

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Isabella Tannús Corrêa

**Analisando o sentimento dos usuários do
Twitter em relação a casos de grande
repercussão de violência contra a mulher**

Uberlândia, Brasil

2020

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Isabella Tannús Corrêa

Analisando o sentimento dos usuários do *Twitter* em relação a casos de grande repercussão de violência contra a mulher

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Universidade Federal de Uberlândia – UFU

Faculdade de Ciência da Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2020

Isabella Tannús Corrêa

Analisando o sentimento dos usuários do *Twitter* em relação a casos de grande repercussão de violência contra a mulher

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Uberlândia, Brasil, 08 de dezembro de 2020:

**Profa. Dra. Elaine Ribeiro de Faria
Paiva**
Orientador

**Profa. Dra. Maria Adriana Vidigal de
Lima**

Profa. Dra. Rita Maria da Silva Julia

Uberlândia, Brasil
2020

Dedico este trabalho a todas as mulheres que já sofreram algum tipo de violência.

Vocês não estão sozinhas.

A culpa nunca é da vítima.

Agradecimentos

À minha orientadora, Elaine, por ter embarcado nessa jornada tão importante comigo. Por me orientar, ser paciente, me fazer pensar fora da caixa e por ser uma grande inspiração profissional para mim. Muito obrigada.

Agradeço aos meus pais, por sempre acreditarem no meu potencial e desde pequena me incentivarem a estudar. “Estudo é a única coisa que ninguém tira de nós” foi uma das frases que eu mais escutei dos meus pais durante toda minha vida acadêmica. Sempre que pensei que eu não me encaixava nesse curso, me mostraram que eu posso ser quem eu quiser. Essa conquista é nossa.

Agradeço aos meus irmãos, por serem os melhores que eu poderia pedir. Sempre acreditaram em mim, levantaram minha auto-estima e me ajudaram. São pessoas que eu me inspiro muito.

Agradeço às minhas avós por me apoiarem e me incentivarem. Aos meus tios, pelas inúmeras reportagens e ideias sobre o tema deste estudo, para que eu pudesse concluí-lo com êxito.

Agradeço ao meu padrinho, que, infelizmente, não está mais presente fisicamente em minha vida. Gostaria que ele tivesse tido a oportunidade de ler este trabalho, afinal grande parte do que sou devo a ele. Espero que ele e meus avós, onde estiverem, estejam orgulhosos de mim.

Agradeço aos meus amigos, que sempre ouviram com atenção quando eu falava sobre o tema. Por sempre estarem presentes quando eu mais precisei. Muito obrigada. Agradeço imensamente ao meu namorado, por ser um parceiro incrível, meu melhor amigo e sempre me encorajar. Todas as vezes que pensei em desistir, estava lá para segurar minha mão.

Aos meus professores, por todo o conhecimento que foi passado. Obrigada pela dedicação e esforço de cada um, que foram essenciais durante minha vida acadêmica.

“I am not free while any woman is unfree, even when her shackles are very different from my own.” Audre Lorde

Resumo

O índice de violência no Brasil é muito alto, em especial, quando se considera a violência contra mulheres. A cada 2 segundos, uma mulher é vítima de violência física ou verbal no Brasil, e a cada 1,4 segundo uma mulher é vítima de assédio. A violência contra as mulheres começou a ser vista como um problema público com o crescimento da antropologia e noções do feminismo. Visto que muitos casos de violência contra mulheres são noticiados e recebem muita atenção das redes sociais, é interessante analisar as opiniões das pessoas sobre os mesmos.

Este trabalho propõe analisar os sentimentos expressos pelos usuários da rede social *Twitter* em relação a casos de violência contra mulheres que obtiveram grande repercussão na mídia. Para isso, dados foram coletados desta rede social, foram pré-processados, divididos em tópicos aplicando o *Latent Dirichlet Allocation* (LDA) e classificados usando o algoritmo *Naive Bayes*.

A partir da aplicação do método proposto aos *tweets* coletados, foi observado que a maior parte dos textos classificados, quando eram comentários referentes às vítimas, continham sentimentos “positivos”, ao passo que, quando relativos aos agressores, continham sentimentos “negativos”.

Palavras-chave: Violência contra mulheres, *Twitter*, Mineração de Texto, Análise de Sentimentos.

Lista de ilustrações

Figura 1 – Consulta ao <i>Twitter</i> realizada por meio da interface de linha de comando do <i>GetOldTweets</i>	29
Figura 2 – Nuvem de palavras contidas na base de dados do Caso Eloá – Gerada a partir do <i>WordItOut</i>	32
Figura 3 – Trecho da implementação do LDA na linguagem <i>Python</i>	33
Figura 4 – Trecho do arquivo ARFF utilizado para alimentar o <i>Weka</i>	36
Figura 5 – Visão do conjunto de treinamento carregado no <i>Weka</i>	36
Figura 6 – Quantidade de <i>tweets</i> coletados a cada 2 anos sobre todos os casos	41
Figura 7 – Quantidade de <i>tweets</i> coletados por ano sobre todos os casos	41
Figura 8 – Nuvem de palavras contidas nos <i>tweets</i> coletados com as palavras-chave do Caso Elaine Caparroz – Gerada a partir do <i>WordItOut</i>	44
Figura 9 – Nuvem de palavras contidas nos <i>tweets</i> coletados com as palavras-chave do Caso Eliza Samúdio – Gerada a partir do <i>WordItOut</i>	45
Figura 10 – Nuvem de palavras contidas nos <i>tweets</i> coletados com as palavras-chave do Caso Eloá Cristina – Gerada a partir do <i>WordItOut</i>	45
Figura 11 – Nuvem de palavras contidas nos <i>tweets</i> coletados com as palavras-chave do Caso Eva Luana – Gerada a partir do <i>WordItOut</i>	46
Figura 12 – Nuvem de palavras contidas nos <i>tweets</i> coletados com as palavras-chave do Caso Mariana Ferrer – Gerada a partir do <i>WordItOut</i>	47
Figura 13 – Nuvem de palavras contidas nos <i>tweets</i> coletados com as palavras-chave do Caso Poliana Bagatini – Gerada a partir do <i>WordItOut</i>	47
Figura 14 – Nuvem de palavras contidas nos <i>tweets</i> coletados com as palavras-chave do Caso Tatiane Spitzner – Gerada a partir do <i>WordItOut</i>	48
Figura 15 – Exemplo de <i>tweet</i> coletado que justifica a rotulação do tópico Culpa no Caso Elaine Caparroz.	49
Figura 16 – Exemplo de <i>tweet</i> coletado que justifica a rotulação do tópico Culpa no Caso Elaine Caparroz.	49
Figura 17 – Exemplo de <i>tweet</i> coletado que justifica a rotulação do tópico Descredibilização da Vítima no Caso Eliza Samúdio.	50
Figura 18 – Exemplo de <i>tweet</i> coletado que justifica a rotulação do tópico Culpa no Caso Eloá Cristina.	51
Figura 19 – Visualização gráfica dos sentimentos expressos pelos usuários do <i>Twitter</i> em relação aos casos	54
Figura 20 – Visualização gráfica dos sentimentos expressos pelos usuários do <i>Twitter</i> em relação aos agressores	55

Figura 21 – Visualização gráfica dos sentimentos expressos pelos usuários do <i>Twitter</i> em relação às vítimas	55
---	----

Lista de tabelas

Tabela 1 – Descrição dos casos usados neste estudo.	20
Tabela 2 – Palavras-chave na coleta pelo <i>Twitter</i> separados por assunto	29
Tabela 3 – Sumarização dos dados contidos na base de <i>tweets</i> rotulados	34
Tabela 4 – Sumarização dos dados contidos nas novas bases de <i>tweets</i> rotulados	35
Tabela 5 – Matriz de confusão para o algoritmo Naive Bayes	38
Tabela 6 – Palavras-chave na coleta pelo <i>Twitter</i> e quantidade de <i>tweets</i> coletados.	40
Tabela 7 – Exemplos de <i>stopwords</i> do algoritmo em <i>Java</i>	42
Tabela 8 – Exemplos de gírias e abreviações com suas traduções no programa em <i>Java</i>	42
Tabela 9 – Exemplo de um <i>tweet</i> antes e depois do pré-processamento	43
Tabela 10 – Assuntos na coleta pelo <i>Twitter</i> e quantidade de <i>tweets</i> coletados.	43
Tabela 11 – Tópicos com 6 palavras referentes ao Caso Elaine Caparroz	49
Tabela 12 – Tópicos com 6 palavras referentes ao Caso Eliza Samúdio	50
Tabela 13 – Tópicos com 6 palavras referentes ao Caso Eloá Cristina	51
Tabela 14 – Tópicos com 6 palavras referentes ao Caso Eva Luana	52
Tabela 15 – Tópicos com 6 palavras referentes ao Caso Mariana Ferrer	52
Tabela 16 – Tópicos com 6 palavras referentes ao Caso Poliana Bagatini	53
Tabela 17 – Tópicos com 6 palavras referentes ao Caso Tatiane Spitzner	53
Tabela 18 – Matriz de confusão para o algoritmo <i>Naive Bayes</i> para o conjunto total de treinamento	56
Tabela 19 – Matriz de confusão para o algoritmo <i>Naive Bayes</i> para o conjunto sobre as vítimas	57
Tabela 20 – Matriz de confusão para o algoritmo <i>Naive Bayes</i> para o conjunto sobre os agressores	57
Tabela 21 – Matriz de confusão para o algoritmo <i>Naive Bayes</i> para o conjunto sobre os casos	57
Tabela 22 – Acurácia calculada para cada uma das bases	57

Sumário

1	INTRODUÇÃO	12
1.1	Motivação	13
1.2	Objetivos	14
1.2.1	Objetivo Geral	14
1.2.2	Objetivos Específicos	15
1.3	Organização do Trabalho	15
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Violência contra a mulher	16
2.1.1	Tipos de Violência	17
2.1.1.1	Violência sexual	17
2.1.1.2	Violência Moral	17
2.1.1.3	Violência Psicológica	17
2.1.1.4	Violência Física	18
2.1.1.5	Violência Patrimonial	18
2.1.2	Cronologia da Violência	18
2.1.2.1	Aumento da Tensão	18
2.1.2.2	Ato de Violência	18
2.1.2.3	Arrependimento e Comportamento Carinhoso	18
2.1.3	Estatísticas	19
2.1.4	Casos de violência e feminicídio	19
2.1.5	Órgãos de Apoio	19
2.2	Mineração de Texto	21
2.3	Etapas da Mineração de Texto	22
2.3.1	Coleta de Dados	22
2.3.2	Pré-processamento dos dados	23
2.3.3	Busca por padrões	24
2.4	Classificação dos textos	24
2.4.1	Aprendizado supervisionado	25
2.4.2	Aprendizado não supervisionado	25
2.5	Naive Bayes	25
2.6	Trabalhos Relacionados	25
2.7	Considerações finais	27
3	MÉTODO PARA ANÁLISE DOS DADOS DO TWITTER SOBRE OS CASOS DE VIOLÊNCIA	28

3.1	Coleta de Dados	28
3.2	Pré-Processamento	30
3.3	Busca por padrões	31
3.3.1	Nuvem de palavras	31
3.3.2	<i>Latent Dirichlet Allocation (LDA)</i>	31
3.4	Classificação dos textos	34
3.4.1	Construção da Base de Dados Rotulada	34
3.4.2	Classificador <i>Naive Bayes</i>	35
3.4.3	Avaliação dos Classificadores	37
4	RESULTADOS	39
4.1	Coleta de Dados	39
4.2	Pré-Processamento	41
4.3	Busca por padrões	43
4.3.1	Nuvens de palavras	43
4.3.2	Modelagem de Tópicos	48
4.4	Análise da base de dados rotulada manualmente	54
4.5	Avaliação do Classificador	55
5	CONCLUSÃO	58
5.1	Trabalhos Futuros	59
ANEXO		60
A	CASOS DE VIOLÊNCIA E FEMINICÍDIO	61
A.1	Caso Elaine Caparroz	61
A.2	Caso Eliza Samúdio	61
A.3	Caso Eloá Cristina	63
A.4	Caso Eva Luana	64
A.5	Caso Mariana Ferrer	64
A.6	Caso Poliana Bagatini	66
A.7	Caso Tatiane Spitzner	67
	REFERÊNCIAS	69

1 Introdução

Segundo a Organização Mundial da Saúde, em 2018, o Brasil ocupava o nono lugar em um *ranking* de países mais violentos. O índice de violência pode ser ainda mais problemático se pensado em relação às mulheres. De acordo com o Ministério da Mulher, da Família e dos Direitos Humanos (MMFDH) (MULHER, 2019b), o canal de denúncias Ligue 180 (Central de Atendimento à Mulher) registrou 17.836 denúncias nos dois primeiros meses de 2019, um aumento de 36,85% em relação ao mesmo período no ano anterior.

A violência contra as mulheres ocorre desde muito tempo, mas começou a ser vista como um problema público somente a partir de 1970, com o crescimento da antropologia e noções de feminismo (DAUPHIN et al., 2000). É possível dizer que muitas pessoas viam, e ainda veem, todo tipo de agressão contra a mulher, principalmente se esta for feita por um cônjuge, como justificável por se tratar de “defesa de honra” (FEDERAL, 2018). As mulheres então são vistas, não como as verdadeiras vítimas, mas como as culpadas por certa brutalidade ter acontecido, e são tratadas muitas vezes sem humanidade, sendo que passaram por um trauma que não se esquecerão (BORGIA, 2016).

A Lei Maria da Penha (MULHERES, 2006) foi criada em 2006 como meio de coibir a violência doméstica e familiar contra a mulher. Em 2015 foi instituída a Lei do Femicídio (ROUSSEFF, 2015), desse modo, um crime contra a vida é considerado feminicídio quando for cometido contra uma vítima por ela ser do sexo feminino. Tais leis não são tão efetivas na realidade, afinal acredita-se que muitas mulheres não denunciam por, em grande parte, medo de seus agressores (BIANCHINI, 2011). Contudo, outros motivos que são usados pelas vítimas são: acreditar que seria a última vez que aconteceria tal atrocidade, dependência financeira, a falta de conhecimento de seus direitos, vergonha de se declarar vítima, percepção de que o agressor sai imune dessas situações, e preocupação com a família (BIANCHINI, 2011).

Devido ao crescente uso da Internet, em especial das redes sociais, nos últimos anos, algumas vítimas começaram a divulgar nas redes sociais, os casos de violência sofrida ou pedidos de ajuda. O objetivo de tal divulgação é dar visibilidade ao relato e conseguir justiça. Como exemplo, pode-se citar os casos de Eva Luana, que relatou os abusos que sofria pelo padrasto (CLAUDIA, 2019) por meio do Instagram¹ e de Mariana Ferrer, que contou no Instagram² e no *Twitter*³ sobre o estupro que sofreu em um *beach club* em Florianópolis. As redes sociais podem ser um refúgio para quem já sofreu ou ainda sofre

¹ <https://www.instagram.com/p/BuEf2-xHl4L/>

² <https://bit.ly/instagram-mariana-ferrer>

³ <http://bit.ly/caso-mariana-ferrer>

violência, seja o tipo que for, doméstica e familiar, assédio, importunação sexual, violência física, moral, sexual, obstétrica ou até virtual. A Internet também pode ser um ambiente em que os agressores postam sobre, mesmo que implicitamente, as razões e insatisfações que os levam a cometer algum crime. Do mesmo modo, por meio do anonimato, ou até mesmo pelo não conhecimento das leis que existem na Constituição Federal e nos Códigos Civil e Penal sobre a Internet (WIZIACK, 2006), usuários usam as redes sociais para tecer comentários sobre casos de violência que acontecem no mundo inteiro.

Além das redes sociais, existem outras fontes de informações que podem ser valiosas para entender melhor os casos de violência contra a mulher e a opinião do público sobre esses casos. Jornais *online* divulgam casos dos diversos tipos de violência, em que estes podem ou não estar anonimizados e, muitas vezes, são ricos em detalhes do fato noticiado. Eles também contêm um conjunto de comentários do público em geral, que expressam seu sentimento sobre o fato, vítima ou agressor. O portal de notícias Globo.com⁴ é um dos sites mais acessados no Brasil de acordo com Alexa (2020), e gera grandes volumes de dados todos os dias.

Desse modo, é inegável que a *Web* é um espaço em que há maior número de informações sobre as pessoas no mundo atual (VASCONCELOS; CARVALHO, 2018). Por esse motivo, trabalhos recentes como o de Rodrigues, Junior e Lobato (2019) e Corrêa (2017) têm automatizado a busca e processamento de textos que, acompanhadas de técnicas de Análise de Sentimentos, tornam possível a descoberta da opinião dos usuários com relação a diversos assuntos. Este trabalho visa avaliar o sentimento dos usuários da rede social *Twitter* sobre casos de violência contra a mulher e feminicídio, que ganharam grande repercussão nas mídias.

1.1 Motivação

A cada 2 segundos, uma mulher é vítima de violência física ou verbal no Brasil, e a cada 1,4 segundo uma mulher é vítima de assédio (PENHA, 2019a). Em 2019, o Ligue 180 (MULHER, 2019a) registrou 85.412 relatos de violência contra a mulher, sendo os maiores números de violência física e psicológica. Deste total, 78,96% foram classificados como violência doméstica. Foram protocolados também 61 feminicídios e 3.715 tentativas de feminicídio no mesmo período. O Ligue 180 recebe as denúncias, faz os encaminhamentos necessários e informa a vítima quanto às leis que podem protegê-la, como Lei Maria da Penha, Lei do Feminicídio e os Direitos da Mulher.

No Brasil, muitos órgãos governamentais têm criado aplicativos para que mulheres relatem e denunciem casos de agressão. A Prefeitura de Uberlândia criou, em 2019 o aplicativo Salve Maria (G1, 2019). Usando esse aplicativo é possível denunciar abusos

⁴ <https://www.globo.com/>

com sigilo e segurança. Ele conta também com o chamado Botão do Pânico para casos de emergência que, quando acionado, envia a localização da vítima em tempo real para a Polícia. Assim, com ênfase na violência doméstica sendo a principal causa de mortes dentro do feminicídio, foi criado o aplicativo Mete a Colher (MAGALHÃES et al., 2016). Por meio desse aplicativo é possível pedir ajuda a outras mulheres que estiverem dispostas a ajudar e prestar apoio emocional, orientação jurídica e ajuda de inserção no mercado de trabalho a vítimas de agressão.

Além dos aplicativos e sistemas para denúncia de casos de violência contra a mulher, as mídias sociais tornaram-se um veículo de destaque para divulgar casos de violência. Alguns casos de violência contra a mulher ganharam destaque nas mídias sociais e são alvo de muitos comentários por parte dos usuários de redes sociais e de sites de notícias. Os comentários giram em torno do ato de agressão, do agressor e da própria vítima. É importante notar, que o sentimento presente nas postagens varia bastante para um mesmo caso de agressão. Enquanto muitas pessoas se solidarizam com a vítima e mostram um sentimento de apoio e de busca por justiça, muitos ainda são coniventes com o ato de violência ou com o agressor.

Considerando o cenário de violência contra a mulher, o trabalho de Rodrigues, Junior e Lobato (2019) analisa, a partir de métodos de modelagem de tópicos e análise de sentimentos, comentários de notícias sobre a tentativa de feminicídio de Elaine Caparroz, fato que ocorreu no dia 17 de fevereiro de 2019 no Rio de Janeiro. Foram coletados comentários de notícias publicadas e feita a análise de sentimentos desses comentários. Os resultados indicaram uma tendência em responsabilizar Elaine Caparroz pela tentativa de feminicídio por ela sofrida.

Além do trabalho de Rodrigues, Junior e Lobato (2019), não foram encontrados outros trabalhos de Mineração de Texto ou Análise de Sentimentos com foco na violência contra a mulher. Este estudo pretende descobrir o que os usuários do *Twitter* pensam a respeito de vítimas e agressores de casos de violência contra a mulher e feminicídio, que ganharam destaque nas mídias sociais. Para este trabalho, técnicas de análise de sentimentos e mineração de textos serão usadas.

1.2 Objetivos

1.2.1 Objetivo Geral

Este trabalho tem como objetivo analisar postagens da rede social *Twitter* sobre casos de violência feminina que obtiveram grande repercussão na mídia desde 2008 e, assim, verificar o sentimento das postagens sobre o assunto. Foram escolhidos casos de feminicídio e casos envolvendo diferentes tipos de violência sofrida: física, psicológica,

sexual, patrimonial ou moral.

Serão utilizadas técnicas de pré-processamento e classificação dos textos, que indicam se o sentimento expresso no *tweet* é “positivo”, “neutro” ou “negativo”. Espera-se identificar o sentimento em relação à vítima, ao agressor ou ao caso em si.

1.2.2 Objetivos Específicos

- Criar uma base de dados distinta para cada caso de violência contra mulher e feminicídio construída a partir de postagens no *Twitter*.
- Sumarizar os *tweets* sobre cada caso de violência por meio de nuvem de palavras e *Latent Dirichlet Allocation* (LDA), permitindo assim uma análise sobre quais assuntos ou palavras foram mais mencionadas pelos internautas para cada caso de violência.
- Identificar o sentimento quanto à vítima, agressor ou quanto ao caso em si por meio de *tweets*.
- Tecer conclusões sobre os cenários nos quais as vítimas ou o ato de violência são discriminadas pelos internautas.

1.3 Organização do Trabalho

O trabalho é organizado da seguinte maneira.

- **Capítulo 2 — Revisão Bibliográfica:** a fundamentação teórica necessária para o desenvolvimento e entendimento do estudo é apresentada, incluindo uma visão geral sobre violência contra mulher e casos que serão usados para o estudo, além da descrição de conceitos de Análise de Sentimentos e citação de trabalhos relacionados.
- **Capítulo 3 — Desenvolvimento:** expõe todas as etapas realizadas para o desenvolvimento deste trabalho, incluindo a justificativa para escolha das ferramentas usadas.
- **Capítulo 4 — Resultados:** os dados obtidos a partir das técnicas de Análise de Sentimentos são sumarizados e analisados de modo a obter o sentimento dos usuários quanto aos casos e às vítimas.
- **Capítulo 5 — Conclusão:** as principais conclusões deste estudo são apresentadas, além de sugestões para trabalhos futuros.

2 Revisão Bibliográfica

O objetivo deste capítulo é apresentar a fundamentação teórica necessária para o desenvolvimento e entendimento deste estudo, além de trabalhos relacionados ao tema.

A Seção 2.1 apresenta uma visão geral sobre violência contra a mulher, expondo um resumo dos casos escolhidos, os tipos de violência, cronologia da violência, estatísticas e órgãos de apoio.

A Seção 2.2 descreve a Mineração de Texto. A Seção 2.3 expõe as etapas referentes à tarefa de Mineração de Texto e como podem ser executadas. São elas: coleta de dados, pré-processamento dos dados, busca por padrões e classificação dos textos. A Seção 2.6 cita trabalhos relacionados e suas principais colaborações para área de Mineração de Texto e Análise de Sentimentos.

2.1 Violência contra a mulher

As primeiras ações de enfrentamento à violência contra a mulher são datadas desde a década de 1980 (FEDERAL, 2011). Essas buscavam o anúncio de políticas nas áreas de segurança pública, saúde, justiça e assistência social. Em 1988 foi promulgada a Constituição Federal que previa a promoção dessas medidas. Em 2006, no âmbito do combate a violência contra a mulher, é criada a Lei nº 11.340 - Lei Maria Da Penha (MULHERES, 2006), e em 2015 foi instituída a Lei do Feminicídio (ROUSSEFF, 2015). Mesmo com medidas legislativas para a diminuição de violação em desfavor das mulheres, a realidade ainda continua desastrosa. Com base nas denúncias feitas no Ligue 180 (MULHER, 2020b), é possível traçar um panorama geral do perfil das vítimas e agressores, e dados sobre a violência no ano de 2019.

Em 84% dos registros do Ligue 180, pessoas do sexo masculino foram os agressores. Dentre os agressores, os casos mais recorrentes de relação com a vítima são companheiros(as), ex-companheiros(as), cônjugues e filhos(as)/enteados(as). É possível perceber que a situação é alarmante, afinal há predominância de suspeitos com relações íntimas e/ou familiares com a vítima.

Visto que muitas mulheres não denunciam seu agressor (CARASCO, 2019), seja por medo de que o mesmo volte a agir de maneira violenta, por vergonha do que a sociedade irá pensar, medo de perder a guarda dos filhos ou até mesmo por descrença das políticas de combate à violência. Em aproximadamente 55,47% dos casos relatados pelo Ligue 180 (MULHER, 2020b), o demandante possui relação direta com a vítima.

2.1.1 Tipos de Violência

Estão previstos cinco tipos de violência doméstica e familiar contra a mulher na Lei Maria da Penha: física, psicológica, moral, sexual e patrimonial - Capítulo II, art. 7º, incisos I, II, III, IV e V.

2.1.1.1 Violência sexual

De acordo com a Organização Mundial da Saúde (SAÚDE, 2018) a violência sexual é “qualquer ato sexual ou tentativa de obter ato sexual, investidas ou comentários sexuais indesejáveis, tráfico ou qualquer outra forma contra a sexualidade de uma pessoa usando coerção” e pode ser praticada por qualquer pessoa, independentemente de relação com a vítima. Além disso, pode ser cometido em diversos cenários, como na rua, casa ou local de trabalho.

A violência sexual engloba subtipos como estupro, assédio sexual, exploração sexual, importunação sexual e estupro coletivo. As vítimas desse tipo de violação podem sofrer consequências comportamentais, sociais, e de saúde mental. A OMS destaca gravidez indesejada, depressão, ansiedade, aborto inseguro, comportamento suicida e transtorno de pânico, como exemplos de implicações de tal ato.

2.1.1.2 Violência Moral

É considerada qualquer conduta que configure calúnia, difamação ou injúria. Abusos como expor a vida íntima, desvalorizá-la pelo modo como se veste e emitir juízos morais sobre a conduta são exemplos de violência moral (PENHA, 2019b).

2.1.1.3 Violência Psicológica

A OMS define esse tipo de violência como “qualquer conduta que lhe cause dano emocional e diminuição da autoestima ou que lhe prejudique e perturbe o pleno desenvolvimento ou que vise degradar ou controlar suas ações, comportamentos, crenças e decisões, mediante ameaça, constrangimento, humilhação, manipulação, isolamento, vigilância constante, perseguição contumaz, insulto, chantagem, ridicularização, exploração e limitação do direito de ir e vir ou qualquer outro meio que lhe cause prejuízo à saúde psicológica e à autodeterminação”. É a forma mais subjetiva e difícil de identificar. Desse modo, essa violação, na maioria dos casos, é mascarada pelo ciúme, humilhação, controle e ofensas. A violência psicológica muitas vezes precede a agressão física por ser tolerada e praticada constantemente (MARTINELLI, 2014).

2.1.1.4 Violência Física

É aquela entendida como qualquer conduta que ofenda a integridade ou a saúde corporal da mulher. É praticada com uso de força física do agressor, como espancamento, tortura, estrangulamento ou sufocamento (PENHA, 2019b).

2.1.1.5 Violência Patrimonial

Segundo Penha (2019b), é “entendida como qualquer conduta que configure retenção, subtração, destruição parcial ou total de seus objetos, instrumentos de trabalho, documentos pessoais, bens, valores e direitos ou recursos econômicos, incluindo os destinados a satisfazer suas necessidades”. Pode implicar estelionato, privação de bens, valores ou recursos econômicos, furto, extorsão ou dano.

2.1.2 Cronologia da Violência

Segundo Walker (1987), atos de violência doméstica parecem ocorrer em ciclos que consistem em três fases.

2.1.2.1 Aumento da Tensão

Na primeira fase, ocorre o aumento de tensão do agressor, isso é visto em atos discretos, como xingamentos e abuso físico suave, como empurrar. É como ele expressa suas insatisfações, mas não ocorre de forma explosiva. A mulher então tenta acalmar o agressor, ela acredita que assim pode controlar a situação e não agravá-la ainda mais. No entanto, à medida que a tensão aumenta, o agressor chega a ter acessos de raiva e evita qualquer conduta que possa provocá-lo. Na maior parte das vezes a vítima tenta negar a situação, acha que fez algo errado para justificar o comportamento violento do agressor e tende a esconder o fato das pessoas.

2.1.2.2 Ato de Violência

Este é o momento em que ocorre a descarga de tensões, materializadas em violência verbal, física, psicológica, moral ou patrimonial, que foram acumuladas na fase anterior. Muitas vezes, aqui a vítima é ferida e, possivelmente, sofre uma tensão psicológica severa e sente um misto de sentimentos, como medo, dor, pena de si mesma, vergonha, solidão. É nesta fase que ela pode tomar algumas decisões, como denunciar, pedir ajuda e esconder-se na casa de familiares ou amigos. Geralmente, há um distanciamento com o agressor.

2.1.2.3 Arrependimento e Comportamento Carinhoso

Essa fase se caracteriza pela lamentação e remissão do agressor. Ele promete mudar, mostra remorso com presentes e se torna amável para que aconteça a reconciliação.

A mulher tem esperanças de que não aconteça mais nenhuma agressão, e pela pressão de manter seu relacionamento diante da sociedade, ela abre mão de seus direitos. O agressor parece encantador e manipulador, levando a vítima a acreditar que o problema já passou. Pelo sentimento de remorso, ela se sente responsável pelo agressor, e isso estreita a dependência entre os dois. Com o tempo, os intervalos entre uma fase e outra diminuem, e as agressões ficam frequentes, sem obedecer a ordem das fases. Em alguns casos, o ciclo termina com feminicídio.

2.1.3 Estatísticas

Em 2019, o Sistema Integrado de Atendimento ([MULHER, 2020a](#)) à Mulher registrou denúncias de ameaças (3.256), cárcere privado (2.511), feminicídio (61), tentativa de feminicídio (3.715), 74,65% a mais que em 2018, trabalho escravo (17), tráfico de mulheres (38), violência contra diversidade religiosa (15), violência doméstica e familiar (67.438), violência física (1.897), moral (3.482), obstétrica (164), policial (566), aumento de 471,72% em relação ao ano anterior, sexual (1.978) e virtual (274). Em 2019, o Ligue 180 registrou um total de 1,3 milhão de atendimentos. Desses, 6,3% foram denúncias. Foram 85.412 denúncias registradas e 1.314.113 ligações atendidas.

A violência doméstica e familiar engloba violência física, psicológica, moral, sexual, patrimonial e descumprimento de medidas protetivas. Ela teve o maior índice, com 78,96%, em relatório que demonstra as violações mais recorrentes registradas pelo Ligue 180 em 2019. Em seguida, tem-se tentativas de feminicídio, 4,35% e, posteriormente, violência moral, com 5,08% ([MULHER, 2020b](#)).

2.1.4 Casos de violência e feminicídio

Para este estudo, foram selecionados sete casos de violência contra a mulher e feminicídio desde 2008. Os acontecimentos escolhidos são brevemente descritos na Tabela 1. Os detalhes de cada caso podem ser encontrados no Anexo A.

2.1.5 Órgãos de Apoio

Desde que a Lei Maria da Penha foi sancionada em 2006, alguns mecanismos de assistência e proteção à mulher foram criados, tais como:

- Centros Especializado de Atendimento à Mulher, fazem acolhimento, acompanhamento psicológico e social e prestam orientação jurídica às mulheres que sofreram violência doméstica;

Caso	Descrição
Elaine Caparroz	A paisagista Elaine Peres Caparroz, de 55 anos, foi vítima de tentativa de feminicídio em fevereiro de 2019. A agressão ocorreu durante um encontro com o lutador Vinicius Batista Serra, de 28 anos, no apartamento da vítima.
Eliza Samúdio	Eliza Silva Samúdio foi assassinada em junho de 2010. O caso teve repercussão internacional, pois o goleiro Bruno Fernandes foi um dos colaboradores do crime.
Eloá Cristina	O caso se refere a um sequestro em cárcere privado com duração de mais de 100 horas, seguido de feminicídio que obteve grande repercussão nacional e internacional em 2008. O agressor, Lindemberg Fernandes Alves, era ex-namorado da vítima Eloá Cristina Pimentel.
Eva Luana	Eva Luana da Silva publicou em uma rede social sua história, contando as agressões que vinha sofrendo de seu padrasto Thiago Oliveira Alves, em fevereiro de 2019.
Mariana Ferrer	A influenciadora digital Mariana Ferrer acusa o empresário André Aranha de estupro em um <i>beach club</i> em Florianópolis. O fato ocorreu em dezembro de 2018.
Poliana Bagatini	Poliana Bagatini, grávida na época da agressão, foi jogada no chão e recebeu vários chutes de seu ex-marido Victor Chaves, conhecido por participar da dupla sertaneja Victor & Leo. O caso ocorreu em 2017.
Tatiane Spitzner	A advogada Tatiane Spitzner foi encontrada morta, em julho de 2018, após cair do 4º andar do prédio em que morava. O agressor foi Luís Felipe Manvailier, marido da vítima.

Tabela 1 – Descrição dos casos usados neste estudo.

- Casas-Abrigo também é um serviço que oferecem local protegido e atendimento integral a mulheres em situação de violência doméstica (acompanhadas ou não de filhos) sob risco de morte. Elas podem permanecer nos abrigos de 90 a 180 dias;
- Delegacias Especializadas de Atendimento à Mulher (DEAMs) são unidades especializadas da Polícia Civil, que realizam ações de prevenção, apuração, investigação e enquadramento legal. Nessas unidades, é possível registrar boletim de ocorrência e solicitar medidas de proteção de urgência (FEDERAL, 2019).

Juizados e varas especializados em violência doméstica e familiar contra mulher foram criados com a finalidade de poder contar com uma equipe de atendimento multidisciplinar a ser integrada por profissionais especializados nas áreas psicossocial, jurídica e da saúde. A Casa da Mulher Brasileira integra no mesmo espaço serviços especializados para os mais diversos tipos de violência contra as mulheres: acolhimento e triagem, apoio psicossocial, delegacia, Juizado, Ministério Público, Defensoria Pública, promoção de autonomia econômica, cuidado das crianças, alojamento de passagem e central de transportes (FEDERAL, 2019).

Além de aplicativos com a finalidade de apoiar as vítimas de violência como o Salve Maria (G1, 2019) e o Mete a Colher (MAGALHÃES et al., 2016), o Prêmio Maria da Penha vai à Escola (TERRITÓRIOS, 2019) busca incentivar profissionais da educação da rede pública do Distrito Federal a desenvolverem práticas relacionadas à prevenção e ao combate da violência contra a mulher. Em 2019, o projeto que foi contemplado com o 1º lugar, A violência doméstica e familiar contra a mulher e leis que promovem a proteção da mulher, desenvolvido, desde 2018, pela educadora Luana Nery Moraes, busca envolver toda a comunidade por meio de seminários abertos às famílias dos alunos.

2.2 Mineração de Texto

A Mineração de Texto é uma tecnologia que inclui conhecimentos de áreas como a Informática, Estatística, Linguística e Ciência Cognitiva. Inspirado pela Mineração de Dados, que procura descobrir padrões emergentes de um banco de dados estruturado, a Mineração de Texto busca extrair conhecimentos úteis de dados não estruturados e semi-estruturados. É um conjunto de métodos usados para navegar, organizar e descobrir informações em bases textuais. Diferente de *chatbot*, não pretende simular comportamento humano (ARANHA; PASSOS, 2006). O processo de minerar texto surgiu da necessidade de se descobrir, de forma automática, padrões ou anomalias, em textos.

Extração de Regras é um exemplo de aplicação da Mineração de Textos. Ela é aplicada a partir de um conjunto de documentos. O sistema compara a frequência das palavras de cada documento à frequência de palavras em um conjunto de treinamento. Alguns softwares como o *Cortex Competitiva*, *Text Mining* aplicada à Inteligência competitiva, e *Text Analyst*, que gera rede semântica do texto baseada em algum algoritmo de Hopfield, contribuem para a área de Mineração de Texto (ARANHA; PASSOS, 2006).

Também é importante destacar que, com o avanço da tecnologia e a popularização da *Internet* e dos dispositivos móveis, as redes sociais se tornaram fonte muito comum e rica de textos não estruturados. São exemplos de aplicações que usam dados de redes sociais: detecção de *bullying* (SILVA; SILVA; DIAS, 2018), política (JUNIOR, 2016) e copa do mundo (LIMIRO; CORDEIRO; CASSIANO, 2017). Desse modo, é evidente a relevância de estudos com ênfase em Mineração de Textos aplicada a redes sociais.

A Análise de Sentimentos é uma área de estudo que analisa as opiniões, sentimentos, avaliações e atitudes das pessoas com relação a diferentes entidades (LIU, 2012). Essa área também tem sido largamente relacionada a redes sociais. Dentre os inúmeros trabalhos que fazem Análise de Sentimento de dados em redes sociais, podemos citar como por exemplo (SILVA et al., 2014), (CORRÊA, 2017) e (LIMIRO; CORDEIRO; CASSIANO, 2017). Silva et al. (2014) usa a rede social Facebook¹ para explorar a dimensão emocional

¹ <http://www.facebook.com>

de um indivíduo a partir dos conteúdos postados. Outros trabalhos utilizam a rede social *Twitter*², como [Corrêa \(2017\)](#), que visa realizar a análise dos sentimentos expressos pelos usuários em relação aos filmes indicados à categoria de Melhor Filme do *Oscar* 2017, e [Limiro, Cordeiro e Cassiano \(2017\)](#), que aplica a análise de sentimentos no contexto Copa do Mundo de Futebol.

2.3 Etapas da Mineração de Texto

A mineração de texto é uma tarefa complexa, composta por um conjunto de etapas, as quais são descritas a seguir.

2.3.1 Coleta de Dados

Primeiramente, é preciso definir a fonte de dados que será utilizada. As fontes fornecedoras dos dados podem vir de diversos locais e possuem diversos formatos. Um dos principais ambientes propícios para uma coleta de dados eficaz é a *Web*. É possível encontrar muitas informações em formato de textos em redes sociais, como o *Twitter* e *Facebook*, em jornais eletrônicos, fóruns, blogs, entre outros. Ambientes fora da *Internet* também são fontes ricas para construir uma base de qualidade, como bancos de dados de ONGs e de órgãos governamentais.

No caso de redes sociais, deve-se escolher a rede social e a forma como os dados serão coletados. O *Facebook* fornece uma API³ para auxiliar nesta etapa, assim como o *Twitter*⁴. No entanto, nestas APIs existem algumas limitações, como a restrição de tempo e quantidade de informação. No caso do *Twitter*, por exemplo, a API limita a coleta a *tweets* publicados há, no máximo, sete dias em relação à data em que a consulta for realizada.

Para lidar com essas restrições, muitos pesquisadores preferem criar seus próprios robôs para coleta de dados ou utilizar outros disponíveis online, como o *GetOldTweets*⁵, que supera algumas limitações da API oficial do *Twitter*, como por exemplo, não restringir o intervalo de tempo da busca por *tweets*. Ao realizar a recuperação de *tweets* a partir de uma palavra-chave, para cada instância retornada, além do texto do *tweet*, também são retornados alguns metadados do mesmo, como data, autor, número de *retweets*, número de *likes*, entre outros.

² <http://www.twitter.com>

³ <https://developers.facebook.com/docs/graph-api>

⁴ <https://developer.twitter.com/en/docs>

⁵ <https://github.com/Jefferson-Henrique/GetOldTweets-python>

2.3.2 Pré-processamento dos dados

A etapa de Pré-Processamento é de suma importância para a Mineração de Textos. Ela tem como finalidade descartar o que é considerado irrelevante à etapa de classificação e/ou alterar o formato dos dados de forma a auxiliar nessa etapa. O pré-processamento é composto por várias tarefas, que variam conforme as particularidades dos dados que estão sendo processados, isto é, cada fonte de dados exige o uso de diferentes tarefas para que um resultado satisfatório seja obtido.

Este é um estágio de preparação dos dados. É importante ressaltar que textos são dados não estruturados, então é necessário estruturá-los. Nessa fase, extraem-se entidades e informações importantes, sendo feita a sumarização, seleção do conteúdo relevante de uma mensagem e organização coerente desse material. O objetivo desta etapa é a eliminação de dados irrelevantes e a padronização de termos utilizando algoritmos específicos, de modo que as etapas seguintes tenham uma melhor desempenho (ARANHA; PASSOS, 2006).

Nessa etapa ocorre a análise léxica, que converte uma sequência de caracteres numa sequência de palavras, que serão palavras candidatas a serem termos de índice (BARION; LAGO, 2006). A seguir ocorre a remoção de *stopwords*, que remove um conjunto de palavras que aparecem com muita frequência no texto. Já o *stemming*, remove todas as variações de palavras permanecendo somente a raiz. É importante que essas palavras sejam eliminadas para não sobrecarregarem o processo e/ou para não interferirem na análise das informações (BARION; LAGO, 2006).

Algumas tarefas relacionadas ao pré-processamento de dados obtidos a partir do *Twitter* são apresentadas a seguir.

- Remoção de *links*, pois esses termos não possuem conteúdo semântico.
- Remoção de caracteres não alfabéticos e pontuação, já que não agregam valor à classificação.
- Substituição de caracteres acentuados pelos correspondentes sem acentuação, com a finalidade de padronizar o texto.
- Remoção de repetição de letras: para dar mais ênfase em seu sentimento, muitos usuários usam a repetição de letras, por exemplo, se a pessoa está muito indignada com alguma situação, ela pode escrever “*estou muitoooooooo indignada*”.
- Remoção de citação a outros usuários: no *Twitter*, usa-se o símbolo “@” para mencionar outros usuários da rede social.
- Conversão de letras maiúsculas em minúsculas, com o objetivo de padronizar o texto.

- Substituição de gírias por expressões completas. Desse modo, um maior número de palavras e expressões pode ser reconhecido pelo classificador.
- Remoção de *stopwords*, palavras que são muito comuns em um idioma, e portanto, não possuem valor semântico.
- Lematização, que consiste no processo de deflexionar uma palavra para determinar o seu lema. Assim, palavras como “esperando”, “esperei” e “esperou” seriam todas reduzidas a “espera” após a lematização. Isso resultaria em um maior peso para esse termo, mas também poderia anular o sentido original da palavra.
- Remoção de termos de busca, já que são constantes em todos os resultados.

Essas tarefas podem ser realizadas a partir da criação de algoritmos próprios ou utilizando *softwares*, como o *NLTK* (LOPER; BIRD, 2002) e *Weka* (SMITH; FRANK, 2016).

Ao final deste processo, tem-se uma base de dados mais consistente que pode ser usada nas etapas seguintes.

2.3.3 Busca por padrões

Para ajudar na visualização e validação dos dados, e com o objetivo de estabelecer correlações, diferentes estratégias são usadas, como por exemplo, as descritas a seguir.

- **Nuvem de palavras:** Representa graficamente a frequência de ocorrência de palavras contidas em certa base de dados, a partir de uma nuvem composta por esses termos. Quanto maior for a ocorrência do termo, maior será seu tamanho na nuvem⁶.
- ***Latent Dirichlet Allocation* (LDA):** é um modelo altamente estatístico, ele se baseia em acreditar que cada tópico é uma mistura de um conjunto de palavras e que cada documento é uma mistura de um conjunto de tópicos. São enviados para o modelo uma quantidade M de documentos, que em si geram N palavras, que são tratadas no LDA gerando K tópicos, que podem ser entendidos como grupos de palavras. Também é extraída a frequência de palavras por tópico e o quanto percentualmente aquele tópico é relevante para o documento (CASTRO, 2020).

2.4 Classificação dos textos

É uma etapa em que os algoritmos de classificação são utilizados para classificar um texto, ou seja, atribuir um sentimento como “positivo”, “neutro” ou “negativo”. Abor-

⁶ <https://worditout.com/word-cloud/create>

dagens como aprendizado supervisionado e não-supervisionado têm sido usadas na tarefa de classificação (FELIX, 2016).

2.4.1 Aprendizado supervisionado

Segundo Baeza-Yates e Ribeiro-Neto (2013), um algoritmo de aprendizado é considerado supervisionado quando um conjunto de treinamento é usado para treinar o classificador. Nesse tipo de aprendizagem pode-se estimar números reais ou valores dentro de um conjunto finito.

2.4.2 Aprendizado não supervisionado

De acordo com Filho (2017), no aprendizado não supervisionado, o dados não são pré-definidos. Diferentemente do algoritmo de classificação supervisionado, não se sabe quantos e quais classes existem.

2.5 Naive Bayes

O *Naive Bayes* é um algoritmo probabilístico de classificação (aprendizado supervisionado) que produz estimativas de probabilidade para as classes. Ele usa de aprendizagem supervisionada, pois o *dataset* que será utilizado tem um agente supervisor, ou seja, alguma pessoa que analisou os dados e os classificou de acordo com suas características, manualmente. Em Mineração de texto, esse algoritmo tem como objetivo calcular a probabilidade de um dado texto pertencer a cada uma das classes do problema e escolher a melhor classe em que o mesmo se encaixa. Ele é chamado de classificador ingênuo, pois assume que os atributos são independentes uns dos outros, dado um valor da classe (BAEZA-YATES; RIBEIRO-NETO, 2013).

Dentro dos modelos de classificação bayesiana, dois são mais conhecidos: o modelo binário e o modelo multinomial. No modelo binário, também chamado de básico, cada atributo pode indicar ou não a ocorrência de um evento (palavra) no documento. O modelo multinomial trabalha com o número de vezes que cada evento ocorre no documento.

2.6 Trabalhos Relacionados

Até o momento deste estudo, foi encontrado somente um trabalho de Mineração de Texto com foco na violência contra mulher, o de Rodrigues, Junior e Lobato (2019). Esse trabalho analisou comentários de notícias sobre a tentativa de feminicídio de Elaine Caparroz, fato que ocorreu no dia 17 de fevereiro de 2019 no Rio de Janeiro. Foram coletados 1.120 comentários de 31 notícias publicadas e 754 foram selecionados para análises.

Na análise de sentimentos via *Polyglot*⁷, foi possível observar que há uma predominância de comentários negativos e neutros sobre os positivos. Assim, foi feita a correlação entre análise de sentimento e a quantidade de *likes* e *dislikes*. Após essa etapa, realizou-se a análise dos cenários em relação ao método *Latent Semantic Analysis* (LSA) ou *Latent Dirichlet Allocation* (LDA). A partir do método LSA, foram definidos cinco tópicos: relacionamento, preferências, internet/social, violência doméstica e delito, e foram determinadas dez palavras por tópico. Mediante a uma análise de legibilidade, grande parte dos autores pertencem à classe de quinta a oitava série, os quais apresentam mais sentimentos negativos. Ao final do estudo pôde-se identificar uma tendência em responsabilizar Elaine Caparroz pela tentativa de feminicídio por ela sofrida.

Também existem diversas pesquisas que tratam de violência em geral usando a técnica de Mineração de Textos. Em [Basilio e Pereira \(2018\)](#), que teve como objetivo desenvolver um modelo de ordenação das estratégias de policiamento em função dos delitos recorrentes em uma determinada localidade, foi considerada uma base de dados com registros de chamadas de emergência realizadas pelo serviço 190 da Polícia Militar do Rio de Janeiro, no período de 2013 ao final de 2016. O método desenvolvido na pesquisa consistiu inicialmente na extração de relatos de atendimento das ocorrências policiais do banco de dados do serviço de emergência, que após o processo de mineração de texto, permitiu a utilização do *Latent Dirichlet Allocation* (LDA) com *Gibbs Sampling Colapsado* culminando com a extração dos topic model das áreas de policiamento estudadas. Foram identificados dez tópicos de cada área pesquisa. Com a rotulação dos tópicos, identificaram-se os tipos latentes de demandas do serviço de emergência em cada uma das áreas estudadas. Sendo assim, este resultado corrobora com o argumento de que o crime não ocorre de forma linear, necessitando, com isso, de estratégias diferenciadas para o seu combate.

Foram encontrados trabalhos que utilizaram a Mineração de Textos em redes sociais, como por exemplo o estudo de [Silva, Silva e Dias \(2018\)](#) criou um corpus, construído a partir do *Twitter*, contendo textos que possuam traços de *bullying* para o português brasileiro, bem como a identificação do papel do escritor, neste caso foram considerados o *Bully* (quem pratica a ação), a vítima (quem sofre a violência) e o relator (quem relata o episódio, pode ou não estar presente). Para analisar e classificar o papel dos envolvidos, foram usados os classificadores *Naive Bayes*, Regressão Logística (*Logistic Regression*) e o SVM (*Support Vector Machine*). Utilizando unigramas e bigramas, foi possível ter uma acurácia de 79,54%. Já fazendo o uso do SVM com kernel RBF, foi obtido 79,45% de acurácia, precisão de 60%, revocação de 58% e *F-score* de 56%. Com o kernel linear, a acurácia foi de 78,92% e com o classificador *Naive Bayes*, 70,91%.

⁷ <https://polyglot.readthedocs.io/en/latest/>

2.7 Considerações finais

Atualmente, o número de registros de violência contra mulher têm aumentado muito. Os sites de notícias e as redes sociais têm sido ambientes propícios para discussão sobre o caso e mensagens de apoio à vítima ou ao agressor. Diante da quantidade de textos disponíveis, a Mineração de Texto torna-se imprescindível para extrair conhecimentos e padrões de grandes volumes de textos em linguagem natural.

Apesar de ser uma área inovadora, que resulta em informações significativas, devido ao grande número de relatos de violência, há uma ausência de trabalhos que aplicam a Mineração de Texto ou Análise de Sentimentos em casos de violência, principalmente de agressões contra mulheres. Neste estudo serão analisados comentários de usuários da rede social *Twitter* sobre casos de violência contra a mulher e feminicídio e, a partir do uso de técnicas de Análise de Sentimentos, espera-se encontrar o sentimento dos usuários em relação aos componentes do caso. Para isso, a tarefa de Mineração de Texto será dividida em cinco etapas: coleta de dados, pré-processamento dos dados, construção da base de dados rotulada, classificação dos textos e, por fim, validação dos resultados.

3 Método para análise dos dados do *Twitter* sobre os casos de violência

Este capítulo expõe o método utilizado para o desenvolvimento deste estudo, que tem como objetivo analisar *tweets* sobre casos de violência contra mulher que tiveram muita repercussão na mídia. A Seção 3.1 apresenta como foi realizada a coleta dos dados do *Twitter* para a construção da base de dados utilizada para o estudo. A Seção 3.2 lista as etapas de pré-processamento realizadas neste trabalho, que foram executadas com o objetivo de tratar os dados, descartando o que é irrelevante para a etapa de classificação. A Seção 3.3 expõe as estratégias utilizadas para auxiliar na validação e análise dos dados. A Seção 3.4 descreve como foi realizado o processo de rotular *tweets* com um dos sentimentos (“positivo”, “negativo” ou “neutro”), com o objetivo de construir a base rotulada para ser utilizada como conjunto de treinamento para algoritmos de aprendizado supervisionado.

3.1 Coleta de Dados

Para começar o processo de Análise de Sentimentos, é necessário realizar a coleta, que consiste em compor uma base de textos a ser trabalhada. É uma etapa a qual necessita muito esforço, afinal é preciso coletar material de qualidade e que sirva de matéria-prima para aquisição de conhecimento (JUNIOR, 2007).

A coleta foi realizada por meio da rede social *Twitter* com as palavras-chave apresentadas na Tabela 2. O *Twitter* foi escolhido por ser uma fonte que gera grandes volumes de dados a todo momento.

A ferramenta `GetOldTweets`¹ foi utilizado para este processo, pois supera algumas limitações da API oficial do *Twitter*, como por exemplo, não restringir o intervalo de tempo da busca. O código utilizado retorna os *tweets* a partir de palavras-chave. Para cada instância retornada, além do texto do *tweet*, também são retornados a data, o autor, número de *retweets* e número de *likes*. Contudo, não foi levado em conta estes números, como também os autores dos textos. É importante destacar que os autores dos relatos não foram considerados neste trabalho, sendo que a identificação do autor foi removida. A busca foi realizada na interface de linha de comando, com os argumentos presentes na Figura 1.

Os argumentos especificados nas consultas realizadas para este trabalho são descritos a seguir.

¹ <https://github.com/Jefferson-Henrique/GetOldTweets-python>

Assunto	Período de Coleta	Palavra-Chave
Caso Elaine Caparroz	16/02/2019 a 06/06/2020	“Elaine Caparroz” “Elaine 4 horas” “Paisagista agredida” “Elaine tentativa feminicidio”
Caso Eliza Samúdio	13/10/2009 a 06/06/2020	“Eliza Samudio” “Eliza goleiro bruno” “Eliza esquartejada” “Eliza bruno cachorros”
Caso Eloá Cristina	13/10/2008 a 06/06/2020	“Eloa Lindemberg” “Eloa sequestro” “Eloa morte” “Eloa carcere”
Caso Eva Luana	19/08/2019 a 06/06/2020	“Eva Luana” “Eva Luana violencia” “Eva Luana padrasto” “SomosTodosEva”
Caso Mariana Ferrer	20/05/2019 a 06/06/2020	“Mariana Ferrer” “Blogueira Cafe de la Musique” “Blogueira estuprada” “maribferrer estupro”
Caso Poliana Bagatini	24/02/2017 a 06/06/2020	“Poliana Bagatini” “Victor chaves esposa” “Victor chaves mulher gravida” “Victor agredir mulher”
Caso Tatiane Spitzner	22/07/2018 a 06/06/2020	“Tatiane Spitzner” “Tatiane jogada predio” “Spitzner feminicidio” “Spitzner morte”

Tabela 2 – Palavras-chave na coleta pelo *Twitter* separados por assunto

```
Python Exporter.py -querysearch "eliza samudio lang:pt-br" -since 2008-10-13 -until 2020-06-06
```

Figura 1 – Consulta ao *Twitter* realizada por meio da interface de linha de comando do *GetOldTweets*

- **querysearch**: palavras-chave a serem consideradas na busca por *tweets*, seguidas do parâmetro *lang:pt-br*, que refina a consulta para *tweets* escritos apenas na língua portuguesa.
- **since**: indica a data inicial a ser considerada na consulta. A data “2018-01-01” foi utilizada como parâmetro.
- **until**: indica a data final a ser considerada na consulta. A data “2020-06-06” foi utilizada como parâmetro.

A execução do programa foi realizada vinte e oito vezes, cada uma com as palavras-

chave expostas na Tabela 2. Após cada execução do programa, um arquivo era gerado, incluindo os *tweets* encontrados e seus respectivos metadados. Os *tweets* foram agrupados de acordo com os casos, sendo assim, os arquivos de cada caso contêm dados de todas as palavras-chave relacionadas a ele.

É importante destacar que o período de início da coleta varia de acordo com a data do caso em questão e o período final não difere. Além disso, não foram selecionados fatos que não aconteceram no Brasil.

3.2 Pré-Processamento

Neste trabalho, o processo utilizado se deu pelo uso de um algoritmo na linguagem *Java*, o mesmo utilizado em Corrêa (2017). É importante destacar que algumas adaptações foram feitas, afinal este algoritmo foi criado para pré-processamento na língua inglesa. O algoritmo realiza todas as etapas:

1. **Remoção de *stopwords*:** a lista do programa original foi criada com *stopwords* na língua inglesa, como foi citado anteriormente, desse modo não contribuiria para este trabalho, já que toda a base é em língua portuguesa. Não existe biblioteca com entradas suficientes para ter uma lista de *stopwords* completa em português. Desse modo, foi construída pela autora deste estudo uma lista manualmente².
2. **Remoção de caracteres não alfabéticos, pontuação e acentuação:** estes não agregam valor à classificação.
3. **Conversão de letras maiúsculas para minúsculas:** com o objetivo de padronizar o texto.
4. **Substituição de gírias e abreviações:** Um dicionário contendo 324 gírias juntamente com abreviações e suas respectivas traduções foi construído³ para incorporar novos termos aos relatos e garantir que a semântica do texto fosse preservada. O dicionário foi construído a partir de experiência da autora e de listas de gírias disponíveis na *Internet*⁴.
5. **Junção de palavras compostas:** termos compostos são colocados como uma palavra só, para ficar mais claro a referência das palavras.

Além das fases descritas acima, foi feita também a remoção de *tweets* que contêm *links* de reportagens, afinal estas não expressam sentimentos, apenas contêm informações sobre os casos.

² A lista de *stopwords* utilizada neste estudo está disponível em <http://bit.ly/ListStopwords>

³ O dicionário utilizada neste estudo está disponível em <http://bit.ly/ListDictionary>

⁴ <http://bit.ly/DicionarioInternetes>

Após a realização destas etapas, foi realizada a lematização por um programa também em *Java*, disponibilizado pelo Núcleo Interinstitucional de Linguística Computacional⁵. É importante destacar que o pré-processamento foi validado pelas nuvens de palavras que são descritas na próxima seção, desse modo, ele teve que ser refeito algumas vezes por serem encontradas falhas e melhorias a serem feitas.

3.3 Busca por padrões

Com o objetivo de auxiliar a entender sobre quais temas apareciam nos *tweets*, este trabalho usou a técnica de nuvens de palavras e o modelo estatístico para modelagem de tópicos *Latent Dirichlet Allocation* (LDA). As nuvens de palavras também foram usadas para ajudar na validação do pré-processamento.

3.3.1 Nuvem de palavras

As nuvens de palavras são imagens compostas por palavras, que demonstram de maneira visual a frequência de ocorrência das palavras em um texto, ou seja, quanto maior for a palavra na nuvem, maior foi a ocorrência da mesma no relato. Elas foram também usadas neste trabalho para validar a eficácia do pré-processamento.

Um exemplo de uso da nuvem de palavras é ilustrado na Figura 2. Nessa figura é possível visualizar que houve falha no pré-processamento, afinal aparecem palavras que deveriam ter sido eliminadas pelas etapas citadas na seção 3.2. Na nuvem são mostrados termos com acentuação, como por exemplo “mês”, o qual deveria ter sido removido no pré-processamento. Também são apresentadas palavras como “dias” e “dia”, que não foram lematizados corretamente. Depois de analisada a nuvem, o pré-processamento foi refeito.

3.3.2 *Latent Dirichlet Allocation* (LDA)

O *Latent Dirichlet Allocation* (LDA) é um dos métodos de modelagem de tópicos mais populares. Cada documento é composto de várias palavras e cada tópico também possui várias palavras pertencentes a ele. O objetivo do LDA é encontrar tópicos aos quais um documento pertence, com base nas palavras nele contidas. É um modelo de probabilidade generativo, o que significa que ele tenta fornecer um modelo para a distribuição de saídas e entradas com base em variáveis latentes.

Para a construção da matriz de pesos dos termos, foi utilizado o *Term Frequency-Inverse Document Frequency*, ou TF-IDF (NETO et al., 2000). Este método é uma medida estatística que tem o intuito de indicar a importância de uma palavra de um documento em relação a uma coleção deles ou em um corpus linguístico. Enquanto o IDF busca indicar

⁵ <http://www.nilc.icmc.usp.br/nilc/index.php>



Figura 2 – Nuvem de palavras contidas na base de dados do Caso Eloá – Gerada a partir do WordItOut

a importância da palavra de forma a ajustar esse item do TF, o TF mede quão frequente um termo aparece dentro de um documento. Para gerar um resultado normalizado, este valor é dividido pelo tamanho do mesmo. Para isso é feito o cálculo do logaritmo do total de documentos pelo número dos que contém o termo neles. Cada tópico resultante do LDA é definido a partir de uma *Bag of Words*, um modelo de representação de informações textuais que define esses dados como uma matriz, o qual assimila pesos para cada termo ou palavra encontrada em todo o conjunto (RODRIGUES; JUNIOR; LOBATO, 2019). Contudo, os algoritmos de modelagem não definem categorias ou rotulam os tópicos. Esta tarefa ficou a cargo de uma análise feita pela autora.

Foi utilizada uma implementação do LDA na linguagem *Python*⁶, ilustrada na Figura 3.

Os parâmetros empregados são descritos a seguir:

- **corpus**: Fluxo de vetores de documentos. O corpus neste estudo recebe todos os *tweets* pré-processados.
- **id2word**: Mapeando de IDs de palavras para palavras. É usado para determinar o tamanho do vocabulário, bem como para depuração e impressão de tópicos.
- **num_topics**: O número de tópicos latentes solicitados a serem extraídos do corpus de treinamento. Neste caso, optou-se por 5 tópicos para obter mais facilidade na

⁶ <https://bit.ly/3f9BzuI>

```
lda_model = gensim.models.ldamodel.LdaModel(corpus=corpus,
                                             id2word=id2word,
                                             num_topics=5,
                                             random_state=100,
                                             update_every=1,
                                             chunksize=100,
                                             passes=4,
                                             alpha='auto',
                                             per_word_topics=True)

pprint(lda_model.print_topics(num_words=6))
doc_lda = lda_model[corpus]
```

Figura 3 – Trecho da implementação do LDA na linguagem *Python*

etapa da rotulação dos mesmos.

- ***random_state***: Controla o embaralhamento aplicado aos dados antes de aplicar a divisão. Passe um inteiro para saída reproduzível em várias chamadas de função.
- ***update_every***: Número de documentos a serem iterados para cada atualização. Defina como 0 para aprendizado em lote, maior que 1 para aprendizado iterativo online.
- ***chunksize***: Número de documentos a serem usados em cada bloco de treinamento.
- ***passes***: Número de passes pelo corpus durante o treinamento.
- ***alpha***: Pode ser definido como uma matriz 1D de comprimento igual ao número de tópicos esperados que expressa nossa crença a priori para a probabilidade de cada tópico.
- ***per_word_topics***: Se Verdadeiro, o modelo também calcula uma lista de tópicos, classificados em ordem decrescente dos tópicos mais prováveis para cada palavra, juntamente com seus valores π multiplicados pelo comprimento do recurso, ou seja, contagem de palavras.
- ***num_words***: O número de palavras a serem incluídas por tópicos (ordenadas por significância). No estudo foram escolhidas 6 palavras a serem contidas em cada tópico.

O programa foi executado duas vezes para cada conjunto de *tweets*: Caso Elaine Caparroz, Caso Eliza Samúdio, Caso Eloá Cristina, Caso Eva Luana, Caso Mariana Ferrer, Caso Poliana Bagatini e Caso Tatiane Spitzner. Na primeira execução, optou-se por 5 tópicos com 6 palavras cada, já na segunda execução, 10 tópicos com 6 palavras cada. Esses cenários foram executados com o objetivo de encontrar tópicos mais fáceis de serem

rotulados. Uma vez que os tópicos foram criados, foi realizada uma análise a fim de rotulá-los. Experimentos com mais tópicos e mais palavras foram feitos. No entanto, houve uma grande dificuldade em rotular tais tópicos.

3.4 Classificação dos textos

Para a Classificação dos Textos, o algoritmo de classificação *Naive Bayes* (BAEZA-YATES; RIBEIRO-NETO, 2013) foi usado. Sua escolha se deve ao fato de que ele é um classificador que apresenta bons resultados, em especial em texto, sendo muito usado na literatura, como em (CORRÊA, 2017), (SILVA; SILVA; DIAS, 2018) e (ANDRADE, 2015). O objetivo da classificação neste trabalho é categorizar os *tweets* sobre os casos de violência em uma dentre três classes: positivo, negativo e neutro, considerando a vítima, agressor e o caso.

3.4.1 Construção da Base de Dados Rotulada

Para a execução do algoritmo, é importante ressaltar que, é necessário construir uma base rotulada para treinar o classificador. Esta etapa foi desempenhada pela autora deste estudo, a qual exigiu bastante cuidado e tempo.

Uma amostra de 100 *tweets* de cada caso foi selecionada de forma aleatória a partir da base de dados coletada originalmente. No primeiro momento, não foi possível rotular cada *tweet* de acordo com o sentimento, afinal o usuário pode ter um sentimento sobre a vítima, agressor ou sobre o caso em si. Desse modo, cada texto foi rotulado de acordo com a classe que estava se referindo. A Tabela 3 expõe os dados contidos na base de dados rotulada, que representa 1,015% da base total de *tweets*.

	Caso	Agressor	Vítima	Total
Caso Elaine Caparroz	18	18	64	100
Caso Eliza Samúdio	18	47	35	100
Caso Eloá Cristina	52	28	20	100
Caso Eva Luana	40	06	54	100
Caso Mariana Ferrer	48	09	43	100
Caso Poliana Bagatini	27	65	08	100
Caso Tatiane Spitzner	54	19	27	100
Total	257	182	251	700

Tabela 3 – Sumarização dos dados contidos na base de *tweets* rotulados

Posteriormente, cada *tweet* da base rotulada foi julgado "positivo", "neutro" ou "negativo". Desse modo, é possível entender a quem o sentimento foi designado, como é mostrado na Tabela 4.

	Positivo	Neutro	Negativo
Caso Elaine Caparroz			
Vítima	36	22	6
Agressor	0	5	13
Caso	0	13	5
Caso Eliza Samúdio			
Vítima	4	8	23
Agressor	1	28	18
Caso	0	17	1
Caso Eloá Cristina			
Vítima	11	4	5
Agressor	4	16	8
Caso	0	27	25
Caso Eva Luana			
Vítima	48	5	1
Agressor	0	0	6
Caso	0	6	34
Caso Mariana Ferrer			
Vítima	34	6	3
Agressor	0	1	08
Caso	0	18	30
Caso Poliana Bagatini			
Vítima	1	1	6
Agressor	3	24	38
Caso	0	19	8
Caso Tatiane Spitzner			
Vítima	23	2	2
Agressor	0	2	17
Caso	0	20	34
Total			
	165	244	291

Tabela 4 – Sumarização dos dados contidos nas novas bases de *tweets* rotulados

3.4.2 Classificador *Naive Bayes*

Para aplicar o *Naive Bayes*, foi utilizado a ferramenta *Weka* (WITTEN et al., 2016), um software escrito na linguagem *Java* que fornece uma coleção de algoritmos de aprendizagem de máquina para realização de tarefas relacionadas à mineração de dados. Os dados das bases rotuladas de textos foram convertidos para o formato ARFF (*Attribute Relation File Format*), um arquivo de texto específico para o *Weka*, que descreve uma lista de instâncias que compartilham os mesmos atributos. A Figura 4 mostra um exemplo de um trecho do arquivo ARFF. Os dois atributos da relação *Vitima* são descritos adiante.

- *Vitima*: uma string contendo o texto do tweet
- *sentimento*: o rótulo assinalado manualmente para o *tweet* (positivo, neutro ou ne-

gativo).

```
@relation Vitima
@attribute tweets string
@attribute sentimento {Positivo,Neutro,Negativo}
@data
'espancar viniciusbatista importa velho conhecer internet cerca mulher acionar lei maria penha dia
contar conseguir denunciar precisar acabar violencia',Positivo
'espancar estudante direito lutadorviniciusbatistaprestar depoimento primeiro delegar responsavel
investigacoes aguardar depoimento vitima conclusao inquerito estudioi',Neutro
'conhecer internet comentario espancar apartamento aprontar cara viver',Negativo
'misoginia doenca mental aprender cultivado disseminar estereotipos genero vitima link',Positivo
'vinicius conversar mes torturar violentar matar confiavel conseguir intuito primitivo portaria
nome falso premeditar culpar vitima',Positivo
```

Figura 4 – Trecho do arquivo ARFF utilizado para alimentar o *Weka*

Após o carregamento do conjunto de treinamento na interface *Explorer* do *Weka*, foi aplicado o filtro *StringToWordVector*, que converte a string representando o *tweet* em um conjunto de atributos numéricos, que representam informações de ocorrência de palavras do texto contido nas sequências.

Na Figura 5 é mostrada uma visão geral dos dados na aba de pré-processamento, após a aplicação do filtro. Na aba de Classificação, foi escolhido o classificador *Naive Bayes* com a estratégia de validação *10-fold cross-validation*, que consiste na divisão do conjunto de dados em 10 partes aleatórias, que são testadas individualmente, utilizando as outras nove partes restantes como conjunto de treinamento.

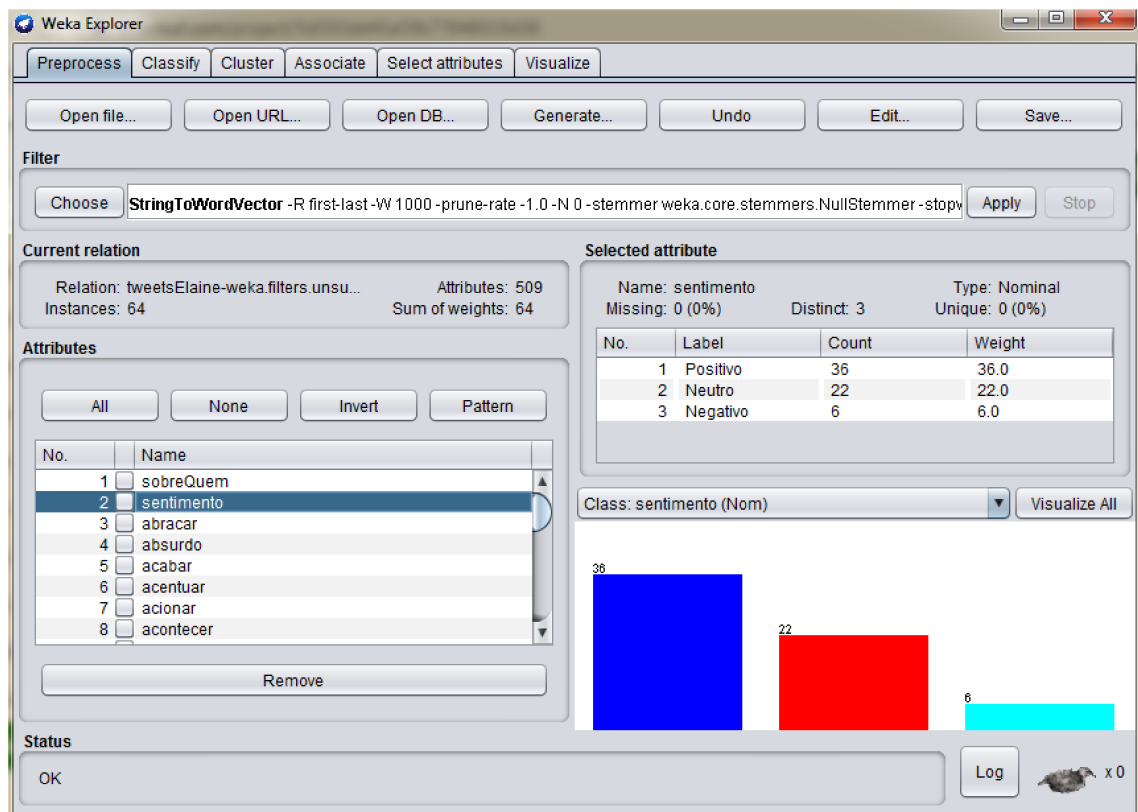


Figura 5 – Visão do conjunto de treinamento carregado no *Weka*

3.4.3 Avaliação dos Classificadores

Após realizar os testes com o classificador utilizando a base de dados rotulada como conjunto de treinamento, é possível obter as seguintes medidas, considerando um cenário em que o classificador é treinado usando as classes “positivo”, “negativo” e “neutro”.

- **VP** (verdadeiros positivos): instâncias positivas corretamente classificadas.
- **VN** (verdadeiros negativos): instâncias negativas corretamente classificadas.
- **VE** (verdadeiros neutros): instâncias neutras corretamente classificadas.
- **FP** (falsos positivos): instâncias negativas ou neutras classificadas como positivas.
- **FN** (falsos negativos): instâncias positivas ou neutras classificadas como negativas.
- **FE** (falsos neutros): instâncias positivas ou negativas classificadas como neutras.

Para avaliar o resultado gerado pelo classificador, torna-se necessária a aplicação de métricas de avaliação. Em mineração de textos e aprendizagem de máquina, a validação dos resultados é feita utilizando métricas de desempenho, como acurácia (Equação 3.1), precisão, Revocação, Medida-F, dentre outras (ANDRADE, 2015). A acurácia mede os acertos realizados pelo classificador. A precisão mede, dentre todas as classificações de classe Positivo que o modelo fez, quantas estão corretas. Já Revocação mede, dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas. Medida-F é uma média harmônica ponderada entre o Revocação e a precisão. Por questões de simplicidade e por não ser o foco deste trabalho, somente a acurácia foi utilizada para realizar a validação do classificador.

$$acurácia = \frac{VP + VN + VE}{VP + VN + VE + FP + FN + FE} \quad (3.1)$$

Todas essas medidas podem ser calculadas a partir da matriz de confusão. Essa matriz mostra os valores classificados, verificando, para cada valor previsto, se o mesmo corresponde ao valor real. Ao indicar uma classe, os resultados serão validados de acordo com os índices de acerto para cada categoria e em relação ao conjunto todo. Com as medidas citadas na subseção 3.4.2 é possível criar uma matriz de confusão. As linhas da matriz representam as instâncias reais das classes, enquanto as colunas indicam como as instâncias foram classificadas pelo algoritmo, como mostrado na Tabela 5.

Neste trabalho, optou-se por testar e utilizar somente o classificador *Naive Bayes*. Trabalhos futuros incluem o estudo de outros classificadores em busca de melhores resultados.

Classificado como			
pos	neu	neg	
VP	FE	FN	pos
FP	VE	FN	neu
FP	FE	VN	neg

Tabela 5 – Matriz de confusão para o algoritmo Naive Bayes

4 Resultados

Este capítulo tem como objetivo analisar os dados obtidos pelo pré-processamento dos *tweets* e apresentar resultados da classificação dos *tweets*.

A Seção 4.1 sumariza a quantidade de *tweets* coletados e tenta estabelecer relações entre esses resultados e os casos de violência contra mulher escolhidos. Observações são realizadas a partir da análise de gráficos que resumem esses dados.

A Seção 4.2 tem como objetivo expor os resultados do pré-processamento, e expõe a uma quantificação dos *tweets* antes e depois desta etapa.

A Seção 4.3 apresenta nuvens de palavras e resultados da modelagem de tópicos referentes aos *tweets* dos casos e discute sobre a qualidade do pré-processamento na base.

A Seção 4.4 apresenta os resultados da classificação dos *tweets* com o classificador escolhido.

4.1 Coleta de Dados

A coleta foi realizada por meio da rede social *Twitter* com as palavras-chave apresentadas na Tabela 6. Foram coletados *tweets* na língua portuguesa, a data de início da publicação dos dados foi relacionado às particularidades de cada caso e a data final foi 06 de junho de 2020. No total, foram encontrados 270.218 *tweets*.

As Figuras 6 e 7 a seguir apresentam a quantidade de *tweets* coletados em cada caso por período. É possível perceber que nos anos 2012 e 2013, houve uma grande quantidade de *tweets* sobre os casos Eliza e Eloá. Este aumento se deu pois nesses anos aconteceram os julgamentos do goleiro Bruno e outros envolvidos no crime contra Eliza Samúdio, e houve o julgamento do agressor de Eloá Cristina em 2012 e no ano seguinte a redução da pena do mesmo. Também é interessante ressaltar que em 2016 e 2017 houve um grande volume de *tweets* relacionados ao caso Eliza Samúdio novamente e o motivo disso pode ser relacionado à notícia em que o agressor se casa com outra mulher, mesmo estando preso, e pelo fato de Bruno conseguir o *habeas corpus* em 2017.

Em 2018 pode-se perceber um grande número de *tweets* relacionados ao caso Eloá Cristina, visto que ele não era muito comentado desde 2014, é possível que este aumento esteja relacionado ao ano, ao fato de que em 2018 completou-se 10 anos deste caso. É válido ressaltar que em 2019 teve uma quantidade considerável de *tweets* sobre os casos Eva Luana e Elaine Caparroz. Ambos aconteceram nesse ano e tiveram bastante repercussão no *Twitter*. No mesmo ano completou-se 10 anos desde o caso Eliza Samúdio, desse modo

foi o caso que mais obteve *tweets*.

É importante destacar que Mariana Ferrer foi agredida em 2018, mas o caso só foi publicado em maio de 2019. No entanto, no período de 1 de janeiro de 2020 a 6 de junho de 2020, houve um aumento considerável do caso Mariana. O julgamento do réu estava previsto para acontecer neste ano, e o assunto voltou a ser repercutido no *Twitter*.

Assunto	Palavra-Chave	Quantidade de <i>tweets</i>
Caso Elaine Caparroz	“Elaine Caparroz”	687
	“Elaine 4 horas”	720
	“Paisagista agredida”	203
	“Elaine tentativa feminicídio”	66
		1.676
Caso Eliza Samúdio	“Eliza Samudio”	117.207
	“Eliza goleiro bruno”	25.500
	“Eliza esquartejada”	423
	“Eliza bruno cachorros”	816
		143.946
Caso Eloá Cristina	“Eloa Lindemberg”	16.274
	“Eloa sequestro”	1.686
	“Eloa morte”	4.696
	“Eloa carcere”	339
		22.995
Caso Eva Luana	“Eva Luana”	1.543
	“Eva Luana violencia”	67
	“Eva Luana padrasto”	308
	“SomosTodosEva”	611
		2.529
Caso Mariana Ferrer	“Mariana Ferrer”	2.872
	“Blogueira Cafe de la Musique”	51
	“Blogueira estuprada”	255
	“maribferrer estupro”	616
		3.794
Caso Poliana Bagatini	“Poliana Bagatini”	447
	“Victor chaves esposa”	192
	“Victor chaves mulher gravida”	300
	“Victor agredir mulher”	583
		1.522
Caso Tatiane Spitzner	“Tatiane Spitzner”	2.701
	“Tatiane jogada predio”	62
	“Spitzner feminicídio”	354
	“Spitzner morte”	464
		3.581

Tabela 6 – Palavras-chave na coleta pelo *Twitter* e quantidade de *tweets* coletados.

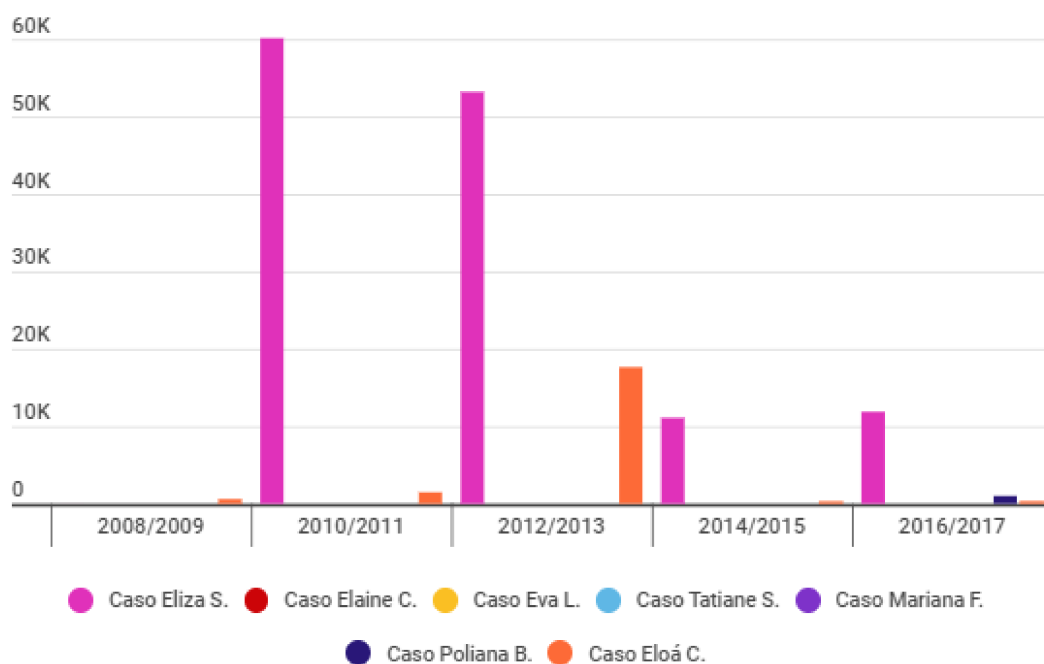


Figura 6 – Quantidade de *tweets* coletados a cada 2 anos sobre todos os casos

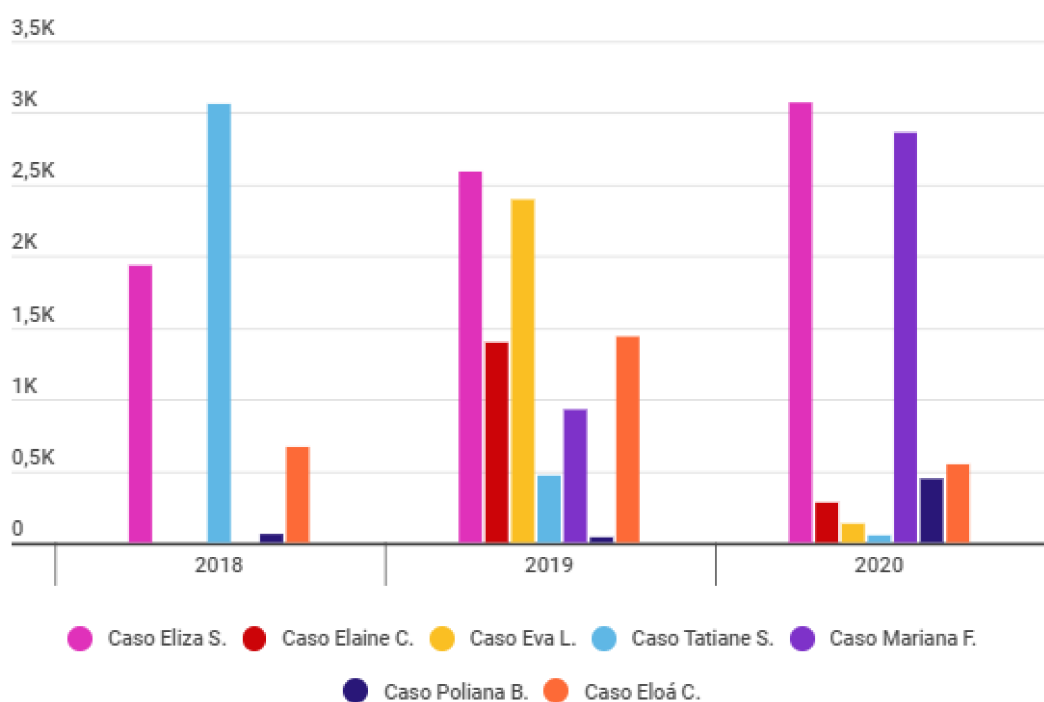


Figura 7 – Quantidade de *tweets* coletados por ano sobre todos os casos

4.2 Pré-Processamento

O algoritmo utilizado realizou as etapas seguintes e obteve os seguintes resultados:

1. **Remoção de *stopwords*:** na Tabela 7 estão alguns exemplos de *stopwords* utilizadas no algoritmo.

aqui	agora	alguém	não
após	nossos	nossa	eu
ninguém	novamente	qual	são
seja	é	serão	só
sou	dele	contudo	manter

Tabela 7 – Exemplos de *stopwords* do algoritmo em *Java*

- Remoção de caracteres não alfabéticos, pontuação e acentuação:** estes não agregam valor à classificação.
- Conversão de letras maiúsculas para minúsculas:** com o objetivo de padronizar o texto.
- Substituição de gírias e abreviações:** Um dicionário contendo 324 gírias juntamente com abreviações e suas respectivas traduções foi construído para incorporar novos termos aos relatos e garantir que a semântica do texto fosse preservada. O dicionário foi construído a partir de experiência da autora e de listas de gírias disponíveis na *Internet*. Alguns dos termos presentes no dicionário, construído para este estudo, são apresentados na Tabela 8.

Gíria	Tradução	Abreviação	Tradução
brinks	brincadeira	hr	hora
bora	vamos	bj	beijo
btf	boto fé	ksa	casa
coments	comentário	vc	você
mano	irmão	qdo	quando

Tabela 8 – Exemplos de gírias e abreviações com suas traduções no programa em *Java*

- Junção de palavras compostas:** termos como “Rio de Janeiro” e nomes compostos são colocadas como uma palavra só, para ficar mais claro a referência das palavras.
- Remoção de *tweets* duplicados.**
- Remoção de *tweets* de reportagens:** estes não agregam valor à classificação.

Após a realização destas etapas, foi realizada a lematização por um programa também em *Java*. A Tabela 9 mostra um exemplo de *tweet* escolhido aleatoriamente para exemplificar as etapas do pré-processamento.

A Tabela 10 condensa os dados, mostrando a quantidade de *tweets* antes e após o pré-processamento para cada caso escolhido. É possível que a diminuição considerável de *tweets* aconteceu pois houve a remoção de reportagens.

Antes do Pré-Processamento	Depois do Pré-Processamento
“Revoltante. Não há outra forma de escrever esse caso absurdo de violência. Elaine foi espancada por quatro horas e teve seu rosto desfigurado por um homem que conheceu pela internet. Ele foi preso em flagrante e responde por tentativa de feminicídio.”	“revoltante forma descrever absurdo violencia espancar rosto desfigurado conhecer internet preso flagrante responder”

Tabela 9 – Exemplo de um *tweet* antes e depois do pré-processamento

Assunto	Quantidade de <i>tweets</i>		
	Antes do pré-processamento	Depois do pré-processamento	Diferença
Caso Elaine Caparroz	1.676	356	-79%
Caso Eliza Samúdio	143.946	53.733	-63%
Caso Eloá Cristina	22.995	8.749	-62%
Caso Eva Luana	2.529	1.518	-40%
Caso Mariana Ferrer	3.794	3.130	-18%
Caso Poliana Bagatini	1.522	415	-73%
Caso Tatiane Spitzner	3.581	1.051	-71%

Tabela 10 – Assuntos na coleta pelo *Twitter* e quantidade de *tweets* coletados.

4.3 Busca por padrões

Com a finalidade de auxiliar no entendimento dos temas que aparecem nos *tweets* coletados, foram usadas nuvens de palavras e o LDA para modelagem de tópicos. Para ajudar na validação do pré-processamento as nuvens de palavras também foram usadas.

4.3.1 Nuvens de palavras

Para validar o pré-processamento e auxiliar na busca por padrões no texto, foram utilizadas nuvens de palavras, que são imagens compostas por palavras e demonstram de maneira visual a frequência da ocorrência dos termos em um texto. Desse modo, quanto maior a palavra na nuvem, maior foi a ocorrência dela na base de dados de cada caso.

A Figura 8 refere-se aos *tweets* coletados com as palavras-chave do caso Elaine Caparroz e é possível perceber que as palavras que mais aparecem são “*espancar*” e “*mulher*”. Isso provavelmente ocorre, pois faz-se uma referência ao caso, no qual Elaine foi espancada. As palavras “*apartamento*”, “*riodejaneiro*”, “*barradatijuca*” também aparecem, por se tratarem do local do crime. O termo “*agressor*” aparece ao lado do nome “*viniciusbatista*”, o homem que espancou Elaine. Também são mostradas as palavras “*covarde*” e “*monstro*”, que podem referenciar o agressor. As palavras “*absurdo*” e “*brutal*” são usadas, possivelmente, pela indignação dos usuários quanto o crime.

A Figura 9 expõe a nuvem de palavras contidas nos *tweets* coletados que fazem



Figura 8 – Nuvem de palavras contidas nos *tweets* coletados com as palavras-chave do Caso Elaine Caparroz – Gerada a partir do WordItOut

referência ao Caso Eliza Samúdio. É possível perceber que “*achar*”, “*corpo*”, “*morte*” e “*matar*” são as palavras que mais aparecem na nuvem, dando a entender que os *tweets* falam sobre a morte e achar o corpo da vítima, que até a data deste estudo não foi encontrado. A nuvem contém as palavras “*filme*” e “*porno*”, o que pode indicar comentários sobre a profissão da vítima, atriz pornô. Os termos “*ex*” e “*flamengo*” podem estar referenciando o agressor de Eliza, que era seu ex-namorado e famoso por ser goleiro do time carioca Flamengo na época do crime. A palavra “*comer*” também aparece com muita frequência, o que pode ter ligação com a morte de Eliza, afinal Bruno afirma que a vítima foi esquartejada e jogada para os cachorros comerem (HERINGER, 2013).

Já a Figura 10 mostra a nuvem de palavras contidas na base de *tweets* coletados sobre o Caso Eloá Cristina. A palavra com maior ocorrência é “*matar*”, o que indica o resultado do sequestro: a morte da vítima. Palavras como “*condenar*”, “*juízo*”, “*prisão*” e “*pena*” também aparece com frequência, referenciando o julgamento do agressor. É interessante destacar que “*mídia*”, “*polícia*” e “*soniaabrao*” são palavras presentes na nuvem. Isso se deve ao fato de que as ações dos mesmos foram bastante criticadas. Muitas pessoas alegam que a mídia, principalmente a apresentadora Sônia Abrão, juntamente com as decisões dos policiais que estavam responsáveis pelas negociações com o sequestrador, não estavam de acordo com o que devia ser feito para salvar Eloá.

Na Figura 11 estão expostos os termos que apareceram com maior frequência nos *tweets* sobre o caso Eva Luana. Termos como “*ler*” e “*chorar*” também aparecem, o que pode complementar a palavra que aparece com maior frequência “*historia*”, estas podem estar referenciando o caso que foi contado por Eva em sua rede social. São expostas mui-



Figura 9 – Nuvem de palavras contidas nos *tweets* coletados com as palavras-chave do Caso Eliza Samúdio – Gerada a partir do WordItOut



Figura 10 – Nuvem de palavras contidas nos *tweets* coletados com as palavras-chave do Caso Eloá Cristina – Gerada a partir do WordItOut

tas palavras que denotam sentimento negativo, como “*triste*”, “*sofrer*”, “*odio*”, “*medo*”, “*horror*” e “*sofrimento*”, indicando indignação quanto ao caso. “*Instagram*” também está presente na nuvem, esta, por sua vez, foi a rede social usada pela vítima para denunciar publicamente as agressões sofridas. Palavras como “*justica*”, “*forca*” e a hashtag “*lutecoumagarota*” indicam solidariedade à vítima.

A Figura 12 apresenta a nuvem de palavras contidas nos *tweets* do caso Mariana Ferrer. É possível perceber claramente que os usuários falaram bastante sobre justiça neste



Figura 11 – Nuvem de palavras contidas nos *tweets* coletados com as palavras-chave do Caso Eva Luana – Gerada a partir do WordItOut

caso, afinal ela é a palavra mais aparece na nuvem. Em complemento, as palavras “*impune*”, “*odio*” e “*prender*” são expostas, podendo inferir que as pessoas estão indignadas com a impunidade do caso. A palavra “*estuprador*” também aparece com muita frequência fazendo referência ao crime, além do nome do réu no processo, “*andrecamargoaranha*”. Termos que remetem ao local da agressão podem ser vistos, como “*florianopolis*” e “*beachclub*”. Palavras de natureza negativa, como “*infelizmente*”, “*triste*”, “*sofrer*”, “*medo*” e “*chorar*” também estão presentes nos *tweets*, podendo indicar solidariedade à vítima. Como na nuvem anterior (Figura 11, “*instagram*” é vista na nuvem, por se tratar de um caso em que a vítima expôs seu relato nessa rede social. Além disso, *hashtags* como “*ajudemarianaferrer*” e “*justicapormariferrer*” também estão presentes.

A nuvem de palavras para os *tweets* relacionados ao caso Poliana Bagatini é exibida na Figura 13 e a palavra que aparece com maior frequência é “*dia*”. Juntamente com “*regime*”, “*condenar*”, “*prisao*” e “*regimesemiaberto*”, palavras que também podem ser vistas na nuvem, é possível inferir que um dos assuntos mais falados foi sobre a condenação do agressor, 18 dias em regime semi aberto. Palavras que se remetem à vida do agressor como “*cantor*”, “*duplaleo*” e “*thevoicekids*” podem ser vistas na nuvem. Termos como “*goleirobruno*”, “*goleirojean*” e “*josemayer*” podem indicar a relação que os usuários fizeram com outros casos de agressão contra mulher. Além disso, termos que indicam a agressão sofrida pela vítima podem ser vistos, como por exemplo “*chute*”, “*espancar*” e “*empurrar*”.



Figura 12 – Nuvem de palavras contidas nos *tweets* coletados com as palavras-chave do Caso Mariana Ferrer – Gerada a partir do WordItOut



Figura 13 – Nuvem de palavras contidas nos *tweets* coletados com as palavras-chave do Caso Poliana Bagatini – Gerada a partir do WordItOut

A Figura 14 apresenta a nuvem de palavras contidas nos *tweets* sobre o caso Tatiane Spitzner. As palavras que aparecem com maior frequência são “mulher”, “marido” e “video”, o que referenciam a algumas características do caso, como o fato da vítima ser uma mulher e o agressor ser seu marido. Vídeos das câmeras de segurança do prédio em que aconteceu o crime foram divulgados e é possível ver as agressões sofridas pela mulher. O nome do agressor, “luisfelipemanvailer” também aparece com frequência nos

tweets. O termo “*metercolher*” também aparece, referenciando o ditado “Em briga de marido e mulher, ninguém mete a colher”. Foi relatado que os vizinhos da vítima ouviram os gritos de socorro da mulher, no entanto não ajudaram, acreditando se tratar de uma briga “normal” de casal, afinal estas palavras são expostas na figura (“*vizinhos*”, “*pedir*” e “*socorro*”). Palavras que indicam indignação quanto ao caso também são vistas na nuvem, como “*triste*”, “*chorar*”, “*revoltante*” e “*monstro*”.



Figura 14 – Nuvem de palavras contidas nos *tweets* coletados com as palavras-chave do Caso Tatiane Spitzner – Gerada a partir do WordItOut

4.3.2 Modelagem de Tópicos

Para a modelagem de tópicos foi usado o LDA a partir de um algoritmo na linguagem *Python*, e foi aplicado aos *tweets* sobre todos os casos analisados separadamente. Foram escolhidos alguns cenários para execução do LDA: 5 tópicos com 6 palavras cada e 10 tópicos com 6 palavras cada. Desse modo, foi adotado somente a perspectiva de 5 tópicos com 6 palavras, isso se deu pois houve maior facilidade de rotulação desses tópicos. Dessa visão, foi possível rotular os 5 tópicos com 6 palavras cada. A rotulação foi feita baseada nos conhecimentos e experiências da autora.

Na Tabela 11 são mostrados os resultados para a modelagem de tópicos a partir de *tweets* sobre o Caso Elaine Caparroz. O primeiro grupo pode estar associado ao encontro da vítima com o agressor, pois a palavra “*apartamento*” aparece. Já o tópico Culpa da violência pode fazer referência ao culpado da agressão, por aparecer a palavra “*culpar*”. A palavra “*merecer*” neste tópico indica duas classes de *tweets* com frequências similares de ocorrência: os que responsabilizam a vítima pelo ocorrido, como mostrado nas Figuras 15 e 16, e os que indicam que Elaine não merecia apanhar. O tópico Violência sofrida pode ser

associado com a agressão cometida, a vítima foi espancada. A partir do tópico Vingança pode-se inferir que o crime foi feito por vingança ou até mesmo alguém vai se vingar pela agressão. As palavras “filho”, “lutador”, “espancar” e “marcar” aparecem juntas neste tópico e é importante destacar que o filho de Elaine é um lutador bem sucedido. Por fim, no último tópico é possível ver a palavra “deus” e “justiça”, o que pode sinalizar um pedido de justiça do caso.

Caso Elaine Caparroz	
Tópicos	Termos
Encontro	0.033*“espancar” + 0.022*“mulher” + 0.014*“acontecer” + 0.013*“covarde” + 0.010*“apartamento” + 0.010*“internet”
Culpa da violência	0.018*“espancar” + 0.016*“gente” + 0.015*“cara” + 0.013*“culpar” + 0.012*“merecer” + 0.011*“casa”
Violência sofrida	0.020*“mulher” + 0.014*“espancar” + 0.013*“vitima” + 0.012*“encontro” + 0.012*“monstro” + 0.012*“deixar”
Vingança	0.010*“filho” + 0.010*“lutador” + 0.010*“espancar” + 0.009*“logo” + 0.009*“querer” + 0.008*“marcar”
Pedido de justiça	0.020*“espancar” + 0.014*“agressor” + 0.014*“deus” + 0.014*“justica” + 0.013*“mulher” + 0.013*“apartamento”

Tabela 11 – Tópicos com 6 palavras referentes ao Caso Elaine Caparroz

Sério mesmo que tem mulher dizendo que a Elaine Caparróz merecia apanhar por ela ter declarado voto no Bolsonaro? É isso? Que doença é essa?

[Translate Tweet](#)

3:28 PM · Feb 18, 2019 · Twitter for Android

Figura 15 – Exemplo de *tweet* coletado que justifica a rotulação do tópico Culpa no Caso Elaine Caparroz.

Sentada na sala de casa ouço: "A Globo fica no pé de Bolsonaro, o homem não pode nem trabalhar direito." da mesma pessoa que disse: "Essa velha (Elaine Caparroz) mereceu apanhar, onde já se viu levar um homem que não conhece pra dentro de casa" CANCELA O MUNDO MINHA GENTE!

[Translate Tweet](#)

9:03 PM · Feb 25, 2019 · Twitter Web Client

Figura 16 – Exemplo de *tweet* coletado que justifica a rotulação do tópico Culpa no Caso Elaine Caparroz.

Já na Tabela 12 podem ser vistos os tópicos sobre o Caso Eliza Samúdio. O primeiro tópico foi rotulado como Crime, pois as palavras que aparecem fazem referência à agressão. É válido lembrar que palavras como “morto” podem estar referenciadas como

substantivos tanto no masculino quanto no feminino, afinal na etapa de lematização elas são consideradas como uma só palavra. Com isso, infere-se que, dado o contexto em que a vítima foi morta, a palavra neste tópico está ligada a ela. O segundo tópico encontrado foi Relacionamento, aparecem palavras que fazem referência ao agressor, que era jogador do clube Flamengo. Eliza Samúdio era ex-amante de Bruno. Já no tópico Caso na Justiça, são vistas palavras como “advogado” e “suspeito”, podendo fazer ligação ao julgamento do caso. Já no próximo tópico, “mae” e “filho” aparecem, desse modo foi rotulado como Família. No último tópico, Descrédibilização da Vítima, assim como na nuvem de palavras, podem ser vistas palavras que se remetem à profissão da vítima. É possível encontrar diversos *tweets* que falam sobre isso em tom de brincadeira ou deboche do crime, como é o caso da Figura 17.

Caso Eliza Samúdio	
Tópicos	Termos
Crime	0.061*“matar” + 0.035*“falar” + 0.025*“policia” + 0.023*“morto” + 0.018*“gente” + 0.018*“mulher”
Relacionamento	0.067*“ex” + 0.060*“morte” + 0.051*“flamengo” + 0.029*“amante” + 0.018*“historia” + 0.018*“condenar”
Caso na justiça	0.042*“advogado” + 0.032*“pai” + 0.022*“acabar” + 0.021*“afirmar” + 0.018*“suspeito” + 0.017*“estar”
Família	0.038*“filho” + 0.018*“mae” + 0.018*“viver” + 0.015*“desaparecimento” + 0.015*“ouvir” + 0.015*“vida”
Descrédibilização da vítima	0.045*“corpo” + 0.044*“comer” + 0.034*“achar” + 0.029*“porno” + 0.025*“filme” + 0.024*“ver”

Tabela 12 – Tópicos com 6 palavras referentes ao Caso Eliza Samúdio

Eliza Samudio previu a sua morte... Em um filme pornô ela disse "Me come cachorro"

[Translate Tweet](#)

5:51 PM · Jun 22, 2011 · Twitter Web Client

Figura 17 – Exemplo de *tweet* coletado que justifica a rotulação do tópico Descrédibilização da Vítima no Caso Eliza Samúdio.

Na Tabela 13 podem ser vistos os resultados para a modelagem de tópicos a partir dos *tweets* sobre o Caso Eloá Cristina. O grupo Crime é composto por palavras que remetem a características do caso, como “atirar” e “morrer”. O segundo tópico está associado à Culpa do que se resultou o sequestro. Como foi citado na nuvem de palavras sobre o caso, muitas pessoas culpabilizaram a mídia, principalmente a apresentadora Sônia Abrão, pela morte de Eloá, o que pode ser visto pela Figura 18. O grupo Relacionamento está associado a relação do agressor com a vítima, eram ex-namorados. Já no tópico Justiça, as palavras que aparecem estão ligadas ao julgamento do agressor, como “prisao”,

“justica” e “condenar”. No último grupo rotulado como Relação com outros casos, é possível ver o nome “isabela” e a palavra “midia”, isso pode se dar por relacionarem a cobertura que foi feita pela mídia no caso Isabella Nardoni, menina de 5 anos de idade jogada do sexto andar de um prédio no mesmo ano que ocorreu o crime contra Eloá.

Caso Eloá Cristina	
Tópicos	Termos
Crime	0.038*“policia” + 0.033*“atirar” + 0.031*“morrer” + 0.020*“esquecer” + 0.018*“gente” + 0.016*“chorar”
Culpa	0.065*“pena” + 0.019*“vivo” + 0.019*“vida” + 0.016*“nayara” + 0.015*“soniaabrao” + 0.015*“culpa”
Relacionamento	0.047*“matar” + 0.041*“lembrar” + 0.031*“menino” + 0.026*“ex” + 0.025*“privado” + 0.025*“namorado”
Justiça	0.026*“justica” + 0.026*“mae” + 0.023*“prisao” + 0.023*“acusar” + 0.019*“irmao” + 0.018*“condenar”
Relação com outros casos	0.033*“midia” + 0.032*“cara” + 0.025*“assistir” + 0.021*“ver” + 0.019*“espero” + 0.018*“isabela”

Tabela 13 – Tópicos com 6 palavras referentes ao Caso Eloá Cristina

e falo mais foi uma das culpadas na morte da elo
sim!!! ficou fazendo circo em volta de homem
psicopata pra ganhar audiência e isso custou a vida
da elo. se essa mulher tivesse o minimo de
profissionalismo nao tinha entrado nessa. vai tomar
no c* sonia abrao

[Translate Tweet](#)

11:06 AM · Oct 10, 2019 · Twitter Web App

Figura 18 – Exemplo de *tweet* coletado que justifica a rotulação do tópico Culpa no Caso Eloá Cristina.

É possível ver que os tópicos resultantes do LDA aplicado ao Caso Eva Luana também condizem com a nuvem de palavras. Esses tópicos são expostos na Tabela 14. O primeiro tópico, Entrevista, pode se referir a uma entrevista que Eva Luana deu ao programa *Encontro com Fátima Bernardes*. No segundo grupo, palavras que expressam intensidade podem ser vistas, assim foi rotulado como Indignação. Já o terceiro grupo, palavras como “ajudar”, “dor” e “chocada” exprimem empatia com a vítima. No tópico rotulado Divulgação do Relato é possível ver palavras que apareceram na nuvem, como “historia”, “relato” e “ler”, desse modo é possível inferir que estão associadas ao relato divulgado por Eva em uma rede social. Por fim, no tópico Solidariedade à vítima são expostas palavras que exaltam a vítima, como “forte”, e pedidos de “justiça”.

Na Tabela 15 é mostrado os resultados do LDA para o Caso Mariana Ferrer. O primeiro tópico foi rotulado como Apoio à vítima por aparecerem palavras de intensi-

Caso Eva Luana	
Tópicos	Termos
Entrevista	0.022*“encontro” + 0.021*“forçar” + 0.020*“querer” + 0.018*“chorar” + 0.017*“crianca” + 0.016*“coragem”
Indignação	0.037*“dia” + 0.028*“esperar” + 0.022*“caralho” + 0.019*“chao” + 0.019*“fe” + 0.017*“difícil”
Empatia	0.029*“ajudar” + 0.027*“igual” + 0.023*“chocada” + 0.020*“nojo” + 0.018*“coracao” + 0.017*“dor”
Divulgação do relato	0.090*“historia” + 0.057*“ler” + 0.029*“relato” + 0.028*“conseguir” + 0.026*“passar” + 0.023*“gente”
Solidariedade à vítima	0.062*“chorar” + 0.043*“lutar” + 0.041*“forte” + 0.032*“garota” + 0.030*“justica” + 0.029*“mulher”

Tabela 14 – Tópicos com 6 palavras referentes ao Caso Eva Luana

dade juntamente com “lembrar” e “relato”. No tópico Agressão é possível ver a palavra “estuprador”, fazendo referência ao crime. Já no grupo Revolta, palavras como “nojo”, “prender” e “mentiroso” podem estar relacionados ao réu do processo e exprimem revolta sobre o caso. No tópico Agressor é possível ver palavras que os internautas provalvemente caracterizam o agressor, como “famoso” e “rico”. O último tópico é composto por termos que estão associados à luta por justiça e fazem referência ao relato da vítima divulgado em seu “Instagram”, assim, foi rotulado como Solidariedade à vítima.

Caso Mariana Ferrer	
Tópicos	Termos
Apoio à vítima	0.032*“falar” + 0.023*“xingamento” + 0.019*“relato” + 0.018*“puto” + 0.017*“jurere” + 0.017*“lembrar”
Agressão	0.072*“justica” + 0.029*“mulher” + 0.023*“florianopolis” + 0.020*“ler” + 0.018*“vitima” + 0.014*“estuprador”
Revolta	0.032*“nojo” + 0.032*“deus” + 0.021*“prender” + 0.016*“abuso” + 0.015*“forma” + 0.013*“mentiroso”
Agressor	0.040*“amigo” + 0.030*“ajudar” + 0.022*“cultura” + 0.021*“passar” + 0.018*“famoso” + 0.017*“rico”
Solidariedade à vítima	0.031*“historia” + 0.030*“precisar” + 0.024*“instagram” + 0.022*“triste” + 0.019*“favor” + 0.019*“lutar”

Tabela 15 – Tópicos com 6 palavras referentes ao Caso Mariana Ferrer

Na Tabela 16 são expostos os tópicos sobre o Caso Poliana Bagatini. O primeiro grupo é composto de palavras que se referem à Acusação do agressor, pois “thevoicekids” se remete ao programa em que Victor Chaves participava, e “leo” pode se referir ao seu irmão Léo Chaves, com o qual formava uma dupla sertaneja. Já no grupo Delito, aparecem palavras que caracterizam a denúncia do crime, como “acusar”, “agressao”, “boletim” e “ocorrencia”. O tópico Agressor se refere a características relacionadas a Victor Chaves, como a dupla que formava com seu irmão, a profissão de cantor, além de “regime”, possivelmente relacionada à pena que recebeu: 18 dias em regime semi aberto.

No tópico Indignação estão presentes palavra de intensidade como “*chocada*” e palavras que podem se remeter a Victor, como “*embuste*”, gíria usada para se referir a uma pessoa inconveniente. Assim como na nuvem de palavras, nos tópicos também aparecem palavras que estão associadas a outros casos, como “*goleirobruno*”. Desse modo, o último tópico foi rotulado como Relação com outros casos.

Caso Poliana Bagatini	
Tópicos	Termos
Acusação	0.028*“thevoicekids” + 0.027*“acusar” + 0.023*“passar” + 0.018*“ler” + 0.017*“leo” + 0.015*“cabeca”
Delito	0.027*“agressao” + 0.026*“acusar” + 0.024*“duplaleo” + 0.021*“achar” + 0.018*“ocorrencia” + 0.018*“boletim”
Agressor	0.064*“dia” + 0.056*“duplaleo” + 0.033*“agredir” + 0.031*“regime” + 0.026*“cantor” + 0.025*“cara”
Indignação	0.030*“cortar” + 0.020*“chocada” + 0.020*“querer” + 0.020*“conseguir” + 0.017*“cara” + 0.016*“embuste”
Relação com outros casos	0.039*“acusar” + 0.029*“goleirobruno” + 0.022*“cantor” + 0.017*“tentar” + 0.015*“mao” + 0.015*“indiciado”

Tabela 16 – Tópicos com 6 palavras referentes ao Caso Poliana Bagatini

Já a Tabela 17 mostra os tópicos do Caso Tatiane Spitzner. No tópico Vídeo do caso estão presentes palavras que estão relacionadas aos vídeos das câmeras de segurança do prédio em que ocorreu o crime. No segundo tópico, Solidariedade à vítima, palavras que expressam sentimentos negativos com o caso podem ser encontradas, como “*odio*” e “*triste*”. Já no tópico Características do caso, é possível ver palavras que estão ligadas ao caso em questão, Tatiane foi agredida e morta por seu marido, além de que imagens da agressão foram divulgadas. O grupo Justiça é composto por termos que podem ser associados a justiça do caso, também é possível perceber palavra de indignação, como por exemplo “*puto*”. No último tópico, Indignação, a palavra de intensidade “*absurdo*” caracteriza o caso.

Caso Tatiane Spitzner	
Tópicos	Termos
Vídeo do caso	0.021*“mulher” + 0.019*“familia” + 0.019*“sofrer” + 0.018*“assistir” + 0.017*“video” + 0.016*“imagem”
Solidariedade à vítima	0.035*“video” + 0.023*“deixar” + 0.022*“morrer” + 0.020*“odio” + 0.018*“gente” + 0.018*“triste”
Características do caso	0.054*“marido” + 0.024*“mulher” + 0.021*“imagem” + 0.021*“video” + 0.017*“agredir” + 0.016*“morto”
Justiça	0.030*“ver” + 0.026*“cara” + 0.020*“irmao” + 0.016*“policia” + 0.015*“ouvir” + 0.012*“puto”
Indignação	0.023*“matar” + 0.022*“mulher” + 0.020*“agredido” + 0.019*“passar” + 0.017*“absurdo” + 0.016*“sair”

Tabela 17 – Tópicos com 6 palavras referentes ao Caso Tatiane Spitzner

4.4 Análise da base de dados rotulada manualmente

Para o treinamento do classificador, uma base de dados rotulada foi construída, conforme apresentado nas Tabelas 3 e 4. Considerando a base de dados rotulada, composta por 700 *tweets* (100 *tweets* para cada caso de violência), pôde-se observar os sentimentos dos usuários em relação a cada uma das partes dos casos de acordo com a classificação feita manualmente.

É possível perceber que, considerando os *tweets* selecionados, nenhum dos casos tem sentimento “positivo” em relação ao acontecimento. Há uma grande maioria de *tweets* classificados como “neutros”. Esse fato também pode ser constatado claramente por meio do gráfico exposto na Figura 19. Isso é um indicativo de que os usuários publicaram mais *tweets* neutros, o que acaba não agregando informações relevantes ao estudo. Sabe-se que o caso foi comentado, mas não se tem conhecimento sobre a natureza deste comentário, ou seja, se o sentimento expresso no *tweet* é “positivo” ou “negativo”. Alguns casos como o de Eva Luana, Mariana Ferrer e Tatiane Spitzner têm predominância de *tweets* classificados como “negativos”.

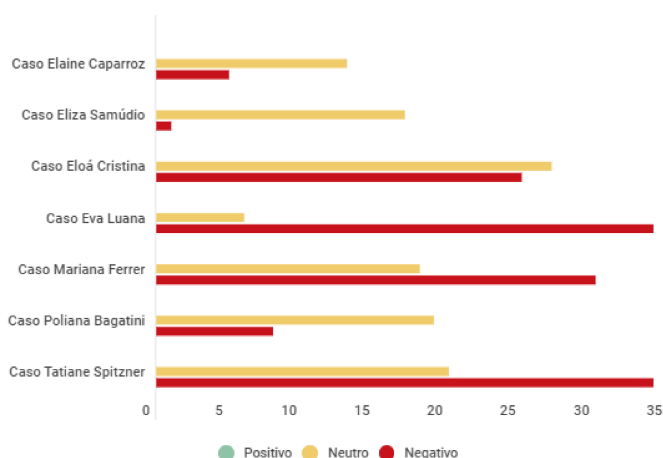


Figura 19 – Visualização gráfica dos sentimentos expressos pelos usuários do *Twitter* em relação aos casos

Já sobre os sentimentos expressos pelos usuários sobre os agressores dos casos, há um alto índice de *tweets* classificados como “negativos”, isto é, os usuários postam mais fazendo críticas sobre quem cometeu os crimes, como exposto na Figura 20. É interessante destacar que existem algumas postagens “positivas”, como nos Casos Eloá Cristina, Eliza Samúdio e Poliana Bagatini. Estes dois últimos têm características em comum sobre os agressores, afinal os dois são conhecidos no país, um deles era um jogador de futebol famoso e o outro, cantor sertanejo.

Analisando os *tweets* sobre as vítimas, na Figura 21, há uma predominância de sentimentos classificados como “positivos”, ou seja, a maioria dos usuários expressam sentimentos bons e de apoio às vítimas. É possível notar que se tratando dos Casos Eliza

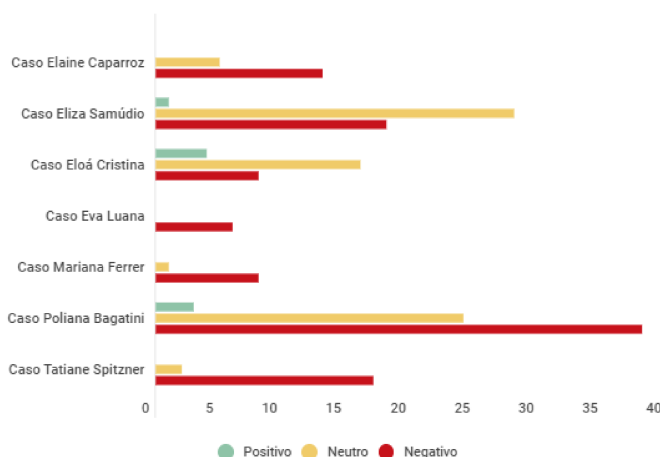


Figura 20 – Visualização gráfica dos sentimentos expressos pelos usuários do *Twitter* em relação aos agressores

Samúdio e Poliana Bagatini, os mesmos casos que tiveram *tweets* classificados “positivos” quanto ao agressor, os sentimentos que prevalecem quanto às vítimas são “negativos”. Pode ser que pelo fato de seus agressores serem conhecidos, muitos usuários descredibilizam a vítima.

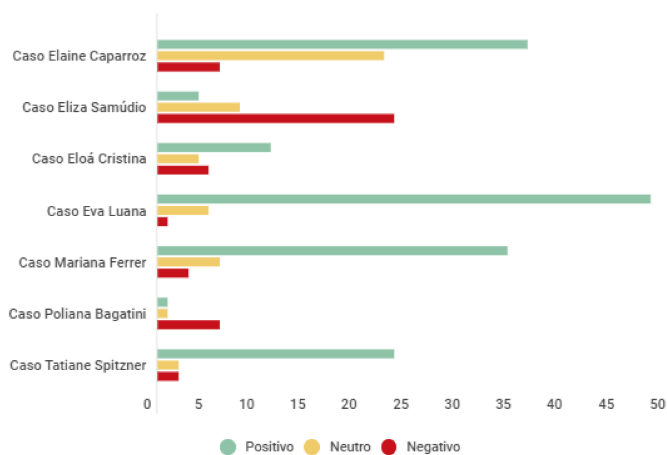


Figura 21 – Visualização gráfica dos sentimentos expressos pelos usuários do *Twitter* em relação às vítimas

4.5 Avaliação do Classificador

Para fazer uma análise dos sentimentos de toda a base de dados, assim como foi feito com a amostra de dados, torna-se necessário classificar o sentimento dos *tweets* (positivo, negativo e neutro) em relação a vítima, caso e agressor. Os experimentos para análise do desempenho do classificador usou a base de dados que manualmente rotulada. Foram executados 4 experimentos variando o conjunto de treinamento e teste a fim de identificar o comportamento do algoritmo *Naive Bayes* frente a diferentes cenários.

O primeiro experimento usou toda a amostra da base de dados, que foi manualmente rotulada, contendo 700 *tweets* (100 *tweets* para cada caso). O objetivo é verificar se o classificador induzido usando o algoritmo *Naive Bayes* consegue distinguir bem os sentimentos, considerando que há *tweets* que falam da vítima, do agressor e do caso. A Tabela 18 mostra a matriz de confusão deste cenário. Dentre as 700 instâncias, 405 foram classificadas corretamente. Os resultados indicam que o desempenho do classificador não foi bom. Um dos possíveis problemas é que neste experimento a base contém instâncias positivas, negativas e neutras de todas os casos de violência e considerando as 3 categorias, agressor, vítima e caso. Assim, as palavras que descrevem um cenário positivo em relação à vítima, podem não ser as mesmas que descrevem o agressor. Além disso, os *tweets* podem conter palavras específicas de um de violência, que podem não aparecer em outro. Com o intuito de melhorar o classificador, 3 novos experimentos foram realizados, considerando a base dividida em 3 categorias, *tweets* que falam da vítima, do agressor e do caso. Aqui novamente o objetivo é classificar o sentimento dos *tweets*, mas criando um classificador distinto para cada uma dessas 3 categorias.

Para realizar tal experimento, a base de dados foi manualmente rotulada nas categorias agressor, vítima e caso. Cada uma dessas sub-bases foi usada em um experimento. Para a sub-base que contém apenas dados da vítima, a matriz de confusão é mostrada na Tabela 19. O classificador acertou 160 instâncias de um total de 251, ou seja, 63,75%. Pode-se perceber que para a classe positiva, a taxa de acerto foi melhor (63,75% em comparação a 57,86% do primeiro experimento).

Os resultados do terceiro experimento, que contém apenas os dados do agressor são mostrados na Tabela 20. De um total de 192 instâncias, o *Naive Bayes* acertou 116, ou seja, 60,42%. Neste experimento, como o número de exemplos da classe positiva são poucos, o desempenho do classificador pode ter sido prejudicado para estes cenários.

Considerando o quarto experimento, cuja base de dados contém apenas instâncias sobre o caso, os resultados são mostrados na Tabela 21. Neste cenário, dentre as 257 instâncias da base, 167 foram corretamente classificadas, obtendo 64,98% de acertos. Esta base de dados não contém *tweets* positivos e portanto a tarefa de classificação envolvia apenas duas classes, o que acaba contribuindo para os melhores resultados do classificador.

Classificado como			
pos	neu	neg	
96	20	49	pos
37	128	79	neu
53	57	181	neg

Tabela 18 – Matriz de confusão para o algoritmo *Naive Bayes* para o conjunto total de treinamento

A partir da Equação 3.1 e dos dados das Tabelas 18, 19, 20 e 21, construiu-se a

Classificado como			
pos	neu	neg	
120	22	15	pos
24	12	12	neu
16	2	28	neg

Tabela 19 – Matriz de confusão para o algoritmo *Naive Bayes* para o conjunto sobre as vítimas

Classificado como			
pos	neu	neg	
2	2	4	pos
4	51	21	neu
17	28	63	neg

Tabela 20 – Matriz de confusão para o algoritmo *Naive Bayes* para o conjunto sobre os agressores

Classificado como			
pos	neu	neg	
0	0	0	pos
3	61	56	neu
2	29	106	neg

Tabela 21 – Matriz de confusão para o algoritmo *Naive Bayes* para o conjunto sobre os casos

Tabela 22, onde é possível observar que a base com *tweets* que se referem aos casos obteve maior acurácia dentre os demais cenários.

Bases	Geral	Vítimas	Agressores	Casos
Acurácia	57,86%	63,75%	60,42%	64,98%

Tabela 22 – Acurácia calculada para cada uma das bases

5 Conclusão

Nesta pesquisa foram apresentados dados estatísticos sobre violência contra mulheres e quão problemáticos são esses índices. O *Twitter* é uma fonte muito rica de dados e a base usada para este estudo foi coletada a partir dele.

O objetivo deste trabalho foi realizar a análise de *tweets* sobre casos de violência contra a mulher que alcançaram grande repercussão na mídia, e verificar o sentimento das postagens sobre o assunto. Para isso, foram realizadas as etapas do processo de Mineração de Texto: coleta dos relatos, pré-processamento dos textos, construção de nuvens de palavras, modelagem de tópicos, criação de base de dados rotulada, classificação dos textos e validação dos resultados. O algoritmo de classificação utilizado foi o *Naive Bayes* e a modelagem de tópicos foi realizado usando o LDA.

A partir dos experimentos realizados, foi possível notar que é difícil classificar os *tweets* em um sentimento geral, afinal os usuários podem estar falando da vítima, do agressor ou do caso em si. Assim, a base de dados também foi dividida em 3 categorias, agressor, vítima e caso, um classificador foi induzido para cada caso. Os resultados indicaram uma melhora no desempenho do classificador, mas em alguns desses casos houve ausência de dados de sentimento positivo, o que pode justificar a melhora nos resultados.

O resultado do classificador poderia ter sido melhor se houvesse um avanço no pré-processamento, como por exemplo retirando palavras específicas de cada caso. Também seria interessante aumentar a base de dados rotulada a fim de ter mais exemplos de cada uma das classes, e assim melhorar a fase de treinamento. Outro ponto a ser destacado é que o desenvolvimento de um classificador para separar os tweets de acordo com o agressor, vítima e caso, seria também muito importante para o trabalho.

Considerando a base de dados manualmente rotulada, foi analisado que 36% dos *tweets* classificados diziam respeito às vítimas, enquanto 37% eram sobre o caso e 27% sobre os agressores. E na maior parte de *tweets* sobre as vítimas continha sentimentos classificados como “positivos”, ao passo que, sobre os agressores, sentimentos “negativos”.

É válida a comparação entre os resultados obtidos com as conclusões de [Rodrigues, Junior e Lobato \(2019\)](#), em que eram analisados comentários sobre o Caso Elaine Caparoz, um dos casos deste estudo. Ao contrário do artigo citado, esta pesquisa concluiu que a maior parte dos comentários feitos no *Twitter* sobre as vítimas de violência contra mulher, são positivos. Essa divergência pode ter ocorrido pelo fato de que em [Rodrigues, Junior e Lobato \(2019\)](#), são analisados comentários extraídos de um site de notícias, no qual não é necessário se identificar para expor alguma opinião, ou seja, tanto os comentários como os *likes* e *dislikes* podem ser feitos anonimamente. Na rede social utilizada neste estudo, é

necessário criar uma conta para que os *tweets* sejam publicados e os *retweets*, que podem ser entendidos como uma medida de aceitação do comentário, sejam expostos no perfil do usuário. Desse modo, é possível inferir que alguns internautas que fazem comentários negativos sobre as vítimas preferem permanecer anônimos.

Vale ressaltar que, em casos envolvendo celebridades como o caso de Eliza Samúdio e da Poliana Bagatini, é possível encontrar *tweets* com sentimento positivo em relação ao agressor ou sentimento negativo em relação à vítima, o que reafirma o machismo ainda presente na sociedade, em especial quando o homem é conhecido. As pessoas tendem a dar mais credibilidade à celebridade do que à vítima. Também é interessante pontuar que foram encontrados *tweets* negativos com relação a vítima no caso Elaine Caparroz, atribuindo a culpa a Elaine por ter conhecido seu agressor pela *Internet* e ter o primeiro encontro em seu apartamento.

5.1 Trabalhos Futuros

Neste trabalho foram utilizados postagens da rede social *Twitter* que fazem referência a casos de grande destaque nas mídias a respeito de violência contra a mulher e feminicídio. Uma predição mais minuciosa pode ser feita obtendo-se uma base mais rica, composta também por comentários publicados nas reportagens dos próprios casos em *sites* de notícias.

Além disso, seria interessante refinar o pré-processamento para que não apareçam palavras que se referem a um caso individualmente. Assim, o classificador receberia como conjunto de treinamento uma base com textos que dizem respeito a todas as vítimas dos casos escolhidos, assim como a todos os agressores. Desse modo, seria possível ter um resultado melhor.

Ao final deste estudo, em novembro de 2020, o Caso Mariana Ferrer voltou a movimentar as redes sociais, e isso se deu pois o site *The Intercept* publicou uma reportagem¹ com o título “JULGAMENTO DE INFLUENCER MARIANA FERRER TERMINA COM SENTENÇA INÉDITA DE ‘ESTUPRO CULPOSO’ E ADVOGADO HUMILHANDO JOVEM”. A expressão utilizada “estupro culposo” e imagens da audiência de instrução e julgamento, que ocorreu em setembro de 2020, causaram indignação e revolta entre a população brasileira. Não só no *Twitter*, mas o caso repercutiu em vários veículos de comunicação, como emissoras de televisão e rádios. Os assuntos mais comentados no *Twitter* no dia da publicação da reportagem eram relacionados ao caso, como “ESTUPRO CULPOSO”, “#justicapormariferrer” e “Mariana”. Assim, para um trabalho futuro, seria interessante analisar o sentimento dos usuários quanto a este caso e comparar se condizem com a classificação encontrada antes da notícia.

¹ <https://theintercept.com/2020/11/03/influencer-mariana-ferrer-estupro-culposo/>

Como o classificador utilizado não foi muito eficaz, seria significativo avaliar outros algoritmos de classificação supervisionados, bem como algoritmos baseados em aprendizado por supervisão à distância, utilizando a ferramenta *Sentiment140* e função de polaridade – utilizando a biblioteca *TextBlob*. Desse modo, outro algoritmo poderia ter um resultado mais favorável para este estudo.

A Casos de violência e feminicídio

Para este estudo, foram selecionados sete casos de violência contra a mulher e feminicídio desde 2008, os quais são detalhados a seguir.

A.1 Caso Elaine Caparroz

A agressão ocorreu na madrugada do dia 16 de fevereiro de 2019 durante um encontro no apartamento da vítima, localizado na Barra da Tijuca, Zona Oeste do Rio de Janeiro. A paisagista Elaine Peres Caparroz, de 55 anos, e o lutador de *jiu-jitsu* Vinicius Batista Serra, 28 anos, se conheceram em uma rede social e, à época da agressão, estavam conversando há oito meses pela *Internet* (SIQUEIRA, 2020).

Durante o encontro, após deitar-se no peito de Vinicius, Elaine caiu no sono (acredita ter sido dopada) e acordou sendo agredida violentamente, o que durou cerca de quatro horas. O agressor desferiu socos e pontapés nela e a arrastou pelos cabelos no chão. Elaine levou mais de cinquenta pontos na boca e teve raízes de dentes destruídas. De acordo com o Ministério Público do Rio, Vinicius não só teve a intenção de matar Elaine como também acreditava que ela estava morta quando deixou seu apartamento (SIQUEIRA, 2020).

O irmão da vítima, Rogério Caparroz, relata que assim que os pedidos de socorro por parte de Elaine começaram, os vizinhos acreditaram que se tratava de uma briga de casal. Só depois de algum tempo prestaram apoio à Elaine, enquanto isso, o agressor foi detido na portaria pelo funcionário do condomínio. Elaine precisou de quase 80 pontos na boca, fraturou o nariz, os ossos que cercam os olhos e perdeu um dente. Vinicius Batista foi preso em flagrante por tentativa de feminicídio (MARTINELLI, 2019).

A.2 Caso Eliza Samúdio

Em 2009, Bruno Fernandes de Souza, conhecido como Goleiro Bruno, conhece Eliza Samúdio em um churrasco no Rio de Janeiro e começam um relacionamento extraconjugal, afinal Bruno era casado. Em maio do mesmo ano, Eliza engravida. Cinco meses depois, o relacionamento entre Bruno e Eliza tornou-se público com o registro da queixa de ameaça de morte contra Bruno, e assim, ele passa a ser investigado criminalmente (UCHÔA, 2009). Em fevereiro de 2010, nasce o filho de Eliza e Bruno não reconhece a paternidade, o que faz com que ela mova um processo na justiça para que Bruno reconhecesse a paternidade e pagamento da pensão.

No dia 4 de junho de 2010, Eliza comunica a família que iria até Minas Gerais para conversar com Bruno, a pedido dele, e este é o último contato dela com família e amigos. Eliza é supostamente assassinada em 10 de junho de 2010 por Marcos Aparecido dos Santos, o “Bola”, e nos dias 24 e 25 do mesmo mês, a polícia recebe denúncias de que Eliza teria sido agredida, morta e o corpo escondido em um sítio do goleiro em Esmeraldas, Minas Gerais. O filho da vítima é encontrado pela polícia no dia 26 na Região Metropolitana de Belo Horizonte, e o pai de Eliza, Luiz Carlos Samúdio, vai ao encontro do neto que estava em um abrigo. A polícia faz buscas no sítio de Bruno em busca do corpo de Eliza, no local foram encontradas roupas de mulher, objetos de crianças e fraldas. Peritos também examinam o carro de Bruno e, segundo a polícia, vestígios de sangue de Eliza foram encontrados no veículo.

Em 6 de julho, um adolescente de 17 anos, que foi apreendido no apartamento de Bruno, depõe na Delegacia de Homicídios da Barra da Tijuca. Ele conta que participou do sequestro de Eliza, e deu coronhadas na vítima. No inquérito que relata o crime, Eliza teve os braços amarrados com uma corda e foi estrangulada por Bola. Ainda segundo o adolescente, após ter estrangulado Eliza, Bola pediu para que todos deixassem o local. Depois, seguiu em direção a um canil, carregando um saco que supostamente continha o cadáver esquartejado de Eliza. No dia seguinte, a prisão preventiva de Bruno e de mais 7 pessoas foram expedidas pela Justiça de Minas. A Justiça do Rio de Janeiro também havia expedido a prisão preventiva de Bruno e Luiz Henrique Mourão, conhecido como Macarrão, por sequestro e cárcere privado de Eliza Samúdio em outubro de 2009.

Sônia Fátima Moura, mãe de Eliza, consegue guarda provisória do neto de 4 meses. O Flamengo, clube no qual Bruno era goleiro, decidiu demitir Bruno por justa causa. Em 30 de julho de 2010, Bruno, Macarrão, Bola e mais seis pessoas são indiciados, e em outubro um exame de DNA comprova que Bruno é pai do filho de Eliza.

Em junho de 2012, mãe de Eliza recebe carta anônima falando sobre possível local onde foram deixados os restos mortais de sua filha. Sérgio Rosa Sales, considerado testemunha-chave do caso e primo de Bruno, é encontrado morto em Belo Horizonte. No dia 19 de novembro de 2012, foi iniciado o julgamento dos réus. No entanto, o julgamento de Bruno foi desmembrado e adiado para março de 2013. Macarrão é condenado a 15 anos de prisão por homicídio qualificado, e Fernanda Gomes de Castro, ex-namorada de Bruno, a 5 anos de prisão por participação no crime.

No início do ano de 2013, a juíza determina expedição de certidão de óbito de Eliza Samúdio, no mandado, a causa da morte foi declarada como asfixia. Em março, Bruno é condenado a 17 anos e 6 meses em regime fechado por homicídio triplamente qualificado, a outros 3 anos e 3 meses em regime aberto por sequestro e cárcere privado, e ainda mais 1 ano e 6 meses por ocultação de cadáver. Bola foi condenado a 22 anos de prisão pelo crime.

Em 2014, mesmo em regime fechado, Bruno assinou contrato com o Montes Claros Esporte Clube. A justiça negou a permissão para que Bruno pudesse jogar, pois segundo o juiz, as atividades de um atleta profissional são incompatíveis com a legislação para trabalho externo. Em 20 de junho de 2016, Bruno se casou com Ingrid Calheiros, com quem iniciou relacionamento amoroso durante o processo sobre o caso, em cerimônia realizada dentro da APAC (Associação de Proteção e Assistência ao Condenado) de Santa Luzia. No dia 10 de março de 2017, Bruno conseguiu um *habeas corpus* por uma liminar deferida pelo ministro do STF Marco Aurélio Mello. No mesmo mês, Bruno se apresentou oficialmente no Boa Esporte, clube com o qual assinou contrato.

Em 25 de abril de 2017, Bruno retorna à prisão, após decisão do Supremo Tribunal Federal (STF). Em agosto, a Justiça autorizou o goleiro a sair da cadeia durante o dia para dar aulas de futebol em uma entidade, o que reduziu um pouco sua pena. E em setembro, o Tribunal diminuiu pena de Bruno para 20 anos e 9 meses, porque o crime de ocultação de cadáver foi prescrito.

Dez anos depois da ocorrência do crime, o corpo de Eliza Samudio ainda não foi encontrado, e Bruno não paga pensão para seu filho, que mora com a avó Sônia (REZENDE, 2020).

A.3 Caso Eloá Cristina

No dia 13 de outubro de 2008, no município de Santo André, em São Paulo, Lindemberg Fernandes Alves invadiu o apartamento de Eloá Cristina Pereira Pimentel, sua ex-namorada. Na época, Lindemberg tinha 22 anos e Eloá, 15. A vítima estava com mais três colegas fazendo trabalhos escolares quando o ex-namorado anunciou o sequestro. Nas primeiras horas, dois jovens foram liberados e restaram no interior da residência Eloá, Nayara Silva e o sequestrador.

No dia seguinte, depois de negociações entre Lindemberg e o Grupo de Ações Táticas Especiais (GATE), Nayara foi libertada. No entanto, no dia 15 foi chamada pela polícia para voltar ao local do sequestro e ajudar na negociação à distância, desse modo, Nayara voltou ao apartamento. Após mais de 100 horas de cárcere privado, os policiais do GATE e da Tropa de Choque da Polícia Militar de São Paulo explodiram a porta do local alegando que ouviram um disparo de arma de fogo no apartamento. Os policiais começaram uma luta corporal com o sequestrador, mas o mesmo conseguiu atirar em direção às reféns. Nayara foi ferida com um tiro no rosto, e Eloá foi levada inconsciente para o Centro Hospitalar de Santo André.

Lindemberg foi levado para delegacia e encaminhado ao Centro de Detenção Provisória de Pinheiros, em São Paulo. Eloá foi baleada na cabeça e na virilha, não resistiu aos ferimentos e morreu por morte cerebral, confirmada no dia 18 de outubro. O julga-

mento do agressor aconteceu em 2012, e foi culpado pelos 12 crimes que foi acusado: um homicídio, duas tentativas de homicídio, cinco cárceres privados e quatro disparos de arma de fogo. Ele foi condenado a 98 anos e 10 meses de prisão, entretanto, o Código Penal limitava, naquela época, a 30 anos o prazo máximo de cumprimento de pena de prisão. Em 2013, sua pena foi reduzida para 39 anos e 3 meses de prisão (WIKIPEDIA, 2008).

É interessante destacar que a ação dos policiais foi muito criticada por especialistas em segurança, principalmente pela volta de Nayara ao cativo (PINHO, 2012). Outro ponto muito questionado e reprovado pelas pessoas foi o papel da mídia brasileira, o crime foi transmitido pelos canais de TV, fazendo uma espetacularização da violência. O documentário *Quem Matou Eloá?*¹ faz uma crítica sobre o caso, apontando as irresponsabilidades cometidas pela cobertura da televisão brasileira e pelos policiais.

A.4 Caso Eva Luana

Em 19 de fevereiro de 2019, Eva Luana da Silva, de 21 anos, publicou em sua rede social Instagram² sua história, contando as violências que sofreu por seu padrasto, Thiago Oliveira Alves. A mãe de Eva era abusada, violentada e torturada por Thiago quase todos os dias. Quando Eva completou 12 anos, seu padrasto começou a abusar dela sexualmente. Aos 13, ela o denunciou, mas foi obrigada a retirar a queixa por ameaças de Thiago. Desde então, os abusos e agressões aumentaram, a ponto de Eva não ter mais vida social.

Eva era obrigada a fazer os trabalhos e responder todas as provas da faculdade de Thiago. Para tudo haviam castigos e punições. Seu celular era vigiado todos os dias e suas mensagens eram respondidas por ele como se fosse ela respondendo. Thiago vigiava Eva na porta da sala da faculdade e a tratava mal em público. Depois de muitos estupros, já abortou diversas vezes, era agredida verbalmente e obrigada a repetir os xingamentos para si mesma. Eva dormia na casinha da cachorra e já passou várias horas sem comer.

No dia 14 de agosto de 2019, Thiago Oliveira Alves foi condenado pela Justiça a 35 anos e 21 dias de reclusão em regime fechado e a um 1 ano e três meses de detenção em regime aberto. Ele foi condenado pelos crimes de lesão corporal no âmbito da violência doméstica, tortura e estupro de vulnerável (HORAS, 2019).

A.5 Caso Mariana Ferrer

Mariana Borges Ferreira, conhecida como Mariana Ferrer, blogueira e influenciadora digital, foi convidada em 2018 para ser embaixadora do *beach club* Café de La

¹ <http://bit.ly/Doc-QuemMatouEloa>

² <https://www.instagram.com/evalluana/>

Musique, em Florianópolis, Santa Catarina. No dia 15 de dezembro de 2018, Mariana relata que foi estuprada pelo empresário André Camargo Aranha, de 42 anos. No entanto, o caso só foi revelado ao público no dia 20 de maio de 2019, quando Mariana divulgou seu relato, vídeos do circuito de segurança da boate e foto do vestido que usava naquela noite todo ensanguentado na rede social Instagram. A blogueira conta que na data do ocorrido tinha 21 anos e estava exercendo seu trabalho no beach club quando sua amiga a chamou para tirar foto em um dos bangalôs exclusivo para sócios. Depois disso, relata que teve uma amnésia temporária, ela menciona também que só tomou um copo de gin. Mariana conta que André não se aproximou quando ela estava lúcida, ela foi levada dopada para um lugar desconhecido por ela e foi estuprada - Mariana era virgem naquela época. A lembrança que ela tem daquela noite, é estar descendo a escada do lugar e alega estar sensações estranhas, como se não sentisse seu corpo.

Depois que a influenciadora saiu do local, mandou mensagens para seus amigos que estavam com ela anteriormente, mas negaram socorro e a deixaram sozinha no beach club. Os efeitos de drogas ficam perceptíveis pelas conversas que Mariana divulgou. Ela conseguiu pedir um carro por um aplicativo para levá-la para casa, e no caminho ligou para sua mãe, Luciane Aparecida Borges, que ficou sem paciência e brava, por achar que a filha estava bêbada ou brincando, desse modo, gravou a conversa para que pudesse mostrar pra ela no dia seguinte. A conversa também foi divulgada por Mariana em seu Instagram, e pelos áudios é possível perceber que a blogueira estava chorando muito e desesperada. Só quando Mariana chegou em casa que sua mãe viu a gravidade do acontecimento. Luciane relata que a filha estava com o corpo mole e olhos vidrados, e foi quando tirou as roupas de Mariana para lhe dar um banho que viu que suas roupas estavam ensanguentadas e com forte cheiro de esperma.

No dia seguinte, 16 de dezembro, o boletim de ocorrência e todos os procedimentos foram feitos. Mariana conta que voltou a sentir seu corpo aos poucos, e com isso vieram as fortes dores entre as pernas e na região íntima. Teve que tomar por 30 dias um coquetel de remédios para evitar doenças sexualmente transmissíveis. Desde a denúncia até o corpo de delito, Mariana foi atendida, examinada, tocada e questionada por homens, se sentiu humilhada e constrangida. No dia 17 de dezembro, o delegado que não estava mais no caso invadiu o condomínio da vítima a paisana, fez questionamentos e pediu mais provas, Mariana se sentiu muito desconfortável com aquela situação.

De acordo com [Alves \(2019\)](#), escritor de um jornal de Santa Catarina, o motorista que a levou para casa naquele dia, afirma em depoimento para a polícia que Mariana parecia estar “sob efeito de algum entorpecente”. Os laudos periciais ([ND, 2019](#)) confirmam que houve rompimento do hímen e conjunção carnal, porém os exames toxicológicos não detectaram a presença de álcool e drogas. Os amigos que estavam com Mariana naquele dia foram ouvidos e no dia 23 de maio, o empresário André Aranha prestou esclarecimentos

à polícia. O investigado afirmou que Mariana gostaria de ir ao banheiro e ele também, desse modo subiram no mesmo instante para o camarim, mas que só havia a visto há uns 5 minutos e a ajudou a subir as escadas porque ela estava de salto. Os investigadores solicitaram o material genético de André para confronto com o sêmen colhido na roupa da vítima, mas seus defensores negaram o pedido. A delegada Caroline Monavique Pedreira, da Dpcami, responsável pela investigação, colheu, durante o interrogatório de André, as impressões digitais do investigado e a saliva dele, em um copo no qual ele bebeu água. De acordo com o laudo, a amostra questionada e a amostra confrontada indicam se tratar da mesma pessoa. A delegada pediu então, decreto de prisão temporária do empresário. Os advogados de Aranha pediram uma liminar de habeas corpus, uma vez que o cliente é réu primário, o que foi acatado pela 1ª Câmara Crimila do Tribunal de Justiça, no dia 18 de julho. Caroline relatou convicção no sentido de indiciar André pelo crime de estupro de vulnerável e representou pela conversão do pedido de prisão temporária em prisão preventiva. Porém, o advogado Cláudio Gastão da Rosa Filho, que defende André, conseguiu reverter a decisão com habeas corpus na segunda instância.

Em audiência de instrução e julgamento, em julho, o réu afirma que a motivação de Mariana de incriminá-lo seria financeira. Além disso, Cláudio Gastão apresenta fotos profissionais publicadas por Mariana nas redes sociais antes da data do crime. Em uma delas, o advogado narra que ela está “com o dedinho na boquinha” e julga que as posições são “ginecológicas”. Também questiona: “por que você apaga essas fotos e deixa só a carinha de choro como se fosse uma santa, só falta uma auréola na cabeça”. As falas dele não são interrompidas na sessão (ALVES, 2020).

Embora tenha concluído que o fato ocorreu, o promotor sustentou que não foi possível comprovar a incapacidade da vítima de evitar o ato sexual. Em 09 de setembro de 2020, o juiz Rudson Marcos, da 3ª Vara Criminal de Florianópolis, absolveu André Aranha ante a ausência de provas. O magistrado aponta que “como as provas acerca da autoria delitiva são conflitantes em si, não há como impor ao acusado a responsabilidade penal, pois, repetindo um antigo dito liberal, “melhor absolver cem culpados do que condenar um inocente”. A absolvição, portanto, é a decisão mais acertada no caso em análise, em respeito ao princípio da dúvida, em favor do réu (*in dubio pro reo*), com base no art. 386, VII, do Código de Processo Penal” (JUSCATARINA, 2020). A decisão causou muita indignação e foi um dos assuntos mais comentados no *Twitter*.

A.6 Caso Poliana Bagatini

No dia 24 de fevereiro de 2017, em Belo Horizonte, Poliana Bagatini prestou queixa na Delegacia de Mulheres contra o marido Victor Chaves, da dupla Victor & Léo. Poliana, que estava grávida, relata ter sido jogada no chão e chutada por Victor, conta também

que foi impedida de deixar o apartamento do casal, mas com a ajuda de uma vizinha conseguiu fugir e pediu socorro. No dia seguinte, a sogra de Poliana registrou boletim de ocorrência, contando que Poliana chegou em sua casa transtornada e que discutiram. Quando Victor chegou, a mulher teria se jogado no chão e se debatido. Nesse mesmo dia, Poliana compareceu à delegacia e fez o exame de corpo de delito, porém os resultados não foram publicados.

Em 25 de fevereiro, Poliana publicou, por meio do Instagram do cunhado, uma carta de apoio ao marido, na qual negou que estivesse sido agredida e explicou estar abalada. Mesmo com a mudança de versão, a Polícia Civil continuou investigando o caso. Após a divulgação do caso, Victor foi afastado do reality *The Voice Kids*, da emissora Globo. No dia 13 de março, o laudo do exame de corpo de delito deu negativo, mas as investigações continuaram, pela possibilidade de agressões que não deixaram marcas corporais. Em 4 de abril, a Polícia Civil concluiu pelo indiciamento de Victor Chaves (VALE, 2017).

O cantor Victor, no dia 11 de fevereiro de 2019, postou um vídeo em seu canal do YouTube, na qual ironiza as acusações de agressão à Poliana, o que gerou muita indignação nas redes sociais (QUEM, 2019). No dia 13 de janeiro de 2020, houve a divulgação das imagens da briga entre o cantor e Poliana. As imagens filmadas pelas câmeras do elevador mostra Poliana tentando sair de casa às pressas, pedindo socorro ao porteiro. Em um momento, é empurrada por Victor, cai no chão e continua sendo agredida com chutes. O sertanejo então foi condenado pelo Tribunal de Justiça de Minas Gerais por vias de fato, a 18 dias em regime aberto, além de uma indenização de 20 mil reais à Poliana. A punição, julgada por muitos injusta, gerou revolta nas redes sociais (BRAZILIENSE, 2020).

A.7 Caso Tatiane Spitzner

Tatiane Spitzner, de 29 anos, foi encontrada sem vida em 22 de julho de 2018, após queda do 4º andar de um prédio em Guarapuava, região central do Paraná. Ao chegar no local, a polícia ouviu testemunhas que disseram ter visto um homem carregando o corpo para dentro do edifício. Conforme a Polícia Militar, o corpo de Tatiane estava dentro do apartamento. O marido de Tatiane, Luis Felipe Manvailer foi preso horas depois ao se envolver em um acidente, em São Miguel do Iguçu, aproximadamente 340 quilômetros do local do crime (PAVANELI; GIMENES, 2018).

Imagens de câmeras de segurança do edifício mostram agressões físicas e discussões do agressor com Tatiane. Luis Felipe foi indiciado pela Polícia Civil, em 31 de julho, por homicídio qualificado. O laudo do exame de necropsia do Instituto Médico-Legal (IML) confirmou que a morte de Tatiane foi por asfixia mecânica, causada por esganadura. Manvailer se tornou réu pelos crimes de homicídio com as qualificadoras de asfixia mecânica,

dificultar defesa da vítima, motivo torpe e feminicídio, além de cárcere privado e fraude processual. No dia 17 de julho de 2020, o Tribunal de Justiça mandou que júri popular de Luis Felipe Manvailer fosse marcado ([HISING, 2020](#)).

Referências

- ALEXA. Top sites brasil. 2020. Acesso em: 13 mar. 2020. Citado na página 13.
- ALVES, S. Caso mariana ferrer: Empresário do ramo esportivo é indiciado por estupro de vulnerável. 2019. Disponível em: <<https://ndmais.com.br/seguranca/caso-mariana-ferrer-empresario-do-ramo-esportivo-e-indiciado-por-estupro-de-vulneravel/>>. Acesso em: 20 set. 2020. Citado na página 65.
- ALVES, S. Caso mariana ferrer: conheça os detalhes do processo que absolveu empresário. 2020. Disponível em: <<https://ndmais.com.br/seguranca/policia/exclusivo-os-detalhes-do-processo-que-absolveu-acusado-de-estuprar-mariana-ferrer/>>. Acesso em: 20 set. 2020. Citado na página 66.
- ANDRADE, P. H. M. A. de. Aplicação de técnicas de mineração de textos para classificação de documentos: um estudo da automatização da triagem de denúncias na cgu. 2015. Disponível em: <https://repositorio.unb.br/bitstream/10482/21004/1/2015_Patr%C3%ADciaHelenaMaiaAlvesdeAndrade.pdf>. Acesso em: 03 abr. 2019. Citado 2 vezes nas páginas 34 e 37.
- ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. *Revista Eletrônica de Sistemas de Informação*, 2006. Acesso em: 03 abr. 2019. Citado 2 vezes nas páginas 21 e 23.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. *Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca*. 2. ed. Porto Alegre: Bookman Editora, 2013. 590 p. Citado 2 vezes nas páginas 25 e 34.
- BARION, E. C. N.; LAGO, D. Mineração de textos. *Revista de Ciências Exatas e Tecnologia*, 2006. Acesso em: 03 abr. 2019. Citado na página 23.
- BASILIO, M. P.; PEREIRA, V. A escolha da estratégica de policiamento em função da demanda criminal: um modelo probabilístico de tópicos integrado à metodologia multicritério. 2018. Disponível em: <https://www.researchgate.net/publication/328102989_A_ESCOLHA_DA ESTRATEGIA_DE POLICIAMENTO_EM_FUNCAO_DA_DEMANDA_CRIMINAL_UM_MODELO_PROBABILISTICO_DE_TOPICOS_INTEGRADO_A_METODOLOGIA_MULTICRITERIO>. Acesso em: 28 ago. 2019. Citado na página 26.
- BIANCHINI, A. Por que as mulheres não denunciam seus agressores? com a palavra, a vítima. 2011. Disponível em: <<https://professoraalice.jusbrasil.com.br/artigos/121814070/por-que-as-mulheres-nao-denunciam-seus-agressores-com-a-palavra-a-vitima>>. Acesso em: 03 abr. 2019. Citado na página 12.
- BORGA, M. 'eu não vou te matar': um retrato da violência doméstica. 2016. Disponível em: <<http://www.ihu.unisinos.br/noticias/550570-eu-nao-vou-te-matar-um-retrato-da-violencia-domestica>>. Acesso em: 03 abr. 2019. Citado na página 12.

- BRAZILIENSE, C. Cantor sertanejo victor chaves recorre de sentença em caso de agressão contra ex. 2020. Disponível em: <<https://www.uai.com.br/app/noticia/e-mais/2020/01/13/noticia-e-mais,254887/cantor-sertanejo-victor-chaves-recorre-de-sentenca-em-caso-de-agressao.shtml>>. Acesso em: 20 set. 2020. Citado na página 67.
- CARASCO, D. Por que muitas mulheres não denunciam a violência doméstica que sofrem? 2019. Disponível em: <<https://www.uol.com.br/universa/noticias/redacao/2017/10/31/por-que-muitas-mulheres-nao-denunciam-a-violencia-domestica-que-sofrem.htm>>. Acesso em: 28 ago. 2019. Citado na página 16.
- CASTRO, B. Y. S. de. Como modelar tópicos através de latent dirichlet allocation (lda) através da biblioteca gensim. 2020. Disponível em: <<https://bit.ly/2PD9LEz>>. Acesso em: 09 ago. 2020. Citado na página 24.
- CLAUDIA. Eva luana: Jovem denuncia padrasto pelo instagram: “me fazia comer o próprio vômito”. 2019. Disponível em: <<https://claudia.abril.com.br/noticias/jovem-denuncia-padrasto-pelo-instagram-me-fazia-comer-o-proprio-vomito/>>. Acesso em: 03 abr. 2019. Citado na página 12.
- CORRÊA, I. T. Análise dos sentimentos expressos na rede social twitter em relação aos filmes indicados ao oscar 2017. 2017. Disponível em: <<https://igoortc.github.io/files/sentimentos-oscar-2017.pdf>>. Citado 5 vezes nas páginas 13, 21, 22, 30 e 34.
- DAUPHIN, C. et al. A história das mulheres. cultura e poder das mulheres: ensaio de historiografia. *Revista do Núcleo Transdisciplinar de Estudos de Gênero*, v.2, n.1, p.7 - 30, 2000. Disponível em: <http://www.marilia.unesp.br/Home/Pesquisa/cultgen/Documentos/historia_das_mulheres_nuteg.pdf>. Acesso em: 03 abr. 2019. Citado na página 12.
- FEDERAL, S. Política nacional de enfrentamento à violência contra as mulheres. 2011. Disponível em: <<https://www12.senado.leg.br/institucional/omv/entenda-a-violencia/pdfs/politica-nacional-de-enfrentamento-a-violencia-contra-as-mulheres>>. Acesso em: 03 abr. 2019. Citado na página 16.
- FEDERAL, S. Panorama da violência contra as mulheres no brasil. 2018. Disponível em: <<http://www.senado.gov.br/institucional/datasenado/omv/indicadores/relatorios/BR-2018.pdf>>. Acesso em: 03 abr. 2019. Citado na página 12.
- FEDERAL, S. Serviços especializados de atendimento à mulher. 2019. Disponível em: <<https://www12.senado.leg.br/institucional/omv/acoes-contraviolencia/servicos-especializados-de-atendimento-a-mulher>>. Acesso em: 28 ago. 2019. Citado na página 20.
- FELIX, N. *Análise de sentimentos em textos curtos provenientes de redes sociais*. 138 p. Tese (Doutorado) — Universidade de São Paulo - São Carlos, 2016. Citado na página 25.
- FILHO, C. H. P. *Técnicas de aprendizado não supervisionado baseadas no algoritmo da caminhada do turista*. Dissertação (Mestrado) — Universidade de São Paulo, 2017. Disponível em: <https://www.teses.usp.br/teses/disponiveis/82/82131/tde-20082018-122603/publico/Dissert_CarlosPortoFilho_corrigida.pdf>. Acesso em: 20 set. 2020. Citado na página 25.

- G1. Nova casa da mulher é inaugurada e aplicativo 'salve maria' para denúncias de violência é lançado em uberlândia. 2019. Disponível em: <<https://glo.bo/2QQUZLx>>. Acesso em: 03 abr. 2019. Citado 2 vezes nas páginas 13 e 21.
- HERINGER, C. Em lágrimas, goleiro bruno confirma que eliza samudio foi esquartejada e jogada para cachorros. 2013. Disponível em: <<https://glo.bo/2HxByp9>>. Citado na página 44.
- HISING, E. O que se sabe do caso de tatiane spitzner, encontrada morta em prédio de guarapuava. 2020. Disponível em: <<https://glo.bo/34rO0j7>>. Acesso em: 20 set. 2020. Citado na página 68.
- HORAS, R. C. . Após abusos, padrasto de eva luana é condenado a 35 anos de prisão. 2019. Disponível em: <<https://www.correio24horas.com.br/noticia/nid/apos-abusos-padrasto-de-eva-luana-e-condenado-a-35-anos-de-prisao/>>. Acesso em: 20 set. 2020. Citado na página 64.
- JUNIOR, J. R. C. *Desenvolvimento de uma Metodologia para Mineração de Textos*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro, Engenharia Eletrica, 2007. Citado na página 28.
- JUNIOR, R. de A. L. Um framework para mineração de textos de redes sociais. 2016. Disponível em: <<http://www.repositoriobib.ufc.br/000033/000033ab.pdf>>. Acesso em: 28 ago. 2019. Citado na página 21.
- JUSCATARINA. Empresário acusado por estupro em beach club de florianópolis é absolvido. 2020. Disponível em: <<https://ndmais.com.br/justica-sc/empresario-acusado-por-estupro-em-beach-club-de-florianopolis-e-absolvido/>>. Acesso em: 20 set. 2020. Citado na página 66.
- LIMIRO, R. M.; CORDEIRO, D. F.; CASSIANO, K. K. Análise de sentimentos em tweets: um estudo de caso da copa do mundo de futebol. 2017. Disponível em: <https://cappa.fic.ufg.br/up/971/o/4gt4_tweets.pdf>. Acesso em: 28 ago. 2019. Citado 2 vezes nas páginas 21 e 22.
- LIU, B. Sentiment analysis and opinion mining. 2012. Disponível em: <<https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>>. Acesso em: 28 ago. 2019. Citado na página 21.
- LOPER, E.; BIRD, S. Nltk: The natural language toolkit. In: *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002. (ETMTNLP '02), p. 63–70. Citado na página 24.
- MAGALHÃES, M. et al. Mete a colher. 2016. Disponível em: <<https://meteacolher.org/>>. Acesso em: 03 abr. 2019. Citado 2 vezes nas páginas 14 e 21.
- MARTINELLI, A. Violência psicológica é a forma mais subjetiva de agressão contra a mulher; saiba como identificar. 2014. Disponível em: <https://www.huffpostbrasil.com/2014/11/25/violencia-psicologica-e-a-forma-mais-subjetiva-de-agressao-contr_a_21676045/>. Acesso em: 28 ago. 2019. Citado na página 17.

MARTINELLI, A. *O caso Elaine Caparroz e a violência que coloca a vítima no banco dos réus*. 2019. HuffPost Brasil. Disponível em: <https://www.huffpostbrasil.com/entry/elaine-caparroz-femicidio_br_5c715b26e4b06cf6bb26451f>. Acesso em: 23 mar. 2020. Citado na página 61.

MULHER, d. F. e. d. D. H. Ministério da. Ligue 180. 2019. Disponível em: <<https://www.mdh.gov.br/mdh/ligue180>>. Acesso em: 03 abr. 2019. Citado na página 13.

MULHER, d. F. e. d. D. H. Ministério da. Ligue 180 recebe quase 18 mil denúncias nos dois primeiros meses do ano. 2019. Disponível em: <<https://www.mdh.gov.br/todas-as-noticias/2019/marco/ligue-180-recebe-quase-18-mil-denuncias-nos-dois-primeiros-meses-do-ano>>. Acesso em: 03 abr. 2019. Citado na página 12.

MULHER, d. F. e. d. D. H. Ministério da. Balanço anual: Ligue 180 registra 1,3 milhão de ligações em 2019. Ago. 2020. Disponível em: <<https://www.gov.br/mdh/pt-br/assuntos/noticias/2020-2/maio/balanco-anual-ligue-180-registra-1-3-milhao-de-ligacoes-em-2019>>. Acesso em: 07 jul. 2020. Citado na página 19.

MULHER, d. F. e. d. D. H. Ministério da. Relatório 2019. 2020. Disponível em: <<https://www.gov.br/mdh/pt-br/assuntos/noticias/2020-2/maio/balanco-anual-ligue-180-registra-1-3-milhao-de-ligacoes-em-2019>>. Acesso em: 07 jul. 2020. Citado 2 vezes nas páginas 16 e 19.

MULHERES, S. E. de Políticas para as. Lei maria da penha. *Lei Maria da Penha - Lei 11340/06*, 2006. Disponível em: <<https://presrepublica.jusbrasil.com.br/legislacao/95552/lei-maria-da-penha-lei-11340-06>>. Acesso em: 03 abr. 2019. Citado 2 vezes nas páginas 12 e 16.

ND, R. Blogueira vítima de estupro em beach club divulga resultado de laudo pericial. 2019. Disponível em: <<https://ndmais.com.br/seguranca/blogueira-vitima-de-estupro-em-beach-club-divulga-resultado-de-laudo-pericial/>>. Acesso em: 20 set. 2020. Citado na página 65.

NETO, J. L. et al. *Document Clustering and Text Summarization*. 2000. Citado na página 31.

PAVANELI, A.; GIMENES, E. Advogada é encontrada morta após queda de 4º andar de prédio em guarapuava, diz pm. 2018. Disponível em: <<https://g1.globo.com/pr/campos-gerais-sul/noticia/2018/07/22/advogada-e-encontrada-morta-em-predio-no-centro-de-guarapuava-diz-pm.ghtml>>. Acesso em: 20 set. 2020. Citado na página 67.

PENHA, I. M. da. Relógios da violência. 2019. Disponível em: <<https://www.relogiosdaviolencia.com.br/#>>. Acesso em: 03 abr. 2019. Citado na página 13.

PENHA, I. M. da. Tipos de violência. 2019. Disponível em: <<http://www.institutomariadapenha.org.br/lei-11340/tipos-de-violencia.html>>. Acesso em: 28 ago. 2019. Citado 2 vezes nas páginas 17 e 18.

- PINHO, M. Inquérito que apurou culpa de policiais no caso eloá está arquivado. 2012. Disponível em: <<http://g1.globo.com/sao-paulo/noticia/2012/02/inquerito-que-apurou-culpa-de-policiais-no-caso-elo-a-esta-arquivado.html>>. Acesso em: 20 set. 2020. Citado na página 64.
- QUEM, R. Victor chaves ironiza acusação de agressão à ex grávida e briga com o irmão leo. 2019. Disponível em: <<https://revistaquem.globo.com/QUEM-News/noticia/2019/02/victor-chaves-ironiza-acusacao-de-agressao-ex-gravida-e-briga-com-o-irmao-leo.html>>. Acesso em: 20 set. 2020. Citado na página 67.
- REZENDE, G. *Mais de 10 anos após morte, mãe de Eliza Samúdio ainda espera achar corpo: 'Penso várias vezes ao dia'*. 2020. G1. Disponível em: <<https://glo.bo/3kvB0P8>>. Acesso em: 31 ago. 2020. Citado na página 63.
- RODRIGUES, L. D. F.; JUNIOR, J. L. F. da S.; LOBATO, F. M. F. A culpa e dela! e isso o que dizem nos comentários da sncias sobre a tentativa de feminicídio de elaine caparroz. 2019. Disponível em: <<https://sol.sbc.org.br/index.php/brasnam/article/view/6547>>. Acesso em: 28 ago. 2019. Citado 5 vezes nas páginas 13, 14, 25, 32 e 58.
- ROUSSEFF, D. Lei nº 13.104. 2015. Disponível em: <http://www.planalto.gov.br/Civil_03/_Ato2015-2018/2015/Lei/L13104.htm>. Acesso em: 03 abr. 2019. Citado 2 vezes nas páginas 12 e 16.
- SAÚDE, O. M. da. Oms aborda consequências da violência sexual para saúde das mulheres. 2018. Disponível em: <<https://nacoesunidas.org/oms-aborda-consequencias-da-violencia-sexual-para-saude-das-mulheres/>>. Acesso em: 28 ago. 2019. Citado na página 17.
- SILVA, G. M. e; SILVA, N. F. F. da; DIAS, M. de S. Detecção de bullying: Como identificar automaticamente essa prática em redes sociais? *Revista de Sistemas de Informação da FSMA*, Jun. 2018. Disponível em: <<http://oaji.net/articles/2017/601-1530795595.pdf>>. Acesso em: 03 abr. 2019. Citado 3 vezes nas páginas 21, 26 e 34.
- SILVA, W. et al. Identificando emoções em redes sociais: Um estudo de caso no facebook. 2014. Citado na página 21.
- SIQUEIRA, A. *Justiça nega liberdade a lutador que espancou paisagista por quatro horas*. 2020. Veja. Disponível em: <<https://veja.abril.com.br/brasil/justica-nega-liberdade-a-lutador-que-espancou-paisagista-por-quatro-horas/>>. Acesso em: 23 mar. 2020. Citado na página 61.
- SMITH, T. C.; FRANK, E. *Statistical Genomics: Methods and Protocols*. New York, NY: Springer, 2016. 353–378 p. Disponível em: <http://dx.doi.org/10.1007/978-1-4939-3578-9_17>. Citado na página 24.
- TERRITÓRIOS, T. de Justiça do Distrito Federal e dos. Tjdft divulga vencedores do prêmio maria da penha vai à escola. 2019. Disponível em: <<https://www.tjdft.jus.br/institucional/imprensa/noticias/2019/julho/conheca-os-vencedores-do-premio-maria-da-penha-vai-a-escola>>. Acesso em: 28 ago. 2019. Citado na página 21.

- UCHÔA, A. *Ex acusa goleiro do Flamengo de sequestro e ameaça, e Bruno nega*. 2009. G1. Disponível em: <<http://g1.globo.com/Noticias/Rio/0,,MUL1340770-5606,00-EX+ACUSA+GOLEIRO+DO+FLAMENGO+DE+SEQUESTRO+E+AMEACA+E+BRUNO+NEGA.html>>. Acesso em: 31 ago. 2020. Citado na página 61.
- VALE, J. H. do. Cantor victor vira réu em processo que apura agressão contra a mulher em bh. 2017. Disponível em: <https://www.em.com.br/app/noticia/gerais/2017/04/10/interna_gerais,861184/cantor-victor-vira-reu-em-processo-que-apura-agressao-contra-a-mulher.shtml>. Acesso em: 20 set. 2020. Citado na página 67.
- VASCONCELOS, L. M. R. de; CARVALHO, C. L. de. Aplicação de regras de associação para mineração de dados na web. 2018. Disponível em: <<https://telematicafactal.com.br/revista/index.php/telfract/article/view/8/5>>. Acesso em: 03 jan. 2020. Citado na página 13.
- WALKER, L. The battered woman. 1987. Disponível em: <<https://www.nlm.nih.gov/exhibition/confrontingviolence/materials/OB11101.pdf>>. Acesso em: 28 ago. 2019. Citado na página 18.
- WIKIPEDIA. Caso eloá cristina. 2008. Disponível em: <<https://bit.ly/2FP12Oa>>. Acesso em: 20 set. 2020. Citado na página 64.
- WITTEN, I. et al. Data mining: Practical machine learning tools and techniques. 2016. Citado na página 35.
- WIZIACK, J. ?a internet não é uma terra sem lei? 2006. Disponível em: <https://istoe.com.br/5032_ainternetnaoematerrasemlei/>. Acesso em: 30 jun. 2020. Citado na página 13.