



CONCURSO PÚBLICO DA EMPRESA DE TECNOLOGIA E
INFORMAÇÕES DA PREVIDÊNCIA
DATAPREV

NÍVEL SUPERIOR – TARDE
**ANALISTA DE TECNOLOGIA DA INFORMAÇÃO
INTELIGÊNCIA DA INFORMAÇÃO**

PROVA OBJETIVA - TIPO 1



SUA PROVA

Além deste caderno de questões contendo **70 (setenta)** questões objetivas, você receberá do fiscal de sala o cartão de respostas.

As questões objetivas têm **5 (cinco)** opções de resposta (A, B, C, D e E) e somente uma delas está correta.



TEMPO

- **4 (quatro) horas** é o período disponível para a realização da prova, **já incluído o tempo para a marcação da folha de respostas.**
- **2 (duas) horas** após o início da prova, é possível retirar-se da sala, sem levar o caderno de questões nem qualquer tipo de anotação de suas respostas.
- **30 (trinta) minutos** antes do término do período de prova, é possível retirar-se da sala, **levando o caderno de questões.**



NÃO SERÁ PERMITIDO

- Qualquer tipo de comunicação entre os candidatos durante a aplicação da prova.
- Usar o sanitário ao término da prova, após deixar a sala.
- Anotar informações relativas às respostas em qualquer outro meio que não seja este caderno de questões.



INFORMAÇÕES GERAIS

- Verifique se este caderno de questões está completo e sem falhas de impressão. Caso contrário, **notifique imediatamente o fiscal da sala**, para que sejam tomadas as devidas providências.
- No cartão de respostas confira seus dados pessoais, especialmente nome, número de inscrição e documento de identidade, e leia atentamente as instruções para preenchê-lo.
- **Use somente caneta esferográfica, fabricada em material transparente, com tinta preta ou azul.**
- Assine seu nome apenas no espaço reservado na folha de respostas.
- Confira o tipo do seu caderno de questões. Caso tenha recebido caderno de questões com tipo diferente do impresso em seu cartão de respostas, o fiscal deve ser **obrigatoriamente** informado para o devido registro na ata da sala.
- O preenchimento das respostas é de sua responsabilidade e não será permitida a substituição do cartão de respostas em caso de erro cometido por você.
- Para fins de avaliação, serão levadas em consideração apenas as marcações realizadas no cartão de respostas.
- Os candidatos serão submetidos ao sistema de detecção de metais quando do ingresso e da saída das salas e dos sanitários durante a realização das provas.

Boa prova!

Conhecimentos Gerais

Língua Portuguesa

1

Leia com atenção o texto a seguir.

Processamento de dados é a conversão de dados brutos em informações úteis por meio de etapas estruturadas, como coleta, preparação, análise e armazenamento de dados. As organizações podem ter insights praticáveis e informar a tomada de decisão processando dados de forma eficaz.

Historicamente, as empresas dependiam de processamento de dados e calculadoras para gerenciar conjuntos de dados menores. Como as empresas geravam volumes cada vez maiores de dados, tornaram-se essenciais os métodos avançados de processamento de dados.

KRANTZ, Tom. JONKER, Alexandra. *O que é processamento de dados?* IBM.2025.

As opções a seguir apresentam observações corretas sobre esse pequeno texto, **à exceção de uma**. Assinale-a.

- (A) O texto se inicia pela definição de “processamento de dados”.
- (B) As duas ocorrências do conectivo “como” possuem o mesmo significado.
- (C) Por meio do processamento de dados, informações úteis resultam da conversão de dados brutos.
- (D) De acordo com o texto, os métodos avançados de processamento são respostas a uma necessidade.
- (E) O advérbio “historicamente” se refere a realidades passadas no terreno do processamento de dados.

2

Leia com atenção o trecho a seguir.

“Se os navios geralmente vão melhor do que os Estados, isso ocorre pela única razão de que naqueles cada um aceita a parte que lhe compete, enquanto nestes, na maioria das vezes, quanto menos se sabe, mais se tem o desejo de comandar.”

Sobre esse trecho, assinale a afirmativa correta.

- (A) A conjunção “Se” inicial indica uma condição.
- (B) O pronome demonstrativo “naqueles” tem como referente os navios.
- (C) O advérbio “geralmente” se opõe semanticamente à expressão “na maioria das vezes”.
- (D) Na comparação entre navios e Estados, o trecho indica a superioridade dos Estados sobre os navios.
- (E) Enquanto o demonstrativo “isso” se refere a algo já dito, o pronome “cada um” se refere a algo futuro do trecho.

3

No texto a seguir, os pronomes demonstrativos sublinhados estão bem empregados.

Se os navios geralmente vão melhor do que os Estados, isso ocorre pela única razão de que naqueles cada um aceita a parte que lhe compete, enquanto nestes, na maioria das vezes, quanto menos se sabe, mais se tem o desejo de comandar.

Assinale a opção que mostra a frase em que o demonstrativo está **erradamente** empregado.

- (A) Esta é a regra de ouro dos governantes: respeitem os votos dos eleitores.
- (B) Antigamente, as leis eram observadas com mais rigor, mas, naqueles tempos tudo era diferente.
- (C) Os advogados devem manter sigilo sobre seus clientes e isso deve ser rigorosamente respeitado.
- (D) Dicionários e enciclopédias fazem parte de boas bibliotecas, mas entre estes e aqueles há diferença de volume.
- (E) Para os que não conseguem governar a própria casa nem os povos, uma boa desculpa é dizer que estes são ingovernáveis.

4

Leia o texto a seguir.

Quem governa uma instituição pública ou privada é sempre objeto do ódio e das queixas dos governados. Os homens estão sempre descontentes porque estão sempre infelizes. Por isso ficam descontentes com o próprio Estado, e também por isso ficam descontentes com quem o governa... Entretanto, no que concerne ao ato de governar, infelizmente existem apenas duas alternativas realmente sensatas: ou abster-se do governo, seja ele público ou privado, ou administrá-lo totalmente em vantagem própria e não dos governados.

Leopardi, Giacomo. Ensaísta italiano.

As opções a seguir apresentam observações de caráter linguístico sobre o texto.

Assinale a única observação **incorreta**.

- (A) Os adjetivos “pública ou privada” são de conteúdo subjetivo, expressando uma opinião do enunciador do texto.
- (B) O segundo período do texto traz uma justificativa para o que é afirmado no período anterior.
- (C) O termo “e também” traz uma redundância, que é normalmente aceita.
- (D) As duas ocorrências de “por isso” mostram o mesmo antecedente.
- (E) O termo “Entretanto” traz o valor de oposição.

5

Os que trabalham em espaços profissionais em que inserem textos devem preocupar-se com a correção gramatical.

Entre as frases a seguir, assinale a que se mostra integralmente correta.

- (A) Vimos hoje aqui depois das aulas.
- (B) Vieram ela e eu para a festa no mesmo táxi.
- (C) Os policiais requisaram nossos documentos.
- (D) Não restou nenhum segredo entre eu e meu advogado.
- (E) As autoridades interviram nos problemas políticos surgidos na semana passada.

6

Assinale a opção que apresenta o único texto coerente.

- (A) Na Europa, nesta época, faz muito frio: vamos comprar roupas mais leves.
- (B) A verdade é que ninguém pretende descobrir nada com essas investigações.
- (C) A despeito dos altos preços dos restaurantes, não comemos fora durante a semana.
- (D) Nessas situações há duas coisas a fazer: dirigir-se imediatamente a uma delegacia.
- (E) Ela não sabia desde que momento de sua vida ela trazia aquela marca de nascença na testa.

7

As frases a seguir mostram um termo sublinhado que se refere a um termo anterior, estabelecendo coesão na frase.

Assinale a opção em que a identificação do antecedente do termo sublinhado está correta.

- (A) A morte é um jeito de a natureza lhe dizer para pegar leve. / a natureza.
- (B) Que apertada é a porta e que estreito o caminho que conduz à vida. / a porta.
- (C) Todo homem deveria ser incendiário até os 25 anos de idade e bombeiro daí em diante. / 25 anos.
- (D) O jovem não raciocina, age; o velho raciocina e põe os outros para agirem em seu lugar. / o jovem.
- (E) Qual pode ser a vida que começa pelos gritos da mãe, que a dá, e os choros do filho que a recebe. / a vida.

8

As opções a seguir apresentam frases em que há a repetição de uma mesma palavra.

Assinale aquela em que as palavras mostram sentidos diferentes.

- (A) Morrer por uma causa não faz com que essa causa seja justa.
- (B) Para mim, velho é aquele sujeito 15 anos mais velho do que eu.
- (C) Depois dos 70, tudo se resume a remendar, remendar, remendar.
- (D) Fala-se muito em loucuras da juventude; não sei, porém, porque nada se diz das loucuras da velhice.
- (E) Quarenta anos é uma idade terrível. Porque é a idade em que chegamos a ser o que somos.

ATENÇÃO: O texto a seguir refere-se às duas próximas questões.

Leia o texto abaixo, sobre gestão e governança.

A governança não se confunde com gestão: enquanto a gestão executa atividades e processos, a governança direciona, define políticas, estabelece prioridades e supervisiona resultados, garantindo que os objetivos organizacionais da empresa sejam alcançados de forma ética e eficiente. Ela envolve a definição de responsabilidades, a alocação de recursos e a criação de mecanismos de controle de atividades e monitoramento.

9

Sobre a estruturação desse texto, assinale a afirmativa correta.

- (A) O texto tem a finalidade de definir os dois processos.
- (B) O texto pretende mostrar a superioridade da governança sobre a gestão.
- (C) O texto pretende mostrar objetivamente a diferença entre gestão e governança.
- (D) O texto pode ser definido como argumentativo, pois expõe opiniões do enunciador do texto.
- (E) O texto indica uma série de atividades da governança em ordem cronológica, montando um texto narrativo.

10

Nas opções a seguir, assinale aquela que traz uma função diferente do conectivo de.

- (A) objetivos organizacionais da empresa.
- (B) definição de responsabilidades.
- (C) alocação de recursos.
- (D) criação de mecanismos.
- (E) controle de atividades.

ATENÇÃO: O texto a seguir refere-se às duas próximas questões.

Um Analista de Negócios de TI desempenha um papel crucial em unir a tecnologia e os objetivos empresariais. Este profissional é responsável por garantir que as soluções de TI estejam alinhadas com as necessidades e objetivos empresariais, promovendo assim eficiência e inovação. Dentro da categoria de alinhamento entre TI e negócios, o Analista de Negócios de TI (IT BA) identifica áreas de melhoria, documenta requisitos e garante que as soluções tecnológicas sejam implementadas de forma eficaz para aprimorar os processos estruturais das empresas modernas.

11

Pela sua estrutura, esse pequeno texto deve ser classificado, quanto ao seu tipo, como

- (A) injuntivo.
- (B) informativo.
- (C) descritivo e narrativo.
- (D) descritivo e expositivo.
- (E) argumentativo e expositivo.

12

Entre os adjetivos que aparecem nas opções abaixo, assinale aquele que mostra uma opinião do enunciador, indicando uma qualidade.

- (A) papel crucial.
- (B) objetivos empresariais.
- (C) soluções tecnológicas.
- (D) processos estruturais.
- (E) empresas modernas.

Língua Inglesa

READ TEXT I AND ANSWER NEXT SEVEN QUESTIONS.

TEXT I

AI is getting better at writing. Humans must get better at editing.

Writing is hard. For Franz Kafka, putting words on the page was a process of “unending torments”. George Orwell said it felt like “a long bout of some painful illness”. Hunter S. Thompson, a journalist, saw it as the opposite of sex: “only good when it’s over”. Now AI has come to the rescue. Large language models (LLMs) eagerly offer reams of text without any pain or torment. Bots can spew out emails (excellently), longer prose (passably) and poetry (poorly). To them, writer’s block is a 500-sheet pack of paper: any user of an LLM can produce as many words in a day as Kafka did in his life.

For human writers this is a blessing and a curse. If you claim to be an author but use AI, you look like a fraud—bot-busting tools like Pangram are good at spotting when someone has had an AI helping hand. [...]

If LLMs keep improving, in a few years they could be as good as professional wordsmiths. In the meantime, though, their baggy prose puts a premium on editing skills. After all, even the best writers need a good editor to keep them sharp. Editing is something *The Economist* deals with day in, day out, so here are our tips for rewriting AI prose.

First, commission well. George R.R. Martin once said the job of editors is to “clip the wings” of writers, and “tell them in which direction to fly”. When it comes to LLMs, that means giving clear instructions about exactly what you want, and asking them to refine their output as many times as necessary. You’ll get less pushback from machines than men. Mr Martin said editors were “the writer’s natural enemy”; AI would probably say “great observation!” and “love the enemy analogy!”

Second, pay attention to detail. AI writers will make mistakes and offer phrases that sound good but mean nothing. They like to lavish their prose with Latinate words, talking about the “bi-directional flow” and how to “expedite grid enhancements”; remind them that Anglo-Saxon words are often punchier. They like clichéd metaphors: there are lots of “Achilles’ heels”. And AI tends to use the “rule of three” far more than humans. *The Economist* also likes this rhetorical trick. But whereas we try to group three sensible bits of advice, AI offers such redundant sentences as: “The mystery has been solved—not partially, not ambiguously, but definitively.” Such lines should be cut—definitively.

Finally, know your audience. The editor is an intermediary between writer and reader. LLMs are sycophants and their smarmy tone is often off-key. If you are writing an email to break some bad news to your boss, tell your bot to skip the exclamation marks. AI can be pretentious or pontificating (it was trained on human writing, after all). Most people are short of time, but a verbose AI pins you to the wall like a bore at a party.

“Write drunk, edit sober,” goes the old writers’ maxim. If that’s right, when the AI-human team gets down to work, the AIs will have all the fun.

Adapted from <https://www.economist.com/leaders/2026/07/30/ai-is-getting-better-at-writing-humans-must-get-better-at-editing>

13

Based on Text I, mark the statements below as true (T) or false (F).

- () The first paragraph refers to authors who assert that writing is an easy task.
- () At times, the tone of LLM-generated prose can be rather inappropriate.
- () It is argued that nowadays editing has lost its relevance.

The statements are, respectively,

- (A) F – F – T.
- (B) T – F – T.
- (C) F – T – T.
- (D) T – T – F.
- (E) F – T – F.

14

In the expression “*bout of some painful illness*” (1st paragraph), “*bout*” can be replaced, without significant change in meaning, by

- (A) cause.
- (B) period.
- (C) recovery.
- (D) symptom.
- (E) diagnosis.

15

In the first paragraph, the sentence “*Now AI has come to the rescue*” means that LLMs can

- (A) reduce memory loss.
- (B) discard important data.
- (C) win literary competitions.
- (D) provide considerable help.
- (E) avoid using personal input.

16

“*For human writers this is a blessing and a curse*” (2nd paragraph) contains a(n)

- (A) personification.
- (B) repetition.
- (C) command.
- (D) antithesis.
- (E) solution.

17

The word “*like*” in “*bot-busting tools like Pangram are good at spotting*” (2nd paragraph) is used in the same way as in

- (A) AI writers like to indulge in words that come from Latin.
- (B) The likes of AI bots are changing the way humans write.
- (C) Some AI programs tend to use words like “leveraging”.
- (D) Editors know what texts written by AI tools look like.
- (E) LLMs like to use predictable figurative expressions.

18

In the sentence “*Anglo-Saxon words are often punchier*” (5th paragraph), “*punchier*” indicates that these words tend to be more

- (A) complex and obscure.
- (B) powerful and effective.
- (C) mysterious and intricate.
- (D) formal and sophisticated.
- (E) ambiguous and confusing.

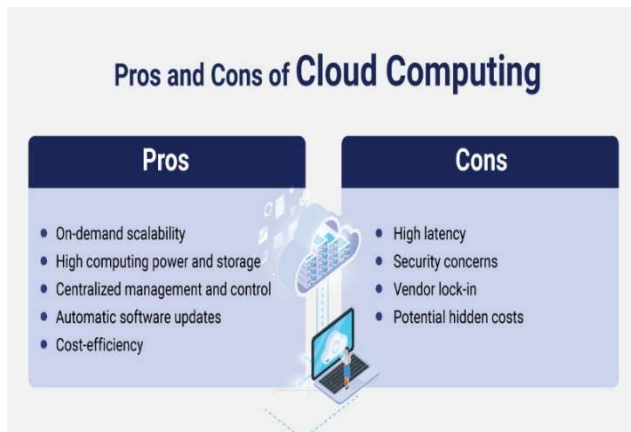
19

The action mentioned in "skip the exclamation marks" (6th paragraph) is semantically equivalent to

- (A) bring in.
- (B) carry on.
- (C) put back.
- (D) leave out.
- (E) move around.

READ TEXT II AND ANSWER NEXT TWO QUESTIONS.

TEXT II



From: https://www.rishabhsoft.com/wp-content/uploads/2024/01/Blog\Pros-and-Cons-of-Cloud-Computing_Image.jpg

20

As regards the pros and cons of cloud computing, analyse the assertions below:

- I. It may adjust resources depending on the needs.
- II. It can reduce spending according to usage.
- III. It is immune to all types of cyberattacks.

Choose the correct answer.

- (A) Only I is correct.
- (B) Both I and II are correct.
- (C) Both I and III are correct.
- (D) Both II and III are correct.
- (E) All three assertions are correct.

21

The term "vendor lock-in" indicates the customer is in a situation that is

- (A) binding.
- (B) negligible.
- (C) forbidden.
- (D) disengaged.
- (E) unrestricted.

READ TEXT III AND ANSWER NEXT THREE QUESTIONS.

TEXT III



"We were looking for a non-hackable, energy efficient data center. Thanks for the file cabinet."

From: https://images.cartoonstock.com/lowres/computers-file_cabinet-file-paperwork-unhackable-security-CS630961_low.jpg

22

The saying that best expresses the point of this cartoon is

- (A) beauty is in the eye of the beholder.
- (B) the new broom sweeps clean.
- (C) every cloud has a silver lining.
- (D) health is wealth.
- (E) old is gold.

23

The phrasal verb "looking for" can be replaced, without significant change of meaning, by

- (A) seeking.
- (B) finding.
- (C) picking.
- (D) acquiring.
- (E) surveying.

24

In the cartoon, the adjective "non-hackable" refers to a system that cannot be

- (A) set up.
- (B) turned off.
- (C) shut down.
- (D) broken into.
- (E) brought back.

Raciocínio lógico-matemático

25

Em determinado período de monitoramento, a equipe responsável por um centro de processamento adotou a regra a seguir.

“Se a taxa de ocupação dos servidores for superior a 80%, então as rotinas não prioritárias serão transferidas para o período noturno.”

Considerando a lógica proposicional, essa regra é logicamente equivalente a

- (A) A taxa de ocupação dos servidores é inferior a 80% ou as rotinas não prioritárias serão transferidas para o período noturno.
- (B) A taxa de ocupação dos servidores é menor ou igual a 80% ou as rotinas não prioritárias serão transferidas para o período noturno.
- (C) A taxa de ocupação dos servidores é menor ou igual a 80% e as rotinas não prioritárias serão transferidas para o período noturno.
- (D) A taxa de ocupação dos servidores é superior a 80% ou as rotinas não prioritárias não serão transferidas para o período noturno.
- (E) Se as rotinas não prioritárias forem transferidas para o período noturno, então a taxa de ocupação dos servidores será superior a 80%.

26

Em relação às alterações processadas por determinada organização, considere verdadeiras as seguintes afirmações:

- toda alteração classificada como emergencial foi implantada no mesmo dia em que foi autorizada;
- toda alteração implantada no mesmo dia em que foi autorizada foi submetida a uma revisão posterior.

Uma dessas alterações, denominada *Alfa*, não foi submetida a uma revisão posterior.

Com base exclusivamente nessas informações, conclui-se necessariamente que a alteração *Alfa*

- (A) não foi autorizada.
- (B) foi autorizada, mas não foi implantada.
- (C) não foi classificada como emergencial.
- (D) foi implantada em dia posterior ao de sua autorização.
- (E) foi classificada como emergencial, mas não foi implantada no mesmo dia em que foi autorizada.

27

Um sistema utiliza um contador que apresenta inicialmente o número 1. A cada comando executado, o número exibido é substituído por um dos seguintes valores:

- o número exibido acrescido de uma unidade;
- o dobro do número exibido.

O menor número de comandos necessários para que o contador passe a exibir o número 100 é

- (A) 6.
- (B) 7.
- (C) 8.
- (D) 9.
- (E) 10.

28

Um sistema de monitoramento utiliza dois sensores e pode operar em modo de manutenção.

Considere as proposições:

- p : o sensor A detecta uma anomalia;
- q : o sensor B detecta uma anomalia; e
- r : o modo de manutenção está ativado.

O sistema aciona um alerta quando pelo menos um dos sensores detecta uma anomalia e o modo de manutenção não está ativado, ou quando ambos os sensores detectam uma anomalia.

A proposição que representa o acionamento do alerta é

- (A) $[(p \vee q) \wedge \neg r] \vee (p \wedge q)$
- (B) $[(p \wedge q) \wedge \neg r] \vee (p \wedge q)$
- (C) $[(p \vee q) \wedge r] \vee (p \wedge \neg q)$
- (D) $[(p \vee q) \vee \neg r] \vee (p \wedge q)$
- (E) $[(p \vee q) \wedge \neg r] \wedge (p \vee q)$

29

Considere o conjunto não vazio dos equipamentos examinados durante uma conferência patrimonial. Ao término do procedimento, foi apresentada a afirmação a seguir.

“Todos os equipamentos examinados receberam etiqueta patrimonial e tiveram sua localização registrada no inventário.”

Assinale a opção que apresenta a negação lógica dessa afirmação.

- (A) Pelo menos um equipamento examinado recebeu etiqueta patrimonial e teve sua localização registrada no inventário.
- (B) Nenhum equipamento examinado recebeu etiqueta patrimonial nem teve sua localização registrada no inventário.
- (C) Nenhum equipamento examinado recebeu simultaneamente etiqueta patrimonial e teve sua localização registrada no inventário.
- (D) Pelo menos um equipamento examinado não recebeu etiqueta patrimonial e não teve sua localização registrada no inventário.
- (E) Pelo menos um equipamento examinado não recebeu etiqueta patrimonial ou não teve sua localização registrada no inventário.

Atualidades e Inteligência Artificial

30

A expansão dos modelos de inteligência artificial tem impulsionado a construção de data centers de grande porte. Isso tem suscitado conflitos em razão dos impactos socioambientais desses empreendimentos, sobretudo quando instalados em regiões vulneráveis ou próximas a comunidades tradicionais.

Assinale a opção que identifica o impacto ambiental mais diretamente relacionado aos data centers.

- (A) A redução da cobertura vegetal pela instalação de cabos, cuja transmissão exige faixas desmatadas entre o data center e seus usuários.
- (B) O descarte frequente de eletrônicos junto aos resíduos domiciliares, o que sobrecarrega os serviços municipais de coleta urbana.
- (C) A contaminação do solo por fluidos de refrigeração, liberados regularmente durante a transferência de dados entre os equipamentos.
- (D) O aumento das emissões locais de gases pela atividade de processamento, que depende da combustão realizada no interior dos servidores.
- (E) A pressão sobre a disponibilidade de água e energia, demandadas pelo processamento contínuo e por certos sistemas de refrigeração.

31

Em 28 de fevereiro de 2026, os Estados Unidos e Israel iniciaram uma ofensiva militar conjunta contra o Irã. No centro do conflito está o Estreito de Ormuz, passagem marítima utilizada por importantes exportadores de petróleo e gás do Golfo Pérsico e com poucas rotas alternativas.

Assinale a opção que apresenta o impacto imediato dessa situação sobre o comércio internacional de energia.

- (A) O redirecionamento dos navios pelo Canal de Suez, que passou a substituir Ormuz nas rotas comerciais entre o Golfo e a Ásia.
- (B) O deslocamento dos embarques para oleodutos terrestres, preservando o volume de petróleo destinado aos mercados internacionais.
- (C) A retração do tráfego de petroleiros, acompanhada pela elevação dos custos de transporte e das cotações internacionais.
- (D) A concentração dos efeitos nas exportações iranianas, sem alteração relevante nos embarques realizados pelos demais países do Golfo.
- (E) A compensação imediata da oferta pelos demais membros da OPEP, que ampliaram a produção para estabilizar o mercado internacional.

32

Em julho de 2026, o Brasil acionou a Organização Mundial do Comércio (OMC) para contestar as tarifas impostas pelos Estados Unidos a produtos brasileiros. Os norte-americanos aceitaram participar da fase de consultas, e a China solicitou ingresso nessa etapa por considerar que também possui interesses comerciais afetados.

Nesse caso, a atuação da OMC pode

- (A) suspender os Estados Unidos da organização até que o governo retire as tarifas questionadas pelo Brasil.
- (B) determinar o pagamento de indenizações às empresas brasileiras afetadas pelas tarifas norte-americanas.
- (C) instaurar um painel para avaliar se as tarifas norte-americanas são compatíveis com os acordos em vigor.
- (D) impor aos produtos norte-americanos tarifas equivalentes às aplicadas pelos Estados Unidos às exportações brasileiras.
- (E) cancelar as tarifas norte-americanas e restabelecer as condições comerciais vigentes antes da decisão do governo.

33

Acerca dos conceitos de Inteligência Artificial (IA), analise as afirmativas a seguir.

- I. As redes neurais artificiais são sistemas computacionais distribuídos compostos de unidades de processamento interconectadas que computam funções matemáticas.
- II. No contexto de IA, o processamento de linguagem natural envolve o desenvolvimento de métodos computacionais para permitir que sistemas reproduzam a habilidade humana de compreender a linguagem falada ou escrita.
- III. O teste de Turing é uma metodologia utilizada para medir a capacidade de processamento numérico e a velocidade de cálculo de um sistema de IA.

Está correto o que se afirma em

- (A) I, apenas.
- (B) I e II, apenas.
- (C) I e III, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

34

No contexto da Inteligência Artificial (IA), os modelos de linguagem de grande porte (*Large Language Models*) pertencem à categoria dos modelos generativos.

O funcionamento básico desses modelos, com objetivo de geração de textos, fundamenta-se em

- (A) consultar um banco de dados fixo, contendo respostas pré-programadas por um humano que correspondam à entrada recebida.
- (B) limitar-se à correção ortográfica e gramatical de textos já existentes, sendo incapazes de estruturar novas frases ou argumentos de forma autônoma.
- (C) realizar buscas em tempo real em enciclopédias virtuais, para coletar trechos exatos de artigos acadêmicos que correspondam à pergunta do usuário.
- (D) calcular estatisticamente a probabilidade de ocorrência das próximas palavras ou termos em uma sequência, baseando-se nos padrões identificados em seu treinamento.
- (E) executar um escaneamento do disco rígido do computador do usuário, para identificar preferências pessoais e replicar o estilo de escrita do proprietário da máquina.

35

Sobre métodos de Aprendizado de Máquina, analise as afirmativas a seguir e assinale (V) para a verdadeira e (F) para a falsa.

- () Na aprendizagem bayesiana, as previsões sobre um dado novo são feitas considerando todo o espaço de hipóteses, ponderadas por suas probabilidades.
- () No aprendizado por reforço ativo, o agente deve decidir autonomamente quais ações tomar em cada estado para minimizar sua função de utilidade a longo prazo.
- () No aprendizado profundo, as hipóteses são estruturadas como redes complexas de funções algébricas interconectadas por pesos numéricos ajustáveis.

As afirmativas são, respectivamente,

- (A) V – V – F.
- (B) F – V – V.
- (C) V – F – V.
- (D) V – V – V.
- (E) F – F – V.

Legislação Acerca de Segurança da Informação e Proteção de Dados

36

Uma universidade federal, visando modernizar e dar celeridade ao processo de inscrição de seu novo concurso público, decidiu desenvolver um sistema de inscrição inteligente. O objetivo é que o candidato digite apenas o seu CPF e, a partir de uma integração de banco de dados, o sistema preencha automaticamente informações cadastrais básicas, como nome completo e data de nascimento, oriundas do Cadastro Base do Cidadão (ou da base da Receita Federal).

Considerando as normativas de compartilhamento de dados na Administração Pública, a Lei de Acesso à Informação (Lei nº 12.527/2011) e a Lei Geral de Proteção de Dados (Lei nº 13.709/2018), assinale a opção que indica corretamente como a universidade deve proceder para ter acesso a esses dados para o seu sistema.

- (A) Deve formular um pedido via Serviço de Informações ao Cidadão (SIC) da Receita Federal com base na transparência passiva da LAI, solicitando o envio da base de dados completa dos cidadãos para alimentar o seu sistema local de inscrições.
- (B) Está impedida de realizar essa integração automatizada, pois o cruzamento de dados pessoais (como CPF, nome e data de nascimento) entre órgãos da Administração Pública exige, obrigatoriamente, o consentimento prévio, expresso e inequívoco de cada cidadão titular dos dados, não havendo exceção legal.
- (C) Deve solicitar que o ente gestor do banco de dados publique a relação de CPFs, nomes e datas de nascimento no Portal da Transparência, garantindo a transparência ativa da LAI, para que então o sistema da universidade possa extrair os dados livremente por meio de ferramentas de automação.
- (D) Deve utilizar os mecanismos oficiais de interoperabilidade governamental (ou firmar convênio/acordo de compartilhamento de dados), justificando que o acesso aos dados pessoais é estritamente necessário para o atendimento da finalidade pública e para a execução de suas competências legais, dispensando-se, neste caso, o consentimento do candidato.
- (E) Deve ser concedido, de forma ampla e irrestrita, o acesso ao banco de dados a todos os servidores da universidade, sob a justificativa de que, por serem dados geridos pelo Estado, aplica-se o princípio constitucional da publicidade absoluta sobre qualquer informação não classificada como ultrassecreta.

37

Roberto é Analista de Tecnologia da Informação em um órgão federal e, em razão de seu cargo, possui credenciais legítimas (*login*, senha e *token*) com autorização expressa para acessar, consultar e gerir o banco de dados sigiloso de licitações do órgão. Aproveitando-se de seu acesso regular e cotidiano, Roberto extraiu dados sigilosos de propostas comerciais de empresas concorrentes e os vendeu para uma organização criminosa, visando fraudar um certame.

Considerando estritamente a tipificação do crime de "*Invasão de Dispositivo Informático*" (Art. 154-A do Código Penal), introduzido pela Lei nº 12.737/2012, sobre a conduta de Roberto assinale a afirmativa correta.

- (A) Roberto responderá pelo crime de invasão de dispositivo informático, na sua forma qualificada, pois a lei prevê que o repasse de informações sigilosas a terceiros configura o delito mesmo quando o agente possui credenciais legítimas de acesso.
- (B) A conduta configura o crime de invasão de dispositivo informático apenas se Roberto tiver precisado instalar alguma vulnerabilidade (*vírus* ou *malware*) no sistema para baixar os arquivos; caso contrário, a conduta é atípica no Direito Penal brasileiro.
- (C) Roberto praticou o crime de invasão de dispositivo informático com causa de aumento de pena de um terço à metade, devendo responder por meio de ação penal pública condicionada à representação do Ministério Público, por se tratar de crime cibernético.
- (D) Roberto responderá pelo crime da Lei nº 12.737/2012 porque o crime se consumou no momento em que ele conectou o banco de dados à rede mundial de computadores (*internet*) para extrair as propostas comerciais, configurando a violação de segurança do órgão.
- (E) A conduta de Roberto não configura o crime de invasão de dispositivo informático (Art. 154-A do CP), pois a elementar do tipo penal exige que a ação ocorra "sem autorização expressa ou tácita". Como Roberto possuía acesso legítimo, ele responderá por outros crimes (como violação de sigilo funcional) mas não por este, amparado na Lei nº 12.737/2012.

38

Nos termos do Marco Civil da Internet (Lei nº 12.965/2014), a lei estabelece obrigações claras sobre a guarda de dados dos usuários para equilibrar a privacidade e a segurança.

Sobre os deveres de retenção de dados aplicáveis aos diferentes tipos de provedores (conexão e aplicação), assinale a afirmativa correta.

- (A) O provedor de conexão à internet (como as operadoras de telecomunicações) deve manter os registros de conexão, sob sigilo e em ambiente controlado, pelo prazo obrigatório de 6 (seis) meses.
- (B) O provedor de conexão à internet tem o dever de guardar, além dos registros de conexão, o histórico detalhado de navegação do usuário (registros de acesso a aplicações) pelo prazo de 1 (um) ano.
- (C) O provedor de aplicações de internet é proibido de guardar registros de acesso preventivamente, só podendo reter esses dados caso receba uma ordem judicial prévia determinando a guarda.
- (D) O provedor de aplicações de internet (como redes sociais e portais), desde que constituído de forma empresarial e com fins econômicos, deve manter os registros de acesso a aplicações de internet pelo prazo obrigatório de 6 (seis) meses.
- (E) Para facilitar investigações criminais, o Marco Civil da Internet unificou as regras, determinando que tanto provedores de conexão quanto de aplicação guardem todos os registros de uso e conexão pelo prazo comum de 5 (cinco) anos.

39

Um órgão federal está organizando um concurso público para provimento de cargos de nível superior e precisa publicar o resultado preliminar das provas objetivas e a lista de candidatos que concorrem às vagas reservadas para Pessoas com Deficiência (PcD), bem como daqueles que solicitaram atendimento especial por razões de saúde. De forma a garantir a transparência do certame, a Comissão do Concurso planejou publicar na internet uma planilha contendo o nome completo, o número completo do CPF, o relatório médico detalhado e o diagnóstico CID (Classificação Internacional de Doenças) de cada candidato aprovado nessa cota.

O Encarregado pelo Tratamento de Dados Pessoais (DPO) do órgão interveio, ressaltando que o diagnóstico de saúde constitui dado pessoal sensível e que a exposição integral de dados de identificação extrapola o limite necessário para a finalidade de controle social do concurso.

Considerando os preceitos da Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018 – LGPD), assinale a opção que apresenta a solução correta que o órgão deve adotar para conciliar a transparência do concurso com a proteção de dados dos candidatos:

- (A) A Comissão do Concurso pode manter a publicação integral com laudos e CPFs completos sem qualquer alteração, tendo em vista que a LGPD não se aplica às autarquias federais quando agem no exercício do poder de império e na realização de concursos públicos.
- (B) O órgão deve decretar o sigilo absoluto de todo o resultado da cota de Pessoas com Deficiência, divulgando apenas o resultado geral da ampla concorrência no Portal da Transparência, encaminhando o resultado da cota de forma individual e privada ao e-mail de cada candidato.
- (C) O órgão deve aplicar os princípios da necessidade e da minimização de dados, publicando a lista oficial apenas com informações indispensáveis (como nome/inscrição, CPF parcialmente mascarado e a condição de 'deferido/aprovado' na cota), franqueando o acesso ao laudo médico detalhado estritamente ao próprio candidato e à banca examinadora.
- (D) O órgão deve exigir o consentimento prévio, escrito e inequívoco de todos os candidatos participantes da cota para publicar seus laudos e diagnósticos médicos completos, sob pena de eliminação sumária do concurso caso o candidato se recuse a assinar o termo de autorização.
- (E) O órgão deve promover a anonimização irreversível e definitiva de todos os dados dos candidatos na lista pública e no sistema interno do concurso, de modo que nem mesmo a comissão examinadora consiga vincular a nota do exame ao indivíduo titular.

40

Com base na Lei Geral de Proteção de Dados Pessoais – LGPD, (Lei nº 13.709/2018), no que diz respeito à segurança e às boas práticas, à fiscalização e à Autoridade Nacional de Proteção de Dados (ANPD), analise as afirmativas a seguir.

- I. Os controladores e operadores podem formular regras de boas práticas e de governança relacionadas ao tratamento de dados pessoais.
- II. Compete à ANPD, entre outras atribuições previstas na LGPD, zelar pela proteção dos dados pessoais, fiscalizar e aplicar sanções nos casos de tratamento realizado em descumprimento à legislação e promover o conhecimento sobre proteção de dados pessoais e privacidade.
- III. Na aplicação das sanções administrativas, a ANPD considera a gravidade da infração e o dano causado, desconsiderando a adoção de mecanismos internos de prevenção e de medidas corretivas pelo infrator.

Está correto o que se afirma em

- (A) I, apenas.
- (B) I e II, apenas.
- (C) I e III, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

Conhecimentos Específicos

41

Os bancos de dados NoSQL (*Not Only SQL*) ganharam popularidade por sua capacidade de lidar com grandes volumes de dados não estruturados ou semiestruturados. Eles são frequentemente classificados em quatro categorias principais, com base no modelo empregado para representar e organizar os dados.

Assinale a opção que apresenta corretamente essas quatro categorias.

- (A) Chave-Valor, Orientado a Documentos, Família de Colunas Amplas e Orientado a Grafos.
- (B) Estruturado, Semiestruturado, Não estruturado e Dinâmico.
- (C) Relacional, Objeto-Relacional, Hierárquico e de Rede.
- (D) Relacional, Orientado a Documentos, Hierárquico e Orientado a Objetos.
- (E) SQL, MySQL, PostgreSQL e Oracle.

42

Considere a tabela *Projeto_Engenheiro* com a seguinte estrutura de colunas: *CodigoEngenheiro*, *CodigoProjeto*, *HorasAlocadas*, *NomeEngenheiro* e *EspecialidadeEngenheiro*.

Sabe-se que a chave primária da tabela é composta conjuntamente por (*CodigoEngenheiro*, *CodigoProjeto*). Ao analisar as dependências funcionais, verifica-se que *NomeEngenheiro* e *EspecialidadeEngenheiro* são funcionalmente determinados apenas por *CodigoEngenheiro*, que corresponde à parte da chave primária composta.

Assinale a opção que avalia corretamente a adequação dessa estrutura à 2ª Forma Normal (2FN).

- (A) a tabela já se encontra na 2FN, dado que a única exigência estrutural desse nível de normalização é a existência de uma chave primária composta bem definida.
- (B) a tabela já se encontra na 2FN, uma vez que a dependência transitiva existente deve ser sanada apenas na 3FN.
- (C) a estrutura viola a 2FN, devendo-se corrigir o problema movendo o atributo de horas alocadas para uma tabela à parte.
- (D) a estrutura viola a 2FN devido à dependência parcial dos atributos do engenheiro.
- (E) a estrutura viola a 2FN, enquanto a chave composta não for substituída por um identificador sequencial simples.

43

Considere as tabelas *Cliente* e *Pedido*, em que *Cliente.CodigoCliente* é a chave primária, e *Pedido.CodigoCliente* é uma chave estrangeira que referencia essa coluna.

Analise as duas definições a seguir:

I - FOREIGN KEY (*CodigoCliente*)
REFERENCES *Cliente*(*CodigoCliente*)
ON DELETE CASCADE

II - FOREIGN KEY (*CodigoCliente*)
REFERENCES *Cliente*(*CodigoCliente*)
ON DELETE NO ACTION

Assinale a opção que apresenta corretamente a diferença entre os comportamentos das duas definições.

- (A) Na definição I, a exclusão de um pedido pode provocar a exclusão do cliente; na definição II, essa operação é impedida.
- (B) Na definição I, a exclusão de um cliente provoca a exclusão dos pedidos que o referenciam; na definição II, a exclusão do cliente é impedida enquanto existirem pedidos que o referenciem.
- (C) Na definição I e na definição II, a exclusão do cliente produz o mesmo resultado, pois CASCADE e NO ACTION apenas documentam a intenção da regra.
- (D) Na definição I, os pedidos permanecem e passam a referenciar automaticamente outro cliente; na definição II, os pedidos são excluídos juntamente com o cliente.
- (E) Na definição I, a exclusão do cliente é permitida apenas quando não houver pedidos associados; na definição II, os pedidos associados são removidos automaticamente.

44

Uma organização possui um *data warehouse*, utilizado para produzir indicadores financeiros a partir de dados provenientes de diferentes sistemas operacionais.

Durante uma auditoria, a equipe responsável analisou algumas informações relacionadas ao ambiente.

- I. A tabela *FATO_VENDAS* possui o atributo *valor_liquido*, definido como um número decimal com duas casas decimais, e esse atributo não pode receber valores nulos.
- II. O indicador “*venda líquida*” representa o valor bruto das vendas após a dedução dos descontos e das devoluções, conforme a regra estabelecida pela organização.
- III. O processo de carga da tabela *FATO_VENDAS* é executado diariamente às 2h. Na última execução, foram extraídos 1.253.428 registros, dos quais 1.251.936 foram carregados com sucesso.

Considerando as características dos metadados em ambientes de *data warehouse*, assinale a opção que apresenta a classificação correta das informações I, II e III, respectivamente.

- (A) de negócio; de processo; técnico.
- (B) de negócio; técnico; de processo.
- (C) de processo; de negócio; técnico.
- (D) técnico; de negócio; de processo.
- (E) técnico; de processo; de negócio.

45

Considere um modelo dimensional destinado à análise de vendas, composto pelas tabelas *FatoVenda*, *DimProduto*, *DimCliente* e *DimData*.

A tabela *FatoVenda* possui as colunas: *CodigoProduto*, *CodigoCliente*, *CodigoData*, *Quantidade* e *ValorVenda*. As três primeiras colunas referenciam, respectivamente, as chaves das tabelas *DimProduto*, *DimCliente* e *DimData*.

Assinale a alternativa que contém a afirmação correta sobre a forma como os relacionamentos entre a tabela fato e as dimensões são materializados nesse modelo.

- (A) As dimensões devem armazenar as chaves estrangeiras que identificam as ocorrências correspondentes na tabela fato, estabelecendo a associação entre as tabelas.
- (B) Cada dimensão deve possuir uma chave primária composta pela combinação de sua própria chave e das chaves das demais dimensões utilizadas pela tabela fato.
- (C) Deve ser criada uma tabela associativa para cada relacionamento entre *FatoVenda* e uma dimensão, contendo as chaves primárias das duas tabelas relacionadas.
- (D) Os relacionamentos entre fato e dimensões são definidos apenas conceitualmente e não precisam ser representados por chaves nas tabelas do modelo dimensional.
- (E) Os relacionamentos são materializados pelas chaves estrangeiras existentes na tabela fato, que referenciam as chaves das respectivas dimensões.

46

Uma instituição pretende implantar uma aplicação projetada para armazenar informações sobre itens pertencentes ao seu patrimônio. Os itens pertencem a diferentes categorias e podem apresentar conjuntos distintos de atributos.

Além disso, novos tipos de itens e novos atributos podem ser incorporados frequentemente, sem que seja desejável modificar continuamente a estrutura global do banco de dados.

A aplicação realizará principalmente operações de inserção, consulta e atualização de itens completos, sendo conveniente recuperar as informações de cada item como uma unidade de dados.

Considerando essas características, assinale a opção que apresenta a abordagem de banco de dados mais adequada ao cenário descrito.

- (A) Objeto-relacional, utilizando tipos definidos pelo usuário para representar os diferentes itens.
- (B) Orientado a documentos, permitindo armazenar documentos com estruturas de atributos distintas e flexíveis.
- (C) Orientado a grafos, utilizando nós e arestas para representar os itens e seus relacionamentos.
- (D) Relacional, utilizando uma tabela única para os produtos, com colunas opcionais para acomodar os atributos específicos de cada categoria.
- (E) Relacional, utilizando tabelas normalizadas e relacionamentos entre elas.

47

Considere o modelo relacional de uma transportadora composto pelas tabelas *Veiculo* e *Viagem*.

A tabela *Veiculo* possui as colunas *CodigoVeiculo* (PK) e *KmAtual*, contendo 50 veículos cadastrados. O atributo *kmAtual* aceita valores nulos. Já a tabela *Viagem* possui as colunas *CodigoViagem* (PK), *CodigoVeiculo* (FK) e *KmFinal*, contendo os registros de viagens finalizadas. Sabe-se que apenas 30 desses veículos possuem registros na tabela *Viagem* (os outros 20 são veículos recém-adquiridos que nunca viajaram).

Um desenvolvedor executou a seguinte instrução SQL no banco de dados:

```
UPDATE Veiculo V
SET KmAtual =
(SELECT MAX(Vi.KmFinal)
 FROM Viagem Vi
 WHERE Vi.CodigoVeiculo = V.CodigoVeiculo
);
```

Assinale a opção que descreve corretamente o resultado da execução dessa instrução.

- (A) Ocorrerá um erro de sintaxe, pois uma subconsulta não pode ser utilizada diretamente para definir o valor de uma coluna em uma instrução UPDATE.
- (B) Ocorrerá um erro de sintaxe, pois o padrão ANSI SQL não permite o uso de funções de agregação, como MAX, dentro de uma subconsulta na cláusula SET.
- (C) Os 30 veículos que possuem registros na tabela *Viagem* receberão a sua maior quilometragem, enquanto os 20 restantes manterão o valor original de *KmAtual* inalterado.
- (D) Os 30 veículos que possuem registros na tabela *Viagem* receberão a sua maior quilometragem, enquanto os 20 restantes terão o valor de *KmAtual* sobrescrito como nulo (NULL).
- (E) Os 30 veículos que possuem registros na tabela *Viagem* receberão a sua maior quilometragem, enquanto os 20 veículos restantes receberão automaticamente o valor 0 (zero) na coluna *KmAtual*, pois MAX() converte o resultado de um conjunto vazio para zero durante uma operação UPDATE.

48

A modelagem conceitual de bancos de dados é uma representação abstrata do negócio que visa identificar as entidades envolvidas e os relacionamentos entre elas.

No mapeamento de um relacionamento N:M identificado no modelo conceitual para o modelo relacional, deve-se

- (A) adicionar uma coluna do tipo *array* em uma das tabelas envolvidas.
- (B) converter uma das entidades em um atributo multivalorado na outra entidade.
- (C) criar uma tabela associativa contendo as chaves primárias das entidades originais como chaves estrangeiras.
- (D) inserir a chave primária de qualquer uma das duas entidades diretamente como chave estrangeira na outra.
- (E) mesclar os atributos das duas entidades em uma única tabela para eliminar a necessidade de relacionamento.

49

Em uma rotina diária de Extração, Transformação e Carga (ETL) destinada à atualização de um *Data Warehouse* (DW) da DATAPREV, são extraídos 120.000 registros de um sistema de atendimento e 80.000 registros de um sistema de gestão de serviços.

Durante a etapa de transformação, 10% do total de registros extraídos são identificados como duplicados e eliminados. Em seguida, 5% dos registros restantes são rejeitados por não atenderem às regras de validação estabelecidas.

Dos registros válidos ao final dessas etapas, 30% correspondem a registros já existentes no DW e são atualizados, enquanto os demais são inseridos como novos registros.

O número de novos registros inseridos no *Data Warehouse*, ao final da execução do ETL, é de

- (A) 71820.
- (B) 119700.
- (C) 126000.
- (D) 140000.
- (E) 171000.

50

Na análise de 10.000 registros de incidentes de serviços de TIC da DATAPREV, foi utilizada a técnica de regras de associação para identificar ocorrências que aparecem conjuntamente nos registros. A equipe deseja avaliar a regra $A \rightarrow B$.

Dentre os registros, destacam-se:

- 2.000 incidentes apresentaram a ocorrência A;
- 1.500 incidentes apresentaram a ocorrência B; e
- 1.200 incidentes apresentaram simultaneamente as ocorrências A e B.

Considerando as métricas empregadas em regras de associação, os valores de suporte e confiança para a regra $A \rightarrow B$ são, respectivamente,

- (A) 12% e 60%.
- (B) 12% e 80%.
- (C) 15% e 60%.
- (D) 20% e 60%.
- (E) 15% e 80%.

51

No desenvolvimento de modelos de Aprendizado de Máquina, a análise do espaço de hipóteses é importante para compreender a capacidade de generalização de um modelo.

Acerca dos conceitos de viés (*bias*), variância (*variance*) e dos fenômenos de subajuste (*underfitting*) e sobreajuste (*overfitting*), analise os itens a seguir:

- I. Um viés alto pode ser provocado por restrições estruturais no próprio espaço de hipóteses, induzindo ao subajuste do modelo.
- II. Uma das causas de sobreajuste do modelo é uma variância alta, prejudicando seu desempenho em novos dados mesmo quando o desempenho no treinamento é satisfatório.
- III. O compromisso viés-variância (*bias-variance trade-off*) estabelece um dilema entre optar por hipóteses mais complexas com baixo viés ou hipóteses mais simples com baixa variância.

Está correto o que se afirma em

- (A) I, apenas.
- (B) I e II, apenas.
- (C) I e III, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

52

Um cientista de dados possui um conjunto de dados com exatamente 1.500 amostras e decide avaliar seu modelo de aprendizado de máquina utilizando a técnica de *k-Fold Cross-Validation* (validação cruzada com k-repetições).

Considerando o funcionamento padrão dessa técnica, o número total de vezes em que o modelo será treinado e a quantidade exata de amostras utilizadas para treinamento em cada uma das rodadas (*folds*) de validação são, para $k = 5$, respectivamente,

- (A) 4 vezes e 300 amostras.
- (B) 4 vezes e 1.200 amostras.
- (C) 5 vezes e 300 amostras.
- (D) 5 vezes e 1.200 amostras.
- (E) 5 vezes e 1.500 amostras.

53

Uma das métricas utilizadas no algoritmo K-NN (*K-Nearest Neighbors*), no desenvolvimento de modelos de Aprendizado de Máquina, é a distância de Minkowski, definida pela norma- p , onde p é a ordem.

Nesse contexto, analise os itens a seguir.

- I. Quando $p = 1$, a distância de Minkowski calcula o número de atributos diferentes entre dois vetores no espaço.
- II. Quando $p = 2$, a distância de Minkowski calcula a distância euclidiana entre dois vetores no espaço.
- III. Quando $p = K$, a distância de Minkowski calcula a distância média para os K vizinhos mais próximos.

Está correto o que se afirma em

- (A) I, apenas.
- (B) II, apenas.
- (C) I e II, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

54

O problema da alta dimensionalidade, comum em modelos de Aprendizado de Máquina, pode ser tratado com técnicas de redução de dimensionalidade.

Sobre esse assunto, analise as afirmativas a seguir e assinale (V) para a verdadeira e (F) para a falsa.

- () A redução de dimensionalidade tem como principal objetivo escolher dados apenas do espaço bidimensional.
- () As técnicas de agregação são aplicadas para descartar parte dos atributos originais por meio da criação de grupos de atributos.
- () A técnica de Análise de Componentes Principais (*Principal Component Analysis*) permite eliminar redundâncias geradas pela alta correlação entre as variáveis originais.

As afirmativas são, respectivamente,

- (A) V – F – V.
- (B) V – F – F.
- (C) F – V – V.
- (D) F – F – F.
- (E) F – F – V.

55

O algoritmo *K-Means* (K-Médias, em português), usado em Aprendizado de Máquina, pode ser descrito como um método

- (A) de classificação que calcula a probabilidade de um registro pertencer a uma classe, com base nas médias dos registros considerando K atributos.
- (B) que cria uma estrutura hierárquica com K nós de decisão, em que cada nó representa um teste em um atributo, e as médias dos valores das folhas definem as classificações.
- (C) que realiza previsões estimando a classe ou o valor de um novo registro, com base nas médias obtidas de seus K vizinhos mais próximos.
- (D) de particionamento, que agrupa dados iterativamente, associando cada ponto ao centróide mais próximo e recalculando a posição dos pontos centrais pela média dos pontos atribuídos.
- (E) não supervisionado, baseado em densidade que identifica agrupamentos de formatos arbitrários, ao isolar regiões com K pontos de alta concentração.

56

No contexto do Aprendizado de Máquina não supervisionado, os Modelos de Mistura Gaussiana (*Gaussian Mixture Models*) representam uma abordagem probabilística para a tarefa de agrupamento (*clustering*).

Acerca das características e do funcionamento dessa técnica, analise os itens a seguir.

- I. De acordo com a técnica, cada gaussiana no modelo de mistura possui três parâmetros principais: média, peso e matriz de covariância.
- II. O modelo calcula para cada instância a probabilidade de ela ter sido gerada por cada um dos componentes da mistura.
- III. Os agrupamentos gerados por essa técnica são matematicamente limitados a formatos esféricos, impossibilitando a modelagem de distribuições elípticas ou com volumes distintos no espaço de dados.

Está correto o que se afirma em

- (A) I, apenas.
- (B) II, apenas.
- (C) I e II, apenas.
- (D) I e III, apenas.
- (E) II e III, apenas.

57

A equipe de IA de uma *FinTech* foi encarregada de refatorar o *pipeline* de treinamento da rede neural responsável pela detecção de fraudes em transações de cartão de crédito. Atualmente, o sistema processa cerca de 50 milhões de transações diárias. No relatório técnico entregue à diretoria, os engenheiros precisam justificar a escolha da arquitetura de treinamento, detalhando as vantagens e o papel de cada algoritmo envolvido na otimização da rede.

Considerando o funcionamento, a eficiência computacional e as vantagens comparativas dos algoritmos de otimização, assinale a afirmativa correta.

- (A) O Método do Gradiente (Gradiente Descendente em Lote) é o mais eficiente para grandes volumes de dados, pois o Retropropagação (*Backpropagation*) calcula a perda de forma estocástica a cada transação, eliminando o uso do Método do Gradiente Estocástico (SGD).
- (B) O Método do Gradiente Estocástico (SGD) atualiza os pesos da rede com base em poucas amostras, reduzindo a memória necessária e o tempo por iteração em relação ao Método do Gradiente, enquanto a Retropropagação (*Backpropagation*) é o algoritmo responsável por calcular com eficiência os gradientes da função de perda em relação a cada peso.
- (C) A Retropropagação (*Backpropagation*) é um algoritmo de otimização alternativo ao Método do Gradiente Estocástico (SGD) que dispensa o cálculo de derivadas parciais, permitindo que o Método do Gradiente encontre o mínimo global sem risco de ficar preso em mínimos locais.
- (D) O Método do Gradiente garante atualização rápida a cada transação individual, sendo ideal para sistemas em tempo real, enquanto o Método do Gradiente Estocástico (SGD) requer o processamento de todo o banco de dados antes de atualizar qualquer parâmetro da rede.
- (E) O Método do Gradiente Estocástico (SGD) substitui a necessidade da Retropropagação (*Backpropagation*), pois calcula diretamente o erro da camada de saída e ajusta a camada de entrada, sem precisar propagar o sinal de erro pelas camadas ocultas.

58

Uma equipe de Ciência de Dados construiu uma rede neural profunda para prever a taxa de cancelamento de clientes em uma grande plataforma. Durante a fase de validação, os engenheiros notaram que o modelo apresentava uma acurácia próxima de 99% no conjunto de dados de treino, mas o desempenho despencava para 62% no conjunto de testes. Para combater esse problema de sobreajuste, a liderança técnica solicitou a aplicação de técnicas de regularização baseadas na penalização da função de perda por meio de normas (L1 e L2), visando melhorar a capacidade de generalização do modelo e limitar o crescimento excessivo dos pesos.

Considerando o comportamento e o impacto teórico das penalizações pela norma L1 (Lasso) e pela norma L2 (Decaimento de Pesos) nas redes neurais artificiais, assinale a afirmativa correta.

- (A) A regularização L1 adiciona a soma dos quadrados dos pesos à função de perda, o que reduz os pesos de forma suave e contínua, enquanto a regularização L2 adiciona a soma dos valores absolutos, forçando muitos coeficientes a zerarem exatamente e gerando modelos mais esparsos.
- (B) O principal objetivo da norma L1 é eliminar o problema do desaparecimento do gradiente, enquanto a norma L2 é aplicada exclusivamente na camada de entrada para ignorar outliers do conjunto de dados de teste.
- (C) A regularização L2 penaliza o quadrado da magnitude dos pesos, empurrando os coeficientes em direção a zero sem zerá-los completamente; já a regularização L1 penaliza o valor absoluto dos pesos, o que pode zerar parâmetros irrelevantes e atuar como um mecanismo automático de seleção de recursos.
- (D) Ambas as regularizações (L1 e L2) alteram a função de ativação das camadas ocultas para funções lineares, garantindo que o erro da rede diminua tanto no treino quanto no teste sem afetar o valor dos pesos conectados.
- (E) Ao aplicar a norma L2, a rede neural gera um modelo com matrizes de pesos esparsas (com muitos valores nulos), o que torna a inferência mais rápida na nuvem, enquanto a norma L1 apenas distribui os pesos igualmente entre todos os neurônios da rede.

59

Um especialista foi contratado por uma multinacional do setor logístico para desenvolver um sistema automatizado de inspeção de qualidade. O objetivo é identificar pequenas fissuras e defeitos superficiais em caixas de papelão que passam em alta velocidade por uma esteira industrial.

Para resolver o problema, a equipe optou por utilizar uma Rede Neural Convolutiva (CNN). Durante a etapa de arquitetura do modelo, os engenheiros precisam definir o papel de cada camada na extração de características.

Considerando o funcionamento interno de uma CNN aplicada a esse cenário industrial, assinale a afirmativa correta.

- (A) As primeiras camadas convolucionais da rede são responsáveis por extrair representações abstratas e complexas, como o formato completo da caixa ou o tipo de embalagem, enquanto as camadas mais profundas identificam padrões simples, como bordas e variações bruscas de contraste.
- (B) A camada de agrupamento, como o *Max Pooling*, é utilizada para aumentar a resolução espacial do mapa de características (feature map), garantindo que a rede não perca detalhes minuciosos das fissuras ao longo do processamento.
- (C) A função de ativação ReLU (*Rectified Linear Unit*) é aplicada após as convoluções com o objetivo principal de transformar a saída da rede em uma distribuição de probabilidade para a classificação final dos defeitos.
- (D) Os filtros das camadas convolucionais operam localmente sobre a imagem, permitindo que a rede aprenda características com invariância à translação, ou seja, um defeito pode ser detectado independentemente de sua posição na imagem da caixa.
- (E) As camadas totalmente conectadas (*Fully Connected*) ocorrem no início da arquitetura para extrair as características espaciais da imagem antes que os filtros convolucionais façam a redução de dimensionalidade.

60

A equipe de Engenharia de Dados de uma multinacional está otimizando o pipeline de PLN de um sistema de atendimento ao cliente em tempo real. O objetivo é classificar os chamados recebidos em categorias de urgência e encaminhá-los ao suporte técnico correto.

Durante o desenvolvimento do modelo de inteligência artificial, os engenheiros debateram sobre o impacto de diferentes técnicas de pré-processamento e vetorização na captura do significado semântico e contextual das mensagens dos usuários.

Considerando os conceitos fundamentais de pré-processamento, vetorização e arquitetura de PLN aplicados a esse cenário, assinale a afirmativa correta.

- (A) A aplicação severa da técnica de *Stemming* é recomendada para substituir totalmente a Lematização em análises de contexto avançadas, pois o *Stemming* garante a criação de palavras gramaticalmente válidas e preserva o contexto exato do verbo conjugado.
- (B) O modelo tradicional de *Bag-of-Words* (BoW) é a escolha ideal para capturar ambiguidades e dependências de longo alcance na frase, uma vez que sua representação matricial preserva rigorosamente a ordem sequencial das palavras na mensagem.
- (C) A remoção de *Stop Words* deve ser aplicada obrigatoriamente em todas as tarefas de PLN, inclusive em modelos baseados em *Transformers* para tradução automática, pois palavras como "não" ou "com" apenas geram ruído e nunca alteram o sentido de uma sentença.
- (D) A métrica TF-IDF (*Term Frequency-Inverse Document Frequency*) gera *embeddings* densos e de baixa dimensão que aprendem as relações semânticas entre sinônimos, superando o algoritmo *Word2Vec* em tarefas de representação contextual contínua.
- (E) As técnicas de *Word Embeddings* densos, como *Word2Vec* ou *embeddings* de *Transformers*, mapeiam palavras ou tokens em vetores contínuos em um espaço multidimensional, permitindo que palavras semanticamente semelhantes fiquem geometricamente próximas umas das outras.

61

Uma equipe de Ciência de Dados da DATAPREV avalia *frameworks* de *deep learning* para dois projetos: um de pesquisa exploratória, com necessidade de depurar a arquitetura da rede a cada iteração, e outro já validado, a ser levado a produção com *pipeline* de treinamento reprodutível.

Avalie as afirmativas a seguir.

- I. O *PyTorch* utiliza grafos de computação dinâmicos, construídos durante a própria execução do código, o que facilita a depuração incremental do modelo, já que cada operação pode ser inspecionada logo após ser executada.
- II. O ecossistema do *TensorFlow* oferece ferramentas nativas voltadas à implantação de modelos em diferentes ambientes de produção, como servidores dedicados e dispositivos com recursos computacionais limitados.
- III. O uso de grafos de computação dinâmicos pelo *PyTorch* impede que modelos desenvolvidos nesse *framework* sejam utilizados para inferência em sistemas implantados para usuários finais.

Está correto o que se afirma em

- (A) I, apenas.
- (B) I e II, apenas.
- (C) I e III, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

62

Uma equipe de ciência de dados está estruturando a entrega de seus resultados: um aplicativo interativo para *stakeholders* não técnicos, gráficos estatísticos para relatórios e rotinas de agregação de dados compartilhadas com uma equipe parceira que trabalha em R.

Sobre as ferramentas envolvidas, considere as afirmativas a seguir.

- I. O *Streamlit* permite transformar um *script* de análise em *Python* em uma aplicação *web* interativa, sem exigir do desenvolvedor conhecimento de *HTML*, *CSS* ou *JavaScript*.
- II. O *Seaborn* é uma biblioteca de visualização estatística independente do *Matplotlib*, funcionando sem depender de sua infraestrutura de renderização de gráficos.
- III. O agrupamento de dados por uma variável categórica, seguido do cálculo de estatísticas resumidas, obtido em R com *group_by()* e *summarise()* do pacote *dplyr*, tem equivalente funcional em *Python* por meio do método *groupby()* do *Pandas*, combinado a funções de agregação.

Está correto o que se afirma em:

- (A) I, apenas.
- (B) I e II, apenas.
- (C) I e III, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

63

Um analista recebeu um arquivo com registros de atendimentos contendo colunas de datas, categorias e valores numéricos, alguns deles ausentes. Ele precisa selecionar os registros que atendam a determinadas condições, tratar os valores ausentes de forma distinta em cada coluna e calcular estatísticas resumidas dos valores numéricos para cada categoria.

Assinale a opção que apresenta a estrutura de dados mais apropriada para realizar, de forma integrada, operações de filtragem de registros, tratamento de valores ausentes por coluna e agrupamento com agregação.

- (A) Um *array* multidimensional do *NumPy*, pela eficiência das operações vetoriais aplicadas a toda a matriz de dados.
- (B) Um dicionário de listas em *Python* puro, pela flexibilidade de armazenar colunas de tipos diferentes sem depender de bibliotecas externas.
- (C) Um *DataFrame* do *Pandas*, que permite selecionar registros por rótulo ou condição, tratar valores ausentes por coluna e agrupar por categoria em poucas operações.
- (D) Uma matriz esparsa do *SciPy*, pela economia de memória ao representar os valores ausentes como elementos não preenchidos.
- (E) Um *array* estruturado do *NumPy*, que atribui um tipo fixo a cada coluna nomeada, de forma semelhante a uma tabela.

64

No âmbito das tecnologias para *Big Data*, o *Apache Hadoop* e o *Apache Spark* apresentam características distintas de execução e tolerância a falhas.

Análise as afirmativas a seguir e assinale (V) para a verdadeira e (F) para a falsa.

- () O *Hadoop MapReduce* grava dados intermediários em disco entre as fases de *Map* e *Reduce*, enquanto o *Spark* processa e retém dados preferencialmente na memória RAM.
- () No *Apache Spark*, as transformações disparam imediatamente a computação física no *cluster*, enquanto as ações apenas registram as etapas no Grafo Acíclico Dirigido (DAG).
- () A tolerância a falhas no HDFS ocorre pela replicação de blocos em múltiplos nós, enquanto no *Spark* ela se baseia no histórico de linhagem (*lineage*) para reconstituir partições perdidas.

As afirmativas são, respectivamente,

- (A) V – V – F.
- (B) V – F – V.
- (C) F – V – V.
- (D) F – F – V.
- (E) V – F – F.

65

Uma curva S é formada pela interseção do plano $z = \frac{1}{4}x$ com o cilindro elipsoidal reto $x^2/16 + y^2/9 = 1$ para todo z .

Sua área é

- (A) 15π .
- (B) 18π .
- (C) 16π .
- (D) 20π .
- (E) 21π .

66

Projetos relacionados à inteligência de dados geralmente envolvem elaboração de modelos que simplifiquem *datasets* de alta dimensionalidade, permitindo identificação e extração de informação de maior significância.

Nesse contexto, diversas ferramentas e métodos utilizados para esses fins se fundamentam em conceitos basilares de álgebra linear, tais como autovalores, autovetores e normas.

Considerando que uma matriz quadrada A de ordem $n \times n$ possui autovalores reais distintos $\lambda_1, \lambda_2, \dots, \lambda_n$ e matriz dos respectivos autovetores à direita unitários $V = [v_1, v_2, \dots, v_n]$, e ainda um vetor coluna unitário u de ordem n , tem-se que

- (A) $AV = LV$, onde $L = \text{diagonal}(\lambda_1, \lambda_2, \dots, \lambda_n)$.
- (B) $V^T A = V^T L$, onde $L = \text{diagonal}(\lambda_1, \lambda_2, \dots, \lambda_n)$.
- (C) as normas L_1 de cada coluna de V são idênticas entre si.
- (D) a norma L_1 de Au é constante para qualquer u .
- (E) a norma L_2 de Au é $|\lambda_i|$ quando $u = v_i$, para $i = 1, 2, \dots, n$.

67

Sejam A e B dois eventos de um mesmo espaço amostral, tais que $P(A) > 0$ e $P(B) > 0$.

De acordo com a teoria da probabilidade, analise os itens a seguir.

- I. Se $P(A) + P(B) > 1$, então $P(A \cap B) > 0$.
- II. Se $P(A \cup B) = P(A) + P(B)$, então A e B não são independentes.
- III. $P(A|B)P(B) = P(B|A)P(A)$

Está correto o que se afirma em

- (A) I, apenas.
- (B) III, apenas.
- (C) I e III, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

68

Dois dados honestos de seis faces, numeradas de 1 a 6, são lançados independentemente.

Considere os seguintes eventos:

A - O resultado do primeiro dado é par.

B - A soma dos resultados dos dois dados é par.

C - O resultado do primeiro dado é maior ou igual a 2.

D - A soma dos resultados dos dois dados é igual a 7.

Com base nos eventos descritos, é correto afirmar que

- (A) os eventos A e C são independentes.
- (B) os eventos C e D são disjuntos.
- (C) $P(A \cap B \cap C) = 1/4$.
- (D) $P(D|A) = 1/3$.
- (E) $P(A \cap C) = 5/12$.

69

Uma pesquisa em uma região do Brasil revelou que o intervalo de confiança de 95% da renda média da população local é de

R\$ 3.500,00 a R\$ 4.200,00.

De acordo com a pesquisa é correto afirmar que

- (A) aproximadamente 95% da população local ganha entre R\$ 3.500,00 e R\$ 4.200,00.
- (B) aproximadamente 2,5% da população local ganha acima de R\$ 4.200,00.
- (C) a renda média da população local é de R\$ 3.850,00.
- (D) no máximo 5% da população local tem renda inferior a R\$ 3.500,00.
- (E) se o procedimento de amostragem e construção do intervalo fosse repetido muitas vezes, aproximadamente 95% dos intervalos assim obtidos conteriam a verdadeira renda média da população.

70

Um levantamento com 100 crianças sobre a idade em que completaram a alfabetização e a renda média familiar obteve os seguintes resultados:

Idade em que completou a alfabetização (anos)	Quantitativo	Renda Familiar Média
5	7	R\$ 6.500,00
6	35	R\$ 5.000,00
7	25	R\$ 3.500,00
8	20	R\$ 3.200,00
9	10	R\$ 2.850,00
10	3	R\$ 2.750,00

De acordo com o levantamento, é correto afirmar que

- (A) a mediana da idade em que as crianças concluíram a alfabetização é de 6,5 anos.
- (B) a moda da idade em que as crianças concluíram a alfabetização é de 7,0 anos.
- (C) a média da idade em que as crianças concluíram a alfabetização é de 7,5 anos.
- (D) o coeficiente de correlação entre a idade de conclusão da alfabetização e a renda familiar média apresentada é positivo e próximo de 1,0.
- (E) o coeficiente de correlação entre a idade de conclusão da alfabetização e a renda familiar média apresentada pertence ao intervalo $[-1; 0)$.

Realização

