

# A culpa é dela! É isso o que dizem nos comentários das notícias sobre a tentativa de feminicídio de Elaine Caparroz

Lucas D. F. Rodrigues<sup>1</sup>, Jorge L. F. da Silva Junior<sup>1</sup>, Fábio M. F. Lobato<sup>1</sup>

<sup>1</sup>Instituto de Engenharia e Geociências - Universidade Federal do Oeste do Pará  
Santarém, Pará, Brasil

{lucas.darlindo, jorgeluzifigueira}@gmail.com

fabio.lobato@ufopa.edu.br

**Abstract.** *Social media plays an important role in the feminist agenda by providing a channel for denouncing and also inducing support networks. However, hate speech is conveyed in this type of media. Aiming to study this counterpoint, this paper analyzes news commentary on the femicide attempt of the landscape artist Elaine Perez Caparroz. The approach was carried out with the collection of news and their comments, which were later analyzed using topic modeling, sentiment analysis, correlation analysis of likes and dislikes and the comments scholarly level. The results show that most of the comments share the opinion that it was the fault of Elaine Caparroz, highlighting the chauvinism that still existing in our society.*

**Resumo.** *As mídias sociais têm um importante papel na agenda de lutas feministas por proporcionarem um canal de denúncia e induzirem à redes de apoio. No entanto, discursos de ódio também são veiculados neste tipo de mídia. Visando estudar este contraponto, o presente trabalho analisa comentários de notícias sobre a tentativa de feminicídio da paisagista Elaine Perez Caparroz. A abordagem realizada foi conduzida com a coleta de dados de notícias e seus comentários que, posteriormente, foram analisados a partir de métodos para modelagem de tópicos, análise de sentimentos, correlação com likes & dislikes e também análise do nível de escolaridade dos comentários. Os resultados mostram que a maior parte dos comentários comunga da opinião que a culpa foi de Elaine Caparroz, ressaltando o machismo que ainda há em nossa sociedade.*

## 1. Introdução

As mídias sociais têm um importante papel em movimentos sociais proporcionarem um canal de denúncia e induzirem à redes de apoio [Caldas et al. 2018]. Elas também contribuem para a propagação de informações relevantes para a população em geral [Figueira and Guimarães 2017]. Essa difusão de informação é proporcionada pela Web 2.0 e provocou alterações na forma de consumo dessas fontes pelos usuários, além da substituição de meios tradicionais de notícias, como jornais e revistas por suas versões digitais [Ramos et al. 2016, Lobato et al. 2017].

No entanto, os discursos de ódio também são veiculados neste tipo de mídia por permitirem a interação dos usuários por meio do sistema de comentários e respostas

[Djuric et al. 2015]. Sendo que o anonimato é um importante aspecto que incita esses discursos dentro das plataformas [Mondal et al. 2017]. São duas as principais esferas de direcionamento destes: i) direcionado a grupos; ii) individualmente [ElSherief et al. 2018]. Nesse contexto, os casos de feminicídio no Brasil que são veiculados nesses espaços são passíveis de receberem variadas formas de expressão dos usuários.

No país, a primeira lei com enfoque em punir os casos de violência doméstica e familiar contra a mulher foi a Lei nº 11.340/2006, também conhecida como Lei Maria da Penha. O feminicídio é definido como o caso mais extremo de violência contra mulheres e pode ser efetivada das mais diversas formas, além disso os fatores sociais também exercem uma forte influência, como a cultura patriarcal [Boira et al. 2017]. A violência compreende um amplo espectro, desde a agressão verbal e outras formas de abuso emocional ou psicológica até o nível físico ou sexual, com seu extremo na morte intencional de uma mulher [Garcia et al. 2013]. Em 2015, incluiu-se o feminicídio no rol de crimes hediondos (Lei nº 13.104/2015), crimes estes definidos pela Lei nº 8.072/1990 como insuscetíveis de graça, indulto ou fiança. Essa inclusão gera penas mais severas e possui agravantes pois é classificada como uma das modalidades de homicídio qualificado.

Porém, mesmo com ambas as leis, os dados presentes no Atlas da Violência de 2018 indicam que no ano de 2016, 4.645 mulheres foram assassinadas no país [Cerqueira et al. 2018]. Isso compreende em uma taxa de 4,5 homicídios para cada 100 mil brasileiras em dez anos (2006-2016). Além disso, quando relacionado ao número de assassinatos de mulheres negras, a taxa sobe para 5,3 homicídios para cada 100 mil brasileiras e resulta em um aumento de 15,4%. Diante deste cenário, o Brasil ocupa a quinta posição dentre os países com maior taxa de feminicídio no mundo [Meneghel and Portella 2017]. A partir disso, foram levantadas as seguintes perguntas de pesquisa:

1. Qual o posicionamento das pessoas diante de notícias sobre o tema?
2. Qual a correlação entre o nível de instrução e o posicionamento?

Nesse panorama, o presente estudo analisa postagens presentes em mídias sociais, especificamente, sobre notícias do caso da paisagista Elaine Perez Caparroz, de 55 anos, vítima de uma tentativa de feminicídio, fato ocorrido no dia 17 de Fevereiro de 2019, no Rio de Janeiro. A principal motivação surgiu devido a proximidade do fato com o Dia Internacional da Mulher e para analisar como seria a forma de expressão dos internautas brasileiros quanto ao ocorrido. Os resultados indicam que a culpa é atribuída a ela por via de comentários negativos e pejorativos, em sua maioria aceitos pelos demais usuários.

O restante deste artigo está organizado da seguinte forma: A Metodologia adotada nos experimentos é descrita na Seção 2. Os resultados obtidos são discutidos na Seção 3. Por fim, as considerações finais, ameaças a validade do estudo e projeções para trabalhos futuros são apresentadas na Seção 4.

## **2. Metodologia**

Nesta seção a metodologia utilizada na condução do estudo é apresentada, cobrindo desde a coleta de dados até a interpretação dos resultados, perpassando pelo pré-processamento e métodos de análise, segundo um fluxo semelhante ao descrito em [Silva et al. 2017].

## 2.1. Coleta de Dados

Para determinar a principal fonte para coleta, verificou-se o *ranking* da *Alexa*<sup>1</sup>, o qual opera de forma a ranquear as páginas mais acessadas por via dos dados de tráfego fornecidos pelo painel de dados globais dos usuários em um período de três meses, atualizado diariamente [Yesbeck 2018]. Diante disto, o Portal G1 foi delimitado com verificação feita também nos portais das demais regiões do país (*e.g.* Norte, Sudeste). Foram coletadas as informações básicas referentes as notícias, além das relacionadas aos comentários nelas presentes. O intervalo para exibição das notícias foi delimitado na própria página, ajustes estes disponibilizados pelo próprio portal para filtrar o conteúdo exibido (*i.e.* período, tipo de conteúdo). Os seguintes termos de busca foram utilizados:

- Elaine Perez
- Elaine Caparroz
- Elaine Perez Caparroz
- Tentativa de feminicídio no Rio de Janeiro

Para tal, criou-se um *web crawler* desenvolvido sobre a linguagem *Python* a partir do *Scrapy Framework*<sup>2</sup> para extrair em arquivos de formato *Comma-Separated Values* (CSV), tal como em [Almeida et al. 2017, Teixeira et al. 2018]. Para a interação com essas páginas, utilizou-se o *Selenium*<sup>3</sup> em combinação para controle autônomo do navegador (*Geckodriver*<sup>4</sup>) para o correto carregamento de todos os itens presentes, pela utilização de requisições *AJAX* ou *JavaScript* durante os acessos. Os dados extraídos estão presentes na Tabela 1.

**Tabela 1. Conteúdo extraído das notícias.**

| Notícias           | Comentários                 |
|--------------------|-----------------------------|
| Título             | Autor                       |
| Resumo             | Data de Publicação          |
| Data de Publicação | Conteúdo                    |
| <i>Link</i>        | <i>Likes &amp; Dislikes</i> |

## 2.2. Pré-processamento

A fim de melhorar o desempenho sobre os dados coletados, a etapa de pré-processamento servirá para a remoção de certos itens que podem influenciar no resultado final das análises [Cirqueira et al. 2018]. Para tal, utilizou-se a biblioteca de processamento em linguagem natural *NLTK*<sup>5</sup> disponível para *Python* e configurada para o Português. Ademais, os seguintes métodos de pré-processamento foram executados:

- Remoção de *stopwords* e normalização;
- Remoção de pontuações ou caracteres especiais (*e.g.* !, ?, +);
- Remoção de saudações ou cumprimentos (*e.g.* “bom dia”, “oi”);
- Remoção de comentários duplicados;
- Remoção de *emojis* (*i.e.* uma imagem que transmite a ideia de uma palavra);
- Remoção de acentuação.

<sup>1</sup><https://www.alexa.com/topsites/countries/BR>

<sup>2</sup><https://scrapy.org/>

<sup>3</sup><https://www.seleniumhq.org/>

<sup>4</sup><https://firefox-source-docs.mozilla.org/testing/geckodriver/>

<sup>5</sup><https://www.nltk.org/>

### 2.3. Análise de Sentimento

A análise de sentimento é o campo de estudo que visa classificar dados textuais de acordo com a polaridade presente no texto, como positivo, negativo e algumas vezes, um sentimento neutro [Cirqueira et al. 2017]. No presente trabalho a categorização dos sentimentos presentes nos comentários baseou-se em três intervalos numéricos. Para a realização dessa etapa, é utilizada a biblioteca *Polyglot*<sup>6</sup> por suportar essas análises no idioma Português de forma nativa e por disponibilizar ao pesquisador diversas ferramentas baseada em sua utilização de *pipeline* de linguagem natural. Os dados foram estratificados a fim de uma melhor análise para cada tipo de sentimento, onde intervalos foram definidos para a definição do tipo assumido pelo algoritmo e apresentados na Tabela 2.

Tabela 2. Intervalos de pontuações para os sentimentos.

| Pontuação (p)   | Sentimento |
|-----------------|------------|
| $0 < p \leq 1$  | Positivo   |
| $p = 0$         | Neutro     |
| $-1 \leq p < 0$ | Negativo   |

### 2.4. Métrica de Aceitação

Os comentários disponíveis na plataforma utilizada como estudo de caso são passíveis de interação com usuários por meio de *likes* e *dislikes*. Para fins de análise, comparou-se o sentimento dos comentários com a prevalência das interações - se elas tinham mais *likes* ou mais *dislikes*. É importante citar que essa métrica tem relevância por permitir

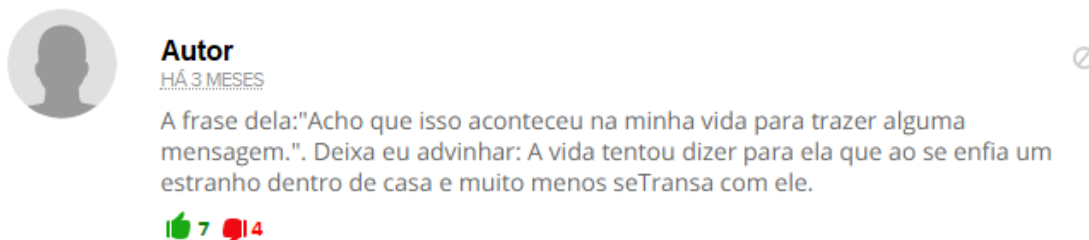


Figura 1. Comentário e suas possíveis interações.

verificar o nível de aceitação de um determinado comentário ou um grupo deles conforme a interação dos demais usuários no mesmo ambiente. Além disto, esse grau também permite a análise do quão verossímil pode ser a aceitação de discursos, de ódio ou não, direcionados ou individuais.

### 2.5. Modelagem de Tópicos

Para a modelagem de tópicos foram utilizados o *Latent Semantic Analysis* (LSA) e o *Latent Dirichlet Allocation* (LDA), tal como em [Lobato et al. 2018] e [Melville et al. 2019]. Os dois métodos foram escolhidos por serem bem aceitos pela comunidade e estarem presentes em diversos trabalhos que visam analisar dados advindos

<sup>6</sup><https://polyglot.readthedocs.io/en/latest/>

de redes sociais. Para se construir a matriz de pesos utilizou-se o *Term Frequency-Inverse Document Frequency*, ou TF-IDF. Tal método permite verificar estatisticamente o quão importante é uma palavra dentro de documentos ou uma coleção deles. Deste modo, o TF mede o quão frequente um termo aparece dentro de um documento que, para gerar um resultado normalizado, este valor é dividido pelo tamanho do documento (*i.e* o número total de termos). Enquanto que o IDF busca indicar a importância de palavra de forma a ajustar esse item do TF, haja vista que ele considera os vocábulos como igualmente importantes. Para isso é feito o cálculo sobre o logaritmo do total de documentos pelo número dos que contém o termo neles.

Cada tópico é definido a partir de uma *Bag of Words*, um modelo de representação de informações textuais por via computacional e que define esses dados como uma matriz, o qual assimila pesos para cada termo ou palavra encontrada em todo o conjunto. Contudo, os algoritmos de modelagem não definem categorias ou rotulam os tópicos. Por conseguinte, a presente tarefa fica a cargo de uma análise subjetiva a ser efetuada pelo autor. Para os tópicos mais utilizados dentro dos comentários, foram adotados diferentes cenários para as suas composições. Em suma, buscou-se executar para mais de uma configuração a fim de encontrar a melhor para os dados utilizados. As iterações foram executadas para sete diferentes parâmetros, os quais modificaram-se: a quantidade de tópicos e de palavras. Também foi considerada a possibilidade de rotulação desses grupos.

Foram verificadas as seguintes opções para construção de tópicos e palavras (T-P), respectivamente: 3-5; 3-10; 5-3; 5-5; 5-10; 10-5; 10-10. Esta parametrização baseou-se em trabalhos que realizaram modelagem de tópicos para dados de mídias sociais como [Lobato et al. 2018] e [Melville et al. 2019] e também na experiência dos pesquisadores. Ao alternar entre os parâmetros disponíveis, ou a dispersão entre esses ficava muito grande, o que ocasionava uma grande repetição de tópicos ou palavras; ou então a quantidade e composições eram insuficientes para permitir a rotulação de todos. Sendo assim, os autores avaliaram as três dimensões de parametrização para modelagem de tópicos, a saber: i) método aplicado (LSA ou LDA); ii) quantidade de tópicos (T); iii) quantidade de palavras (P). O resultado da parametrização é apresentado na Seção 3.

## 2.6. Legibilidade

A legibilidade avalia a dificuldade de leitura de um texto [Schriver 1989]. Há diversos métodos que exploram como avaliar essa dificuldade representando-as como uma pontuação por meio de equações matemáticas que representam a relação entre características linguísticas simples. Essas equações são conhecidas como fórmulas de legibilidade e entre as comumente aplicadas são as fórmulas de Flesch-Kincaid.

O índice de Flesch possui variações para diferentes finalidades, a saber, *Reading Ease* e *Grade Level*, sendo que a primeira possui uma adaptação em Português Brasileiro validada [Martins et al. 1996] e emprega o conceito de séries escolares do *Grade Level*. O índice de Flesch para o português brasileiro é denotado pela seguinte equação:

$$Flesch_{PT} = 248.835 - (1.015 \times ASL) - (84.6 \times ASW). \quad (1)$$

Sendo ASL a média de palavras por sentença e ASW a média de sílabas por palavra. Conforme [Martins et al. 1996] os índices obtidos pela equação podem ser classificados em quatro níveis de legibilidade representados na Tabela 4.

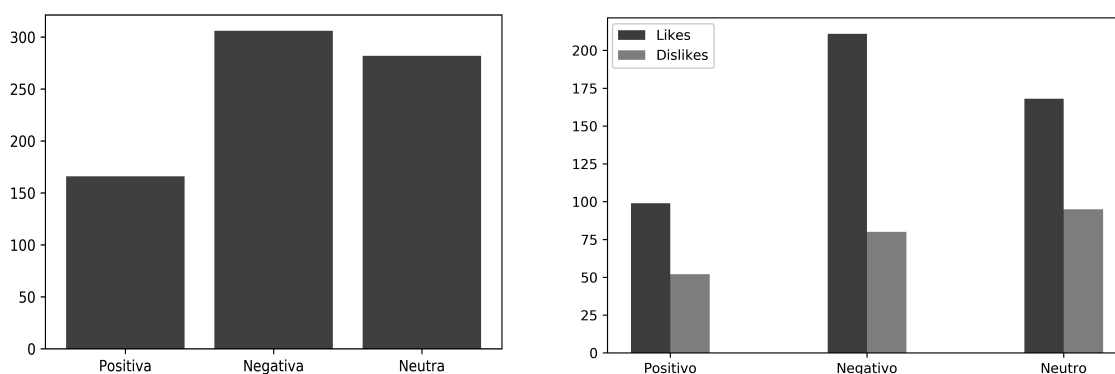
**Tabela 3. Níveis de legibilidade por escolaridade.**

| Pontuação    | Legibilidade  | Escolaridade                          |
|--------------|---------------|---------------------------------------|
| 75 à 100     | Muito Fácil   | 1 <sup>a</sup> à 4 <sup>a</sup> série |
| 50 à 75      | Fácil         | 5 <sup>a</sup> à 8 <sup>a</sup> série |
| 25 à 50      | Pouco difícil | Ensino Médio e Nível Superior         |
| Abaixo de 25 | Muito difícil | Textos Acadêmicos                     |

Na metodologia de avaliação da métrica proposta pelos autores, foi utilizado amostras textuais retiradas de livros escolares e acadêmicos obtendo um resultado satisfatório. Como os dados avaliados neste artigo são comentários, alguns deles possuem características pobres que ao mensurar sua dificuldade acarretariam em um resultado duvidoso. Para contornar esse problema foram filtrados e reduzidos os números de comentários apoiados na média geral de sentenças e palavras por sentença, de maneira que o novo conjunto de dados fosse constituído por no mínimo duas sentenças e 13 palavras, assim enriquecido quanto à estrutura de textos encontrados em livros.

### 3. Resultados Obtidos

O período de coleta das notícias e comentários corresponde a uma janela cronológica de 17 de Fevereiro à 11 de Março. Este intervalo foi escolhido considerando o dia da ocorrência do fato (17/02) e o tempo até as notícias saírem da página inicial do portal, com isso, o número de interações (comentários e *likes & dislikes*) cai significativamente. Nesse tempo, foram coletados 1.120 comentários a partir de 31 notícias publicadas. A partir da execução de todas as etapas para coleta e estruturação dos dados para a formulação de resultados, o novo conjunto gerado contabilizou um total de 754 comentários.

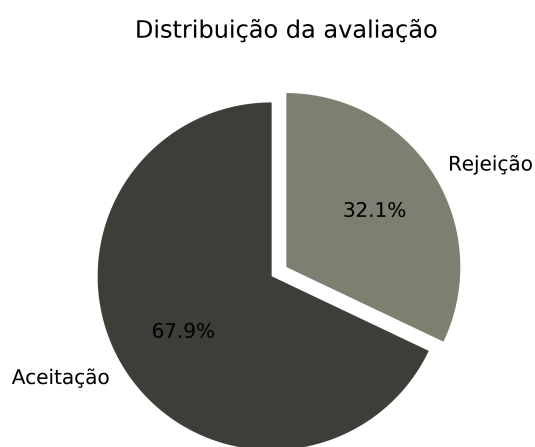


**Figura 2. a) Sentimentos dos comentários; b) Interações de avaliação por classe.**

Na análise de sentimento via *Polyglot* (Figura 2.a), é possível observar que há uma predominância quanto a quantidade de comentários de classificação negativa (306) e neutra (282) sobre os de positiva (166), equivalente a uma percentagem de 40.58% e 37.4% contra 22.01%. Um dos principais motivos para a redução drástica quando comparados os valores, é a utilização de termos que visam denegrir a imagem ou responsabilizar a vítima pelo ocorrido, por exemplo: “inacreditável”, “mulher”, “vivida”; ou a disseminação de

discursos de ódio entre os usuários, característica comum para plataformas que permitam a interação, principalmente por conta da presente cultura retrógrada e machista perceptível na sociedade brasileira.

Realizou-se também a correlação entre a análise de sentimento e a quantidade de *likes & dislikes*, tal como apresentado na Figura 2.b. É importante destacar que comentários que possuíam a mesma quantidade de avaliações (*e.g.* 20 *likes* e 20 *dislikes*) foram desconsiderados. Por meio da análise da 2.b é possível verificar que mesmo com a predominância de aprovação nas três classes, a diferença é notória dentro da categoria negativa. Realizou-se também uma análise da distribuição dos *likes & dislikes* (aceitação e rejeição), tal como apresentado na Figura 3.



**Figura 3. Percentagens de avaliações gerais.**

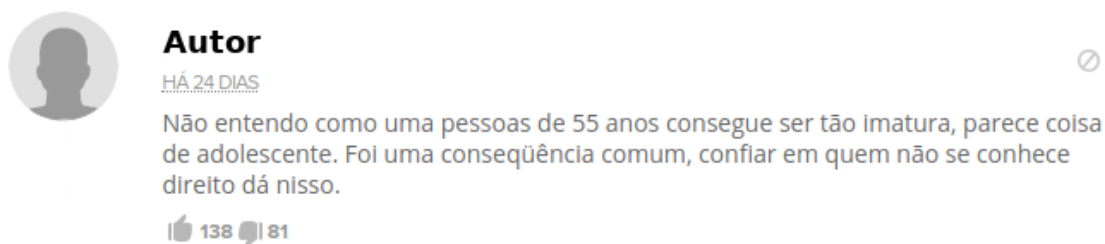
Por meio da análise da Figura 3 é possível afirmar que há uma convergência quanto aos resultados apresentados, com uma maior taxa de aceitação dos comentários publicados nas notícias do caso (67,9%); em detrimento da rejeição aos demais que podem estar posicionados de forma neutra ou crítica (32,1%). A Tabela 4 apresenta alguns comentários coletados.

Os comentários dispostos na Tabela 4 foram selecionados por possuírem grande quantidade de interações. Por exemplo, na categoria de rejeição, o Item 1 contabilizou 122 *likes* contra 510 *dislikes*; e, na de aceitação, o Item 2 obteve 271 *likes* contra 234 *dislikes*. Os dois objetivam responsabilizar a Elaine Caparroz pelo ocorrido, a exemplo das expressões “*se tivesse um comportamento de mulher direita*” e “*conhece o cara na net e já leva pra casa*”. No exemplo há uma orientação mais direcionada ao individual, a ela, com quantidade de *dislikes* superior; e no segundo, tem-se o direcionamento também para as mulheres em geral, com uma margem de aproximadamente 15,8% superior em *likes*. Além disso, é possível identificar comentários com o viés de responsabilizá-la por se relacionar pelo intermédio da internet, com tal imprudência e, subentendido, com uma pessoa mais nova, como por exemplo, no Item 3.

Na mesma linha de raciocínio também há as publicações escritas por outras mulheres, como no caso dos Itens 4 e 5 da Tabela 4. Na Figura 4, com seu respectivo autor e foto ocultados (anonimato), é apresentado um exemplo na forma de captura de tela extraída diretamente de uma das notícias e presente no conjunto de dados.

**Tabela 4. Exemplos de comentários extraídos das notícias.**

| Comentário |   |
|------------|---|
| 1          | <i>“Pagou o preço pela promiscuidade, se tivesse um comportamento de mulher direita...”</i>   |
| 2          | <i>“Conhece o cara na net e ja leva pra casa. Infelizmente deu azar de ser um psicopata. Mulheres se dêem o valor.”</i>   |
| 3          | <i>“Meu Deus como pode uma pessoa de 55 anos, ser tão ingênua de colocar um cara que nunca viu, conhece somente pela internet, dentro de casa (...)”</i>  |
| 4          | <i>“com essa violência toda e essa criatura põe um estranho pra dentro de casa... sinceramente, que imprudência (...)”</i>  |
| 5          | <i>“Me desculpem, mas essa moça foi muito ingênua, se o cara já chegou dando nome falso ela já deveria ter desconfiado, mais ainda não deveria ter marcado primeiro encontro na casa dela e sim em local publico (...)”</i> |
| 6          | <i>“tem gente otária demais. como que pode uma mulher acreditar nessas redes sociais e marcar encontro”</i>   |



**Figura 4. Captura de tela de um comentário.**

Seguindo a metodologia descrita na Seção 2, realizou-se a modelagem de tópicos. Após a análise dos cenários em relação ao método (LDA ou LSA), quantidade de tópicos (3, 5 e 10) e quantidade de palavras por tópicos (3, 5 e 10), optou-se pelo seguinte cenário: LSA, com cinco tópicos e 10 palavras por tópico, pois os avaliadores reportaram maior facilidade na anotação dos dados e também houve um maior consenso na rotulação dos tópicos identificados. Na Tabela 5 os resultados para a modelagem de tópicos são apresentados.

**Tabela 5. Tópicos extraídos a partir dos comentários.**

| Tópico              | Termos   |
|---------------------|--|
| Relacionamento      | cariocas sao mulher cara tao so ja gostam vai pessoas                |
| Preferências        | cariocas sao gostam tao atentos sinal fechado espertos motivos vazia |
| Internet/Social     | pessoas vai ja redes sociais pessoa conhecer pode estao ate          |
| Violência Doméstica | mulher vai homem covarde ai ate bate familia virar uns               |
| Delito              | vai cara ai agora ja ver ate cariocas cadeia mae                     |

Os tópicos apresentados na Tabela 5 foram identificados e discutidos entre os autores, que chegaram à um consenso ao final do processo. O grupo "Internet/Social", por



exemplo, pode ser associado a comentários que criticam a atitude de Elaine Caparroz ao utilizar as redes sociais para um encontro, exemplo citado no Item 6 da Tabela 4. A abordagem utilizada foi semelhante ao de [Melville et al. 2019].

Mediante a análise de legibilidade, foram selecionados os comentários a partir dos critérios predefinidos na Seção 2, Como resultado obteve-se 107 textos avaliados (redução de 85.8% a partir do *dataset* utilizado). Os escores gerados foram distribuídos conforme seus respectivos significados de escolaridade, dados apresentados na Figura 5. Percebe-se que grande parte dos autores dos comentários do caso pertencem à classe de 5ª à 8ª série, o equivalente à 52.3%; seguida de Ensino Médio e Nível Superior, com 31.8%; 1ª à 4ª série, com 14%; e Textos Acadêmicos (mais elevado grau de escolaridade), com 1.9%.

Segmentações por nível de escolaridade

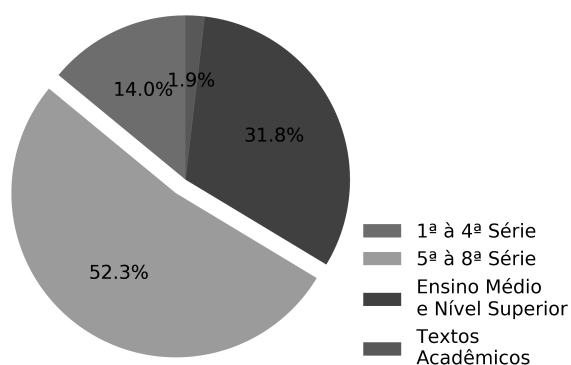


Figura 5. Distribuição da escolaridade dada legibilidade.

Com o intuito de visualizar quais classes possuem sentimentos mais negativos ou positivos, verificou-se a correlação dessa métrica com a análise de sentimentos. Para isso, foi calculada a ocorrência de cada um e sem incluir os comentários rotulados como neutros. Por conseguinte, na Figura 6 é exibido que a ocorrência de negativos é maior nos níveis escolares de 5ª à 8ª série e Ensino Médio e Nível Superior.

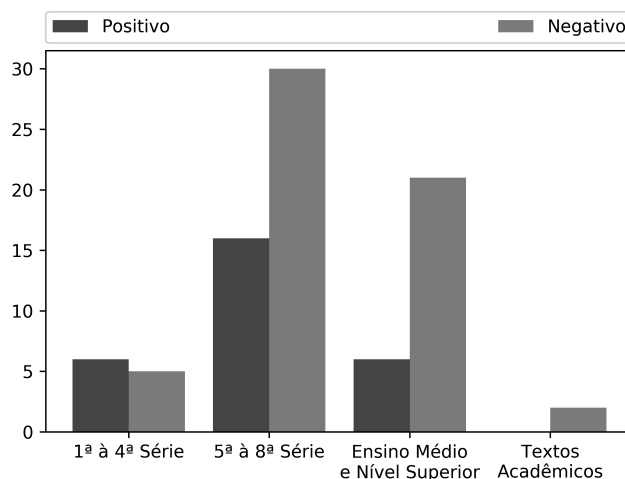


Figura 6. Ocorrência de sentimento por escolaridade.

Por meio da análise dos resultados dispostos na Figura 6 é possível identificar que embora a maior parte dos comentários que foram classificados como negativos foram feitos por pessoas categorizadas como ensino médio ou superior, eles não fazem menções contra a vítima e sim contra o agressor, ou manifestando indignação ao fato. Por meio de análise exploratória observou-se que alguns dos comentários tratavam-se de argumentos que defendiam a vítima por meio de críticas ao agressor e ao caso de forma geral, ponto este embasado pela análise apresentada na Figura 1-b, que discorre a maior aceitação por comentários classificados como negativos. Releva-se também um posicionamento preconceituoso, ofensivo e imaturo por parte de um grupo que, em teoria, tem a capacidade para discernir o que, em termos de sociedade, é eticamente aceitável ou não.

#### 4. Considerações Finais

Neste artigo apresentou-se a análise de comentários das notícias sobre o caso da tentativa de feminicídio de Elaine Caparroz, 55 anos, ocorrido em 17 de fevereiro de 2019. Este caso obteve notoriedade pois a vítima foi agredida por aproximadamente quatro horas até que houvesse a denúncia por parte de seus vizinhos. Além disso, aspectos do discurso das notícias chamaram atenção devido ressaltarem a diferença de idade entre a vítima e o agressor e pelo contexto pelo qual se conheceram - por meio de aplicativos de relacionamento. Para realizar o estudo, 1.120 comentários de 31 notícias foram coletados e 754 foram selecionados para as análises subsequentes, que consistiram na análise de sentimento, correlação com as interações de *likes & dislikes* disponibilizados pela plataforma de notícias, modelagem de tópicos e estudo de legibilidade.

Os resultados obtidos permitiu-nos responder as perguntas de pesquisa anteriormente levantadas. Acerca da primeira, “Qual o posicionamento das pessoas diante de notícias sobre o tema?” é possível identificar uma tendência em responsabilizar Elaine Caparroz pela tentativa de feminicídio por ela sofrida. Alguns convergiam para uma crítica mais severa a partir de argumentos de base mista, isto é, com a assimilação de características presentes em discursos de ódio e afirmações que entram em uma escala de degradação tanto das redes, quanto da própria vítima. A partir da métrica verificada pelo nível de escolaridade, também é possível responder a segunda pergunta de pesquisa, “Qual a correlação entre o nível de instrução e posicionamento?”. Como resposta desta pergunta constatou-se que a maioria está em um grupo abrangido da quinta até a oitava série do ensino fundamental. Identificou-se também que grande parte desses comentários são gerados, na categoria de sentimento negativo, a partir do grupo de quinta a oitava série do ensino fundamental.

Em uma avaliação geral é possível perceber algumas características retrógradas nos discursos dos comentários, refletindo a alguns posicionamentos machistas ainda presente em parte da sociedade, mesmo em um ambiente com relações de escala global e de rápida propagação, como a internet. Um estudo mais aprofundado faz-se necessário a fim aumentar a confiabilidade das conclusões. Neste ensejo convém pontuar algumas ameaças a validade da pesquisa, como por exemplo a não remoção de *emojis* na etapa de pré-processamento. No presente estudo os *emojis* foram removidos e tal fato pode causar impactos nos resultados, haja vista que eles podem representar sentimentos de modo não verbal [Kralj Novak et al. 2015]. Outro ponto importante diz respeito à interpretação do espectro de aceitação de um dado comentário, com uma classificação passível de ser erroneamente definida como positiva, quando esta deveria ser negativa e vice-versa.

Sendo assim, os autores pretendem, como trabalhos futuros, reestruturar o *framework* experimental para diminuir as ameaças a validade da pesquisa, além de conduzir um estudo nas notícias em si para verificar se há uma tendência a provocar ou não um sentimento negativo ao leitor antes dele realizar o acesso.

## Agradecimentos

Este trabalho foi parcialmente financiado pelo Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) e Fundação Amazônia Paraense de Amparo à Pesquisa (FAPESPA). Agradecemos também aos revisores pelos comentários e sugestões enriquecedoras que muito auxiliaram na melhora do manuscrito.

## Referências

- Almeida, G. R. d., Cirqueira, D. R., and Lobato, F. M. (2017). Improving social crm through eletronic word-of-mouth: a case study of reclameaqui. *XIV Workshop de Trabalhos de Iniciação Científic*.
- Boira, S., Tomas-Aragones, L., and Rivera, N. (2017). Intimate partner violence and femicide in Ecuador. *Qualitative Sociology Review*, 13(3):30–47.
- Caldas, L. V. A., Jacob, A. F. L., Silva, S. S. C., Pontes, F. A. R., and Lobato, F. M. F. (2018). Development of a social network for research support and individual well-being improvement. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 383–386.
- Cirqueira, D., de Lima, R. S., Bueno, S., Neme, C., Ferreira, H., Coelho, D., Alves, O. P., Pinheiro, M., Astolfi, R., and Marques, D. (2018). Atlas da Violência 2018. *IPEA e FBSP*, page 93.
- 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.
- Cirqueira, D., Jacob, A., Lobato, F., de Santana, A. L., and Pinheiro, M. (2017). Performance Evaluation of Sentiment Analysis Methods for Brazilian Portuguese. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, pages 245–251. Springer International Publishing, Cham.
- Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V., and Bhamidipati, N. (2015). Hate Speech Detection with Comment Embeddings. In *Proceedings of the 24th International Conference on World Wide Web - WWW '15 Companion*, pages 29–30, New York, New York, USA. ACM Press.
- ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., and Belding, E. (2018). Hate Lingo: A Target-based Linguistic Analysis of Hate Speech in Social Media. *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*, pages 42–51.
- Figueira, Á. and Guimarães, N. (2017). Detecting Journalistic Relevance on Social Media. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17*, pages 1136–1139.

- Garcia, L. P., de Freitas, L. R. S., da Silva, G. D. M., and Höfelmann, D. A. (2013). Violência contra a mulher: feminicídios no Brasil. *Instituto de Pesquisa Econômica Aplicada*, page 5.
- Kralj Novak, P., Smailović, J., Sluban, B., and Mozetič, I. (2015). Sentiment of Emojis. *PLOS ONE*, 10(12):e0144296.
- Lobato, F., Pinheiro, M., Jacob, A., Reinhold, O., and Santana, Á. (2017). Social crm: Biggest challenges to make it work in the real world. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, pages 221–232. Springer International Publishing, Cham.
- Lobato, F. M. F., da Silva, M., Coelho, K., da Costa Silva, S., and Pontes, F. (2018). Vamos falar sobre deficiência? Uma análise dos Tweets sobre este tema no Brasil. *Proceedings of the 7th Brazilian Workshop on Social Network Analysis and Mining*, pages 274–279.
- Martins, T. B. F., Ghiraldelo, C. M., Nunes, M. d. G. V., and de Oliveira Junior, O. N. (1996). Readability formulas applied to textbooks in brazilian portuguese.
- Melville, S., Eccles, K., and Yasseri, T. (2019). Topic Modeling of Everyday Sexism Project Entries. *Frontiers in Digital Humanities*, 5:5:28.
- Meneghel, S. N. and Portella, A. P. (2017). Feminicídios: conceitos, tipos e cenários. *Ciência & Saúde Coletiva*, 22(9):3077–3086.
- Mondal, M., Silva, L. A., and Benevenuto, F. (2017). A Measurement Study of Hate Speech in Social Media. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media - HT '17*, pages 85–94, New York, New York, USA. ACM Press.
- Ramos, P., Reis, J., and Benevenuto, F. (2016). Uma Análise da Polaridade Expressa nas Manchetes de Notícias Brasileiras. *Proceedings of the 5th Brazilian Workshop on Social Network Analysis and Mining*, pages 187–198.
- Schrifer, K. A. (1989). Evaluating text quality: The continuum from text-focused to reader-focused methods. *IEEE Transactions on professional communication*, 32(4):238–255.
- 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.
- Teixeira, M. A. M., Lobato, F. M. F., Chagas, B. N. R., and Jacob Junior, A. F. L. (2018). A System of Acquisition and Analysis of Data for Extraction of Knowledge of the Ebit Platform. In *15th International Conference on Information Systems & Technology Management*, pages 4195–4206.
- Yesbeck, J. (2018). How are Alexa’s traffic rankings determined?