



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**UNIDADE ACADÊMICA DE SERRA TALHADA**  
**BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

# **Comitê de Classificador para Mineração de Opinião de Eleitores Brasileiros**

**Por**

**Guilherme Henrique Pereira de Souza**

Serra Talhada,  
Junho/2019



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
UNIDADE ACADÊMICA DE SERRA TALHADA  
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**GUILHERME HENRIQUE PEREIRA DE SOUZA**

## **Comitê de Classificador para Mineração de Opinião de Eleitores Brasileiros**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação da Unidade Acadêmica de Serra Talhada da Universidade Federal Rural de Pernambuco como requisito parcial à obtenção do grau de Bacharel.

Orientador: Profa. Dra. Ellen Polliana Ramos Souza

Serra Talhada,  
Junho/2019

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema Integrado de Bibliotecas da UFRPE  
Biblioteca da UAST, Serra Talhada - PE, Brasil.

S729c Souza, Guilherme Henrique Pereira de  
Comitê de Classificador para Mineração de Opinião de Eleitores  
Brasileiros / Guilherme Henrique Pereira de Souza. – Serra Talhada,  
2019.  
48 f.: il.

Orientadora: Ellen Polliana Ramos Souza

Trabalho de Conclusão de Curso (Graduação em Bacharelado em Sistemas  
de Informação) – Universidade Federal Rural de Pernambuco.  
Unidade Acadêmica de Serra Talhada, 2019.  
Inclui referências e apêndices.

1. Redes sociais. 2. Twitter (Rede social on-line). 3. Monitorização  
eleitoral. I. Souza, Ellen Polliana Ramos, orient. II. Título.

CDD 004

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
UNIDADE ACADÊMICA DE SERRA TALHADA  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**GUILHERME HENRIQUE PEREIRA DE SOUZA**

**Comitê de Classificador para Mineração de Opinião de Eleitores Brasileiros**

Trabalho de Conclusão de Curso julgado adequado para obtenção do título de Bacharel em Sistemas de Informação, defendida e aprovada por unanimidade em 13/06/2019 pela banca examinadora.

Banca Examinadora:

---

Profa. Dra. Ellen Polliana Ramos Souza  
Orientador  
Universidade Federal Rural de Pernambuco

---

Arthur Diego de Godoy Barbosa  
Universidade Federal Rural de Pernambuco

---

Prof. Dr. Sérgio de Sá Leitão Paiva Junior  
Universidade Federal Rural de Pernambuco

*“I used to be an adventurer like you,  
then I took an arrow in the knee”  
(The Elder Scrolls V: Skyrim)*

## RESUMO

O crescimento recente e exponencial das mídias sociais e do conteúdo gerado pelo usuário (CGU) na Internet fornece uma grande quantidade de dados que permite descobrir os sentimentos, as opiniões e as experiências de usuários e clientes. Dentre as diversas mídias sociais existentes atualmente, destaca-se o Twitter, possuindo mais de 336 milhões de usuários mensais ativos. Compreender o que as pessoas estão pensando ou suas opiniões é fundamental para a tomada de decisão, principalmente no contexto no qual essas pessoas exprimem seus comentários voluntariamente. Para esse tipo de análise, utiliza-se a mineração de opinião, também conhecida como análise de sentimento, que analisa, opiniões, sentimentos e emoções das pessoas acerca de entidades como serviços, produtos, marcas, indivíduos, problemas, eventos e tópicos, podendo ser expressos de forma não estruturada, como em textos, por exemplo. A difusão das mídias sociais também levanta a possibilidade de se aprofundar na web para explorar e rastrear as preferências políticas e eleitorais dos cidadãos. Neste sentido, estudos recentes exploram as mídias sociais como um dispositivo para avaliar a popularidade dos políticos durante seu percurso eleitoral. Analisar as mídias sociais durante uma campanha eleitoral pode ser um suplemento ou complemento útil de pesquisas tradicionais e, dentre as razões, destacam-se: o baixo custo e a velocidade de análise em tempo real. Este trabalho tem como objetivo analisar a eficiência de um comitê de classificador para aplicações de mineração de opinião, utilizando dados extraídos de eleitores brasileiros. Neste estudo, foi desenvolvida uma aplicação de mineração de opinião a nível de documento utilizando comitês de classificador, com o objetivo de analisar a eficiência dos comitês em aplicações desse tipo. Para isso, utilizou-se dados extraídos do Twitter contendo opiniões de eleitores brasileiros. Foram avaliadas seis configurações de pré-processamento e três abordagens de comitê de classificador, sendo elas: Voting, Bagging e Boosting. Os resultados foram obtidos em termos de Acurácia. De acordo com os experimentos realizados, a configuração 3, que utiliza de remoção de stopwords, n-gram e TF-IDF alcançou os melhores resultados, com acurácia de 90,26% para o comitê que utilizava a abordagem Boosting.

**Palavras-chave:** Comitê de Classificador. Mineração de Opinião. Análise de Sentimento. Twitter.

# ABSTRACT

The recent and exponential growth of social media and user-generated content on the Internet, provide a wealth of data to uncover the feelings, opinions, and experiences of users and customers. Among the various social media currently available, Twitter stands out, with more than 336 million monthly active users. Understanding what people are thinking or their opinions is critical to decision making, especially in the context in which these people express their comments voluntarily. For this type of analysis, opinion mining is used, also known as sentiment analysis, which analyzes people's opinions, feelings and emotions about entities such as services, products, brands, individuals, problems, events and topics, be expressed in an unstructured form, as in texts, for example. The spread of social media also raises the possibility of delving deeper into the web to explore and track citizens' political and electoral preferences. In this sense, recent studies explore social media as a device to assess the popularity of politicians during the campaign. Analyzing social media during an election campaign can be a useful supplement or supplement to traditional surveys, and among the reasons include the low cost and speed of real-time analysis. This paper aims to analyze the efficiency of an ensemble for opinion mining applications, using data extracted from Brazilian voters. In this study, an application of opinion mining was developed at the document level using ensembles, with the objective of analyzing the efficiency of the ensembles in such applications. For that, data extracted from Twitter containing opinions of Brazilian voters were used. Six pre-processing configurations and three ensemble approaches were evaluated: Voting, Bagging and Boosting. The results were obtained in terms of Accuracy. According to the experiments, configuration 3, which uses stopwords, n-gram and TF-IDF, achieved the best results, with a 90.26% accuracy for the committee that used the Boosting approach.

**Keywords:** Ensemble. Opinion Mining. Sentiment Analysis. Twitter

## LISTA DE FIGURAS

Figura 2.1 – Tipos de Aprendizado . . . . .	18
Figura 2.2 – Representação das abordagens Bagging e Boosting . . . . .	23
Figura 3.1 – Método adotado no trabalho . . . . .	27

## LISTA DE QUADROS

Quadro 2.1 – Esboço do algoritmo . . . . .	21
Quadro 2.2 – Visão geral e comparação dos trabalhos relacionados com este trabalho. . .	26
Quadro 3.1 – Interpretação dos valores do coeficiente Kappa de Fleiss. . . . .	29
Quadro 3.2 – Exemplo do processo de Tokenização. . . . .	30
Quadro 3.3 – Exemplo do processo de Remoção de Stopwords. . . . .	31
Quadro 3.4 – Exemplo do processo de bigram. . . . .	31
Quadro 3.5 – Métricas utilizadas e suas fórmulas . . . . .	33
Quadro 4.1 – Configurações adotadas do trabalho. . . . .	35
Quadro 4.2 – Visão geral e comparação dos trabalhos relacionados com este trabalho. . .	41

## LISTA DE TABELAS

Tabela 3.1 – Mentions e Hashtags utilizadas para coleta. . . . .	28
Tabela 4.1 – Configuração da corpora . . . . .	36
Tabela 4.2 – Resultados obtidos para o corpus PT-BR desbalanceado . . . . .	37
Tabela 4.3 – Resultados obtidos para o corpus PT-BR desbalanceado com classificadores simples . . . . .	37
Tabela 4.4 – Resultados obtidos para o corpus PT-BR balanceado . . . . .	38
Tabela 4.5 – Resultados obtidos para o corpus EN desbalanceado . . . . .	39
Tabela A.1 – Exemplo de anotação manual. . . . .	48
Tabela A.2 – Quantidade de cada polaridade anotada. . . . .	48

## **LISTA DE ABREVIATURAS E SIGLAS**

AG	Algoritmo Genético
CC	Comitê de Classificador
HMM	Modelo Hidden Markov
MNB	Multinomial Naïve-Bayes
PA	Passive Agressive
PNL	Programação Neurolinguística
SVM	Support Vector Machines
RL	Regressão Logística

# SUMÁRIO

1	INTRODUÇÃO . . . . .	13
1.1	Motivação . . . . .	13
1.2	Justificativa . . . . .	14
1.3	Objetivos . . . . .	14
1.3.1	Objetivos Gerais . . . . .	14
1.3.2	Objetivos Específicos . . . . .	15
1.4	Organização do Trabalho . . . . .	15
2	FUNDAMENTAÇÃO TEÓRICA . . . . .	16
2.1	Mineração de Opinião . . . . .	16
2.2	Aprendizagem de Máquina . . . . .	17
2.2.1	Algoritmos Supervisionados . . . . .	18
2.2.1.1	Multinomial Naïve-Bayes (MNB) . . . . .	18
2.2.1.2	Support Vector Machines (SVM) . . . . .	19
2.2.1.3	Regressão Logística (RL) . . . . .	20
2.2.1.4	Passive Agressive (PA) . . . . .	20
2.3	Comitê de Classificador (CC) . . . . .	21
2.4	Trabalhos Relacionados . . . . .	23
2.4.1	Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach . . . . .	23
2.4.2	Analysis of Bagged Ensemble Classifiers for Blogger Data . . . . .	24
2.4.3	Sentiment Analysis of Tweets Using Machine Learning Approach . . . . .	25
2.4.4	Comparativo . . . . .	25
3	MÉTODO . . . . .	27
3.1	Coleta . . . . .	27
3.1.1	Anotação Manual . . . . .	28
3.2	Pré-processamento . . . . .	29
3.2.1	Tokenização . . . . .	30
3.2.2	Filtragem . . . . .	30
3.2.3	Remoção de Stopwords . . . . .	30

3.2.4	Modelo N-Gram	31
3.2.5	TF-IDF	32
3.3	Processamento	32
3.3.1	Comitê de Classificador	33
3.4	Avaliação	33
4	RESULTADOS E DISCUSSÃO	35
4.1	Configuração do Experimento	35
4.2	Resultados do Experimento	36
4.2.1	Corpus em Português do Brasil (PT-BR)	36
4.2.2	Corpus em Inglês (EN)	39
4.3	Discussão	39
5	CONCLUSÃO	42
5.1	Contribuições deste trabalho	43
5.2	Proposta para trabalhos futuros	43
5.3	Limitações e Ameaças	44
	REFERÊNCIAS BIBLIOGRÁFICAS	45
	APÊNDICE A – ANOTAÇÃO MANUAL	48

# 1 Introdução

## 1.1 Motivação

O crescimento recente e exponencial das mídias sociais e do conteúdo gerado pelo usuário (CGU) na Internet fornece uma grande quantidade de dados que permite descobrir os sentimentos, as opiniões e as experiências de usuários e clientes. O volume desse tipo de dado tem crescido de terabytes a petabytes (MARINE-ROIG; CLAVÉ, 2015)

Como as mídias sociais são ricas fontes de informação em tempo real, muitas entidades, como empresas, têm demonstrado interesse em saber as opiniões das pessoas acerca de produtos e serviços (SOUZA et al., 2016a). Dentre as diversas mídias sociais existentes atualmente, destaca-se o Twitter, possuindo mais de 336 milhões de usuários mensais ativos<sup>1</sup>. As mensagens do Twitter, chamadas *tweets*, têm um limite de 280 caracteres, tornando seu conteúdo mais objetivo.

Compreender o que as pessoas estão pensando ou suas opiniões é fundamental para a tomada de decisão, principalmente no contexto no qual estas pessoas exprimem seus comentários voluntariamente (ALVES et al., 2014). Entretanto, é impossível para o ser humano compreender completamente CGU em uma quantidade razoável de tempo, o que despertou um interesse na comunidade científica por sistemas capazes de extrair informações desse tipo de dado de forma automática (BALAZS; VELÁSQUEZ, 2016).

A difusão das mídias sociais também levanta a possibilidade de se aprofundar na web para explorar e rastrear as preferências políticas e eleitorais dos cidadãos. Neste sentido, estudos recentes exploram as mídias sociais como um dispositivo para avaliar a popularidade dos políticos durante a campanha (CERON; CURINI; IACUS, 2015). Analisar as mídias sociais durante uma campanha eleitoral pode ser um suplemento ou complemento útil de pesquisas tradicionais por várias razões (JIN et al., 2010).

Para esse tipo de abordagem utiliza-se a mineração de opinião, também conhecida como análise de sentimento, que, segundo Liu e Zhang (2012), é a área de estudo que analisa,

<sup>1</sup> Disponível em: <http://idgnow.com.br/internet/2018/04/25/twitter-cresce-base-de-usuarios-e-reporta-lucro-pela-segunda-vez-consecutiva/>. Acesso em 17 de Dezembro de 2018

opiniões, sentimentos e emoções das pessoas acerca de entidades como serviços, produtos, marcas, indivíduos, problemas, eventos e tópicos, podendo ser expressos de forma não estruturada, como em textos, por exemplo. Essa análise é feita através da classificação de opinião de um documento, sentença ou característica em categorias, tais como: positiva, negativa ou neutra. Esse tipo de classificação é referida na literatura como “polaridade de sentimento” ou “classificação de polaridade” (VITÓRIO; SOUZA, 2017).

Além de ser mais barata e mais rápida em comparação com pesquisas tradicionais, uma análise de mídia social permite monitorar uma campanha eleitoral em tempo real. Com isso, a possibilidade de deixar a campanha em dia, ou seja, rastrear tendências em tempo real e captar qualquer mudança repentina nas opiniões coletadas (CERON; CURINI; IACUS, 2015).

## 1.2 Justificativa

Atualmente, a mineração de opinião tem tornado-se mais importante, pois fornece aos tomadores de decisão a estimativa do sucesso de novas técnicas propostas, novas campanhas publicitárias ou lançamento de novos produtos (KALARANI; BRUNDA, 2017).

Para Castro et al. (2017), a utilização de um comitê de classificador, oferece bons resultados para construção de uma aplicação de identificação de língua. Assim como Kumar e Singh (2018) que buscaram mostrar os benefícios dos comitês utilizando três tipos de abordagens e obtendo ótimos resultados. Segundo Kalarani e Brunda (2017), os resultados experimentais mostram que o conjunto de técnicas na construção de um comitê de classificador fornecem altas precisões.

## 1.3 Objetivos

### 1.3.1 Objetivos Gerais

Este trabalho tem como objetivo geral analisar a eficiência de um comitê de classificador para mineração de opinião de eleitores brasileiros.

### 1.3.2 Objetivos Específicos

Foram estabelecidos os seguintes objetivos específicos para atingir o objetivo geral deste trabalho:

- Construir corpus para análise de opiniões de eleitores brasileiros;
- Realizar análise comparativa de técnicas de pré-processamento e processamento de texto;
- Construir e validar aplicação de mineração de opinião de texto utilizando comitê de classificador.

## 1.4 Organização do Trabalho

O restante deste trabalho está estruturado da seguinte forma:

- O Capítulo 2 apresenta o referencial teórico utilizado no trabalho, explicando a mineração de opinião, algoritmos de aprendizado supervisionado, os tipos de comitês, além dos trabalhos relacionados;
- No Capítulo 3, o método utilizado neste trabalho é apresentado, composto por quatro etapas, são elas: coleta, pré-processamento, processamento, avaliação;
- O Capítulo 4 apresenta os resultados alcançados e uma discussão acerca deles, além da configuração do experimento;
- Por fim, o Capítulo 5 apresenta as considerações finais deste trabalho, sobre as suas contribuições e apresenta propostas para trabalhos futuros.

## 2 Fundamentação Teórica

Neste Capítulo, foram abordados os conceitos de Mineração de Opinião, Aprendizagem de Máquina, Comitê de Classificador e por fim os Trabalhos Relacionados, que contribuíram para realização deste trabalho. No final do Capítulo foi realizado um comparativo entre os Trabalhos Relacionados e Este Trabalho.

### 2.1 Mineração de Opinião

Mineração de opinião, também conhecida na literatura como análise de sentimento, é considerada uma atividade de classificação de texto, já que discrimina a orientação de um sentimento, ou opinião, de um determinado texto em duas ou mais classes (RAVI; RAVI, 2015a). No que se diz respeito à quantidade de classes, ela tem sido realizada de várias formas: binária, ternária, n-ária na forma de estrelas, etc.(RAVI; RAVI, 2015b). Um tipo comum de classificação de texto é a classificação de rótulo único, na qual cada texto pertence a exatamente uma classe ou categoria (CARDOSO-CACHOPO; OLIVEIRA, 2007). Pesquisadores também têm levado em consideração vários tipos de emoção, como, por exemplo, as seis emoções universais: raiva, desgosto, medo, felicidade, tristeza e surpresa (PANG; LEE, 2008). Na mineração de opinião, as opiniões, são classificadas em categorias como, por exemplo, “positiva”, “negativa” ou “neutra”. Esse tipo de classificação também é referida na literatura como “polaridade de sentimento” ou “classificação de polaridade”.

Mineração de opinião costuma ser realizada em três diferentes níveis de análise: documento, sentença ou aspecto (FELDMAN, 2013). A classificação a nível de documento determina se a opinião de todo o documento é positiva, negativa ou neutra, por exemplo. A classificação a nível de sentença determina se uma determinada sentença apresenta uma opinião positiva, negativa ou neutra. Enquanto a nível de aspecto, ou característica, foca em todas as expressões de sentimentos presentes em um documento e o aspecto que elas se referem.

Revisões da literatura sugerem que a maioria das aplicações de mineração de opinião pode ser classificada em quatro categorias distintas: avaliações de produtos, avaliações de filmes, extração de orientação política e predições do mercado de ações (RAVI; RAVI, 2015b). Facebook

e Twitter são importantes fontes de opiniões, porém o primeiro é menos utilizado em mineração de texto porque geralmente contém dados não-textuais, como imagens, e a análise apenas do texto acaba não sendo efetiva (PAK; PAROUBEK, 2010).

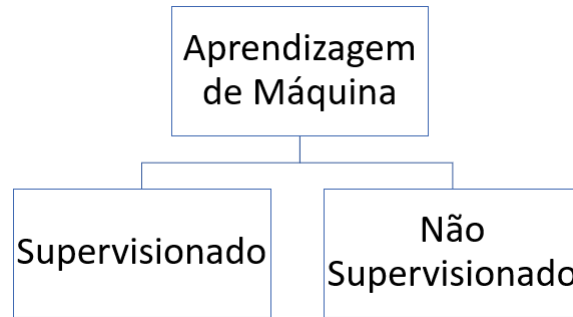
As abordagens de mineração de opinião mais frequentemente utilizadas são as de aprendizado de máquina supervisionado e as baseadas em léxicos, mas também há abordagens híbridas, que usam tanto aprendizado de máquina quanto léxicos (PANG; LEE, 2008; RAVI; RAVI, 2015b). Os métodos de aprendizado de máquina supervisionado aplicam algoritmos de classificação para aprender padrões subjacentes a partir de dados de exemplo com o objetivo de classificar novos dados não rotulados (BALAZS; VELÁSQUEZ, 2016). Para esses métodos, são necessários dois conjuntos de dados anotados: um para treinamento e outro para teste. Um bom número de algoritmos de aprendizado de máquina supervisionado têm sido utilizado para classificar opiniões. Dentre eles, se destacam o Naïve-Bayes e Support Vector Machines (SVM), que têm alcançado bons resultados com mineração de opinião (RAVI; RAVI, 2015b; SOUZA et al., 2016b).

A abordagem baseada em léxicos, também conhecida como semântica ou simbólica, faz uso de palavras de opinião positivas, usadas para expressar estados desejados, e palavras de opinião negativas, usadas para expressar estados indesejados (MEDHAT; HASSAN; KORASHY, 2014). Três abordagens principais são utilizadas para construir léxicos de opinião: abordagem manual, embora consuma muito tempo; baseada em dicionários, onde um conjunto inicial, construído manualmente, é incrementado procurando os seus antônimos e sinônimos em corpora; e baseado em corpus, que começa com uma lista de semente de palavras de opinião para encontrar outras palavras de opinião em um grande corpus com orientações específicas de contexto (VITÓRIO; SOUZA, 2017).

## 2.2 Aprendizagem de Máquina

De modo geral, as técnicas de aprendizado de máquina podem ser divididas em dois tipos: aprendizado supervisionado e aprendizado não supervisionado (TAN et al., 2007), como mostrado na Figura 2.1, ilustra os tipos de aprendizado de máquina:

**Figura 2.1 – Tipos de Aprendizado**



Elaborada pelo autor (2019)

Os métodos de aprendizado de máquina supervisionado aplicam algoritmos de classificação para aprender padrões subjacentes a partir de dados de exemplo com o objetivo de classificar novos dados não rotulados (BALAZS; VELÁSQUEZ, 2016). Para esses métodos, são necessários dois conjuntos de dados anotados: um para treinamento e outro para teste. Um bom número de algoritmos de aprendizado de máquina supervisionado têm sido utilizados para classificar opiniões. Dentre eles, se destacam o MNB e o SVM, que têm alcançado grande sucesso com mineração de opinião (RAVI; RAVI, 2015a; SOUZA et al., 2016b).

Entretanto, se o algoritmo não recebe dados rotulados na etapa de treinamento, o aprendizado é não-supervisionado. Geralmente usam-se algoritmos de clusterização para identificar as classes dos dados, já que essas também não são providas ao algoritmo (HAN; PEI; KAMBER, 2011).

## 2.2.1 Algoritmos Supervisionados

### 2.2.1.1 Multinomial Naïve-Bayes (MNB)

O Multinomial Naïve-Bayes é um classificador probabilístico baseado no teorema de Bayes, que usa a probabilidade condicional para classificar os dados em classes predeterminadas. Essa abordagem é chamada naïve (ingênua) porque assume a independência entre os vários valores dos atributos (KUMARI, 2014).

O MNB usa a informação da frequência de palavras nos documentos para análise. Cada documento é considerado um conjunto ordenado de palavras obtido a partir de um vocabulário  $V$ . A probabilidade de um evento de palavra é independente do contexto da palavra e de sua posição no documento (MCCALLUM; NIGAM et al., 1998).

Também é assumido que o comprimento dos documentos é independente de classe e cada documento  $d_i$  é desenhado a partir de uma distribuição multinomial de palavras com tantas tentativas independentes quanto o tamanho do documento  $d_i$ . Então, sendo  $N_{it}$  a contagem do número de vezes que a palavra  $W_t$  aparece no documento  $d_i$ , a probabilidade de um dado documento para sua classe é dada pela Equação 2.1:

$$P(d_i|c_j; \theta) = P(|d_i|)|d_i|! \prod_{t=1}^{|V|} \frac{P(W_t|c_j; \theta)^{N_{it}}}{N_{it}!}, \quad (2.1)$$

onde  $P(d_i|c_j; \theta)$  é a probabilidade de  $d_i$  pertencer a classe  $c_j$ ,  $P(|d_i|)$  é a probabilidade do documento  $d_i$  e  $P(W_t|c_j; \theta)$  é a probabilidade da ocorrência da palavra  $w_t$  na classe  $c_j$  (MCCALLUM; NIGAM et al., 1998).

### 2.2.1.2 Support Vector Machines (SVM)

Este método foi desenvolvido por Cortes e Vapnik (1995) e utiliza hiperplanos como limites de decisão. O classificador foi desenvolvido para classificações binárias usando um hiperplano de separação ótima entre as duas classes pela maximização da margem entre os pontos mais próximos de cada classe. Um hiperplano é uma superfície no espaço com várias dimensões, que é separado em duas metades de espaço. Uma vez que o SVM foi treinado, ele está apto a avaliar novas entradas relativas ao hiperplano divisor e classificá-las em uma categoria (VITÓRIO; SOUZA, 2017).

Dado um conjunto de treino com pares de dados anotados  $(x_i, y_i)$ , onde  $i = \{1, 2, 3, \dots, l\}$ ,  $x_i \in R^n$  e  $y_i \in \{1, -1\}^l$ , o SVM precisa resolver o problema de otimização presente na Equação 2.2:

$$\begin{aligned} \min_{x, b, \xi} \quad & \frac{1}{2} W^T W + C \sum_{i=1}^l \xi_i \\ \text{sujeito a} \quad & y_i (W_t \phi(X_i) + b) \geq -\xi_i, \\ & \xi_i \geq 0, \end{aligned} \quad (2.2)$$

onde  $W$  é o parâmetro do peso atribuído às variáveis,  $\xi_i$  é a folga, ou correção de erro, adicionada e  $C$  é o fator de regularização (HSU et al., 2003). O objetivo do problema é minimizar  $\frac{1}{2} W^T W +$

$C \sum_{i=1}^l \xi_i$ , onde o valor de  $y_i(W_t \phi(X_i) + b)$  precisa ser maior ou igual a  $1 - \xi_i$  e o valor de  $\xi$  é bem pequeno.

A efetividade do classificador é definida pelo núcleo utilizado. O núcleo define a forma que os limites serão desenhados, ou seja, usando um núcleo linear, como o SVC Linear utilizado neste trabalho, por exemplo, os limites serão todos linhas retas (VITÓRIO; SOUZA, 2017).

### 2.2.1.3 Regressão Logística (RL)

O modelo de Regressão Logística é um classificador linear probabilístico que usa um hiperplano para descrever o conjunto de dados. O classificador é parametrizado por uma matriz de peso  $W$  e um vetor de viés  $b$ . A classificação é realizada projetando um vetor de entrada em um conjunto de hiperplanos, cada um deles correspondendo a uma classe na base de dados. A distância de uma entrada até um hiperplano reflete a probabilidade dessa entrada ser um membro da classe correspondente àquele hiperplano (DEEPLARNING.NET, 2018).

A probabilidade de um vetor de entrada  $x$  ser um membro da classe  $i$ , que, por sua vez, é um valor da variável estocástica  $Y$ , pode ser escrito como na Equação 2.3:

$$P(Y = i|x, W, b) = \text{softmax}_i(Wx + b) = \frac{e^{w_i x + b_i}}{\sum_j e^{w_j x + B_j}} \quad (2.3)$$

A saída do modelo, ou seja, a predição  $y_{pred}$ , será a classe que obtiver a máxima probabilidade (DEEPLARNING.NET, 2018), como descrito na Equação 2.4:

$$y_{pred} = \text{argmax}_i P(Y = i|x, W, b) \quad (2.4)$$

### 2.2.1.4 Passive Agressive (PA)

Segundo Cui, Mittal e Datar (2006), os algoritmos Passive-Aggressive (PA), foram introduzidos por Gzik, Jozsko e Pieniżek (2014), e fazem parte da família de algoritmos de aprendizado *online* baseados em margem para classificação binária. Os algoritmos de PA funcionam de forma semelhante ao SVM e podem ser vistos como uma versão *online* de um

SVM.

Algoritmos PA tentam encontrar um hiperplano que separa as instâncias em dois meios espaços. A margem de um exemplo é proporcional à distância do exemplo ao hiperplano. Ao cometer erros na previsão de exemplos, o algoritmo PA utiliza a margem para modificar o classificador atual. A atualização do classificador segue as restrições: o novo classificador deve estar muito próximo do atual (atualização passiva) enquanto atinge pelo menos uma margem unitária no exemplo mais recente (atualização agressiva) (CUI; MITTAL; DATAR, 2006). O Quadro 2.1 ilustra o algoritmo PA:

**Quadro 2.1 – Esboço do algoritmo**

<p>INITIALIZE: <math>w_1 = (0 \dots 0)</math> as parameters of the classifier</p> <p>For <math>t = 1, 2, \dots</math></p> <ul style="list-style-type: none"> <li>• receive instance: <math>x_t \in R^n</math></li> <li>• predict: <math>\hat{y}_t = \text{sign}(w_t \cdot x_t)</math></li> <li>• receive correct label: <math>y_t \in \{-1, +1\}</math></li> <li>• suffer loss: <math>l_t = \max\{0, 1 - y_t(w_t \cdot x_t)\}</math></li> <li>• update:           <ol style="list-style-type: none"> <li>1. set: <math>\tau_t = \frac{l_t}{\ x_t\ ^2}</math></li> <li>2. update: <math>w_{t+1} = w_t + \tau_t y_t x_t</math></li> </ol> </li> </ul>
---

**Fonte: Cui, Mittal e Datar (2006)**

## 2.3 Comitê de Classificador (CC)

Segundo Kumar e Singh (2018), comitê de classificador, do inglês, *ensemble*, são usados para combinar os resultados de vários classificadores base para finalmente obter o resultado final. A combinação de resultados dos classificadores base é o passo determinante para a construção de um eficiente comitê. A maneira como o algoritmo combina os resultados afeta o desempenho dos CC.

Pesquisas apontam que para projetar um algoritmo, que combine o resultado dos classificadores base, de tal maneira que a precisão do resultado final se torne superior a qualquer um de seus classificadores base. Bagging e Boosting e Voto Majoritário (Voting) são alguns dos algoritmos para combinar os resultados dos classificadores base (KUMAR; SINGH, 2018).

Voto Majoritário (Voting) é o conceito mais geral de CC. Votar por maioria é uma regra simples e comum na vida social humana para resolver vários tipos de conflitos. Na aprendizagem baseada em votação simples, a classe que recebe o máximo de votos dos classificadores base torna-se o resultado final do classificador de conjunto (KUMAR; SINGH, 2018).

No Bagging, os classificadores são treinados de forma independente por diferentes conjuntos de treinamento através do método de inicialização. Para construí-los é necessário montar  $k$  conjuntos de treinamento idênticos e replicar esses dados de treinamento de forma aleatória para construir  $k$  redes independentes por re-amostragem com reposição. Em seguida, deve-se agregar as  $k$  redes através de um método de combinação apropriada, tal como a maioria de votos (BREIMAN, 1996).

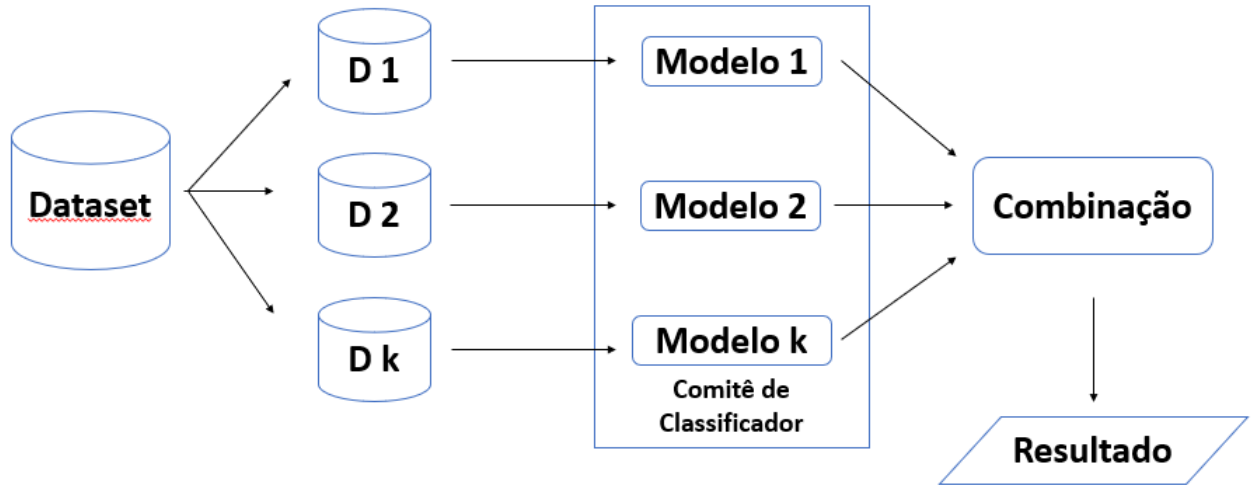
Para garantir que há amostras de treinamento suficientes em cada subconjunto, grandes porções de amostras (75-100%) são colocadas em cada subconjunto. Com isso, os subconjuntos individuais de formação se sobrepõem de forma significativa, com muitos casos fazendo parte da maioria dos subconjuntos e podendo até mesmo aparecer várias vezes num mesmo subconjunto. A fim de assegurar a diversidade de situações, um classificador de base relativamente instável é usado para que limites de decisão diferentes possam ser obtidos, considerando-se pequenas perturbações em diferentes amostras de treinamento (WANG et al., 2011).

Segundo Lantz (2013), o Boosting trabalha de forma semelhante ao Bagging, cada classificador é treinado usando um conjunto de treinamento diferente. A principal diferença em relação ao Bagging é que os conjuntos de dados re-amostrados são construídos especificamente para gerar aprendizados complementares e a importância do voto é ponderado com base no desempenho de cada modelo, em vez da atribuição de mesmo peso para todos os votos. Essencialmente, esse procedimento permite aumentar o desempenho de um limiar arbitrário simplesmente adicionando classificadores mais fracos. Dada a utilidade desse achado, Boosting é considerado uma das descobertas mais significativas em aprendizado de máquina (LANTZ, 2013).

Tanto para o Bagging e o Boosting, uma base de dados (*dataset*) é dividida em  $m$  partes, os conjuntos de treinamento, onde cada conjunto é atribuído a um dos  $m$  modelos de classificadores base presentes no comitê de classificador, para logo após, ser realizada uma forma de combinação dos resultados dos  $m$  modelos.

A Figura 2.2 ilustra as abordagens Bagging e Boosting:

Figura 2.2 – Representação das abordagens Bagging e Boosting



Fonte: Elaborada pelo Autor (2019)

## 2.4 Trabalhos Relacionados

Uma revisão da literatura foi realizada, foram encontrados vários trabalhos que tratassem de mineração de opinião e comitê de classificador. A pesquisa foi realizada nas bases *IEEE Xplore*, *Portal de Periódicos CAPES*, *Google Scholar*, *ACM Digital Library* e *ScienceDirect*, utilizando os termos de busca: (“ensemble” or “comitê de classificador”) and (“mineração de opinião” or “opinion mining”).

Sendo assim, os três trabalhos apresentados nesta seção, foram escolhidos pois se relacionam com este por se tratarem de aplicações de mineração de opinião utilizando comitê de classificador. Além disso, o trabalho de Jose e Chooralil (2016), por ser uma aplicação de mineração de opinião utilizando comitê de classificador, utiliza um corpus que possui opinião de eleitores.

### 2.4.1 Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach

Jose e Chooralil (2016) procuram reduzir os riscos de selecionar um classificador inadequado, combinando as saídas de um conjunto de classificadores. Nesse trabalho, foi utilizada uma abordagem que classifica automaticamente os sentimentos dos *tweets*, combinando classificadores de aprendizado de máquina com classificador léxico. Essa combinação foi realizada

entre os classificadores SentiWordNet, MNB e Modelo Hidden Markov (HMM).

Neste trabalho, adotaram um método com seis etapas, sendo elas: coleta, pré-processamento, classificação usando SentiWordNet, classificação usando MNB, classificação usando HMM e classificação usando comitê de classificador. Durante o pré-processamento, realizaram uma filtragem removendo: *URLs*, *Mentions*, *Hashtags* e caracteres especiais.

Jose e Chooralil (2016), extraíram 12.000 *tweets* sobre Arvind Kejriwal por 3 semanas durante os dias de eleição em Deli. Usaram três classificadores para uma classificação precisa dos dados políticos. Aqui a positividade/negatividade de cada *tweet* é determinada usando o princípio de votação majoritária no resultado desses três classificadores.

Foram utilizadas as técnicas de pré-processamento, a remoção de URLs, hashtags e mentions e o tratamento da palavra “não”, como é demonstrado no seu estudo. A polaridade dos *tweets* foi determinada usando o método de Voting. Para assim, encontrar os sentimentos políticos nos *tweets* em tempo real. Para essa abordagem de comitê foi obtida uma acurácia de 71.48%.

## 2.4.2 Analysis of Bagged Ensemble Classifiers for Blogger Data

Para Govindarajan (2016), *Blogs*, sites de resenhas, Twitter e outras redes sociais são as plataformas mais comuns usadas por pessoas e organizações para postar seus pontos de vista. Foi observado através deste estudo que uma maior precisão foi alcançada através da construção de um ensemble utilizando ensacamento com classificadores simples como SVM, MNB, Algoritmo Genético (AG).

Neste trabalho, a base de dados possui um domínio misto entre política, mídia e educação. O autor realiza uma avaliação comparativa no conjunto de dados do *blogger* que é comumente usado para mineração de sentimentos. Como também, realiza no pré-processamento, remoção de stopword, delimitação de sentenças e stemming. Os resultados experimentais exibem o melhor desempenho dos comitês de classificador propostos sobre os classificadores únicos. O melhor resultado obtido foi 81% de acurácia.

### 2.4.3 Sentiment Analysis of Tweets Using Machine Learning Approach

Para Rathi et al. (2018), mesclando mineração de dados com outros campos, como mineração de texto, PNL e inteligência computacional, pode-se classificar os *tweets* como bons, ruins ou neutros. A ênfase principal desta pesquisa está na classificação das emoções dos dados dos *tweets* coletados no Twitter.

Os autores adotaram no pré-processamento, conversão de todas as letras para minúsculas (lower case), ajustes/transformações de URLs, mentions e hashtag, stemming e tratamento de emoticons.

Para melhorar os resultados da classificação da análise de sentimentos, os autores utilizaram técnicas de aprendizado de máquinas para aumentar a eficiência e a confiabilidade da abordagem proposta. Sendo assim, fundindo SVM com Decision Tree e os resultados experimentais provam que nossa abordagem proposta está fornecendo melhores resultados de classificação em termos de f-medida e precisão em contraste com os classificadores individuais. Onde o melhor resultado obtido foi 84% de acurácia para o comitê que utiliza a abordagem Boosting.

### 2.4.4 Comparativo

Observando de forma geral, os trabalhos relacionados, apresentam pontos que podem contribuir para realização deste estudo. O comparativo mostrado nesta seção procura comparar as técnicas de pré-processamento e processamento utilizadas no Trabalhos Relacionados para realização deste trabalho.

Levando em consideração as comparações com este trabalho, diversas técnicas foram utilizadas, sendo elas: remoção de URLs, mention e hashtags, remoção de stopwords, lower case, SVM, MNB, Voting, Bagging e Boosting. Além destas técnicas, foram incluídas, o Modelo N-gram, TF-IDF, RL e PA. O conjunto destas técnicas resultou nas configurações deste trabalho para cumprir seu objetivo.

O Quadro 2.2 apresenta uma visão geral comparando os trabalhos relacionados apresentados nesse trabalho nessa seção:

**Quadro 2.2 – Visão geral e comparação dos trabalhos relacionados com este trabalho.**

Estudo	Pré-processamento	Algoritmos	Comitê	Domínio	Corpus	Resultados
Jose e Chooralil (2016)	remoção de URLs, hashtags e mentions	SentiWordNet, MNB e HMM	Voting	Política	<i>tweets</i>	71,48% de acurácia
Govindarajan (2016)	remoção de stopword, determinar o limite da sentença e stemming	SVM, MNB e AG	Bagging	Política, mídia e educação	<i>UCI Machine Learning Repository</i>	81,00% de acurácia
Rathi et al. (2018)	letras minúsculas, ajuste de URLs, hashtags e mentions, stemming e tratamento de emoticon	SVM e Decision Tree	AdaBoost	Variado	<i>Stanford Sentiment140, Polarity Dataset e University of Michigam</i>	84,00% de acurácia
Este Trabalho (Corpus Português)	Tokenização, lower case, remoção de URLs, hashtags, mentions e stopwords, modelo n-gram e TF-IDF	SVM, MNB, RL e PA	Voting, Bagging e Boosting	Política	<i>tweets</i>	90,26% de acurácia (Boosting)
Este Trabalho (Corpus Inglês)	Tokenização, lower case, remoção de URLs, hashtags, mentions e stopwords, modelo n-gram e TF-IDF	SVM, MNB, RL e PA	Voting, Bagging e Boosting	Política	<i>tweets</i>	77,94% de acurácia (Boosting)

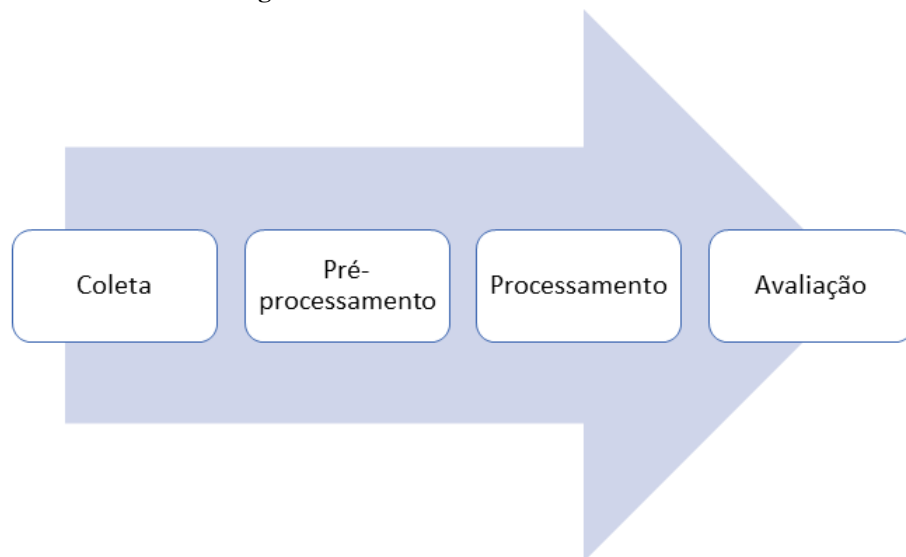
**Fonte: Elaborada pelo autor (2019)**

## 3 Método

Neste trabalho, foi realizada análise de sentimento a nível de documento, uma vez que é utilizada uma base de dados composta por *tweets*, que, apesar de curtos, podem conter mais de uma sentença.

A Figura 3.1 apresenta o método da aplicação de mineração de opinião proposta. Ele é dividido em quatro etapas: “coleta”, na qual os *tweets* foram extraídos e anotados; “pré-processamento”, na qual os *tweets* coletados foram estruturados antes de serem processados; “processamento”, na qual ocorreu a classificação dos *tweets* por parte dos algoritmos; e “avaliação”, responsável pela obtenção dos resultados das classificações. Nas seções subsequentes, cada uma das etapas é descrita com mais detalhes.

**Figura 3.1 – Método adotado no trabalho**



**Fonte: Elaborada pelo autor (2018)**

### 3.1 Coleta

O corpus utilizado para esse trabalho foi construído utilizando a biblioteca *Tweepy*, uma biblioteca Python para acessar a API do Twitter. Essa API pesquisa amostras dos dados publicados nos últimos sete dias (TWITTER, 2017).

O período selecionado para coleta dos *tweets* compreende os anos de 2017 e 2018, todo o período de pré-candidatura e campanha das possíveis candidatas para eleição de presidente. Para coleta dos *tweets*, a busca utilizada buscava por usuários específicos, *hashtags* ou palavras-chave. Como também, ignorando *retweets*, por se tratarem de *tweets* que foram publicados por um usuário e compartilhados por outros, logo, provavelmente já estando entre os *tweets* coletados.

A Tabela 3.1, apresenta as Mentions e Hashtags utilizadas para busca e coleta dos *tweets*:

**Tabela 3.1 – Mentions e Hashtags utilizadas para coleta.**

Mentions	Hashtags
@jairsolsonaro	#bozonaro
@cirogomes	#LulaLivre
@Haddad_Fernando	#LulaNaCadeia
@joaoamoedonovo	#EleNao
@CaboDaciolo	#EleSim
@MarinaSilva	#ForaTemer

**Fonte: Elaborado pelo autor (2019)**

### 3.1.1 Anotação Manual

Os *tweets*, após o processo de coleta, foram anotados manualmente como uma das três seguintes polaridades:

- “positiva”: *tweets* contendo opiniões ou sentimentos positivos;
- “negativa”: *tweets* contendo opiniões ou sentimentos negativos;
- “neutra”: *tweets* que não continha sentimentos nem opiniões, isto é, textos objetivos.

Sete anotadores participaram do processo. Onde foram organizadas três duplas para o processo de anotação. A anotação foi realizada em dois momentos: no primeiro deles, as duplas classificaram todos os 2670 *tweets*; no segundo momento, o autor deste trabalho classificou os 2670 *tweets*. Logo, cada tweet foi classificado por três anotadores diferentes.

A polaridade final de cada tweet foi definida como a polaridade atribuída a ele pela maioria dos três anotadores. No caso de discordância entre os três anotadores, isto é, nos

casos onde cada anotador classificou o tweet como uma polaridade diferente, o autor decidiu a polaridade final do *tweet*. O Apêndice A explica como foi realizada a anotação manual.

De acordo com Wiebe, Wilson e Cardie (2005) e Vitória et al. (2017), a capacidade humana para avaliação correta de subjetividade de um texto gira em torno de 72% a 85%, então cada *tweet* foi classificado por três anotadores como uma forma de minimizar classificações erradas. Para o corpus, foi calculado o coeficiente Kappa de Fleiss (FLEISS, 1971) com o objetivo de descobrir o nível de concordância entre os três anotadores. O corpus obteve um coeficiente de 72,89%.

Segundo o Quadro 3.1, adaptada de Seljan, Brkic e Vivic (2012), os valores do coeficiente Kappa de Fleiss obtidos indicam uma concordância substancial entre os anotadores para o corpus.

**Quadro 3.1 – Interpretação dos valores do coeficiente Kappa de Fleiss.**

Valor	Interpretação
<0%	Concordância pobre
1% a 20%	Concordância fraca
21% a 40%	Concordância razoável
41% a 60%	Concordância moderada
61% a 80%	Concordância substancial
81% a 100%	Concordância quase perfeita

**Fonte: Adaptado de Seljan, Brkic e Vivic (2012)**

## 3.2 Pré-processamento

Informações textuais são, geralmente, não estruturadas e sem regras de padronização. Por este motivo, é necessário preparar os textos para a aplicação dos métodos de aprendizagem (WEISS, 2005). Para transformar essas informações para que os classificadores possam entendê-las e trabalhar com as mesmas, são utilizadas as técnicas de pré-processamento. Neste trabalho foram utilizadas técnicas de tokenização, filtragem, remoção de stopword e modelo n-gram.

Uma aplicação Python foi desenvolvida para executar todas as etapas de mineração de opinião. Para etapa de pré-processamento, foi utilizada a biblioteca NLTK (Natural Language Toolkit), essa biblioteca fornece diversas ferramentas e funcionalidades para processar as informações textuais.

### 3.2.1 Tokenização

A primeira técnica utilizada para tratar um texto é a tokenização, que consiste na divisão de um documento em elementos denominados *tokens* (WEISS, 2005). Essa técnica divide o texto em vários pedaços usando um delimitador, que pode ser espaços em branco entre a palavra, sinais de pontuação ou regras de negócio.

Nesta aplicação, foi utilizado o TweetTokenizer da biblioteca NLTK, logo, por ser próprio para o Twitter, realiza a separação das partes específicas do *tweet*, como as *hashtags*, *mentions*, pontuações, *emoticons*, *emojis*, etc. O Quadro 3.2 demonstra como é realizada a técnica de Tokenização:

<b>Quadro 3.2 – Exemplo do processo de Tokenização.</b>	
<b>Entrada</b>	"Este jornal é de dar nojo!"
<b>Saída</b>	["Este", "jornal", "é", "dar", "nojo!"]

**Fonte: Elaborado pelo autor (2019)**

### 3.2.2 Filtragem

Filtragem é a técnica de remoção dos *tokens* que são considerados irrelevantes para a aplicação. Nesta etapa, foram removidos os seguintes *tokens*:

- todas as *mentions*, isto é, *tokens* iniciados por '@';
- todas as *hashtags*, isto é, *tokens* iniciados por '#';
- e todas as URLs.

Após isso, os *tokens* restantes tiveram suas letras convertidas para letras minúsculas.

### 3.2.3 Remoção de Stopwords

A técnica de Remoção de Stopwords é utilizada para remover um conjunto de palavras que aparecem com muita frequência no texto. Estas palavras, chamadas de Stopwords, geralmente são preposições, artigos, conjunções, alguns verbos, nomes, adjetivos e advérbios. Para isto, deve

ser criada uma lista, denominada Stoplist, no idioma em que se está trabalhando, contendo estas palavras consideradas irrelevantes. Este processo faz-se necessário para retirar do texto palavras que não tem nenhuma importância, diminuindo assim o tamanho das estruturas de indexação e facilitando a mineração (BARION; LAGO, 2015). O Quadro 3.3 demonstra como é realizada a Remoção de Stopwords:

**Quadro 3.3 – Exemplo do processo de Remoção de Stopwords.**

<b>Entrada</b>	"Este jornal é de dar nojo!"
<b>Saída</b>	["jornal", "dar", "nojo"]

**Fonte: Elaborado pelo autor (2019)**

Neste trabalho a palavra “não” foi removida da Stoplist, por se tratar de análise de opinião a palavra “não” é relevante para o estudo.

### 3.2.4 Modelo N-Gram

N-gram é um modelo de linguagem probabilística amplamente utilizado na identificação de língua (VATANEN; VÄYRYNEN; VIRPIOJA, 2010). Um n-gram é uma combinação de “n” itens de uma determinada sequência de texto. Esses itens podem ser fonemas, caracteres, palavras e outros. Um modelo de linguagem tenta refletir a frequência com que cada item ocorre como uma sentença em um texto. Unigram refere-se a n-gram de tamanho 1, bigram refere-se a n-gram de tamanho 2, trigram refere-se a n-gram de tamanho 3 e assim por diante (CASTRO et al., 2017).

Neste trabalho foi adotado n-gram de tamanho 2, bigram, para a aspecto da palavra. O Quadro 3.4 traz uma demonstração do bigram:

**Quadro 3.4 – Exemplo do processo de bigram.**

<b>Entrada</b>	"Este jornal é de dar nojo!"
<b>Saída</b>	["Este jornal", "jornal é", "é de", "de dar", "dar nojo"]

**Fonte: Elaborado pelo autor (2019)**

### 3.2.5 TF-IDF

O Termo Frequência-Frequência de Documento Inverso, do inglês, Term Frequency-Inverse Document Frequency (TF-IDF) é o esquema de ponderação de termo mais popular na recuperação de informação e classificação de texto (HUANG, 2016).

De acordo com Wu et al. (2004), TF-IDF pode ser descrito da seguinte forma: dado um conjunto de documentos  $D = d_1, d_2, \dots, d_n | d_i \in D$ . A frequência do termo  $TF(d, t)$  é o número de vezes, um termo  $t$  (n-gram) ocorre em um documento. Por sua vez,  $IDF(t)$  é o número de documentos em que o termo ocorre pelo menos uma vez. O inverso do termo frequência na coleção de documentos (IDF) pode ser calculado pela Equação 3.1, pode-se inferir que o peso do termo TF-IDF, como mostrado na Equação 3.2. Como resultado, TF atribui um peso maior a um n-grama frequente enquanto o IDF diminui o peso se o n-gram ocorrer em muitos documentos.

$$\log\left(\frac{|D|}{DF(t)}\right) \quad (3.1)$$

$$W(t) = TF(d, t) * IDF(t) \quad (3.2)$$

## 3.3 Processamento

A etapa de processamento, consiste na realização da classificação de polaridade dos *tweets*, foram utilizados quatro classificadores para construção do comitê de classificador:

- Multinomial Naïve-Bayes (MNB)
- Support Vector Machines (SVM)
- Regressão Logística (RL)
- Passive Agressive (PA)

### 3.3.1 Comitê de Classificador

Neste trabalho, foram utilizadas três abordagens para construção do comitê de classificador: Bagging, Boost e Voting, apresentados nos Referencial Teórico, Capítulo 2. Para cada abordagem foi construído um comitê de classificador, utilizando os quatro classificadores apresentados.

## 3.4 Avaliação

Neste trabalho, para a etapa de avaliação, foi adotado o método de validação cruzada *10-fold*, logo o corpus foi dividido em 10 partes iguais e a classificação foi realizada 10 vezes para cada configuração. Que geralmente apresenta melhores resultados que outras técnicas de validação cruzada mais custosas (KOHAVI et al., 1995). Os resultados foram obtidos utilizando a medida de *acurácia*, *precisão*, *f-measure* e *revocação*. Por fim, os resultados obtidos foram comparados utilizando a medida de *acurácia*, comparando os comitês e as configurações utilizadas.

O Quadro 3.5, demonstra as métricas e suas fórmulas utilizadas para obtenção dos resultados deste trabalho:

Quadro 3.5 – Métricas utilizadas e suas fórmulas	
Métrica	Formula
Acurácia	$Acuracia = \frac{CC}{Total}$
Precisão	$Precisao = \frac{TP}{TP+FP}$
Revocação	$Revocacao = \frac{TP}{TP+FN}$
F-measure	$F - measure = \frac{2*Precisao*Revocacao}{Precisao+Revocacao}$

Fonte: Elaborado pelo autor (2019)

Logo:

- Classificados Corretamente, *CC*, são os *tweets* classificados corretamente;
- Verdadeiros Positivos *TP*, são os *tweets* positivos classificados corretamente;

- Falsos Positivos,  $FP$ , são os *tweets* positivos classificados incorretamente;
- Falsos Negativos,  $FN$ , são os *tweets* negativos classificados incorretamente.

## 4 Resultados e Discussão

Para realização do experimento foram montados dois corpus, sendo eles desbalanceado e balanceado, contendo *tweets* de opiniões de eleitores brasileiros, sendo essas opiniões classificadas em três classes, no idioma Português do Brasil, o Apêndice A demonstra como foi realizada a anotação manual para montagem do corpus desbalanceado. Também foi utilizado o corpus de Diakopoulos e Shamma (2010), contendo opiniões sobre o debate de Barack Obama and John McCain, sendo essas opiniões apresentadas em *tweets* na língua inglesa e sendo classificadas em três categorias, o corpus foi construído de forma desbalanceada.

O corpus balanceado foi montado a partir do corpus desbalanceado, a quantidade de opiniões de cada classe é escolhida pela classe que possui a menor quantidade de documentos e as opiniões das outras classes são escolhidas aleatoriamente para essa quantidade.

### 4.1 Configuração do Experimento

Neste trabalho foram adotadas as seguintes configurações para o experimento, apresentadas no Quadro 4.1:

**Quadro 4.1 – Configurações adotadas do trabalho.**

Pré-processamento	Config 1	Config 2	Config 3	Config 4	Config 5	Config 6
Filtragem	x	x		x		
Stopwords	x	x	x			
N-gram	x		x	x	x	
TF-IDF	x	x	x	x	x	x

**Fonte: Elaborada pelo autor (2019)**

- Configuração 1: utiliza todas as técnicas de pré-processamento;
- Configuração 2: utiliza somente as técnicas de filtragem, remoção de stopwords e TF-IDF;
- Configuração 3: utiliza somente as técnicas de remoção de stopwords, n-gram e TF-IDF;
- Configuração 4: utiliza somente as técnicas de filtragem, n-gram e TF-IDF;
- Configuração 5: utiliza somente a técnica de n-gram e TF-IDF;

- Configuração 6: utiliza somente a técnica de TF-IDF.

A Tabela 4.1 apresenta a configuração da corpora utilizada neste trabalho:

**Tabela 4.1 – Configuração da corpora**

<b>Corpus</b>	<b>Classe</b>	<b>Quantidade</b>
<b>PT-BR Desbalanceado</b>	1	978
	2	1313
	3	379
<b>PT-BR Balanceado</b>	1	379
	2	379
	3	379
<b>EN Desbalanceado</b>	1	1246
	2	741
	3	290
<b>EN Balanceado</b>	1	290
	2	290
	3	290

**Fonte: Elaborado pelo autor (2019)**

O balanceamento da corpora foi realizado de forma aleatória para as classes apresentadas.

- PT-BR Desbalanceado: corpus desbalanceado no idioma Português do Brasil;
- PT-BR Balanceado: corpus balanceado no idioma Português do Brasil;
- EN Desbalanceado: corpus desbalanceado no idioma Inglês;
- EN Balanceado: corpus balanceado no idioma Inglês.

## 4.2 Resultados do Experimento

### 4.2.1 Corpus em Português do Brasil (PT-BR)

A Tabela 4.2 apresenta os resultados obtidos pelos algoritmos e as configurações utilizadas para o corpus desbalanceado:

Tabela 4.2 – Resultados obtidos para o corpus PT-BR desbalanceado

		Configurações					
Comitê		1	2	3	4	5	6
Acurácia	Voting	71,91%	69,28%	72,65%	69,28%	68,91%	<b>74,90%</b>
	Bagging	<b>79,85%</b>	55,80%	74,15%	66,66%	67,04%	71,16%
	Boosting	88,76%	75,65%	<b>90,26%</b>	84,26%	87,64%	73,03%
Precisão	Voting	61,54%	59,35%	62,53%	57,33%	56,57%	<b>63,56%</b>
	Bagging	<b>67,43%</b>	45,85%	57,37%	55,55%	52,90%	51,51%
	Boosting	81,69%	61,41%	<b>82,68%</b>	71,39%	72,45%	59,49%
F-measure	Voting	0,4811	0,4915	0,4912	0,4735	0,4769	<b>0,4999</b>
	Bagging	<b>0,6189</b>	0,4367	0,5876	0,4936	0,4832	0,4365
	Boosting	0,5683	0,4868	<b>0,6323</b>	0,4978	0,3957	0,4476
Revocação	Voting	71,91%	69,28%	72,65%	69,28%	68,91%	<b>74,90%</b>
	Bagging	<b>79,85%</b>	55,80%	74,15%	66,66%	67,04%	71,16%
	Boosting	88,76%	75,65%	<b>90,26%</b>	84,26%	87,64%	73,03%

Fonte: Elaborada pelo autor (2018)

A Tabela 4.3 apresenta os resultados obtidos pelos algoritmos e as configurações utilizadas para o corpus desbalanceado utilizando classificadores simples:

Tabela 4.3 – Resultados obtidos para o corpus PT-BR desbalanceado com classificadores simples

		Configurações					
Classificador		1	2	3	4	5	6
Acurácia	SVM	<b>49,17%</b>	<b>49,17%</b>	<b>49,17%</b>	<b>49,17%</b>	<b>49,17%</b>	<b>49,17%</b>
	MNB	55,58%	53,37%	55,28%	51,94%	51,87%	<b>56,03%</b>
	RL	55,19%	55,69%	57,26%	55,13%	55,24%	<b>58,95%</b>
	PA	57,82%	55,13%	<b>59,10%</b>	58,35%	58,72%	58,83%
Precisão	SVM	<b>24,18%</b>	<b>24,18%</b>	<b>24,18%</b>	<b>24,18%</b>	<b>24,18%</b>	<b>24,18%</b>
	MNB	55,73%	51,94%	53,73%	50,88%	50,72%	<b>62,31%</b>
	RL	<b>62,32%</b>	55,22%	56,61%	59,24%	59,10%	58,19%
	PA	<b>60,53%</b>	53,65%	57,33%	58,57	60,13%	57,84%
F-measure	SVM	<b>0,3242</b>	<b>0,3242</b>	<b>0,3242</b>	<b>0,3242</b>	<b>0,3242</b>	<b>0,3242</b>
	MNB	0,5010	0,4961	<b>0,5161</b>	0,4797	0,4752	0,5075
	RL	0,4710	0,5123	0,5293	0,4670	0,4703	<b>0,5487</b>
	PA	0,5257	0,5415	0,5787	0,5485	0,5535	<b>0,5883</b>
Revocação	SVM	<b>49,17%</b>	<b>49,17%</b>	<b>49,17%</b>	<b>49,17%</b>	<b>49,17%</b>	<b>49,17%</b>
	MNB	55,58%	53,37%	55,28%	51,94%	51,87%	<b>56,03%</b>
	RL	55,19%	55,69%	57,26%	55,13%	55,24%	<b>58,95%</b>
	PA	57,82%	55,13%	<b>59,10%</b>	58,35%	58,72%	58,83%

Fonte: Elaborada pelo autor (2019)

Os Trabalhos Relacionados realizaram comparações entre os Comitês de Classificador e os classificadores simples, as abordagens utilizadas nos trabalhos ofereceram resultados superiores aos dos classificadores simples. Sendo assim, a utilização dos Comitês se mostrou eficiente para os experimentos.

Assim como os Trabalhos relacionados, as abordagens de comitê de classificador utilizadas neste trabalho também ofereceram melhores resultados em comparação aos classificadores simples. As comparações entre os classificadores e os comitês foram realizadas com o corpus PT-BR desbalanceado, pois neste trabalho apresentou os melhores resultados durante as execuções. Sendo assim, as comparações consideraram os melhores casos para o experimento ser realizado.

Pode-se observar que a configuração 3 apresentou o melhor resultado para o corpus desbalanceado. Analisando os comitês, pode-se notar que Voting e Boosting obtiveram as melhores acurácias nas configurações adotadas.

A configuração 6 apresentou o melhor resultado utilizando Voting com 74,90% de acurácia. Enquanto isso, a configuração 1, com Bagging de 79,85% de acurácia. Logo mais a frente, a configuração 3, com 90,26% de acurácia utilizando Boosting.

A Tabela 4.4 apresenta os resultados obtidos pelos algoritmos e as configurações utilizadas para o corpus balanceado:

**Tabela 4.4 – Resultados obtidos para o corpus PT-BR balanceado**

		Configurações					
Comitê		1	2	3	4	5	6
Acurácia	Voting	56,51%	55,46%	56,47%	55,24%	54,94%	<b>59,10%</b>
	Bagging	46,47%	46,10%	49,17%	47,94%	48,31%	<b>51,61%</b>
	Boosting	54,83%	49,02%	<b>55,16%</b>	52,99%	54,71%	52,24%
Precisão	Voting	58,39%	54,30%	58,16%	55,13%	54,54%	<b>60,24%</b>
	Bagging	<b>52,03%</b>	46,13%	50,58%	48,15%	48,43%	50,21%
	Boosting	65,87%	45,50%	<b>66,58%</b>	60,75%	62,27%	47,50%
F-measure	Voting	0,4998	0,5153	0,5003	0,4963	0,4957	<b>0,5465</b>
	Bagging	0,4413	0,4545	0,4777	0,4708	0,4760	<b>0,5055</b>
	Boosting	0,4458	0,4334	0,4484	0,4291	0,4490	<b>0,4664</b>
Revocação	Voting	56,51%	55,46%	56,47%	55,24%	54,94%	<b>59,10%</b>
	Bagging	46,47%	46,10%	49,17%	47,94%	48,31%	<b>51,61%</b>
	Boosting	54,83%	49,02%	<b>55,16%</b>	52,99%	54,71%	52,24%

**Fonte: Elaborada pelo autor (2019)**

Os resultados obtidos pelo corpus balanceado se mostraram inferiores ao corpus desbalanceado, mas mesmo assim, o comitê que utiliza a abordagem Boosting, ainda aparece com o melhor resultado, 55,16% de acurácia. Devido a quantidade inferior de *tweets* por classe apresentada no corpus balanceado, foi observada uma drástica queda nos resultados obtidos. Mesmo assim, o corpus balanceado apresenta precisões obtidas razoáveis, a abordagem Boosting obteve 66,58% de precisão.

## 4.2.2 Corpus em Inglês (EN)

Utilizando o corpus de Diakopoulos e Shamma (2010), foram executadas as configurações apresentadas no Quadro 4.1, sendo assim, a Tabela 4.5 apresenta os resultados:

**Tabela 4.5 – Resultados obtidos para o corpus EN desbalanceado**

		Configurações					
Comitê		1	2	3	4	5	6
Acurácia	Voting	<b>74,63%</b>	68,55%	73,31%	73,88	73,48%	71,38%
	Bagging	61,65%	63,44%	69,69%	72,18%	<b>73,83%</b>	71,55%
	Boosting	63,67%	64,72%	72,78%	<b>77,94%</b>	77,00%	72,78%
Precisão	Voting	45,44%	<b>48,56%</b>	41,99%	47,97%	44,31%	44,58%
	Bagging	52,44%	<b>57,65%</b>	49,84%	52,53%	56,56%	56,28%
	Boosting	49,52%	47,43%	47,66%	<b>62,92%</b>	62,09%	55,04%
F-measure	Voting	0,3887	<b>0,4008</b>	0,3834	0,3926	0,3879	0,3863
	Bagging	0,4010	<b>0,4995</b>	0,4104	0,4099	0,4737	0,4801
	Boosting	0,3971	<b>0,5005</b>	0,3943	0,4778	0,4787	0,4756
Revocação	Voting	<b>74,63%</b>	68,55%	73,31%	73,88	73,48%	71,38%
	Bagging	61,65%	63,44%	69,69%	72,18%	<b>73,83%</b>	71,55%
	Boosting	63,67%	64,72%	72,78%	<b>77,94%</b>	77,00%	72,78%

Fonte: Elaborada pelo autor (2019)

Assim como nos outros resultados do corpus PT-BT, a abordagem Boosting se mostrou mais eficiente em comparação ao Voting e a Bagging. Para o corpus de Diakopoulos e Shamma (2010), o Boosting obteve 77,94% de acurácia na configuração 4, enquanto o Voting, 74,63% na configuração 1 e a Bagging, 73,83% de acurácia na configuração 5.

Um corpus balanceado não foi utilizado, devido ao fato dos resultados obtidos pelo corpus PT-BR e pelos resultados mostrados por esse corpus desbalanceado apresentarem inferiores aos da configuração utilizada no outro experimento anterior.

## 4.3 Discussão

De acordo com os resultados apresentados, a união do n-gram com TF-IDF fornece bons resultados quando utilizado com uma abordagem Boosting. A configuração 6 obteve acurácias bastantes próximas utilizando somente TF-IDF para os três comitês utilizados no corpus desbalanceado. Pode-se notar os impactos causados por cada configuração, especialmente na Configuração 2, na qual a abordagem Bagging obteve a menor acurácia do experimento.

Também foi observado, a diferença nos resultados obtidos do corpus desbalanceado para o corpus balanceado, que pode ser causada pela quantidade de opiniões de cada classe.

O corpus de Diakopoulos e Shamma (2010) mostrou resultados satisfatórios, pois suas opiniões pertencem a outro idioma, o inglês. A única mudança realizada na configuração deste corpus, foi a mudança no idioma da Stoplist, pois a mesma estava no idioma Português Brasileiro. Este corpus obteve uma acurácia de 77,94% para abordagem Boosting, como melhor resultado dos testes. Os autores não fornecem métricas para comparações dos resultados, logo, o corpus foi utilizado para fim experimental, pois, seu idioma é o inglês.

Assim sendo, com base nos resultados apresentados, foi possível analisar a eficiência de um comitê de classificador para aplicações de mineração de opinião, utilizando dados extraídos do Twitter de eleitores brasileiros e do corpus de 2010. A Tabela 4.4 apresenta os resultados obtidos pelos algoritmos e as configurações utilizadas para o corpus desbalanceado de Diakopoulos e Shamma (2010). Entretanto, as configurações da etapa de pré-processamento podem influenciar ou prejudicar os resultados obtidos durante a avaliação.

Tanto na base de dados utilizado por Jose e Chooralil (2016), Rathi et al. (2018) e Castro, Souza e Oliveira (2016), utilizam *tweets* para construção dos seus corpus, especialmente Jose e Chooralil (2016) que os *tweets* são de eleitores, assim como este estudo. Assim é possível realizar comparativos mais significativos.

Pode ser observado diferentes tipos de pré-processamentos utilizados ao longo dos experimentos. Assim como este trabalho, Jose e Chooralil (2016), realiza a técnica de filtragem, removendo URLs, *mentions* e *hashtags*. Govindarajan (2016), realizou a técnica de remoção de stopwords. Rathi et al. (2018), também realiza a conversão para letras minúsculas.

Levando em consideração aos algoritmos e abordagens utilizadas, Govindarajan (2016), utiliza SVM e MNB como classificadores base, para construção do comitê que utiliza a abordagem *Bagging*, Jose e Chooralil (2016), utilizam MNB, para abordagem *Voting* e Rathi et al. (2018), utilizam SVM e Decision Tree para abordagem *Boosting* com o AdaBoost. Castro, Souza e Oliveira (2016), utiliza SVM, MNB e RL, para construção de um comitê com abordagem de combinação algébrica.

Apesar de Jose e Chooralil (2016), Govindarajan (2016), Rathi et al. (2018), Castro, Souza e Oliveira (2016), Kumar e Singh (2018) e Kalarani e Brunda (2017) terem realizado estudos avaliando comitê de classificador, que possuem algumas diferenças a este trabalho, é possível comparações entre os resultados obtidos deles neste estudo. O Quadro 4.2 apresenta um comparativo dos trabalhos relacionados e seus resultados com este trabalho:

**Quadro 4.2 – Visão geral e comparação dos trabalhos relacionados com este trabalho.**

Estudo	Pré-processamento	Algoritmos	Comitê	Domínio	Corpus	Resultados
Jose e Chooralil (2016)	remoção de URLs, hashtags e mentions	SentiWordNet, MNB e HMM	Voting	Política	<i>tweets</i>	71,48% de acurácia
Govindarajan (2016)	remoção de stopwords, determinar o limite da sentença e stemming	SVM, MNB e AG	Bagging	Política, mídia e educação	<i>UCI Machine Learning Repository</i>	81,00% de acurácia
Rathi et al. (2018)	letras minúsculas, ajuste de URLs, hashtags e mentions, stemming e tratamento de emoticon	SVM e Decision Tree	AdaBoost	Variado	<i>Stanford Sentiment140, Polarity Dataset e University of Michigam</i>	84,00% de acurácia
Este Trabalho (Corpus Português)	Tokenização, lower case, remoção de URLs, hashtags, mentions e stopwords, modelo n-gram e TF-IDF	SVM, MNB, RL e PA	Voting, Bagging e Boosting	Política	<i>tweets</i>	90,26% de acurácia (Boosting)
Este Trabalho (Corpus Inglês)	Tokenização, lower case, remoção de URLs, hashtags, mentions e stopwords, modelo n-gram e TF-IDF	SVM, MNB, RL e PA	Voting, Bagging e Boosting	Política	<i>tweets</i>	77,94% de acurácia (Boosting)

**Fonte: Elaborada pelo autor (2019)**

Ainda assim, é possível compreender o impacto causado pelos pré-processamentos, classificadores e abordagens utilizados ao longo dos resultados obtidos por este trabalho e dos trabalhos citados.

## 5 Conclusão

Neste estudo foi desenvolvida uma aplicação de mineração de opinião a nível de documento utilizando comitês de classificador, com o objetivo de analisar a eficiência dos comitês em aplicações desse tipo. Para isso, utilizou-se dados extraídos do Twitter contendo opiniões de eleitores brasileiros.

O estudo foi realizado utilizando as opiniões coletadas e anotadas durante as etapas de treinamento e teste dos comitês de classificadores, para esse fim, foram utilizadas seis configurações de pré-processamento.

De acordo com os experimentos realizados, a configuração 3, a qual utilizava-se de remoção de stopwords, n-gram e TF-IDF, alcançou os melhores resultados, com o melhor de todos (acurácia de 90,26%) para o comitê que utilizava a abordagem Boosting.

Neste trabalho foi construído um corpus desbalanceado contendo *tweets* de opiniões de eleitores brasileiros classificadas em três classes, a construção do corpus é demonstrada no Apêndice A. A partir do corpus desbalanceado foi montado o corpus balanceado, contendo a mesma quantidade de *tweets* para cada classe.

Também foram desenvolvidas seis configurações de pré-processamento e processamento de texto, sendo essas configurações contendo técnicas bastantes utilizadas na literatura de Mineração de Opinião.

Para este estudo, foram utilizadas três abordagens do Comitê de Classificador, sendo elas, Voting, Bagging e Boosting. A aplicação desenvolvida avaliou os resultados obtidos através das seis configurações e das quatro abordagens utilizadas para este fim. O algoritmo, com as configurações e corpus utilizados podem ser encontrados no GitHub <sup>1</sup>.

Neste estudo foi desenvolvida uma aplicação de mineração de opinião a nível de documento utilizando comitês de classificador, com o objetivo de analisar a eficiência dos comitês em aplicações desse tipo. Para isso, utilizou-se dados extraídos do Twitter contendo opiniões de eleitores brasileiros.

O estudo foi realizado utilizando as opiniões coletadas e anotadas durante as etapas de treinamento e teste dos comitês de classificadores, para esse fim, foram utilizadas seis

---

<sup>1</sup> Disponível: <https://github.com/gsubp/ensemble-tcc> . Acesso em: 26 de Junho de 2019

configurações de pré-processamento.

De acordo com os experimentos realizados, a configuração 3, a qual utilizava-se de remoção de stopwords, n-gram e TF-IDF, alcançou os melhores resultados, com o melhor de todos (acurácia de 90,26%) para o comitê que utilizava a abordagem Boosting.

Pode-se concluir que as configurações utilizadas refletem nos resultados da aplicação, já que o uso de várias técnicas de pré-processamento podem ajudar ou prejudicar o aprendizado dos classificadores.

## 5.1 Contribuições deste trabalho

As contribuições deste trabalho foram, a investigação do uso de comitê de classificador em aplicações de mineração de opinião, chegando a conclusão que o pré-processamento tem impacto no aprendizado dos classificadores, a disponibilização de uma base de dados contendo *tweets* de eleitores brasileiros para o treinamento e teste do comitê de classificador e a criação de um corpus desbalanceado contendo *tweets* de opiniões de eleitores brasileiros, classificadas em três classes.

## 5.2 Proposta para trabalhos futuros

Como foram encontrados uma quantidade relativamente pequena de trabalhos analisando comitê de classificador para mineração de opinião, estudos adicionais devem ser realizados. O domínio do texto, é uma variável que pode afetar o aprendizado, logo novos estudos devem ser feitos utilizando diversos domínios.

A etapa de pré-processamento se mostrou impactante na realização do estudo, logo outras técnicas de pré-processamento devem ser analisadas a fim de verificar o impacto destas nos resultados da aplicação.

A utilização da matriz de confusão pode se mostrar eficiente para o entendimento e compreensão dos resultados, pois poderão ser verificadas as quantidades de cada classe classificada corretamente e incorretamente.

Assim como foram obtidos resultados utilizando as medidas de *precisão*, *f-measure* e *revocação*, um estudo mais profunda pode ser realizado, utilizando uma Análise de Curva ROC.

Assim como, um estudo em mais base de dados, para assim, comparar os resultados dessas bases utilizando a aplicação criada para justificar o alto valor da acurácia obtida.

### 5.3 Limitações e Ameaças

Devido as opiniões se tratarem de política, umas das limitações é o não tratamento de ironia e detecção de discurso de ódio na construção do corpus. Onde, esses fatores apresentam um grau de dificuldade na identificação, tanto para seres humanos, quanto para os classificadores.

Outro fator que pode ser ressaltado, é a quantidade de *tweets* que possuíam sentimentos (“positvos”, “negativos”, e “neutros”) foi relativamente baixa, o que também pode prejudicar o classificador, diminuindo seu desempenho.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, A. L. F. et al. A comparison of svm versus naive-bayes techniques for sentiment analysis in tweets: A case study with the 2013 fifa confederations cup. In: *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*. [S.l.]: ACM, 2014. (WebMedia '14), p. 123—130.
- BALAZS, J. A.; VELÁSQUEZ, J. D. Opinion Mining and Information Fusion: A survey. *Information Fusion*, v. 27, p. 95–110, 2016.
- BARION, E. C. N.; LAGO, D. Mineração de textos. *Revista de Ciências Exatas e Tecnologia*, v. 3, n. 3, p. 123–140, 2015.
- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996.
- CARDOSO-CACHOPO, A.; OLIVEIRA, A. L. Semi-supervised single-label text categorization using centroid-based classifiers. In: ACM. *Proceedings of the 2007 ACM symposium on Applied computing*. [S.l.], 2007. p. 844–851.
- CASTRO, D.; SOUZA, E.; OLIVEIRA, A. L. I. Discriminating between brazilian and european portuguese national varieties on twitter textdiscriminating between brazilian and european portuguese national varieties on twitter textss. *Proceedings of 5th Brazilian Conference on Intelligent Systems (BRACIS'2016)*, p. 265–270, 2016.
- CASTRO, D. W. et al. Smoothed n-gram based models for tweet language identification: A case study of the brazilian and european portuguese national varieties. *Applied Soft Computing*, v. 61, p. 1160 – 1172, 2017. ISSN 1568-4946.
- CERON, A.; CURINI, L.; IACUS, S. M. Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence From the United States and Italy. *Social Science Computer Review*, v. 33, n. 1, p. 3–20, 2015. ISSN 15528286.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.
- CUI, H.; MITTAL, V.; DATAR, M. Comparative experiments on sentiment classification for online product reviews. In: *AAAI*. [S.l.: s.n.], 2006. v. 6, p. 1265–1270.
- DEEPLARNING.NET. *Classifying MNIST digits using Logistic Regression*. 2018. <http://deeplearning.net/tutorial/logreg.html>. Dezembro, 2018.
- DIAKOPOULOS, N. A.; SHAMMA, D. A. Characterizing debate performance via aggregated twitter sentiment. In: ACM. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. [S.l.], 2010. p. 1195–1198.
- FELDMAN, R. Techniques and applications for sentiment analysis. *Communications of the ACM*, ACM, v. 56, n. 4, p. 82–89, 2013.
- FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, v. 76, n. 5, p. 378—382, 1971.

- GOVINDARAJAN, M. Analysis of bagged ensemble classifiers for blogger data. In: *IEEE. Computing Technologies and Intelligent Data Engineering (ICCTIDE), International Conference on*. [S.l.], 2016. p. 1–5.
- GZIK, M.; JOSZKO, K.; PIENIAŻEK, J. Engineering support for the spine with spondylolisthesis treatment. *Computational Vision and Medical Image Processing IV - Proceedings of Eccomas Thematic Conference on Computational Vision and Medical Image Processing, VIPIMAGE 2013*, v. 7, p. 349–354, 2014. ISSN 1532-4435.
- HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011.
- HSU, C.-W. et al. *A practical guide to support vector classification*. Taipei, 2003.
- HUANG, Y. Conceptually categorizing geographic features from text based on latent semantic analysis and ontologies. *Annals of GIS*, Taylor & Francis, v. 22, n. 2, p. 113–127, 2016.
- JIN, X. et al. The wisdom of social multimedia: using flickr for prediction and forecast. In: *ACM. Proceedings of the 18th ACM international conference on Multimedia*. [S.l.], 2010. p. 1235–1244.
- JOSE, R.; CHOORALIL, V. S. Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach. *Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016*, p. 64–67, 2016. ISSN 0362-028X.
- KALARANI, P.; BRUNDA, S. S. An efficient approach for ensemble of SVM and ANN for sentiment classification. *2016 IEEE International Conference on Advances in Computer Applications, ICACA 2016*, p. 99–103, 2017.
- KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *MONTREAL, CANADA. Ijcai*. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.
- KUMAR, S.; SINGH, R. Comparative analysis of ensemble classifiers for sentiment analysis and opinion mining. *Proceedings - 2017 3rd International Conference on Advances in Computing, Communication and Automation (Fall), ICACCA 2017*, v. 2018-Janua, p. 1–5, 2018.
- KUMARI, A. Study on naive bayesian classifier and its relation to information gain. *International Journal on Recent and Innovation Trends in Computing and Communication*, v. 2, n. 3, p. 601–603, 2014.
- LANTZ, B. *Machine learning with R*. [S.l.]: Packt Publishing Ltd, 2013.
- LIU, B.; ZHANG, L. A survey of opinion mining and sentiment analysis. In: \_\_\_\_\_. *Mining Text Data*. Boston, MA: Springer US, 2012. p. 415–463. ISBN 978-1-4614-3223-4.
- MARINE-ROIG, E.; CLAVÉ, S. A. Tourism analytics with massive user-generated content: A case study of barcelona. *Journal of Destination Marketing and Management*, v. 4, n. 3, p. 162—172, 2015.
- MCCALLUM, A.; NIGAM, K. et al. A comparison of event models for naive bayes text classification. In: *CITeseer. AAAI-98 workshop on learning for text categorization*. [S.l.], 1998. v. 752, n. 1, p. 41–48.
- MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, Elsevier, v. 5, n. 4, p. 1093–1113, 2014.

- PAK, A.; PAROUBEK, P. Twitter as a corpus for sentiment analysis and opinion mining. In: *LREC*. [S.l.: s.n.], 2010. v. 10, n. 2010, p. 1320–1326.
- PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, v. 2, n. 1-2, p. 1–135, 2008.
- RATHI, M. et al. Sentiment analysis of tweets using machine learning approach. In: IEEE. *2018 Eleventh International Conference on Contemporary Computing (IC3)*. [S.l.], 2018. p. 1–3.
- RAVI, K.; RAVI, V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, v. 89, p. 14–46, 2015.
- RAVI, K.; RAVI, V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, Elsevier, v. 89, p. 14–46, 2015.
- SELJAN, S.; BRKIC, M.; VICIC, T. Bleu evaluation of machine-translated english-croatian legislation. In: *LREC*. [S.l.: s.n.], 2012. p. 2143–2148.
- SOUZA, E. et al. Topie: An open-source opinion mining pipeline to analyze consumers' sentiment in brazilian portuguese. In: \_\_\_\_\_. *Computational Processing of the Portuguese Language: 12th International Conference, PROPOR 2016, Tomar, Portugal, July 13-15, 2016, Proceedings*. [S.l.]: Springer International Publishing, 2016. p. 95–105. ISBN 978-3-319-41552-9.
- SOUZA, E. et al. Characterizing opinion mining: A systematic mapping study of the portuguese language. In: SPRINGER. *International Conference on Computational Processing of the Portuguese Language*. [S.l.], 2016. p. 122–127.
- TAN, P.-N. et al. *Introduction to data mining*. [S.l.]: Pearson Education India, 2007.
- TWITTER. *The Search API*. 2017. <https://dev.twitter.com/rest/public/search/>. February, 2017.
- VATANEN, T.; VÄYRYNEN, J. J.; VIRPIOJA, S. Language identification of short text segments with n-gram models. In: *LREC*. [S.l.: s.n.], 2010.
- VITÓRIO, D. et al. Investigating opinion mining through language varieties: a case study of brazilian and european portuguese tweets. In: *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology*. [S.l.: s.n.], 2017. p. 43–52.
- VITÓRIO, D. W.; SOUZA, E. *Analisando Mineração de Opinião com Variantes Linguísticas: Um Estudo com Tweets do Português Brasileiro e do Português Europeu*. 2017.
- WANG, G. et al. A comparative assessment of ensemble learning for credit scoring. *Expert systems with applications*, Elsevier, v. 38, n. 1, p. 223–230, 2011.
- WEISS, S. M. *Text mining : predictive methods for analyzing unstructured information*. New York (N.Y.) : Springer, 2005. ISBN 0387954333. Disponível em: <<http://lib.ugent.be/catalog/rug01:000859115>>.
- WIEBE, J.; WILSON, T.; CARDIE, C. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, Springer, v. 39, n. 2-3, p. 165–210, 2005.
- WU, S.-T. et al. Automatic pattern-taxonomy extraction for web mining. In: IEEE COMPUTER SOCIETY. *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*. [S.l.], 2004. p. 242–248.

## APÊNDICE A – Anotação Manual

Foram coletados 2670 *tweets*, os quais foram anotados manualmente em uma das três polaridades, a Tabela A.1 exemplifica como foi realizada a anotação:

**Tabela A.1 – Exemplo de anotação manual.**

Polaridade	Opinião
2	Né defendendo ele não, mas infelizmente é a realidade.
1	aquele momento em que Jair Messias Bolsonaro ganha até em pagina petista não tem preço =D
3	Gustavo já te expliquei sobre isso Gustavo Paes

**Fonte: Elaborado pelo autor (2019)**

Segundo as instruções transmitidas pelo autor, os anotadores deviam classificar as opiniões utilizando números, sendo eles:

- 1 - *tweets* contendo opiniões ou sentimentos positivos;
- 2 - *tweets* contendo opiniões ou sentimentos negativos;
- 3 - *tweets* que não continha sentimentos nem opiniões, isto é, textos objetivos.

Após a anotação manual, a quantidade de *tweets* anotados para cada polaridade está representada na Tabela A.2:

**Tabela A.2 – Quantidade de cada polaridade anotada.**

Polaridade	Quantidade
1	978
2	1313
3	379

**Fonte: Elaborado pelo autor (2019)**