

Amanda Cordeiro Staben

**Desenvolvimento de um *dashboard* sobre o
Perfil dos Clientes de uma empresa de
telecomunicações do Paraná**

Curitiba

2025

Amanda Cordeiro Staben

Desenvolvimento de um *dashboard* sobre o Perfil dos Clientes de uma empresa de telecomunicações do Paraná

Trabalho de Conclusão de Curso apresentada ao Curso de Graduação em Matemática Industrial, Setor de Exatas, da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Matemática Industrial.

Universidade do Federal do Paraná – UFPR

Bacharelado em Matemática Industrial

Orientador: Profa. Dra. Mariana Kleina

Curitiba

2025

Resumo

A área de telecomunicações no setor de tecnologia do mercado global pode ser resumida em coleta e transmissão de dados, disponíveis para inúmeros consumidores durante as décadas. Com a evolução tecnológica, vêm problemas de visualização dos grandes volumes de informações propagadas, surgindo a necessidade de aprimorar o processo de tomada de decisões com dados históricos integrados em uma interface uniforme, o *dashboard*. Para sua aplicação intuitiva, é essencial tratar a base de dados escolhida. Nesse estudo, o Processamento de Linguagem Natural (PLN) foi fundamental para a melhor classificação de uma variável muito importante para a análise sobre o perfil dos clientes da empresa estudada, o motivo de cancelamento dos contratos. A partir disso, foram avaliados modelos para a categorização da variável e criados gráficos, cartões informativos e uma visualização geográfica sobre os consumidores, com o intuito de auxiliar a tomada de decisões, otimizar tempo e possivelmente cortar gastos.

Palavras-chaves: *Dashboard*. Telecomunicações. Processamento de Linguagem Natural. Dados.

Abstract

The telecommunications area within the technology sector of the global market can be summarized as the collection and transmission of data, made available to countless consumers over the decades. With technological evolution, challenges arise in visualizing the large volumes of disseminated information, creating the need to improve the decision-making process through historical data integrated into a uniform interface, the dashboard. For its intuitive application, it is essential to process the chosen database. In this study, Natural Language Processing (NLP) was fundamental for the improved classification of a highly important variable in the analysis of the customer profile of the company under review, the reason for the cancellation of the contracts. Based on this, models were evaluated for the categorization of the variable, and charts, informational cards, and a geographic visualization of consumers were created, aiming to support decision-making, optimize time, and potentially reduce costs.

Key-words: Dashboard. Telecommunications. Natural Language Processing. Data.

Lista de ilustrações

Figura 1 – Fluxograma da Pesquisa.	16
Figura 2 – Matriz de Confusão.	19
Figura 3 – Cartão do Total de Contratos.	24
Figura 4 – Cartão da Receita Total Arrecadada.	24
Figura 5 – Cartão do Contrato Ativo.	25
Figura 6 – Cartão do Tempo para Instalação.	25
Figura 7 – Cartão do Tempo para Cancelamento.	25
Figura 8 – Gráfico da Faixa Etária.	26
Figura 9 – Gráfico do Tipo Pessoa.	26
Figura 10 – Gráfico do Status do Contrato.	27
Figura 11 – Gráfico do Motivo de Cancelamento (Especialista).	27
Figura 12 – Gráfico do Motivo de Cancelamento (PLN).	28
Figura 13 – Mapa do Paraná com o Volume de Contratos por Cidade.	28
Figura 14 – Gráfico das Cidades com mais Contratos.	29
Figura 15 – Filtros.	29
Figura 16 – <i>Dashboard</i> Completo.	30

Lista de tabelas

Tabela 1 – Matriz de Confusão Binária.	14
Tabela 2 – Classes usadas na Matriz de Confusão.	19
Tabela 3 – Resultados dos Classificadores de Performance e Métricas.	20

Sumário

1	INTRODUÇÃO	7
1.1	Objetivos	8
1.1.1	Objetivo geral	8
1.1.2	Objetivos específicos	8
2	REFERENCIAL TEÓRICO	9
2.1	História da telecomunicação e ascensão da <i>internet</i> no Brasil	9
2.2	Contexto e evolução de <i>dashboards</i> e suas características	10
2.3	O uso de PLN e a escolha da classificação <i>Zero-Shot</i>	11
2.4	Matriz de confusão	13
2.4.1	Classificadores de Performance e Métricas	14
3	METODOLOGIA	16
4	RESULTADOS	19
4.1	Matriz de Confusão	19
4.2	Criação do <i>Dashboard</i>	21
4.2.1	Total de Contratos	24
4.2.2	Receita Total Arrecadada	24
4.2.3	Contrato Ativo	24
4.2.4	Tempo para Instalação	25
4.2.5	Tempo para Cancelamento	25
4.2.6	Faixa Etária	25
4.2.7	Tipo Pessoa	26
4.2.8	Status do Contrato	26
4.2.9	Motivo de Cancelamento (Especialista) e (PLN)	27
4.2.10	Volume de Contratos por Cidade	28
4.2.11	Cidades com mais Contratos	28
4.2.12	Filtros	29
4.2.13	<i>Dashboard</i> Completo	30
5	CONCLUSÃO	31
	REFERÊNCIAS	32

1 Introdução

Com o crescimento exponencial do uso da *internet* como meio de comunicação, novas operadoras por assinatura foram surgindo, evoluindo com o tempo e ampliando cada vez mais seus clientes. A tecnologia avançada do século XXI, leva a um grande acúmulo de dados e informações importantes sobre cada cliente e seu tipo de consumo. Ainda mais, surge a necessidade de uma análise concisa e lúdica para o fácil entendimento, com o objetivo de tomada de decisões corretas e rápidas.

Ao afirmar que as empresas que não investirem em tecnologias emergentes, comprometerão sua capacidade de adaptação no cenário global das comunicações, Monteiro e Oliveira (2015) deixam claro que essas instituições se tornarão vulneráveis à obsolescência e à perda de competitividade no mundo atual. De forma complementar, Hansoti (2010) defende a otimização dos processos de tomadas de decisões, principalmente considerando o avanço das tecnologias de coleta de dados e a intensificação da personalização nas dinâmicas comerciais globais.

A fim de reestruturar e aprimorar uma grande quantidade de dados, foram surgindo estudos e métodos para refiná-los. Um exemplo indispensável são os *dashboards*, que moldam decisões e promovem abordagens baseadas em fatos, devido à visualização de dados organizados e dinâmicos, permitindo ações baseadas neles, especialmente em cargos de liderança (MATTERN, 2015). Além disso, é crescente o uso de técnicas de aprendizagem de máquina para caracterizar e reduzir a complexidade das informações analisadas (PERNA; DELGADO; FINATTO, 2010).

Este trabalho apresenta um estudo de como um tipo de processamento de linguagem natural (PLN) ajuda a estruturar melhor uma das variáveis (motivo cancelamento) da base de dados sobre clientes de uma operadora de internet no estado do Paraná. Por essa empresa ter iniciado as atividades no final de 1993, passando por várias mudanças ao longo do tempo e ainda estar presente nos dias atuais, os dados dos clientes foram se acumulando em um banco de forma extensa e disjunta.

Diante do exposto, surgiu a ideia de criar um *dashboard* com o intuito de apresentar essas informações sobre os clientes com uma visão diferente da óbvia. Fazer análises que inicialmente parecem não ter ligação direta, mas desencadeiam outras ações e comportamentos dos consumidores, criando a possibilidade de ajudar futuramente com projetos sobre como melhorar a experiência do cliente.

Para alcançar esse objetivo, ao analisar o problema e uma possível solução, o Processamento de Linguagem Natural surge como uma das ferramentas principais para auxiliar e automatizar o processo de classificação textual. Para a melhor aplicação do método,

história e teoria sobre a telecomunicação no Brasil, *dashboards* e métodos computacionais e estatísticos foram estudados.

1.1 Objetivos

1.1.1 Objetivo geral

Criar um *dashboard* sobre o perfil do cliente de uma empresa de telecomunicações do Paraná, para melhor compreensão dos consumidores e ajudar na tomada de decisões para melhor suas experiências de consumo.

1.1.2 Objetivos específicos

- Obter e pré-processar uma base de dados;
- Usar o Processamento de Linguagem Natural com o método *Zero-Shot*, utilizando um modelo de classificação para uma variável imprecisa;
- Avaliar os resultados de diferentes testes por meio da matriz de confusão;
- Criar um *dashboard* com as principais informações da base de dados, e adicionando o resultado do método do Processamento de Linguagem Natural.

2 Referencial Teórico

2.1 História da telecomunicação e ascensão da *internet* no Brasil

Com o intuito de transmitir dados para objetivos militares nos Estados Unidos, foi implantado em 1968 um sistema de informação em rede, que evoluiu e se transformou na *internet* que conhecemos hoje (ABREU, 2009). Ao decorrer dos anos 60 e 70, caracterizados por grandes avanços tecnológicos, essa nova tecnologia foi se popularizando e, em 1975, a Embratel foi encarregada de instalar e explorar uma rede nacional de transmissão de dados no Brasil (BENAKOUCHE, 1997).

Ainda de acordo com Benakouche (1997), por mais que houvesse sucesso nas ações demandadas pelo governo da época para a Embratel, que disponibilizou fácil acesso à teleinformática, como era chamada, a principal clientela era a elite do país. Para assinar os serviços oferecidos, era necessário ter ao menos uma linha telefônica e um microcomputador – considerados luxos devido ao alto custo. Além de empresas de grande e médio porte, a maioria dos usuários eram engenheiros e médicos, profissões que sempre se caracterizaram por estabilidade financeira e grande poder de compra.

Mesmo sendo uma tecnologia inovadora, as redes clássicas de telecomunicação estavam se tornando ultrapassadas para o volume intenso de informações sendo geradas. Assim, foram implementadas uma série de medidas comandadas pelo poder público no setor de telecomunicações, tentando superar o atraso nessa área em comparação com outros países e, por isso, foram instaladas novas redes destinadas exclusivamente à transmissão de dados (BENAKOUCHE, 1997). Apesar da iniciativa, mudanças na política nos anos 90 desaceleraram o desenvolvimento consideravelmente, mesmo assim, o número de usuários cresceu de forma exponencial (VARGAS; GONÇALVES, 2020).

Junto de muitas dificuldades que tiveram que ser resolvidas, a expansão massiva da utilização da *internet* foi impulsionada pela Rede Globo com a novela Explode Coração (1995), na qual os personagens principais se conhecem pela internet e acabam se apaixonando (Memória Globo, 2021) — marcando a entrada da internet na cultura popular brasileira (BENAKOUCHE, 1997). Com essa ascensão, foram criados inúmeros *softwares* para várias áreas. Um estudo baseado na pedagogia mostrou que a afinidade de crianças e jovens com o mundo digital desde cedo cria futuros consumidores. Assim, surgiu no final do século XX a tríade: informação, educação e entretenimento, que segue até hoje como os principais motivos do uso da *internet* (ABREU, 2009).

A partir desse momento decisivo para o desenvolvimento da popularidade desse espaço virtual, o custo de processamento, transmissão e armazenamento de dados caiu

muito, enquanto a quantidade de dados criados crescia exponencialmente, o que reduziu os custos em geral. Na tentativa de otimizar ainda mais o consumo e compartilhamento de dados, foi feita a implementação de cabos de fibra óptica, que são fios feitos de vidro que transmitem dados com grande velocidade, suportam grande fluxo de informações e têm baixo custo (KNIGHT, 2014).

A fibra óptica é usada no mundo todo e vem sendo uma forma mais econômica e rápida comparada com tecnologias anteriores, como o fio de cobre. Desde que chegou ao Brasil, seu principal objetivo foi melhorar a velocidade e a estabilidade das redes. Isso possibilitou novas formas de usar a *internet*, conectar pessoas e, como consequência, proporcionou o crescimento da rede em geral e barateou os serviços oferecidos (VARGAS; GONÇALVES, 2020). Como a *internet* é um instrumento aberto e de fácil acesso ao público, ela afeta diretamente os setores socioeconômicos e políticos do mundo todo (KNIGHT, 2014).

A base da indústria de telecomunicações é o desenvolvimento tecnológico. Por causa disso, as empresas estão remodelando seus comportamentos e estratégias para uma versão mais competitiva, levando ao aprimoramento das atividades nas áreas de desenvolvimento e pesquisa (GALINA; PLONSKI, 2005). “As empresas que não investirem em novas tecnologias não terão condições de sobreviver no mundo globalizado das comunicações” (MONTEIRO; OLIVEIRA, 2015).

Com isso em mente, e considerando a quantidade exorbitante de usuários da *internet*, o aprimoramento, organização e exposição de dados de forma concisa e factual vêm se tornando cada vez mais necessários. Consequentemente, foram criadas ferramentas para auxiliar essa visualização, a principal delas sendo *dashboards*.

2.2 Contexto e evolução de *dashboards* e suas características

Inicialmente, em 1846, o termo *dashboard* se referia a uma placa de couro na frente de veículos impedindo que lama fosse jogada para dentro. A partir dos anos 90, o termo passou a ser usado como conhecemos hoje, pois vem da similaridade com os painéis de controle como os de carros e aviões (velocidade, combustível, etc.). Essa origem se conecta à ideia de que, ao criar um *dashboard*, separamos os dados mais importantes, descartando os “dados sujos” — como a lama que o painel original evitava que entrasse (MATTERN, 2015).

Stephen Few é um dos principais nomes no desenvolvimento da governança de *dashboards*. Ele relata que entre os anos 80 e 90, as grandes empresas estavam mais focadas em acumular dados, sem considerar quais indicadores eram essenciais ou como analisá-los. Apenas no começo do século XXI, os executivos de cargos de liderança nessas empresas começaram a adotar *dashboards* como ferramentas estratégicas, buscando uma

visão panorâmica dos dados (MATTERN, 2015).

Hansoti (2010) alega que toda organização, pequena ou grande, deve otimizar seu processo de tomada de decisão, especialmente diante do crescimento da coleta de dados e da personalização no mercado global. Muitas vezes os dados vêm em formatos diferentes, o que gera problemas de compatibilidade. O *dashboard* resolve isso ao integrar dados de vários departamentos em uma interface uniforme, sendo ele uma estrutura com um padrão definido para suas características visuais, que disponibiliza informações no momento por meio de indicadores, sendo altamente personalizável para diferentes usuários e cenários. Para Mattern (2015), quando bem feito, um *dashboard* é uma aplicação intuitiva que representa dados qualitativos e quantitativos de forma gráfica e compreensível. Sendo assim, é um tipo de investimento valioso em tempos de crise econômica, ajudando a cortar gastos e otimizar tempo (HANSOTI, 2010).

O *display* ideal oferece uma visão macro em tempo real, combinada com tendências históricas, permitindo aos usuários entender o como e o porquê dos eventos e tomar decisões futuras com base nisso. Os desenvolvedores de *dashboards* empenham-se para tornar dados representáveis, contextualizáveis e inteligíveis, traduzindo percepção em performance, gerando *insights* que analistas usam para otimizar sistemas ou ainda para públicos com pouca familiaridade com a origem dos dados (MATTERN, 2015).

Um exemplo disso pode ser encontrado em Vasco (2021), que tem como objetivo principal: “desenvolver uma sistemática para melhorar o gerenciamento da saúde dos colaboradores e previsão de alocação de recursos financeiros na área da saúde, oferecendo um *dashboard*, que por meio de recursos visuais auxilie no processo de tomada de decisão e planos de ação do setor de recursos humanos, financeiro e saúde de uma multinacional no ramo de soluções de tecnologia”. Que traz um *dashboard* sobre a área da saúde para a otimização de tomadas de decisões de vários setores de uma empresa.

A prevalência e acessibilidade dos dados estão mudando nossa forma de ver o mundo (MATTERN, 2015). Com isso, é possível concluir que é essencial simplificar a visualização de dados, garantindo que os usuários compreendam as informações dadas e as interpretem de maneira correta. Ao decorrer do tempo e da evolução de modelos de programação e processamento de dados, foram criados métodos de classificação de dados que auxiliam os desenvolvedores de *dashboards* a refiná-los para o enriquecimento dos painéis criados. Um desses métodos, sendo ele o Processamento de Linguagem Natural (PLN), foi usado no desenvolvimento do *dashboard* apresentado em seguida.

2.3 O uso de PLN e a escolha da classificação *Zero-Shot*

As primeiras instâncias do uso de Processamento de Linguagem Natural (PLN) são notadas na década de 40, durante a Segunda Guerra Mundial, com tentativas de tradução

automática para quebra de códigos. Como não obtiveram resultados bons, pesquisadores começaram a criar teorias mais robustas (JOSEPH et al., 2016) e nos anos seguintes das tentativas iniciais, a pesquisa sobre PLN foi se multiplicando, assim estabelecendo diversas subáreas que são estudadas até hoje. Na virada do século, houve a incorporação de aprendizado de máquinas, e mais adiante, especificamente a mineração de dados (PERNA; DELGADO; FINATTO, 2010).

A essência de PLN vem da intenção de caracterizar e definir o que é chamado de Expressões Multipalavras (EMs): construções linguísticas com mais de uma palavra para expressar com precisão ideias e conceitos complexos, como uma situação ou ação que não pode ser descrita apenas com palavras simples. Logo, surgiu a necessidade do processamento de maneira uniforme, tentando minimizar a ambiguidade e facilitar a transmissão de conhecimento, reduzindo a complexidade de tratamento dos dados (PERNA; DELGADO; FINATTO, 2010).

De forma computacional, o PLN analisa e sintetiza linguagens faladas ou escritas, desenvolvendo ferramentas e técnicas apropriadas, nas quais fazem o sistema de computador entender e manipular linguagens naturais para performar várias ações úteis. O uso da palavra “natural” vem para distinguir linguagens humanas de notações matemáticas ou de programação, onde o vocabulário é mais restrito. Portanto, são construídos sistemas bem direcionados para fins específicos, como extração e normalização de informações com intuito de construir bases de dados (JOSEPH et al., 2016).

Dando a uma máquina a capacidade de aprender com poucas informações e se ajustando a novas tarefas rapidamente, o PLN generaliza dados, alavancando modelos de linguagem pré-treinada. Por usar modelos de alta escala treinados em volumes enormes de dados de texto, esses ajudam imensamente em capturar padrões de linguística e conhecimento, deixando-os generalizar classes não vistas. Assim, aumentando as suas capacidades de reconhecimento além dos dados de treinamento, rapidamente se adapta e performa muito bem com poucos exemplos, que são usados para extrair informações úteis e ajudar a máquina a fazer decisões (RAMESH et al., 2025).

Ao introduzir o aprendizado de máquina nos estudos de PLN, foram criados vários métodos de caracterização e identificação de dados como textos, imagens, áudios, etc. Para o problema encontrado na base de dados utilizada para desenvolver o *dashboard* deste estudo, o escolhido foi o *Zero-Shot Text Classification*, de acordo com Ramesh et al. (2025), é um método com bom desempenho especialmente para situações nas quais não é prático nem economicamente viável obter um grande volume de dados rotulados. O *Zero-Shot* é uma inovação e avanço em *machine learning* do começo de 2010, que hoje nutre uma nova era com Inteligência Artificial (IA), onde as máquinas não só aprendem com os dados, mas também possuem a engenhosidade para abordar o desconhecido com confiança e adaptabilidade (RAMESH et al., 2025).

Esse método classifica uma *string* de texto não vista durante o treinamento usando incorporação semântica, considerada um tipo de *embedding* (BARNARD, 2025), que é uma representação vetorial de palavras e conceitos em um espaço contínuo, no qual ideias semanticamente relacionadas tendem a se agrupar neste espaço. Conseguir que o modelo categorize classes que ele não encontrou durante o treinamento é o desafio do aprendizado *Zero-Shot*. Ao coletar ligações semânticas, elas ajudam o conhecimento do modelo a ser aplicado de maneira mais ampla a classes das quais ele não tinha conhecimento anteriormente (RAMESH et al., 2025).

Em PLN, isso tipicamente envolve o refinamento de modelos pré-treinados como o BERT, que prevê a próxima palavra de uma frase ou uma palavra baseada em outras ao redor e também se um par de frases são sequenciais. Sua base é do tipo *Transformer*, que usa um mecanismo que pesa a importância de palavras em dados de entrada e processa dependências de longo alcance no texto, sendo muito efetivo para várias tarefas do PLN. O desenvolvimento do BERT envolve pré-treinamento e refinamento, ao ser treinado em um grande corpo de texto, posteriormente pode ser refinado para uma tarefa específica com uma quantidade significativamente menor de dados (GHAFARI; GAZIT, 2024).

O modelo escolhido (*mDeBERTa-v3-base*), usado como molde, é uma das inúmeras ramificações do BERT e foi encontrado na plataforma *Hugging Face*, onde a comunidade de aprendizagem de máquina colabora em modelos, conjuntos de dados e aplicações (Hugging Face, 2024). Para atender com melhor utilidade, o modelo selecionado é poliglota e atende 100 línguas diferentes, sua base foi treinada pela *Microsoft* e refinada em datasets contendo mais de 2,7 milhões de pares de hipóteses e premissas em 27 línguas faladas por mais de 4 bilhões de pessoas. Além disso, foi o modelo de *transformer* poliglota de tamanho base de melhor desempenho apresentado pela *Microsoft* em 2021. Como foi treinado principalmente em inglês, ao usar outras línguas sua performance pode diminuir, mas em geral, é muito bem condicionado, apresentando bons resultados baseados na vasta coleção de informações dos treinamentos (LAURER, 2024).

2.4 Matriz de confusão

Uma matriz de confusão (ou matriz de erro) ajuda a avaliar o desempenho do modelo de classificação no aprendizado de máquina, comparando os valores previstos com os valores reais de um conjunto de dados usando uma tabela. Essa estrutura de grade é uma ferramenta conveniente para visualizar a precisão da classificação do modelo, exibindo o número de previsões corretas e incorretas de todas as classes, uma ao lado da outra (MUREL; KAVLAKOGLU, 2024).

Seu *layout* consiste em linhas que representam os valores previstos de uma determinada classe e colunas que representam os valores reais de uma determinada classe (ou as

linhas como os valores reais e as colunas com os valores previstos). Para um classificador binário temos: A caixa superior esquerda informa o número de verdadeiros positivos (V_P), sendo o número de previsões corretas da classe positiva. A caixa abaixo apresenta o número de falsos negativos (F_N), as instâncias positivas reais previstas incorretamente como negativas. A caixa superior direita apresenta falsos positivos (F_P), são instâncias de classe negativa identificadas incorretamente como casos positivos. Por fim, a caixa inferior direita exibe o número de verdadeiros negativos (V_N), ou as instâncias de classe negativa reais previstas corretamente (MUREL; KAVLAKOGLU, 2024).

Tabela 1 – Matriz de Confusão Binária.

		Esperado	
		Positivo	Negativo
Previsto	Positivo	V_P	F_P
	Negativo	F_N	V_N

Essa matriz também pode ser usada para problemas de classificações multiclases, nas quais todas as caixas diagonais indicam verdadeiros positivos previstos. As outras caixas informam quantidades de falsos positivos, falsos negativos e verdadeiros negativos, dependendo da classe escolhida para prever. Para análises mais profundas, são usadas métricas de precisão para o modelo (MUREL; KAVLAKOGLU, 2024).

Para o caso estudado (multiclases), são calculados os Classificadores de Performance para as classes individualmente e, para avaliar o modelo, é feita a média aritmética de cada. Analogamente para a métrica escolhida. De acordo com Sathyanarayanan e Tantri (2024) e Grandini, Bagli e Visani (2020) temos, então:

2.4.1 Classificadores de Performance e Métricas

A escolha dos classificadores de performance e/ou métricas depende do problema de interesse e dos objetivos do estudo, para o modelo estudado, foram selecionadas as seguintes.

- **Acurácia:** Mede com qual frequência uma previsão está correta, uma medida geral de quão corretamente o modelo está prevendo em todo o conjunto de dados.

$$A = \frac{V_P + V_N}{V_P + F_P + V_N + F_N} . \quad (2.1)$$

- **Precisão:** É a fração dos resultados positivos corretamente previstos. É útil quando os F_P passam dos F_N . Um alto valor de precisão sugere que o modelo faz previsões positivas com acuracidade. De forma contrária, um baixo valor de precisão indica que o modelo faz mais previsões positivas incorretas.

$$P = \frac{V_P}{V_P + F_P} . \quad (2.2)$$

- **Sensibilidade:** Mede a proporção de casos positivos reais previstos corretamente, ou a precisão com que o modelo prevê casos positivos. É interessante usá-lo quando os falsos negativos precisam ser minimizados. Um alto valor de sensibilidade implica que o modelo captura efetivamente a maioria das instâncias positivas. Por outro lado, um baixo valor indica que o modelo está com instâncias positivas ausentes.

$$S = \frac{V_P}{V_P + F_N} . \quad (2.3)$$

- **Taxa de Falsos Positivos (TFP):** Identifica a proporção de negativos classificados como positivos. O ideal é zero ou o mais próximo possível, quanto menor a taxa, menor as instâncias de negativos classificados de forma incorreta como positivos.

$$TFP = \frac{F_P}{F_P + V_N} . \quad (2.4)$$

- **Taxa de Classificação Incorreta (TCI):** Também chamada de “taxa de erro”, mostra o quão incorreto o modelo está. O ideal é zero ou o mais próximo possível, minimizando os erros cometidos pelo modelo.

$$TCI = \frac{F_P + F_N}{V_P + F_P + V_N + F_N} . \quad (2.5)$$

- **F1 Score:** Pode ser interpretada como uma média harmônica entre Precisão e Sensibilidade, em que o F1 *score* atinge seu melhor valor em 1 e seu pior em 0.

$$F1S = \frac{2 \times (\text{precisão} \times \text{sensibilidade})}{(\text{precisão} + \text{sensibilidade})} . \quad (2.6)$$

3 Metodologia

Para melhor descrever a metodologia empregada neste trabalho, foi elaborada a Figura 1, que representa o fluxograma da pesquisa.

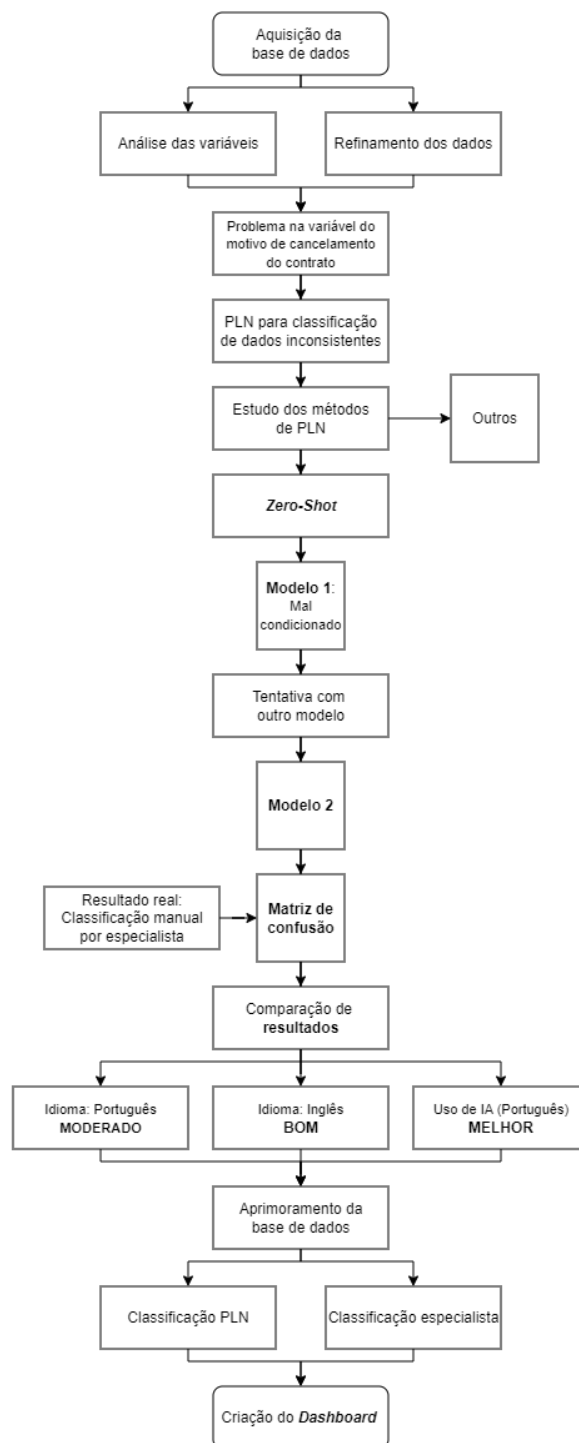


Figura 1 – Fluxograma da Pesquisa.

Fonte: A autora

Com a intenção de criar um *dashboard* sobre o perfil dos clientes de uma empresa de telecomunicações, foi feita uma extração de dados em meados de outubro de 2024. Originalmente, havia dados sensíveis que não foram incluídos, então a primeira versão apresentava apenas informações, confirmadas pela empresa, que poderiam ser usadas para a análise sem que houvesse problemas de confidencialidade. Ao examinar essa base de dados, foi feita uma análise das variáveis e um refinamento dos dados julgados dispensáveis para o objetivo apresentado, como a exclusão de outros estados que tinham clientes, sendo escolhido apenas o Paraná. A base de dados representa as informações dos clientes a partir do primeiro, em 1993, até a última extração feita. No total são 17 variáveis com 735203 linhas, cada uma representando um contrato diferente, as quais serão detalhadas posteriormente.

A variável que se destacou pela maior necessidade de torná-la mais organizada e menos complexa foi o *motivo cancelamento* (motivo pelo qual foi cancelado o contrato com a companhia), que apresentava mais de 90 condições para o cancelamento, sendo grande parte conceitos repetidos e/ou muito similares. Sendo uma empresa que começou no início dos anos 90, a última extração foi no final de 2024, e estando presente em quase todas as cidades do estado do Paraná, conclui-se que muitas informações foram coletadas. Assim, esses motivos de cancelamento foram concatenados e não substituídos a cada mudança de sistema e estratégia de negócio, gerando um acúmulo supérfluo de dados e um empecilho para uma visualização coerente do *dashboard*.

Tendo em conta a necessidade de clarificar essa variável, foi proposta a ideia de usar o PLN para classificá-la. Ao fazer esse estudo, foram vistas múltiplas subcategorias do Processamento de Linguagem Natural, mas a que mais se adequou foi a classificação *Zero-Shot*, que não precisava de treinamento prévio e tinha modelos facilmente disponíveis *on-line*. Utilizando a plataforma Hugging Face (2024), foi selecionado e testado um primeiro modelo (*facebook/bart-large-mnli*), que, ao modificá-lo com as classes criadas, mostrou-se com resultados instáveis, logo, o modelo 1 era mal condicionado e, por esse motivo, foi descartado. Em seguida, ainda na mesma plataforma, foi selecionado um segundo modelo para teste (*MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7*), no qual tinha uma base de treinamento e autor diferentes, aparentando ser melhor treinada e com resultados mais satisfatórios.

O método principal para verificar se o modelo está classificando corretamente as variáveis é a matriz de confusão, que mostra as instâncias previstas correta e incorretamente. Mas, para poder utilizá-la, primeiramente deve existir uma “classificação real”, uma avaliação inicial feita por um especialista. Foram criados 9 motivos para o cancelamento dos contratos, com o intuito de simplificar as respostas, agindo como novas classificações. Essa distribuição de classes foi feita manualmente pelo especialista, deixando-as com os valores reais para comparar com os valores previstos do modelo. Assim que foi feita a

análise, ela se tornou o valor esperado, e a classificação do modelo 2 se tornou o valor previsto.

Inicialmente, os resultados analisados iriam ser apenas na língua portuguesa, mas, pelo treinamento do modelo ser em inglês, também foi testado na língua inglesa. Além disso, foi feita uma classificação usando uma IA (*Copilot*), que utiliza o mesmo tipo de método, *Zero-Shot*. Tendo as três previsões, foram construídas as matrizes de confusão das classificações em português, inglês e usando IA, apresentando resultados considerados moderado, bom e melhor, respectivamente.

Após classificar essa variável e condensar os motivos de cancelamento usando o método *Zero-Shot*, foi possível aprimorar a base de dados, criando duas novas variáveis com os agrupamentos feitos pelo especialista e pelo PLN, respectivamente. Diante disso, o *dashboard* foi desenvolvido de maneira mais objetiva e visualmente agradável.

4 Resultados

4.1 Matriz de Confusão

Para visualização mais simples, as classes usadas serão designadas de acordo com a definição presente na Tabela 2. Essas classes foram definidas pelo especialista e representam, no contexto geral, os principais motivos de cancelamento de contrato pelos clientes.

Número	Classe
1	Tempo
2	Custo
3	Desafios Locacionais
4	Impedimento do cliente
5	Processo de Transição
6	Desafios Operacionais
7	Concorrência
8	Insatisfação
9	Temporário

Tabela 2 – Classes usadas na Matriz de Confusão.

A partir dessas classes, foi usado o modelo com a classificação em inglês e, com seus resultados, foi gerada a matriz de confusão, apresentada na Figura 2.

		ESPERADO								
		1	2	3	4	5	6	7	8	9
PREVISÃO	1	3	0	0	0	0	0	0	0	0
	2	0	5	0	0	0	0	0	1	0
	3	0	0	3	2	1	0	0	1	0
	4	0	0	0	5	0	1	0	1	1
	5	2	0	0	1	12	1	1	5	1
	6	1	0	3	0	0	1	0	1	2
	7	0	0	0	0	0	1	1	0	1
	8	0	0	0	4	0	1	0	4	0
	9	2	0	0	9	0	9	0	3	3

Figura 2 – Matriz de Confusão.

Fonte: A autora

Em sequência, foram calculados os classificadores de performance e métrica. Usando as expressões 2.1 até 2.6 para cada classe e calculando sua média aritmética foram encontrados os resultados apresentados na Tabela 3.

Classificadores e Métrica	Resultado
Acurácia	0,8681
Precisão	0,4929
Sensibilidade	0,0524
Taxa de Falsos Positivos	0,1319
Taxa de Classificação Incorreta	0,0749
F1 <i>Score</i>	0,0895

Tabela 3 – Resultados dos Classificadores de Performance e Métricas.

Como o modelo apresenta classes desbalanceadas (cada uma com números diferentes de exemplos), iremos estabelecer o ponto de comparação em 0,50, esperando que o modelo acerte, ao menos, metade dos valores.

Acurácia: Está bem acima do ponto de comparação, mostra que o modelo prevê mais de 85% dos exemplos corretamente. Apesar de o valor obtido ser muito bom, a acurácia pode mostrar resultados enganosos, mascarando problemas ao favorecer a classe majoritária.

Precisão: Não apresenta um resultado satisfatório, estando logo abaixo do ponto de comparação, o modelo está fazendo muitas previsões positivas incorretamente, classificando dados nas categorias erradas.

Sensibilidade: Apresenta um resultado fraco, muito perto de zero. Implica que suas previsões positivas estão ausentes, com muitas estimativas incorretas.

Taxa de Falsos Positivos: Por mais que o resultado apresentado seja baixo em relação aos outros classificadores, é um bom sinal, o modelo apresenta uma baixa proporção de casos negativos classificados como positivos comparados ao total que realmente eram negativos, mas em geral ainda está elevado, sendo que seu melhor resultado é zero.

Taxa de Classificação Incorreta: Apresenta seu resultado próximo de zero, mostra que o modelo raramente classifica incorretamente os exemplos. É o valor complementar da acurácia, então também pode ser enganoso.

F1 *Score*: Para um modelo ter seu desempenho considerado excelente, seu F1 *score* deve ser o mais próximo de 1 possível. No caso estudado, o modelo tem resultado muito próximo de zero, tornando-o fraco.

Ainda que sua acurácia e TCI apresentem valores desejáveis, pela métrica utilizada, o modelo é insatisfatório. Embora o modelo em si seja bem condicionado, com treinamento extenso e completo, o caso estudado pode ser considerado muito complexo. Por ser um assunto extremamente específico de uma área desconhecida pelo senso comum, o modelo

se torna inapto para esse tipo de categorização. Além disso, foram usadas muitas classes, 9 no total, tornando ainda mais complexa a classificação.

4.2 Criação do *Dashboard*

Depois do refinamento e de novas variáveis adicionadas (motivo cancelamento (especialista) e motivo cancelamento (PLN)), a base de dados apresenta 17 variáveis, sendo elas:

circuito: número de identificação do circuito usado, também para identificação do cliente pelo fato de ser único. Originalmente modificados, pois são dados confidenciais, são então números sequenciais começando com 1 no primeiro contrato documentado até 735203, sendo o contrato mais recente até o momento da extração.

status: mostra a categoria que o contrato está de forma detalhada. Se está ativo, cancelado, cancelado por inadimplência, etc.

data aceite: data na qual o cliente, ao receber o contrato, aceita os termos e condições e o assina.

data instalação: data na qual foi feita a instalação do equipamento no local contratado pelo cliente.

data pedido cancelamento: data na qual houve um pedido de cancelamento ou terminação por parte da empresa, em casos como inadimplência ou inatividade, por exemplo.

data fim: data na qual o contrato é cancelado oficialmente, dada a tentativa de renovação e o tempo de processamento do sistema interno.

motivo cancelamento: variável usada no estudo de PLN para classificação dos dados. Apresenta 91 alternativas do possível motivo para cancelamento, acumuladas durante os anos.

tipo pessoa: física ou jurídica. Ao preencher os dados cadastrais, se for usado um CPF para o titular, é considerado pessoa física. Se for usado CNPJ, é considerado pessoa jurídica.

origem troca produto: indica se houve troca de produto/pacote contratado, campo preenchido por sim ou não.

produto final: nome do produto/pacote contratado.

velocidade: velocidade da internet contratada, entre 300 *bps* (bits por segundo) até 100 *Gbps* (gigabits por segundo).

valor circuito: valor pago por mês, de acordo com o produto contratado.

cidade: a cidade que está no endereço do contrato.

bairro: bairro da cidade do endereço no contrato.

data nascimento: data de nascimento que os clientes colocam ao preencher o contrato.

motivo cancelamento (especialista): classificação, feita por um especialista, da variável motivo cancelamento, transformando-a em apenas 9 classes.

motivo cancelamento (PLN): classificação, feita pelo PLN em inglês e traduzida para português, da variável motivo cancelamento, transformando-a em apenas 9 classes.

Desse modo, a base foi importada no *Power BI*, uma plataforma unificada e escalonável para autoatendimento e BI (*business intelligence*) empresarial, podendo visualizar quaisquer dados e incorporar perfeitamente os recursos visuais nos aplicativos usados diariamente (MICROSOFT, 2025). Foram adicionadas colunas calculadas, usando fórmulas DAX (*Data Analysis Expressions*), criando expressões para otimizar a visualização, sendo elas:

Idade: Usa a variável *data nascimento* e calcula a idade em anos no dia de hoje.

`Idade = DATEDIFF([data nascimento], TODAY(), YEAR)`

Tempo Instalação: Diferença das variáveis *data aceite* e *data instalação* em dias. Mostra a quantidade de dias desde o aceite do contrato até o dia da instalação do equipamento.

`Tempo Instalação = DATEDIFF([data aceite], [data instalação], DAY)`

Tempo Contrato: Diferença das variáveis *data aceite* e *data fim* em meses. Mostra a quantidade de meses que um cliente ficou ativo, desde o aceite do contrato até o dia em que o cancelamento foi computado no sistema.

`Tempo Contrato = DATEDIFF([data aceite], [data fim], MONTH)`

Gasto Total: Calcula o quanto um cliente já gastou desde o primeiro mês até seu fim, ou até o dia presente, caso não tenha sido cancelado.

```

Gasto Total =
VAR DataFimContrato =
    IF( ISBLANK([data fim]), TODAY(), [data fim] )
VAR MesesAtivo =
    DATEDIFF([data aceite], DataFimContrato, MONTH)
RETURN
    MesesAtivo * [valor circuito]

```

Classe Status: Reorganiza os campos da variável *status* de forma mais condensada, para melhor entendimento dos dados.

```

Classe Status =
SWITCH(
    TRUE(),
    -- Ativo
    [status] IN {"Habilitado", "Velocidade reduzida", "Em Reabilitação"},
    "Ativo",

    -- Cancelado / Encerrado
    [status] IN {"Cancelado Espont.", "Cancelado por Inadimplência",
    "Em Retirada de Equip.", "Desabilitado", "Em Desabilitação"},
    "Cancelado",

    -- Em Processo / Pré-Contrato
    [status] IN {"Proposta", "Desistente", "Em Análise de Crédito",
    "Aguardando Aceite", "Crédito Reprovado", "Aguardando Instalacao",
    "Suspensao", "Em Suspensão"}, "Processamento",

    -- Caso não se encaixe
    "Outro")

```

Tempo Retirada: Diferença das variáveis *data pedido cancelamento* e *data fim* em dias. Mostra a quantidade de dias desde o pedido de cancelamento até o dia em que foi computado no sistema.

```

Tempo Retirada = DATEDIFF([data pedido cancelamento],
    [data fim], DAY)

```

Além disso, para representar a localização correta das cidades, foram coletadas as latitudes e longitudes de cada, vindo de um arquivo disponibilizado na plataforma *GitHub*.

Dessas informações, foi criada uma planilha no *Excel* das cidades com suas respectivas coordenadas, em seguida importada para o *Power BI* e relacionada com a coluna cidade da base original, tornando possível usá-las em um mapa e transmitir dados corretos.

Com todas as informações organizadas, foi possível criar algumas visualizações de valores no *dashboard*.

4.2.1 Total de Contratos

Cartão que mostra a contagem distinta de cada contrato, totalizando o número de linhas da base de dados.



Figura 3 – Cartão do Total de Contratos.

Fonte: A autora

4.2.2 Receita Total Arrecadada

Cartão que mostra o valor acumulado arrecadado durante o período de tempo selecionado. É considerado apenas o que foi recebido durante os meses de contrato ativo, sem a redução dos custos de material e mão de obra, essas informações não estavam disponíveis para utilizar na base de dados.

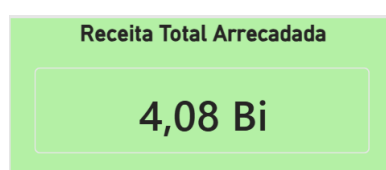


Figura 4 – Cartão da Receita Total Arrecadada.

Fonte: A autora

4.2.3 Contrato Ativo

Cartão que mostra a média em meses que um contrato fica ativo, o padrão são 12 meses. A média sendo em torno de 20 meses mostra que há renovação do contrato, consequentemente, os clientes tendem a continuar com o produto por mais de um ano.



Figura 5 – Cartão do Contrato Ativo.

Fonte: A autora

4.2.4 Tempo para Instalação

Cartão que mostra a média em dias do tempo do aceite do contrato até o dia oficial da instalação. É considerado levemente elevado, pois é necessário marcar um horário amplo (algumas horas) no qual o cliente ou um responsável esteja presente para acompanhar a instalação. Normalmente demora, pois depende da disponibilidade dos equipamentos mais próximos do endereço e se há necessidade de alguma mudança de estrutura.



Figura 6 – Cartão do Tempo para Instalação.

Fonte: A autora

4.2.5 Tempo para Cancelamento

Cartão que mostra a média em dias do tempo entre o contrato terminar/ser cancelado até a retirada do aparelho do endereço cadastrado. É um processo rápido, pois é recolhido apenas o equipamento usado, não precisa instalar cabeamento de rede.



Figura 7 – Cartão do Tempo para Cancelamento.

Fonte: A autora

4.2.6 Faixa Etária

Gráfico de barras que mostra o volume de pessoas pela idade informada. O pico está para adultos de meia idade que, em geral, têm a responsabilidade pelo pagamento das contas, especialmente por terem maior estabilidade financeira.

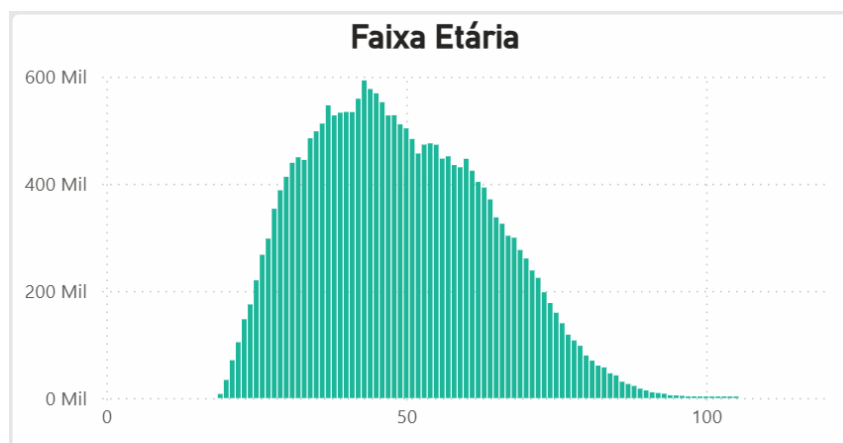


Figura 8 – Gráfico da Faixa Etária.

Fonte: A autora

4.2.7 Tipo Pessoa

Gráfico de pizza que mostra os campos pessoa física ou jurídica. A maior parte é de pessoa física, o que indica que a maioria dos clientes se encaixa na modalidade *B2C* (*business-to-consumer*) mas ainda são atendidos clientes *B2B* (*business-to-business*).

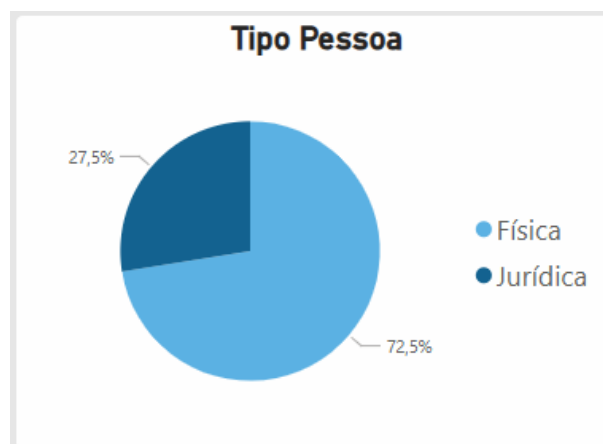


Figura 9 – Gráfico do Tipo Pessoa.

Fonte: A autora

4.2.8 Status do Contrato

Gráfico de rosca que mostra a separação dos status como ativo, cancelado ou em processamento. Essa visualização permite entender o volume de contratos encontrados em cada categoria, desencadeando ações para melhorar o número de contratos ativos e, conseqüentemente, alterar os em processamento e diminuir os cancelados em comparação com o total.

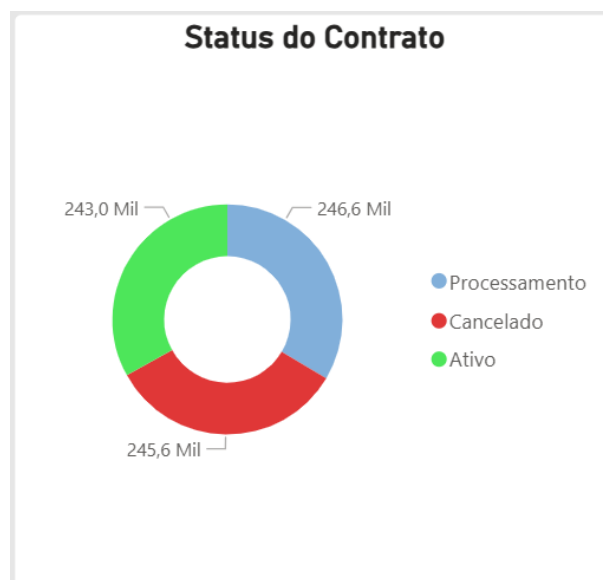


Figura 10 – Gráfico do Status do Contrato.

Fonte: A autora

4.2.9 Motivo de Cancelamento (Especialista) e (PLN)

Gráficos de linhas que mostram o volume da classificação de um especialista e do PLN dos motivos de cancelamento, feitos para serem comparados diretamente, já que têm as mesmas classes, mas com volumes diferentes, representando as diferenças de cada classificação.

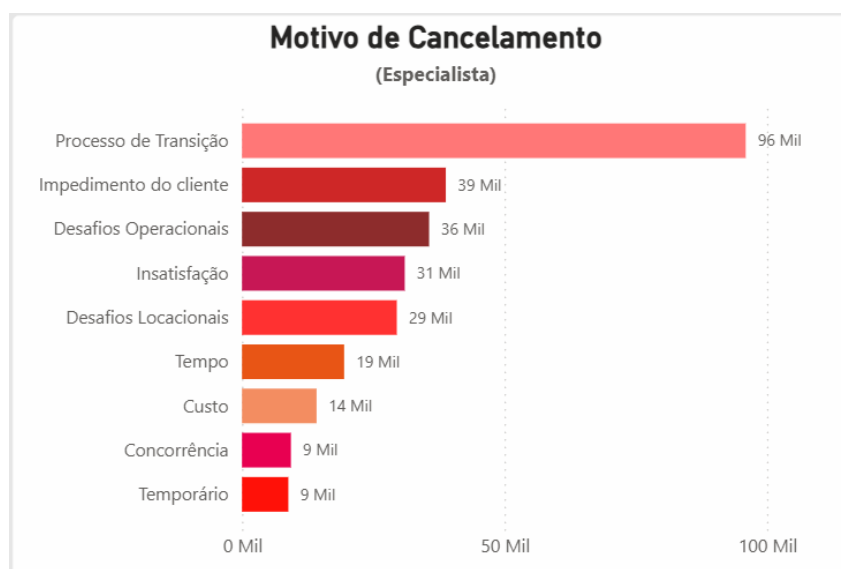


Figura 11 – Gráfico do Motivo de Cancelamento (Especialista).

Fonte: A autora

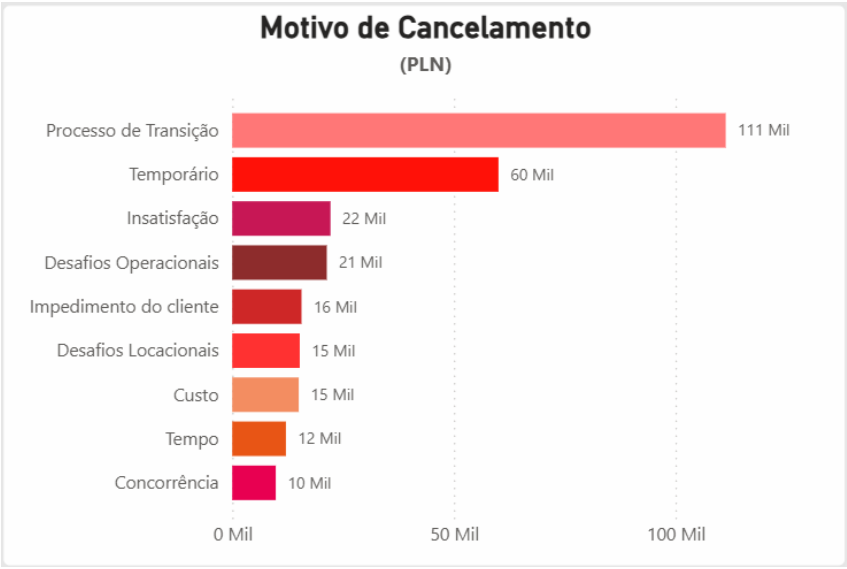


Figura 12 – Gráfico do Motivo de Cancelamento (PLN).

Fonte: A autora

4.2.10 Volume de Contratos por Cidade

No mapa do Paraná, cada círculo representa o volume de contratos por cada cidade. Quanto maior a bola, maior seu volume.

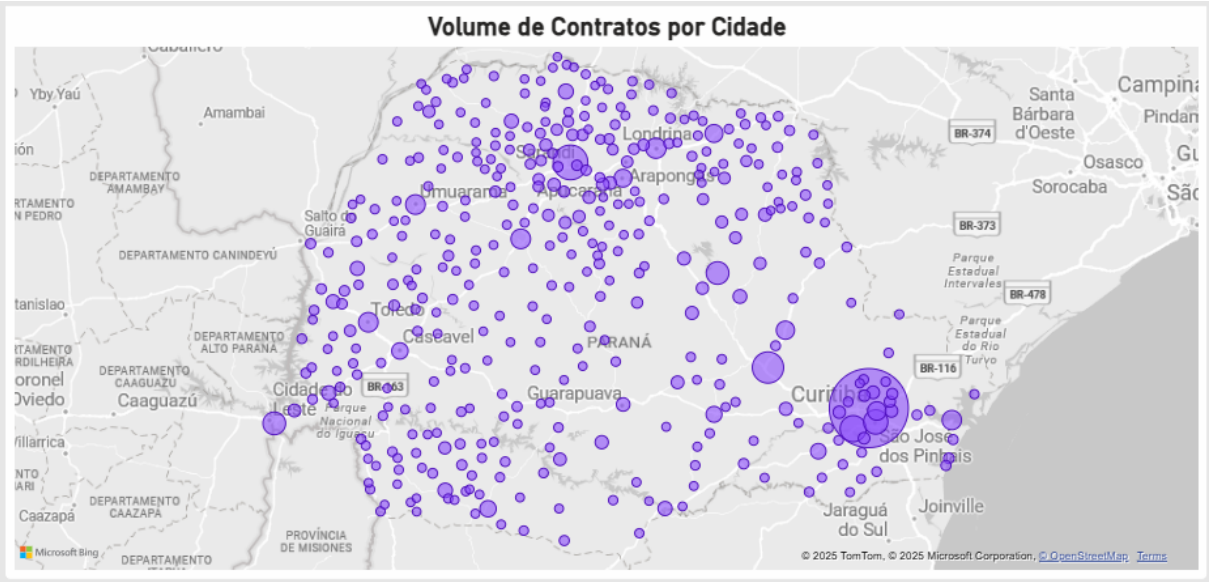


Figura 13 – Mapa do Paraná com o Volume de Contratos por Cidade.

Fonte: A autora

4.2.11 Cidades com mais Contratos

Gráfico de colunas mostrando as 5 cidades com maior volume de contratos, Curitiba a maior de todas, não só por ser a capital do estado e ter a maior população em comparação

com as outras, mas também por ser onde está instituída a sede atual da empresa. Em seguida, as cidades com mais clientes são não só da área metropolitana de Curitiba, mas também grandes centros urbanos do estado.



Figura 14 – Gráfico das Cidades com mais Contratos.

Fonte: A autora

4.2.12 Filtros

Para visualização mais precisa, foram adicionados filtros para selecionar uma ou mais cidades e também uma data ou período de tempo específico.

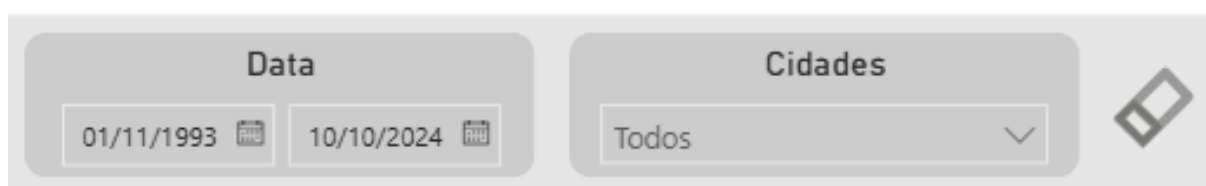


Figura 15 – Filtros.

Fonte: A autora

4.2.13 Dashboard Completo

A visão completa do *dashboard* pode ser observada na Figura 16.

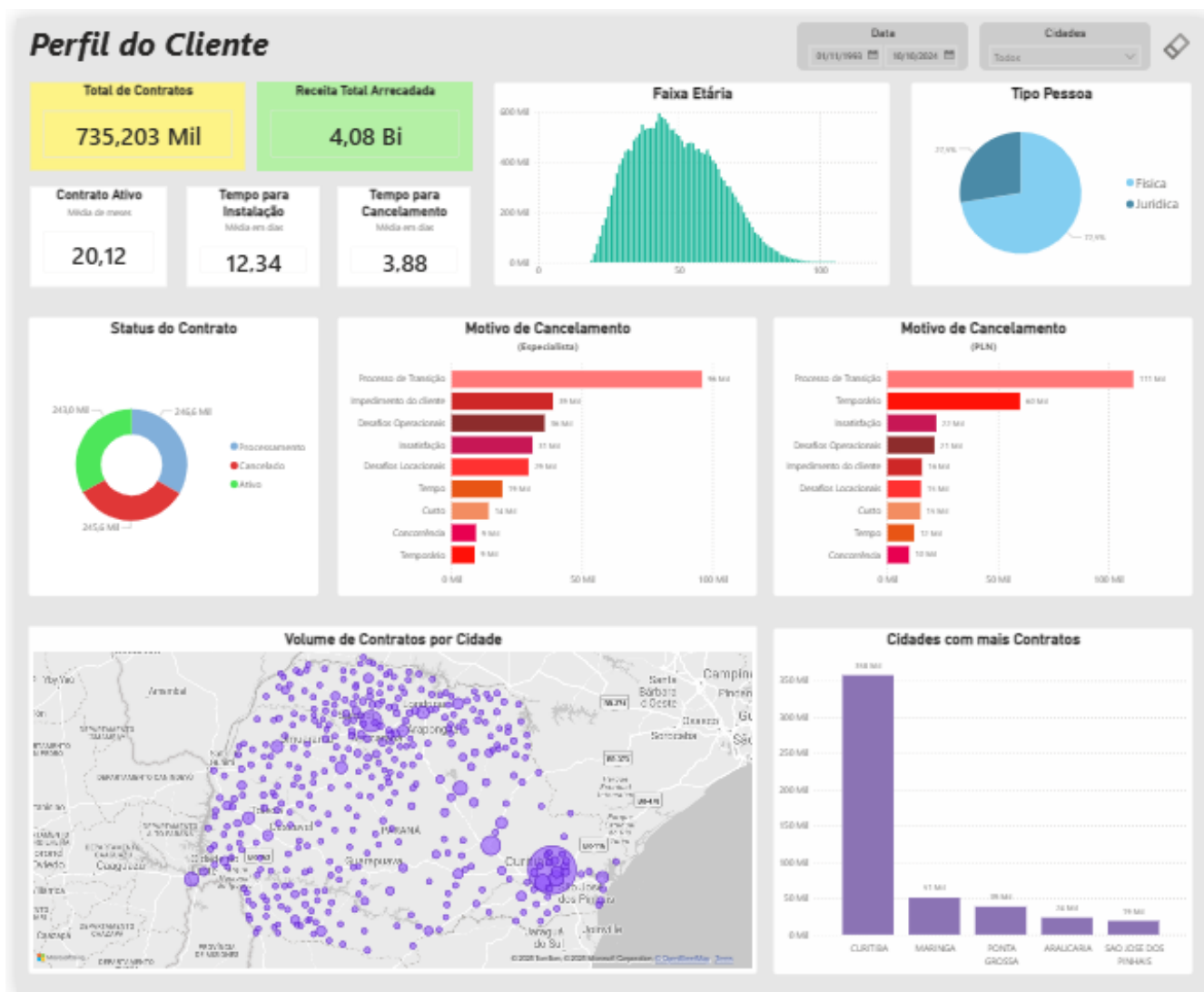


Figura 16 – *Dashboard* Completo.

Fonte: A autora

O *dashboard* é um facilitador de visualização rápida e objetiva de uma grande quantidade de dados. É uma ferramenta que pode auxiliar na tomada de decisão, tal como possibilidade de expansão da rede em uma certa cidade, entre outros. Além disso, a classificação do motivo de cancelamento possibilitou identificar onde residem os maiores problemas da companhia, pois esta nunca quer perder seus clientes.

5 Conclusão

O estudo possibilitou a criação de um painel sobre o perfil dos clientes de uma empresa de telecomunicações, com o intuito de evidenciar pontos importantes que possam servir para auxiliar a tomada de decisão para melhorar a experiência do consumidor.

A pesquisa usou conhecimentos aprendidos durante o curso de Matemática Industrial, como a linguagem *Python* utilizada para programar e transformar o modelo de PLN para o escopo principal, conhecimentos de estatística para fazer a análise crítica dos resultados da matriz de confusão e também certificar que o produto final está correto, sem cálculos errados.

No universo da tecnologia, volumes intensos de informações são constantes, tornando-se necessário estar sempre atualizado das mudanças tecnológicas e socioeconômicas, além de criar um ambiente organizado de dados, com uma visão concisa e acessível para todos. A veracidade das informações é fundamental para um *dashboard* completo e confiável.

Inicialmente o propósito principal foi definido para criar um *dashboard* sobre o perfil de clientes com a base de dados extraída de uma empresa de telecomunicações. Ao fazer um refinamento, foi notado que uma variável poderia ser simplificada, assim surgiu a ideia de usar o Processamento de Linguagem Natural, por consequência a classificação *Zero-Shot* para aprimorar as possíveis classes de caracterização do motivo de cancelamento dos contratos. Aplicando o modelo escolhido no problema encontrado, foi criada a Matriz de Confusão para avaliá-lo, mostrando resultados insatisfatórios de acordo com os classificadores de performance e métrica usados.

Mesmo a classificação sendo, de certa forma, insuficiente, ao criar o painel e seus gráficos no *Power BI*, foram exibidos os resultados do PLN e também do especialista, além dos pontos sobre o perfil do cliente. Possibilitando, assim, a visualização de tópicos importantes como o tempo entre um pedido e sua realização, o volume de clientes por cidade, etc. A partir disso, podem ser criadas medidas para otimização do tempo e de possíveis estratégias de vendas em cidades com menos clientes ativos.

Referências

- ABREU, K. C. K. *História e usos da Internet*. 2009. <<https://www.bocc.ubi.pt/pag/abreu-karen-historia-usos-internet.pdf>>. BOCC – Biblioteca Online de Ciências da Comunicação. 9
- BARNARD, J. *Incorporações de palavras*. 2025. <<https://www.ibm.com/br-pt/think/topics/word-embeddings>>. Acesso em: 17 dez. 2025. 13
- BENAKOUCHE, T. Redes técnicas/redes sociais: pré-história da internet no brasil. *Revista USP*, p. 124–133, 1997. 9
- GALINA, S. V. R.; PLONSKI, G. A. Inovação no setor de telecomunicações no brasil: uma análise do comportamento empresarial. *Revista Brasileira de Inovação*, v. 4, p. 129–155, 2005. 10
- GHAFFARI, M.; GAZIT, L. *Mastering NLP from Foundations to LLMs*. [S.l.]: Packt Publishing Ltd, 2024. 13
- GRANDINI, M.; BAGLI, E.; VISANI, G. Metrics for multi-class classification: an overview. *Cornell University*, v. 1, ago. 2020. 14
- HANSOTI, B. *Business Intelligence Dashboard in Decision Making*. 2–17 p. Tese (Doutorado) — Instituição não especificada, 2010. 7, 11
- Hugging Face. *Hugging Face – On a mission to solve NLP, one commit at a time*. 2024. <<https://huggingface.co/>>. Huggingface.co. 13, 17
- JOSEPH, S. et al. Natural language processing: A review. *International Journal of Research in Engineering and Applied Sciences*, v. 6, 2016. <<https://www.icts.res.in/sites/default/files/media/media-library/NLPIntro.pdf>>. 12
- KNIGHT, P. T. *A Internet no Brasil*. [S.l.]: Author House, 2014. 10
- LAURER, M. *Model card for mDeBERTa-v3-base-xnli-multilingual-nli-2mil7*. 2024. <<https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7>>. Huggingface.co. 13
- MATTERN, S. Mission control: A history of the urban dashboard. *Places Journal*, mar. 2015. Acesso em 20 set. 2019. 7, 10, 11
- Memória Globo. *Trama Principal*. 2021. <<https://memoriaglobo.globo.com/entretenimento/novelas/explode-coracao/noticia/tramas.ghtml>>. Acesso em 21 ago. 2025. 9
- MICROSOFT. *Power BI - Visualização de Dados / Microsoft Power Platform*. 2025. <<https://www.microsoft.com/pt-br/power-platform/products/power-bi>>. Acesso em: 17 dez. 2025. 22
- MONTEIRO, M. P. B.; OLIVEIRA, A. de. Processo de privatização e expansão das telecomunicações no brasil. *Revista Avaliação de Políticas Públicas - AVAL*, v. 3, 2015. 7, 10

- MUREL, J.; KAVLAKOGLU, E. *Matriz de confusão*. 2024. <<https://www.ibm.com/br-pt/think/topics/confusion-matrix>>. IBM.com. 13, 14
- PERNA, C. L.; DELGADO, H. K.; FINATTO, M. J. *Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa*. [S.l.]: EDIPUCRS, 2010. 7, 12
- RAMESH, G. et al. A review on nlp zero-shot and few-shot learning: methods and applications. *Deleted Journal*, v. 7, ago. 2025. 12, 13
- SATHYANARAYANAN, S.; TANTRI, B. R. *View of Confusion Matrix-Based Performance Evaluation Metrics*. 2024. <<https://africanjournalofbiomedicalresearch.com/index.php/AJBR/article/view/4345/3327>>. African Journal of Biomedical Research. 14
- VARGAS, A. V.; GONÇALVES, G. *Fibra óptica e um estudo dirigido aos planos de internet disponíveis no Brasil e em outros países*. Tese (Doutorado) — Instituto Federal Farroupilha, 2020. <<https://arandu.iffarroupilha.edu.br/handle/itemid/142>>. 9, 10
- VASCO, L. *Elaboração de uma sistemática de gerenciamento da qualidade de saúde dos colaboradores de uma empresa de grande porte*. Tese (Doutorado) — Instituição não especificada, 2021. 11