

CONCURSO PÚBLICO

COMPANHIA DE PROCESSAMENTO DE DADOS
DO MUNICÍPIO DE PORTO ALEGRE - PROCEMPA

ANALISTA EM TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO

Analista de Programação e Suporte de Sistemas Informatizados e Georreferenciados, Infraestruturas e Redes Computacionais: Ciência de Dados

- Confira devidamente o material (CARTÃO DE RESPOSTAS e CADERNO DE QUESTÕES); se houver falha, solicite a substituição ao fiscal.
- Este caderno de questões está assim constituído: **10 questões de Língua Portuguesa e 40 questões de Conhecimentos Específicos.**
- Após conferir seus dados no cartão de respostas (NOME, Nº da INSCRIÇÃO e CARGO), assine no local indicado.

• O CARTÃO DE RESPOSTAS é o único documento de correção, devendo ser preenchido com bastante atenção, à caneta esferográfica, azul ou preta, ponta grossa, marcando uma única alternativa em cada questão. Verifique no cartão de respostas as instruções para preenchimento. Não rasure, não amasse nem dobre o CARTÃO DE RESPOSTAS. É vedada a substituição do CARTÃO DE RESPOSTAS decorrente de erro cometido por candidato.

• **ATENÇÃO, MUITO IMPORTANTE!** No campo indicado no seu cartão de respostas, FAÇA A MARCAÇÃO DO NÚMERO CORRESPONDENTE AO SEU TIPO DE CADERNO DE PROVA, CONFORME CONSTA NO CANTO SUPERIOR DIREITO DESTA CAPA.

• Durante a realização da prova, não é permitida a comunicação entre candidatos nem a utilização de máquina calculadora e/ou similares, livros, anotações, réguas de cálculo, impressos ou qualquer outro material de consulta.

• O tempo de duração da prova objetiva será de **até 4 horas**, incluído o tempo para preenchimento do Cartão de Respostas. O candidato somente poderá levar consigo seu caderno de provas **após 1 hora** contada do efetivo início das provas. Ao candidato que sair antes de uma hora contada do efetivo início das provas, será permitido copiar seu cartão de respostas, para conferência com o gabarito oficial.

• Em todas as questões de legislação desta prova (se houver), serão consideradas as emendas e/ou alterações da legislação com entrada em vigor até a data do início das inscrições. Na prova de Português será exigida a norma ortográfica atualmente em vigor. Para responder às questões de Informática (se houver), a menos que seja informado o contrário, considerar que os programas mencionados se encontram na versão Português-BR e em sua configuração padrão de instalação, possuem licença de uso, o mouse está configurado para destros, um clique ou duplo clique correspondem ao botão esquerdo do mouse, e teclar corresponde à operação de pressionar uma tecla e, rapidamente, liberá-la, acionando-a apenas uma vez.

• Ao final dos trabalhos, **DEVOLVA** ao Fiscal da sala o CARTÃO DE RESPOSTAS devidamente preenchido e assinado.

• Ao final da prova objetiva, os três últimos candidatos de cada sala de prova deverão permanecer no recinto, a fim de acompanhar os fiscais até a coordenação para o lacre dos envelopes, quando, então, poderão retirar-se do local, simultaneamente, após concluído.

Direitos autorais reservados à Objetiva Concursos Ltda. Proibida a reprodução, ainda que parcial, sem autorização prévia.

Boa Prova!

Seriedade e ética: nós acreditamos nesses valores.

Nome do(a) Candidato(a)

Nº de inscrição

✂ Esta parte destacável serve para copiar as suas respostas, somente poderá ser destacada no ato da entrega do material na mesa do fiscal.

CADERNO DE PROVA TIPO 01

01	☐	02	☐	03	☐	04	☐	05	☐	06	☐	07	☐	08	☐	09	☐	10	☐
11	☐	12	☐	13	☐	14	☐	15	☐	16	☐	17	☐	18	☐	19	☐	20	☐
21	☐	22	☐	23	☐	24	☐	25	☐	26	☐	27	☐	28	☐	29	☐	30	☐
31	☐	32	☐	33	☐	34	☐	35	☐	36	☐	37	☐	38	☐	39	☐	40	☐
41	☐	42	☐	43	☐	44	☐	45	☐	46	☐	47	☐	48	☐	49	☐	50	☐

LÍNGUA PORTUGUESA

Fósseis de 14 mil anos achados na China têm identidade revelada após análise de DNA

Um mistério em torno de fósseis humanos encontrados em uma caverna na China foi resolvido pelo sequenciamento de DNA, de acordo com um novo estudo, e pode abrir espaço para discussões sobre a ascendência dos primeiros americanos.

Em 1989, um osso da coxa e parte de um crânio foram encontrados em uma caverna na província chinesa de Yunnan, no sudoeste do país.

A datação por radiocarbono realizada em 2008 nos sedimentos onde os fósseis foram encontrados indicou que eles tinham cerca de 14 mil anos — o que significa que eles eram de um período de tempo em que o *Homo sapiens* (humanos modernos) havia migrado para muitas partes do mundo.

No entanto, as características primitivas dos ossos incomodaram os cientistas, que questionaram a que espécie de humano os fósseis pertenciam.

A forma do crânio lembrava a dos neandertais — uma população humana arcaica que desapareceu há cerca de 40 mil anos — e parecia que o cérebro seria menor do que o dos humanos modernos.

Como resultado, alguns especialistas em evolução humana pensaram que o crânio provavelmente pertencia a uma população híbrida de humanos arcaicos e modernos ou talvez a uma espécie humana anteriormente desconhecida que existia ao lado da nossa. Pesquisadores deram ao grupo o nome *Red Deer People*.

Agora, cientistas chineses extraíram material genético da calota craniana e sequenciaram o DNA. Eles descobriram que o crânio pertencia a um indivíduo do sexo feminino, que provavelmente era um ancestral humano direto — um membro do *Homo sapiens* — e não um tipo de humano anteriormente desconhecido.

“A técnica de DNA antigo é uma ferramenta realmente poderosa”, disse Bing Su, professor do Instituto Kunming de Zoologia da Academia Chinesa de Ciências em Yunnan, que esteve envolvido na pesquisa, em um comunicado à imprensa.

“Isso nos diz definitivamente que o povo da Caverna *Red Deer* eram humanos modernos em vez de uma espécie arcaica, como neandertais ou denisovanos, apesar de suas características morfológicas incomuns”. Su e seus colegas compartilharam suas descobertas em um estudo publicado na revista *Current Biology*.

A análise do genoma revelou que o indivíduo ao qual os ossos pertenciam tinha níveis de ancestralidade neandertal e denisovano semelhantes aos encontrados em humanos modernos — sugerindo que eles não faziam parte de uma população híbrida que cruzava entre si.

O DNA dos denisovanos, um grupo pouco compreendido de humanos arcaicos, e dos neandertais vive em alguns humanos hoje. Isso porque há muito tempo nossos ancestrais *Homo sapiens* encontraram esses grupos à medida que se espalhavam pelo mundo e se reproduziam com eles.

(Fonte: CNN Brasil - adaptado.)

1) Em conformidade com o texto, marcar C para as sentenças Certas, E para as Erradas e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- () Em um comunicado à imprensa, Bing Su, professor do Instituto Kunming de Zoologia da Academia Chinesa de Ciências em Yunnan, que esteve envolvido na pesquisa, disse que a técnica de DNA antigo é uma ferramenta realmente poderosa.
- () Ao analisar o genoma, foi revelado que o indivíduo ao qual os ossos pertenciam tinha níveis de ancestralidade neandertal e denisovano distinto dos encontrados em humanos modernos.
- () O DNA dos denisovanos, um grupo bastante compreendido de humanos arcaicos, e dos neandertais não vive em alguns humanos hoje. Isso aconteceu porque, à medida que os denisovanos se espalhavam pelo mundo e se reproduziam com neandertais.

- a) C - C - C.
- b) E - C - E.
- c) C - E - E.
- d) E - E - C.
- e) E - C - C.

2) No trecho “No entanto, as características primitivas dos ossos incomodaram os cientistas, [...]” o verbo sublinhado é classificado, quanto à transitividade, como:

- a) Transitivo nominal.
- b) Transitivo direto.
- c) Transitivo indireto.
- d) Transitivo de ligação.
- e) Bitransitivo.

3) Assinalar a alternativa que apresenta o mesmo tempo e modo verbal que o verbo sublinhado em “Fósseis de 14 mil anos achados na China têm identidade revelada após análise de DNA”:

- a) As encomendas foram entregues durante a tarde.
- b) Ninguém acreditou na sua inocência.
- c) Em breve, cantaremos no festival.
- d) Os meninos procuram o brinquedo.
- e) Analisarei todas as propostas.

4) No trecho “Em 1989, um osso da coxa e parte de um crânio foram encontrados em uma caverna na província chinesa de Yunnan, no sudoeste do país.”, temos:

- a) Uma oração.
- b) Duas orações.
- c) Três orações.
- d) Uma frase interrogativa.
- e) Um período composto.

5) O trecho “Um mistério em torno de fósseis humanos encontrados em uma caverna na China foi resolvido pelo sequenciamento de DNA, de acordo com um novo estudo, e pode abrir espaço para discussões sobre a ascendência dos primeiros americanos.” é reescrito, mantendo o sentido original, em:

- a) Deixando de abrir espaço para discussões sobre a ascendência dos primeiros americanos, foi resolvido um mistério em torno de fósseis humanos encontrados em uma caverna na China pelo sequenciamento de DNA, devido a um estudo antigo.
- b) Em razão de um estudo que abriu espaço para discussões sobre a ascendência dos primeiros americanos, foi resolvido um mistério em torno de fósseis humanos encontrados em uma caverna de DNA pelo sequenciamento na China.
- c) Na China, devido a um estudo sobre DNA, uma caverna abriu espaço para discussões sobre a ascendência dos primeiros americanos, sendo resolvido um mistério em torno de fósseis humanos pelo sequenciamento genético.
- d) O sequenciamento de DNA abriu espaço para discussões sobre a ascendência dos primeiros americanos, sendo um mistério os fósseis humanos encontrados em uma caverna na China.
- e) Segundo um novo estudo, um mistério em torno de fósseis humanos encontrados em uma caverna na China foi resolvido pelo sequenciamento de DNA, e pode abrir espaço para discussões sobre a ascendência dos primeiros americanos.

6) Em relação à pontuação, assinalar a alternativa CORRETA:

- a) Os dias pareciam, passar demasiadamente lentos.
- b) Os mamíferos se alimentavam de brotos frutas, e sementes.
- c) Os alunos estavam cansados desanimados, tristes.
- d) Ao primeiro pinga de chuva foi se esconder na, marquise.
- e) Aquela tarde, ou melhor, aquele dia foi desagradável.

7) Em relação à regência nominal, marcar C para as sentenças Certas, E para as Erradas e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- () O artigo estava anteposto a um substantivo comum.
 - () Um soldado caiu, esvaído em sangue.
 - () Aqueles versos eram uma reminiscência de outro autor.
- a) C - C - C.
 - b) E - C - E.
 - c) C - E - E.
 - d) E - E - C.
 - e) E - C - C.

8) Em relação ao uso da crase, assinalar a alternativa CORRETA:

- a) O aluno tinha certeza de que havia muito à fazer.
- b) Durante a apresentação, fez alusão à músicas trovadorescas.
- c) Naquela tarde, o julgamento foi desfavorável à ele.
- d) Na reunião, sem titubear, ele se referiu à diretora da empresa.
- e) Por ter visto o tutorial, seguiu o passo à passo da receita.

9) Na oração “Embora soubesse da previsão climática, foi navegar.”, a expressão sublinhada exprime ideia de:

- a) Tempo.
- b) Causa.
- c) Proporção.
- d) Concessão.
- e) Conformidade.

10) Sobre o uso dos porquês, assinalar a alternativa CORRETA:

- a) Ninguém sabia ao certo por quê ele foi embora.
- b) Faltou à aula por que?
- c) Sabíamos que sua ausência tinha um porquê.
- d) Ele não disse a ninguém o porque viajaria tão logo.
- e) Porque solicitou a dispensa do cargo?

CONHECIMENTOS ESPECÍFICOS

11) No raciocínio lógico, as proposições devem obedecer sempre a três princípios fundamentais: princípio da identidade, princípio da não contradição, princípio do terceiro excluído. Sobre esses princípios, analisar os itens abaixo:

- I. Princípio da identidade: uma proposição verdadeira é verdadeira; uma proposição falsa é falsa.
- II. Princípio da não contradição: uma proposição ou será verdadeira, ou será falsa; não há outra possibilidade.
- III. Princípio do terceiro excluído: nenhuma proposição poderá ser verdadeira e falsa ao mesmo tempo.

Está(ão) CORRETO(S):

- a) Somente o item I.
- b) Somente o item II.
- c) Somente os itens I e III.
- d) Somente os itens I e II.
- e) Todos os itens.

12) Sobre o sistema de controle de versão usando o GIT, analisar a sentença abaixo:

A principal função do sistema de controle de versão é armazenar todo o histórico de desenvolvimento do documento, desde o primeiro envio até sua última versão (1ª parte). O histórico armazenado permite que seja possível resgatar uma determinada versão de qualquer data mais antiga, evitando desperdício de tempo no desenvolvimento para desfazer alterações quando se toma algum rumo equivocado (2ª parte). O envio das alterações é feito com autorização do cliente, quando ele desejar; mas, para minimizar conflitos de versões, facilitar no desfazer de alterações e também no controle do histórico (3ª parte).

A sentença está:

- a) Totalmente correta.
- b) Correta somente em sua 1ª parte.
- c) Correta somente em sua 2ª parte.
- d) Correta somente em suas 1ª e 2ª partes.
- e) Totalmente incorreta.

13) Desde os primórdios da era da computação, já em trabalhos de Von Neumann (1945), as Arquiteturas Paralelas foram apresentadas como formas de obter uma maior capacidade de processamento para a execução de tarefas com altas cargas de processamento. Sobre a temática, analisar a sentença abaixo:

Processamento paralelo e processamento distribuído são métodos semelhantes para aumentar a quantidade de poder de processamento disponível para uma tarefa específica. Geralmente, estes métodos são usados para resolver os problemas que envolvem grandes quantidades de cálculo repetitivo (1ª parte). Quando o problema é dividido em muitos pedaços grandes e complexos, seções individuais do problema podem ser calculadas com muitos processadores simultaneamente (2ª parte). Dividir o problema em partes traz uma maior capacidade de processamento a ser exercida sobre o problema de que pode ser fornecido por um único processador (3ª parte).

A sentença está:

- a) Totalmente correta.
- b) Correta somente em sua 1ª parte.
- c) Correta somente em sua 2ª parte.
- d) Correta somente em suas 1ª e 3ª partes.
- e) Totalmente incorreta.

14) A função *hash* é utilizada para inserir ou buscar um elemento, mas podem ocorrer colisões quando a função *hash* produz o mesmo endereço para chaves diferentes.

$$p = \frac{M-1}{M} \times \frac{M-2}{M} \times \dots \times \frac{M-N+1}{M} = \prod_{i=1}^N \frac{M-i+1}{M} = \frac{M!}{(M-N)!M^N}$$

Assinalar a alternativa que explica corretamente o que a imagem acima representa:

- a) A probabilidade p de se inserir N itens consecutivos sem colisão em uma tabela de tamanho M .
- b) Resolve o uso ineficiente do espaço de armazenamento.
- c) A probabilidade de ocorrer colisão ao se inserir M itens aleatórios em uma tabela de tamanho N .
- d) Uma estratégia eficiente para tratamento de colisões.
- e) Uma função de transformação que mapeia chaves em inteiros dentro do intervalo $[0..M-1]$, onde N é o tamanho da tabela.

15) Uma matriz é uma variável composta, homogênea e multidimensional, formada por variáveis, todas do mesmo tipo de dados, com o mesmo identificador (mesmo nome). Considerando-se o exposto, marcar C para as afirmativas Certas, E para as Erradas e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- () Em pseudocódigo, é possível expressar que uma matriz deve ser totalmente lida por meio do comando *Leia*.
- () Um elemento de uma matriz pode ser referenciado com o nome da matriz, seguido do seu índice correspondente, colocado entre parênteses.
- () Em pseudocódigo, é possível expressar que uma matriz deve ser totalmente impressa por meio do comando *Escreva*.

- a) C - C - E.
- b) E - E - C.
- c) E - C - E.
- d) C - E - C.
- e) C - C - C.

16) À medida que as necessidades de rede evoluíram, também evoluíram os tipos de rede de computadores que atendem a essas necessidades. Considerando-se alguns dos tipos de rede de computadores mais comuns e amplamente usados, no que se refere à rede e sua respectiva definição, numerar a 2ª coluna de acordo com a 1ª e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- (1) LAN.
- (2) WLAN.
- (3) VPN.
- (4) WAN.

- () Conecta computadores em uma área ampla, como de região a região ou mesmo de continente a continente.
- () Conecta computadores a uma distância relativamente curta, permitindo que eles compartilhem dados, arquivos e recursos.
- () É uma conexão segura, ponto a ponto, entre dois terminais de rede.
- () É como uma rede local, mas as conexões entre os dispositivos da rede são feitas sem fio.

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

17) Os atributos são os elementos que definem a estrutura de uma classe. Os atributos também são conhecidos como variáveis de classe e podem ser divididos em dois tipos básicos: atributos de instância e de classe. Com base no exposto, analisar os itens abaixo:

- I. Os valores dos atributos de instância determinam o estado de cada objeto.
- II. Um atributo de classe possui um estado que é compartilhado por todos os objetos de uma classe, também conhecido como propriedades e descrevem um intervalo de valores que as instâncias da classe podem apresentar.
- III. Atributos de classe podem ser chamados também de atributos estáticos ou dinâmicos.

Está(ão) CORRETO(S):

- a) Somente o item I.
- b) Somente o item II.
- c) Somente os itens I e III.
- d) Somente os itens I e II.
- e) Todos os itens.

18) Considerando-se a classificação dos algoritmos abaixo quanto ao tipo de aprendizagem, numerar a 2ª coluna de acordo com a 1ª e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- (1) Aprendizagem supervisionada.
- (2) Aprendizagem não supervisionada.

- () *Random Forest*.
- () *k-Nearest Neighbors*.
- () *Naive Bayes*.

- a) 1 - 2 - 1.
- b) 2 - 1 - 2.
- c) 2 - 2 - 2.
- d) 1 - 1 - 2.
- e) 1 - 1 - 1.

19) MongoDB é um *software* de banco de dados orientado a documentos livre, de código aberto e multiplataforma, escrito na linguagem C++. Sobre o MongoDB, analisar a sentença abaixo:

Uma de suas principais características consiste em oferecer alta disponibilidade, usando conjuntos duplicados. Um conjunto duplicado consiste em uma ou mais cópias dos dados. Cada cópia pode agir como primária ou secundária (1ª parte). Por padrão, operações de leitura e escrita são executadas na cópia primária, enquanto a secundária mantém uma cópia da primária (2ª parte). Se houver uma falha na cópia primária, o conjunto replicado automaticamente conduz um processo de eleição para determinar qual das cópias secundárias deve assumir o papel de primária. Existe a opção de realizar operações de leitura, mas a consistência dos dados ocorrerá de forma eventual (3ª parte).

A sentença está:

- a) Totalmente correta.
- b) Correta somente em sua 1ª parte.
- c) Correta somente em sua 3ª parte.
- d) Correta somente em suas 1ª e 2ª partes.
- e) Totalmente incorreta.

20) Em Jenkins, as compilações podem ser acionadas por vários meios. Considerando-se estes meios, marcar C para as afirmativas Certas, E para as Erradas e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- () Um *webhook* que é acionado após *commits* serem enviados em um sistema de controle de versão.
- () Agendamento por meio de um mecanismo semelhante ao *cron*.
- () Depois que as outras compilações na fila forem concluídas.

- a) C - C - E.
- b) E - E - C.
- c) E - C - E.
- d) C - E - C.
- e) C - C - C.

21) Sobre a *Sprint* do método *Scrum*, marcar C para as afirmativas Certas, E para as Erradas e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- () Não são feitas mudanças que possam por em perigo o objetivo da *Sprint*.
- () As metas de qualidade podem diminuir no decorrer do processo.
- () O escopo pode ser clarificado e renegociado entre o *Product Owner* e o Time de Desenvolvimento.

- a) C - C - E.
- b) E - E - C.
- c) E - C - E.
- d) C - E - C.
- e) C - C - C.

22) A imagem abaixo representa um quadro Kanban.



A definição de pronto no sistema Kanban está resumida nas seguintes etapas:

- I. Especificação: item foi subdivido em tarefas que podem ser finalizadas em menos de uma semana, cada uma.
- II. Desenvolvimento: revisado e testado por um colega.
- III. Validação: o trabalho foi validado por usuários reais e todos os problemas encontrados foram tratados.

Está(ão) CORRETO(S):

- a) Somente o item I.
- b) Somente o item II.
- c) Somente os itens I e II.
- d) Somente os itens I e III.
- e) Todos os itens.

23) A Modelagem de Dados é um instrumento de grande importância na formatação de estruturas capazes de serem implantadas e compreendidas pelos Sistemas Gerenciadores de Banco de Dados. Analisar a sentença abaixo:

O Modelo multidimensional é uma técnica de modelagem conceitual de negócios, que facilita a investigação, o resumo e a organização de dados para a análise de negócios (1ª parte). O cubo é uma estrutura de dados multidimensional para analisar e recuperar dados (2ª parte). As hierarquias em dimensões e fórmulas são de primeira classe, ou seja, um cubo não duplica suas definições, o cubo pode ter qualquer n dimensões, também denotado de hipercubo ao invés de cubo, sendo que uma coleção de cubos é relacionada ao banco de dados multidimensional ou armazém de dados (Data Warehouse) multidimensional (3ª parte).

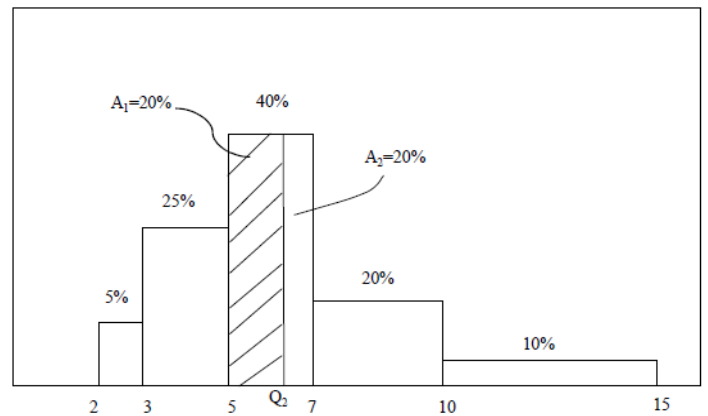
A sentença está:

- a) Totalmente correta.
- b) Correta somente em sua 1ª parte.
- c) Correta somente em sua 3ª parte.
- d) Correta somente em suas 1ª e 2ª partes.
- e) Totalmente incorreta.

24) Quando a relação linear entre variáveis resposta e preditoras for compatível com um modelo linear (no caso de uma relação polinomial, por exemplo), o uso de regressão linear é conveniente e obtemos modelos com interpretabilidade e poder preditivo satisfatórios. Uma forma de determinar se a regressão foi bem sucedida em explicar uma parcela significativa da variância de y é testar a hipótese nula de que a proporção da variabilidade de y explicada por x é igual a zero. Esse trabalho é realizado por meio do:

- a) Teste Z.
- b) Teste Qui Quadrado.
- c) Teste F.
- d) Métodos dos Mínimos Quadrados.
- e) Teste T.

25) Considerar o histograma abaixo e assinalar o valor correspondente a Q2 sinalizado pela seta:



- a) $Q2 = 6,0$.
- b) $Q2 = 6,2$.
- c) $Q2 = 6,4$.
- d) $Q2 = 6,6$.
- e) $Q2 = 6,8$.

26) As fases do CRISP-DM não são necessariamente executadas em sequência, uma vez que durante o processo podem ocorrer alterações e ser modificada a sequência. A formação de hipóteses em cima do que se aprendeu com os dados, é executada em qual fase do CRISP-DM?

- a) *Data Preparation.*
- b) *Business Understanding.*
- c) *Modeling.*
- d) *Evaluation.*
- e) *Data Understanding.*

27) A respeito dos conceitos de *Deep Learning*, analisar os itens abaixo:

- I. Redes profundas híbridas se destacam pela discriminação que é assistida, muitas vezes de forma significativa, com os resultados de redes profundas generativas ou não supervisionadas. Isso pode ser alcançado por meio de uma melhor otimização e regularização das redes profundas supervisionadas. A proposta dessa rede também pode ser alcançada quando critérios discriminativos para a aprendizagem supervisionada são usados para estimar os parâmetros em qualquer das redes profundas generativas ou não supervisionadas.
- II. Em termos de topologia, para implementar uma RNA, deve-se definir diferentes variáveis, dentre as quais: o número de nós na camada de entrada, o número de camadas escondidas, o número de neurônios a serem colocados nessas camadas, e o número de neurônios na camada de saída.
- III. As redes neurais com treinamento não supervisionado, os algoritmos de agrupamento e a mineração de dados com regras de associação possuem a grande vantagem de que não é necessário conhecer a classe de saída de cada item de dado e, alguns casos, nem a quantidade de classes de saída.

Está(ão) CORRETO(S):

- a) Somente os itens I e II.
- b) Somente os itens I e III.
- c) Somente os itens II e III.
- d) Todos os itens.
- e) Nenhum dos itens.

28) Entre as principais tarefas de pré-processamento de dados fortemente dependentes de domínio, assinalar a alternativa que é exemplificada na situação descrita abaixo:

Uma empresa de cartão de crédito cujo banco de dados possuía um campo "gender" para armazenar o sexo de seus clientes. Alguns registros assumiam, entretanto, o valor B para esse atributo, o qual, posteriormente, descobriu-se que correspondia à informação Business. Originalmente, o sistema tinha sido projetado somente para cadastrar cartões para pessoas físicas, porém, quando cartões para empresas foram permitidos, não havia um campo específico para indicar que o cadastrado era uma empresa. Essa informação foi então armazenada no campo "gender".

- a) Identificação de Poluição.
- b) Identificação de atributos redundantes.
- c) Verificação de Integridade.
- d) Identificação e Descrição de Valores Externos.
- e) Tratamento de Valores Desconhecidos.

29) Existem diversas metodologias já desenvolvidas, e em desenvolvimento, de modelos de classificação supervisionada. Nesse sentido, a melhor metodologia é aquela que se ajusta melhor aos seus dados e que é capaz de trazer as melhores respostas de acordo com o seu objetivo. A vantagem e desvantagem listada abaixo nos itens I e II, respectivamente, entre os diversos modelos de Classificação Supervisionada, referem-se a:

- I. Controle de *Overfitting*.
- II. Interpretabilidade das variáveis.

- a) I. KNN; II. Regressão Logística.
- b) I. *Randon Forest*; II. KNN.
- c) I. Regressão Logística; II. *Randon Forest*.
- d) I. Árvore de Decisão ; II. KNN.
- e) I. *Randon Forest*; II. Árvores de Decisão.

30) Ao construir um classificador usando *machine learning*, um desenvolvedor deve se perguntar o quão bom é seu modelo para predição. Assim, ao treinar um modelo de aprendizagem, algumas métricas podem ser utilizadas para avaliação. Considerando a legenda abaixo, pode-se afirmar que a Especificidade e a Precisão, são calculadas pelas seguintes fórmulas, respectivamente:

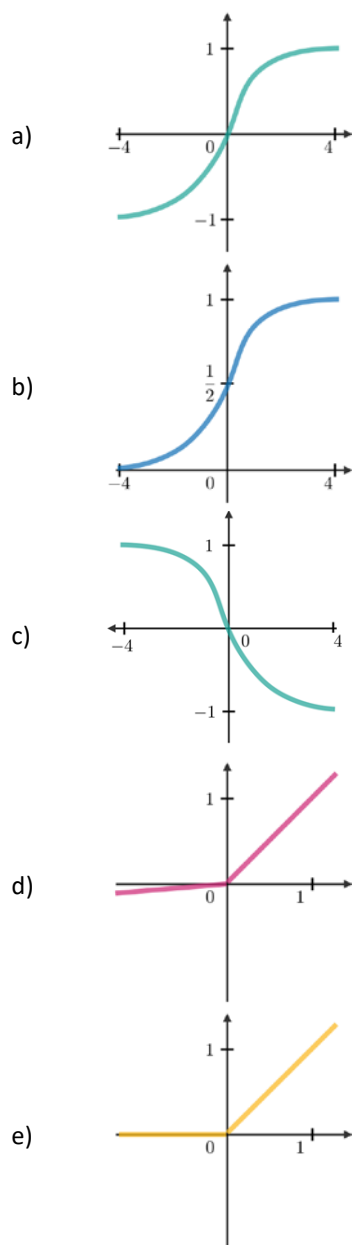
VP: verdadeiros positivos; FN: falsos negativos; FP: falsos positivos; VN: verdadeiros negativos; P: precisão; S: sensibilidade; N: total de elementos.

- a) $VP / (VP+FN)$; $2 \times (PxS) / (P+S)$.
- b) $2 \times (PxS) / (P+S)$; $(VP+VN) / N$.
- c) $VN / (FP+VN)$; $VP / (VP+FP)$.
- d) $VP / (VP+FN)$; $VP / (VP+FP)$.
- e) $VN / (FP+VN)$; $2 \times (PxS) / (P+S)$.

31) Em relação à arquitetura de Redes Neurais Convolucionais e Recorrentes, assinalar a alternativa CORRETA:

- a) As redes recorrentes ingerem e processam imagens como tensores; e tensores são matrizes de números com várias dimensões.
- b) As redes recorrentes possuem a vantagem de poder processar sequências muito longas, se usar *tanh* ou *relu* como uma função de ativação.
- c) As redes convolucionais codificam a posição e orientação do objeto.
- d) As redes convolucionais possuem a desvantagem de não detectar automaticamente os recursos importantes sem qualquer supervisão humana.
- e) As redes recorrentes diferem das redes *feed forward* porque incluem um *loop* de *feedback*, pelo qual a saída do passo n-1 é alimentada de volta à rede para afetar o resultado do passo n, e assim por diante para cada etapa subsequente.

32) Redes neurais são uma classe de modelos que são construídos com camadas, entre as quais as mais comumente utilizadas incluem redes neurais convolucionais (CNN) e recorrentes (RNN). Funções de ativação (*activation function*) são usadas no fim de uma unidade escondida para introduzir complexidades não lineares ao modelo. Assinalar a alternativa que se refere à função de ativação que tem sido mais utilizada em *Deep Learning*, por facilitar o processo de treinamento, por não saturar a partir de um determinado ponto, e ser simplesmente a função identidade para valores positivos:



33) Ao utilizar o algoritmo *K-means*, a forma mais simples de verificarmos a “qualidade” dos *clusters* encontrados, a fim de que se possa dizer o quão distantes estão os *clusters* dos outros, é por meio da:

- a) Distância Euclidiana.
- b) Distância inter-clusters.
- c) Distância entre centroides.
- d) Distância intra-cluster.
- e) Distância entre as médias.

34) Para ajustar um modelo de regressão logística, é necessário estimar os parâmetros β_0 e β_1 do modelo. Para isso, utiliza-se o método:

- a) Da verossimilhança.
- b) Dos Mínimos Quadrados.
- c) *Byesian*.
- d) De Momentos.
- e) Dos Mínimos Quadrados Ponderados.

35) Em relação ao modelo clássico de regressão linear, assinalar a alternativa INCORRETA:

- a) Relação Linear entre X e Y.
- b) Os valores de X são fixos em repetidas amostras, não aleatórios.
- c) Os erros são autocorrelacionados.
- d) Esperança condicional dos erros, igual a zero.
- e) Os erros apresentam distribuição normal.

36) A Mineração de Dados é comumente classificada por sua capacidade em realizar determinadas tarefas. Em uma auditoria, por exemplo, separar comportamentos suspeitos, é uma tarefa realizada pelo(a):

- a) Estimação.
- b) Regressão.
- c) Predição.
- d) Agrupamento.
- e) Classificação.

37) Considerar o código Python abaixo:

```
1 x = range(1, 15, 2)
2 for n in x:
3     print(n)
```

A execução desse código exibe:

- a) A sequência de números de 1 a 15 duas vezes.
- b) A sequência de números pares contidos no intervalo de 1 a 15.
- c) O termo médio da sequência de 1 a 15.
- d) O segundo número da sequência de 1 a 15.
- e) A sequência de números ímpares de 1 a 13, inclusive.

38) Considerando-se as funções DAX disponíveis no Power BI Desktop, assinalar a fórmula que soma todos os números contidos na coluna “Frete” da tabela “Vendas”:

- a) = COUNT(Vendas[Frete])
- b) = REPLACE(Vendas[Frete])
- c) = SUM(Vendas[Frete])
- d) = MAXA(Vendas[Frete])
- e) = SOFTMAX(Vendas[Frete])

39) No R, para obter subconjuntos de um vetor, é necessário realizar operações de indexação. Isto é, associar ao vetor um outro vetor de mesmo tamanho com os índices dos elementos selecionados. Nesse sentido, para indexar subconjuntos (*subsets*) de objetos no R, utiliza-se o operador:

- a) ()
- b) []
- c) {}
- d) <-
- e) #

40) Tendo em vista que vetores são estruturas muito importantes em R, em relação ao caso abaixo, a saída do código será:

```
vetor1 <- c(11, 12)
vetor2 <- c(10, 20, 30, 40)

vetor1 + vetor2
```

- a) Um erro, pois os vetores não possuem o mesmo tamanho.
- b) 21 32 30 40
- c) 21 32 41 52
- d) 11 12 10 20 30 40
- e) 10 11 12 20 30 40

41) No R, a função *mean()* calcula a média de um conjunto de valores. Sendo assim, considerando-se as saídas dos códigos indicados abaixo, numerar a 2ª coluna de acordo com a 1ª e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- (1) `mean(1, 3)`
- (2) `mean(c(1, 3, NA))`
- (3) `mean(c(1, 3, NA), na.rm = TRUE)`

- () NA
- () 1
- () 2

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

42) Considerando-se alguns dos comandos SQL, marcar C para as afirmativas Certas, E para as Erradas e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- () DROP DATABASE Procempa: duplica o Banco de Dados “Procempa”.
- () CREATE TABLE Analistas (id INT PRIMARY KEY, cpf VARCHAR(50)): cria uma tabela chamada “Analistas”, a qual terá uma coluna “id” idêntica para todos os analistas (chave-primária) e uma coluna “cpf” que aceita no máximo 50 caracteres e somente do tipo número.
- () ALTER TABLE Analistas ADD idade INT: promove uma alteração da tabela “Analistas” com a inserção de um novo campo (coluna) chamado “idade”.

- a) C - C - E.
- b) E - E - C.
- c) C - E - C.
- d) E - C - E.
- e) C - C - C.

43) Tendo em vista os comandos SQL, o comando SELECT talvez seja o comando mais utilizado e o que possui mais parâmetros e opções. Ao executar o comando abaixo, serão apresentados todos os registros cujo valor do campo NOMECLIENTE:

```
SELECT NOMECLIENTE FROM CAD_CLIENTES WHERE  
NOMECLIENTE LIKE 'INFO%';
```

- a) Comecem com as letras "INFO".
- b) Não contenham as letras "INFO".
- c) Terminem com as letras "INFO".
- d) Contenham as letras "INFO", em qualquer que seja a ordem.
- e) Contenham as letras "INFO", em qualquer que seja a posição.

44) Considerar o banco de dados da empresa Objetiva, que possui empregados e está organizada em departamentos. Os empregados possuem nome, CPF, endereço, telefones (residencial e comercial) e, os departamentos, possuem código e o nome do departamento. Com base nos conhecimentos sobre banco de dados e modelagem de dados, numerar a 2ª coluna de acordo com a 1ª e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- (1) Entidade
- (2) Atributo.
- (3) Relacionamento.
- () Empregados.
- () CPF.
- () Departamentos.

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

45) Considerando-se as características dos modelos *Star Schema* e *Snow Flake* para a modelagem multidimensional de dados, numerar a 2ª coluna de acordo com a 1ª e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- (1) Modelo *Star Schema*.
- (2) Modelo *Snow Flake*.

- () Todas as dimensões são ligadas diretamente à tabela fato.
- () Algumas dimensões não estão ligadas diretamente à tabela fato.
- () Consultas mais complexas e lentas quando comparadas ao outro modelo.
- () As tabelas dimensões são desnormalizadas.

- a) 1 - 2 - 2 - 2.
- b) 1 - 2 - 1 - 2.
- c) 1 - 2 - 2 - 1.
- d) 2 - 1 - 1 - 1.
- e) 2 - 1 - 1 - 2.

46) O ETL é um processo tradicional de transformação de dados formado por três etapas, sendo, a última delas o(a):

- a) Coleta e extração de dados brutos de uma ou diversas fontes.
- b) Transformação dos dados brutos em dados modelados dentro de um *Data Warehouse* para a aplicação de *Business Intelligence*, análise de dados e *Advanced Analytics*.
- c) Carregamento dos dados coletados em um *Data Warehouse* ou repositório de dados.
- d) Organização e conversão dos dados em um formato único.
- e) Adaptação das informações que foram analisadas e padronizadas.

47) Um *Data Warehouse* (DW) geralmente é um componente central de uma infraestrutura *Business Intelligence* (BI) em uma organização, e funciona como um repositório de armazenagem e análise de informações numéricas. Considerando-se as características dos tipos principais de tabelas utilizadas na modelagem multidimensional, marcar C para as afirmativas Certas, E para as Erradas e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- () Uma tabela fato armazena o aspecto quantitativo do DW.
- () Uma tabela dimensão possui um relacionamento de “um para um” com a tabela fato.
- () Um DW normalmente terá uma única tabela de dimensão para todas as dimensões representadas em uma tabela de fatos.

- a) C - C - E.
- b) E - E - C.
- c) C - C - C.
- d) E - C - E.
- e) C - E - E.

48) Acerca da containerização de aplicações com *Docker*, analisar os itens abaixo:

- I. Impede que muitos contêineres sejam executados simultaneamente em um único *hardware* hospedeiro.
- II. É possível utilizar um único sistema operacional e compartilhar seus recursos entre diversas máquinas.
- III. Utiliza o *hypervisor* das máquinas virtuais e o sistema operacional é virtualizado por completo.

Está(ão) CORRETO(S):

- a) Somente o item I.
- b) Somente o item II.
- c) Somente os itens I e II.
- d) Somente os itens I e III.
- e) Todos os itens.

49) Considerando-se alguns dos principais comandos utilizados na containerização de aplicações com *Docker*, o comando utilizado para criar um container, de forma que o processo seja executado em *background*, evitando, assim, que o terminal fique travado no contexto do processo do container é:

- a) `docker run -d nome-imagem`
- b) `docker rm container-id`
- c) `docker stats container-id`
- d) `docker rmi imagem-id`
- e) `docker push user/nome-imagem`

50) Um objeto, quando é criado, pode herdar atributos e ações de outros já existentes, não começando do zero. Em relação a isso, marcar C para as afirmativas Certas, E para as Erradas e, após, assinalar a alternativa que apresenta a sequência CORRETA:

- () Herança representa e propicia uma relação de hierarquia entre classes.
- () Uma subclasse herda características da classe da qual foi derivada.
- () Cada subclasse pode ter, além das características herdadas, características particulares a ela.

- a) C - C - C.
- b) C - C - E.
- c) E - C - C.
- d) E - E - C.
- e) E - E - E.