (Network Attached Storage) baseado em protocolos de compartilhamento de arquivos como NFS e SMB, restrito ao ambiente da rede local.
- e) o armazenamento DAS (Direct Attached Storage), conectado diretamente ao servidor por meio de barramentos como SATA ou SAS, sem distribuição entre múltiplos nós.

51. Sobre os paradigmas e aplicações do Processamento de Linguagem Natural (PLN), analise as afirmativas a seguir.

- I. No paradigma simbólico, o PLN baseia-se em regras linguísticas explicitamente definidas por especialistas, modelando a linguagem por meio de estruturas gramaticais, sintáticas e semânticas codificadas em lógica formal.
- II. No paradigma estatístico, modelos probabilísticos são treinados em grandes corpora textuais, estimando a probabilidade de ocorrência de palavras e frases em determinados contextos.
- III. A análise de sentimentos é uma aplicação do PLN voltada exclusivamente para classificar textos em duas categorias rigidamente binárias: spam ou não spam.
- IV. A sumarização extrativa gera o resumo a partir da seleção de trechos do texto original, sem reescrita ou geração de novas frases, enquanto a sumarização abstrativa produz frases inéditas a partir da representação semântica do conteúdo.

Estão corretas apenas as afirmativas

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

52. Em um projeto de mineração de textos jurídicos, a equipe deseja reduzir as variações morfológicas das palavras presentes nos documentos antes de aplicar algoritmos de classificação. Sobre as técnicas de lematização e stemização, é correto afirmar que

- a) ambas produzem sempre a mesma saída quando aplicadas a uma mesma palavra, sendo diferenciadas apenas pela performance computacional.
- b) a lematização converte palavras à sua forma canônica (lema), produzindo uma palavra que existe no idioma, enquanto a stemização remove afixos para obter o radical, podendo resultar em uma sequência não dicionarizada.
- c) a stemização exige análise contextual semântica profunda, sendo mais cara computacionalmente que a lematização, que apenas corta sufixos.
- d) a lematização é específica para a língua inglesa, enquanto a stemização é exclusivamente aplicável ao português por meio do algoritmo de Porter.
- e) tanto a lematização quanto a stemização preservam integralmente as flexões verbais e nominais, distinguindo-se apenas pelo tipo de algoritmo utilizado.

53. A tokenização é uma das primeiras etapas do pré-processamento textual em PLN. Sobre essa técnica, é correto afirmar que ela

- a) atribui rótulos morfossintáticos a cada palavra, identificando classes gramaticais como substantivos, verbos e adjetivos.
- b) elimina palavras de baixo valor semântico do texto, como preposições, artigos e conjunções.
- c) reduz palavras flexionadas à sua forma canônica, considerando o contexto gramatical.
- d) divide o texto em unidades menores (palavras, subpalavras, caracteres ou pontuação), associando a cada uma um identificador único.
- e) converte representações textuais em vetores densos de baixa dimensionalidade que capturam relações semânticas latentes.

54. Sobre as principais técnicas de representação vetorial de textos no Processamento de Linguagem Natural, analise as afirmativas a seguir.

- I. O modelo Bag of Words (BoW) representa o texto como vetor de frequências brutas das palavras, preservando a ordem em que elas aparecem no documento e capturando relações semânticas entre os termos.
- II. O modelo TF-IDF atribui maior peso a palavras frequentes em um documento específico e simultaneamente raras no conjunto geral de documentos do corpus, sendo amplamente utilizado em sistemas de recuperação de informação.
- III. O modelo N-gram com $n \geq 2$ captura a ordem das palavras ao representar o texto como sequências contíguas de n termos, considerando o contexto local de ocorrência.
- IV. O modelo Word2Vec, em sua variante CBOW, prevê a palavra-alvo a partir das palavras de contexto ao redor, enquanto a variante Skip-gram realiza o caminho inverso, prevendo o contexto a partir da palavra-alvo.

Estão corretas apenas as afirmativas

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

55. Em uma reunião técnica sobre os fundamentos da arquitetura Transformer, um engenheiro de IA explica que o mecanismo central de atenção opera com três projeções vetoriais sobre cada token de entrada, cujo papel é, respectivamente, representar "quem está perguntando", "quem está apto a responder" e "qual conteúdo será efetivamente agregado à representação contextual final". Essas três projeções vetoriais correspondem, na ordem mencionada, a

- a) Embedding, Tokenizer e Encoder.
- b) Logit, Softmax e Loss.
- c) Query, Key e Value.
- d) Encoder, Decoder e Classifier.
- e) Input, Output e Hidden State.

BLOCO III

Carlos Henrique

56. Durante a calibração de sensores utilizados em uma malha de monitoramento de pressão em dutos submarinos, um engenheiro de dados propôs o seguinte modelo matemático para representar a eficiência relativa do sistema em função do parâmetro operacional x :

$$E(x) = \frac{x^2 - 4x + 7}{x - 1}$$

Sabe-se que valores de x próximos de 1 correspondem a uma faixa crítica de operação do equipamento.

Com base no comportamento da função nessa região, o valor do limite de $E(x)$ quando $x \rightarrow 1$ é:

- a) $-\infty$
- b) $+\infty$
- c) 2
- d) O limite não existe, pois os limites laterais são distintos.
- e) 4

57. Durante o treinamento de um algoritmo de regressão logística utilizado para prever falhas em bombas centrífugas, um cientista de dados definiu a seguinte função custo:

$$J(x, y) = x^2 + y^2 - 4x + 6y + 13$$

em que x e y representam parâmetros ajustáveis do modelo.

Sabendo que a convergência ideal ocorre no ponto de mínimo global da função custo, o par ordenado correspondente a esse ponto é:

- a) $(2, -3)$
- b) $(-2, 3)$
- c) $(4, -6)$
- d) $(1, -3)$
- e) $(2, 3)$

58. Considere a matriz

$$A = [2 \ 1 \ 1 \ 2]$$

Os autovalores de A são:

- a) 1 e 4
- b) 2 e 2
- c) 3 e 1
- d) 0 e 3
- e) -1 e 3

59. Em um centro de processamento de dados utilizado para monitoramento em tempo real de plataformas offshore, o número diário de falhas críticas detectadas em determinado servidor segue uma distribuição de Poisson com média igual a 4 ocorrências por dia.

Considere que, em um determinado dia, a equipe deseje estimar a probabilidade de ocorrerem exatamente 2 falhas críticas.

Nesse caso, a probabilidade procurada é:

- a) $\frac{8e^{-4}}{3}$
- b) $8e^{-4}$

c) $4e^{-2}$

d) $\frac{16e^{-4}}{4}$

e) $\frac{16e^{-2}}{2}$

60. Em um processo de análise estatística, duas variáveis aleatórias X e Y possuem:

$$\text{Var}(X) = 9$$

$$\text{Var}(Y) = 16$$

$$\text{Cov}(X, Y) = -6$$

O coeficiente de correlação linear de Pearson entre X e Y é:

a) $-\frac{1}{2}$

b) $\frac{1}{2}$

c) $-\frac{3}{4}$

d) $\frac{3}{8}$

e) $-\frac{3}{8}$

61. Em um estudo sobre comportamento de consumo energético de servidores utilizados em aplicações de aprendizado de máquina, a taxa de desempenho de um sistema foi modelada pela função

$$f(x) = x^3 - 6x^2 + 9x + 1$$

em que x representa um parâmetro de ajuste do sistema.

A equipe responsável deseja identificar o ponto em que ocorre o máximo local da função, pois ele representa a condição operacional mais eficiente antes do início da perda de desempenho.

O valor de x correspondente ao máximo local é:

- a) 0
- b) 1
- c) 2
- d) 3
- e) 4

62. Um cientista de dados utiliza a norma L1 para medir o erro entre os vetores

$$u = (2, -1,4, -3)$$

e

$$v = (1,3, -2,5)$$

A distância L1 entre u e v é:

- a) 18
- b) 19
- c) 20
- d) 21
- e) 22

63. Considere a integral

$$\int_0^1 (3x^2 + 2x - 1)dx$$

O valor dessa integral é:

- a) 0
- b) 1
- c) $\frac{3}{2}$
- d) 2
- e) $\frac{5}{2}$

64. Uma equipe de ciência de dados da PETROBRÁS está avaliando o tempo médio de resposta de um sistema de análise sísmica baseado em processamento distribuído.

Após coletar uma amostra aleatória de 144 execuções, obteve-se média amostral igual a 80 ms. Estudos anteriores indicam que o desvio padrão populacional pode ser considerado igual a 12 ms.

Assumindo distribuição aproximadamente normal e utilizando nível de confiança de 95%, determine o intervalo de confiança aproximado para a média populacional.

(Dado: $z_{0,025} = 1,96$)

- a) (78,04 ; 81,96)

b) (77,00 ; 83,00)

c) (79,02 ; 80,98)

d) (76,08 ; 83,92)

e) (78,50 ; 81,50)

65. Considere as matrizes

$$A = [1 \ 2 \ 3 \ 4]$$

e

$$B = [2 \ 0 \ 1 \ -1]$$

O produto matricial AB é:

a) [4 - 2 10 - 4]

b) [2 - 2 6 - 4]

c) [4 0 10 - 1]

d) [2 2 4 - 1]

e) [3 - 2 7 - 4]

66. Considere a função de duas variáveis:

$$f(x, y) = x^2y + 3xy^2$$

A derivada parcial de f em relação a x é:

a) $2xy + 3y^2$

b) $x^2 + 6xy$

c) $2x + 3y$

d) $2xy^2 + 3y$

e) $x^2y + 6xy$

67. Em um sistema automatizado de monitoramento industrial, dois alarmes independentes podem ser acionados durante o processamento de dados:

- evento A: detecção de sobrecarga computacional;
- evento B: detecção de inconsistência nos dados de sensores.

Sabe-se que:

$$P(A) = 0,6$$

$$P(B) = 0,5$$

Considerando a independência entre os eventos, a probabilidade de ocorrer simultaneamente $A \cap B$ é:

- a) 0,10
- b) 0,20
- c) 0,25
- d) 0,30
- e) 0,55

68. Uma equipe de engenharia de dados avalia o comportamento da função

$$f(x) = \ln \ln(x) - x$$

para $x > 0$.

O valor de x no qual ocorre o ponto crítico da função é:

- a) 0
- b) $\frac{1}{2}$
- c) 1
- d) e
- e) Não existe ponto crítico.

69. Considere o vetor

$$v = (3, -4, 12)$$

A norma euclidiana L2 desse vetor é:

- a) 12
- b) 13
- c) 14
- d) 15
- e) 16

70. Uma equipe de ciência de dados está analisando o consumo diário de processamento computacional de diferentes unidades operacionais.

Após a construção de um *boxplot*, verificou-se que:

$$Q_1 = 20$$

$$Q_3 = 44$$

A equipe deseja identificar automaticamente valores atípicos superiores utilizando o critério usual baseado no intervalo interquartil.

Segundo esse critério, pode ser classificado como outlier superior qualquer valor acima de:

- a) 56
- b) 68
- c) 72
- d) 74
- e) 80

O QUE VOCÊ ACHOU DESTE SIMULADO?

*Conte-nos como foi sua experiência ao fazer este simulado.
Sua opinião é muito importante para nós!*

<https://forms.gle/3RC8pkHLNAnpSnxQ9>

NÃO É ASSINANTE?

Confira nossos planos, tenha acesso a milhares de cursos e participe gratuitamente dos projetos exclusivos. Clique no link!

<http://estrategi.ac/assinaturas>

CONHEÇA NOSSO SISTEMA DE QUESTÕES

Estratégia Questões nasceu maior do que todos os concorrentes, com mais questões cadastradas e mais soluções por professores. Clique no link e conheça!

<http://estrategi.ac/ok1zt0>
