



INSTITUTO NACIONAL DE PESQUISA ESPACIAL – INPE

INTELIGÊNCIA ARTIFICIAL APLICADA À ANÁLISE E MODELAGEM DE DADOS DE FENÔMENOS NATURAIS (PQ04)



SUA PROVA

- Além deste caderno contendo **5 (cinco)** questões discursivas **com as respectivas folhas de rascunho**, você receberá do fiscal de prova as folhas de textos definitivos;



TEMPO

- Você dispõe de **4 (quatro) horas** para a realização da prova;
- **2 (duas) horas** após o início da prova, é possível retirar-se da sala, sem levar o caderno de questões;
- A partir dos **30 (trinta) minutos** anteriores ao término da 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;
- Anotar informações relativas às respostas em qualquer outro meio que não seja no caderno de questões e nas folhas de textos definitivos;
- Levantar da cadeira sem autorização do fiscal de sala;
- Usar o sanitário ao término da prova, após deixar a sala.



INFORMAÇÕES GERAIS

- Verifique se seu caderno de questões está completo, sem repetição de questões ou falhas. Caso contrário, **notifique imediatamente o fiscal da sala**, para que sejam tomadas as devidas providências;
- Confira seus dados pessoais, especialmente nome, número de inscrição e documento de identidade e leia atentamente as instruções para preencher as folhas de textos definitivos;
- Para o preenchimento das folhas de textos definitivos, use somente caneta esferográfica, fabricada em material transparente, com tinta preta ou azul;
- Assine seu nome apenas no(s) espaço(s) reservado(s) no cartão de respostas;
- Caso você tenha recebido caderno de cargo **diferente** do impresso em suas folhas de textos definitivos, o fiscal deve ser **obrigatoriamente** informado para o devido registro na ata da sala;
- O preenchimento das folhas de textos definitivos é de sua responsabilidade e **não será permitida a troca de folha de texto definitivo em caso de erro cometido pelo candidato**;
- Para fins de avaliação, serão levadas em consideração apenas os textos das folhas de textos definitivos;
- A FGV coletará as impressões digitais dos candidatos na lista de presença;
- Os candidatos serão submetidos ao sistema de detecção de metais quando do ingresso e da saída de sanitários durante a realização das provas.
- **Boa prova!**

QUESTÃO 1

Inteligência Artificial (IA) tem como objetivo fazer com que os computadores executem tarefas que até recentemente requereriam conhecimentos e habilidades humanas para executá-las. Em geral, são realizadas análises de grandes volumes de dados e a identificação de padrões, o que é feito por meio de métodos diversos.

Sobre o tema, responda aos itens a seguir.

- A) Apresente a diferença entre IA e aprendizado de máquina.
- B) Diferencie *IA Tradicional* de *IA Generativa*.
- C) Explique o que são *Aprendizado de Máquina Supervisionado* e *Aprendizado de Máquina Não-Supervisionado*. Dê exemplos de aplicações.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 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

51

52

53

54

55

56

57

58

59

60

QUESTÃO 2

Fenômenos naturais são muitas vezes registrados e mapeados em imagens digitais. Tais imagens podem ser obtidas por sensores locais ou remotos, como estações meteorológicas, drones e satélites. O volume de dados gerado por tais sensores tende a crescer rapidamente, dependendo da frequência em que os dados/imagens são coletados.

Desse modo, há uma grande preocupação da área de inteligência artificial e de reconhecimento de padrões em desenvolver mecanismos escaláveis que permitam armazenar e recuperar tais dados de modo a atender às consultas que os especialistas ou usuários do domínio dos dados necessitem fazer. Para isso, são desenvolvidas representações sucintas desses dados, muitas vezes denominadas de “assinaturas” que são utilizadas no processo de armazenar e recuperar os itens de dados (imagens, por exemplo). Tais assinaturas podem trazer ambiguidade, isto é, uma mesma assinatura representar mais do que uma imagem, gerando falsos positivos.

Considerando que as imagens obtidas possuem tamanho de 1024 por 1024 pixels, ou seja, são matrizes bidimensionais de 1024 x 1024 (linhas $M = 1024$ e colunas $N = 1024$), com 128 níveis de quantização monocromática, e que a base possua 1 milhão de imagens e que cresça com proporção de 15% ao ano,

- A) indique uma representação sintética (assinatura) para cada uma dessas imagens que seja escalável, ou seja, que possua custo linear no tamanho da imagem;
- B) indique como pode ser feito o tratamento de ambiguidade (falsos positivos) dessa representação escolhida.
- C) Este cenário de uso da base deve permitir o acesso rápido às imagens utilizando essa assinatura. Justifique sua resposta e as escolhas tomadas nessa decisão.



Exemplo de imagem monocromática da base de imagens indicada no enunciado da questão.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 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

51

52

53

54

55

56

57

58

59

60

QUESTÃO 3

A partir da utilização de imagens de satélites orbitais que monitoram a superfície da Terra, como, por exemplo, os da série *Landsat*, foram obtidos polígonos que delimitam mudanças no uso e na cobertura da Terra. Estes polígonos representam áreas relacionadas a mudanças como desmatamento, queimadas e colheitas.

Cada polígono possui o valor médio das bandas individuais dos diferentes canais do sensor imageador (TM, TM+ ou OLI), bem como índices espectrais como NDVI (Índice de Vegetação por Diferença Normalizada), NBR (Índice de Queimadas Normalizado) e BAI (Índice de Área Queimada) da data da imagem que foi utilizada para o mapeamento, além da informação sobre qual imagem foi utilizada.

A) Com base nessas informações, responda como a Inteligência Artificial pode ser utilizada para esse tipo de problema, nas seguintes aplicações:

A₁ geração de novos polígonos de mudança.

A₂ validação da classificação dos polígonos de mudança.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 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

51

52

53

54

55

56

57

58

59

60

Questão 4

É sabido que a análise de *agrupamentos de dados*, vulgarmente conhecida como *clusterização de dados*, é uma tarefa de descoberta do conhecimento por excelência. Por ser uma tarefa intrinsecamente não-supervisionada, a clusterização tem uma ampla gama de aplicações em aprendizado de máquinas e reconhecimento de padrões, tais como mineração de dados, segmentação de imagens, redução de grandes volumes de dados etc., sendo o algoritmo K-médias um dos métodos de clusterização particional mais conhecido e utilizado.

Isto posto, com relação ao algoritmo K-médias responda aos itens a seguir.

- A) **Descreva seu funcionamento.**
- B) **Liste 2 (duas) vantagens e, também, 2 (duas) limitações do algoritmo e possíveis formas de mitigá-las.**
- C) **Disserte sobre a etapa de validação do número de *clusters*.**
- D) **O algoritmo K-médias é capaz de lidar adequadamente com *outliers* presentes nos dados? Justifique sua resposta.**

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 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

51

52

53

54

55

56

57

58

59

60

QUESTÃO 5

Com o avanço das técnicas de aprendizado em profundidade, cresce também a busca por sua implementação em dispositivos portáteis, como drones e, inclusive, satélites. É sabido que redes neurais profundas possuem desempenho promissor em diversas tarefas importantes na área de sensoriamento remoto que envolvam imagens, tais como classificação, segmentação e detecção de mudanças.

Suponha o seguinte cenário: queremos embarcar um modelo de aprendizado profundo em um drone com recursos limitados de armazenamento e processamento para classificação de determinados objetos em tempo real. No entanto, dispomos de algumas poucas imagens rotuladas para o seu treinamento e uma grande quantidade de imagens sem anotação. Ademais, a base conta com um bom nível de desbalanceamento entre as classes.

Com base neste cenário, responda às seguintes questões:

- A) No que diz respeito à *arquitetura da rede neural*, responda aos itens a seguir.**
1. Apresente uma proposta de arquitetura de rede neural que enderece a situação acima.
 2. Dê exemplos de redes neurais convolucionais do tipo "leve" e justifique a sua escolha.
 3. Indique os critérios usualmente utilizados para comparar tais redes leves.
- B) Apresente uma metodologia de treinamento que consiga mitigar a tarefa apresentada, contendo as seguintes informações:**
1. Como a atualização dos pesos da rede será realizada?
 2. Qual a função de custo?
 3. Como tratar desbalanceamento de classes?
 4. Como tratar a questão do recurso limitado de processamento e armazenamento?
 5. Como fazer uso de todo o conjunto de dados, isto é, imagens anotadas e não anotadas?

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 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

51

52

53

54

55

56

57

58

59

60

Realização

