

CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL

SEMANA DE IMERSÃO 02

PROJETO DE MACHINE LEARNING

Pré-processamento e regressão

Predição do Custo Médio de Internações Hospitalares no SUS

O Sistema Único de Saúde (SUS) atende milhões de brasileiros anualmente, fornecendo serviços hospitalares de forma gratuita. Os custos das internações variam de acordo com diversos fatores, como tipo de procedimento, perfil do paciente e estado da federação. A previsão do custo médio das internações hospitalares é essencial para o planejamento orçamentário e alocação eficiente dos recursos públicos.

Nesta atividade, você irá utilizar os dados do **SIH-SUS de 2024** para um estado específico e construir um modelo de regressão linear para prever o custo médio das internações. A análise ajudará a entender quais fatores mais influenciam esse custo e permitirá explorar diferentes abordagens preditivas.

INSTRUÇÕES

Escolha de um Estado

- Selecione um estado ou município relevante para análise (exemplo: **São Paulo**, devido ao alto volume de internações).

Análise Exploratória dos Dados (EDA)

- Verifique a distribuição das variáveis numéricas e categóricas.
- Analise a correlação entre as variáveis.
- Explore tendências temporais e geográficas.

Seleção de Variáveis

- Identifique as variáveis mais relevantes para prever o custo médio de internação.



INSTRUÇÕES

Modelo de Regressão Linear

- Utilize **Regressão Linear Múltipla** para prever o custo médio de internação.
- Separe os dados em **treino e teste** para avaliar o desempenho do modelo.

Avaliação e Comparação de Modelos

- Utilize métricas como **R^2 , RMSE (Erro Quadrático Médio) e MAE (Erro Absoluto Médio)**.
- Teste outros modelos como **Árvore** ou **KNN** para comparação.


Conclusões e Interpretação

- Apresente os fatores mais impactantes no custo de internação.
- Discuta possíveis intervenções para otimizar custos no SUS.



ALTERNATIVAS

Após a análise exploratória, fique à vontade para redefinir o objetivo do estudo, caso identifique padrões mais interessantes ou queira seguir por um caminho diferente. O foco principal é compreender os dados e estruturar uma abordagem preditiva relevante para a realidade analisada.



ACESSO AOS DADOS

Devido ao grande volume de informações contidas na base do **SIH-SUS**, os dados utilizados nesta análise, bem como dicionário de variáveis, foram disponibilizados através da plataforma **Kaggle**, no seguinte repositório:

[!\[\]\(4729e517bc6a7cd81c8025b9646574fb_img.jpg\) Sistema de Informações Hospitalares do SUS – Kaggle](#)

O conjunto de dados contém registros detalhados sobre internações hospitalares realizadas no **Sistema Único de Saúde (SUS)**, incluindo informações sobre procedimentos médicos, custos, tempo de permanência, desfechos clínicos e distribuição geográfica.

Para acessar os dados e utilizá-los na análise, siga os seguintes passos:

1. **Acesse o link do Kaggle** e faça o download do dataset.
2. Caso ainda não tenha, crie uma conta gratuita no Kaggle para obter permissão de acesso.
3. Utilize a biblioteca **kaggle** para baixar os arquivos diretamente no seu ambiente de trabalho, ou faça o download manualmente.



CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL



MINISTÉRIO DA
SAÚDE

