

Rede Neural Convolutacional aplicada na Análise de Sentimentos em Comentários de Clientes de Empresa do Ramo Varejista

Convolutional Neural Network applied to Sentiment Analysis in Customer Comments of a Retail Company

Maria Sheila Carneiro Mestre em Informática e Gestão do Conhecimento. UNINOVE –
<https://orcid.org/0000-0002-6890-2276> Brasil. msheilacarneiro@uni9.edu.br

Dacyr Dante de Oliveira Gatto Doutor em Informática e Gestão do Conhecimento. UNINOVE –
Brasil. dacyr.gatto@uni9.pro.br
<https://orcid.org/0000-0003-2146-4819>

Renato José Sassi Doutor em Engenharia Elétrica. EPUSP – Brasil. sassi@uni9.pro.br
<https://orcid.org/0000-0001-5276-4895>

Marcos Antonio Gaspar Doutor em Administração – USP – Brasil.
<https://orcid.org/0000-0002-2422-2455> marcos.antonio@uni9.pro.br

RESUMO

A Análise de Sentimentos é utilizada no processamento de linguagem humana para detectar opiniões, emoções ou atitudes positivas, negativas ou neutras em textos. Ela permite que empresas compreendam as emoções expressas pelos clientes por trás das palavras ao abrir possibilidades que vão desde melhorias no atendimento até a apoiar a gestão do cliente. A Rede Neural Convolutacional (RNC) é uma técnica da Inteligência Artificial (IA) aplicável a este tipo de análise. Assim, o objetivo deste trabalho foi aplicar RNC na Análise de Sentimentos em comentários de clientes de uma empresa do ramo varejista para apoiar a gestão do cliente. Em adição, comparou-se os resultados dos experimentos com os resultados dos indicadores detratores do NPS (*Net Promoter Score*) da empresa. Para atingir o objetivo acima foi selecionada uma base de dados de atendimento aos clientes. Os principais resultados da análise estão relacionados a determinados aspectos dos produtos e serviços da empresa, dentre os quais destacam-se: cartão, limite, pagamento, aumento e dificuldade de atendimento nos canais disponibilizados pela empresa. Quanto ao cruzamento dos atributos relativos ao perfil do cliente com os resultados de NPS foi possível identificar os comentários e principais argumentos segregados por gênero, faixa etária, nível de renda e localidade de domicílio de clientes com NPS detratores. Conclui-se que foi possível conhecer o cliente a partir dos comentários elaborados pelos mesmos e que a solução desenvolvida é capaz de auxiliar na gestão do cliente em empresas varejistas.

Palavras-chave: avaliação de clientes; análise de sentimentos; varejo; *Net Promoter Score*; descoberta de conhecimento.

ABSTRACT

Sentiment Analysis is used in natural language processing to detect positive, negative, or neutral opinions, emotions, or attitudes in texts. It allows companies to understand the emotions expressed by customers behind the words, opening possibilities ranging from customer service improvements to supporting customer management. The Convolutional Neural Network (CNN) is an Artificial Intelligence (AI) technique applicable to this type of analysis. Thus, the objective of this study was to apply CNN in Sentiment Analysis on customer comments from a retail company to support customer management. Additionally, the results of the experiments were compared with the company's NPS (Net Promoter Score) detractor indicators. To achieve the above objective, a customer service database was selected. The main results of the analysis are related to certain aspects of the company's products and services, including card, limit, payment, increase, and difficulty in accessing the company's service channels. Regarding the correlation between customer profile attributes and NPS results, it was possible to identify comments and main arguments segmented by gender, age group, income level, and location of residence of customers with NPS detractors. It was concluded that it was possible to understand the customer based on the comments made by them and that the developed solution can assist customer management in retail companies.

Keywords: customer reviews; sentiment analysis; retail; Net Promoter Score; knowledge discovery.

Recebido em 21/09/2024. Aprovado em 04/02/2025. Avaliado pelo sistema *double blind peer review*. Publicado conforme normas da ABNT.
<https://doi.org/10.22279/navus.v16.2028>

1 INTRODUÇÃO

Para se destacarem no mercado competitivo, as empresas que atuam no varejo digital precisam adotar estratégias de diferenciação que utilizem a Inteligência Artificial (IA). Ela disponibiliza ferramentas eficazes para aprimorar a experiência do cliente, otimizar processos operacionais e se ajustar rapidamente às mudanças, permitindo que as empresas se estabeleçam como líderes no setor (Baehre et al., 2022).

Javaid et al. (2022) afirmam que a IA pode proporcionar oportunidades significativas para inovação e aprimoramento no varejo digital, ao apoiar e gerar novas formas de se relacionar com o cliente das mídias sociais. O uso da IA permite a criação de experiências de compra personalizadas que sugerem produtos com base no histórico e nas preferências dos clientes.

Assim, conhecer os clientes é fundamental para o sucesso de uma empresa que atua no varejo digital, isto porque pode proporcionar a adaptação de produtos e serviços às suas necessidades. Uma forma de viabilizar tal desafio é realizar a análise dos sentimentos dos clientes buscando extrair informações úteis para a geração de conhecimento sobre ele (Wankhade; Rao; Kulkarni, 2022).

A Análise de Sentimentos, conforme Chandrasekaran, Nguyen e Hemanth (2021) é uma área da IA que facilita a identificação e quantificação de emoções em textos e mídias, ajudando a entender a opinião pública, orientar decisões de *marketing* e melhorar a interação com clientes digitais.

Na visão dos autores Meena, Mohbey e Kumar (2023) a Análise de Sentimentos tornou-se essencial para as empresas, pois revela as percepções dos clientes sobre produtos e serviços, fortalece relacionamentos e fidelidade, aprimora o atendimento ao cliente e apoia estratégias de no varejo digital.

A Rede Neural Convolutacional (RNC) é uma técnica da IA que pode ser aplicada na análise de sentimentos, devido à sua capacidade de extrair e representar características textuais de maneira eficiente e eficaz (Cao et al., 2021).

De acordo, com Klaus e Maklan (2021), para avaliar a satisfação e a fidelidade do cliente, as empresas utilizam a metodologia NPS (*Net Promoter Score*). Ainda segundo o autor, o principal objetivo da metodologia NPS é definir uma pergunta simples que possa ajudar as organizações a construírem relacionamentos e satisfação com seus clientes, de modo a proporcionar à empresa o gerenciamento do relacionamento com seus clientes e planejar estrategicamente seu crescimento. A aplicação da ferramenta NPS ilustra a relação diretamente proporcional entre qualidade do relacionamento/serviço e crescimento organizacional.

Comparar os resultados obtidos da aplicação de uma técnica da IA com uma metodologia que avalia a satisfação do cliente é importante para avaliar se as respostas dos experimentos são semelhantes as respostas obtidas a partir da aplicação da metodologia. Esta comparação pode fornecer uma visão mais completa e detalhada do sentimento dos clientes, ajudando a tomar decisões mais informadas e a melhorar a experiência do cliente de forma mais eficaz.

Assim, com base no contexto apresentado, o objetivo deste trabalho foi aplicar RNC na Análise de Sentimentos em comentários de clientes de uma empresa do ramo varejista para apoiar a gestão do cliente.

2 REFERENCIAL TEÓRICO

Nesta seção são descritos os conceitos fundamentais da Análise de Sentimentos e das RNCs.

2.1 Análise de Sentimentos

As mídias sociais possibilitam às empresas uma comunicação direta e imediata com os clientes, promovendo interações personalizadas que fortalecem o relacionamento. Elas fornecem um canal digital valioso para *feedback* e *insights*, ajudando a ajustar estratégias de *marketing* e a desenvolver novos produtos. As plataformas digitais oferecem dados e análises para otimizar campanhas em tempo real, possibilitam experiências personalizadas e permitem um gerenciamento eficaz de crises e reputação (Dwivedi *et al.*, 2021).

Segundo Antonelo e Lima (2021), as mídias sociais fortalecem a conexão direta e personalizada com os consumidores, possibilitando o engajamento e a lealdade à marca. Além disso, campanhas e promoções nas redes sociais podem aumentar as vendas, e o *feedback* obtido oferece *insights* valiosos sobre o comportamento dos consumidores. Uma presença ativa nas mídias sociais também contribui para consolidar a identidade da marca, promovendo seus valores e diferenciando-a da concorrência.

Entretanto, a geração de grandes quantidades de dados não é, por si só, suficiente para que a empresa compreenda melhor seu cliente. Isso ocorre porque, quando analisados isoladamente ou em seu formato bruto, esses dados não são capazes de gerar conhecimento significativo para a empresa.

Não obstante, o conjunto desses dados tem enorme potencial para geração de conhecimentos. Jahani, Jain e Ivanov (2023), afirmam que essas análises podem ser realizadas por meio de técnicas de IA essenciais para transformar volumes de dados em informações úteis e acionáveis, apoiando a gestão do cliente.

Nesse contexto, de acordo com Nandwani e Verma (2023), a Análise de Sentimentos, uma área da IA possibilita que empresas entendam as emoções do cliente em relação a produtos, serviços e marcas. Esse entendimento auxilia na adaptação das estratégias e na melhoria da satisfação do cliente.

A Análise de Sentimentos permite identificar áreas para inovação e aprimoramento, além de monitorar a percepção pública em tempo real para respostas rápidas e detectar tendências para estratégias de marketing. Também fornece dados qualitativos que complementam análises quantitativas, apoia pesquisas acadêmicas e sociais, e permite a personalização das interações com clientes (Liu, 2023).

Dessa forma, muitas empresas têm valorizado este tipo de análise de opinião, não apenas por sua rentabilidade, mas também porque os resultados obtidos são altamente relevantes para a aceitação de um produto, serviço ou imagem corporativa. Assim, os dados são coletados sem a necessidade de entrevistas diretas com os autores dos comentários, o que facilita a coleta e a interpretação das opiniões postadas sobre um tema específico (Cho *et al.*, 2020).

Os dados coletados podem ter diferentes origens e devem ser armazenados e analisados posteriormente. Na etapa de pré-processamento, os dados são verificados e padronizados por meio de algoritmos que organizam as palavras e termos presentes nos textos produzidos pelos clientes. Wankhade, Rao e Kulkarni (2022) abordam três níveis principais de análise de sentimentos:

- **Nível de documento:** avalia o sentimento geral de um texto completo, oferecendo uma visão ampla das emoções predominantes, mas pode não captar detalhes específicos.
- **Nível de sentença:** analisa sentimentos em sentenças individuais, proporcionando uma visão detalhada e contextual das emoções expressas.
- **Nível de entidade e aspecto:** examina sentimentos sobre entidades e características específicas, permitindo uma análise detalhada dos aspectos destacados em um texto.

A avaliação dos sentimentos expressos em textos ou comunicações pode identificar padrões que preveem dificuldades ou sucessos futuros. São propostas, então, duas abordagens para a análise de sentimentos: uma do ponto de vista dos autores, que transmitem emoções e sentimentos através das palavras, e outra do ponto de vista dos leitores, que interpretam essas emoções e sentimentos ao ler o texto.

A avaliação dos sentimentos expressos em textos ou comunicações pode identificar padrões que podem prever dificuldades ou sucessos futuros. Dessa forma, são propostas duas abordagens para a análise de sentimentos: uma do ponto de vista dos autores, que transmitem emoções e sentimentos por meio das palavras, e outra do ponto de vista dos leitores, que interpretam essas emoções e sentimentos ao lerem o texto.

De acordo com Kit e Mokji (2022), a Análise de Sentimentos desempenha um papel fundamental na compreensão das opiniões e emoções dos usuários em relação a produtos e serviços, possibilitando que as empresas melhorem a experiência do cliente e ajustem suas estratégias com base no *feedback* recebido. Esse processo também contribui para o monitoramento da percepção da marca e a gestão da reputação *online*, além de identificar tendências e padrões nas opiniões ao longo do tempo.

A utilização de modelos pré-treinados para a análise automatizada economiza tempo e recursos, tornando o processo mais eficiente e viável para análises em larga escala. Além disso, as informações obtidas auxiliam no alinhamento das decisões estratégicas às necessidades e expectativas dos clientes. Por esse motivo, é importante considerar o uso de uma técnica da IA, como a Rede Neural Convolutacional (RNC), na análise de sentimentos (Meena, Mohbey e Kumar, 2023).

2.2 Redes Neurais Convolutacionais

Redes Neurais Convolutacionais (RNCs) são um tipo de arquitetura de rede neural artificial empregada na classificação de padrões bidimensionais, sendo compostas por três camadas principais, cada uma responsável por funções matemáticas específicas que aprimoram a precisão da classificação final (Kim; Park, 2024).

De acordo com Raiaan *et al.*, (2024), as RNCs são formadas por várias camadas convolutacionais que extraem características como bordas e texturas, nas primeiras camadas, e características como formas e padrões, nas camadas mais profundas. Esses filtros destacam aspectos importantes, como cor e textura, indicativos do grau de maturação.

Cada imagem passa por processos como convolução com filtros (*Kernels*), *pooling*, camadas totalmente conectadas e a função ReLU. Quando aplicadas a textos, como opiniões de clientes, essas técnicas exigem a conversão dos textos em matrizes numéricas para viabilizar a convolução.

A camada de agrupamento reduz as dimensões dos mapas de características utilizando funções como *Max Pooling* e *Average Pooling*, que identificam os valores máximos e médios, respectivamente. Isso resulta em uma redução pela metade da altura e largura da imagem original, permitindo a compactação de cada bloco de 4 pixels em um único pixel sem perda significativa de dados (Raiaan et al., 2024).

A camada totalmente conectada, também conhecida como *Fully Connected Layer* (FCL), é responsável pela classificação dos dados utilizando funções como *Softmax* para múltiplas classes e *Sigmoid* para classes binárias. Nessa etapa, a matriz é "achatada" em um vetor, e os vetores resultantes são processados pela camada totalmente conectada para formar o modelo. Finalmente, a função de ativação *Softmax* ou *Sigmoid* é aplicada para classificar as saídas (Raiaan et al., 2024).

Uma revisão sistemática de técnicas de otimização de hiperparâmetros para RNCs é apresentada no trabalho de Raiaan, et al. (2024).

A aplicação eficaz da RNC na análise de sentimentos em comentários pode ser comprovada nos trabalhos de Alawi e Bozkurt (2024), Sharbatian e Moattar (2023), Li, et al., (2023), Wang, et al., (2021), Nandwani e Verma (2021) e Wankhade, et al., (2022).

RNCs também apresentam bons resultados quando aplicadas em sistemas de recomendação (Almaghrabi; Chetty, 2020), na redução da lacuna entre as experiências de varejo *online* e *offline* (Srivastava; et al., 2024), na previsão de preços do mercado varejista (Kim; Park, 2024), e na previsão de vendas de automóveis (Sivabalan; Minu, 2025).

3 MÉTODOS E INSTRUMENTOS DE PESQUISA

Nesta seção a tipologia de pesquisa, a estrutura da base de dados de atendimento e as fases que compõem a metodologia experimental são apresentadas.

3.1 Tipologia da Pesquisa

A tipologia de pesquisa é composta pelas pesquisas bibliográfica, exploratória e experimental. A pesquisa bibliográfica foi realizada com base em consultas a fontes de referência bibliográficas e teóricas, tais como artigos, livros, teses, dissertações e sites com conteúdo sobre análise de sentimentos e descoberta de conhecimento do cliente. Foram consultadas as seguintes bases de dados: *SciELO*, *IEEE Xplore*, *Scopus* e *Google Scholar*.

3.2 Estrutura da Base de Dados de Atendimento

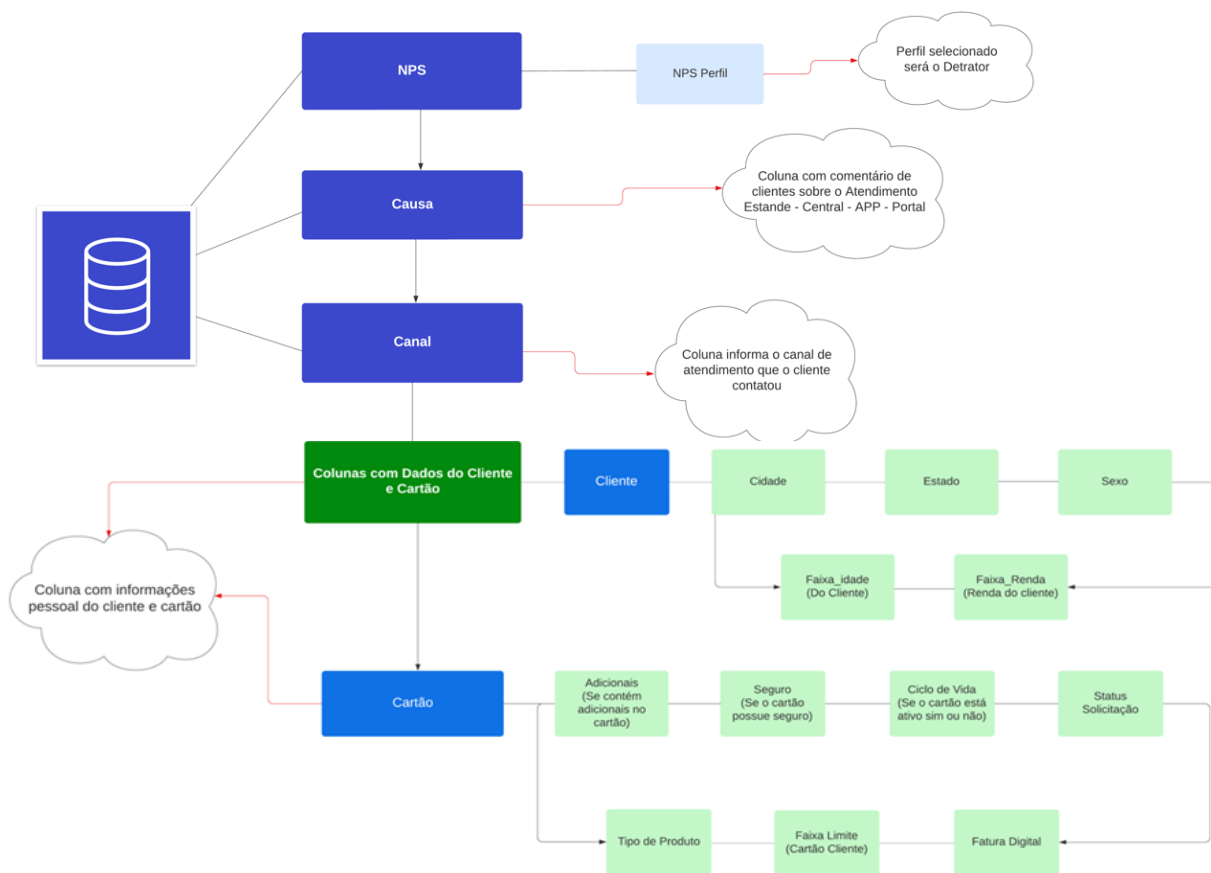
A empresa varejista que forneceu seus dados para a realização deste trabalho possui um sistema interno com registros de comentários de clientes sobre produtos e serviços, os quais são importados das mídias sociais e armazenados sem análise. A organização permitiu o uso da base de dados, mas não autorizou a divulgação do seu nome e marca. Os dados utilizados neste trabalho foram anonimizados, ou seja, foram removidas informações pessoais que permitiriam a associação direta ou indireta a um indivíduo.

Para aprimorar produtos e serviços, diversos canais são utilizados para coletar *feedback* dos consumidores. O banco de dados inclui comentários obtidos por meio de mídias sociais, central de atendimento, lojas físicas, aplicativos

e portais, todos registrados seguindo a metodologia NPS. Esses comentários, que podem ser positivos, negativos ou neutros, muitas vezes contêm erros de ortografia e concordância, dificultando interpretação.

Na Figura 1, apresenta-se as categorias de dados e suas correlações que formam a base de atendimentos. Essas correlações foram estabelecidas com a ajuda de um especialista da empresa, por meio de reuniões *online* e encontros presenciais, que foram fundamentais para compreender a estrutura da base de dados, resolver dúvidas, e discutir os resultados e as necessidades da solução proposta.

Figura 1 - Estrutura da base de dados de atendimento



Fonte: Autores (2024)

Assim, as correlações descritas a seguir têm relevância prática para a empresa, visando proporcionar um atendimento mais eficaz ao cliente com base no conhecimento obtido sobre os perfis específicos.

As quatro principais correlações apontadas como mais expressivas pelo especialista são as seguintes:

- 1) Correlação entre a Causa (comentário do cliente) X Perfil do Cliente, cruzando os atributos sexo, idade, renda e estado.
- 2) Correlação entre a Causa (comentário do cliente) X Produto cartão, cruzando com os atributos faixa limite, tipo de produto, ciclo de vida e adicional.
- 3) Correlação entre a Causa (comentário do cliente) X Canais de atendimento, cruzando com os atributos central de atendimento estandes/lojas, aplicativos, portais e status da solicitação.

4) Correlação entre Causa (comentário de cliente) X Estratificação de notas detratoras do NPS (notas de 0 a 6), segregadas em dois substratos: notas de 0 a 3 (detratores inferiores) e notas de 4 a 6 (detratores superiores).

Para conduzir os experimentos e validar a solução, foram apresentados apenas os resultados dos atributos da categoria perfil do cliente (sexo, idade, renda e estado) cruzados com NPS detratores. Embora outras categorias, como produto cartão e canais de atendimento, pudessem ser incluídas, isso não alteraria a conclusão de que a solução pode ser eficaz para o objetivo proposto.

3.3 Metodologia Experimental

A metodologia experimental foi dividida em cinco fases baseada no processo de Descoberta de Conhecimento em Bases de Dados (Fayyad, Piatetsky-Shapiro e Smyth, 1996): seleção da base de dados de atendimento, pré-processamento e transformação dos dados, aplicação da RNC, comparação dos resultados experimentais com os indicadores detratores do NPS e interpretação e avaliação dos resultados:

Fase 1: Seleção da base de dados de atendimento

A base de dados selecionada é do ano de 2021 com 17.547 registros e 59 atributos, contendo duas classes, masculino e feminino. Os dados são oriundos, dos *feedbacks* de clientes sobre produtos e serviços, coletados pela central de atendimento, lojas físicas, aplicativos e portais, além dos resultados do NPS. A empresa não autorizou a divulgação do seu nome.

Fase 2: Pré-processamento e transformação dos dados

Inicialmente, a base foi extraída pelo sistema *Quest Manager (QWST)* e exportada para uma planilha MSEXcel (.CSV). Utilizou-se o *Google Colaboratory (Colab)*, uma ferramenta do *Google Research*, para a leitura da base de dados. O *Colab* facilita a escrita e execução de código *Python* diretamente no navegador, sendo ideal para aprendizado de máquina e análise de dados. Foram instaladas e importadas no *Colab* as bibliotecas *NLTK*, *Spacy*, *Pandas*, *Random*, *Numpy*, *Re*, *Seaborn*, *Matplotlib* e *PIL*.

O pré-processamento da base incluiu a remoção de atributos, a limpeza, o tratamento dos textos, bem como a padronização dos tamanhos das sequências textuais. A remoção dos atributos foi realizada após análise conduzida por especialistas da empresa que forneceu a base. Esses especialistas identificaram quais atributos não eram relevantes ou essenciais para a análise, ou seja, aqueles que não contribuiriam para os resultados ou que poderiam gerar ruído nos dados. Essa avaliação permitiu a redução do conjunto de dados, mantendo apenas os atributos mais significativos.

Além disso, foi realizado um processo de limpeza e transformação dos textos, que envolveu a remoção de caracteres especiais, de *stopwords* que são palavras irrelevantes para a análise, de números e de pontuações. Para permitir que os textos fossem compreendidos pelas técnicas selecionadas, aplicou-se a tokenização, que consiste na segmentação das frases em palavras ou partes menores, e a vetorização de termos, que converte os textos em representações numéricas adequadas para processamento computacional.

Outro aspecto importante do pré-processamento foi a padronização do comprimento das sequências textuais, uma vez que os comentários possuíam tamanhos variados. Para resolver essa questão, utilizou-se a técnica de *padding*, que ajusta todas as sequências para um tamanho fixo, garantindo que

as técnicas consigam processar os dados de maneira uniforme. Por fim, os textos foram transformados em vetores numéricos.

No caso das notas detratoras do NPS, focou-se nos comentários dos clientes da coluna 'Causa'. Apenas os comentários associados aos detratores ou insatisfeitos que atribuíram notas de 0 a 6 no NPS foram selecionados para a realização dos experimentos computacionais.

Vale destacar que todas as perguntas e respostas presentes na base de dados foram fornecidas pelos clientes diretamente através dos canais de comunicação da empresa, como a central de atendimento, estandes ou lojas, aplicativos e portais.

O cálculo do NPS é feito da seguinte forma (Baquero, 2022):

1. Coleta de Dados: pergunta-se aos clientes o quão provável é que recomendem a empresa, usando uma escala de 0 a 10.

2. Classificação: os clientes são divididos em três grupos:

-Promotores (9-10): Altamente satisfeitos e propensos a recomendar.

-Neutros (7-8): Satisfeitos, mas não entusiásticos.

-Detratores (0-6): Insatisfeitos e podem prejudicar a reputação.

3. Cálculo do NPS: A fórmula é: $NPS = \frac{\text{Promotores} - \text{Detratores}}{\text{Número total de respondente}}$. O NPS varia de -100 a +100 e um valor positivo indica lealdade dos clientes.

Por exemplo, em uma pesquisa com 50 pessoas, 25 deram notas 9 e 10 (promotores), 20 deram notas 7 ou 8 (neutros) e 5 deram notas de 0 a 6 (detratores). Nesse sentido, o cálculo de NPS deve ser: $\frac{25 \text{ (promotores)} - 5 \text{ (detratores)}}{50 \text{ (número total de pessoas que responderam)}} = 40\%$.

Destaca-se novamente que foram considerados apenas as notas dos detratores (0 a 6) para a realização dos experimentos. As notas dos neutros (7 e 8) foram excluídas por serem comentários ambíguos, e as notas dos promotores (9 a 10) não foram incluídas porque se referem a comentários favoráveis sobre produtos e serviços da empresa.

Após a remoção dos atributos em consenso com o especialista da empresa e considerando somente as notas detratoras, a base de dados utilizada nos experimentos ficou com 16 atributos e 17.547 registros.

Apresenta-se na Tabela 1, os nomes e as descrições dos atributos.

Tabela 1 - Nomes e descrições dos atributos selecionados

Número do atributo	Nome do atributo	Descrição do atributo
1	NPS	Notas de 0 a 6 atribuídas pelos clientes
2	Perfil do NPS	Detrator
3	Causa	Comentário do cliente
4	Canal	Informações sobre o canal de atendimento realizado por meio da central de atendimento, de estandes (lojas), de aplicativos ou portais
5 a 9	Cidade, Estado, Sexo, Faixa de idade, Faixa de renda	Dados pessoais dos clientes
10	Faixa limite	Limite do cartão de crédito
11	Adicional	Cliente possui cartão de crédito adicionais
12	Seguro	Cartão do cliente tem seguro oferecido pela empresa
13	Ciclo de vida	Cartão fidelidade está ativo, inativo ou nunca ativo
14	Fatura digital	Cliente ativo possui fatura digital
15	Tipo de produto	Cartão é nacional ou internacional
16	Status solicitação	Status do atendimento: automático, finalizado ou pendente

Fonte: Autores (2024)

Seguindo com o desenvolvimento das atividades das fases 1 e 2 foi necessário normalizar valores, ajustando-os para um intervalo específico. O processo de normalização incluiu a remoção de 'stopwords', caracteres especiais, stemização, lematização, tokenização e *padding* de palavras. De acordo com Benhar e Alemán (2020), a extração eficiente de recursos textuais, como anotações, é essencial para aprimorar a precisão dos modelos de classificação.

Os autores destacam a importância de limpar e normalizar textos, removendo ruído e informações irrelevantes, e de utilizar tokenização e normalização para padronizar o formato do texto. Além disso, o artigo menciona a aplicação de técnicas de redução de dimensionalidade, como a vetorização de termos, para transformar textos em representações numéricas mais adequadas para algoritmos de aprendizado de máquina.

Após essas etapas, os dados foram agrupados em um único local para a aplicação dos modelos de análise. Foi criada a representação *bag of words* com base no texto pré-processado.

Na sequência foram usadas ferramentas de classificação de texto para extrair conhecimento, começando pela identificação de palavras-chave. A ferramenta *Rake* foi utilizada para criar um *ranking* das principais palavras. Em seguida foram aplicadas as ferramentas *Gensim* e *Bertopic* para a representação de tópicos, além da medida TF-IDF (*Term Frequency - Inverse Document Frequency*).

Fase 3: Aplicação da RNC

O *hardware* utilizado foi um computador com placa de vídeo Graphics 620, processador Intel Core i7-8565U, SSD de 512GB, 16GB de memória e Windows 11 Home Single Language.

Os experimentos foram realizados com *Python 3.10.7* e *Google Colaboratory*, utilizando as bibliotecas *Scikit-learn* e *spaCy* para o processamento dos dados. A biblioteca TF-IDF (*Term Frequency-Inverse Document Frequency*) foi utilizada para avaliar a importância das palavras em documentos em relação a um corpus (Tietz et al., 2017).

A base de dados pré-processada e transformada foi dividida, para a realização dos experimentos, em três partes: treinamento, teste e validação. Os resultados da RNC foram comparados com duas técnicas de IA: a rede neural artificial do tipo *Multilayer Perceptron* (MLP) e a Árvore de Decisão, comumente aplicadas ao tipo de problema tratado neste trabalho. A escolha da técnica para dar continuidade aos experimentos recaiu sobre aquela com melhor desempenho, avaliado com base nas seguintes métricas: acurácia, precisão, *recall* e *F1 Score*, descritas na Tabela 2.

Tabela 2 - Métricas de Avaliação de Desempenho

Métrica	Descrição	Fórmula
Acurácia	Indica uma <i>performance</i> geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente;	$\frac{VP + VN}{VP + FV + FP + FN}$
Precisão	Dentre todas as classificações de classe Positivo que o modelo fez, quantas estão corretas;	$\frac{VP}{VP + FP}$
<i>Recall</i>	Dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas;	$\frac{VP}{VP + FN}$
<i>F1-Score</i>	Média harmônica entre precisão e <i>recall</i> .	$\frac{2 * Precisão * Recall}{Precisão + Recall}$

Fonte: Adaptado de Provost e Fawcett (2013)

O desempenho também foi avaliado pela Matriz de Confusão, que é uma tabela que permite obter as métricas de avaliação de desempenho para as técnicas de IA, descritas acima. Apresenta-se na Tabela 3, a Matriz de Confusão.

Tabela 3 - Matriz de Confusão

		Previsto	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Adaptado de Provost e Fawcett (2013)

Fase 4: Comparação dos resultados dos experimentos com os resultados dos indicadores detratores do NPS

A comparação do NPS com os resultados dos experimentos foi realizada para saber se o NPS trouxe respostas de reclamações semelhantes às encontradas nos resultados dos experimentos com a RNC. A análise dos resultados comparativos foi realizada em conjunto com o especialista da empresa varejista.

Fase 5: Interpretação e avaliação dos resultados

Os resultados obtidos com a aplicação da RNC foram avaliados com base nas correlações e cruzamentos de dados estabelecidos para a descobrir conhecimento sobre os clientes.

Apresenta-se no Quadro 1, as correlações das categorias de dados utilizadas para os cruzamentos, com o objetivo de revelar *insights* sobre os clientes da empresa varejista.

Quadro 1 - Correlações de categorias de dados para cruzamentos

Causa (comentário do cliente)	Perfil do cliente (sexo, idade, renda e estado)
	Produto cartão (faixa limite, tipo de produto, ciclo de vida e adicional)
	Canais de atendimento (central de atendimento, estandes/lojas, aplicativos, portais e status da solicitação)
	Estratificação de notas do NPS (notas de 0 a 6, segregadas em dois substratos: notas de 0 a 3 (destratores inferiores) e notas de 4 a 6 (destratores superiores)).

Fonte: Autores (2024)

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Nesta seção os resultados dos experimentos computacionais são apresentados, analisados e discutidos, considerando as fases de desenvolvimento da metodologia experimental.

4.1 Resultados dos experimentos computacionais

Fases 1 e 2: Seleção da base de dados de atendimento, pré-processamento e transformação dos dados

Após selecionar, pré-processar e transformar a base de dados, foram usadas ferramentas de classificação de texto para extrair conhecimento, começando pela identificação de palavras-chave. A ferramenta *Rake* foi utilizada para criar um *ranking* das principais palavras, com as dez primeiras listadas na Tabela 4.

Tabela 4 - *Ranking* de palavras

Ranking	Palavras	Score
1	atendimento / cartão / vencimento/ cancelamento	25.0
2	cobrar/ juro/ cliente	24.5
3	cancelamento / assinatura/ sistema	23.0
4	atendimento / cartão / cobrar/ anuidade	16.8
5	parcelar / cartão / alimentação	15.8
6	aumentar / limite / cancelar	15.8
7	limite / aumentar/ parcelar	15.5
8	senha / cadastro / cartão	9.5
9	diminuíram / atraso	9.5
10	bloquear / cartão	9.0

Fonte: Autores (2024)

A análise da Tabela 4 revela as palavras-chave mais relevantes nos comentários dos clientes, com base no *score*, que combina a frequência, o grau de concorrência com outras palavras e a relação grau/frequência. As palavras como atendimento, cartão, vencimento e cancelamento tiveram os maiores *scores*, sendo as mais citadas e frequentemente associadas a outras palavras importantes. A seguir, descreve-se a aplicação das ferramentas *Gensim* e do *Bertopic* para a modelagem de tópicos.

O texto foi tokenizado e convertido em um dicionário de termos, atribuindo um índice a cada termo. Em seguida foi criada uma matriz de texto baseada nesse dicionário revelando tópicos como: cartão, limite, atendimento, compra e pagamento.

Os tópicos foram extraídos com a aplicação da medida TF-IDF (*Term Frequency - Inverse Document Frequency*). O TF-IDF é um cálculo estatístico utilizado como uma forma de quantificar palavras em conjuntos de documentos. Em seguida, a aplicação do *Bertopic* confirmou os cinco principais tópicos identificados pelo *Gensim*: cartão, limite, atendimento, pagamento e compra. Para validar a classificação de palavras-chave e a modelagem de tópicos, a frequência das palavras foi calculada usando uma biblioteca de probabilidade, excluindo *stopwords* (Kit e Mojik, 2022)

As palavras mais frequentes nos comentários dos clientes estão listadas na Tabela 5.

Tabela 5 - Lista das palavras mais frequentes nos comentários

Palavras frequentes		
1	cartão	4.931
2	limite	3.432
3	comprar	2.115
4	atendimento	1.749
5	conseguir	1.656
6	pagar	1.650
7	nome empresa	1.499
8	aumentar	1.478
9	resolver	1.416
10	baixo	1.322

Fonte: Autores (2024)

A base de dados conta agora um novo atributo denominado *Topic*. Apresenta-se na Figura 2, a base de dados com o atributo *Topic*.

Figura 2 - Base de dados com o atributo *Topic*

	A	B	C	D	E	F
	causa	sexo	idade	renda	estado	topic
2	atendimento eletronico lento confuso	F	Entre 36 e 40 anos	Entre 768,01 e 1.625,00	MG	atendimento
3	atendimento	M	Entre 36 e 40 anos	Entre 768,01 e 1.625,00	PB	atendimento
4	receber cartao ridicula receber receber pessoa problema cartao	F	Entre 36 e 40 anos	Entre 768,01 e 1.625,00	SP	cartao
5	cartao bloquei problema	F	Acima de 60 anos	Entre 768,01 e 1.625,00	SP	cartao
6	fixar anuidade incomodar carrefour cartao principal faco inumeras	F	Entre 36 e 40 anos	Entre 768,01 e 1.625,00	PE	cartao
7	aprovar motivo cancelar comprar	F	Entre 51 e 60 anos	Entre 768,01 e 1.625,00	RN	comprar
8	parcelamento automatico fatura autorizacao atraso efetuar pagar	M	Entre 41 e 45 anos	Entre 2.705,01 e 4.852,00	SP	pagar
11	conseguir resolver pleitear	M	Acima de 60 anos	Entre 1.625,01 e 2.705,00	MT	conseguir
12	limite aumentar dificultar opcoes comprar loja carrefour	M	Entre 25 e 30 anos	Entre 768,01 e 1.625,00	PR	limite
13	aumentar limite conseguir comprar praticamente	F	Entre 18 e 24 anos	Entre 768,01 e 1.625,00	PE	limite
14	cliente carrefour dependente titular pedir aumentar limite negar	M	Entre 31 e 35 anos	Entre 1.625,01 e 2.705,00	RN	limite
15	ligar baixo cobrar pagto fatura vencer natal quando vencto ocorre	F	Acima de 60 anos	Entre 2.705,01 e 4.852,00	SP	baixo
16	horar preciso cartao minitos baixo aumento	F	Entre 46 e 50 anos	Entre 768,01 e 1.625,00	RJ	cartao
17	baixo cancelar nado	M	Entre 36 e 40 anos	Entre 2.705,01 e 4.852,00	SP	baixo
18	receber cartao senha digitos adquirir senha conseguir	M	Entre 36 e 40 anos	Entre 768,01 e 1.625,00	RJ	cartao
19	ligar senha receber cartao senha mandar sms receber desbloq	F	Entre 51 e 60 anos	Entre 1.625,01 e 2.705,00	SP	cartao
20	cartao carrefour experiencia melhorar agente consegui comprar e	M	Entre 41 e 45 anos	Entre 1.625,01 e 2.705,00	GO	cartao
21	contar anuidade receber mail anuidade isentar anuidade efetua	M	Entre 36 e 40 anos	Entre 4.852,01 e 9.254,00	RJ	comprar
22	chato mes antar vencer fatura mandar mails oferecer parcelamer	M	Entre 51 e 60 anos	Entre 1.625,01 e 2.705,00	SP	cartao
23	limite baixo demorar entregar cartao	F	Entre 36 e 40 anos	Entre 1.625,01 e 2.705,00	SP	cartao
24	colocar cartao receber cartao entregar cartao carrefour come	F	Entre 31 e 35 anos	Entre 768,01 e 1.625,00	RJ	cartao

Fonte: Autores (2024)

Fase 3 - Aplicação da RNC na Análise de Sentimentos

Os testes determinaram que o melhor percentual de divisão da base de dados para aplicar as técnicas selecionadas foi de 50% para o treinamento, 30% para teste e 20% validação. A melhor topologia para a RNC foi a seguinte: camada de entrada com 28 x 28 neurônios, camada de saída com 5 x 5 neurônios, duas camadas ocultas de 24 x 24 neurônios cada e o uso da função de ativação *Softmax*. O desempenho da RNC foi comparado com a MLP e com a Árvore de Decisão. Apresenta-se na Tabela 6, os resultados comparativos das técnicas aplicadas, considerando as métricas de avaliação de desempenho selecionadas.

Tabela 6 - Resultado do desempenho das técnicas selecionadas

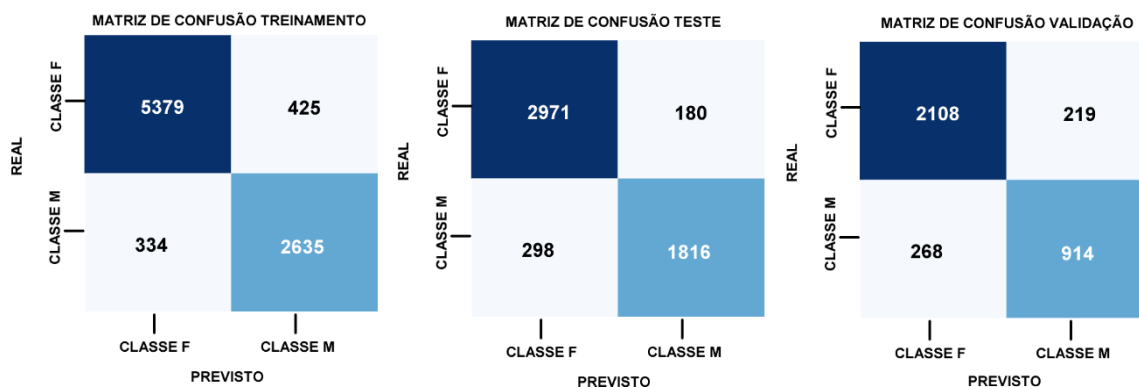
Técnica	Conjunto de dados	Acurácia	Precisão (F/M)	Recall (F/M)	F1-score (F/M)
MLP	Treinamento (50%)	67%	69%/57%	86%/32%	77%/41%
	Teste (30%)	63%	68%/42%	82%/26%	74%/32%
	Validação (20%)	64%	67%/49%	90%/18%	77%/26%
	Treinamento (50%)	75%	70%/59%	80%/48%	88%/39%

Árvore de Decisão	Teste (30%)	70%	62%/55%	78%/28%	80%/19%
	Validação (20%)	69%	65%/32%	75%/42%	70%/40%
Rede Neural Convolucional (RNC)	Treinamento (50%)	97%	99%/98%	95%/93%	95%/92%
	Teste (30%)	99%	99%/99%	100%/98%	99%/98%
	Validação (20%)	99%	99%/99%	99%/97%	99%/98%

Fonte: Autores (2024)

Apresenta-se na Figura 3, as Matrizes de Confusão da RNC de acordo com a divisão da base em treinamento, teste e validação.

Figura 3 - Matrizes de Confusão da RNC



Fonte: Autores (2024)

Analisando-se os resultados da Tabela 6 e as matrizes da Figura 3, verificou-se que a RNC apresentou bom desempenho nas etapas de treinamento, teste e validação. Além disso, foi a técnica que melhor classificou a classe "M". Os resultados apresentaram um ótimo equilíbrio entre precisão e *recall*, com uma baixa margem de erro. Apesar da presença de falsos negativos, o impacto desses erros foi reduzido devido à boa performance geral.

Desta forma, a RNC foi selecionada para a continuação dos experimentos. Os principais motivos que levaram a essa decisão incluem:

1. **Melhor Acurácia:** Obteve a melhor acurácia em todos os conjuntos de dados.
2. **Equilíbrio entre Precisão e Recall:** Conseguiu manter um equilíbrio mais satisfatório entre ambas as classes, garantindo que a classe M (Masculino) fosse melhor identificada do que nas outras técnicas.
3. **Melhor Desempenho no F1-Score:** Demonstrou um desempenho mais consistente no *F1-Score*, com valores próximos para ambas as classes. Isso indica que o modelo não só identifica corretamente os exemplos positivos, mas também reduz a ocorrência de falsos positivos e falsos negativos.
4. **Análise da Matriz de Confusão:** Revelou que a RNC teve a menor taxa de erros comparada aos outros modelos. Embora ainda existam falsos

negativos na classe M, seu desempenho foi consideravelmente melhor do que os modelos MLP e Árvore de Decisão.

5. **Capacidade de Capturar Padrões Complexos:** Redes Neurais são conhecidas por sua habilidade de aprender padrões complexos nos dados, permitindo maior flexibilidade e robustez na classificação. Isso pode justificar o melhor desempenho da RNC em relação às abordagens tradicionais como a MLP e a Árvore de Decisão.

Fase 4 - Comparação dos resultados dos experimentos com os resultados dos indicadores detratores do NPS

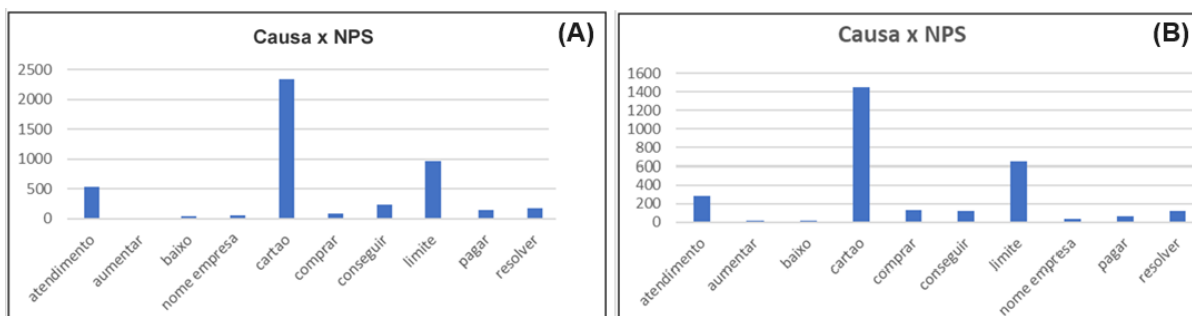
A partir dos resultados auferidos nos experimentos foi realizada a comparação com os indicadores detratores do NPS. Esta atividade ocorreu em conjunto com o especialista da empresa, de modo a aproveitar sua expertise no fenômeno analisado.

Ao comparar os tópicos identificados (cartão, limite, compras, atendimento, pagamentos e aumento) com os resultados do NPS expressos pelos clientes, foi possível estabelecer uma primeira correlação com as notas dos detratores (0 a 3).

Os tópicos que geraram mais reclamações dos clientes foram relacionados ao cartão, limite e atendimento nos canais disponibilizados pela empresa, conforme ilustrado na Figura 4, (A), que apresenta a comparação com os detratores do NPS de 0 a 3.

Por outro lado, os tópicos com menor volume de comentários dos clientes foram: baixo, nome da empresa, comprar, não conseguir atendimento e resolver problema com a empresa. Vale destacar que o tópico "aumentar" não recebeu nenhum comentário dos clientes na correlação entre os experimentos analisados e os detratores do NPS com notas de 0 a 3.

Figura 4 - Comparação dos resultados dos experimentos com os resultados do NPS



Fonte: Autores (2024)

Ao comparar os resultados dos experimentos com as notas de NPS dos detratores, foi realizada uma segunda correlação, desta vez com as notas de 4 a 6.

Pode-se verificar na Figura 4, (B) a comparação dos resultados com o NPS desses detratores. Os tópicos que mais geraram insatisfações entre os clientes foram relacionados a cartão, limite e atendimento. Por outro lado, os tópicos que registraram menos reclamações foram: aumentar, baixo, comprar, pagar, nome da empresa e resolver problemas.

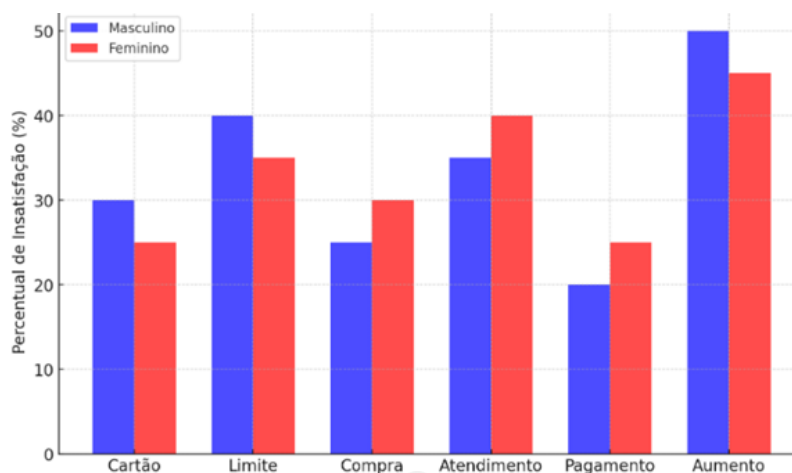
A análise mostrou que muitos comentários dos clientes estavam relacionados a problemas com cartão, atendimento, limite, pagamento e questões não resolvidas pelos canais da empresa. Além disso, foi observado que alguns clientes expressaram insatisfação em relação a múltiplos problemas

identificados pela solução desenvolvida, abordando dois ou até três dos principais tópicos em um único comentário.

Fase 5: Interpretação e avaliação dos resultados

Após a classificação dos sentimentos pela RNC, foi realizada a correlação entre o atributo causa (comentário do cliente) e o atributo sexo em cruzamento com 'idade', 'renda', e 'estado do cliente'. Os resultados obtidos foram organizados nos principais tópicos identificados como os mais relevantes para a análise: cartão, limite, compra, atendimento, pagamento e aumento. Essa abordagem permitiu identificar quais grupos de clientes demonstraram maior insatisfação e quais foram os principais problemas relatados. A saída da RNC consistiu em um conjunto de categorias (Tópicos) geradas com base na similaridade semântica dos comentários. Os resultados da correlação são apresentados na Figura 5.

Figura 5 - Identificação dos Tópicos Relevantes



Fonte: Autores (2024)

O próximo passo consistiu, juntamente com os especialistas da empresa, em aplicar as correlações e cruzar os dados para obter *insights* a partir das reclamações dos clientes. Nessa fase, foi correlacionado o comentário do cliente com sexo, idade, renda e estado do cliente. Os resultados são apresentados na Tabela 7, de acordo com os principais tópicos significativos para a

Tabela 7 - Correlação entre os atributos

Tópicos	Resultados
Cartão	A correlação entre os atributos 'causa' e 'sexo' revelou que a maioria dos clientes insatisfeitos com os serviços de cartão tem idade entre 36 e 40 anos e 51 e 60 anos, e renda mensal entre R\$ 768,01 e R\$ 1.625,00. O estado de São Paulo lidera com cerca de 30% das reclamações, tanto para perfis femininos quanto masculinos. Outros estados, como Rio de Janeiro e Minas Gerais, tiveram aproximadamente 15% das reclamações cada, enquanto os demais estados somaram cerca de 40%.
Limite	A correlação entre o tópico 'Limite' e o atributo 'sexo' revelou que 40% dos clientes insatisfeitos com o limite do cartão têm idade entre 36 e 40 anos e 25% têm entre 51 e 60 anos. As principais queixas envolvem limite baixo, dificuldade

	<p>para solicitar aumento e incompatibilidade com a renda. Os 35% restantes são de insatisfações de outras faixas etárias. Boa parte dos clientes insatisfeitos tem renda mensal entre R\$ 768,01 e R\$ 1.625,00. São Paulo concentra 60% das reclamações sobre limite de cartão, enquanto os demais estados representam 40%.</p>
Comprar	<p>A correlação entre o tópico 'Comprar' e o atributo 'sexo' revelou que a maioria dos clientes insatisfeitos com compras, tanto em lojas físicas quanto <i>online</i> tem idade entre 41 e 45 anos. Entre os homens, 47% estão insatisfeitos com o parcelamento, dificuldades no site e no uso do cartão. Entre as mulheres, 53% expressam insatisfação com a aprovação de compras, análise de crédito, tarifas altas, reembolso e vencimento das compras. Além disso, 60% dos clientes com renda entre R\$ 768,01 e R\$ 1.625,00 estão insatisfeitos, enquanto 40% das demais faixas de renda estão insatisfeitos com outros produtos e serviços. A análise também indicou que 60% das reclamações vêm de clientes do estado de São Paulo.</p>
Atendimento	<p>A correlação entre o tópico 'Atendimento' e o atributo 'sexo' mostrou que a maioria dos clientes insatisfeitos com o atendimento da empresa tem entre 51 e 60 anos. Entre os homens, 40% estão insatisfeitos com a demora no atendimento, lentidão da central e falta de qualificação dos atendentes. Entre as mulheres, 60% expressam insatisfação com o atendimento confuso, dificuldades no atendimento por telefone, informações falsas e longos períodos de espera.</p>
Pagamento	<p>Observou-se que 52% das mulheres estão insatisfeitas com os serviços da empresa devido a promoções que consideram como falsas, dificuldade de contato com a central de atendimento, problemas com o aplicativo, mau atendimento via <i>chat</i> e tarifas altas. Entre os homens, 48% estão descontentes com cobranças indevidas, com produtos e com problemas com o aplicativo, atendimento em lojas e questões com a fatura do cartão. A maioria das reclamações sobre pagamentos é de clientes entre 51 e 60 anos, com outras faixas etárias também expressando insatisfação. Entre os homens, 35% relatam dificuldades com atendimento via <i>chat</i>, troca de senha, espera na central, desbloqueio do cartão e acesso ao app, enquanto 65% das mulheres enfrentam problemas semelhantes, além de dificuldades no pagamento de faturas e suporte. Em São Paulo, 60% dos clientes com renda entre R\$ 768,01 e R\$ 1.625,00 apresentaram reclamações sobre pagamentos. A análise mostra que tanto homens quanto mulheres com o percentual de 50% cada estão insatisfeitos com questões de pagamento, incluindo anuidade, pagamento automático, opções e taxas.</p>
Aumento	<p>A correlação entre o tópico 'aumento' e o atributo 'sexo' revelou que boa parte dos clientes insatisfeitos com o aumento de limite tem entre 31 e 35 anos. Entre os homens, 51% estão insatisfeitos com as dificuldades para aumentar o limite de crédito e o aumento da anuidade, enquanto 49% das mulheres se queixam do aumento da pontuação de compras e dificuldades para elevar o limite do cartão. Clientes com renda entre R\$ 1.625,01 e R\$ 2.705,00 também estão insatisfeitos, alegando que a empresa não aumenta os limites para compras. São Paulo e Amazonas são os estados com o maior número de reclamações sobre limites de crédito.</p>

Fonte: Autores (2024)

4.2 Análise dos Resultados

Apresenta-se se na tabela (Tabela 8) os principais conhecimentos sobre as manifestações de insatisfação dos clientes detratores oriundos da análise dos resultados.

Tabela 8 - Principais conhecimentos descobertos sobre o cliente

Tópico	Conhecimento descoberto
Cartão	Clientes com perfis diversos estão insatisfeitos com os serviços da empresa, principalmente devido à falta de aumento de limite, dificuldades no contato com a central de atendimento, problemas com desbloqueio e cancelamento de cartões, e dificuldade em obter informações sobre os cartões.
Limite	Existência de clientes com perfis diferentes com dificuldade em aumentar o limite do cartão, mesmo quando o cliente tem renda alta.
Compras	Clientes com diversos problemas relacionados a compras efetuadas nas lojas próprias da empresa, problemas com entregas, pois muitos produtos chegam danificados e com tempo de entrega atrasado em relação ao prometido pela empresa.
Atendimento	Clientes reclamaram dos seguintes problemas: atendimento na central, lojas e aplicativos, pois há muita demora no tempo de espera para ser atendido. O estado com mais reclamações no atendimento é o estado de São Paulo, provavelmente em razão do volume de clientes.
Pagamento	Problemas relacionados a pagamento de fatura de cartão e multas altas para pagamentos efetuados em atraso.
Aumento	Clientes de diferentes perfis, estados e idades estão insatisfeito com falta de aumento do limite de crédito e aumento de anuidade do cartão.

Fonte: Autores (2024)

Foi identificado que clientes com renda mensal entre R\$ 768,01 e R\$ 1.625,00 e residentes em São Paulo estão insatisfeitos com questões relacionadas a cartões, limite, atendimento *online* e em lojas físicas. Entre os homens, 47% estão descontentes com parcelamento de compras, dificuldades no site da empresa e problemas ao usar o cartão nas lojas. Para as mulheres, 53% enfrentam dificuldades com análise de crédito, altas taxas para parcelamento e reembolso de compras.

Constatou-se que 47% dos homens estão insatisfeitos com a demora e qualidade do atendimento, enquanto 48% das mulheres têm problemas com atendimento confuso e demorado. Quanto aos pagamentos, 52% das mulheres estão insatisfeitas com promoções incorretas, indisponibilidade da central de atendimento e altos custos, enquanto 48% dos homens reclamam de cobranças indevidas, problemas com o app e atendimento ruim nas lojas físicas.

Foi constatado também que a maioria dos clientes insatisfeitos com pagamentos e com dificuldades para resolver problemas com a empresa têm entre 51 e 60 anos. Entre os homens dessa faixa etária, 46% estão insatisfeitos com a falta de ajuda pelo *chat*, problemas com o aplicativo e a central de atendimento. Entre as mulheres, 49% relataram dificuldades no *call center*, problemas para pagar a conta e suporte inadequado.

Além disso, 51% dos homens estão descontentes com o aumento de limite de crédito e previdência, enquanto 49% das mulheres reclamam do aumento dos pontos de compra e limite do cartão. Clientes com renda entre R\$ 1.625,01 e R\$ 2.705,00 também estão insatisfeitos com o limite de crédito e aumento de previdência, alegando que a empresa não ajusta os limites de forma adequada.

São Paulo e Amazonas foram identificados como os estados com mais reclamações sobre o limite de crédito concedido pela empresa.

Os resultados dos experimentos foram comparados com os indicadores críticos do NPS detratores. A correlação entre os comentários dos clientes e atributos como sexo, idade, renda e estado foi analisada em conjunto com os indicadores de NPS. Verificou-se que 60% dos clientes que deram as notas mais baixas no NPS (entre 0 e 3) são do sexo feminino, indicando uma alta insatisfação desse grupo com os serviços e produtos da empresa. Em contraste, 40% dos homens atribuíram notas mais altas (entre 4 e 6), sugerindo que as mulheres tendem a expressar mais insatisfação nos canais de atendimento.

Constatou-se que 70% dos clientes com renda entre R\$ 768,01 e R\$ 1.625,00 deram as piores notas no NPS (entre 0 e 3), indicando insatisfação com o aumento de limite de crédito e previdência do cartão. Clientes com renda entre R\$ 9.254,01 e R\$ 20.888,00, que representam 30% dos insatisfeitos, estão descontentes principalmente com o cartão e o limite estabelecido. Entre os clientes com mais de 60 anos, 50% deram notas de 0 a 3, relatando problemas com o atendimento ao cliente, enquanto 50% dos clientes entre 18 e 50 anos deram notas de 4 a 6, expressando insatisfação com o limite e anuidade do cartão. Esses resultados indicam que a empresa não ofereceu limites de cartão adequados para diferentes perfis de clientes, variando em renda, idade e sexo.

A análise mostrou que 60% dos clientes insatisfeitos com NPS entre 0 e 3 são de São Paulo, evidenciando uma alta insatisfação com os serviços e produtos da empresa. Em comparação, apenas 40% dos clientes de outros estados deram notas entre 4 e 6, com as principais reclamações relacionadas à falta de atendimento adequado nas lojas e na central de atendimento.

A comparação dos resultados dos experimentos com os NPS detratores (notas 0 a 3) revelou que as principais queixas eram sobre limite e atendimento, enquanto os tópicos com menos comentários incluíram limite baixo, dificuldades em comprar, obter atendimento e resolver problemas. O tópico "aumentar" não recebeu comentários na faixa de NPS detratador entre 0 e 3. Os comentários dos clientes frequentemente abordam problemas com cartão, atendimento, limite e pagamentos, bem como questões não resolvidas pelos canais de atendimento da empresa. Tanto homens quanto mulheres mostraram insatisfação com vários problemas identificados.

Considera-se que a aplicação da RNC na análise de sentimentos em comentários de clientes de uma empresa do ramo varejista descobriu conhecimento nos comentários sobre as manifestações de insatisfação dos clientes detratores, considerado relevante para apoiar a empresa na gestão deste cliente.

5 CONCLUSÃO

A aplicação da RNC buscou analisar os sentimentos expressados pelos clientes da empresa varejista abordada neste estudo para apoiar a gestão do cliente. Para tanto foi realizada a avaliação e interpretação do conhecimento descoberto, além do cruzamento deste conhecimento produzido com os indicadores detratores do NPS.

Para atingir os objetivos indicados foi considerada uma base de dados com registros provenientes da base de atendimento aos clientes, bem como os resultados do NPS. Os dados foram utilizados para a realização de experimentos computacionais baseados em métodos e técnicas voltados à análise e

classificação de sentimentos, visando assim constituir uma solução eficiente e útil para gerir o cliente, considerando os produtos e serviços oferecidos.

A partir da análise realizada, pode-se concluir que muitos comentários dos clientes dizem respeito a problemas com cartão, atendimento, limite, pagamento e problemas que não são solucionados pelos canais de atendimento da empresa. Também foi possível verificar que tanto o perfil feminino, quanto o perfil masculino demonstram insatisfação com mais de um problema (tópicos identificados) indicado pela solução aplicada neste trabalho.

Assim, os resultados alcançados indicam que a solução desenvolvida é capaz de proporcionar descoberta de conhecimento do cliente que apoie a gestão dos clientes na empresa varejista, objeto de estudo neste trabalho.

Como contribuições deste trabalho para a Academia, indica-se a aplicação da RNC voltada à análise de sentimentos de comentários de clientes como um tema de pesquisa atual e relevante. Isto porque as técnicas de análise de sentimentos de clientes denotam sua capacidade de produzir conhecimentos novos acerca do cliente e, portanto, merecem a atenção dos pesquisadores, que poderão aplicar a solução aqui desenvolvida em outros fenômenos e objetos de pesquisa.

Também se indica contribuições desta pesquisa para os profissionais, gestores e organizações de mercado. Isto porque a solução desenvolvida mostrou-se relevante para a viabilização da aplicação de métodos e técnicas inteligentes em demandas e problemas reais das empresas atuais, no que concerne à descoberta de conhecimento a partir da análise de sentimentos em comentários realizados por clientes, podendo ser replicada em outros tipos de empresas (porte, setor de atuação, perfil de clientes).

Segundo Nonaka e Takeuchi (1997) adquirir e descobrir o conhecimento do cliente é vital, principalmente para os negócios, além de ajudar a empresa a aprender mais sobre o cliente que compra seus serviços e produtos. Portanto, ao realizar experimentos aplicando a RNC para a análise de sentimento do cliente e comparando com a metodologia NPS, prevê-se importante contribuição para o campo científico da Inteligência Artificial e seus métodos e técnicas.

Segundo Baquero (2022), o NPS fornece uma visão clara da lealdade dos clientes e do grau de satisfação, ajudando as empresas a entenderem o quão bem estão atendendo às expectativas dos clientes. Os resultados comprovam a importância do NPS quando comparado aos resultados dos experimentos computacionais realizados.

O desenho de pesquisa aqui delineado poderá permitir o desenvolvimento e validação de diferentes tipos de soluções inteligentes para a descoberta de conhecimento do cliente, subsidiando assim uma melhor tomada de decisão dos gestores nas empresas acerca dos clientes.

Como limitações desta pesquisa indica-se a seleção da técnica inteligente inicialmente aplicada para a realização dos experimentos realizados, que se restringiu à RNC. Essa técnica foi escolhida uma vez que trabalhos já publicados indicavam que essa técnica é aplicada para a análise de sentimentos. Outra limitação diz respeito a base de dados analisada. Não obstante sua relevância em função da diversidade e volume dos dados analisados, a base em questão retrata um fenômeno restrito: a insatisfação de clientes em relação ao atendimento prestado por empresa varejista.

Há ainda a ser mencionado o fato de que os experimentos foram realizados em uma única empresa, o que impossibilitou a comparação dos resultados com outros trabalhos da literatura. Embora a empresa abordada nesta pesquisa seja uma multinacional do setor de varejo, há de se considerar as especificidades do negócio de atuação da empresa, bem como suas características únicas.

Não obstante, estima-se que em função do volume e características dos dados disponíveis na empresa em questão, os resultados dos experimentos ora realizados possam ajudar a resolver problemas que ocorrem costumeiramente em empresas varejistas, não sendo possível, entretanto, generalizar os resultados para toda e qualquer empresa deste segmento.

Por fim, como indicação de pesquisas futuras, sugere-se realizar a aplicação desta solução em outras bases de dados, bem como em empresas de diferentes setores de atuação. Indica-se ainda a aplicação de outras técnicas inteligentes além da RNC, realizando-se assim um comparativo de suas métricas e resultados.

REFERÊNCIAS

ALAWI, Abdulfattah Ba; BOZKURT, Ferhat. A hybrid machine learning model for sentiment analysis and satisfaction assessment with Turkish universities using Twitter data. **Decision Analytics Journal**, Volume 11, 100473, 2024. ISSN 2772-6622, DOI: 10.1016/j.dajour.2024.100473.

ALMAGHRABI, Maran; CHETTY, Ginja. Multilingual Sentiment Recommendation System based on Multilayer Convolutional Neural Networks (MCNN) and Collaborative Filtering based Multistage Deep Neural Network Models (CFMDNN). **2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)**, Antalya, Turkey, 2020. p. 1-6. DOI: 10.1109/AICCSA50499.2020.9316492.

ANTONELO, Kamila; LIMA, Gustavo Barbieri. A importância das mídias sociais em empresa varejista do segmento de moda: O Caso Alpha. **Brazilian Journal of Development**, v. 7, n. 1, p. 3684-3703, 2021. DOI 10.34117/bjdv7n1-248.

BENHAR, H.; IDRI, A.; FERNÁNDEZ-ALEMÁN, J. L. Data preprocessing for heart disease classification: A systematic literature review. **Computer Methods and Programs in Biomedicine**, v. 195, 2020. Artigo 105635. ISSN 0169-2607. Disponível em: <https://doi.org/10.1016/j.cmpb.2020.105635>.

BAEHRE, Sven. et al. Customer mindset metrics: A systematic evaluation of the net promoter score (NPS) vs. alternative calculation methods. **Journal of Business Research**, v. 149, p. 353-362, 2022. DOI 10.1016/j.jbusres.2022.04.048.

BAQUERO, Asier. Net Promoter Score (NPS) and Customer Satisfaction: Relationship and Efficient Management. **Sustainability**, v. 14, p. 2011, 2022. DOI 10.3390/su14042011.

CAO, Zixuan et al. Deep transfer learning mechanism for fine-grained cross-domain sentiment classification. **Connection Science**, v. 33, n. 4, p. 911-928, 2021. DOI 10.1080/09540091.2021.1912711.

CHANDRASEKARAN, Ganesh; NGUYEN, Tu N.; HEMANTH, Jude D. Multimodal sentimental analysis for social media applications: A comprehensive review. **ACM Transactions on Information Systems**, v. 39, n. 2, p. 1-38, 2021. DOI 10.1002/widm.1415.

CHO, Young Sik et al. The Impact of Social Media Technologies on Supplier and Customer Relationship Management: An Empirical Analysis in the U.S. **IEEE Access**, v. 8, p. 220878-220888, 2020. DOI 10.1109/ACCESS.2020.3043143.

DWIVEDI, Yogesh. K. et al. Setting the future of digital and social media marketing research: Perspectives and research propositions. **International Journal of Information Management**, v. 59, p. 102168, 2021. DOI 10.1016/j.ijinfomgt.2020.102168.

FAYYAD, Usama. M.; PIATETSKY-SHAPIRO, Gregory.; SMYTH, Padhraic. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, p. 27-34, 1996. DOI doi.org/10.1145/240455.240464.

JAHANI, Hamed; JAIN, Richa; IVANOV, Dmitry. Data science and big data analytics: a systematic review of methodologies used in the supply chain and logistics research. **Annals of Operations Research**, v. 319, p. 335-374, 2023. Disponível em: DOI 10.1007/s10479-023-05390-7.

JAVAID, Kumail, et al. Explainable Artificial Intelligence Solution for Online Retail. **Computers, Materials and Continua**, 71(3), 4425-4442, 2022. DOI 10.32604/cmc.2022.022984.

KIM, Sodom; PARK, Eunil. STAD-GCN: Spatial-Temporal Attention-based Dynamic Graph Convolutional Network for retail market price prediction. **Expert Systems with Applications**, v. 255, 2024, 124553. DOI 10.1016/j.eswa.2024.124553.

KIT, Yuheng; MOKJI, Musa Mohd. Sentiment analysis using pre-trained language model with no fine-tuning and less resource. **IEEE Access**, v. 10, p. 107056-107065, 2022. DOI 10.1109/ACCESS.2022.3212367. Disponível em: <https://ieeexplore.ieee.org/document/9912410>. Acesso em 15 set. 2024.

KLAUS, P.; MAKLAN, S. The Net Promoter Score and customer experience: a longitudinal analysis. **Journal of Service Management**, v. 32, n. 4, p. 507-527, 2021.

LI, Hongchan; LU, Yantong; ZHU, Haodong; ZHU, Haodong; MA, Yu. A Novel AB-CNN Model for Multi-Classification Sentiment Analysis of e-Commerce Comments. **Electronics**, v. 12, n. 8, p. 1880, 2023. DOI: 10.3390/electronics12081880.

LIU, Bing. Sentiment analysis and opinion mining: a survey. **Journal of Computer Science and Technology**, v. 28, n. 4, p. 585-606, 2023.

MEENA, Gourav; MOHBHEY, Krishna Kumar; KUMAR, Sunil. Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach. **International Journal of Information Management Data Insights**, v. 3, i. 3, p. 100174, 2023. DOI: 10.1016/j.jjime.2023.100174.

NANDWANI, Pansy; VERMA, Rupali. A review on sentiment analysis and emotion detection from text. **Social Network Analysis and Mining**, v. 11, p. 81, 2021. DOI 10.1007/s13278-021-00776-6.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka. **Criação de conhecimento na empresa: Como as empresas japonesas geram a dinâmica da inovação**. Rio de Janeiro: Elsevier, 1997.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. O'Reilly Media, Inc., 2013.

RAIAAN, Mohaimenul Azam Khan *et al.* A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks. **Decision Analytics Journal**, v. 11, 2024, 100470. DOI 10.1016/j.dajour.2024.100470.

SHARBATIAN, Kamal; MOATTAR, Mohammad Hossein. Deep aspect extraction and classification for opinion mining in e-commerce applications using convolutional neural network feature extraction followed by long short term memory attention model. **Applied AI Letters**, v. 4, n. 3, e86, 2023. DOI: 10.1002/ail2.86.

SIVABALAN, S.; MINU, R. I. Optimized multi-anchor space-aware temporal convolutional neural network for automobile sales prediction. **Knowledge-Based Systems**, v. 311, 113000, 2025. DOI: 10.1016/j.knosys.2025.113000.

SRIVASTAVA, Saurabh; NATARAJAN, Sundarapandiyan; BHARATHI, Sivagnana; YADAV, Archana; SINGH, Rashmita; ARUN, R. Omnichannel Marketing Strategies: Bridging the Gap Between Online and Offline Retail Experiences Using DCNN-BiLSTM Technique. **2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)**, Kalaburagi, India, 2024. p. 1-6. DOI: 10.1109/ICIICS63763.2024.10859835.

TIETZ, Marian *et al.*; **skorch Developers**. **skorch**: A scikit-learn compatible neural network library that wraps PyTorch. [S.l.], 2017. Disponível em: <https://skorch.readthedocs.io/en/stable/>.

WANG, Xinyi; LI, Feng; ZHANG, Zequn; XU, Guangluan; ZHANG, Jingyuan; SUN, Xian. A unified position-aware convolutional neural network for aspect-based sentiment analysis. **Neurocomputing**, v. 450, p. 91-103, 2021. ISSN 0925-2312. DOI: 10.1016/j.neucom.2021.03.092.

WANKHADE, Mayur.; RAO, Annavarapu Chandra Sekhara.; KULKARNI, Chaitanya. A survey on sentiment analysis methods, applications, and challenges. **Artificial Intelligence Review**, v. 55, p. 5731-5780, 2022. DOI 10.1007/s10462-022-10144-1.