

Sistema de Detecção de Fraudes em Consultas Odontológicas com Inteligência Artificial

Autores:

- Arthur Fenili (RM 552752)
 - Enzo Antunes Oliveira (RM 553185)
 - Vinício Raphael (RM 553813)
-

1. Reflexão Sobre o Processo de Desenvolvimento

O desenvolvimento deste projeto foi uma jornada de aprendizado prático sobre a aplicação da Inteligência Artificial no setor da saúde. Desde o início, nosso objetivo foi entender como tecnologias de machine learning podem ajudar na detecção de fraudes — um problema recorrente em sistemas de reembolso odontológico.

Optamos por uma abordagem iterativa, começando por construir uma base sólida de conhecimento em Python, manipulação de dados com Pandas, e utilização de algoritmos clássicos de machine learning com scikit-learn. A decisão de utilizar Flask para a interface web foi estratégica: trata-se de um framework leve e eficiente, ideal para prototipagem rápida e fácil integração com modelos treinados localmente.

Durante o processo, enfrentamos desafios técnicos como o tratamento de dados ausentes, balanceamento das classes (fraudulento/não fraudulento), e avaliação adequada dos resultados do modelo. Isso nos levou a explorar não só métricas como acurácia, mas também **precision**, **recall** e **f1-score**, que são fundamentais quando há desequilíbrio entre as classes.

Refletindo sobre o processo, aprendemos que não basta apenas treinar modelos — é necessário entender a origem e a estrutura dos dados, saber interpretar os resultados e apresentar de forma clara as evidências de eficácia (ou limitação) do modelo. Também percebemos a importância de interfaces acessíveis, pois uma boa solução técnica perde valor se não puder ser utilizada de forma simples por usuários finais.

O projeto proporcionou uma visão completa de como uma aplicação de IA nasce: da concepção, passando por desenvolvimento, até chegar na apresentação e na usabilidade.

2. Descrição Final do Projeto

Trata-se de um sistema web simples que aplica um modelo de aprendizado de máquina para **identificar consultas odontológicas possivelmente fraudulentas**. O usuário pode fazer upload de um arquivo CSV contendo dados históricos de atendimentos, que são então processados por um classificador (Random Forest). O sistema devolve, na interface, os resultados da análise em tempo real.

A solução permite:

- Classificação binária de consultas (fraude ou não).
- Visualização das métricas de desempenho do modelo (acurácia, matriz de confusão e relatório de classificação).
- Upload flexível de diferentes bases, promovendo adaptabilidade.

Toda a aplicação roda localmente e depende apenas de um ambiente Python com as bibliotecas essenciais instaladas.

3. Arquitetura da Inteligência Artificial Utilizada

Modelo Escolhido: Random Forest Classifier

Escolhemos a **Random Forest** por sua capacidade de lidar bem com:

- Dados tabulares com variáveis categóricas e numéricas;
- Conjuntos com ruído e outliers;
- Problemas de overfitting, devido à sua estrutura baseada em múltiplas árvores de decisão.

Além disso, o Random Forest fornece **importância das variáveis**, o que é essencial para entender quais atributos estão mais associados à classificação de fraudes.

Fluxo da Arquitetura:

1. **Entrada de dados (CSV):** O usuário envia um conjunto de dados contendo consultas odontológicas.
 2. **Pré-processamento:**
 - Limpeza de dados (remoção de nulos, conversão de tipos);
 - Codificação de variáveis categóricas;
 - Normalização/Padronização, se necessário.
 3. **Treinamento do modelo:**
 - Os dados são divididos em treino e teste (por exemplo, 70/30).
 - O modelo Random Forest é treinado no subconjunto de treino.
 4. **Avaliação:**
 - O modelo é testado nos dados restantes e avaliado com `classification_report` e `confusion_matrix`.
 5. **Apresentação dos Resultados na Interface Flask.**
-

4. Base de Dados Utilizada

Utilizamos um conjunto simulado de dados em formato `.csv`, contendo atributos como:

- Código da consulta,
- Código do dentista,
- Data do atendimento,
- Procedimento realizado,
- Valor reembolsado,
- Frequência do paciente,
- Indicação de fraude (coluna-alvo).

Esses dados foram preparados para representar um cenário realista de atendimentos odontológicos, com a classe “fraude” presente de forma minoritária (aproximadamente 10-20% dos registros), simulando o desequilíbrio normalmente encontrado em bases reais.

Futuramente, o projeto pode ser facilmente adaptado para utilizar dados de bancos relacionais (MySQL, PostgreSQL) ou bancos NoSQL como MongoDB, o que permitirá maior escalabilidade.

5. Possíveis Expansões Futuras

- **Dashboard com Gráficos Interativos:** Utilizar Plotly ou Dash para tornar a visualização mais rica e dinâmica.
 - **Modelos mais complexos:** Como redes neurais (MLP), ou uso de algoritmos de detecção de anomalias (Isolation Forest, Autoencoders).
 - **Automação do pipeline:** Criação de rotinas para atualizar modelos periodicamente com novos dados.
 - **Integração com banco de dados:** Tornar a solução mais robusta e pronta para produção real.
-

Conclusão

Este projeto demonstra o potencial da IA para apoiar auditorias e reduzir fraudes na área da saúde. Ele combina simplicidade de uso, eficácia na modelagem e possibilidade real de expansão. O processo de desenvolvimento foi rico em aprendizados e desafiador, e nos deixou mais preparados para aplicar IA em problemas do mundo real.

Vídeo

[Link do Vídeo](#)