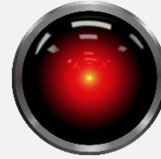


HAL9000



Turning data into revenue

Diogo Dias Rodrigues Gallina - RM98605 – 2TDSS

João Dubas Leal Kleye Souza - RM76153 – 2TDSS

Pedro Henrique Couto Archilha - RM550450 – 2TDSS

Pedro Henrique Lourenço Rodrigues - RM98402 – 2TDSS

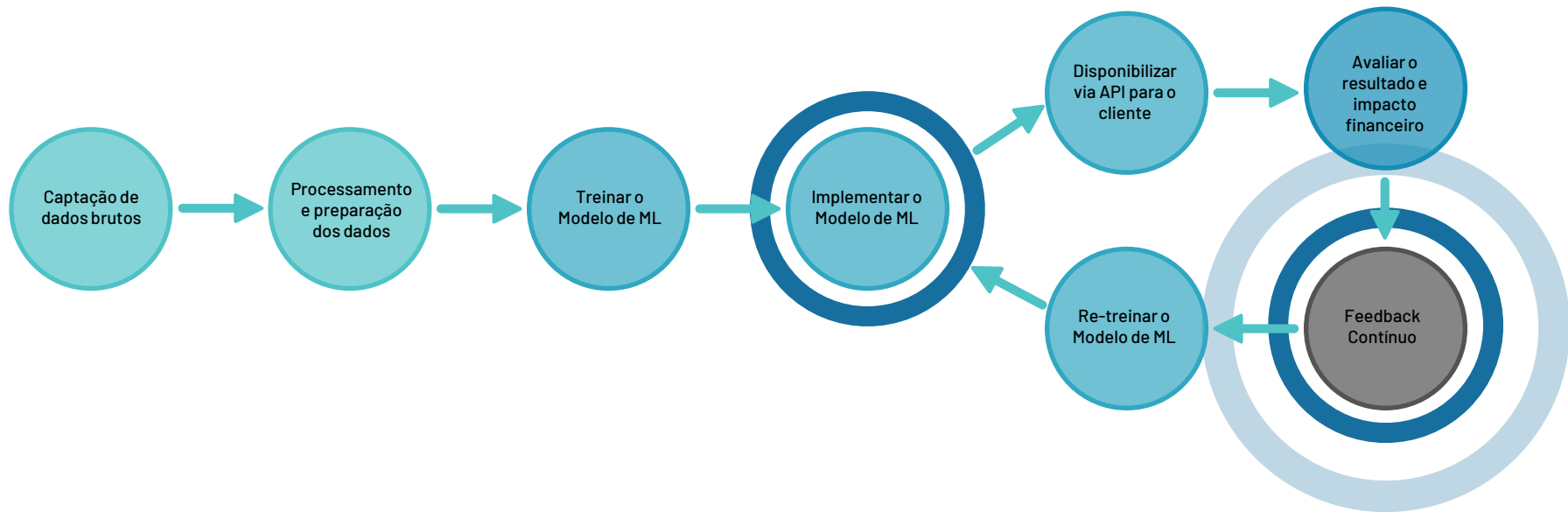
Rafael Klanfer Nunes - RM99791 – 2TDSA

Link para vídeo explicativo desta etapa da entrega:

<https://www.youtube.com/watch?v=tqabNOd-p8E>

Como gerar receita para empresas que capturam **dados** de atendimento ao cliente?

Solução: Data Science



Arquitetura da Solução

Website

Portal para o consumidor inserir reclamações e cliente avaliar via modelos

Modelo HAL 9000

Modelo treinado e implementado para avaliação de atendimentos

APIs

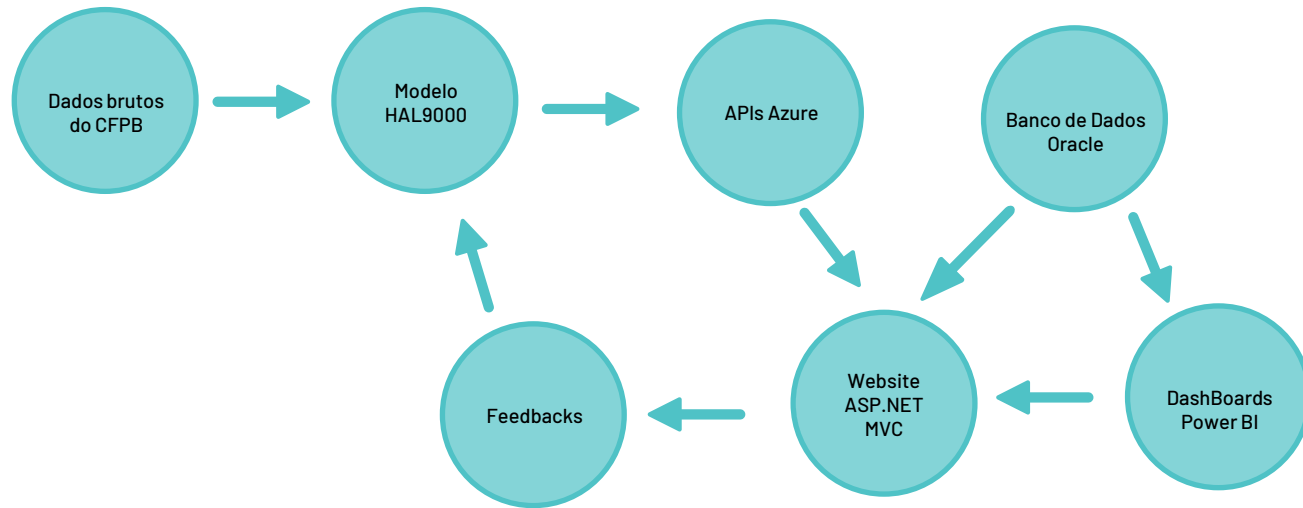
Apis que integra o Modelo HAL9000 com o portal

Power BI

Relatórios

Dashboards em Power BI para analisar o impacto financeiro da solução

Arquitetura da Solução



Arquitetura da Solução

Recurso	Descrição	Tecnologias Utilizadas
Dashboard	Relatório dinâmico para análise da eficiência do modelo em termos financeiros, seja na redução de custos ou no aumento de receitas.	<ul style="list-style-type: none">• Microsoft Power BI e gatilhos do Azure Flow
Website	Um portal onde consumidores poderão registrar reclamações, enquanto empresas poderão acionar o modelo de ML para analisar as reclamações e identificar potenciais oportunidades de vendas ou redução de custos.	<ul style="list-style-type: none">• Desenvolvido em ASP.NET MVC com .NET 8. Usando Entity Framework Core e Oracle Entity Framework Core.• Também usamos BCrypt.NET-Next para a implementação de senha seguras.
Banco de Dados	Banco de dados que suporta tanto a implementação do website, quanto o treinamento dos modelos de ML.	<ul style="list-style-type: none">• Oracle Database 19c Enterprise Edition - FIAP
APIs	Apis que integram o modelo treinado ao portal, para uso do cliente empresarial	<ul style="list-style-type: none">• Construído em python e o deploy realizado com Azure Functions. Ou então, incorporado ao projeto ASP.NET MVC do portal.
Modelo de ML HAL 9000	Modelo HAL9000, treinado após o pré-processamento dos dados brutos captados. Ele é composto por modelos de classificação de reclamações, predição de potenciais oportunidades de vendas de serviços, bem como classificação supervisionada para categorização de novas reclamações. Também será feita análise de sentimento dos textos.	<ul style="list-style-type: none">• Desenvolvido em Python 3.11.2. Utilizando os seguintes frameworks principais:<ul style="list-style-type: none">• NumPy: Para manipulação eficiente de arrays e matrizes.• Pandas: Manipulação e análise de dados estruturados.• Matplotlib e Seaborn: Para visualização de dados.• Scikit-learn / TensorFlow: Oferece ferramentas robustas para pré-processamento de dados, modelagem através de algoritmos de machine learning como K-Means, Árvore de Decisão, Floresta Aleatória, e métodos de avaliação de modelos.• Existe também a possibilidade de se construir análises a partir de APIs da OpenAI

A Escolha da Base de Dados

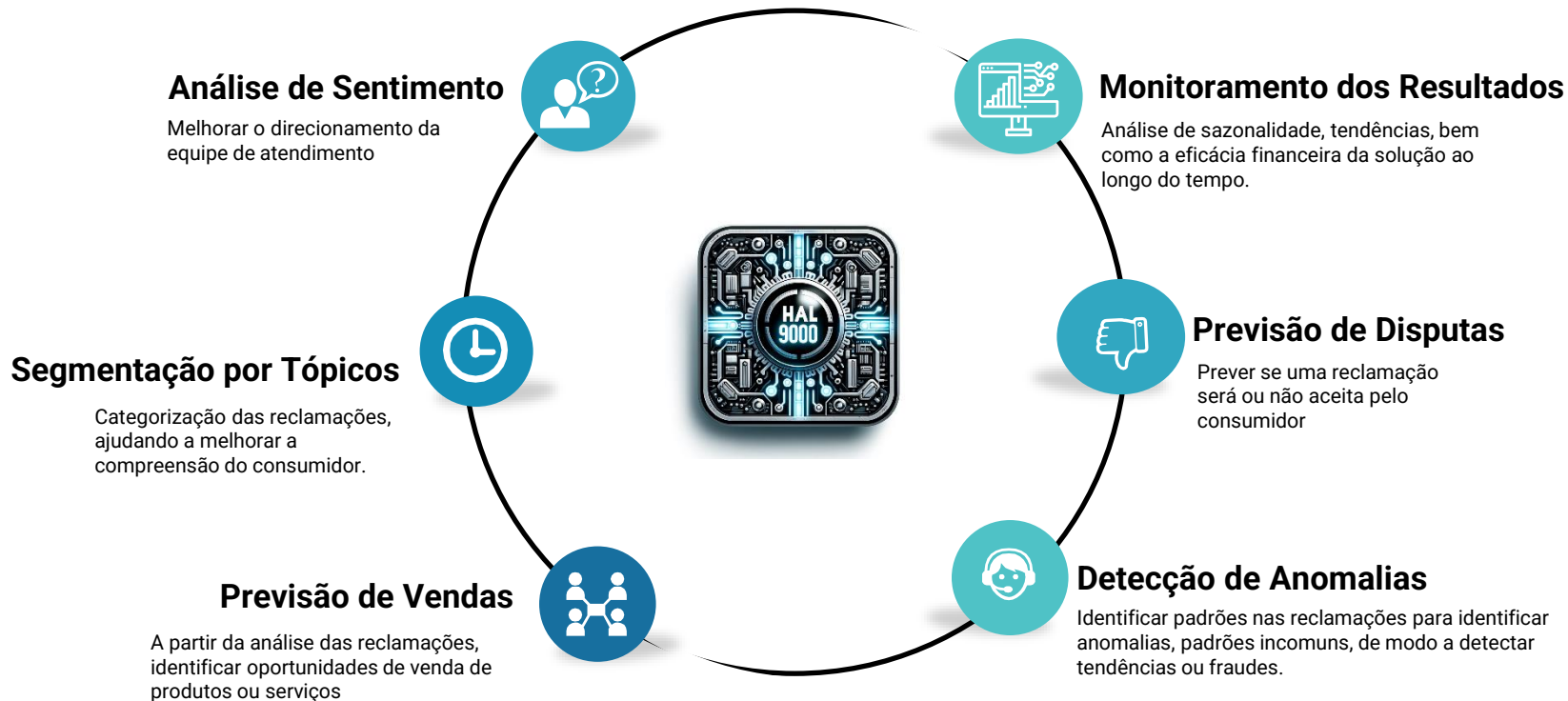
Área	Quant.	%
Demais Produtos	121.211	18,4%
Serviços Financeiros	115.281	17,5%
Telecomunicações	60.298	9,2%
Alimentos	58.948	8,9%
Demais Serviços	55.036	8,4%
Produtos de Telefonia e Informática	42.742	6,5%

Fonte: Procon-SP 2022

Consumer Financial Protection Bureau: Consumer Complaint Database

Agência independente do governo dos Estados Unidos criada em 2010. Disponibiliza cerca de 5MM de reclamações de consumidores relacionadas a práticas financeiras abusivas. Espécie de PROCON americano.

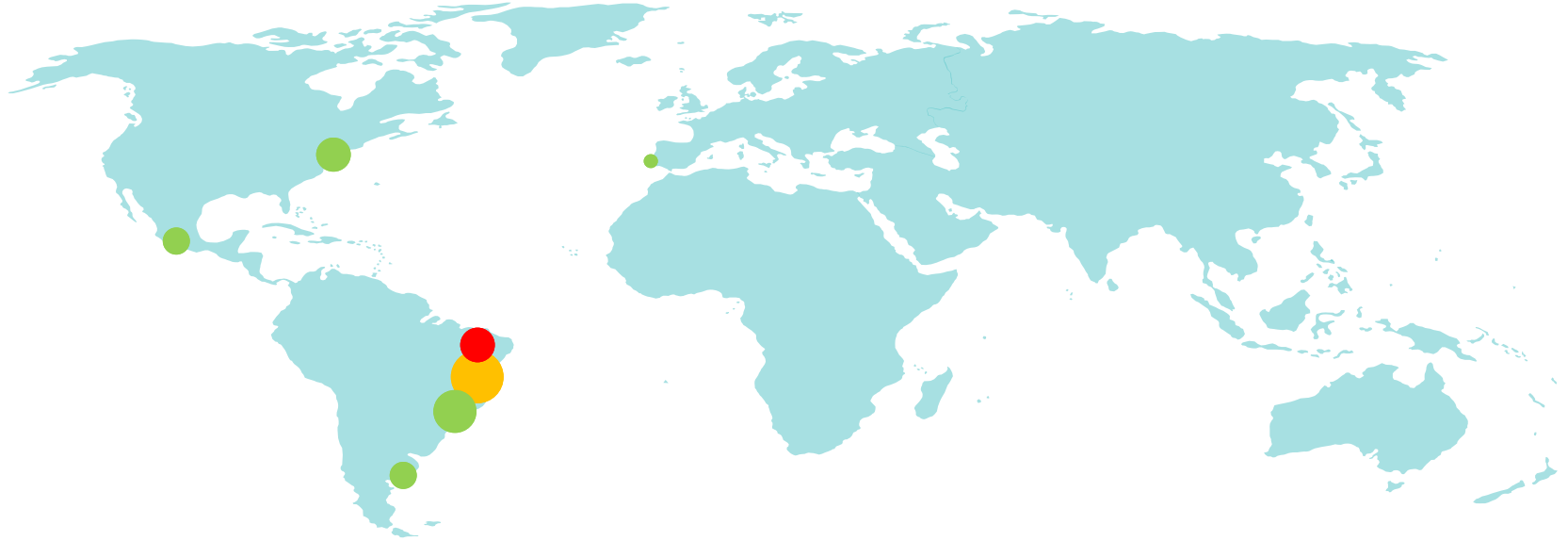
Modelo HAL9000



Detalhamento dos Modelos

	Tipo de Modelo	Descrição	Objetivo	Valor Gerado
Análise de Sentimento	Deep Learning / NLP	Este modelo aplica técnicas de processamento de linguagem natural para analisar o conteúdo textual das reclamações ("Consumer complaint narrative"), identificando sentimentos positivos, negativos ou neutros.	O resultado pode ajudar a priorizar reclamações que requerem atenção mais imediata, utilizando frameworks como NLTK ou spaCy.	Redução de custos com telemarketing. Impacto na satisfação do consumidor. Redução de litígios.
Segmentação por Tópicos	Clustering / Dendograma/ K-means	Uso de técnicas de análise de tópicos e embeddings de texto para identificar padrões e agrupar reclamações semelhantes, facilitando a categorização por conteúdo.	Ajuda a detectar tópicos recorrentes e anomalias, facilitando a compreensão das principais preocupações dos consumidores.	Melhora no entendimento do conteúdo das chamadas, permitindo uma resposta mais personalizada e eficaz.
Previsão de Vendas	Random Forest	Por meio da análise textual das reclamações, este modelo identifica expressões chave que sinalizam oportunidades de venda de produtos ou serviços adicionais, empregando algoritmos de machine learning para classificação e análise de texto.	Identificar potencial de vendas de serviços financeiros existentes ou criação de novos produtos com base nos padrões identificados.	Aumentar a receita/faturamento da empresa
Detecção de Anomalias	Lei de Benford	Implementando a lei de Benford, este modelo analisa a frequência de palavras ou caracteres para detectar anomalias e padrões incomuns nos textos das reclamações	Oferecer ferramenta adicional para identificação de tendências ou fraudes.	Redução de fraudes, bem como identificar novas tendências e padrões que geram receitas ou reduzem custos.
Previsão de Disputas	Classificação / SVM (RBF)	Este modelo utiliza "Consumer Disputed?" como variável alvo para prever se uma reclamação será contestada, baseando-se em fatores como tipo de produto, questões específicas, análise de texto detalhada e a resposta inicial da empresa, usando técnicas de classificação através de algoritmos como Regressão Logística ou Árvores de Decisão.	Prever a probabilidade de uma reclamação não ser aceita com base no texto da interação.	Redução do risco legal e melhor gestão de relações com clientes potencialmente problemáticos.
Monitoramento dos resultados	Business Intelligence	Desenvolver um dashboard em Power BI para que o cliente possa acompanhar a evolução da eficácia dos modelos e o impacto financeiro gerado após a implementação das sugestões oferecidas.	Análise de sazonalidade, padrões e resultados financeiros.	Ajuda na tomada de decisões da gerência com base nos resultados analisados.

Público Alvo



Empresas que fornecem serviços para o setor financeiro, como:

- Empresas de análise de crédito.
- Empresas de marketing.
- Empresas de consultoria.

Fintechs: Startups e empresas de tecnologia financeira.

Instituições financeiras tradicionais:

- Bancos
- Cooperativas de crédito
- Companhias de Seguro

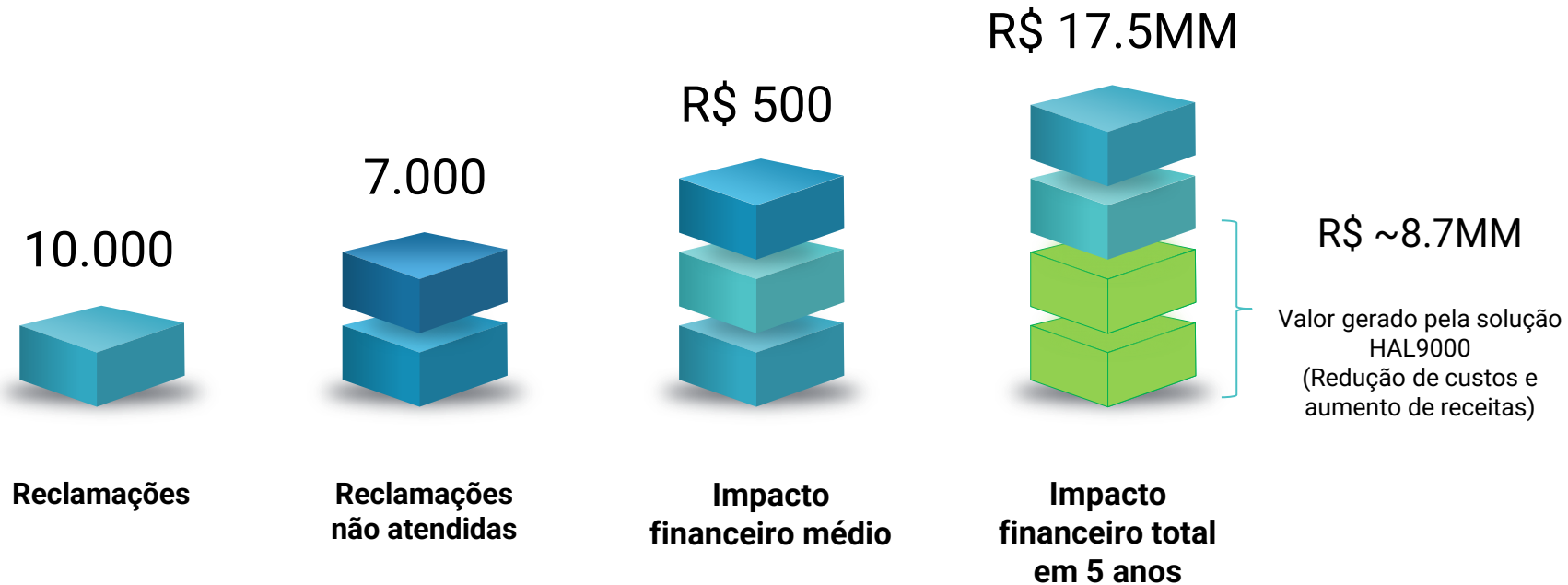
Análise Financeira

Fator	Descrição	Estimativa
Tipo de Reclamação	Natureza da reclamação influencia o impacto financeiro.	<ul style="list-style-type: none">Cobrança indevida de R\$ 100,00.Falha em serviço que causa dano de R\$ 5.000,00.
Custo de Resolução	Gastos para resolver a reclamação.	<ul style="list-style-type: none">Reembolso ao cliente: R\$ 100,00 (cobrança indevida) ou R\$ 5.000,00 (falha em serviço).Indenização por danos: R\$ 2.000,00 (falha em serviço).Custos legais: R\$ 1.000,00 (falha em serviço).Custos administrativos: R\$ 20,00 (cobrança indevida).
Perda de Receita	Diminuição da receita por cancelamento de produtos/serviços.	<ul style="list-style-type: none">Cancelamento de contrato que gera R\$ 50,00 por mês (cobrança indevida).
Dano à Reputação	Prejuízo à imagem da empresa por reclamações negativas.	<ul style="list-style-type: none">Valor indeterminado, depende da gravidade da reclamação e da cobertura da mídia.
Custos com Marketing e Publicidade	Gastos para recuperar a reputação após reclamações negativas.	<ul style="list-style-type: none">Valor indeterminado, depende do alcance da campanha e da estratégia de marketing.
Total de custos estimado	Soma dos custos diretos e indiretos.	<ul style="list-style-type: none">R\$ 170,00 (sem considerar dano à reputação) para cobrança indevida.R\$ 8.200,00 (sem considerar dano à reputação) para falha em serviço.
Potencial de receitas	Receitas geradas a partir de modelos de predição de vendas	<ul style="list-style-type: none">Valor médio da venda de produto financeiro: R\$ 300
Total estimado em termos de receita	Soma das receitas mais os custos evitados	<ul style="list-style-type: none">Entre R\$ 470 a R\$ 8.500

Serviços Financeiros

Grupo / Empresa		Total	Índice de Atendidas
1º	BRADESCO	2.460	30%
2º	ITAU UNIBANCO	2.350	31%
3º	MERCADO LIVRE	2.163	27%
4º	SANTANDER	1.916	29%
5º	PAN	1.507	24%

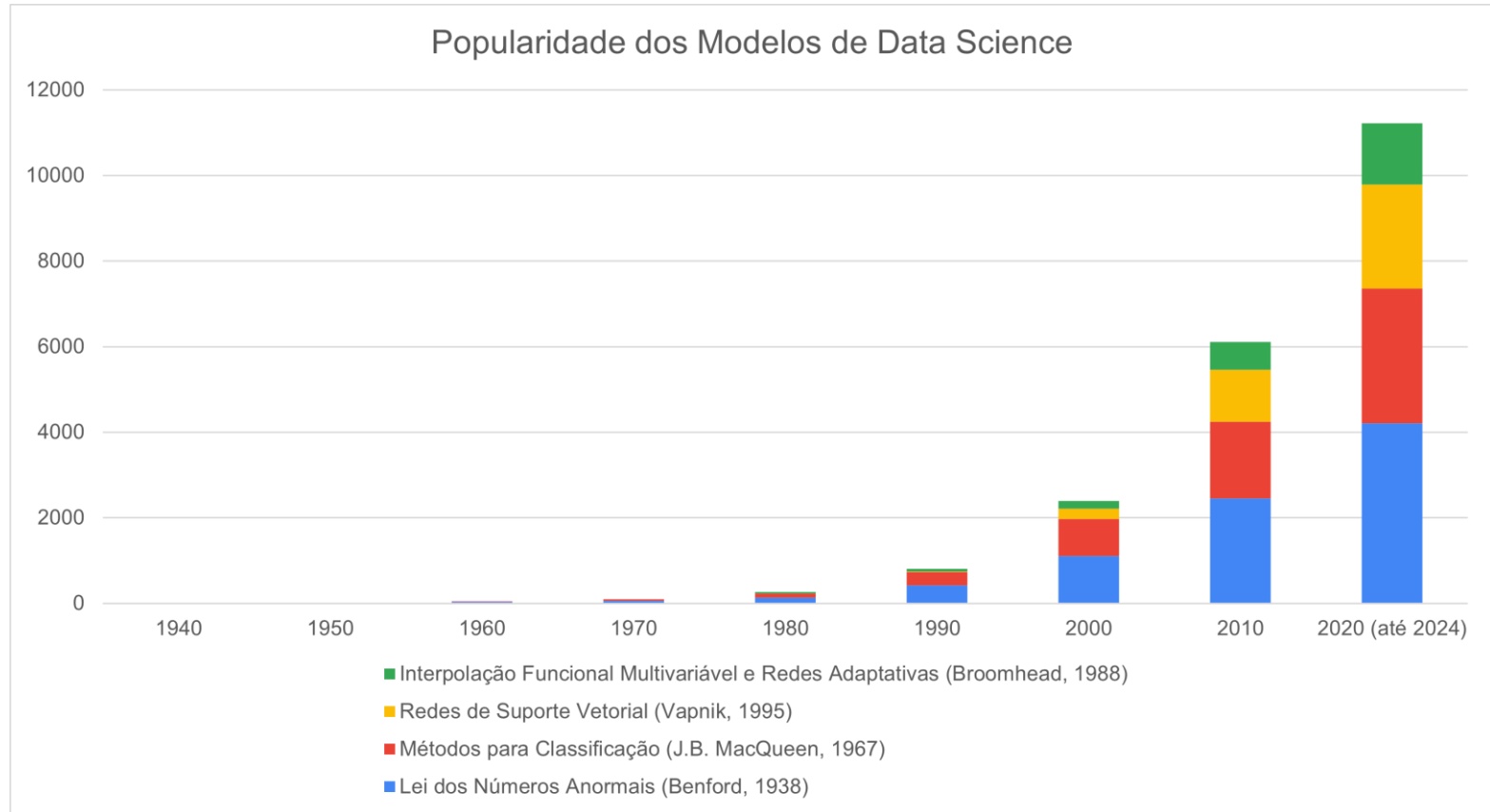
Fonte: Procon-SP 2022



* Considerando apenas top 5 empresas de acordo com o Procon-SP no ano de 2022

** Considerando uma taxa de redução das reclamações de 50%

*** Estimamos que uma reclamação não atendida tem impacto de R\$ 470 a R\$ 8.500. Levando em consideração prejuízos e/ou receitas não aproveitadas.



Etapas da Modelagem

Etapa	Descrição
Captação de dados brutos	Coleta de dados de CRM e atendimento ao cliente, incluindo interações multicanais.
Processamento e preparação dos dados	Preparação dos dados para garantir precisão, incluindo remoção de duplicatas e correções de erros.
Treinar o Modelo de ML	Criação de modelos de machine learning para prever comportamentos de clientes e encontrar oportunidades de geração de receita. Vendas, otimização de processos.
Implementar o Modelo de ML	Implementação de fato do modelo, de maneira a ser capaz de receber dados novos e gerar resultados em termos de previsão/classificação.
Disponibilizar via API para o cliente	Desenvolver plataforma (website, mobile, etc) ou API para que o cliente possa acionar o modelo e obter os resultados.
Feedback Contínuo	Incorporação de feedback para aprimoramento constante dos modelos, garantindo que eles se adaptem a mudanças e novas informações.
Avaliar o resultado e impacto financeiro	Avaliação do impacto das decisões baseadas nos modelos, focando em melhorias de eficiência e aumento de receita.
Re-treinar o Modelo de ML	Após avaliar o desempenho do modelo e comparar as previsões vis a vis aos resultados obtidos, re-treinar o modelo de modo a deixá-lo cada vez mais preciso e sempre atualizado.

Etapas do Projeto

Captação dos Dados	Processamento dos Dados	Modelo HAL 9000	Website	Dashboard
Obter dados do Consumer Financial Protection Bureau	Utilizar modelos de classificação para acrescentar mais variáveis ao dataset original	Modelo de Previsão de Disputas	Desenvolvimento de portal onde cliente final e empresa poderão acessar	Construir dashboard no Power BI para melhor entendimento dos dados
Usar ou não dados simulados para complementar a análise	Utilizar modelos de classificação não supervisionada para agrupamento das reclamações	Análise de Sentimento da Narrativa do Consumidor	Consumidor final poderá se logar, e criar reclamações para uma determinada empresa	Visão de sazonalidade das reclamações
	Sanitização dos dados	Identificação de Oportunidades de Vendas	A empresa poderá ver as reclamações	Visão do impacto financeiro das ações tomadas após o uso dos modelos
	Análise exploratória dos dados	Segmentação de Categorias	A empresa poderá acionar o modelo HAL 9000 para avaliar o texto da reclamação, obtendo previsões diversas acerca daquela interação	Análise da eficácia dos modelos
		. Análise da Distribuição de Frequência de Palavras	Criação das APIs que permitem a integração entre portal e modelo	Outras percepções e insights dos dados