

IFPI

CURSO: TÉCNICO EM DESENVOLVIMENTO DE SISTEMAS

DISCIPLINA: TÓPICOS ESPECIAIS EM PROGRAMAÇÃO

EXERCICIO – Introdução à ciência de dados, big data e aprendizado de máquina

1. Qual das opções a seguir é um dos pilares fundamentais da Ciência de Dados?

- a) Design Gráfico
- b) Estatística
- c) Marketing Digital
- d) Fotografia

2. Qual dos seguintes elementos faz parte dos 3 V's do Big Data?

- a) Velocidade, Variedade, Valor
- b) Valor, Volume, Verdade
- c) Volume, Velocidade, Variedade
- d) Variedade, Velocidade, Visão

3. Em relação ao conceito de "Variedade" no Big Data, o que ele significa?

- a) O número total de registros armazenados
- b) A diversidade de tipos de dados, como texto, imagem e vídeo
- c) A rapidez com que os dados são gerados
- d) A quantidade de dados que uma empresa coleta

4. Qual das seguintes opções define corretamente o Machine Learning?

- a) Método manual de análise de grandes volumes de dados
- b) Técnica de programação baseada em regras rígidas
- c) Algoritmos que permitem que máquinas aprendam com dados e melhorem ao longo do tempo
- d) Tecnologia exclusiva da área de robótica

5. O que é aprendizado supervisionado?

- a) Aprendizado sem a necessidade de dados rotulados
- b) Treinamento de um modelo com dados rotulados
- c) Aprendizado em tempo real a partir de novos dados
- d) Aprendizado baseado em redes neurais profundas

6. Qual das seguintes alternativas é um exemplo de uso da aprendizagem não supervisionada?

- a) Diagnóstico médico baseado em histórico de pacientes
- b) Previsão de faturamento mensal de uma empresa
- c) Agrupamento de clientes com base em padrões de comportamento
- d) Detecção de fraudes financeiras

7. Em qual das seguintes situações é mais comum usar aprendizado por reforço?

- a) Previsão de vendas
- b) Classificação de e-mails
- c) Treinamento de robôs autônomos para navegar em um ambiente
- d) Agrupamento de documentos semelhantes

8. Qual das opções melhor descreve "Overfitting" em um modelo de aprendizado de máquina?

- a) O modelo se ajusta tão bem aos dados de treinamento que tem dificuldade em generalizar para novos dados
- b) O modelo tem desempenho excepcional tanto nos dados de treinamento quanto nos dados de teste
- c) O modelo é muito simples e não consegue capturar padrões complexos nos dados
- d) O modelo se ajusta adequadamente aos dados de treinamento e generaliza bem para novos dados

9. Qual é o objetivo principal do pré-processamento de dados em projetos de Ciência de Dados?

- a) Aumentar o número de variáveis em um conjunto de dados
- b) Limpar e organizar os dados para melhorar o desempenho do modelo
- c) Reduzir o tamanho dos dados para economizar espaço de armazenamento

- d) Criar visualizações mais eficazes para relatórios

10. Qual é a principal diferença entre classificação e regressão em aprendizado de máquina?

- a) A classificação prevê valores contínuos, enquanto a regressão prevê categorias
- b) A classificação prevê categorias, enquanto a regressão prevê valores contínuos
- c) Classificação e regressão são a mesma coisa
- d) Nenhuma das anteriores

11. Qual das seguintes alternativas não é um algoritmo de aprendizado supervisionado?

- a) Regressão linear
- b) Redes neurais convolucionais
- c) K-Means
- d) Árvores de decisão

12. O que é aprendizado não supervisionado?

- a) Aprendizado a partir de dados sem rótulos ou categorias predefinidas
- b) Aprendizado que requer supervisão contínua de um analista de dados
- c) Um método que se aplica apenas a dados rotulados
- d) Um algoritmo que só funciona em grandes volumes de dados

13. Qual das seguintes opções é um exemplo de aprendizado supervisionado?

- a) Previsão do valor de imóveis com base em características como tamanho e localização
- b) Agrupamento de produtos semelhantes em categorias
- c) Detecção de padrões desconhecidos em dados não rotulados
- d) Segmentação de clientes em grupos de interesse

14. O que são "features" em um conjunto de dados para aprendizado de máquina?

- a) Exemplos de dados rotulados
- b) Variáveis ou atributos usados para fazer previsões
- c) Os valores de saída que o modelo tenta prever
- d) As partes do código usadas no treinamento do modelo

15. Qual das alternativas abaixo representa uma técnica de pré-processamento de dados?

- a) Normalização de dados
- b) Execução de testes A/B
- c) Treinamento de um modelo de IA
- d) Criação de algoritmos de recomendação

16. O que é "subconjunto de validação" em um modelo de aprendizado de máquina?

- a) Um conjunto de dados usado apenas para testar o modelo após o treinamento
- b) Um conjunto de dados usado para ajustar os parâmetros do modelo durante o treinamento
- c) Um conjunto de dados usado exclusivamente para otimizar a velocidade do treinamento
- d) Um conjunto de exemplos de código para validar o desempenho do modelo

17. Qual é o principal benefício do uso de algoritmos de aprendizado por reforço?

- a) Alta precisão em previsões lineares
- b) Aprendizado a partir de tentativas e erros, com recompensas e punições
- c) Redução do número de dados necessários para treinar o modelo
- d) Exclusividade para o uso em jogos de vídeo

18. Qual das opções a seguir é um dos maiores desafios em projetos de Big Data?

- a) Dificuldade em encontrar fontes de dados
- b) Armazenamento e processamento de grandes volumes de dados
- c) Falta de interesse das empresas em usar Big Data
- d) Desinteresse dos clientes em fornecer dados

19. O que são outliers em um conjunto de dados?

- a) Dados que representam a média de um conjunto de variáveis
- b) Dados que se desviam significativamente dos outros dados do conjunto
- c) Dados que são ignorados durante o pré-processamento
- d) Dados categóricos que aparecem com mais frequência

20. Qual é a função principal de um algoritmo de clusterização?

- a) Prever uma variável de saída específica
- b) Agrupar dados semelhantes sem rótulos predefinidos
- c) Testar a precisão de um modelo de aprendizado
- d) Identificar outliers em um conjunto de dados