

14º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2023

MINERAÇÃO DE DADOS DE FEEDBACKS DE CLIENTES POR MEIO DA APRESENTAÇÃO DA NUVEM DE PALAVRAS.

YASMIM C. GAMBOA¹, NATALY S. DIAS², VITÓRIA S. A. OLIVEIRA³, CAMILA B. DE SOUZA⁴, REGIVALDO S. FERREIRA⁵, ADRIANO J. FERRUZZI⁶

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, nataly.souza@aluno.ifsp.edu.br

³ Estudante do curso Técnico em Logística integrado ao Ensino Médio, IFSP, Campus Pirituba, v.amaral@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

⁶ Docente do IFSP, Campus Pirituba, adrianoferruzzi@ifsp.edu.br

Área do conhecimento (Tabela CNPq) 1.03.03.04-9 Sistemas de Informação

RESUMO: Visto que a avaliação dos consumidores é de suma importância para qualquer empresa, este trabalho visa extrair essas informações a respeito dos produtos, a partir da ferramenta de mineração de dados, KNIME, e através dele criar filtros/nichos a fim de retirar informações desnecessárias para esse trabalho, formando uma visualização dessas opiniões de uma forma simples, transformando em uma nuvem de palavras. A escolha do KNIME é respaldada por análise comparativa em outros trabalhos, A nuvem de palavras é utilizada pela sua viabilidade visual na identificação de frequências de palavras podendo carregar consigo percepções iniciais do processo de mudanças no modo de ação no processo comercial. Ao compreender a dimensão da grande escala, torna-se indispensável ter uma ferramenta de processamento de Linguagem Natural para otimizar a obtenção de informações uma vez que se torna inviável depender da análise manual.

PALAVRAS-CHAVE: Avaliação do consumidor; Mineração de Dados; KNIME; Nuvem de Palavras

CUSTOMER FEEDBACK DATA MINING THROUGH WORD CLOUD PRESENTATION

ABSTRACT: Given that consumer feedback is of paramount importance to any company, this study aims to extract this information regarding products using the data mining tool, KNIME, and employ it to create filters/niches to remove unnecessary information for this study. This process results in a visualization of these opinions in a straightforward manner, ultimately transforming them into a word cloud. The choice of KNIME is supported by comparative analysis in other studies. The word cloud is utilized due to its visual feasibility in identifying word frequencies that may carry insights that engender potential shifts in the course of action in the business process. Understanding the scale of the data, it becomes indispensable to have a Natural Language Processing tool to optimize information extraction, as relying solely on manual analysis becomes unfeasible.

KEYWORDS: Consumer assessment; Data mining, KNIME; Word Cloud.

INTRODUÇÃO

Considerando o crescimento do comércio virtual, obter informações das opiniões voluntárias dos clientes pode ser um diferencial na busca por melhoria. De acordo com (Muduc, et al, 2014), o *feedback* é fundamental para a garantia do crescimento permitindo direcionar atitudes, comportamentos e desempenhos. Com o passar do tempo esses *feedbacks* tendem a se acumular tornando inviável a avaliação humana. Visando trabalhar em bases de dados massivas, faz-se necessário a utilização de ferramentas que possibilitam a mineração de dados.

Esta pesquisa tem como finalidade verificar a funcionalidade da aplicação de técnicas de processamento de linguagem natural a fim de compreender as opiniões dos clientes de produtos Amazon avaliando uma base de dados de *feedbacks* disponibilizada na base de dados públicos do Kaggle, utilizando a geração de nuvem de palavras que apresenta de forma visual palavras mais comuns no texto permitindo entender as informações nelas contidas. A combinação dessas técnicas pode ser aplicada em diversas áreas, como análise de sentimento, monitoramento da reputação da marca, entre outros. Este trabalho mostra como a análise de texto, com foco em elaborar nuvens de palavras, pode ser aplicada na prática quando o objetivo é extrair informações úteis de grandes conjuntos de dados.

MATERIAL E MÉTODOS

Para a realização desta pesquisa foi utilizado o KNIME, cujo nome corresponde a uma abreviação de Konstanz Information Miner. Ele é uma plataforma de análise e mineração de dados de código aberto (Berthold et al., 2009). A escolha do Knime foi baseada no trabalho de (Naik; Samant, 2016), que apresenta o resultado comparativo entre os *softwares* WEKA, Orange, KNIME, RapidMiner e Tanagra e expõe o Knime como melhor em medidas de acurácia quando comparado aos outros e conclui sobre a escolha da ferramenta e a possibilidade de afetar significativamente os resultados.

Além disso, o KNIME oferece uma gama de ferramentas cujo objetivo é a análise de dados, incluindo recursos de processamento de linguagem natural. Baseado no trabalho de Fayyad, Piatetsky-Shapiro e Smyth (1996, p. 37), a mineração de dados é o processo de descoberta de conhecimento a partir de informações valiosas, extraídas por meio de padrões e conhecimentos úteis em determinado conjuntos de dado. A pesquisa de cunho descritiva e explicativa, buscará minerar os dados existentes em uma tabela de dados com um total aproximado de 725628 palavras, disponibilizada no site Kaggle contendo os *feedbacks* dos clientes. Para obter uma análise sobre o produto adquirido na visão do cliente, foi utilizado a estrutura de *pipeline* permitindo integrar, transformar, analisar, minerar e visualizar dados de maneira fácil e intuitiva. A ferramenta usa uma abordagem visual com a finalidade de construir fluxos de trabalho, facilitando a criação de uma série de etapas ou processos pelos quais os dados passam. Os dados coletados e transformados em fluxo serão analisados e caracterizados por seu nicho, e finalmente será criada a nuvem de palavras.

A nuvem de palavras é uma técnica que possibilita encontrar e compreender informações relativas à opinião dos clientes em frases e comentários que usam a linguagem natural e que, normalmente, não são encontradas de forma usual. Com o uso dessa técnica é possível criar uma representação visual das palavras mais frequentes em um conjunto de textos, possibilitando encontrar os termos mais usados (Heimerl, 2014).

RESULTADOS E DISCUSSÃO

A avaliação dos clientes é de suma importância em qualquer empresa, com propósito de melhoria e desenvolvimento do produto no mercado, levando em consideração as preferências e necessidades dos consumidores. Nos últimos anos, o avanço da tecnologia e o surgimento de plataformas online de comércio têm gerado uma quantidade massiva de dados de *feedbacks* de clientes. Empresas de vários setores têm investido em estratégias para compreender melhor esses *insights* de clientes, a fim de melhorar seus produtos e serviços. Nesse contexto, o processamento de

linguagem natural (NLP - Natural Language Processing) surge como uma ferramenta poderosa na extração de informações relevantes a partir desses dados textuais.

As ferramentas especializadas em processamento de linguagem natural podem ser uma abordagem eficiente na busca por percepções e sentimentos dos clientes. A partir do uso do KNIME, é possível extrair informações significativas de grandes volumes de dados textuais, sendo especialmente útil quando é usado para fazer análises de *feedbacks* pós-compra. Esses dados são normalmente não-estruturados e requerem tratamento com o propósito de serem compreendidos e analisados. Após a coleta dos dados, o próximo passo é pré-processá-los. A utilização do KNIME permite oferecer uma série de opções almejando pré-processar dados textuais, como remoção de *stopwords* (palavras consideradas irrelevantes), lematização (transformar palavras na sua forma base) e tokenização (dividir o texto em unidades significativas, como palavras ou frases). Essas etapas são cruciais na melhoria da qualidade dos dados e eliminar informações irrelevantes que podem afetar a análise.

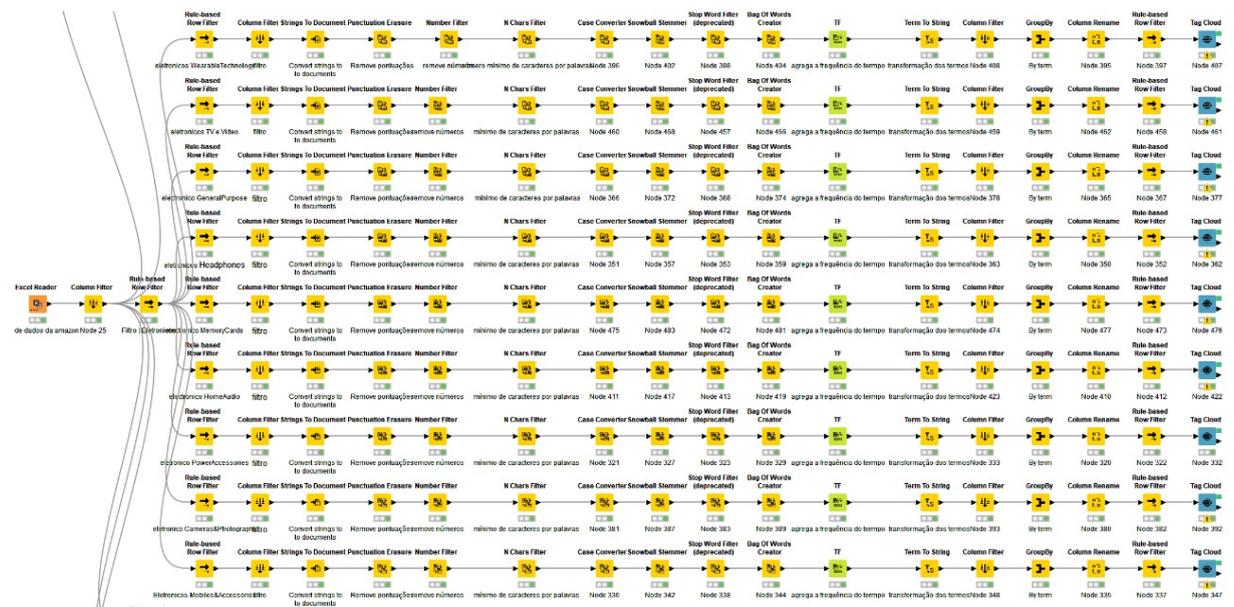


FIGURA 1. Etapa do processamento da mineração de textos no *workflow* na plataforma KNIME.

A Figura 1 apresenta 8 diferentes fluxos de trabalho, cada um deles resulta em uma nuvem de palavras diferentes seguindo as próprias especificidades. Cada nó foi configurado individualmente no objetivo de garantir as peculiaridades a partir de cada filtro realizado. Foram consideradas 8 categorias de produtos destinados a serem analisados no formato de nuvem de palavras possíveis de serem visualizadas na *Tag Cloud* existente no final de cada ligação.

Após os dados serem pré-processados, o KNIME permite a análise dos textos através da elaboração de uma nuvem de palavras. O objetivo principal é resumir de forma estatística o conteúdo de um texto baseado no tamanho da apresentação das palavras. Quanto maior a frequência de uma palavra, maior sua representação na nuvem. A Figura 2 apresenta a nuvem de palavras criada e seus destaques.

Essa representação visual permite identificar as palavras-chave mais comuns nos *feedbacks* dos clientes, fornecendo *insights* valiosos sobre suas preferências, expectativas e necessidades. Além da nuvem de palavras, o KNIME também possibilita a análise de sentimentos presentes nos *feedbacks*. A análise de sentimentos é uma técnica NLP que utiliza algoritmos de aprendizado de máquina almejando identificar a polaridade do sentimento expresso em um texto, como positivo, negativo ou neutro. Com essa análise, é possível compreender a satisfação geral dos clientes em relação aos produtos e serviços oferecidos.

REFERÊNCIAS

BERTHOLD, M. R. et al. KNIME - the Konstanz information miner: Version 2.0 and beyond. SIGKDD explorations: newsletter of the Special Interest Group (SIG) on **Knowledge Discovery & Data Mining**, v. 11, n. 1, p. 26–31, 2009.

BILGIN, M. ; KOKTAS, H. Sentiment Analysis with Term Weighting and Word Vectors. **The International Arab Journal of Information Technology**, Vol. 16, No. 5, September 2019

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37-37, 1996.

HEIMERL, Florian et al. Word cloud explorer: Text analytics based on word clouds. In: 2014 47th Hawaii international conference on system sciences. **IEEE**, 2014. p. 1833-1842.

MUDUC, Sabina et al. A importância do feedback no processo de avaliação de desempenho, como forma de gerar valor. **Órgãos da PROELIUM**, 2014.

NAIK, A.; SAMANT, L. Correlation review of classification algorithms using data mining tools: WEKA, rapidminer, tanagra, orange and KNIME. **Procedia computer science**, v. 85, p. 662–668, 2016.