UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CONCURSO PÚBLICO DE PROVAS E TÍTULOS PARA O PROVIMENTO DE CARGO DE PROFESSOR DO MAGISTÉRIO SUPERIOR NAS CLASSES ADJUNTO-A, ASSISTENTE-A E AUXILIAR

EDITAL Nº 101/2021-PROGESP

INTELIGÊNCIA ARTIFICIAL

Leia estas instruções:

1	Informe seu nome nos dois espaços indicados na parte inferior desta capa. Ao finalizar sua prova, as duas partes onde constam seu nome e o código numérico serão destacadas pelo fiscal. Uma parte será entregue a você e a outra será guardada em um envelope que será lacrado no fim da aplicação.
2	Em atendimento ao Art. 18 da Resolução nº 150/2019-CONSEPE, sua prova será identificada unicamente por esse código numérico, gerado por sorteio na ocasião da impressão da prova.
3	Quando o Fiscal autorizar, verifique se o Caderno está completo e sem imperfeições gráficas que impeçam a leitura. Detectado algum problema, comunique-o, imediatamente, ao Fiscal.
4	Este caderno contém **20 questões** de múltipla escolha e **quatro** questões discursivas, cujas respostas serão avaliadas considerando-se apenas o que estiver escrito no espaço reservado para o texto definitivo. Para rascunho, utilize as folhas fornecidas pelo fiscal destinadas a esse fim.
5	Escreva de modo legível, pois dúvida gerada por grafia ou rasura implicará redução de pontos.
6	Cada questão de múltipla escolha apresenta quatro opções de resposta, das quais apenas uma é correta.
7	Interpretar as questões faz parte da avaliação, portanto não peça esclarecimentos aos fiscais.
8	A prova escrita **deverá** ser respondida com **caneta esferográfica de tinta preta**, **sob pena de eliminação** no concurso.
9	Os rascunhos e as marcações que você fizer neste Caderno não serão considerados para efeito de avaliação.
10	Você dispõe de, no máximo, **quatro horas** para redigir a resposta da questão discursiva **no espaço definitivo** deste caderno, responder às questões de múltipla escolha e preencher a **Folha de Respostas**.
11	O preenchimento da Folha de Respostas é de sua inteira responsabilidade.
12	Antes de se retirar definitivamente da sala, **devolva** ao Fiscal **este Caderno** e a **Folha de Respostas**.

✂ Corte aqui

VIA DO ENVELOPE DE SEGURANÇA

Informe seu nome completo: _____

_____

✂ Corte aqui

VIA DO CANDIDATO

Informe seu nome completo: _____

_____

COMPROVANTE DO TEMA SORTEADO PARA A PROVA DIDÁTICA

Concurso Público para Professor Efetivo – Edital nº 101/2021-PROGESP

ÁREA: INTELIGÊNCIA ARTIFICIAL

NOME DO CANDIDATO: _____

TEMA SORTEADO: ____ (_____) - Preenchido pelo chefe de sala

CHEFE DE SALA: _____

FISCAL: _____

QUESTÕES DISCURSIVAS

Questão 1 (2,5 pontos)

As redes neurais convolucionais (CNNs) são tipos particulares de redes neurais que têm conseguido excelentes resultados em muitas aplicações práticas de processamento de imagens. Considerando uma situação em que seja necessário classificar imagens em uma dentre 100 (cem) categorias de imagens coloridas com rótulos de garrafas utilizando uma CNN, explique de forma detalhada a arquitetura da rede neural, destacando suas particularidades em cada etapa do pré-processamento e processamento.

Espaço destinado à Resposta

Fim do Espaço destinado à Resposta

Questão 2 (2,5 pontos)

Os métodos supervisionados podem ser usados para resolver problemas de classificação e regressão. Explique a diferença entre esses tipos de problemas e exemplifique-os. Cite, também, um exemplo que pode ser resolvido com ambos os métodos, destacando as diferenças em como aplicá-los.

Espaço destinado à Resposta

Fim do Espaço destinado à Resposta

Questão 3 (2,5 pontos)

Algumas das aplicações de Inteligência Artificial pertencem à categoria dos problemas não computáveis. Estes problemas são geralmente resolvidos por meio da busca por uma solução. Considere que um professor de um dos cursos de formação técnica do Instituto Metrópole Digital precisa sair de Natal e chegar a um dos pólos do curso, no Município de Mossoró. Ajude o professor a construir a melhor solução para este problema, usando um algoritmo de busca. Apresente as suposições que você utilizou para obter a sua solução, identificando o maior número de detalhes possível acerca do algoritmo que você escolheu.

Espaço destinado à Resposta

Fim do Espaço destinado à Resposta

Questão 4 (2,5 pontos)

Na aplicação de algoritmos de aprendizado de máquina a problemas reais, em geral, conhecer o domínio dos dados torna-se um aliado para a escolha de um modelo mais adequado aos objetivos definidos. Existem diversos critérios para a avaliação de modelos preditivos, bem como diversas medidas de desempenho associadas a esses modelos. Descreva pelo menos uma medida de desempenho, especificando onde é aplicada, como é definida, o que significa, e qual a forma de analisá-la.

Espaço destinado à Resposta

Fim do Espaço destinado à Resposta

QUESTÕES DE MÚLTIPLA ESCOLHA

01. A mineração de dados exige um processo bem definido que pode aumentar a probabilidade de um resultado bem-sucedido. Nesse contexto, analise as seguintes afirmações:

I	Não é importante a compreensão do negócio porque a solução está exclusivamente na análise de dados.
II	A redução de dimensionalidade também pode ser usada para visualização de dados.
III	O objetivo da fase de avaliação é estimar os resultados de mineração de dados e obter confiança de que são válidos.
IV	Ajustar os tipos das variáveis e selecionar atributos com ganho de informação, não são etapas da mineração de dados.

Estão corretas apenas as seguintes alternativas:

A) II, III e IV.

B) I,II e IV.

C) I e IV.

D) II e III.

02. Implantar modelos inteligentes em ambientes de produção envolve várias ações. São ações e/ou serviços relevantes para inserir o modelo em um ambiente de produção para ser acessado via requisição web:

A) pré-processamento, limpeza dos dados e testes automatizados.

B) treinamento, validação e servidor web.

C) treinamento, teste de desempenho e ambiente de hospedagem.

D) ambiente de hospedagem, servidor web e configuração do domínio.

03. Um dos problemas na análise de dados, durante o processo de extração e seleção de atributos, é a alta dimensionalidade do espaço de características. Uma técnica para minimizar esse problema é

A) o Coeficiente R2.

B) a Entropia Cruzada Binária.

C) o Erro Médio Quadrático.

D) a Análise de Componentes Principais.

04. No contexto de aprendizado profundo, o papel da função de ativação softmax é forçar a saída da rede neural

A) entre -1 e 1.

B) na representação de probabilidade.

C) normalizada entre 0 e um valor predefinido x.

D) a ser simétrica ao redor de um valor predefinido.

05. Durante um treinamento de uma rede neural convolucional para classificar imagens, a camada responsável por aplicar filtros na imagem é a

A) camada de ativação.

B) camada de entrada.

C) camada convolucional.

D) camada densa.

06. Sobre Redes Neurais Recorrentes (RNN), analise as afirmativas abaixo.

I	LSTM (Long short-term memory) é uma arquitetura de rede neural recorrente.
II	Apresentam uma boa eficiência no processamento de linguagem natural.
III	Possuem estados que permitem registrar informações que foram processadas até o momento atual.
IV	As sequências de entrada sempre precisam ter tamanho fixo.

Estão corretas apenas as afirmativas:

A) I,II e III.

B) I,II e IV.

C) II, III e IV.

D) I, III e IV.

07. Existe um método que objetiva treinar vários modelos, geralmente com diferentes tipos de algoritmos, nos dados do treinamento e, em seguida, em vez de escolher o melhor modelo, todos eles são agregados usando outro (meta-aprendizagem), para fazer a previsão final. Estamos nos referindo ao método

A) Bagging.

B) Bootstrap.

C) Stacking.

D) Cross Validation.

08. Acerca do processo de fuzzificação de um problema, analise as afirmativas abaixo.

I	A base de conhecimento (experiência) não é necessária.
II	Valores linguísticos são transformados em graus de pertinência.
III	A etapa de defuzzificação engloba: definição das proposições, análise das regras e criação da região fuzzy resultante.
IV	Quando as regras inferem a mesma variável linguística, os resultados são agregados.

Estão corretas somente as afirmativas

A) I e III.

B) III e IV.

C) II e IV.

D) I e II.

09. Sobre o grau de pertinência de elementos, é verdade dizer que, na teoria tradicional dos conjuntos e no conjunto fuzzy, eles são definidos, respectivamente, por

A) 0 ou 1, entre 0 e 1.

B) entre 0 e 1, 0 ou 1.

C) entre 0 e 1, entre 0 e 1.

D) 0 ou 1, 0 ou 1.

10. Com as métricas de avaliação, temos um cenário mais completo sobre o desempenho de um modelo. Nesse contexto, é verdade dizer que:

A) A medida F1 score só pode ser aplicada em dados balanceados.

B) O coeficiente de Matthews é apropriado para avaliar dados desbalanceados.

C) A acurácia é a métrica mais eficiente para avaliar um modelo.

D) O desbalanceamento entre as classes não é reconhecido a partir de métricas de avaliação.

11. Analise a afirmativas a seguir acerca dNo método de agrupamento K-means:

I	Encontra-se automaticamente o número de clusters.
II	Cada centro do cluster é definido pela média dos pontos atribuídos a ele em cada iteração.
III	Um número muito pequeno de clusters pode levar a overfitting.
IV	O algoritmo converge quando a mudança no cluster atribuído é inferior a um limiar.

Estão corretas apenas as afirmativas

A) II e IV. **C)** I e IV.

B) I e III. **D)** II e III.

12. Sobre agentes inteligentes e seus tipos, é CORRETO afirmar que:

A) Quando existem várias alternativas possíveis, para decidir qual é a melhor, são usados agentes baseados em utilidade.

B) Um agente baseado em modelo pode lidar apenas com ambientes totalmente observáveis pelo uso de um modelo sobre o mundo.

C) No modelo reflexivo, caso ocorra alguma alteração no ambiente, o conjunto de regras não precisa ser atualizado.

D) No agente baseado em modelo, o modelo não depende do histórico de percepção e, portanto, reflete, pelo menos, alguns dos aspectos não observados do estado atual.

13. Após a publicação, em 11 de janeiro de 2022, da Resolução de nº. 004/2022 – CONSEPE, o Instituto Metrópole Digital (IMD) tem preparado o seu retorno às aulas presenciais. Neste sentido, realizou uma pesquisa com os seus alunos a respeito da preferência sobre o ensino no formato remoto ou presencial (Y). Observou-se, dentre outras características, o tempo em que o aluno se encontra matriculado no Instituto (X). Os resultados foram resumidos na seguinte Tabela:

		Você prefere estudar em qual formato de ensino? (Y)	
		Remoto	Presencial
Há quanto tempo você está matriculado no IMD? (X)	Menos de 2 anos	95	35
	2 anos ou mais	95	125

Fonte: Dados hipotéticos.

Tendo por base os dados desta pesquisa, considere as afirmativas a seguir:

I	É possível dizer que X e Y são independentes, visto que não existe diferença entre P(Y| X) e P(X| Y).
II	É possível dizer que as distribuições marginais de X e Y são iguais.
III	É possível dizer que P(X, Y) assume o mesmo valor independente de X para um dado nível de Y.
IV	É possível dizer que, P(X |Y)× P(Y) = P(Y |X)× P(X).

Estão corretas as afirmativas:

A) II e IV. **C)** I e III.

B) I e II. **D)** III e IV.

14. Uma cidade inteligente (ou *Smart City*) é um conceito amplo que sugere o uso de Tecnologia da Informação e Comunicação (TIC) e ferramentas de inteligência artificial no sentido de melhorar a sua eficiência operacional, prezando conceitos como transparência e equidade, além de possibilitar uma melhor qualidade dos serviços públicos e o bem-estar dos seus cidadãos. Em uma cidade inteligente,

A) além do planejamento urbano, é essencial que as soluções tecnológicas sejam aceitas e utilizadas pelos seus moradores e visitantes.

B) todos os moradores dispõem de dispositivos de IoT para garantirem o acesso ao que está disponível no serviço público.

C) tudo é monitorado por meio de sensores, sendo praticamente impossível um morador agir sem ser acompanhado pelo poder público.

D) a maior preocupação, necessariamente, deve ser o seu desenvolvimento tecnológico.

15. Avaliar a possibilidade de uma decisão errada deve estar sempre presente ao analisarmos dados. Quando utilizamos o raciocínio probabilístico, é comum nos deparamos com a realização de um teste de hipóteses. A respeito do teste de hipóteses, considere as afirmativas abaixo.

I	Nos testes de hipóteses, normalmente há duas suposições concordantes em consideração, no caso, H_0 e H_1, devendo-se decidir qual delas é a correta.
II	Um erro do tipo I ocorre quando a hipótese nula (ou seja, H_0) é rejeitada apesar de ela ser verdadeira. A probabilidade de um erro do tipo I é chamada de nível de significância do teste.
III	Um erro do tipo II ocorre quando a hipótese nula (ou seja, H_0) é rejeitada apesar de ela ser verdadeira. A probabilidade de um erro do tipo II é chamada de poder do teste.
IV	Para decidir se H_0 é rejeitada, precisamos de uma estatística de teste em função dos dados da amostra e de uma região de rejeição.

Estão corretas as afirmativas:

A) I e III.

B) I e II.

C) II e IV.

D) III e IV.

16. Dentre as diversas aplicações do Processamento da Linguagem Natural (PLN) está a análise de sentimentos, cujo objetivo é identificar, extrair e quantificar a polaridade expressada nos dados. A partir da combinação de PLN com elementos de aprendizado de máquina, é possível identificar conteúdo positivo, negativo ou neutro em textos, áudios, imagens e vídeos. A respeito das etapas de um processo de análise de sentimentos a partir de textos, é correto afirmar que:

A) O One Hot Encoding é um processo bem conhecido de transformar um texto em um token e, a partir deste, a representação de cada token é aprendida durante o treino, facilitando a otimização do uso da memória.

B) Como as camadas iniciais das redes neurais trabalham com vetores de entrada de tamanho fixo, é necessário aplicar a Tokenização, ou seja, uma forma de garantir que os comprimentos dos vetores de entrada sejam todos iguais.

C) Uma das maneiras de se trabalhar com dados de texto é separá-los em unidades menores chamadas tokens (unidades mínimas em que o texto pode ser fragmentado). Este processo é chamado de Tokenização.

D) O Word Embeddings é um processo utilizado para evitar ambiguidades na aplicação da Tokenização, por exemplo, nos casos em que um ponto pode identificar uma abreviação, uma separação numérica ou o final da sentença.

17. As estratégias de busca sem informação (também chamada de busca cega) utilizam apenas a informação disponível na definição do problema. Elas se distinguem pela ordem em que os nós são expandidos. São exemplos de estratégias de busca sem informação:

A) Busca de custo uniforme, Busca com profundidade iterativa, Busca gulosa.

B) Busca em largura, Busca de custo uniforme, Busca com profundidade iterativa.

C) Busca em largura, Busca gulosa, Busca A*.

D) Busca A*, Busca competitiva, Busca em extensão.

18. O classificador Naive Bayes (ou Bayes Ingênuo) é um dos métodos mais populares para lidar com tarefas preditivas em aprendizado de máquina, principalmente quando as informações disponíveis são incompletas ou imprecisas. Sobre o classificador Naive Bayes, considere as afirmativas a seguir:

I	O termo *naive* vem da hipótese de que os valores dos atributos de um exemplo são independentes entre si, dada a classe.
II	Uma desvantagem importante do Naive Bayes está no fato de que um atributo redundante terá um peso maior nas decisões do modelo.
III	Uma desvantagem importante do Naive Bayes está no fato de que todas as probabilidades podem ser calculadas em n passos a partir do conjunto de treinamento.
IV	O desempenho do classificador Naive Bayes tende a decrescer na presença de atributos irrelevantes.

Estão corretas as afirmativas:

A) II e IV. **B)** I e II. **C)** I e III. **D)** III e IV.

19. Em Inteligência Artificial orientada a dados, diversos podem ser os desafios e dificuldades das tarefas necessárias de se realizar: preparação de dados, seleção de técnicas, ajuste de hiperparâmetros, testes e validações. Uma alternativa que surge para lidar com essas dificuldades é o uso do AutoML ou Aprendizado de Máquina Automatizado. Sobre o AutoML é correto afirmar:

A) Surge a chamada meta-learning, que observa diferentes abordagens de aprendizado de máquina, executadas em diversas tarefas de aprendizado e, então, buscam aprender a partir dessa experiência para se ajustar a novas tarefas mais rápido.

B) Alguns módulos do AutoML resolveram o problema da visualização de dados, ajustes e resultados, sendo apontados como solução de algoritmos baseados em qualquer princípio que possa ser utilizado para aproximação de funções.

C) São apontadas como principais vantagens do AutoML: a possibilidade de execução em paralelo; a possibilidade de se trabalhar com streaming de dados durante o treinamento; e o melhor desempenho para conjuntos de dados muito grandes.

D) O AutoML permite ajuste a dados desbalanceados e soluções ótimas para o caso de superdispersão. Como desvantagem, pode ser citado o custo elevado em termos de tempo e processamento.

20. Precisão e Acurácia são duas medidas que se destacam quando se deseja avaliar o desempenho de aplicações que utilizam inteligência artificial. Isso se dá devido a ser fundamental que algumas medidas sejam definidas para avaliar a qualidade dos modelos. A respeito dessas duas medidas, considere as seguintes afirmativas:

I	A precisão está relacionada a um conjunto de repetições de estimativas realizadas, dando uma ideia de variabilidade dos resultados.
II	A precisão e a acurácia estão sempre relacionadas de modo diretamente proporcional, sendo seus valores próximos um do outro.
III	A acurácia está relacionada ao valor de um resultado conforme a sua proximidade com o seu valor de referência real.
IV	Ter uma alta precisão implica em ter uma alta dispersão, indicando que o modelo testado contém um baixo erro sistemático.

Estão corretas as afirmativas:

A) III e IV. **B)** I e II. **C)** II e IV. **D)** I e III.