Revista Multidisciplinar do Nordeste Mineiro, v.12, 2024
ISSN 2178-6925

ANÁLISE DIAGNÓSTICA DE DADOS EM TEXTOS NÃO ESTRUTURADOS POR PROCESSAMENTO DE LINGUAGEM NATURAL PARA IDENTIFICAÇÃO E MITIGAÇÃO DE CAUSAS DE PERDA DE CLIENTES EM EMPRESAS

DIAGNOSTIC DATA ANALYSIS OF UNSTRUCTURED TEXTS USING NATURAL LANGUAGE PROCESSING FOR THE IDENTIFICATION AND MITIGATION OF CAUSES OF CUSTOMER CHURN IN COMPANIES

Jhonny Anthony Ribeiro
Cursando Ciências da Computação,
Universidade do Estado de Mato Grosso - UNEMAT, Brasil,
E-mail: jhonny.anthony@unemat.br

Leandro Avelino Mazurek
Professor Orientador: Mestre em Ciência da Computação,
Especialista em *Business Intelligence*,
Universidade do Estado de Mato Grosso - UNEMAT, Brasil
E-mail: leandro.mazurek@unemat.br

Raquel Da Silva Vieira Coelho
Professora da Disciplina TCC II: Mestre em Ciência da Computação,
Universidade do Estado de Mato Grosso - UNEMAT, Brasil
E-mail: raquelcoelho@unemat.br

Resumo

A perda de cliente resulta em redução da receita, aumento dos custos e diminuição da reputação, o que torna mais difícil a aquisição de novos clientes. Com a realização de uma análise dos dados referente as causas das perdas de clientes, é possível desenvolver estratégias para mitigá-las, e proteger receitas e os recursos. Ao compreender as razões subjacentes, as empresas podem direcionar seus esforços para áreas específicas que demandam atenção, contribuindo assim para a retenção de clientes e a sustentabilidade do negócio. Este artigo tem como base a utilização de técnicas de processamento de linguagem natural e análise diagnóstica em registros textuais de dados simulados, referente a motivos ligados a saídas de clientes, para identificá-los e compreende-los, bem como estipular soluções e estratégias para prevenir futuros cancelamentos.

Palavras-chave: Análise Diagnóstica, Processamento de Linguagem Natural, Retenção de Clientes, Perda de Clientes.

Abstract

Customer loss results in revenue reduction, increased costs, and diminished reputation, making it more difficult to acquire new customers. By conducting an analysis of the data regarding the causes of customer churn, it is possible to develop strategies to mitigate them and protect revenues and resources. Understanding the underlying reasons allows companies to focus their efforts on specific areas that demand attention, thus contributing to customer retention and business sustainability. This article is based on the use of natural language processing techniques and diagnostic analysis applied to textual records of simulated data related to reasons for customer churn, aiming to identify and understand them, as well as to propose solutions and strategies to prevent future cancellations.
Keywords: Diagnostic Analysis, Natural Language Processing, Customer Retention, Customer Churn.

1. Introdução

A análise de dados de perda de clientes é de grande importância para o sucesso e a sustentabilidade da empresa, uma vez que a perda resulta em prejuízos financeiros, custos elevados e danos à reputação, dessa forma se torna necessário identificar e compreender seus principais motivos, bem como estipular soluções e estratégias para prevenir futuros cancelamentos. De acordo com Kotler (2006 apud Rodrigues, 2020, p. 24):

> O nível de fidelização do cliente leva a um comportamento positivo do mesmo quando ele ajuda a fechar novos negócios por meio da recomendação, assim a empresa tende a evitar investimentos de propaganda, por exemplo, uma vez que a boca a boca confere alta credibilidade à empresa o custo da manutenção em atividade do cliente é inferior ao custo de captação de novos clientes, diminuindo assim os efeitos da perda de clientes.

Dessa forma a análise não apenas oferece *insights* cruciais para a retenção de clientes e o crescimento sustentável das empresas, mas também desempenha um papel fundamental na melhoria contínua dos produtos e serviços oferecidos. E para identificar as causas ligadas é necessário compreender as relações da empresa com o cliente, e essas relações por muitas vezes não são registradas de modo categorizadas ou estão registrados de forma descritiva e individualizado para cada relação gerada, esses dados textuais não estruturados em um âmbito geral não tem utilidade e muitas vezes é trato como irrelevantes, contudo, são extremamente importantes, pois através dos mesmos, os motivos destacados por cada indivíduo

referente a sua desistência da utilização dos produtos ou serviços da empresa estão registrado.

E de modo a compreender e categorizar tais dados é necessário realizar o Processamento de Linguagem Natural (PLN), uma vez que através do mesmo se torna possível permitir que os computadores compreendam a linguagem humana.

Com esses dados processados previamente as relações e causas podem ser interpretadas através da Análise Diagnóstica, que consiste de acordo com Bolívar, Castro e Herreño (2020, tradução nossa) em um processo essencial de identificação e compreensão de um problema ou fenômeno específico.

Ao compreender as razões pelas quais os clientes optam por cancelar as relações com uma empresa, pode-se não apenas implementar medidas para reter clientes existentes, mas também identificar áreas de oportunidade para inovação e aprimoramento. Portanto, esta pesquisa não só visa resolver problemas imediatos de perda de clientes, mas também contribuir para a evolução, melhorar práticas de retenção de clientes, garantir o crescimento sustentável e competitividade contínua no mercado.

2. Processamento de Linguagem Natural

O PLN é um campo da ciência da computação e da inteligência artificial que se concentra no desenvolvimento de algoritmos e técnicas para permitir que os computadores compreendam, interpretem e gerem linguagem humana de forma natural. De acordo com Barbosa, *et al.* (2017), seu propósito central é extrair representações e significados mais abrangentes de textos escritos em linguagem natural, que se referem a idiomas usados em comunicações cotidianas, como o Português, Inglês ou Mandarim.

2.1. Python

Python é uma linguagem de programação de alto nível criada no início dos anos 1990 por Guido van Rossum, utilizada para solucionar demandas operacionais em diversos meios (Lambert, 2022, p. 18). Ela pode ser utilizada no desenvolvimento web, análise de dados, inteligência artificial, automação de tarefas e muitas outras

aplicações.

Sua sintaxe simples e legibilidade tornam-a popular entre programadores de todos os níveis de experiência. Maciel (2020, p. 3) salienta que:

> Discutir a popularidade de qualquer coisa é um assunto traiçoeiro, sempre se esbarra em preferências pessoais. Porém, pode-se dizer que, entre as diversas justificativas para a linguagem Python ser muito utilizada, destacam-se: Eficiência, Comunidade ativa, Simplicidade, Forte presença na academia, Tendência.

Além disso, tem um vasto conjunto de bibliotecas e *frameworks* para diversas aplicações no geral, Python é uma linguagem versátil e poderosa, amplamente utilizada em uma variedade de domínios, desde desenvolvimento de software até ciência de dados e automação de tarefas, dessa forma a sua utilização para realização do PLN, se torna prático e intuitivo.

2.2. Bibliotecas Python

As bibliotecas pertencentes ao Python são extensões facilitam o desenvolvimento de operações, sem ser necessário estar recriar funções ou códigos já existentes, permitindo um desenvolvimento mais rápido e preciso, além de permitir integrar mecanismos mais robustos e próprios para o PLN.

2.2.1. Pandas

Para realizar a leitura e análise dos dados de dados relacionais ou rotulados, a biblioteca Pandas é oferece estruturas de dados rápidas, flexíveis e expressivas, projetadas para facilitar esse trabalho de maneira simples e intuitiva. Conforme Reback *et al.* (2020, tradução nossa), seu objetivo é ser o principal bloco de construção de alto nível para a realização de análise de dados prática e do mundo real no Python.

Desse modo sua utilização permite a leitura de diversos formatos de arquivos, a criação de estruturas para análise, a transformação e limpeza dos dados, a filtragem e seleção de observações com base nos rótulos, e a exportação dos dados processados para diferentes formatos. A versatilidade do Pandas torna-o uma

ferramenta indispensável para o pré-processamento de dados rotulados em diversas aplicações.

2.2.2. Gemini API

Large Language Models (LLM), Modelos de Linguagem de Grande Escala, é a junção entre diversos desenvolvimentos no campo das redes neurais e do processamento de linguagem natural. Rachmat e Kesuma (2024, p. 41, tradução nossa) definem:

> Trata-se de um tipo de modelo de inteligência artificial capaz de processar e gerar texto em linguagem natural. Os modelos de LLM são geralmente treinados com uma quantidade enorme de dados textuais e utilizam técnicas de aprendizado profundo para aprender padrões e estruturas linguísticas. Essa capacidade permite que o LLM execute uma variedade de tarefas relacionadas ao uso da linguagem, incluindo a geração de texto, tradução de idiomas, resumo de textos, resposta a perguntas, escrita de código e análise de sentimentos.

O Gemini API é um LLM que se destaca pela sua versatilidade e poder computacional, sendo adaptativo a diferentes ambientes, desde servidores até dispositivos móveis, o Gemini é multimodal, processando e gerando diversos tipos de dados, e apresenta um desempenho superior a outros modelos de LLMs, conforme ao estudo feito por Raposo (2024), ao comparar o Gemini com outros dois modelos, ChatGPT 3.5 e Llama, para responder questões do ENEM, apresentou superioridade no desempenho e ainda vale ressaltar conforme que em comparativos de realizado por Moraes *et al.(*2024, p. 9), “Na detecção de ambiguidade, o ChatGPT conseguiu uma acurácia de 28,75% e o Gemini 49,58%”. Desse modo a utilização da Gemini API se torna apropriada tanto pelo seu desempenho quando por sua praticidade.

3. Análise Exploratória de Dados

Mesmo com a classificação e desenvolvimento do PLN ainda se torna necessário agrupar e transformar todas os dados, gerar gráficos e tabelas, para que

seja possível realizar a análise, esse processo conforme Ferreira, Miranda e Pinto (2021, p. 15), conceituam, se denomina de análise exploratória de dados (AED), sendo o meio que se obtém o entendimento sobre os dados coletados e, principalmente, sobre as relações existentes entre as variáveis analisadas.

Orange Data Mining é um *software* que permite a análise de dados, se destaca por ser uma ferramenta de código aberto, intuitiva e voltada para a mineração de dados e aprendizado de máquina. Sua interface gráfica é baseada em blocos que facilitam a construção de fluxos de trabalho para análise de dados, sem a necessidade de escrever código, conforme Rodrigues (2023, p. 32) explica:

> Por meio de sua interface gráfica denominada Orange Canvas é possível, criar todo fluxo de trabalho de um projeto de Data Mining, sem necessidade de código, de forma simples, visto que este apresenta um design visual e inovador. Além do suporte às tarefas que vão desde o pré-processamento de dados até modelagem e avaliação.

O Orange permite a importação de dados de diversas fontes, integra-se com bibliotecas, oferecendo um ecossistema flexível para analistas de dados e cientistas de dados.

4. Análise de Dados Diagnóstica

Análise se refere ao estudo individual de um todo, com o intuito de esclarecer suas relações, origens e proporções. Mcfedries (2020, p. 21), define que “análise de dados é a aplicação de ferramentas e técnicas para organizar, estudar, chegar a conclusões e, às vezes, também fazer previsões sobre uma coleção específica de informações”. Bolívar, Castro e Herreño (2020, tradução nossa) define diagnóstico como um processo essencial de identificação e compreensão de um problema ou fenômeno específico.

Relacionando a descrição de diagnóstico à de análise de dados, se tem a definição de análise diagnóstica, um processo sistemático de investigação usado para identificar problemas, padrões e causas subjacentes em uma determinada situação. Minayo e Deslandes (2007, p. 32) enfatiza que a Análise Diagnóstica:

> [...] busca as causas (complexas) dos problemas para medir e compreender o sucesso e o insucesso de determinada ação. Por sua vez, o processo diagnóstico vincula-se estreitamente ao mecanismo de retorno, tendo, portanto, um caráter estratégico.

A análise de dados diagnóstica concentra-se em identificar as causas fundamentais de um evento. Ela envolve a coleta e análise de dados relevantes, como sintomas, resultados de testes ou comportamentos, a fim de formular um entendimento claro do problema em questão. Seus propósitos abrangem a resolução de problemas, a melhoria de processos e a orientação das tomadas de decisões. Conforme, Sarker (2021, p. 7, tradução nossa) a "Análise diagnóstica: é uma forma de análise avançada que examina dados ou conteúdo para responder à pergunta 'por que isso aconteceu?' O objetivo da análise diagnóstica é ajudar a encontrar a causa raiz do problema". Nesse mesmo sentindo,

> No geral, a análise de diagnóstico permite extrair valor dos dados, colocando as perguntas certas e conduzindo investigações aprofundadas sobre as respostas. É caracterizado por técnicas como detalhamento, descoberta de dados, mineração de dados e correlações. (SARKER, 2021, p. 7, tradução nossa).

Ao examinar minuciosamente os dados históricos, essa abordagem analítica permite uma compreensão profunda dos fatores que contribuíram para determinados resultados, oferecendo uma base sólida para a implementação de soluções eficazes, aprimoramento contínuo de operações e escolhas informadas.

4.1. Motivos Comuns para a Perda de Clientes

A deserção de um cliente é por muitas vezes ligados a diversos fatores, dentre esses fatores Oliveira (2018), evidencia oito principais: preços, inconveniência, falhas no serviço central, falhas nos encontros de serviços, resposta dos funcionários às falhas do serviço, atração por concorrentes, problemas éticos, mudança e ainda destaca que dentre esses fatores os cinco primeiros são mais relevantes. Bateson e

Hoffman (2016, p. 414) reiteram que:

> Os desertores por preço mudam para concorrentes de bens e serviços com preços mais baixos [...]. Os desertores por produto mudam para concorrentes que oferecem produtos e serviços de qualidade superior. Os desertores por serviço migram em função do mau serviço prestado ao cliente [...]. Os desertores do mercado saem do mercado por razões de mudança/transferência ou insucesso empresarial.

Considerando as informações fornecidas pelos autores, observa-se que os motivos centrais estão frequentemente associados à negligência das empresas em relação à gestão do cliente e também à falta de atenção na necessidade de manter-se constantemente atualizado no mercado, bem como na competição contínua com outros concorrentes.

4.1.1. Impacto da Perda de Clientes na Receita

De acordo com Bateson e Hoffman (2016), o custo para adquirir novos clientes é de três a cinco vezes menor em relação a obtenção de novos e nesse mesmo âmbito um cliente insatisfeito gera custos na receita e prejuízos a imagem da empresa,

> [...] o impacto de reclamações não tratadas é a divulgação negativa da experiência, direto impacto na reputação e imagem da empresa e a consequente perda de clientes, mercado e receita para as empresas. [...] (Couto, 2015, p. 16).

Além da perda de potenciais clientes e receitas, isso também acarreta uma deterioração da reputação da empresa, impactando sua credibilidade no mercado, conforme Sparemberger (2019, p. 84) “Assim, o boca a boca negativo representa um grande problema para a empresa, ocasionando a perda de clientes para a concorrência, bem como a destruição da sua imagem”. Desse modo, assim como clientes engajados geram marketing positivo e promovem a aquisição de novos clientes por meio de recomendações, clientes insatisfeitos resultam em marketing negativo.

4.1.2. Estratégias de Retenção e Fidelização de Clientes

Para gerenciar as perdas de clientes empresas adotam medidas para a retenção e fidelização de clientes, uma vez que conforme Sparemberger (2019, p. 94) afirma a perda de um cliente não influencia somente na perda da próxima transação com a empresa mais sim todas as possíveis transações que poderia ocorrer, de mesmo modo,

> A gestão do relacionamento com o cliente é o processo de construir e manter relacionamentos lucrativos com esse cliente, entregando-lhe valor superior e atendendo suas necessidades, o que resulta em compras maiores e continuas, gerando retornos de longo prazo para a empresa (Alvarez, 2015, p. 19).

Partindo desse ponto a retenção e fidelização emprega que manter o cliente trará um potencial a empresa e controle sobre seus lucros e para o mesmo Nepomucena (2023) enfatiza que o uso de *marketing*, enfocando na criação de valor nas relações com o cliente; banco de dados para gerenciar informações do cliente; *marketing* de relacionamento para manter a satisfação do cliente; e *Customer Relationship Management* (CRM), que significa Gerenciamento da Relação com o Cliente.

Em relação ao último Nepomucena (2023, p. 101) ainda destaca:

> O CRM é uma ferramenta que vai muito além de apenas conhecer o cliente, ela se dedica em aprofundar seu conhecimento sobre o consumidor, suas necessidades e desejos, a fim de realmente satisfazer o cliente, transformando essa relação em fidelidade com marca, facilitar na previsão de hábitos dos clientes e auxilia na administração do relacionamento com o mesmo.

Conforme Milan (2008, p. 153), “Analisar apenas a taxa de deserção não é suficiente, mas sim entender quais motivos levam o cliente a trocar de fornecedor”. As estratégias de retenção de clientes são muitos benéficas a empresa e em trabalho

conjunto com a análise das causas de perda de clientes, geram uma gestão direcionada as lacunas que a empresa tem em seu gerenciamento atual.

5. Metodologia

Para o desenvolvimento inicial do protótipo serão utilizadas quatro etapas: PLN, AED, análise diagnóstica e mitigação das causas de perda de clientes. Será utilizado dados fictícios que simulam informações de uma empresa de desenvolvimento de software especializada em sistemas de *Enterprise Resource Planning* (ERP). Esses dados incluem registros detalhados sobre os motivos de cancelamento do serviço pelos clientes, tendo um total de 538 registros e os mesmos tem sua classificação feita manualmente para efeito de comparação com os tópicos que serão classificados pela etapa PLN.

A primeira etapa, denominada PLN, terá como abordagem central a classificação dos registros em tópicos, com a utilização do Pandas, para leitura e gravação dos dados, e o Gemini API, que realizará a análise e classificação através de um *prompt* de comando, contendo um contexto para cada possível tópico pertencente a frase que está sendo analisada. Esses contextos são definições adaptadas em relação aos motivos comuns para a perda de clientes, conforme descrito anteriormente na seção Motivos Comuns para a Perda de Clientes, essa adaptação se deve ao fato de ser necessário simplificar e ser mais direto para fácil entendimento da API, os mesmos foram analisados e considerados mais importantes dentro do contexto dos dados a serem examinados, conforme estão dispostos a seguir:

- Preço: Inclui a busca por alternativas mais baratas, com preços mais competitivos que levam os clientes a mudar para concorrentes. Isso envolve taxas, tarifas de serviço, e custos gerais percebidos como injustos.
- Qualidade do Serviço: Abrange falhas no serviço principal, como erros em transações ou entrega de produtos abaixo do esperado, e nos encontros de serviço, como atendimento ineficiente ou

comportamentos inadequados de funcionários, bem como a incapacidade de resolver problemas e reclamações de forma eficaz.

- Concorrência: Clientes migram atraídos por promoções, inovação tecnológica ou serviços mais sofisticados oferecidos por concorrentes. Isso inclui a atração por concorrentes novos com serviços mais modernos e atrativos.
- Mudanças Pessoais: Ocorre quando clientes mudam de cidade, encerram suas atividades ou deixam de necessitar dos serviços.

Em seguida através do Pandas será adicionado ao *prompt* 10 registros com os motivos a serem classificados e após a classificação os tópicos pertencentes a cada motivo será registrado na coluna a frente.

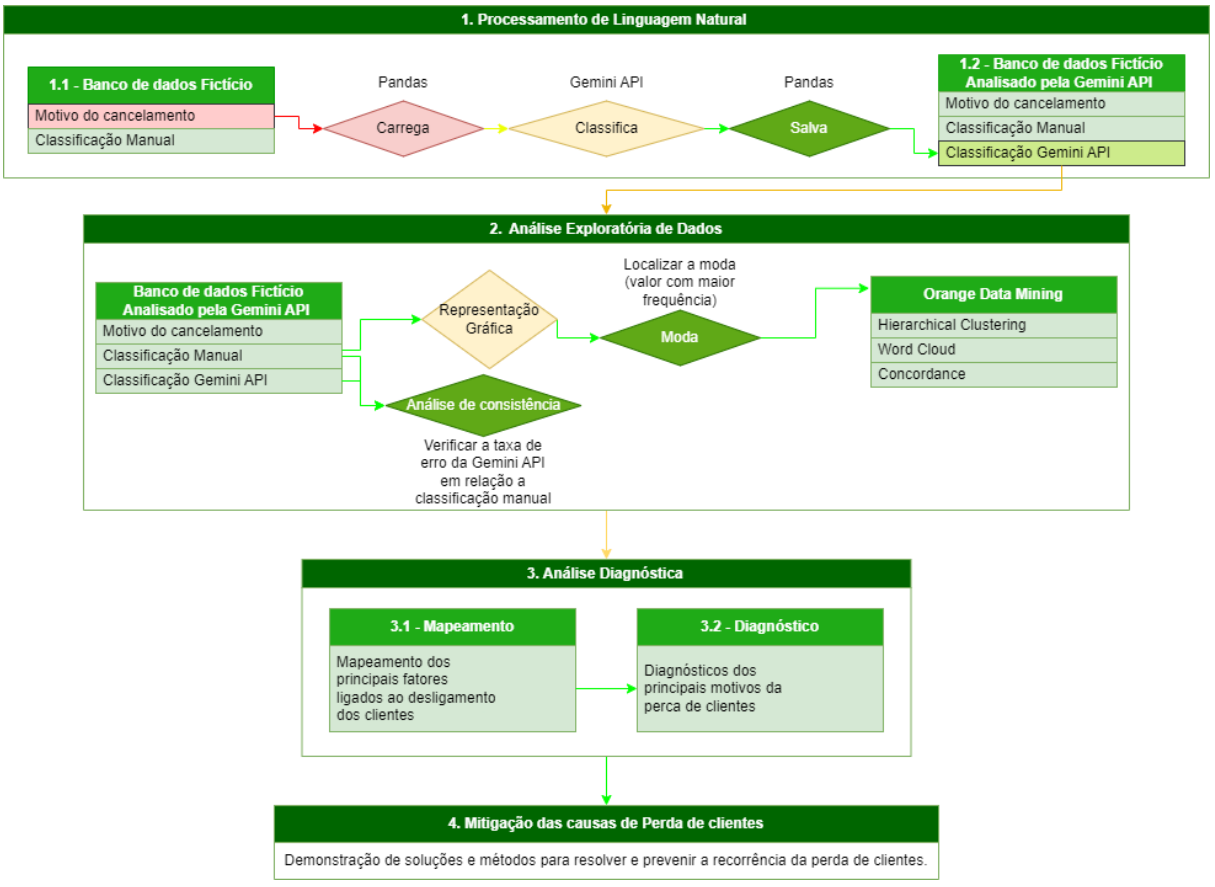
Através do banco de dados classificado será realizado para avaliar o desempenho da Gemini API uma análise de consistência, comparando os tópicos classificados manualmente pelo autor com os da API, resultando na taxa total de precisão. Em seguida será feito a AED com a criação de um gráfico de barras para representar as classificações, também será realizado a Moda, conforme definido Guedes *et al.* (2005, p. 31), “A moda é o valor que apresenta a maior freqüência da variável entre os valores observados. Para o caso de valores individuais, a moda pode ser determinada imediatamente observando-se o rol ou a freqüência absoluta dos dados”, dessa forma identificando qual tópico contém mais registros e com o Orange Data Mining se realizado uma nova analise com os principais registros dentro desse tópico, pois apesar de já se saber qual a maior causa de cancelamento, ainda se torna necessário identificar quais os principais motivos dentro dessa causa.

Serão utilizados três *widget* o *Hierarchical Clustering*, que atuara no agrupamento dos registros por similaridade textual, o *Word Cloud*, que permite verificar quais palavras são mais recorrentes dentro dos grupos do *Hierarchical Clustering*, e o *Concordance*, que permite buscar uma palavra do *Word Cloud* dentro do *Hierarchical Clustering* e verificar o contexto e numero de vezes que a palavra apresenta dentro do grupo, desse modo se torna possível identificar dentro do tópico quais os principais motivos dos cancelamentos, permitindo uma análise completa e interativa.

No processo da análise diagnóstica, será realizado o mapeamento dos principais fatores ligados ao desligamento dos clientes, conforme os dados gerados pelo Orange Data Mining, em seguida através do mapeamento será realizado os diagnósticos. Para Minayo e Deslandes (2007, p. 32) “este tipo de pesquisa está comprometido com respostas para o presente do sistema, buscando o aperfeiçoamento das ações a favor da população atendida”.

Por fim a partir dos métodos e soluções encontradas na primeira etapa, junto a análise diagnosticas será realizado a mitigação das causas de perda de clientes, conforme a Figura 1, que representa as etapas que serão seguidas:

Figura 1 - Proposta das etapas de desenvolvimento do PLN, AED, análise diagnóstica e mitigação das causas de perda de clientes.



Fonte: Do autor, 2024

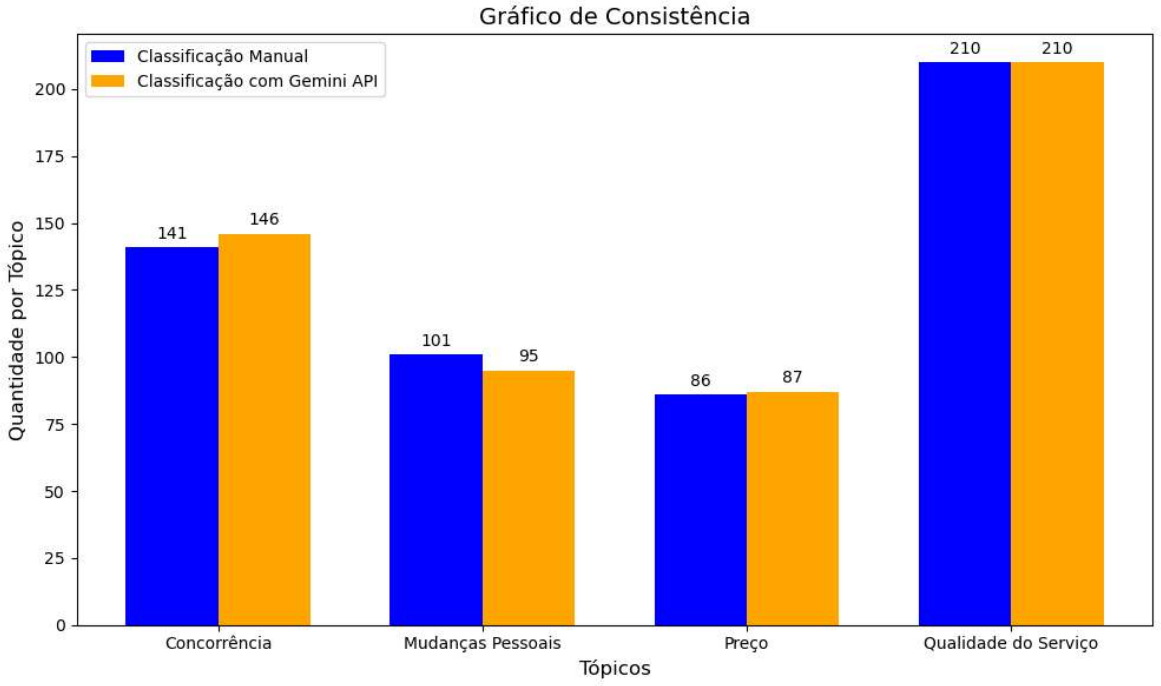
Mediante as etapas descritas e ilustradas, a aplicação prática destas técnicas visa não apenas entender os motivos de cancelamento, mas também fornecerá soluções práticas para aumentar a retenção de clientes, além de oferecer

informações e meios para a realização de PLN em outras aplicações.

6. Resultados e Discussão

A utilização de Inteligência Artificial para a realização de atividades já se tornou uma prática rotineira. Com o avanço das novas LLMs (Modelos de Linguagem de Grande Escala), cada vez mais acessíveis, seu uso no dia a dia para auxiliar ou até mesmo automatizar tarefas tornou-se comum em diversas áreas. A aplicação do Gemini API, por exemplo, mostrou-se extremamente eficaz na classificação de causas de perda de clientes. Conforme apresentado na Figura 2, que compara o resultado da classificação automatizada com uma análise manual:

Figura 2 – Gráfico de Consistência, comparando os valores da classificação manual com a realizado por Gemini API, gerado pelo Orange Data Mining.



Fonte: Do autor, 2024

O Gemini API alcançou uma taxa 92,93% de consistência, com um total de 38 classificações divergentes somente, além disso, conseguiu realizar a classificação dos 538 registros em 304,92s ou 1,76s por registro, em comparação com a análise manual

realizada pelo próprio autor, que demorou aproximadamente um total de 2h para os 538, ou 13,38s por registro. Se utiliza a Tabela 1 citada por Serrano (2019, p. 131)[1]:

Tabela 1 – Probabilidades de erro humano

Tarefa:	Descrição	5%	50%	95%
Não familiar	Tarefa não familiar executada à mesma velocidade do que se fosse uma tarefa habitual.	0,35	0,55	0,97
Mudança sem procedimentos	Mudar ou restaurar o sistema para um novo estado ou estado original numa única tentativa sem supervisão e/ou sem procedimentos definidos.	0,14	0,26	0,42
Complexa	Tarefa complexa que requira um alto nível de compreensão e destreza.	0,12	0,16	0,28
Simples	Tarefa simples executada muito rapidamente e/ou sem a devida atenção.	0,06	0,09	0,13
Rotineira, pouco exigente	Tarefa rotineira, praticada muitas vezes e de execução rápida que envolva um nível relativamente baixo de competências.	0,007	0,02	0,045
Restaurar com procedimentos	Mudar ou restaurar o sistema para um novo estado ou estado original seguindo os procedimentos, com alguma supervisão.	0,0008	0,003	0,007
Familiar, praticada	Tarefa completamente familiar, bem desenhada, altamente praticada e rotineira, que ocorra várias vezes numa hora, executada segundo as regras mais exigentes e por pessoas altamente motivadas, bem formadas e altamente experientes, totalmente conscientes das implicações da falha, com tempo para corrigir possíveis erros, mas sem	0,00008	0,0004	0,009
Responder corretamente	Responder corretamente a comandos do sistema mesmo quando existe um sistema supervisor aumentado ou automático que providencie uma detalhada interpretação do estado do sistema.	0,000006	0,00002	0,0009
Várias	Tarefas várias para as quais não foi encontrada nenhuma descrição	0,008	0,03	0,11

Fonte: Serrano (2019, p. 131)

Em um cenário de probabilidade de erro médio, sendo uma tarefa simples podemos considerar que a margem de erro tanto para a classificação manual de até 9%, e considerando a classificação realizada pelo Gemini API, resultou somente em 7,07% de erro, estando ainda abaixo do esperado, quando feita por um humano.

[1] Citação faz parte do Anexo C. HEART: Tabelas Auxiliares de Cálculo

Por meio da análise de moda, foi constatado que a principal causa de perda de clientes está relacionada à Qualidade dos Serviços, com 210 registros. Dentro dessa categoria, utilizando técnicas de *Hierarchical Clustering*, *Word Cloud* e *Concordance*, foi possível identificar as palavras mais importantes e o contexto em que cada uma se insere. Esse processo permitiu o mapeamento dos principais fatores associados ao desligamento de clientes e o diagnóstico das principais causas de perda de clientes, que incluem:

- Problemas de instabilidade, inconsistência: perda de informações ou dados decorrente de falhas no próprio sistema e demorar em processos realizados diariamente.
- Suporte: falta de clareza nas informações, demora no atendimento e ineficácia na resolução de problemas.

Para resolver problemas de instabilidade ou inconsistência, é essencial realizar testes regulares de stress e carga, identificando possíveis pontos de falha em condições de alto tráfego. Além disso, monitorar clientes que relatam os mesmos erros pode revelar se os problemas são de natureza local, como comunicação ou hardware, ou se realmente estão relacionados ao sistema ERP.

A criação de ambientes de teste dedicados é igualmente fundamental, permitindo verificar constantemente novas versões do sistema e identificar erros. Analisar os pontos de alta utilização no código-fonte também contribui para a prevenção, especialmente ao avaliar a eficiência do algoritmo com anotações de complexidade Big O, essa análise permite identificar possíveis gargalos e lentidões nos processos, orientando melhorias no desempenho geral do sistema.

Para resolver os problemas identificados no suporte ao cliente, é essencial investir em um programa contínuo de treinamento e capacitação, capacitando a equipe para responder de forma rápida e precisa às solicitações. Esse treinamento deve incluir simulações de atendimento, permitindo que os agentes lidem com situações comuns de forma prática e eficiente.

Além disso, implementar um Sistema de Gestão de Conhecimento é crucial para que o suporte tenha acesso rápido a respostas e soluções padronizadas, facilitando a resolução de problemas. Coletar *feedback* dos clientes após o

atendimento e realizar análises periódicas dos dados obtidos com métricas também contribui para melhorias constantes e um atendimento mais eficiente e resolutivo.

Para assegurar que a equipe de suporte tem capacidade suficiente para atender à demanda, é importante realizar análises periódicas de carga de trabalho e volume de chamados. Esse processo inclui monitorar o número médio de atendimentos realizados por agente, o tempo médio de espera dos clientes e a taxa de resolução dentro dos prazos estabelecidos. Ferramentas de análise de demanda, como *dashboards* em tempo real, permitem identificar picos de solicitações e períodos de maior carga, possibilitando ajustes na equipe para manter o atendimento ágil e eficiente.

Além disso, uma revisão trimestral da quantidade de suporte em relação ao crescimento da base de clientes e à complexidade dos chamados pode ajudar a decidir se é necessário aumentar o quadro de funcionários. Essas estratégias garantem que o suporte não apenas atenda à demanda, mas também mantenha a qualidade e a satisfação do cliente em alto nível.

7. Conclusão

A análise diagnóstica juntamente com as técnicas de PLN, revelou-se uma ferramenta essencial para identificar e compreender as causas subjacentes da perda de clientes. A aplicação de modelos avançados como o Gemini API, aliada a metodologias como AED e ferramentas como *Orange Data Mining*, demonstrou-se eficaz para classificar e mapear os fatores críticos de desligamento.

Esses métodos possibilitam não apenas um diagnóstico preciso, mas também uma base sólida para implementar estratégias preventivas e corretivas, proporcionando à empresa uma visão aprofundada dos desafios de retenção e direcionando ações para garantir a satisfação e lealdade dos clientes, fortalecendo assim a competitividade e a sustentabilidade do negócio.

Referências

BARBOSA, J., *et al.* Introdução ao processamento de linguagem natural usando

python. In: Escola Regional de Informatica do Piauí, v. 1, n. 1, 2017, Picos. **Livro Anais**. Picos: ERIPI, 2017. p. 336-360. Disponível em: https://www.facom.ufu.br/~wendelmelo/terceiros/tutorial_nltk.pdf. Acesso em: 20 maio 2024.

BATESON, J. E. G.; HOFFMAN, K. D. **Princípios de Marketing de Serviços:** Conceitos, estratégias e casos. Tradução: Cristina Bacellar. São Paulo: Cengage Learning, 2016. E-book. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9788522124039/. Acesso em: 20 abr. 2024.

BOLÍVAR, D. M. C.; CASTRO, J. A. B.; HERREÑO, P. A. B. ***Análisis de factores internos y externos que influyen en el desarrollo de las Mipymes del sector comercial en Girardot***. 2020. Tese de Doutorado – Corporación Universitaria Minuto de Dios, Girardot, 2020. Disponível em: .Acesso em: https://repository.uniminuto.edu/items/5ec0dd8c-93a9-447e-a434-6f33eddcfcdc/. 21 nov. 2024.

COUTO, A. F. **Satisfação e lealdade de clientes:** Índice NPS. 2015. Monografía (Especialista em Pesquisa de Mercado) – Escola de Comunicações e Artes da Universidade de São Paulo – ECA/USP, São Paulo, 2015. Disponível em: https://pospesquisa.eca.usp.br/monografias/Amanda%20Falchetti%20Couto.pdf. Acesso em: 20 maio. 2024.

FERREIRA, R. G. C.; MIRANDA, L. B. A. de; PINTO, R. A.; *et al.* **Preparação e Análise Exploratória de Dados**. Porto Alegre: Grupo A, 2021. E-book. ISBN 9786556902890. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786556902890/. Acesso em: 18 set. 2024.

GUEDES, T. A. *et al*. Estatística descritiva. **Projeto de ensino aprender fazendo estatística**, p. 1-49, 2005. Disponível em: https://www.ime.usp.br/~rvicente/Guedes_etal_Estatistica_Descritiva.pdf. Acesso

em: 11 nov. 2024.

LAMBERT, K. A. **Fundamentos de Python: Primeiros programas**. Tradução: Edson Furmankiewicz. São Paulo: Cengage Learning Brasil, 2022. E-book. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786555584301/. Acesso em: 21 abr. 2024.

MACIEL, F. M. B. **Python e Django:** Desenvolvimento Web Moderno e Ágil. Rio de Janeiro: Alta Books, 2020. E-book. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786555200973/. Acesso em: 21 abr. 2024.

MILAN, G. S. Um estudo exploratório sobre a satisfação de clientes em um ambiente de serviços de fisioterapia. **Revista de Ciências da Administração**, Florianópolis, v. 10, n. 22, p. 146–170, 2008. Disponível em: https://periodicos.ufsc.br/index.php/adm/article/view/2175-8069.2008v10n22p146. Acesso em: 20 maio 2024.

MCFEDRIES, P. **Análise de dados com excel Para Leigos**. Rio de Janeiro: Alta Books, 2020. E-book. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786555201932/. Acesso em: 11 jun. 2024.

MINAYO, M. C. S.; DESLANDES, S.F., orgs. **Análise diagnóstica da política nacional de saúde para redução de acidentes e violências**. Rio de Janeiro: Editora FIOCRUZ, 2007, E-book. Disponível em: http://books.scielo.org/id/fx9hn/epub/minayo-9788575415412.epub. Acesso em: 20 maio 2024.

MORAES, L. C., *et al.* Análise de ambiguidade linguística em modelos de linguagem de grande escala (LLMs). **arXiv e-prints**, p. 9, 2024. Disponível em: https://arxiv.org/abs/2404.16653. Acesso em: 11 nov. 2024.

NEPOMUCENA, F. V. Marketing de relacionamento para retenção de clientes. **Revista Multidisciplinar do Nordeste Mineiro**, Teófilo Otoni, v. 1, n. 1, p. 91-108, 2023. Disponível em: https://revista.unipacto.com.br/index.php/multidisciplinar/article/view/377. Acesso em: 20 maio 2024.

OLIVEIRA, S. A. D. **Determinantes da deserção de clientes bancários**. 2018. Tese de Doutorado. (Mestre em Administração) – Curso de Mestrado em Administração da Faculdade de Ciências Empresariais da Universidade – FUMEC, Belo Horizonte, 2018. Disponível em: https://repositorio.fumec.br/handle/123456789/126. Acesso em: 20 maio. 2024.

RACHMAT, N.; KESUMA D. P. *Implementasi Large Language Models Gemini Pada Pengembangan Aplikasi Chatbot BerbasisAndroid*. **Jurnal Ilmu Komputer**, Gorontalo, v. 4, n. 1, p. 41, 2024. Disponível em: https://journal.umgo.ac.id/index.php/juik/article/view/2831/1550. Acesso em: 21 nov. 2024.

RAPOSO, L. B. **Avaliação de LLMs na resolução de questões do ENEM**. 2024. Trabalho de Conclusão de Curso. (Bacharel em Ciência da Computação) – Curso Bacharelado em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande – UFCG, Campina Grande, 2024. Disponível em: .Acesso em: http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/38059. 21 nov. 2024.

REBACK, Jeff *et al.* pandas-dev/pandas: Pandas 1.0.5 (v1.0.5). **Zenodo**, 2020. Disponível em: https://doi.org/10.5281/zenodo.3898987. Acesso em: 21 out. 2024.

RODRIGUES, C. C. **A importância do marketing de relacionamento nas pequenas empresas:** uma revisão bibliográfica. 2020. Monografia (Bacharel em Administração) – Centro Universitário Atenas, Paracatu, 2021. Disponível em:

.Acesso em: https://www.atenas.edu.br/uniatenas/assets/files/spic/monography/A_IMPORTANCIA_DO_MARKETING_DE_RELACIONAMENTO_NAS_PEQUENAS_EMPRESAS_UMA_REVISAO_BIBLIOGRAFICA.pdf. 21 nov. 2024.

RODRIGUES, R. L. **Uma abordagem de *data mining* e *business intelligence* para otimização do processo produtivo e apoio à tomada de decisão na agroindústria**. 2023. Monografia. (Bacharel em Engenharia de Produção) – Curso Superior de Engenharia de Produção do Centro de Desenvolvimento Sustentável do Semiárido da Universidade Federal de Campina Grande – UFCG, Sumé, 2023. Disponível em: http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/29114.Acesso em: 21 nov. 2024.

SARKER, I.H. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. 2021. **SN Computer Science**, Online, v. 2, n. 377, p. 1-22, 2021. Disponível em: https://doi.org/10.1007/s42979-021-00765-8. Acesso em: 20 maio 2024.

SERRANO, E. J. M. Conceito, **Classificação e Quantificação da Fiabilidade Humana na Relação Homem-Máquina**. Dissertação de Mestrado para obtenção do grau de Mestre em Engenharia Mecânica, ISEL - INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA Departamento de Engenharia Mecânica. 2009. Disponível em: https://repositorio.ipl.pt/bitstream/10400.21/434/1/Dissertação.pdf . Acesso em: 05 nov. 2024.

SPAREMBERGER, A. **Vendas:** Fundamentos e Relacionamento com os Clientes. Ijuí: Unijuí, 2019. E-book. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9788541903073/. Acesso em: 11 jun. 2024.