



# Concurso Público Fiocruz 2023

## Pesquisador em Saúde Pública

### Prova Discursiva

#### PE32

## Ciência de Dados em Saúde

#### Espelho de Resposta

**Pontuação de cada Questão Discursiva conforme Anexo II do Edital nº 3, de acordo com a Unidade detentora da vaga.**

Espera-se que o candidato, no desenvolvimento do tema, tenha feito considerações técnicas adequadas sobre os seguintes pontos:

#### Questão 01

A Ciência de Dados representa um processo interdisciplinar e iterativo que combina técnicas estatísticas, computacionais e conhecimento especializado do domínio para extrair conhecimentos valiosos e significativos sobre os dados. A identificação precisa do problema, a análise exploratória, a aplicação dos paradigmas de aprendizado e a apresentação dos resultados constituem etapas essenciais nesse processo, com aplicações cada vez mais abrangentes e impactantes em diversas áreas da saúde.

#### 1. Identificação do problema a ser resolvido

A Ciência de Dados é uma disciplina interdisciplinar que integra conhecimentos provenientes de áreas como estatística, computação, matemática e expertise no domínio específico, com o propósito de extrair informações valiosas e conhecimentos significativos a partir de conjuntos de dados complexos. Requer a adoção de um processo rigoroso e detalhado, que abrange diversas etapas e técnicas especializadas para lidar com os desafios e oportunidades inerentes à análise de grandes quantidades de dados.

O estágio inicial e fundamental na prática da Ciência de Dados consiste na identificação precisa e detalhada do problema a ser abordado. Este processo demanda uma compreensão aprofundada das necessidades do contexto de pesquisa a ser investigado, a definição clara de objetivos específicos e a formulação precisa das questões a serem respondidas. Subsequentemente, é essencial realizar a coleta dos dados pertinentes para a resolução do problema em questão, os quais podem se apresentar em formatos estruturados ou não estruturados.

#### 2. Dados estruturados e não estruturados

Dados estruturados são dados organizados em formatos predefinidos e facilmente identificáveis, como tabelas em bancos de dados ou planilhas. Nesse formato, as informações são dispostas em linhas e colunas, o que facilita a análise e o processamento por meio de algoritmos e técnicas estatísticas. Dados não estruturados, por outro lado, não seguem um formato pré-determinado e podem ser mais complexos de serem interpretados. Exemplos de dados não estruturados

incluem textos, imagens, vídeos, áudios, e-mails, posts em redes sociais, entre outros. Esses tipos de dados requerem técnicas avançadas de processamento de linguagem natural e visão computacional para extrair informações relevantes e significativas.

### **3. Análise exploratória de dados**

Após a aquisição dos dados, a etapa de análise exploratória desempenha um papel fundamental no processo. Neste estágio, os dados são explorados com o intuito de identificar padrões, tendências, valores discrepantes e relações entre as variáveis. Diversas técnicas são empregadas nesse contexto, incluindo visualização de dados, estatísticas descritivas e análise de correlação, visando obter uma compreensão abrangente e aprofundada do conjunto de dados em análise.

### **4. Paradigmas de aprendizado supervisionado e não supervisionado**

Durante a análise de dados na Ciência de Dados, são empregados os paradigmas de aprendizado supervisionado e não supervisionado para extrair conhecimentos e informações significativas sobre os dados. No aprendizado supervisionado, o modelo é treinado com um conjunto de dados rotulados, ou seja, em que as saídas desejadas já são conhecidas. O objetivo é aprender a mapear as entradas para as saídas corretas e, posteriormente, ser capaz de fazer previsões ou classificações em novos dados. Algoritmos de aprendizado supervisionado incluem regressão linear, regressão logística, árvores de decisão, entre outros. Já no aprendizado não supervisionado, o modelo é treinado com dados não rotulados, ou seja, sem informações prévias sobre as saídas desejadas. O objetivo é explorar a estrutura dos dados, identificar padrões, *clusters* ou relações intrínsecas entre as variáveis. Algoritmos de aprendizado não supervisionado incluem análise de *cluster*, análise de componentes principais, entre outros.

Ao aplicar esses paradigmas, os cientistas de dados conseguem extrair informações valiosas dos dados, identificar tendências, padrões ocultos e conhecimentos que podem ser utilizados para tomadas de decisão informadas e estratégicas em diversos contextos da pesquisa em saúde.

### **5. Apresentação dos resultados e possíveis aplicações em saúde**

Por fim, a apresentação dos resultados obtidos desempenha um papel crucial na comunicação eficaz das descobertas realizadas. Neste sentido, ferramentas como visualizações de dados, relatórios detalhados e *dashboards* interativos são comumente utilizadas para transmitir de forma clara e compreensível os resultados derivados da análise de dados. A Ciência de Dados tem desempenhado um papel cada vez mais relevante na área da saúde pública e coletiva, oferecendo oportunidades para melhorar a eficiência, a precisão e a eficácia das intervenções e políticas de saúde.

Alguns exemplos de aplicação da Ciência de Dados nesse contexto incluem: (i) Monitoramento de epidemias e surto de doenças: a análise de dados epidemiológicos em tempo real pode ajudar na detecção precoce de surtos de doenças infecciosas. Modelos preditivos baseados em dados podem auxiliar na previsão de padrões de propagação de doenças e na identificação de áreas de risco, permitindo uma resposta mais rápida e eficaz por parte das autoridades de saúde; (ii) Análise de dados demográficos e sociais: a utilização de dados demográficos e sociais em conjunto com informações de saúde pode fornecer informações valiosas sobre os determinantes sociais da saúde e as desigualdades existentes. Essa análise pode ajudar na identificação de grupos populacionais vulneráveis, orientando a formulação de políticas públicas mais direcionadas e eficazes; (iii) Predição de internações hospitalares: Modelos de aprendizado de máquina podem ser empregados para prever a probabilidade de internações hospitalares com base em dados clínicos e demográficos dos pacientes. Essa abordagem pode auxiliar na alocação de recursos de forma mais eficiente, na identificação de pacientes com maior risco de complicações e na melhoria do planejamento de cuidados de saúde; (iv) Monitoramento de doenças crônicas: a análise de dados longitudinais de pacientes com doenças crônicas, como diabetes, hipertensão e doenças cardiovasculares, pode ajudar na identificação de padrões de progressão da doença, na avaliação da eficácia de tratamentos e na personalização de intervenções para melhorar os resultados de saúde; e (v) Análise de Sentimento em mídias sociais: a mineração de dados em plataformas de mídias sociais pode ser utilizada para

monitorar o sentimento e as preocupações da população em relação a questões de saúde pública, como vacinação, políticas de saúde e epidemias. Essa análise de sentimentos pode fornecer informação sobre a percepção pública e orientar campanhas de conscientização e comunicação.

Esses são apenas alguns exemplos de como a Ciência de Dados pode ser aplicada de forma inovadora e impactante na área da saúde pública e coletiva, contribuindo para a melhoria da qualidade dos serviços de saúde, a prevenção de doenças e a promoção do bem-estar da população.

## **Questão 02**

A adoção de técnicas de Ciência de Dados e IA pelo SUS enfrenta diversos desafios éticos que requerem atenção e ação para garantir que a tecnologia seja implementada de forma ética, transparente e responsável, visando benefícios reais para a saúde pública e o bem-estar da população.

A governança e segurança de dados sensíveis são pilares essenciais para a adoção segura e ética dessas tecnologias no contexto do SUS, assegurando a confidencialidade, integridade e disponibilidade das informações de saúde dos usuários.

A privacidade dos dados pessoais e sensíveis dos pacientes é um princípio fundamental que deve ser respeitado e protegido na implementação dessas tecnologias no SUS, assegurando a confiança da população e o cumprimento de normas éticas e legais de privacidade.

O racismo algorítmico é um desafio ético importante que deve ser abordado na implementação de Inteligência Artificial, incluindo no contexto do Sistema Único de Saúde, a fim de garantir que as decisões automatizadas sejam justas, imparciais e respeitem a diversidade e dignidade de todos os indivíduos.

### **1) Desafios éticos na adoção de Inteligência Artificial**

Na adoção de Inteligência Artificial (IA) pelo Sistema Único de Saúde (SUS), alguns dos principais desafios éticos incluem:

**Viés e discriminação:** a presença de viés nos algoritmos de IA pode resultar em decisões discriminatórias, especialmente em relação a grupos minoritários, ampliando desigualdades e injustiças sociais.

**Privacidade e proteção de dados:** o uso de dados sensíveis de saúde dos usuários levanta preocupações sobre a privacidade, segurança e consentimento no processamento dessas informações.

**Transparência e interpretabilidade:** a opacidade dos modelos de IA pode dificultar a compreensão de como as decisões são tomadas, tornando desafiador responsabilizar algoritmos por suas ações.

**Equidade no acesso e tratamento:** garantir que a IA no SUS promova a equidade no acesso aos serviços de saúde e no tratamento dos pacientes, evitando discriminações e disparidades injustas.

**Responsabilidade e prestação de contas:** estabelecer mecanismos claros de responsabilidade e prestação de contas para as decisões tomadas por sistemas de IA assegurando a transparência e a ética em seu uso.

**Treinamento e capacitação:** capacitar profissionais de saúde e gestores do SUS para compreender e utilizar de forma ética as tecnologias de IA garantindo que sejam aplicadas de maneira responsável.

### **2) Governança e segurança de dados sensíveis**

A governança e segurança de dados sensíveis são aspectos fundamentais em qualquer ambiente que lide com informações críticas de saúde. Alguns pontos importantes sobre governança e segurança de dados sensíveis incluem:

**Proteção da privacidade:** garantir a privacidade dos dados sensíveis dos pacientes é essencial para preservar a confidencialidade e a integridade das informações de saúde.

Conformidade com regulamentações: cumprir as regulamentações de proteção de dados, como a Lei Geral de Proteção de Dados (LGPD) no Brasil, é crucial para garantir a legalidade e a ética no tratamento das informações sensíveis.

Controle de acesso: implementar medidas de controle de acesso e autenticação para garantir que apenas pessoas autorizadas possam acessar e manipular os dados sensíveis, reduzindo o risco de violações de segurança.

Criptografia e anonimização: utilizar técnicas de criptografia e anonimização de dados sensíveis pode ajudar a proteger as informações durante o armazenamento, transmissão e processamento, minimizando o risco de exposição indevida.

Monitoramento e auditoria: estabelecer mecanismos de monitoramento e auditoria dos dados sensíveis permite identificar e responder rapidamente a possíveis violações de segurança, garantindo a conformidade e a proteção das informações.

### **3) Privacidade**

A privacidade é um aspecto crucial a ser considerado na adoção de tecnologias baseadas em Ciência de Dados e Inteligência Artificial, especialmente no contexto do Sistema Único de Saúde (SUS) e no tratamento de dados sensíveis de saúde. Alguns pontos relevantes sobre privacidade incluem:

Consentimento informado: garantir que os pacientes forneçam consentimento informado para o uso de seus dados pessoais e de saúde em aplicações tecnológicas, respeitando sua autonomia e privacidade.

Minimização de dados: adotar práticas de minimização de dados, ou seja, coletar apenas as informações necessárias para a finalidade específica, a fim de reduzir o risco de exposição e proteger a privacidade dos usuários.

Anonimização e pseudonimização: utilizar técnicas de anonimização e pseudonimização para proteger a identidade dos pacientes nos dados utilizados para treinar modelos, preservando sua privacidade.

Segurança da informação: implementar medidas robustas de segurança da informação, como criptografia, controle de acesso e monitoramento, para proteger os dados pessoais e sensíveis contra acessos não autorizados e violações de segurança.

Transparência e *accountability*: promover a transparência sobre como os dados são coletados, armazenados e utilizados em sistemas digitais, além de estabelecer mecanismos de prestação de contas para garantir a responsabilidade no tratamento das informações.

### **4) Racismo algorítmico**

O termo "racismo algorítmico" refere-se à discriminação ou viés racial que pode ser perpetuado por algoritmos de Inteligência Artificial, resultando em decisões injustas ou discriminatórias com base em características raciais. Alguns pontos relevantes sobre o racismo algorítmico incluem:

Viés nos dados: algoritmos de IA podem refletir e amplificar preconceitos existentes na sociedade se forem treinados com conjuntos de dados que contenham viés racial, levando a decisões discriminatórias.

Impacto desproporcional: o racismo algorítmico pode afetar de forma desproporcional grupos minoritários, resultando em disparidades no acesso a serviços de saúde, emprego, crédito, entre outros, e perpetuando injustiças sociais.

Necessidade de mitigação: identificar e mitigar o viés racial nos algoritmos é essencial para garantir a equidade e justiça nas decisões automatizadas, exigindo transparência, auditoria e correção permanente dos modelos.

Responsabilidade e ética: as organizações que desenvolvem e implementam sistemas de IA têm a responsabilidade ética de garantir que seus algoritmos não perpetuem o racismo ou qualquer forma de discriminação, adotando práticas que promovam a equidade e a inclusão.