**Documentação Disruptive architerures**

**Integrantes:**

Gabriel Pescarolli Galiza – RM 554012

Guilherme Gambarão Baptista – RM 554258

**Objetivo do trabalho**

Algo muito recorrente nos dias de hoje, são as fraudes de seguros, o que acontece, muitas pessoas contratam serviços de seguro e utilizam de má-fé, de forma a receber os ganhos, sem realmente necessitar usar. Trazendo o problema para a Odonto prev, há algumas formas de burlar um atendimento e a guia de pagamento por tal ser emitida, fazendo com que a seguradora realize o pagamento ao médico sem efetiva produção de tal. Isso pode ser feito só pela clínica de forma unitária, quanto em conjunto com o cliente, que via lá várias vezes sem necessidade, e assim dividem os ganhos.

O nosso grupo tem como objetivo diminuir essas fraudes, assim consequentemente diminuindo os gastos da seguradora, assim aumentando os lucros da empresa, e a sua segurança. Para tal objetivo temos como ideia criar um aplicativo, que será usado pela clínica do dentista e a odonto prev. Neste aplicativo iremos implementar um “formulário” obrigatório para cada atendimento feito pela clínica. A ideia é que o paciente envie fotos ou vídeos da sua situação bucal e descreva seus sintomas, estes serão analisados por um dentista ou algum estagiário na área odontológica da clínica, que então produzirá um formulário com sua análise com base nas fotos e ou vídeos e os sintomas descritos pelo paciente, este formulário deverá conter a descrição completa sobre o paciente, sintomas, tratamento sugerido, urgência, gravidade do caso, entre outros.

Esse formulário será enviado para uma análise tanto de especialistas, mas principalmente da inteligência artificial que iremos implementar, onde ela analisará o formulário, e trará uma porcentagem de “veracidade” do atendimento, mostrando a necessidade e urgência do atendimento. Importante ressaltar que todas as informações irão ao banco de dados da empresa, portanto não será somente utilizado as informações daquele formulário para a porcentagem, como também os outros, então por exemplo se a mesma pessoa vai ao dentista 10 vezes ao mês, é um comportamento suspeito que será analisado pela IA e pelos especialistas.

### **Análise da Arquitetura de IA Utilizada**

#### **Objetivo do Modelo:**

O objetivo principal do modelo desenvolvido é prever a necessidade de consulta urgente e classificar os níveis de urgência (baixa, média, alta) para pacientes de uma clínica odontológica com base em informações relacionadas à sua saúde bucal. A previsão é baseada em dados coletados de relatórios (como gravidade do caso, idade, presença de comorbidades e urgência do caso).

#### **Arquitetura Utilizada:**

A arquitetura de inteligência artificial utilizada no modelo foi baseada em **Random Forest** (Floresta Aleatória), uma técnica de aprendizado supervisionado, que foi implementada para resolver duas tarefas:

1. **Regressão:** Para prever a "necessidade urgente" do paciente, que é uma variável contínua.
2. **Classificação:** Para categorizar a necessidade de consulta em três classes: baixa, média e alta.

#### **Escolha do Modelo:**

**Random Forest - Regressão:**

Para a tarefa de regressão, foi utilizado o **Random Forest Regressor**, uma técnica robusta e poderosa que combina múltiplas árvores de decisão, produzindo um modelo mais preciso e menos propenso ao overfitting (sobreajuste) em relação a uma única árvore de decisão.

**Por que foi escolhido?**

1. O **Random Forest** é um modelo de ensemble (conjunto), que gera múltiplas árvores de decisão e agrega suas previsões, o que ajuda a aumentar a precisão e a robustez do modelo.
2. A capacidade de lidar com dados não lineares e com uma grande quantidade de variáveis de entrada, como no seu caso (idade, gravidade do caso, comorbidades), faz dele uma boa escolha para prever valores contínuos, como a necessidade urgente.

**Random Forest - Classificação:**

Para a tarefa de classificação, foi utilizado o **Random Forest Classifier**, uma variação do modelo que classifica os dados em categorias predefinidas. A classificação foi feita com base na variável "necessidade de consulta", que foi categorizada em três níveis: **baixa**, **média** e **alta**.

**Por que foi escolhido?**

1. O **Random Forest Classifier** é ideal para problemas de classificação com múltiplas classes, como o caso em que a necessidade de consulta pode ser categorizada em mais de duas classes (baixa, média e alta).
2. Assim como o modelo de regressão, ele lida bem com variáveis categóricas e numéricas, o que é vantajoso para seu caso, onde os dados incluem tanto valores numéricos (como a idade e gravidade do caso) quanto categóricos (como urgência e comorbidades).
3. Ele também oferece boas métricas de desempenho, como a matriz de confusão e o relatório de classificação, que ajudam a avaliar a precisão do modelo e identificar possíveis melhorias.

#### **Pré-processamento dos Dados:**

* **Transformação de Variáveis Categóricas:**
  + Antes de aplicar o modelo, algumas variáveis categóricas, como Urgência e Comorbidades, foram transformadas para variáveisnuméricas. O **LabelEncoder** foi utilizado para converter as categorias de urgência ("Baixa", "Média", "Alta") em valores numéricos (1, 2, 3), e a variável Comorbidades foi binarizada (0 para "Nenhuma" e 1 para outros tipos), na parte do modelo de classificação foi criada a função (def) **necessidade\_consulta** onde com base na idade da pessoa ela é organizada em 1- necessidade baixa, ou 2– necessidade média ou 3– necessidade alta.
* **Criação de Novas Variáveis:**
  + A variável **Necessidade\_Urgente** foi criada com base em uma ponderação dos fatores (idade, gravidade do caso, urgência e comorbidades). Esse processo ajudou a definir um valor numérico para a urgência do paciente, que foi utilizado como variável dependente para o modelo de regressão.
  + Posteriormente, a variável **Necessidade\_Consulta** também foi criada com base em uma ponderação dos fatores (idade, gravidade do caso, urgência e comorbidades), e foi transformada em uma classificação (1, 2 ou 3), o que foi utilizado como variável de destino para o modelo de classificação.

#### **Treinamento do Modelo:**

* **Divisão dos Dados:**
  + Os dados foram divididos em conjuntos de treinamento e teste para evitar overfitting e garantir que o modelo fosse capaz de generalizar bem para novos dados.
  + A divisão foi feita utilizando a função train\_test\_split do scikit-learn, com 80% dos dados para treinamento e 20% para teste no modelo de regressão, e 70% para treinamento e 30% para teste no modelo de classificação.

#### **Avaliação do Modelo:**

1. **Modelo de Regressão:**

A avaliação do modelo de regressão foi feita utilizando duas métricas principais:

**Erro Quadrático Médio (MSE):** Mede a média dos quadrados das diferenças entre os valores reais e previstos, indicando a precisão do modelo.

**Coeficiente de Determinação (R²):** Indica a proporção da variabilidade nos dados que é explicada pelo modelo. Quanto mais próximo de 1, melhor a explicação do modelo sobre os dados.

Uma visualização de dispersão foi criada para comparar os valores reais com os valores previstos e entender melhor o desempenho do modelo.

1. **Modelo de Classificação:**

Para o modelo de classificação, a matriz de confusão e o classification\_report foram utilizados. A matriz de confusão mostra a quantidade de previsões corretas e incorretas, enquanto o relatório de classificação apresenta métricas como **precisão**, **recall** e **F1-score**, que ajudam a entender como o modelo está classificando corretamente cada categoria.

#### **Conclusão sobre a Arquitetura:**

* O **Random Forest** foi escolhido devido à sua robustez, flexibilidade e capacidade de lidar bem com dados mistos (numéricos e categóricos). Além disso, o fato de ser um modelo não linear e baseado em múltiplas árvores de decisão torna-o particularmente útil para o problema proposto, onde há interações complexas entre as variáveis.
* A combinação de **RandomForestRegressor** para regressão e **RandomForestClassifier** para classificação permite que o modelo seja utilizado tanto para prever a necessidade urgente de um paciente quanto para classificá-lo em categorias de urgência, oferecendo uma solução completa para a análise dos dados odontológicos.