

OdontoVision - Inteligência Artificial

OdontoVision - Inteligência Artificial (Versão Final)

Objetivo do Projeto

Construir um sistema inteligente de detecção de fraudes em sinistros odontológicos, baseado em dados clínicos, comportamentais e históricos, utilizando aprendizado supervisionado para prever a ocorrência de fraudes com precisão e confiabilidade.

Arquitetura de IA Utilizada

O modelo escolhido foi o XGBoost (Extreme Gradient Boosting), uma arquitetura de árvores de decisão em forma de boosting. Diferente de algoritmos como Random Forest (que faz bagging), o XGBoost cria árvores sequenciais com foco na minimização do erro residual a cada iteração.

Motivos da escolha:

- Altamente eficaz para classificação com dados tabulares.
- Lida bem com dados desbalanceados (via `scale_pos_weight`).
- Capaz de capturar interações não lineares entre variáveis.
- Robusto contra overfitting com tuning mínimo.
- Permite análise de importância de features.

Melhorias de Robustez na Versão 2

- Substituição de modelos lineares ou básicos da versão anterior por XGBoost.
- Criação de pipeline completo: `ColumnTransformer` + `OneHotEncoder` + `StandardScaler` + XGBoost.
- Geração de 5.000 registros sintéticos com distribuição ajustada para enriquecer a base.
- Feature engineering com variáveis derivadas de contexto:
 - Média de custo por consulta (custo / consultas).
 - Frequência de paciente e dentista na base (histórico).
 - Categoria de idade por faixa etária (faixas discretas).
- Tratamento de dados faltantes e valores extremos (infinitos, NaNs).
- Backend com validação robusta para campos vazios e valores inconsistentes.

OdontoVision - Inteligência Artificial

- Compatibilidade garantida com scikit-learn ≥ 1.4 (model retrain completo).
- Teste e refinamento para evitar erros de serialização, ambiente e inconsistência de input.
- Integração do modelo a uma interface interativa com feedback imediato ao usuário final.

Features Novas Implementadas

- media_custo_por_consulta: Derivada do custo / número de consultas.
- freq_dentista: Quantidade de sinistros associados ao dentista.
- freq_paciente: Frequência de uso do plano pelo paciente.
- categoria_idade: Transformação da idade contínua em grupos (jovem, adulto, etc.).

Reflexão Técnica sobre a Versão Anterior

A versão anterior apresentava limitações estruturais sérias:

- Modelo simples, sem pipeline de pré-processamento.
- Baixa explicabilidade e risco de overfitting.
- Falta de verificação de campos no backend, propensa a erros.
- Dificuldade de teste devido à complexidade desnecessária do React.

A V2 resolve todos esses problemas com uma stack enxuta, validada e segura para testes reais. A IA agora tem fundamentos sólidos de ML moderno com foco em detecção de fraude de maneira confiável e interpretável.