

---

**Análises de Publicações da Rede Social**  
***Instagram* em Contraste com o Temperamento**  
**Humano**

---

**Lara Mondini Martins**



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia  
2022



Lara Mondini Martins

**Análises de Publicações da Rede Social**  
***Instagram* em Contraste com o Temperamento**  
**Humano**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Prof<sup>ª</sup> Dr<sup>ª</sup> Elaine Ribeiro de Faria Paiva

Coorientador: Prof<sup>ª</sup> Dr<sup>ª</sup> Maria Camila Nardini Barioni

Uberlândia

2022

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

M386  
2022

Martins, Lara Mondini, 1998-  
Análises de Publicações da Rede Social Instagram em  
Contraste com o Temperamento Humano [recurso eletrônico]  
/ Lara Mondini Martins. - 2022.

Orientadora: Elaine Ribeiro de Faria Paiva.  
Coorientadora: Maria Camila Nardini Barioni.  
Dissertação (Mestrado) - Universidade Federal de  
Uberlândia, Pós-graduação em Ciência da Computação.  
Modo de acesso: Internet.  
Disponível em: <http://doi.org/10.14393/ufu.di.2022.664>  
Inclui bibliografia.  
Inclui ilustrações.

1. Computação. I. Paiva, Elaine Ribeiro de Faria, 1980-  
, (Orient.). II. Barioni, Maria Camila Nardini, 1978-  
, (Coorient.). III. Universidade Federal de Uberlândia.  
Pós-graduação em Ciência da Computação. IV. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:  
Gizele Cristine Nunes do Couto - CRB6/2091  
Nelson Marcos Ferreira - CRB6/3074




**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**

Coordenação do Programa de Pós-Graduação em Ciência da Computação  
 Av. João Naves de Ávila, 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902  
 Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpqfacom@ufu.br


**ATA DE DEFESA - PÓS-GRADUAÇÃO**

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Dissertação de Mestrado 24/2022, PPGCO				
Data:	09 de dezembro de 2022	Hora de início:	09:00	Hora de encerramento:	12:10
Matrícula do Discente:	12022CCP005				
Nome do Discente:	Lara Mondini Martins				
Título do Trabalho:	Análises de Publicações da Rede Social Instagram em Contraste com o Temperamento Humano				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Inteligência Artificial				
Projeto de Pesquisa de vinculação:	-				

Reuniu-se, por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Prof. Dr. Paulo Henrique Ribeiro Gabriel - FACOM/UFU; Prof.<sup>ª</sup> Dr.<sup>ª</sup> Maria Camila Nardini Barioni - FACOM/UFU (Coorientadora); Prof. Dr. Luiz Henrique de Campos Merschmann - UFLA e Prof.<sup>ª</sup> Dr.<sup>ª</sup> Elaine Ribeiro de Faria Paiva - FACOM/UFU, orientadora da candidata.

Os examinadores participaram desde as seguintes localidades: Luiz Henrique de Campos Merschmann - Lavras/MG; Paulo Henrique Ribeiro Gabriel, Maria Camila Nardini Barioni e Elaine Ribeiro de Faria Paiva - Uberlândia/MG. A discente participou da cidade de Uberlândia/MG.

Iniciando os trabalhos a presidente da mesa, Prof.<sup>ª</sup> Dr.<sup>ª</sup> Elaine Ribeiro de Faria Paiva, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu a Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir a senhora presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir a candidata. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando a candidata:

**Aprovada**

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Elaine Ribeiro de Faria Paiva, Professor(a) do Magistério Superior**, em 12/12/2022, às 09:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Paulo Henrique Ribeiro Gabriel, Professor(a) do Magistério Superior**, em 12/12/2022, às 10:43, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Maria Camila Nardini Barioni, Professor(a) do Magistério Superior**, em 12/12/2022, às 11:53, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Luiz Henrique de Campos Merschmann, Usuário Externo**, em 14/12/2022, às 16:28, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **4111915** e o código CRC **6897CA56**.

*Este trabalho é dedicado às crianças adultas que,  
quando pequenas, sonharam em se tornar cientistas.*



---

# Agradecimentos

A Deus, por ter me dado força, sabedoria e saúde para cursar o mestrado.

A minha família e amigos, pelo apoio de sempre.

Aos professores que me apoiaram e compartilharam sua experiência e conhecimento com maestria e dedicação.

Ao Programa de Pós Graduação da Universidade Federal de Uberlândia, por ter me dado a oportunidade e todo o suporte necessário durante o curso de mestrado.

A FAPEMIG, pelo financiamento do projeto.



*“A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original.”*  
*(Albert Einstein)*





---

# Resumo

Com o crescimento do uso das redes sociais, as possibilidades de estudos sobre relações e interações sociais têm crescido significativamente. Compreender como os usuários expressam seus sentimentos e manifestam seus temperamentos nas redes sociais pode ser um passo para antecipar os transtornos psicológicos. O *Instagram* tem bilhões de usuários e está entre as redes sociais mais utilizadas atualmente. No entanto, essa rede ainda é pouco explorada como fonte de estudo para o temperamento humano. Este trabalho tem o objetivo de analisar as relações entre o temperamento dos usuários e seus dados coletados na rede social *Instagram*. Para a análise dos dados textuais das postagens dos usuários são utilizadas duas estratégias de classificação de sentimentos. A primeira delas é um *ensemble* de classificadores e a segunda é um classificador em níveis. Ambas as estratégias foram avaliadas em três bases de dados e alcançaram resultados de acurácia acima de 77%. O classificador *ensemble* obteve melhores resultados nas três bases com acurácia acima de 80%. Para a análise da relação entre temperamento e os dados do *Instagram*, cada usuário é representado pelo temperamento (atributo alvo), quantidade e média de curtidas, quantidade de publicações e pelas proporções de uso de *emojis* nas publicações, de publicações positivas, de publicações negativas, de publicações durante a semana, de publicações aos finais de semana e de seguidos e seguidores. Nessa análise, foram utilizados testes estatísticos, gráficos e árvore de decisão. Os resultados indicam que usuários com temperamento depressivo postam mais legendas com sentimento positivo do que usuários hipertímicos, irritados e preocupados. O comportamento de publicações durante a semana e aos finais de semana dos temperamentos irritável e preocupado são similares, assim como o comportamento dos usuários ciclotímicos e depressivos. E finalmente, os usuários ansiosos usam mais *emojis* nas legendas do Instagram do que usuários deprimidos e irritados.

**Palavras-chave:** redes sociais, temperamento, análise de sentimentos, *Instagram*.



---

# Abstract

With the growth in the use of social networks, the possibilities for studies on social relationships and interactions have grown significantly. Understanding how users express their feelings and manifest their temperaments on social networks can be a step for anticipating psychological disorders. Instagram has billions of users and is the most used social networks today. However, this network is still little explored as a source of study for the human temperament. This work aims to analyze the relationships between the temperament of users and their data collected on the social network Instagram. For the analysis of the textual data of the users' posts, two sentiment classification strategies are used. The first one is an ensemble of classifiers and the second one is a layered classifier. Both strategies were evaluated in three databases and achieved accuracy results above 77%. The ensemble classifier obtained better results in the three bases with accuracy above 80%. For the analysis of the relationship between temperament and Instagram data, each user is represented by temperament (target attribute), number and average of likes, number of publications and proportions of use of emojis in publications, positive publications, negative publications, publications during the week, publications on weekends and followers. In this analysis, statistical tests, graphs and decision tree were used. The results indicate that users with a depressive mood post more positive-feeling captions than hyperthymic, irritable, and worried users. The behavior of publications during the week and on weekends of irritable and worried temperaments are similar, as well as the behavior of cyclothymic and depressive users. And finally, anxious users use more emojis in Instagram captions than depressed and angry users.

**Keywords:** social networks, temperament, sentiment analysis, Instagram.



---

## Lista de ilustrações

Figura 1 – Exemplo de Frase Para Análise de Sentimento - <i>Tokenização</i> . . . . .	36
Figura 2 – Exemplo de Frase Para Análise de Sentimento — Remoção de <i>Stop Words</i> . Fonte: elaborado pela autora (2022) . . . . .	37
Figura 3 – <i>Exemplo de Bag-of-Words</i> . Fonte: elaborado pela autora (2022) . . . . .	38
Figura 4 – Exemplo de Aprendizado-Não Supervisionado. Fonte: Adaptado de Benevenuto, Ribeiro e Araújo (2015) . . . . .	40
Figura 5 – Árvore de Decisão Sobre Ir à Praia. Fonte: elaborado pela autora (2022)	42
Figura 6 – Representação de uma <i>Random Forest</i> . Fonte: elaborado pela autora (2022) . . . . .	42
Figura 7 – Linear SVM. Fonte: Adaptado de (LEARN, 2021) . . . . .	43
Figura 8 – SVM Kernel. Fonte: Adaptado de (WIKIPEDIA, 2021) . . . . .	44
Figura 9 – Método de Avaliação <i>Hold-out</i> . Fonte: elaborado pela autora (2022) . .	44
Figura 10 – Método de Avaliação <i>K-Fold Cross-Validation</i> . Fonte: elaborado pela autora (2022) . . . . .	45
Figura 11 – Matriz de Confusão. Fonte: Adaptado de (LEAL, 2017) . . . . .	46
Figura 12 – Tabela de Relação Entre Homens e Mulheres Graduados. Fonte: Adap- tado de (FISHER, 1922) . . . . .	49
Figura 13 – Método Proposto Para Análise de Temperamento a Partir de Redes Sociais. Fonte: Elaborado pela autora (2022) . . . . .	63
Figura 14 – Site Equipe Social Temp. Fonte: Elaborado pela autora (2022) . . . . .	64
Figura 15 – Exemplo de Publicação no <i>Instagram</i> . Fonte: Elaborado pela autora (2022) . . . . .	66
Figura 16 – Exemplo de classificação de uma sentença usando um <i>ensemble</i> de três classificadores. Fonte: Elaborado pela autora (2022) . . . . .	68
Figura 17 – Classificação de Sentimento em Níveis. Fonte: Elaborado pela autora (2022) . . . . .	69
Figura 18 – Representação de um Usuário. Fonte: Elaborado pela autora (2022) . .	71

Figura 19 – Relação de Temperamentos e Usuários por Temperamento. Fonte: Elaborado pela autora (2022) . . . . .	74
Figura 20 – De-para Temperamentos. Fonte: Elaborado pela autora (2022) . . . . .	75
Figura 21 – Pré-processamento de Legendas. Fonte: Elaborado pela autora (2022) .	76
Figura 22 – Gráfico <i>Emojis</i> e Temperamento. Fonte: Elaborado pela autora (2022)	84
Figura 23 – Gráfico Publicações com Sentimento Positivo e Temperamento. Fonte: Elaborado pela autora (2022) . . . . .	85
Figura 24 – Gráfico Quantidade de Publicações e Temperamento. Fonte: Elaborado pela autora (2022) . . . . .	86
Figura 25 – Gráfico Média de Curtidas e Temperamento. Fonte: Elaborado pela autora (2022) . . . . .	87
Figura 26 – Gráfico Final de Semana e Temperamento. Fonte: Elaborado pela autora (2022) . . . . .	87
Figura 27 – Gráfico Proporção Publicações Durante a Semana e Temperamento. Fonte: Elaborado pela autora (2022) . . . . .	88
Figura 28 – (A) Árvore de Decisão Temperamento e Dados do <i>Instagram</i> . Fonte: Elaborado pela autora (2022) . . . . .	89

---

## Lista de tabelas

Tabela 1 – Resumo Trabalhos Relacionados — Classificação de Texto em Publicações de Redes Sociais em Inglês . . . . .	55
Tabela 2 – Resumo Trabalhos Relacionados — Classificação de Texto em Publicações de Redes Sociais em Português . . . . .	56
Tabela 3 – Resumo Trabalhos Relacionados — Redes Sociais e Temperamento Humano - Parte 1 . . . . .	60
Tabela 4 – Resumo Trabalhos Relacionados — Redes Sociais e Temperamento Humano - Parte 2 . . . . .	61
Tabela 5 – Resumo Bases de Dados . . . . .	76
Tabela 6 – Comparação da acurácia dos diferentes modelos de classificação nas três bases de dados . . . . .	77
Tabela 7 – Medidas de Avaliação: Classificador <i>Ensemble</i> e Classificador em Níveis	77
Tabela 8 – Medidas de Avaliação por Nível do Modelo de Classificação em Níveis .	77
Tabela 9 – Matriz de Confusão Classificação em Níveis - Nível 1 - B1, B2, B3 . . .	80
Tabela 10 – Matriz de Confusão Classificação em Níveis - B1, B2 e B3 . . . . .	80
Tabela 11 – Matriz de Confusão Classificação <i>Ensemble</i> - B1, B2 e B3 . . . . .	81
Tabela 12 – Resultados do Teste t de Student entre temperamento e quantidade de curtidas nas publicações . . . . .	83
Tabela 13 – Resultados do Teste Exato de Fisher entre temperamento e dados da rede social . . . . .	83





---

## Lista de siglas



---

# Sumário

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>25</b>
<b>1.1</b>	<b>Objetivos . . . . .</b>	<b>27</b>
<b>1.2</b>	<b>Hipótese . . . . .</b>	<b>28</b>
<b>1.3</b>	<b>Contribuições . . . . .</b>	<b>28</b>
<b>1.4</b>	<b>Organização da Dissertação . . . . .</b>	<b>29</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA . . . . .</b>	<b>31</b>
<b>2.1</b>	<b>Temperamento Humano e Redes Sociais . . . . .</b>	<b>31</b>
2.1.1	Sentimentos, Emoções e Temperamento . . . . .	31
2.1.2	Questionário de Temperamento . . . . .	32
2.1.3	Influências das Redes Sociais nos Seres Humanos . . . . .	34
<b>2.2</b>	<b>Mineração de Texto em Redes Sociais . . . . .</b>	<b>34</b>
2.2.1	Coleta de Dados . . . . .	35
2.2.2	Pré-processamento . . . . .	36
2.2.3	Normalização de Texto . . . . .	37
2.2.4	Definição de Características . . . . .	37
2.2.5	Técnicas de Aprendizado de Máquina . . . . .	39
2.2.6	Métodos e Medidas de Avaliação . . . . .	44
2.2.7	Outras Tarefas de Mineração de Texto . . . . .	48
<b>2.3</b>	<b>Considerações Finais . . . . .</b>	<b>49</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS . . . . .</b>	<b>51</b>
<b>3.1</b>	<b>Análise de Sentimentos em Publicações de Redes Sociais . . . . .</b>	<b>52</b>
<b>3.2</b>	<b>Redes Sociais e Temperamento Humano . . . . .</b>	<b>57</b>
<b>3.3</b>	<b>Considerações Finais . . . . .</b>	<b>62</b>
<b>4</b>	<b>MÉTODO PARA ANÁLISE DE TEMPERAMENTO A PARTIR DE DADOS DE REDES SOCIAIS . . . . .</b>	<b>63</b>

4.1	Coleta de Dados . . . . .	64
4.2	Rotulação da Base . . . . .	65
4.3	<b>Framework</b> de Extração de Sentimentos . . . . .	65
4.3.1	Pré-processamento . . . . .	65
4.3.2	Representação dos Dados . . . . .	66
4.3.3	Classificação . . . . .	67
4.3.4	Métodos e Medidas de Avaliação . . . . .	69
4.4	<b>Análises de Dados</b> . . . . .	69
4.4.1	Mapeamento do temperamento dos usuários . . . . .	69
4.4.2	Representação do Usuário . . . . .	70
4.4.3	Relação Temperamento e Dados <i>Instagram</i> . . . . .	71
4.5	<b>Considerações Finais</b> . . . . .	72
5	<b>EXPERIMENTOS E ANÁLISE DOS RESULTADOS</b> . . . . .	73
5.1	<b>Base de Dados</b> . . . . .	73
5.1.1	Bases de Sentimentos . . . . .	73
5.1.2	Base de Temperamentos . . . . .	74
5.2	<b>Avaliação do Framework</b> de Extração de Sentimentos . . . . .	75
5.2.1	Pré-processamento e Representação dos Dados . . . . .	75
5.2.2	Avaliação dos Resultados dos Classificadores . . . . .	76
5.3	<b>Relação Entre Dados do Usuário no <i>Instagram</i> e seu Temperamento</b> . . . . .	82
5.3.1	Testes Estatísticos . . . . .	82
5.3.2	Gráficos . . . . .	83
5.3.3	Árvore de Decisão . . . . .	86
5.4	<b>Considerações Finais</b> . . . . .	90
6	<b>CONCLUSÃO</b> . . . . .	91
6.1	<b>Contribuições</b> . . . . .	92
6.2	<b>Trabalhos Futuros</b> . . . . .	93
6.3	<b>Contribuições em Produção Bibliográfica</b> . . . . .	93
	<b>REFERÊNCIAS</b> . . . . .	95

## APÊNDICES 105

APÊNDICE A	–	QUESTIONÁRIO DE TEMPERAMENTO TEMPORÁRIO . . . . .	107
------------	---	---	-----

APÊNDICE B – MANUAL DE CLASSIFICAÇÃO DE LEGEN-  
DAS DO *INSTAGRAM* . . . . . 127



---

## Introdução

As redes sociais contribuem com os relacionamentos interpessoais, sendo considerada um dos meios de comunicação do ser humano (CHAMBERS, 2013). Algumas das ações mais comuns realizadas nas redes sociais são publicação e compartilhamento de assuntos de interesse e interação com pessoas de qualquer lugar do mundo. Essas ações permitem que os usuários façam novos amigos e se conectem cada vez mais com pessoas que estão distantes fisicamente. *Instagram*<sup>1</sup>, *Facebook*<sup>2</sup>, *YouTube*<sup>3</sup>, *TikTok*<sup>4</sup>, *LinkedIn*<sup>5</sup>, *WhatsApp*<sup>6</sup> e *Twitter*<sup>7</sup> são algumas das redes sociais que estão disponíveis na Internet.

De acordo com Social (2020), no período de abril de 2019 a janeiro de 2020, dez milhões de brasileiros se tornaram usuários de redes sociais e passaram em média 3h31m por dia conectados. Ainda segundo o Social (2020), o Brasil ocupa a sexta posição no ranking no uso de redes sociais com 140 milhões de usuários, sendo que no mundo existem cerca de 3.5 bilhões.

Em 2020, o vírus SARS-CoV-2, conhecido popularmente como COVID-19 ou coronavírus, foi responsável pela pandemia do coronavírus que matou milhões de pessoas ao redor do mundo (DUARTE; SILVA; BAGATINI, 2020). Uma das medidas de contenção da contaminação do vírus é o distanciamento social implicando em redução de encontros presenciais (RODRIGUES et al., 2021). Durante a pandemia do coronavírus, 58% dos usuários de Internet no Brasil entre 16 e 64 anos afirmaram ter aumentado o tempo gasto em redes sociais (SOCIEDADE, 2020).

Considerando o elevado número de usuários das redes sociais e o excesso de tempo gasto pelos usuários nelas, questões importantes precisam ser respondidas sobre quais são os impactos positivos e negativos das redes sociais na vida das pessoas. Alguns trabalhos têm investigado se o uso da interação virtual é saudável para os usuários (RAUT; PATIL,

---

<sup>1</sup> <https://www.instagram.com/>

<sup>2</sup> <https://www.facebook.com/>

<sup>3</sup> <https://www.youtube.com/>

<sup>4</sup> <https://www.tiktok.com/>

<sup>5</sup> <https://www.linkedin.com/>

<sup>6</sup> <https://www.twitter.com/>

<sup>7</sup> <https://www.whatsapp.com/>

2016) e a influência das redes sociais na vida do ser humano (WOODS; SCOTT, 2016).

As mídias sociais possuem diversos impactos positivos para seus usuários. Na área de negócios, as redes sociais são um meio de divulgação acessível e atinge muitas pessoas trazendo mais clientes para o empreendimento (SIDDIQUI; SINGH et al., 2016). Na área da educação, as redes sociais deixam o conteúdo mais atrativo para os estudantes e eles se sentem mais motivados a estudarem (RAUT; PATIL, 2016). Além disso, as mídias auxiliam os estudantes a se comunicarem, trocarem ideias, criar e participar de comunidades da sua área de interesse (RAUT; PATIL, 2016). Na área da saúde, as mídias sociais auxiliam no compartilhamento de informações sobre prevenção, sintomas e tratamentos de doenças (AKRAM; KUMAR, 2017).

Por outro lado, as redes sociais também possuem vários impactos negativos. As redes sociais são consideradas mais viciantes do que cigarros e álcool (HEALTH, 2017). O trabalho de Sampasa-Kanyinga e Lewis (2015) mostrou que crianças e adolescentes de Ottawa no Canadá, que disseram estar conectados nas redes por mais de duas horas diárias, também relataram ter problemas psicológicos, como ideação suicida, ansiedade e depressão. O estudo de Woods e Scott (2016) revelou que o uso das redes sociais antes de dormir está relacionado com a má qualidade do sono, além de elevados níveis de ansiedade e depressão.

Segundo a pesquisa de Health (2017), a rede social com impacto mais positivo para os usuários na faixa etária de 14 a 24 anos é o *YouTube* e a com impacto mais negativo é o *Instagram*. O *YouTube* esteve nessa categoria porque tem efeitos positivos nos fatores de suporte emocional, depressão, ansiedade, solidão, auto expressão e identidade própria. Já o *Instagram* teve essa classificação, pois a rede tem alta influência negativa nos aspectos de ansiedade, depressão, solidão, qualidade de sono, *bullying*, vício e imagem corporal.

Em janeiro de 2022, o Brasil possuía 119 milhões de usuários no *Instagram* ficando atrás apenas dos Estados Unidos e Índia, onde cada um tem 140 milhões de usuários. (STATISTA, 2021). Os usuários em suas publicações nas redes sociais podem expressar seus sentimentos e suas emoções (WEISMAYER; GUNTER; ÖNDER, 2021). As emoções têm papéis importantes na vida das pessoas, porque elas influenciam nos relacionamentos interpessoais (PAYNE; COOPER, 2003). As emoções estão relacionadas com o temperamento, pois o temperamento proporciona maneiras de expressar as diferenças individuais nas emoções e como elas se desenvolvem (BATES; GOODNIGHT; FITE, 2008). O estudo das emoções e do temperamento é importante para detectar e analisar sintomas de problemas psicológicos, como depressão e ansiedade (BATES; GOODNIGHT; FITE, 2008).

Alguns trabalhos estudaram a relação entre redes sociais e o temperamento ou personalidade dos usuários (GOLBECK et al., 2011; LIMA; CASTRO, 2019; FERWERDA; SCHEDL; TKALCIC, 2016). O trabalho descrito em Lima e Castro (2019) propõe um modelo de aprendizado que permite prever o temperamento de um usuário a partir de



dados linguísticos e comportamentais do *Twitter*. O questionário de temperamento utilizado pelos autores foi o MBTI (*Myers-Briggs Type Indicator*) e o modelo de *Keirsey*. Já o trabalho descrito em (GOLBECK et al., 2011) propõe um modelo para prever a personalidade de um usuário com base em suas publicações do *Twitter*. O questionário de personalidade utilizado pelos autores foi o *The Big Five Personality Inventory*. Os autores Ferwerda, Schedl e Tkalcic (2016) tiveram o objetivo de descobrir a personalidade de um usuário por meio dos filtros utilizados em nas fotos das publicações no *Instagram*.

Considerando que o temperamento está melhor relacionado ao componente afetivo e mais diretamente ligado ao padrão de respostas emocionais (GOIS et al., 2012) do que a personalidade geral, entender a relação entre o temperamento de um usuário, os sentimentos e emoções presentes nas suas postagens é um objetivo relevante e que pode ser útil para diagnóstico de desordens psicológicas. Além disso, a rede social *Instagram*, é uma das mais usadas no Brasil (JIANG; NGIEN, 2020), mas ainda foi pouca explorada neste contexto. Ademais, os trabalhos realizados sobre temperamento e redes sociais se focaram na língua inglesa e, portanto, podem exigir uma adaptação ou o desenvolvimento de novas técnicas de pré-processamento e transformação de dados para serem utilizadas no idioma português, por exemplo, a remoção de acentos das frases. Por último, o teste de personalidade mais comumente usado na literatura é o MBTI. Apesar deste teste ser bastante popular, segundo (PLANK; HOVY, 2015) ele não representa uma estrutura adequada para compreensão do temperamento. Uma alternativa de questionário de temperamento é o TEMPS-RIO (WOODRUFF et al., 2011). Esse questionário é baseado no instrumento de avaliação TEMPS (AKISKAL et al., 2005) que foi validado em dezenas de países, sendo utilizado para investigar a relação do temperamento com o risco de diversas condições de saúde mental (DEMBIŃSKA-KRAJEWSKA; RYBAKOWSKI, 2014).

Diante das lacunas citadas, este trabalho visa estabelecer relações entre o temperamento humano, medido por meio do questionário de temperamento TEMPS-RIO, e o comportamento dos usuários na rede social *Instagram*.

## 1.1 Objetivos

O objetivo geral deste trabalho é desenvolver um método computacional que permita analisar as postagens e comportamentos dos usuários na rede social *Instagram* contrastando-os com seus temperamentos. Cada usuário é descrito pelos sentimentos presentes nas suas postagens e por outros dados obtidos a partir da rede social. Estes dados extraídos da rede social são contrastados com os diferentes temperamentos, a fim de estabelecer relações entre os mesmos. Os objetivos específicos deste trabalho são:

- Coletar informações a partir de uma amostra de estudantes dos cursos de Tecnologia da Informação e áreas afins da Universidade Federal de Uberlândia. As informações

coletadas são: respostas a um questionário de temperamento, postagens e informação do perfil da rede social *Instagram* destes usuários. O projeto para coleta foi autorizado pelo Comitê de Ética em Pesquisas com Seres Humanos da Universidade Federal de Uberlândia — CAAE 35533420.2.0000.5152;

- ❑ Criar uma base de dados anonimizada a partir das postagens e dos dados do perfil do usuário juntamente com o seu respectivo temperamento;
- ❑ Construir um framework que permita pré-processar e classificar o sentimento presente nas postagens de rede sociais;
- ❑ Criar um classificador em níveis para classificar o sentimento de postagens de rede sociais contrastando-o com o uso de um comitê (*ensemble*) de classificadores;
- ❑ Analisar a relação entre os dados extraídos do *Instagram* com o temperamento do usuário a fim de identificar padrões de comportamentos de um determinado temperamento na rede social.

## 1.2 Hipótese

Este trabalho possui a seguinte hipótese:

- ❑ O temperamento de um usuário tem relação com a forma que este se comporta na rede social. Assim, usuários com um mesmo temperamento possuem comportamentos similares na rede social *Instagram*.

A hipótese foi avaliada a partir de testes estatísticos, gráficos e árvore de decisão utilizando as variáveis coletadas da rede social e o temperamento do usuário.

## 1.3 Contribuições

A principal contribuição deste trabalho é a proposição de um método e um estudo experimental que contrasta temperamento com dados extraídos da rede social *Instagram* a partir de uma base de dados coletada de estudantes da Universidade Federal de Uberlândia. O trabalho também avança a área de Processamento de Linguagem Natural, contribuindo com desenvolvimento de uma estratégia de classificação em níveis para textos em português, que usa aprendizado supervisionado e não-supervisionado e contrasta tal estratégia com algoritmos que usam apenas aprendizado supervisionado.

## 1.4 Organização da Dissertação

Este trabalho está organizado da maneira a seguir. O Capítulo 2 apresenta os principais conceitos que serão utilizados no decorrer do trabalho. O Capítulo 3 discorre os trabalhos do estado da arte relacionados a este trabalho. O Capítulo 4 aborda o desenvolvimento do trabalho. O Capítulo 5 mostra os experimentos e a análise dos resultados obtidos. Por fim, o Capítulo 6 tece as principais conclusões deste presente trabalho e os principais trabalhos futuros.



---

## Fundamentação Teórica

Neste capítulo são apresentados os conceitos fundamentais usados neste trabalho. Estes conceitos foram divididos em três seções. A Seção 2.1 discorre os temperamentos humanos e a relação deles com as redes sociais. A Seção 2.2 aborda o conceito e as etapas de mineração de texto em redes sociais. Por fim, a Seção 2.3 apresenta as considerações finais do capítulo.

### 2.1 Temperamento Humano e Redes Sociais

Esta seção tem o objetivo de apresentar os conceitos fundamentais sobre temperamento humano e a influência das redes sociais na vida dos usuários.

#### 2.1.1 Sentimentos, Emoções e Temperamento

Os seres humanos expressam sentimentos e emoções. Na área de psicologia, o sentimento é definido pela combinação de sensações, gestos, culturas e aprendizados ao longo da vida (STETS, 2006). E as emoções estão conectadas por interpretações situacionais, mudanças fisiológicas, gestos expressivos e por componentes que incluem experiências subjetivas, reações expressivas, padrões de respostas e reações de enfrentamento (GORDON, 2017). Kemper (1987) identificou que as quatro emoções primárias são raiva, medo, depressão e satisfação e elas possuem relação com a evolução de sobrevivência do ser humano. As emoções e o temperamento do ser humano estão diretamente relacionados um com o outro (MACNEILL; PÉREZ-EDGAR, 2019).

Segundo Putnam (1992), as escolhas de uma pessoa estão diretamente relacionadas com seu temperamento. O temperamento tem sido definido como um conjunto de diferenças individuais na reatividade e autorregulação, afetadas pela hereditariedade, maturidade e experiência (ROTHBART; AHADI; EVANS, 2000). Os temperamentos são identificados como ocorrência subclínica ou fenótipos de transtornos de humor e eles representam uma parte de doenças afetivas (VÁZQUEZ et al., 2012). Segundo Rihmer et al. (2010) os

cinco temperamentos são: ciclotímico, irritável, hipertímico, depressivo e ansioso. Esses temperamentos são definidos a seguir:

- ❑ **Ciclotímico:** segundo Brieger e Marneros (1997), o autor Kretschmer (1921) declarou esse temperamento como mudanças de humor frequentes. As pessoas que possuem esse temperamento têm tendência a possuir traços de transtorno bipolar (AKISKAL et al., 1977). Esse temperamento está associado a pessoas que possuem baixo nível de escolaridade, emprego instável, problemas financeiros, vícios em álcool e drogas (WOODRUFF et al., 2011).
- ❑ **Irritável:** esse temperamento se caracteriza com um limiar de sintomas depressivos (AKISKAL et al., 1998).
- ❑ **Hipertímico:** segundo Akiskal e Akiskal (2005) as pessoas que possuem esse temperamento são exageradas, otimistas, extrovertidas, confiantes e possuem muita energia. Essas características são importantes para liderança e essas pessoas defendem seu território dentro e fora do seu grupo social.
- ❑ **Depressivo:** os indivíduos com esse temperamento tem sensibilidade ao sofrimento e possuem maior tendência a desenvolverem depressão. Normalmente, essas pessoas tendem a deixar suas vontades de lado e se dedicam a família e instituições. Além disso, elas se sentem mais confortáveis quanto a harmonia entre as pessoas (AKISKAL; AKISKAL, 2005). Segundo Angst et al. (2004) o estado civil casado é altamente correlacionado com a depressão.
- ❑ **Ansioso:** os ansiosos são filantrópicos, mas também são obsessivos e sempre estão alerta a qualquer perigo. Para os ansiosos, estar relaxado é contra a sua própria natureza e o relaxamento contribui com o aumento da ansiedade (AKISKAL, 1998).

Uma das formas de identificar o temperamento humano é via questionário de temperamento (CORR; MATTHEWS, 2020). Esses questionários possuem perguntas/afirmações onde o usuário deve escolher uma opção de resposta e no final o seu temperamento é descoberto de acordo com suas escolhas (CORR; MATTHEWS, 2020).

### 2.1.2 Questionário de Temperamento

O TEMPS é um instrumento de avaliação do temperamento desenvolvido por Akiskal (AKISKAL et al., 2005) e validado em dezenas de países. Este instrumento é largamente utilizado para investigar a relação do temperamento com a risco de diversas condições de saúde mentais (DEMBIŃSKA-KRAJEWSKA; RYBAKOWSKI, 2014). Estudos utilizando o TEMPS demonstraram, por exemplo, a relação entre o temperamento depressivo e o risco de depressão e diabetes (GOIS et al., 2012), risco de bipolaridade e suicídio

em crianças e adolescentes com temperamento ciclotímico (KOCHMAN et al., 2005), e o risco de suicídio em pacientes deprimidos com temperamentos depressivo e ciclotímico (YIN et al., 2022).

O auto-questionário *Temperament Evaluation of Memphis, Pisa and San Diego* (TEMPS-A) é uma medida de autorrelato projetada para avaliar o temperamento em indivíduos saudáveis e pacientes psiquiátricos por meio de subescalas (AKISKAL; AKISKAL, 2005).

TEMPS-RIO (WOODRUFF et al., 2011) é uma versão resumida da tradução do TEMPS-A para o português. O questionário TEMPS-RIO foi escolhido para identificar o temperamento dos usuários neste trabalho, porque ele faz parte do TEMPS e esse instrumento de avaliação de temperamento foi largamente validado em dezenas de países.

O TEMPS-RIO (detalhado no Apêndice A) possui 45 questões de verdadeiro ou falso que são relacionadas aos cinco temperamentos: ciclotímico, irritável, hipertímico, depressivo e ansioso, mais o temperamento preocupado. O temperamento preocupado é a combinação entre os temperamentos ansioso e depressivo (KARAM et al., 2005). As questões do questionário são utilizadas para identificar o temperamento da pessoa que respondeu o questionário. Cada temperamento possui uma pontuação que é o somatório de questões multiplicado por um peso. O temperamento aplicado ao usuário é o maior valor da pontuação. A pontuação dos temperamentos depressivo (D), ciclotímico (C), irritável (I), hipertímico (H), ansioso (A) e preocupado (P) são calculados pelas respectivas Equações 1, 2, 3, 4, 5 e 6

$$D = (Q1 + Q2 + Q3 + Q4 + Q5 + Q6 + Q7 + Q8) \cdot 12.5 \quad (1)$$

$$C = (Q9 + Q10 + Q11 + Q12 + Q13 + Q14 + Q15 + Q16) \cdot 12.5 \quad (2)$$

$$I = (Q17 + Q18 + Q19 + Q20 + Q21 + Q22 + Q23 + Q24) \cdot 12.5 \quad (3)$$

$$H = (Q25 + Q26 + Q27 + Q28 + Q29 + Q30 + Q31 + Q32) \cdot 12.5 \quad (4)$$

$$A = (Q33 + Q34 + Q35 + Q36 + Q37 + Q38 + Q39 + Q40) \cdot 12.5 \quad (5)$$

$$P = (Q41 + Q42 + Q43 + Q44 + Q45) \cdot 20 \quad (6)$$

onde  $QN$  é o número da questão do questionário de temperamento TEMPS-RIO.

O temperamento final do usuário é calculado pela Equação 7

$$\text{temperamento} = \text{MAIOR}(D, C, I, H, A, P) \quad (7)$$

### 2.1.3 Influências das Redes Sociais nos Seres Humanos

A Internet mudou consideravelmente a forma de interação entre indivíduos (GRAHAM; DUTTON, 2019). Na Internet existem diversas redes sociais que permitem que as pessoas se apresentem de uma nova forma, podendo transformar suas identidades, pois numa rede social a personalidade do usuário é definida pelo seu perfil na rede (CHAMBERS, 2013).

As redes sociais permitem a comunicação entre um usuário e outro de forma escrita e com isso uma nova forma de comunicação foi desenvolvida, a comunicação por meio de *emojis*. Os *emojis* podem ser utilizados para representar um sentimento de uma determinada sentença e eles são muito usados nas redes sociais (SHIHA; AYVAZ, 2017).

Segundo Critcher (2008), o exagero no uso de redes sociais pode causar ansiedade e pânico aos usuários. Na pesquisa feita por Sampasa-Kanyinga e Lewis (2015) foi identificado que o uso de mais de duas horas diárias de redes sociais por crianças e adolescentes podem causar ansiedade e depressão. Entretanto, as redes sociais também podem cultivar relacionamentos pessoais (CRITCHER, 2008). A pesquisa de Health (2017) informou que a rede social *Youtube* ofereceu suporte emocional para seus usuários que possuem depressão e ansiedade. Dado os impactos das redes sociais na vida do ser humano, analisar a relação das redes sociais com o temperamento humano se torna uma tarefa muito importante.

## 2.2 Mineração de Texto em Redes Sociais

Mineração de texto descreve um conjunto de técnicas linguísticas, estatísticas e de aprendizado de máquina que modelam e estruturam o conteúdo das fontes de dados para permitir análises, buscas ou pesquisas (GOLDSCHMIDT; PASSOS; BEZERRA, 2015). Existe um enorme volume de textos de diferentes tipos, como textos publicados em redes sociais, jornais e *blogs*, e cada um tem suas características (AGGARWAL; ZHAI, 2012).

Uma das áreas de mineração de texto é a análise de sentimentos de um texto. A análise de sentimentos tem o objetivo de identificar o sentimento de um determinado texto (LIU, 2012). Na computação, em geral, os sentimentos usados para classificar um texto são positivo, negativo e neutro. Este trabalho irá focar na mineração de texto e análise de sentimentos em redes sociais e irá usar o conceito de sentimento descrito neste parágrafo.

A mineração de texto de redes sociais é utilizada para representar, analisar e extrair dados das redes sociais. A análise dos dados das redes sociais possui alguns desafios (ZAFARANI; ABBASI; LIU, 2014). Um deles é o enorme volume de dados que existem nas redes sociais. Outro é a extração do dado, já que as redes sociais disponibilizam suas próprias APIs, mas normalmente elas são limitadas, não permitindo a coleta de muitos dados. Por fim, os textos não usam a norma culta de escrita, ou seja, há muito uso de gírias e léxicas incorretas.



A análise de dados de redes sociais tem recebido recentemente grande atenção (HOSSEINMARDI et al., 2015; ZHAN; TU; YU, 2018; KANSAON; BRANDÃO; PINTO, 2018). Alguns exemplos incluem os trabalhos de (NAF'AN et al., 2019; HOSSEINMARDI et al., 2015), que exploraram a rede social *Instagram* para identificar *cyberbullying* nas publicações. Ambos utilizaram os comentários das postagens dos usuários para a análise. Os estudos de (KANSAON; BRANDÃO; PINTO, 2018; AGUIAR et al., 2018) utilizaram a rede social *Twitter* e propuseram métodos para análise de sentimentos.

A mineração de texto possui as etapas de coleta de dados, pré-processamento, normalização, definição de características, aprendizado de máquina, métodos e medidas de avaliação (PUJARI, 2001). Além dessas etapas também é necessário realizar a análise dos dados obtidos. Essas etapas estão detalhadas nas próximas subseções.

### 2.2.1 Coleta de Dados

Há duas maneiras de coletar os dados de redes sociais. Uma é utilizando as APIs que são disponibilizadas pelas redes e a outra é por meio de *scraping* (ZAFARANI; ABBASI; LIU, 2014).

Redes sociais, como *Facebook* (FACEBOOK, 2021), *Instagram* (INSTAGRAM, 2021) e *Twitter* (TWITTER, 2021) disponibilizam APIs que permitem buscar as publicações de um determinado usuário ou um determinado termo de busca. Cada rede social tem uma documentação específica para utilização das APIs (DEVELOPERS, 2021a; DEVELOPERS, 2021b; DEVELOPERS, 2021c). Além disso, cada rede social tem as suas políticas de privacidade, que precisam ser analisadas antes de se realizar a coleta.

A API do *Instagram*, Graph API<sup>1</sup>, disponibiliza serviços REST que fazem o gerenciamento de publicações e comentários do usuário, identificação de publicações nas quais o usuário foi mencionado, busca de publicações por *hashtags* e para obtenção métricas (alcance de publicações, quantidade de seguidores por período, média de curtidas, entre outros) da conta. Segundo a documentação, a API é indicada para contas empresariais ou para criadores de conteúdo no *Instagram* e ela não pode ser utilizada por outros tipos de conta. Para os outros tipos de conta do *Instagram* existe a API de exibição básica do *Instagram*<sup>2</sup> que tem seus recursos mais limitados sendo a maioria dos serviços REST apenas para consulta de dados.

*Scraping* é uma técnica de extrair informações de um website automaticamente (VARGIU; URRU, 2013). Normalmente, algumas ferramentas são utilizadas para auxiliar esse processo de automação, por exemplo o *Selenium WebDriver* (WEBDRIVER, 2021).

<sup>1</sup> <https://developers.facebook.com/docs/instagram-api/>

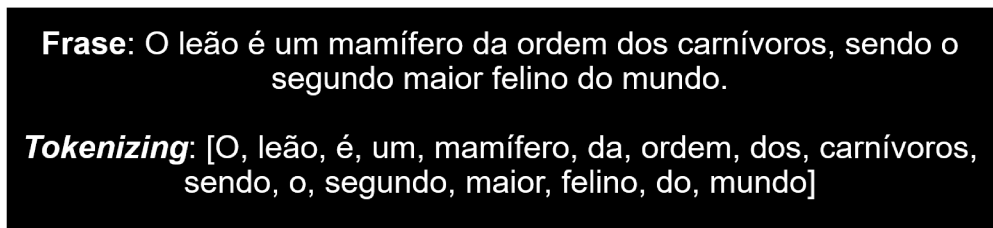
<sup>2</sup> <https://developers.facebook.com/docs/instagram-basic-display-api>

## 2.2.2 Pré-processamento

Em mineração de texto, o pré-processamento de dados é utilizado para transformar os dados brutos para um formato mais compreensível pela máquina (NAYAK et al., 2016). Além disso, os passos de limpeza e padronização do texto também são feitos no pré-processamento (NAYAK et al., 2016). A seguir, os passos de pré-processamento mais comuns em mineração de texto serão detalhados.

### 2.2.2.1 *Tokenizing*

O processo de *tokenizing* consiste em dividir uma frase composta por várias palavras em palavras individuais (VIJAYARANI; JANANI et al., 2016). Na Figura 1 é mostrado um exemplo do processo de *tokenizing*. Uma das ferramentas para separar palavras está disponível na biblioteca NLTK (PROJECT, 2021) do Python por meio da função `word_tokenize`.



**Frase:** O leão é um mamífero da ordem dos carnívoros, sendo o segundo maior felino do mundo.

**Tokenizing:** [O, leão, é, um, mamífero, da, ordem, dos, carnívoros, sendo, o, segundo, maior, felino, do, mundo]

Figura 1 – Exemplo de Frase Para Análise de Sentimento - *Tokenização*

.Fonte: elaborado pela autora (2022)

### 2.2.2.2 Remoção de *Stopwords*

As *stopwords* são palavras irrelevantes que não agregam valor no texto (GERLACH; SHI; AMARAL, 2019). Normalmente as *stopwords* se repetem muito em um texto, não trazendo muita relevância para o contexto principal. Artigos e preposições são alguns tipos comuns de palavras irrelevantes no contexto de mineração de texto.

Na Figura 2 é apresentado um exemplo de frase e a remoção de suas *stopwords*. As palavras destacadas em vermelho são irrelevantes para o contexto geral da frase, logo elas podem ser removidas. Na mineração de texto e análise de sentimentos, as *stopwords* são removidas na etapa de pré-processamento.

### 2.2.2.3 Padronização do texto

Para uma melhor análise do texto, quanto mais padronizado ele estiver, melhor. Os textos de redes sociais costumam ter muitos erros, léxicos e gírias, então a padronização do texto nesse caso é crucial. Algumas técnicas de padronização de texto para as redes

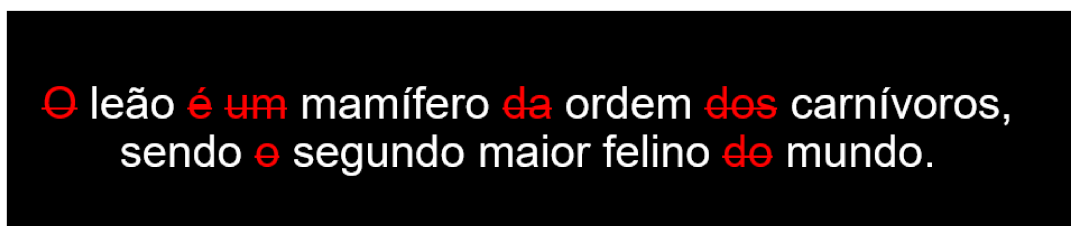


Figura 2 – Exemplo de Frase Para Análise de Sentimento — Remoção de *StopWords*.  
Fonte: elaborado pela autora (2022)

sociais são (NAF'AN et al., 2019; KANSAON; BRANDÃO; PINTO, 2018; AGUIAR et al., 2018):

- ❑ Passar todas as letras do texto para minúsculo;
- ❑ Eliminar todas as pontuações, URLs, e-mails e citações dos usuários;
- ❑ Substituir *emoticons* por palavras;
- ❑ Substituir gírias;
- ❑ Corrigir palavras escritas incorretamente.

### 2.2.3 Normalização de Texto

Em mineração de texto, técnicas de *stemming* e lematização são utilizadas para a normalização de texto. Uma palavra escrita no masculino, feminino, presente, passado ou futuro possuem o mesmo significado, então essas duas técnicas fazem reduções de palavras para facilitar no processo de aprendizagem da máquina (NICOLAI; KONDRAK, 2016).

O processo de *stemming* consiste em reduzir a palavra para o seu radical (KORENIUS et al., 2004). As palavras 'meninos', 'meninas' e 'menininhos', por exemplo, serão reduzidas para 'menin'.

A lematização reduz a palavra ao seu lema, ou seja, reduz uma palavra para o verbo infinitivo masculino (KORENIUS et al., 2004). As palavras 'tinha', 'tenho', 'terei' serão reduzidas para o verbo 'ter'.

### 2.2.4 Definição de Características

A tarefa de definição de características consiste em extrair características de um texto de forma que seja possível a separação dos dados para facilitar a classificação (BENEVENUTO; RIBEIRO; ARAÚJO, 2015). As técnicas de *N-gramas*, *Bag-of-words* e *TF-IDF* são muito utilizadas para definir características em mineração de texto.

### 2.2.4.1 Bag-of-Words

O *Bag-of-Words* (BOW) é a transformação do texto para números. O BOW é a apresentação do texto de forma vetorizada criando um vocabulário de palavras que conta a frequência de ocorrência separadamente cada palavra do documento (YAN et al., 2020). Na Figura 3 é mostrado um exemplo de (BOW). A primeira linha da tabela apresentada na Figura 3 apresenta todas as palavras que estão presentes em todas as frases da esquerda da Figura. Embaixo de cada palavra possui números que representam a quantidade de vezes que aquela palavra apareceu na frase.

	Bag-of-Words									
	o	leão	vive	na	selva	tigre	existem	muitos	e	animais
O leão vive na selva	1	1	1	1	1	0	0	0	0	0
O tigre vive na selva	1	0	1	1	1	1	0	0	0	0
Na selva existem muitos e muitos animais	0	0	0	1	1	0	1	2	1	1

Figura 3 – *Exemplo de Bag-of-Words*. Fonte: elaborado pela autora (2022)

### 2.2.4.2 N-gramas

O N-gramas é a união de  $N$  palavras. Um unigrama é apenas uma palavra, por exemplo a palavra 'muito'. O bigrama é a união de duas palavras, por exemplo, 'muito pouco'. O trigramas a união de três palavras e assim sucessivamente. (ROSA, 2015).

O unigrama da frase 'Eu gosto muito pouco de filme' é uma lista das palavras: ['eu', 'gosto', 'muito', 'pouco', 'de', 'filme']. O bigrama dessa mesma frase é a lista de tuplas: [('Eu', 'gosto'), ('gosto', 'muito'), ('muito', 'pouco'), ('pouco', 'de'), ('de', 'filme')]. Sendo assim, o N-gramas é formado por  $N$  palavras consecutivas e quanto maior o valor de  $N$  menor o tamanho da lista formada.

Podemos observar que a interpretação da palavra 'muito' pode mudar conforme as palavras que a acompanham. No caso, o unigrama com o termo 'muito' o significado é de intensidade de alto grau. Já no caso do bigrama 'muito pouco' o sentido muda para intensidade de baixo grau.

### 2.2.4.3 TF-IDF

TF-IDF é um método muito utilizado no Processamento de Linguagem Natural (PLN), que determina o quão relevante é uma palavra em um determinado texto (TRSTENJAK; MIKAC; DONKO, 2014). Esse método utiliza o TF — *Term Frequency*, que determina

a frequência de uma palavra em um texto e o IDF — *Inverse Document Frequency*, que determina o inverso da frequência de uma palavra no texto. No TF-IDF o texto é chamado de documento e a palavra é chamada de termo.

O peso de um termo em um documento pode ser calculado pela Equação 8

$$a_{ij} = tf_{ij}idf_{ij} = tf_{ij} \times \log_2\left(\frac{N}{df_i}\right) \quad (8)$$

Onde,  $a_{ij}$  é o peso do termo  $i$  no documento  $j$ ,  $N$  é o tamanho da coleção de documentos,  $tf_{ij}$  é a frequência do termo  $i$  no documento  $j$  e  $df_i$  é o número de documentos onde  $i$  está presente.

O TF e o IDF são utilizados juntos para manter o equilíbrio dos pesos das palavras. Por exemplo, na frase “Gosto de andar de bicicleta” o maior TF é o da palavra “de”, porém ela não tem nenhuma relevância nessa frase. Então, o IDF é incorporado e diminui o peso das palavras que mais se repetem e aumenta o peso das palavras que quase não aparecem no texto (CASAROTTO, 2019). Então nessa frase as palavras “gosto”, “andar” e “bicicleta” apesar de aparecer menos vezes são mais relevantes.

### 2.2.5 Técnicas de Aprendizado de Máquina

Em aprendizado de máquina existem duas formas de aprendizado para mineração de texto, o aprendizado supervisionado e não-supervisionado (BENEVENUTO; RIBEIRO; ARAÚJO, 2015).

O aprendizado não-supervisionado acontece quando não há instâncias rotuladas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Um dos exemplos de aprendizado não supervisionado é prever o sentimento de um texto. Para isso é utilizado um dicionário de palavras, onde cada palavra possui uma polaridade de sentimento, isto é, positivo, negativo ou neutro. Dado uma frase de entrada e o dicionário de palavras, o algoritmo não-supervisionado consegue prever a polaridade da frase de entrada (BENEVENUTO; RIBEIRO; ARAÚJO, 2015). A Figura 4 mostra um exemplo de análise de sentimentos de um texto utilizando a técnica de aprendizado não supervisionado. O Léxico de Sentimentos é o dicionário de palavras com seus respectivos sentimentos. O sinal (+) indica que a palavra possui sentimento positivo e o sinal (-) indica que a palavra possui sentimento negativo. LIWC (PENNEBAKER; FRANCIS; BOOTH, 2001) e *Opinion Lexicon* (SOUZA; VIEIRA, 2011) são dois dos dicionários mais utilizados na literatura para essa tarefa.

Já no aprendizado supervisionado ou classificação um conjunto de instâncias rotulada está disponível (TAN et al., 2013). Os algoritmos de classificação são treinados com uma parte dos dados rotulados e posteriormente conseguem classificar os dados que não estão rotulados. Cada algoritmo de classificação tem sua estratégia e nesta subseção os

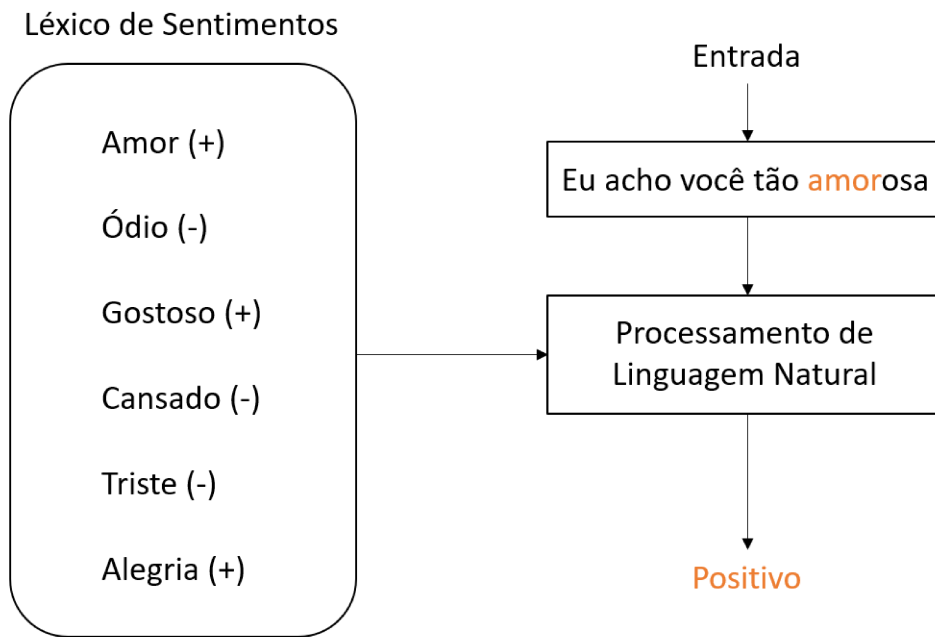


Figura 4 – Exemplo de Aprendizado-Não Supervisionado. Fonte: Adaptado de Benevenuto, Ribeiro e Araújo (2015)

algoritmos de classificação *Naive Bayes*, *Árvore de Decisão*, *Random Forest* e *Support Vector Machine* (SVM) são detalhados.

### 2.2.5.1 *Naive Bayes*

O classificador probabilístico *Naive Bayes* é baseado no teorema de *Bayes* (EFRON, 2013) e estima a probabilidade condicional de um conjunto de atributos pertencer a uma classe, assumindo que os atributos são independentes (TAN et al., 2013). Para os cálculos de probabilidade, o algoritmo utiliza a Equação 9

$$P(x|Y) = \frac{P(Y|x) \cdot P(x)}{P(Y)} \quad (9)$$

onde  $P(x|Y)$  é a probabilidade de ocorrência da classe  $x$  dado o evento  $Y$ ,  $P(Y|x)$  a probabilidade de ocorrência de  $Y$  dado a classe  $x$  sendo calculado pela Equação 11,  $P(x)$  a probabilidade de ocorrência da classe  $x$  no conjunto de dados de treinamento calculado pela Equação 10 e  $P(Y)$  a probabilidade do evento  $Y$  no conjunto de treinamento. A probabilidade  $P(Y)$  pode ser desconsiderada nesse cálculo, pois ela é igual para todas as classes (MITCHELL, 1997).

$$P(x) = \frac{N_i}{N} \quad (10)$$

onde,  $N$  é o número total de instâncias e  $N_i$  o número de total de instâncias associadas a classe  $x$ .

$$P(Y|x) = \prod_{i=0}^n P(t_i|x) \quad (11)$$

onde,  $P(t_i|x)$  representa a probabilidade de cada termo sendo calculado pela Equação 12,  $t$  é um atributo pertencente ao evento  $Y$ ,  $n$  é o total de atributos que estão no evento  $Y$  e  $x$  a classe em questão.

$$P(t|x) = \frac{T(t_i|x)}{\sum_{i=0}^n T(t_i|x)} \quad (12)$$

onde,  $T(t_i|x)$  representa a frequência que o atributo  $t_i$  é encontrado na classe  $x$  e  $n$  o total de atributos do evento. O somatório de um na equação é feito para evitar que a probabilidade do termo sempre difira de zero (PIVETTA, 2013).

O classificador constrói uma tabela de probabilidades de  $P(x|Y)$  e escolhe a classe que possui maior chance de ocorrência dado o evento e atribui a classe a esse evento (TAN et al., 2013). A Equação 13 se refere ao cálculo de escolha da classe para um determinado evento

$$\max_{x \in X} (P(x|Y)) \quad (13)$$

### 2.2.5.2 *Árvore de Decisão*

As árvores de decisão tomam uma decisão através de um caminho (CAMPOS, 2021). As árvores são compostas por ramos que possuem nós. Os ramos são rotulados com valores ou com intervalos. O nó folha é o último nó do ramo sendo rotulado com a classe. O nó raiz é o primeiro nó da árvore e os demais são os nós internos que se dividem conforme os valores dos atributos (LAURETTO, 2010). A Figura 5 ilustra um exemplo de árvore de decisão que decidi se o indivíduo vai à praia ou não. O nó raiz é o tempo, se estiver chovendo o indivíduo decide que não vai à praia. Se estiver com sol a condição de vento é verificada. Se estiver ventando a escolha é de não ir à praia, e caso contrário a escolha é de ir à praia.

### 2.2.5.3 *Random Forest*

É uma técnica que usa um comitê de classificadores, sendo a combinação de árvores de decisão que são independentes entre si (YIU, 2021). Segundo Breiman (2001) os passos para a construção da árvore de decisão são três. O primeiro passo desse algoritmo é a criação do *bootstrap dataset*, isto é, uma base de dados aleatória gerada a partir dos dados originais. Essa base de dados possui um conjunto de instâncias. O segundo passo é a seleção de subconjuntos de atributos que melhor caracterizam os dados para dar início a criação da árvore de decisão. O terceiro passo é a seleção dos demais atributos de forma aleatória para serem utilizados na construção da árvore. A construção das demais árvores

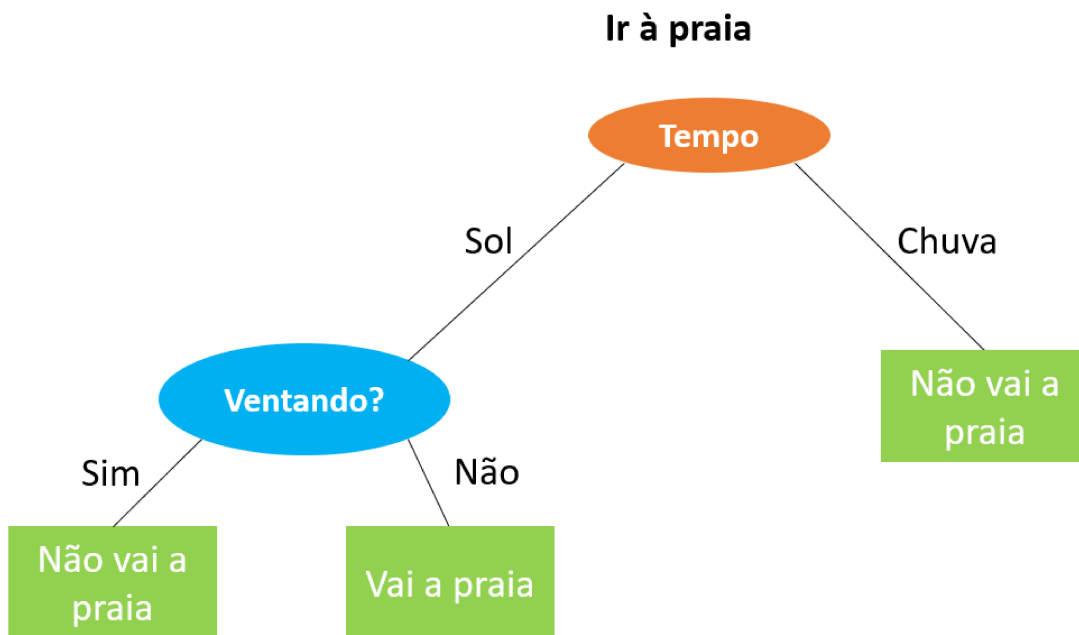


Figura 5 – Árvore de Decisão Sobre Ir à Praia. Fonte: elaborado pela autora (2022)

de decisão que compõe o *Random Forest* é feita considerando o segundo e o terceiro passo. Dessa maneira, cada árvore de decisão é construída de forma aleatória. Com todas as árvores de decisão construídas, cada árvore toma a sua decisão e a decisão final pode ser feita utilizando a média ou a moda das respostas das árvores independentes, ver Figura 6.

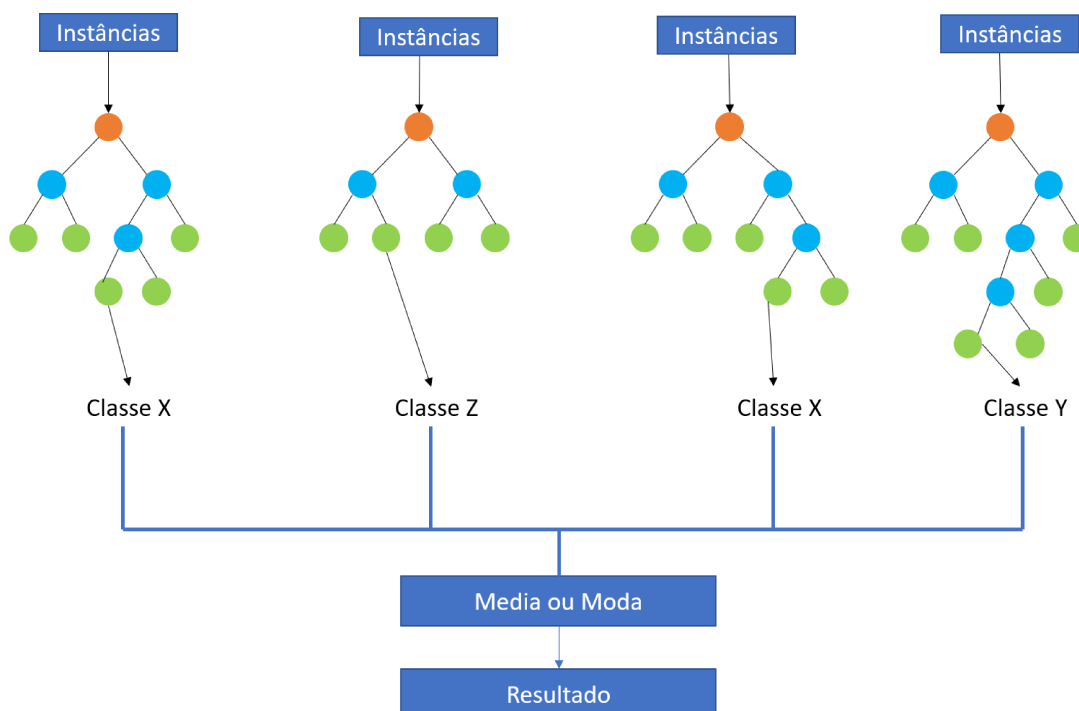


Figura 6 – Representação de uma *Random Forest*. Fonte: elaborado pela autora (2022)



#### 2.2.5.4 Support Vector Machine (SVM)

É um algoritmo de aprendizagem supervisionada que pode ser utilizado em problemas de regressão e classificação. O algoritmo classifica o conjunto de dados tentando separar linearmente as classes que estão presentes naquele conjunto. Para isso, o SVM encontra um hiperplano que maximiza a distância dos dados mais próximos das classes (BENEVENUTO; RIBEIRO; ARAÚJO, 2015). O hiperplano é definido pela Equação 14

$$ax + b = 0 \quad (14)$$

onde,  $x$  é a instância a ser classificada,  $a$  é um vetor e  $b$  uma constante.  $a$  e  $b$  serão definidos a partir dos dados de treino das instâncias linearmente separáveis.

Na Figura 7 é ilustrado o hiperplano que o algoritmo traça entre as classes. A reta linear é conhecida como hiperplano de separação ótima, retas tracejadas são conhecidas como vetores de suporte, e o espaço entre os vetores de suporte e o hiperplano com separação ótima é a margem.

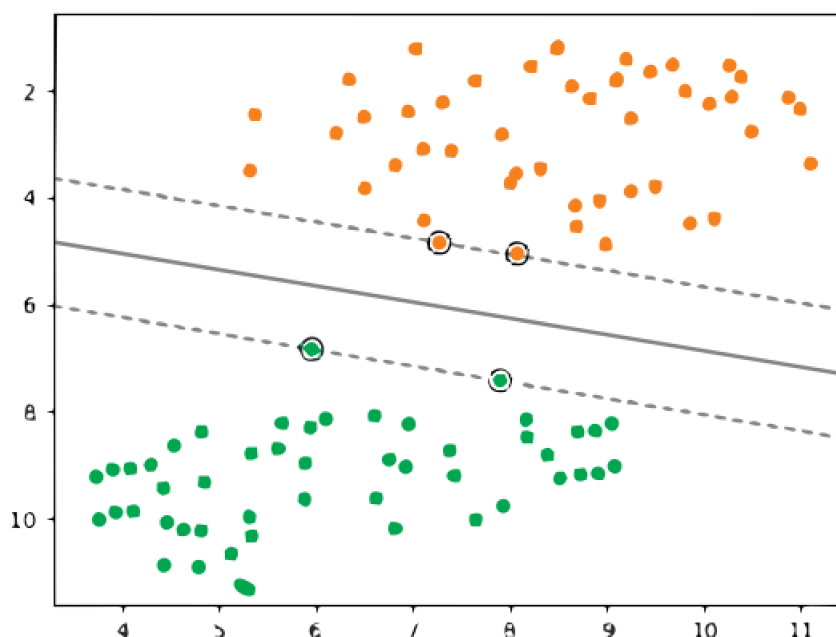


Figura 7 – Linear SVM. Fonte: Adaptado de (LEARN, 2021)

O *Support Vector Machine* também pode ser utilizado para um conjunto de dados não linear. Para essa tarefa, as funções de *Kernel* podem ser utilizadas. Essas funções aumentam a dimensionalidade e permitem maior variedade de espaços no plano (BURBIDGE; BUXTON, 2001). Um exemplo do comportamento do SVM para um conjunto de dados não linear é apresentado na Figura 8.

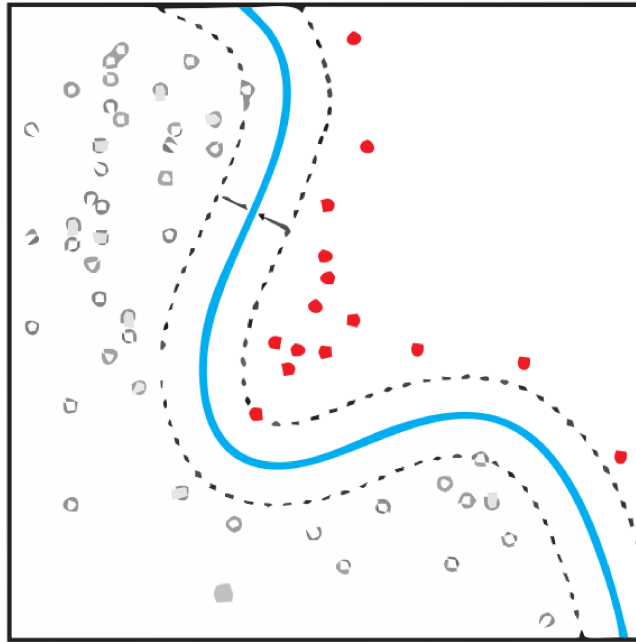


Figura 8 – SVM Kernel. Fonte: Adaptado de (WIKIPEDIA, 2021)

## 2.2.6 Métodos e Medidas de Avaliação

Os métodos de avaliação são usados para definir a melhor forma de distribuir a base de dados entre treino e teste para avaliação do classificador (TAN et al., 2013). As medidas de avaliação são aplicadas para identificar a qualidade do classificador para uma determinada tarefa (TAN et al., 2013).

### 2.2.6.1 *Hold-out*

É um método de avaliação que divide a base de dados em duas partições, treino e teste (BISHOP; NASRABADI, 2006) (ver Figura 9). Segundo Tan et al. (2013) a proporção da divisão dos dados de teste e treino fica a critério dos analistas.

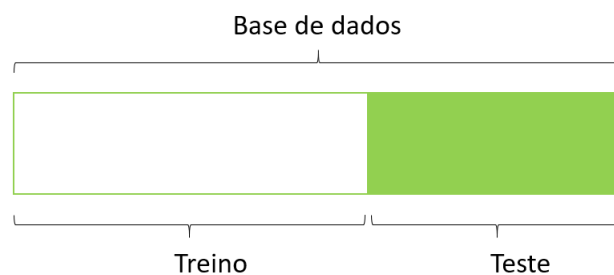


Figura 9 – Método de Avaliação *Hold-out*. Fonte: elaborado pela autora (2022)

### 2.2.6.2 *K-Fold Cross-Validation*

É um método de avaliação que divide a base em  $K$  blocos de tamanhos iguais de treino e teste iterando-os  $K$  vezes cruzadamente, sendo um bloco para teste e os demais para treino (ZAKI; JR; MEIRA, 2014) (ver Figura 10). Normalmente na literatura, o valor de  $K$  para esse método de avaliação é igual a 10. O valor de  $K$  pode ser igual à quantidade total de dados que a base de dados possui, esse evento é conhecido por *leave-one-out*, isto é, quando cada partição de treino e teste possui apenas um exemplo (ZAKI; JR; MEIRA, 2014).

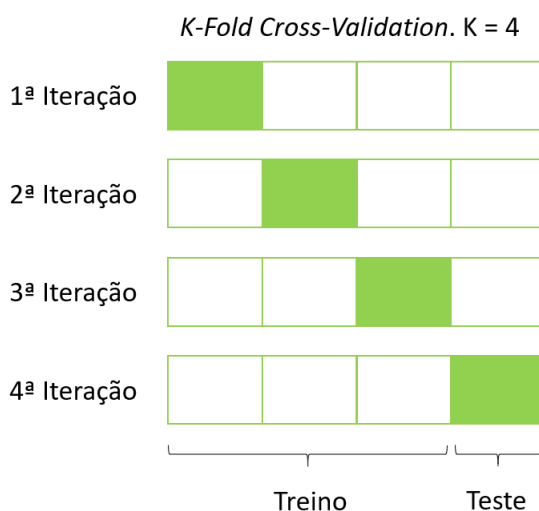


Figura 10 – Método de Avaliação *K-Fold Cross-Validation*. Fonte: elaborado pela autora (2022)

### 2.2.6.3 Matriz de Confusão

A matriz de confusão é representada por uma tabela que resume os resultados do modelo de classificação (BURKOV, 2019). Essa tabela é composta por um eixo que possui as classes preditas e outro eixo com os rótulos reais. A Figura 11 mostra a matriz de confusão de um modelo que possui as classes positivo, negativo e neutro. Na primeira tabela as colunas destacadas em verde representam as classes classificadas corretamente e as demais colunas as classes classificadas incorretamente, podendo elas serem falso negativo ou falso positivo. Nas demais tabelas são apresentados os valores de  $VN$  verdadeiros negativos,  $VP$  verdadeiros positivos,  $FP$  falsos positivos e  $FN$  falsos negativos.

Os valores  $VN$  são valores das classes que não estão sendo analisadas e que foram previstas corretamente. Os valores  $VP$  são os valores previstos corretamente da classe analisada.  $FP$  são os valores previstos incorretamente da classe analisada.  $FN$  são os valores das classes que não estão sendo analisadas e que foram previstos incorretamente.

		Classe Prevista			Falso Negativo
		Positivo	Negativo	Neutro	
Classe Atual	Positivo	6	4	0	4
	Negativo	3	4	2	5
	Neutro	0	1	11	1
Falso Positivo		3	5	2	10

Positivo		Negativo		Neutro	
6 (VP)	4 (FN)	4 (VP)	5 (FN)	11 (VP)	1 (FN)
3 (FP)	15 (VN)	5 (FP)	17 (VN)	2 (FP)	10 (VN)

Figura 11 – Matriz de Confusão. Fonte: Adaptado de (LEAL, 2017)

#### 2.2.6.4 Acurácia

É uma medida de avaliação que indica o desempenho geral do modelo (BURKOV, 2019). A validação do modelo apenas pelo uso da acurácia pode não ser suficiente para tarefas de classificação desbalanceadas, por exemplo, um modelo de classificação de e-mails como *spam* e não *spam*. Como existem mais e-mails classificados como não *spam*, se o classificador classificar todos os e-mails como não *spam* a acurácia dele será alta, mesmo ele tendo errado todos os e-mails não *spam*. A acurácia para bases de dados com duas classes é calculada utilizando a Equação 15

$$ACC = \frac{VP + VN}{VP + FN + FP + VN} \quad (15)$$

onde,  $ACC$  é o resultado da acurácia,  $VN$  verdadeiros negativos,  $VP$  verdadeiros positivos,  $FP$  falsos positivos e  $FN$  falsos negativos.

Caso a base de dados analisada no modelo tenha mais que duas classes, a acurácia macro pode ser utilizada para analisar o modelo (SOKOLOVA; LAPALME, 2009). A acurácia macro é calculada conforme a equação 16

$$ACCM = \frac{\sum_{i=1}^n \frac{VP_i + VN_i}{VP_i + FN_i + FP_i + VN_i}}{n} \quad (16)$$

onde  $ACCM$  é o resultado da acurácia macro calculada pela média do somatório da acurácia de todas as classes presentes na base de dados e  $n$  é o número de classes.

#### 2.2.6.5 Precisão

É uma medida de avaliação que analisa o quão preciso é o modelo de classificação (BURKOV, 2019). Ela pode ser utilizada em situações onde os valores falsos positivos

são considerados mais graves do que os falsos negativos, por exemplo, classificação de vídeos adequados para crianças.

O cálculo da precisão para classificação binária se dá pela quantidade de valores verdadeiros positivos dividido pela soma dos valores verdadeiros positivos e falsos positivos, veja Equação 17

$$P = \frac{VP}{VP + FP} \quad (17)$$

Para o caso da classificação multi-classe a precisão macro pode ser calculada (SOKOLOVA; LAPALME, 2009) conforme a equação 18

$$P_M = \frac{\sum_{i=1}^n \frac{VP_i}{VP_i + FP_i}}{n} \quad (18)$$

onde  $P_M$  é o resultado da precisão macro calculada pela média do somatório da precisão de todas as classes presentes na base de dados.

#### 2.2.6.6 Revocação

É uma medida de avaliação que analisa a quantidade de acertos do modelo de classificação dentre as classes com o rótulo positivo como o valor esperado (BUCKLAND; GEY, 1994). A revocação pode ser utilizada em situações onde os valores falsos negativos são considerados mais graves do que os falsos positivos, por exemplo diagnóstico de doenças. Ela é calculada utilizando a Equação 19

$$R = \frac{VP}{VP + FN} \quad (19)$$

onde,  $R$  é o resultado da revocação,  $VP$  os valores positivos e  $FN$  os falsos negativos.

Para o caso da classificação multi-classe a revocação macro pode ser calculada (SOKOLOVA; LAPALME, 2009) conforme a equação 20

$$R_M = \frac{\sum_{i=1}^n \frac{VP}{VP + FN}}{n} \quad (20)$$

onde  $R_M$  é o resultado da revocação macro calculada pela média do somatório da revocação de todas as classes presentes na base de dados.

#### 2.2.6.7 F1-Score

É uma medida de avaliação que faz a média harmônica entre a precisão e a revocação (GOUTTE; GAUSSIER, 2005). Se o valor *F1-Score* for alto significa que os valores da precisão e da revocação estão equilibrados e o resultado da acurácia é confiável. O *F1-Score* para classificação multi-classe é calculado (SOKOLOVA; LAPALME, 2009) utilizando a Equação 21

$$F_1Score_M = \frac{\sum_{i=1}^n 2 \cdot \frac{P \cdot R}{P+R}}{n} \quad (21)$$

onde,  $F_1Score_M$  é calculado pela média dos resultados de  $F1-Score$  de cada classe. E o  $F1-Score$  é o dobro da divisão entre a multiplicação da precisão e revocação e a soma entre precisão e revocação.

## 2.2.7 Outras Tarefas de Mineração de Texto

Além das tarefas já descritas nas seções anteriores, é comum que um dos passos da mineração de textos seja a análise descritiva dos dados. Essas análises visam identificar, comparar e analisar as características de uma base de dados (REIS; REIS, 2002). Além disso, as análises descritivas são importantes para identificar padrões de comportamento de determinado conjunto de dados. Nas análises descritivas, além de comportamentos padrões, podem ser identificados anomalias, ou seja, comportamentos que se diferem do padrão da maioria dos dados.

A análise descritiva dos dados usa todos os dados da base de dados analisada (SILVESTRE, 2007). A partir disso, é possível extrair informações dos dados e sintetizá-las por meio de tabelas, gráficos e medidas descritivas (GUEDES et al., 2005). Neste trabalho as análises descritivas utilizadas foram testes estatísticos, árvore de decisão e gráficos.

Os testes estatísticos são utilizados para ajudar na tomada de decisão dado uma amostra de dados de uma determinada população (VAYEGO, 2022). Nesses testes uma hipótese é dada e são feitos cálculos estatísticos, de acordo com cada teste, para identificar se a hipótese é aceita ou não. Dois exemplos de testes estatísticos são Teste T de *Student* e Teste Exato de Fisher.

O teste T de *Student* avalia a diferença da média entre duas amostras (KIM, 2015). O valor de  $t$  utilizado nesse teste é calculado conforme a Equação 22

$$t = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (22)$$

onde  $M_1$  e  $M_2$  são as médias de cada amostra analisada,  $s_1$  e  $s_2$  o desvio padrão de cada amostra e  $n_1$  e  $n_2$  a quantidade total de cada amostra.

O Teste Exato de Fisher é normalmente utilizado para amostras pequenas e compara dois grupos de acordo com uma característica (BOWER, 2003). Esse teste utiliza uma tabela com os valores de cada amostra (FISHER, 1922), conforme mostra a Figura 12. Todas as possibilidades de combinações das células dessa tabela são consideradas no cálculo do  $p$  valor.

O  $p$  valor pode ser calculado conforme a equação 23

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!} \quad (23)$$

	Homem	Mulher	Total
Graduação	1 (a)	9 (b)	10 (a+b)
Sem Graduação	11 (c)	3 (d)	14 (c+d)
<b>Total</b>	<b>12 (a+c)</b>	<b>12 (b+d)</b>	<b>24 (a+b+c+d) = n</b>

Figura 12 – Tabela de Relação Entre Homens e Mulheres Graduados. Fonte: Adaptado de (FISHER, 1922)

onde,  $N$  é o total das frequências e  $a$ ,  $b$ ,  $c$  e  $d$  são a frequência individual de cada célula da tabela.

A construção da árvore de decisão auxilia na identificação das características de uma determinada classe, porque essa árvore toma uma decisão através de um caminho. (CAMPOS, 2021). Conforme ilustrado na Figura 5 a árvore de decisão mostra o caminho que foi percorrido para chegar em um determinado rótulo.

## 2.3 Considerações Finais

Este capítulo teve o objetivo de descrever os principais conceitos que serão utilizados neste trabalho. A primeira seção apresentou os conceitos de sentimentos, emoções e temperamento. Descreveu sobre o questionário de temperamento TEMPS-A, o qual é uma das formas de identificar o temperamento humano e por fim dissertou sobre a influência das redes sociais na vida dos seres humanos.

A segunda seção descreveu sobre mineração de texto em redes sociais e detalhou todas as suas etapas, sendo elas, coleta de dados, pré-processamento, normalização de texto, definição de características, técnicas de aprendizado de máquina, métodos e medidas de avaliação e por fim, outras tarefas de mineração de texto.

O próximo capítulo apresenta os trabalhos do estado da arte que estão relacionados ao tema deste trabalho.





---

## Trabalhos Relacionados

Este capítulo apresenta os trabalhos relacionados que estão presentes na literatura. A fim de organizar os trabalhos mais relacionados ao tema deste projeto, este capítulo está dividido em duas seções: i) trabalhos de análise de sentimentos que usaram postagens em língua inglesa e língua portuguesa; ii) e os trabalhos mais relacionados ao proposto, que analisam temperamento/personalidade dos usuários.

Na literatura existem diversos trabalhos com extração de conhecimento a partir de redes sociais, mas poucos utilizam o idioma português nas análises e os que estudam temperamento são ainda mais raros. O primeiro grupo de trabalhos está diretamente relacionado com a primeira parte da proposta deste trabalho, que consiste na análise de sentimentos das postagens do *Instagram*. O segundo grupo está relacionado com a segunda parte do trabalho que consiste em contrastar temperamento e comportamento na rede social.

A pesquisa dos trabalhos relacionados foi feita por palavras-chave em português e inglês no contexto de análise de sentimentos em redes sociais, análise de sentimentos em redes sociais em português e temperamento e redes sociais nos sites Google Acadêmico<sup>1</sup>, Biblioteca Digital da Sociedade Brasileira de Computação<sup>2</sup> *ACM Digital Library*<sup>3</sup> e *Research Gate*<sup>4</sup>. Algumas das palavras-chave utilizadas foram 'análise de sentimentos', '*Instagram* e temperamento', 'análise de sentimentos em português', '*emotions and sentiments*', '*human temperament and sentiment analysis*'.

---

<sup>1</sup> <https://scholar.google.com.br/>

<sup>2</sup> <https://sol.sbc.org.br/index.php/indice>

<sup>3</sup> <https://dl.acm.org/>

<sup>4</sup> <https://www.researchgate.net/>

## 3.1 Análise de Sentimentos em Publicações de Redes Sociais

Existem diversos trabalhos na literatura sobre análise de sentimentos em redes sociais no idioma em inglês. Já no idioma em português eles são mais incomuns. Os trabalhos descritos nessa seção foram selecionados, porque os mesmos possuem as técnicas de pré-processamento e/ou classificação de texto das redes sociais similares. Ademais, alguns dos estudos descritos utilizam um comitê (*ensemble*) de classificadores, assim como uma das estratégias propostas neste trabalho.

O estudo de (SALEENA et al., 2018) propôs um *ensemble* de classificadores utilizando os algoritmos *Naive Bayes* (NB), *Random Forest* (RF) e SVM para a tarefa de classificação de sentimentos na rede social Twitter. O trabalho utilizou quatro bases de dados públicas do Twitter, *Stanford - Sentiment140 corpus*<sup>5</sup>, *Health Care Reform (HCR)* (SPERIOSU et al., 2011), *First GOP debate twitter sentiment dataset*<sup>6</sup> e *Twitter sentiment analysis dataset*. Todas as bases estavam no idioma inglês. Os tweets com sentimentos positivo e negativo foram considerados nas análises. As técnicas de pré-processamento utilizadas foram substituição de *emojis* por palavras que representassem o *emoji*, *stemming*, substituição de gírias por palavras com significado equivalente e por fim remoção dos seguintes itens: menção de usuários, URLs, *retweets* que começassem com "RT", *hashtags* e *stopwords*. A técnica de extração de características utilizada foi o *Bag-of-Words* (BOW). Os classificadores utilizados foram *Naive Bayes*, *Random Forest*, SVM, *Logistic Regression* (LR) e comitê composto por *Naive Bayes*, *Random Forest* e SVM. Os autores utilizaram o Python na implementação. O *ensemble* de classificadores trouxe os melhores resultados em todas as bases de dados utilizadas, com acurácia de 75,81%, 73,68%, 85,83% e 74,67%, respectivamente.

No trabalho de Zhan, Tu e Yu (2018) foram analisadas as *hashtags* #reading e #read da rede social *Instagram* com objetivo de identificar a polaridade e as emoções das legendas da rede social. Apenas legendas em inglês foram analisadas neste trabalho. As classes de emoções usadas foram: amor, alegria, tristeza, medo, raiva e surpresa. O estudo relacionou as emoções com a polaridade, por exemplo, publicações que expressavam emoções positivas (amor, surpresa e alegria) estavam em sua maioria relacionadas as publicações classificadas como positivas. As bibliotecas do *Python Vader* e *TextBlob* foram usadas na extração de atributos e o algoritmo *Random Forest* na classificação. A base de dados utilizada no trabalho tinha 2.035.706 de publicações e o treinamento do algoritmo foi feito com base em 500 publicações com polaridade neutra, 500 positiva e 500 negativa. A medida de validação *5-fold cross-validation* teve acurácia de 86,12%. Os autores concluíram que 50.5% dos posts foram positivos, 5% foram negativos e 44.5% neutros. Os

<sup>5</sup> <https://www.kaggle.com/datasets/kazanov/sentiment140>. Acesso em 02/11/2022.

<sup>6</sup> <https://www.kaggle.com/datasets/crowdfower/first-gop-debate-twitter-sentiment>. Acesso em 02/11/2022.

*posts* positivos expressaram mais a emoção alegria e os negativos as emoções de tristeza e raiva.

O estudo de Pinto et al. (2020) tem como um de seus objetivos realizar testes com algoritmos de análise de sentimento de texto no contexto da pandemia do COVID-19. Foi utilizada uma base de dados pública na língua inglesa com 41.157 *tweets* utilizados na fase de treino e 3.798 *tweets* na fase de teste. As classes utilizadas foram positivo, negativo, neutro, extremamente positivo e extremamente negativo. No pré-processamento dos dados foi feita a transformação de todas as palavras para minúsculo, remoção de *stopwords*, pontuações, números e caracteres especiais. Na análise de sentimentos foi utilizada a rede neural do tipo MLP, *Universal Sentence Encoder*, *Polyglot* e *TextBlob*. Os autores concluíram que para o contexto do trabalho, o modelo de *Universal Sentence Encoder* em conjunto com a rede neural do tipo MLP teve melhores resultados com F1-Score de 68,77%.

O trabalho de Aguiar et al. (2018) teve o objetivo de apresentar um modelo composto por um conjunto de algoritmos de classificação para estimar os sentimentos de *tweets* que estavam em português da rede social *Twitter*. A base de dados pública, com 2.516 *tweets* com sentimentos positivos, negativos e neutros do grupo *MiningBR* foi utilizada. Na extração de características do texto foi feita a normalização que teve as seguintes etapas: limpeza do texto, extração de *tokens* e remoção de *stopwords*. Na normalização, foi utilizado o *bag of words*. Para a otimização dos algoritmos foi utilizado o *Grid Search* da biblioteca SKLEARN do Python<sup>7</sup>. A técnica de classificação usada foi baseada em um comitê de classificadores, composto pelos algoritmos *Naive Bayes*, SVM, Árvore de decisão, *Random Forest* e Regressão Logística. Para a validação do desempenho foram utilizadas as medidas de acurácia, erro, precisão, revocação e *F1-score*. Para a validação dos classificadores também foi utilizada a curva ROC (*Receiver Operating Characteristic*) e, por fim, para a amostragem foi utilizado o método *cross-validation* com K igual a 10. Além disso, foram executados os testes estatísticos *Shapiro Wilk*, *Friedman* e *post-hoc* para verificação da eficácia do comitê. Esse trabalho concluiu que, no geral, o melhor desempenho foi do comitê, porém conforme o gráfico ROC, o *Naive Bayes* gera os melhores modelos de classificação. Após os testes estatísticos, foi concluído que para conjuntos de dados pequenos é melhor utilizar o comitê, mas para base de dados maiores o *Naive Bayes* é o mais indicado.

O objetivo do trabalho (KANSAON; BRANDÃO; PINTO, 2018) era apresentar uma análise de sentimentos dos *tweets* em português brasileiro para comparar as técnicas de classificação de sentimentos. Para a coleta de dados foi utilizada a API do *Twitter* sendo feitas buscas pelas *hashtags* #Triste, #Chateado, #Feliz, #Amor, #Raiva, #Inveja, #Ironia. Essas *hashtags* representam os sentimentos analisados nesse estudo. No pré-processamento do texto, foram removidas as *hashtags* utilizadas na busca, *stop words*,

<sup>7</sup> <https://scikit-learn.org/stable/>

links, *retweets* e menções. Os *emoticons* foram substituídos por palavras com a inicial 'E', por exemplo 'ECoracao'. Os sentimentos de ironia, inveja e raiva foram excluídos da análise, porque tinham menos de 1.000 *tweets*. Com isso, as análises foram feitas em cima de 9.631 *tweets*. A análise foi feita com base nas relações de sentimentos positivos x negativos, negativos x negativos e positivos x positivos. Os algoritmos de classificação utilizados foram: *Naive Bayes*, *Naive Bayes Multinomial*, *Naive Bayes Multinomial Updateable*, *Sparse Generative Model*, *DMNB Text*, *Complement Naive Bayes*, *Bayesian Logistic Regression*, *IBK*, *Forest* e *Random Committee*, todos executados no framework WEKA. As medidas de validação foram *F1-Score* e ROC e a estratégia de separação dos dados para treino e teste foi o *cross-validation* com K igual a 10. O trabalho concluiu que os algoritmos baseados no *Naive Bayes* tiveram melhores resultados e a relação dos sentimentos positivos x negativos tiveram os resultados mais satisfatórios com acurácia de 85% com o algoritmo *Complement Naive Bayes*.

As Tabelas 1 e 2 apresentam o resumo dos trabalhos relacionados descritos nesta seção.

Tabela 1 – Resumo Trabalhos Relacionados — Classificação de Texto em Publicações de Redes Sociais em Inglês

Referência	Objetivo	Classificador	Medidas de Validação	Estratégia Treino e Teste	Base de Dados	Resultado
(SALEENA et al., 2018)	Propor um <i>ensemble</i> de classificadores na análise de sentimentos do Twitter	NB, RF SVM, LR	Acurácia Precisão Revocação F1-Score	-	Bases de dados Pública	O <i>ensemble</i> de classificadores trouxe os melhores resultados em todas as bases de dados
(ZHAN; TU; YU, 2018)	Analisar sentimentos e emoções dos posts dos usuários com hastags #reading e #read	Random Forest (RF)	Acurácia	K Fold cross validation	Coleta de dados via código PHP	50.5% dos posts foram positivos, 5% foram negativos e 44.5% neutros. Os posts positivos expressaram mais a emoção de alegria e os negativos as emoções tristeza e raiva
(PINTO et al., 2020)	Realizar testes com algoritmos de análise de sentimentos de texto no contexto da COVID-19	Rede Neural MLP, Universal Sentence Encoder, Polyglot TextBlob	F1-Score Precisão Revocação	-	Base Pública <sup>8</sup>	Modelo Universal Sentence Encoder com a rede MLP teve melhores resultados de F1-Score

Tabela 2 – Resumo Trabalhos Relacionados — Classificação de Texto em Publicações de Redes Sociais em Português

Referência	Objetivo	Classificador	Medidas de Validação	Estratégia Treino e Teste	Base de Dados	Resultado
(AGUIAR et al., 2018)	Estimar o sentimento em comentários do Twitter para a língua portuguesa com base em algoritmos de classificação	Comitê composto por: NB, SVM, Árvore de decisão, RF e Regressão Logística	Acurácia, erro, precisão, recall, F1-score e ROC	K Fold cross validation	Base de dados pública MiningBR	O melhor desempenho dos teste foi do comitê. Porém, após os testes estatísticos, foi concluído que para conjuntos de dados pequenos é melhor utilizar o comitê, mas para base de dados maiores o NB é o mais indicado
(KANSAON; BRANDÃO; PINTO, 2018)	Apresentar uma análise de sentimentos dos tweets em português de um usuário	NB, NB Multinomial, NB Multinomial Updateable, Sparge Generative Model, DMNB Text, Complement NB, Bayesian Logistic Regression, IBK, Forest e Random Committee	F-Measure e ROC	K Fold cross validation	API do Twitter	Algoritmos baseados no NB tiveram melhores resultados

## 3.2 Redes Sociais e Temperamento Humano

Poucos trabalhos na literatura trata a relação entre temperamento humano (e/ou personalidade) e redes sociais. Considerando a língua portuguesa, nenhum trabalho foi encontrado.

O trabalho de Ferwerda, Schedl e Tkalcic (2016) analisou as imagens das publicações do *Instagram* para prever a personalidade do usuário. A coleta dos dados foi feita utilizando a API da rede social e a pesquisa teve 113 participantes com 22.398 imagens. Os participantes da pesquisa responderam um questionário da escala de *Likert* (JOSHI et al., 2015) sobre sua personalidade. A análise das fotos das publicações foram baseadas nos filtros aplicados, como saturação e brilho, extraindo as cores que mais se destacavam. O modelo de previsão de personalidade criado foi treinado utilizando o método de validação *10-fold cross validation* e os classificadores *Radial Basis Function Network*, *Random Forest* e *M5 Rules* do *framework WEKA*. Para relacionar as imagens e a personalidade do usuário foi utilizada a correlação de *Pearson* ( $r [-1,1]$ ). As classes de personalidade foram: aberto à experiência, consciência, extroversão, amabilidade, neuroticismo. Os autores concluíram que as características das imagens estão correlacionadas com a personalidade do usuário. As figuras mais correlacionadas foram a da classe aberto à experiência que continha mais tons de verde, alta saturação, pouco brilho, mais cores frias e menos rostos de pessoas na foto e a menos correlacionada foi a classe neuroticismo que apresentava uma imagem com muito brilho.

Os autores de (FERWERDA; TKALCIC, 2018) também fizeram a análise de imagens do *Instagram* com o intuito de descobrir a relação da foto publicada e a personalidade do usuário. A coleta de dados do *Instagram* foi feita utilizando a API da rede social e o questionário seguiu a escala de *Likert*. Ao todo, 193 usuários, que juntos tinham 54.962 fotos, participaram da pesquisa. Para analisar as fotos, foi utilizada a *API do Google Vision* e para relacionar as imagens com a personalidade do usuário os autores usaram a escala de Spearman's. As fotos foram separadas em categorias, como animais, comida, esportes, partes do corpo etc., utilizando o algoritmo de agrupamento *K-MEANS*. As classes da personalidade foram: aberto à experiência, consciência, extroversão, amabilidade e neuroticismo. Esse estudo concluiu que a personalidade aberto a experiência está relacionada com imagens da categoria instrumentos musicais, a classe consciência com a categoria de roupas e esportes, extroversão com eletrônicos, amabilidade com roupas e lazer, e por fim neuroticismo relacionado a roupas e joias.

Em Skowron et al. (2016) foi proposto um modelo para integrar as análises de texto, imagens e personalidades de usuários das redes sociais *Instagram* e *Twitter*. A motivação foi que um mesmo usuário pode demonstrar comportamentos distintos em diferentes redes sociais. Os participantes da pesquisa foram recrutados na *Amazon Mechanical Turk*. A pesquisa teve 62 usuários e todos estavam com mais de 30 publicações em cada rede social. Nas imagens do *Instagram* os autores utilizaram o modelo de estado emocional

PAD (*Pleasure, Arousal and Dominance*), brilho, saturação, tons e conteúdo da imagem. Nos textos, tanto do *Instagram* quanto do *Twitter*, os autores utilizaram o sistema de Skowron et al. (2011), que faz a integração entre processamento de linguagem natural e classificadores. O algoritmo *Random Forest* foi utilizado para construir um modelo de baixa variância e baixo viés das características dos traços de personalidade. Além disso, também foi considerado o número de seguidores e seguidos, reputação e influência do usuário. O questionário aplicado tinha 44 perguntas de compreensão e verificações cruzadas com base nos cinco grandes traços de personalidades (aberto à experiência, consciência, extroversão, amabilidade e neuroticismo). Os resultados preliminares da pesquisa indicaram que a análise das redes sociais simultâneas parece levar a diminuição consistente dos erros na previsão dos traços de personalidade.

O trabalho de (GOLBECK et al., 2011) analisou a rede social *Twitter* para prever a personalidade de um usuário com base no seu perfil e nos *tweets*. Os participantes da pesquisa responderam um formulário de 45 perguntas baseadas nas cinco personalidades: aberto à experiência, consciência, extroversão, amabilidade e neuroticismo. A base de dados utilizada foi a mesma de (PENNEBAKER; KING, 1999), que possui 2.000 *tweets* mais recentes do participante, ou então todos, caso ele tivesse menos que o estipulado. Os seguintes dados do *Twitter* foram considerados: número de seguidores e seguidos, densidade do perfil, número de citações de usuários, número de respostas dos *tweets*, quantidade de *hashtags*, *links* e palavras por tweet. Os números de citações, respostas, *hashtags* e *links* foram utilizados apenas para contagem e média. Para analisar o conteúdo os *tweets* foi utilizada a ferramenta LIWC que produziu características das seguintes categorias: contagem de palavras, processos psicológicos, tempo, preocupações pessoais e outras dimensões como ocupação, situação financeira e palavras. Os autores optaram por excluir as categorias de contagem de palavras e outras dimensões para evitar ruídos no texto. Além do LIWC, foi utilizada a base de dados *MRC Psycholinguistic* que contém uma lista de mais de 150.000 palavras com características linguísticas e psicolinguísticas de cada palavra. Para analisar os sentimentos das palavras foi utilizada a base de dados *General Inquirer* que é um dicionário de palavras que classifica as palavras em sentimentos numa escala de -1 a 1. Para correlacionar a personalidade com as palavras dos *tweets* foi utilizada a correlação de Pearson. Na previsão da personalidade foi utilizado os algoritmos *Gaussian Process* e *ZeroR* com validação cruzada de 10 pastas com 10 iterações, no WEKA. Essa pesquisa concluiu que usando os dados do perfil como um conjunto de recursos, é possível treinar os algoritmos *Gaussian Process* e *ZeroR* para prever pontuações em cada um dos cinco traços de personalidade com uma variação de 11% a 18% de seu valor real.

Em (LIMA; CASTRO, 2019) os autores têm o objetivo de identificar se existem padrões de comportamento nas publicações do *Twitter*, que podem ser mapeados com alta precisão nos tipos psicológicos do MBTI ou nos temperamentos de *Keirsey* (CLARO et al., 2018). Para essa tarefa foi desenvolvido o *framework* TECLA, que usa técnicas de



processamento de linguagem natural e mineração de texto para classificar temperamentos ou tipos psicológicos. A base de dados utilizada foi a do trabalho de (PLANK; HOVY, 2015) com mais de um milhão de mensagens de 1500 usuários. As classes do modelo de MBTI foram introversão/extroversão, intuição/sensação, pensamento/sentimento e julgamento/percepção. As classes do modelo de Keirsey foram artesão, guardião, idealista e racional. Os algoritmos de classificação utilizados foram *AdaBoost*, *Bagging*, *J48*, *Naive Bayes*, *Random Forest* e SVM. O desempenho dos classificadores foi comparado com as características (MBSP, n-gramas, LIWC, oNLP, POS-tags e características léxicas) e com MRC, LIWC e *Apache OpenNLP* para encontrar uma combinação de temperamento e tipos psicológicos com base nos dados do *Twitter*. As medidas de validação utilizadas foram acurácia, F1-score e ROC. A estratégia de separação de treino e teste foi *K-fold cross-validation* com 10 iterações e com *K* igual a 5. Os resultados do framework TECLA foram comparados com os resultados obtidos pelos demais trabalhos da literatura. Nos testes do modelo de MBTI, o TECLA sempre teve o melhor resultado. Nos testes do modelo de *Keirsey* o *Random Forest* com LIWC teve melhor desempenho. Os resultados obtidos na predição do temperamento foram maiores que 80%, na acurácia, para a maioria dos temperamentos.

As Tabelas 3 e 4 apresentam o resumo dos trabalhos relacionados descritos nesta seção.

Tabela 3 – Resumo Trabalhos Relacionados — Redes Sociais e Temperamento Humano - Parte 1

Referência	Objetivo	Classificador	Medidas de Validação	Estratégia Treino e Teste	Base de Dados	Resultado
(FERWERDA; SCHEDL; TKALCIC,2016)	Descobrir a personalidade do usuário com base no filtro que era usado em suas fotos	Radial Basis Function Network, Random Forest e M5 Rules	Escala de Person	K Fold cross validation	API Instagram e questionário personalidade	As características das imagens estão correlacionadas com a personalidade do usuário
(FERWERDA; TKALCIC, 2018)	Descobrir relações das fotos postadas no Instagram com a personalidade dos usuários	K-means com 17 clusters	Escala de Spearman's	-	API Instagram e questionário personalidade	A personalidade aberto a experiência está relacionada com imagens instrumentos musicais, a classe consciência com roupas e esportes, extroversão com eletrônicos, amabilidade com roupas e lazer, e por fim neuroticismo relacionado a roupas e joias
(SKOWRON et al., 2016)	Relacionar imagem, texto das redes sociais Instagram e Twitter para prever a personalidade	Sistema de (SKOWRON et al.,2011)	-	-	Amazon Mechanical Turk e questionário personalidade	A análise das redes sociais simultâneas parece levar a diminuição consistente dos erros na previsão dos traços de personalidade

Tabela 4 – Resumo Trabalhos Relacionados — Redes Sociais e Temperamento Humano - Parte 2

Referência	Objetivo	Classificador	Medidas de Validação	Estratégia Treino e Teste	Base de Dados	Resultado
(GOLBECK et al., 2011)	Predizer a personalidade de um usuário com base no seu perfil e tweets no twitter.	Gaussian Process e ZeroR	Escala de Person	K Fold cross validation	API do Twitter	É possível treinar os algoritmos Gaussian Process e ZeroR para prever pontuações em cada um dos cinco traços de personalidade de 11% a 18% da base total
(LIMA; CASTRO, 2019)	Identificar padrões de comportamento nas publicações do Twitter que podem ser mapeados com alta precisão nos tipos psicológicos do MBTI ou nos temperamentos de Keirsey	AdaBoost, Bagging, J48, NB, RF e SVM	Acurácia, F-Measure e ROC	K Fold cross validation	Base de (PLANK; HOVY,2015)	Nos testes do modelo de Keirsey o RF com LIWC e para o modelo de MBTI RF com LIWC ou oNLP tiveram melhor desempenho.

### 3.3 Considerações Finais

Este capítulo teve o objetivo de descrever os principais trabalhos relacionados de mineração de texto e análise de sentimentos em redes sociais e sobre temperamento humano e redes sociais.

O trabalho proposto se diferencia dos trabalhos relacionados nos seguintes pontos: i) usa dados de postagens em língua portuguesa; ii) usa uma base de dados específica de discentes da Universidade Federal de Uberlândia; iii) usa os dados comportamentais e postagens dos usuários da rede *Instagram* para analisar temperamento; iv) o teste de temperamento usado será o TEMPS-RIO, que identifica o temperamento do indivíduo por meio de questões de verdadeiro ou falso.

## Método para Análise de Temperamento a Partir de Dados de Redes Sociais

Este capítulo apresenta o método proposto para relacionar o temperamento de um usuário medido pelo questionário de temperamento TEMPS-RIO com os dados extraídos da rede social *Instagram*. O trabalho foi dividido nas etapas de coleta de dados, rotulação da base, construção de um *framework* para classificar o sentimento das postagens e análises da relação entre temperamento e dados da rede social. A Figura 13 apresenta um resumo do método para análise de temperamento a partir de redes sociais. Cada uma das etapas serão descritas detalhadamente a seguir.

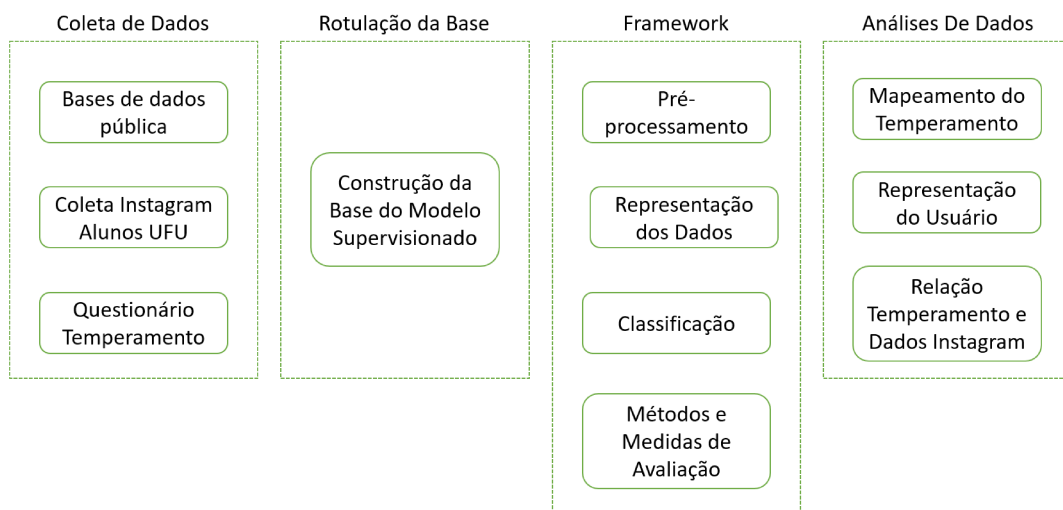


Figura 13 – Método Proposto Para Análise de Temperamento a Partir de Redes Sociais.  
Fonte: Elaborado pela autora (2022)

## 4.1 Coleta de Dados

Este trabalho criou uma base de dados própria a partir de dados de voluntários discentes da UFU. O projeto foi submetido e aprovado pelo Comitê de Ética em Pesquisas com Seres Humanos da Universidade Federal de Uberlândia — CAAE 35533420.2.0000.5152 — que autorizou a criação da base.

Para a coleta de dados, o website<sup>1</sup> foi desenvolvido pela autora para coletar os dados dos discentes da Universidade Federal de Uberlândia. Os discentes são alunos dos cursos de Tecnologia da Informação e áreas afins. O site possuía três passos: coleta de dados do *Instagram*, coleta de dados do *Twitter* e questionário de temperamento, conforme ilustrado na Figura 14. Para este trabalho, apenas os passos de coleta de dados do *Instagram* e questionário de temperamento foram utilizados. A técnica de coleta do *Instagram* utilizada foi o *scraping*. Apenas os discentes que aceitaram participar da pesquisa (assinando o Termo de Compromisso Livre e Esclarecido) e possuíam conta aberta do *Instagram* foram considerados nas análises. A divulgação da pesquisa para conseguir voluntários foi realizada por meio de e-mails, divulgação na rede social do grupo de pesquisa, divulgação na rede social da UFU, conversa com os alunos nas salas de aula online e presencial e abordagem feita nos demais espaços da universidade. Um perfil no Instagram<sup>1</sup> foi criado para a divulgação de informações sobre o projeto da Equipe Social Temp.



Figura 14 – Site Equipe Social Temp. Fonte: Elaborado pela autora (2022)

O questionário de temperamento utilizado foi o TEMPS-RIO que possui 45 questões de verdadeiro ou falso para identificar um dos seguintes temperamentos do usuário: ciclotímico, irritável, preocupado, hipertímico, depressivo e ansioso (WOODRUFF et al., 2011). Os participantes desta pesquisa responderam o questionário utilizando o formulário presente no Apêndice A.

<sup>1</sup> <https://www.socialtemp.facom.ufu.br/>

<sup>1</sup> <https://www.instagram.com/socialtempufu/>

## 4.2 Rotulação da Base

Duas bases de dados foram coletadas e elas são compostas por publicações de usuários em uma rede social. Como este trabalho utiliza algoritmos supervisionados, uma amostra da base de dados precisou ser rotulada. Essa amostra foi manualmente rotulada com o sentimento positivo, negativo ou neutro.

A Base de Dados I (B1) foi extraída da rede social *Instagram* de usuários selecionados aleatoriamente que possuíam conta pública no *Instagram* e ela foi rotulada por completo por um aluno da Faculdade de Medicina da Universidade Federal de Uberlândia. O aluno da faculdade de Medicina foi escolhido, porque ele possuía mais conhecimento técnico sobre os conceitos de sentimentos.

A Base de Dados III (B3) também foi extraída do *Instagram* de usuários que aceitaram participar da pesquisa e ela foi rotulada conforme o manual de classificação (Apêndice B) criado para este projeto. Esse manual foi criado por três alunos de Iniciação Científica (IC) da Universidade Federal de Uberlândia (UFU), para que a rotulação fosse feita padronizadamente por todos os rotuladores. Quando não foi possível rotular a postagem olhando apenas o texto do *Instagram*, ela foi marcada como inconclusiva. Dois discentes de IC da UFU rotularam a base de forma independente, seguindo o manual e um terceiro aluno de IC da UFU rotulou somente os casos de divergência. Nos casos de divergência, o terceiro rotulador concordava com um dos dois primeiros rotuladores. Os dois alunos que rotularam a base de forma independente tiveram concordância em 69% na rotulagem.

## 4.3 *Framework* de Extração de Sentimentos

O *framework* de extração de sentimentos faz o processamento dos dados ponta a ponta, ou seja, executa os passos de pré-processamento, representação, e classificação de sentimentos das legendas das publicações do *Instagram*.

Uma publicação do Instagram é composta por 1) foto, 2) legenda, 3) curtidas e 4) comentários, conforme exemplo mostrado na Figura 15. Neste trabalho o escopo principal foi a análise das legendas.

As legendas analisadas neste trabalho são da rede social *Instagram*, mas o *framework* desenvolvido não está limitado apenas a esta rede social. Ele pode ser usado para analisar as frases de outras redes sociais, como *Twitter*.

### 4.3.1 Pré-processamento

Na etapa de pré-processamento foi feita a extração das legendas das publicações dos usuários e a partir delas, aplicadas as técnicas de substituição de emojis por palavras



Figura 15 – Exemplo de Publicação no *Instagram*. Fonte: Elaborado pela autora (2022)

utilizando a biblioteca EMOJI<sup>2</sup> do Python, padronização das palavras para minúsculo e remoção de *stopwords*, URLs, e-mails, marcações de usuários, números e acentos.

O dicionário de *stopwords* utilizado foi da biblioteca do Python NLTK<sup>3</sup> com as palavras em português. Esse dicionário foi incrementado com algumas palavras comuns identificadas no texto e que não trazem nenhuma relevância, como as palavras "eu", "ser", "você", "nós".

### 4.3.2 Representação dos Dados

Para a representação do texto foi utilizado o método TF-IDF e o N-gramas. O TF-IDF foi escolhido porque ele mede a relevância de uma palavra no texto, sendo que para o contexto mineração de texto é uma técnica muito relevante. A escolha do N-gramas se deve porque essa técnica consegue analisar o contexto de mais de uma palavra no texto. Foram feitos experimentos com os valores de  $N$  igual a 2 e 3. Para o contexto deste trabalho, o N-gramas com  $N$  igual a 2 foi utilizado porque teve melhores resultados.

<sup>2</sup> <https://pypi.org/project/emoji/>

<sup>3</sup> <https://www.nltk.org/search.html?q=stopwords>



### 4.3.3 Classificação

Na etapa de classificação das legendas dos usuários do *Instagram* foram utilizadas duas estratégias de classificação: *ensemble* e em níveis.

#### 4.3.3.1 Classificação *Ensemble*

É uma estratégia que usa um conjunto de classificadores, treinados individualmente, combinados por um operador de consenso. Nesta proposta, os algoritmos usados são *Naive Bayes* (NB), *Random Forest* (RF) e *Support Vector Machine* (SVM), que são alguns dos algoritmos mais comuns na literatura para a tarefa de mineração de texto (HOTH; NÜRNBERGER; PAASS, 2005). Segundo os trabalhos do estado da arte (NAF'AN et al., 2019; ZHAN; TU; YU, 2018; AGUIAR et al., 2018; KANSAON; BRANDÃO; PINTO, 2018; LIMA; CASTRO, 2019), esses algoritmos são também bons candidatos para classificação de texto de redes sociais. Além disso, o trabalho de (RUZ; HENRÍQUEZ; MASCAREÑO, 2020) apresentou que os algoritmos escolhidos nesta estratégia *ensemble* são os mais utilizados na literatura para a tarefa de classificação de sentimentos. Vale ressaltar também que os trabalhos de (KANSAON; BRANDÃO; PINTO, 2018) e (AGUIAR et al., 2018) utilizaram a estratégia de *ensemble* para a tarefa de classificação de sentimentos de texto em redes sociais.

Na estratégia *ensemble* utilizada neste trabalho, cada classificador faz a sua classificação independentemente e a legenda foi classificada com o voto da maioria, veja Figura 16. Caso, cada classificador escolha uma classe, positivo, negativo e neutro, então a classe positivo é escolhida para o rótulo da legenda. Os três classificadores escolhidos também foram analisados separadamente.

#### 4.3.3.2 Classificação de Sentimento - Proposta baseada em Níveis

Os textos das redes sociais podem ter muitos *emojis* e esses podem ter muito peso no sentimento do texto da publicação da rede social. Além disso, uma palavra, por si só ou duas palavras em conjunto, podem expressar o sentimento de toda a frase publicada na rede social. Dado esses dois fatos, este trabalho propõe um método baseado em níveis para classificação de sentimentos, composto por três etapas: i) *emojis*, ii) palavras e iii) *ensemble*. As duas primeiras etapas do classificador de sentimentos em níveis são não-supervisionadas, ou seja, não precisam de um aprendizado prévio para atribuir o sentimento da legenda. Já a última etapa do classificador em níveis é supervisionada e utiliza a mesma técnica de classificação *ensemble* descrita na subseção 4.3.3.1.

A etapa i) analisa se a legenda possui algum *emoji* que está no dicionário de *emojis*<sup>4</sup> (criado para este trabalho). Esse dicionário é composto por *emojis* feitos de pontuações, exemplo ':)' e por *emojis* de imagem e seu respectivo sentimento. Os *emojis* selecionados

<sup>4</sup> <https://tinyurl.com/dicionarioEmojis>

## Classificação Ensemble

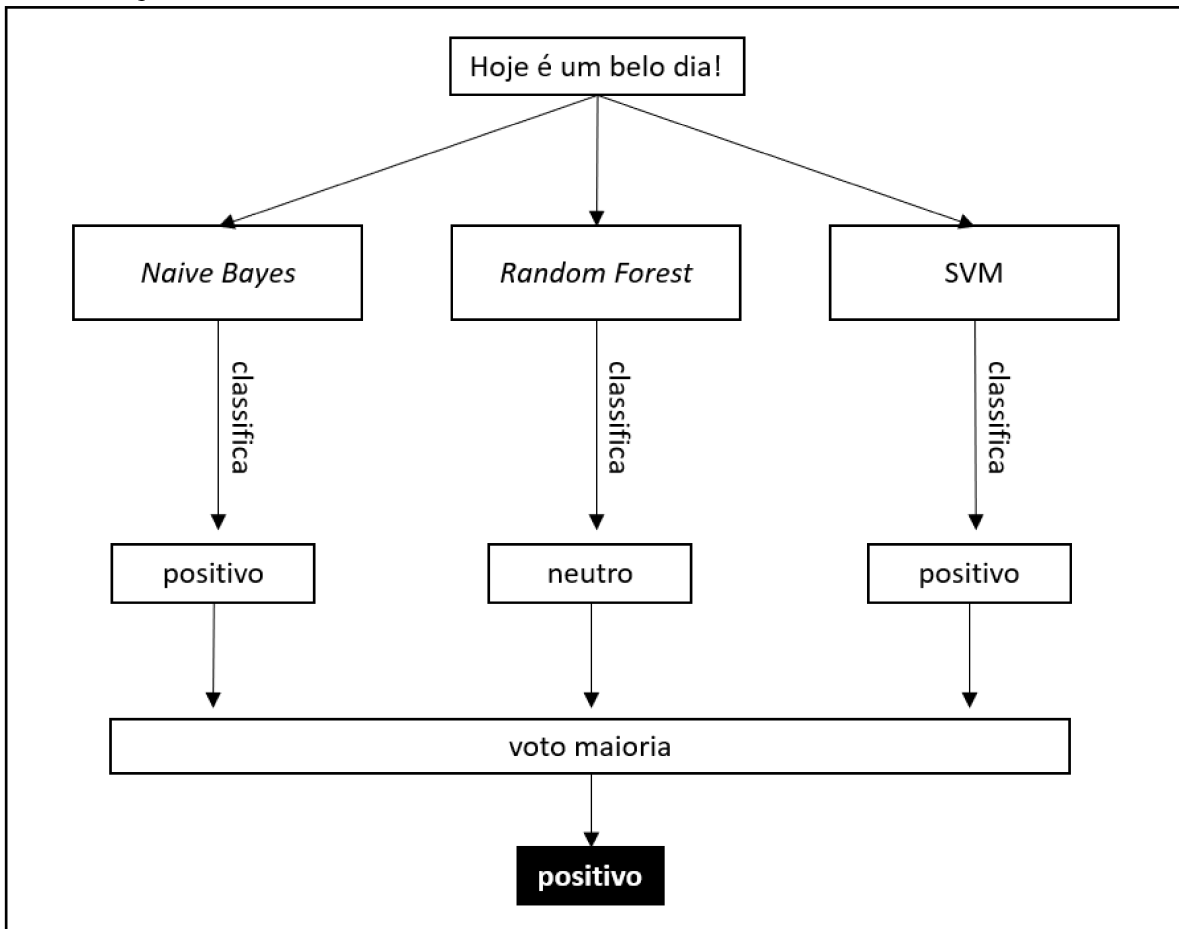


Figura 16 – Exemplo de classificação de uma sentença usando um *ensemble* de três classificadores. Fonte: Elaborado pela autora (2022)

para esse dicionário possuem sentimento positivo ou negativo. Caso exista pelo menos um *emoji* negativo na legenda, então ela é classificada com sentimento negativo independente da quantidade de *emojis* positivos que tenha na legenda. Caso exista apenas *emoji(s)* positivo(s) na legenda, então ela é classificada com sentimento positivo. O próximo nível da classificação é acionado caso a legenda não possua nenhum *emoji* ou caso o *emoji* encontrado na legenda não foi encontrado no dicionário de *emojis*. A etapa ii) avalia se a legenda possui pelo menos duas palavras no dicionário de palavras<sup>5</sup> (criado para este trabalho), que indicam um sentimento. O respectivo sentimento daquelas palavras é atribuído à legenda, somente se os sentimentos obtidos são iguais. No caso de mais de duas palavras presentes no dicionário, o voto da maioria é realizado. No caso de empate, ou caso nenhuma palavra da legenda seja encontrada no dicionário, vai para o próximo nível. A etapa iii) é feita caso a legenda ainda não tenha sido classificada nos níveis anteriores, então o *ensemble* de classificadores é usado para classificá-la. A Figura

<sup>5</sup> <https://tinyurl.com/dicionarioPalavraSentimento>

17 mostra o funcionamento da classificação de sentimento em níveis. Nos níveis 1 e 2 dessa estratégia a legenda original é utilizada para classificação e no nível 3 a legenda pré-processada é utilizada.

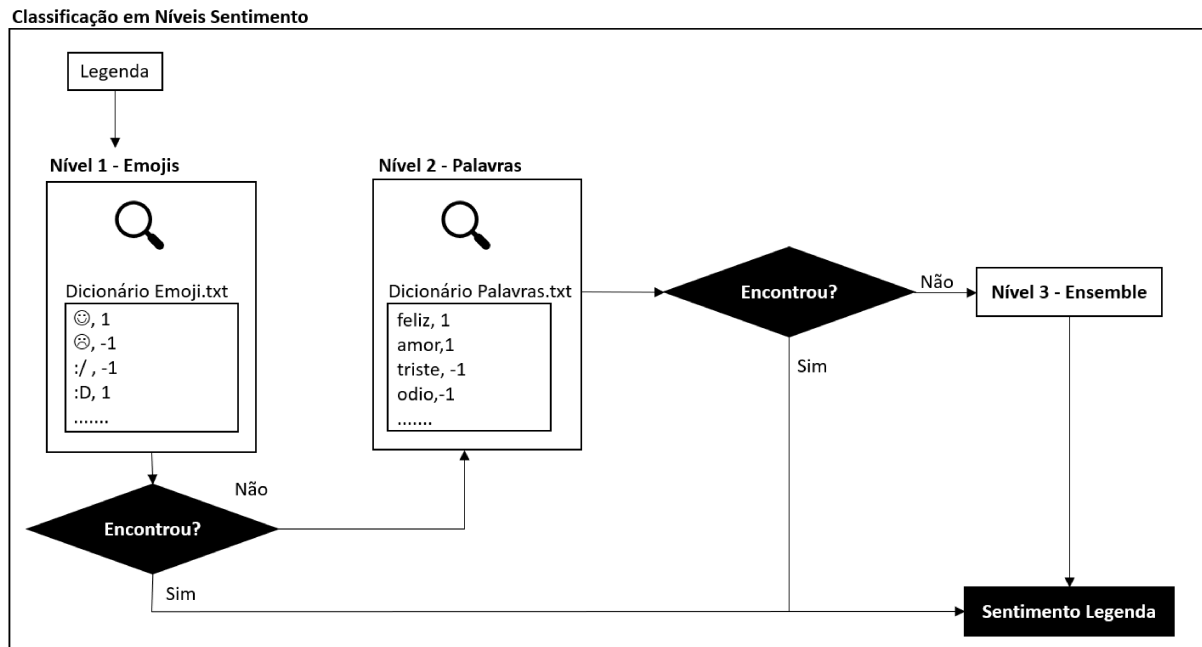


Figura 17 – Classificação de Sentimento em Níveis. Fonte: Elaborado pela autora (2022)

#### 4.3.4 Métodos e Medidas de Avaliação

O método de avaliação utilizado no *framework* foi o *K-Fold Cross-Validation* com  $K$  igual a 10. Esse valor de  $K$  foi escolhido porque é um dos mais utilizados na literatura (KANSAON; BRANDÃO; PINTO, 2018), (AGUIAR et al., 2018), (GOLBECK et al., 2011). As medidas de avaliação utilizadas foram acurácia macro e *F1-Score* macro.

## 4.4 Análises de Dados

A análise dos dados consistiu em fazer o mapeamento do temperamento dos usuários, representar um usuário a partir dos dados coletados do *Instagram* e relacionar esses dados do *Instagram* com o temperamento.

### 4.4.1 Mapeamento do temperamento dos usuários

Cada participante da pesquisa respondeu ao questionário de temperamento TEMPS-RIO. A fim de mapear o temperamento de cada participante, foram realizados os cálculos

indicados na Equação 7 da Seção 2.1. No caso de empate entre dois ou mais temperamentos, todos os temperamentos de empate foram considerados, ou seja, o usuário se repetiu em todos os temperamentos de empate. Os temperamentos que não tiveram empate são considerados temperamentos puros.

#### 4.4.2 Representação do Usuário

Após a classificação do sentimento das postagens do usuário e da identificação do seu temperamento, este trabalho propôs uma representação do usuário por meio de um conjunto de atributos. Os atributos foram extraídos a partir das postagens do usuário e de dados obtidos do seu perfil na rede social. O atributo alvo é o temperamento do usuário. A Figura 18 ilustra os atributos extraídos da rede social.

Os atributos usados para representar um usuário foram:

- ❑ proporção do uso de *emojis*, que indica a porcentagem das postagens do usuário que usaram o recurso de *emojis*;
- ❑ proporção de publicações positivas, que indica a porcentagem de publicações positivas do usuário em relação ao total de publicações. O sentimento das postagens foi identificado por meio do *framework* de extração de sentimentos criado neste trabalho;
- ❑ proporção de publicações negativas, que indica a porcentagem de publicações negativas do usuário em relação ao total de publicações.
- ❑ proporção de publicações no meio de semana e proporção de publicações fim de semana. Esses dados foram extraídos por meio da data da publicação da postagem extraída do *Instagram*;
- ❑ média de curtidas das publicações, que se deu pelo cálculo da média de curtidas de todas as publicações do usuário;
- ❑ quantidade de curtidas, extraído de cada publicação do usuário;
- ❑ quantidade de posts, extraído do perfil do usuário que representa a quantidade total de publicações que ele fez;
- ❑ proporção de seguidores e seguidos, que se deu pela divisão entre quantidade de seguidos e a quantidade de seguidores.

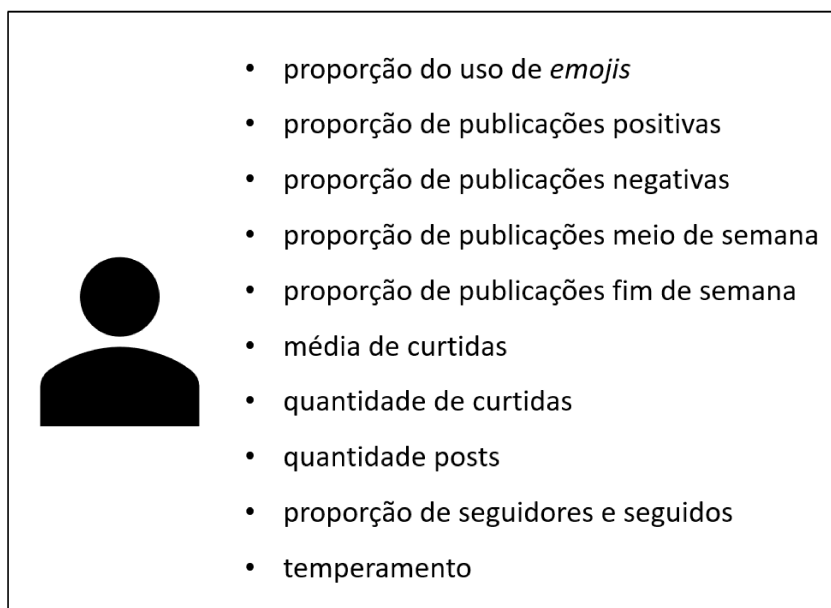


Figura 18 – Representação de um Usuário. Fonte: Elaborado pela autora (2022)

### 4.4.3 Relação Temperamento e Dados *Instagram*

Após a criação da base de dados composta por atributos que descrevem o usuário na rede social *Instagram* e o atributo alvo, temperamento, a próxima etapa objetivou em verificar se há uma relação entre os atributos extraídos da rede social e o temperamento do usuário. Para esta etapa diferentes abordagens foram usadas. A primeira delas foi o uso de testes estatísticos, que permitam relacionar um temperamento com os diferentes atributos que descrevem um usuário. A segunda abordagem foi a análise de gráficos *boxplot*, que permitem relacionar o temperamento com os diferentes atributos que descrevem o usuário observando as variações dentro de um mesmo atributo. Por último, a árvore de decisão foi usada como um modelo descritivo que permita entender quais atributos podem ser usados para descrever um dado temperamento.

Os testes estatísticos foram utilizados para responder as seguintes questões: Q1) um determinado temperamento utiliza *emojis* com frequência em suas postagens? Q2) Existe diferença na quantidade média de curtidas das publicações entre um temperamento e outro? Q3) Quais temperamentos publicam mais postagens positivas?

Os testes utilizados foram o Teste t de *Student* (KIM, 2015) e o Teste Exato de Fisher (FISHER, 1922) ao nível de 0,05 de significância. O Teste t de *Student* foi utilizado porque ele avalia a diferença da média de duas amostras. O Teste Exato de Fisher foi escolhido porque ele compara a proporção de duas amostras e também porque o tamanho das amostras analisadas eram pequenas e independentes.

Os gráficos foram utilizados para identificar o comportamento de um determinado temperamento em relação a um atributo. Os gráficos *boxplot* foram escolhidos porque eles resumem a distribuição dos dados da base de dados, indicando os valores máximo

e mínimo, mediana e quartis (POTTER et al., 2006). Na montagem dos gráficos foram utilizadas as bibliotecas GGLOT2<sup>6</sup> e RPART<sup>7</sup> da linguagem R. Essa análise dos dados tem o objetivo de responder as seguintes questões: Q4) Qual é o comportamento dos temperamentos em relação ao uso de *emojis*? Q5) Qual o comportamento dos temperamentos em relação a publicações nos finais de semana e durante a semana? Q6) Qual é o comportamento dos temperamentos em relação a publicação de postagens com sentimento positivo? Q7) Qual é o comportamento dos temperamentos em relação à quantidade de publicações? Q8) E em relação à média de curtidas das publicações?

A árvore de decisão, além de ser um algoritmo de classificação, pode ser utilizada como um modelo descritivo, já que o caminho da árvore mostra quais atributos e valores foram usados para classificar as instâncias de uma determinada classe. Dado isso, a árvore de decisão foi utilizada para entender quais atributos estão envolvidos na classificação de um dado temperamento. A árvore de decisão foi montada utilizando uma biblioteca do SKLEARN do Python<sup>8</sup>. Na construção da árvore foi escolhido o índice de GINI e os demais parâmetros da árvore foram os padrões da biblioteca.

## 4.5 Considerações Finais

Este capítulo teve o objetivo de apresentar a proposta desta dissertação de mestrado. Este trabalho propõe a construção de um *framework* para extração de sentimentos das publicações da rede social *Instagram*, não sendo limitado apenas a esta rede social. Esse *framework* possui as etapas de pré-processamento, representação dos dados, classificação dos dados, métodos e medidas de avaliação. Além disso, foram descritas as análises que fazem o mapeamento do temperamento dos usuários, representação dos usuários a partir dos dados coletados do *Instagram* e a relação dos dados do *Instagram* com o temperamento por meio de testes estatísticos, gráficos e árvore de decisão. O próximo capítulo apresenta os experimentos realizados e a análise dos resultados.

---

<sup>6</sup> <https://ggplot2.tidyverse.org/>

<sup>7</sup> <https://www.rdocumentation.org/packages/rpart/versions/4.1.16/topics/rpart>

<sup>8</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>

---

## Experimentos e Análise dos Resultados

Este capítulo descreve os experimentos executados neste trabalho e seus resultados. A Seção 5.1 detalha as bases de dados utilizadas. A Seção 5.2 descreve sobre a qualidade das estratégias de classificação utilizadas neste trabalho, em contrapartida, com os classificadores utilizados na literatura. E por fim, a Seção 5.3 apresenta os resultados da relação dos dados do usuário do *Instagram* com o temperamento.

### 5.1 Base de Dados

Esta Seção tem o objetivo de descrever as bases de dados utilizadas no trabalho. A Subseção 5.1.1 descreve sobre as bases utilizadas e classificadas com os rótulos, positivo, negativo e neutro e a Subseção 5.1.2 descreve sobre a base de temperamento.

#### 5.1.1 Bases de Sentimentos

Para os experimentos foram utilizadas três bases de dados diferentes com dados em português, sendo duas delas do *Instagram* e uma do *Twitter*. A validação das estratégias de classificação foram feitas utilizando as três bases de dados, já que as estratégias utilizadas neste trabalho podem ser usadas em dados de diferentes redes sociais.

A Base de Dados I (B1) possui um total de 925 publicações de 6 usuários selecionados aleatoriamente, que possuíam contas públicas no *Instagram*. Essa base contém 31,14% de legendas classificadas como negativas, 59,24% de positivas e 9,62% de neutras.

A Base de Dados II (B2) é pública<sup>1</sup> e contém 2.787 *tweets* rotulados com os sentimentos neutro, negativo e positivo. Neste trabalho, foram utilizados 1.656 *tweets*, devido à limitação da API do *Twitter* ao recuperar os *tweets* relacionados ao ID presente na base, sendo 31,22% negativos, 30,56% positivos e 38,22% neutros.

Por fim, a Base de Dados III (B3) é composta por 1.028 publicações de 57 usuários do *Instagram*, os quais aceitaram participar da pesquisa, possuíam conta pública e

---

<sup>1</sup> <https://github.com/arialab/tash-pt>

responderam o questionário de temperamento. Na rotulação dessa base foram selecionados 20 publicações de cada usuário de forma aleatória. Existiam usuários que possuíam menos que 20 publicações, então nesses casos, todas as publicações do usuário foram selecionadas. A base de dados rotulada é composta por 523 publicações, representando aproximadamente 50,87% da base total. Os sentimentos usados são positivo, negativo e neutro, rotulados por alunos de IC da Faculdade de Computação da Universidade Federal de Uberlândia. As legendas rotuladas como inconclusivas foram retiradas da análise, portanto a base de dados final contém 495 legendas. Essa base possui 6,87% legendas classificadas como negativas, 59,8% positivas e 33,33% neutras.

### 5.1.2 Base de Temperamentos

Para a análise da relação entre temperamento e dados de redes sociais, somente a base de dados B3 foi utilizada, já que ela é a única que possui o temperamento de um conjunto de usuários e suas respectivas postagens na rede social *Instagram*. A Figura 19 mostra os temperamentos identificados e a quantidade de usuários para cada temperamento.

	<b>Temperamento</b>	<b>Quantidade Usuários</b>
1	ansioso	2
2	ansioso,ciclotímico,depressivo	1
3	ansioso,ciclotímico,depressivo,preocupado	1
4	ansioso,depressivo	2
5	ansioso,hipertímico	1
6	ciclotímico,depressivo	2
7	ciclotímico,depressivo,irritável	1
8	ciclotímico,depressivo,irritável,preocupado	2
9	ciclotímico,irritável	1
10	depressivo	9
11	depressivo,hipertímico	1
12	depressivo,irritável	6
13	depressivo,preocupado	3
14	hipertímico	11
15	hipertímico,irritável	3
16	irritável	9
17	irritável,preocupado	1
18	preocupado	1
<b>Total Usuários</b>		<b>57</b>

Figura 19 – Relação de Temperamentos e Usuários por Temperamento. Fonte: Elaborado pela autora (2022)

O questionário teve respostas de 57 usuários, sendo 32 usuários com temperamentos puros. Dos 32 usuários, 2 possuem o temperamento ansioso, 9 depressivo, 10 hipertímico, 9 irritável e 1 preocupado. Nenhum usuário foi identificado apenas com o temperamento



ciclotímico. Os outros 25 usuários possuem temperamentos mistos, ou seja, o mesmo usuário tem características de dois ou mais temperamentos. Isso acontece quando durante mapeamento do temperamento do usuário, os cálculos mencionados na Seção 2.1.2 dão empate. Esses usuários com temperamentos mistos foram transformados em usuários com temperamentos puros, considerando o usuário em todos os temperamentos que houveram empate, conforme ilustra a Figura 20. Após esse tratamento dos dados foram obtidos 7 instâncias de usuários com temperamento ansioso, 8 com ciclotímico, 28 com depressivo, 16 com hipertímico, 23 com irritável e 8 com preocupado, totalizando 90 instâncias com temperamentos puros.

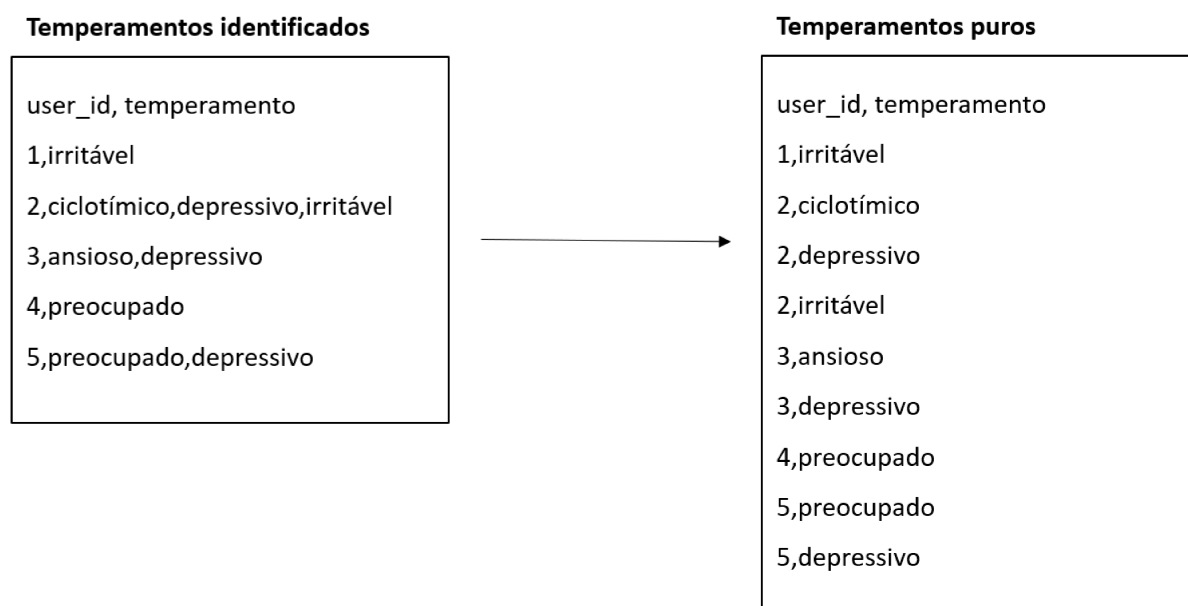


Figura 20 – De-para Temperamentos. Fonte: Elaborado pela autora (2022)

## 5.2 Avaliação do *Framework* de Extração de Sentimentos

Esta Seção apresenta a avaliação do framework de extração de sentimento desenvolvido neste trabalho. A Seção está dividida em duas partes. A primeira detalha os resultados do pré-processamento e representação dos dados e a segunda os resultados dos classificadores e das estratégias de classificação utilizadas neste trabalho.

### 5.2.1 Pré-processamento e Representação dos Dados

Em todas as bases de dados foram aplicados pré-processamentos. Conforme descrito na Subseção 4.3.1, os pré-processamentos aplicados nas legendas foram substituição de *emojis*

por palavras, padronização das palavras para minúsculo e remoção de *stopwords*, URLs, e-mails, marcações de usuários, números e acentos. A Figura 21 mostra um exemplo de legenda, onde a primeira frase é a legenda original (como o usuário publicou na rede) e a segunda frase é como a legenda ficou após a aplicação das técnicas de pré-processamento.

**Legenda Original**

RT @ancalado: Quando você já tem cartão amarelo e não pode ser expulso !!!!! 😁  
<https://t.co/huufkfKFXh>

**Legenda Pré-processada**

rt cartao amarelo nao pode expulso  
 !!!!! rosto contente olhos sorridentes

Figura 21 – Pré-processamento de Legendas. Fonte: Elaborado pela autora (2022)

Após o pré-processamento foram aplicados os métodos TDF-IDF e N-gramas para a preparação do texto para o classificador. As bases de dados após essa etapa ficaram com a seguinte quantidade de atributos i) B1: 8.281 atributos ii) B2: 18.687 atributos iii) B3: 8.215 atributos.

O resumo dos dados das bases de dados B1, B2 e B3 estão apresentados na tabela 5.

Tabela 5 – Resumo Bases de Dados

Base Dados	Qtd Instâncias	Qtd Positivo	Qtd Negativo	Qtd Neutro	Qtd Atributos
B1	925	548	288	89	8.281
B2	1.656	506	517	633	18.687
B3	495	296	34	165	8.215

### 5.2.2 Avaliação dos Resultados dos Classificadores

Esta Seção apresenta os resultados do classificador de sentimentos proposto neste trabalho.

A Tabela 6 apresenta os resultados de acurácia macro para a classificação de sentimentos. Nela aparecem os resultados dos algoritmos comumente usados na literatura para a classificação de sentimentos e das duas estratégias avaliadas neste trabalho, classificação em níveis e classificação *ensemble*. O modelo de classificação *ensemble* teve melhor desempenho nas bases B1, B2 e B3, com acurácia de 86,16%, 80,01% e 89,1%, respectivamente. Os algoritmos utilizados no comitê (*Naive Bayes*, *Random Forest* e *SVM*) também foram

avaliados separadamente, porém se comparados ao comitê e ao classificador em níveis, os resultados não são satisfatórios.

Tabela 6 – Comparação da acurácia dos diferentes modelos de classificação nas três bases de dados

Base de Dados	Naive Bayes	SVM	Random Forest	Classif. em Níveis	Classif. <i>Ensemble</i>
B1	62,36	67,69	62,35	84,65	<b>86,16</b>
B2	38,94	39,30	38,39	77,47	<b>80,01</b>
B3	63,41	72,56	62,80	84,65	<b>89,1</b>

Considerando que as duas estratégias utilizadas neste trabalho, classificação em níveis e *ensemble*, tiverem os melhores resultados, elas serão analisadas em mais detalhes a seguir. A Tabela 7 compara os resultados das medidas de avaliação Acurácia e F1-score para as duas estratégias. Pode-se observar que tanto o *F1-Score* quanto a acurácia para as três bases de dados analisadas tiveram resultados satisfatórios para classificação de sentimento.

Tabela 7 – Medidas de Avaliação: Classificador *Ensemble* e Classificador em Níveis

Base	Classif. <i>Ensemble</i>		Classif. em Níveis	
	Acurácia	<i>F1-Score</i>	Acurácia	<i>F1-Score</i>
B1	86,16	79,39	84,65	76,98
B2	80,01	80,18	77,47	77,43
B3	89,1	78,48	84,65	74,91

A Tabela 8 detalha os resultados da estratégia de classificação em níveis. Ela apresenta os resultados de cada nível dessa estratégia e entre parenteses a porcentagem da base de dados que foi rotulada naquele nível. O classificador em níveis para a base de dados B1 classificou 18,05% da base no nível 1 (N1), 1,19% no nível 2 (N2) e 80,76% no nível 3 (N3). Na base B2 7,55% dos dados foram classificados no nível 1, 3,38% no nível 2 e 89,07% no nível 3. E por último, a base de dados B3 47,27% dos dados foram classificados no nível 1, 2,22% no nível 2 e 50,51% no nível 3.

Tabela 8 – Medidas de Avaliação por Nível do Modelo de Classificação em Níveis

Base/ Medida/ Nível	Classificação em Níveis					
	Acurácia			<i>F1-Score</i>		
	N1	N2	N3	N1	N2	N3
B1	85,63 (18,05)	54,54 (1,19)	85 (80,76)	48,83	35,30	79,84
B2	55,2 (7,55)	67,86 (3,38)	79,73 (89,07)	42,74	48,05	79,53
B3	87,18 (47,27)	81,82 (2,22)	82,4 (50,51)	45,63	45,00	75,78

Pode-se observar que o primeiro nível de classificação teve bons resultados de acurácia nas bases B1 e B3 e em relação a base B2 o resultado pode ser melhorado. Apesar de a acurácia nas bases B1 e B3 estarem altas, o *F1-Score* está baixo. Isso aconteceu

porque esse nível acerta o rótulo de muitas legendas com sentimento positivo e erra na classificação das legendas com sentimento negativo. A tabela 9 apresenta a matriz de confusão da classificação em níveis no N1. Além disso, dessas bases analisadas, vê-se que os usuários do *Instagram* usam mais *emojis* em suas publicações do que os usuários do *Twitter*. Ademais, o N1 é uma importante etapa na classificação de texto que possuem *emojis*, já que esses, em grande parte das vezes, expressam o sentimento da sentença. Porém, existem trabalhos que estudaram que nem sempre o *emoji* por si só representa o sentimento de toda a frase (CHEN et al., 2018). Um usuário pode colocar um *emoji* com sentimento negativo na legenda e a frase possuir sentimento positivo, por exemplo. Esse fato, pode explicar o motivo do N1 na base B2 ter acertado apenas 55,2% do sentimento dos tweets, já que esse nível se baseia apenas no sentimento dos emojis para a classificação.

O segundo nível teve bons resultados de acurácia para a base B3 e em relação as outras duas bases há oportunidades de melhoria. Na base de dados B1 1,19%, na B2 3,38% e na B3 2,22% da base foi classificada no N2. Isto significa que pouquíssimas legendas/*tweets* foram classificadas nesse nível do classificador em níveis. Uma oportunidade de melhoria nesse nível é aumentar o dicionário de palavras que expressam um sentimento por si só.

O nível três nas bases B1, B2 e B3 tiveram bons resultados tanto em relação à acurácia quanto em relação ao *F1-Score*, com valores de 79,84%, 79,53% e 75,78%, respectivamente. A maioria das legendas da B1 e B2 foram classificadas no último nível do classificador em níveis. Já na B3, um pouco mais da metade da base foi classificada no N3.

A base de dados B3, foi rotulada conforme o manual de rotulação (Apêndice B) criado para esse trabalho. Esse manual tem o mesmo passo a passo de rotulação da legenda utilizado no classificador em níveis, portanto, os rotuladores primeiro olhavam o sentimento dos *emojis*, caso não fosse possível identificar o sentimento por meio do *emoji*, era analisado a presença de palavras-chave que identificam o sentimento da legenda, caso ainda não fosse possível identificar o sentimento uma análise geral era feita para identificar o sentimento da legenda. Além disso, três rotuladores foram selecionados, sendo dois rotulando a mesma legenda e o terceiro rotulando apenas os casos de divergência. Pode-se observar que na base B3 o classificador em níveis teve resultados parecidos com a rotulação manual. Dado isso, ressalta-se a importância da qualidade da rotulação da base de dados que o classificador de sentimentos usa no aprendizado.

Para mais detalhes dos resultados da classificação em níveis, a matriz de confusão das bases de dados B1, B2 e B3 estão apresentadas na tabela 10. Na B1 o modelo de classificação em níveis acertou 72,92% do rótulo da classe negativo, 44,94% da classe neutro e 97,44% da classe positivo. Na B2 o modelo de classificação em níveis acertou o rótulo de 72,34% das legendas pertencentes a classe negativo. Para a classe neutro, o modelo acertou 87,83%. Por fim, para a classe positivo, o classificador teve 69,76% de taxa de acerto. Na base de dados B3 o modelo de classificação em níveis acertou o rótulo de 94,59% das legendas da classe positivo, 47,06% da classe negativo e 74,54% da classe

neutro. Pode-se observar que o *Instagram* possui mais legendas com sentimento positivo, logo, o classificador aprendeu mais sobre o sentimento positivo. Já o *Twitter* tende a ser uma rede social de manifestações, reclamações e protestos, sendo assim existem muitos *tweets* com sentimentos negativos na rede social e o classificador consegue aprender mais sobre o sentimento negativo.

O modelo de classificação *ensemble* trouxe os melhores resultados nas três bases de dados que possuem legendas retiradas das redes sociais *Instagram* e *Twitter* com acurácia superior a 80%. Comparando o modelo de classificação *ensemble* e o modelo de classificação em níveis com os algoritmos de classificação mais utilizados na literatura (*Naive Bayes*, *Random Forest* e *SVM*), estes obtiveram os melhores resultados nas três bases de dados.

Para mais detalhes dos resultados da classificação *ensemble* a matriz de confusão das bases de dados B1, B2 e B3 estão apresentadas na tabela 11. Na B1 o modelo de classificação *ensemble* acertou 75% do rótulo da classe negativo, 49,44% da classe neutro e 97,99% da classe positivo. Na B2 o modelo de classificação *ensemble* acertou o rótulo de 69,63% das legendas pertencentes a classe negativo, sendo que a maior parte das legendas classificadas erroneamente foram para a classe neutro. Para a classe neutro, o modelo acertou 93,68%. Por fim, para a classe positivo, o classificador teve 73,52% de taxa de acerto, sendo que a maior parte das legendas classificadas incorretamente foram para a classe neutro. Na base de dados B3 o modelo de classificação *ensemble* acertou o rótulo de 96,96% das legendas da classe positivo, 38,23% da classe negativo e 85,45% da classe neutro.

Observando as tabelas da matriz de confusão dos modelos de classificação *ensemble* e em níveis, para a base B1 nota-se que o classificador *ensemble* teve melhor desempenho de classificação em todas as classes (positivo, negativo e neutro). Já para as bases B2 e B3 o classificador *ensemble* se sobressaiu na classificação das classes positivo e neutro e o classificador em níveis se destacou na classe negativo.

Tabela 9 – Matriz de Confusão Classificação em Níveis - Nível 1 - B1, B2, B3

Matriz de Confusão - Classificação em Níveis - Nível 1												
Classe Atual	Classe Predita											
	B1				B2				B3			
	Negativo	Neutro	Positivo	Total	Negativo	Neutro	Positivo	Total	Negativo	Neutro	Positivo	Total
Negativo	<b>6</b>	0	5	11	<b>27</b>	0	9	36	<b>5</b>	0	4	9
Neutro	0	<b>0</b>	14	14	13	<b>0</b>	23	36	2	<b>0</b>	17	19
Positivo	5	0	<b>137</b>	142	11	0	<b>42</b>	53	7	0	<b>199</b>	206
Total	11	0	156	<b>167</b>	51	0	74	<b>125</b>	14	0	220	<b>234</b>

Tabela 10 – Matriz de Confusão Classificação em Níveis - B1, B2 e B3

Matriz de Confusão - Classificação em Níveis												
Classe Atual	Classe Predita											
	B1				B2				B3			
	Negativo	Neutro	Positivo	Total	Negativo	Neutro	Positivo	Total	Negativo	Neutro	Positivo	Total
Negativo	<b>209</b>	2	77	288	<b>374</b>	122	21	517	<b>16</b>	5	13	34
Neutro	4	<b>40</b>	45	89	33	<b>556</b>	44	633	2	<b>123</b>	40	165
Positivo	14	0	<b>534</b>	548	32	121	<b>353</b>	506	7	9	<b>280</b>	296
Total	227	42	656	<b>925</b>	439	799	418	<b>1656</b>	25	137	333	<b>495</b>

Tabela 11 – Matriz de Confusão Classificação *Ensemble* - B1, B2 e B3

Matriz de Confusão - Classificação Ensemble												
Classe Atual	Classe Preditada											
	B1				B2				B3			
	Negativo	Neutro	Positivo	Total	Negativo	Neutro	Positivo	Total	Negativo	Neutro	Positivo	Total
Negativo	<b>215</b>	2	71	288	<b>360</b>	142	15	517	<b>13</b>	6	15	34
Neutro	4	<b>44</b>	41	89	24	<b>593</b>	16	633	0	<b>141</b>	24	165
Positivo	10	0	<b>538</b>	548	5	129	<b>372</b>	506	0	9	<b>287</b>	296
Total	229	46	650	<b>925</b>	389	864	403	<b>1656</b>	13	156	326	<b>495</b>

## 5.3 Relação Entre Dados do Usuário no *Instagram* e seu Temperamento

Essa Seção apresenta os modelos utilizados na identificação dos padrões e a relação entre os dados do usuário no *Instagram* e seu temperamento. Essa tarefa foi realizada apenas na base B3, porque apenas ela possui os dados do *Instagram* de usuários que responderam o questionário de temperamento. Além disso, apenas os de sentimentos das legendas que foram rotulados manualmente foram considerados nas análises.

### 5.3.1 Testes Estatísticos

Os testes estatísticos analisaram a relação entre o temperamento do usuário e a quantidade de curtidas, proporção de publicações com *emoji* e proporção de publicações positivas e negativas. Nesse método de relação de dados foi utilizado apenas os 32 usuários com temperamentos puros devido à alta complexidade de interpretação dos dados com muitos níveis.

Para relacionar a quantidade de curtidas das publicações do *Instagram* e o temperamento foi utilizado o Teste  $t$  de *Student*, veja Tabela 12. Essa tabela apresenta as hipóteses  $H_0$  e  $H_1$ , o valor de  $\mu$  e de  $t$  calculados conforme a métrica analisada, que neste caso é a quantidade de curtidas. Por fim, apresenta se a hipótese  $H_0$  foi aceita pelo teste estatístico. Os resultados indicaram que os usuários ansiosos (A) possuem mais curtidas em suas publicações do que usuários depressivos (D), hipertímicos (H), irritados (I) e preocupados (P). Além disso, o teste mostrou que o temperamento depressivo possui mais curtidas que o temperamento preocupado e que o temperamento hipertímico tem mais curtidas que os depressivos.

O Teste Exato de Fisher foi utilizado para relacionar o temperamento do usuário e os sentimentos expressos na rede social *Instagram*, veja Tabela 13. Essa tabela apresenta as métricas analisadas, as hipóteses  $H_0$  e  $H_1$ , o valor de  $p$  e de  $t$  calculados conforme a métrica analisada, que neste caso são proporção de *emojis*, proporção de sentimento positivo e proporção de sentimento negativo. Por fim, apresenta se a hipótese  $H_0$  foi aceita pelo teste estatístico. Os resultados indicaram que usuários depressivos publicam mais legendas com sentimento positivo do que usuários hipertímicos, irritados e preocupados e que usuários hipertímicos postam mais legendas com sentimento negativo que usuários depressivos. Além disso, os resultados indicaram que o temperamento ansioso usa mais *emojis* em suas legendas do *Instagram* do que o temperamento depressivo e irritado e que o temperamento hipertímico usa mais *emojis* do que o depressivo.



Tabela 12 – Resultados do Teste t de Student entre temperamento e quantidade de curtidas nas publicações

	Métrica	H0	H1	$\mu$	t	Aceita H0
<b>Teste t de Student</b>	Qtd Curtidas	A=D	A>D	0,000	6,530	não
		A=H	A>H	0,000	3,741	não
		A=I	A>I	0,000	3,986	não
		A=P	A>P	0,035	1,882	não
		D=P	D>P	0,041	1,762	não
		H=D	H>D	0,013	2,254	não

Tabela 13 – Resultados do Teste Exato de Fisher entre temperamento e dados da rede social

	Métrica	H0	H1	p	t	Aceita H0
<b>Teste Exato de Fisher</b>	Proporção Emojis	A=D	A>D	0,012	3,472	não
		A=I	A>I	0,037	2,875	não
		H=D	H>D	0,037	1,887	não
	Proporção Sent. Positivo	D=H	D>H	0,001	3,040	não
		D=I	D>I	0,016	2,095	não
		D=P	D>P	0,028	inf	não
	Proporção Sent. Negativo	A=D	A>D	0,229	3,500	sim
		A=H	A>H	0,750	0,769	sim
		H=D	H>D	0,047	4,550	não

### 5.3.2 Gráficos

Para identificar o comportamento de cada temperamento em relação às medidas extraídas na rede social *Instagram* foi utilizado os gráficos de *boxplot*. As medidas foram quantidade de posts, média de curtidas recebidas, proporção de *emoji*, proporção de publicações positivas, proporção de publicações durante a semana e final de semana e proporção de seguidos e seguidores. Nessa análise foram considerados 90 usuários, sendo os usuários com temperamentos puros e usuários com temperamentos mistos transformados em temperamentos puros, conforme descrito na Subseção 5.1.2.

O gráfico apresentado na Figura 22 mostra o comportamento de cada temperamento em relação ao uso de *emojis* nas publicações no *Instagram*. Pode-se observar que em todos os temperamentos a variabilidade do uso de *emojis* nas publicações é grande. Além disso, pelo menos 25% dos usuários ansiosos utilizam *emojis* em aproximadamente 50% de suas publicações. Pelo menos a metade dos preocupados utilizam *emojis* em quase todas, se não todas, suas publicações na rede.

O gráfico apresentado na Figura 23 mostra o comportamento de cada temperamento em relação à proporção de publicações que expressam sentimento positivo. Podemos observar que todos os usuários ansiosos possuem no mínimo a metade de suas publicações com sentimento positivo. Já os hipertímicos são os que menos publicam legendas com sentimento positivo no *Instagram*. 75% dos usuários ciclotímicos e depressivos possuem

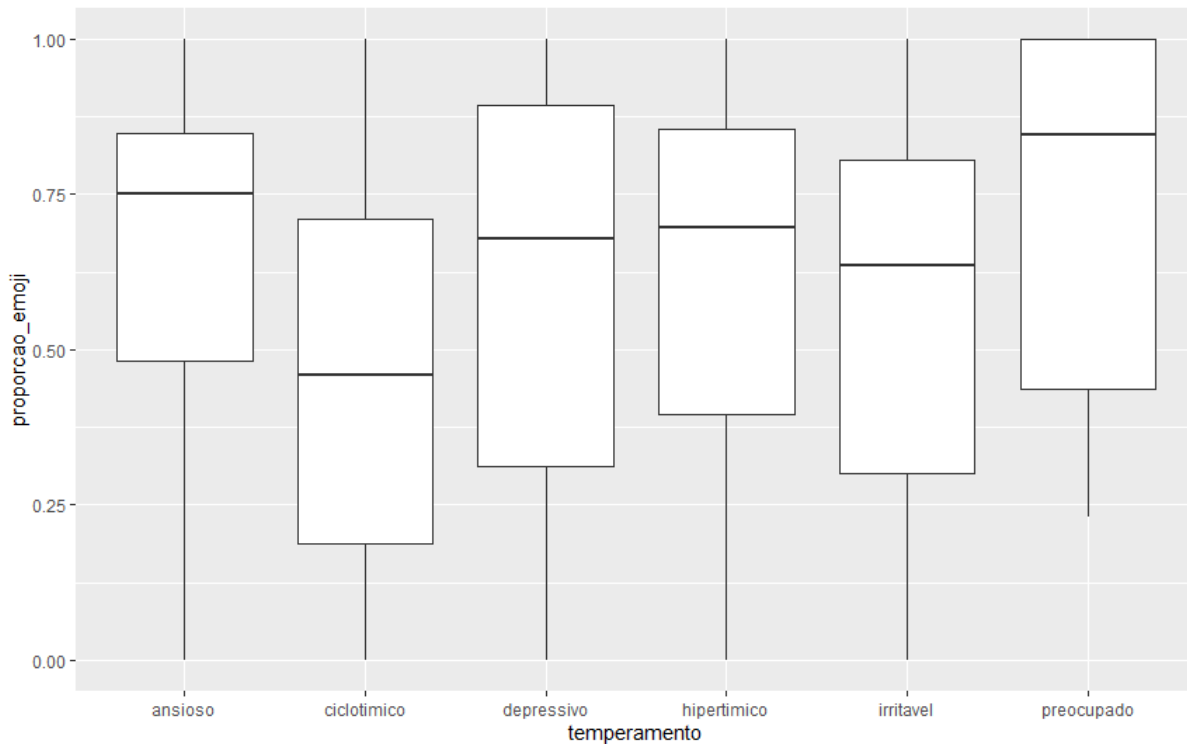


Figura 22 – Gráfico *Emojis* e Temperamento. Fonte: Elaborado pela autora (2022)

pelo menos a metade das legendas com sentimento positivo em seu perfil. Outro ponto de análise é que os usuários depressivos publicam mais legendas com sentimento positivo do que os hipertímicos, preocupados e irritados.

O gráfico apresentado na Figura 24 mostra o comportamento de cada temperamento em relação à quantidade de publicações de cada usuário. Podemos observar que a variabilidade da quantidade de publicações nos usuários com temperamento preocupado é pequena. Os temperamentos ciclotímico, depressivo e irritável apresentaram alguns desvios de quantidade de publicações em relação à maioria dos usuários da amostra analisada. Esse desvio é conhecido como *outlier*. 75% dos usuários hipertímicos da amostra publicaram no máximo 25 posts.

O gráfico apresentado na Figura 25 mostra o comportamento de cada temperamento em relação à média de curtidas nas publicações no *Instagram*. Pode-se observar que os temperamentos ansioso, depressivo, hipertímico e irritável possuem *outliers* nessa medida. Com exceção do temperamento preocupado, os demais não possuem muita variabilidade na média de curtidas. A mediana dos temperamentos ciclotímico, depressivo e hipertímico são próximas, isso significa que a metade dos usuários desse temperamento possuem aproximadamente a mesma média de curtidas em suas publicações.

O gráfico apresentado na Figura 26 mostra o comportamento de cada temperamento em relação a novas publicações na rede social durante o final semana. Podemos observar que o comportamento de publicações aos finais de semana é muito variável e pelo menos

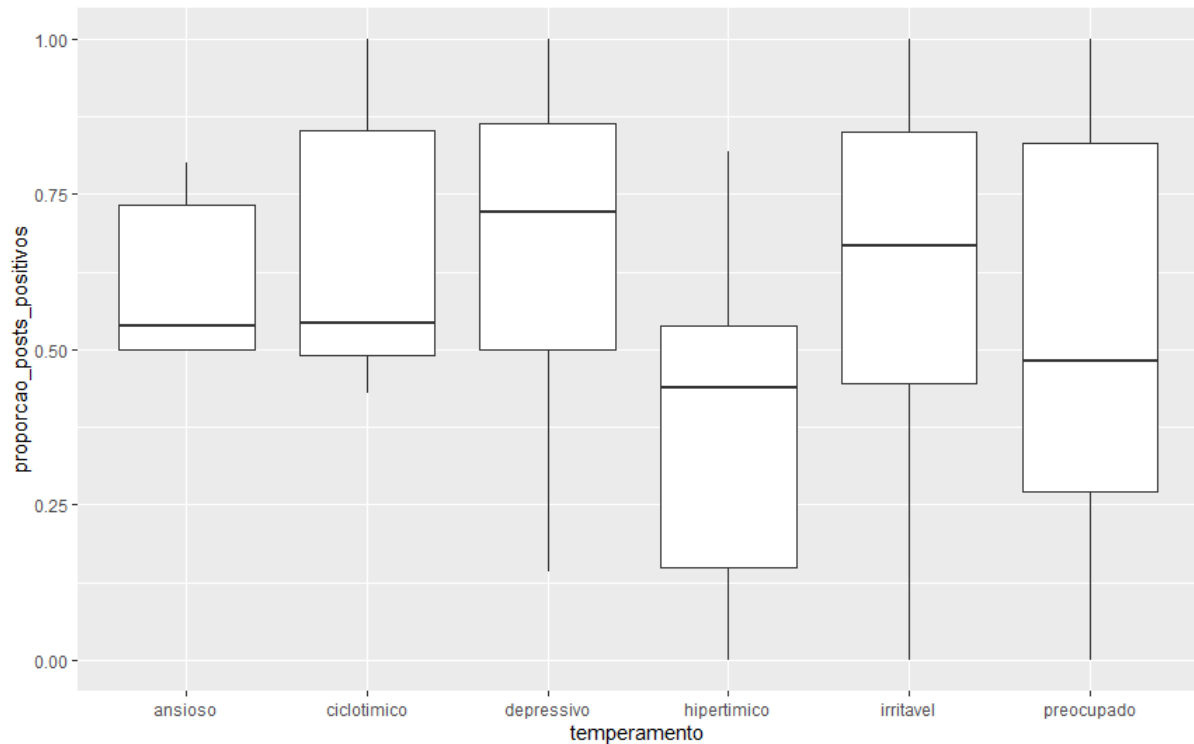


Figura 23 – Gráfico Publicações com Sentimento Positivo e Temperamento. Fonte: Elaborado pela autora (2022)

um usuário de cada temperamento não fez nenhuma publicação no final de semana. Ademais, no máximo 50% dos usuários com temperamento ansioso possuem até 25% das suas postagens publicadas nos finais de semana. Além disso, pelo menos 50% dos ciclotímicos tiveram no mínimo 45% das suas publicações feitas no final de semana. Mais uma observação é que o comportamento de publicações aos finais de semana dos temperamentos irritável e preocupado são similares, assim como o comportamento dos usuários ciclotímicos e depressivos. Por fim, o temperamento que menos publica aos finais de semana é o hipertímico.

O gráfico apresentado na Figura 27 mostra o comportamento de cada temperamento em relação a novas publicações na rede social durante a semana. Podemos observar que o comportamento de publicações durante a semana é variável assim como as publicações aos finais de semana e que todos os temperamentos fizeram pelo menos uma publicação durante a semana. O temperamento hipertímico é o que mais faz postagens no *Instagram* durante a semana. Além disso, todos os temperamentos tiveram pelo menos uma publicação em seu perfil na rede durante a semana. Ademais, o comportamento de publicações aos durante a semana dos temperamentos irritável e preocupado são similares, assim como o comportamento dos usuários ciclotímicos e depressivos. Por fim, pelo menos 75% dos ansiosos possuem 75% de suas publicações no *Instagram* feitas durante a semana.

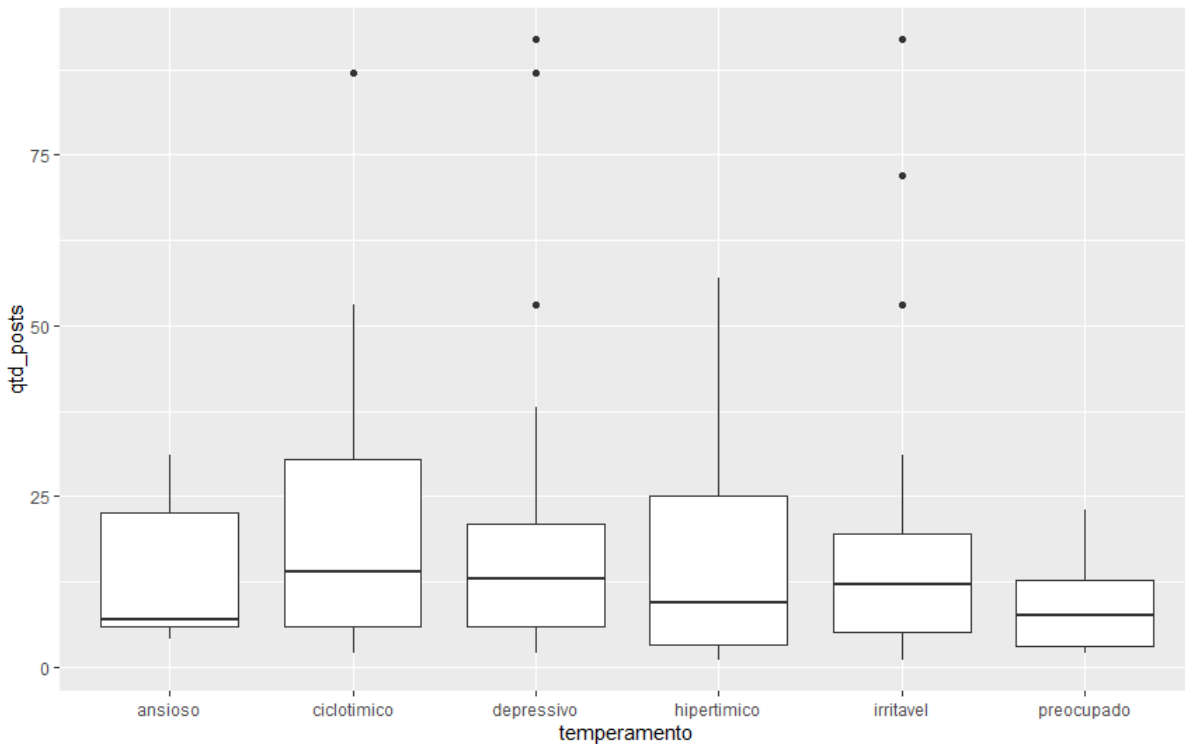


Figura 24 – Gráfico Quantidade de Publicações e Temperamento. Fonte: Elaborado pela autora (2022)

### 5.3.3 Árvore de Decisão

A árvore de decisão foi utilizada para identificar as características de comportamento do usuário na rede social *Instagram* de um determinado temperamento. Nessa análise foram utilizados os 32 usuários com temperamentos puros. Os dados utilizados na relação foram quantidade de posts, média de curtidas, proporção de emoji, proporção de publicações positivas, proporção de publicações durante a semana e final de semana e proporção de seguidos e seguidores.

A Figura 28 mostra a árvore de decisão dos dados dos 32 usuários do *Instagram* que possuem temperamentos puros. Na raiz da árvore, a medida proporção de publicações positivas foi selecionado, e isso significa que ele é um importante atributo para identificação do comportamento de um temperamento. Em relação aos demais atributos, não se pode tirar conclusões já que a árvore apresenta que um mesmo temperamento possui comportamentos diferentes um dos outros e similares a outros temperamentos. Essa árvore de decisão gerada apresenta um estudo preliminar do comportamento de um determinado temperamento com base dos dados da rede social *Instagram*. Mais estudos devem ser desenvolvidos a fim de melhorar o critério de poda da árvore e também é necessário mais usuários de um mesmo temperamento para identificar um padrão de comportamento.

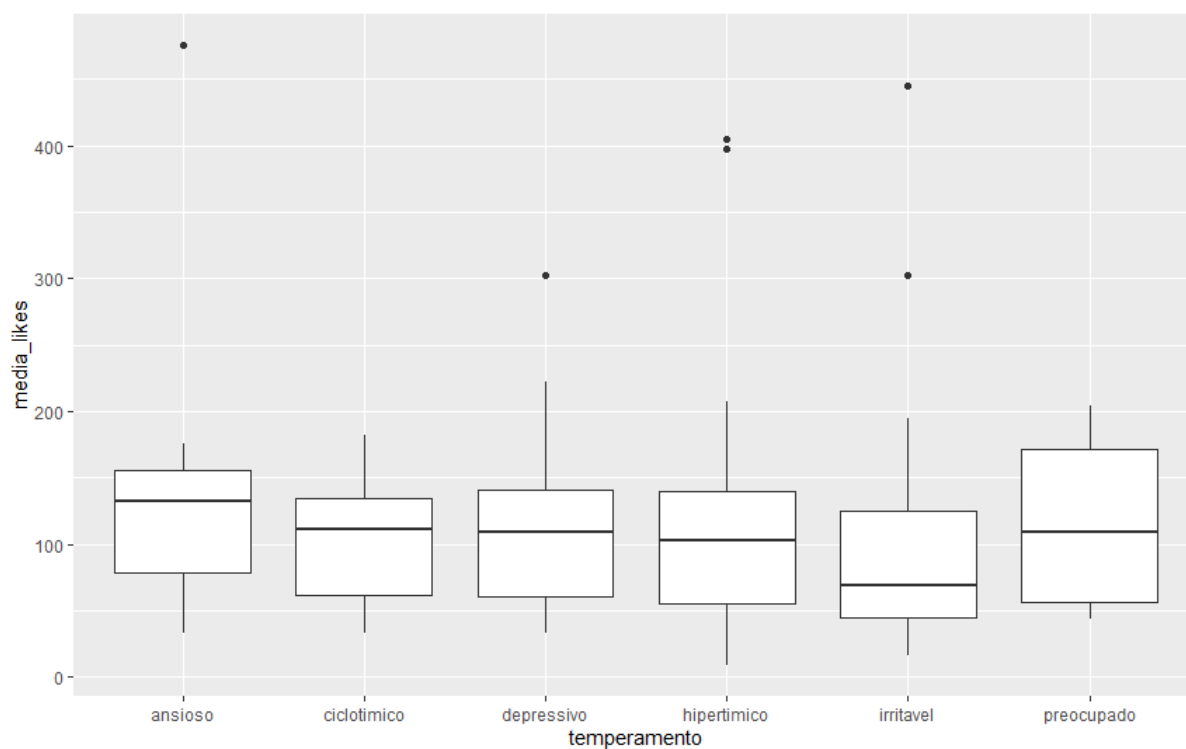


Figura 25 – Gráfico Média de Curtidas e Temperamento. Fonte: Elaborado pela autora (2022)

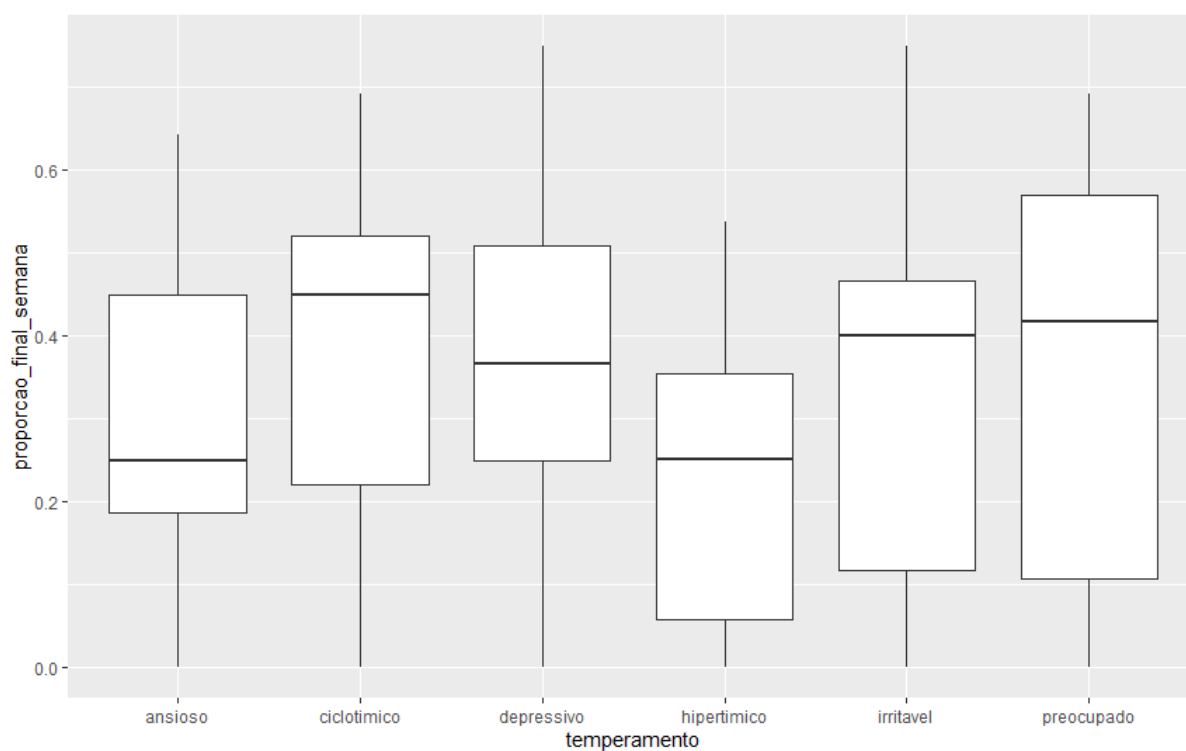


Figura 26 – Gráfico Final de Semana e Temperamento. Fonte: Elaborado pela autora (2022)

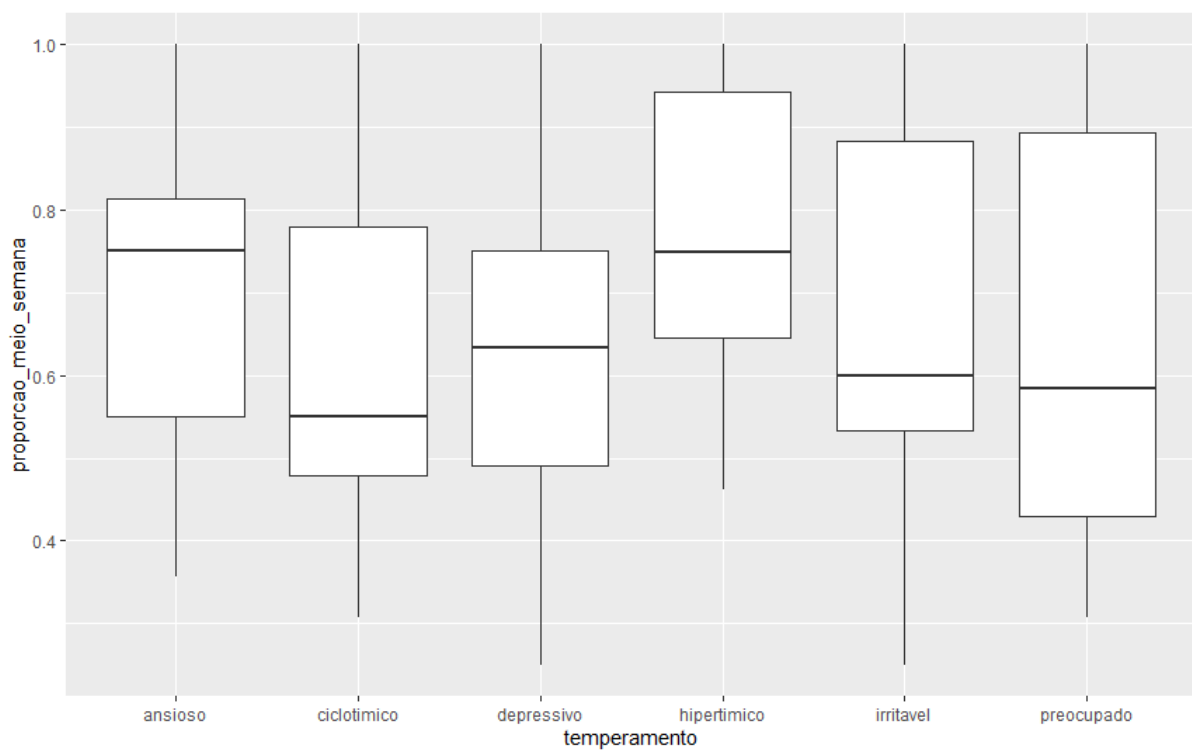


Figura 27 – Gráfico Proporção Publicações Durante a Semana e Temperamento. Fonte: Elaborado pela autora (2022)

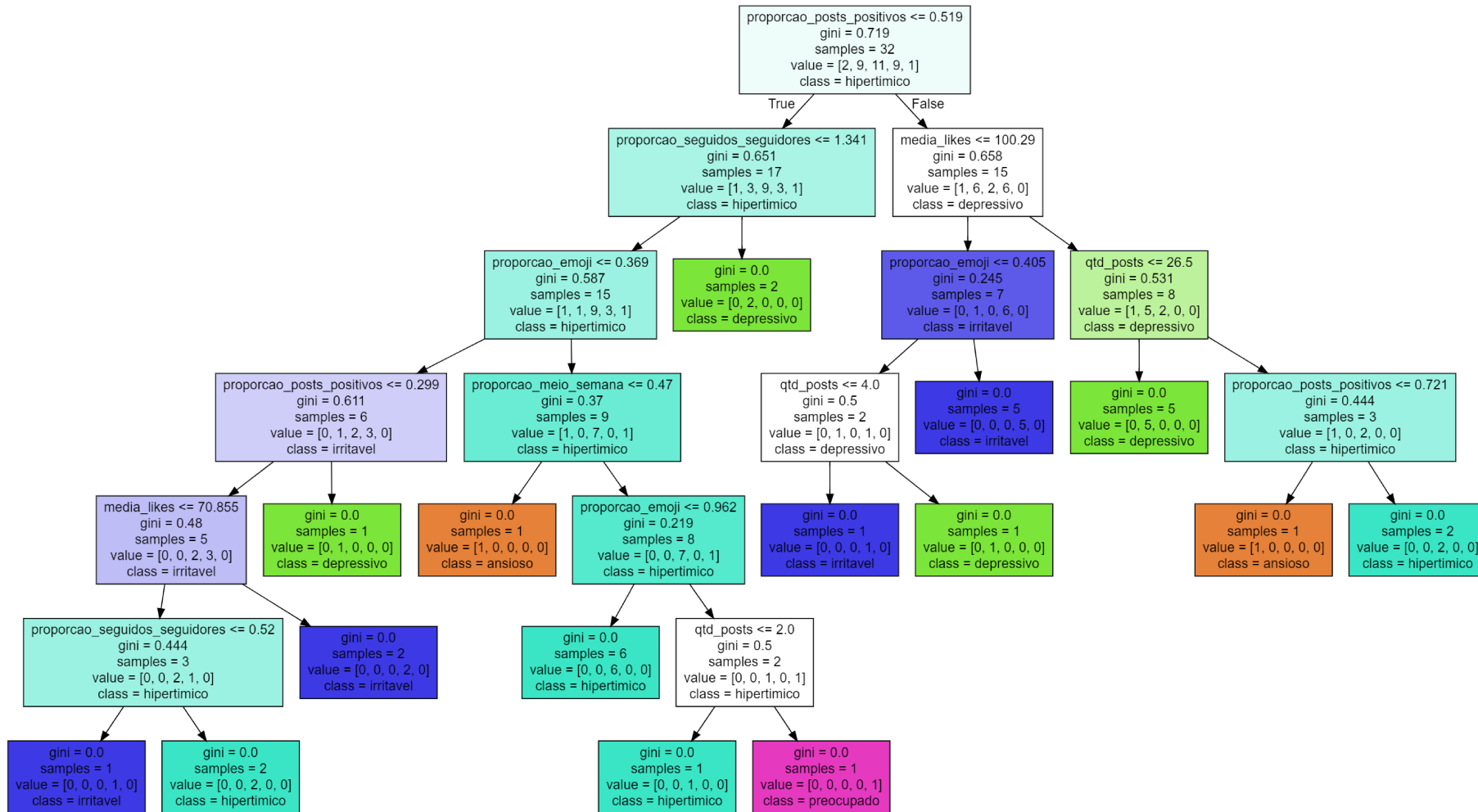


Figura 28 – (A) Árvore de Decisão Temperamento e Dados do *Instagram*. Fonte: Elaborado pela autora (2022)

## 5.4 Considerações Finais

Essa Seção apresentou dos experimentos e as análises dos resultados do método proposto neste trabalho. Primeiro foi descrito os detalhes das bases de dados utilizadas nos experimentos, depois os detalhes do pré-processamento e a representação dos dados, em seguida o resultado dos classificadores de sentimento e por fim, foi mostrado a relação entre os dados do *Instagram* do usuário e seu temperamento por meio dos testes estatísticos, gráficos e árvore de decisão.

O próximo capítulo apresenta a conclusão do trabalho.



---

## Conclusão

O objetivo deste trabalho foi propor e desenvolver um método computacional que permita analisar as postagens e comportamentos dos usuários na rede social *Instagram* contrastando-os com seus temperamentos. Cada usuário foi descrito pelos sentimentos presentes nas suas postagens e por outros dados obtidos a partir da rede social. Estes dados extraídos da rede social foram contrastados com os diferentes temperamentos, a fim de estabelecer relações entre os mesmos.

Neste trabalho foram utilizadas duas estratégias de classificação de sentimentos em redes sociais, a classificação em níveis e a classificação *ensemble*. Essas estratégias foram contrastadas entre si e entre os classificadores comumente utilizados na literatura, *Naive Bayes*, *Random Forest* e SVM. A estratégia de classificação *ensemble* trouxe os melhores resultados nas bases de dados B1, B2 e B3 que possuem legendas retiradas das redes sociais *Instagram* e *Twitter* com acurácia superior a 80%. Os classificadores comumente utilizados na literatura trouxeram resultados bem inferiores as duas estratégias utilizadas neste trabalho, com acurácia máxima de 72,6% para o SVM, 63,4% no *Naive Bayes* e 62,8% no *Random Forest* nas bases de dados analisadas. O classificador em níveis trouxe bons resultados de acurácia em todas as bases de dados com valores superiores a 77%.

Este trabalho também analisou a relação entre os dados do usuário no *Instagram* e seu temperamento. Para tal análise de relações foi utilizado três métodos: i) testes estatísticos, ii) gráficos de *boxplot* e iii) árvore de decisão descritiva.

Os testes estatísticos mostraram que o temperamento ansioso usa mais *emojis* em suas legendas do *Instagram* do que os temperamentos depressivo e irritado. Ademais, os hipertímicos usam mais *emojis* do que os depressivos em suas publicações. Eles também mostraram que os usuários ansiosos possuem mais curtidas em suas publicações do que usuários depressivos, hipertímicos, irritados e preocupados. Por fim, os usuários depressivos publicam mais legendas com sentimento positivo do que usuários hipertímicos, irritados e preocupados e usuários hipertímicos postam mais legendas com sentimento negativo que usuários depressivos.

Os gráficos de *boxplot* apresentaram que o comportamento dos temperamentos em

relação a publicações durante a semana e no final de semana é muito variável dentro do mesmo temperamento e que pelo menos um usuário de cada temperamento não fez nenhuma publicação no final de semana e que todos os temperamentos publicaram pelo menos uma postagem durante a semana. Além disso, o comportamento de publicações durante a semana e aos finais de semana dos temperamentos irritável e preocupado são similares, assim como o comportamento dos usuários ciclotímicos e depressivos. Outra relação apresentada pelos gráficos foi que os usuários ansiosos possuem no mínimo metade de suas publicações com sentimento positivo e que os hipertímicos são os que menos publicam legendas com sentimento positivo no *Instagram*. Assim como os testes estatísticos, os gráficos mostraram que os usuários depressivos publicam mais legendas com sentimento positivo do que os hipertímicos, preocupados e irritados. Em relação à quantidade de publicações, o gráfico mostrou que os temperamentos ciclotímico, depressivo e irritável possuem alguns desvios na quantidade de publicação em relação à maioria dos usuários com esses temperamentos e que 75% dos usuários hipertímicos publicaram no máximo 25 postagens. Por fim, a relação entre a média de curtidas e o temperamento mostrou que os temperamentos ciclotímico, depressivo e irritável possuem aproximadamente a mesma média de curtidas em suas publicações na rede social *Instagram*.

Por último, foi utilizada a árvore de decisão descritiva preliminar que mostrou que o atributo proporção de publicações positivas é um importante atributo na identificação do comportamento de um temperamento.

## 6.1 Contribuições

Esse trabalho teve contribuições as seguintes contribuições:

- Foi desenvolvido um ferramental teórico e prático que auxilia na análise de comportamento na rede social *Instagram*. O uso desse ferramental pode auxiliar a mitigar importantes problemas de saúde pública relacionados a desordens psicológicas no uso de redes sociais. O método desenvolvido foi apresentado no capítulo 4;
- Duas estratégias de classificação de sentimentos em redes sociais que trouxeram resultados superiores aos algoritmos analisados (*Naive Bayes*, *Random Forest* e SVM) os quais são comumente utilizados na literatura. Os experimentos que discutem o desempenho de tais estratégias foram discutidos na seção 5.2.2;
- Discussão de resultados a partir de um estudo experimental transversal retrospectivo que contrasta o comportamento dos usuários na rede social *Instagram* com o temperamento de estudantes dos cursos de Tecnologia da Informação e áreas afins. Os resultados da relação entre temperamento e dados de redes sociais foram discutidos na seção 5.3.

## 6.2 Trabalhos Futuros

Os seguintes itens foram identificados como trabalhos futuros para aperfeiçoar os resultados obtidos neste trabalho:

- ❑ Ampliar da base de dados de publicações do *Instagram* e de temperamento dos usuários do *Instagram*;
- ❑ Selecionar outras bases de dados públicas de redes sociais para experimentação;
- ❑ Aprimorar a estratégia de classificação de sentimentos em níveis, em especial o nível 2, cujo obteve os piores resultados;
- ❑ Adicionar uma nova estratégia de classificação de sentimento baseada em Redes Neurais, como, por exemplo, o BERT;
- ❑ Testar o *ensemble* de outros classificadores;
- ❑ Extrair outros atributos do usuário da rede social *Instagram*, como, por exemplo, a imagem usada na postagem para compor com os dados textuais e o período do dia em que os usuários fazem as publicações;
- ❑ Explorar outros métodos descritivos, como, por exemplo, técnicas de visualização da informação para extrair padrões dos dados;
- ❑ Explorar as técnicas de montagem da árvore de decisão como estratégias de poda, para melhorar o modelo descritivo preliminar que foi avaliado;
- ❑ Avaliar não só o temperamento dominante, que pode ser composto por um ou mais temperamentos (no caso de empate), mas todo o espectro de possíveis temperamentos, que considera aqueles que tenham atingido menor valor no questionário de temperamento.

## 6.3 Contribuições em Produção Bibliográfica

O artigo "*A method for analysis of human temperament in contrast to social network data*" (MARTINS et al., 2022) que apresenta uma estratégia para a análise do relacionamento entre o temperamento dos usuários e os seus dados coletados da rede social *Instagram* foi publicado no XXVIII Simpósio Brasileiro de Sistemas Multimídia e Web 2022.



---

## Referências

AGGARWAL, C. C.; ZHAI, C. **Mining text data**. [S.l.]: Springer Science & Business Media, 2012.

AGUIAR, E. J. de; FAIÇAL, B. S.; UEYAMA, J.; SILVA, G. C.; MENOLLI, A. Análise de sentimento em redes sociais para a língua portuguesa utilizando algoritmos de classificação. In: SBC. **Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**. [S.l.], 2018.

AKISKAL, H. S. Toward a definition of generalized anxiety disorder as an anxious temperament type. **Acta Psychiatrica Scandinavica**, Wiley Online Library, v. 98, p. 66–73, 1998.

AKISKAL, H. S.; AKISKAL, K. K. Temps: Temperament evaluation of memphis, pisa, paris and san diego. **Journal of Affective Disorders**, v. 85, n. 1, p. 1–2, 2005. ISSN 0165-0327. TEMPS: Temperament Evaluation of Memphis, Pisa, Paris and San Diego.

AKISKAL, H. S.; AKISKAL, K. K.; HAYKAL, R. F.; MANNING, J. S.; CONNOR, P. D. Temps-a: progress towards validation of a self-rated clinical version of the temperament evaluation of the memphis, pisa, paris, and san diego autoquestionnaire. **Journal of affective disorders**, Elsevier, v. 85, n. 1-2, p. 3–16, 2005.

AKISKAL, H. S.; DJENDEREDJIAN, A. H.; ROSENTHAL, R. H.; KHANI, M. K. Cyclothymic disorder: validating criteria for inclusion in the bipolar affective group. **The American journal of psychiatry**, American Psychiatric Assn, 1977.

AKISKAL, H. S.; HANTOUCHE, E. G.; BOURGEOIS, M. L.; AZORIN, J.-M.; SECHTER, D.; ALLILAIRE, J.-F.; LANCRENON, S.; FRAUD, J.-P.; CHÂTENET-DUCHÊNE, L. Gender, temperament, and the clinical picture in dysphoric mixed mania: findings from a french national study (epiman). **Journal of affective disorders**, Elsevier, v. 50, n. 2-3, p. 175–186, 1998.

AKISKAL, K. K.; AKISKAL, H. S. The theoretical underpinnings of affective temperaments: implications for evolutionary foundations of bipolar disorder and human nature. **Journal of affective disorders**, Elsevier, v. 85, n. 1-2, p. 231–239, 2005.

AKRAM, W.; KUMAR, R. A study on positive and negative effects of social media on society. **International Journal of Computer Sciences and Engineering**, v. 5, n. 10, p. 347–354, 2017.

- ANGST, J.; GAMMA, A.; AJDACIC, V.; EICH, D.; ROSSLER, W. Prevalence and clinical significance of subthreshold mood disorders. In: CAMBRIDGE UNIV PRESS 40 WEST 20TH ST, NEW YORK, NY 10011-4211 USA. **INTERNATIONAL JOURNAL OF NEUROPSYCHOPHARMACOLOGY**. [S.l.], 2004. v. 7, p. S50–S50.
- BATES, J. E.; GOODNIGHT, J. A.; FITE, J. E. Temperament and emotion. **Handbook of emotions**, Citeseer, v. 3, p. 485–496, 2008.
- BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. **Sociedade Brasileira de Computação**, 2015.
- BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [S.l.]: Springer, 2006. v. 4.
- BOWER, K. M. When to use fisher’s exact test. In: **American Society for Quality, Six Sigma Forum Magazine**. [S.l.: s.n.], 2003. v. 2, n. 4, p. 35–37.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001.
- BRIEGER, P.; MARNEROS, A. Dysthymia and cyclothymia: historical origins and contemporary development. **Journal of Affective Disorders**, Elsevier, v. 45, n. 3, p. 117–126, 1997.
- BUCKLAND, M.; GEY, F. The relationship between recall and precision. **Journal of the American society for information science**, Wiley Online Library, v. 45, n. 1, p. 12–19, 1994.
- BURBIDGE, R.; BUXTON, B. An introduction to support vector machines for data mining. **Keynote papers, young OR12**, Univ. of Nottingham, Operational Research Society, p. 3–15, 2001.
- BURKOV, A. **The hundred-page machine learning book**. [S.l.]: Andriy Burkov Quebec City, QC, Canada, 2019. v. 1.
- CAMPOS, R. **Árvores de Decisão**. 2021. Medium. Disponível em: <<https://medium.com/machine-learning-beyond-deep-learning/%C3%A1rvores-de-decis%C3%A3o-3f52f6420b69>>. Acesso em: 09.07.2021.
- CASAROTTO, C. **TF-IDF: a abordagem de otimização on page que seu blog precisa**. 2019. TF-IDF: a abordagem de otimização on page que seu blog precisa. Disponível em: <<https://rockcontent.com/br/blog/tf-idf/>>. Acesso em: 09.01.2021.
- CHAMBERS, D. **Social media and personal relationships: Online intimacies and networked friendship**. [S.l.]: Springer, 2013.
- CHEN, Y.; YUAN, J.; YOU, Q.; LUO, J. Twitter sentiment analysis via bi-sense emoji embedding and attention-based lstm. In: **Proceedings of the 26th ACM international conference on Multimedia**. [S.l.: s.n.], 2018. p. 117–125.
- CLARO, C. F. et al. Um estudo de caso sobre o modelo de temperamento de keirse. Universidade Presbiteriana Mackenzie, 2018.

CORR, P. J.; MATTHEWS, G. E. **The Cambridge handbook of personality psychology**. [S.l.]: Cambridge University Press, 2020.

CRITCHER, C. Making waves: Historical aspects of public debates about children and mass media. **The international handbook of children, media and culture**, Sage Los Angeles, CA, p. 91–104, 2008.

DEMBIŃSKA-KRAJEWSKA, D.; RYBAKOWSKI, J. The temperament evaluation of memphis, pisa and san diego autoquestionnaire (temps-a)—an important tool to study affective temperaments. **Psychiatr Pol**, Citeseer, v. 48, n. 2, p. 261–76, 2014.

DEVELOPERS, F. for. **Ferramentas para desenvolvedores**. 2021. Ferramentas para desenvolvedores. Disponível em: <<https://developers.facebook.com/>>. Acesso em: 09.01.2021.

\_\_\_\_\_. **Graph API do Instagram**. 2021. Graph API do Instagram. Disponível em: <<https://developers.facebook.com/docs/instagram-api/>>. Acesso em: 09.01.2021.

DEVELOPERS, T. **Publish and analyze Tweets, optimize ads, and create unique customer experiences**. 2021. Publish and analyze Tweets, optimize ads, and create unique customer experiences. Disponível em: <<https://developer.twitter.com/>>. Acesso em: 09.01.2021.

DUARTE, M. d. L. C.; SILVA, D. G. d.; BAGATINI, M. M. C. Enfermagem e saúde mental: uma reflexão em meio à pandemia de coronavírus. **Revista Gaúcha de Enfermagem**, SciELO Brasil, v. 42, 2020.

EFRON, B. Bayes' theorem in the 21st century. **Science**, American Association for the Advancement of Science, v. 340, n. 6137, p. 1177–1178, 2013.

FACEBOOK. **Facebook**. 2021. Facebook. Disponível em: <<https://www.facebook.com/>>. Acesso em: 09.01.2021.

FERWERDA, B.; SCHEDL, M.; TKALCIC, M. Using instagram picture features to predict users' personality. In: SPRINGER. **International Conference on Multimedia Modeling**. [S.l.], 2016. p. 850–861.

FERWERDA, B.; TKALCIC, M. You are what you post: What the content of instagram pictures tells about users' personality. In: CEUR-WS. **The 23rd International on Intelligent User Interfaces, March 7-11, Tokyo, Japan**. [S.l.], 2018.

FISHER, R. A. On the interpretation of  $\chi^2$  from contingency tables, and the calculation of p. **Journal of the royal statistical society**, JSTOR, v. 85, n. 1, p. 87–94, 1922.

GERLACH, M.; SHI, H.; AMARAL, L. A. N. A universal information theoretic approach to the identification of stopwords. **Nature Machine Intelligence**, Nature Publishing Group, v. 1, n. 12, p. 606–612, 2019.

GOIS, C.; AKISKAL, H.; AKISKAL, K.; FIGUEIRA, M. L. The relationship between temperament, diabetes and depression. **Journal of affective disorders**, Elsevier, v. 142, p. S67–S71, 2012.

- GOLBECK, J.; ROBLES, C.; EDMONDSON, M.; TURNER, K. Predicting personality from twitter. In: IEEE. **2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing**. [S.l.], 2011. p. 149–156.
- GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. **Data Mining**. [S.l.]: Elsevier Brasil, 2015.
- GORDON, S. L. The sociology of sentiments and emotion. In: **Social psychology**. [S.l.]: Routledge, 2017.
- GOUTTE, C.; GAUSSIÉ, E. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In: SPRINGER. **European conference on information retrieval**. [S.l.], 2005. p. 345–359.
- GRAHAM, M.; DUTTON, W. H. **Society and the internet: How networks of information and communication are changing our lives**. [S.l.]: Oxford University Press, 2019.
- GUEDES, T. A.; MARTINS, A. B. T.; ACORSI, C. R. L.; JANEIRO, V. Estatística descritiva. **Projeto de ensino aprender fazendo estatística**, Universidade Estadual de Maringá Maringá, p. 1–49, 2005.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. Unsupervised learning. In: **The elements of statistical learning**. [S.l.]: Springer, 2009. p. 485–585.
- HEALTH, R. S. for P. **Social media and young people’s mental health and wellbeing**. 2017. Disponível em: <<https://www.rsph.org.uk/static/uploaded/d125b27c-0b62-41c5-a2e0155a8887cd01.pdf>>. Acesso em: 07.03.2021.
- HOSSEINMARDI, H.; MATTSON, S. A.; RAFIQ, R. I.; HAN, R.; LV, Q.; MISHRA, S. Analyzing labeled cyberbullying incidents on the instagram social network. In: SPRINGER. **International conference on social informatics**. [S.l.], 2015. p. 49–66.
- HOTH, A.; NÜRNBERGER, A.; PAASS, G. A brief survey of text mining. In: CITESEER. **Ldv Forum**. [S.l.], 2005. v. 20, n. 1, p. 19–62.
- INSTAGRAM. **Instagram**. 2021. Instagram. Disponível em: <<https://www.instagram.com/>>. Acesso em: 09.01.2021.
- JIANG, S.; NGIEN, A. The effects of instagram use, social comparison, and self-esteem on social anxiety: A survey study in singapore. **Social Media+ Society**, Sage publications Sage UK: London, England, v. 6, n. 2, p. 2056305120912488, 2020.
- JOSHI, A.; KALE, S.; CHANDEL, S.; PAL, D. K. Likert scale: Explored and explained. **British Journal of Applied Science & Technology**, SCIENCEDOMAIN International, v. 7, n. 4, p. 396, 2015.
- KANSAON, D. P.; BRANDÃO, M. A.; PINTO, S. A. de P. Análise de sentimentos em tweets em português brasileiro. In: SBC. **Anais do VII Brazilian Workshop on Social Network Analysis and Mining**. [S.l.], 2018.



- KARAM, E. G.; MNEIMNEH, Z.; SALAMOUN, M.; AKISKAL, K. K.; AKISKAL, H. S. Psychometric properties of the lebanese–arabic temps-a: a national epidemiologic study. **Journal of Affective Disorders**, Elsevier, v. 87, n. 2-3, 2005.
- KEMPER, T. D. How many emotions are there? wedding the social and the autonomic components. **American journal of Sociology**, University of Chicago Press, v. 93, n. 2, 1987.
- KIM, T. K. T test as a parametric statistic. **Korean journal of anesthesiology**, The Korean Society of Anesthesiologists, v. 68, n. 6, p. 540–546, 2015.
- KOCHMAN, F. J.; HANTOUCHE, E. G.; FERRARI, P.; LANCRENON, S.; BAYART, D.; AKISKAL, H. S. Cyclothymic temperament as a prospective predictor of bipolarity and suicidality in children and adolescents with major depressive disorder. **Journal of affective disorders**, Elsevier, v. 85, n. 1-2, p. 181–189, 2005.
- KORENIUS, T.; LAURIKKALA, J.; JÄRVELIN, K.; JUHOLA, M. Stemming and lemmatization in the clustering of finnish text documents. In: **Proceedings of the thirteenth ACM international conference on Information and knowledge management**. [S.l.: s.n.], 2004. p. 625–633.
- LAURETTO, M. S. **Árvores de Decisão**. 2010. EACH-USP. Disponível em: <[https://edisciplinas.usp.br/pluginfile.php/4469825/mod\\_resource/content/1/ArvoresDecisao\\_normalsize.pdf](https://edisciplinas.usp.br/pluginfile.php/4469825/mod_resource/content/1/ArvoresDecisao_normalsize.pdf)>. Acesso em: 07.09.2021.
- LEAL, R. dos S. **Métricas Comuns em Machine Learning: como analisar a qualidade de chat bots inteligentes — métricas (3 de 4)**. 2017. Medium.com. Disponível em: <<https://medium.com/as-m%C3%A1quinas-que-pensam/m%C3%A9tricas-comuns-em-machine-learning-como-analisar-a-qualidade-de-chat-bots-inteligentes-m%C3%A9tricas-1ba580d7cc96>>. Acesso em: 30.09.2022.
- LEARN scikit. **Support Vector Machines**. 2021. Scikit-learn. Disponível em: <<https://scikit-learn.org/stable/modules/svm.html#>>. Acesso em: 09.07.2021.
- LIMA, A. C. E.; CASTRO, L. N. de. Tecla: A temperament and psychological type prediction framework from twitter data. **Plos one**, Public Library of Science San Francisco, CA USA, v. 14, n. 3, p. e0212844, 2019.
- LIU, B. Sentiment analysis and opinion mining. **Synthesis lectures on human language technologies**, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.
- MACNEILL, L. A.; PÉREZ-EDGAR, K. Temperament and emotion. **The Encyclopedia of Child and Adolescent Development**, Wiley Online Library, 2019.
- MARTINS, L. M.; ALCANTARA, C. D.; BARIONI, M. C. N.; JÚNIOR, L. C. D. O.; FARIA, E. R. A method for analysis of human temperament in contrast to social network data. In: **Brazilian Symposium on Multimedia and Web**. [S.l.: s.n.], 2022. p. 19–27.
- MITCHELL, T. M. **Machine Learning**. [S.l.]: McGraw-Hill Science/Engineering/Math, 1997.

- NAF'AN, M. Z.; BIMANTARA, A.; LARASATI, A.; RISONDANG, E.; NUGRAHA, N. Sentiment analysis of cyberbullying on instagram user comments. **Journal of Data Science and Its Applications**, v. 2, p. 88–98, 04 2019.
- NAYAK, A. S.; KANIVE, A. P.; CHANDAVEKAR, N.; BALASUBRAMANI, R. Survey on pre-processing techniques for text mining. **International Journal of Engineering and Computer Science**, v. 5, n. 6, p. 16875–16879, 2016.
- NICOLAI, G.; KONDRAK, G. Leveraging inflection tables for stemming and lemmatization. In: **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. [S.l.: s.n.], 2016. p. 1138–1147.
- PAYNE, R. L.; COOPER, C. **Emotions at work: Theory, research and applications for management**. [S.l.]: John Wiley & Sons, 2003.
- PENNEBAKER, J. W.; FRANCIS, M. E.; BOOTH, R. J. Linguistic inquiry and word count: Liwc 2001. **Mahway: Lawrence Erlbaum Associates**, v. 71, n. 2001, p. 2001, 2001.
- PENNEBAKER, J. W.; KING, L. A. Linguistic styles: language use as an individual difference. **Journal of personality and social psychology**, American Psychological Association, v. 77, n. 6, p. 1296, 1999.
- PINTO, M. A. S.; JUNIOR, A. F. L. J.; BUSSON, A. J. G.; COLCHER, S. Relacionando modelagem de tópicos e classificação de sentimentos para análise de mensagens do twitter durante a pandemia da covid-19. In: SBC. **Anais Estendidos do XXVI Simpósio Brasileiro de Sistemas Multimídia e Web**. [S.l.], 2020. p. 61–64.
- PIVETTA, S. P. Classificação de documentos do exército brasileiro utilizando o classificador naive bayes e técnicas de seleção de sentenças. Universidade Federal do Pampa, 2013.
- PLANK, B.; HOVY, D. Personality traits on twitter—or—how to get 1,500 personality tests in a week. In: **Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis**. [S.l.: s.n.], 2015. p. 92–98.
- POTTER, K.; HAGEN, H.; KERREN, A.; DANNENMANN, P. Methods for presenting statistical information: The box plot. In: **VLUDS**. [S.l.: s.n.], 2006. p. 97–106.
- PROJECT, N. **Natural Language Toolkit**. 2021. Disponível em: <<https://www.nltk.org/>>. Acesso em: 20.10.2021.
- PUJARI, A. K. **Data mining techniques**. [S.l.]: Universities press, 2001.
- PUTNAM, H. **Realism with a human face**. [S.l.]: Harvard University Press, 1992.
- RAUT, V.; PATIL, P. Use of social media in education: Positive and negative impact on the students. **International Journal on Recent and Innovation Trends in Computing and Communication**, v. 4, n. 1, p. 281–285, 2016.
- REIS, E. A.; REIS, I. A. Análise descritiva de dados. **Relatório Técnico do Departamento de Estatística da UFMG**, v. 1, 2002.

RIHMER, Z.; AKISKAL, K. K.; RIHMER, A.; AKISKAL, H. S. Current research on affective temperaments. **Current opinion in psychiatry**, LWW, v. 23, n. 1, 2010.

RODRIGUES, N. P. A.; SILVA, D. R. da; JÚNIOR, É. A. G.; JÚNIOR, E. F. da S.; GOMES, R. da S.; FERNANDES, K. F. D.; OLIVEIRA, K. Á. R. de. Divulgação de informações sobre higiene e mudança de hábitos durante a pandemia da covid-19. **Research, Society and Development**, v. 10, n. 1, p. e30910111739–e30910111739, 2021.

ROSA, R. L. **Análise de sentimentos e afetividade de textos extraídos das redes sociais**. Tese (Doutorado) — Universidade de São Paulo, 2015.

ROTHBART, M. K.; AHADI, S. A.; EVANS, D. E. Temperament and personality: origins and outcomes. **Journal of personality and social psychology**, American Psychological Association, v. 78, n. 1, p. 122, 2000.

RUZ, G. A.; HENRÍQUEZ, P. A.; MASCAREÑO, A. Sentiment analysis of twitter data during critical events through bayesian networks classifiers. **Future Generation Computer Systems**, Elsevier, v. 106, p. 92–104, 2020.

SALEENA, N. et al. An ensemble classification system for twitter sentiment analysis. **Procedia computer science**, Elsevier, v. 132, p. 937–946, 2018.

SAMPASA-KANYINGA, H.; LEWIS, R. F. Frequent use of social networking sites is associated with poor psychological functioning among children and adolescents. **Cyberpsychology, Behavior, and Social Networking**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 18, n. 7, p. 380–385, 2015.

SHIHA, M.; AYVAZ, S. The effects of emoji in sentiment analysis. **Int. J. Comput. Electr. Eng.(IJCEE.)**, v. 9, n. 1, p. 360–369, 2017.

SIDDIQUI, S.; SINGH, T. et al. Social media its impact with positive and negative aspects. **International Journal of Computer Applications Technology and Research**, v. 5, n. 2, p. 71–75, 2016.

SILVESTRE, A. **Análise de dados e estatística descritiva**. [S.l.]: Escolar editora, 2007.

SKOWRON, M.; THEUNIS, M.; RANK, S.; BOROWIEC, A. Effect of affective profile on communication patterns and affective expressions in interactions with a dialog system. In: SPRINGER. **International Conference on Affective Computing and Intelligent Interaction**. [S.l.], 2011. p. 347–356.

SKOWRON, M.; TKALČIČ, M.; FERWERDA, B.; SCHEDL, M. Fusing social media cues: personality prediction from twitter and instagram. In: **Proceedings of the 25th international conference companion on world wide web**. [S.l.: s.n.], 2016. p. 107–108.

SOCIAL, W. A. **Global Digital Overview Essential Insights Into How People Around The World Use The Internet, Mobile Devices, Social Media, And Ecommerce**. 2020. Disponível em: <<https://wearesocial.com/digital-2020>>. Acesso em: 07.03.2021.

- SOCIEDADE, S. com a. **Pandemia E O Consumo De Notícias Nas Redes Sociais**. 2020. Disponível em: <<https://gente.globo.com/pandemia-e-o-consumo-de-noticias-nas-redes-sociais/>>. Acesso em: 07.03.2021.
- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information processing & management**, Elsevier, v. 45, n. 4, p. 427–437, 2009.
- SOUZA, M.; VIEIRA, R. Construction of a portuguese opinion lexicon from multiple resources. **Anais do Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana, 2011, Brasil**, 2011.
- SPERIOSU, M.; SUDAN, N.; UPADHYAY, S.; BALDRIDGE, J. Twitter polarity classification with label propagation over lexical links and the follower graph. In: **Proceedings of the First workshop on Unsupervised Learning in NLP**. [S.l.: s.n.], 2011. p. 53–63.
- STATISTA. **Leading countries based on Instagram audience size as of January 2021**. 2021. Disponível em: <<https://www.statista.com/statistics/578364/countries-with-most-instagram-users/>>. Acesso em: 09.03.2021.
- STETS, J. E. Emotions and sentiments. In: **Handbook of social psychology**. [S.l.]: Springer, 2006. p. 309–335.
- TAN, P.-N.; STEINBACH, M.; KARPATNE, A.; KUMAR, V. **Introduction to Data Mining, (Second Edition)**. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2013. ISBN 0133128903.
- TRSTENJAK, B.; MIKAC, S.; DONKO, D. Knn with tf-idf based framework for text categorization. **Procedia Engineering**, Elsevier, v. 69, p. 1356–1364, 2014.
- TWITTER. **Twitter**. 2021. Twitter. Disponível em: <<https://twitter.com/>>. Acesso em: 09.01.2021.
- VARGIU, E.; URRU, M. Exploiting web scraping in a collaborative filtering-based approach to web advertising. **Artif. Intell. Research**, v. 2, n. 1, p. 44–54, 2013.
- VAYEGO, S. A. **Resumo 11 - Testes de Hipóteses**. 2022. Ur-  
[lhttps://docs.ufpr.br/vayego/pedeeefes/resumo1.pdf](https://docs.ufpr.br/vayego/pedeeefes/resumo1.pdf).
- VÁZQUEZ, G. H.; TONDO, L.; MAZZARINI, L.; GONDA, X. Affective temperaments in general population: a review and combined analysis from national studies. **Journal of affective disorders**, Elsevier, v. 139, n. 1, 2012.
- VIJAYARANI, S.; JANANI, R. et al. Text mining: open source tokenization tools-an analysis. **Advanced Computational Intelligence: An International Journal (ACII)**, v. 3, n. 1, p. 37–47, 2016.
- WEBDRIVER, S. **Selenium automates browsers. That's it!** 2021. Selenium automates browsers. That's it! Disponível em: <<https://www.selenium.dev/>>. Acesso em: 09.01.2021.

WEISMAYER, C.; GUNTER, U.; ÖNDER, I. Temporal variability of emotions in social media posts. **Technological Forecasting and Social Change**, Elsevier, v. 167, p. 120699, 2021.

WIKIPEDIA, t. f. e. **Support-vector machine**. 2021. Wikipedia, the free encyclopedia. Disponível em: <[https://en.wikipedia.org/wiki/Support-vector\\_machine](https://en.wikipedia.org/wiki/Support-vector_machine)>. Acesso em: 10.07.2021.

WOODRUFF, E.; GENARO, L. T.; LANDEIRA-FERNANDEZ, J.; CHENIAUX, E.; LAKS, J.; JEAN-LOUIS, G.; NARDI, A. E.; VERSIANI, M. C.; AKISKAL, H. S.; MENDLOWICZ, M. V. Validation of the brazilian brief version of the temperament auto-questionnaire temps-a: the brief temps-rio de janeiro. **Journal of affective disorders**, Elsevier, v. 134, n. 1-3, p. 65–76, 2011.

WOODS, H.; SCOTT, H. #sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. **Journal of Adolescence**, v. 51, p. 41–49, 08 2016.

YAN, D.; LI, K.; GU, S.; YANG, L. Network-based bag-of-words model for text classification. **IEEE Access**, IEEE, v. 8, p. 82641–82652, 2020.

YIN, L.; SONG, T.-H.; WEI, Y.-Y.; ZHANG, L.-G.; ZHOU, S.-J.; YU, J.-J.; ZHANG, L.-Y.; LI, H.-J.; CHEN, J.-X. Relationship between affective temperaments and suicide risk in patients with first-onset major depressive disorder. **Frontiers in psychiatry**, Frontiers Media SA, v. 13, 2022.

YIU, T. **Understanding Random Forest - How the Algorithm Works and Why it Is So Effective**. 2021. Towards Data Science. Disponível em: <<https://towardsdatascience.com/understanding-random-forest-58381e0602d2>>. Acesso em: 09.07.2021.

ZAFARANI, R.; ABBASI, M. A.; LIU, H. **Social media mining: an introduction**. [S.l.]: Cambridge University Press, 2014.

ZAKI, M. J.; JR, W. M.; MEIRA, W. **Data mining and analysis: fundamental concepts and algorithms**. [S.l.]: Cambridge University Press, 2014.

ZHAN, M.; TU, R.; YU, Q. Understanding readers: Conducting sentiment analysis of instagram captions. In: **Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence**. [S.l.: s.n.], 2018. p. 33–40.



# Apêndices





APÊNDICE **A**

---

**Questionário de Temperamento  
TEMPS-RIO**

# Estudo Observacional Transversal Retrospectivo sobre a Influência das Redes Sociais no Temperamento dos Usuários

Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisas com Seres Humanos da Universidade Federal de Uberlândia CAAE 35533420.2.0000.5152

---

## \*Obrigatório

1. E-mail (usado apenas para retirar o voluntário da pesquisa, caso seja solicitado) \*

---

2. Idade \*

*Marcar apenas uma oval.*

- 15 a 19 anos
- 20 a 24 anos
- 25 a 29 anos
- 30 a 34 anos
- 35 a 39 anos
- 40 a 44 anos
- 45 a 49 anos
- 50 a 54 anos
- 55 a 59 anos
- 60 a 64 anos
- 65 a 69 anos
- 70 anos ou mais

3. Curso \*

---

4. Gênero \*

Marcar apenas uma oval.

- Feminino
- Masculino
- Prefiro não dizer
- Outro: \_\_\_\_\_

Coleta de Dados

Para coletar seus dados da rede social Instagram é necessário que seu PERFIL seja PUBLICO

5. Para coletar seus dados da rede social Instagram é necessário que seu perfil seja público e que você informe seu usuário no campo abaixo, conforme exemplo sublinhado de vermelho na imagem



---

---

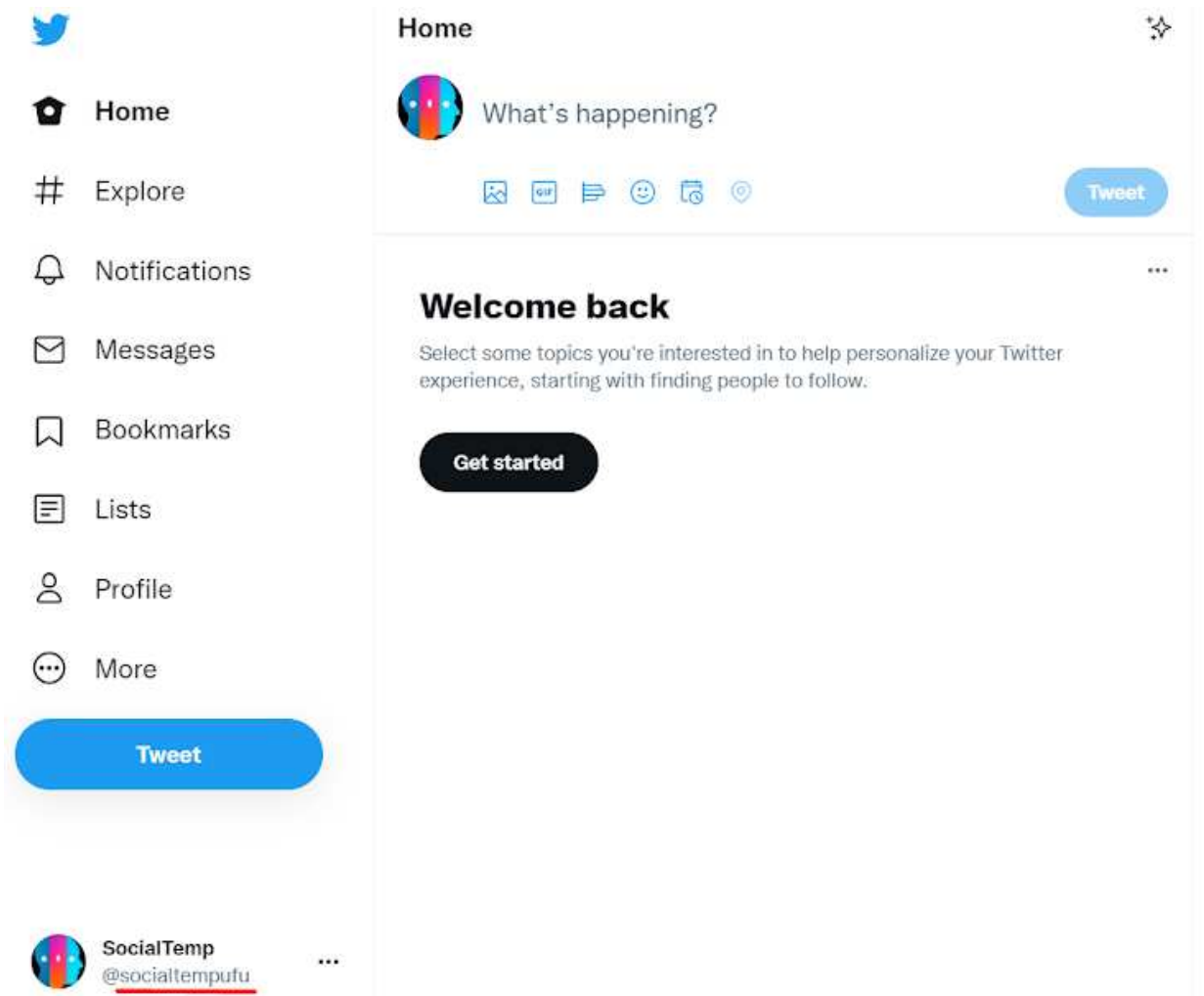
---

---

---

---

- 6. Para coletar seus dados da rede social Twitter é necessário que seu perfil seja público e que você informe seu usuário no campo abaixo, conforme exemplo sublinhado de vermelho na imagem



Suporte Social

Responda às afirmações a seguir de acordo com o seu nível de concordância:

discordo

- 1- Concordo totalmente
- 2- Concordo na maior parte
- 3- Não concordo nem
- 4- Discordo na maior parte
- 5- Discordo totalmente

7. Meus amigos não me procuram tantas vezes quanto eu gostaria. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

8. Estou Satisfeito com a quantidade de amigos que tenho. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

9. Estou satisfeito com a quantidade de tempo que passo com meus amigos. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

10. Estou satisfeito com o tipo de amigos que tenho. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

11. Por vezes sinto-me só no mundo e sem apoio. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

12. Encontro facilmente amigos com quem desabafar quando preciso. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

13. Se precisar de apoio nas situações de emergência, tenho a quem recorrer mesmo nas situações mais embaraçosas. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

14. Sinto falta de alguém verdadeiramente íntimo que me compreenda e com quem eu possa desabafar. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

15. Estou satisfeito com o meu relacionamento com a minha família. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

16. Estou satisfeito com a quantidade de tempo que passo com a minha família. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

17. Estou satisfeito com o que faço em conjunto com a minha família. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

18. Não saio com amigos tantas vezes quanto eu gostaria. \*

*Marcar apenas uma oval.*

	1	2	3	4	5	
Concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Discordo totalmente

19. Sinto falta de atividades sociais que me satisfaçam. \*

Marcar apenas uma oval.

1      2      3      4      5

---

Concordo totalmente      Discordo totalmente

---

20. Gostaria de estar mais envolvido em atividades de organizações (clubes esportivos, partidos políticos, etc) \*

Marcar apenas uma oval.

1      2      3      4      5

---

Concordo totalmente      Discordo totalmente

---

Temperamento

Responda às afirmações a seguir de acordo com o seu nível de concordância:

1- Verdadeiro  
2- Falso

21. Sou uma pessoa triste, infeliz. \*

Marcar apenas uma oval.

1      2

---

Verdadeiro   Falso

---

22. As pessoas dizem que não consigo ver o lado mais positivo das coisas. \*

Marcar apenas uma oval.

1      2

---

Verdadeiro   Falso

---



23. Tenho sofrido muito na vida. \*

*Marcar apenas uma oval.*

1 2

Verdadeiro   Falso

24. Acho que as coisas sempre acabam do pior modo. \*

*Marcar apenas uma oval.*

1 2

Verdadeiro   Falso

25. Que eu me lembre, sempre me senti um fracasso. \*

*Marcar apenas uma oval.*

1 2

Verdadeiro   Falso

26. Dizem que muitas vezes fico pessimista com as coisas e esqueço os momentos felizes do passado. \*

*Marcar apenas uma oval.*

1 2

Verdadeiro   Falso

27. Sou, por natureza, uma pessoa insatisfeita. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

28. Sou impulsionado (a) por uma agitação desagradável que não consigo entender. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

29. Tenho mudanças bruscas no humor e no nível de energia. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

30. Meu humor e meu nível de energia ora estão altos, ora estão baixos; raramente estão no meio. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

31. Minha capacidade de pensar varia bastante, de muito precisa a pobre, sem motivo aparente. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

32. Muitas vezes, começo coisas, mas perco o interesse nelas antes de acabá-las. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

33. Eu oscilo constantemente entre me sentir energizado (a) e quase parando, lento (a). \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

34. Vario muito entre me sentir superconfiante e me sentir inseguro (a) de mim mesmo (a). \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

35. O modo como vejo as coisas às vezes é vibrante e às vezes é sem graça. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

36. Sou o tipo de pessoa que pode ficar triste e alegre ao mesmo tempo. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

37. Frequentemente, estouro com as pessoas e depois me sinto culpado (a) por isso. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

38. Sou uma pessoa rabugenta (irritável). \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

39. Sou muito crítico (a) em relação aos outros. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

40. Muitas vezes eu fico tão aborrecido (a) a ponto de quebrar tudo. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

41. Quando contrariado (a), eu posso entrar numa briga. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

42. As pessoas me dizem que eu estouro por nada. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

43. Quando irritado (a), eu sou áspero (a) com as pessoas. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

44. Meu humor ferino já me causou problemas. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

45. Sinto todas as emoções intensamente. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

46. A vida é uma festa que aproveito ao máximo. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

47. Tenho muita confiança em mim mesmo. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

48. Tenho com freqüência grandes ideias. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

49. Tenho o dom de falar, convencer e inspirar os outros. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

50. Adoro iniciar novos projetos, mesmo que arriscados. \*

*Marcar apenas uma oval.*

1      2

---

Verdadeiro   Falso

---

51. Uma vez que me decida a fazer algo, nada me detém. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

52. Sinto-me totalmente à vontade mesmo com pessoas que mal conheço. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

53. Quando estou estressado (a), minhas mãos muitas vezes tremem. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

54. Muitas vezes meu estômago fica embrulhado. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso



55. Quando estou nervoso (a), eu posso ter diarreia. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

56. Quando estou nervoso (a), muitas vezes sinto náuseas. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

57. Tenho frequentemente medo que alguém na minha família pegue uma doença séria. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

58. Estou sempre na expectativa de que alguém possa me trazer más notícias sobre alguém da minha família. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

59. Quando estou estressado (a), fico com uma sensação desconfortável no peito. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

60. Ruídos bruscos facilmente me deixam sobressaltado (a). \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

61. Que eu me lembre, sempre fui uma pessoa preocupada. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

62. Estou sempre preocupado (a) com uma coisa ou outra. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

63. Estou sempre preocupado (a) com assuntos do dia-a-dia que os outros consideram sem importância. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

64. Não consigo deixar de ficar preocupado (a). \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

65. Muitas pessoas já me disseram para não me preocupar tanto. \*

*Marcar apenas uma oval.*

1      2

Verdadeiro   Falso

---

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários



APÊNDICE **B**

---

# Manual de Classificação de Legendas do *Instagram*

**Manual de Classificação**  
Diretrizes para avaliadores de *posts* online  
para extração de sentimentos

Pedro Henrique Marra Araújo,  
Gean Fernandes da Silva,  
Gabriel Cardoso Brandão de Sousa Amaral

2022

## Sumário

1	Do Objetivo do Manual	3
2	Dos Sentimentos	4
3	Da Metodologia	6
4	Dos Exemplos	9

# 1 Do Objetivo do Manual

A priori, este manual tem como objetivo fornecer as diretrizes necessárias para a classificação manual dos sentimentos presentes em *posts* coletados das redes sociais, tais como *Instagram* e *Twitter*. Tais excertos textuais, nesse contexto, trazem consigo evidentes contrastantes narrativas sustentadas pelas singularidades encontradas em cada plataforma: uma falsa factualidade que visa retratar sentimentos hiper felizes e um extremo pessimismo que ironiza os sofrimentos da vida, encontrados predominantemente no *Instagram*, ao passo que, no *Twitter*, comumente, revelam-se posições cínicas, sarcásticas e, até mesmo, hostis acerca de tópicos, muitas vezes, desconexos com o contexto pessoal de cada usuário, evidenciando as idiossincrasias de cada plataforma e como estas modelam a maneira a qual o usuário comunica-se nesses espaços virtuais. Ademais, o denso volume de figuras de linguagem, regionalismos e *emojis* presentes nestes ambientes, os quais caracterizam um uso informal e desregrado da linguagem, dificulta a detecção precisa de sentimentos que poderiam vir a ser determinantes para a análise.

Ou seja, são notórias as barreiras presentes na definição de um padrão classificatório para diferentes manuais que visam extrair, manualmente, um determinado sentimento — ou qualquer outra informação relevante — nesse grande volume de excertos oriundos destas plataformas digitais. Nesse sentido, uma vez que os algoritmos supervisionados tem como base para predição a classificação manual destes, constata-se, conseqüentemente, que essas dificuldades acarretam em uma menor acurácia por parte destes mesmos algoritmos. Sendo assim, existe uma necessidade relacionada à construção e à estruturação de um manual competente, em primeira instância, na observação de tais nuances, que consiga, em segunda instância, utilizar a informação apreendida anteriormente para, de forma geral e levando em consideração, na medida do possível, a maior diversidade de contextos possíveis, estabelecer definições que sumariza estas características observadas. A seguir, trataremos dos sentimentos os quais o avaliador deverá ser capaz de extrair de cada excerto e de uma metodologia de classificação manual desses — a priori, *tweets* retirados do *Twitter* e *posts* retirados do *Instagram* — a fim de padronizar tal classificação.



## 2 Dos Sentimentos

O sentimento expresso por determinado excerto refere-se ao sentimento ao qual o autor estava experienciando no momento da postagem e/ou ao sentimento ao qual pretendia externalizar ao publicá-lo. Isto é, pelo fato de ativamente elaborar tal excerto, ou seja, ao redigi-lo e, sobretudo, ao postá-lo em uma rede social, o autor manifesta clara intenção de compartilhar tal sentimento, ou seja, sua experiência subjetiva no momento da postagem. Dessa forma, delimitada essa intenção, esse sentimento pode, e neste manual será distinguido, em três classes distintas, as quais serão explicitadas abaixo: sentimento negativo — rotulado, também, como “-1”; sentimento neutro — “0”; sentimento positivo — “1”.

- (i) **Sentimento negativo (-1):** o excerto em análise expressa a clara intenção do autor em externar um sentimento negativo, isto é, de sua vivência subjetiva em relação a algum fato ou acontecimento pessoal que o proporcionou uma experiência de insatisfação, de raiva, de ódio, de hostilidade, de tristeza, de melancolia, de saudade, de culpa, de nojo, de surpresa negativa, de medo, de frustração, de aversão e/ou de ciúme, por exemplo.
- (ii) **Sentimento neutro (0):** o excerto em análise expressa a clara intenção do autor em externar um sentimento neutro, isto é, de sua vivência subjetiva em relação a algum fato ou acontecimento pessoal que o proporcionou uma experiência de apatia, ou seja, de indiferença emocional e/ou de dúvida, por exemplo. Ou, também, o excerto expressa alguma notícia, narração de algum fato, descrição objetiva de algo, opinião objetiva e/ou reportagens em geral que não possui consigo opinião pessoal do autor carregada de sentimentalismos descritos pelo tópico acima, nem pelo tópico abaixo.
- (iii) **Sentimento positivo (1):** o excerto em análise expressa a clara intenção do autor em externar um sentimento positivo, isto é, de sua vivência subjetiva em relação a algum fato ou acontecimento pessoal que o proporcionou uma experiência de felicidade, de humor, de alegria, de amor, de gratidão, de esperança, de surpresa positiva, de admiração, de empatia, de excitação, de nostalgia, de apreço visual e/ou de triunfo, isto é, de orgulho de si próprio, por exemplo.

Todavia, existem excertos que possuem determinada ambiguidade entre os sentimentos explícitos acima, os quais denotaremos, neste manual, como excertos inconclusivos. Isto é, no contexto da classificação manual desses, o classificador há de ficar, em alguns casos, em dúvida quanto à intenção do autor em externar seu sentimento, isto é, de sua vivência subjetiva, ao publicar aquele determinado excerto, seja por falta de elementos textuais capazes de limitarem o verdadeiro sentimento do autor, seja pela falta de clareza do autor em externar seu sentimento, ou seja pela presença múltipla de diferentes sentimentos. Tanto no caso de convicção do avaliador sobre o sentimento expresso no excerto, quanto no caso de ambiguidade entre os sentimentos, o avaliador deverá seguir os passos de classificação manual descritos na próxima seção.

### 3 Da Metodologia

A respeito da metodologia, procuramos estabelecer, hierarquicamente, três pilares centrais para edificar este manual, de tal forma que, juntos, estes possam servir como um guia direto e sucinto para aqueles que desejam realizar classificações manuais de objetos semelhantes. Ou seja, ao deparar com um excerto textual virtual, o classificador deve se dar maior relevância para os primeiros critérios, e menor para os últimos. Isto é, ao analisar determinado excerto, o avaliador deverá primeiramente analisar separadamente os *emojis*, caso a análise seja inconclusiva a partir dessas elementos textuais, o avaliador deverá partir para a identificação de palavras-chave e, caso a análise dessas seja, também, inconclusiva, o avaliador deverão então partir para a análise do contexto geral do excerto.

- (i) ***Emojis***: sua importância se dá pois, uma vez utilizado, este revela forte intenção em expressar algum sentimento ou emoção específica, haja vista se tratar de uma linguagem visual que o usuário teve que ativamente buscar. Além disso, em diferentes contextos, por exemplo no contexto social, um mesmo *emoticon* pode possuir significados totalmente diferentes, impossibilitando, assim, uma classificação não-contextual extensiva de todos os *emojis*. Ainda nesse sentido, é notório que determinados *emojis* têm uma popularidade maior entre algumas classes de usuários em determinada realidade histórica. Por exemplo, dado que, geralmente, o corpo desses excertos são recuperados em relação a um contexto específico (uma *hashtag* ou um período fixo de tempo, por exemplo), pode ser que apareça um *emoticon* que seja usado de forma não convencional nesse conjunto de excertos.

Assim, o avaliador deve levar em consideração seu conhecimento do uso padrão e do uso não convencional dos *emojis*, dado o contexto do corpo a ser classificado, para extrair o sentimento geral do excerto, dando preferência ao significado contextual do *emoticon* em questão. Caso o excerto seja composto somente por *emojis*, o avaliador deverá considerar o significado desses *emojis* em conjunto e avaliar seus significados inerentes de acordo com o uso padrão e o uso não convencional, como exemplificado na próxima seção. Todavia, caso o sentimento desse *emoticon* extravie gravemente do sentimento geral do excerto, seja por in-

congruência com o restante do excerto ou por esse *emoticon* tratar-se de uma figura de linguagem dissonante (ironia, por exemplo), o avaliador deve seguir as seguintes diretrizes;

- (ii) **Palavras-Chave:** estas se resumem a palavras ou expressões que descrevem um sentimento muito claramente, de tal maneira que, mesmo isoladas, estas corroboram significativamente para a definição do sentimento geral presente. Essas palavras possuem, nesse sentido, grande valor inerente na tarefa de classificação de um excerto, pois a própria escolha do autor de expressá-las denota a decisão consciente dele de externar determinado sentimento. São consideradas, assim, as palavras que possuem grande valor semântico e personalidade do autor, isto é, que expressam seu sentimento de maneira contundente, ainda que sem o contexto geral, o qual somente é gerado pela totalidade do excerto. Assim, o avaliador deve identificar, quando possível, essas palavras-chave, levando em consideração seu conhecimento prático do uso da linguagem, o contexto do corpo de excertos e os exemplos disponibilizados na próxima seção. O avaliador, nesse sentido, deve basear-se em sua experiência na utilização desses meios eletrônicos, ou seja, dessas redes sociais, para a identificação dessas palavras, seja em sua forma padrão (norma culta), seja em suas formas abreviadas, seja em forma de gírias. Todavia, caso o sentimento dessas palavras e/ou expressões extraviem gravemente do sentimento geral do excerto, seja por incongruência com o restante do excerto ou por essas tratarem-se de uma figura de linguagem dissonante (ironia, por exemplo), o avaliador deve seguir a diretriz seguinte;
- (iii) **Análise Geral:** trata-se da análise da totalidade do excerto, a fim de obter alguma conclusão relevante, caso esta não esteja óbvia de acordo com os critérios acima. Acontece que, em diversos casos, somente o excerto em sua integralidade, caracterizado no contexto social e temporal o qual foi postado, possui sentimento evidente, isto é, demarca nitidamente qual o sentimento que o autor estava experienciando ao redigi-lo, sendo que somente a descrição de outro indivíduo (no caso, o avaliador) consegue extrair. Assim, o avaliador deve ter a total descrição de avaliar qual o sentimento geral do excerto, levando em consideração sua vivência pessoal, em seu contexto social, e sua capacidade de interpretação. Caso, ainda assim, o avaliador fique em dúvida quanto ao sentimento

do excerto, esse deve marcar um dos sentimentos entre esses e marcar o texto como inconclusivo.

Dessa forma, o avaliador, ao analisar determinado excerto, deve assinalar, na coluna “Sentimento”, o respectivo sentimento dentre os descritos em 2. Dos Sentimentos. Ou seja, assinalar com “-1” caso o excerto exprima um sentimento negativo, “0” caso o excerto exprima um sentimento neutro ou “1” caso o excerto exprima um sentimento positivo, caso possível. Em caso de ambiguidade, isto é, de excerto inconclusivo como descrito na seção acima, o avaliador deverá escolher um sentimento dentre as possibilidades ambíguas e assinalar com “X” na coluna “Inconclusivo”.

## 4 Dos Exemplos

Abaixo, estão alguns exemplos de *posts* retirados da plataforma *Instagram* e *Twitter*, os quais foram rotulados de acordo com as diretrizes da seção Da Metodologia. Para cada exemplo, segue uma breve explicação da motivação, seguindo as diretrizes acima, de cada sentimento, ou seja, como o avaliador deveria interpretá-lo. Abaixo, encontra-se, também, alguns exemplos de excertos ambíguos, igualmente elucidados.



**Figura 1:** No trecho acima, nota-se o uso de *emojis*, sem nenhuma outra estrutura textual. Dessa forma, o sentimento do excerto pode, e deve, ser extraído exclusivamente dos *emojis*. Assim, há gotas e um sol, inerentemente positivo, os quais remetem, por exemplo, a um dia quente de sol, de alegria e de sensações positivas. Portanto, a publicação é classificada como positiva.



**Figura 2:** No trecho acima, nota-se o uso de *emojis*, sem nenhuma outra estrutura textual. Dessa forma, o sentimento do excerto pode, e deve, ser extraído exclusivamente do *emoticon*. Assim, há um único *emoticon* de uma mulher dançando, uma atividade geralmente relacionada à sentimentos positivos. Portanto, a publicação é classificada como positiva.



**Figura 3:** No trecho acima, nota-se o uso de *emojis*, sem nenhuma outra estrutura textual. Dessa forma, o sentimento do excerto pode, e deve, ser extraído exclusivamente do *emoticon*. Assim, há um único *emoticon* de uma árvore, a qual, sem nenhum outro contexto, não possui conotação negativa, nem conotação positiva. Portanto, a publicação é classificada como neutra.



**Figura 4:** No trecho acima, nota-se o uso de *emojis*, sem nenhuma outra estrutura textual. Dessa forma, o sentimento do excerto pode, e deve, ser extraído exclusivamente do *emoticon*. Assim, há um único *emoticon* de uma casa, a qual, sem nenhum outro contexto, não possui conotação negativa, nem conotação positiva. Portanto, a publicação é classificada como neutra.



**Figura 5:** No trecho acima, nota-se o uso de *emojis*, sem nenhuma outra estrutura textual. Dessa forma, o sentimento do excerto pode, e deve, ser extraído exclusivamente do *emojis*. Assim, como há somente rostos expressando sentimentos negativos, a publicação é classificada como negativa.



**Figura 6:** No trecho acima, nota-se o uso de *emojis*, sem nenhuma outra estrutura textual. Dessa forma, o sentimento do excerto pode, e deve, ser extraído exclusivamente do *emojis*. Assim, como um de um coração partido e outro de um coração em chamas, suponha-se que o autor esteja passando por uma situação amorosa de coração partido. Portanto, a publicação é classificada como negativa.



**Figura 7:** No trecho acima, nota-se o uso de *emojis*, sem nenhuma outra estrutura textual. Dessa forma, o sentimento do excerto pode, e deve, ser extraído exclusivamente do *emojis*. Assim, como há uma mistura uniforme entre *emojis* de conotação negativa e *emojis* de conotação positiva, não há uma clara demarcação do sentimento proposto pelo autor. Portanto, o autor deve marcar o excerto com um desses sentimentos e assinalá-lo como “Inconclusivo”.

“Acabou 🎓🙌”

**Figura 8:** Na frase em questão, pode-se notar que foi utilizado um *emoticon* de um capelo (que remete à formatura) juntamente com um *emoticon* de mãos juntas (o qual, nesse contexto, remete ao sentimento de alívio ao se formar, de felicidade – substitui “finalmente formei”). Analisando os *emojis* podemos inferir que a pessoa que realizou a publicação externa sentimento positivo, uma vez que os *emojis* remetem a um sentimento de alegria, de satisfação.

*“Tem q passar essa entrevista em telões nas praças #TheNoite”*

**Figura 9:** Não há a presença de *emojis* que sugeririam algum sentimento, tampouco palavras-chave e/ou expressões que denotem sentimento. Nesse caso, uma das interpretações implícitas da publicação é que o autor gostou tanto do programa, e portanto passa por um sentimento positivo, que sugeriu a disseminação desse. Porém, a grande sutileza dessa interpretação faz com que, nesse caso, o excerto expresse, na verdade, um sentimento neutro, de opinião pessoal sem sentimentalismos impressos. Portanto, a publicação é classificada como neutra.

*“te odeio microbio maldito, nota pra hj é 7... qual foi a nota da terça por aí?? tô exausto”*

**Figura 10:** Na frase em questão nota-se que não foi publicado nenhum *emoticon*. Sendo assim, continuamos a rotulação analisando as palavras-chaves no texto publicado. Podemos identificar que na publicação ele utiliza algumas palavras que remetem ao sentimento negativo, como “maldito” e “exausto”, tornando evidente, assim, o sentimento negativo expresso pelo autor na publicação.



*“nossa hoje meu dia foi horrível, mas pelo menos nao peguei cheguei atrasado no trampo, entao fiquei feliz”*



**Figura 11:** Nessa publicação, não há nenhum *emoticon*. Todavia, existem palavras-chaves conflitantes no excerto, isto é, “horrível” e “feliz”, as quais incorrem em uma ambiguidade, pois o avaliador não é capaz de avaliar se, ao publicá-lo, ele estava cansado por um dia “horrível” de trabalho ou estava “feliz” por não ter chegado atrasado no trabalho. Portanto, o avaliador deverá escolher entre o sentimento positivo ou o sentimento negativo e assinalar “X” na coluna “Inconclusivo”.

*“se arrependimento matasse, eu estaria vivíssimo agr... marca quem merece o tapa do dia kkk”*



**Figura 12:** Assim como nos exemplos acima, não temos nenhum *emoticon* presente na publicação, sendo assim partimos para a análise das palavras-chaves, de tal forma que podemos identificar algumas palavras que remetem ao sentimento negativo, como “arrependimento” e “matasse”. Porém, compreendendo a utilização da gíria “kkk”, ou seja, risadas, podemos identificar que a palavra “matasse” é colocada de maneira irônica, sendo que a intenção principal do autor é de externar seu sentimento de humor quanto a esse arrependimento. Portanto, a publicação é classificada como sendo positiva.

*“Pq o programa não ta ao vivo? @ANAMARIABRAGA #MaisVocê”*

**Figura 13:** Não há a presença de *emojis*, nem de palavras-chave e/ou expressões que denotem sentimento. Analisando a postagem em sua totalidade, percebe-se que trata-se de uma dúvida do autor, da narração de um acontecimento (a impossibilidade do autor da postagem em assistir o programa), sem nenhuma marca linguística que denotaria sentimento. Portanto, a publicação é classificada como neutra.

*“Saudade que não passa...   ”*

**Figura 14:** Analisando os *emojis* contidos na publicação podemos notar o *emoticon* “coração azul” que remete a um sentimento positivo. Mas analisando a frase como um todo, percebemos que a palavra “saudade” está presente na publicação e que remete a um sentimento de melancolia. Sendo assim, a publicação é classificada como sendo um sentimento negativo.

*“hoje meu dia foi maravilhoso    ”*

**Figura 15:** Há tanto a presença de *emojis* com carga sintática negativa, quanto a presença da palavra-chave “maravilhoso” positivo. Assim, o avaliador deverá interpretar que, tanto o autor foi incapaz de externar seu sentimento de maneira clara, quanto também existem elementos textuais discrepantes no excerto. Dessa forma, o avaliador deverá escolher entre o sentimento positivo ou o sentimento negativo e assinalar “X” na coluna “Inconclusivo”.

*“Te amo belezura!”*

**Figura 16:** Identificamos a palavra “amor”, que remete a um sentimento positivo. Como na publicação não tem nenhuma outra palavra ou *emoticon* que pode mudar o sentido da publicação, especialmente pela presença do ponto de exclamação — utilizado comumente para evidenciar um sentimento presente em determinada frase —, a publicação é classificada como positiva.

*“o mundo so vai mudar quando as pessoas entenderem que não deve atacar/xingar/zomba quem tem um pensamento ignorante como o desse cara e sim que tu deve ensinar e explicar primeiro, esse debate foi inutil tendo em vista que o cara não mudou”*

**Figura 17:** Não há presença de *emojis*, nem de palavras-chave e/ou expressões que denotem sentimento. Percebe-se claramente, porém, que se trata de uma opinião pessoal do autor sem nenhuma outra marca linguística de sentimento. Portanto, a publicação é classificada como neutra.

*“Amo vocês e sempre vou amar!!! 😞❤️”*

**Figura 18:** Identificamos um *emoticon* de “cara triste” que remete ao sentimento negativo e um de “coração” que remete ao sentimento positivo, não sendo possível, assim, identificar um sentimento predominante. Porém, ao analisar as palavras-chaves, encontramos palavras que remetem ao sentimento positivo, como “amo” e “amar”. Mas após analisar o texto como um todo, podemos perceber que a publicação expressa saudade e melancolia. Sendo assim, a mesma foi classificada como negativa.

*“adorei o capítulo da novela de ontem, porem não entendi porque ainda não mataram aquela certa personagem 😞😞😞”*

**Figura 19:** Há tanto a presença de *emojis* com carga sintática negativa, quanto a presença da palavra-chave “adorei” positivo. Nesse caso, o avaliador deverá interpretar que o autor não conseguiu deixar claro seu sentimento ao publicar esse excerto, pois tanto pode estar feliz pelo bom capítulo da novela, quanto pode estar com raiva pela presença da personagem ainda na novela. Dessa forma, o avaliador deverá escolher entre o sentimento positivo ou o sentimento negativo e assinalar “X” na coluna “Inconclusivo”.