

Integrantes:

Beatriz Rocha 553455

Luis Alberto 553507

Isabelle Torricelli 552806

Rafael Nascimento, 553117

Challenge CarePlus

IA e Machine Learning

São Paulo

2025

Sumário:

Introdução

1.1 Contexto

1.2 Objetivos

Metodologia

2.1 Preparação dos dados

2.2 Variáveis

2.3 Modelo de ML

Discussão

3.1 É possível prever alta adoção de telemedicina?

3.2 Quais fatores têm maior impacto?

3.3 Como evoluiu a adoção ao longo do tempo?

Análise de desempenho dos modelos

4.1 Classificação Binária

4.2 Regressão Linear

Gráficos Gerados

5.1 Distribuição das variáveis

5.2 Heatmap de correlação

5.3 Curvas ROC

5.4 Predição x valores reais

5.5 Evolução temporal

Conclusão

Introdução

1.1 Contexto

A telemedicina emergiu como componente crucial dos sistemas de saúde durante a pandemia de COVID-19, com o Medicare expandindo significativamente a cobertura desses serviços a partir de 2020. Compreender os padrões de adoção e os fatores que influenciam a utilização é essencial para otimizar a alocação de recursos e reduzir disparidades no acesso à saúde.

1.2 Objetivos

1. Desenvolver modelos preditivos para identificar combinações de características regionais e demográficas associadas à alta adoção de telemedicina
2. Identificar os fatores demográficos e regionais com maior impacto nas taxas de adoção
3. Analisar a evolução temporal da adoção e identificar grupos demográficos com maior crescimento

Metodologia

2.1 Preparação dos Dados

O dataset inicial continha 31.304 registros com 13 variáveis. Após remoção de valores ausentes na variável alvo e filtragem de agregações totais, a amostra final para análise foi de 27.927 registros.

Variável binária de classificação: Alta_Adocao foi criada usando a mediana de Pct_Telehealth (0,1776) como threshold, resultando em distribuição balanceada (50% alta, 50% baixa adoção).

2.2 Variáveis utilizadas

| Categoria | Variáveis | Tipo |
|--------------|--|------------|
| Demográficas | Bene_Race_Desc, Bene_Sex_Desc, Bene_Age_Desc | Categórica |
| Geográficas | Bene_Geo_Desc, Bene_RUCA_Desc | Categórica |
| Temporais | Year, Quarter | Numérica |

| | | |
|----------------------|--|------------|
| Elegibilidade | Bene_Mdcr_Entlmt_Stus | Categórica |
| Volume | Total_Bene_TH_Elig, Total_PartB_Enrl, Total_Bene_Telehealth | Numérica |

2.3 Modelos de ML

- **Classificação:** Random Forest e Regressão Logística
- **Regressão:** Random Forest Regressor e Linear Regression
- **Validação:** Train-test split (80-20) com 5-fold cross-validation
- **Métricas:** Accuracy, ROC-AUC, R², RMSE, MAE

Discussão

3.1 É possível prever se uma combinação de características regionais e demográficas leva uma região a estar entre os grupos de alta adoção de telemedicina?

R:

Sim, é possível prever com boa acurácia.

Resultados dos Modelos:

- Random Forest Classifier:
 - Accuracy: 78.9%
 - ROC-AUC: 0.863
 - Cross-validation ROC-AUC: 0.859 (± 0.009)
- Logistic Regression:
 - Accuracy: 68.6%
 - ROC-AUC: 0.748
 - Cross-validation ROC-AUC: 0.740 (± 0.010)

Top 5 Features Mais Importantes:

1. Ano (Year): 30.1% - Reflete a tendência temporal de declínio pós-pandemia
2. Região Geográfica (Geo_Encoded): 16.8% - Diferenças entre estados/regiões
3. Status de Elegibilidade (Status_Encoded): 13.4% - Aged, Disabled, ESRD

4. Total de Beneficiários Elegíveis: 13.1% - Tamanho da população
5. Faixa Etária (Age_Encoded): 11.7% - Idade dos beneficiários

Conclusão: O modelo Random Forest demonstra excelente capacidade preditiva, com ROC-AUC de 0.863, indicando que características regionais e demográficas são fortemente preditivas da alta adoção de telemedicina.

3.2 Quais fatores demográficos e regionais têm maior impacto na taxa de adoção de telemedicina?

R:

Análise revelou padrões claros por diferentes dimensões:

1. Faixa Etária (Maior Impacto)

- 0-64 anos: 31.8%
- 65-74 anos: 18.1%
- 75-84 anos: 17.5%
- 85+ anos: 17.2%

Insight: Beneficiários mais jovens têm quase o dobro da taxa de adoção comparado aos mais velhos.

2. Região (RUCA)

- Áreas Urbanas: 20.1%
- Áreas Rurais: 16.9%
- Diferença: 3.2 pontos percentuais

Insight: Áreas urbanas têm maior adoção, possivelmente devido a melhor infraestrutura de internet e acesso a tecnologia.

3. Raça/Etnia

- American Indian/Alaska Native: 23.1%
- Black/African American: 22.2%
- Hispanic: 21.8%
- Asian/Pacific Islander: 19.5%
- Non-Hispanic White: 19.3%

Insight: Grupos minoritários mostram taxas ligeiramente mais altas, possivelmente devido a barreiras de acesso a cuidados presenciais.

4. Sexo

- Mulheres: 20.5%
- Homens: 17.9%
- Diferença: 2.6 pontos percentuais

Insight: Mulheres demonstram maior propensão ao uso de telemedicina.

Conclusão: A faixa etária é o fator de maior impacto, seguida por região e características demográficas. Beneficiários mais jovens em áreas urbanas têm maior probabilidade de alta adoção.

3.3 Como a taxa de adoção de telemedicina varia ao longo do tempo e quais grupos demográficos mostraram maior crescimento?

R: Padrão claro de declínio após pico inicial em 2020.

Evolução Geral (Média Nacional)

- 2020: 32.4% (pico - pandemia COVID-19)
- 2021: 23.5%
- 2022: 19.3%
- 2023: 16.2%
- 2024: 15.9%
- 2025: 14.1%

Tendência: Declínio de 18.3 pontos percentuais de 2020 a 2025, com estabilização a partir de 2023.

Evolução por Região (RUCA)

- Urban: 36.1% → 15.9%
- Rural: 28.4% → 12.3%

Insight: Áreas urbanas tiveram declínio mais acentuado, mas mantiveram taxas mais altas.

Evolução por Faixa Etária

- 0-64: 42.0% → 25.7%
- 65-74: 32.7% → 14.1%
- 75-84: 33.4% → 13.0%)

- 85+: 33.2% → 13.2%

Insight: Grupos mais jovens mantiveram taxas mais altas mesmo após o declínio.

Evolução por Raça/Etnia

- Hispanic: 38.4% → 18.3%
- American Indian/Alaska Native: 36.6% → 12.7%
- Asian/Pacific Islander: 37.3% → 18.3%
- Black/African American: 34.1% → 15.9%
- Non-Hispanic White: 32.8% → 15.7%

Conclusão: Todos os grupos demográficos mostraram declínio após 2020, mas grupos mais jovens e áreas urbanas mantiveram taxas relativamente mais altas. O padrão sugere que a telemedicina foi amplamente adotada durante a pandemia, mas seu uso diminuiu com o retorno aos cuidados presenciais.

Análise de desempenho dos modelos

4.1 Classificação Binária (Alta vs Baixa Adoção)

| Métrica | Random Forest | Logistic Regression |
|------------|----------------|---------------------|
| Accuracy | 78.9% | 68.6% |
| ROC-AUC | 0.863 | 0.748 |
| Precision | 0.79 | 0.69 |
| Recall | 0.79 | 0.69 |
| F1-Score | 0.79 | 0.69 |
| CV ROC-AUC | 0.859 (±0.009) | 0.740 (±0.010) |

Vencedor: Random Forest demonstra melhor desempenho geral.

4.2 Regressão (Predição de Pct_Telehealth)

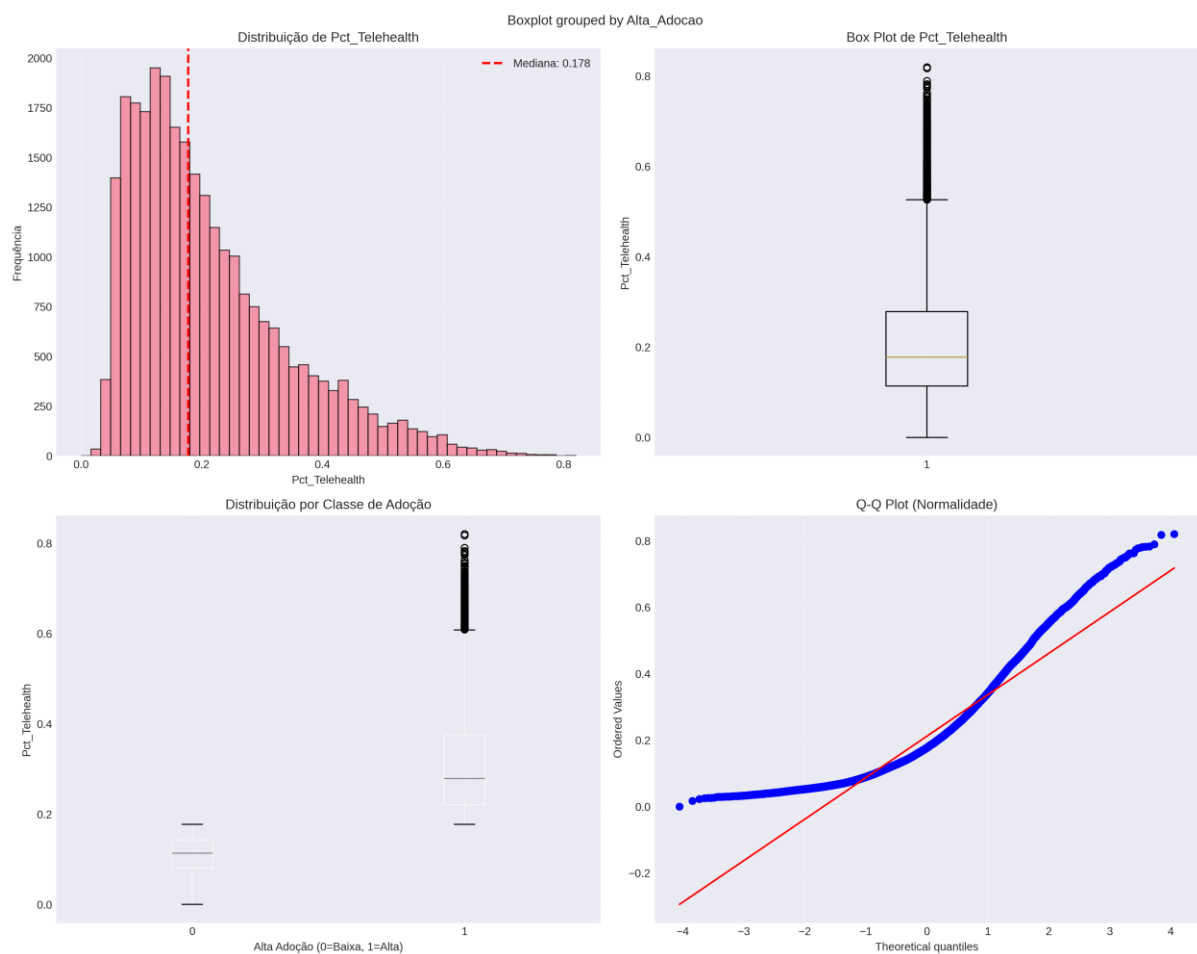
| Métrica | Random Forest | Linear Regression |
|---------|---------------|-------------------|
|---------|---------------|-------------------|

| | | |
|-------------------------|-----------------------|-----------------------|
| R² | 0.469 | 0.259 |
| RMSE | 0.095 | 0.112 |
| MAE | 0.068 | 0.086 |
| CV R² | 0.456 (±0.022) | 0.249 (±0.015) |

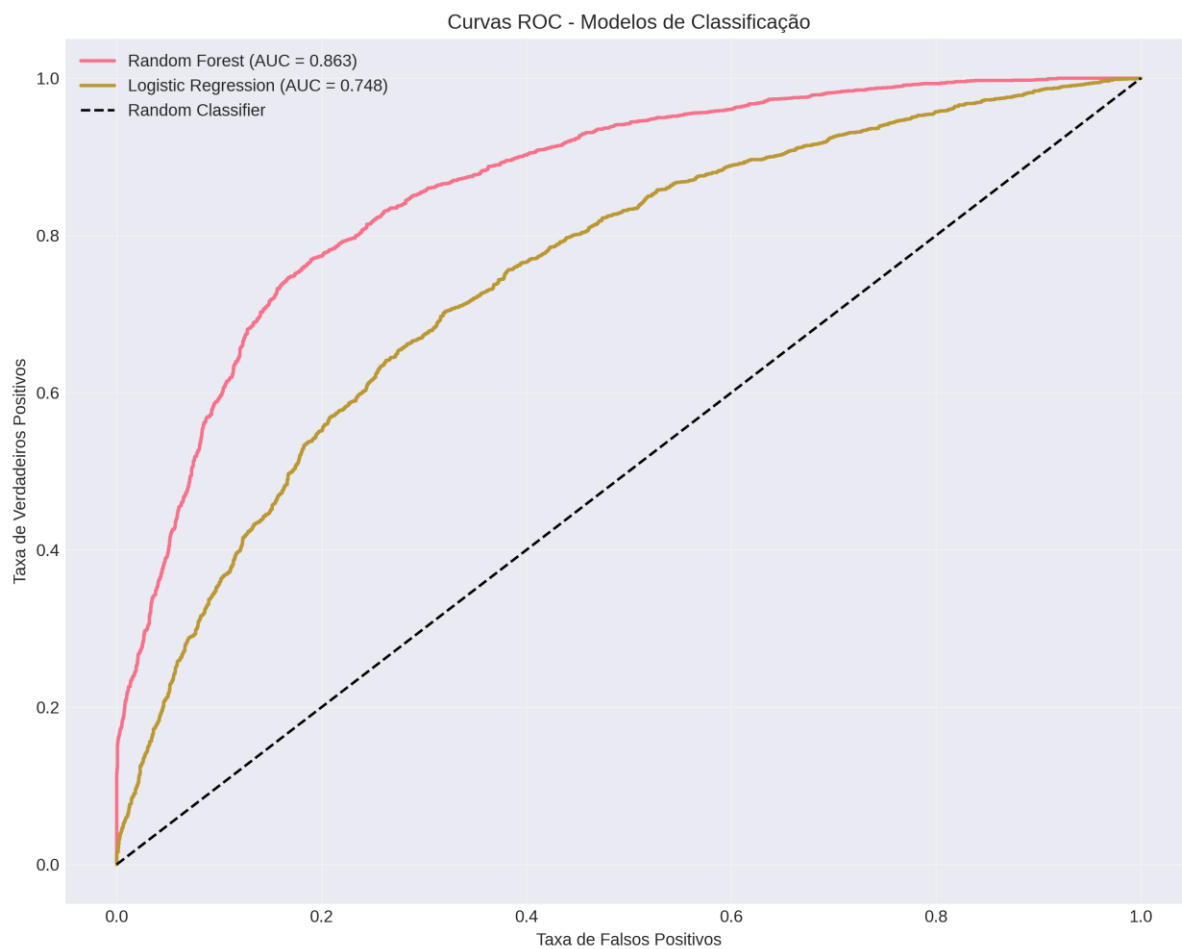
Vencedor: Random Forest explica 46.9% da variância, significativamente melhor que regressão linear.

Gráficos gerados

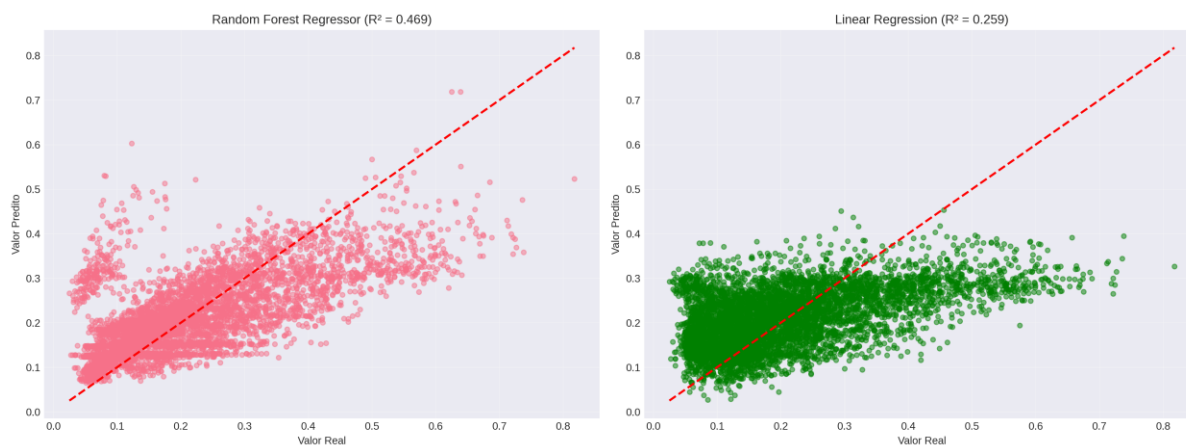
5.1 Distribuição de variáveis



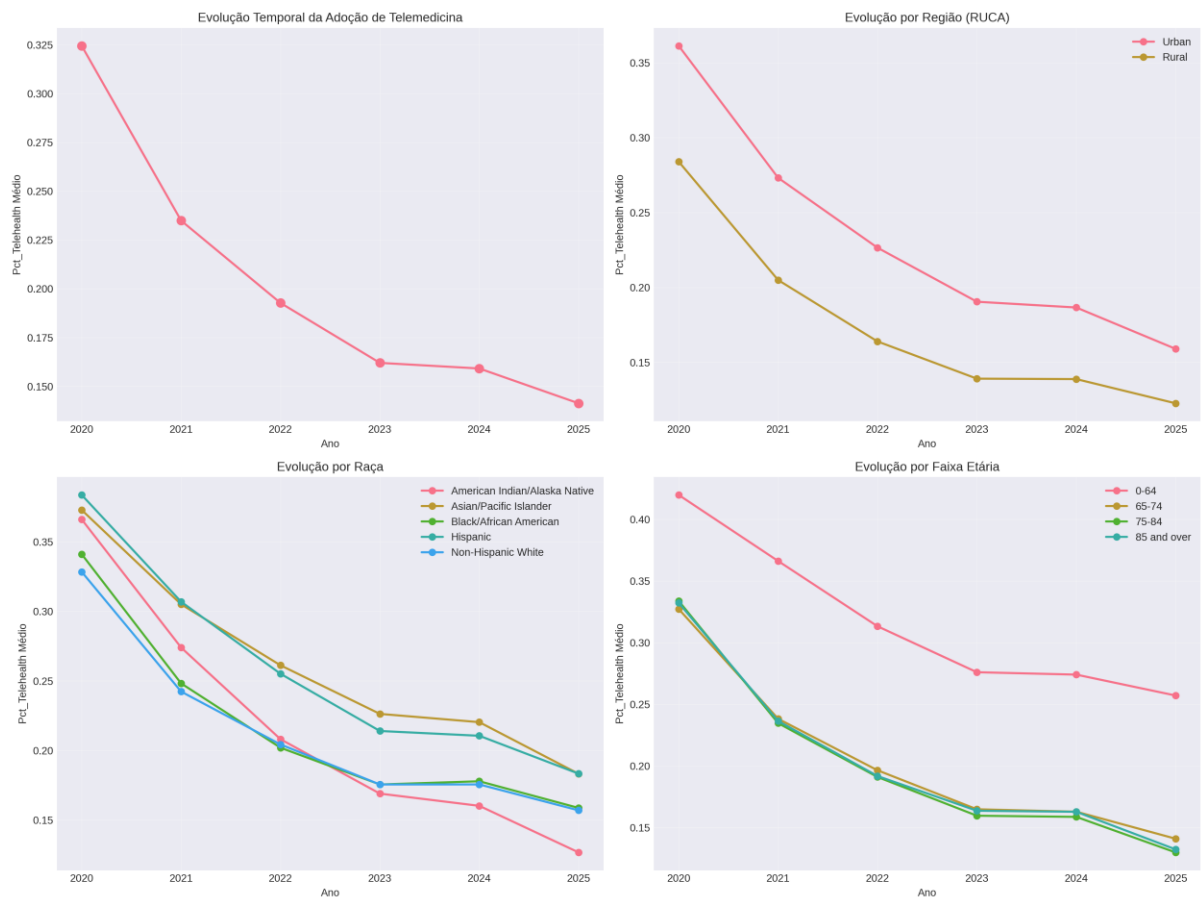
5.2 Correlação



5.4 Predição vs valores reais



5.5 Evolução Temporal



Conclusão

Os resultados obtidos respondem bem às perguntas propostas pela pesquisa, e abrem caminho para demais inovações baseadas na tecnologia. A telemedicina demonstrou ser uma ferramenta importante durante a pandemia, e entender seus pontos atraentes traçará o caminho para a implementação de novas tecnologias inovadoras dentro da área da saúde.