

A mineração de opinião em mídias sociais como ferramenta para medir a (in)satisfação do consumidor

Opinion mining in social media as a tool to measure consumer (in)satisfaction

La minería de opinión en los medios sociales como herramienta para medir la (in)satisfacción del consumidor

Luis Sancliment Iglesias
Universidade Federal do Paraná
Brasil

Denise Fukumi Tsunoda
Universidade Federal do Paraná
Brasil

Submetido em: 26/03/2021

Aceito em: 14/06/2021

Publicado em: 28/10/2021

Licença:



Autor para correspondência: Luis Sancliment Iglesias

Email: luissiglesias@hotmail.com

ORCID: <https://orcid.org/0000-0002-0256-8869>

Como citar este artigo:

IGLESIAS, Luis Sancliment; TSUNODA, Denise Fukumi. A mineração de opinião em mídias sociais como ferramenta para medir a (in)satisfação do consumidor. **REBECIN**, São Paulo, v. 8, edição especial, p. 01-13, 2021. DOI: 10.24208/rebecin.v8i.235

RESUMO

Estudo de natureza quantitativa que objetiva explorar a contribuição da mineração de opinião em bases de dados extraídas do Facebook para a medição da (in)satisfação dos consumidores. Visa propor um fluxo que auxilie nas etapas do processo de descoberta de conhecimento em texto e selecionar ferramentas para a mineração de opinião a nível de sentença, onde se analisa o sentimento positivo, negativo e neutro. Submete-se ao proposto fluxo uma base de dados, de marca de carros, extraída do Facebook com quatro tratamentos de pré-processamento. Utilizam-se os algoritmos Naïve Bayes, SMO e J48 na ferramenta Weka para a etapa de processamento. Apresenta resultados satisfatórios na mineração de opinião com melhor taxa de acerto obtida usando o algoritmo SMO. Propõe trabalhos futuros em bases de dados SAC (Serviço de Atendimento ao Consumidor) com a aplicação desta metodologia desenvolvida e estudos de descobrimento das causas de (in)satisfação dos consumidores encontradas em bases de dados SAC e SAC 2.0.

Palavras-Chave: Árvore de decisão; Análise de sentimento; Mineração de Texto; Mineração de dados.

ABSTRACT

A quantitative study that aims to explore the contribution of opinion mining in databases extracted from Facebook for the measurement of consumer (in)satisfaction. It aims to propose a flow that assists in the steps of the process of knowledge discovery in text and select tools for opinion mining at the sentence level, where the positive, negative and neutral sentiment is analyzed. A car brand database extracted from Facebook with four pre-processing treatments is submitted to the proposed flow. The Naïve Bayes, SMO and J48 algorithms in the Weka tool are used for the processing stage. It presents satisfactory results in opinion mining with the best hit rate obtained using the SMO algorithm. It proposes future work in CSC (Consumer Care Service) databases with the application of this developed methodology and studies to discover the causes of consumer (in)satisfaction found in CSC and CSC 2.0 databases.

Keywords: Decision tree; Sentiment analysis; Text mining; Data mining.

RESUMEN

Estudio de carácter cuantitativo que pretende explorar la contribución de la minería de opinión en bases de datos extraídas de Facebook para la medición de la (in)satisfacción del consumidor. Su objetivo es proponer un flujo que ayude en los pasos del proceso de descubrimiento de conocimiento en el texto y seleccionar herramientas para la minería de opinión a nivel de frase, donde se analiza el sentimiento positivo, negativo y neutro. Se somete al flujo propuesto una base de datos, de marca de coches, extraída de Facebook con cuatro tratamientos de preprocesamiento. Los algoritmos Naïve Bayes, SMO y J48 se utilizan en la herramienta Weka para la etapa de procesamiento. Presenta resultados satisfactorios en la minería de opinión, con la mejor tasa de aciertos obtenida con el algoritmo SMO. Propone trabajos futuros en bases de datos SAC (Servicio de Atención al Consumidor) con la aplicación de esta metodología desarrollada y estudios para descubrir las causas de la (in)satisfacción del consumidor encontradas en las bases de datos SAC y SAC 2.0.

Palabras clave: Árbol de decisión; Análisis de sentimientos; Minería de textos; Minería de datos.

1 INTRODUÇÃO

Com o avanço das tecnologias de comunicação e a globalização mundial em proporções cada vez maiores, as mídias sociais são ferramentas de descentralização e veiculação de dados e informações. Dentro destas, as redes sociais digitais têm o objetivo de aproximar pessoas com interesses comuns e permitem que estas expressem suas opiniões e sentimentos a respeito dos mais diversos assuntos: política, religião, livros, gastronomia, educação e outros. Estas opiniões podem por exemplo, expressar (in)satisfação sobre os produtos ou serviços consumidos e até mesmo questionar ou buscar informações para a tomada de decisão na aquisição (ou não) de produtos e serviços ainda não conhecidos.

O conjunto de opiniões viabiliza criação de bases de dados advindos destas mídias sociais pode ser transformado em valor para as organizações. Com isso, forma-se um canal direto para que as empresas que estão atentas e presentes nas diferentes mídias sociais se beneficiem destas opiniões para aproximar-se de seus consumidores. As opiniões permitem além de medir o grau de (in)satisfação dos mesmos, responder mais rapidamente às solicitações, questionamentos e reclamações, bem como aumentar a competitividade das organizações.

No dia 25 de março de 2021 foi realizada uma busca na base Dimensions¹ com os descritores “mineração de opinião” e “mídias sociais”. Foram recuperados 23 trabalhos, dos quais 19 artigos em periódicos e 4 em anais de eventos. Destes 23, 10 (43,48%) são de 2019 e 2020, mas todos estão no período de 2015 até 2020. No mesmo dia e ainda na Dimensions, com os descritores “mineração de opinião” e “mídia social” foram recuperados 8 artigos mas, todos eles, já haviam sido recuperados anteriormente. Com os termos em inglês “opinion mining” e “social media” foram recuperados 19.053 artigos, dos quais 108 trabalhos do professor Erik Cambria da Nanyang Technological University de Singapura. De todos os artigos, 10.203 são da grande área Information and Computing Sciences e 1.631 foram publicados na Lecture Notes in Computer Science.

Este artigo sumariza uma pesquisa realizada para investigar a eficácia da utilização da mineração de opinião como ferramenta para medir a satisfação consumidores manifestadas em redes sociais. Foi escolhido Facebook por ter o maior número de usuários ativos (2,74

¹ DIMENSIONS. Disponível em: <https://www.dimensions.ai/>. Acesso em: 26 mar. 2021.

bilhões), seguido pelo YouTube (2,29 bilhões) e WhatsApp (2,00 bilhões)².

2 REFERENCIAL TEÓRICO

Lima *et al.* (2019) afirmam que as mídias sociais mudaram a forma de comunicação tanto entre as pessoas quanto com as empresas, uma vez que a comunicação online passou a ser utilizada e, inclusive, novos papéis como os dos influenciadores digitais foram popularizados.

Recuero (2009) afirma que o advento da comunicação mediada pelo computador e congêneres influencia não apenas na sociedade, mas também na vida cotidiana por meio das novas formas de conexões, estabelecer relações e formar comunidades. Principalmente com o advento da COVID-19, as pessoas passaram a utilizar cada vez mais estes espaços de interação e relacionamento online para se comunicar, buscar informações, formar opiniões, negociar (inúmeras empresas atuam principalmente pelas mídias sociais) e muito mais.

Segundo Lima *et al.* (2019), satisfação é um conceito multidimensional pois é pessoal, complexo e sujeito a diversas variáveis. Uma simples pergunta tal como: você prefere verão ou inverno? Pode admitir diversas respostas, com suas diversas justificativas. Por este motivo, Merlo e Ceribeli (2014) apresentam um constructo multidimensional que envolve atendimento às expectativas, ausência de arrependimento, emoções positivas percebidas, desempenho do produto e equidade percebida quando o consumidor percebe que realizou um bom negócio ao comparar sua experiência com outros indivíduos.

² STATISTA. Disponível em: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>. Acesso em: 25 mar. 2021.

A satisfação das pessoas está diretamente relacionada às expectativas sobre este “algo” que pode ser produto, serviço, pessoa e outros. Uma empresa com foco no consumidor, mantém preocupação com a identificação das expectativas dos seus clientes e estabelece estas expectativas como objetivos para a definição de estratégias de atuação, dos produtos e serviços a serem ofertados, experiência de compra, troca, devolução e diversos outros aspectos relacionados ao planejamento estratégico de marketing.

Segundo Kotler (2012), a satisfação tanto pode consistir em um sentimento de prazer como de desapontamento, este sentimento resulta da comparação entre o desempenho de um produto e as expectativas existentes no consumidor. Quando o desempenho do produto vai além das expectativas o consumidor ficará altamente satisfeito. A fidelidade dos consumidores para com uma marca fica diretamente ligada às percepções a respeito de um produto de uma marca à qual eles constroem sentimentos favoráveis.

As mídias sociais podem representar diferencial competitivo com informações, notícias, imagens, vídeos e em alguns casos mais específicos como o Facebook, conteúdo significativo em texto, tanto de pessoas como de organizações que escrevem mensagens contendo opiniões, comentários e reclamações dos mais variados assuntos. A análise de sentimentos se usa com frequência na mineração de opinião para poder identificar sentimentos, afetos, subjetividade e demais estados emocionais destes textos online. Uma dificuldade é que estes textos se encontram em linguagem natural, isto é, na linguagem razoavelmente inteligível para o ser humano, mas não para as máquinas. Para Schiessl e Bräscher (2011), devido à complexidade da linguagem natural para a interpretação direta das máquinas, é necessário fazer uma extração de

conhecimento das bases textuais e criar agrupamentos e modelos de classificação automatizados para que possam ser interpretados por computadores.

Conforme Liu (2015), a análise de sentimentos se conduz em três níveis. O primeiro nível é o de documento, onde se classifica primeiramente todo o documento para saber se está expressando um sentimento positivo ou negativo. O segundo nível é o de sentença, onde se analisa se cada frase expressa uma opinião positiva, negativa ou neutra. O terceiro nível é o de aspecto, onde diferentemente dos níveis um e dois, que em nenhum caso as análises denotam se as pessoas gostam precisamente ou não, no nível três a análise realiza esta diferenciação.

3 RESULTADOS E DISCUSSÃO

Seguindo a classificação de Liu (2015), o nível de análise de sentimento que foi adotado neste trabalho é o segundo nível denominado nível de sentença, que visa a análise das frases de consumidores com opiniões de subjetividade positiva, negativa e neutra. Para as opiniões em que os consumidores não expressam claramente uma opinião, serão consideradas como opiniões neutras.

A base de dados utilizada para este trabalho foi retirada da página oficial da empresa Ford Brasil no Facebook, especificamente dos comentários da postagem de 30 de julho de 2018, sobre a nova linha Ford Ka 2018 que contava quando foi capturada a base com 8 mil visualizações, 64 mil curtidas, 1.169 compartilhamentos e 1.114 comentários.

Desta base original foram extraídas quatro bases, conforme Tabela 1. Todas as bases contêm dois atributos, sendo o primeiro do tipo string (texto a analisar) e o segundo do tipo *phrasesentiment* (positivo, neutro ou negativo).

Os algoritmos escolhidos para a mineração da base selecionada dentro do software Weka foram o Naïve Bayes que usa o modelo de classificação probabilístico, o SMO (Sequential Minimal Optimization) que utiliza o modelo baseado em função e o J48 que é um algoritmo baseado em árvores de decisão. Estes algoritmos foram selecionados devido a serem mais comumente utilizados e mais adequados ao tipo de base de dados que está sendo analisada. Os experimentos em todas as bases analisadas foram realizados com validação cruzada de 10 partições, com a utilização dos filtros “FilteredClassifier” e “StringToWordVector” e com o uso de “tokenizer”.

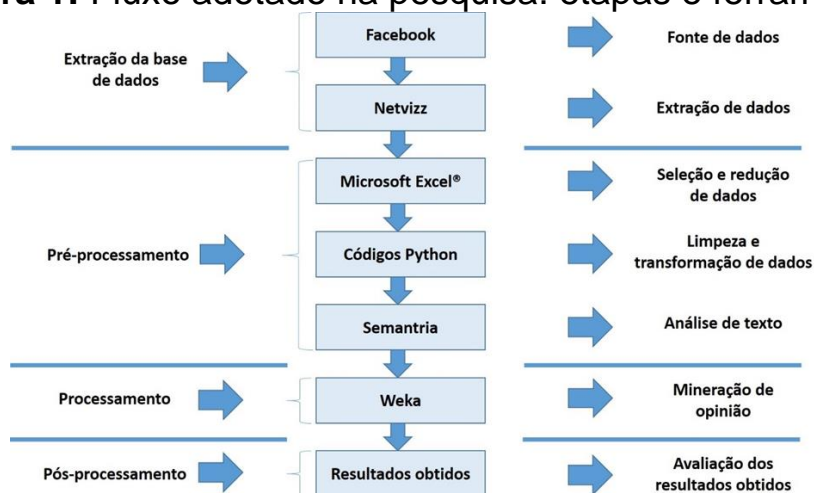
Tabela 1: características das bases de dados analisadas

Base de Dados	Descrição da Base de Dados	Nº de Instâncias	Nº de Atributos
Ford_Dados_Brutos	Base de dados sem nenhum tratamento dos dados.	430	2
Ford_Com_StopWords	Base de dados com padronização de minúsculas, abreviações, emoticons, retirada de caracteres especiais e de acentuação. Mantendo-se os StopWords.	1241	2
Ford_Sem_StopWords _sem_ao	Base de dados com padronização de minúsculas, abreviações, emoticons, retirada de caracteres especiais e de acentuação. Retirando-se os StopWords, inclusive o "ao".	806	2
Ford_Sem_StopWords _com_ao	Base de dados com padronização de minúsculas, abreviações, emoticons, retirada de caracteres especiais e de acentuação. Retirando-se os StopWords, exceto o "ao".	843	2

Fonte: Os autores (2018).

O fluxo do trabalho explicitando etapas, ferramentas e as funcionalidades destes está apresentado na Figura 1.

Figura 1: Fluxo adotado na pesquisa: etapas e ferramentas



Fonte: Os autores (2018).

A Tabela 2 apresenta as taxas de acertos de cada um dos métodos utilizados e percebe-se o algoritmo SMO obteve melhores resultados de taxas de acertos nas 4 bases submetidas para análise. Para a base bruta o percentual de taxa de acerto de instâncias classificadas corretamente foi de 73,3% utilizando o algoritmo SMO, 68,4% com o algoritmo J48 e 67,4% com o algoritmo Naïve Bayes.

Tabela 2: Taxas de acertos dos três métodos utilizados

Bases de Dados	Naïve Bayes	SMO	J48
Ford_Dados_Brutos	67,4%	73,3%	68,4%
Ford_Com_StopWords	76,2%	93,4%	86,7%
Ford_Sem_StopWords_sem_nao	69,9%	83,9%	77,9%
Ford_Sem_StopWords_com_nao	68,6%	87,3%	76,6%

Fonte: Os autores (2018)

A respeito dos resultados de grau de confiabilidade intermediária atribuído pela estatística Kappa, para a base de dados bruta o grau encontrado foi “moderado” para os 3 algoritmos. Para a base em que se

manteve os *stopwords* o grau foi “muito bom” para o algoritmo SMO, “bom” para o algoritmo J48 e “moderado para o algoritmo Naïve Bayes. Para a base que teve a retirada de *stopwords* inclusive o “nao” o grau encontrado foi “bom” para os algoritmos SMO e J48 e “moderado” para o algoritmo Naïve Bayes. Para a base que teve a retirada de *stopwords* excetuando o “não”, o grau foi “bom” unicamente para o algoritmo SMO e “moderado” para os algoritmos J48 e Naïve Bayes.

Para verificação dos indicadores de satisfação do consumidor, foram avaliados os termos negativos e positivos classificados na análise do Semantria por meio da frequência das palavras e a presença de alguns objetivos de desempenho.

Tomando como exemplo a Base Ford_Dados_Brutos que gerou 430 instâncias, sendo 137 negativas, 196 neutras e 97 positivas, as maiores frequências são encontradas nas palavras: “problema” com 27 ocorrências (19,7%), “problemas” com 7 ocorrências (5%), “defeito” com 6 ocorrências (4%), “reclamações” com 5 ocorrências (3%), “falta” com 4 ocorrências (3%), “péssimo” com 3 ocorrências (2%) e “prejuízo” com 3 ocorrências (2%). Destes 7 (sete) termos com maior frequência, 2 deles denotam relação direta com objetivos de desempenho: “defeito”, atrelado ao objetivo qualidade, que se relaciona com a não conformidade de especificações de produto e “prejuízo”, que pode estar relacionado com custo, que também é um dos cinco objetivos de desempenho nos indicadores de satisfação.

Uma vez que a teoria afirma que a qualidade está relacionada às expectativas do consumidor, a constatação dos termos “lindo”, “bom”, “confortável” e “conforto” (quando gerada a nuvem de palavras com os 97 termos positivos) pode indicar a satisfação do consumidor com os produtos e serviços da empresa em questão.

Figura 2: Nuvem de palavras dos 97 termos positivos da base Ford_Dados_Brutos



Fonte: Os autores (2018)

Desta forma, a fidelidade dos consumidores para com uma marca, diretamente relacionada às percepções a respeito de um produto de uma marca à qual eles constroem sentimentos favoráveis (KOTLER, 2012) pode ser observada pela presença de termos “bom”, “lindo”, “sonho”, “maravilhoso” e “amo”, entre outros na nuvem de palavras (FIGURA 2).

4 CONSIDERAÇÕES FINAIS

Uma vez que a quantidade de informações contidas na Web tem um crescimento constante e diante desse enorme volume de dados a dificuldade de pessoas poderem explorá-las é cada vez maior, a utilização de máquinas e processamento de linguagem natural por meio do uso da mineração de opinião pode ser uma alternativa muito útil para mensurar a (in)satisfação dos consumidores de forma mais ágil e prática. Percebe-se, pela pesquisa realizada nas bases de periódicos, o interesse

crescente de pesquisadores no tema mas a pouca representatividade dos pesquisadores brasileiros no panorama mundial.

Considera-se que, para fins acadêmicos, as bases utilizadas e os diversos procedimentos que envolveram o pré-processamento, a mineração de opinião e a avaliação dos resultados obtidos, poderá destacar a importância do monitoramento do grau de (in) satisfação dos clientes quanto aos produtos e serviços fornecidos pelas organizações. Ao mesmo tempo, com a descoberta de conhecimento em dados pela mineração de opinião e os padrões encontrados nas análises efetuadas, é possível serem geradas informações que podem ser de utilidade e valor para as empresas, tais como: a busca de tendências de mercado, requisitos de clientes, aprimoramentos e melhorias, bem como recomendações e sugestões que os consumidores deixam registrados nas mídias sociais.

Como sugestão para trabalhos futuros, recomenda-se aplicar este estudo utilizando-se uma base de dados SAC (Serviço de Atendimento ao Consumidor), onde existem algumas características distintas das encontradas em bases SAC 2.0 como linguagens mais padronizadas e especializadas (linguagens documentais) e registros textuais que permitam um menor esforço de pré-processamento por serem efetuados por profissionais especializados e treinados para conduzir este trabalho.

REFERÊNCIAS

KOTLER, P. **Administração de marketing**. 10. ed. São Paulo: Afiliada, 2012.

LIMA, A. P. L. *et al.* **Comportamento do consumidor**. Porto Alegre: SAGAH, 2019.

LIU, B. **Sentiment analysis**: mining opinions, sentiments and emotions. New York: Cambridge University, 2015.

MERLO, E.; CERIBELI, H. **Comportamento do consumidor**. Rio de Janeiro: LTC, 2014.

RECUERO, R. **Redes Sociais na Internet**. Porto Alegre: Ed. Sulina, 2009.

SCHIESSL, M.; BRÄSCHER, M. Descoberta de conhecimento em texto aplicada a um sistema de atendimento ao consumidor. **Revista Ibero-Americana de Ciência da Informação**, v. 4, n. 2, 2011. Disponível em: <http://periodicos.unb.br/ojs311/index.php/RICI/article/view/1682/1481>. Acesso em: 23 fev. 2021.