Projeto de Gerenciamento de Dados em Estudo Clínico Simulado sobre Hipertensão

1. Objetivo do Projeto
Desenvolver um sistema integrado para coleta, tratamento e análise de dados clínicos simulados, com foco em hipertensão. O projeto visa demonstrar competências em pesquisa clínica, análise de dados e aplicação de boas práticas na área da saúde.

2. Visão Geral
O projeto abrange todas as etapas essenciais de um pipeline clínico:

- **Banco de Dados no REDCap:** Estruturação de formulários e campos para coleta de dados clínicos padronizados.
- **ETL e Análise com Python:** Exportação dos dados do REDCap, aplicação de scripts de limpeza, transformação e enriquecimento.
- **Modelos de IA:** Implementação de modelos preditivos simples para estimar riscos de complicações associadas à hipertensão.

Passo 1: Configuração do Banco de Dados Clínico no REDCap

1. Objetivo
Estruturar um banco de dados clínico no REDCap para coleta padronizada de informações sobre pacientes hipertensos.

2. Estrutura dos Campos Criados

- **Informações Demográficas:**
  - `patient_id` – Identificador do paciente (texto)
  - `idade` – Idade em anos (número)
  - `sexo` – Sexo biológico (radio: Masculino, Feminino)
  - `imc` – Índice de Massa Corporal (número)
- **Medidas Clínicas:**
  - `pressao_sistolica` – Pressão arterial sistólica (número)
  - `pressao_diastolica` – Pressão arterial diastólica (número)
  - `frequencia_cardiaca` – Frequência cardíaca em bpm (número)
  - `medicacao` – Medicamentos em uso (checkbox: Diurético, Beta-bloqueador, IECA, ARA II, Cálcio antagonista)
- **Desfecho Clínico:**
  - `complicacao` – Presença de complicações (radio: Sim, Não)
  - `tipo_complicacao` – Tipo de complicação registrada (texto)

Resultado Final:
Uma base de dados bem estruturada e compatível com exportação para análise

estatística e preditiva, promovendo padronização e qualidade na coleta de dados clínicos.

Passo 2: Geração de Dados Clínicos Fictícios para Testes

1. Inserção Manual de Dados

- **Volume:** Criação manual de 20 a 30 registros fictícios diretamente no REDCap.
- **Objetivo:** Simular um conjunto inicial de pacientes para testes e validações da base de dados.

2. Geração Automática com Python

- **Script Desenvolvido:** Código em Python para gerar dados clínicos aleatórios seguindo a estrutura do banco.
- **Finalidade:** Ampliar rapidamente a base de dados com registros sintéticos para testes de ETL, análise e modelagem preditiva.

Resultado Final:
Base de dados enriquecida com registros realistas e variados, garantindo volume e diversidade suficientes para validação dos processos analíticos e do pipeline completo.

Passo 3: ETL e Análise de Dados Clínicos com REDCap, Python e Power BI

1. Coleta de Dados

- **Fonte:** Exportação dos dados clínicos do sistema REDCap para arquivos CSV.
- **Objetivo:** Integrar os dados ao ambiente Python para análise e visualização.

2. ETL com Python

- **Processamento:** Desenvolvimento de script para limpeza, transformação e enriquecimento dos dados.
- **Tratamento:** Padronização de formatos, tratamento de dados faltantes e preparação para análise.

3. Visualização com Power BI

- **Objetivo:** Criar um dashboard interativo para explorar os dados clínicos de forma dinâmica e acessível.
- **Métricas Apresentadas:**
    - Distribuição de pacientes por sexo e faixa etária
    - Médias de pressão arterial por grupo
    - Relação entre IMC e pressão arterial
    - Taxa de complicações por tipo de medicamento

- Tendência temporal da pressão arterial

Resultado Final:
Uma solução completa de ETL e visualização que transforma dados clínicos brutos em insights valiosos, apoiando a análise exploratória e a tomada de decisão em estudos de saúde.

Passo 4: Modelo Preditivo (Python - IA)

Modelo Preditivo de Complicações em Pacientes Hipertensos

1. Objetivo
Desenvolver um modelo de Machine Learning para prever o risco de complicações em pacientes com hipertensão, auxiliando decisões médicas por meio de previsões probabilísticas.

2. Preparação dos Dados

- **Seleção de Variáveis:** Escolha de 5 variáveis clínicas relevantes.
- **Tratamento de Dados:** Preenchimento de valores ausentes e normalização dos dados.
- **Divisão dos Dados:** Separação em conjuntos de treino e teste.

3. Modelagem Preditiva

- **Algoritmo:** Random Forest com ajustes específicos para lidar com desbalanceamento de classes.
- **Treinamento:** Execução do modelo com validação cruzada.
- **Avaliação:** Uso de métricas como acurácia, recall, precisão e F1-score.
- **Importância das Variáveis:** Identificação das variáveis com maior impacto na previsão.

4. Deploy e Uso Futuro

- **Persistência do Modelo:** Salvamento do modelo treinado para reutilização.
- **Função de Previsão:** Função pronta para prever novos casos, com tratamento de erros e mensagens explicativas.

5. Boas Práticas

- **Mensagens Claras:** Feedback informativo em caso de erro ou entrada inválida.
- **Reutilização Segura:** Modelo pronto para integração com sistemas médicos.

Resultado Final:
Uma solução preditiva confiável e interpretável para apoiar a prevenção de complicações em pacientes hipertensos, com foco em robustez, clareza e aplicação prática.

Resumo do Projeto: Pipeline de Dados Clínicos com Previsão de Riscos

1. Fluxo de Trabalho

- **Coleta:** Exportação de dados clínicos do REDCap em CSV.
- **Processamento (ETL):** Limpeza, transformação e enriquecimento com Python.
- **Análise:** Visualização em dashboard interativo no Power BI.
- **Predição:** Modelo de Machine Learning em Python para prever riscos de complicações.

2. Boas Práticas Aplicadas

- **Documentação:** Código comentado e README explicativo.
- **Qualidade dos Dados:** Tratamento de valores ausentes, checagem de inconsistências e validação estatística.
- **Segurança:** Dados anonimizados.
- **Reprodutibilidade:** Uso de seed fixa e arquivo requirements.txt com dependências.

3. Apresentação do Projeto

- **ETL:** Comparação entre dados brutos e tratados; explicação das transformações.
- **Dashboard Power BI:** Filtros por idade, sexo e risco; insights como correlação entre IMC e pressão arterial.
- **Modelo Preditivo:** Acurácia, classificação e simulação de risco; destaque das variáveis mais influentes.
- **Boas Práticas:** Exemplos de tratamento de dados e justificativas técnicas (ex: uso de Random Forest).

Objetivo Final:
Criar um pipeline de ponta a ponta para dados clínicos, garantindo qualidade, segurança e reprodutibilidade, com entrega de valor por meio de visualizações e modelos preditivos.