

# **Análise comparativa das principais plataformas de reclamações online: implicações para análise de mídia social em negócios**

**Gustavo Nogueira de Sousa<sup>1</sup>, Isabelle Guimarães<sup>2</sup>, Antonio F. L. Jacob Jr.<sup>1</sup>, Fábio M. F. Lobato<sup>1,2</sup>**

<sup>1</sup>Programa de Pós-Graduação em Engenharia da Computação e Sistemas –  
Universidade Estadual do Maranhão (UEMA)  
São Luís – Maranhão – Brasil

<sup>2</sup>Instituto de Engenharia e Geociências – Universidade Federal do  
Oeste do Pará (UFOPA) – Santarém – PA – Brasil

{sougusta, isabelle.guimaraes2}@gmail.com,

antoniojunior@professor.uema.br, fabio.lobato@ufopa.edu.br

**Abstract.** *New forms of relationship between companies and customers have been introduced through the massive use of social media, and have transformed the way in which customers make purchasing decisions. This new reality explained the importance of content analysis related to brands and products published by consumers on social media platforms. In this sense, this article presents a comparative analysis of the two largest online complaint platforms in Brazil, ReclameAqui and Consumidor.gov. The analyzes use content from textual data mining of consumer complaints from five major companies in the Brazilian ecommerce sector, in order to provide a basis for understanding the challenges and opportunities in the analysis of these social media in business. The results show that the way in which companies operate on these platforms must be specific to each platform, since each platform has its particularities, such as consumer group, content and main types of problems.*

**Keywords.** *Mídias Sociais; Mineração de Dados; Modelagem de Tópicos.*

**Resumo.** *Novas formas de relacionamento entre empresas e clientes foram introduzidas através do uso massivo de mídias sociais, e têm transformado a forma com a qual clientes tomam decisões de compras. Esta nova realidade explicitou a importância da análise do conteúdo relacionado a marcas e produtos publicados por consumidores em plataformas de mídias sociais. Nesse sentido, este artigo apresenta uma análise comparativa das duas maiores plataformas de reclamação online no Brasil, o ReclameAqui e o Consumidor.gov. Nas análises são utilizados os conteúdos provenientes de mineração de dados textuais de reclamações de cinco grandes empresas do setor de ecommerce brasileiro, com o objetivo de fornecer uma base para compreender os desafios e oportunidades nas análises destas mídias sociais em negócios. Os resultados mostram que a forma de atuação das empresas nessas plataformas deve ser específica para cada plataforma, pois existem particularidades, tais como grupo de consumidores, conteúdo e principais tipos de problemas.*

**Palavras-chave.** *Social Media; Data Mining; Topic Modeling.*

## 1. Introdução

O surgimento de novas ferramentas de comunicação modificaram a forma com a qual as pessoas tomam suas decisões de compra [Orenga-Roglá and Chalmeta 2016, Lobato et al. 2017]. Os canais de compartilhamento de informações online facilitam o diálogo entre consumidores de todos os lugares, independentemente de suas posições geográficas [Constantinides and Holleschovsky 2016]. E, assim como o boca-a-boca tradicional se faz presente na sociedade, o ambiente virtual tornou-se um dos meios de disseminação e troca de experiências a respeito de produtos e serviços [Kim and Johnson 2016]. Como consequência, surgiu o conceito de boca-a-boca virtual (*Electronic Word-of-Mouth, eWoM*), definido como o ato de compartilhar opiniões sobre produtos e serviços em mídias sociais [Schmäh et al. 2017].

O eWoM tem se mostrado ainda mais eficaz que o "boca-a-boca" tradicional, devido a sua facilidade de propagação e de engajamento [Constantinides and Holleschovsky 2016]. Por este motivo, os consumidores passaram a ser chamados de *prosumers*, uma vez que são responsáveis tanto pelo consumo quanto pela produção de conteúdo [Roy et al. 2019]. Ao estudar conteúdos de eWoM, as empresas têm uma rica fonte de conhecimento sobre seus clientes e quais são suas preferências [Aldous et al. 2019]. Porém, a sobrecarga de conteúdo torna o entendimento individual dos clientes um grande desafio [Lobato et al. 2017].

Na literatura, há diversos estudos que utilizam dados de revisões e de reclamações online obtidos de diferentes plataformas de mídias sociais tais como *Twitter, Facebook, TripAdvisor, Booking.com* e entre outras [Bahtar and Muda 2016, Silva et al. 2017, Xiang et al. 2017, Vermeer et al. 2019]. Geralmente, estes estudos utilizam em suas análises um conjunto amostral de dados para extrair informações que possibilitam a detecção, descrição ou previsão de padrões que influenciam na tomada de decisões teóricas e/ou práticas. Entretanto, há poucos estudos tal como o de [de Almeida et al. 2019] que utilizam em suas análises plataformas específicas para publicações de reclamações online de marcas e/ou produtos.

Com isto em mente, este estudo analisa comparativamente o conteúdo textual das reclamações de duas plataformas de reclamações online, a saber o ReclameAqui (RA) e o Consumidor.gov (CGOV). O objetivo é prover uma base para a compreensão dos desafios e oportunidades que estas plataformas proporcionam para o desenvolvimento de análises de mídias sociais que orientem a tomada de decisões em diversos segmentos do mercado.

O restante do artigo encontra-se organizado como segue. Na Seção 2 são apresentados os trabalhos relacionados, na Seção 3 é descrito processo de condução da pesquisa, na Seção 4 são apresentados os resultados decorrentes das análises e por fim na Seção 5 o autor apresenta suas considerações finais ao desenvolvimento do estudo.

## 2. Trabalhos Relacionados

O engajamento de consumidores em múltiplas plataformas de mídias sociais é um grande desafio que empresas têm enfrentado. Neste sentido, [Aldous et al. 2019] buscou entender como o engajamento de usuários ocorre em diferentes plataformas de mídias sociais e como isto afeta o conteúdo produzido, através da modelagem de tópicos em dados de publicações, comentários e curtidas extraídos de perfis de empresas no *Facebook, Instagram, Twitter, YouTube e Reddit*.

[Quattrone et al. 2018] propôs a junção de vários métodos para analisar as revisões em plataformas de economia compartilhada, com o objetivos de gerar conhecimento e valor para o setor por meio da classificação e análise semântica de milhares de revisões. E ao analisar mídias sociais em busca de padrões que determinam o comportamento do usuário, [Vydiswaran et al. 2018] examinou a viabilidade dos uso dos dados textuais presentes no *Twitter* para determinar padrões alimentícios de usuários. Através da coleta e classificação de dados textuais relacionados a alimentação, verificou-se que é possível utilizar as mídias sociais para determinar tendências, tal como determinar o nível de consumo de alimentos saudáveis e não saudáveis. Outro ponto que merece destaque quanto a busca de padrões é o apresentado por [Fang et al. 2016c], que estudaram os fatores que influenciam na percepção de valor em *reviews*. Através de análises de perspectiva dos textos e de seus autores, verificou-se que a legibilidade e sentimento do texto exerce influência significativa na importância que as pessoas atribuem a um *review*.

Há diferentes tipos de plataformas de mídias sociais com diferentes formas de interagir. [Xiang et al. 2017] propõem um estudo que examina comparativamente três plataformas de *reviews* de consumidores em termos de qualidade das informações relacionadas às avaliações *online* nesses sites. O objetivo deste trabalho foi fornecer uma base para a compreensão dos desafios metodológicos e para identificar oportunidades para o desenvolvimento da hotelaria e turismo. Foram extraídos 1.491 *reviews* em três das principais plataformas *online* no mundo da área, a saber: *TripAdvisor*, *Expedia* e a *Yelp*. Com as análises das características dos dados, foi verificado que, apesar de grande parte da literatura apresentar os *reviews* como fontes primárias de conhecimento para consumidores, há, no entanto, diferenças significativas nos *reviews* em diferentes plataformas.

Assim como [Aldous et al. 2019], este trabalho apresenta técnicas de mineração de texto provindos de mídias sociais, juntamente com classificações e análises semânticas presentes em [Quattrone et al. 2018]. E em consonância com o trabalho de [Fang et al. 2016a], também são verificados fatores que influenciam na percepção de valor dos *reviews*, além de medir a viabilidade de determinação de padrões e tendências a partir da coleta e classificação de dados textuais originados de mídias sociais [Vydiswaran et al. 2018]. E os resultados corroboram com a pesquisa de [Xiang et al. 2017] que observou a partir de características dos dados de *reviews* que existem diferenças notáveis entre as plataformas de clientes.

### **3. Materiais e Métodos**

Nesta seção são apresentados os processos de coleta e análise de dados, baseados nos trabalhos de [Xiang et al. 2017, Fang et al. 2016a]. Os trabalhos foram escolhidos devido sua relevância em publicações na área de mineração de dados textuais. Na primeira etapa, os dados de reclamações foram coletados. Posteriormente, foram pré-processados e extraídas as métricas textuais. Por fim, foram conduzidos os cálculos de coerência e de modelagem de tópicos.

#### **3.1. Coleta de Dados e Pré-Processamento**

Devido à sua importância econômica, foi escolhido para a mineração dos dados o setor de *ecommerce* brasileiro [ABCMM 2019]. Neste caso, selecionou-se as cinco empresas com maiores números de acessos em suas lojas virtuais [Netrica 2020]. Os dados coleta-

dos incluem nome da empresa, relato da reclamação, tempo de resposta, estado, cidade e data da reclamação.

Todos os dados de reclamações foram pré-processados utilizando os procedimentos de remoção de saudações, de URLs, *stopwords*, números, acentuação e caracteres especiais [Cirqueira et al. 2018]. A remoção de saudações e de URLs significa que toda *string* que representa uma saudação (e.g. "Olá", "Oi") ou um endereço de algum site (e.g. "www.site.com.br") foram removidos. Da mesma forma, ocorreu com números, acentuações e caracteres especiais, visto que são desnecessários para as análises. Palavras que são consideradas *stopwords*, que não contribuem para o significado do texto, também, foram removidas (e.g. "e", "de", "em"). Este processo é importante, uma vez que garante uma maior qualidade dos dados para executar as análises propostas.

### 3.2. Extração de Características Textuais

As características textuais são importantes para detectar padrões que possibilitem a identificação e comparação de autores de textos em diferentes locais [Hirsch et al. 2017]. Os dados textuais das reclamações foram processados e avaliados, levando em consideração a quantidade de caracteres de cada reclamação e do grau de legibilidade através do *Flesch Readability Ease Score (FRES)*. O FRES é uma fórmula desenvolvida por Rudolph Flesch em 1948 utilizada para calcular o grau de legibilidade de um texto [Flesch 1948]. O valor de FRES é obtido por meio da seguinte fórmula [Othman et al. 2012]:

$$FRES = 206.835 - (1.015 \times ASL) - (84.6 \times ASW) \quad (1)$$

A variável *ASL (Average Sentence Length, em português "tamanho médio da sentença")* é obtida a partir da divisão do número de palavras pelo número de sentenças. O *ASW (Average number of Syllables per Word)* é obtido a partir do número médio de sílabas pelo número de palavras. Os valores de FRES representam o grau de legibilidade de um texto e pode variar dentro do intervalo de zero a cem, sendo que quanto maior for o valor mais fácil é a leitura de um texto [Othman et al. 2012]. Esses valores são classificados de acordo com a Tabela 1, construída a partir dos trabalhos de [Porto et al. 2014, Othman et al. 2012, Hirsch et al. 2017].

Tabela 1. Classificação dos graus de legibilidade

Classificação	Grau
1º ao 5º ano	75 à 100
6º ao 9º ano	50 à 75
Ensino Médio e Nível Superior	25 à 50
Textos Acadêmicos	0 à 25

### 3.3. Coerência e Modelagem de Tópicos

A tarefa de determinar o melhor número de tópicos em modelagens de tópicos têm se mostrado muito subjetiva em diversos trabalhos na literatura. Visando evitar este viés, foi utilizado um método apresentado por [Damani 2013, Fang et al. 2016b, Aldous et al. 2019] que possibilita a avaliação da modelagem de tópicos por meio do cálculo da coerência de cada tópico. A etapa inicial na análise de coerência consiste em calcular para cada tópico

o *Pointwise Mutual Information (PMI)*, também conhecido como similaridade semântica (Equação 2). A etapa final na análise de coerência ocorre por meio do cálculo da média dos resultados do PMI, Equação 3, através do qual é obtido o valor da coerência do tópico analisado  $C(t)$ .

$$PMI(w_i w_j) = \log \frac{PMI(w_i w_j)}{p(w_i) \times p(w_j)} \quad (2)$$

$$C(t) = \frac{1}{\sum_{m=1}^{n-1}} \sum_{i=1}^n \sum_{j=i+1}^n PMI(w_i w_j) \quad (3)$$

A coerência média de um modelo é dada a partir da pontuação de coerência em todos os tópicos [Aldous et al. 2019]. De acordo com [Fang et al. 2016b], usuários finais normalmente estão interessados apenas nos tópicos mais coerentes de um modelo, em vez do modelo inteiro com todos os tópicos. Para isso, dado um modelo com um total de  $k$  tópicos, seleciona-se uma quantidade  $n$  (*coerência@n*) de tópicos de maior coerência e calcula-se a média destes valores. A fim de evitar o viés neste cálculo, o ideal é comparar as médias de diferentes valores de  $n$  e ampliar o espaço de busca pelos tópicos mais coerentes do modelo.

A identificação de tópicos nas reclamações foi realizada através do uso do algoritmo *Latent Dirichlet Allocation (LDA)* [Li et al. 2019]. A implementação deste algoritmo foi realizado por meio da biblioteca *scikit-learn*<sup>1</sup> da linguagem Python.

O LDA é um modelo bastante utilizado por estudos relacionados na literatura [Genc-Nayebi and Abran 2017], sendo que diversos trabalhos demonstraram a eficácia do modelo na extração de tópicos em diferentes conjuntos de dados textuais [Charles-Smith et al. 2015, Li et al. 2019, Fang et al. 2016a]. Na execução, o algoritmo assume que existem estruturas de tópicos que não estão visíveis no conjunto de dados e usa a coocorrência de palavras observadas em diferentes instâncias [Gómez et al. 2015]. Dado um conjunto de dados textuais, a execução deste algoritmo retorna uma lista de tópicos associados a um conjunto de palavras. Este conjunto de palavras são associados ao tópico por meio de pesos e uma lista de registros (no caso deste trabalho, reclamações) com um vetor de valores de peso, que definem a probabilidade de haver um documento que contém um tópico específico [Xiang et al. 2017].

#### 4. Resultados

Foram coletados dados de empresas que atuam no setor de *ecommerce* e que tinham reclamações de consumidores nas duas plataformas. No total foram coletadas, no período de 26/01/2019 a 26/01/2020, 289.566 reclamações nas duas plataformas. Na Figura 2 são apresentadas as taxas de reclamações proporcionais ao tamanho da população de cada estado. Sendo os estados da região sudeste tem maiores taxas de reclamações por habitante.

Na Tabela 2 são expostos a quantidade de reclamações coletadas, juntamente com a taxa de resposta e de resoluções fornecidas pelas plataformas no período de 6 meses. Ao observar a tabela, é possível constatar que o CGOV ter menor quantidade de reclamações quando comparado com o RA, porém, proporcionalmente no CGOV há uma alta taxa de respostas às reclamações por parte das empresas analisadas.

<sup>1</sup><https://scikit-learn.org/stable/>

Tabela 2. Número de reclamações por empresa em cada plataforma.

	Consumidor.gov			ReclameAqui		
	Reclamações	Respondidas	Resolvidas	Reclamações	Respondidas	Resolvidas
Americanas.com	16.456	99,9%	67,8%	33.297	95,9%	85,8%
Magazine Luíza	5.694	98,7%	79,8%	50.493	94,2%	90,2%
Mercado Livre	26.346	100%	77,9%	71.018	0%	0%
Netshoes	4.377	99,8%	84,5%	47.237	0%	0%
Submarino	5.169	99,9%	75,5%	29.479	93,8%	84,2%

As plataformas analisadas são similares em termos de propósitos e conteúdo dos textos. Entretanto, há particularidades em cada plataforma que as diferem e influenciam as reclamações e taxa de resposta. O RA é aberto para reclamação de todas as empresas disponíveis no mercado, qualquer cliente pode acessar a plataforma e enviar seu relato. Já o CGOV só aceita reclamações de empresas que aderiram voluntariamente e se cadastraram na plataforma. Isto tem impacto na quantidade de empresas presentes nas plataformas e é o principal motivo de haver reclamações sem resposta e resolução no RA, Tabela 2.

As métricas textuais são pontos que destacam as diferenças entre as plataformas, tal como a *readability* e tamanho das reclamações. Na Figura 1 são apresentadas as distribuições do tamanho das reclamações do CGOV e RA. A partir dos dados expostos, foi possível verificar que os relatos de problemas no CGOV (em azul) tem maior variação no tamanho dos relatos no RA (em vermelho), indicando a presença de duas bases de usuários com características distintas nas duas plataformas.

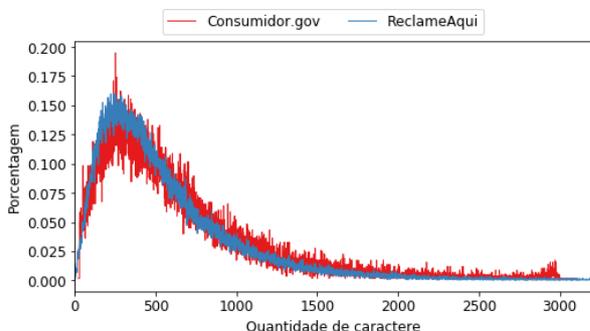


Figura 1. Comparação do tamanho das reclamações nas duas plataformas.

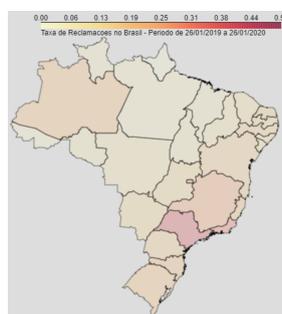


Figura 2. Taxa de distribuição das reclamações no Brasil.

Esta diferença fica mais evidente na Figura 3, que representa todas as reclamações coletadas das empresas nas duas plataformas divididas em inferidos níveis de escolaridade. A segregação das reclamações em níveis de escolaridade nos revelam um pouco mais sobre perfis dos consumidores de cada plataforma de reclamação online. A seguir são descritas as características encontradas. No caso do CGOV identificou-se como principal grupo as pessoas cuja escolaridades é de 6º à 9º ano do ensino fundamental, seguido por "Ensino Médio e Nível Superior", "Textos Acadêmicos" e, por fim, pessoas com 1º à 5º ano do ensino fundamental. No caso do RA encontrou-se um público de maior nível de escolaridade, sendo o principal grupo de pessoas identificados como que cursaram o "Ensino Médio e Nível Superior", seguido por textos de nível "Textos Acadêmicos", 6º à 9º ano e 1º à 5º ano do ensino fundamental. Pode-se destacar que nas duas plataformas,

o menor grupo de pessoas que realizam reclamações são os que tiveram poucos anos de ensino formal, de 1º à 5º ano.

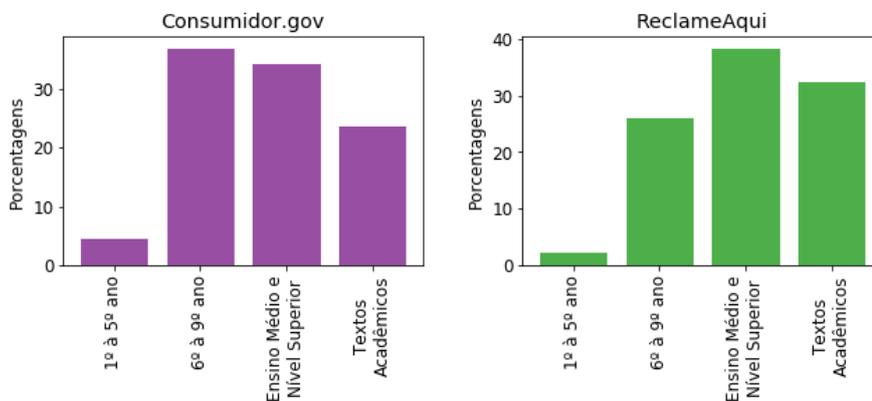


Figura 3. Comparação no nível de escolaridade entre as plataformas.

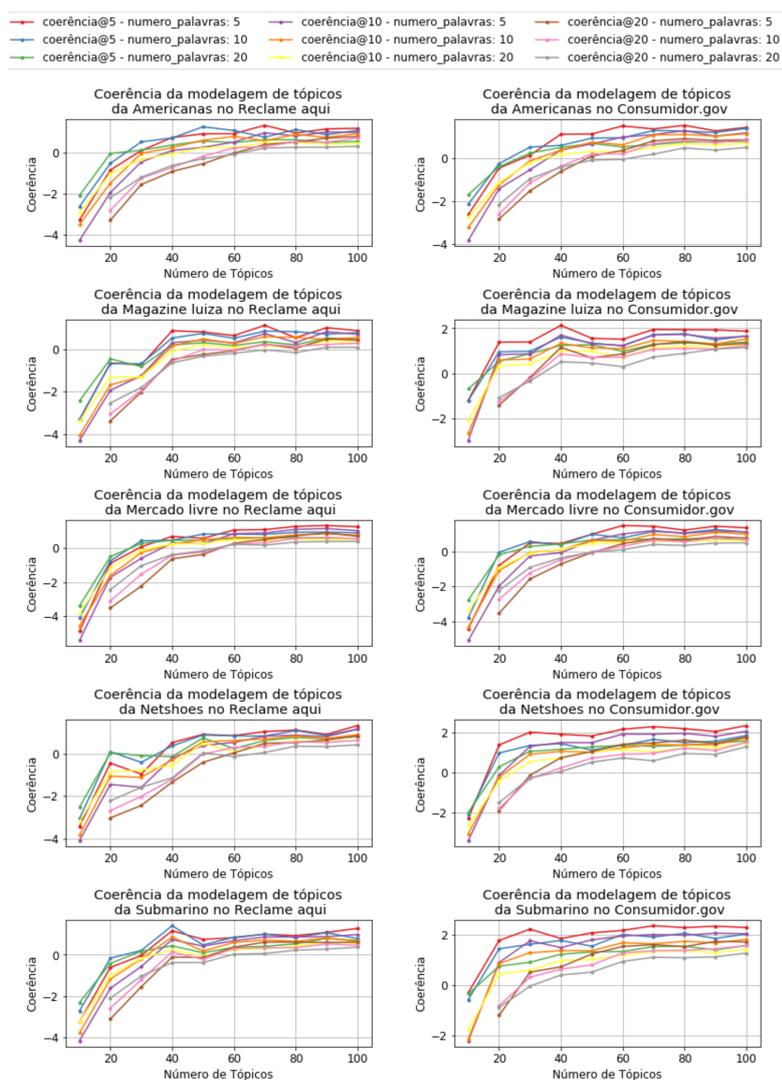


Figura 4. Coerência da modelagem de tópicos.

Os cálculos de coerência das empresas em diferentes cenários são necessários para se ter uma modelagem de tópicos mais acurada e com maior nível de confiabilidade para determinar os principais tipos de problemas. Na Figura 4 são apresentados os resultados dos cálculos de coerência para cada empresa a partir de diferentes cenários que mesclam: i) plataforma; ii) quantidade de tópicos; iii) número de palavras por tópicos. Os maiores valores de coerência são considerados melhores e mais aptos a serem utilizados. Cada linha dos gráficos indica os resultados para um cenário diferente (e.g. "coerência@5: numero\_palavra:5" indica que o cálculo foi realizado para os 5 principais tópicos que continham 5 palavras). A fim de determinar os melhores parâmetros a serem utilizados, os tópicos foram modelados considerando os seguintes valores de número de tópicos de [10, 20, 30, 40, 50, 60, 70, 80, 90, 100] variando a quantidades de palavras por tópico em [5, 10, 20].

Ao final do processo, foi observado que houve variação nas quantidades ideais de tópicos modelados, tópicos principais e números de palavras para cada empresa e plataforma analisada. Na Tabela 3 são apresentados os valores obtidos para cada empresa, sendo: "Nº tópicos" - representa a quantidade de tópicos modelados pelo algoritmo; "Nº tópicos principais" - a quantidade de tópicos com maior coerência média que deverão ser utilizados nos resultados; "Nº palavras" - representa a quantidade de palavras para um tópico ter a melhor coerência.

Conforme pode ser observado, todas as empresas nas duas plataformas obtiveram 5 como número de tópicos e de palavras ideais para serem utilizados na modelagem, com exceção do "Submarino" que tem 10 como número de palavras ideais.

Tabela 3. Melhores valores para modelagem de tópicos nas plataformas analisadas.

	Reclame Aqui			Consumidor.gov		
	Nº tópicos Modelados	Nº tópicos Principais	Nº palavras	Nº tópicos Modelados	Nº tópicos Principais	Nº palavras
Americanas.com	70	5	5	60	5	5
Magazine Luíza	70	5	5	40	5	5
Mercado Livre	90	5	5	60	5	5
Netshoes	100	5	5	100	5	5
Submarino	40	5	10	70	5	5

A Tabela 4 apresenta os principais tópicos das reclamações nas duas plataformas, que representam os cinco principais tipos de problemas com maior valor e importância dentre os demais, considerando cada plataforma.

As células T1 e T5 do RA foram mescladas pois deram origem ao mesmo tópico "Entrega". É importante ressaltar que cada tópico representa um elemento importante para a análise de reclamações, pois fornece bases para o entendimento do contexto da empresa. Os dados estão organizados na tabela de forma a representar a importância através da ordem sequencial dos valores de coerência, sendo que na linha T1 estão os tópicos mais representativos e na linha T5 estão os menos representativos dentre os principais. Os principais tipos de reclamações variam de acordo com a plataforma, e a seguir são exemplificadas essas diferenças obtidas a partir dos tópicos mais coerentes:

- A empresa Americanas.com tem o "CashBack" como problema principal no CGOV, o que indica que os usuários da plataforma têm mais problemas com este

Tabela 4. Tópicos modelados a partir das reclamações nos sites Consumidor.gov e ReclameAqui.

Consumidor gov					
	Americanas.com	Magazine Luiza	Mercado Livre	Netshoes	Submarino
T1	Cashback	Assinatura	Resolução de conflitos	Produtos	Produtos
T2	Atendimento	Expectativas	Publicidade	Entrega	Promoção
T3	Cadastro	Transporte	Custos	Atendimento	Devolução
T4	Solicitações fundamentadas	Atendimentos	Produtos	Chuteira Infantil	Entrega
T5	Fraude	Defeitos	Serviço	Desistência	Compra

Reclame Aqui					
	Americanas.com	Magazine Luiza	Mercado Livre	Netshoes	Submarino
T1	Busca	Entrega	Resolução de conflitos	Entrega	Entrega
T2	Produto	Cobrança	Venda	Devolução	Reembolso
T3	Entrega	Produto	Cancelamento	Acesso	Produto
T4	Atendimento	Retorno produto	Promoção	Cobrança	Desistência
T5	Reembolso	Atendimento	Cobrança	Enganação	

produto da empresa. Porém, no RA a principal reclamação é sobre a "Busca", que pode estar relacionado com os mais variados serviços da empresa, o que significa que é necessário investigar quais e como os serviços lidam com a busca dos consumidores.

- A empresa Magazine Luiza tem o tópico "Assinatura" como problema principal no CGOV, onde o grupo de clientes que são usuários do CGOV tendem a ter mais problemas com os produtos com assinatura. Para o grupo de usuários do RA a principal reclamação é sobre a "Entrega", e devido as características da empresa essas reclamações estão extremamente relacionadas a entrega de produtos.
- A empresa Mercado Live tem o tópico "Resolução de Conflitos" como problema principal no CGOV e no RA, o que indica que independente do grupo de cliente que utilizam as plataformas a principal reclamação é a mesma, indicando que o problema é recorrente e afeta muitos clientes.
- A empresa Netshoes tem o "Produto" como problema principal no CGOV e "Entrega" como principal reclamação no RA. Estes tipos de reclamação são diferentes e devem ser tratadas individualmente, mas devido ao modelo de negócio da empresa estão muito relacionadas entre si.
- A empresa Submarino tem o "Produto" como problema principal no CGOV e "Entrega" como principal tipo de reclamação no RA. Porém, devido ao modelo de negócio da empresa, esses tipos de reclamações não necessariamente estão relacionados entre si.

Os achados aqui apresentados se mostram valiosos e com grande possibilidade de orientar a atuação de empresa nas duas plataformas. Através dos mesmos, é possível obter uma ordem sequencial de problemas recorrentes que desgastam a relação entre empresa e consumidor. Além disto, devido a presença de grupos distintos de usuários nas plataformas, é possível estruturar estratégias de soluções que melhor se adéquem a cada grupo.

## 5. Considerações Finais

Neste artigo, foram analisadas reclamações de cinco empresas do setor de *ecommerce* brasileiro em duas plataformas de reclamações *online*. O objetivo destas análises é prover uma base para a compreensão das oportunidades que elas oferecem para orientar a tomada de decisões em negócios. Foram aplicadas técnicas de mineração de dados textuais a fim de explorar as plataformas e extrair suas características. Os resultados mostram que apesar de as plataformas terem a mesma finalidade, elas divergem em grupos de usuários, confiabilidade de solução, conteúdo e principais tópicos das reclamações. Ademais, as análises se mostraram viáveis mesmo com grandes quantidades de dados, o que torna possível a antecipação e correção de problemas de forma rápida e eficaz, e abre a possibilidade de fazer amplas análises de mercado e concorrência de acordo com a perspectiva de cada plataforma.

Diante disto, a principal contribuição deste estudo está na provisão de uma abordagem de análise de reclamações para as duas plataformas que permite identificar diferentes formas de atuação que uma empresa deve seguir para melhorar o relacionamento com seus clientes. No entanto, não foram realizadas análises de significância estatística nas comparações apresentadas neste estudo. E as análises e resultados obtidos não incluem aplicações práticas nas empresas. Como trabalhos futuros, pretende-se expandir as análises de modo a incluir análise de significância estatística nas comparações, conduzir um estudo prático em empresas que explore tanto as reclamações quanto a respostas.

## 6. Agradecimentos

Os autores agradecem ao Serviço de Intercâmbio Alemão (DAAD), à Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) e a Universidade Federal do Oeste do Pará (UFOPA) por meio da Pró-Reitoria da Cultura, Comunidade e Extensão (PROCCE), pelos fomentos destinados à execução desta pesquisa.

## Referências

- ABCOMM (2019). Crescimento do e-commerce no Brasil. Disponível em: <https://abcomm.org/noticias/crescimento-do-e-commerce-no-brasil>. Acesso em: 28/03/2020.
- Aldous, K. K., An, J., and Jansen, B. J. (2019). View , Like , Comment , Post : Analyzing User Engagement by Topic at 4 Levels across 5 Social Media Platforms for 53 News Organizations. In *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media*, pages 47–57.
- Bahtar, A. Z. and Muda, M. (2016). The Impact of User – Generated Content (UGC) on Product Reviews towards Online Purchasing – A Conceptual Framework. *Procedia Economics and Finance*, 37:337–342.
- Charles-Smith, L. E., Reynolds, T. L., Cameron, M. A., Conway, M., Lau, E. H., Olsen, J. M., Pavlin, J. A., Shigematsu, M., Streichert, L. C., Suda, K. J., and Corley, C. D. (2015). Using social media for actionable disease surveillance and outbreak management: A systematic literature review. *PLoS ONE*, 10(10):1–20.

- Cirqueira, D., Fontes Pinheiro, M., Jacob, A., Lobato, F., and Santana, A. (2018). A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 746–749. IEEE.
- Constantinides, E. and Holleschovsky, N. I. (2016). Impact of Online Product Reviews on Purchasing Decisions. *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, pages 271–278.
- Damani, O. P. (2013). Improving Pointwise Mutual Information (PMI) by incorporating significant co-occurrence. In *CoNLL 2013 - 17th Conference on Computational Natural Language Learning, Proceedings*, pages 20–28.
- de Almeida, J. S. D. N. G. R. T., Lobato, F. M. F., and Junior, A. F. L. J. (2019). Melhorando Sistemas de Social CRM por meio de Eletronic Word-of-Mouth. *Revista Eletrônica de Iniciação Científica em Computação*, 17(4).
- Fang, A., Macdonald, C., Ounis, I., and Habel, P. (2016a). Examining the coherence of the top ranked tweet topics. *SIGIR 2016 - Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 825–828.
- Fang, A., Macdonald, C., Ounis, I., and Habel, P. (2016b). Examining the coherence of the top ranked tweet topics. *SIGIR 2016 - Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 825–828.
- Fang, B., Ye, Q., Kucukusta, D., and Law, R. (2016c). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52:498–506.
- Flesch, R. F. (1948). A new readability yardstick. *The Journal of applied psychology*, 32 3:221–33.
- Genc-Nayebi, N. and Abran, A. (2017). A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software*, 125:207–219.
- Gómez, M., Rouvoy, R., Monperrus, M., and Seinturier, L. (2015). A Recommender System of Buggy App Checkers for App Store Moderators. *Proceedings - 2nd ACM International Conference on Mobile Software Engineering and Systems, MOBILESoft 2015*, pages 1–11.
- Hirsch, M., Aggarwal, S., Barker, C., Davis, C. J., and Duffy, J. M. (2017). Googling endometriosis: a systematic review of information available on the Internet. *American Journal of Obstetrics and Gynecology*, 216(5):451–458.e1.
- Kim, A. J. and Johnson, K. K. (2016). Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook. *Computers in Human Behavior*, 58:98–108.
- Li, D., Zamani, S., Zhang, J., and Li, P. (2019). Integration of Knowledge Graph Embedding Into Topic Modeling with Hierarchical. In *Proceedings of the 2019 Conference of*

- the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 940–950.
- Lobato, F., Pinheiro, M., and Jr, A. J. (2017). Social CRM: Biggest Challenges to Make it Work in the Real World. *Business Information Systems Workshops. BIS 2016. Lecture Notes in Business Information Processing*, 263:221–232.
- Netrica (2020). TOP ECOMMERCE RANKING REPORTS. Disponível em: <https://ecommerce-brasil.rankings.netquest.digital>. Acesso em: 27/01/2020 .
- Orenga-Roglá, S. and Chalmeta, R. (2016). Social customer relationship management: taking advantage of Web 2.0 and Big Data technologies. *SpringerPlus*, 5(1).
- Othman, I. W., Hasan, H., Tapsir, R., Rahman, N. A., Tarmuji, I., Majdi, S., Masuri, S. A., and Omar, N. (2012). Text readability and fraud detection. In *ISBEIA 2012 - IEEE Symposium on Business, Engineering and Industrial Applications*, pages 296–301. IEEE.
- Porto, J. S., Paiva, T. S. d. S., Amaral, C. L. F., Rebouças, T. N. H., and Silva, R. d. A. (2014). Legibilidade de artigos de um periódico nacional na área de melhoramento vegetal. *Cultivando o Saber*, 7(2):205–211.
- Quattrone, G., Nicolazzo, S., Nocera, A., Quercia, D., and Capra, L. (2018). Is the sharing economy about sharing at all? A linguistic analysis of Airbnb reviews. In *12th International AAI Conference on Web and Social Media, ICWSM 2018*, pages 668–671.
- Roy, G., Datta, B., and Mukherjee, S. (2019). Role of electronic word-of-mouth content and valence in influencing online purchase behavior. *Journal of Marketing Communications*, 25(6):661–684.
- Schmäh, M., Wilke, T., and Rossmann, A. (2017). Electronic Word-of-Mouth: A Systematic Literature Analysis. *Lecture Notes in Informatics (LNI)*, page 147.
- Silva, W., Santana, Á., Lobato, F., and Pinheiro, M. (2017). A Methodology for Community Detection in Twitter. In *Proceedings of the International Conference on Web Intelligence*, pages 1006–1009.
- Vermeer, S. A., Araujo, T., Bernitter, S. F., and van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, 36(3):1–17.
- Vydiswaran, V. G., Romero, D. M., Zhao, X., Yu, D., Gomez-Lopez, I., Lu, J. X., Iott, B., Baylin, A., Clarke, P., Berrocal, V., Goodspeed, R., and Veinot, T. (2018). “Bacon bacon bacon”: Food-related tweets and sentiment in metro detroit. In *12th International AAI Conference on Web and Social Media, ICWSM 2018*, pages 692–695.
- Xiang, Z., Du, Q., Ma, Y., and Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58:51–65.