

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

---

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

Isabella Tannús Corrêa

Orientadora: Profa. Dra. Elaine Ribeiro de Faria Paiva

---

---

# 01

## Introdução

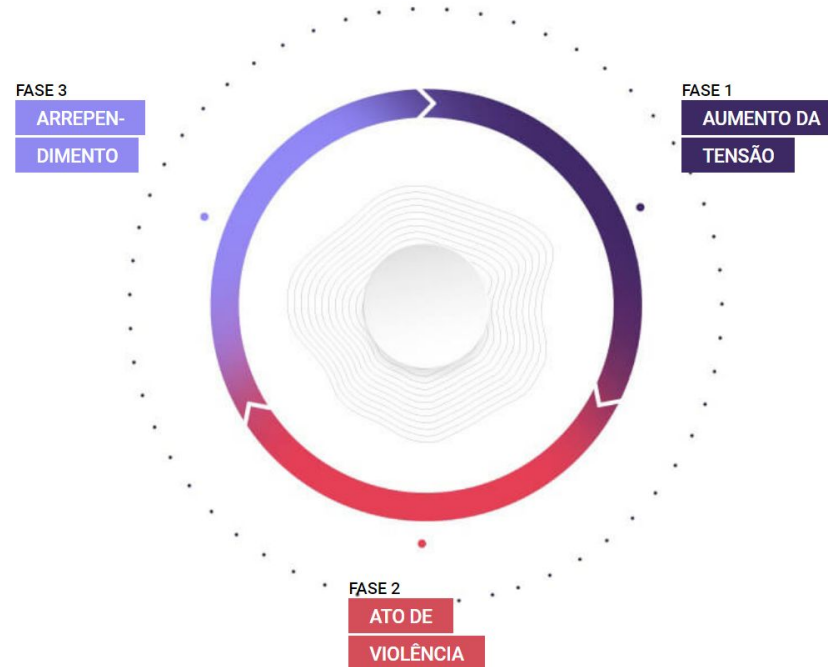
Uma visão geral sobre violência contra  
mulher e Twitter

---

# Violência contra a mulher

- A violência contra as mulheres começou a ser vista como um problema público somente a partir de 1970, com o crescimento da antropologia e noções de feminismo
- Lei Maria da Penha
- Lei do Feminicídio
- Tipos de violência: violência sexual, moral, psicológica, física, patrimonial

# Violência contra a mulher



Fonte: <https://www.institutomariadapenha.org.br/violencia-domestica/ciclo-da-violencia.html>

# Casos de violência contra mulher e feminicídio

## ▲ Caso Elaine Caparroz

A paisagista Elaine 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.

## ▲ Caso 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.

## ▲ Caso 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.

## ▲ Caso 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.

# Casos de violência contra mulher e feminicídio

## ▲ Caso Mariana Ferrer

A influenciadora digital Mariana Ferrer acusa o empresário André Aranha de estupro em um beach club em Florianópolis. O fato ocorreu em dezembro de 2018.

## ▲ Caso 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.

## ▲ Caso 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 Manvailer, marido da vítima.

# Twitter

- Fonte que gera grandes volumes de dados a todo momento
- Usuários geram conteúdo e expressam suas opiniões sobre notícias no mundo inteiro

---

# 02

## Motivação

Apresentação do tema deste estudo

---



# 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.
- Em 2019, o Ligue 180 registrou 85.412 relatos de violência contra a mulher.
- 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.

---

# 03

## Objetivos

Objetivos gerais e específicos

---

# Objetivo Geral

Verificar o sentimento das postagens de usuários no Twitter sobre casos de violência contra a mulher que tiveram grande repercussão na mídia desde 2008.

## Objetivos específicos

- Construção de uma base de dados para cada caso de violência contra mulher e feminicídio construída a partir de postagens no Twitter
- Sumarizar os tweets sobre cada caso por meio de técnicas como nuvem de palavras e LDA
- 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

---

# 04

## Método Proposto

Exposição das etapas do  
desenvolvimento

---

# Etapas da Mineração de Textos

## Coleta de dados

- Twitter utilizado  
como fonte de dados

Casos	Quantidade de tweets
Caso Elaine Caparroz	1.676
Caso Eliza Samúdio	143.946
Caso Eloá Cristina	22.995
Caso Eva Luana	2.529
Caso Mariana Ferrer	3.794
Caso Poliana Bagatini	1.522
Caso Tatiane Spitzner	3.581

# Etapas da Mineração de Textos

## Pré-processamento dos dados

- Tem como finalidade descartar o que é considerado irrelevante à etapa de classificação.
- ▲ **Remoção de stopwords:** palavras que não possuem valor semântico.
- ▲ **Remoção de caracteres não alfabéticos, pontuação e acentuação:** não agregam valor à classificação.
- ▲ **Conversão de letras maiúsculas para minúsculas:** padronização.

# Etapas da Mineração de Textos

Pré-processamento dos dados

- ▲ **Substituição de gírias e abreviações:** garantir a semântica do tweet.
- ▲ **Junção de palavras compostas**
- ▲ **Remoção de tweets de reportagens:** não agregam valor à classificação.
- ▲ **Lematização**



# Etapas da Mineração de Textos

Busca por padrões: Nuvens de palavras e LDA

- Auxílio para entender sobre quais temas aparecem nos tweets.
- Nuvens de palavras: imagens compostas por palavras, que demonstram de maneira visual a frequência de ocorrência das palavras em um texto. Também foram usadas para validar a eficácia do pré-processamento.
- *Latent Dirichlet Allocation* (LDA): encontrar tópicos aos quais um documento pertence, com base nas palavras nele contidas.

# Etapas da Mineração de Textos

## Construção da base rotulada

- Amostra de 100 tweets de cada caso foram selecionados de forma aleatória, a partir da base de dados coletada originalmente.
- Primeiro momento: cada texto foi rotulado de acordo com a classe que se refere: vítima, caso ou agressor.
- Segundo momento: cada tweet foi julgado com “positivo”, “neutro” ou “negativo”.

	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

# Etapas da Mineração de Textos

## Classificador Naive Bayes

- Para aplicar o *Naive Bayes*, foi utilizado a ferramenta *Weka*.
- Os dados das bases rotuladas foram convertidos para o formato ARFF.
- Após o carregamento do conjunto de treinamento, foi aplicado o filtro *StringToWordVector*, que converte a string representando o tweet em um conjunto de atributos numéricos.
- Na aba de Classificação, foi escolhido o classificador Naive Bayes com a estratégia de validação 10-fold cross-validation.
- A acurácia foi utilizada para realizar a validação do classificador.

---

# 05

## Resultados

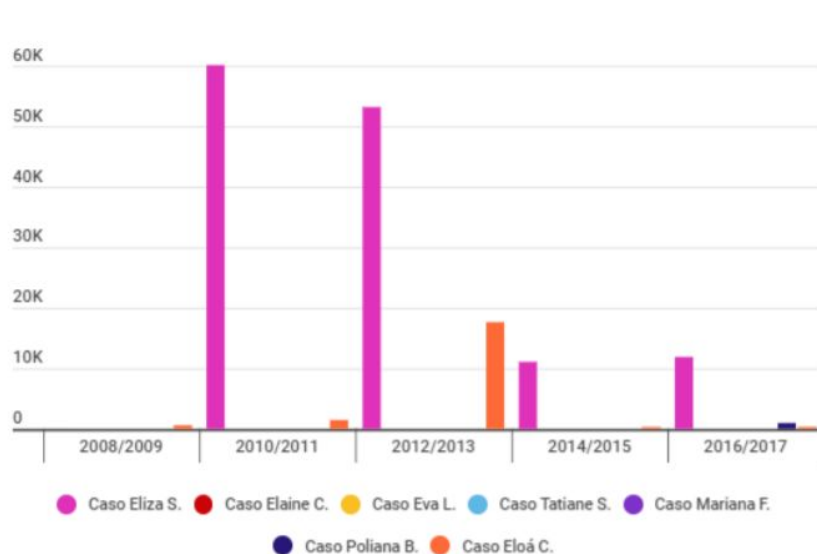
Dados obtidos a partir da Análise de  
Sentimentos são sumarizados e  
analisados

---

# Resultados

## Coleta de dados

- Total de 270.218 tweets



# Resultados

## Pré-processamento dos dados

Assunto	Quantidade de <i>tweets</i>		Diferença
	Antes do pré-processamento	Depois do pré-processamento	
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%

---

## Busca por padrões: Nuvens de palavras

- Validação do pré-processamento



### Caso Eloá Cristina



### Caso Eloá Cristina



## Busca por padrões: Nuvens de palavras

- Entendimento dos temas mais frequentes



## Caso Eliza Samúdio



## Caso Mariana Ferrer

# Resultados

Busca por padrões: LDA

- Modelagem de tópicos: 5 tópicos com 6 palavras cada

Caso Eliza Samúdio	
Tópicos	Termos
Crime	0.061*“matar” + 0.035*“falar” + 0.025*“polícia” + 0.023*“morto” + 0.018*“gente” + 0.018*“mulher”
Relacionamento	0.067*“ex” + 0.060*“morte” + 0.051*“flamengo” + 0.029*“amante” + 0.018*“história” + 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*“mãe” + 0.018*“viver” + 0.015*“desaparecimento” + 0.015*“ouvir” + 0.015*“vida”
Descredibilização da vítima	0.045*“corpo” + 0.044*“comer” + 0.034*“achar” + 0.029*“porno” + 0.025*“filme” + 0.024*“ver”

# Resultados

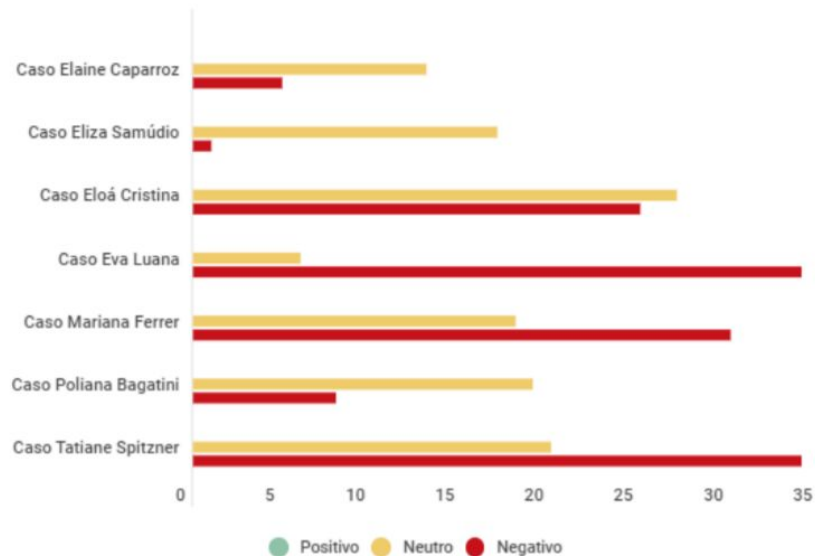
## Busca por padrões: LDA

- Modelagem de tópicos: 5 tópicos com 6 palavras cada

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”

# Resultados

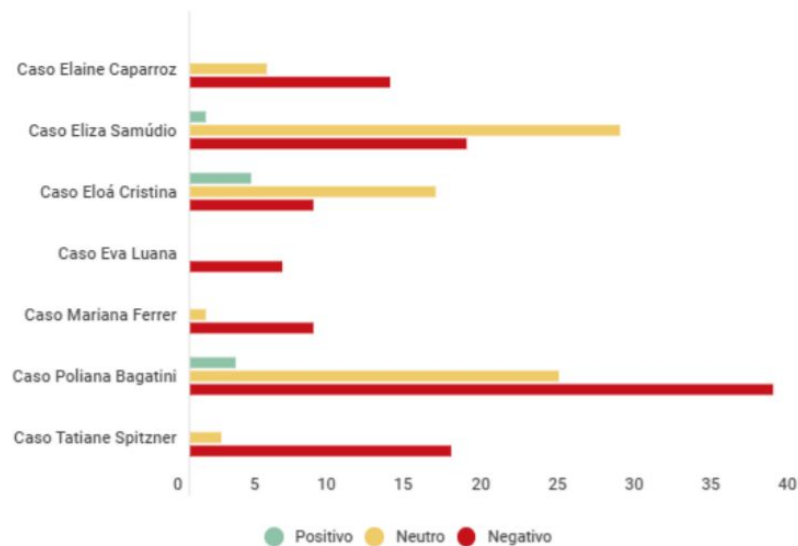
## Análise da base de dados rotulada manualmente



Relação aos casos

# Resultados

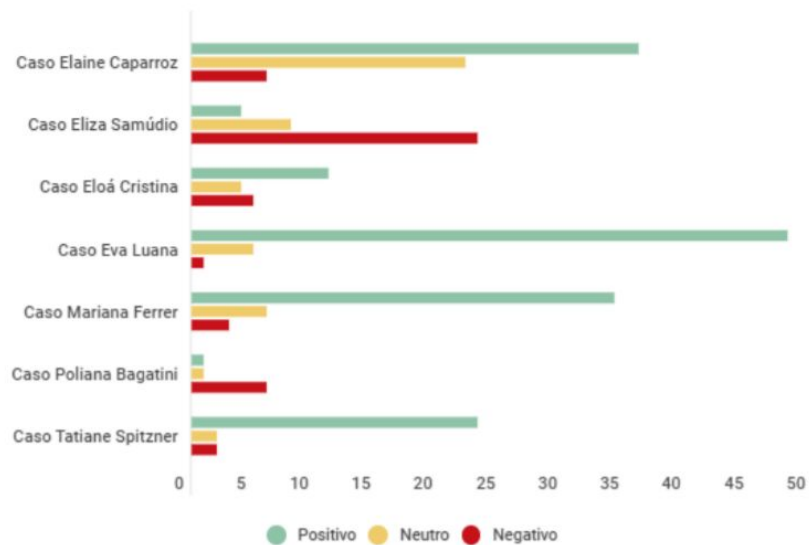
## Análise da base de dados rotulada manualmente



Relação aos agressores

# Resultados

## Análise da base de dados rotulada manualmente



Relação às vítimas

# Resultados

## Avaliação do classificador

- Matriz de confusão para o conjunto total de treinamento

Classificado como			
pos	neu	neg	
<b>96</b>	20	49	pos
37	<b>128</b>	79	neu
53	57	<b>181</b>	neg

# Resultados

## Avaliação do classificador

- Matriz de confusão para o conjunto sobre as vítimas

Classificado como			
pos	neu	neg	
<b>120</b>	22	15	pos
24	<b>12</b>	12	neu
16	2	<b>28</b>	neg

- Matriz de confusão para o conjunto sobre os casos

Classificado como			
pos	neu	neg	
<b>0</b>	0	0	pos
3	<b>61</b>	56	neu
2	29	<b>106</b>	neg

- Matriz de confusão para o conjunto sobre os agressores

Classificado como			
pos	neu	neg	
<b>2</b>	2	4	pos
4	<b>51</b>	21	neu
17	28	<b>63</b>	neg



# Resultados

## Avaliação do classificador

- Acurácia dos cenários

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

---

# 06

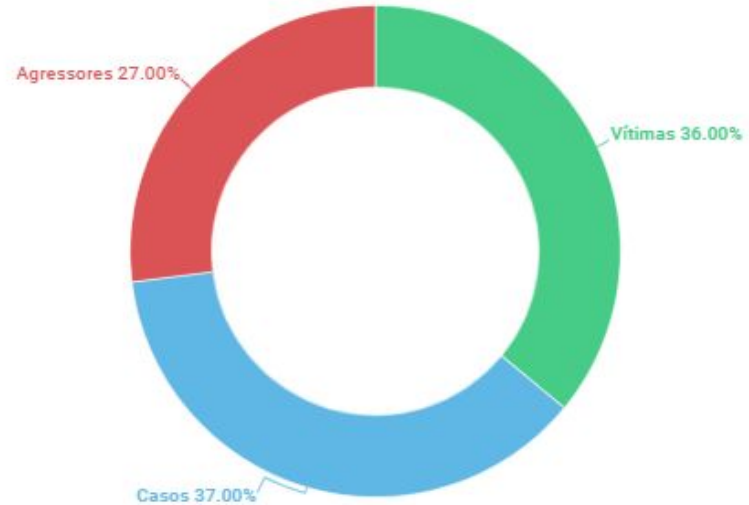
## Conclusões

Conclusões do estudo e trabalhos  
futuros

---

# Conclusões

- Dificuldade na classificação de tweets em um sentimento geral. Assim, a base de dados foi dividida em 3 categorias, agressor, vítima e caso.
- O resultado do classificador poderia ter sido melhor se houvesse um avanço no pré-processamento.
- Desenvolvimento de um classificador para separar os tweets de acordo com o agressor, vítima e caso, seria importante para o trabalho.



## Conclusões

- Comparação dos resultados com as conclusões de Rodrigues, Junior e Lobato (2019): é possível inferir que alguns usuários que fazem comentários negativos sobre as vítimas preferem permanecer anônimos.
- Em casos envolvendo celebridades, como os casos Eliza Samúdio e 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 na sociedade.

## Trabalhos futuros

- Construir uma base composta por comentários publicados nas reportagens dos casos em sites de notícias, além dos tweets.
- Refinamento do pré-processamento para que não apareçam palavras que se referem a um caso individualmente.
- Em novembro de 2020, o Caso Mariana Ferrer voltou a movimentar as mídias sociais. Assim, seria interessante analisar o sentimento dos usuários quanto ao caso atualmente, e comparar se condizem com a classificação encontrada antes das novas notícias sobre o mesmo.
- Avaliar outros classificadores para que tenham resultados mais favoráveis.

---

Obrigada!

---