

CENTRO UNIVERSITÁRIO UNIFECAP
Engenharia da Computação

Gabriela Camarço de Sousa
RA: 14476

**Melhoria do Atendimento ao Paciente através da
Análise de Dados na HealthCare Solutions**

Trabalho apresentado como requisito parcial de avaliação da disciplina **Eletiva: Data Science-EAD** do Curso de Graduação em **Engenharia da Computação** do Centro Universitário UniFECAF.

Tutor(a): **Fernando Leonid.**

Taboão da Serra, SP

2025

Melhoria do Atendimento ao Paciente através da Análise de Dados na HealthCare Solutions

1. Cenário-Problema

A HealthCare Solutions, uma rede hospitalar, enfrenta o desafio de transformar um grande volume de dados dispersos (EHRs, dispositivos, pesquisas) em informações estratégicas. O objetivo deste projeto é utilizar a Ciência de Dados para desenvolver um modelo preditivo capaz de avaliar o risco de readmissão hospitalar em 30 dias. A análise visa identificar os fatores operacionais e clínicos que impulsionam esse risco, permitindo à gestão hospitalar realizar intervenções preventivas e otimizar a qualidade do atendimento.

2. Fundamentos da Ciência de Dados

O projeto seguiu o ciclo de vida completo de um projeto de Data Science:

Coleta e Limpeza: Foi criado um dataset simulado (dataset_healthcare_solutions_v2.csv) com 150 registros, utilizando a biblioteca Faker para garantir a anonimização de dados pessoais como nomes e cidades. O pré-processamento inicial foi realizado para preparar os dados para a análise.

Análise Exploratória (EDA): O script 2_eda_analise.py gerou diversas visualizações (Histogramas, Boxplots, Violin Plots) que permitiram as primeiras descobertas.

Modelagem Preditiva: O script 3_modelagem_preditiva.py aplicou o algoritmo Random Forest Classifier para prever a variável-alvo readmissao_30dias.

Interpretação: Foram gerados gráficos de Matriz de Confusão e de Importância das Features para extração dos insights críticos.

3. Análise de Resultados e Insights

3.1. Descobertas da Análise Exploratória (EDA)

A análise exploratória revelou padrões importantes:

Idade: A distribuição de pacientes se concentra em uma média de aproximadamente 55.6 anos, indicando uma população com risco inerente.

Custo e Readmissão: O Boxplot indica que o custo médio de atendimento é muito semelhante entre pacientes que foram e não foram readmitidos, o que sugere que o custo absoluto não é o único diferencial de risco.

Gargalos Operacionais: O Violin Plot mostra que, apesar da eficiência média, o atendimento de Urgência Alta tem alta dispersão no tempo de espera, sinalizando falhas na alocação de recursos durante picos de demanda.

3.2. Desempenho do Modelo Preditivo

O modelo de Random Forest obteve uma Acurácia Geral de 70%. Embora seja uma base sólida, o principal desafio está no Recall da classe 'Readmitido' (Sim), que foi de $\approx 50\%$. Isso significa que o modelo precisa de otimização para identificar metade dos pacientes que realmente serão readmitidos.

3.3. Fatores de Risco Críticos

Conforme o gráfico de Importância das Features, os três fatores de risco mais importantes na previsão de readmissão são:

Custo de Atendimento: O fator mais importante (com $\approx 16\%$ de peso).

Tempo de Espera (min): O segundo fator mais importante (com $\approx 15\%$ de peso).

Idade: O terceiro fator mais importante (com $\approx 14.5\%$ de peso).

Conclusão do Insight: Para reduzir a readmissão, a HealthCare Solutions deve priorizar a padronização de tratamento de casos de alto custo e a otimização dos fluxos de triagem para diminuir o tempo de espera.

4. Aspectos Éticos e Legais (LGPD)

O tratamento de dados sensíveis na saúde exige conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD). Para atender aos princípios de privacidade e anonimização:

O projeto utilizou uma base de dados 100% simulada (criada via biblioteca Faker), garantindo que não houve coleta ou processamento de dados pessoais sensíveis reais de pacientes.

O uso do modelo é estritamente para a melhoria do cuidado e segurança do paciente.

5. Levantamento de Requisitos (10 Perguntas Simuladas)

As respostas são baseadas na análise dos dados:

- Com base na Modelagem Preditiva, qual o principal fator que deve ser atacado para reduzir a readmissão? O principal fator é o Custo de Atendimento, seguido pelo Tempo de Espera, conforme a importância das features.
- O alto Tempo de Espera afeta a satisfação? Qual a recomendação? Sim, existe uma correlação negativa entre eles. A recomendação é otimizar os fluxos de triagem para reduzir a espera, impactando positivamente a satisfação.

- Qual o perfil de paciente mais sensível? A Idade é um fator crucial, sendo necessária a criação de programas de acompanhamento pós-alta para faixas etárias de risco (idosos).
- O Custo de Atendimento é diferente para pacientes readmitidos? Não de forma clara. As medianas de custo são semelhantes entre os grupos 'Sim' e 'Não', o que sugere que a complexidade do tratamento é a chave, e não apenas o custo.
- Onde estão os gargalos no atendimento de urgência? O Violin Plot indica alta dispersão no tempo de espera da Urgência Alta, sinalizando que o sistema é vulnerável a picos de demanda.
- Qual a principal métrica a ser melhorada no modelo de ML? O Recall da classe 'Readmitido' (Sim), atualmente em 50%, para aumentar a capacidade de identificar todos os pacientes de alto risco.
- Como a HealthCare Solutions deve usar o modelo na prática? O modelo deve ser usado para sinalizar (flag) pacientes de alto risco no momento da alta, garantindo o encaminhamento para acompanhamento clínico especializado (telemonitoramento).
- Qual a vantagem de usar o Random Forest no lugar de outro modelo? O Random Forest oferece a Importância das Features, fornecendo insights claros sobre o que a gestão deve mudar.
- O uso de dispositivos de monitoramento tem alto impacto na readmissão? As variáveis de dispositivo de monitoramento estão na parte inferior da lista de Importância das Features, indicando um impacto direto baixo na previsão de readmissão.
- Que melhoria futura é sugerida para o projeto? O Deploy do modelo em uma plataforma interativa (Streamlit ou Power BI) para que a equipe de gestão possa acompanhar os pacientes de risco em tempo real.

6. Considerações Finais

O projeto demonstra que a aplicação da Ciência de Dados permite à HealthCare Solutions transformar dados dispersos em conhecimento acionável. Ao focar em fatores como Custo de Atendimento e Tempo de Espera, a instituição pode intervir diretamente nos pontos críticos identificados pelo modelo preditivo, resultando em melhoria da qualidade do cuidado e otimização dos recursos.

6. Referências

Registros Eletrônicos de Saúde (EHRs) – Soluções de Saúde. Disponível em: [link]

Dispositivos de Monitoramento – Saúde Inteligente. Disponível em:

<https://www.smarthealth.com/devices>

Pesquisas de Satisfação dos Pacientes – Soluções de Saúde. Disponível em:

<https://www.healthcaresolutions.com.br/pesquisas-satisfacao>

Dados Administrativos – <https://www.who.int/standards>

Legislação – BRASIL. Lei Geral de Proteção de Dados Pessoais (LGPD). Lei nº

13.709/2018. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L13709.htm

Estudos de Caso – SMITH, João. Melhores Práticas em Ciência de Dados na Saúde.

Revista de Ciência de Dados em Saúde, v. 2, p. 123-135, 2023. Disponível em:

<https://www.jhds.com/melhores-praticas>