1. Explique com suas palavras o conceito de Machine Learning.
2. Diferencie modelos descritivos de modelos preditivos.
3. Pesquise aplicações de modelos de machine learning voltados para sua área de interesse e descreva exemplificando o que encontrou.
4. Escolha um das aplicações que citou acima e descreva como você imaginaria que seriam aplicadas as etapas do Crisp-DM a esse projeto.

Relembre as etapas: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem e avaliação dos modelos.

* + **Explique com suas palavras o conceito de Machine Learning**

Machine Learning (Aprendizado de Máquina) é uma área da inteligência artificial onde **algoritmos aprendem padrões** e realizem tarefas **a partir de dados**, sem serem explicitamente programados para cada tarefa.

Em vez de seguir regras fixas, o modelo 'treina' com exemplos passados para **fazer previsões ou tomar decisões** em novos dados.

A lógica do processo envolve três etapas principais:

1. **Disponibilização de dados**: O modelo recebe dados históricos (ex: vendas passadas, imagens médicas).
2. **Identificação de padrões**: O algoritmo analisa estatisticamente os dados para encontrar relações (ex: "clientes com alto gasto tendem a cancelar").
3. **Avaliação de desempenho**: Testa-se o modelo com novos dados para medir sua precisão (ex: acurácia de 90% em prever churn).

Em vez de seguir regras fixas, o Machine Learning **opera com base em conceitos estatísticos** (como distribuições e correlações) para generalizar aprendizados. É como ensinar uma criança a reconhecer animais: mostramos fotos (dados), ela extrai padrões (formato, cores), e depois classifica sozinha — mas com matemática por trás.

**Com Machine Learning, transformamos dados brutos em decisões inteligentes e na saúde, isso pode salvar vidas.**

* + **Diferencie Modelos Descritivos de Modelos Preditivos**

Ambos são essenciais para decisões baseadas em dados.

**Modelos Descritivos** explicam o que já ocorreu;

**Modelos Preditivos** antecipam o que virá.

* **Modelos Descritivos**
* **O que fazem:** Analisam dados históricos para **entender padrões e comportamentos**.
* **Objetivo:** Responder *"O que aconteceu?"*
* **Exemplos práticos:**
  + Segmentação de clientes (agrupamento por similaridade).
  + Identificação de produtos frequentemente comprados juntos.
* **Ferramentas:** Clustering (K-means), Regras de Associação (Apriori).
* **Modelos Preditivos**
* **O que fazem:** Usam dados passados para **prever resultados futuros**.
* **Objetivo:** Responder *"O que pode acontecer?"*
* **Exemplos práticos:**
  + Previsão de vendas.
  + Detecção de fraude em transações.
* **Ferramentas:** Regressão, Classificação (Random Forest, XGBoost).

| **Tipo** | **Pergunta Respondida** | **Exemplo em Saúde** | **Técnica** |
| --- | --- | --- | --- |
| **Descritivo** | "O que aconteceu?" | Segmentação de pacientes em UTI | K-means |
| **Preditivo** | "O que pode acontecer?" | Prever readmissão hospitalar | Random Forest |

* + **Pesquise aplicações de modelos de machine learning voltados para sua área de interesse e descreva exemplificando o que encontrou.**

**Aplicações de Machine Learning na Área da Saúde**

#### ****1. Diagnóstico Médico Automatizado****

* **Modelo Preditivo**: Classificação de imagens (ex: raio-X, ressonância) para detectar doenças.
  + Exemplo: Rede Neural Convolucional (CNN) para identificar câncer de mama em mamografias com 95% de acurácia.
  + Impacto: Reduz tempo de diagnóstico e erros humanos.

#### ****2. Previsão de Risco de Doenças****

* **Modelo Preditivo**: Análise de dados clínicos (ex: idade, histórico familiar, exames).
  + Exemplo: Regressão Logística para prever diabetes em pacientes com base em glicose, IMC e pressão arterial.
  + Impacto: Permite intervenções preventivas.

#### ****3. Monitoramento de Pacientes em Tempo Real****

* **Modelo Descritivo**: Agrupamento de padrões vitais (ex: batimentos cardíacos, saturação de O₂).
  + Exemplo: K-means para classificar pacientes em grupos de risco (baixo, médio, alto) em UTIs.
  + Impacto: Alerta precoce para equipes médicas.

#### ****4. Otimização de Recursos Hospitalares****

* **Modelo Preditivo**: Séries temporais para prever demanda por leitos ou medicamentos.
  + Exemplo: ARIMA para prever picos de internações por COVID-19.
  + Impacto: Evita sobrecarga do sistema de saúde.
  + **Escolha um das aplicações que citou acima e descreva como você imaginaria que seriam aplicadas as etapas do Crisp-DM a esse**

**projeto.**

Previsão de Risco de Readmissão Hospitalar

***Fluxo****: Dados brutos → Limpeza → Modelagem → Alertas no prontuário → Monitoramento contínuo.*

**1. Entendimento do Negócio**

* **Objetivo**: Reduzir readmissões de pacientes em 30 dias (custo alto para hospitais e risco para pacientes).
* **Partes Envolvidas**: Equipe médica, administradores hospitalares, seguradoras.
* **Critérios de Sucesso**: Modelo com pelo menos 85% de recall (identificar a maioria dos pacientes de alto risco).

#### ****2. Entendimento dos Dados****

* **Fontes de Dados**:
  + Prontuários eletrônicos (idade, diagnósticos, medicamentos).
  + Histórico de internações (tempo de permanência, complicações).
  + Dados socioeconômicos (acesso a cuidados pós-alta).
* **Análise Exploratória**:
  + Verificar correlações (ex: pacientes diabéticos têm 2x mais chances de readmissão).
  + Identificar missing values (ex: 15% dos registros sem info sobre acompanhamento pós-alta).

#### ****3. Preparação dos Dados****

* **Limpeza**:
  + Imputar missing values (ex: preencher "acompanhamento pós-alta" com "não" se não houver registro).
  + Remover outliers (ex: tempo de internação > 30 dias).
* **Transformação**:
  + Codificar variáveis categóricas (ex: "tipo de diagnóstico" em one-hot encoding).
  + Normalizar dados numéricos (ex: idade, número de medicamentos).

#### ****4. Modelagem****

* **Algoritmos**:
  + **Regressão Logística** (para interpretabilidade pelos médicos).
  + **Random Forest** (para capturar relações complexas).
* **Treinamento**:
  + Usar 70% dos dados para treino, 30% para teste.
  + Balancear classes (readmitidos/não readmitidos) com SMOTE.

#### ****5. Avaliação****

* **Métricas**:
  + **Recall** (priorizar identificar o máximo de pacientes de risco).
  + **AUC-ROC** (avaliar trade-off entre sensibilidade e especificidade).
* **Validação com Médicos**:
  + Apresentar casos onde o modelo acertou/errou para ajustes clínicos (ex: incluir variável "histórico familiar").

#### ****6. Implantação (Fase Extra do CRISP-DM)****

* **Integração**:
  + Alertas no prontuário eletrônico para pacientes classificados como alto risco.
* **Monitoramento**:
  + Recalibrar o modelo a cada 6 meses com novos dados.