

15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

ANÁLISE DE SENTIMENTOS DOS FEEDBACKS DE CLIENTES DA AMAZON POR MEIO DE TÉCNICAS DE MINERAÇÃO DE DADOS

YASMIM C. GAMBOA¹, VITÓRIA S. A. OLIVEIRA², NATALY S. DIAS³, CAMILA B. DE SOUZA⁴, REGIVALDO S. FERREIRA⁵

1 Estudante do curso Técnico em Logística integrado ao Ensino Médio, IFSP, Campus Pirituba, gamboa.yasmim@aluno.ifsp.edu.br

2 Estudante do curso Técnico em Logística integrado ao Ensino Médio, IFSP, Campus Pirituba, v.amaral@aluno.ifsp.edu.br

3 Estudante do curso Técnico em Logística integrado ao Ensino Médio, IFSP, Campus Pirituba, nataly.souza@aluno.ifsp.edu.br

4 Estudante do curso Técnico em Logística integrado ao Ensino Médio, IFSP, Campus Pirituba, barros.souza@aluno.ifsp.edu.br

5 Docente do IFSP, Campus Pirituba, regivaldoferreira@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.02.02.08-0 Análise de Dados

RESUMO: Com o aumento exponencial de *feedbacks* gerados nas plataformas digitais, a mineração de dados torna-se essencial para uma melhor avaliação dos sentimentos dos consumidores. A análise de sentimentos é uma técnica de processamento de linguagem natural (PLN), que extrai informações de textos de forma prática e sucinta, permitindo identificar opiniões, emoções e avaliações de clientes, podendo determinar se o tom emocional da mensagem é positivo, negativo ou neutro. Essa pesquisa teve como finalidade analisar comentários de clientes sobre produtos ou serviços de forma simplificada, minimizando o tempo levado para verificar uma quantidade massiva de dados, por meio de uma plataforma de código aberto. Os dados foram compilados por meio do software Orange, que proporciona diversos recursos de visualização como a ferramenta VADER, projetada para funcionar com textos curtos, informais e comentários online. Ao receber o *feedback* do cliente foi possível identificar pontos passíveis de melhorias, dos serviços e produtos, para atender as demandas, necessidades e expectativas dos consumidores. Verificou-se através deste trabalho que o processamento de linguagem natural (PLN), permite entender as emoções dos clientes, de forma rápida e eficaz, por meio de gráficos, a partir de seus *feedbacks*, sendo uma ferramenta de solução de melhoria empresarial.

PALAVRAS - CHAVE: comentários; avaliação; programa; informações; visualização; mineração.

SENTIMENT ANALYSIS OF AMAZON CUSTOMER FEEDBACK USING DATA MINING TECHNIQUES

ABSTRACT: With the exponential increase in feedback generated on digital platforms, data mining has become essential for a better evaluation of consumer sentiment. Sentiment analysis is a natural language processing (NLP) technique that extracts information from texts in a practical and concise manner, enabling the identification of customer opinions, emotions, and evaluations, and determining whether the emotional tone of the message is positive, negative, or neutral. This research aimed to

analyze customer comments on products or services in a simplified way, minimizing the time required to review a massive amount of data using an open-source platform. The data was compiled through Orange software, which provides various visualization tools such as the VADER tool, designed to work with short, informal texts and online comments. By gathering customer feedback, it was possible to identify areas for improvement in services and products to meet consumer demands, needs, and expectations. This study found that natural language processing (NLP) enables a quick and effective understanding of customer emotions through graphical representations of their feedback, serving as a tool for business improvement.

KEYWORDS: comments; evaluation; program; information; visualization; mining.

INTRODUÇÃO

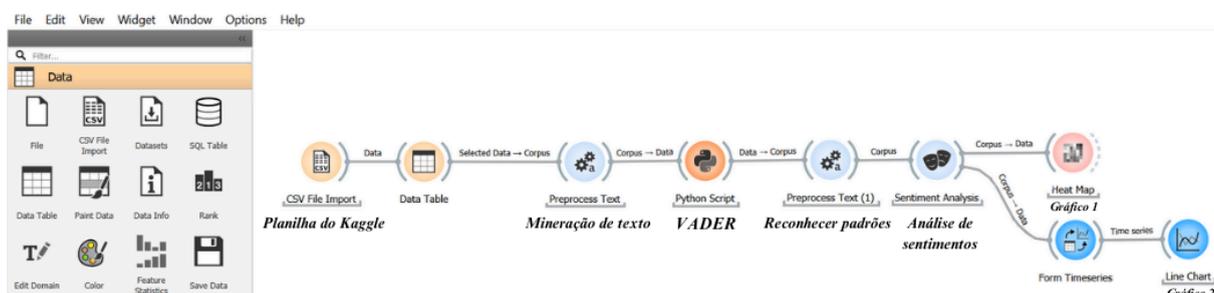
Devido ao aumento exponencial de dados gerados nas plataformas digitais, como em sites e bancos de dados, a análise de sentimentos se tornou uma ferramenta essencial na mineração de texto. A análise de sentimentos se trata de uma técnica de processamento de linguagem natural (PLN), que permite identificar e extrair informações pertinentes de textos de forma prática, sucinta, e visual, utilizando gráficos para interpretar opiniões, emoções e avaliações de clientes, determinando se o tom emocional das mensagens é positivo, negativo ou neutro.

Processamento de Linguagem Natural (PLN) refere-se a uma subárea da inteligência artificial que se concentra na interação entre computadores e linguagem humana. O objetivo do PLN é permitir que as máquinas compreendam, interpretem e gerem textos e falas em linguagem natural de uma maneira útil. (Indurkha and Damerau, 2010). A Mineração de Texto (ou *Text Mining*), por sua vez, é o processo de extrair informações úteis e *insights* a partir de grandes volumes de texto. Ela envolve técnicas de PLN para analisar e processar texto não estruturado e transformá-lo em dados estruturados. (Passos 2006).

Portanto, essa pesquisa tem como finalidade a compreensão dos sentimentos dos clientes em vista aos produtos da Amazon, partindo de uma avaliação de um conjunto de dados de *feedbacks* disponibilizada na base de dados públicos do Kaggle, utilizando a geração de uma análise de sentimentos que apresenta de forma visual um gráfico permitindo entender as informações a respeito dos *feedbacks* dos produtos.

MATERIAL E MÉTODOS

Para esta pesquisa foi elaborado um estudo a partir de uma planilha obtida na base de dados públicos Kaggle, contendo informações como: nome do produto, códigos, comentários e etc, a respeito da empresa Amazon. Após isso foi utilizado o Programa *Orange Data Mining*, o qual é um *software* que usa algoritmos para organizar e visualizar dados e reconhecer padrões. Dentro do programa Orange, através da biblioteca VADER (*Valence Aware Dictionary and Sentiment Reasoner*), uma ferramenta, em linguagem Python de código aberto, projetada para ser usada em tarefas de análise de sentimentos, foi possível analisar os dados a fim de extrair apenas as informações relevantes para a avaliação dos *feedbacks* dos clientes, contendo uma visualização através de gráficos.



Programa Orange - Fluxos até a formação dos gráficos de sentimentos.

Os resultados demonstraram que o VADER, é uma ferramenta eficiente para analisar sentimentos, especialmente em contextos curtos e informais, o que se revelou crucial para a compreensão das nuances nas interações online. A análise revelou padrões significativos, como a predominância de sentimentos positivos ou negativos em relação a determinados produtos, o que pode ajudar empresas a ajustar suas estratégias de marketing e atendimento ao cliente. Após a análise de sentimentos é gerado um gráfico com o intuito de visualização mais clara e de forma mais prática.



Gráfico heatmap (1) Produzido pelo Orange - Média de avaliação dos clientes

A imagem apresentada é um heatmap com dendrograma, que representa a análise de sentimentos realizada com o VADER. O dendrograma mostra a relação de similaridade entre as diferentes instâncias de dados, agrupando-as com base em uma métrica de similaridade.

No caso deste trabalho, ele agrupa sentenças ou documentos que têm perfis de sentimento semelhantes. Os ramos indicam a proximidade entre os grupos; quanto mais próximo dois ramos se encontram, mais semelhantes são os sentimentos desses grupos.

O heatmap encontrado também na bibliografia como mapa de calor, representa os valores de diferentes métricas de sentimento em cores. A escala de cor, que vai de azul até o amarelo, indica a intensidade das emoções ou sentimentos detectados.

As métricas de sentimento do VADER são:

Positive: Probabilidade de o texto ter um sentimento positivo.

Negative: Probabilidade de o texto ter um sentimento negativo.

Neutral: Probabilidade de o texto ser neutro.

Compound: Uma pontuação agregada que considera os valores positivos, negativos e neutros, resultando em uma avaliação final de sentimento (positivo ou negativo).

Considerando as cores utilizadas, teremos:

Azul: Representa valores mais baixos, que podem estar associados a uma alta negatividade ou baixa positividade.

Amarelo: Representa valores mais altos, indicando alta positividade ou baixa negatividade.

Verde: Representa valores intermediários, o que pode indicar neutralidade ou sentimentos mistos.

Cada linha horizontal representa uma instância de dado (um comentário dado pelo cliente). As diferentes colunas de cor mostram como cada instância se comporta em termos de positividade, negatividade, neutralidade e pontuação composta (compound).

O agrupamento hierárquico mostrado pelo dendrograma permite ver quais instâncias compartilham perfis de sentimento semelhantes, ajudando a identificar padrões em grandes conjuntos de dados.

RESULTADOS E DISCUSSÃO

Esta pesquisa investigou a eficácia da mineração de dados e análise de sentimentos, utilizando o VADER (*Valence Aware Dictionary and Sentiment Reasoner*), com o objetivo de avaliar a percepção do público em relação aos produtos da empresa Amazon. Através da aplicação do VADER, foi possível identificar e classificar sentimentos expressos em textos de comentários e postagens, proporcionando uma visão abrangente das opiniões dos consumidores. A investigação desses *feedbacks*, revelou vários *insights* significativos sobre as opiniões dos clientes..

Distribuição dos Sentimentos: Através dos gráficos se revelou uma distribuição de sentimentos em três categorias principais: positivas, negativas e intermediárias.. Cerca de 60 % dos comentários expressam sentimentos intermediários, o que indica no geral clientes satisfeitos parcialmente, seja por algumas falhas na qualidade do produto ou na entrega. Já 30% dos comentários expressam sentimentos positivos, onde os clientes estão plenamente satisfeitos. No entanto 5% dos comentários foram negativos, devido a problemas na entrega do produto ou a qualidade dele. E 5% das avaliações foram classificadas como neutras devido a textos mais descritivos e técnicos, sem uma carga emocional expressa. Os resultados sugerem que a maior parte dos sentimentos foi intermediária, que requer uma maior atenção para os aspectos negativos.

Sentimentos Neutros: As avaliações muitas vezes eram descritivas ou técnicas, sem uma carga emocional forte, indicando que muitos clientes preferem fornecer *feedbacks* mais objetivos e factuais.

Visto isso, a análise de sentimentos demonstrou ser eficaz na identificação e interpretação de sentimentos em *feedbacks* de produtos da Amazon. Evidenciando um pequeno índice de comentários negativos, porém, revela que em certos produtos dessa empresa é necessário uma melhor atenção a melhorias no processo e implementar soluções que possam melhorar a experiência do cliente. Essas recomendações também podem auxiliar a empresa a direcionar recursos para melhorar as áreas com maior impacto na percepção dos consumidores.



Gráfico de linha (2) produzido pelo Orange - Média de avaliação dos clientes.

Limitações

Apesar da eficácia do VADER, a pesquisa enfrentou algumas limitações que devem ser consideradas:

1. **Qualidade dos Dados:** O processo depende da qualidade dos dados coletados. Se os *feedbacks* forem limitados ou não representativos, os resultados podem não refletir com precisão a opinião geral dos clientes. A diversidade e a representatividade dos dados são cruciais para obter uma visão completa e precisa dos sentimentos dos consumidores.
2. **Sentimentos Ambíguos e Sarcasmo:** A identificação de sentimentos ambíguos ou sarcásticos continua sendo um desafio. Embora o VADER seja eficiente na avaliação de textos informais, pode não captar todas as nuances emocionais, especialmente em casos de ironia ou sarcasmo. Isso pode resultar em uma interpretação incorreta dos sentimentos expressos nas avaliações.
3. **Complexidade dos Textos:** Textos mais complexos e elaborados podem apresentar sutilezas emocionais que o VADER não consegue identificar completamente. A interpretação de sentimentos em textos mais complexos pode exigir ferramentas adicionais ou técnicas de PLN mais avançadas para uma compreensão mais precisa.

CONCLUSÕES

A análise de sentimentos, utilizada para fins de extração dos sentimentos percebidos nos *feedbacks* dos clientes se torna de suma importância para o crescimento e desenvolvimento de uma empresa, visto que evidenciam a experiência do consumidor e identificam áreas de melhoria. O programa Orange viabiliza a construção de fluxos que têm o potencial de otimizar as tarefas de análise de dados de uma empresa ou organização, a automatização desse trabalho torna o processo de extração de informações mais simples e prático, assim, otimizando o tempo de empresas e departamentos para uma obter uma devolutiva de informações atrativa e eficiente.

Nesse sentido, este estudo demonstra como a coleta, a transformação e a avaliação dos dados podem oferecer uma abordagem de relevância para as organizações que almejam identificar e compreender as percepções de seus clientes, transformando, assim, os dados em conhecimento estratégico e ampliando as possibilidades de sucesso empresarial.

CONTRIBUIÇÃO DOS AUTORES

R.S.F, como orientador, instruiu em todos os processos do trabalho. R.S.F e Y.C.G, como corretores do artigo. Y.C.G, V.S.A.O, N.S.D, C.B.S, procederam com a escrita do artigo, processamento dos dados, e geração da análise de sentimentos através do programa *Orange Data Mining*.

AGRADECIMENTOS

Nossos mais sinceros agradecimentos primeiramente a Deus, que fez com que nossos objetivos fossem alcançados. Ao Instituto Federal de São Paulo - Pirituba e ao nosso professor Regivaldo Ferreira, por todo auxílio e a oportunidade de desenvolver este projeto. Agradecemos também, ao Grupo de pesquisa de Informática Tecnologia Educação e Sociedade (GITES) por fornecer apoio e sua colaboração para que esse projeto fosse possível.

REFERÊNCIAS

IBM. O que é mineração de textos?. IBM, São Paulo, 2023. Disponível em: <<https://www.ibm.com/br-pt/topics/text-mining>> Acesso em 01 de maio de 2024.

BENEVENUTO, F, et al. Métodos para Análise de Sentimentos em mídias sociais. São Paulo. Disponível em: <<https://homepages.dcc.ufmg.br/fabricio/download/webmedia-short-course.pdf>>. Acesso em 08 de maio de 2024.

REIS, J. et al. Uma Abordagem Multiíngue para Análise de Sentimentos. Disponível em: <<https://sol.sbc.org.br/index.php/brasnam/article/view/6767/6660>>. Acesso 8 de maio de 2024

SALDAÑA, Z. Análise de sentimentos para exploração de dados. Disponível em: <<https://programminghistorian.org/pt/licoes/analise-sentimento-exploracao-dados#:~:text=No%20entanto%2C%20VADER%20tamb%C3%A9m%20foi>>. Acesso em: 9 Junho de 2024.

PASSOS, E. et al. A Tecnologia de Mineração de Textos. Disponível em: <<https://www.periodicosibepes.org.br/index.php/reinfo/article/view/171>>. Acesso 9 de junho de 2024.

INDURKHYA, N. et al. Handbook of Natural Language Processing. New York, 2010.

ISHAK, A. et al. Uso do software Orange no método de classificação de mineração de dados nas lentes do conjunto de dados/ Orange Software Usage in Data Mining Classification Method on The Dataset Lenses. Disponível em <<https://iopscience.iop.org/article/10.1088/1757-899X/1003/1/012113/meta>> . Acesso em 9 junho de 2024.