

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DE MINAS GERAIS – *CAMPUS* SÃO JOÃO EVANGELISTA
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Giovanna Kruk Da Silva Souza

**ANÁLISE DE SENTIMENTOS UTILIZANDO ORANGE DATA MINING:
Uma percepção sobre as Eleições Presidenciais Brasileiras.**

SÃO JOÃO EVANGELISTA

2022

GIOVANNA KRUK DA SILVA SOUZA

**ANÁLISE DE SENTIMENTOS UTILIZANDO ORANGE DATA MINING:
Uma percepção sobre as Eleições Presidenciais Brasileiras.**

Trabalho de Conclusão de Curso apresentado ao curso Bacharelado em Sistemas de Informação do Instituto Federal de Minas Gerais – Campus São João Evangelista para obtenção do grau Bacharel em Sistemas de Informação.

Orientador: Me. Eduardo Augusto Costa
Trindade

São João Evangelista

2022

REDE DE BIBLIOTECAS

FICHA CATALOGRÁFICA PARA TRABALHO DE CONCLUSÃO DE CURSO

-
- S729a Souza, Giovanna Kruk da Silva.
Análise de sentimentos utilizando *orange data mining*: uma percepção sobre as eleições presidenciais brasileiras. / Giovanna Kruk da Silva Souza – 2022.
124f.: il.
Orientador: Me. Eduardo Augusto Costa Trindade.
Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação – Instituto Federal Minas Gerais. *Campus* São João Evangelista, 2022.
1. Mineração de dados textuais. 2. Análise de sentimentos. 3. Predição. 4. Eleição. 5. Twitter. I. Souza, Giovanna Kruk da Silva. II. Instituto Federal de Minas Gerais *Campus* São João Evangelista. III. Título.

CDD 005

Catálogo: Rejane Valéria Santos - CRB-6/2907


Giovanna Kruk da Silva Souza

ANÁLISE DE SENTIMENTOS UTILIZANDO ORANGE DATA MINING:


Uma percepção sobre as Eleições Presidenciais Brasileiras.

Trabalho de Conclusão de Curso apresentado ao curso Bacharelado em Sistemas de Informação do Instituto Federal de Minas Gerais – Campus São João Evangelista para obtenção do grau Bacharel em Sistemas de Informação.


Aprovado em 08/12/2022 pela banca examinadora:

Documento assinado digitalmente
 EDUARDO AUGUSTO COSTA TRINDADE
Data: 14/12/2022 22:17:04-0300
Verifique em <https://verificador.iti.br>

Prof. Me. Eduardo Augusto Costa Trindade (Orientador) – IFMG -SJE

Documento assinado digitalmente
 ITALO MAGNO PEREIRA
Data: 20/12/2022 10:01:07-0300
Verifique em <https://verificador.iti.br>

Prof. Me. Ítalo Magno Pereira – IFMG -SJE

Documento assinado digitalmente
 WESLEY GOMES DE ALMEIDA
Data: 19/12/2022 19:05:22-0300
Verifique em <https://verificador.iti.br>

Prof. Dr. Wesley Gomes de Almeida– IFMG -SJE

Dedico este trabalho a todo o curso de Sistemas de Informação do IFMG, ao corpo docente e discente, o qual fico lisonjeada por ter feito parte. Aos meus familiares e amigos que me apoiaram durante minha trajetória acadêmica. A todas as divindades, por todas as forças dedicadas a mim e a todos aqueles que tem algum interesse pelo tema proposto.

AGRADECIMENTOS

Às divindades por terem me dado saúde e força para superar as dificuldades.

A esta universidade, seu corpo docente, direção e administração que me oportunizou a formação acadêmica.

Aos professores, especialmente ao meu orientador de TCC, prof. Me. Eduardo Trindade. Obrigado por me exigir mais do que eu acreditava que seria capaz de realizar. Declaro aqui minha eterna gratidão pelo compartilhamento de seu conhecimento e tempo. Ao corpo docente, por me proporcionar o conhecimento e afetividade da educação no processo de formação profissional. Pelo tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender.

Aos meus pais, pelo amor, incentivo e apoio incondicional.

Aos meus queridos amigos, pelo apoio, força, amor e assistência inabalável.

A eu mesma, pela perseverança e acreditar no meu potencial.

“No inferno os lugares mais quentes são reservados àqueles que escolheram a neutralidade em tempo de crise”.

Dante Alighieri

RESUMO

Ao longo dos últimos anos, o uso da internet como meio de comunicação se tornou cada vez mais popular. Com a chegada da Web 2.0, o crescimento das redes sociais ocorreu de forma exponencial, tornando o acesso à informação algo democrático, no qual cada usuário pode selecionar assuntos e interesses de acordo com suas preferências. Sendo assim, assuntos de interesse comum, tal como política, se tornou umas das principais pautas de discussão nessa rede, em que cada usuário expõe seu ponto de vista sobre este assunto. Com isso, há uma enorme quantidade de dados gerados diariamente sobre este tópico. Tais dados contém informações importantes para tomadas de decisões. Através delas, é possível concluir como está o índice de aprovação sobre as ações daquele candidato. Além disso, em anos eleitorais, as redes sociais são movimentadas com diversas opiniões políticas. Em paralelo, ainda são realizadas pesquisas eleitorais de forma manual, que acabam sendo lentas e financeiramente inviáveis. Sendo assim, este trabalho visa utilizar técnicas de mineração de dados textuais para analisar comentários da rede social Twitter, a fim de verificar os níveis de popularidade dos candidatos à presidência da República em 2022, verificando como o posicionamento dos candidatos afetam sua popularidade em relação aos eleitores e, ao final, comparar a popularidade construída na rede social com os resultados finais da eleição. Após realizar diversas análises, foi possível concluir que o Twitter pode ser utilizado como meio alternativo de pesquisa eleitoral, bem como um monitor de popularidade para direcionar melhor os candidatos com suas propostas.

Palavras-chave: Mineração de dados textuais. Análise de sentimentos. Predição eleitoral. Twitter.

ABSTRACT

Over the past few years, the use of the internet as a means of communication has become increasingly popular. With the arrival of Web 2.0, the growth of social networks took place exponentially, making access to information something democratic, in which each user can select subjects and interests according to their preferences. Therefore, subjects of common interest, such as politics, have become one of the main topics of discussion in this network, in which each user exposes his point of view on this subject. With that, there is a huge amount of data generated daily on this topic. Such data contains important information for decision making. Through them, it is possible to conclude how is the approval rate on the actions of that candidate. In addition, in election years, social networks are busy with different political opinions. In parallel to this, electoral polls are still carried out manually, which end up being slow and financially unfeasible. Therefore, this work aims to use textual data mining techniques to analyze comments on the social network Twitter, in order to verify the levels of popularity of the candidates for the presidency of the Republic in 2022, verifying how the positioning of the candidates affects their popularity in relation to the voters and, in the end, compare the popularity built on the social network with the final results of the election. After carrying out several analyses, it was possible to conclude that Twitter can be used as an alternative means of electoral research, as well as a popularity monitor to better target candidates with their proposals.

Keywords: Textual data mining. Sentiment analysis. Electoral prediction. Twitter.

LISTA DE FIGURAS

Figura 1 - Processos do KDD	24
Figura 2 - Interface do <i>Orange Data Mining</i>	31
Figura 3 - Metodologia empregada	43
Figura 4 - <i>Widget</i> Twitter	44
Figura 5 - Tela do <i>Widget</i>	44
Figura 6 - Inserção de Dados API	45
Figura 7 - <i>Word Cloud</i> antes do Pré-Processamento.....	46
Figura 8 - <i>Word Cloud</i> depois do Pré-Processamento.....	46
Figura 9 - Modelo Classificador	48
Figura 10 - Resultado do Primeiro Debate - 1º Turno.....	52
Figura 11 - Resultado do Segundo Debate - 1º Turno.....	53
Figura 12 - <i>Heatmap</i> - <i>Ciro Gomes</i>	54
Figura 13 - Gráfico de Emoções - <i>Ciro Gomes</i>	54
Figura 14 - <i>Heatmap</i> - <i>Felipe D'Avila</i>	55
Figura 15 - Gráfico de Emoções - <i>Felipe D'Avila</i>	56
Figura 16 - <i>HeatMap</i> Candidato <i>Jair Bolsonaro</i>	56
Figura 17 - Gráfico de Emoções - <i>Jair Bolsonaro</i>	57
Figura 18 - <i>HeatMap</i> candidato <i>Luís Inácio Lula da Silva</i>	58
Figura 19 - Gráfico de Emoções - <i>Lula</i>	58
Figura 20 - <i>Heatmap</i> - Candidato <i>Padre Kelmon</i>	59
Figura 21 - Gráfico de Emoções - <i>Padre Kelmon</i>	59
Figura 22 - <i>HeatMap</i> - Candidata <i>Simone Tebet</i>	60
Figura 23 - Gráfico de Sentimentos - <i>Simone Tebet</i>	60
Figura 24 - <i>HeatMap</i> - Candidata <i>Soraya Thronicke</i>	61
Figura 25 - Gráfico de Sentimento - <i>Soraya Thronicke</i>	61
Figura 26 - Gráfico Geral dos Candidatos- 1º Turno	62
Figura 27 - Sentimento Isolado - <i>Anger</i>	63
Figura 28 - Sentimento Isolado - <i>Disgust</i>	63
Figura 29 - Sentimento Isolado - <i>Fear</i>	64
Figura 30 - Sentimento Isolado - <i>Joy</i>	64
Figura 31 - Sentimento Isolado - <i>Sadness</i>	65
Figura 32 - Sentimento Isolado - <i>Surprise</i>	65

Figura 33 - Resultado do 1º Turno	66
Figura 34 - <i>HeatMap</i> - Candidato Jair Bolsonaro	67
Figura 35 - Gráfico de Emoções – Jair Bolsonaro.....	67
Figura 36 - <i>HeatMap</i> - Candidato Lula	68
Figura 37 - Gráfico de Emoções - Candidato Lula.....	68
Figura 38 - Gráfico de Barras - Coleta 2º Turno	69
Figura 39 - Sentimento Isolado - Anger	70
Figura 40 - Sentimento Isolado - Disgust.....	70
Figura 41 - Sentimento Isolado - Fear	71
Figura 42 - Sentimento Isolado - Joy.....	71
Figura 43 - Sentimento Isolado – Sadness.....	71
Figura 44 - Sentimento Isolado - Surpeise	72
Figura 45 - Resultado - 2º Turno	73
Figura 46- <i>HeatMap</i> Pré-Debate - Jair Bolsonaro.....	80
Figura 47 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro	80
Figura 48 - <i>WordCloud</i> Pré-Debate - Jair Bolsonaro	81
Figura 49 - <i>WordCloud</i> Pós-Processada - Pré-Debate - Jair Bolsonaro	81
Figura 50 - <i>HeatMap</i> Debate - Jair Bolsonaro.....	82
Figura 51 - <i>WordCloud</i> Debate - Jair Bolsonaro -.....	82
Figura 52 - <i>WordCloud</i> Pós-Processada - Debate - Jair Bolsonaro.....	83
Figura 54 - <i>HeatMap</i> Pós-Debate - Jair Bolsonaro	84
Figura 53 - Gráfico de Emoções Pós-Debate - Jair Bolsonaro.....	84
Figura 55 - <i>WordCloud</i> Pós-Processada - Pré-Debate - Jair Bolsonaro	85
Figura 56 - <i>WordCloud</i> Pós-Debate - Jair Bolsonaro.....	85
Figura 57 - <i>HeatMap</i> Pré-Debate - Jair Bolsonaro	86
Figura 58 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro	86
Figura 59 - <i>WordCloud</i> Pré-Debate - Jair Bolsonaro	87
Figura 60 - <i>WordCloud</i> Pós-Processada - Pré-Debate - Jair Bolsonaro	87
Figura 61 - <i>HeatMap</i> Debate - Jair Bolsonaro.....	88
Figura 62 - Gráfico de Emoções Debate - Jair Bolsonaro.....	88
Figura 63 - <i>WordCloud</i> Debate - Jair Bolsonaro	89
Figura 64 - <i>WordCloud</i> Pós-Processada - Debate - Jair Bolsonaro.....	89
Figura 65 - <i>HeatMap</i> Pós-Debate - Jair Bolsonaro	90
Figura 66 - Gráfico de Emoções Pós-Debate - Jair Bolsonaro.....	90

Figura 67 - <i>WordCloud</i> Pós-Debate - Jair Bolsonaro.....	91
Figura 68 - <i>WordCloud</i> Pós-Processada - Pós-Debate - Jair Bolsonaro.....	91
Figura 69 - <i>HeatMap</i> Pré-Debate - Jair Bolsonaro.....	92
Figura 70 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro	92
Figura 71 - <i>WordCloud</i> Pós-Processada - Pré-Debate - Jair Bolsonaro	93
Figura 72 - <i>WordCloud</i> Pré-Debate - Jair Bolsonaro	93
Figura 73 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro	94
Figura 74 - <i>HeatMap</i> Debate - Jair Bolsonaro	94
Figura 75 - <i>WordCloud</i> Pós-Processada - Debate - Jair Bolsonaro.....	95
Figura 76 - <i>WordCloud</i> Debate - Jair Bolsonaro.....	95
Figura 77 - <i>HeatMap</i> Pós-Debate - Jair Bolsonaro	96
Figura 78 - Gráfico de Emoções Pós-Debate - Jair Bolsonaro.....	96
Figura 79 - <i>WordCloud</i> Pós-Processada - Pós-Debate - Jair Bolsonaro.....	97
Figura 80 - <i>WordCloud</i> Pós-Debate - Jair Bolsonaro.....	97
Figura 81 - <i>HeatMap</i> Pré-Debate - Jair Bolsonaro.....	98
Figura 82 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro	98
Figura 83 - <i>WordCloud</i> Pré-Debate - Jair Bolsonaro	99
Figura 84 - <i>WordCloud</i> Pós-Processada - Pré-Debate - Jair Bolsonaro	99
Figura 85 - <i>HeatMap</i> Debate - Jair Bolsonaro	100
Figura 86 - Gráfico de Emoções Debate - Jair Bolsonaro.....	100
Figura 87 - <i>WordCloud</i> Debate - Jair Bolsonaro.....	101
Figura 88 - <i>WordCloud</i> Pós-Processada - Debate - Jair Bolsonaro.....	101
Figura 89 - Gráfico de Emoções Pós-Debate - Jair Bolsonaro.....	102
Figura 90 - <i>HeatMap</i> Pós-Debate - Jair Bolsonaro	102
Figura 91 - <i>WordCloud</i> Pós-Processada - Pós-Debate - Jair Bolsonaro.....	103
Figura 92 - <i>WordCloud</i> Pós-Debate - Jair Bolsonaro.....	103
Figura 93 - <i>HeatMap</i> Pré-Debate - Luiz Inácio Lula da Silva	104
Figura 94 - Gráfico de Emoções Pré-Debate - Luiz Inácio Lula da Silva.....	104
Figura 95 - <i>WordCloud</i> Pós-Processada - Pré-Debate - Luiz Inácio Lula da Silva	105
Figura 96 - <i>WordCloud</i> Pré-Debate - Luiz Inácio Lula da Silva.....	105
Figura 97 - <i>HeatMap</i> Debate - Luiz Inácio Lula da Silva.....	106
Figura 98 - Gráfico de Emoções Debate - Luiz Inácio Lula da Silva	106
Figura 99 - Gráfico de Emoções Debate - Luiz Inácio Lula da Silva	107
Figura 100 - <i>WordCloud</i> Pós-Processada - Debate - Luiz Inácio Lula da Silva	107

Figura 101 - <i>HeatMap</i> Pós-Debate - Luiz Inácio Lula da Silva.....	108
Figura 102 - Gráfico de Emoções Pós-Debate - Luiz Inácio Lula da Silva	108
Figura 103 - <i>WordCloud</i> Pós-Debate - Luiz Inácio Lula da Silva	109
Figura 104 - <i>WordCloud</i> Pós-Processada - Pós-Debate - Luiz Inácio Lula da Silva.....	109
Figura 105 - <i>HeatMap</i> Pré-Debate - Luiz Inácio Lula da Silva	110
Figura 106 - Gráfico de Emoções Pré-Debate - Luiz Inácio Lula da Silva.....	110
Figura 107 - <i>WordCloud</i> Pré-Debate - Luiz Inácio Lula da Silva.....	111
Figura 108 - <i>WordCloud</i> Pós-Processada - Pré-Debate - Luiz Inácio Lula da Silva.....	111
Figura 109 - <i>HeatMap</i> Debate - Luiz Inácio Lula da Silva.....	112
Figura 110 - <i>WordCloud</i> Debate - Luiz Inácio Lula da Silva	112
Figura 111 - <i>WordCloud</i> Pós-Processada - Debate - Luiz Inácio Lula da Silva	113
Figura 112 - <i>HeatMap</i> Pós-Debate - Luiz Inácio Lula da Silva.....	114
Figura 113 - Gráfico de Emoções Pós-Debate - Luiz Inácio Lula da Silva	114
Figura 114 - <i>WordCloud</i> Pós-Debate - Luiz Inácio Lula da Silva	115
Figura 115 - <i>WordCloud</i> Pós-Processada - Pós-Debate - Luiz Inácio Lula da Silva.....	115
Figura 116 - <i>HeatMap</i> Pré-Debate - Luiz Inácio Lula da Silva	116
Figura 117 - Gráfico de Emoções Pré-Debate - Luiz Inácio Lula da Silva.....	116
Figura 118 - <i>WordCloud</i> Pré-Debate - Luiz Inácio Lula da Silva.....	117
Figura 119 - <i>WordCloud</i> Pós-Processada - Pré-Debate - Luiz Inácio Lula da Silva.....	117
Figura 120 - <i>HeatMap</i> Debate - Luiz Inácio Lula da Silva.....	118
Figura 121 - Gráfico de Emoções Debate - Luiz Inácio Lula da Silva	118
Figura 122 - <i>WordCloud</i> Debate - Luiz Inácio Lula da Silva	119
Figura 123 - <i>WordCloud</i> Pós-Processada - Debate - Luiz Inácio Lula da Silva	119
Figura 124 - <i>HeatMap</i> Pós-Debate - Luiz Inácio Lula da Silva.....	120
Figura 125 - Gráfico de Emoções Pós-Debate - Luiz Inácio Lula da Silva	120
Figura 126 - <i>WordCloud</i> Pós-Processada - Pós-Debate - Luiz Inácio Lula da Silva.....	121
Figura 127 - <i>WordCloud</i> Pós-Debate - Luiz Inácio Lula da Silva	121

LISTA DE TABELAS

Tabela 1 - Comparativo entre os Trabalhos Correlatos	35
Tabela 2 - Métodos e Procedimentos	41
Tabela 3 - Debates Presidenciais 2022	49
Tabela 4 - Lista dos Candidatos à Presidência em 2022	51
Tabela 5 - Dados - 1º Turno	62
Tabela 6 - Coleta de Dados Eleição - Segundo Turno	69

LISTA DE ABREVIATURAS E SIGLAS

API - *Application Programming Interface*

DF – *Document Frequency*

FGV – Fundação Getúlio Vargas

KDD - *Knowledge-Discovery In Databases*

PLN – Processamento De Linguagem Natural

RI – Recuperação Da Informação

SNA - *Social Network Analysis*

SVM - *Support Vector Machine*

TF-IDF - *Term Frequency - Inverse Document Frequency*

TV - Televisor

UFOP - Universidade Federal de Ouro Preto

UFRN - Universidade Federal do Rio Grande do Norte

UX - *User Experience*

WEB - *World Wide Web*

SUMÁRIO

1. INTRODUÇÃO	19
1.1 Justificativa	21
1.2 Questões de Pesquisa	21
1.3 Objetivos	22
1.3.1 Objetivos Específicos	22
1.4 Organização.....	23
2. REVISÃO LITERÁRIA.....	24
2.1 Mineração de Dados.....	24
2.2 Mineração de Texto.....	25
2.3 Recuperação da Informação	26
2.4 Análise de Sentimentos	26
2.5 Redes Sociais	27
2.6 Twitter	28
2.7 Eleições	29
2.8 <i>Orange Data Mining</i>	30
2.9 Trabalhos Correlatos	31
3. METODOLOGIA	38
3.1 Natureza da Pesquisa.....	38
3.2 População e Amostra.....	40
3.3 Instrumentos	40
3.4 Métodos e Procedimentos	40
4. DESENVOLVIMENTO DA PESQUISA	43
4.1 Coleta de dados e pré-processamento	43
4.2 Montagem do modelo classificador	47
4.3 Escopo de Experimentos	49

5. RESULTADOS.....	50
5.1 Dados Analisados.....	52
5.2 Análise de Popularidade Individual – Primeiro Turno	53
5.2.1 Ciro Gomes	53
5.2.2 Felipe D’Avila.....	55
5.2.3 Jair Bolsonaro	56
5.2.4 Luiz Inácio Lula da Silva	57
5.2.5 Padre Kelson.....	59
5.2.6 Simone Tebet	60
5.2.7 Soraya Thronicke	61
5.3 Média de popularidade durante o primeiro turno.....	62
5.4 Comparativos com o Resultado da Eleição – Primeiro Turno.....	65
5.5 Análise de Popularidade Individual – Segundo Turno	66
5.5.1 Jair Bolsonaro	66
5.5.2 Luiz Inácio Lula da Silva	68
5.3 Média de popularidade durante o segundo turno	69
5.4 Comparativos com o Resultado da Eleição – Segundo Turno.....	72
6. CONCLUSÃO	74
6.1 Discussão	74
6.2 Limitações.....	74
6.3 Considerações Finais e trabalhos futuros.....	75
7. REFERÊNCIAS	76
APÊNDICE A – CANDIDATO JAIR BOLSONARO – PRIMEIRO DEBATE - (BAND, UOL, FOLHA DE S. PAULO E TV CULTURA) – 28/08/2022	80
APÊNDICE B – CANDIDATO JAIR BOLSONARO – SEGUNDO DEBATE – (SBT, CNN, VEJA, O ESTADO DE S. PAULO, NOVA BRASIL FM E TERRA) – 24/09/2022.....	86

APÊNDICE C – CANDIDATO JAIR BOLSONARO – TERCEIRO DEBATE – (DEBATE DA BAND, UOL, FOLHA DE S. PAULO E TV CULTURA) – 16/10/2022 ..	92
APÊNDICE D – CANDIDATO JAIR BOLSONARO – QUARTO DEBATE – (TV GLOBO) – 28/10/2022.....	98
APÊNDICE E – CANDIDATO LUIZ INÁCIO LULA DA SILVA – PRIMEIRO DEBATE - (BAND, UOL, FOLHA DE S. PAULO E TV CULTURA) – 28/08/2022	104
APÊNDICE F – CANDIDATO LUIZ INÁCIO LULA DA SILVA – TERCEIRO DEBATE – (DEBATE DA BAND, UOL, FOLHA DE S. PAULO E TV CULTURA) – 16/10/2022	109
APÊNDICE G – CANDIDATO LUIZ INÁCIO LULA DA SILVA – QUARTO DEBATE - (TV GLOBO) – 28/10/2022	115
ANEXO A - <i>STOPWORD</i>.....	121

1. INTRODUÇÃO

De acordo com Bernardo e Goulart (2011), “uma rede social é uma estrutura social composta por pessoas ou organizações, conectadas por um ou vários tipos de relações, que partilham valores e objetivos comuns”. Com a ascensão da rede mundial de computadores e a disseminação da internet, o termo rede social ressignificou-se para o ambiente online, com o constante uso de aplicativos por meio da interação da internet, tornando-se cada vez mais popular.

Esta nova interação do público com a internet gera novos tipos de dados, que vêm chamando atenção justamente por seu conteúdo. Com a crescente demanda pelo uso de redes sociais, os internautas gradualmente demonstraram seus sentimentos sobre os mais determinados assuntos, fazendo com que a opinião do usuário obtivesse mais relevância para as organizações. Tais informações, quando coletadas e analisadas corretamente, são boas perspectivas para empresas acompanharem o andamento de um produto, questões étnico-raciais ou até mesmo o uso governamental.

Segundo Lévy (2007), a inteligência coletiva é uma forma de valorizar as capacidades individuais, unificando indivíduos por meio da tecnologia e permitindo o compartilhamento de conhecimento. Um dos métodos destes compartilhamentos é através de fóruns *online*, *chats*, *blogs* e as próprias redes sociais. Dentre as redes existentes, destaca-se o uso do Twitter, um serviço de microblog que permite aos usuários postarem textos de até 280 caracteres sobre diversos assuntos. Um dos principais assuntos contidos nos *tweets*, como são chamadas as postagens, são conteúdos de cunho político.

Alguns pontos em questão considerando as redes sociais são o compartilhamento de informações, conhecimentos, interesses e esforços em busca de alcançar um objetivo comum (BERNARDO e GOULART, 2011). Estudos realizados por Marques e Sampaio (2011) constatam que a partir de 2010 houve um crescimento considerável do uso das redes sociais para fins políticos, destacando-se principalmente no período eleitoral, palco de debates e apresentações de propostas para cativar o público a escolher um candidato.

A opinião coletiva dos usuários perante atos do governo detém uma certa relevância, podendo extrair dados sobre aprovações ou não das políticas públicas. Além disso, pesquisas obtidas por meios tradicionais têm cada vez mais caído em desuso pela pouca praticidade em obter dados, além de demandar tempo e entregar resultados não muito satisfatórios (BECKER e TUMITAM, 2013 apud PEREIRA, 2019). Sendo assim, com o grande volume de dados gerados pelas redes sociais, a mineração de texto torna-se importante

para viabilizar e estabelecer métricas referentes à opinião pública sobre os candidatos à presidência.

Cada vez mais a classe política brasileira passa a usar a tecnologia com o intuito de atingir uma nova massa de eleitorado (SILVA e JUNIOR, 2014). Segundo Maarek (2014), no passado, atividades como palestras, mídias impressas e discursos de rádio e TV eram suficientes para disseminação de conteúdo. Porém, com a nova massa de eleitorado, faz-se necessário encontrar novas formas de *marketing* político, utilizando as redes sociais como contato direto com o público alvo da campanha.

Assuntos políticos são temas recorrentes em ambientes virtuais, ganhando destaque principalmente no período eleitoral. Por meio da evolução tecnológica, atualmente é possível, por meios de técnicas computacionais, extrair informações potencialmente ser úteis para tomadas de decisões. Para Aggarwal e Zhai (2012), com a criação das mídias sociais nos últimos anos, a mineração de texto alcançou posições importantes, visto que as redes sociais facilitam a criação de conteúdos, fazendo com que a mineração seja cada vez mais necessária.

O *big data*, conjunto de dados em maior volume e complexidade gerado pelas redes, pontua uma linha tênue entre da informação que encontra-se disponível ao usuário e a veracidade da mesma. Para Shao et.al (2017), ao consumir notícias através de redes sociais, é possível estar exposto a grandes quantidades de informações inverídicas, em sua maioria por fomentação pecuniária, ideológica ou até mesmo política. Segundo o Dicionário de Oxford (2022), *fake news* são relatórios falsos de eventos, escritos e lidos em sites. Ressalta-se que a divulgação da mesma acarreta sérios problemas, pois o conteúdo fictício disseminado pode ameaçar os atos democráticos de um país.

Dada a importância das estratégias de *marketing* político digital, como afirma Lévy (2007), torna-se importante analisar as eleições presidenciais dispostas no ano de 2022. Como resultado, foi possível compreender o que se pressupõe acerca dos candidatos presidenciais por meio das mídias sociais, especificamente o Twitter. Ao final realizou-se uma comparação entre os resultados das eleições presidenciais e a popularidade dos candidatos na internet, com o objetivo de compreender a relação entre o desempenho dos candidatos nas urnas e sua aceitação nas redes sociais. Como resultado, foi avaliado a viabilidade dessa rede social como preditor eleitoral.

1.1 Justificativa

A reflexão acerca da opinião dos usuários do Twitter e o resultado final das eleições tem gerado diversos trabalhos com o intuito de obter informações subjetivas e análises contextuais. Este tipo de análise, denominado Análise de Sentimentos, possibilita descobrir tendências e opiniões gerais acerca de diversos tipos de assuntos. Trabalhos como Pereira (2019), Matos et. al (2020) e De Paula Filho e Garcia (2015) são bons exemplos da demonstração da ótica do período eleitoral nas redes sociais de anos anteriores. Ambos estudos têm foco na política inseridas em rede social e como o aprofundamento deste campo de pesquisa pode ser útil para tirar conclusões sobre o tema.

Cada vez mais canais de comunicações governamentais vem se adequando aos meios virtuais, como *sites* e demais outras redes de comunicação. Sendo assim, ressalta-se a importância de aprofundar-se mais ao tema, considerando que a população apta ao voto está cada vez mais conectada ao mundo virtual, minerar dados e analisar os sentimentos dos eleitores é uma boa estratégia para direcionar os candidatos a elaborar propostas governamentais que atendam à população, além de poder obter *feedbacks* instantâneos da população perante as ações do governo.

Neste contexto, o trabalho mostrará como o estudo do posicionamento dos candidatos pode afetar sua popularidade em relação aos eleitores através de sentimentos expressos no Twitter. Este trabalho é justificado pela necessidade e importância em dar continuidade às pesquisas recentes, fazendo-se necessário compreender o envolvimento de eleitores com a tecnologia e as eleições simultaneamente.

1.2 Questões de Pesquisa

Por conseguinte, algumas questões foram levantadas com o decorrer desta pesquisa. Considerando que o embate político da última eleição, ocorrida em 2018, gerou uma grande quantidade de dados, notou-se a necessidade de utilizar a mineração de dados como forma de extrair informações convenientes, de forma ágil e não custosa para ser utilizada tanto pelos candidatos quanto para os eleitores. Como afirmado por Fayyad, Piatetsky-Shapiro e Smyth (1996), a mineração de dados vem com o intuito de solucionar a dificuldade em transformar dados em informações.

Segundo dados divulgados pela Amper Energia Humana (2022), em 2021 o Twitter alcançou a marca de 17,2 milhões de usuários no Brasil, minerar dados advindos dessa

plataforma é algo a se pensar. Uma pesquisa realizada pelo *Digital Marketing Institute*, em 2019, constatou que quatro em cada dez usuários do Twitter realizaram uma compra depois de verem o item sendo anunciado na plataforma via anúncio, pagos ou não, e conteúdos promovidos por um *digital influencer*. Sendo assim, fica evidente como um usuário pode ser influenciado através das opiniões de posicionamento de *influencers* digitais, inclusive com o sua intenção de voto (AZEVEDO, 2022).

Assim, como forma de compreender e enaltecer os resultados finais que serão apresentados, este trabalho visa responder às seguintes Questões de Pesquisa (QP):

QP1: As técnicas de mineração utilizadas são capazes de verificar se os níveis de popularidade dos candidatos às eleições presidenciais de 2022 no Twitter condizem com os resultados das eleições nas urnas?

QP2: É possível realizar uma predição de resultados com base na Análise de Sentimentos aplicada no Twitter em datas ou períodos específicos?

Sendo assim, este trabalho tem por objetivo verificar se o desempenho dos candidatos nas eleições presidenciais brasileiras em 2022 tem por base a popularidade na rede social Twitter.

1.3 Objetivos

O objetivo geral deste trabalho constitui-se utilizar técnicas de mineração de dados textuais para analisar comentários da rede sociais Twitter, a fim de mapear os níveis de popularidade dos candidatos à presidência da República em 2022. Sendo assim, constata como o posicionamento dos candidatos afeta sua popularidade em relação aos eleitores e, ao final, comparar a popularidade construída na rede social com os resultados finais da eleição. Tal verificação será feita através da participação dos candidatos nos debates presidenciais, de forma que o posicionamento dos mesmos influencie em sua popularidade perante aos eleitores.

1.3.1 Objetivos Específicos

Nos tópicos a seguir estão dispostos os objetivos específicos para este trabalho:

- Mapear a popularidade de cada candidato à presidência na rede social Twitter durante debates de TV aberta;

- Comparar os resultados da eleição presidencial com as informações obtidas na mineração de texto e pesquisas eleitorais registradas pelo TSE;
- E analisar a viabilidade da rede social Twitter como ferramenta de predição eleitoral.

1.4 Organização

O presente trabalho se organiza como segue. O Capítulo 1 apresentou a introdução e alguns aspectos de estudo e relevância deste trabalho. O Capítulo 2 apresenta a revisão literária, aprofundando nos conceitos trabalhados e estudados. Os procedimentos e métodos utilizados encontram-se detalhados no Capítulo 3. O Capítulo 4 descreve o desenvolvimento da pesquisa e o Capítulo 5 traz os resultados obtidos com análise detalhada desta pesquisa. Por fim, no Capítulo 6, encontram-se as conclusões deste trabalho.

2. REVISÃO LITERÁRIA

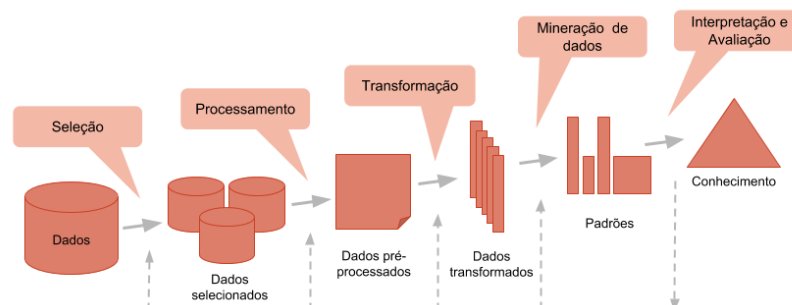
Nesta seção será apresentada os principais conceitos utilizados na construção deste trabalho, seguindo as definições, conceitos e considerações de estudiosos e especialistas em suas determinadas áreas.

2.1 Mineração de Dados

Segundo Amaral (2016), mineração de dados são processos que visam a exploração e análise de grandes volumes de dados a fim de encontrar padrões, previsões, erros e elaborar associações. A mineração de dados tem por base a descoberta de padrões em dados, eliminando tendências e limitações nas quais poderiam ocorrer em avaliações humanas, feitas através de algoritmos de aprendizado de máquina, *machine learning*¹. Além disso, é possível descrever mineração de dados como um conjunto de técnicas para descrição e predição, baseados em grandes massas de dados (BRAGA, 2005).

A mineração de dados está contida num processo denominado Descoberta de Conhecimento em Bases de Dados, do inglês *Knowledge Discovery in Databases* (KDD), processo iterativo para identificar nos dados padrões que sejam válidos, novos, potencialmente úteis e interpretáveis, mapeando dados de baixo nível e transformando-os em formas que possam ser mais completas, abstratas e úteis. Enquanto a KDD, como exemplificado na Figura 1, foca em extrair conhecimento útil a partir de uma base de dados, mineração de dados seria apenas uma parte deste processo. (FAYYAD, PIATETSKY-SHAPIRO e SMITH, 1996).

Figura 1 - Processos do KDD



Fonte : (MOURA, 2019)

¹ O *machine learning* é uma subcategoria da inteligência artificial e automatiza efetivamente o processo de construção de modelos analíticos e permite que as máquinas se adaptem a novos cenários de forma independente (FIA, 2021).

Moura (2019) explica que a primeira etapa consiste na Seleção de um conjunto de dados oriundos de diversas fontes, os quais farão parte da análise. Em seguida, parte-se para a etapa do Processamento, em que consiste em fazer a verificação da qualidade dos dados. Nessa etapa, é feita a limpeza, correção e remoção de dados inconsistentes. Logo após, é feita a etapa de Transformação dos Dados, aplicando técnicas transformação como normalização e agregação.

Na sequência, são realizadas as etapas de Mineração de Dados, em que visa verificar uma hipótese ou descobrir algum padrão nos dados obtidos. Por fim, é feita a interpretação e avaliação, com o intuito de avaliar e validar se o desempenho do modelo criado foi satisfatório. O processo de KDD pode ser descrito como iterativo, pois o resultado de uma atividade depende de outra, e interativo pois pode sofrer interversões externas. Além disso, o processo pode ser repetido diversas vezes (MOURA, 2019).

O processo de Descoberta de Conhecimento em Bases de Dados (KDD), é voltado para análise de dados estruturados, ou seja, possuem algum tipo de estrutura em sua composição. Nas próximas seções, serão descritas técnicas para trabalhar em descobertas de dados não estruturados, que será a principal composição deste trabalho.

2.2 Mineração de Texto

Segundo Ambrosio e Moraes (2007), Mineração de Texto é um Processo de Descoberta de Conhecimento, que utiliza técnicas de análise e extração de dados a partir de textos, frases ou apenas palavras. Ainda segundo os autores, essa atividade envolve a aplicação de algoritmos computacionais aos quais processam textos, identificando informações úteis e implícitas que não seriam tão simples de adquirir utilizando métodos tradicionais de consultas uma vez que, são informações contidas em formatos não estruturados, ou seja, sem um formato definido.

Um das principais contribuições da área são as buscas em documentos de textos, em seus mais variados tipos, como sites Web, e-mail, campos em base de dados e diversos outros documentos que podem ser encontrados. Com isso, pode-se obter informações específicas contidas nestes documentos, podendo inferir análises qualitativas e quantitativas sobre os mesmos, além de poder compreender melhor os textos dispostos nestes documentos. (AMBROSIO, MORAIS, 2007).

2.3 Recuperação da Informação

Mooers (1951) descreve Recuperação da Informação como o processo ou método pelo qual um potencial usuário de informação é capaz de converter a sua necessidade de informação em uma lista real de citações e documentos em um acervo contendo informações úteis para ele. A Recuperação da Informação (RI), como área Ciência da Computação, tem por objetivo prover ao usuário a facilitação de acesso a informações de cunho pessoal.

A Recuperação da Informação trata da representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas Web, catálogos online, registros estruturados e semiestruturados, objetos de multimídia etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de seu interesse. (YATES e NETO, 2013).

Ao longo dos anos, a RI obteve um grande crescimento em termos de escopo, passando o foco de indexações textuais e recomendação de documentos úteis para modelagem, classificação de textos, arquitetura de sistemas, interfaces de usuários, visualização e filtragem de dados e linguagens. Esta área ainda pode ser vista tanto pelo lado computacional, como processamentos, quanto pelo lado do usuário, visando o estudo do comportamento do usuário, com foco na melhoria da UX² (YATES e NETO, 2013).

2.4 Análise de Sentimentos

Matos (2015) diz que Análise de Sentimentos é uma tarefa de identificação para averiguar se a opinião expressada sob algum determinado assunto texto é positiva ou negativa. Vale ressaltar-se que, as opiniões e sentimentos, assim como seus conceitos relativos à avaliação, atitude, emoção e humor, influenciam no comportamento humano, fazendo com o que a visão de mundo tenha influência na percepção e opinião de outras pessoas.

A polarização das palavras trata-se de caracterizar uma frase através do grau de emotividade contido na mesma, através da avaliação de positividade, neutralidade ou negatividade. Para isso, é utilizado métodos léxicos, por meio de dicionários capazes de classificar o grau de polaridade de uma palavra (DE QUEIROZ e ALMEIDA, 2020).

² Experiência do Usuário (*User Experience*)

Na análise de sentimentos, a finalidade é classificar o texto que será analisado. Através do *machine learning*, a classificação busca encontrar uma função que descreva diferentes tipos de dados. Além disso, ainda captura informações contidas em dados textuais através de métodos de Processamento de Linguagem Natural (PLN) (SILVA; STABILE, 2016).

Segundo Vicente (2019), a análise recebe o texto a ser analisado e um identificador único para o mesmo. Assim, a partir da análise feita, é gerado um *score*³ que varia de 0 a 1, sendo 0 referente a um sentimento negativo e 1 a um sentimento positivo. O *score* é gerado por técnicas de classificação de palavras, classificando-as como positivas e negativas, verificando a existência das mesmas no texto e a quantidade presente.

Sendo assim, quanto mais palavras positivas conter no texto, maior a chance de conter sentimentos positivos ligados a ele e, quanto mais palavras negativas, maior chance de ter sentimentos negativos ligados ao texto. Além do *score* de positivo e negativo, é possível exibir os tópicos e palavras mais citados (palavras chaves) para a análise (VICENTE, 2019).

2.5 Redes Sociais

Bernardo e Goulart (2011) definem rede social como uma estrutura composta por pessoas e organizações, conectadas por um ou vários tipos de relações, que partilham valores e objetivos em comum. Uma das características das redes sociais são o compartilhamento de informações, conhecimentos, interesses e esforços em busca de objetivos em comum.

Para Zenha (2018), rede social online se enquadra como um ambiente digital disposto numa interface virtual própria, possuindo perfis humanos que dispõem de afinidades, interesses, pensamentos e maneiras de expressão em comum. Ainda segundo a autora, é interessante considerar as redes sociais como uma malha de relacionamentos, tanto afetivos quanto profissionais entre os indivíduos.

Rodrigues (2022) explica que os termos Rede Social e Mídia Social são termos facilmente confundidos. Ainda segundo o autor, Mídia Social é o uso de tecnologias para tornar interativo o diálogo entre pessoas e Rede Social é uma estrutura social formada por pessoas que compartilham interesses similares. O propósito principal das Redes Sociais é conectar pessoas, enquanto a Mídia Social abrange diferentes tipos mídias, como vídeos e blogs, podendo considerar as redes sociais como categoria da mesma (RODRIGUES, 2022).

³ *Score* é um substantivo da língua inglesa que significa pontuação.

2.6 Twitter

Criado em 2006 por Jack Dorsey, Evan Williams, Biz Stone e Noah Glass, o Twitter é uma rede social com foco em microblog (atualizações breves em poucos caracteres), que permite ao usuário enviar e receber informações curtas, de até 280 caracteres, por meio de sua plataforma. As atualizações são exibidas na tela principal dos usuários em tempo real, em suas multiplataformas, como o site na web e aplicativos em dispositivos móveis (WIKIPEDIA, A ENCICLOPÉDIA LIVRE, s.d.).

Alguns termos encontrados nesta rede social podem ser descritos como:

- *Tweet*: Texto de até 280 caracteres publicado nesta rede social.
- *Retweet*: Compartilhamento de um *tweet*;
- *Hashtag ou #*: *tags* com assuntos ou interesses em comum. Por exemplo, atualmente, no horário nobre da Rede Globo é exibido o programa Big Brother Brasil (BBB) 22. Os usuários utilizam a *tag* #BBB22 para comentar assuntos que consideram relevante deste programa;
- *User*: Usuário da rede social. Pode ser usado como meio de interação ao ser mencionado por outros usuários. Um usuário é descrito através de @ (arroba) único, como se fosse sua identificação;
- *Follower*: Ato de seguir um usuário nesta rede social;
- *Trending Topics*: São os assuntos em alta no momento. São atualizados de acordo com as interações e relevância dados pelos usuários da rede.

Segundo dados apresentados pela Agência Mestre (2022), uma pesquisa realizada pela Opinion Box com 800 pessoas constatou que 11% dos entrevistados consideram o Twitter a rede social mais utilizada; Para 61% dos entrevistados, acessar a rede uma vez ao dia é um hábito; Já 57% utilizam a rede social há mais de 3 anos; 44% dos usuários passaram a utilizar mais a rede social devido à pandemia; E 61% dos entrevistados pretendem continuar usando nos próximos 12 meses (AGÊNCIA MESTRE, 2022).

A pesquisa também levantou dados referentes aos hábitos de utilização dos usuários. Em questões de múltipla escolha, 66% dos entrevistados utilizam a rede para acompanhar as notícias. Além disso, 49% dos usuários utilizam a rede social para interagir com publicações e amigos 47% e consumir conteúdos relacionados à humor.

2.7 Eleições

Um dos primeiros relatos de eleições para governanças locais, segundo Brasil (1988) , ocorreu em 1532 para o Conselho Municipal da Vila de São Vicente/SP. Com a crescente demanda das pressões populares e da economia, ter representantes para atuar nas decisões da corte passou a ser cada vez mais necessário. De acordo com Cajado, Dornelles e Pereira (2014), a primeira eleição para presidente no Brasil ocorreu em 1894, com participação de 2,2% da população.

Paes (2017) diz que a atual constituição adota o regime democrático representativo, o qual um representante é eleito democraticamente pela população. Segundo a Constituição Federal Brasileira de 1988, no cap. IV Atr. 14, a soberania popular será exercida pelo sufrágio universal e pelo voto direto e secreto, com valor igual para todos. Com isso, a Constituição garante que as eleições sejam exercidas de forma justas, contribuindo para os atos democráticos de um país.

Paes (2017) ainda afirma que o processo eleitoral brasileiro sofreu diversas mutações entre os períodos históricos vivenciados pelo país, demonstrando que o poder de votar percorreu um caminho árduo até que se tornasse algo sólido. A atividade de exercer o voto chegou a terras brasileiras com os portugueses, com os primeiros núcleos de povoação. Essa herança obtida do povo lusitano em eleger os administradores dos povoados foi um dos princípios das eleições brasileiras (TSE, 2009).

Com o fim das eleições indiretas exercidas na Ditadura Militar, o cenário eleitoral passou por diversas mudanças. Com a introdução das comunicações sociais, debater sobre o processo eleitoral ficou cada vez mais fundamental. Com os debates televisionados e pesquisas eleitorais, a imagem dos candidatos e propagandas eleitorais ganharam cada vez mais destaque e apoio público (CAJADO, DORNELLES E PEREIRA, 2014).

Ainda segundo os autores, com o advento da era da informação, as eleições não poderiam ficar à mercê da tecnologia. A informatização eleitoral hoje é um traço estruturante da prática democrática brasileira, com a urna eletrônica é um grande exemplo deste marco. Nos anos 80, o desenvolvimento de novos recursos tecnológicos e a ampliação no uso dos meios digitais possibilitaram as condições para o recadastramento geral do eleitorado em meio eletrônico.

2.8 Orange Data Mining

Segundo Larhud (s.d.), *Orange Data Mining*⁴ é um programa de computador que possibilita a mineração de dados de forma simples e intuitiva, trabalhando com aprendizado de máquina (*machine learning*) e visualização de dados (*data visualization*), utilizando software de código aberto. Seus fluxos de trabalho são baseados em análise de dados interativos atrelados a uma grande opção de ferramentas, incluindo uma variedade de técnicas de visualização, exploração, pré-processamento e modelagem de dados.

A utilização do *Orange* cabe tanto a uma interface intuitiva quanto ao módulo mais avançado com a utilização linguagem de programação *Python*. Considerando que, a mineração de dados busca identificar padrões de similaridade ou de não conformidade nos conjuntos de dados, o *Orange* oferece diversos recursos de base estatística para o processamento dos dados (LARHUD, s.d.).

Ainda segundo o autor, o programa trabalha com arquivos nos seguintes formatos:

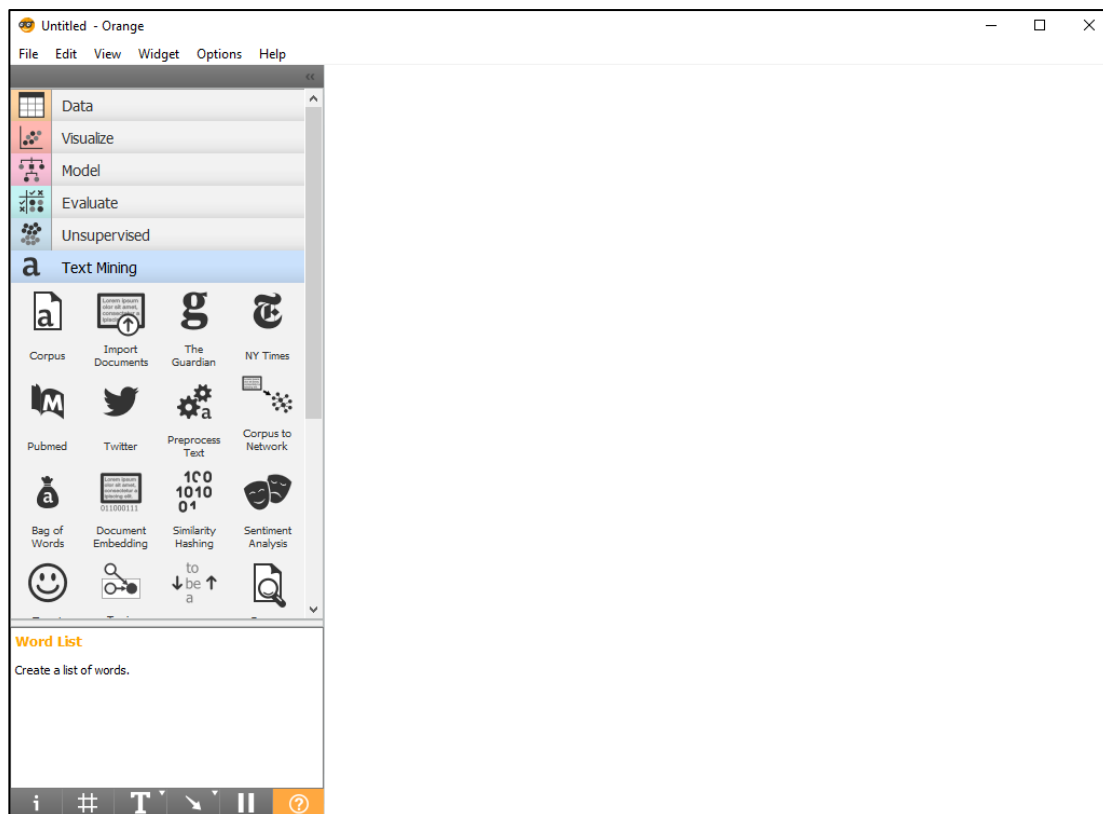
- *Excel*, arquivos delimitados por vírgulas e tabulações (.xlsx, .csv, .tab);
- Dados on-line no formato do Google Planilhas;
- Imagens (.jpg, .tiff, .png) através do complemento *Image Analytics*;
- Arquivos de texto (.txt, .docx, .odt) com o complemento, *Orange3-Corpus*;
- Banco de dados *PostgreSQL* e *MSSQL*, por meio do complemento SQL.

Com a disposição destes arquivos, é possível criar um fluxo de trabalho de análise de dados. A instalação padrão do *Orange* inclui diversos algoritmos de aprendizado de máquina, pré-processamento e visualização de dados, dispostos em seis conjuntos de complementos (dados, visualização, classificação, regressão, avaliação e não supervisionados). Com isso, é possível ler os dados, mostrar uma tabela de dados, selecionar recursos, comparar algoritmos de aprendizado, visualizar elementos de dados, dentre outras diversas possibilidades (LARHUD, s.d.).

A Figura 2 abaixo ilustra a página principal do programa. Na tela inicial tem-se dois módulos, na esquerda estão localizadas as categorias e *widgets* disponibilizadas pelo programa e, na direita, está a área de trabalho do programa. Através das categorias, é possível selecionar itens específicos e arrastá-los até a área de trabalho e conectá-los, gerando um fluxo de informações. Logo após finalizado, o fluxo então é gerado e assim podendo extrair informações do mesmo.

⁴ Disponível em : <<https://orangedatamining.com/>>

Figura 2 - Interface do Orange Data Mining



Fonte: Elaborado pela autora, 2022.

2.9 Trabalhos Correlatos

Os trabalhos desenvolvidos até então são de suma importância para o meio acadêmico, pois essas pesquisas servem de base para compreender os respectivos cenários políticos, levantando questões sobre os mesmos, a fim de obter questionamentos e, sobretudo, métodos para melhoria da política no geral.

Pereira (2019), em sua monografia para Universidade Federal do Rio Grande do Norte (UFRN), “**Análise De Sentimentos Da População Brasileira Em Relação A Eleição Presidencial De 2018 Através Da Rede Social Twitter**”, através da API do Twitter, classificou as publicações referentes aos candidatos à presidente da República nas eleições de 2018. Para isso, utilizou aprendizado de máquina supervisionado e a análise de sentimentos, relacionando-as a fim de verificar qual era o sentimento que um usuário dispunha sobre um determinado candidato. Como objetivo, este trabalho buscou a relação entre apoio demonstrado na mídia social e o desempenho obtido no resultado da eleição e em pesquisas eleitorais da época. Assim, seria possível avaliar a viabilidade ou não da rede social Twitter como uma

ferramenta de predição eleitoral de baixo custo e tempo. Como conclusão, o autor afirma que os resultados obtidos não foram tão próximos, não atingindo uma conclusão clara para comprovar que a Análise de Sentimentos na rede social Twitter seja uma alternativa em relação às pesquisas tradicionais.

De Paula Filho e Garcia (2015), em seu artigo, “**Predição Do Resultado Das Eleições Presidenciais Do Brasil Baseado Em Tuítes**”, utilizam as eleições presidenciais do ano de 2014, baseados em mensagens de texto publicadas na rede social Twitter. Para isso, foram utilizados aproximadamente 3 milhões e 200 mil mensagens, de mais de 460.000 usuários, com referências aos principais candidatos da época. Os autores concluíram que a contagem de tuítes um dia anterior dos dois turnos das eleições para presidente no Brasil foi capaz de prever corretamente os resultados de ambos os turnos. Para tal análise, foi construída uma interface web utilizando a linguagem de programação PHP, com a API do Twitter.

Teixeira (2019), na monografia apresentada à Universidade Federal de Ouro Preto, “**Análise de Sentimento dos Usuários do Twitter em Relação à Atual Situação Política do Brasil**”, teve como principal objetivo pesquisar e aplicar técnicas estatísticas e computacionais a fim de minerar dados vindos do Twitter para a identificação do sentimento das pessoas em relação à atual situação política do Brasil. Para isso, a autora utilizou a linguagem R, aplicando-se numa base contendo 107.240 *tweets*. Como conclusão, a autora diz que, predominantemente, a maior parte dos usuários da rede se sentem positivamente quanto ao governo. Porém, a autora ressalta que, durante sua análise, aconteceu uma manifestação a favor do atual presidente do Brasil e de sua política, o que pode ter influenciado na análise. Além disso, pela localização das origens do *tweets*, seria possível uma robotização das publicações, o que poderia ter influenciado no resultado da análise.

Matos *et al.* (2020), em seu artigo “**Recuperação e Classificação de Sentimentos de Usuários do Twitter em Período Eleitoral**”, buscaram analisar os sentimentos expressos pelos usuários no Twitter relacionados aos candidatos que se elegeram à presidência do Brasil no ano de 2018 e, com estes dados, predizeram o resultado das eleições com base nessas postagens. Para isso, utilizou-se o *software Orange Data Mining*. Como técnicas, as autoras empregaram a técnica de análise de opinião, com objetivo de fazer classificação automática dos sentimentos em positivos, negativos e neutros. Para melhorar a análise e visualização dos dados, foram utilizadas nuvens de palavras e gráficos de distribuição de frequência. Como resultado, foram detectados diversos sentimentos negativos nas postagens e a emoção de surpresa foi a que mais se destacou para ambos os concorrentes. As autoras sentiram bastante dificuldades

em obter um resultado preciso devido às altas taxas de rejeição dos eleitores aos candidatos e a polarização.

De Queiroz e Almeida (2020), em seu artigo, “**Uma metodologia de análise de sentimentos dos candidatos as eleições presidenciais de 2018 no Twitter**” visam apresentar uma metodologia para análise de sentimentos aplicada em tweets realizados pelos candidatos com maior intenção de voto no primeiro turno das eleições presidenciais brasileiras de 2018. Os *tweets* foram submetidos a técnicas de pré-processamento, uso de dicionários léxicos e algoritmos para agrupamento de dados. Os resultados obtidos permitiram a identificação de comportamentos como o grau de positividade ou negatividade dos candidatos, considerando fatores como a divulgação de pesquisas de intenção de votos realizadas, grau de similaridade e a frequência de termos utilizados nas postagens. Os *tweets* foram coletados diretamente nos perfis oficiais dos candidatos as eleições presidenciais de 2018. Assim como no trabalho de Matos *et al.* (2020), os autores também sentiram dificuldades em efetuar a análise.

Dutra e Francisco (2018), em seu artigo “**Text Mining: Análise de Sentimentos nas Eleições 2018**”, fizeram a análise através da extração de dados do Twitter, aplicando as técnicas analíticas de *Text Mining* e *SNA*⁵ para identificar os sentimentos expressos por usuários da rede social em relação às Eleições para Presidente do Brasil de 2018. Para isso, utilizaram a linguagem de programação *Python* juntamente às bibliotecas de Mineração de Dados. Para a análise, foram coletados dados entre os dias 24 a 30 de junho de 2018. Dos 1.204.036 *tweets* coletados, os resultados apontaram a maior parte dos tweets para Jair Bolsonaro e Manuela Dávila, e maior classificação positiva para Bolsonaro que, segundo os autores, teve um destaque pela sua estratégia de uso de redes sociais. Como resultados, os autores classificaram os resultados satisfatórios e apontam potencialidades do uso dessas técnicas para futuras eleições.

Hammes *et al.* (2021), em seu artigo para o Salão do Conhecimento, “**Machine Learning e suas aplicações em Processos Eleitorais**”, os autores visaram compreender como as ferramentas de *Machine Learning* podem contribuir para solucionar questões que dificultam a veracidade das informações que são disseminadas referentes aos candidatos à presidência. Os autores, após analisarem diversas pesquisas, verificaram a existência de algoritmos de *Machine Learning* capazes de identificar contas automatizadas em redes sociais e *Fake News*, bem como analisar o sentimento do eleitorado em relação aos candidatos e suas campanhas eleitorais. Como metodologia, foram utilizadas a linguagem de programação *Python* e bibliotecas de

⁵ Análise de redes sociais (ARS; em inglês: social network analysis, SNA) é uma interpelação da Sociologia, da Psicologia Social e da Antropologia (FREEMAN, 1996), sendo uma análise metódica de Redes Sociais (WIKIPEDIA, s.d.)

Aprendizado de Máquina. Como conclusão, os autores perceberam que os algoritmos de *Machine Learning* são eficazes para identificar contas automatizadas dentro das redes sociais. Os autores ainda afirmaram que os estudos da aplicação de algoritmo para analisar os sentimentos de eleitores se mostram promissores, porém devem ter ajustes e refinamentos para que se possa ter uma aplicação de forma sistemática dentro do *marketing* político.

Esta monografia, diferentemente dos trabalhos citados, visa proporcionar uma análise com maior duração, estendendo-se desde a candidatura oficial dos candidatos à presidência, no dia 16 de agosto de 2022, quando se dão início às campanhas eleitorais até o dia 30 de outubro, data final das eleições com o fim do segundo turno, totalizando assim 75 dias. Pretende-se, ainda, coletar dados relacionados aos candidatos, preferencialmente em períodos de destaque, como debates, a fim de ter-se uma variedade de dados para conclusões futuras.

A coleta de dados ocorreu durante três momentos:

- Período antes do debate;
- Período durante o debate;
- E período posterior ao debate;

A coleta em diferentes momentos permite comparar se houve ou não diferenças na popularidade de um candidato durante o período do debate. Dessa forma, é possível realizar uma análise mais profunda e compreender em qual momento o posicionamento do candidato afetou sua popularidade.

A Tabela 1 faz um comparativo entre os trabalhos citados anteriormente.

Tabela 1 - Comparativo entre os Trabalhos Correlatos

TRABALHOS	PEREIRA (2019)	DE PAULA FILHO E GARCIA(2015)	TEIXEIRA (2019)	MATOS <i>ET AL.</i> (2020)	DE QUEIROZ E ALMEIDA (2020)	DUTRA E FRANCISCO (2018)	HAMMES <i>ET AL.</i> (2021)
REPOSITÓRIO	UFRN	ANAIS DO IV BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING	UFOP	FGV	REVISTA DE ENGENHARIA E PESQUISA APLICADA	CONGRESSO TRANSFORMAÇÃO DIGITAL 2018	SALÃO DO CONHECIMENTO
FERRAMENTAS UTILIZADAS	<i>Python</i> , API do Twitter, Nltk, <i>Scikit-Learn</i> , <i>Naive Bayes</i> E SVM	Php, API do Twitter,	R, API do Twitter	<i>Orange Data Mining</i> , API do Twitter	R, API do Twitter, <i>K-Means</i> , Tf-Idf	Python, API do Twitter, <i>Naive Bayes</i> , Nltk, <i>Scikit-Learn</i> , SNA	Python, API do Twitter, <i>Naive Bayes</i> , Nltk, <i>Scikit-Learn</i> , API do Keras
TWEETS COLETADOS	1.014.752 Total	4.700.000 Total	107.240 Total	1.000 Por Candidato	4.600 Total	1.204.036 Total	15.000.000 Total
OBJETIVO CENTRAL	Analisar comentários advindos da rede social Twitter, a fim de verificar os níveis de popularidade dos candidatos à Presidência da República em 2018 e	Este trabalho se propõe a fazer uma revisão da literatura sobre os trabalhos já publicados a respeito do uso de mensagens públicas do Twitter para predição do resultado de	Aplicar técnicas estatísticas e computacionais de Mineração de Dados para análise dos sentimentos expressos nos <i>tweets</i> minerados a respeito da atual	Analisar os sentimentos expressos pelos usuários no Twitter, referentes aos candidatos que concorreram à presidência do Brasil no ano de 2018, e	Apresentar uma metodologia de análise de sentimentos que realiza a extração, tratamento e classificação dos <i>tweets</i> gerados, interpretando o sentimento transmitido por	Este trabalho aplica técnicas analíticas de <i>Text Mining</i> e SNA para identificar os sentimentos expressos por usuários da rede social Twitter em relação às Eleições para Presidente do Brasil de 2018. DUTRA E FRANCISCO (2018)	O objetivo geral do trabalho foi compreender como a ferramenta de <i>Machine Learning</i> pode contribuir para os processos eleitorais. HAMMES <i>ET AL.</i> (2021)

TRABALHOS	PEREIRA (2019)	DE PAULA FILHO E GARCIA(2015)	TEIXEIRA (2019)	MATOS <i>ET AL.</i> (2020)	DE QUEIROZ E ALMEIDA (2020)	DUTRA E FRANCISCO (2018)	HAMMES <i>ET AL.</i> (2021)
	compará-los com os resultados finais desta eleição. PEREIRA (2019)	eleições presidenciais no Brasil, apresentar as etapas dos modelos mais utilizados atualmente para predição do resultado de eleições presidenciais e comparar o desempenho de cada um deles considerando o cenário político brasileiro. DE PAULA FILHO E GARCIA(2015)	situação política do Brasil. TEIXEIRA (2019)	predizer o resultado das eleições com base nessas postagens. MATOS ET AL. (2020)	cada mensagem e identificando seu grau de polaridade a partir da aplicação de dicionários para classificação das palavras em positivas ou negativas. QUEIROZ E ALMEIDA (2020)		
RESULTADOS	O autor afirma que resultados obtidos tiveram certo distanciamento entre si, afirmando que analisar os	Os autores conseguiram prever o resultado levando em consideração <i>tweets</i> coletados um dia anterior	A pesquisa da autora obteve resultados capazes de atender os objetivos. O sentimento positivo teve	Foi concluído que o Twitter é uma boa ferramenta para usuários expressarem seus sentimentos	Os autores encontraram similaridade nos resultados, sugerindo adaptações para trabalhos futuros.	Apesar de encontrar dificuldades quanto ao uso da aplicação por fins gramaticais, os autores consideraram que obteve-se um resultado satisfatório	Os autores concluíram que algoritmos de <i>Machine Learning</i> são eficazes para identificar contas automatizadas dentro das redes

TRABALHOS	PEREIRA (2019)	DE PAULA FILHO E GARCIA(2015)	TEIXEIRA (2019)	MATOS <i>ET AL.</i> (2020)	DE QUEIROZ E ALMEIDA (2020)	DUTRA E FRANCISCO (2018)	HAMMES <i>ET AL.</i> (2021)
	sentimentos dos usuários do Twitter não seja tão viável como preditor eleitoral.	à eleição. Como sugestão, os autores propõem propor soluções para as etapas de Filtragem de Dados e Redução de viés de dados, aprimorar o modelo baseado em tuítes, além de desenvolver modelos que prevejam 3 ou mais candidatos.	predominância entre a maioria das pessoas com relação à atual situação política do Brasil, considerando assim que a maior parte da população brasileira apoia o governo atual.	no período eleitoral. Porém, não foi possível prever os resultados das eleições, as autoras elucidam que um dos principais motivos para isso seriam as altas taxas de rejeição e a crescente polarização do cenário político.		no processo fim a fim, e que de certa forma retrata a realidade.	sociais. Quanto a identificação de <i>Fake News</i> , foi possível identificar dois diferentes modelos de aprendizagem supervisionado com resultados precisos e satisfatórios e, por fim, os estudos da aplicação de algoritmo para analisar os sentimentos de eleitores se mostram promissores.

Fonte: Elaborado pela autora, 2022.

3. METODOLOGIA

Segundo Oliveira (2013), a mineração de dados deixou de ser apenas um processo de pesquisa científica em meio acadêmico. Atualmente, devido aos seus bons resultados, passou a ser aplicada em ambientes que exigiam tomadas de decisões baseadas em dados, buscando conhecimentos que estão ocultos em bases de dados e tornando-os úteis para gerar conhecimentos.

Sendo assim, este capítulo demonstrará a abordagem utilizada na aplicação da mineração de dados em contexto político, gerando conhecimento para averiguar as questões de pesquisa definidas neste trabalho, bem como as atividades referentes aos processos utilizados para obter as informações resultantes desta pesquisa, as caracterizações deste estudo, a população e amostras, os instrumentos e o processo de coletas de dados, além da estratégia de análises dos dados adotada neste contexto.

3.1 Natureza da Pesquisa

Para Lakatos e Marconi (2003), a pesquisa se descreve como um procedimento formal, que possui o método de pensamento reflexivo, requerendo tratamento científico e constituindo o caminho para conhecer a realidade ou descobrir verdades parciais. Sendo assim, esta monografia busca sua verdade parcial em observar os sentimentos dispostos pelo usuário em prol de obter respostas para validação de uma rede social como meio de prever o resultado de uma eleição.

Ainda segundo a autora, em toda pesquisa deve ter um objetivo determinado, o qual pretende-se alcançar ao final, tornando o problema explícito e aumentando o conhecimento sobre o tema. Um objetivo deve responder as perguntas “Por quê?”, “Para quê?”, “Para quem?”. Além de ter características intrínsecas ou extrínsecas, serem teóricas ou práticas, de modos gerais ou específicos e de curta ou longa duração.

Oliveira (2021) diz que as pesquisas podem ser classificadas em quatro âmbitos:

- quanto à natureza ou finalidade, sendo básica ou aplicada;
- quanto à abordagem, sendo qualitativa, quantitativa ou ambas;
- quanto ao objetivo, sendo descritiva, explicativa ou exploratória;
- quanto ao procedimento técnico, sendo bibliográfica, documental, operacional, experimental, levantamento, dentre outros;

Haja vista as afirmações expostas pelos autores citados anteriormente, pode-se considerar que esta pesquisa, quanto à sua natureza, tem-se finalidade aplicada, pois visa gerar conhecimento para a população sobre um bem público, envolvendo interesses locais.

Quanto à sua abordagem, esta pesquisa encaixa-se como quantitativa – qualitativa, pois além de reunir dados estatísticos sobre as eleições, foram feitas interpretações e análises subjetivas, dando origem a um estudo completo sobre um objeto.

Considerando seus objetivos, pode-se classificá-la como pesquisa explicativa, pois seu objetivo consta relacionar a popularidade de um candidato com o fato de ser eleito ou não, identificando os fatores determinantes ou contribuintes para a análise do fenômeno, como sentimentos dos usuários perante àquele candidato.

Por fim, levando em consideração os procedimentos adotados na coleta de dados, esta pesquisa pode ser classificada em duas categorias:

- levantamento, pois será coletado dados de um grupo restrito, sendo eles os usuários da rede social Twitter, gerando conclusões sobre o todo a partir da análise exploratória de dados;
- experimental, pois controla as variáveis de pesquisa e realiza experimentos em detrimento das questões de pesquisa.
- exploratória, pois o pré-requisito para esta é ter variáveis que serão associadas entre si e comparadas, tema presente neste trabalho.

As ciências são caracterizadas pelo uso de métodos científicos, mesmo sendo ausentes em alguns ramos das mesmas.

Método é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo de produzir conhecimentos válidos e verdadeiros, traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista. (LAKATOS, MARCONI, 2003).

Neste sentido, a presente monografia se encaixa como método hipotético-dedutivo. Segundo Prodanov e Freitas (2013) *apud* (GIL, 2008, p. 12), quando os conhecimentos obtidos sobre algum assunto não são suficientes o bastante para explicar um fenômeno, tem-se um problema. Para tentar explicar as dificuldades dispostas no problema, são formadas então as hipóteses, que conseqüentemente deduzirão as conseqüências que serão testadas e falseadas (tornar falsas as conseqüências deduzidas das hipóteses).

Ainda segundo os autores, o método hipotético-dedutivo inicia-se quando há um problema no conhecimento científico, passa-se pela formulação de hipóteses e por um processo de inferência dedutiva, o qual testará a predição da ocorrência de fenômenos abrangidos pela referida hipótese.

Aplicando-se a este trabalho, tem-se que a problemática engloba verificar se a uma rede social pode servir de base para fazer a predição eleitoral. Partindo deste princípio, formula-se a hipótese de que a popularidade de um candidato na rede social Twitter pode torná-lo ou não o presidente eleito, através dos sentimentos expressados pelos usuários da rede social sobre este candidato.

Assim, passa-se pelo processo de inferência dedutiva, o qual a partir da teoria é feita a previsão de suas consequências, no caso deste trabalho, a partir do processo de KDD será extraído as informações necessárias para averiguar ou não se a hipótese é certa. Caso o objetivo não seja atingido, a hipótese é falseada e são sugeridas melhorias para a mesma.

3.2 População e Amostra

Neste trabalho, a população designada foram os usuários da rede social Twitter. Para coleta de dados, foram adotados como amostra usuários aleatórios, preferencialmente brasileiros, que façam uso desta rede.

3.3 Instrumentos

Para o desenvolvimento deste trabalho utilizou-se computadores portados com acesso à internet e ferramentas necessárias para a construção deste estudo. Primeiramente, será feita a instalação e configuração do software *Orange Data Mining*, adicionando a API do *Twitter* para extração dos dados que serão utilizados adiante.

Para consolidação do trabalho, a entrada de dados serão os *tweets* postados pelos usuários, recuperados através da API e, os resultados após a mineração, serão armazenados na própria máquina para compará-los posteriormente.

3.4 Métodos e Procedimentos

Como procedimentos, este trabalho será conduzido, através de experimentos, a verificar a análise da viabilidade da rede social *Twitter* como uma ferramenta de predição

eleitoral ágil e acessível, traçando um paralelo entre a aprovação de cada candidato na rede com seu desempenho na eleição para presidente, além de mensurar o quão fiel as informações obtidas conseguem predizer de fato popularidade leva um candidato a ser eleito.

O primeiro passo para a mineração de dados é a seleção. Para isso, foi utilizado a API do Twitter e, através do *Orange Data Mining*, serão recuperadas uma coleção de comentários sobre o candidato selecionado. Por conseguinte, é necessário informar um *tag*⁶ com o conteúdo a ser pesquisado e a quantidade de *tweets* que são retornados. Para este estudo, delimitaremos em 1.000 *tweets* por candidato.

Por conseguinte, será feita a limpeza dos dados. Ao recuperar uma base de dados, os mesmos tendem a vir incompletos, inconsistentes e com valores fora do padrão. Com este propósito, utiliza-se técnicas para tornar tais dados úteis para fazer a análise. Neste caso, a limpeza ocorrerá com a remoção de *stopwords*⁷, caracteres especiais como @ (arroba), # (hashtag), artigos e conectivos, dentre outros tipos de caracteres que não serão úteis para essa análise.

Seguindo para a fase de transformação dos dados, nesta é feita a consolidação dos dados de forma que sejam atrativos para a mineração em si. Nesta etapa, serão verificados se os dados estão nulos, se há inconsistência e valores aberrantes. Caso haja, são tratados para que a base de dados possua valores úteis para a análise.

Então, ao atingir a etapa de mineração de dados, serão aplicados métodos para extrair informações, como análise de sentimentos e análise do perfil de usuário. Logo, é avaliado se os padrões foram válidos e se representam algum conhecimento. Por fim, serão apresentados os conhecimentos extraídos deste processo, a fim de tomar-se decisões baseados no resultado apresentado.

Os métodos e procedimentos foram propostos baseados nos objetivos específicos presentes neste trabalho, conforme descrito na Tabela 2:

Tabela 2 - Métodos e Procedimentos

Objetivo	Tarefa
Pré-processar e transformar as publicações obtidas na rede social Twitter, utilizando a técnica de mineração de texto (<i>Text Mining</i>) sobre cada candidato;	a) Instalar e compreender as funções do <i>Orange Data Mining</i> ; b) Fazer o processo de Knowledge Discovery in Databases (KDD);

⁶ *Tag* é um substantivo da língua inglesa que significa etiqueta.

⁷ *Stopwords*, ou palavras vazias, são palavras que são removidas antes ou após o processamento em um linguagem natural.

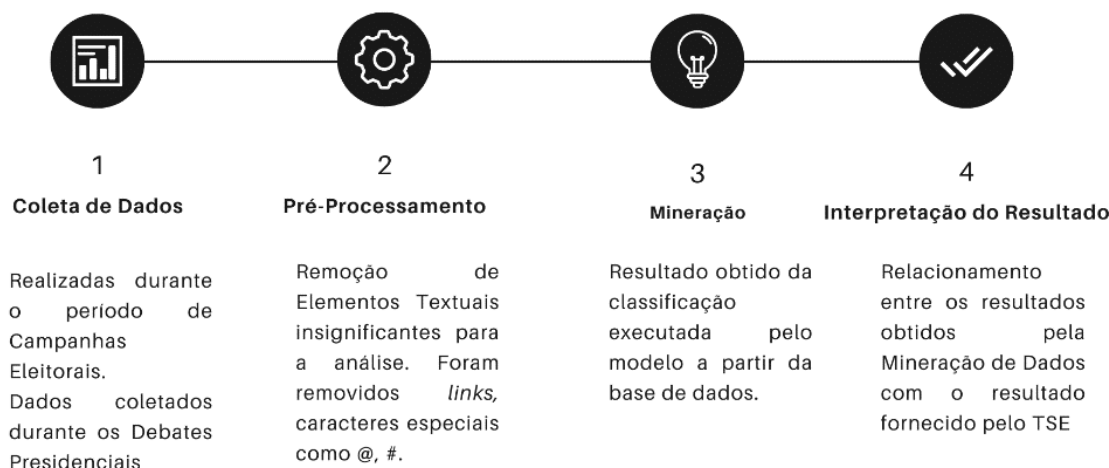
	c) Elaborar e treinar o modelo para classificação.
Mapear a popularidade de cada candidato à presidência na rede social Twitter durante debates de TV aberta;	a) Aplicar o modelo e recolher as análises; b) Armazenar dados recolhidos.
Comparar os resultados da eleição presidencial com as informações obtidas na mineração de texto e pesquisas eleitorais registradas pelo TSE;	a) Traçar relações entre os dados obtidos da mineração e comparar com o resultado do TSE para validar a hipótese.
Analisar a viabilidade da rede social Twitter como ferramenta de predição eleitoral.	a) Verificar, através dos resultados obtidos, se redes sociais podem ou não ter efeitos preditores como pesquisa eleitoral.

Fonte: Elaborado pela autora, 2022.

4. DESENVOLVIMENTO DA PESQUISA

Nesta seção será abordada, de forma sequencial, toda a estrutura disposta na pesquisa dessa monografia. A metodologia proposta busca esclarecer, transpassando desde a coleta até os resultados, o objetivo geral proposto neste trabalho e, ao final, realizar uma análise dos resultados junto às conclusões e considerações finais.

Figura 3 - Metodologia empregada



Fonte: Elaborado pela autora, 2022.

Como exibido na Figura 3 acima, este trabalho segue-se uma metodologia de trabalho. Nas próximas seções, serão detalhadas o funcionamento de cada passo realizado.

4.1 Coleta de dados e pré-processamento

Como exemplificado na Figura 3, primeiramente ocorre a coleta de dados. Para este fim, foi necessário utilizar uma *API* disponibilizada pelo próprio aplicativo do Twitter. Esta *API* fornece uma chave única para acesso, denominada *API KEY* e uma chave secreta, sendo a *API KEY SECRET*. Após obter estes dados, então foi necessário inseri-los no Orange Data Mining para coletar os dados.

Para acrescentar a chaves da *API* no programa, primeiramente criou-se um novo projeto e inseriu o *widget* do Twitter (Figura 4).

Figura 4 - Widget



Fonte: Elaborado pela autora, 2022.

Logo após, ao clicar sobre o widget, será exibida uma tela, Figura 5, em que será possível informar a *query* e informar as chaves disponibilizadas pela API (Figura 6) . Esta *query* deverá conter as principais informações necessárias, sendo elas:

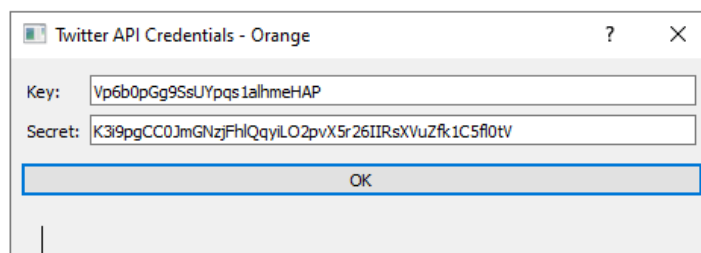
- Lista de palavras relacionadas ao tema;
- Tipo de conteúdo a ser buscado (conteúdo ou autor);
- A linguagem em que o *tweet* foi publicada;
- A quantidade máxima de *tweets*;
- *Retweets*, se necessário;
- E o conteúdo a ser coletado do texto (Conteúdo ou Descrição do Autor);

Figura 5 - Tela do Widget

The image is a screenshot of a software window titled "Twitter - Orange". The window contains a form for configuring a Twitter widget. At the top, there is a field for "Twitter API Key". Below that is a "Query" input field. Underneath the query field is a "Query word list:" label and an empty text area. The "Search by:" dropdown menu is set to "Content". The "Language:" dropdown menu is set to "Portuguese". The "Max tweets:" field has a checked checkbox and a value of "1000". There are two unchecked checkboxes: "Allow retweets:" and "Collect results:". The "Text includes" section has two checkboxes: "Content" (checked) and "Author Description" (unchecked). At the bottom of the form is a "Search" button. The window also has standard OS window controls (minimize, maximize, close) and a help icon.

Fonte: Elaborado pela autora, 2022.

Figura 6 - Inserção de Dados API



Fonte: Elaborado pela autora, 2022.

Para este trabalho, serão considerados que *string* de busca será o nome como o qual o candidato lançou-se à campanha; a busca como Conteúdo; a linguagem do *tweet* como Português; a quantidade máxima de tweet como 1.000 por candidato; e o texto coletado como conteúdo.

Ao recolher todos os dados, então é feita a etapa de pré-processamento. Nesta etapa, os dados foram preparados, através da limpeza e transformação, para receber a análise. Toda e qualquer limpeza acarreta no sucesso da mineração de dados. Sendo assim, o primeiro ajuste a ser realizados é a conversão dos *tweets* coletados para a versão em letras minúsculas, logo após foram removidas as *URL's* e *tags* HTML, para considerar apenas o texto dessas marcações. Optou-se por não remover os acentos para evitar discrepâncias no sentido do tweet.

Em seguida, dividiu-se a coleção de sentimentos em unidades menores, denominada *tokens*, utilizando as expressões regulares como métodos de separação. Por conseguinte, para que obtiveste um melhor resultado da análise, optou-se por uma técnica normalização linguística que consiste na redução das palavras em seu radical, denominada *stemming*. Sendo assim, os prefixos, sufixos e características de gênero, número e grau das palavras são removidos, reduzindo o número de atributos em até 50% (AMBRÓSIO e MORAIS, 2007).

Seguidamente, removeram-se as *stopwords*, palavras consideradas irrelevantes para a análise, como artigos, conjunções e preposições, que comumente aparecem nos comentários para dar sentido ao *post*, mas não detém de nenhuma informação relevante. Sendo assim, elaborou-se uma lista de *stopwords* e a anexou-se ao conjunto já disponibilizado pelo *Orange*. A lista com *stopwords* pode ser encontrada no ANEXO A disponibilizado ao final deste trabalho. Abaixo, encontra-se duas figuras ilustrando o resultado do Pré-Processamento.

Ademais, utilizou-se expressões regulares para remover os caracteres especiais. E, por fim, para captar apenas os termos mais importantes do conjunto, utilizou-se a *Document Frequency* (DF). A DF visa a eliminação de palavras em determinados graus de frequência, sendo mínima ou máxima, por não contribuir com a análise dos sentimentos.

4.2 Montagem do modelo classificador

Após fazer toda a preparação dos dados, então é realizada então a análise de opiniões. A plataforma *Orange* disponibiliza diversos módulos de análise da NLTK (*Natural Language Toolkit*) para o diagnóstico. Para este trabalho, decidiu-se utilizar o *Vader* (*Valence Aware Dictionary for sEntiment Reasoning*), que é baseado em abordagem de dicionários léxicos de sentimento, regras gramaticais e convenções sintáticas, além de trabalhar com diversas linguagens. Este componente computa os sentimentos com pontuação positiva, negativa e neutra.

Como descrito por Ma (2020), a abordagem léxica se caracteriza como:

(...) a abordagem léxica significa que este algoritmo (*Vader*) constrói um dicionário que contém uma lista abrangente de recursos de sentimentos. Este dicionário lexical não contém apenas palavras, mas também frases. Todas as características lexicais foram classificadas quanto à polaridade e intensidade em uma escala de “-4: Extremamente negativo” a “+4 Extremamente positivo ” por 10 avaliadores humanos independentes. A pontuação média é então usada como indicador de sentimento para cada recurso lexical no dicionário.

A autora ainda descreve a utilização de regras gramaticais como:

Além do dicionário léxico do sentimento, existem estruturas que são neutras inerentemente, mas podem alterar a polaridade do sentimento ou modificar a intensidade de toda a frase. Em *Vader*, os desenvolvedores incorporaram várias regras heurísticas que tratam dos casos de pontuação, capitalização, advérbios e conjunções contrastantes (MA,2020).

Para calcular a pontuação de sentimento, o *Vader* examina o texto em busca de características sentimentais conhecidas, modifica a intensidade e a polaridade de acordo com as regras, resume as pontuações das características encontradas no texto e calcula o resultado final, utilizando a normalização. E, por fim, atribui uma pontuação que varia de (- 1, 1). Dessa forma, para calcular a pontuação final, é utilizado a seguinte fórmula matemática:

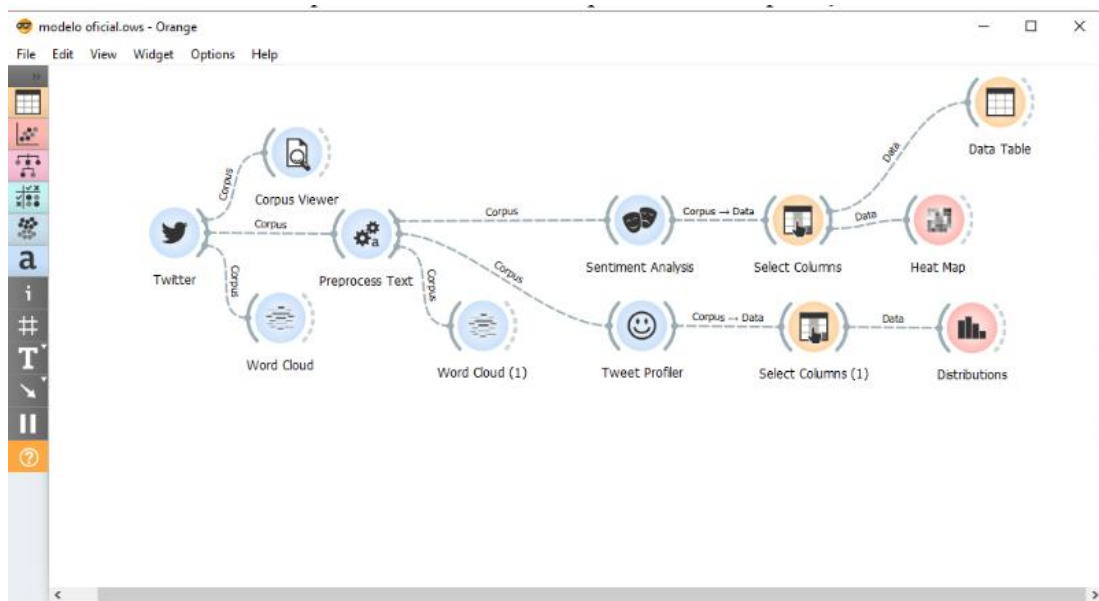
$$\frac{x}{\sqrt{x^2 + a}}$$

No *Vader*, o *alfa* é definido como 15, o que aproxima o valor máximo esperado de x . Além da pontuação composta da frase, *Vader* também retorna a porcentagem de recursos de sentimentos positivos, negativos e neutros (MA, 2020).

Sendo assim, os sentimentos foram classificados por pontuação numa escala de -1 a 1, o qual -1 representa o sentimento mais negativo e 1, o sentimento mais positivo. Além disso, ainda optou-se pela análise das opiniões dos usuários usando o módulo *Tweet Profiler*. Essa análise resulta numa decomposição de emoções referentes à *Anger* (raiva), *Disgust* (descontentamento), *Fear* (medo), *Joy* (alegria), *Sadness* (tristeza) e *Surprise* (surpresa). O resultado dessa avaliação foi exibido num gráfico de distribuição de frequência.

Além disso, foram inseridos *widgets* de *Word Cloud* como meio de identificar palavras frequentes, antes e depois do processamento de texto. Dando continuidade ao modelo, os sentimentos dos usuários foram exibidos num gráfico de calor, *heatmap*, facilitando a identificação dos sentimentos predominantes. A Figura 9 a seguir ilustra a montagem do modelo classificador.

Figura 9 - Modelo Classificador



Fonte: Elaborado pela autora, 2022.

4.3 Escopo de Experimentos

Considerando que o objetivo deste trabalho é a análise da viabilidade da rede social Twitter como meio de predição eleitoral, não foi levada em consideração a possibilidade de que uma parcela das publicações capturadas pelo *script* de busca tenha partido de *bots* (robôs) presentes na rede social. Além disso, foi considerado que todos os usuários coletados são eleitores aptos à votação, desconsiderando a idade, nacionalidade ou situações de pendência com o TSE.

Para tanto, foram analisados os debates da TV aberta com abrangência nacional, e também, a coleta de dados durante o dia da votação em ambos turnos como meio de verificação da proposta do trabalho, realizadas durante o primeiro e segundo turnos da eleição de 2022, que somados, representam 05 eventos. Tais eventos, foram escolhidos pela sua grande repercussão, que geraram enormes volumes de publicações na rede social Twitter, publicações essas que foram capturadas e armazenadas para fins de análise.

A Tabela 3, mostra a lista de emissoras de TV e a data em que ocorreu cada evento monitorado.

Tabela 3 - Debates Presidenciais 2022

Debate	Ordem do	Local	Data
	1º	Band, UOL, Folha de S. Paulo e TV Cultura	28/08
	2º	SBT, CNN, Veja, O Estado de S. Paulo, Nova Brasil FM e Terra.	24/09
	3º	TV Globo	29/09
	4º	Estadão, Rádio Eldorado, SBT, CNN, Veja, Nova Brasil FM e Terra	21/10
	5ª	TV Globo	28/10

Fonte: Elaborado pela autora, 2022.

5. RESULTADOS

Após a etapa de desenvolvimento da pesquisa, os dados foram gerados e então, analisados. Os resultados foram apresentados em três tipos de gráficos: Gráficos de barras; Gráficos de pizza; Gráficos Mapa de Calor (*heatmap*). Neste trabalho contém dois tipos de gráficos de barras, os que contém dados gerais das emoções e os que contém dados isolados por candidatos. No Gráfico de Barras que possui os dados gerais, é possível fazer um comparativo entre todos os candidatos presentes nos debates enquanto que, no que possui os dados individuais, pode-se fazer um comparativo entre as emoções dispostas sobre um único candidato.

O Gráfico Mapa de calor é um tipo de gráfico utilizado para visualizar os valores de atributos por classe, numa matriz bidirecional. Os valores são representados por cores: quanto maior um determinado valor, mais escura a cor representada (ORANGE DATA MINING, 2015). Dessa forma, há três cores que ilustram os resultados: Azul, para ilustrar sentimentos positivos; Rosa, para sentimentos negativos; Amarelo, para sentimentos neutros.

Dessa forma, é gerada uma imagem representando quais são os pontos em que ocorreram determinada atividade. Assim, quando há combinação entre classe e atributos em x e y eixos, é possível observar os locais onde os atributos são mais fortes e onde mais fraco, o que permite encontrar características típicas (discretas) ou valor intervalo (contínuo) para cada classe (ORANGE DATA MINING, 2015).

Ainda, no gráfico de Mapa de Calor, os *tweets* que continham dados parecidos foram agrupados em *clusters*⁸, utilizando o algoritmo *k-means*. Segundo Anastacio (2020), *K-Means* é um algoritmo de clusterização, disponível na biblioteca *Scikit-Learn*. É um algoritmo de aprendizado não supervisionado que avalia e agrupa os dados de acordo com suas características, sendo necessário definir um número de *clusters*. Para este trabalho, optou-se pela utilização de 100 *clusters*.

Por fim, os gráficos de Pizza apresentam os resultados isolados das emoções dispostas pelos usuários. Dessa forma, é possível comparar qual candidato teve predominância numa determinada emoção. Tanto neste gráfico quando no gráfico de Barras, as emoções ilustradas referem-se à *Anger* (raiva), *Disgust* (descontentamento), *Fear* (medo), *Joy* (alegria), *Sadness* (tristeza) e *Surprise* (surpresa).

⁸ *Cluster* pode ser definido como uma categorização e agrupamento de dados que contém padrões e características em comum entre as informações de maneira autônoma (COUTINHO, 2021).

Na eleição para presidente da República Federativa do Brasil, contou-se com treze inscrições para a disputa do cargo de presidente. Porém, houve dois participantes inaptos aos cargos, um por cancelamento de sua candidatura e outro por indeferimento, restando assim, onze candidatos às vagas. A Tabela 4 lista todas as informações dos candidatos à presidência. Os dados foram retirados do portal do TSE e adaptados pela autora .

Tabela 4 - Lista dos Candidatos à Presidência em 2022

Partido	Número do Candidato	Nome do Candidato	Nome na Urna	Situação
PDT	12	Ciro Ferreira Gomes	Ciro Gomes	Deferido
DC	27	José Maria Eymael	Constituinte Eymael	Deferido
NOVO	30	Luiz Felipe Chaves D'Avila	Felipe D'Avila	Deferido
PL	22	Jair Messias Bolsonaro	Jair Bolsonaro	Deferido
UP	80	Leonardo Péricles Vieira Roque	Léo Péricles	Deferido
PT	13	Luiz Inácio Lula da Silva	Lula	Deferido
PROS	90	Pablo Henrique Costa Marçal	Pablo Marçal	Cancelado
PTB	14	Kelmon Luis Da Silva Souza	Padre Kelmon	Deferido
PTB	14	Roberto Jefferson Monteiro Francisco	Roberto Jefferson	Indeferido
MDB	15	Simone Nassar Tebet	Simone Tebet	Deferido
PCB	21	Sofia Padua Manzano	Sofia Manzano	Deferido
UNIÃO	44	Soraya Viera Thronicke	Soraya Thronicke	Deferido
PSTU	16	Vera Lucia Pereira da Silva Salgado	Vera	Deferido

Fonte: TSE, adaptado pela autora, 2022.

Segundo a Lei 13.488/2017, é garantida a participação do candidato nos debates eleitorais, se o partido onde o mesmo é vinculado possuir ao menos cinco representantes

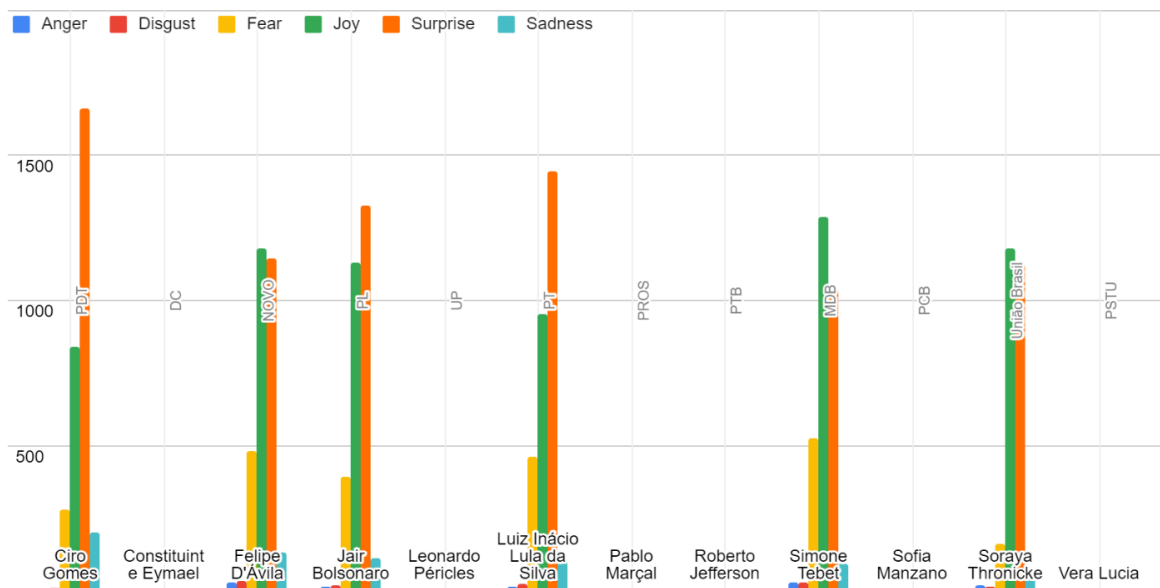
(somados entre câmara dos deputados e senado federal), os demais candidatos podem ou não serem convidados pelas emissoras. Os convites geralmente levam em conta a posição atual do presidenciável nas pesquisas eleitorais (PEREIRA, 2019).

Com base nisso, o presente trabalho realizou a análise de sentimentos de sete dos onze candidatos inscritos e deferidos na disputa presidencial do Brasil em 2022, tendo em vista suas participações nos debates em emissoras de rede aberta a nível nacional, sendo eles: Ciro Gomes, Felipe D'Ávila, Jair Bolsonaro, Lula, Padre Kelmon, Simone Tebet e Soraya Thronicke.

5.1 Dados Analisados

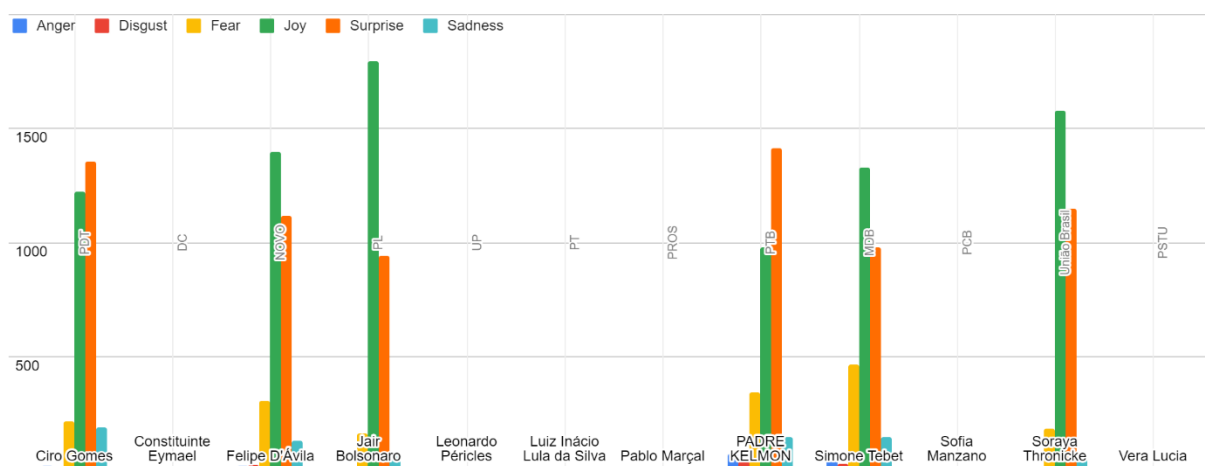
Os gráficos apresentados nas Figuras 10 e 11 demonstram a segregação de emoções dispersas pelos usuários perante aos candidatos, capturados através da busca de *tweets* e utilizados no processo de mineração de texto e análise de sentimentos. Por problemas técnicos, não foi possível realizar a coleta do terceiro debate realizado na TV Globo.

Figura 10 - Resultado do Primeiro Debate - 1º Turno



Fonte: Elaborado pela autora, 2022.

Figura 11 - Resultado do Segundo Debate - 1º Turno



Fonte: Elaborado pela autora, 2022.

5.2 Análise de Popularidade Individual – Primeiro Turno

Todas as publicações foram classificadas numa escala que vai de 1, para *tweets* positivos, 0 para imparciais ou ambíguos e -1 para publicações negativas, e apresentadas num gráfico de Mapa de Calor (*heatmap*), agregando em *clusters* os *tweets* que contém publicações com mesmo contexto. Além disso, foi utilizado gráficos de distribuição para demonstrar as emoções dispostas nos *tweets* dos internautas. Os dados apresentados nos tópicos abaixo são referentes ao dia 02/10/2022, referente ao primeiro turno da eleição.

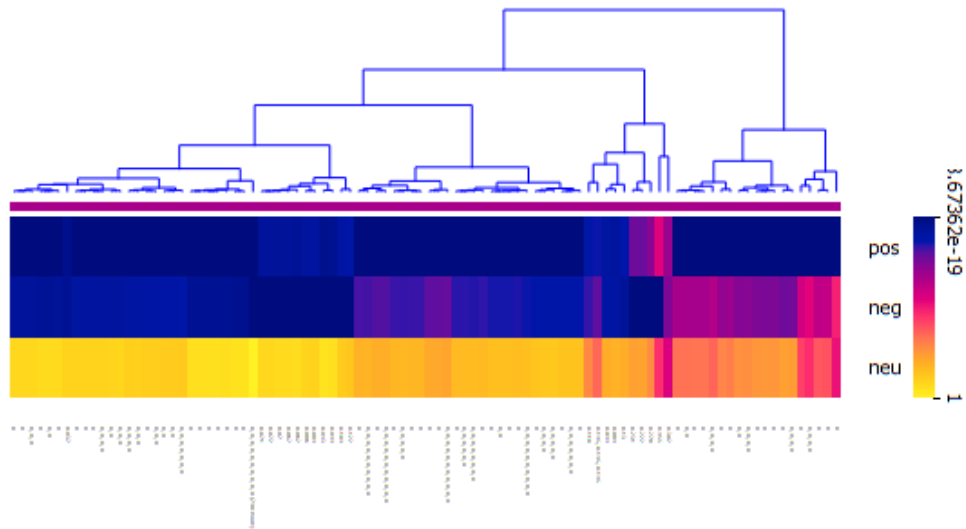
Para este trabalho, foram disponibilizados os dados individuais dos candidatos concorrentes ao Segundo Turno uma vez que, foi gerada uma quantidade massiva de dados. Dessa forma, os mesmos podem ser encontrados nos apêndices localizados ao final desta monografia. Nesta seção, decidiu-se apresentar dados isolados num único dia, sendo o dia da votação, como meio de avaliar se os *tweets* podem ser considerados como preditores eleitorais e para melhor visualização e compreensão dos resultados.

5.2.1 *Ciro Gomes*

Ciro Gomes esteve presente em todos os debates ocorridos no primeiro turno, mantendo sua popularidade estável. A seguir, representado pela Figura 12, encontra-se o *heatmap* do candidato referente ao dia 02/10/2022, data que então ocorreu o primeiro turno das eleições. Nota-se que o candidato *Ciro Gomes*, em cerca de 30% dos *clusters*, possui opiniões

negativas vinculadas à sua candidatura, o que pode ter explicado o 4º lugar no resultado final do primeiro turno.

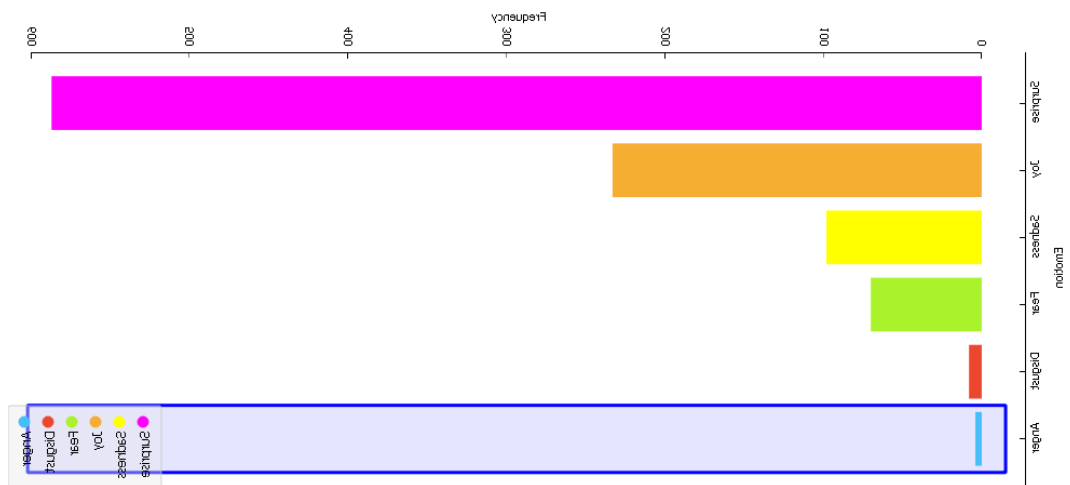
Figura 12 - Heatmap - Ciro Gomes



Fonte: Elaborado pela autora, 2022.

Além disso, como exemplificado na Figura 13 abaixo, o candidato deteve *surprise* como maior emoção ligada à sua candidatura, não demonstrando apoio ou rejeição pela população brasileira. O candidato deteve *joy* como segunda maior emoção presente nos *tweets* e *sadness* como terceira opção.

Figura 13 - Gráfico de Emoções - Ciro Gomes



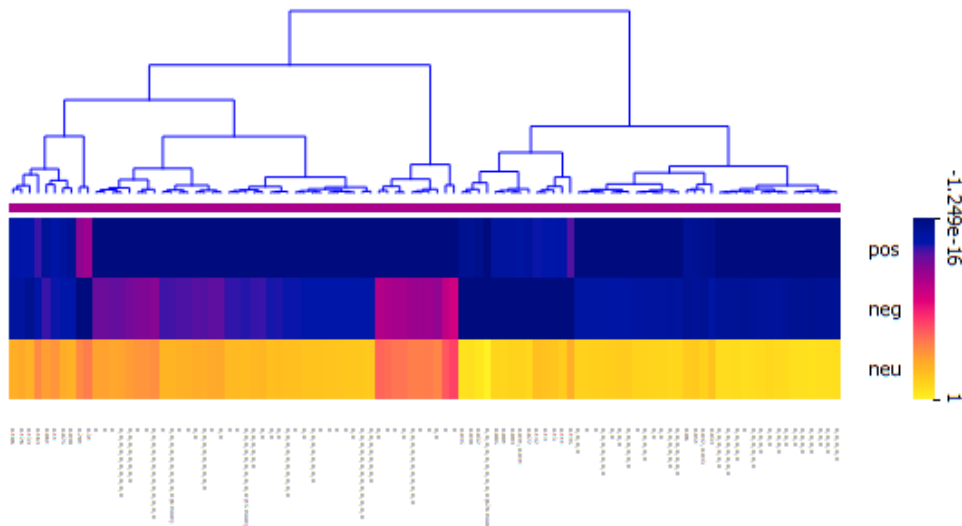
Fonte: Elaborado pela autora, 2022.

5.2.2 Felipe D'Avila

Felipe D'Avila, assim como Ciro Gomes, esteve presente em todos os debates ocorridos no primeiro turno, mantendo sua popularidade pouco estável no primeiro debate e notou-se uma melhora no segundo debate.

A seguir, representado pela Figura 14, encontra-se o *heatmap* do candidato referente ao dia 02/10/2022, data que então ocorreu o primeiro turno das eleições. Nota-se que o candidato Felipe D'Avila, em cerca de pouco mais de 20% dos *clusters*, possui opiniões negativas vinculadas à sua candidatura.

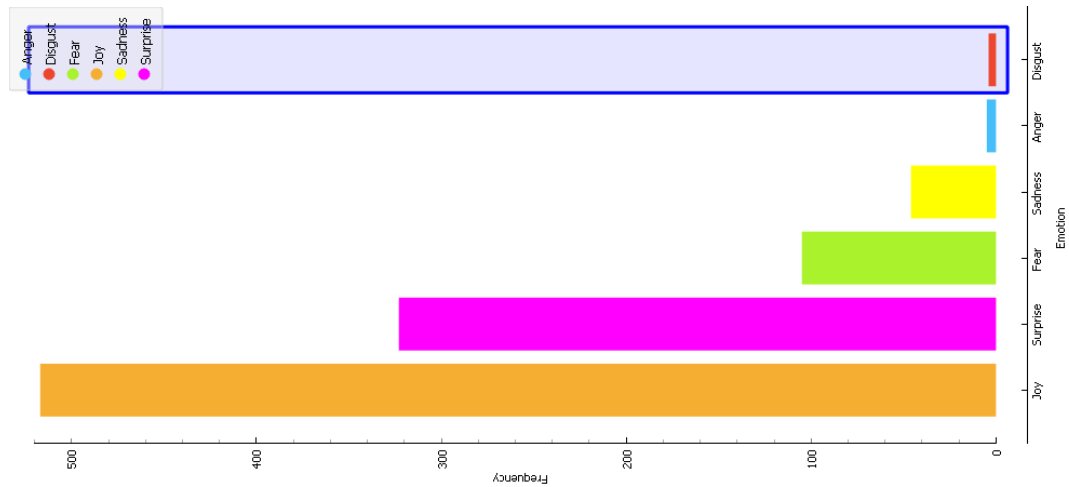
Figura 14 - Heatmap - Felipe D'Avila



Fonte: Elaborado pela autora, 2022.

Porém, ainda como apresentado no gráfico da Figura 15, com a melhoria na estabilização a partir do segundo debate, nota-se que a maior parte das emoções dispostas por este candidato eferem-se a *joy*, seguida por *surprise* e *fear*. Ainda que o sentimento negativo domine pouco menos da metade do gráfico, D'Avila conseguiu obter uma certa aprovação de seus eleitores, justificando assim que a maior emoção relacionada a si seja *joy*.

Figura 15 - Gráfico de Emoções - Felipe D'Avila



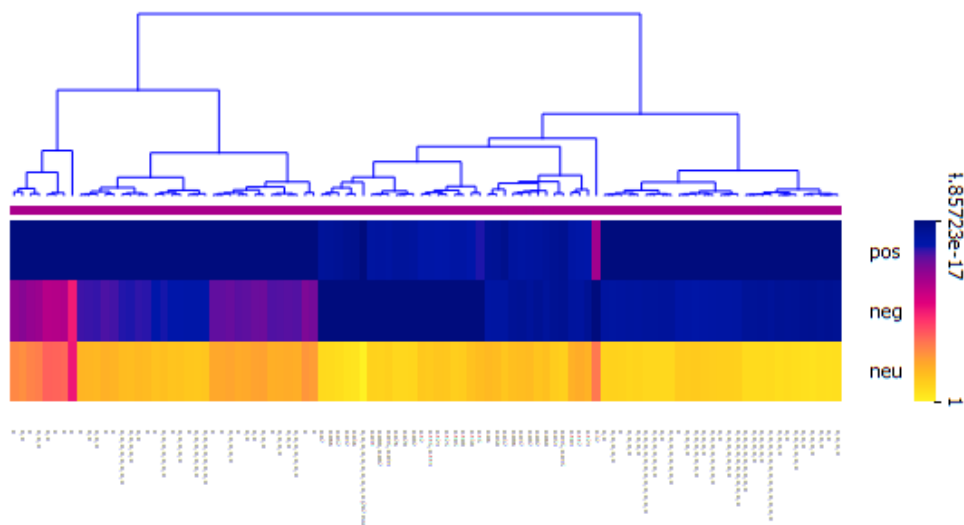
Fonte: Elaborado pela autora, 2022.

5.2.3 Jair Bolsonaro

Jair Bolsonaro, assim como os candidatos já apresentados, teve participação em todos os debates ocorridos, tendo uma boa aceitação pelo público. Os dados detalhados deste candidato podem ser encontrados APÊNDICE A e APÊNDICE B.

Na Figura 16, encontra-se o *heatmap* do candidato, indicando que o mesmo possui muitos sentimentos positivos vinculados à sua candidatura. Ainda pode-se observar alguns trechos com sentimentos negativos distribuídos ao longo do gráfico, indicando um certo nível de desaprovação quanto à sua campanha.

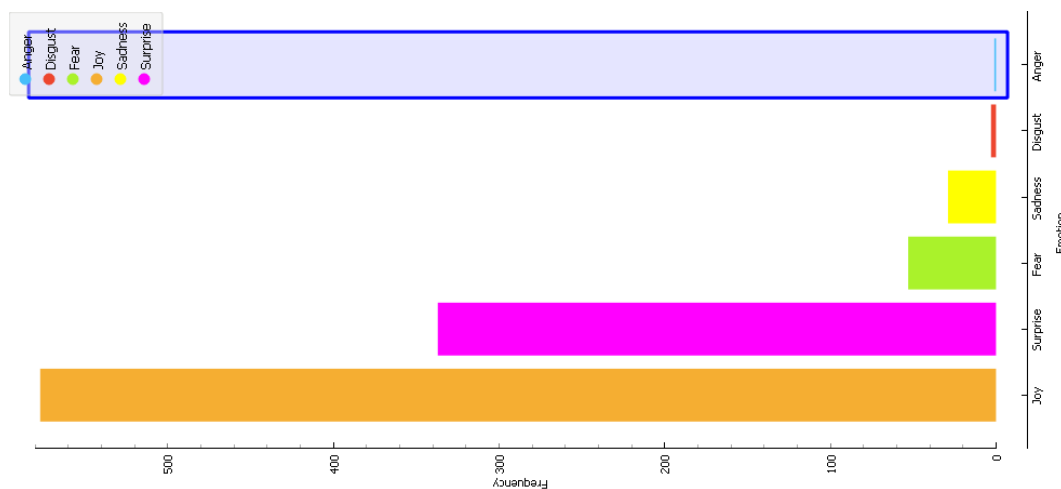
Figura 16 - HeatMap Candidato Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

Além disso, como retratado na Figura 17, observa-se ainda que o candidato Jair Bolsonaro possui um número considerável de emoções *joy*, o que se repetiu durante toda sua campanha como candidato à presidente da República.

Figura 17 - Gráfico de Emoções - Jair Bolsonaro



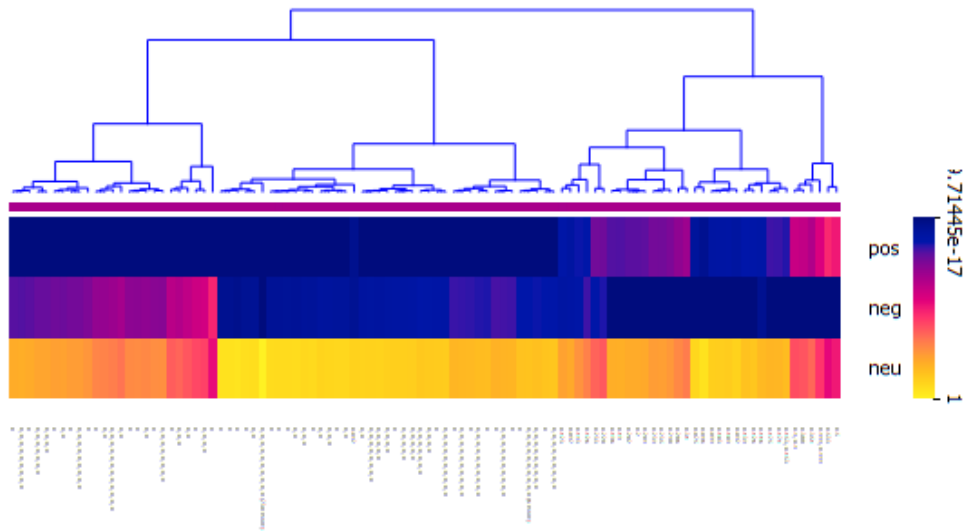
Fonte: Elaborado pela autora, 2022.

5.2.4 Luiz Inácio Lula da Silva

O candidato Luís Inácio Lula da Silva esteve presente em 2 dos 3 debates ocorridos durante o período eleitoral referente ao primeiro turno. Os dados detalhados deste candidato podem ser encontrados APÊNDICE E.

A Figura 18 apresenta o *heatmap* do sentimento coletados sobre este candidato no primeiro turno. É possível observar que o candidato possui os dois extremos, além de possuir muitos sentimentos positivos nas publicações, também é possível observar que uma parte considerável do gráfico indica cenários de rejeição por este candidato.

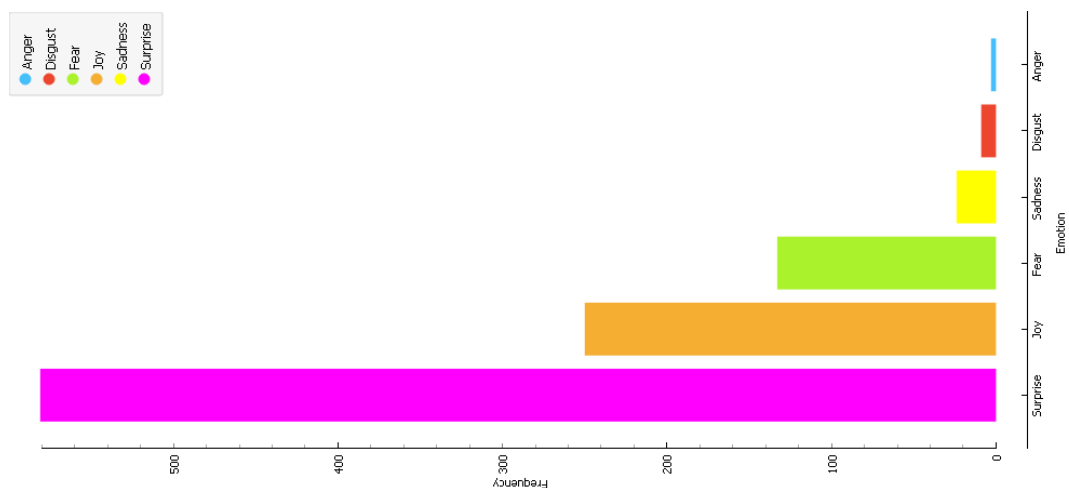
Figura 18 - *HeatMap* candidato Luís Inácio Lula da Silva



Fonte: Elaborado pela autora, 2022.

Ainda, como mostra a Figura 19 abaixo, a maior emoção disposta sobre este candidato seria *surprise*, não necessariamente indicando um sentimento positivo ou negativo. Além disso, o candidato deteve um número mediano da emoção *joy* e, pode-se destacar, uma quantidade considerável de sentimento *fear*.

Figura 19 - Gráfico de Emoções – *Luiz Inácio Lula da Silva*

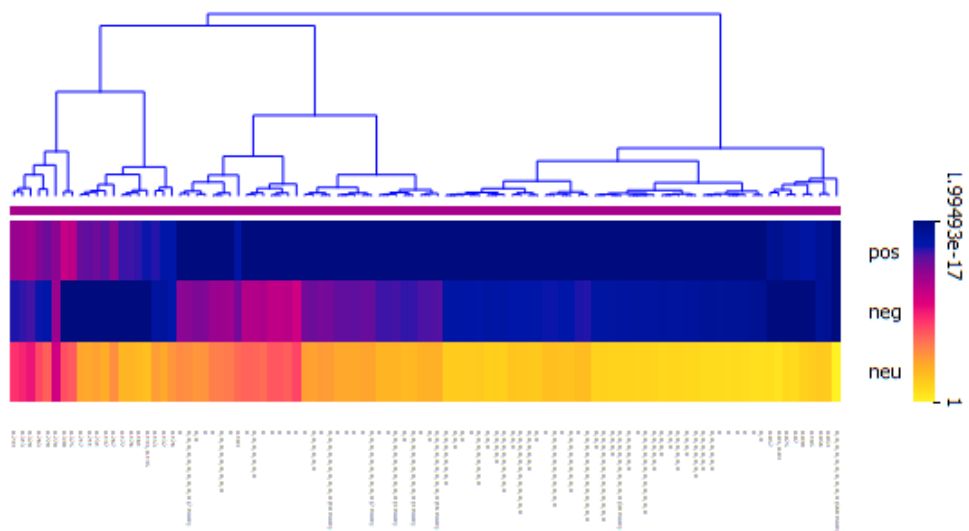


Fonte: Elaborado pela autora, 2022.

5.2.5 Padre Kelmon

O candidato Padre Kelmon esteve presente em 2 dos 3 debates ocorridos durante o período eleitoral referente ao primeiro turno. A candidatura do Padre foi inesperada, pois surgiu como substituto de Roberto Jefferson, que teve sua candidatura impedida pelo TSE. A Figura 20 apresenta o *heatmap* do candidato. Nota-se que o Padre possui muitos sentimentos negativos relacionados à sua candidatura, chegando a cerca de 50% do gráfico.

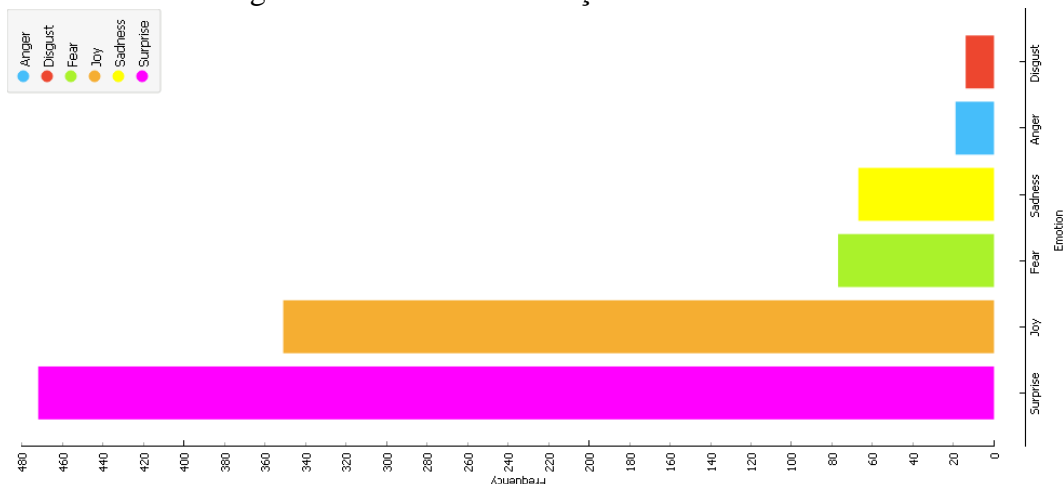
Figura 20 - *Heatmap* - Candidato Padre Kelmon



Fonte: Elaborado pela autora, 2022.

Além disso, a Figura 21 indica que a maior emoção presente foi a *surprise, seguida por joy*. Chama-se atenção para os sentimentos de *Fear, Sadness, Anger* e *Disgust* que possuem níveis consideráveis comparados a outros candidatos.

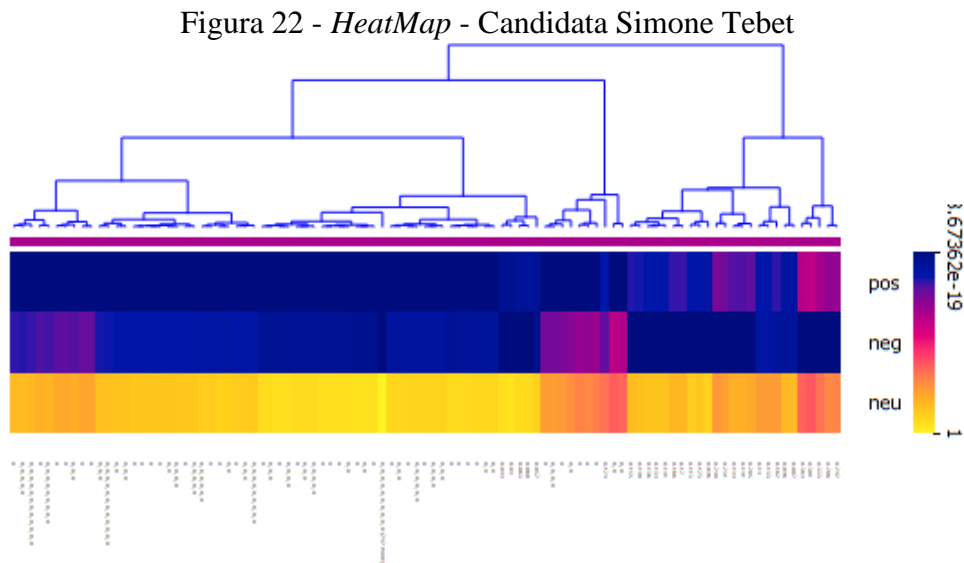
Figura 21 - Gráfico de Emoções - Padre Kelmon



Fonte: Elaborado pela autora, 2022.

5.2.6 Simone Tebet

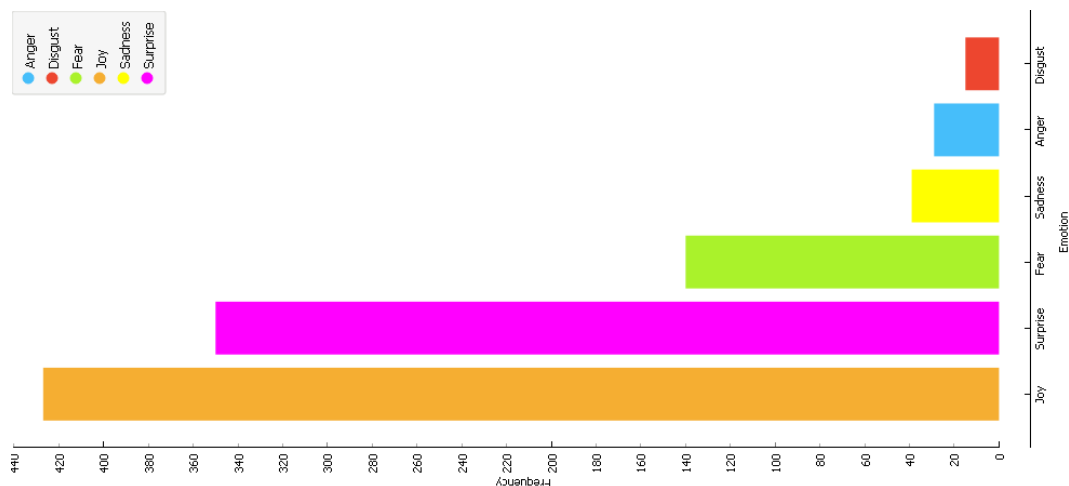
A candidata Simone Tebet esteve presente em todos os debates ocorridos durante o período eleitoral referente ao primeiro turno. A Figura 22 refere-se ao *heatmap* da candidata, observa-se que Simone possui um bom índice de aceitação pelos usuários do Twitter, visto que os sentimentos negativos ocupam pouca porcentagem do mapa.



Fonte: Elaborado pela autora, 2022.

Além disso, como retratado na Figura 23, observa-se ainda que a candidata Simone Tebet possui um número considerável de emoções *joy*. Simone obteve bons resultados na sua campanha, fazendo com que fosse vista como uma alternativa entre os principais candidatos.

Figura 23 - Gráfico de Sentimentos - Simone Tebet

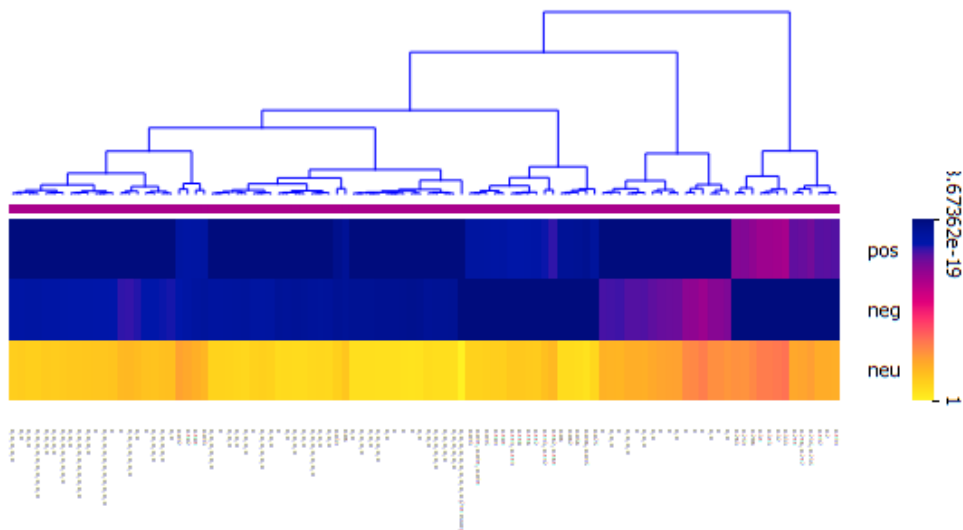


Fonte: Elaborado pela autora, 2022.

5.2.7 Soraya Thronicke

A candidata Soraya Thronicke foi anunciada pouco antes da data limite para candidatura, esteve presente em todos os debates ocorridos durante o período eleitoral referente ao primeiro turno. Como descrito na Figura 24, Soraya obteve uma boa popularidade, concentrando o maior número de reações positivas à sua candidatura. Nos debates, Soraya gostava de indagar os candidatos populares com polêmicas, o que despertou interesse do público. (BRASIL, 2022)

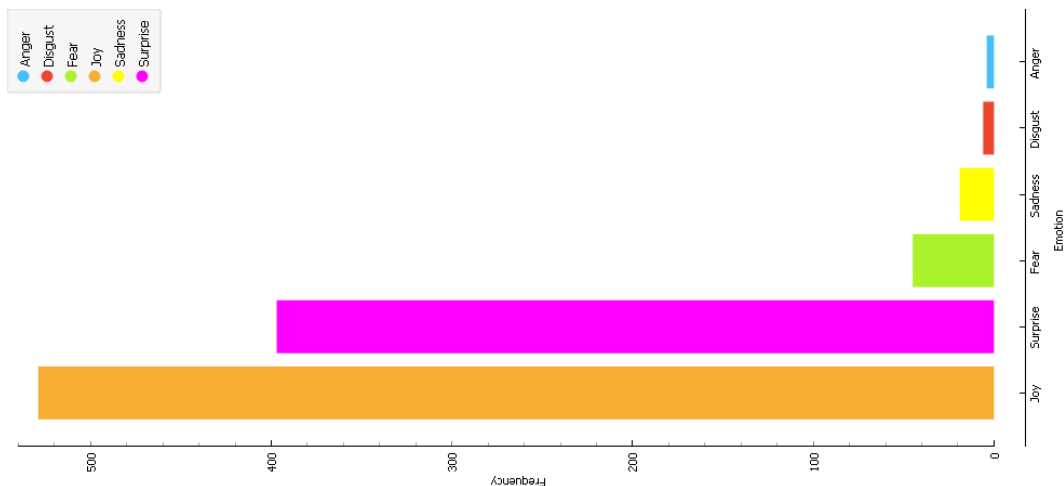
Figura 24 - HeatMap - Candidata Soraya Thronicke



Fonte: Elaborado pela autora, 2022.

Além disso, como retratado na Figura 25, observa-se ainda que a candidata possui um número considerável de emoções *joy*, o que se repetiu durante toda sua campanha como candidata à presidente da República.

Figura 25 - Gráfico de Sentimento - Soraya Thronicke

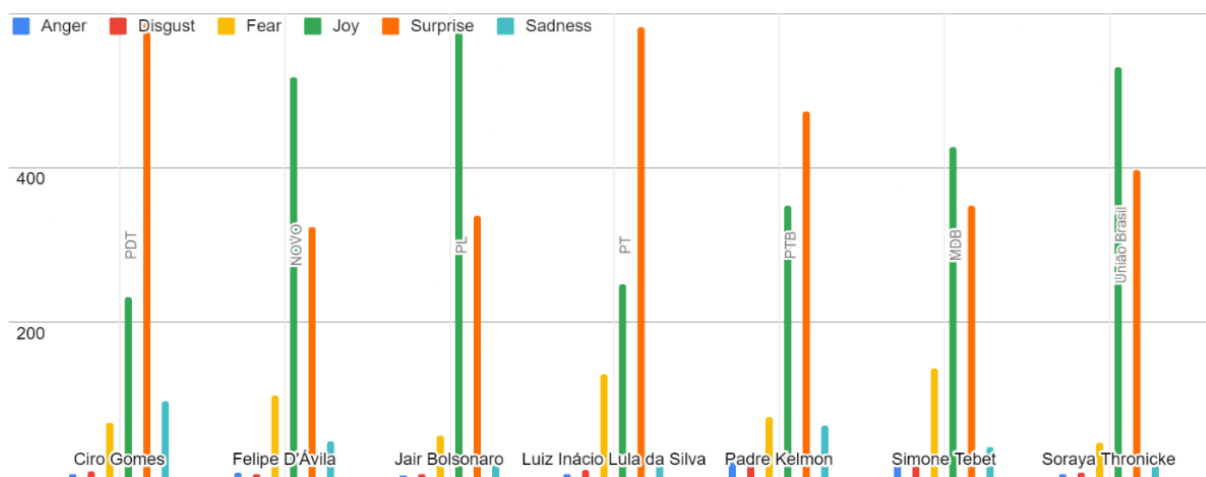


Fonte: Elaborado pela autora, 2022.

5.3 Média de popularidade durante o primeiro turno

A Figura 26 representa a coleta de dados gerais que ocorreu durante o dia 02/10/2022, data em que ocorreu o 1º turno das Eleições Presidenciais. Nota-se que, os candidatos com maiores emoções *joy* são Jair Bolsonaro, Soraya Thronicke e Vera Lúcia, enquanto os candidatos com maior número de emoção *surprise* são os candidatos Ciro Gomes, Padre Kelmon e Lula.

Figura 26 - Gráfico Geral dos Candidatos- 1º Turno



Fonte: Elaborado pela autora, 2022.

A Tabela 5 diz-se respeito à tabela utilizada para gerar os gráficos desta seção. Os dados preenchidos na tabela foram obtidos através dos números fornecidos pelo *Orange Data Mining*, resultados da pesquisa das emoções dispostas em cada tweet.

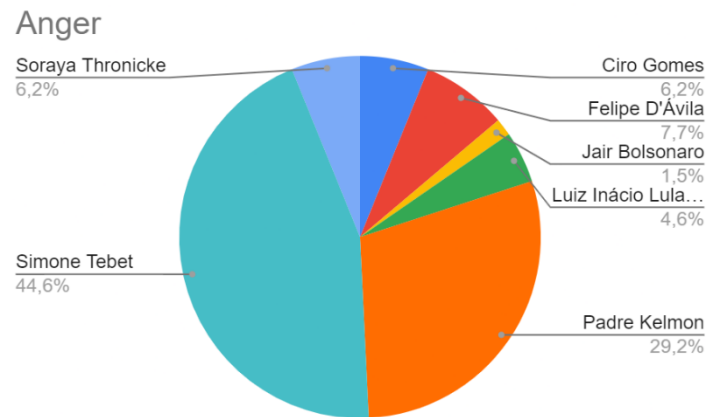
Tabela 5 - Coleta de Dados Eleição - Primeiro Turno

Carga de Dados - Eleição								
Candidato	Partido	Anger	Disgust	Fear	Joy	Surprise	Sadness	Soma
Ciro Gomes	PDT	4	8	70	233	587	98	1000
Felipe D'Ávila	NOVO	5	4	105	517	323	46	1000
Jair Bolsonaro	PL	1	3	53	577	337	29	1000
Luiz Inácio Lula da Silva	PT	3	9	133	250	581	24	1000
Padre Kelmon	PTB	19	14	77	351	472	67	1000
Simone Tebet	MDB	29	15	140	427	350	39	1000
Soraya Thronicke	União Brasil	4	6	45	529	397	19	1000

Fonte: Elaborado pela autora, 2022.

Nas Figuras apresentadas a seguir, pode-se analisar cada emoção separadamente. Primeiramente, na emoção *Anger*, representada na Figura 27, o candidato que maior obteve o número dessa emoção foi Simone Tebet, contando com 29 *tweets* contendo essa emoção, enquanto o detentor do menor número foi o candidato Jair Bolsonaro.

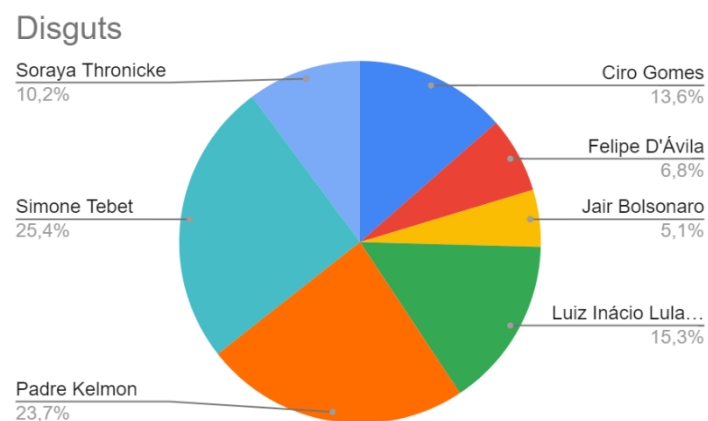
Figura 27 - Sentimento Isolado - Anger



Fonte: Elaborado pela autora, 2022.

No que diz respeito à emoção *Disgust*, representada na Figura 28, o candidato que maior apresentou o número de descontentamento, contando com 24 *tweets*, foi a candidata Simone Tebet. Enquanto isso, o detentor do menor número foi o candidato Jair Bolsonaro.

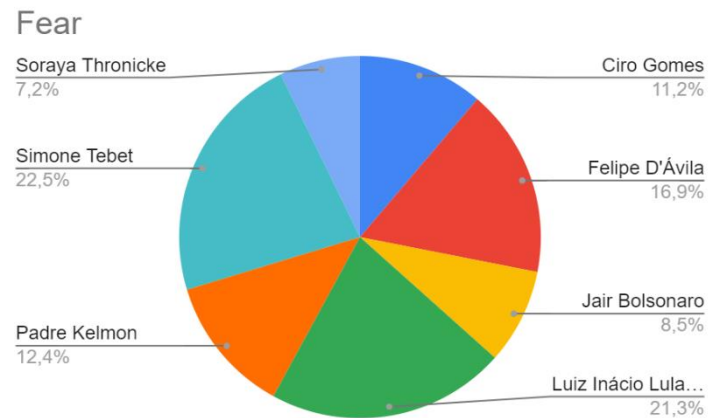
Figura 28 - Sentimento Isolado - Disgust



Fonte: Elaborado pela autora, 2022.

No que diz respeito à emoção *Fear*, representada na Figura 29, o candidato que maior obteve o número de medo, contando com 140 *tweets*, foi a candidata Simone Tebet. Enquanto isso, o detentor do menor número foi a candidata Soraya Thronicke.

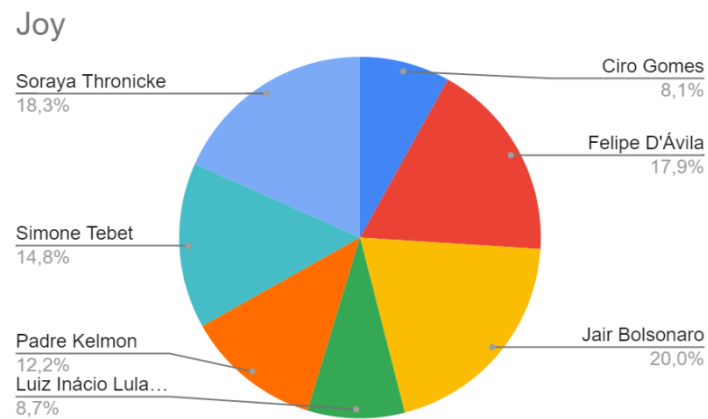
Figura 29 - Sentimento Isolado - Fear



Fonte: Elaborado pela autora, 2022.

No que diz respeito à emoção *Joy*, representada na Figura 30, o candidato que maior obteve o número de alegria, contando com 577 *tweets*, foi o candidato Jair Bolsonaro. Enquanto isso, o detentor do menor número foi o candidato *Ciro Gomes*.

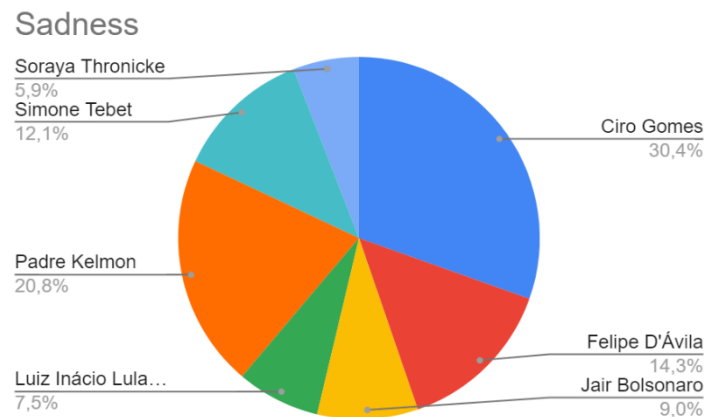
Figura 30 - Sentimento Isolado - Joy



Fonte: Elaborado pela autora, 2022.

No que diz respeito à emoção *Sadness*, representada na Figura 31, o candidato que maior obteve o número de tristeza, contando com 98 *tweets*, foi o candidato *Ciro Gomes*. Enquanto isso, o detentor do menor número foi a candidata *Soraya Thronicke*.

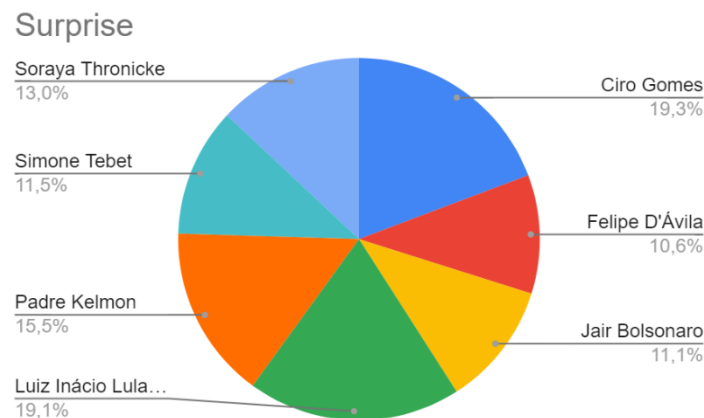
Figura 31 - Sentimento Isolado - Sadness



Fonte: Elaborado pela autora, 2022.

No que diz respeito à emoção *Surprise*, representada na Figura 33, o candidato que maior obteve o número de descontentamento, contando com 587 *tweets*, foi o candidato Ciro Gomes. Enquanto isso, o detentor do menor número foi o candidato Pablo Marçal.

Figura 32 - Sentimento Isolado - Surprise

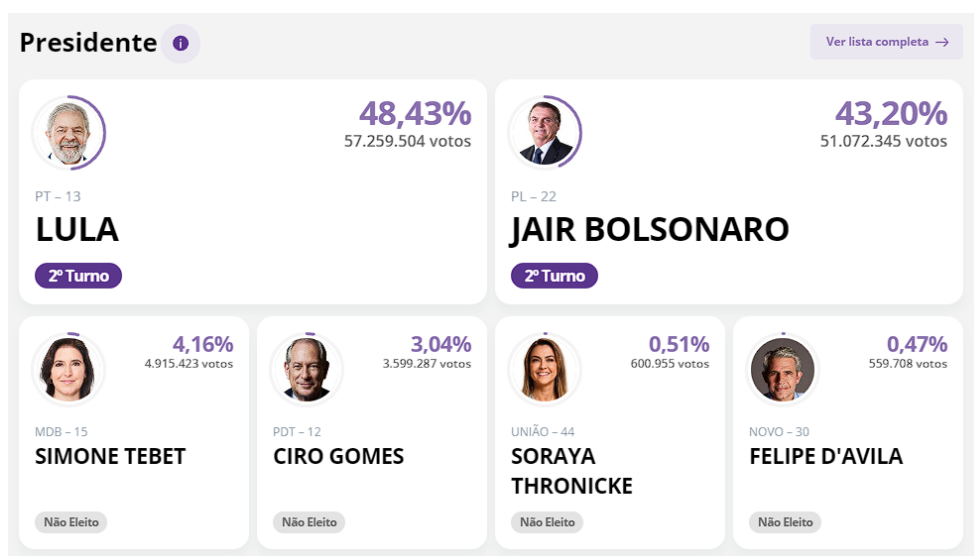


Fonte: Elaborado pela autora, 2022.

5.4 Comparativos com o Resultado da Eleição – Primeiro Turno

No Capítulo 1 desta monografia foi apresentada a seguinte questão de pesquisa: “As técnicas de mineração utilizadas são capazes de verificar se os níveis de popularidade dos candidatos às eleições presidenciais de 2022 no Twitter condizem com os resultados das eleições nas urnas?”. Dessa forma, buscou-se analisar e comprara os resultados obtidos com os resultados do TSE. A Figura 33 abaixo ilustra o resultado obtido no primeiro turno das eleições.

Figura 33 - Resultado do 1º Turno



Fonte: TSE, 2022.

Como informado na Figura 33, os candidatos Luiz Inácio Lula da Silva e Jair Bolsonaro conseguiram obter o maior quantitativo de votos, dando continuidade para o segundo turno uma vez que, ambos não atingiram 50% + 1 voto. Em 3º lugar, a candidata Simone Tebet levou com 4,16% dos votos, seguida por Ciro Gomes.

Na análise feita no dia 02/10/2022, dia da eleição, os candidatos com maiores possibilidade de ser eleitos eram Jair Bolsonaro, Simone Tebet e Soraya Thronicke, sendo que tais candidatos contavam com baixa rejeição, além de todos terem a emoção *joy* como maior emoção refletida em suas candidaturas.

Porém, como mostrado na Figura 33, os candidatos com maior número de votos foram Luiz Inácio Lula da Silva, Jair Bolsonaro e Simone Tebet, respectivamente. Parcialmente, a questão de pesquisa foi respondida corretamente, pois indicou que 2 dos 3 candidatos com maior aceitação pelos internautas alcançaram boas colocações.

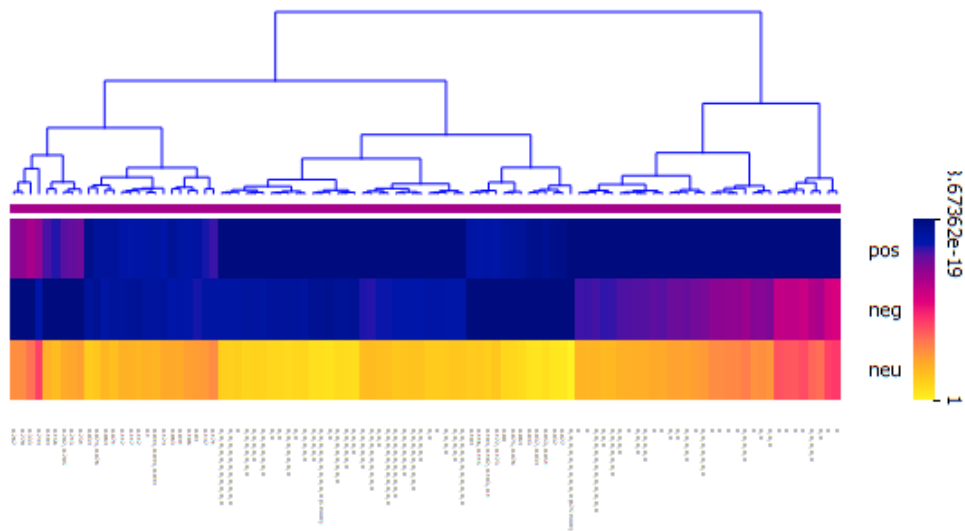
5.5 Análise de Popularidade Individual – Segundo Turno

5.5.1 Jair Bolsonaro

Durante o segundo turno, ocorreram dois debates, o qual o candidato Jair Bolsonaro participou de todos. Na Figura 34 a seguir, apresentada num *heatmap*, pode-se observar o

nível de rejeição do candidato teve um leve aumento se comparado ao primeiro turno. Os dados detalhados deste candidato podem ser vistos nos APÊNDICE C e APÊNDICE D.

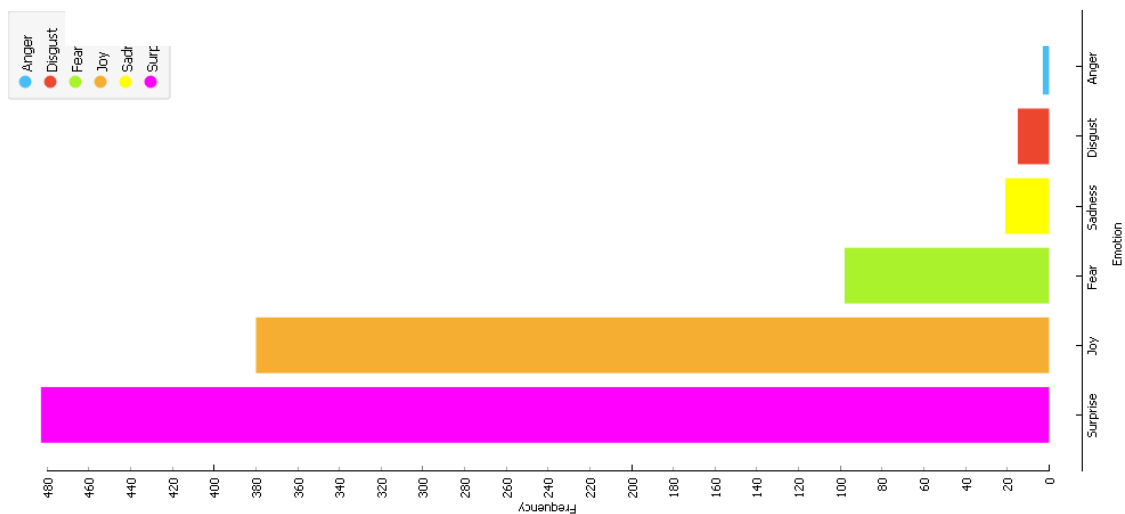
Figura 34 - *HeatMap* - Candidato Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

Enquanto isso, no gráfico da Figura 35 abaixo, nota-se que o maior sentimento expressado pelos internautas era *joy*, indicando que, mesmo com o nível de rejeição tendo um aumento, o candidato ainda recebe uma boa aceitação do público.

Figura 35 - Gráfico de Emoções – Jair Bolsonaro

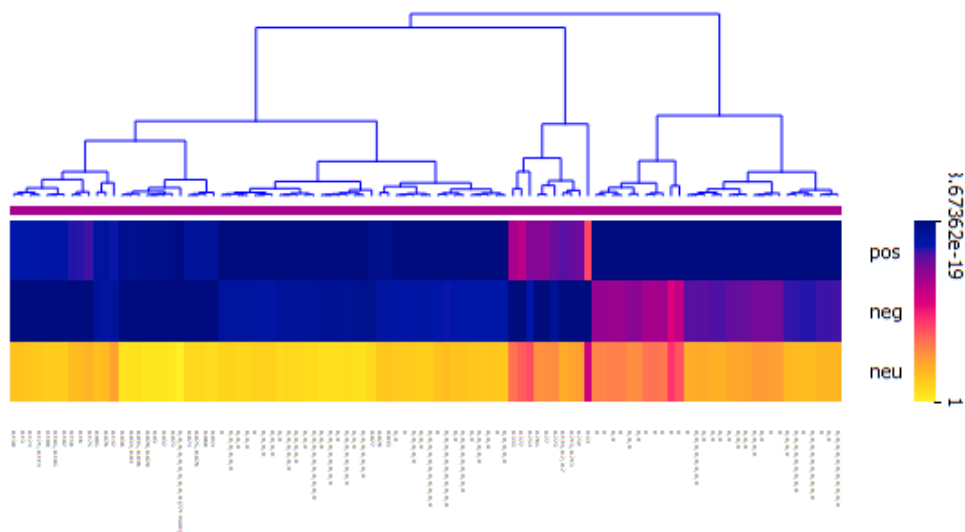


Fonte: Elaborado pela autora, 2022.

5.5.2 Luiz Inácio Lula da Silva

Assim como seu concorrente, o candidato Lula participou dos dois debates ocorridos no segundo turno. Os dados detalhados deste candidato podem ser vistos nos APÊNDICE F e APÊNDICE G. Comparado ao primeiro turno, o candidato conseguiu diminuir seu número de rejeições, tornando sua campanha tanto quanto favorável. A Figura 36 ademonstra o *heatmap* coletado durante a eleição do segundo turno.

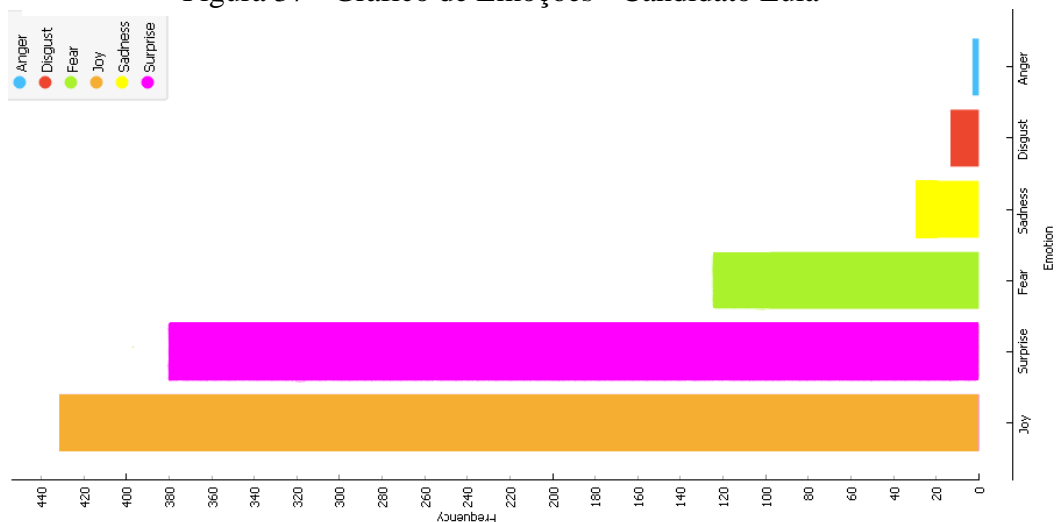
Figura 36 - *HeatMap* - Candidato Lula



Fonte: Elaborado pela autora, 2022.

A Figura 37 apresentada demonstra o Gráfico das Emoções contidas nos *tweets* relacionados ao candidato. Nota-se que, o maior sentimento presente é o *surprise*, não indicando contentamento ou não dos eleitores.

Figura 37 - Gráfico de Emoções - Candidato Lula



Fonte: Elaborado pela autora, 2022.

5.6 Média de popularidade durante o segundo turno

O segundo turno contou com 28 dias para que os candidatos pudessem realizar suas campanhas presidenciais. Ao final do primeiro turno, os candidatos Jair Bolsonaro e Luiz Inácio Lula da Silva foram os mais votados e, como nenhum obteve 50% + 1 dos votos, ambos foram disputar o segundo turno das eleições.

Dessa forma, esta seção mostra uma análise geral do segundo turno das eleições, exibindo os gráficos coletados durante o dia 30/10/2022, em que ocorreu a votação para a escolha do presidente. Os dados dispostos foram obtidos através da ferramenta de mineração de dados Orange. A Figura 39 retrata a tabela em que utilizou-se para gerar os gráficos.

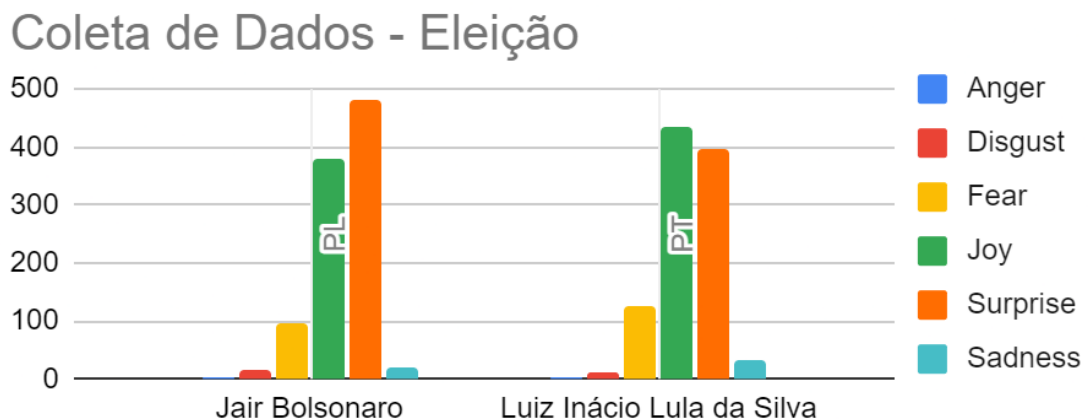
Tabela 6 - Coleta de Dados Eleição - Segundo Turno

Candidato	Partido	Anger	Disgust	Fear	Joy	Surprise	Sadness	Soma
Jair Bolsonaro	PL	3	15	98	380	483	21	1000
Luiz Inácio Lula da Silva	PT	2	10	125	434	395	34	1000

Fonte: Elaborado pela autora, 2022.

Através dela é possível notar um empate técnico⁹ entre os candidatos, visto que ambos possuem números próximos das emoções dispostas nos *tweets*. A Figura 39 possibilita observar melhor a proximidade entre os candidatos.

Figura 38 - Gráfico de Barras - Coleta 2º Turno

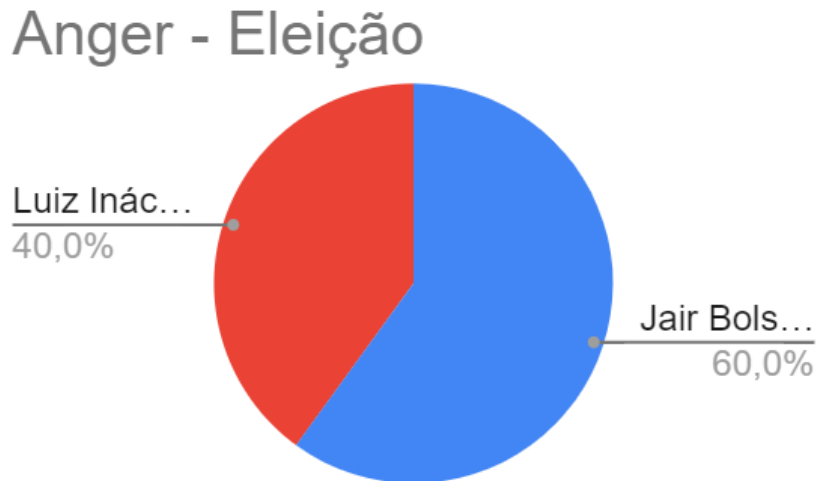


Fonte: Elaborado pela autora, 2022.

⁹ Empate técnico é ocorre quando a diferença entre os candidatos se encontra dentro das margens de erro das pesquisas (POVO, 2020).

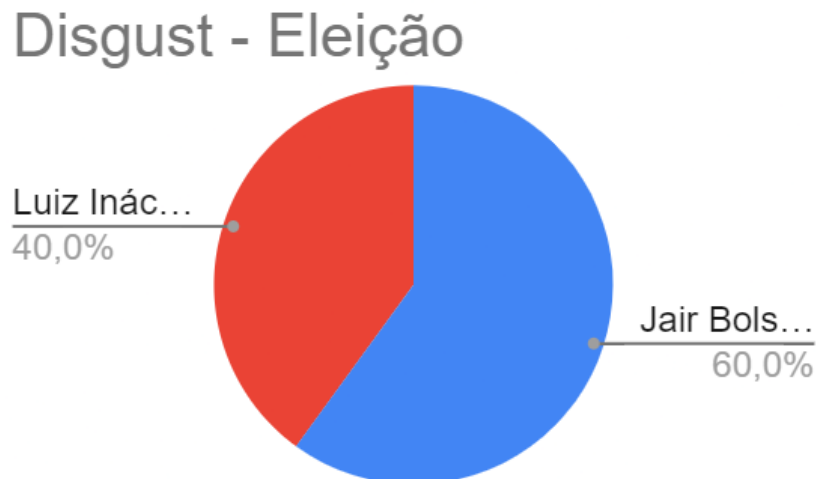
As imagens a seguir permitem fazer um comparativo entre as emoções dispostas pelos candidatos. Na Figura 39, o qual é demonstrado o sentimento *anger*, o candidato Jair Bolsonaro obteve maior quantidade desta emoção, assim como no sentimentos *disgust*, ilustrado na Figura 40.

Figura 39 - Sentimento Isolado - Anger



Fonte: Elaborado pela autora, 2022.

Figura 40 - Sentimento Isolado - Disgust



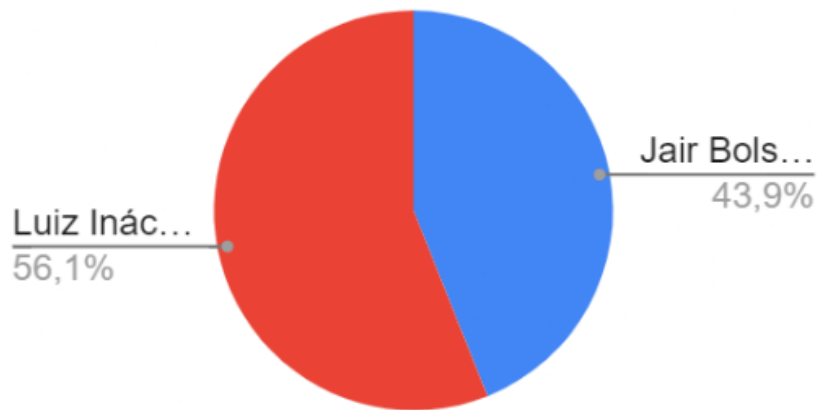
Fonte: Elaborado pela autora, 2022.

Apesar disso, o candidato Lula deteve maior porcentagem da emoção *fear*, como demonstrado na Figura 41. Além de conter, maior número de emoção *joy* atrelado a si, também possui *surprise* e *sadness* com maior porcentagem atrelado à sua candidatura, como é mostrado nas Figuras 42, 43 e 44, respectivamente. Mesmo que um dos candidato possua maior

porcentagem num sentimento, ambos possuem números muito próximos, dificultando a análise individual dos mesmos.

Figura 41 - Sentimento Isolado - Fear

Fear - Eleição



Fonte: Elaborado pela autora, 2022.

Figura 42 - Sentimento Isolado - Joy

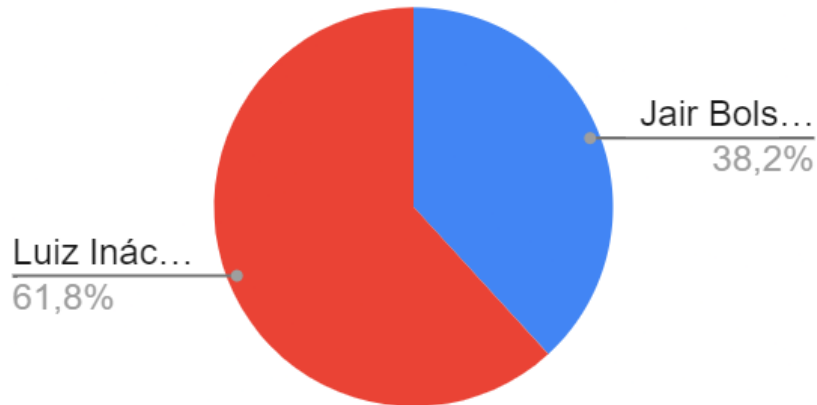
Joy - Eleição



Fonte: Elaborado pela autora, 2022.

Figura 43 - Sentimento Isolado – Sadness

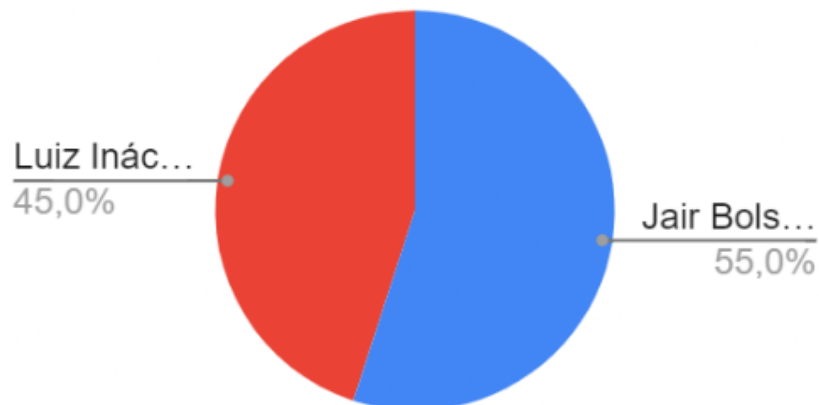
Sadness - Eleição



Fonte: Elaborado pela autora, 2022.

Figura 44 - Sentimento Isolado - Surpeise

Surprise - Eleição

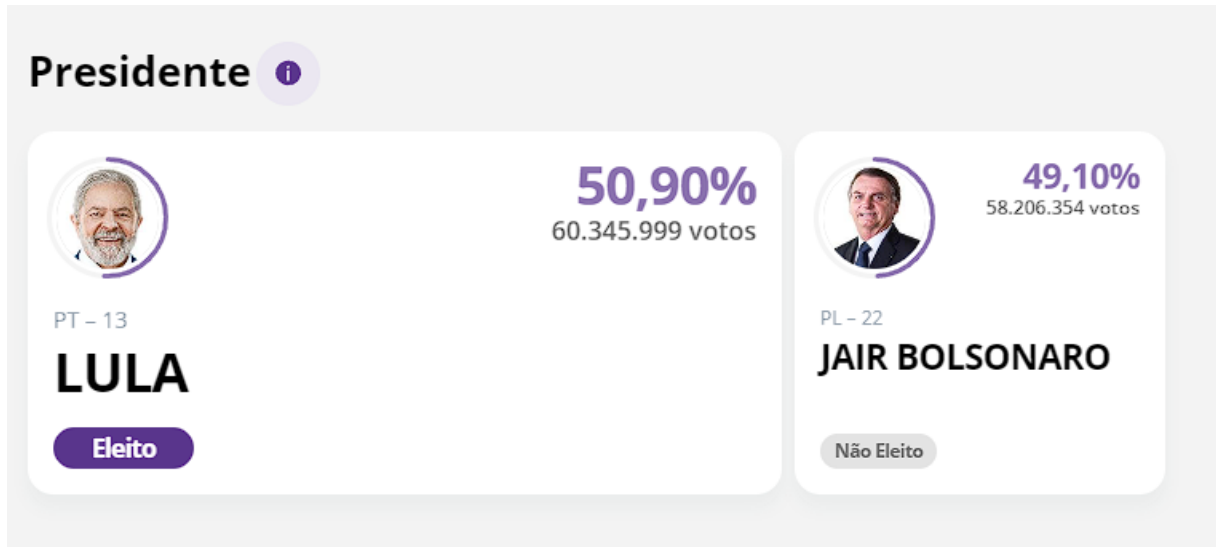


Fonte: Elaborado pela autora, 2022.

5.3 Comparativos com o Resultado da Eleição – Segundo Turno

Retomando a primeira questão de pesquisa sugerida neste trabalho, também analisou-se e comparou-se os resultados do segundo turno disponibilizados pelo TSE. A Figura 46 ilustra o resultado obtido no segundo turno das eleições.

Figura 45 - Resultado - 2º Turno



Fonte: TSE, 2022.

Ambos candidatos possuíam números próximos de sentimentos positivos e negativos, demonstrando a polarização vivida durante o período eleitoral. Além disso, o candidato Lula possuía, no dia da eleição, um quantitativo superior de emoções *joy* vinculadas a si, comparados a seu concorrente Jair Bolsonaro.

Na análise feita no dia 30/10/2022, dia da eleição, o candidato com maior possibilidade de ser eleitos era Luiz Inácio Lula da Silva e, de fato, foi eleito com 50,90% dos votos. Dessa forma, a questão de pesquisa foi respondida, já que o candidato mais popular venceu a eleição presidencial.

6. CONCLUSÃO

Nesta seção encontra-se a conclusão perante à análise disposta neste trabalho. Além disso, descreve as limitações encontradas e, também, faz-se considerações finais e comentários sobre trabalhos futuros.

6.1 Discussão

Como descrito durante o trabalho, a proposta consiste em apresentar a viabilidade de rede social Twitter enquanto uma ferramenta de predição eleitoral. Sendo assim, durante o desenvolvimento deste trabalho, utilizou-se técnicas de mineração de dados textuais e análise de sentimentos como meio de validação da pesquisa. Neste sentido, comparando a popularidade dos candidatos com o resultado final do primeiro turno, os resultados obtidos foram precisos, o que valida a análise sentimental na rede social Twitter como meio alternativo para as pesquisas tradicionais.

Analisando os gráficos do segundo turno e comparando-os com o resultado disponibilizado pelo TSE, viu-se que os votos ficaram divididos, assim como o resultados da segunda análise desta pesquisa. Com base nisso, é possível afimar que a mineração de dados feita na rede social é interessante, também, ser utilizada como um meio de monitorar o comportamento popular em casos de intenção de voto. Isso ganha força e viabilidade, já que estes resultados são gerados a qualquer momento.

6.2 Limitações

O presente trabalho, durante sua construção, encontrou algumas dificuldades as quais poderiam interferir na construção do mesmo. Primeiramente, vale destacar a imensa quantidade de *Fake News* divulgadas nas redes sociais. Mesmo com toda campanha de combate ao compartilhamento de informações falsas, a propagação ainda é de grande quantidade, tornando difícil ter uma certeza dos sentimentos analisados.

Além disso, como as eleições estavam polarizadas em dois candidatos, a análise do segundo turno obteve resultados bem próximos, tendo uma dificuldade em apontar um vencedor, visto que o resultado apontado pelo TSE indica uma diferença de 1,8% entre os candidatos.

Por fim, a última limitação encontrada foi a larga utilização de *bots* encontrados nesta rede. A utilização de *bots*¹⁰ prejudica a análise, pois não são pessoas reais expressando seus sentimentos, fazendo com o que os candidatos recebestem mais ou menos sentimentos reais.

6.3 Considerações Finais e Trabalhos Futuros

Considerando os resultados obtidos e as limitações do estudo anteriormente abordadas, é possível concluir que, em suma, a utilização de redes sociais como preditor eleitoral é uma forma de se obter respostas. Ainda é possível utilizá-las como meio de auxílio em tomadas de decisão, e na compreensão de como o público está respondendo às promessas e ações dos políticos. Mesmo que, em alguns casos houvesse distanciamento, a pesquisa se mostrou condizente com os resultados dos candidatos que mais obtiveram votos no segundo turno, por exemplo. Assim, a utilização de *tweets* para verificar a popularidade de candidatos políticos, é uma alternativa interessante às pesquisas eleitorais, podendo intercalar-se entre si, de modo a construir um perfil do desempenho dos presidencialistas de forma mais contínua.

Por fim, com relação a trabalhos futuros, recomenda-se o estudo com o foco em detecção de *bots* e *fake news*, de forma que possa ter uma melhor filtragem dos dados, além de obter dados regionais para aprofundamento da análise. Tais melhorias acarretam trabalhos que geram um grande valor para a sociedade e academia, trazendo maior credibilidade para estes estudos.

¹⁰ *Bots* são programas de computador automatizados que são projetados para imitar usuários humanos (CLOUDFARE, 2016).

7. REFERÊNCIAS

- AGÊNCIA MESTRE. **Estatísticas do Twitter no Brasil | Agência Mestre**. 2022. Disponível em: <<https://www.agenciamestre.com/redes-sociais/estatisticas-twitter-brasil/#:~:text=61%25%20dos%20entrevistados%20acessam%20a,de%20uso%20ou%20aument%C3%A1%20D1a.>>. Acesso em: 18 abr. 2022.
- AGGARWAL, C. C.; ZHAI, C. **Mining text data**. [S.l.]: Springer Science & Business Media, 2012.
- AMARAL, Fernando. **Aprenda mineração de dados: teoria e prática**. Alta Books Editora, 2016.
- AMBROSIO, Ana Paula L. MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L. **Mineração de textos**. Relatório Técnico–Instituto de Informática (UFG), 2007.
- AMPER ENERGIA HUMANA. **We Are Social e HootSuite - Digital 2021 [Resumo e Relatório Completo]**. Disponível em: <<https://www.amper.ag/post/we-are-social-e-hootsuite-digital-2021-resumo-e-relat%C3%B3rio-completo>>. Acesso em: 12 dez. 2022.
- ANASTACIO, B. **K-means: o que é, como funciona, aplicações e exemplo em Python**. 2020. Disponível em: <<https://medium.com/programadores-ajudando-programadores/k-means-o-que-%C3%A9-como-funciona-aplica%C3%A7%C3%B5es-e-exemplo-em-python-6021df6e2572>>. Acesso em: 21 nov. 2022.
- AZEVEDO, J. **Influenciadores digitais e seu impacto nas decisões de consumo**. Disponível em: <<https://bagy.com.br/blog/influenciadores-digitais-e-seu-impacto-nas-decisoes-de-consumo/#:~:text=A1%C3%A9m%20disso%2C%20a%20principal%20caracter%C3%ADstica,que%20acompanham%20eles%20nas%20redes.>>. Acesso em: 16 nov. 2022.
- BECKER, K.; TUMITAN, D. **Introdução à mineração de opiniões: Conceitos, aplicações e desafios**. Simpósio brasileiro de banco de dados, v. 75, 2013
- BERNARDO, D. S.; GOULART, E. E. **Evolução na comunicação: estudos nas redes sociais**. Sub-projeto de pesquisa em Iniciação Científica. Programa de Mestrado em Comunicação. Universidade Municipal de São Caetano do Sul, USCS. São Caetano do Sul, 2011. Acesso em: 6 de março.
- BRAGA, LUIS PAULO VIEIRA. **Introdução à Mineração de Dados-2a edição: Edição ampliada e revisada**. Editora E-papers, 2005.
- BRASIL, Luana Melody. **Debate: Soraya pergunta a Bolsonaro se pretende dar golpe de Estado**. 2022. Disponível em: <<https://www.otempo.com.br/eleicoes/debate-soraya-pergunta-a-bolsonaro-se-pretende-dar-golpe-de-estado-1.2742290>>. Acesso em: 15 nov. 2022.
- BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil**: promulgada em 5 de outubro de 1988. [S. l.: s. n.], 1990. Disponível em: <https://www.tse.jus.br/legislacao/compilada/constituicao-federal/1988/constituicao-federal-de-1988>. Acesso em: 18 abr. 2022.
- CAJADO, A. F. R.; DORNELLES, T.; PEREIRA, A. C. **Eleições no Brasil: uma história de 500 anos**. Tribunal Superior Eleitoral, 2014
- CLOUDFLARE. **O que é um bot de redes sociais? | Definição de bot de redes sociais**. [S. l.], 2016. Disponível em: <https://www.cloudflare.com/pt-br/learning/bots/what-is-a-social-media-bot/>. Acesso em: 18 nov. 2022.
- COUTINHO, T. **O que é clustering? Veja como funciona essa análise!**, 2021. Disponível em: <<https://www.voitto.com.br/blog/artigo/clustering>>. Acesso em: 21 nov. 2022.
- DE PAULA FILHO, Wilton; GARCIA, Ana Cristina. **Predição do resultado das eleições presidenciais do Brasil baseado em tuítes**. In: **Anais do IV Brazilian Workshop on Social Network Analysis and Mining**. SBC, 2015.

DE QUEIROZ, Guilherme Guimarães; ALMEIDA, Leandro. Uma metodologia de análise de sentimentos dos candidatos as eleições presidenciais de 2018 no Twitter. **Revista de Engenharia e Pesquisa Aplicada**, v. 5, n. 1, p. 21-30, 2020.

DICIONÁRIO DE OXFORD. **Oxford University Press**. [S. l.: s. n.], 2022. Disponível em: <https://www.oxfordlearnersdictionaries.com/definition/english/fake-news?q=FAKE+NEWS>. Acesso em: 12 dez. 2022.

DUTRA, Daniela Aparecida Monteiro; FRANCISCO, Eduardo de Rezende. Text mining: análise de sentimentos nas eleições 2018. In: **Congresso Transformação Digital 2018**. 2018. Acesso em: 7 maio 2022.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37-37, 1996.

FIA. **Machine Learning: como funciona, benefícios, tipos e exemplos**. [S. l.], 12 nov. 2021. Disponível em: <https://fia.com.br/blog/machine-learning/>. Acesso em: 16 nov. 2022.

FREEMAN, L. C. Some Antecedents of Social Network Analysis. **Connections**, v.19, n.1, p.39-42, 1996

GIL, Antonio Carlos. Como elaborar projetos de pesquisa. 4. ed. São Paulo: Atlas, 2008.

HAMMES, MAICON RAFAEL et al. MACHINE LEARNING E SUAS APLICAÇÕES EM PROCESSOS ELEITORAIS. **Salão do Conhecimento**, v. 7, n. 7, 2021.

L13488. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/lei/L13488.htm>. Acesso em: 12 dez. 2022.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de Metodologia Científica**. 5º. ed. São Paulo: ATLAS, 2003.

LARHUD - Laboratório em Rede de Humanidades Digitais. **LARHUD**. [S.l.] , [s.d.]. Disponível em Orange - LARHUD (ibict.br)>. Acesso em: 11 maio. 2022.

LEANDRO, J. et al. Capítulo 5 Introdução ao Processamento de Linguagem Natural usando Python. n. 1, p. 336–360, 2017.

Levando a IA e o ML até a empresa. Disponível em: <[https://www.hpe.com/br/pt/what-is/machinelearning.html#:~:text=O%20machine%20learning%20\(ML\)%20%C3%A9,serem%20especificamente%20programados%20para%20isso.](https://www.hpe.com/br/pt/what-is/machinelearning.html#:~:text=O%20machine%20learning%20(ML)%20%C3%A9,serem%20especificamente%20programados%20para%20isso.)>. Acesso em: 17 abr. 2022.

LÉVY, P. Inteligência coletiva (A). [S.l.]: Edições Loyola, 2007. Acesso em: 6 março 2022.

MAAREK, P. J. Politics 2.0: New forms of digital political marketing and political communication. *Trípodos*, v. 1, n. 34, p. 13-22, 2014.

MARQUES, F. P. J. A.; SAMPAIO, R. C. Internet e eleições 2010 no brasil: rupturas e continuidades nos padrões midiáticos das campanhas políticas online. *Galáxia. Revista do Programa de Pós-Graduação em Comunicação e Semiótica*. ISSN 1982-2553, n. 22, 2011

MA, Ying. **PNL: Como o NLTK.Vader calcula o sentimento?**. [S. l.], 5 fev. 2020. Disponível em: <https://medium.com/@mystery0116/nlp-how-does-nltk-vader-calculate-sentiment-6c32d0f5046b>. Acesso em: 12 dez. 2022.

MATOS, DAVID. **Análise de Sentimentos e Machine Learning**. [S. l.], 9 out. 2015. Disponível em: <https://www.cienciaedados.com/analise-de-sentimentos-e-machine-learning/>. Acesso em: 12 abr. 2022.

MATOS, Fernanda Fernandes; MAGALHÃES, Lúcia Helena de; SOUZA, Renato Rocha. Recuperação e classificação de sentimentos de usuários do Twitter em período eleitoral. 2020.

MOOERS, C. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v.2, n.1, 1951, p.20-32.

MOURA, K. **Ciclo de vida dos dados #1 - Karina Moura – Medium**, 2019. Disponível em: <https://medium.com/@kvmoura/kdd-process-9b8e3062142>. Acesso em: 12 abr. 2022.

OLIVEIRA, Fernando Elias de. **BUSCA DE CONHECIMENTO SOBRE O PROCESSO SELETIVO DO INSTITUTO FEDERAL DE MINAS GERAIS – CAMPUS SÃO JOÃO EVANGELISTA::** a mineração de dados como tecnologia para desvelar tendências e padrões. 2013. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) - INSTITUTO FEDERAL DE MINAS GERAIS, [S. l.], 2013.

OLIVEIRA, Sofia Cisneiros Alves de. **Metodologia científica: tipos de pesquisa**. [S. l.], 4 nov. 2021. Disponível em: <https://www.sanarmed.com/metodologia-cientifica-tipos-de-pesquisa-colunistas>. Acesso em: 7 maio 2022.

ORANGE DATA MINING. **Heat Map**. [S. l.], 2015. Disponível em: <https://orange3.readthedocs.io/en/3.4.0/widgets/visualize/heat.html>. Acesso em: 18 nov. 2022.

PAES, Janiere Portela Leite. O sufrágio e o voto no Brasil: direito ou obrigação. **Tribunal Superior Eleitoral**, 2017. Disponível em: <https://www.tse.jus.br/o-tse/escola-judiciaria-eleitoral/publicacoes/revistas-da-eje/artigos/revista-eletronica-eje-n.-3-ano-3/o-sufragio-e-o-voto-no-brasil-direito-ou-obrigacao>. Acesso em: 18 abr. 2022.

PEREIRA, Janailton Galvão. **Análise de sentimentos da população brasileira em relação a eleição presidencial de 2018 através da rede social twitter**. 2019. Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Norte.

POVO, R. O. **Eleições 2020: entenda o que é um empate técnico nas pesquisas**. 2020. Disponível em: <https://www.opovo.com.br/noticias/politica/2020/10/28/eleicoes-2020--entenda-o-que-e-um-empate-tecnico-nas-pesquisas.html>. Acesso em: 21 nov. 2022.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico::** Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico. 2ª. ed. Novo Hamburgo: Feevale, 2013. 277 p. ISBN 978-85-7717-158-3. Disponível em: <https://www.feevale.br/institucional/editora-feevale/metodologia-do-trabalho-cientifico---2-edicao>. Acesso em: 11 maio 2022.

RAFAEL NEVES;LETÍCIA REGINALDO. **É padre ou não? Aliado de Bolsonaro em debate, Kelson tem cargo questionado**. Disponível em: <https://noticias.uol.com.br/eleicoes/2022/09/29/quem-e-o-candidato-padre-keelson-que-tem-chamado-atencao-dos-eleitores.htm>. Acesso em: 15 nov. 2022.

RODRIGUES, J. **Redes Sociais: saiba tudo sobre Instagram, TikTok e mais**. Disponível em: <https://resultadosdigitais.com.br/marketing/redes-sociais/>. Acesso em: 12 dez. 2022.

SHAO, C.; CIAMPAGLIA, G. L.; VAROL, O.; FLAMMINI, A.; MENCZER, F. **The spread of fake news by social bots**. arXiv preprint arXiv:1707.07592, arXiv, 2017.

SILVA, L. T. da; JUNIOR, A. B. F. Marketing político e sua importância através das mídias sociais. *Temática*, v. 9, n. 8, 2014. Acesso em: 6 março 2022.

SILVA, T.; STABILE, M. Monitoramento e pesquisa em mídias sociais: metodologias, aplicações e inovações. *Análise de redes em Mídias Sociais*. São Paulo: Uva Limão, p. 237–260, 2016.

TEIXEIRA, Carolina Barcelos. **Análise de Sentimento dos Usuários do Twitter em Relação à Atual Situação Política do Brasil**. Orientador: Prof. Dr. Tiago Martins Pereira. 2019. Monografia (Bacharelado em Estatística) - Universidade Federal de Ouro Preto, Ouro Preto - MG, 2019. Disponível em: <https://monografias.ufop.br/handle/35400000/1872>. Acesso em: 7 maio 2022.

TSE. Divulgação de Candidaturas e Contas Eleitorais. 2022. Disponível em: <https://divulgacandcontas.tse.jus.br/divulga/#/estados/2022/2040602022/BR/candidatos>. Acesso em: 23 out. 2022.

TSE. Eleições no Brasil: uma história de 500 anos. **2014**.

TSE. **Resultados**. Disponível em:

<https://resultados.tse.jus.br/oficial/app/index.html#/eleicao;e=e545/resultados>. Acesso em: 12 dez. 2022.

VICENTE, Eduarda. **Análise de Sentimentos, uma aplicação da Inteligência Artificial que avalia o comportamento do seu público**. 2019. Disponível em: <https://www.programmers.com.br/blog/analise-de-sentimentos-uma-aplicacao-da-inteligencia-artificial-que-avalia-o-comportamento-do-seu-publico/>. Acesso em: 26 maio 2022.

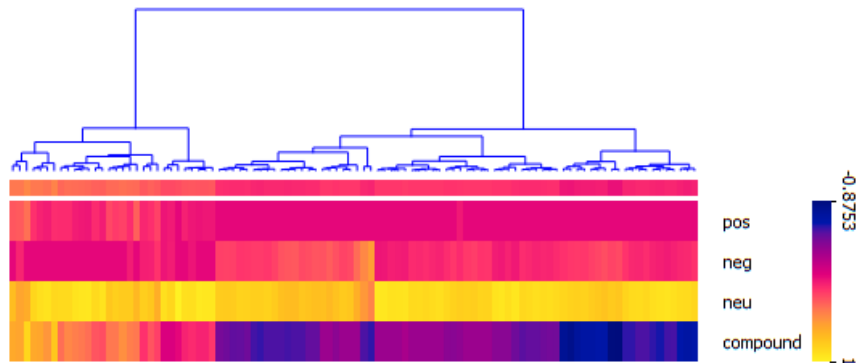
WIKIPEDIA, A ENCICLOPÉDIA LIVRE. **Twitter**. [S. l.], 30 maio 2022. Disponível em: <https://pt.wikipedia.org/wiki/Twitter>. Acesso em: 16 maio 2022.

YATES, Ricardo Baeza; NETO, Berthier Ribeiro. **Recuperação de Informação::** Conceitos e Tecnologia das Máquinas de Busca. 2ed. ed. [S. l.]: Bookman Editora, 2013. 612 p.

ZENHA, Luciana. Redes sociais online: o que são as redes sociais e como se organizam? **Caderno de Educação**, [s. l.], ano 20, ed. 49,v.1,2017/2018, p. 19-42, 18 abr. 2022.

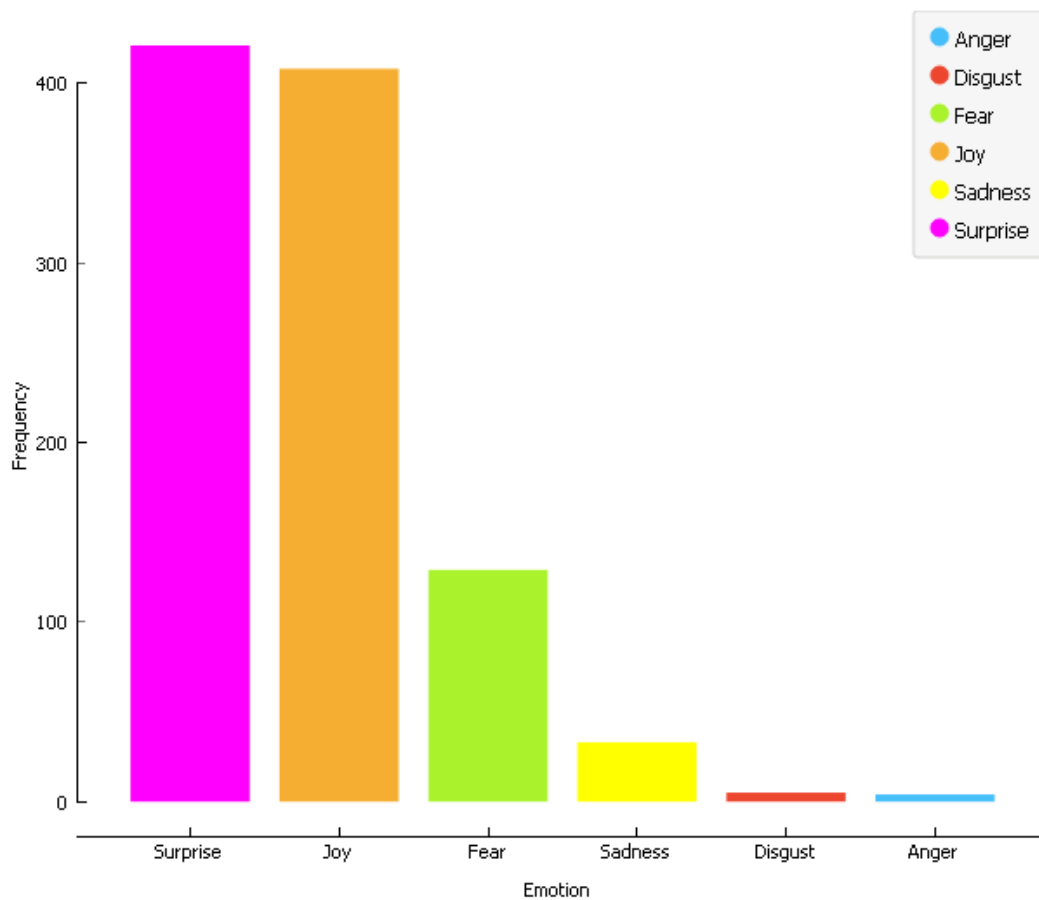
APÊNDICE A – CANDIDATO JAIR BOLSONARO – PRIMEIRO DEBATE - (BAND,
UOL, FOLHA DE S. PAULO E TV CULTURA) – 28/08/2022

Figura 46- *HeatMap* Pré-Debate - Jair Bolsonaro



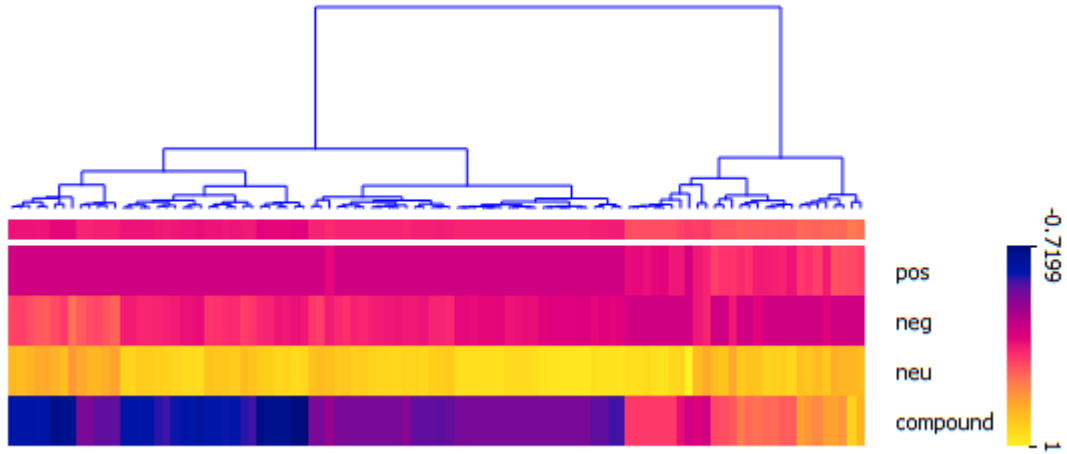
Fonte: Elaborado pela autora, 2022.

Figura 47 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro



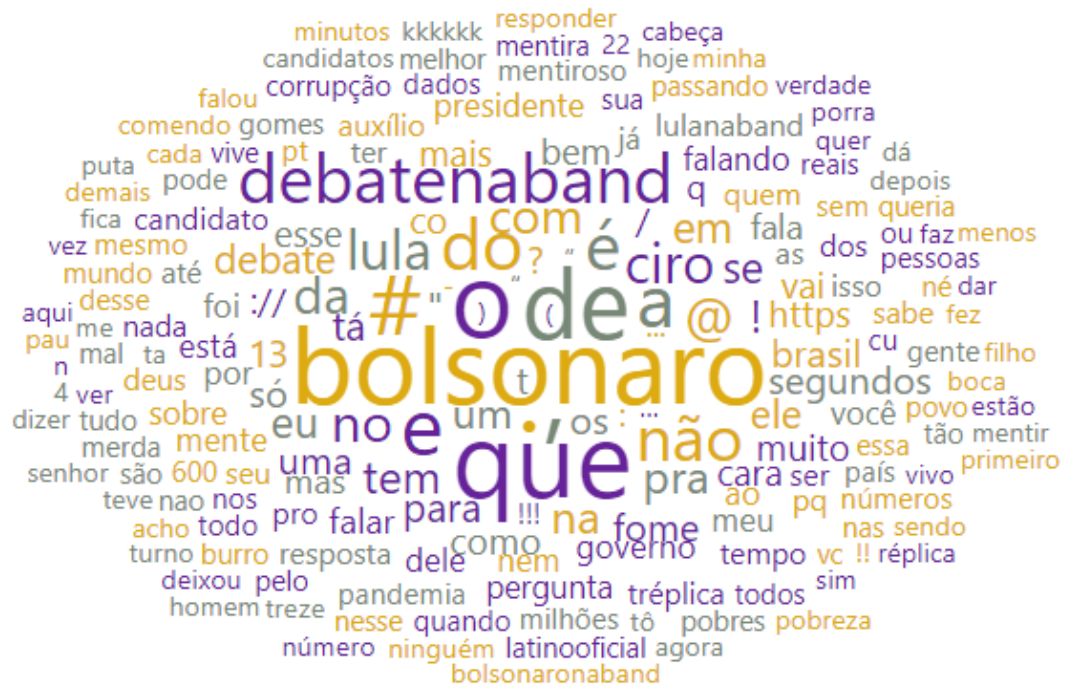
Fonte: Elaborado pela autora, 2022.

Figura 50 - HeatMap Debate - Jair Bolsonaro



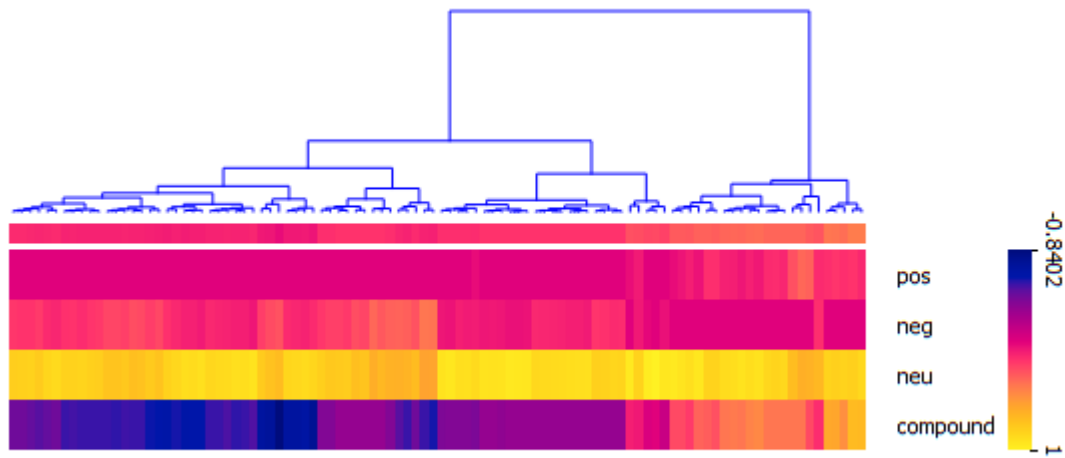
Fonte: Elaborado pela autora, 2022.

Figura 51 - WordCloud Debate - Jair Bolsonaro -



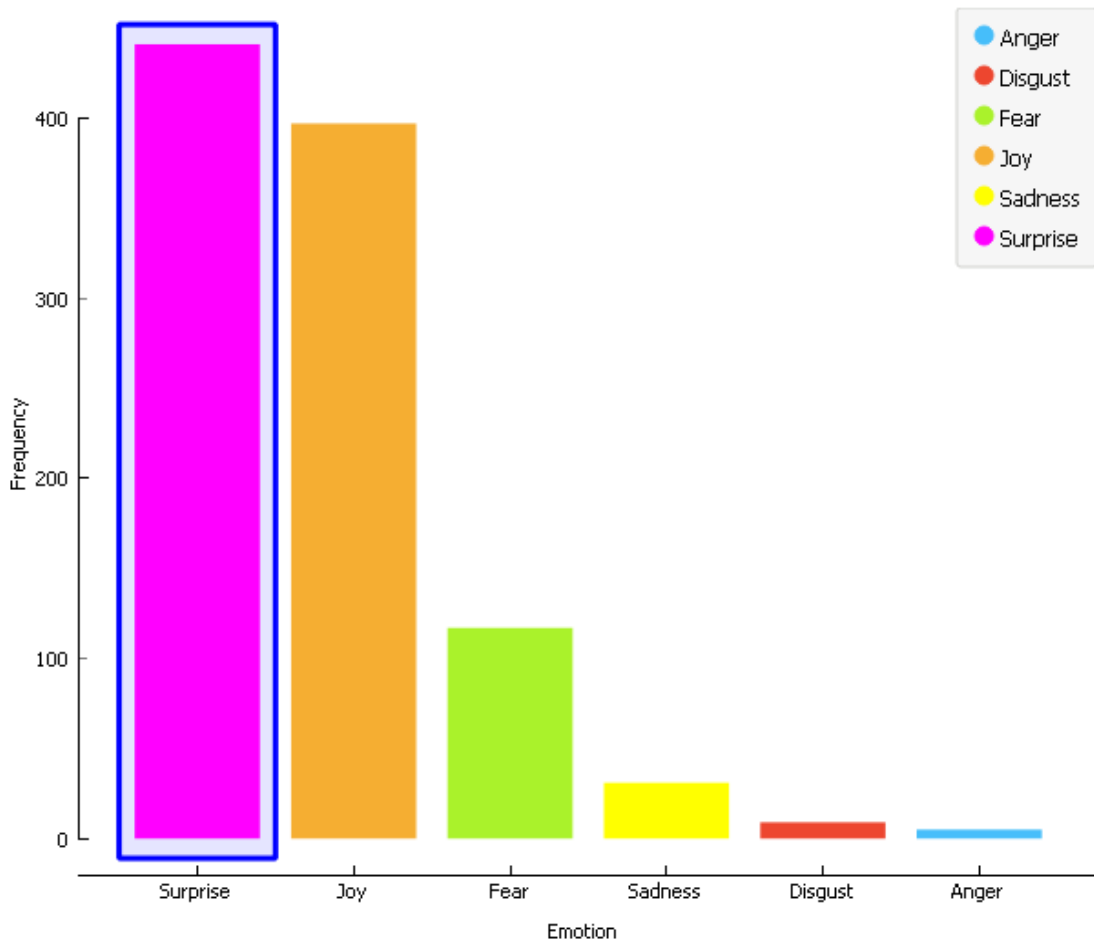
Fonte: Elaborado pela autora, 2022.

Figura 53 - HeatMap Pós-Debate - Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

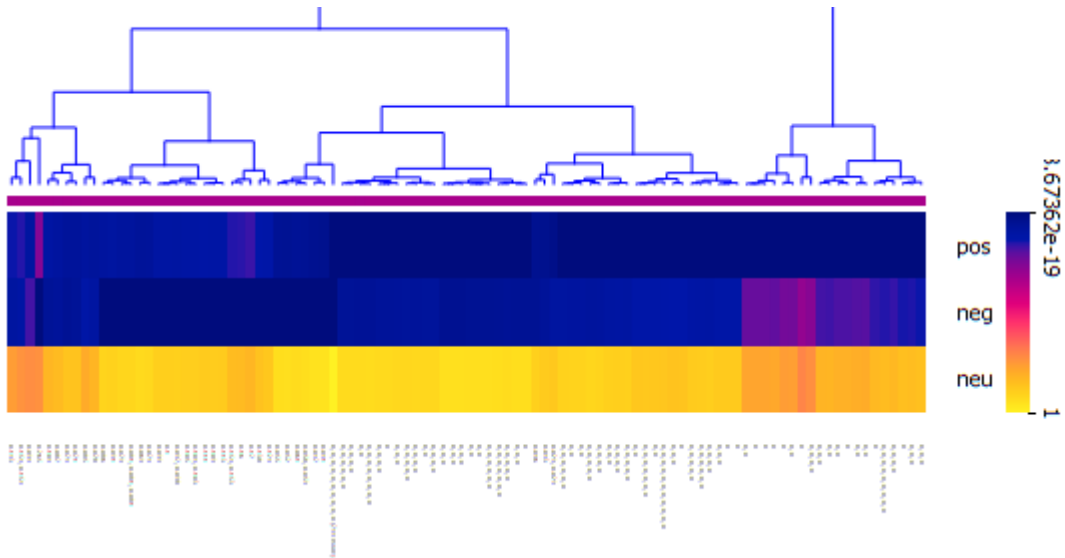
Figura 54 - Gráfico de Emoções Pós-Debate - Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

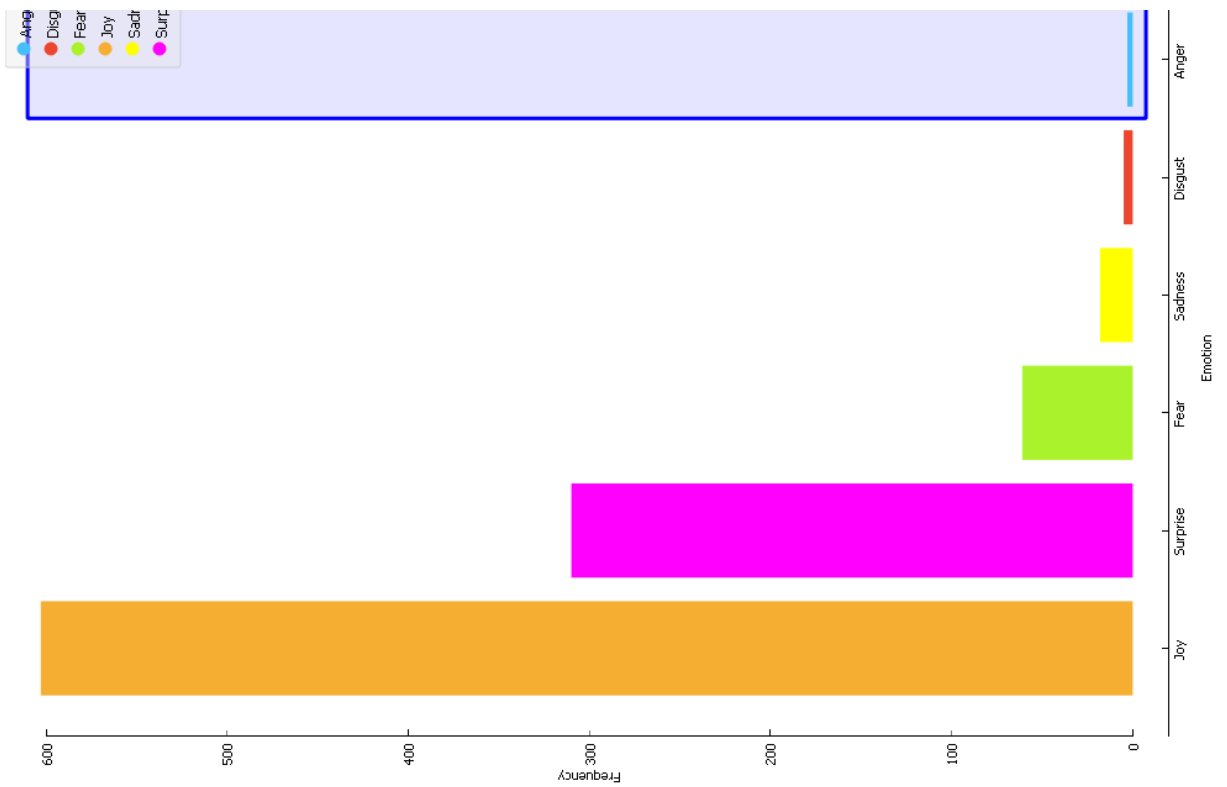
APÊNDICE B – CANDIDATO JAIR BOLSONARO – SEGUNDO DEBATE – (SBT, CNN, VEJA, O ESTADO DE S. PAULO, NOVA BRASIL FM E TERRA) – 24/09/2022

Figura 53 - HeatMap Pré-Debate - Jair Bolsonaro



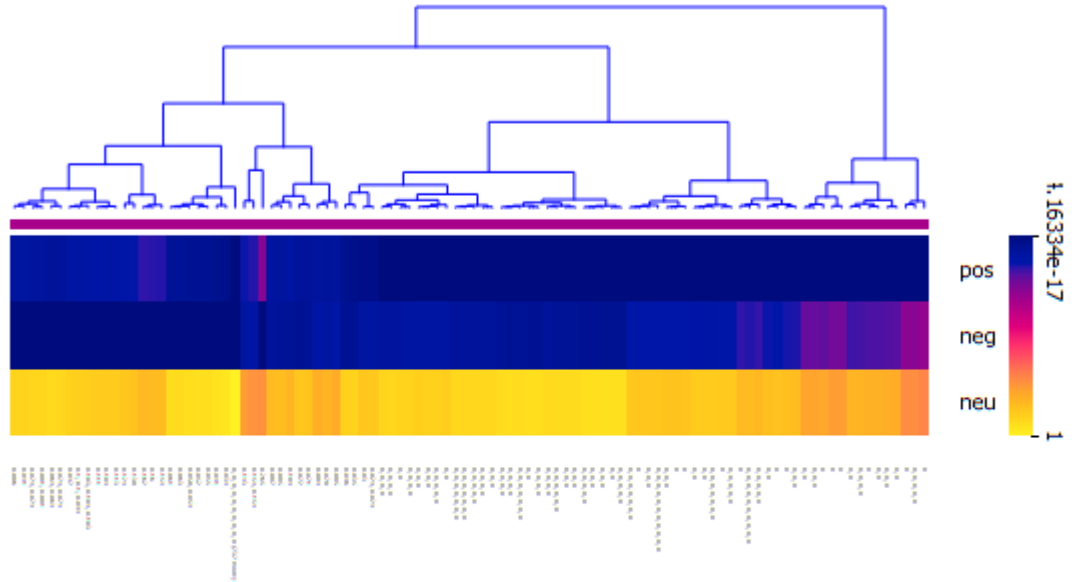
Fonte: Elaborado pela autora, 2022.

Figura 54 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro



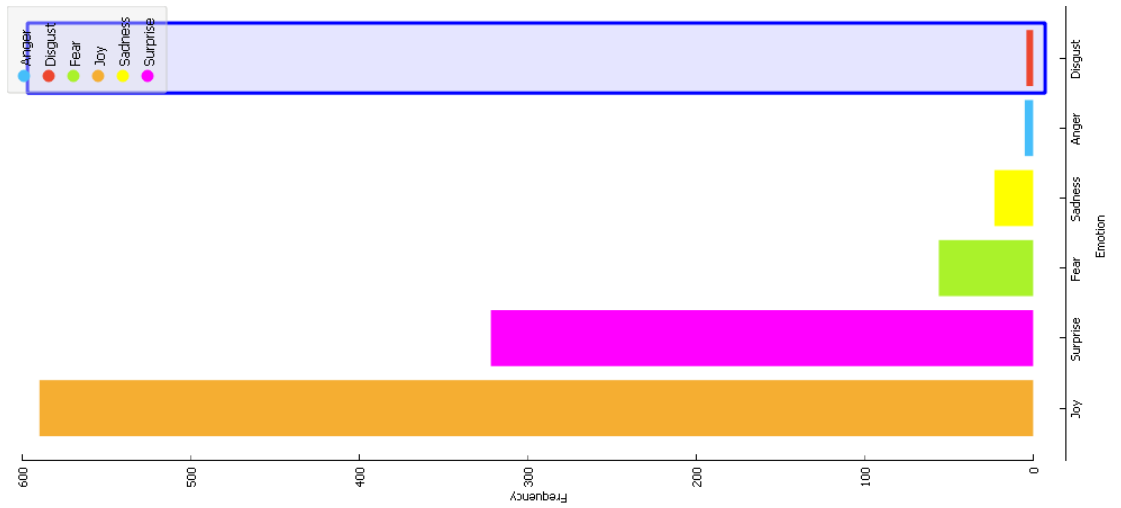
Fonte: Elaborado pela autora, 2022.

Figura 57 - HeatMap Debate - Jair Bolsonaro



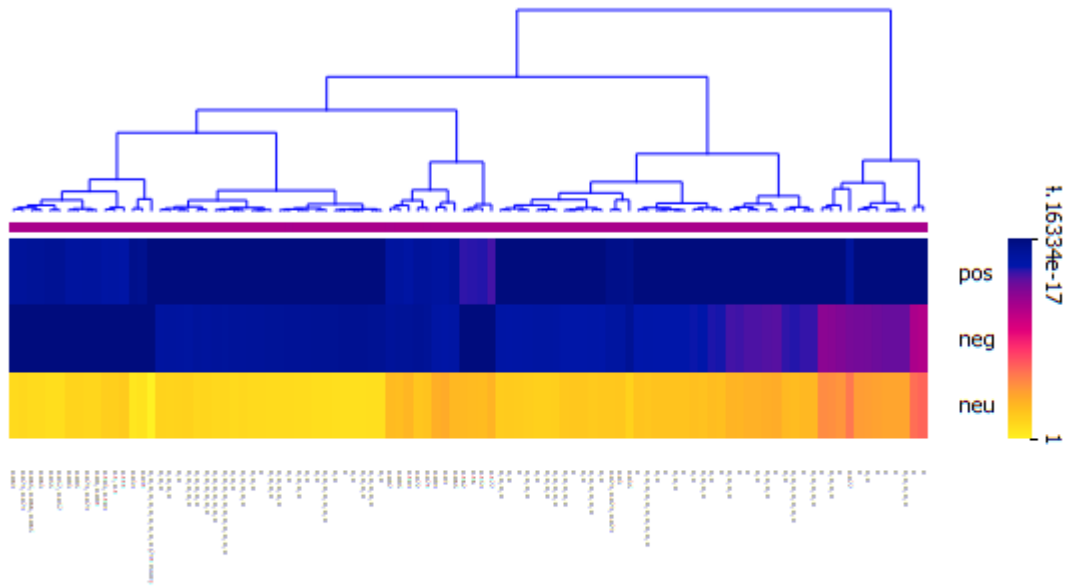
Fonte: Elaborado pela autora, 2022.

Figura 58 - Gráfico de Emoções Debate - Jair Bolsonaro



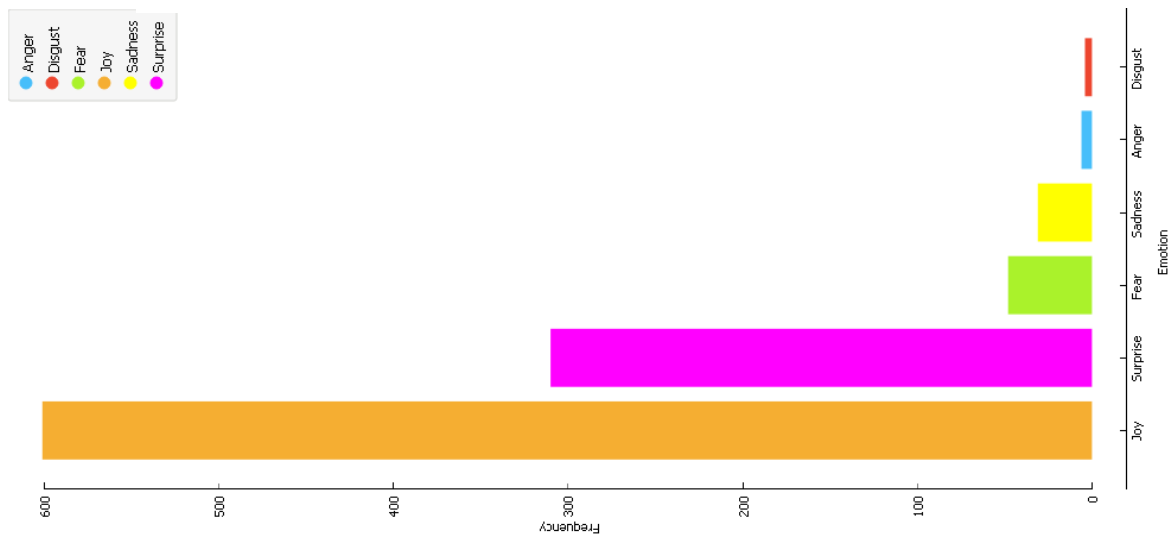
Fonte: Elaborado pela autora, 2022.

Figura 61 - HeatMap Pós-Debate - Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

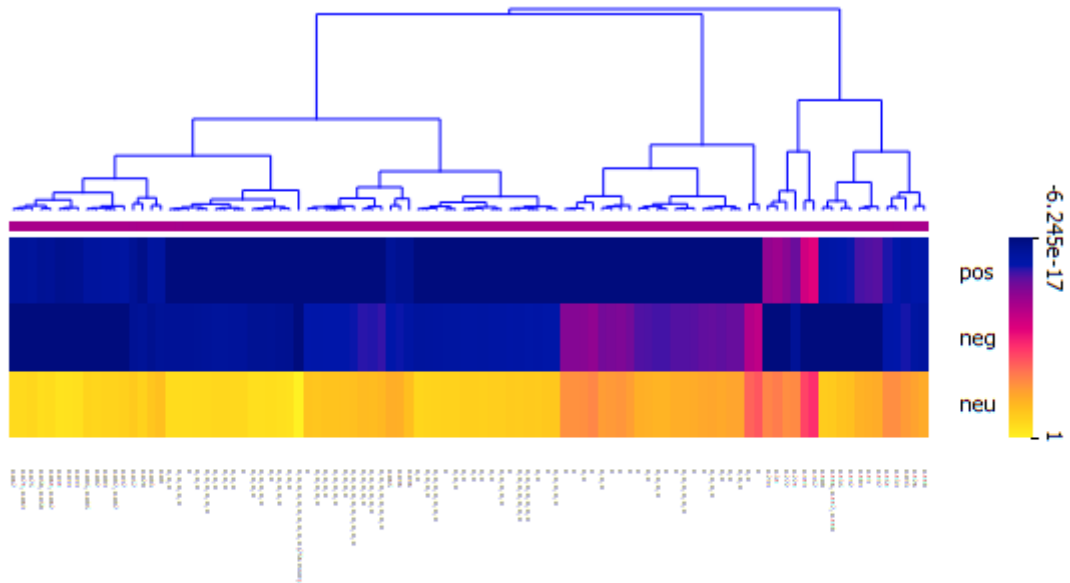
Figura 62 - Gráfico de Emoções Pós-Debate - Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

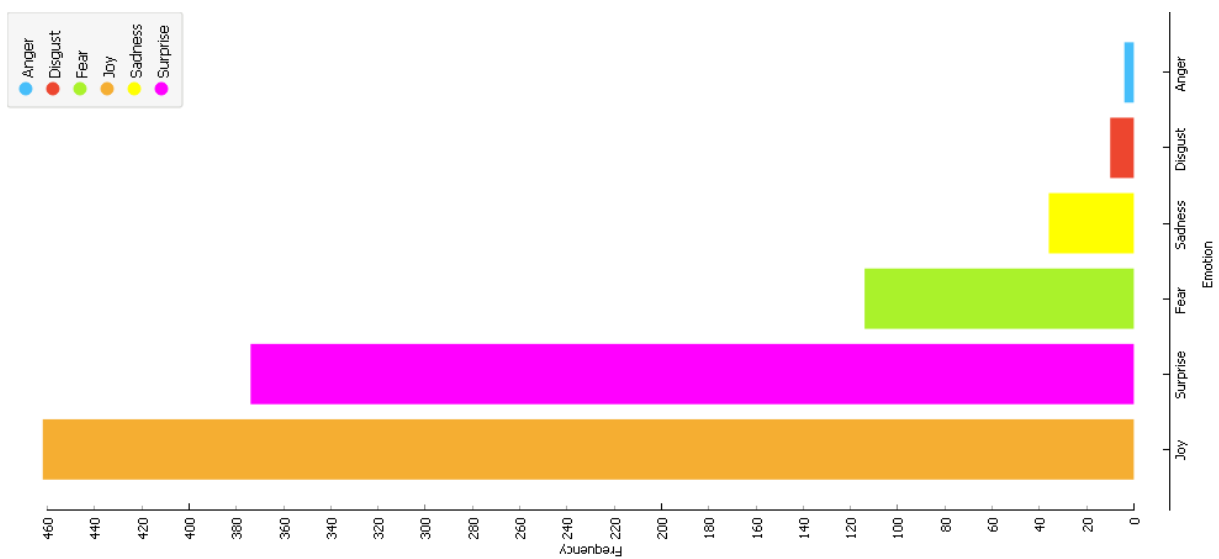
**APÊNDICE C – CANDIDATO JAIR BOLSONARO – TERCEIRO DEBATE –
(DEBATE DA BAND, UOL, FOLHA DE S. PAULO E TV CULTURA) – 16/10/2022**

Figura 65 - HeatMap Pré-Debate - Jair Bolsonaro



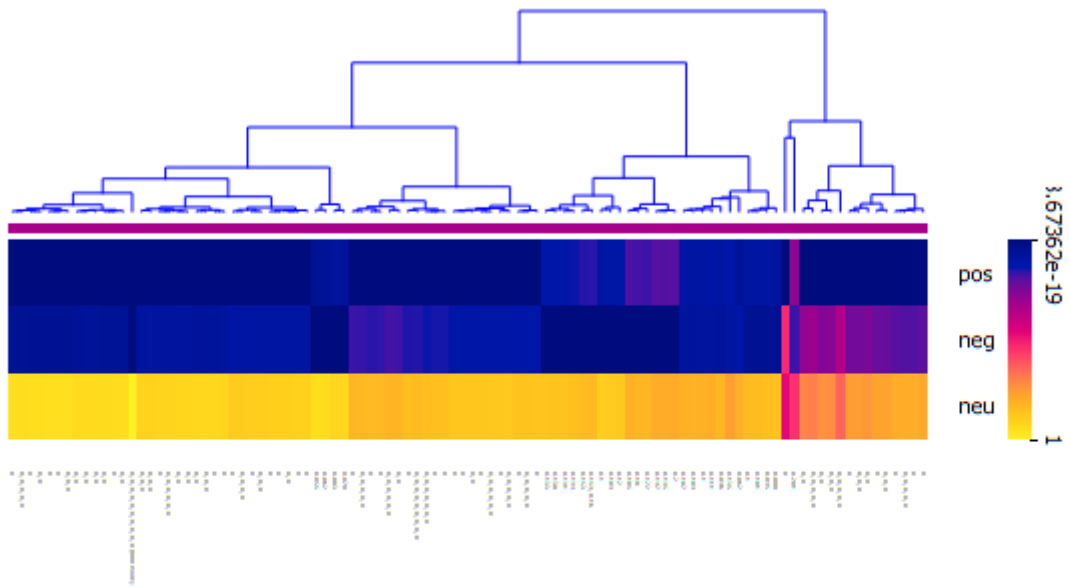
Fonte: Elaborado pela autora, 2022.

Figura 66 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro



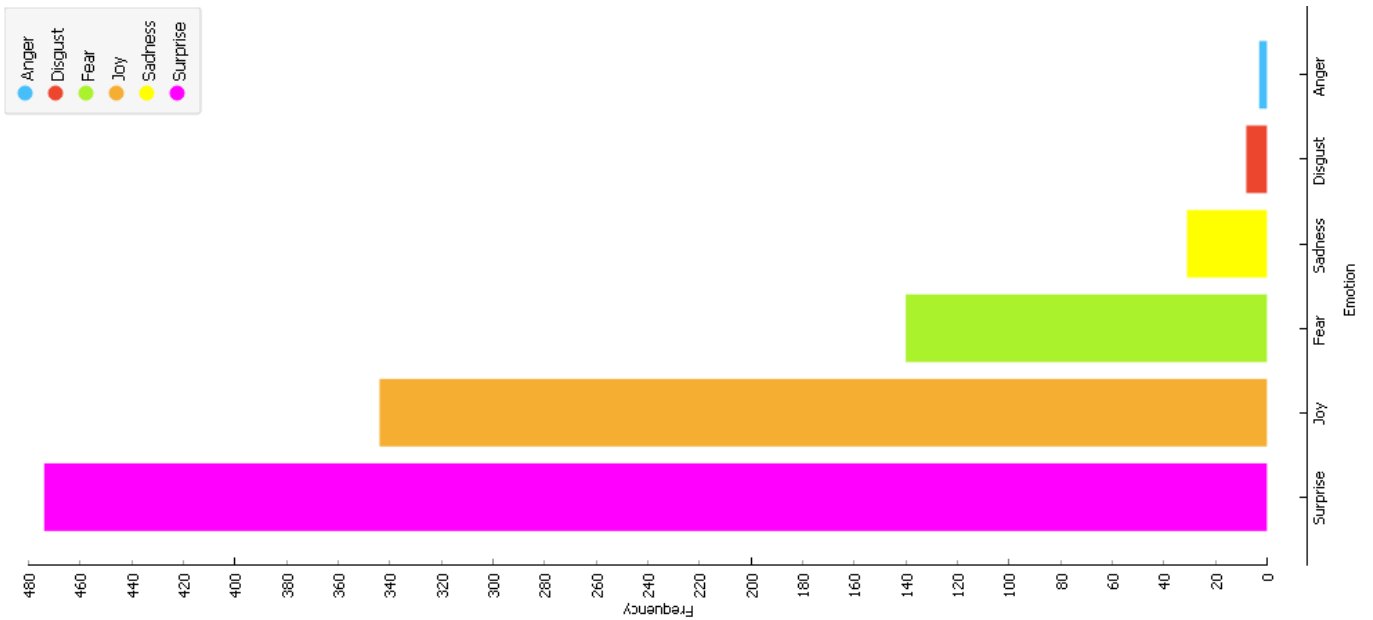
Fonte: Elaborado pela autora, 2022.

Figura 67 - HeatMap Debate - Jair Bolsonaro73



Fonte: Elaborado pela autora, 2022.

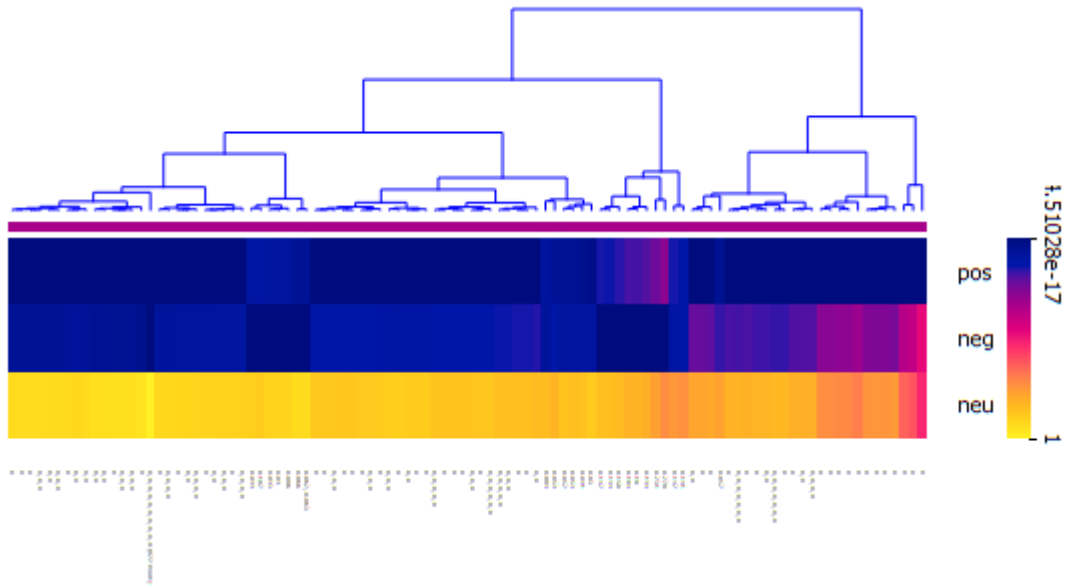
Figura 74 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

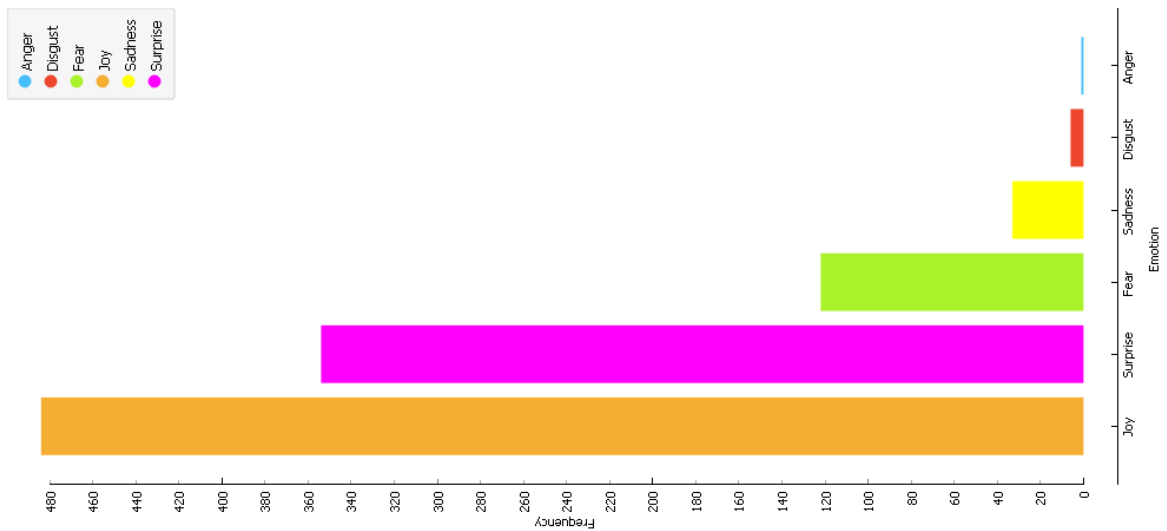
APÊNDICE D – CANDIDATO JAIR BOLSONARO – QUARTO DEBATE – (TV GLOBO) – 28/10/2022

Figura 70 - HeatMap Pré-Debate - Jair Bolsonaro



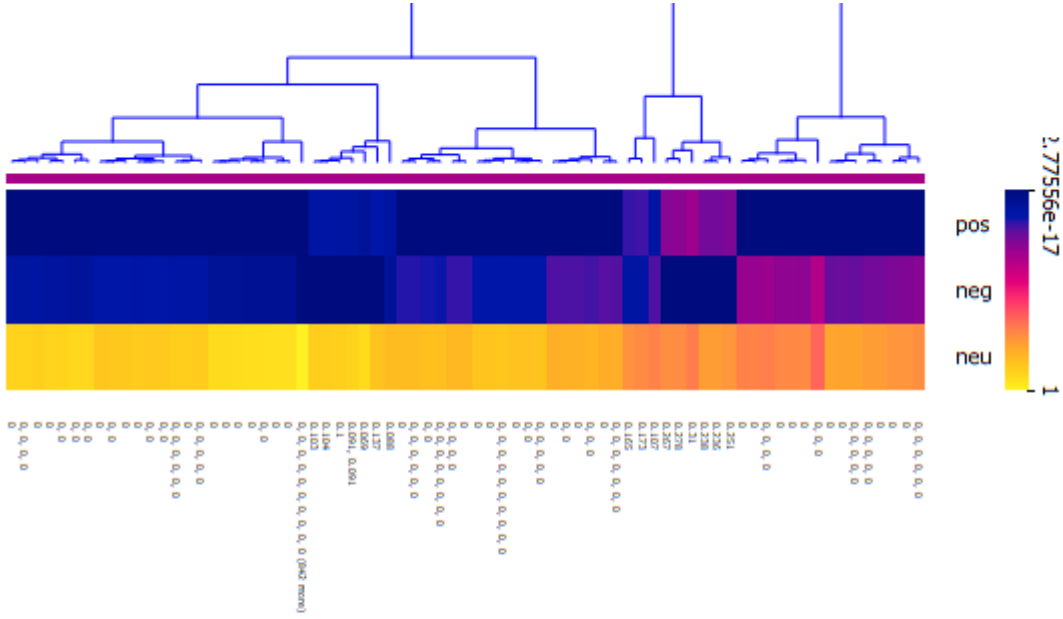
Fonte: Elaborado pela autora, 2022.

Figura 71 - Gráfico de Emoções Pré-Debate - Jair Bolsonaro



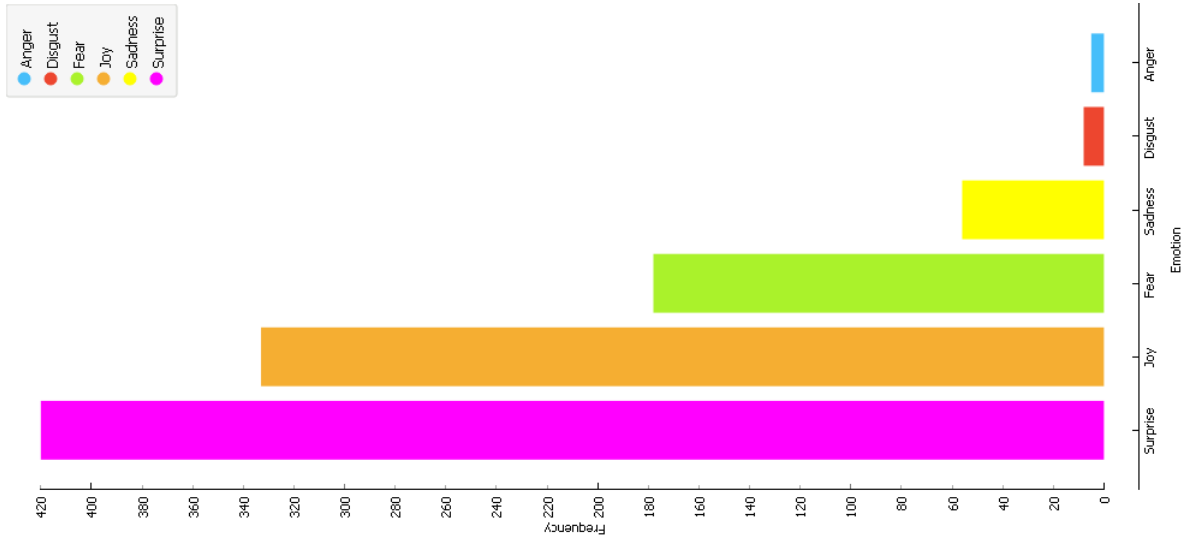
Fonte: Elaborado pela autora, 2022.

Figura 85 - HeatMap Debate - Jair Bolsonaro



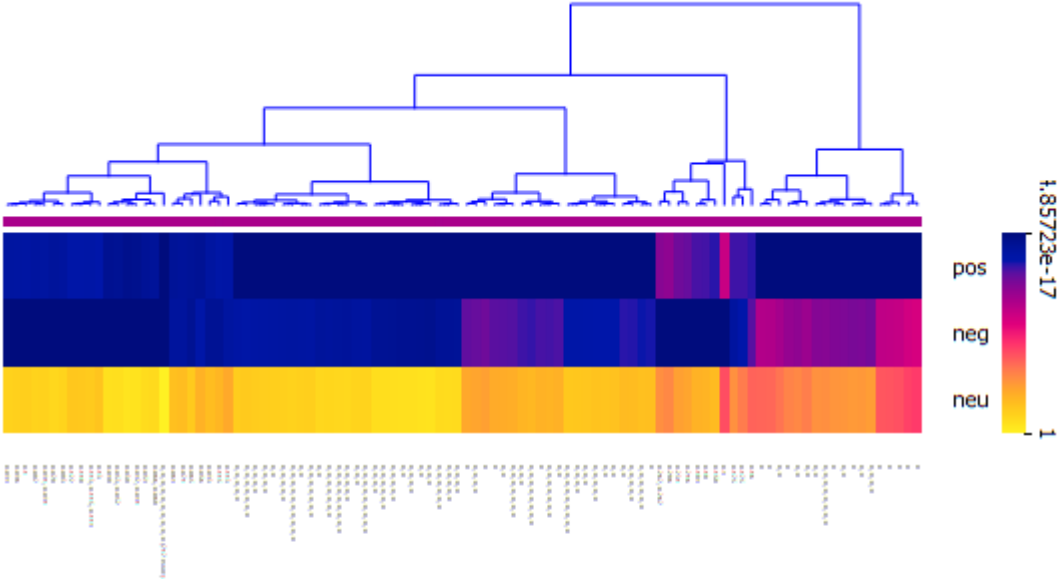
Fonte: Elaborado pela autora, 2022.

Figura 86 - Gráfico de Emoções Debate - Jair Bolsonaro



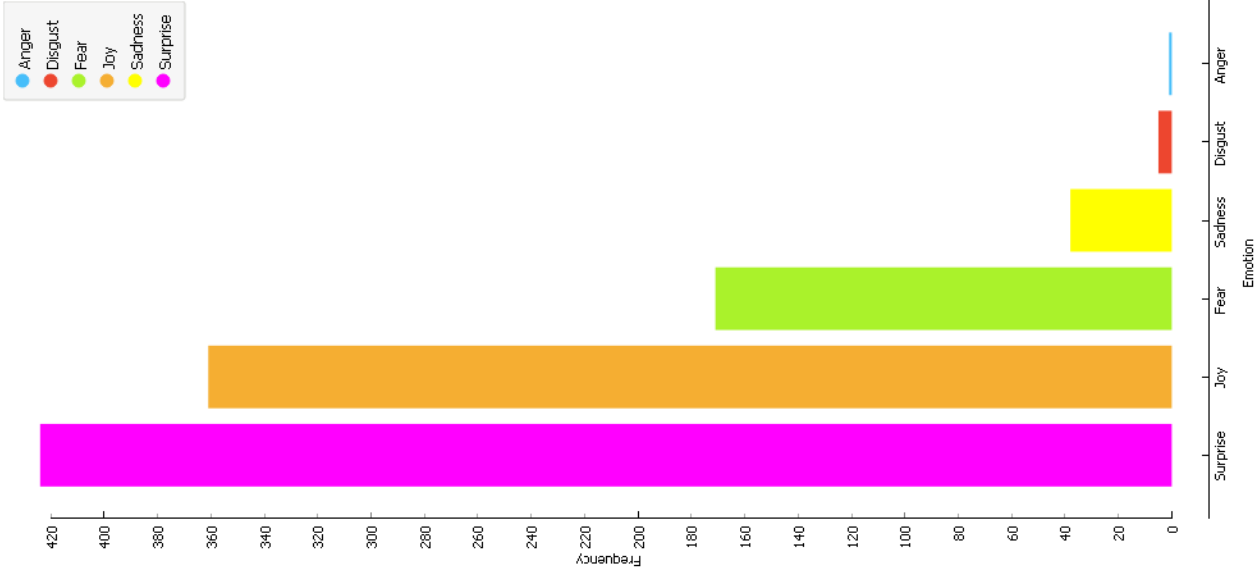
Fonte: Elaborado pela autora, 2022.

Figura 89 - HeatMap Pós-Debate - Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

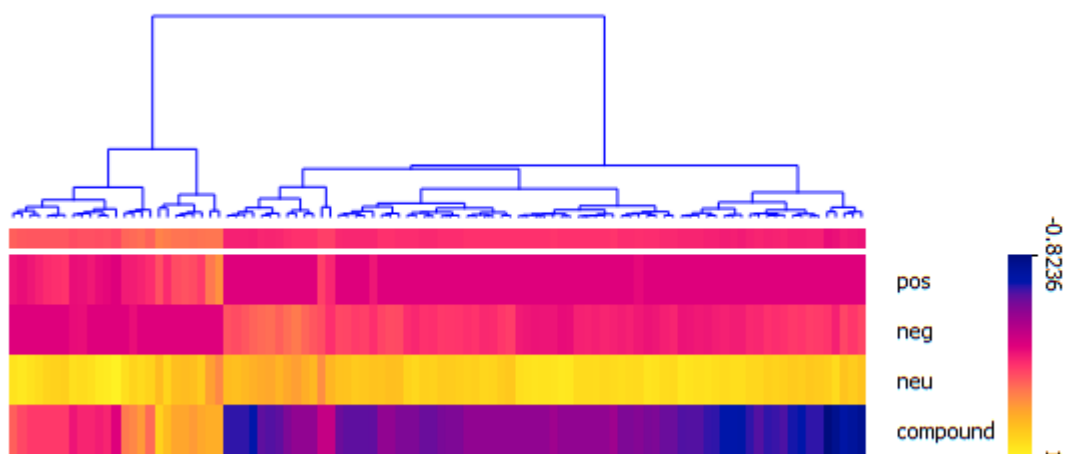
Figura 90 - Gráfico de Emoções Pós-Debate - Jair Bolsonaro



Fonte: Elaborado pela autora, 2022.

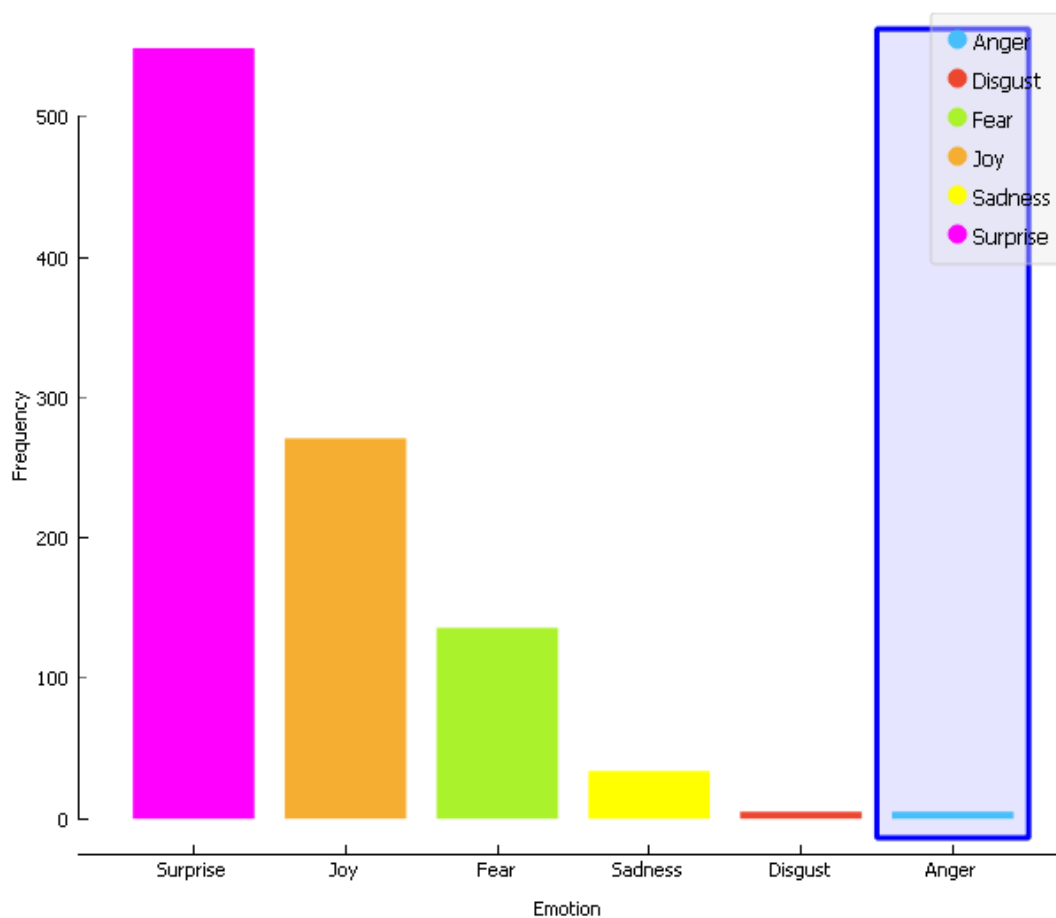
APÊNDICE E – CANDIDATO LUIZ INÁCIO LULA DA SILVA – PRIMEIRO DEBATE - (BAND, UOL, FOLHA DE S. PAULO E TV CULTURA) – 28/08/2022

Figura 74 - *HeatMap* Pré-Debate - Luiz Inácio Lula da Silva



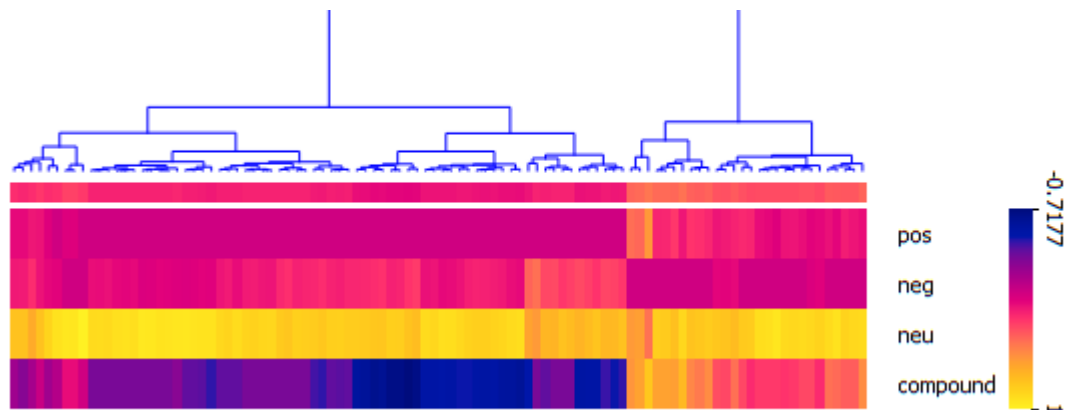
Fonte: Elaborado pela autora, 2022.

Figura 75 - Gráfico de Emoções Pré-Debate - Luiz Inácio Lula da Silva



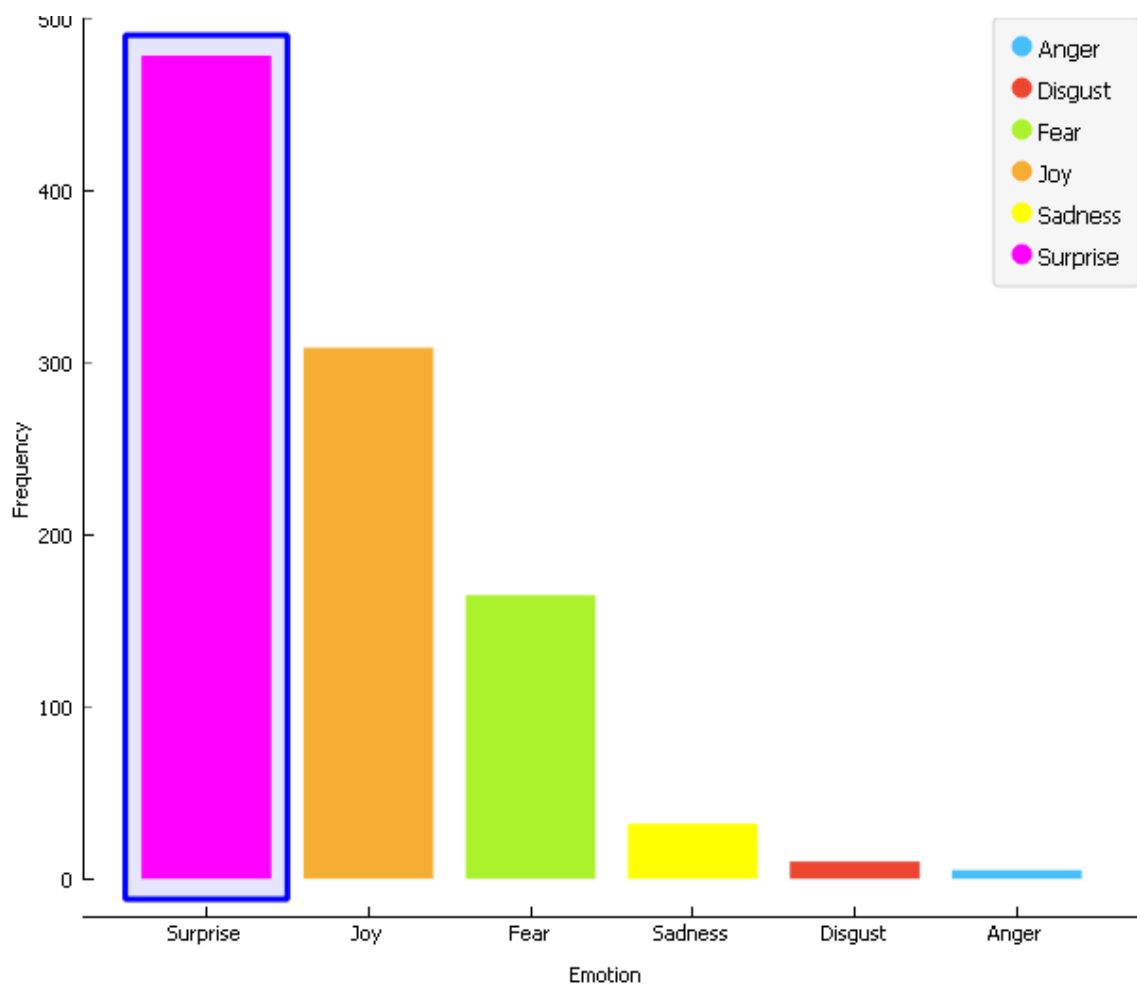
Fonte: Elaborado pela autora, 2022.

Figura 76 - *HeatMap* Debate - Luiz Inácio Lula da Silva



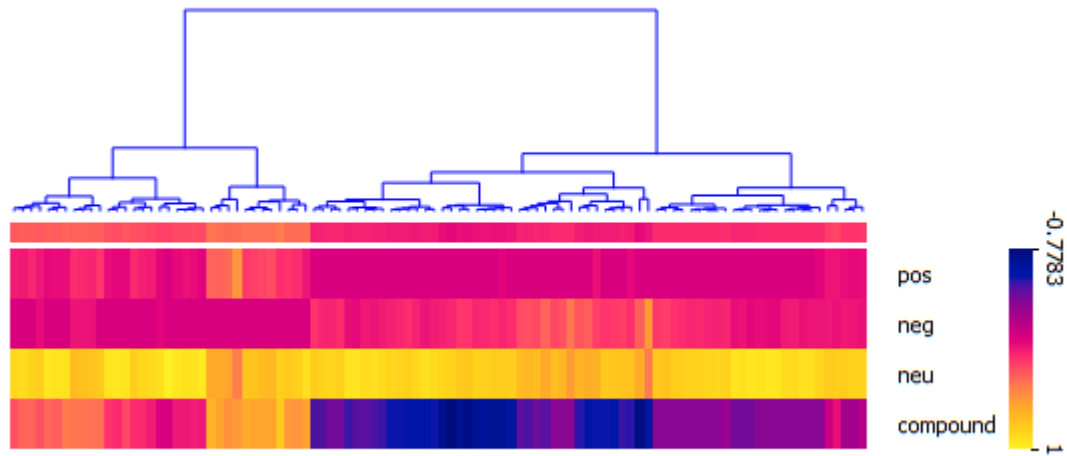
Fonte: Elaborado pela autora, 2022.

Figura 77 - Gráfico de Emoções Debate - Luiz Inácio Lula da Silva



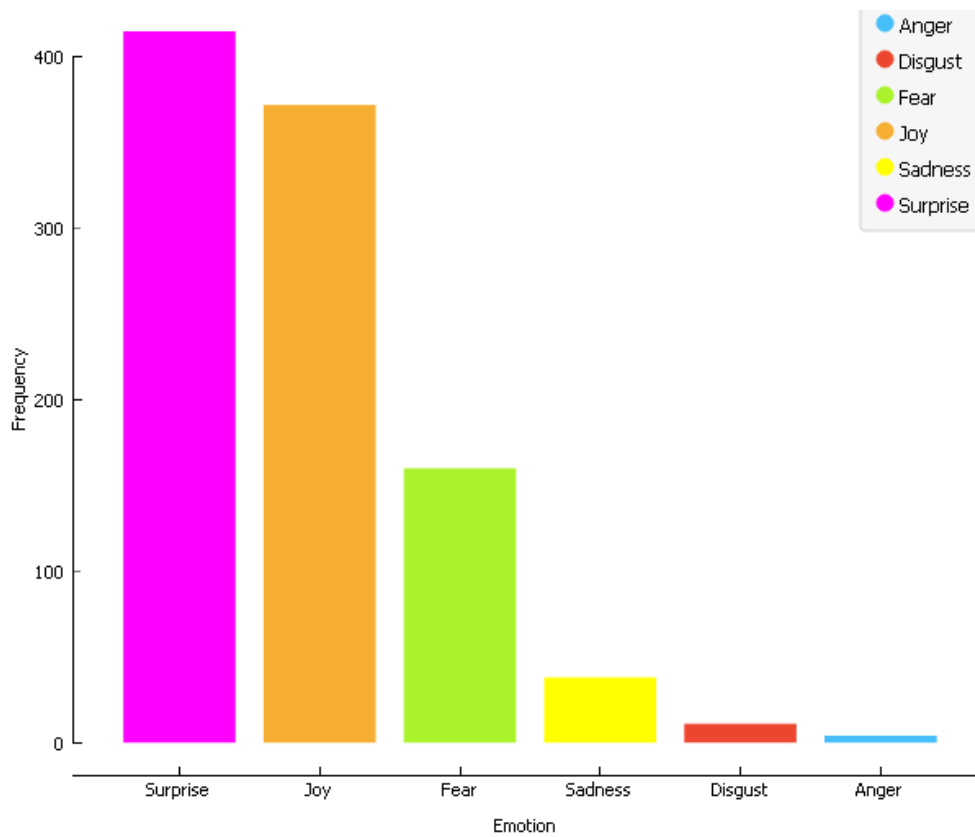
Fonte: Elaborado pela autora, 2022.

Figura 78 - *HeatMap* Pós-Debate - Luiz Inácio Lula da Silva



Fonte: Elaborado pela autora, 2022.

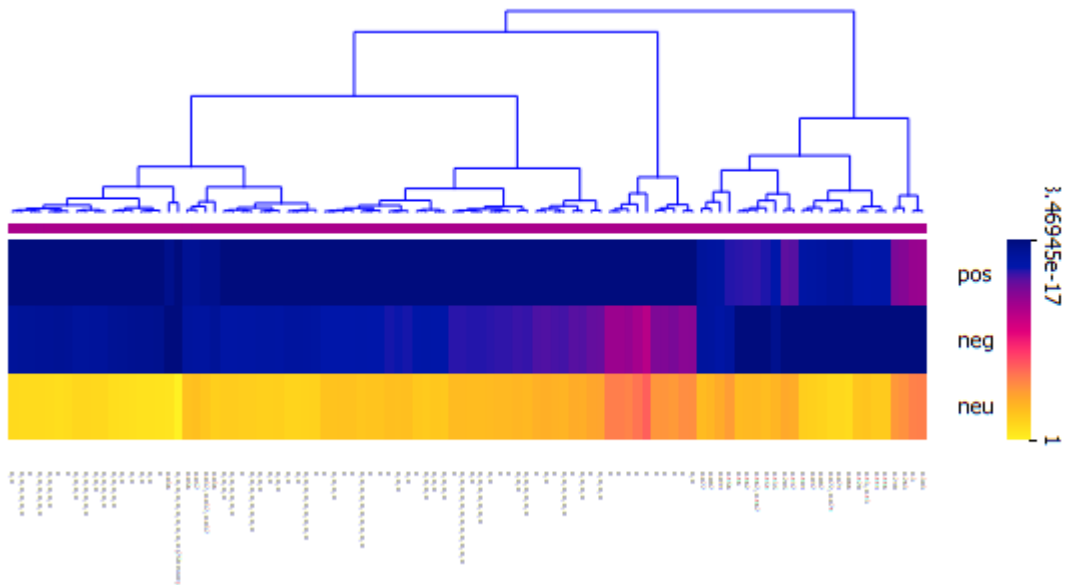
Figura 79 - Gráfico de Emoções Pós-Debate - Luiz Inácio Lula da Silva



Fonte: Elaborado pela autora, 2022.

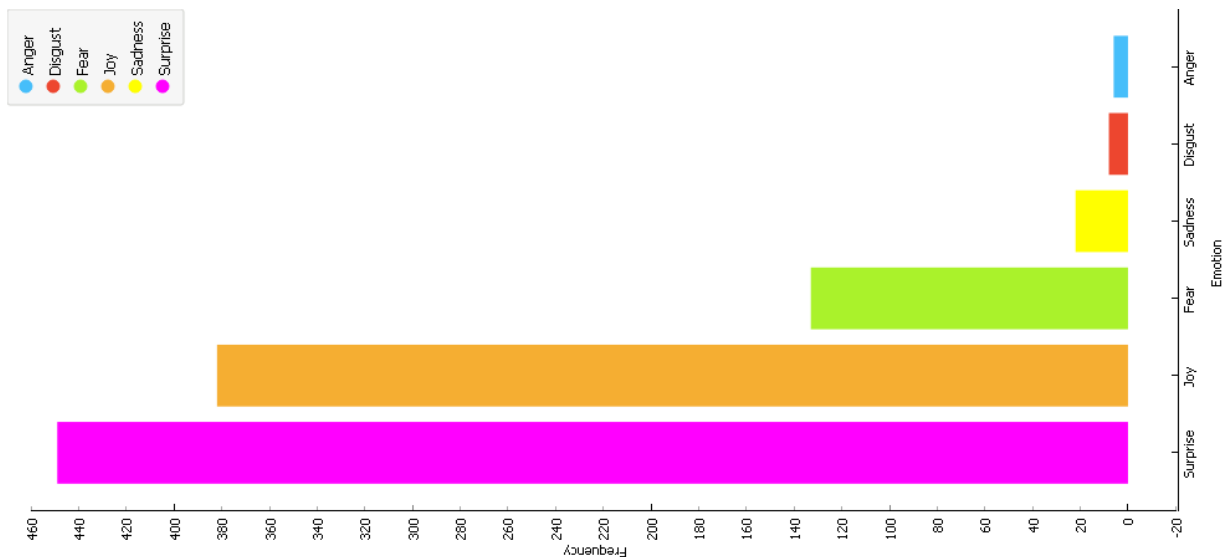
APÊNDICE F – CANDIDATO LUIZ INÁCIO LULA DA SILVA – TERCEIRO DEBATE – (DEBATE DA BAND, UOL, FOLHA DE S. PAULO E TV CULTURA) – 16/10/2022

Figura 82 - HeatMap Pré-Debate - Luiz Inácio Lula da Silva



Fonte: Elaborado pela autora, 2022.

Figura 83 - Gráfico de Emoções Pré-Debate - Luiz Inácio Lula da Silva



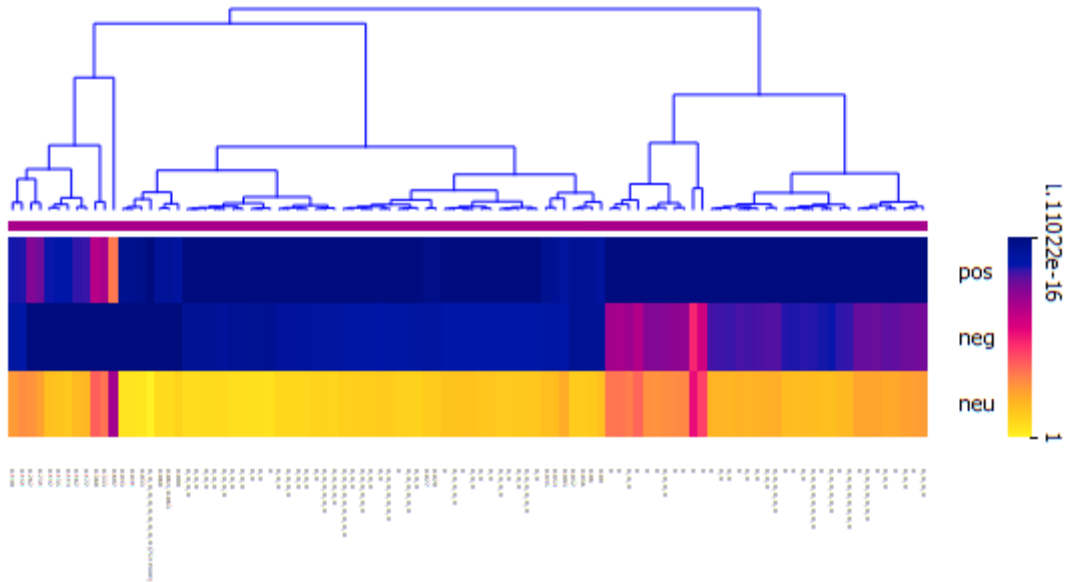
Fonte: Elaborado pela autora, 2022.

Figura 84 - WordCloud Pós-Processada - Debate - Luiz Inácio Lula da Silva



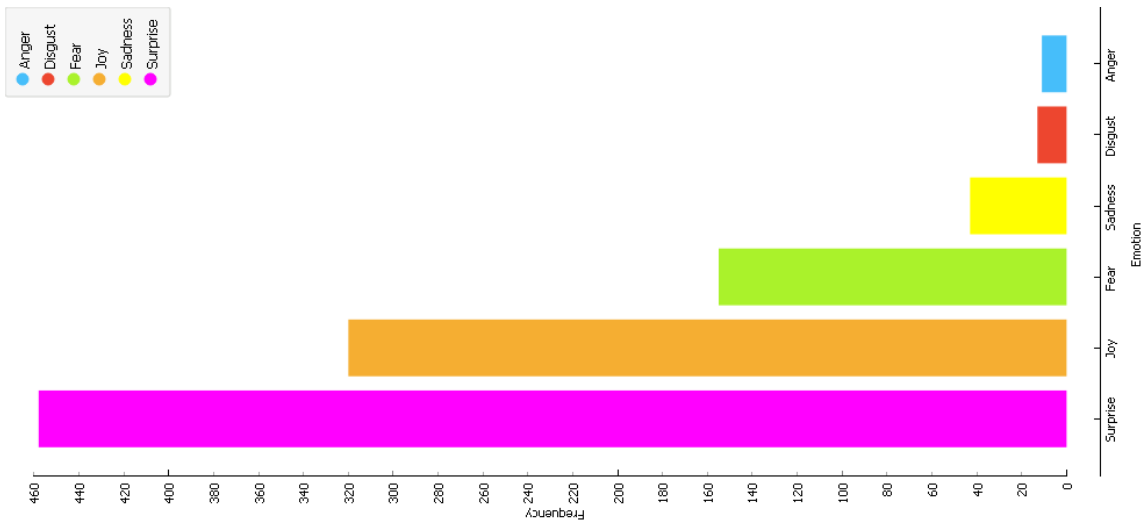
Fonte: Elaborado pela autora, 2022.

Figura 85 - HeatMap Pós-Debate - Luiz Inácio Lula da Silva



Fonte: Elaborado pela autora, 2022.

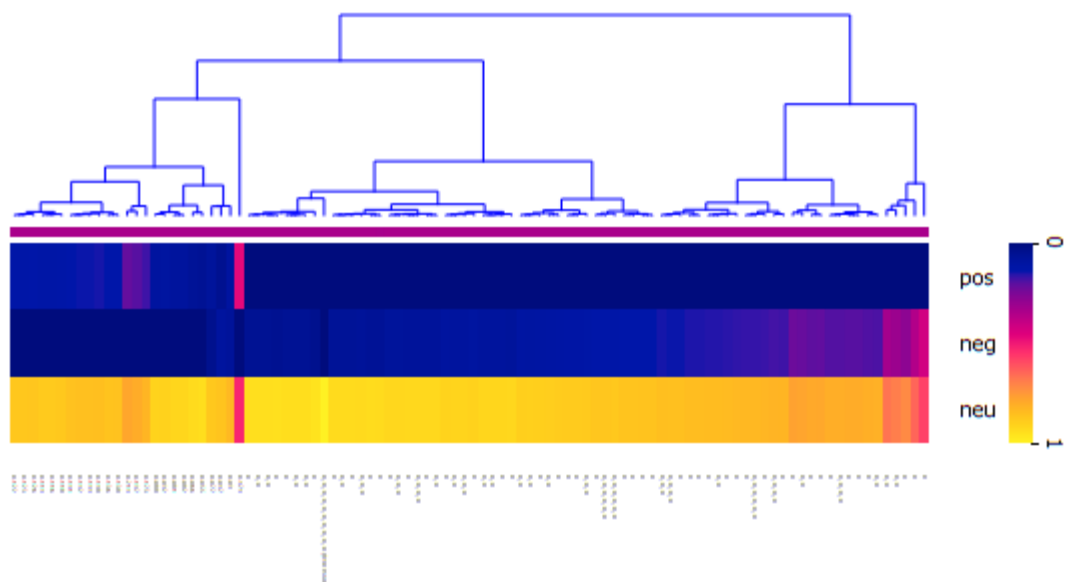
Figura 86 - Gráfico de Emoções Pós-Debate - Luiz Inácio Lula da Silva



Fonte: Elaborado pela autora, 2022.

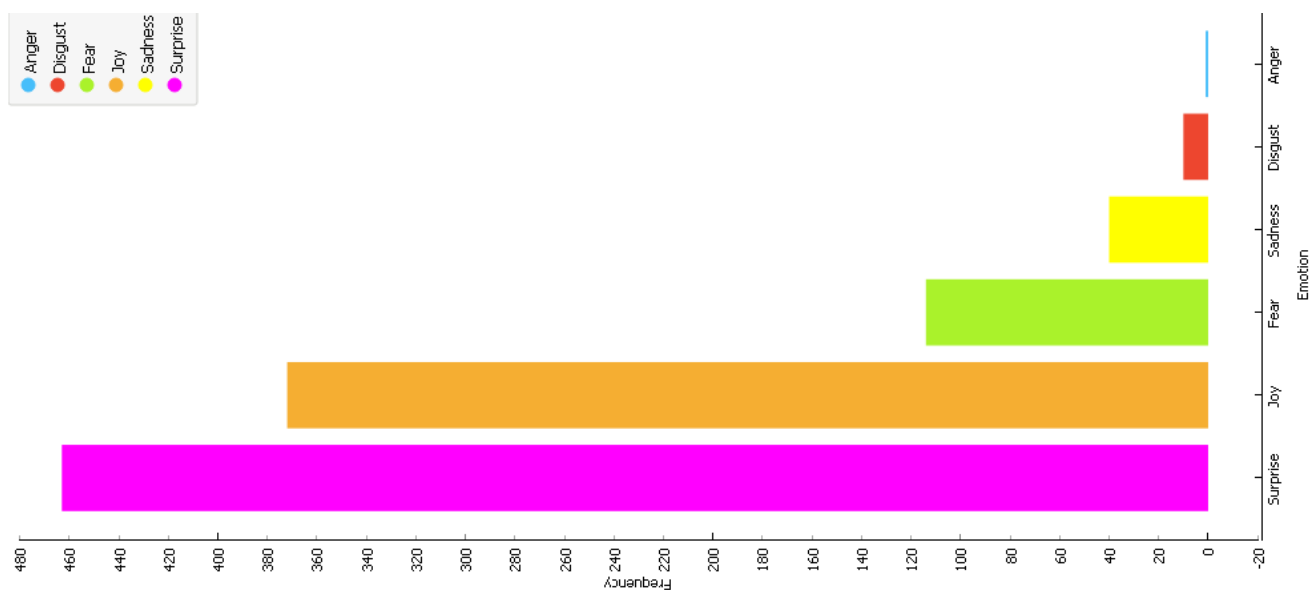
APÊNDICE G – CANDIDATO LUIZ INÁCIO LULA DA SILVA – QUARTO DEBATE – (TV GLOBO) – 28/10/2022

Figura 87 - *HeatMap* Pré-Debate - Luiz Inácio Lula da Silva



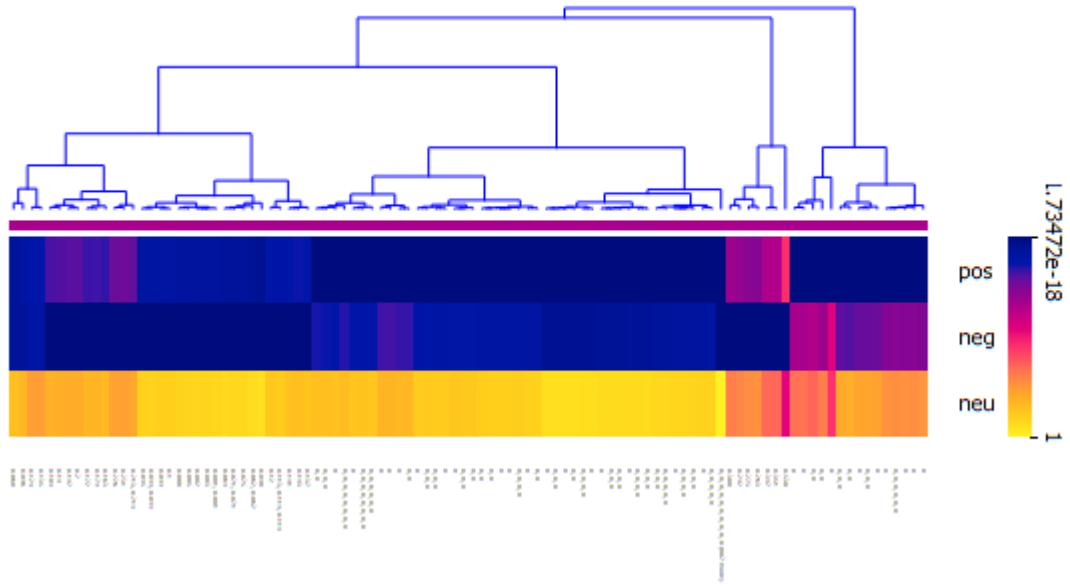
Fonte: Elaborado pela autora, 2022.

Figura 88 - Gráfico de Emoções Pré-Debate - Luiz Inácio Lula da Silva



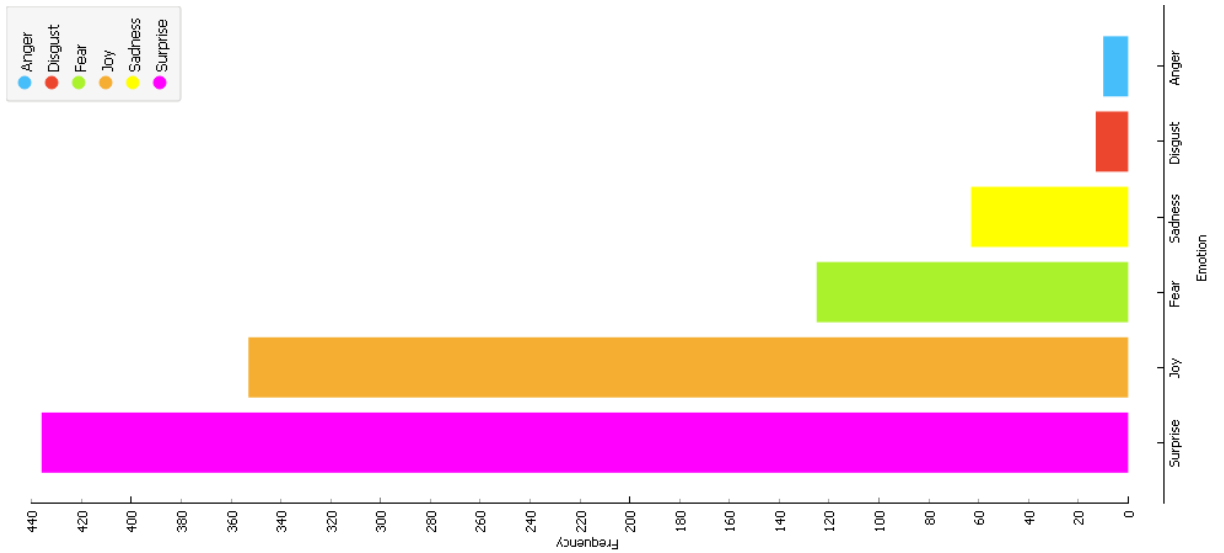
Fonte: Elaborado pela autora, 2022.

Figura 91 - HeatMap Debate - Luiz Inácio Lula da Silva



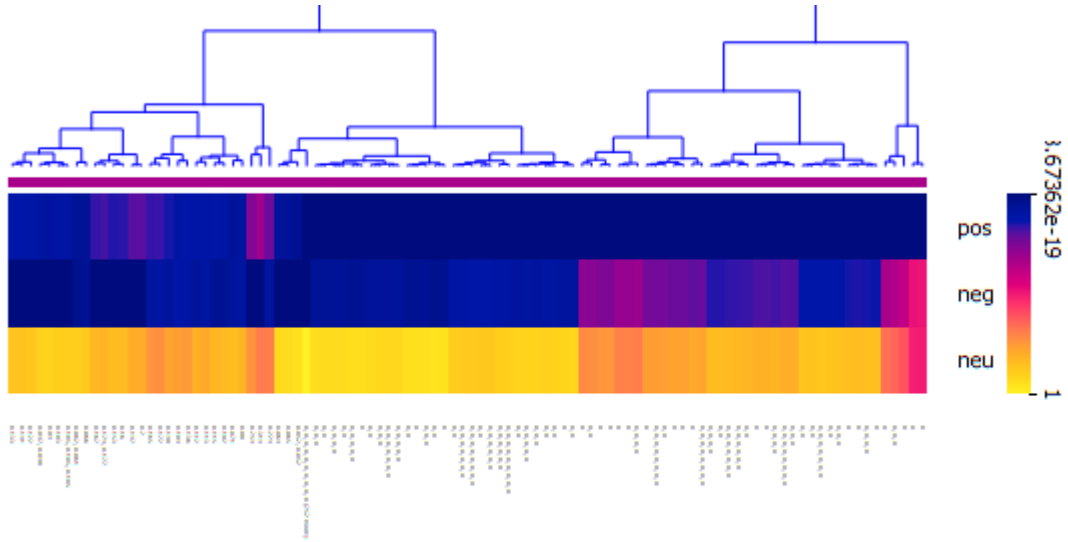
Fonte: Elaborado pela autora, 2022.

Figura 92 - Gráfico de Emoções Debate - Luiz Inácio Lula da Silva



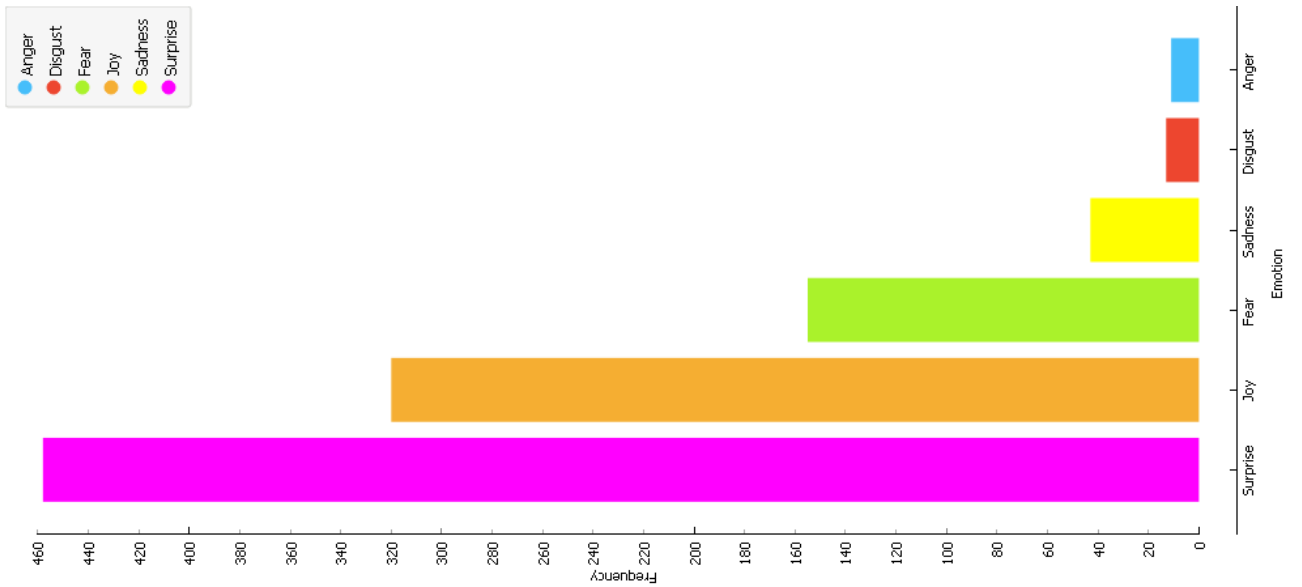
Fonte: Elaborado pela autora, 2022.

Figura 93 - HeatMap Pós-Debate - Luiz Inácio Lula da Silva



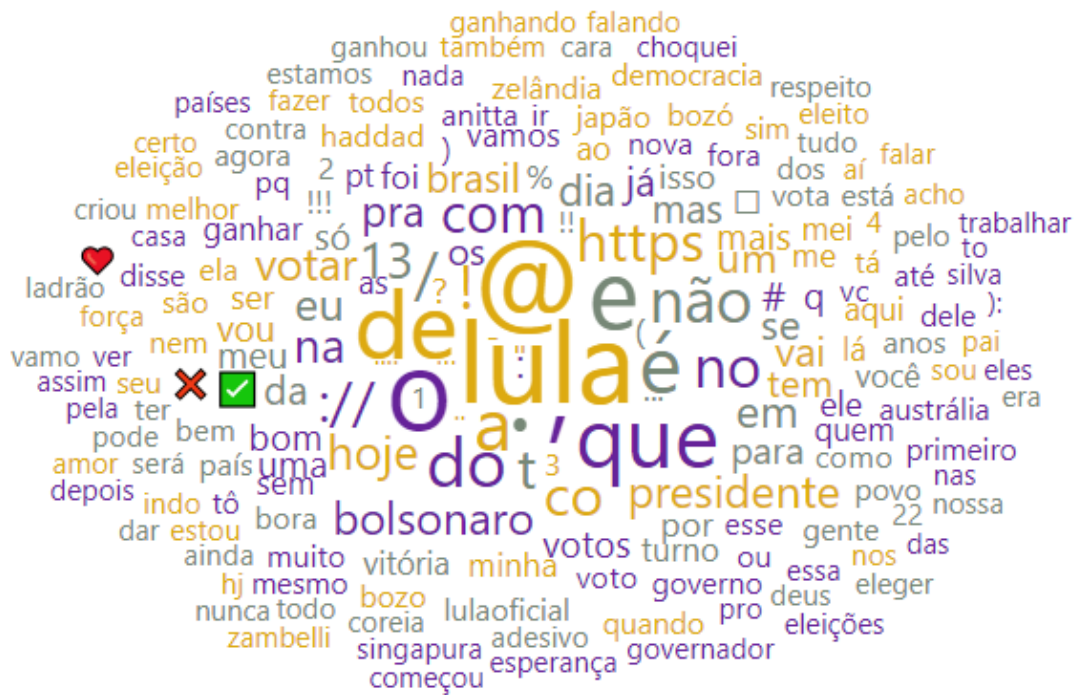
Fonte: Elaborado pela autora, 2022.

Figura 94 - Gráfico de Emoções Pós-Debate - Luiz Inácio Lula da Silva



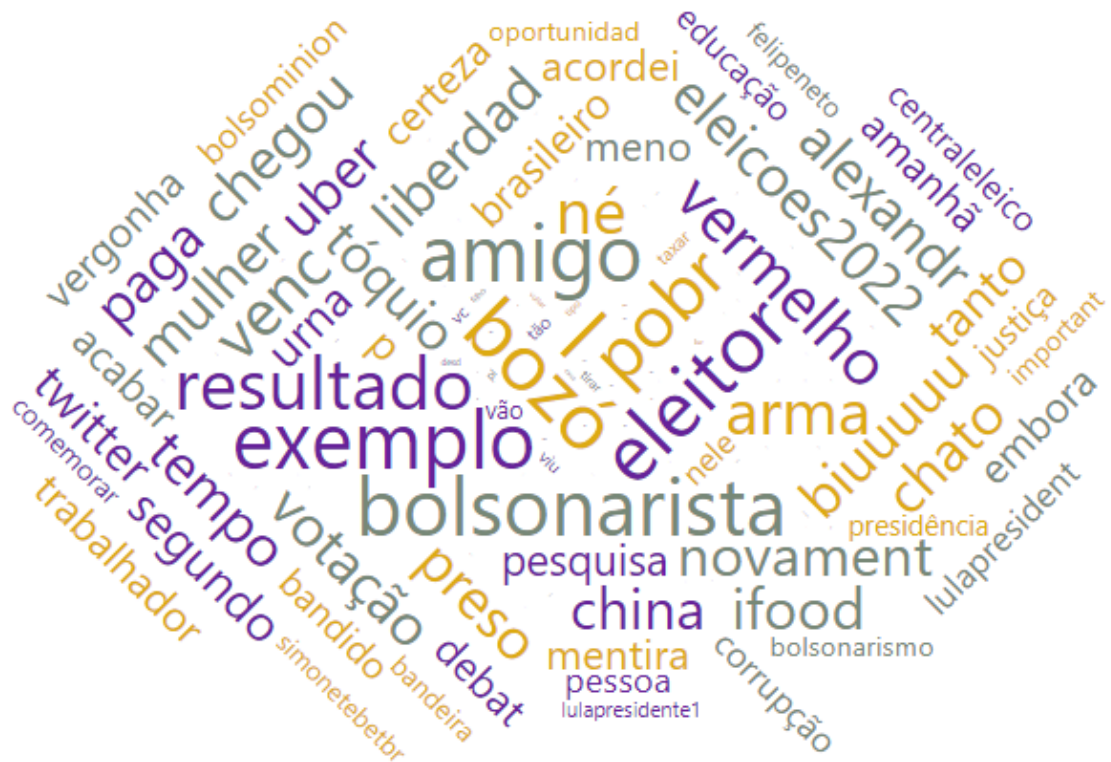
Fonte: Elaborado pela autora, 2022.

Figura 126 - WordCloud Pós-Debate - Luiz Inácio Lula da Silva



Fonte: Elaborado pela autora, 2022.

Figura 127 - WordCloud Pós-Processada - Pós-Debate - Luiz Inácio Lula da Silva



Fonte: Elaborado pela autora, 2022.

ANEXO A – STOPWORDS

a	aqueles	depois	disso	estes
à	aquilo	dessa	disto	esta
agora	as	dessas	dito	está
ainda	até	deste	diz	estamos
alguém	através	destes	dizem	estão
algum	cada	desta	do	estas
alguma	coisa	destas	dos	estava
algumas	coisas	deste	e	estavam
alguns	com	deste	é	estávamos
ampla	como	destes	e'	este
amplas	contra	deve	ela	estes
amplo	contudo	devem	elas	estou
amplos	da	devendo	ele	eu
ante	daquele	dever	eles	fazendo
antes	daqueles	deverá	em	fazer
ao	das	deverão	enquanto	feita
aos	de	deveria	entre	feitas
após	dela	deveriam	era	feito
aquela	delas	devia	essa	feitos
aquelas	dele	deviam	essas	foi
aquele	deles	disse	este	for

foram	meus	nosso	perante	próprias
fosse	minha	nossos	pode	próprio
fossem	minhas	num	pôde	próprios
grande	muita	numa	podendo	q
grandes	muitas	nunca	poder	quais
há	muito	o	poderia	qual
isso	muitos	os	poderiam	quando
isto	na	ou	podia	quanto
já	não	outra	podiam	quantos
la	nas	outras	pois	que
la	nem	outro	por	quem
lá	nenhum	outros	porém	são
lhe	nessa	para	porque	se
lhes	nessas	pela	posso	seja
lo	nesta	pelas	pouca	sejam
mas	nestas	pelo	poucas	sem
me	ninguém	pelos	pouco	sempre
mesma	no	pequena	poucos	sendo
mesmas	nos	pequenas	pra	será
mesmo	nós	pequeno	primeiro	serão
mesmos	nossa	pequenos	primeiros	seu
meu	nossas	per	própria	seus

si	todas	vir
sido	todavia	vos
só	todo	vós
sob	todos	1
sobre	tu	2
sua	tua	3
suas	tuas	4
talvez	tudo	5
também	última	6
tampouco	últimas	7
te	último	8
tem	últimos	9
tendo	um	0
tenha	uma	
ter	umas	
teu	uns	
teus	vendo	
ti	vai	
tido	vamos	
tinha	ver	
tinham	vez	
toda	vindo	