

DataView – HealthCare Solutions

Aluno: Matheus Ernesto dos Santos.

RA: 43202.

Turma: Engenharia da Computação - 7º semestre.

Disciplina: Eletiva – Data Science.

Cenário-problema

A HealthCare Solutions, uma rede de hospitais, está passando por uma transformação digital com o objetivo de aprimorar a qualidade do atendimento ao paciente. A empresa lida com grandes volumes de dados provenientes de registros eletrônicos de saúde (EHRs), dispositivos de monitoramento, pesquisas de satisfação e sistemas administrativos.

Essas informações, porém, estão dispersas e desorganizadas, o que dificulta a extração de insights que poderiam otimizar a experiência do paciente e a eficiência operacional. A análise de dados surge, portanto, como uma ferramenta essencial para identificar padrões, prever necessidades e apoiar decisões estratégicas que impactem diretamente a jornada do paciente.

Fontes de dados

Os principais tipos de dados utilizados no projeto incluem:

- **Registros eletrônicos de saúde (EHRs):** informações clínicas e histórico médico.
- **Dispositivos de monitoramento:** sensores e wearables que acompanham sinais vitais em tempo real.
- **Pesquisas de satisfação:** feedback dos pacientes sobre atendimento, conforto e tempo de espera.
- **Dados administrativos:** gestão de leitos, recursos humanos e custos operacionais.

Fundamentos

O projeto segue as etapas clássicas de um processo completo de ciência de dados:

1. **Coleta de dados:** integração de informações provenientes de múltiplas fontes.
2. **Tratamento e limpeza:** eliminação de duplicatas, correção de inconsistências e tratamento de valores ausentes.
3. **Análise exploratória (EDA):** uso de estatísticas descritivas e visualizações para entender padrões e detectar anomalias.
4. **Modelagem preditiva:** aplicação de algoritmos de aprendizado de máquina para prever tendências e comportamentos (ex: risco de readmissão hospitalar).
5. **Visualização e interpretação:** construção de dashboards e relatórios interativos que permitam decisões baseadas em evidências.
6. **Avaliação contínua:** validação dos resultados e atualização dos modelos conforme novas informações são coletadas.

LGPD

O projeto considera os princípios da **Lei Geral de Proteção de Dados (LGPD)**, garantindo o uso ético e responsável de dados sensíveis de pacientes.

Entre as medidas adotadas estão:

- **Anonimização** de informações pessoais;
- **Controle de acesso** a dados confidenciais;
- **Consentimento informado** dos pacientes quando aplicável;
- **Transparência** nas práticas de armazenamento e uso dos dados.

Essas ações asseguram conformidade legal e reforçam a confiança entre a instituição e os pacientes.

Requisitos

10 perguntas simuladas à equipe de gestão da HealthCare Solutions (com respostas simuladas):

1. **Quais são os principais desafios no atendimento ao paciente atualmente?**
Dificuldade em integrar dados de diferentes unidades hospitalares e acompanhar métricas em tempo real.
2. **Como é feita atualmente a coleta de dados clínicos e administrativos?**
Por meio de sistemas independentes, com pouca padronização entre as áreas.
3. **Há interesse em prever readmissões hospitalares?**
Sim, especialmente em pacientes com doenças crônicas.
4. **Quais indicadores são considerados prioritários?**
Tempo médio de internação, taxa de readmissão e nível de satisfação dos pacientes.
5. **A instituição já utiliza alguma ferramenta de Business Intelligence?**
Usamos relatórios no Power BI, mas ainda sem integração total com os dados clínicos.
6. **Existem restrições de privacidade específicas além da LGPD?**
Sim, seguimos políticas internas de segurança da informação e anonimização total de dados sensíveis.
7. **Os dados dos dispositivos de monitoramento são integrados em tempo real?**
Não ainda, mas há planos para integrar via API até o próximo semestre.
8. **Como é feita a validação dos dados inseridos pelos profissionais de saúde?**
Através de revisões manuais e auditorias automatizadas em amostras de registros.

9. Qual o principal objetivo da análise preditiva no curto prazo?

Reducir o tempo médio de internação e melhorar o fluxo de pacientes.

10. Há planos de treinamento interno sobre análise de dados?

Sim, a empresa pretende capacitar equipes médicas e administrativas em interpretação de dashboards.

Projeto

Coleta de dados

Serão utilizados datasets simulados de saúde disponíveis em plataformas como **Kaggle** e **GitHub**, complementados por arquivos **CSV** anonimizados com dados realistas de pacientes, sinais vitais e satisfação.

Limpeza e pré-processamento

As etapas incluem:

- Remoção de duplicatas e inconsistências;
- Tratamento de valores ausentes com médias ou interpolação;
- Padronização de variáveis como idade, gênero e níveis de pressão arterial;
- Conversão de datas e horários para formatos uniformes.

Análise exploratória (EDA)

Serão criadas visualizações como:

- **Histogramas** de distribuição de idade e tempo de internação;
- **Heatmaps** de correlação entre variáveis clínicas e satisfação;
- **Boxplots** para detecção de outliers;
- **Gráficos de barras e linhas** mostrando tendências de readmissão e taxa de ocupação.

Modelagem preditiva

Aplicação de algoritmos supervisionados, como:

- **Regressão Logística** para prever risco de readmissão;
- **Árvore de Decisão** para identificar fatores que influenciam a satisfação do paciente;
- **Random Forest** para aumentar a precisão preditiva.

O desempenho será avaliado por métricas como **acurácia**, **precisão**, **recall** e **AUC-ROC**.

Visualização interativa

Os resultados serão apresentados em painéis interativos desenvolvidos com:

- **Dashboard em Site:** Dashboard intuitivo e simples em formato de site, usando o Tailwind CSS como ferramenta de design e Javascript para leitura de dados já organizados e treinados.

Essas visualizações facilitarão a interpretação e a tomada de decisão pela equipe gestora.