

UNIVERSIDADE NOVE DE JULHO - UNINOVE
PROGRAMA DE PÓS GRADUAÇÃO EM INFORMÁTICA E GESTÃO DO
CONHECIMENTO – PPGI

ANDRÉ LUIS MARQUES FERREIRA DOS SANTOS

Desenvolvimento de um Modelo Epidemiológico Aprimorado com Dados de Redes Sociais e Processamento de Linguagem Natural para a Predição de Casos de COVID-19

São Paulo
2023

ANDRÉ LUIS MARQUES FERREIRA DOS SANTOS

Desenvolvimento de um Modelo Epidemiológico Aprimorado com Dados de Redes Sociais e Processamento de Linguagem Natural para a Predição de Casos de COVID-19

Exame de Defesa apresentado a Universidade Nove de Julho - UNINOVE, como parte dos requisitos para a obtenção do título de Doutor em Informática e Gestão do Conhecimento.

Prof. Orientador: Dr. José Eduardo Storopoli
Profa. Coorientadora: Dra. Alessandra Cristina Guedes Pellini

São Paulo
2023



ATA DE DEFESA DA TESE





Ao trigésimo primeiro dia do mês de maio de dois mil e vinte e três, às 14h00, reuniu-se em sessão pública e virtual a Comissão Julgadora da tese de doutorado de André Luis Marques Ferreira dos Santos sob o título "Desenvolvimento de um Modelo Epidemiológico Aprimorado com Dados de Redes Sociais e Processamento de Linguagem Natural para a Predição de Casos de COVID-19". Integraram a comissão os professores Doutores: José Eduardo Storopoli (PPGI/UNINOVE), Alessandra Cristina Guedes Pellini (PPGCIS/UNINOVE), Ana Freitas Ribeiro (USCS), André Felipe Henriques Librantz (PPGI/UNINOVE), sob a presidência do primeiro, orientador da tese. A banca examinadora, tendo decidido aceitar a tese, passou à arguição pública do candidato. Encerrados os trabalhos, os examinadores deram parecer final sobre a Tese.

	Parecer
Prof(a). Dr(a). José Eduardo Storopoli	Aprovado
Prof(a). Dr(a). Alessandra Cristina Guedes Pellini	Aprovado
Prof(a). Dr(a). Ana Freitas Ribeiro	Aprovado
Prof(a). Dr(a). André Felipe Henriques Librantz	Aprovado

Parecer:

Em conclusão, o candidato foi considerado aprovado, no grau de Doutor em Informática e Gestão do Conhecimento. E, para constar, eu, Prof. Dr. André Felipe Henriques Librantz, diretor do Programa de Mestrado e Doutorado em Informática e Gestão do Conhecimento, lavrei a presente ata que assino juntamente com os membros da banca examinadora.

São Paulo, 31 de maio de 2023.

 _____ Prof(a). Dr(a). José Eduardo Storopoli	 _____ Prof(a). Dr(a). Alessandra Cristina Guedes Pellini
 _____ Prof(a). Dr(a). Ana Freitas Ribeiro	 _____ Prof(a). Dr(a). André Felipe Henriques Librantz

Santos, André Luis Marques Ferreira dos.

Desenvolvimento de um modelo epidemiológico aprimorado com dados de redes sociais e processamento de linguagem natural para a predição de casos de COVID-19. / André Luis Marques Ferreira dos Santos. 2023.

162 f.

Tese (Doutorado) - Universidade Nove de Julho - UNINOVE, São Paulo, 2023.

Orientador (a): Prof. Dr. José Eduardo Storopoli.

1. Modelo Epidemiológico Bayesiano, Processamento de Linguagem natural. 2. Mídias sociais. 3. COVID-19. 4. Modelo de crença em saúde.

I. Storopoli, José Eduardo. II. Título.

CDU 004

Este trabalho é dedicado aos meus pais Daura e Gregório, responsáveis por eu ter conseguido as condições necessárias para chegar até aqui; aos meus filhos Andrey e Arthur, espero que esse trabalho os inspirem para irem ainda mais longe; e a minha esposa Karinne, minha parceira de todos os dias.

AGRADECIMENTOS

A Deus, por ter me dado forças para conseguir superar todos os obstáculos ao longo dessa pesquisa.

Em especial, aos meus orientadores José Eduardo Storopoli e Alessandra Cristina Guedes Pellini por terem me guiado nessa jornada. À minha coorientadora Alessandra Pellini, pela parceria e apoio sempre que eu precisei; e gratidão eterna ao meu orientador José Storopoli que esteve comigo desde o começo desse projeto, me motivando, mostrando os melhores caminhos e destravando trancas e portas, fazendo assim com que minha caminhada fosse muito mais fácil, agradável e produtiva.

Aos professores do programa pelos ensinamentos que me permitiram tanto o desenvolvimento técnico, quanto o desenvolvimento pessoal e profissional.

À minha esposa Karinne Marques e meus filhos Andrey Marques e Arthur Marques por compreenderem a minha ausência enquanto eu me dedicava ao desenvolvimento desta pesquisa.

Aos meus familiares e amigos que por vezes me ajudaram com palavras de conforto, inclusive atuando como patrocinadores nessa jornada.

À Universidade Nove de Julho pela excelente estrutura e financiamento que garantiram as condições necessárias para o desenvolvimento das minhas atividades de pesquisa.

À CAPES pela bolsa de doutoramento que permitiu custear eventos, softwares e demais atividades acadêmicas.

“Se eu enxerguei mais longe, foi porque me apoiei sobre os ombros de gigantes.”

Isaac Newton

Introdução: Este trabalho aborda os desafios enfrentados pelos sistemas de saúde do Brasil durante a pandemia de COVID-19, com destaque para o papel das mídias sociais no apoio à saúde pública durante a crise pandêmica. Ressalta-se que, com a necessidade do isolamento físico e distanciamento social para reduzir a propagação do novo coronavírus durante a pandemia de COVID-19, a sociedade se voltou cada vez mais para os ambientes virtuais, levando a um aumento significativo no uso das mídias sociais. Em consequência, a análise de informações provenientes das mídias sociais ganhou grande relevância, impulsionando o uso e o desenvolvimento de várias técnicas de inteligência artificial para a criação de modelos preditivos. Dessa forma, a utilização desses modelos se tornou uma ferramenta relevante para aprimorar as ações de combate a surtos, epidemias, pandemias e outras situações críticas de saúde. **Objetivo:** O principal objetivo desta pesquisa foi desenvolver um modelo epidemiológico estatístico Bayesiano, que incorporou dados de redes sociais e processamento de linguagem natural para prever casos de COVID-19. **Método:** Esta pesquisa foi desenvolvida sob o paradigma da ciência de dados, na qual foram utilizados dados extraídos de redes sociais (Twitter) e do Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe), de Janeiro/2020 até Dezembro/2020. Como base teórica para a análise contextual dos dados de redes sociais foi utilizado um modelo teórico da área da comunicação em saúde – o Modelo de Crença em Saúde. As técnicas utilizadas foram: (1) processamento de linguagem natural; (2) análise de redes sociais; e (3) modelo epidemiológico Bayesiano. Por fim, para analisar o poder preditivo do modelo estatístico proposto, foram desenvolvidos dois modelos epidemiológicos, um que não utilizou dados de redes sociais e o outro que incorporou esses dados. **Resultados:** Como resultado principal, foi observado que o modelo epidemiológico que incorporou dados de redes sociais conseguiu prever melhor os casos de COVID-19, comparado modelo que não utilizou dados de redes sociais. Outro resultado relevante foi a contribuição para a literatura da comunicação em saúde, onde se apresentou a extensão do Modelo de Crença em Saúde no contexto da pandemia de COVID-19. **Conclusão:** Esta pesquisa fornece informações relevantes para pesquisadores e outras partes interessadas no combate a pandemias, ao entender como as mídias sociais podem ser usadas para melhorar modelos epidemiológicos, permitindo o aprimoramento das estimativas de predição, o que, por sua vez, contribui para mitigar os riscos associados às pandemias, e, consequentemente, subsidiar os formuladores de políticas no desenvolvimento e na implementação de estratégias mais efetivas de saúde pública.

Palavras-chave: Modelo Epidemiológico Bayesiano, Processamento de Linguagem Natural, Mídias Sociais, COVID-19, Modelo de Crença em Saúde.

ABSTRACT

Introduction: Este trabalho aborda os desafios enfrentados pelos sistemas de saúde no Brasil durante a pandemia de COVID-19, com foco no papel das mídias sociais no apoio à saúde pública durante a crise pandêmica. Ressalta-se que, com a necessidade de isolamento físico e distanciamento social para reduzir a propagação do novo coronavírus durante a pandemia de COVID-19, a sociedade tem se voltado cada vez mais para os ambientes virtuais, levando a um aumento significativo do uso das redes sociais. Consequentemente, a análise de informações das mídias sociais ganhou considerável relevância, impulsionando o uso e desenvolvimento de diversas técnicas de inteligência artificial para criar modelos preditivos. Assim, a utilização desses modelos tornou-se uma ferramenta relevante para potencializar ações no combate a surtos, epidemias, pandemias e outras situações críticas de saúde. **Objetivo:** Esta pesquisa teve como objetivo desenvolver um modelo epidemiológico estatístico bayesiano que incorporou dados de mídia social e processamento de linguagem natural para prever casos de COVID-19. **Objective:** The main objective of this research was to develop a Bayesian statistical epidemiological model that incorporated social media data and natural language processing to predict COVID-19 cases. **Method:** This research was developed under the data science paradigm, in which data extracted from social networks (Twitter) and the Influenza Epidemiological Surveillance Information System (SIVEP-Gripe) were used from January/2020 to December/2020. As a theoretical basis for the contextual analysis of social media data, a theoretical model from the field of health communication was used – the Health Belief Model. The techniques used were: (1) natural language processing; (2) social network analysis; and (3) Bayesian epidemiological model. Finally, to analyze the predictive power of the proposed statistical model, two epidemiological models were developed, one that did not use social media data and the other that incorporated this data. **Results:** As a main result, it was observed that the epidemiological model that incorporated social media data could better predict COVID-19 cases compared to the model that did not use social media data. Another relevant result was the contribution to the health communication literature, where the extension of the Health Belief Model in the context of the COVID-19 pandemic was presented. **Conclusion:** This research provides relevant information for researchers and other stakeholders in the pandemic response by understanding how social media can be used to improve epidemiological models, enabling the enhancement of prediction estimates, which, in turn, contributes to mitigating risks associated with pandemics. Consequently, it supports policymakers in developing and implementing more effective public health strategies. **Keywords:** Bayesian, Natural Language Processing, Social Media, COVID-19, Health Belief Model.

SUMÁRIO

	Páginas
Lista de Figuras	12
Lista de Algoritmos	13
Lista de Equações	14
Lista de Quadros	16
Lista de Abreviaturas	17
Lista de Abreviaturas	18
Lista de Tabelas	19
Lista de Símbolos	20
Lista de Mapas	21
1 Introdução	22
1.1 Problema de Pesquisa	26
1.2 Questão de Pesquisa	27
1.3 Objetivos	28
1.3.1 Objetivo Geral	28
1.3.2 Objetivos Específicos	28
1.4 Justificativa para o Estudo do Tema	29
1.5 Estrutura da Tese	33
2 Referencial Teórico	34
2.1 Comunicação em Saúde	34
2.1.1 Teorias de Mudança de Comportamento em Saúde no Contexto das Mídias Sociais	34
2.1.2 Modelo de Crença em Saúde	38
2.2 Mídias Sociais	40
2.2.1 <i>Social Media Analytics (SMA)</i>	40
2.2.2 Mídias sociais para a comunicação em saúde	41
2.3 Modelagem Matemática	43
3 Metodologia	47
3.1 Desafios Científicos e Tecnológicos	47
3.2 Delineamento da Pesquisa	48

3.2.1	Paradigma da Ciência de Dados	48
3.2.2	<i>Framework</i> da Pesquisa	49
3.3	Dados da Pesquisa	50
3.3.1	Fontes de Dados	50
3.3.2	Etapas para o Tratamento dos Dados	61
3.4	Modelo de Processamento de Linguagem Natural	65
3.4.1	Expressões Matemáticas do Classificador Logístico	73
3.5	Modelo Epidemiológico Bayesiano	80
3.6	Limitações	85
4	Resultados	86
4.1	Produção Científica e Tecnológica da Tese	86
4.2	Estatística Descritiva da Amostra	87
4.3	Modelo de <i>Machine Learning</i>	95
4.4	Modelo Epidemiológico Bayesiano	97
5	Conclusão	103
	Referências Bibliográficas	105
	Apêndices	130
Apêndice A	Dicionário de expressões de tweets	132
Apêndice B	Emojis presentes nos dados da pesquisa	151
Apêndice C	Reprodução do <i>Framework</i> da Pesquisa	156
	Anexos	160
Anexo A – Teorias de Mudança de Comportamento		160

LISTA DE FIGURAS

Figura 1 – Frequências de pesquisas publicadas sobre Comunicação em Saúde no mundo e no Brasil, distribuídas por ano de publicação, de 1950 a 2020.	30
Figura 2 – Diagrama do modelo conceitual integrado com base no modelo de crença em saúde.	39
Figura 3 – Principais etapas para o desenvolvimento desta pesquisa.	50
Figura 4 – Visão geral do processo de coleta, curadoria e verificação de dados de COVID-19, Brasil.	55
Figura 5 – Definições de caso operacionais para a COVID-19, Brasil, março de 2020. . .	58
Figura 6 – Fluxo de formação de redes sociais no Twitter.	61
Figura 7 – Etapas de tratamento para a classificação dos dados da pesquisa.	62
Figura 8 – Exemplo de tweet considerado na Etapa 2, no modelo de classificação. . . .	63
Figura 9 – Emojis relacionados com o tema da pesquisa encontrados na amostra (Etapa 3).	64
Figura 10 – Tweet com <i>multilabel</i> classificado com construtos da <i>Health Belief Model</i> (Etapa 4).	69
Figura 11 – Esquema do classificador de regressão logística.	73
Figura 12 – Função Logística.	75
Figura 13 – Função Custo.	77
Figura 14 – SIRT Modelo Compartimental (STOROPOLI et al., 2021a).	82
Figura 15 – Tweet relatando casos de COVID-19 em viajantes	89
Figura 16 – Representação das relações entre os construtos da <i>Health Belief Model</i> . . .	91
Figura 17 – Frequência com que os construtos da <i>Health Belief Model</i> são mencionados (aos pares) em um mesmo <i>tweet</i> . N = 4.000	92
Figura 18 – Nuvem de palavras dos tweets classificados errados pelo modelo LR (<i>tf-idf</i>). . .	97
Figura 19 – Novos casos de SRAG por COVID-19 em 2020.	98
Figura 20 – Tweets (em Língua Portuguesa) sobre sinais e sintomas de COVID-19 em 2020. N = 48.281	99
Figura 21 – Campanha nacional de combate à COVID-19 "Fique Em Casa", março de 2020.	100
Figura 22 – Matriz de correlação entre casos, mortes e tweets de COVID-19 em 2020. . .	100
Figura 23 – Percepção da COVID-19 nas redes sociais de acordo com os construtos da <i>Health Belief Model</i> , 2020.	102

LISTA DE ALGORITMOS

Algoritmo 1 – PSEUDOCÓDIGO DO CLASSIFICADOR LOGÍSTICO LR (TF-IDF) . . .	65
---	----

LISTA DE EQUAÇÕES

Equação 1 – SIR	44
Equação 2 – SIS	44
Equação 3 – SEIR	44
Equação 4 – Dígrafos	59
Equação 5 – Ridge Regression	67
Equação 6 – Lasso Regression	67
Equação 7 – Função Logística	74
Equação 8 – Net input function	74
Equação 9 – Função Logit	74
Equação 10 – Função $g(z)$	75
Equação 11 – Equivalente da função $g(z)$	75
Equação 12 – Likelihood	76
Equação 13 – Log-Likelihood	76
Equação 14 – Função Custo da Log-likelihood	76
Equação 15 – Derivada Parcial da log-likelihood	77
Equação 16 – Derivada parcial da função sigmoide	78
Equação 17 – substituir Equação 16 em 15	78
Equação 18 – Pesos do modelo	78
Equação 19 – Atualização para o peso j	79
Equação 20 – Delta da função peso	79
Equação 21 – Regra de atualização para o peso j	79
Equação 22 – Atualização simultânea dos pesos	79
Equação 23 – Delta dos pesos da função custo	79
Equação 24 – Termo de regularização L_2	80
Equação 25 – Aplicação do termo L_2 na função custo	80
Equação 26 – Atualização da Regularização para único peso	80
Equação 27 – Atualização simultânea de pesos	80
Equação 28 – Delta para atualização simultânea de peso	80
Equação 29 – ODE DS	81
Equação 30 – ODE DI	81
Equação 31 – ODE DR	81
Equação 32 – ODE DT	81
Equação 33 – ODE DD	81
Equação 34 – Prior β	84
Equação 35 – Prior ω	84
Equação 36 – Prior d_I	84
Equação 37 – Prior d_T	84

Equação 38 –Prior Tweets	84
Equação 39 –Prior ϕ	84
Equação 40 –Prior ϕ Tweets	84
Equação 41 –Likelihood Deaths	84
Equação 42 –Likelihood Tweets	84
Equação 43 – F_1 score	96

LISTA DE QUADROS

Quadro 1 – Visão geral de modelos teóricos de mudança de comportamento em saúde selecionados ligados à mídia social.	37
Quadro 2 – Limitações das mídias sociais para a comunicação em saúde entre o público em geral, pacientes e profissionais de saúde.	43
Quadro 3 – Os 56 principais sinais e sintomas da COVID-19 em português do Brasil. . .	53
Quadro 4 – Serviços e procedimentos utilizados para classificar procura de serviços por sinais e sintomas de COVID-19 (Etapa 2).	63
Quadro 5 – Construtos da HBM utilizados para classificar os tweets (Etapa 4).	71
Quadro 6 – Extensão dos Construtos da HBM para a COVID-19, para os tweets da amostra.	93

LISTA DE ABREVIATURAS

AIDS	<i>Acquired Immunodeficiency Syndrome</i>
API	Interface de Programação Aplicada (do inglês, <i>Application Programming Interface</i>)
BR	Brasil
CD	Ciências de Dados
CoDatMo	<i>Covid Data Model</i>
Dr	Doutor
Dra	Doutora
EN	<i>Ego networks</i>
EPIs	Equipamentos de Proteção Individual
ESPIN	Emergência de Saúde Pública de Importância Nacional
EUA	Estados Unidos da América
H1N1	Influenza A(H1N1)
HIV	<i>Human Immunodeficiency Virus</i>
Inc	<i>Incorporated</i>
MMS	Monitoramento de Mídias Sociais
NSF	<i>National Science Foundation</i>
ODE	<i>Ordinary Differential Equations</i>
OMS	Organização Mundial da Saúde
ONG	Organização não Governamental
PCR	Proteína C-Reativa
PIB	Produto Interno Bruto
Prof	Professor
SARS-CoV-2	<i>Severe Acute Respiratory Syndrome causada pelo Coronavirus 2</i>
SCA	<i>Social Capital Approach</i>
SG	Síndrome Gripal
SINAN	Sistema de Informação de Agravos de Notificação
SIR	<i>Susceptible, Infected and Recovered</i>
SIRD	<i>Susceptible Infected, Recovered and Deceased</i>
SIRTD	<i>Susceptible Infected, Recovered, Terminally Ill and Deceased</i>
SIVEP-Gripe	Sistema de Informação da Vigilância Epidemiológica da Gripe
SNA	<i>Social Network Analysis</i>
SMA	<i>Social Media Analytics</i>
SRAG	Síndrome Respiratória Aguda Grave

LISTA DE ABREVIATURAS

SSA	<i>Social System Approach</i>
SNVS	Secretária Nacional de Vigilância em Saúde
TMC	Teorias de Mudança de Comportamento
UK	<i>United Kingdom</i>
UNINOVE	Universidade Nove de Julho
WN	<i>Whole Networks</i>
WoS	<i>Web of Science</i>
WWW	<i>World Wide Web</i>

LISTA DE TABELAS

Tabela 1 – Casos e Óbitos de COVID-19 no Brasil, anos 2020 até Dezembro de 2022. . .	51
Tabela 2 – Ajuste dos parâmetros do classificador RL.	66
Tabela 3 – Amostra com dados rotulados como sinal ou ruído (Etapa 3) – ano 2020. . .	88
Tabela 4 – Distribuição (%) das pessoas mencionadas nos tweets segundo os construtos do Modelo de Crenças em Saúde. N = 4.000	90
Tabela 5 – Estatísticas dos modelos de <i>NLP</i> testados na pesquisa	95
Tabela 6 – Estatísticas dos modelo RL com dados de teste. N = 1.001	96
Tabela 7 – Matriz de confusão para os dados de teste do modelo de classificação.	97
Tabela 8 – Comparação entre o modelo sem tweets (1) e o modelo com tweets (2) . . .	101

LISTA DE SÍMBOLOS

LETRAS GREGAS

α Letra grega alfa

β Letra grega beta

SÍMBOLOS MATEMÁTICOS

\in É elemento (ou pertence a) de um conjunto

MOEDAS

\pounds Libra esterlina

LISTA DE MAPAS

Mapa 1 – Casos de SRAG por COVID-19 no Brasil em 2020.	52
Mapa 2 – Casos de de SRAG por COVID-19 (por milhão de habitantes) no Brasil em 2020.	52

INTRODUÇÃO

Resumo do capítulo

Neste capítulo é contextualizada a situação problema que envolve esta pesquisa. Na sequência, é apresentada a questão de pesquisa, além dos objetivos e a relevância do tema.

No final de 2019, o mundo presenciou o surgimento da maior crise sanitária desde a gripe espanhola de 1918, a qual levou à morte aproximadamente 50 milhões de pessoas (JOHNSON; MUELLER, 2002), a pandemia de COVID-19. O primeiro caso conhecido de COVID-19 surgiu em Wuhan, na China, em dezembro de 2019, e rapidamente a doença se espalhou pelo mundo, tornando-se uma pandemia que já levou à morte 6,92 milhões de pessoas (OMS, 2023a). Nos meses subsequentes, a COVID-19 sobrecarregou os sistemas de saúde pública e privada em quase todos os países, obrigando representantes de diversas nações a tomarem medidas de emergência para conter o avanço da doença, como proibir viagens, restringir a circulação em lugares públicos, entre outras medidas sanitárias. Essas medidas levaram ao fechamento de escolas e comércios, falência de empresas, lotação de hospitais, restrição de viagens, interrupções da vida cotidiana e profissional (CHIPIDZA et al., 2021). Não obstante, em meio à pandemia da COVID-19, pesquisadores já advertiam sobre as próximas pandemias (OMS, 2020; WWF, 2021; FUTURE AGENDA, 2021). A discussão não é se haverá outras pandemias, mas quando a próxima irá chegar.

No Brasil, o primeiro caso de COVID-19, confirmado pelo Ministério da Saúde (MS), aconteceu em 26 de fevereiro de 2020. O paciente, um homem com 61 anos de idade, era residente no estado de São Paulo e foi diagnosticado com o novo coronavírus (SARS-CoV-2) (Agência Brasil, 2020). Em 20 de março do mesmo ano foi declarada, em todo o território brasileiro, a transmissão comunitária da doença COVID-19 causada pelo novo coronavírus (Ministério da Saúde, 2020c). Desse modo, o Ministério da Saúde, por intermédio de sua Secretaria de Vigilância em Saúde (SVS), precisou realizar adaptações no Sistema de Vigilância de Síndromes Respiratórias Agudas, com o objetivo de orientar o Sistema Nacional de Vigilância em Saúde para a circulação simultânea do novo coronavírus (SARS-CoV-2), do vírus da influenza e de outros vírus respiratórios no âmbito da Emergência de Saúde Pública de Importância Nacional (ESPIN) (Ministério da Saúde, 2020c). Contudo, diante das dificuldades para conter o avanço da COVID-19, após um ano e meio de pandemia já tinham sido confirmados, no Brasil, mais de 21 milhões de casos e mais de 600 mil mortes pela COVID-19 (DATASUS, 2021).

Compreender como a sociedade se engaja em campanhas de saúde e quais fatores contribuem para uma comunicação eficaz, que gere mudanças de comportamento em saúde, é fundamental para prevenir e minimizar os impactos de surtos e epidemias. Mudanças de comportamento em saúde vão além de lavar as mãos, usar máscaras, praticar o distanciamento social e evitar aglomerações, apesar dessas medidas, durante a pandemia, serem as mais eficazes para a

redução da transmissão do vírus. As mudanças de comportamento em saúde dizem respeito ao apoio às políticas e aos programas executados pelas autoridades sanitárias oficiais no combate a surtos e propagação de doenças infecciosas (GULJAŠ et al., 2021).

A Organização Mundial da Saúde (OMS) ressalta que, para isso, deve-se considerar o nível de educação em saúde dos indivíduos e as complexidades dos contextos nos quais as pessoas atuam ou vivem (KICKBUSCH et al., 2013). Para a OMS, baixos níveis de educação em saúde estão associados a comportamentos de maior risco, saúde precária, maior nível de hospitalização e custos com a saúde. Portanto, é importante identificar maneiras efetivas de comunicar à sociedade sobre saúde, usando tanto as mídias tradicionais, como rádio e televisão, quanto as novas mídias, como *streaming* e redes sociais.

O fato é que as questões relacionadas à saúde têm ocupado uma posição central na sociedade e se tornaram o foco da atenção da população, empresas e governos (PAIM, 2020). As agências governamentais estão conscientes de que, para aprimorar as políticas de vigilância em saúde pública, necessitam aperfeiçoar a tomada de decisão baseada em dados e evidências (KRAHN, 2019), o que inclui melhorar a comunicação de saúde para combater eventos como a pandemia de COVID-19. Essa vigilância em saúde pública trata da coleta sistemática, contínua, armazenamento, análise e interpretação dos dados necessários para planejar, implementar e avaliar a prática de saúde pública (JAJOSKY; GROSECLOSE, 2004; OMS, 2017; GOSTIN; WILEY, 2018).

No Brasil, a legislação da vigilância em saúde pública foi introduzida na era pré-industrial, quando foi criado o Sistema Nacional de Vigilância Epidemiológica (SNVE) em 1975 (BRASIL, 1975). Nesse período, a lista de doenças era mais restrita e abordava políticas nacionais sobre saúde escolar, prevenção de doenças, saúde da mãe e da criança, água para consumo humano e tratamento de esgoto (NETTO et al., 2017; TEIXEIRA et al., 2018). No entanto, as chamadas doenças endêmicas, como esquistossomose, doença de Chagas e malária, entre outras, estavam sob responsabilidade da Superintendência de Campanhas (SUCAM), órgão sob a responsabilidade direta da administração do Ministério da Saúde (TEIXEIRA et al., 2018). Somente em 1991, com o desenvolvimento de um novo modelo de vigilância para o Sistema Único de Saúde (SUS), o sistema de saúde brasileiro passou a ter uma área de inteligência epidemiológica, separada das áreas de controle e prevenção de doenças (TEIXEIRA et al., 2018; PAIM, 2020).

Compete ao SUS, dentre outras atribuições, realizar ações de vigilância em saúde, conforme determina a Constituição Federal de 1988 (BRASIL, 2016; BRASIL, 2018). Em uma emergência de saúde pública, como a pandemia de COVID-19, que demanda medidas urgentes para prevenir, controlar e conter riscos, danos e agravos à saúde pública, o SUS deve implementar um conjunto de ações que possibilitem o conhecimento e a detecção de mudanças nos fatores determinantes e condicionantes da saúde individual e coletiva, de modo a recomendar e adotar medidas para a prevenção de problemas de saúde pública (BRASIL, 2018).

Considerando a pandemia de COVID-19, um dos principais objetivos da vigilância em saúde

pública é gerenciar os casos de COVID-19, incluindo a detecção, o isolamento e o monitoramento da evolução do vírus SARS-CoV-2 (SEKALALA et al., 2020). Nesse cenário, aplicações digitais começaram a emergir como uma ferramenta essencial para a resposta a surtos. Assim, com o objetivo de apoiar e facilitar o tratamento e o manejo da saúde individual e coletiva, e processos administrativos, atores de saúde têm adotado cada vez mais novas aplicações digitais que promovem a digitalização e interconectividade de processos, produtos, serviços e pessoas, como *big data*, computação em nuvem e Internet-das-Coisas (TORTORELLA et al., 2022). Em particular, as mídias sociais podem ser de grande utilidade como fontes de dados para complementar informações para a vigilância em saúde pública. Inclusive, já existem várias propostas para o uso de mídias sociais no apoio aos sistemas de vigilância de saúde pública, como o uso de dados de mídias sociais para realizar o monitoramento de informações oficiais (ZHANG et al., 2015); a detecção de surtos a partir dos sintomas mencionados (BRONIATOWSKI; PAUL; DREDZE, 2013; STOROPOLI et al., 2021a); ou a identificação de informações não oficiais, como *fake news* (SALATHÉ et al., 2013).

Na comunicação em saúde, há uma suposição de que o padrão de comunicação foi transformado pelos avanços da mídia social (CHOU et al., 2009). As mídias sociais estão mudando a natureza e a velocidade da interação entre indivíduos e organizações de saúde, e as pessoas estão usando essas mídias para se comunicarem sobre diversas questões de saúde. Tem havido muitas aplicações de mídias sociais em contextos de saúde, como, por exemplo, profissionais de saúde buscando informações para tomar ciência da qualidade de sua prática profissional (MOORHEAD et al., 2013), ou para o monitoramento de surtos de doenças, tais como o HIV/AIDS (NIELSEN et al., 2017) e a Dengue (ANTUNES et al., 2014), e na definição de políticas de saúde (SCHIAVO, 2014).

Para que haja uma comunicação em saúde eficaz, com o apoio das mídias sociais, as organizações e os profissionais de saúde precisam reconhecer e compreender o panorama atual da mídia social. Para isso, precisam ter certo grau de acesso sobre a comunicação *online* relativa à saúde, a fim de manter a validade e a confiabilidade dessa informação (MOORHEAD et al., 2013). Durante uma emergência em saúde pública, monitorar as mídias sociais envolve um processo de buscas sistemáticas e contínuas para recuperar conteúdo incidente relevante. Isso pode ser feito por meio de várias ferramentas, sejam elas gratuitas ou pagas. Com o uso de uma combinação de palavras-chave estratégicas e pesquisas geográficas, as informações das mídias sociais podem fornecer uma compreensão contínua das atividades de interesse, bem como avaliar, e, em alguns casos, responder diretamente às perguntas da população afetada (HADI; FLESHLER, 2016).

Interagir em plataformas de mídia social é a atividade *online* mais popular no mundo. Atualmente, há 3,78 bilhões de usuários de mídia social no mundo, um crescimento de 21% nos últimos quatro anos (STATISTA, 2021). Facebook, YouTube e Twitter são as plataformas mais populares no mundo. Os usuários de mídia social passam mais de 20% do tempo *online* interagindo, compartilhando e consumindo conteúdos nessas plataformas (FAN; GORDON, 2014).

Essa tendência de crescimento deve continuar, muito em função das ações dos órgãos públicos e da pressão da sociedade para a inclusão digital das comunidades mais carentes, do surgimento de novas plataformas, e devido à pandemia da COVID-19, que obrigou a população mundial a permanecer em isolamento físico em suas residências, o que contribuiu para acelerar a mudança de comportamento no que diz respeito ao consumo digital (SRISATHAN; NARUETHARADHOL, 2022).

Apesar da popularização de plataformas de mídia social, como Facebook e Twitter, fornecer uma ideia sobre o que são as mídias sociais, é importante não confundir "mídias sociais" com "redes sociais". O termo "mídia social" é geralmente aplicado para descrever as várias formas de conteúdo de mídia que estão publicamente disponíveis e são criadas por usuários finais na *Internet* em plataformas que permitem aos usuários interagirem e se comunicarem uns com os outros (KAPLAN; HAENLEIN, 2010; HADI; FLESHLER, 2016). Por outro lado, as "redes sociais" são as conexões, os grupos formados dentro dessas plataformas *online* (mídias), que são projetadas para permitir a interação social entre os seus usuários, que utilizam tecnologias móveis baseadas na *web* para criar ambientes altamente interativos. Por conseguinte, essas plataformas permitem que seus usuários criem as redes sociais de acordo com os seus interesses, para se comunicar, compartilhar conteúdo, criarem e modificarem o conteúdo criado por outros usuários (KIETZMANN et al., 2011).

O desenvolvimento e o uso crescente e exponencial da *Internet*, e o crescimento acelerado das plataformas de mídia social para a publicação de conteúdos relacionados a políticas públicas e troca de opiniões políticas relevantes, tem criado grandes oportunidades para o setor público explorar a riqueza do conhecimento disponível nessas mídias, a fim de apoiar o processo de inovação aberta neste setor (LOUKIS; CHARALABIDIS; ANDROUTSOPOULOU, 2017). Essas plataformas facilitam uma mudança de escala que torna a organização da ação coletiva, com grande número de participantes, mais eficiente (BEKKERS; EDWARDS; KOOL, 2013).

Recentemente, tem havido um interesse crescente em aproveitar o conhecimento compartilhado nas mídias sociais, a fim de explorar a "sabedoria coletiva" e minerar novas ideias de um grande número de indivíduos. Segundo Loukis, Charalabidis e Androutsopoulou (2017), as primeiras ações do setor público com vistas a aproveitar o conhecimento compartilhado nas mídias sociais foram realizadas de forma ativa, apresentando um problema social específico em sites ou mídias sociais do governo e solicitando que os cidadãos fornecessem seus conhecimentos e opiniões sobre o assunto. Posteriormente, segundo os autores, o processo passou a ser realizado de forma passiva, monitorando assuntos relevantes nas plataformas de mídia social, como *blogs* políticos, sites de notícias, Twitter, Facebook etc. Em seguida, são recuperados e analisados conteúdos sobre tópicos específicos livremente gerados pelos cidadãos, sem qualquer orientação ou estímulo governamental, a fim de extrair conhecimentos e opiniões relevantes dos cidadãos.

Na era do *Social Media Big Data*, as mídias sociais desempenham um papel fundamental para a nossa sociedade, fornecendo uma plataforma para comunicar, colaborar e compartilhar ideias e pensamentos. Dados de mídia social, como opiniões de consumidores, experiências e

sentimentos sobre vários produtos e serviços, ajudam a identificar as necessidades das pessoas e melhorar os valores das organizações. Porém, segundo Stieglitz e colegas, esses dados precisam ser melhor explorados (STIEGLITZ et al., 2018).

A coleta dos dados de mídia social e a análise desses dados para a tomada de decisões de negócios é chamada de *Social Media Analytics* (SMA). A SMA tem ganhado importância devido aos mais de 1,5 bilhão de usuários que utilizam regularmente plataformas como Facebook, Instagram, Twitter e LinkedIn (MURUGANANTHAM; GANDHI, 2020). A SMA tem como propósito desenvolver e avaliar ferramentas computacionais para coletar, monitorar, analisar, resumir e visualizar dados de mídia social. Os padrões encontrados nas informações provenientes desses dados podem facilitar as conversas e interações entre o público em geral e as organizações. A SMA pode fornecer uma quantidade significativa de informações relevantes sobre os usuários de mídias sociais. É possível, assim, identificar onde os usuários vivem, quais os seus hábitos, como interagem com as organizações etc. Esse tipo de estratégia baseada em SMA pode ser utilizada para os gestores se comunicarem e responderem rapidamente às preocupações do público (GÓMEZ-CARRASCO; GUILLAMÓN-SAORÍN; OSMA, 2021).

Com o aumento significativo de informações nessas plataformas de redes sociais, usar moderadores humanos para rastrear conteúdo não é mais uma estratégia viável, o que tem incentivado cada vez mais o uso de ferramentas de processamento de linguagem natural (*Natural Language Processing* – NLP) para automatizar os processos de identificação de linguagem, bem como monitorar a comunicação nessas plataformas (ANAND et al., 2023). O Processamento de Linguagem Natural contribui para a análise de dados de redes sociais, pois pode ser usado para extrair informações significativas de dados de texto não estruturados, como postagens e comentários. As técnicas de NLP podem ajudar a identificar sentimentos, tópicos e entidades mencionados no texto, que podem ser usados para análise e modelagem adicionais. O NLP também pode auxiliar na tradução de idiomas e resumo de texto, tornando-se uma ferramenta valiosa para analisar dados de mídia social de diversas fontes e idiomas (MARQUEZ; CARRASCO; CUADRADO, 2018). Nesta pesquisa, o NLP é utilizado com o propósito analisar dados de mídia social (Twitter) relacionados a sinais e sintomas de COVID-19.

1.1 PROBLEMA DE PESQUISA

O avanço, ao longo da história, das doenças transmissíveis trouxe a necessidade de novos estudos, que busquem entender a propagação desse tipo de doença, por meio de modelos matemáticos e estatísticos para descrever e prever a dinâmica dessas doenças, avaliar a eficácia de estratégias de controle e entender os fatores que influenciam suas propagações (YADAV; AKHTER, 2021). Dessa necessidade, surgiram os modelos matemáticos ou modelos epidemiológicos que permitiram descrever probabilisticamente as interações entre os indivíduos em uma população, levando em consideração fatores como taxa de transmissão da doença, período de incubação, suscetibilidade, recuperação e mortalidade. Esses modelos podem ser usados para

prever a progressão de uma doença, estimar o impacto de intervenções de saúde pública, como vacinação e quarentena, e identificar estratégias mais eficazes para o controle de doenças. Sobre tudo, quando surge uma nova doença como a COVID-19, onde tudo é muito novo, a única coisa que resta é lançar mão desses modelos matemáticos para fazer projeções. Por exemplo, se quiser saber quando os leitos de Unidade de Terapia Intensiva (UTI) vão saturar, não adianta pegar os dados registrados até o momento, traçar uma curva e tentar estender essa curva infinitamente para fazer projeções porque não é assim que a doença se comporta. Então, é preciso desenvolver um modelo que tenha capacidade de fazer previsões para auxiliar os tomadores de decisão (DAHER, 2020).

Ter maior compreensão do comportamento dinâmico das pandemias é a primeira etapa para poder controlá-las e os modelos matemáticos que mostram como acontece a disseminação dessas doenças em diferentes redes sociais é um fator importante na modelagem epidemiológica. Assim, uma das motivações em realizar estudos de modelos epidemiológicos e redes sociais é auxiliar os governos a impedir a disseminação da doença em grande escala (MANRIQUE, 2016). Contudo, o crescimento acelerado das redes sociais apresentam desafios conceituais e teóricos importantes para a área da saúde. As plataformas de redes sociais oferece uma forma dinâmica de disseminar informações de saúde, com um certo grau de interatividade e a comunicação para diferentes públicos, o que pode criar desafios particulares.

Ademais, as redes sociais contribuem para as estratégias em saúde pública, pois permitem atingir o público-alvo a um custo marginal relativamente mais baixo do que outras mídias, como por exemplo, a televisão (GÓMEZ-CARRASCO; GUILLAMÓN-SAORÍN; OSMÁ, 2021). Assim, a utilização das mídias sociais torna-se uma ferramenta importante para apoiar a gestão pública, com vistas à redução de custos, à maximização de resultados e ao desenvolvimento de novas políticas públicas. Para isso, é importante ter a atenção da academia e dos formuladores de políticas públicas, uma vez que o estudo sobre novas mídias, entre elas as mídias sociais, no combate a pandemias e crises sanitárias ainda é incipiente, principalmente no Brasil. Também não há um consenso, na literatura, sobre como utilizar de forma eficaz as mídias sociais no enfrentamento às pandemias ou como mensurar os resultados dessas ações.

1.2 QUESTÃO DE PESQUISA

Posto a problemática relacionada a esta pesquisa, este projeto visa responder à seguinte pergunta de pesquisa: **Como o processamento de linguagem natural e os dados de redes sociais podem ser utilizados para aprimorar modelos epidemiológicos?**

1.3 OBJETIVOS

1.3.1 OBJETIVO GERAL

Propor um modelo estatístico epidemiológico aprimorado com dados de redes sociais e aplicações de aprendizado de máquina para a predição da ocorrência de casos (ou de casos graves) de COVID-19.

1.3.2 OBJETIVOS ESPECÍFICOS

- I Analisar os fatores que levam as pessoas a assumirem determinados comportamentos de saúde, sob a ótica do Modelo de Crença em Saúde.
- II Propor um modelo epidemiológico otimizado com a incorporação de dados de mídias sociais para a análise da variação de casos graves notificados de COVID-19.
- III Fornecer ferramentas analíticas para o monitoramento de menções sobre sinais e sintomas relacionados à COVID-19 nas mídias sociais.

1.4 JUSTIFICATIVA PARA O ESTUDO DO TEMA

Prevenir é a chave para melhorar o bem-estar da população e reduzir os custos de saúde a longo prazo. O setor de saúde, na maioria dos países, representa uma área crítica de responsabilidade governamental, e o gasto mundial com saúde equivale a aproximadamente 10% do PIB global (AL-DMOUR *et al.*, 2020). Apenas nos Estados Unidos, 1.570.000 mortes poderiam ser evitadas por ano com programas de prevenção: 465.000 por tabagismo, 395.000 por hipertensão, 216.000 por sedentarismo, 191.000 por diabetes, 190.000 por colesterol alto e 113.000 por outros riscos (MURRAY; FRENK, 2010).

Uma das principais medidas preventivas adotadas durante a pandemia de COVID-19 foi o distanciamento social. Os benefícios dessa medida vão além de vidas salvas; por exemplo, nos Estados Unidos, as vidas salvas com o distanciamento social superaram o valor das perdas projetadas do PIB em cerca de US\$5,2 trilhões, além de reduzir o pico da curva de infecção pelo novo coronavírus em mais da metade (THUNSTRÖM *et al.*, 2020).

As estratégias de prevenção pautadas em campanhas de comunicação em saúde possuem um lugar de destaque nas políticas de saúde pública, seja por meio de programas de vacinação, desinfecção de espaços públicos e domiciliares ou divulgação de informações relevantes à população para a aplicação de medidas que garantam a proteção contra doenças (MARTINS; SPINK, 2019). Para desenvolver programas de comunicação em saúde eficazes, apenas a produção de mensagens não basta. Eles devem ser baseados na combinação de *marketing* social, educação em saúde e avaliação da comunicação (FLEISHER *et al.*, 2008). A comunicação em saúde é um componente essencial às ações de promoção da saúde, e, devido à sua relevância, as pesquisas sobre esse assunto têm se intensificado em diversos países.

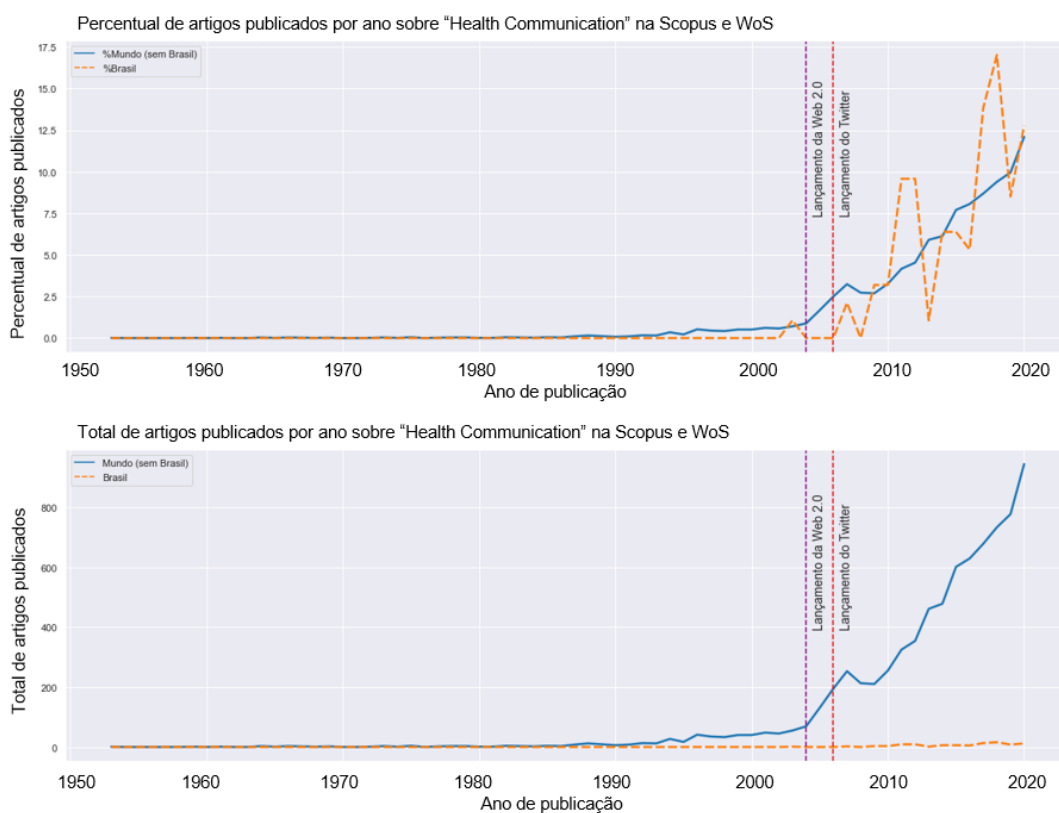
Independentemente do tipo de governo ou do estágio de desenvolvimento em que o país se encontra, em todos os continentes têm sido realizadas pesquisas sobre comunicação em saúde, por exemplo, na Bélgica (VERBEKE JACQUES VIAENE, 1999), na Índia (DANIEL; MASILAMANI; RAHMAN, 2008), nos Estados Unidos (SNYDER *et al.*, 2004; CHOU *et al.*, 2009), na Rússia (BRONIATOWSKI *et al.*, 2018) e no Brasil (NIELSEN *et al.*, 2017; MARTINS; SPINK, 2019). Porém, conforme exibido na Figura 1, é possível observar, no primeiro gráfico, (Percentual de artigos sobre Comunicação em Saúde publicados por ano em relação ao total de artigos publicados), que o Brasil tem acompanhado a tendência mundial, e a cada ano a quantidade de pesquisas publicadas sobre Comunicação em Saúde tem crescido. Contudo, quando se analisa o segundo gráfico (Total de artigos sobre Comunicação em Saúde publicados por ano), observa-se que a quantidade de publicações no Brasil é baixa, principalmente levando-se em conta a dimensão continental do país.

Considerada como um marco importante para a literatura da Comunicação em Saúde, a evolução da *Web* 1.0 para a *Web* 2.0 foi responsável pelo surgimento das plataformas de mídia social. Conforme se observa na Figura 1, a evolução no número de artigos publicados sobre comunicação em saúde após o surgimento da *Web* 2.0 cresceu significativamente a partir da década

de 2000. Especificamente o ano de 2004, considerado o ano de origem da *Web 2.0*, representa um marco importante para a comunicação em saúde (BERRY, 2007; BOULOS; WHEELER, 2007) e, posteriormente, o surgimento do Twitter e de outras plataformas contribuiu para o desenvolvimento de teorias e aplicações para a análise das comunicações em redes sociais que estão relacionadas à promoção da saúde.

A *Web 2.0* é a segunda geração da *World Wide Web* (WWW), e a principal diferença em relação à sua antecessora é que na *Web 2.0* o ambiente é mais dinâmico, permitindo que os usuários se comuniquem mais facilmente (GREENHOW; ROBELIA; HUGHES, 2009). As tecnologias que surgiram com a *Web 2.0* são revolucionárias, agindo como facilitadoras dos cuidados de saúde para todas as partes envolvidas, de organizações ao público mais leigo. Contudo, pesquisadores advertem que, apesar dessas tecnologias emergentes serem promissoras e potencialmente adequadas para muitas aplicações em saúde, testes e pesquisas adicionais são necessários para estabelecer modelos mais adequados para os diversos cenários de saúde (BOULOS; WHEELER, 2007).

Figura 1 – Frequências de pesquisas publicadas sobre Comunicação em Saúde no mundo e no Brasil, distribuídas por ano de publicação, de 1950 a 2020.



Fonte: *Scopus & Web of Science* (WoS), 2020.

Além das pesquisas com mídias sociais estarem aumentando ao longo do tempo, os órgãos públicos também estão gradativamente incorporando os dados dessas plataformas às estratégias

de saúde pública, principalmente para acompanhar a evolução de surtos de doenças infecciosas, como dengue (ANTUNES et al., 2014), HIV (NIELSEN et al., 2017), Ebola (HAZEL et al., 2022), Hanseníase (LAWRENCE et al., 2018) e a própria COVID-19 (ENDOMBA; BIGNA; NOUBIAP, 2020; XUE et al., 2021), entre outras doenças respiratórias infecciosas (STEIJVERS et al., 2021). Por exemplo, no estudo sobre a dengue (ANTUNES et al., 2014), observa-se que o número de tweets acompanha o crescimento do número de casos de dengue, o que sugere uma relação entre os rumores sobre a dengue e o aumento de número de casos registrados nos sistemas de vigilância da doença. O mesmo comportamento pôde ser observado durante a pandemia de COVID-19, objeto de análise desta pesquisa. Todavia, no Brasil, a exploração das mídias sociais pelos órgãos públicos para esse fim ainda é muito limitada, ou são feitas tentativas de replicar o que outros países já fizeram. Além disso, a mídia social não pode ser considerada como um fenômeno global, pois existem diferenças significativas entre as culturas de cada país que precisam ser respeitadas e consideradas nas pesquisas (BRANDÃO, 2014; AL-DMOUR et al., 2020), o que coloca o Brasil como um caso paradigmático a ser investigado (BRANDÃO, 2014).

Com o número de usuários de mídias sociais no Brasil equivalendo a 70% da população (havia 150 milhões de usuários de mídia social no Brasil em janeiro de 2021) (HOOTSUITE, 2021), dedicar recursos a essas mídias durante uma emergência não deve mais ser considerado apenas como uma alternativa, mas sim uma operação crítica para garantir que os órgãos competentes estejam cientes de todas as informações sobre os eventos à medida que ocorrem (HADI; FLESHLER, 2016). Um dos principais benefícios da utilização de mídias sociais na vigilância em saúde pública é que estas podem fornecer comunicação em tempo real a um custo relativamente baixo (MOORHEAD et al., 2013; LARANJO et al., 2015; HADI; FLESHLER, 2016; SU; XIAO, 2020; AL-DMOUR et al., 2020).

Outra vantagem dos dados de mídia social é que são altamente contextuais e em rede (AL-DMOUR et al., 2020), permitindo com que os usuários gerem discussões ponto a ponto de uma forma que não é possível com *sites* da geração da *Web* 1.0. No entanto, há uma forte necessidade de os provedores de saúde manterem um papel na comunicação em saúde nas mídias sociais, que é diferente do papel dos pacientes e do público em geral. Assim, as organizações precisam reconhecer e compreender o contexto da mídia social e onde as conversas sobre saúde já estão ocorrendo, desenvolver estratégias que sejam adequadas, descobrir com que frequência e quando devem entrar nas conversas e ficar atentos aos comportamentos dos usuários (MOORHEAD et al., 2013). Porém, apesar dos benefícios à saúde pública, o uso dessas plataformas por profissionais de saúde pública ainda enfrenta diversos desafios, dentre os quais os mais importantes são a detecção de surtos de doenças infecciosas, o monitoramento de emergências, a predição de tendências de doenças e a mensuração da conscientização e das respostas do público (AL-DMOUR et al., 2020).

Claramente, as mídias sociais têm benefícios: (1) durante as pandemias as pessoas estão, em geral, mais afastadas de eventos sociais, então têm mais tempo para acessar essas mídias; (2)

oferecem comunicação instantânea; (3) permitem encontrar rapidamente informações, notícias e eventos; (4) permitem encontrar pessoas, grupos e assuntos de interesse; (5) pode-se encontrar trabalhos (inclusive remotos), estabelecer ligações profissionais, divulgar trabalhos e até vender produtos. Mas, não se pode ignorar as limitações relacionadas à saúde pública. Quanto a essas limitações, ainda não se conhecem os efeitos de longo prazo para a saúde pública, e também há falta de informações sobre os usos, benefícios e restrições dessas plataformas à comunicação em saúde (MOORHEAD et al., 2013), à promoção e a vigilância em saúde pública (AL-DMOUR et al., 2020).

Destaca-se, ainda, que esta tese é um dos primeiros trabalhos no Brasil que se propõe a analisar as menções de sinais e sintomas relacionados à COVID-19 nas redes sociais como ferramenta para auxiliar os gestores da saúde pública no combate à pandemia da COVID-19. Sobre as evidências apresentadas e considerando, até o presente, o que já foi investigado, esta pesquisa se faz relevante pelos motivos a seguir:

- I São escassos, na literatura, trabalhos com a incorporação de dados de mídia social na previsão de surtos e epidemias.
- II No Brasil, os estudos sobre Comunicação em Saúde são limitados.
- III A mídia social é uma fonte auxiliar eficiente para o monitoramento de surtos e epidemias a um baixo custo.
- IV A utilização da mídia social possibilita uma comunicação em saúde eficaz ao permitir direcionar a comunicação a um grande público de forma rápida e segmentada.
- V Não há consenso, na literatura, de quais são as melhores maneiras de utilizar as mídias sociais na saúde pública.

Assim, acredita-se que as descobertas desta pesquisa são relevantes para as autoridades públicas de saúde, de modo que possam compreender melhor o impacto das intervenções em saúde. Além de as mensagens nas redes sociais determinarem como a informação, os sinais e sintomas e a doença são percebidos pela população.

1.5 ESTRUTURA DA TESE

Esta tese está estruturada em cinco capítulos. No capítulo inicial (Introdução) é apresentado o problema e a questão de pesquisa, junto com os objetivos e a relevância do tema. Na sequência, no capítulo 2, apresenta-se a fundamentação teórica que norteia esta pesquisa: comunicação em saúde, mídias sociais, modelos epidemiológicos e modelos de *machine learning* para processamento de linguagem natural.

No capítulo 3, são elencados os procedimentos metodológicos e as ferramentas utilizadas para a análise dos dados coletados, que serviram de apoio para atingir os objetivos propostos. No capítulo 4, são apresentados e discutidos os resultados e achados da pesquisa, tendo como referência o arcabouço teórico do capítulo dois.

No capítulo 5, são apresentadas as considerações finais, limitações e principais contribuições da pesquisa, assim como sugestões para trabalhos futuros.

Por fim, seguem os apêndices e anexos da tese. Nos apêndices encontram-se a lista de emojis e outra lista com as expressões encontradas nos dados da pesquisa que foram utilizadas para otimizar o modelo de *machine learning*. No apêndice são apresentados, ainda os programas computacionais utilizados para executar os modelos. No anexo A são relacionados as teorias de mudança de comportamento investigadas nesta pesquisa.

REFERENCIAL TEÓRICO

Resumo do capítulo

A seção a seguir apresenta uma revisão da literatura sobre mídias sociais no apoio à comunicação em saúde no combate às pandemias e outras crises sanitárias. Na primeira subseção, apresento a revisão da literatura sobre a comunicação em saúde, destacando como esta tem sido utilizada na gestão de riscos e no enfrentamento às crises públicas, tais como desastres naturais, epidemias e pandemias. Na subseção seguinte, discorro sobre a literatura que corrobora a utilização das mídias sociais como apoio à tomada de decisão, mostrando os prós e as limitações da utilização desse meio de comunicação. Por fim, apresento como os modelos epidemiológicos têm sido utilizados no combate às pandemias, em particular a COVID-19.

2.1 COMUNICAÇÃO EM SAÚDE**2.1.1 TEORIAS DE MUDANÇA DE COMPORTAMENTO EM SAÚDE NO CONTEXTO DAS MÍDIAS SOCIAIS**

Estilos de vida não saudáveis, como tabagismo, sedentarismo, dieta inadequada, entre outros, têm sido considerados alguns dos principais fatores de riscos para a mortalidade em todo o mundo, e hoje há evidências de que otimizar comportamentos ligados à saúde é um fator fundamental para a prevenção de doenças. No entanto, a mudança de comportamento em saúde ainda é um dos maiores desafios da atualidade, e a melhor abordagem para promover mudanças de comportamentos em saúde ainda não foi identificada. Porém, já se sabe que a mudança de comportamento depende das motivações de cada pessoa. Assim, é importante criar estratégias que promovam o envolvimento e o engajamento dos indivíduos com os cuidados de saúde (LARANJO, 2016).

As pessoas estão cada vez mais aptas a usar as tecnologias da informação para ajudá-las a tomar decisões informadas sobre cuidados de saúde, e o número de pessoas que estão utilizando as mídias sociais para esse tipo de finalidade está aumentando rapidamente (LARANJO, 2016). Com a popularização das mídias sociais digitais, surgiu um interesse crescente em utilizar essas plataformas na promoção da saúde pública. O comportamento e as decisões de saúde dos usuários dessas mídias são influenciados por fatores sociais, políticos e culturais demograficamente específicos, especialmente em termos de costumes, crenças e opiniões dos seus círculos sociais, bem como por experiências pessoais e estado de saúde atual. Nesse sentido, as teorias de mudança de comportamento em saúde, quando aplicadas a iniciativas de mídia social, podem ajudar as autoridades de saúde a compreender os motivos pelos quais alguns grupos sociode-

mográficos resistem em seguir algumas medidas preventivas (GULJAŠ et al., 2021).

A mudança de comportamento em saúde é relevante em diversas áreas, desde a prevenção primária de doenças crônicas até o combate a surtos de doenças, e as teorias de mudança de comportamento (TMC) conseguem explicar por que, quando e como ocorre ou não um comportamento, bem como seus fatores influenciadores (LARANJO, 2016). Essas teorias em saúde estão fortemente ligadas à promoção em saúde (NIEUWENHUIJSEN et al., 2006), e são teorias desenvolvidas para explicar as diferenças observadas no comportamento em saúde de maneira a ser possível orientar o processo de intervenções em saúde (RAVESLOOT et al., 2011).

Intervenção no tipo de comportamento em saúde pode ser definida como um conjunto de ações coordenadas para alterar condutas de risco à saúde, a fim de aumentar a adesão às políticas públicas de saúde e, conseqüentemente, aumentar a adoção de comportamentos de saúde (NIEUWENHUIJSEN et al., 2006; PALMEIRA et al., 2007; WALSH; GROARKE, 2019). Apesar de as TMC em saúde terem como objetivo identificar os fatores que levam uma pessoa ou um grupo a assumirem um determinado padrão de saúde, até o presente, poucos estudos mencionaram teorias ou modelos de comportamento de saúde subjacentes às intervenções em mídias sociais (LARANJO et al., 2015; NIEUWENHUIJSEN et al., 2006; WALSH; GROARKE, 2019). Contudo, há evidências de que intervenções em saúde baseadas em teorias levam a melhores resultados (WALSH; GROARKE, 2019). Assim, considerando a importância dessas teorias em saúde, e que vários modelos já foram propostos para explicar a mudança de comportamento em intervenções em saúde – Modelo de Crença em Saúde, Teoria Cognitiva Social¹, Modelo Trans-teórico de Mudança de Comportamento (PROCHASKA; VELICER, 1997), Teoria do Comportamento Planejado (AJZEN, 1991), Teoria da Autodeterminação¹ (RYAN; DECI, 2000), entre outros –, a expectativa é que o número de pesquisas sobre intervenções em saúde apoiadas em mídias sociais e fundamentadas em teorias aumentem.

Há várias razões para a utilização de TMC na saúde pública, podendo ser divididas de acordo como as formas de prevenção em (i) prevenção primária, (ii) prevenção secundária, (iii) prevenção terciária e (iv) prevenção quartenária. A prevenção primária diz respeito aos aspectos preventivos dos comportamentos em saúde que levam à promoção da saúde, ao bem-estar e à prevenção primária de uma doença, desordem ou lesão. Enquanto a prevenção primária tem como objetivo promover ações focadas na promoção de saúde, a prevenção secundária tem como objetivo realizar a detecção precoce de doenças possivelmente já existentes nos indivíduos, além do tratamento, com vistas a evitar complicações, sequelas e óbitos. Contudo, em certas situações que não é mais possível realizar o trabalho de prevenção primária ou secundária, em que, por exemplo, já foi constatado a transmissão comunitária de uma doença, a forma imediata de prevenção possível é a terciária. Por fim, a última forma de prevenção, a quartenária tem como objetivo identificar pacientes em risco de sobremedicalização e sugerir intervenções eticamente aceitáveis (NIEUWENHUIJSEN et al., 2006; TESSER; NORMAN, 2021).

¹Teorias mais abrangentes do comportamento humano e não são especificamente teorias da comunicação em saúde (PALMEIRA et al., 2007).

Como apresentado na Figura 1 (ver introdução, pg. 30), o número de pesquisas que utilizaram teorias de comunicação em saúde, entre elas as TMC, aumentaram significativamente. Destas pesquisas a maioria tem sido realizada no escopo da atenção primária à saúde, com os mais diversos objetivos, entre eles o de identificar os fatores que influenciam a mudança de comportamento em saúde a fim de projetar intervenções efetivas (NIEUWENHUIJSEN et al., 2006), e normalmente focam no estudo de uma única teoria. Contudo, pesquisadores ressaltam que pesquisas baseadas em evidências sobre comunicação em saúde não suportam uma teoria exclusiva de TMC (NIEUWENHUIJSEN et al., 2006; BERRY, 2007; SCHIAVO, 2014; LARANJO, 2016). Assim, comparar e contrastar teorias existentes para analisar os diferentes pontos de vista é promissor na identificação de fatores que influenciam o processo de mudança de comportamento e na segmentação desses fatores para intervenção (NIEUWENHUIJSEN et al., 2006).

De acordo com o tipo de intervenção a ser aplicada, é necessário utilizar uma ou outra teoria (ver exemplo no Quadro 1), ou mesmo vários construtos de teorias de comportamento de saúde diferentes (PALMEIRA et al., 2007). Inclusive, na literatura da comunicação em saúde, há modelos que utilizam construtos de outras teorias de comunicação em saúde (PROCHASKA; VELICER, 1997). Por exemplo, o Modelo Trans-teórico de Mudança de Comportamento (PALMEIRA et al., 2007) usa vários construtos de outras teorias, assim, esse modelo consegue oferecer uma visão mais ampla, identificando quando, como e por que os indivíduos mudam seu comportamento. Todavia, as maneiras informais com que as TMC são a ser utilizadas limitam sua comparação direta e sua integração. Assim, o avanço na aplicação dessas teorias (ver lista das teorias no anexo A) requer que sejamos capazes de compará-las em termos de seu conteúdo, escopo e previsões a fim de selecioná-las, integrá-las e modificá-las (WEST et al., 2019).

Dada essa breve introdução sobre as teorias de mudança de comportamento em saúde, a seguir é apresentado o Modelo de Crença em Saúde (ROSENSTOCK, 1974a; ROSENSTOCK, 1974b; ROSENSTOCK; STRECHER; BECKER, 1988) que tem apoiado pesquisas sobre intervenções em saúde nas mídias sociais digitais, bem como este trabalho.

Quadro 1 – *Visão geral de modelos teóricos de mudança de comportamento em saúde selecionados ligados à mídia social.*

Modelo teórico	Foco	Conceitos-chave	Questões relacionadas à mídia social
Modelo de Crenças em Saúde	Os comportamentos preventivos podem ser explicados pela percepção da ameaça de um problema de saúde e pela avaliação do comportamento recomendado para prevenir ou gerenciar o problema.	<ul style="list-style-type: none"> – Suscetibilidade percebida – Gravidade percebida da condição – Benefícios perceptíveis – Barreiras percebidas – Dicas para ação – Auto-eficácia 	<ul style="list-style-type: none"> – Explore o nível de suscetibilidade. – Quais são os benefícios e as barreiras para a implementação do programa? – Que plano pode ser implementado para superar essas barreiras? – Que tipo de reforços podem ser usados para estimular o comportamento desejado?
Teoria Sócio-Cognitiva	A mudança de comportamento de saúde é o resultado de uma interação recíproca entre fatores pessoais, influências ambientais e comportamentos	<ul style="list-style-type: none"> – Determinismo recíproco – Capacidade de comportamento – Auto-eficácia – Expectativas de resultado – Aprendizagem por observação – Reforço 	<ul style="list-style-type: none"> – Discuta o que fazer e como fazer. – A pessoa acredita na capacidade de controlar sua condição de saúde? – Incentive a crença na capacidade de ter sucesso ao realizar ações de saúde. – Estabeleça pequenas metas incrementais. – Observe os outros e use comportamentos de "modelos" que são desejáveis.
Modelo Trans-teórico de Mudança	Mudança de comportamento de saúde é um processo de prontidão dos indivíduos para mudar ou tentar mudar em direção a comportamentos saudáveis.	<ul style="list-style-type: none"> – Pré-contemplação – Contemplação – Preparação – Ação – Manutenção 	<ul style="list-style-type: none"> – Avalie a prontidão de uma pessoa para a mudança e sustentação de comportamentos saudáveis. – A pessoa está ciente dos fatores de risco relacionados à saúde ao realizar certas tarefas? – Em caso afirmativo, o que a pessoa está fazendo a respeito? – Quais são os prós e contras do tratamento ou intervenção? – Adapte as intervenções ao estágio de mudança da pessoa.

Fonte: Adaptado de Nieuwenhuijsen et al. (2006).

2.1.2 MODELO DE CRENÇA EM SAÚDE

As ações de uma pessoa dentro de um ambiente são um elemento crítico à mudança de comportamento em saúde (NIEUWENHUIJSEN et al., 2006). Da qualidade do ar, às predisposições genéticas e escolhas pessoais levam a um desfecho de saúde específico, fatores ambientais e pessoais que o modelo de crença em saúde considera como determinantes à promoção da saúde (RAVESLOOT et al., 2011).

O modelo de crença em saúde estabelece que as pessoas são mais propensas a seguirem medidas preventivas se acreditarem que: (i) a doença pode ser um risco à vida, (ii) são suscetíveis a contraírem a doença, (iii) as medidas preventivas as beneficiariam ao reduzir o risco de desenvolver uma doença ou a forma grave de uma doença, e (iv) que os benefícios das medidas preventivas superam os riscos à saúde (LARANJO, 2016).

Com base nos conceitos do modelo de crença em saúde, Guljaš et al. (2021) avaliaram a autopercepção de um grupo de 276 pessoas (56 homens e 220 mulheres) residentes na cidade de Osijek, na Croácia, em relação à vacinação contra a COVID-19, considerando o período compreendido antes do início da vacinação na cidade, mas quando a vacina já estava sendo disponibilizada rapidamente em todo o mundo. Uma análise de regressão múltipla foi aplicada para demonstrar a influência de fatores individuais na decisão de vacinar contra a COVID-19, e os autores observaram percepções distintas sobre o risco de contrair a doença e sobre a decisão de tomar a vacina.

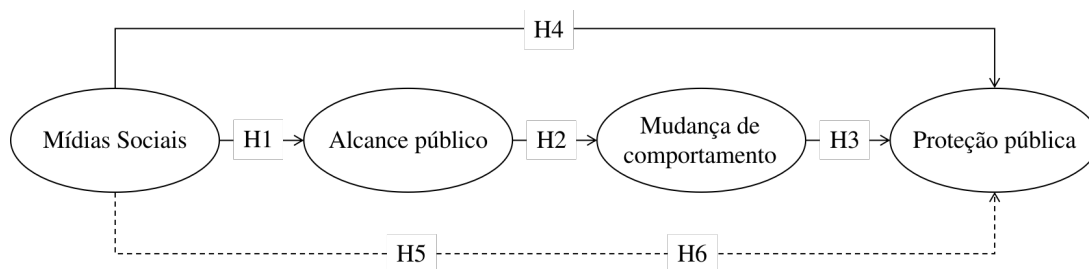
No que diz respeito ao perigo de contrair a doença COVID-19, Guljaš et al. (2021) observaram que a percepção dos entrevistados é que conseguiriam superar a pandemia sem consequências mais graves. Assim, apesar de considerarem a vacinação contra a COVID-19 importante, a percepção sobre a vacina foi negativa, ou seja, acreditavam que haveria mais malefícios do que benefícios em se vacinarem. Além disso, os autores observaram que os padrões cognitivos mentais relativos à associação entre o risco de consequências da doença e a percepção da segurança da vacina influenciam na decisão de tomar a vacina. Assim, com base nos construtos do modelo de crença em saúde, os resultados mostraram que, para grupos específicos, é importante adaptar a estratégia de abordagem, a fim de promover uma mudança de comportamento em favor de tomar a vacina.

Comportamentos adversos à promoção de saúde, por exemplo, a decisão de não tomar vacina, como observado no estudo de Guljaš et al. (2021), podem ser estimulados a mudanças com o auxílio de campanhas de promoção à saúde nas mídias sociais e de acordo com o modelo de crença em saúde (LARANJO, 2016). Para isso, é importante que a mensagem consiga transmitir claramente os benefícios que uma determinada mudança de comportamento pode trazer para a saúde, ou quais os perigos presentes em não se adotar determinados comportamentos em saúde.

A fim de analisar os contextos nos quais os indivíduos assumem um determinado comportamento de saúde, Al-Dmour et al. (2020) desenvolveram um modelo conceitual integrado

(Figura 2), com base nos construtos do modelo de crença em saúde. Nesse modelo, as variáveis de interesse estudadas foram as plataformas de mídia social (variáveis independentes) e a proteção pública contra a COVID-19 (variável dependente), e por meio de fatores mediadores (consciência pública e mudanças comportamentais), foi estudada a influência das variáveis independentes no combate à COVID-19.

Figura 2 – Diagrama do modelo conceitual integrado com base no modelo de crença em saúde.



H1: O uso de plataformas de mídia social está aumentando significativamente a conscientização da saúde pública.

H2: A conscientização da saúde pública está contribuindo significativamente para a mudança de comportamento em saúde pública.

H3: A mudança de comportamento em saúde pública está aumentando significativamente a proteção da saúde pública.

H4: O uso de plataformas de mídia social está aumentando significativamente a proteção da saúde pública.

H5: Alcance público está mediando significativamente a relação entre as plataformas de mídia social e a conscientização da população sobre a proteção da saúde pública.

H6: A mudança de comportamento da saúde pública está mediando significativamente a relação entre o uso de plataformas de mídia social e a conscientização da saúde pública.

Fonte: adaptado de Al-Dmour et al. (2020).

Para confirmar o modelo conceitual (Figura 2), uma pesquisa foi aplicada em uma amostra de 2.555 usuários de mídias sociais. Os respondentes tinham entre 18 e 44 anos, sendo 50,2% mulheres e, do total, 47,7% possuindo nível superior, todos residentes em Amã, capital da Jordânia. As mídias sociais avaliadas foram Facebook, Instagram, Twitter, WhatsApp e YouTube. Das seis hipóteses da pesquisa, os resultados forneceram evidências empíricas sobre o impacto do uso de plataformas de mídia social na conscientização da saúde pública, mudanças de comportamento em saúde pública e proteção à saúde contra a COVID-19 para as hipóteses H1, H3 e H4.

2.2 MÍDIAS SOCIAIS

2.2.1 *Social Media Analytics (SMA)*

O uso crescente das mídias sociais faz com que uma grande quantidade de dados seja gerado e acumulada nessas plataformas, o que é descrito como *Social Media Big Data* ou simplesmente *Big Data*, e que tem aberto novas oportunidades para analisar vários aspectos e padrões de comunicação (STIEGLITZ et al., 2018). Na era do *Social Media Big Data*, as mídias sociais desempenham um papel fundamental para a nossa sociedade, fornecendo uma plataforma para comunicar, colaborar e compartilhar pensamentos. Dados de mídia social, como opiniões de consumidores, experiências e sentimentos sobre vários produtos e serviços, ajudam a identificar as necessidades das pessoas e a melhorar os valores das organizações. Porém, segundo os pesquisadores, esses dados precisam ser explorados com mais detalhamento (STIEGLITZ et al., 2018).

A coleta de dados de mídia social e a análise desses dados para tomar decisões de negócios é chamada de *Social Media Analytics (SMA)*. A SMA tem ganhado importância devido aos mais de 1,5 bilhões de usuários que utilizam regularmente plataformas como Facebook, Instagram, Twitter e LinkedIn (MURUGANANTHAM; GANDHI, 2020). A SMA tem como propósito desenvolver e avaliar ferramentas computacionais e estruturas para coletar, monitorar, analisar, resumir e visualizar dados de mídia social. Os padrões encontrados nas informações provenientes desses dados podem facilitar conversas e interações entre o público em geral e as organizações. Além disso, permitem identificar onde esses usuários de mídias sociais vivem, quais os seus hábitos, o que os influenciam, como interagem com as organizações etc. Esses *insights* permitem que os tomadores de decisão refinem suas estratégias e tenham uma percepção mais apurada sobre sua audiência (FAN; GORDON, 2014).

Até ao momento, os gestores podem usar as mídias sociais para abordar, remodelar e reenquadrar as preocupações do público em geral, e mídias sociais como o Twitter são ferramentas poderosas para a gestão dos *stakeholders*² externos (GÓMEZ-CARRASCO; GUILLAMÓN-SAORÍN; OSMA, 2021). Dentre essas plataformas, considera-se que o Twitter pode desempenhar um papel importante na gestão da informação ao permitir identificar usuários que podem atuar como filtro de informação, permitindo encontrar diretamente a informação mais relevante para um determinado tema de interesse (ANTUNES et al., 2014). Para fins de comunicação corporativa, o Twitter é particularmente popular, pois oferece três vantagens. Primeiro, os tweets são públicos por padrão: as conversas acontecem na esfera pública, tornando-se de conhecimento comum. Em segundo lugar, as postagens são limitadas em número de caracteres, facilitando a identificação do conteúdo principal. Por fim, as funções são limitadas, o que simplifica a compreensão dos processos de comunicação (GÓMEZ-CARRASCO; GUILLAMÓN-SAORÍN; OSMA, 2021).

²São todas as pessoas que são afetados, diretamente ou indiretamente, pelas ações dos tomadores de decisão.

2.2.2 MÍDIAS SOCIAIS PARA A COMUNICAÇÃO EM SAÚDE

Moorhead et al. (2013) identificaram os usos, benefícios e as limitações das mídias sociais para a comunicação em saúde. Por meio de uma revisão sistemática da literatura em 11 bases de dados (CSA Illumina, Cochrane Library, Communication Abstracts, EBSCO Host CINAHL, ISI Web of Knowledge, Web of Science, OvidSP Embase, OvidSP MEDLINE, OVIDSP PsycINFO, PubMeb Central e *Journal of Medical Internet Research*), e utilizando como termo de busca as expressões: “social media” OR “social network” OR “social networking” OR “Web 2.0” OR “Facebook” OR “Twitter” OR “MySpace” AND “Health”, os autores identificaram 9.773 artigos relevantes. Os principais critérios para a seleção dos artigos foram: (1) que tivessem o foco principal em todas as interações de comunicação dentro e entre o público em geral e/ou pacientes e/ou profissionais de saúde sobre questões de saúde usando a mídia social; (2) que incluíssem os usos e/ou benefícios e/ou limitações da mídia social para a comunicação em saúde; e (3) estudos com foco principal em marketing ou publicidade; todos em inglês, publicados entre 2002 e 2012. Desses artigos, 98 foram selecionados para a amostra, dos quais 48 utilizaram um método de pesquisa qualitativo, 40 quantitativo e 10 utilizaram uma abordagem mista. Considerando apenas os estudos quantitativos e mistos, os autores avaliaram que, no geral, esses estudos eram principalmente exploratórios e descritivos, assim, receberam uma baixa pontuação.

Moorhead et al. (2013) também identificaram uma característica bem específica em relação ao público e à comunicação nas mídias sociais. As principais mídias sociais utilizadas nas pesquisas foram Facebook, blogs, Twitter e YouTube, sendo que nos blogs e fóruns os participantes incluíam, além do público em geral, pacientes e profissionais de saúde. A faixa etária nessas mídias sociais estava concentrada entre 11 e 34 anos, com a maior presença do público feminino. Em relação aos assuntos, havia uma ampla variedade de tópicos de saúde em discussão, contudo, os mais frequentes eram sobre saúde sexual, diabetes, gripe/H1N1 e problemas de saúde mental, como depressão ou estresse. Os autores também relatam algumas controvérsias em relação ao perfil dos públicos que acessam as redes sociais. Enquanto algumas pesquisas apontam características demográficas e socioeconômicas específicas na utilização das mídias sociais (KONTOS et al., 2010; SCANFELD; SCANFELD; LARSON, 2010), outras concluíram que o acesso às mídias sociais independe dos aspectos culturais ou econômicos (CHOU et al., 2009). Os resultados apontados por Moorhead e coautores, em relação às características do público que utiliza as mídias sociais, evidenciam a necessidade de mais estudos sobre o tema em questão.

Para Moorhead et al. (2013), cada tipo de mídia social pode servir para um propósito na comunicação em saúde. Os autores identificaram sete objetivos de se utilizar as mídias sociais para esse tipo de comunicação:

1. Fornecer informações de saúde em uma variedade de condições para o público em geral, pacientes e profissionais de saúde (SANFORD, 2010; HU; SUNDAR, 2010; LIANG;

SCAMMON, 2011).

2. Fornecer respostas às perguntas médicas (BEARD et al., 2009).
3. Facilitar o diálogo entre pacientes e entre pacientes e profissionais de saúde (COLINEAU; PARIS, 2010; BOSSLET et al., 2011).
4. Coletar dados sobre as experiências e opiniões dos pacientes (ADAMS, 2010; LAGU et al., 2010).
5. Utilizar para intervenção em saúde, promoção e educação em saúde (DOWDELL; BURGESS; FLORES, 2011; USHER, 2011; YOUNG; RICE, 2011).
6. Reduzir o estigma sobre certas condições de saúde (LO; ESSER; GORDON, 2010).
7. Fornecer consultas *online* (GAJARIA et al., 2011).

Além dos propósitos das mídias sociais para a comunicação em saúde, Moorhead et al. (2013) destacaram seis benefícios do uso dessas mídias para a comunicação em saúde, considerando o público em geral, pacientes e profissionais de saúde. Segundo os autores, o fluxo de comunicação nas mídias sociais pode auxiliar no monitoramento à resposta pública aos problemas de saúde, rastrear e monitorar surtos de doenças, identificar áreas-alvo para esforços de intervenção e disseminar informações de saúde pertinentes às comunidades-alvo. Outro benefício dessas mídias é a possibilidade de difundir mensagens personalizadas, tornando o alcance mais eficaz. Para os autores há, ainda, a possibilidade de se beneficiar das mídias sociais para o desenvolvimento de políticas de saúde, uma vez que os perfis dos médicos nessas mídias sociais são seguidos por um público expressivo.

Apesar dos benefícios das mídias sociais para a comunicação em saúde, é importante que os gestores, ao implementarem estratégias de comunicação em saúde, fiquem atentos às limitações e aos problemas em potencial que essas mídias podem causar. Moorhead et al. (2013) alertam para doze principais limitações desse meio de comunicação (Quadro 2). Primeiro, a falta de confiabilidade das informações de saúde. Como não há um órgão regulatório que fiscalize a qualidade das informações, torna-se difícil para os indivíduos discernirem sobre a confiabilidade das informações encontradas *online*. Deve-se tomar cuidado com a vulnerabilidade das informações, principalmente no que diz respeito a registros médicos oficiais. Para que haja uma comunicação em saúde eficaz, é necessário que haja treinamento em como usar e navegar pelas tecnologias de mídia social e treinamento para mostrar como validar as informações encontradas, ou como buscar fontes mais confiáveis.

Ainda, a grande quantidade de informações disponíveis nos sites de mídias sociais é outro desafio para a validação das informações. Somado a isso, os autores advertem sobre os problemas de privacidade, confidencialidade e segurança dos dados, principalmente dados pessoais. Há também a possibilidade de sobrecarga de informações para o usuário, e o público pode aplicar de forma inapropriada para a saúde as informações encontradas *online*. Além disso, pode

Quadro 2 – Limitações das mídias sociais para a comunicação em saúde entre o público em geral, pacientes e profissionais de saúde.

Limitações das mídias sociais para a comunicação em saúde	Usuários de mídia social		
	Público em geral	Pacientes	Profissionais de saúde
Falta de confiabilidade	✓	✓	✓
Preocupações com a qualidade	✓	✓	✓
Falta de confiabilidade e privacidade	✓	✓	✓
Risco em divulgar informações pessoais	✓	✓	
Sobrecarga de informação	✓	✓	
Informações prejudiciais ou incorretas	✓	✓	
Incerteza em como aplicar as informações encontradas	✓	✓	
Certas mídias sociais podem ser mais eficazes na mudança de comportamento do que outras	✓	✓	
Consequências adversas para a saúde	✓	✓	
Comportamentos de saúde negativos	✓	✓	
A mídia social pode atuar como pretexto para não ir ao médico		✓	✓

Fonte: adaptado de Moorhead et al. (2013).

haver comportamentos negativos de risco à saúde exibidos *online*, inclusive, há evidências limitadas de que o envolvimento em comunidades *online* influenciam positivamente a saúde das pessoas, muito em virtude de que os profissionais de saúde, médicos, por exemplo, não costumam interagir com seus pacientes e o público em geral.

2.3 MODELAGEM MATEMÁTICA

A maioria dos modelos de doenças é baseada na compartimentação de indivíduos de acordo com o *status* da doença. Os modelos compartimentais surgiram no início do século XX, com o modelo *Susceptible-Infectious-Recovered* (SIR) e, desde então, têm sido utilizados no controle de doenças infecciosas (HERNANDEZ-CERON; FENG; CASTILLO-CHAVEZ, 2013; GOUNANE et al., 2021). Essa compartimentação permite gerar uma representação matemática

tica simplificada da progressão da infecção, onde muitos detalhes são negligenciados, como as diferenças de respostas entre os indivíduos. As etapas mais básicas, presentes na maioria dos modelos, procuram descrever o número de indivíduos que são suscetíveis, infectados e recuperados de uma determinada doença. Esse tipo de modelo permite simplificar a complexidade de uma epidemia ao subdividir a população total em grupos homogêneos, denominados compartimentos. Assim, esses modelos dividem a população em três compartimentos dependentes: suscetíveis, infectados e recuperados; dessa forma, indivíduos que estão em um mesmo compartimento são considerados em um mesmo estágio quanto à progressão da doença (KERMACK; MCKENDRICK, 1927). As suposições básicas presentes nos modelos geram dois conjuntos padrão de equações diferenciais que fornecem os fundamentos de quase toda a modelagem matemática: o modelo suscetível-infeccioso-recuperado (SIR)

$$SIR = \begin{cases} \frac{dS}{dt} &= bN - \lambda S - dS, \\ \frac{dI}{dt} &= \lambda S - gI - dI, \\ \frac{dR}{dt} &= gI - dR, \end{cases} \quad (1)$$

e o modelo suscetível-infeccioso-suscetível (SIS).

$$SIS = \begin{cases} \frac{dS}{dt} &= gI - \lambda S, \\ \frac{dI}{dt} &= \lambda S - gI. \end{cases} \quad (2)$$

O modelo SIR é apropriado para doenças infecciosas, enquanto o modelo SIS é predominantemente utilizado para doenças sexualmente transmissíveis (W.O.KERMACK; MCKENDRICK, 1927; KEELING; EAMES, 2005). Nas equações SIR e SIS, S , I e R referem-se, respectivamente, ao número de indivíduos suscetíveis, infecciosos e recuperados em uma população de tamanho N . Já os outros parâmetros são a taxa de natalidade, b , a taxa de mortalidade natural, d , e a taxa de recuperação da infecção g . Por fim, lambda (λ) é a taxa na qual os indivíduos que são suscetíveis acabaram se infectando. No caso desta pesquisa, lambda contém informações sobre as interações entre os indivíduos que levam à transmissão da infecção por SARS-CoV-2.

Outro modelo clássico da literatura da matemática epidemiológica é o modelo SEIR (susceptível, exposto, infectado, recuperado) (AREA et al., 2015), no qual temos as seguintes equações diferenciais:

$$SEIR = \begin{cases} S'(t) = -\frac{\beta S(t)(qE(t)+I(t))}{N}, \\ E'(t) = \frac{\beta S(t)(qE(t)+I(t))}{N} - \delta E(t), \\ I'(t) = \delta E(t) - \gamma I(t), \\ R'(t) = \gamma I(t). \end{cases} \quad (3)$$

A primeira diferença do modelo SEIR para os outros dois modelos (SIR e o SIS) é a inclusão do compartimento de pessoas expostas (E) ao vírus. Outro parâmetro incorporado neste modelo é parâmetro beta (β), que é a média do número de pessoas infectadas a partir do contato com um indivíduo sintomático que foi infectado pelo vírus. No entanto, o vírus também pode ser transmitido por pessoas assintomáticas que estão infectadas ($E(t)$). Além disso, o parâmetro gamma (γ) é a taxa de mortalidade *per capita* da região (AREA et al., 2015), no Brasil, em dezembro de 2020, por COVID-19, era de 110 por 100 mil habitantes (RONCATO; ANDRADE, 2021). Já o parâmetro delta (δ) representa a taxa de infecção *per capita* e, no Brasil, essa taxa de infecção pelo novo coronavírus chegou a 5,3% na população em geral e 7,3% nos profissionais da área da saúde (COFEN, 2021).

Contudo, algumas doenças infecciosas, como a COVID-19, podem ser fatais. Para esse tipo de doença, é importante diferenciar os casos recuperados daqueles que foram a óbito. Para a modelagem estatística desses tipos de doenças, foi desenvolvido o modelo *Susceptible-Infectious-Recovered-Deceased* (SIRD) (BAILEY et al., 1975). Considerando que a COVID-19 pode esgotar rapidamente o sistema de saúde de um país, como, por exemplo, o SUS no caso do Brasil, sobrecarregando os leitos de Unidade de Terapia Intensiva (UTI) (NETO et al., 2021), nesta tese foi atribuído um estado que representa os pacientes terminais (que vão a óbito). Assim, o modelo proposto inclui também um estado T para os indivíduos com doença terminal que foram infectados e possuem alta probabilidade de falecerem. A sigla torna-se, então, *Susceptible-Infectious-Recovered-Terminally-Ill-Deceased* (SIRTD).

Foi por meio dos modelos de doenças apresentados que os governos em diversas partes do mundo conseguiram definir suas estratégias para lutar contra a pandemia de COVID-19, e as pesquisas científicas foram fundamentais para ajudar nesse processo. Por exemplo, usando dados de casos notificados de COVID-19 na China, Liu et al. (2020a) examinaram o período de latência da infecção por SARS-CoV-2, ou seja, o período de tempo em que os indivíduos recém-infectados permanecem assintomáticos e não infecciosos. Em seu estudo, os autores utilizaram como base o modelo SEIR e usando equações diferenciais desenvolveram dois modelos. Um modelo contenha uma classe exposta de indivíduos infectados, que ainda não são infecciosos, e outro com um atraso de tempo em indivíduos recém-infectados, antes de se tornarem infecciosos. O desenvolvimento dos modelos tinha como objetivo entender como as medidas de isolamento, quarentena e fechamentos públicos, reduzem o tamanho da pandemia. Para atingir esse objetivo, Liu et al. (2020a) consideraram um período de latência nos dois modelos, com indivíduos recém-infectados, mas que não eram nem sintomáticos nem infecciosos. Dentre os resultados da pesquisa os autores identificaram que o período médio de exposição varia entre 6 e 12 horas, o que, segundo os autores, é consistente com Liu et al. (2020b), onde a alta carga viral observada para a COVID-19 foi usada para explicar que a transmissão do vírus SARS-CoV-2 pode acontecer no estágio inicial da infecção.

Em suma, os modelos epidemiológicos desempenham um papel fundamental no estudo e no controle de doenças infecciosas, fornecendo uma base matemática para entender a dinâmica da

disseminação de doenças em uma população. Ao longo do tempo, vários modelos foram desenvolvidos para representar diferentes aspectos das doenças, desde os modelos básicos até os mais complexos que levam em consideração fatores demográficos, comportamentais e ambientais.

A abordagem Bayesiana tem se destacado na estatística aplicada e na modelagem epidemiológica devido ao avanço computacional e à disponibilidade de dados. Essa abordagem permite incorporar informações prévias e atualizar o conhecimento à medida que novos dados são coletados, fornecendo uma maneira flexível e poderosa de analisar e estimar parâmetros em um modelo estatístico.

Apesar do progresso na utilização de modelos epidemiológicos e da estatística Bayesiana, há ainda lacunas a serem exploradas. Portanto, o avanço contínuo na modelagem epidemiológica, a incorporação da abordagem Bayesiana e a exploração de novas áreas de pesquisa são essenciais para melhorar nossa compreensão das doenças infecciosas e desenvolver estratégias eficazes de prevenção e controle. O uso de abordagens integradas, que combinam dados de mídias sociais, teorias de comunicação em saúde, informações demográficas e de saúde, bem como técnicas estatísticas avançadas, pode fornecer *insights* valiosos para lidar com desafios futuros na área da epidemiologia.

Resumo do capítulo

Para superar os desafios científicos e tecnológicos, nesta tese foram utilizados diferentes meios e métodos. Contudo, em alinhamento com as pesquisas que abordaram o mesmo tema (por exemplo, Yin et al. (2020), Melo e Figueiredo (2021)), ela foi fundamentada na ciência de dados, considerando que: (1) é adequada para dar sentido a grandes conjuntos de dados interconectados; (2) é direcionada à pesquisa interdisciplinar, uma vez que é menos restrita a limites teóricos a priori; (3) o raciocínio adutivo é encorajado; e (4) modelos holísticos e teorias sobre sistemas complexos, ao invés de seus elementos, são possíveis (RIZK; ELRAGAL, 2020).

3.1 DESAFIOS CIENTÍFICOS E TECNOLÓGICOS

Em termos de sistemas de vigilância para a detecção de doenças, pesquisadores acreditam que é neste campo que veremos um aumento significativo de investimentos (AL-DMOUR et al., 2020; FUTURE AGENDA, 2021). Porém, economias emergentes consideram o monitoramento e o rastreamento de doenças particularmente desafiadores. São vários os obstáculos, dentre eles a escassez de profissionais capacitados e a pobreza extrema. Como resultado, em alguns casos, os recursos para monitorar doenças são extremamente limitados. Especialistas advertem, ainda, que as próximas pandemias acontecerão na próxima década, mas a dificuldade é saber onde surgirão e que forma tomarão.

Contudo, a análise de redes sociais, como proposto nesta pesquisa, têm se mostrado um campo promissor para os sistemas de vigilância, visto que cada vez mais as pessoas vão às redes sociais para compartilhar fatos e experiências, o que pode ser útil para os sistemas de vigilância em saúde, como, por exemplo, a identificação das regiões com maior incidência de casos suspeitos. Assim, o desenvolvimento de sistemas inteligentes com dados de redes sociais pode contribuir para identificar as origens, as áreas de concentração e a propagação dos surtos. Diante do exposto, os principais desafios estão na resposta rápida e eficaz ao surto inicial, e em garantir que as epidemias não se transformem em pandemias (FUTURE AGENDA, 2021).

Na ocorrência de surtos de novas doenças, a compreensão sobre a doença aumenta com o tempo, o que pode evidenciar lacunas no conhecimento inicial, que, por sua vez, pode contribuir para a disseminação de desinformação e notícias falsas. A desinformação pode ser uma fonte de riscos em pandemias, com as inúmeras opções de fontes de notícias representando um desafio para a comunicação em saúde pública (CHIPIDZA et al., 2021). Para mitigar a desinformação em saúde, é importante compreender os gatilhos que levam as pessoas a acreditarem em uma determinada comunicação. Um dos mecanismos que pode contribuir para entender es-

ses gatilhos é analisar a comunicação em saúde nas redes sociais sob a ótica das teorias de comunicação em saúde. Nesta pesquisa, isso é feito por meio da teoria de crença em saúde, que permite identificar os tipos de crenças que as pessoas possuem sobre uma determinada doença ou tratamento.

Para superar os desafios apresentados, este projeto conta com uma parceria internacional pelo consórcio CoDatMo, entre as Universidades UNINOVE/BR, Columbia/EUA e Liverpool/UK (CODATMO, 2021). Cada instituição é independente para realizar suas próprias pesquisas e inovações, porém, há uma colaboração mútua e troca de conhecimento e tecnologia, o que faz deste um consórcio colaborativo e transparente. O consórcio CoDatMo possui custeio internacional nos EUA por meio do *grant* da *National Science Foundation* (NSF, 2020) e no Reino Unido por meio do *grant* da *COVID-19 Task Force*. No Brasil, é apoiado pela UNINOVE e recebe auxílio da CAPES (CAPES, 2021). A equipe do Brasil possui 10 membros, entre eles a epidemiologista, professora Dra. Alessandra Pellini, pesquisadora associada e coorientadora deste projeto de Doutorado, e o professor Dr. José Eduardo Storopoli, orientador deste projeto.

3.2 DELINEAMENTO DA PESQUISA

3.2.1 PARADIGMA DA CIÊNCIA DE DADOS

Novos paradigmas científicos surgem devido a: (a) revolução na medição, ou seja, como novos mecanismos usados para medir o desempenho da pesquisa podem promover o desenvolvimento de condições propícias à crescente estagnação na produção de novos conhecimentos; (b) avanços na geração de dados, coleta, gerenciamento e métodos analíticos; e (c) limitações existentes na habilidade dos paradigmas em avançar na descoberta do conhecimento e no desenvolvimento de teorias, ou não abordar questões-chave do assunto, como acontece com as abordagens tradicionais que não são mais suficientes para analisar grandes quantidades de dados (*big-data*) (GENDRON, 2013). Dentre as razões destacadas, o avanço na geração de dados é o principal fator para o surgimento do paradigma da ciência baseadas em dados, onde se encontra a Análise de Redes Sociais (do inglês: *Social Network Analysis – SNA*) como uma subárea de pesquisa, a qual é aplicada neste trabalho.

A SNA emerge da Ciência de Dados (CD), onde se pressupõe que nem todas as disciplinas desfrutam de uma visão unificada das teorias, dos pressupostos filosóficos e metodológicos aceitos. Especialmente nas disciplinas de Ciências Sociais e Sistema da Informação (SI), uma gama diversificada de suposições filosóficas são empregadas para dar sentido aos fenômenos sociotécnicos. São nessas suposições, tradições das ciências naturais, sociais, artificiais e nas suposições das epistemologias positivista, interpretativa e crítica que as pesquisas de CD se baseiam (RIZK; ELRAGAL, 2020), incluindo esta pesquisa.

Estudiosos defendem duas visões bem distintas que envolvem a SNA, uma focada na análise total da rede – *whole networks* (WN) e outra focada no indivíduo – *ego networks* (EN)

(KIRSCHBAUM, 2019; RIZK; ELRAGAL, 2020). O WN é anterior ao EN, e é focado na abordagem que investiga o sistema social – *social system approach* (SSA). Os primeiros estudos envolvendo a SSA ocorreram na década de 1970; contudo, foi a consolidação da utilização das mídias sociais digitais que permitiu que pesquisas com a SSA ganhassem força e fossem realizados estudos com inteligência artificial em uma série de direções, entre elas análise de conflitos *online* e leis sociais (TENNENHOLTZ, 1994).

Todavia, umas das principais barreiras em utilizar a SSA é o fato dessa abordagem ter como objetivo estudar toda a rede social (seus atores e conexões), ou seja, todo o sistema, o que dificulta que ela ganhe escala no conceito de *big data* em vista do alto custo computacional. Na década de 1970, quando começou a ganhar escala e ser mais utilizada nas ciências sociais, como as populações de interesse eram relativamente pequenas, se comparadas com o período pós *big data*, então utilizar essa abordagem se justificava, pois era relativamente fácil de aplicar em um conjunto de dados pequeno. Porém, com o surgimento das plataformas de mídias sociais e a geração de dados na grandeza numérica nas casas dos *terabytes*, utilizar essa abordagem tornou-se algo complexo, que demanda grande esforço, além de gerar um elevado custo computacional (KIRSCHBAUM, 2019; RIZK; ELRAGAL, 2020).

Diante das limitações iniciais da SSA, surgiu a abordagem focada no capital social – *social capital approach* (SCA). A SCA é uma abordagem estruturada para análises focadas nos indivíduos (atores), e não em todo o sistema, como acontece na SSA; e esta é a principal vantagem da SCA (ADLER; KWON, 2002). O fato de ser estruturada no conteúdo das relações sociais do ator, ou seja, concentrar-se no estudo do indivíduo ou em grupos específicos, reduz o investimento em recursos, facilita o reconhecimento de padrões, a geração de conhecimento, a replicação e ampliação para novas pesquisas. Assim, esta pesquisa se vale da SCA para analisar individualmente as características das postagens realizadas no Twitter por cada ator da amostra; com isso procura-se reconhecer os padrões presentes com base nos construtos do Modelo de Crença em Saúde (ROSENSTOCK, 1974a). Por outro lado, ao analisar um grande conjunto de dados coletados nesta pesquisa, é possível inferir sobre a população de interesse, ou seja, casos de COVID-19 no Brasil.

3.2.2 *Framework* DA PESQUISA

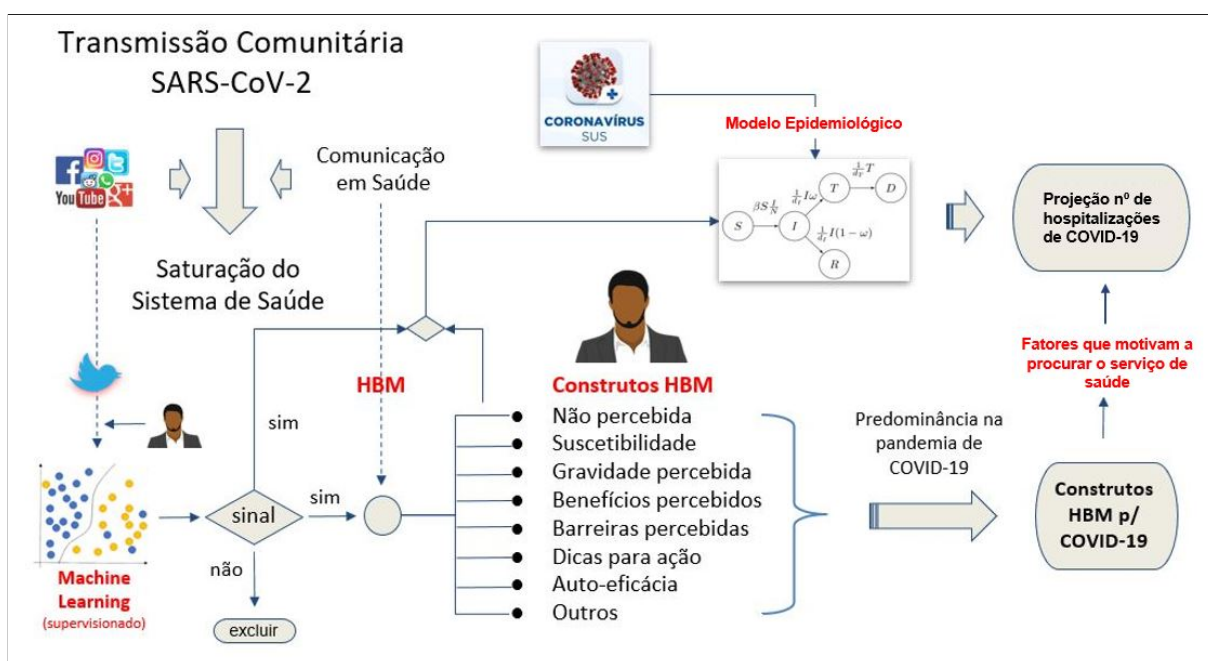
Esta pesquisa pode ser representada de forma integrada, conforme exhibe a Figura 3. Primeiro, dada a transmissão comunitária do SARS-CoV-2 no Brasil, buscou-se os dados de redes sociais, especificamente do Twitter, sobre menções de sintomas relacionados à COVID-19. Do total dos tweets coletados foi gerada uma amostra para a classificação manual, que posteriormente foi utilizada no modelo de processamento de linguagem natural (PLN) para classificar, automaticamente, todos os demais dados coletados.

Esses dados foram classificados em *signal* – dados de interesse para pesquisa, ou *ruído* – dados que deveriam ser excluídos da pesquisa. Após a aplicação do modelo supervisionado

de classificação, foi gerada uma nova amostra dos tweets rotulados como *signal*. Essa nova amostra foi analisada com base na teoria de comunicação em saúde, a *Health Belief Model* (ROSENSTOCK, 1974a). Com isso, o objetivo era entender quais construtos da HBM eram predominantes nos dados, e com isso entender quais eram os principais fatores (construtos da HBM) que motivaram as pessoas a procurarem ou não os serviços de saúde e que, consequentemente, podem ter levado à saturação dos sistemas de saúde público e privado. Da mesma forma pretendia-se avançar na literatura da HBM estendendo os seus construtos no contexto da pandemia de COVID-19.

Por fim, os dados de mídia social previamente classificados como *signal* foram incorporados em um modelo estatístico epidemiológico Bayesiano (STOROPOLI et al., 2021a), o qual teve como objetivo prever o número de hospitalizações por COVID-19, com base nos registros de casos confirmados de COVID-19 da Vigilância.

Figura 3 – Principais etapas para o desenvolvimento desta pesquisa.



Nota: Na seção Apêndice está disponível um tutorial de como implementar, via linguagem de programação, os modelos propostos neste *Framework* da pesquisa.

Fonte: Elaborado pelo autor.

3.3 DADOS DA PESQUISA

3.3.1 FONTES DE DADOS

Nesta pesquisa foram utilizados dados secundários, estruturados e semiestruturados. Os dados semiestruturados são comumente usados para oferecer suporte aos dados de mídia social

em plataformas de *Internet*, como o Twitter (CHANG; CHUA, 2019), no caso desta pesquisa seriam os tweets. Já os dados estruturados auxiliam os epidemiologistas a entenderem as relações entre os potenciais fatores de risco para doenças e sua morbimortalidade, permitindo o aprimoramento de investigações epidemiológicas, o controle de doenças e eficientemente, ajuda os gestores de políticas públicas a direcionarem recursos de saúde com base em cenários e relações concretas de morbidade (INÁCIO et al., 2016). Nesta pesquisa em questão representaram os dados de casos graves de COVID-19.

O estudo abrangeu o período de 1 de janeiro de 2020 até 31 de dezembro de 2020. No dia 26 de fevereiro de 2020, foi registrado o primeiro caso de COVID-19 no Brasil (Ministério da Saúde, 2020b). Vale ressaltar que esse é o período que antecedeu o início da vacinação contra a COVID-19 no Brasil. Comparando-se os casos e óbitos no ano referente a esta pesquisa, 2020, com os anos subsequentes (2021 e 2022) os casos e óbitos (Tabela 1), o ano de 2020 representou 28% de todas as mortes por COVID-19 no Brasil, e 21% do total de casos. Esses resultados mostram o quão representativo foi o ano de 2020 na pandemia de COVID-19 no Brasil, e que, mesmo após a campanha de vacinação, os números de casos e óbitos continuaram a crescer.

Tabela 1 – Casos e Óbitos de COVID-19 no Brasil, anos 2020 até Dezembro de 2022.

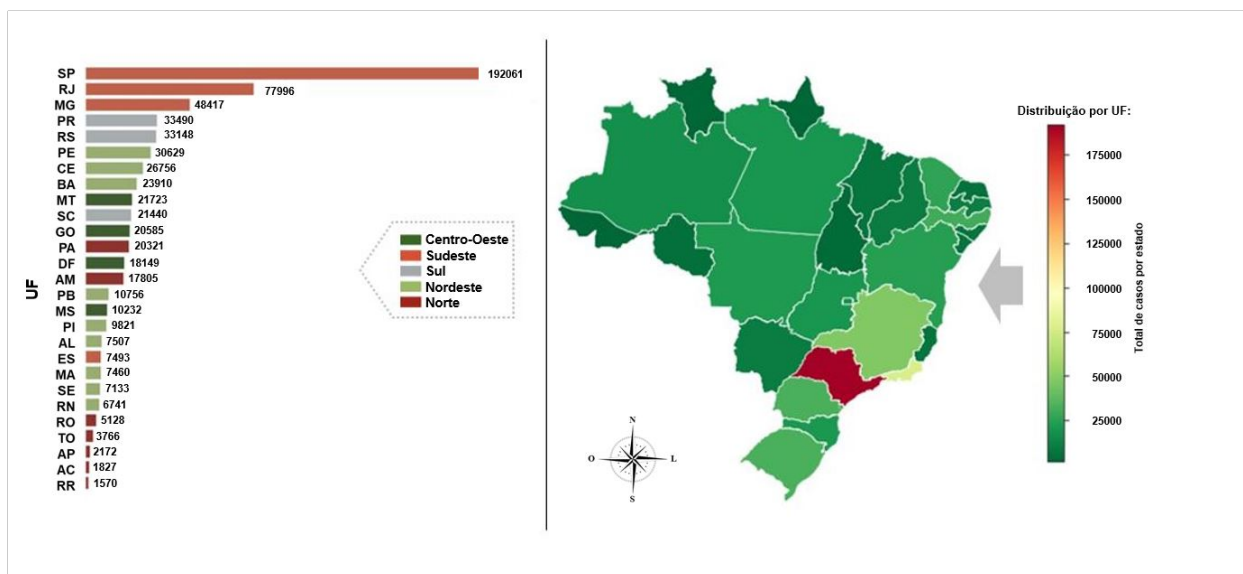
Região	População	Casos Acumulados		Óbitos Acumulados	
		2020	(2021 e 2022)	2020	(2021 e 2022)
Totais	210,1M	7,6M	36,3M	194,9K	693,8K
Sudeste	88,3M	2,6M	14,4M	89,2K	333,3K
Nordeste	57,0M	1,8M	7,2M	47,7K	133,8K
Sul	29,9M	1,3M	7,7M	22,0K	109,8K
Norte	18,4M	859,1K	2,8M	18,0K	51,3K
Centro-Oeste	16,2M	873,7K	4,1M	17,8K	65,5K

$K = 10^3$; $M = 10^6$; 2020 = valores registrados no ano de 2020;
(2021 e 2022) = valores registrados entre janeiro de 2021 até dezembro de 2022.

Fonte: Ministério da Saúde (2023).

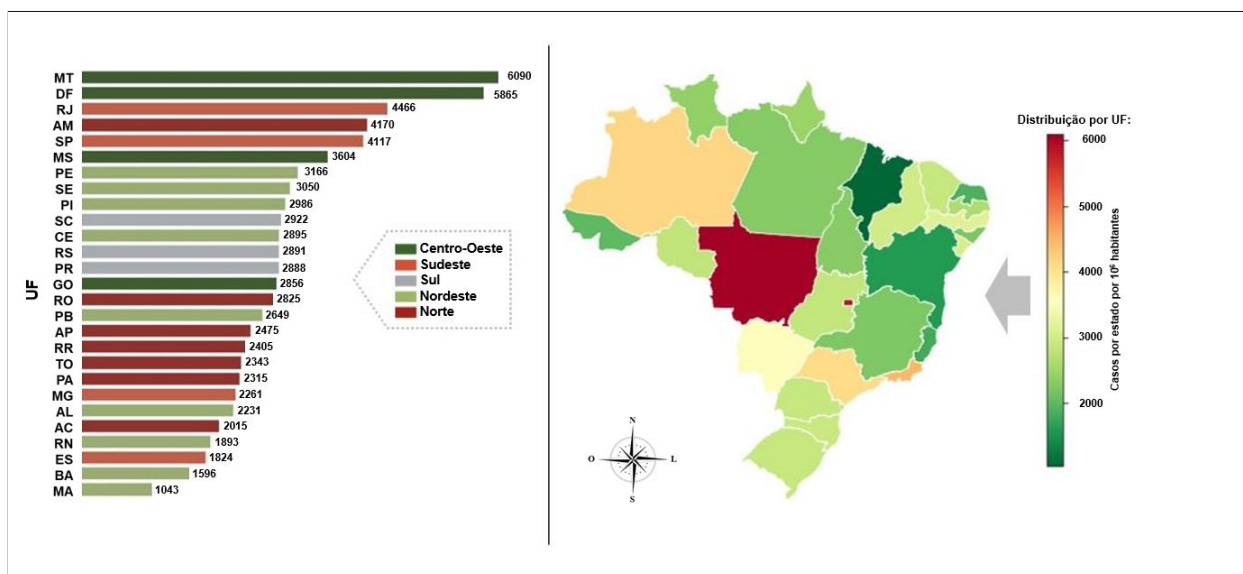
A área de estudo desta pesquisa é composta pelas 26 Estados do Brasil e o e o Distrito Federal (todos com casos confirmados de COVID-19). A maioria dos casos registrados por COVID-19 estão na região Sudeste (Tabela 1), sendo que, em 2020, os estados com maior número de notificações por COVID-19 (Mapa 1) foram São Paulo, com 30%; Rio de Janeiro, 10%; e Minas Gerais, com 9,4% (Ministério da Saúde, 2021b). Porém, ao considerar o total de casos de SRAG por COVID-19 nas Unidades Federativas por habitantes, é possível observar (Mapa 2) que o maior índice, em 2020, foi na região Centro-Oeste (Mato Grosso e Distrito Federal).

Mapa 1 – Casos de SRAG por COVID-19 no Brasil em 2020.



Fonte: Ministério da Saúde (2023).

Mapa 2 – Casos de de SRAG por COVID-19 (por milhão de habitantes) no Brasil em 2020.



Fonte: Ministério da Saúde (2023).

Todos os dados utilizados nesta pesquisa foram obtidos de fontes públicas, incluindo dados de governo – Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe) (Ministério da Saúde, Secretaria de Vigilância em Saúde, 2023) – e dados de mídia social – Twitter (TWITTER, 2020). O SIVEP-Gripe foi implantado em 2000 para o monitoramento do vírus influenza no país, a partir de uma rede de vigilância sentinela da síndrome gripal (SG) e, a partir disso, o Ministério da Saúde vem fortalecendo a vigilância dos vírus respiratórios. Com a declaração da transmissão comunitária da doença pelo SARS-CoV-2 em todo o território nacional, a Secretaria de Vigilância de Síndromes Respiratórias Agudas realizou a adaptação do

SIVEP-Gripe para adequá-lo à circulação no novo coronavírus. Posto isso, em primeiro lugar, nesta pesquisa, as fontes oficiais do governo foram analisadas e apenas os registros de casos graves e óbitos confirmados e registrados de COVID-19 no SIVEP-Gripe foram considerados. Cada caso e óbito foi agregado em séries temporais de listas por região e foi indicado onde as informações demográficas (por exemplo, idade, etnia, sexo) e clínicas (por exemplo, sintomas, hospitalização) estavam disponíveis. Em segundo lugar, foram incluídos dados de contas de usuários de acesso público de mídia social (Twitter) e seus respectivos tweets sobre os sinais e sintomas de COVID-19, em português do Brasil. Os textos livres sobre sinais e sintomas de COVID-19 foram identificados e compilados em variáveis padronizadas de acordo com a lista de sinais e sintomas presentes na Quadro 3.

Quadro 3 – *Os 56 principais sinais e sintomas da COVID-19 em português do Brasil.*

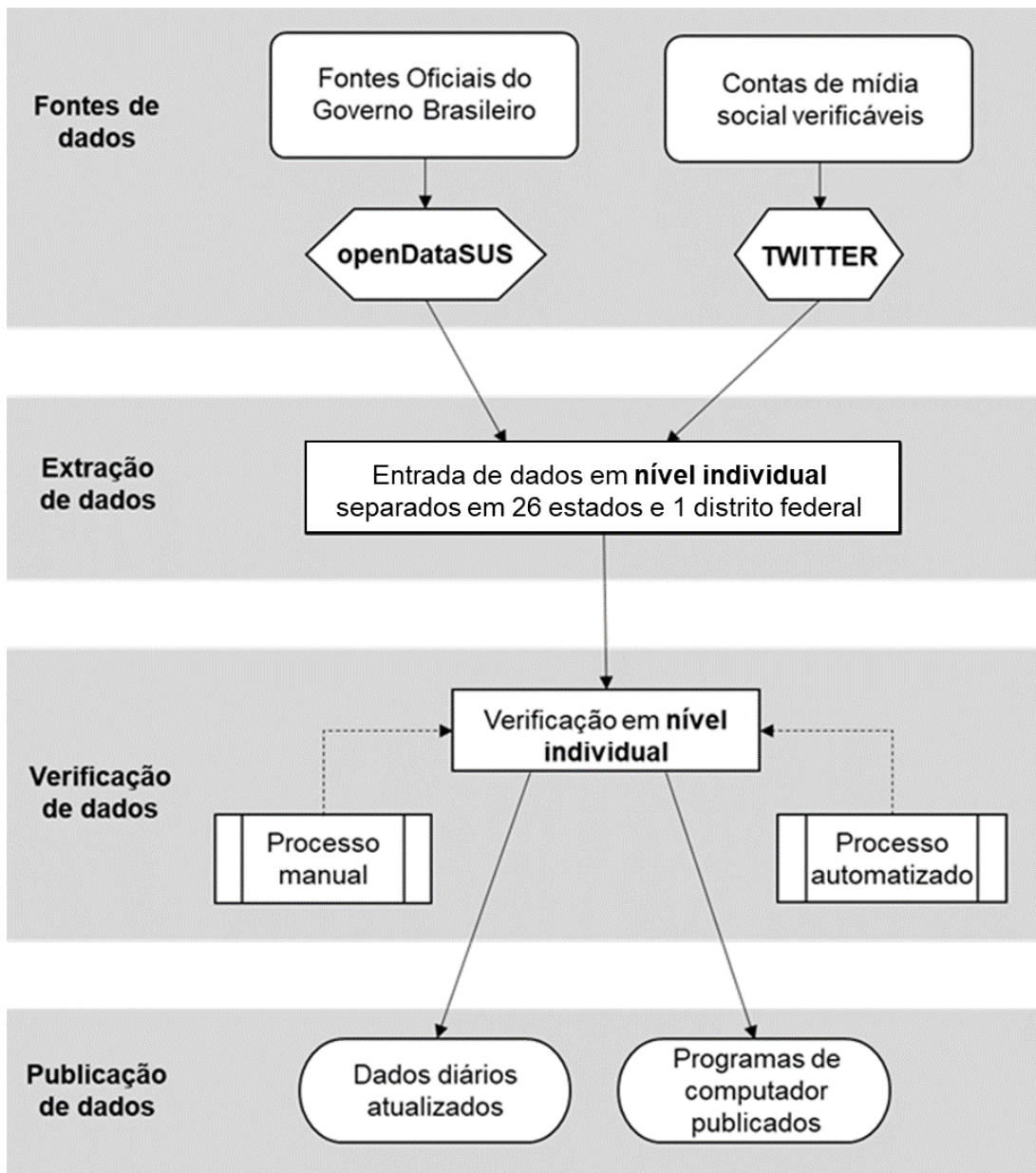
Sintomas	Código	Sintomas	Código
Adinamia	sym01	Febre	sym29
Ageusia	sym02	Gripe	sym30
Anosmia	sym03	Hiporexia	sym31
Boca azulada	sym04	Inapetência	sym32
Calafrio	sym05	Infecção respiratória	sym33
Cansaço	sym06	Lábio azulado	sym34
Cefaleia	sym07	Mialgia	sym35
Cianose	sym08	Nariz entupido	sym36
Coloração azul no rosto	sym09	Náusea	sym37
Congestão nasal	sym10	Obstrução nasal	sym38
Conjuntivite	sym11	Perda de apetite	sym39
Coriza	sym12	Perda de olfato	sym40
Desconforto respiratório	sym13	Perda do paladar	sym41
Diarreia	sym14	Pneumonia	sym42
Dificuldade para respirar	sym15	Pressão no peito	sym43
Diminuição do apetite	sym16	Pressão no tórax	sym44
Dispneia	sym17	Prostração	sym45
Distúrbio gustativo	sym18	Quadro gripal	sym46
Distúrbio olfativo	sym19	Quadro respiratório	sym47
Dor abdominal	sym20	Queda da saturação	sym48
Dor de cabeça	sym21	Resfriado	sym49
Dor de garganta	sym22	Rosto azulado	sym50
Dor no corpo	sym23	Saturação baixa	sym51
Dor no peito	sym24	Saturação de O_2 menor que 95%	sym52
Dor persistente no tórax	sym25	Síndrome respiratória aguda grave	sym53
Erupção cutânea na pele	sym26	SRAG	sym54
Fadiga	sym27	Tosse	sym55
Falta de ar	sym28	Vômito	sym56

Códigos: são as expressões usadas no modelo matemático desta pesquisa para representar cada sinal ou sintoma de COVID-19.

Fontes: OMS (2023b), Pan American Health Organization (2021), ANVISA (2020), Ministério da Saúde (2021a).

Para inserir os dados na pesquisa, foram necessários: (1) classificar os dados manualmente e (2) utilizar ferramentas (linguagem de programação) para a manipulação dos dados (Figura 4). Isso foi fundamental para conseguir um conjunto de dados uniforme. Assim, para assegurar a consistência na curadoria de dados, foi incorporado um protocolo de entrada de dados (por exemplo, formato de dados, dados ausentes, nomes diferentes para o mesmo objeto). Para auxiliar nesse processo, foi utilizada a biblioteca `pandas` (Pandas Development Team, 2021), uma biblioteca de código aberto que fornece estruturas de dados e ferramentas de dados de alto desempenho e fáceis de usar para a linguagem de programação *Python* (ROSSUM; WARSAW; COGHLAN, 2001; Pandas Development Team, 2021). Uma versão estática atual do conjunto de dados de sinais e sintomas de *tweets*, de 01 de Janeiro de 2019 até 30 de setembro de 2021, está disponível *online*, de forma pública, na plataforma Zenodo (SANTOS et al., 2021). Esses dados estão disponíveis em formato `.csv`, e no banco de dados, cada observação (linha) representa um único caso individual, e os códigos utilizados nesta pesquisa estão disponíveis no repositório no site *Github* (LABCIDADE, 2021).

Figura 4 – Visão geral do processo de coleta, curadoria e verificação de dados de COVID-19, Brasil.



Fonte: Adaptado de [Berry et al. \(2021\)](#).

3.3.1.1 Variáveis para os bancos de dados de SRAG e Twitter

Os dados do openDataSUS referem-se aos casos notificados de síndrome respiratória grave (SRAG). São casos de SRAG registrados no Sistema de Informação de Agravos de Notificação, cujo processo é realizado desde 2009, período da pandemia de Influenza A/H1N1 ([Ministério da Saúde, 2021c](#)). No sistema do SIVEP-Gripe registram-se informações sobre casos graves, suspeitos e confirmados de COVID-19 no Brasil, bem como óbitos nos níveis nacional e regiões (por exemplo, estados, cidades). Para incorporação dos dados de SRAG (casos de COVID-19) no modelo epidemiológico desenvolvido nesta pesquisa, foram identificados, no sistema do openDataSUS, seis tipos de registros de interesse:

- DT_NOTIFIC: data da entrada na unidade de saúde.
- DT_SIN_PRI: data que a pessoa declarou ter tido o primeiro sintoma.
- DT_INTERNA: data da hospitalização por SRAG.
- DT_ENTUTI: data da admissão na unidade de tratamento intensivo (UTI)
- DT_SAIDUTI: data de saída da UTI.
- DT_EVOLUCA: data da alta ou morte.

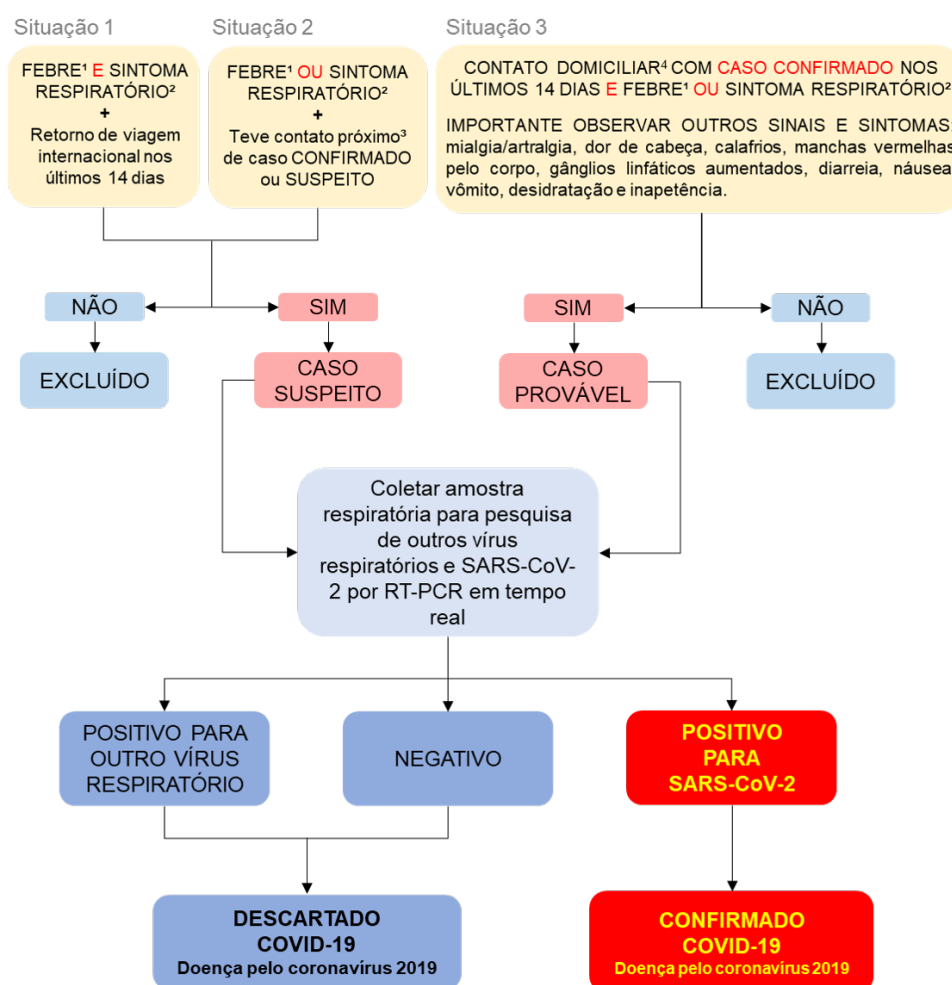
Dentre esses registros foi selecionada a variável DT_NOTIFIC. Essa variável foi escolhida pois, o objetivo da pesquisa era analisar o total de notificações de COVID-19 e essa variável representa 100% de todos os casos registrados no sistema. Definida a variável para a modelagem estatística o próximo passo foi selecionar a dimensão relevante para pesquisa presente na respectiva variável. Para isso, outra variável foi analisada, avariável CLASSI_FIN, a qual representa o registro de todos os tipos de SRAG dentro do banco de dados, são elas:

1. Influenza
2. Outro vírus
3. Outro agente etiológico
4. SRAG não especificado
5. COVID-19

Assim, para composição da amostra, foram selecionados todos registros válidos de acordo com a variável DT_NOTIFIC e de acordo com a variável CLASSI_FIN foram filtradas todas as observações de número 5. Ou seja, foram selecionados todas pessoas que deram entrada em uma unidade de saúde e que foram registradas como caso de COVID-19.

Para coleta dos *tweets* foi utilizado, como referência, o protocolo de vigilância para SRAG que estava disponível no site do Ministério da Saúde, no início da pandemia no Brasil, em março de 2020. Para a coleta dos dados epidemiológicos de SRAG, foi utilizada uma interface de programação de aplicativo (API) para o ano epidemiológico de SRAG 2020 (Ministério da Saúde, 2021d). Antes de disponibilizar os dados no portal openDataSUS, é realizado um tratamento da base de dados que envolve o anonimato dos pacientes, em conformidade com a legislação vigente (Ministério da Saúde, 2021f). Vale relatar que até dezembro de 2020 não havia disponível, para a população brasileira, vacinas contra a COVID-19. Assim, eram utilizadas apenas medidas não farmacológicas para a prevenção e o controle da disseminação do SARS-CoV-2, como equipamentos de proteção individual (EPIs), isolamento, quarentena e distanciamento social.

No momento da coleta de dados para esta pesquisa, os registros de casos de COVID-19 eram realizados com base no fluxograma da Figura 5 (Ministério da Saúde, 2020a). Esse fluxograma de definição de casos de COVID-19 foi disponibilizado online no início da transmissão comunitária da COVID-19 no Brasil, em 13 de março de 2020, dois dias após a Organização Mundial da Saúde ter classificado a doença pelo novo coronavírus (COVID-19) como uma pandemia (Ministério da Saúde, 2020a). Contudo, é importante ressaltar que, desde o início da pandemia, algumas definições que estão presentes nesta pesquisa podem ter sido atualizadas; assim, se outro pesquisador, porventura, quiser replicar esse trabalho, é importante atualizar essas informações, o que pode levar a resultados e conclusões diferentes daqueles desta pesquisa.

Figura 5 – Definições de caso operacionais para a COVID-19, Brasil, março de 2020.

- ¹FEBRE: Temperatura acima de 37,8°C é considerada febre.
- ²SINTOMAS RESPIRATÓRIOS: Tosse, dificuldade para respirar, produção de secreção, congestão nasal ou conjuntival, dificuldade em engolir, dor de garganta, coriza, saturação do oxigênio baixa, sinais de cianose, coriza, secreção intercostal e dispneia ou febre.
- ³CONTATO PRÓXIMO
 - Cerca de 2 metros de um paciente suspeito ou confirmado por 15 minutos ou mais.
 - Morar no mesmo ambiente com CASO SUSPEITO, no ambiente de trabalho, sala de atendimento, aeronave e outro meio de transporte, escola ou pré-escola.
 - Houve contato ocasional (horas) com o CASO CONFIRMADO.
 - Este contato inclui visitar ou compartilhar uma área de saúde ou sala de espera.
- ⁴CONTATO DOMICILIAR
 - Contato íntimo.
 - Contato prolongado na residência de CASO CONFIRMADO, incluindo morar ou cuidar.

Fonte: adaptado de [Ministério da Saúde \(2020a\)](#)

3.3.1.2 TWITTER

O Twitter é uma plataforma *online* de mídia social onde há finitas redes sociais orientadas, cujos nós representam os perfis de diferentes atores (por exemplo, pessoas com sintomas). O conjunto de arestas representam conexões (por exemplo, mãe e filho, médico e paciente, colegas etc.) entre os atores, e as setas representam a direção dos relacionamentos em cada rede. Essas redes orientadas são formadas por grafos orientados ou dígrafos do tipo:

$$G = (N, E) \quad (4)$$

No qual N é um conjunto finito, cujos elementos são chamados de nós (ou vértices) e E é um conjunto de pares de nós, cujos elementos $(i \in N, j \in N)$ são chamados de arestas. As arestas são pares ordenados, subconjuntos de $N \times N$ (TUTTE; TUTTE; NASH-WILLIAMS, 2005; NEWMAN, 2018).

Como dito, no Twitter, as redes sociais que são formadas são orientadas, o que é diferente de outras mídias sociais, como, por exemplo, o Facebook e o Instagram, onde há predominância de redes não orientadas. Esse entendimento de como as redes se formam nessas mídias é importante, pois permite identificar como se dá o comportamento das pessoas, quem influencia e quem é influenciado, e quais estratégias de comunicação as organizações devem adotar. No Twitter, as redes mostram como se disseminam os relacionamentos, sendo o(s) vértice(s) “ i ” considerado(s) os ponto(s) de origem da mensagem, e “ j ” o(s) ponto(s) de destino. O Twitter possui várias ferramentas que possibilitam a comunicação e a criação de redes, a saber:

- a) *Tweets*
- b) *Followers*
- c) *Replies*
- d) *Retweets*
- e) *Mentions*
- f) *Hashtags*
- g) *Direct message*

Tweets é a principal ferramenta do Twitter, sem a qual a plataforma não existiria da maneira como é conhecida. Os tweets contêm mensagens e tópicos que impulsionam relacionamentos e criações de redes. Um *tweet* pode conter, além da mensagem principal, marcações que permitem relacionar cada mensagem a determinados atores (*mentions*). *Mentions* são menções que podem ser atribuídas a membros internos ou externos no Twitter. Porém, normalmente são

utilizadas para mencionar um membro interno na plataforma, e sua principal função é "sinalizar" para determinados atores um assunto que pode ser de seu interesse, ou associar esses atores com o assunto em pauta, ou ainda, para promover um diálogo.

Se um ator mencionado em um *tweet* responder (*replie*), forma-se uma rede que pode iniciar com apenas dois atores e, a partir daí, se expandir para mais atores, começando a formar redes complexas de relacionamentos, com subgrupos discutindo particularidades sobre um mesmo assunto ou sobre temas distintos. Contudo, por ser uma comunicação mais direcionada, o alcance das *mentions* para a construção de redes tende a ser mais limitado do que a utilização de *hashtag*.

As *hashtags*, representadas pelo símbolo #, permitem marcar palavras ou temas específicos. Assim, qualquer um que utilizar o símbolo # junto e do lado esquerdo de uma palavra ou expressão em um *tweet*, fará parte de uma rede que tem como interesse principal a respectiva expressão. A *hashtag* é uma das principais ferramentas para a formação e ampliação de uma rede. Diferentemente de outras marcações, as *hashtags* não necessitam que os atores tenham qualquer grau de proximidade; podem possuir como único fator comum o assunto relacionado à *hashtag*. Dentre todas as métricas que contribuem para formação de redes, *hashtag* é a que permite a formação das maiores redes e, por ter maior alcance, é uma das métricas mais utilizadas para analisar os temas sociais mais relevantes.

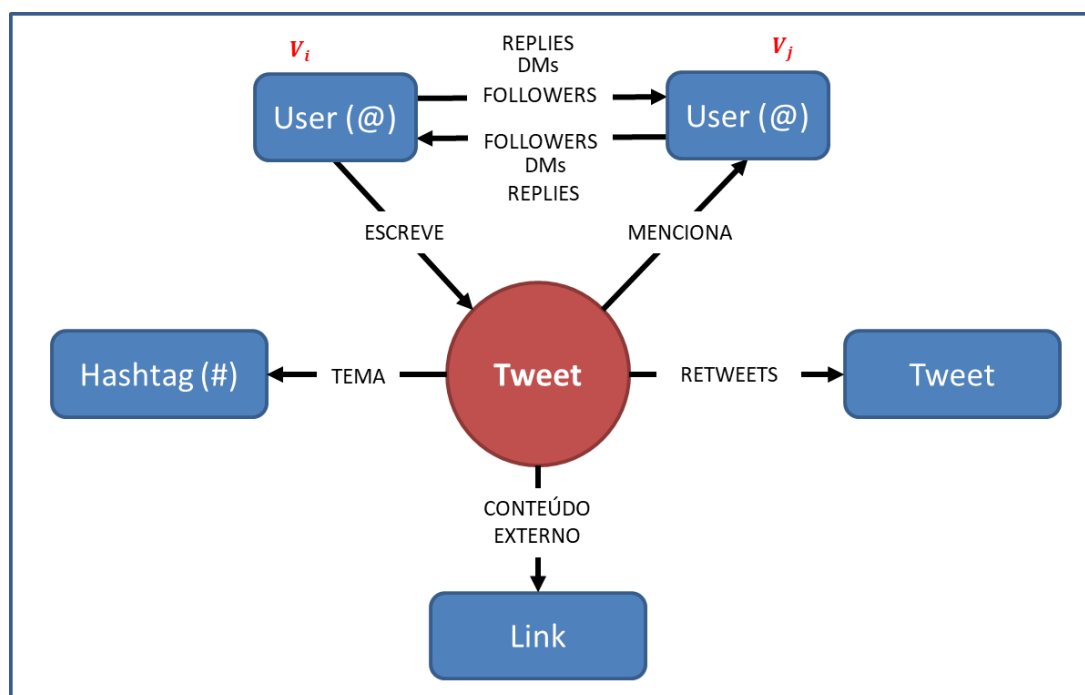
Outra métrica de grande relevância no Twitter é o RT(s) ou Retweets. O número de RTs demonstra a influência que tem um ator (pessoa, organização, marca etc.), e refere-se às mensagens que são compartilhadas por diferentes atores. Quanto mais os usuários no Twitter compartilham uma mensagem, mais relevantes são os assuntos e os atores relacionados aos temas. Porém, apesar de ser uma ferramenta relevante para determinar a popularidade de temas e atores, o RTs tem uma limitação. Seu alcance é diretamente relacionado aos perfis que os usuários seguem (*followers*). Visto que uma mensagem do tipo RT pode conter #, isso faz com que os RT sejam uma das ferramentas mais ricas para análises.

As formações de redes sociais no Twitter podem ser representadas pela Figura 6. Esta formação começa com um usuário (V_i) escrevendo uma mensagem (*tweet*). Essa mensagem pode conter diversas métricas (#, RTs, links etc.), que são os principais instrumentos para fazer com que a mensagem se dissemine na plataforma e, conseqüentemente, aumente o número de conexões nas redes. Porém, para se formar uma rede, é necessário que a mensagem que teve origem em V_i chegue até V_j , e esses usuários (V_i, V_j) podem, mas não precisam, ser seguidores (*followers*) um do outro.

Apenas para a troca de mensagens privadas – *direct message* (DMs) – faz-se necessário que V_i e V_j estejam na mesma rede. Outra característica na formação da rede é que ela não é limitada apenas à comunicação dentro do Twitter, sendo possível incluir nas mensagens links externos (ou internos) ou referenciar atores que não estejam presentes no Twitter. Posto isso, os tweets (e RTs) nesta pesquisa foram selecionados como ferramenta para análise, visto que o objetivo principal é o estudo sobre as percepções dos atores relacionadas aos sinais e sintomas

da COVID-19, como essas percepções se relacionam e variam no decorrer da pandemia e como esses *tweets* podem ajudar a compreender a dinâmica das hospitalizações de COVID-19 no Brasil.

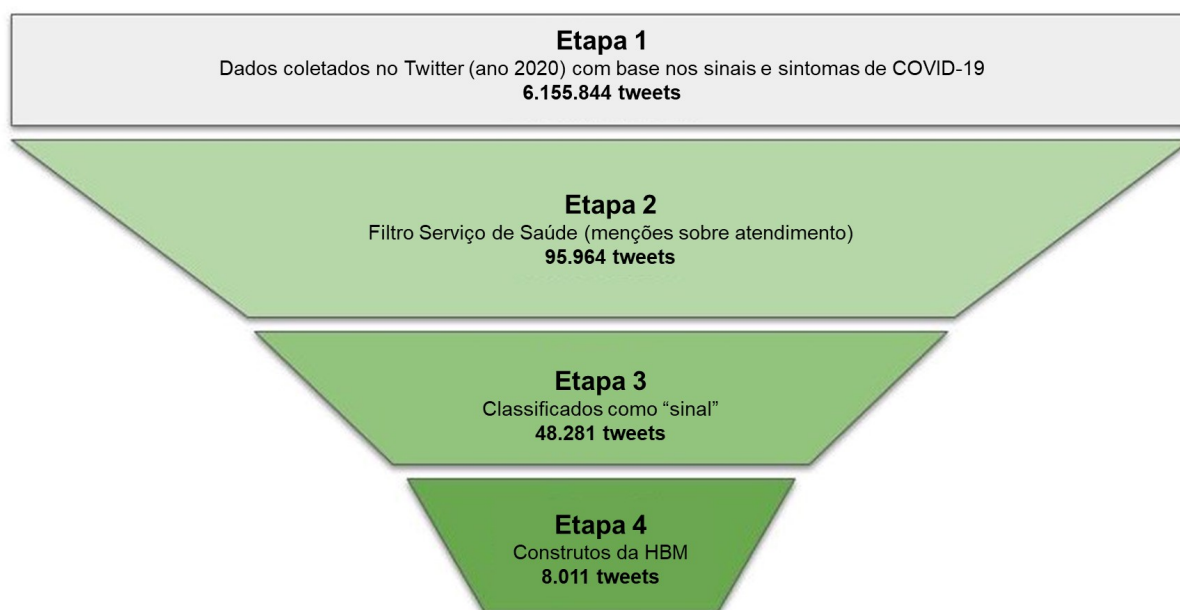
Figura 6 – Fluxo de formação de redes sociais no Twitter.



Fonte: desenvolvido pelo autor.

3.3.2 ETAPAS PARA O TRATAMENTO DOS DADOS

O processo de tratamento para a classificação dos dados desta pesquisa pode ser dividido em quatro etapas principais (Figura 7). A primeira (Etapa 1) consiste na mineração de tweets que fizeram menções a pelo menos um dos 56 sinais ou sintomas relacionados à doença COVID-19 (ver lista no Quadro 3), compreendidos no ano de 2020. Para realizar a coleta de dados foi utilizada a biblioteca para mineração de tweets *twint* (POLDI et al., 2022). Essa ferramenta para a mineração de dados em mídias sociais é escrita na linguagem de programação *Python* (ROSSUM; WARSAW; COGHLAN, 2001), e permite extrair apenas dados do Twitter de contas públicas, ou seja, dados que qualquer pessoa pode acessar direto pelo navegador *web* (por exemplo, *Google Chrome* e *Internet Explorer*). Assim, a utilização desses dados está em conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD), que dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, e permite a utilização de dados abertos para fins de pesquisa (BRASIL, 2019).

Figura 7 – *Etapas de tratamento para a classificação dos dados da pesquisa.*

Nota: HBM: *Health Belief Model*

Fonte: desenvolvido pelo autor.

Na etapa 2, com o objetivo de eliminar dados não relevantes à pesquisa, os cerca de 6 milhões de tweets coletados receberam o primeiro tratamento. Para isso, mais uma vez, foi utilizada como ambiente de desenvolvimento a Linguagem de Programação Python (ROSSUM; WARSAW; COGHLAN, 2001). Contudo, nessa etapa, a principal biblioteca utilizada para a manipulação dos dados foi a *pandas* (Pandas Development Team, 2021). Essa fase consistiu em filtrar os tweets da Etapa 1 que indicassem que os usuários (ou alguém mencionado no *tweet*) procuraram ou teriam interesse em procurar atendimento médico para tratar os sintomas relacionados à COVID-19. A seleção dos termos-chave utilizados para filtrar os tweets foi realizada com o auxílio de uma médica especialista em epidemiologia, tendo como base os termos mais frequentes que apareceram na amostra e expressões comumente utilizadas na área médica para designar serviços de saúde, como, por exemplo "PA" (expressão para "pronto atendimento"). A lista completa dos termos utilizados pode ser vista no Quadro 4.

A aplicação deste filtro é importante, pois um dos objetivos da pesquisa é prever, com dados de mídias sociais, o número de pessoas que foram até algum serviço de saúde com suspeita de estar com COVID-19, ou que eram casos confirmados e precisaram de atendimento médico. Assim, só foram considerados os tweets que mencionaram algum tipo de atendimento médico, como, por exemplo, o *tweet* exibido na Figura 8. A aplicação desse filtro fez com que, de 6 milhões de tweets a amostra reduzisse para apenas 95 mil *tweets*, aproximadamente.

Quadro 4 – *Serviços e procedimentos utilizados para classificar procura de serviços por sinais e sintomas de COVID-19 (Etapa 2).*

Termos	
1. ama	20. pronto-socorro
2. ambulatório	21. ps
3. atendimento online	22. pt-pcr
4. clinica	23. ptpcr
5. cotonete	24. radiografia
6. enfermaria	25. radiologista
7. enfermeira	26. raiox
8. enfermeiro	27. rt-pcr
9. exame de covid	28. rtpcr
10. hospital	29. serviço de saúde
11. interna	30. sus
12. farmácia	31. remédio
13. médica	32. telemedicina
14. médico	33. teste covid
15. pa	34. tomografia
16. pcr	35. tomógrafo
17. posto de saúde	36. ubs
18. postinho	37. upa
19. pronto atendimento	38. uti

Fonte: desenvolvido pelo autor.

Figura 8 – *Exemplo de tweet considerado na Etapa 2, no modelo de classificação.*



Fonte: Twitter

Após a coleta (Etapa 1) e a aplicação do filtro com os termos de interesse (Etapa 2) na Etapa 3 foi gerada uma amostra aleatória por proporção¹, com nível de confiança de 95% e margem

¹Foi preservada a representatividade de cada mês nos dados coletados.

de erro de 1%, contendo 10.004 *tweets* com possível duplicidade de perfis, visto que uma pessoa poderia se reinfectar com o SARS-CoV-2; logo, em um momento, ela poderia escrever que estava com sintomas e posteriormente retornar na rede social e mencionar que foi reinfectada. Contudo, o estudo das pessoas reinfectadas foge do escopo desta pesquisa, mas pode ser analisado em estudo futuros.

Nesta etapa, os *tweets* passaram por três processos. Primeiro, foram classificados com dois rótulos: *senal* = 1; *ruído* = 0. O rótulo “senal” foi atribuído aos *tweets* que eram pertinentes à pesquisa e o rótulo “ruído” caso não tivessem relação com a pesquisa. A distribuição dos dados referentes aos meses de 2020 pode ser visto na Tabela 3. Segundo, foi criado um dicionário com expressões que foram comumente utilizadas nas redes sociais, no ano de 2020, durante a pandemia do novo coronavírus. Por exemplo, “coronga” e “cvd”, ambas expressões que significavam COVID-19. Terceiro, foram selecionados os emojis² presentes nos tweets da amostra que tinham relação com a pandemia de COVID-19 (Figura 9). Com esses emojis foi criado um outro dicionário de significados que foi utilizado para substituir cada emoji por uma expressão por extenso. Dos 10.004 *tweets* rotulados, cerca de 55% foram classificados como sinal (dados aderentes à pesquisa), e cerca de 45% classificados como ruído (dados que foram excluídos da pesquisa). Esses dados rotulados foram utilizados no modelo de classificação para rotular de forma automática todos os tweets presentes na Etapa 2 (95.964 tweets), o que gerou um total de 48.281 tweets classificados como sinal, 50% do total da amostra da Etapa 2, que é um valor bem próximo do que foi rotulado nos 10.004 (55% de tweets rotulados como sinal).

Figura 9 – Emojis relacionados com o tema da pesquisa encontrados na amostra (Etapa 3).



Fonte: elaborado pelo autor.

²A lista completa dos emojis encontrados nos dados da pesquisa encontra-se no Apêndice B

3.4 MODELO DE PROCESSAMENTO DE LINGUAGEM NATURAL

Esta seção compreende uma descrição da aplicação de um modelo de Processamento de Linguagem Natural (do inglês, *Natural Language Processing* – NLP), que foi usado nesta pesquisa. O NLP é um dos tópicos de *Machine Learning* (BALLI et al., 2022), e foi utilizado para classificar os dados coletados no Twitter sobre as menções de sinais e sintomas de COVID-19 em Língua Portuguesa do Brasil (Quadro 3).

Para classificar de forma automática os 95 mil tweets que foram tratados na Etapa 2 (Figura 7), na Etapa 3 foi implementado o Classificador de Regressão Logística³ LR (tf-idf) (Algoritmo 11), um dos algoritmos de classificação mais populares e amplamente utilizados (BENCKE; CECHINEL; MUNOZ, 2020). Porém, outros algoritmos do Scikit-learn também foram testados (ver Tabela 5, no capítulo de resultados); contudo, o desempenho foi inferior ao LR (tf-idf). Os benefícios de utilizar o Scikit-learn é que essa é uma biblioteca de *machine learning* de código aberto que oferece suporte ao aprendizado supervisionado e não supervisionado, a qual fornece ferramentas para o ajuste de modelos, pré-processamento de dados, seleção de modelos e avaliação de modelos, entre outras funcionalidades (PEDREGOSA et al., 2011).

Algoritmo 1: PSEUDOCÓDIGO DO CLASSIFICADOR LOGÍSTICO LR (TF-IDF)

Entrada: Parâmetros da função, dados rotulados, dados não rotulados.

Saída: Treino → Teste → Validação.

1 **início**

2 **Preparar dados rotulados:** *standardization*, *Word2vec*, *stemming*.

3 **Dividir dados:** bases de treino e teste.

4 **Converter bases em vetor:** TF-IDF.

5 **Definir função:** *LogisticRegression*(*penalty*, ... , *l1_ratio*).

6 **Ajustar modelo para treino:** *load_log_reg.fit*(*X_train*, *Y_train*).

7 **Estimar \hat{Y} em base teste:** *lr_tfidf.predict*(*X_val_vectors_tfidf*).

8 **Validar modelo:** *precision*, *recall*, *f1-score*, *support*, *accuracy*,
 macro avg, *weightde avg*, *Confusion Matrix*.

9 **Aplicar LR (tf-idf):** dados não rotulados.

10 **fim**

11 **retorna** *tweet = sinal*, se $p(\text{tweet}) > 0,5$, caso contrário *ruído*.

Fonte: adaptado de Bilandi, Verma e Dhir (2021).

A etapa inicial da implementação do algoritmo de classificação consiste no pré-processamento dos dados (preparar os dados). Assim, foi realizada a discretização e a limpeza de texto, normalização, remoção de *stop-words* e a lematização. Na sequência, a variável-alvo (texto dos

³As expressões matemáticas do classificador binário de Regressão Logística adaptado de Raschka et al. (2022) encontra-se na seção 3.4.1.

tweets) foi discretizada e *hiperlinks* e caracteres especiais foram removidos. Depois as palavras foram normalizadas e o *stemming*, que é o processo de reduzir as palavras aos seus radicais, foi aplicado. Na sequência, foi extraída a matriz de recursos de frequência de documento inversa de frequência de termo (do inglês, *term frequency-inverse document frequency* – TF-IDF), utilizando o módulo `TfidfVectorizer` do `Scikit-learn`. Este módulo, que contribui para o processo de tratamento dos dados, ignora os *outliers* – termos que possuem uma frequência de documento maior ou menor que um limite mínimo e um máximo (HAMDI; SHABAN; ZAINAL, 2018).

Terminada a fase de pré-processamento dos dados, iniciou-se o processo de definição dos *n*-gramas do modelo. Ou seja, um conjunto consecutivo de palavras cujo intervalo pode variar produzindo bigramas, trigramas, e assim por diante. A seleção do *n*-gramas é importante para reduzir a dimensionalidade e, conseqüentemente, reduzir custos computacionais, e evitar ajustes excessivos do modelo de aprendizagem aos dados de treinamento (HAMDI; SHABAN; ZAINAL, 2018). Para tanto, assim como na proposta de Hamdi, Shaban e Zainal (2018), foi utilizado o método de limite de variação implementado na biblioteca `Scikit-learn` para remover todos os recursos cuja variação está abaixo de um determinado limite. Feito isso, empiricamente foram avaliadas extrinsecamente várias configurações de parâmetros. O classificador foi treinado e testado. Assim, os parâmetros TF-IDF foram corrigidos e os limites de variação alterados. A Tabela 2 lista os parâmetros originais da biblioteca `Scikit-learn` (PEDREGOSA et al., 2011), bem como os ajustes empíricos realizados nesta pesquisa.

Tabela 2 – Ajuste dos parâmetros do classificador *RL*.

Parâmetros	Scikit-learn	Ajuste empírico
penalty	l2	l2
dual	False	False
tol	0,0001	0,0001
C	1,0	1,4
fit_intercept	True	True
intercept_scaling	1	0
class_weight	None	balanced
random_state	None	0
solver	lbfgs	lbfgs
max_iter	100	100
multi_class	auto	ovr
verbose	0	0
warm_start	False	False
n_jobs	None	None
l1_ratio	None	None

Fontes: Pedregosa et al. (2011); dados da pesquisa.

O Classificador de Regressão Logística utilizado nesta pesquisa foi o classificador implementado na biblioteca `Scikit-learn`, que possui quinze parâmetros. No intuito de melhorar o poder preditivo do modelo, foram feitos testes nos parâmetros, dentro das variações permitidas na configuração da biblioteca `Scikit-learn`, e a melhor opção encontrada está representada na Tabela 2, na coluna "Ajuste empírico". O parâmetro `penalty` refere-se à função custo e pode assumir um dos quatro valores: `l1`, `l2`, `elasticnet`, `none`. As funções custos `l1` e `l2` são técnicas de regularização, sendo que `l1` é chamada de *Lasso Regression* (Equação 6) e `l2` é chamado de *Ridge Regression* (Equação 5). A principal diferença entre essas funções é o tipo de penalidade sobre a função custo. Enquanto a *Ridge Regression* adiciona "magnitude" ao quadrado do coeficiente como termo de penalidade, a *Lasso Regression* adiciona "valor absoluto de magnitude" do coeficiente. Na prática, `l2` é uma técnica que funciona melhor para evitar *overfitting*, e `l1` funciona bem caso haja um grande número de recursos (NAGPAL, 2017; TIBSHIRANI, 1997; TIBSHIRANI, 1996). Ao aplicar `elasticnet`, ambos os termos de penalidade (`l1` e `l2`) são adicionados ao modelo, e em `none` nenhuma penalidade é aplicada.

$$\sum_{i=1} (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (5)$$

$$\sum_{i=1} (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (6)$$

O segundo parâmetro do classificador RL presente na Tabela 2 é o `dual`. A formulação dual é implementada no classificador para `l2` com o solucionador linear (*liblinear*) e, por padrão, a formulação `primal` (`False`) está sempre ativa no classificador (PEDREGOSA et al., 2011). No caso linear, no problema primal existe subespaço de direções que aumenta a função objetivo que remove a folga entre a solução candidata e uma ou mais restrições. No problema dual, o vetor dual multiplica as restrições que determinam as posições das restrições no primal. Em suma, optar pelo problema dual é equivalente a revisar os limites superiores no problema primal (MIELE; WANG; BASAPUR, 1986).

Na sequência, o próximo parâmetro a ser ajustado no classificador é o que define a tolerância para a otimização (`tol`). Sempre que as atualizações forem menores que `tol`, o código de otimização verifica a otimalidade⁴ do `dual gap` e continua até que seja menor que `tol`. Depois, temos o parâmetro `C`, que é o inverso da força de regularização e deve assumir um valor real positivo. O quinto parâmetro do modelo é o `fit_intercept`, que nesta pesquisa corresponde ao log da probabilidade para os tweets serem classificados em uma determinada classe (sinal ou ruído) e, por padrão, no `Scikit-learn` esse parâmetro está configurado para

⁴Método simplex para problemas lineares multiobjetivo.

ser otimizado (`True`). Outro parâmetro associado com as classes é o `class_weight`, que usa os valores de y para ajustar os pesos inversamente proporcionais às frequências das classes (`balanced`), caso contrário, é atribuído o mesmo peso às classes (`None`). O oitavo parâmetro na Tabela 2 é o `random_state`. A implementação desse parâmetro garante que toda vez que o modelo for executado será obtido o mesmo resultado; contudo, se o objetivo é ter alguma produção aleatória, deve-se fixar com o valor `None` (PEDREGOSA et al., 2011).

Um dos parâmetros mais significativos para o ajuste do modelo de classificação é o `solver`. Esse parâmetro é utilizado para encontrar os pesos dos parâmetros que minimizam a função custo. O `Scikit-learn` oferece cinco opções para esse solucionador: `newton-cg`, `lbfgs`, `liblinear`, `sag`, e `saga`. Os métodos `newton-cg`, `lbfgs` e `sag` lidam apenas com o tipo de penalidade L2. Para conjuntos de dados pequenos, `liblinear` é o mais indicado, enquanto `sag` é mais rápido para conjuntos grandes. Já para problemas multi-classe, apenas `newton-cg` e `lbfgs` conseguem lidar com a perda multinomial (PEDREGOSA et al., 2011). Essa perda está associada ao desbalanceamento das classes, e quando há um desequilíbrio grande entre as classes, muitos algoritmos são menos eficazes, e, quando isso acontece, uma solução é tentar balancear os dados, seja removendo amostras de classes sobre-representadas ou adicionando mais amostras de classes sub-representadas (LEMAÎTRE; NOGUEIRA; ARIDAS, 2017).

Além desses parâmetros, outros podem ser usados para a otimização adicional e evitar *overfitting*. O parâmetro `max_iter` controla a iteração máxima permitida para a convergência do `solver`. O `warm_start` é útil apenas com solucionadores diferentes do `liblinear`, e é usado para determinar a inicialização da solução anterior à fase de treinamento. Se o `warm_start` é ajustado como `Falso`, a solução anterior é descartada, e se é ajustado como `Verdadeiro`, a solução anterior será reutilizada. Por padrão, o parâmetro `multi_class` é configurado como automático, o que significa que se o `solver` for `liblinear`, o seletor será `ovr` (*One-versus-rest*), o que faz com que cada problema seja considerado como binário e, em outros casos, `multinomial` será selecionado automaticamente. Já o parâmetro `verbose` controla a impressão das informações durante a execução do algoritmo, e se selecionado como 0 (zero) as informações não serão exibidas, se selecionado 1 (um) algumas informações serão exibidas, e 2 (dois) mais informações serão exibidas.

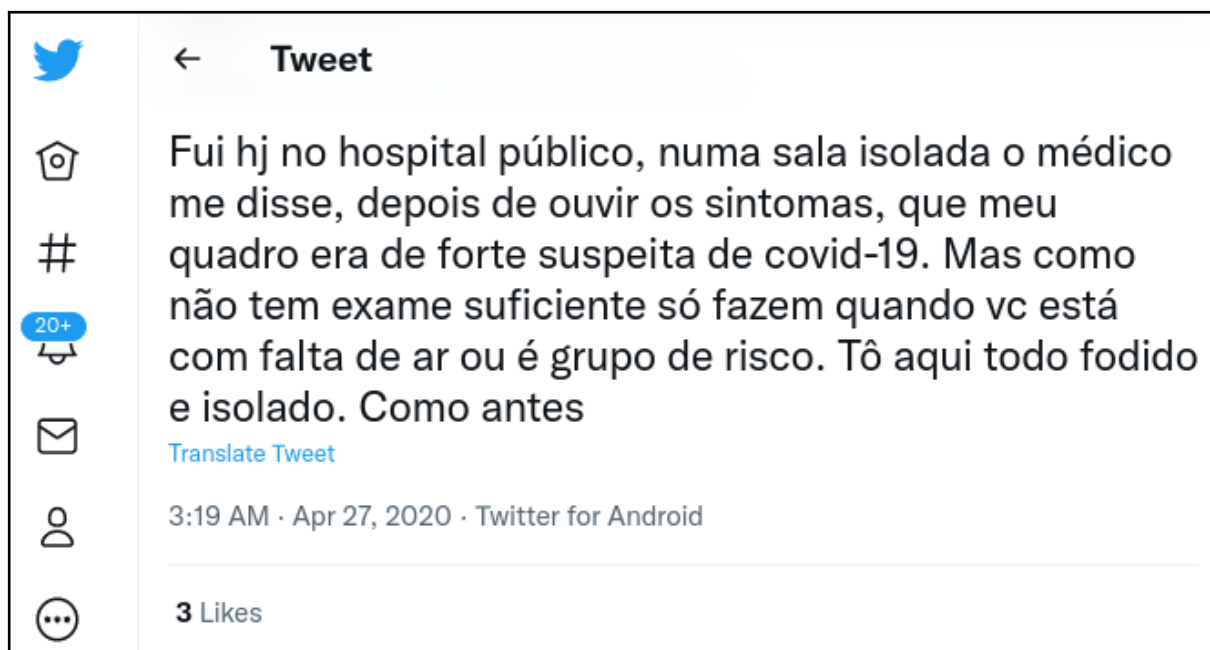
O penúltimo parâmetro a ser configurado é o `n_jobs`, e o padrão no `scikit-learn` é `None`, ou seja, os núcleos podem funcionar em paralelo; outra característica desse parâmetro é que só funciona quando o `solver` não for `liblinear` e a `multi_class` for `ovr`. Se `n_jobs` for configurado como `None`, apenas 1 núcleo da CPU funcionará; se for configurado com `-1`, todos núcleos presentes, quando possível, serão atribuídos; e `int` os núcleos da CPU poderão funcionar em paralelo com base no valor inteiro atribuído durante a regressão logística (PEDREGOSA et al., 2011). Para o último parâmetro da lista, o `l1_ratio`, se o `solver` selecionado for `saga` e a penalidade for `elasticnet`, então este parâmetro pode oferecer uma otimização adicional.

Por fim, como resultado da implementação do algoritmo de classificação, na Etapa 4, a

amostra ficou com cerca de 48 mil tweets classificados como sinal. Com esses tweets, obteve-se uma nova amostra de forma aleatória simples, com nível de confiança de 95% e margem de erro de 1%, o que resultou em um total de 8.011 tweets. Nesta fase, esses oito mil tweets passaram por um novo tratamento manual. Porém, dessa vez o objetivo foi classificá-los com base nos construtos da HBM (ROSENSTOCK; STRECHER; BECKER, 1988; ROSENSTOCK, 1974a; ROSENSTOCK, 1974b). Essa classificação é importante, pois permite identificar o nível de percepção das pessoas sobre a pandemia do novo coronavírus, e, junto com o modelo Bayesiano, inferir sobre quais construtos da HBM estão mais associados aos atendimentos nos serviços de saúde.

Além dos construtos da *Health Belief Model* (ROSENSTOCK; STRECHER; BECKER, 1988; ROSENSTOCK, 1974a; ROSENSTOCK, 1974b), também foram considerados outros atributos para classificar os tweets na Etapa 4. No total, os tweets foram rotulados considerando-se doze atributos, dos quais oito são os construtos da HBM (Quadro 5). Diferente de pesquisas realizadas em ambientes controlados, onde é possível direcionar as perguntas para identificar um único atributo ou construto da HBM no ambiente virtual, nas plataformas de mídias sociais isso não é possível. Um único tweet pode ser classificado com um único ou vários atributos (*multilabel*), como pode ser visto na Figura 10.

Figura 10 – Tweet com multilabel classificado com construtos da *Health Belief Model* (Etapa 4).



Fonte: dados da pesquisa.

Na Figura 10 é possível observar que no tweet há dois atributos (ou construtos) relacionados à *Health Belief Model*:

- Gravidade: a expressão “fodido” indica que a pessoa considera a situação grave.

- Barreira: não ter exames suficientes para a COVID-19 pode prejudicar o tratamento.

Além dos construtos da HBM, também é possível se observar, na Figura 10, outros atributos relacionados à pesquisa. A percepção sobre a gravidade da doença pode estar associada ao grau de relacionamento entre as pessoas. Por isso, nessa pesquisa, também foi analisado o nível do relacionamento entre a pessoa que está com o sinal/sintoma com quem escreve a mensagem. Por exemplo, neste *tweet*, é a própria pessoa que está com os sintomas da COVID-19, o que poderia gerar uma percepção diferente se os mesmos sintomas estivessem associados a uma terceira pessoa, como um filho ou uma mãe. Outro atributo importante que está sendo considerado nesta pesquisa é quando a pessoa menciona estar (ou alguém estar) com COVID-19.

Quadro 5 – *Construtos da HBM utilizados para classificar os tweets (Etapa 4).*

Construtos	Critérios para classificação
Ameaça não percebida	Quando a ameaça não é percebida. Pode ser em relação à doença ou ao tratamento (medicação). Por exemplo, quando não percebe a ineficácia do "kit covid". Não há necessidade de ser o doente, por exemplo, a pessoa pode ter todos os sintomas, mas o médico achar que não é COVID-19 (nesse caso, o médico não percebeu). Na ausência de um sintoma, a pessoa pode considerar improvável ter a doença. Quando a pessoa não percebe que outros podem ter dificuldade em seguir uma recomendação, por exemplo, usar máscara por ser claustrofóbica (não confundir com "eficácia percebida", que é quando a pessoa declara que não consegue seguir um tratamento, por exemplo, por ser claustrofóbica). Quando a pessoa menciona que deve ter pego COVID-19 em um período anterior ao primeiro caso registrado, por exemplo, quando fala que deve ter pego na infância, pois os sintomas eram parecidos; isso indica que a pessoa não entendeu o que é o novo coronavírus.
Suscetibilidade percebida	Quando a mensagem indica um sintoma e por causa disso a pessoa (ou terceiro) indica que vai tomar uma ação preventiva. Ou o teste deu positivo. Contudo, nada na mensagem indica que a pessoa (ou terceiros) acha ser grave.
Gravidade percebida	Quando a mensagem indica que a pessoa (a mesma ou terceiros) acredita que a COVID-19 é grave. Se a pessoa pegou a COVID-19 e acha que pode evoluir e se tornar grave. Analisar também emojis de sentimentos de gravidade alta e se no texto tem expressões como: "corre pro hospital"; "Jesus...". Se a pessoa relata que foi várias vezes no médico, considera-se que ela acredita que é grave e que possivelmente os sintomas estão piorando. Se a pessoa relata morte por COVID-19, desde que não haja indicação de que pode ser fraude. Se a pessoa fala que apesar de não ter ficado internada, passou o dia todo ou a noite toda no hospital.
Benefícios percebidos	Quando a pessoa acredita que a intervenção médica que ela (ou um terceiro) recebeu (ou pode receber) foi positiva.

Continua na próxima página.

Quadro 5 – *Construtos da HBM utilizados para classificar os tweets (Etapa 4).*

Construtos	Critérios para classificação
Barreiras percebidas	Quando a pessoa acredita que a intervenção médica tem mais aspectos negativos do que positivos. Ou ainda, quando não tomou uma ação por fatores externos, por exemplo, hospital que só atende emergência. Ou referência a não conseguir saber se teve COVID-19. Ou quando comenta que o fato de as pessoas não respeitarem os protocolos pode prejudicar a contenção da pandemia. Ou quando a pessoa manifesta que há problema em falar para outros que está com COVID-19 (medo do preconceito da doença). Quando não é possível identificar o vírus; por exemplo quando a pessoa tem pneumonia (que segundo os médicos pode ser em decorrência do coronavírus), mas o exame deu negativo. Ter a doença e não poder ver os entes queridos. Quando a pessoa indica que foi várias vezes no médico e não adiantou.
Dicas para ação	Quando a mensagem é sobre conselhos, sugestões, recomendações de tratamento (ir ao médico) e orientações relacionadas à COVID-19. Se a pessoa diz que conversou com outras pessoas para saber mais sobre a doença ou sobre como agir na pandemia. Ou quando a pessoa coloca de forma explícita seu caso para servir de exemplo e orientação para pessoas na mesma situação.
Auto-eficácia	Crença de uma pessoa sobre se ela mesma é ou não capaz de tomar a ação preventiva. Aqui estamos considerando se a pessoa manifestar que tomou todos os cuidados recomendados (isolamento social, não aglomerar, etc) ou se não conseguiria seguir as recomendações. Se aplica também a outras pessoas. Consideramos também se a pessoa mencionou um tratamento que deve seguir (independentemente de mencionar a chance de seguir ou não).
Outras variáveis	Influências por variáveis externas relacionadas à COVID-19: política, crenças religiosas, outros aspectos culturais etc.
COVID-19	Se a pessoa falar que o exame deu positivo para COVID-19 ou se pelo exame clínico o médico falar que está com COVID-19. Caso não esteja explícito no texto que está com COVID-19, mas se falar que o médico mandou ficar em isolamento, então iremos admitir que está com COVID-19 (por exemplo, licença de 14 dias).

Continua na próxima página.

Quadro 5 – Construtos da HBM utilizados para classificar os tweets (Etapa 4).

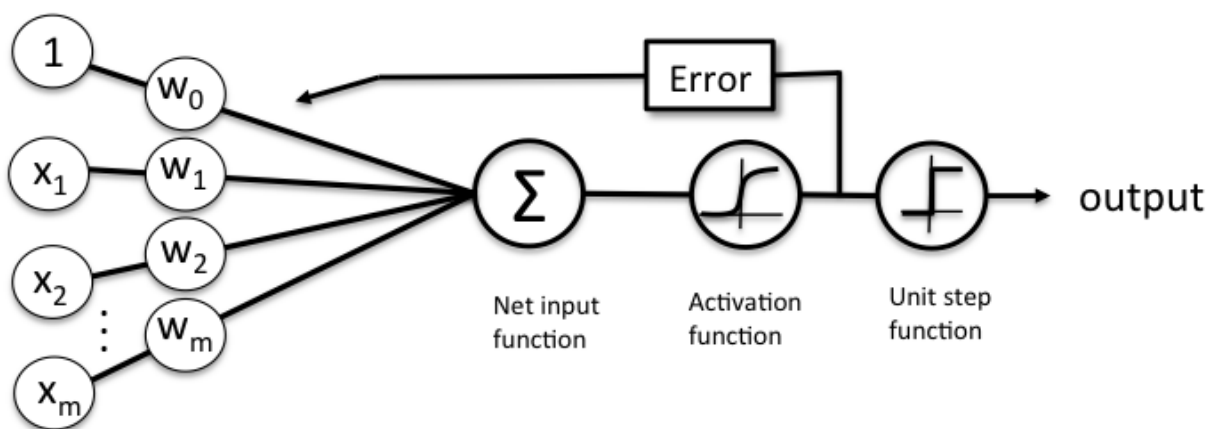
Construtos	Cr�terios para classifica�o
Interven�o (antes do <i>tweet</i>)	Se o <i>tweet</i> foi realizado antes de alguma interven�o relacionada � COVID-19 (ida ao m�dico, consulta <i>online</i> , medica�o orientada, vacina etc.). Ou se fica evidente que a pessoa n�o foi ao m�dico at� aquele momento em que escreveu a mensagem.
Interven�o (durante ou depois do <i>tweet</i>)	Se o <i>tweet</i> foi realizado ap�s ou durante alguma interven�o relacionada � COVID-19 (ida no m�dico, consulta <i>online</i> , medica�o orientada, vacina etc.).
Pessoas	Rela�o com a pessoa mencionada na mensagem que est� com sintomas. Por exemplo, m�e, amigo, esposa, desconhecido etc.
Ru�do	<i>Tweets</i> n�o relacionados � pesquisa. Exemplo: quando algu�m relata sintomas de animais. Ou menciona um evento que aconteceu antes da pandemia, por exemplo, menciona que n�o sentia cheiro e tinha falta de ar no ano de 2018, e por isso acha que naquele ano teve COVID-19.

Fonte: desenvolvido pelo autor com base no Modelo de Cren as em Sa de (Health Belief Model).

3.4.1 EXPRESS ES MATEM TICAS DO CLASSIFICADOR LOG STICO

A seguir   apresentado o modelo matem tico do classificador bin rio de regress o log stica descrito por Raschka et al. (2022):

Figura 11 – Esquema do classificador de regress o log stica.



Fonte: adaptado de Raschka et al. (2022).

O modelo de Regressão Logística é um modelo de classificação binária que tem como objetivo minimizar a função de custo linear:

$$\phi(z) = \frac{1}{1 + e^{-z}}, \quad (7)$$

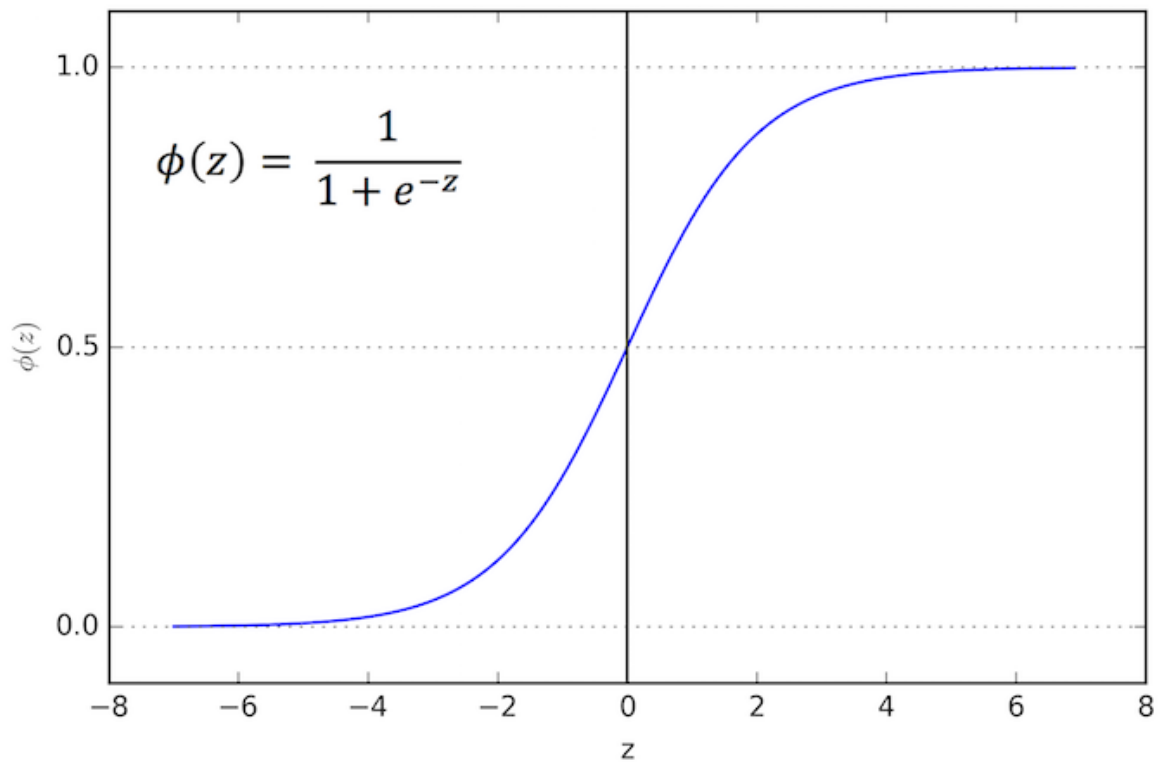
Onde z é definido como *net input* – uma equação matemática que define o sinal total que está entrando em uma unidade de outras unidades em uma rede (DAWSON, 2008)

$$z = w_0x_0 + w_1x_1 + \dots + w_mx_m = \sum_{j=0}^m w_jx_j = \mathbf{w}^T \mathbf{x}. \quad (8)$$

A entrada de rede, por sua vez, é baseada na função *logit*

$$\text{logit}(p(y = 1 | \mathbf{x})) = z. \quad (9)$$

Aqui, $p(y = 1 | \mathbf{x})$ é uma probabilidade condicional de que uma determinada amostra pertence à classe 1 dadas as suas características x . A função *logit* recebe entradas no intervalo $[0, 1]$ e as transforma em valores ao longo de todo o intervalo de números reais. Em contraste, a função logística recebe valores de entrada em todo o intervalo de números reais e os transforma em valores no intervalo $[0, 1]$. Em outras palavras, a função logística é o inverso da função *logit*, e permite prever a probabilidade condicional de que uma determinada amostra pertença à classe 1 (ou classe 0).

Figura 12 – Função Logística.

Fonte: Raschka et al. (2022).

Após o ajuste do modelo, a probabilidade condicional $p(y = 1 \mid \mathbf{x})$ é convertida em um rótulo de classe binária por meio de uma função limite $g(\cdot)$:

$$y = g(z) = \begin{cases} 1 & \text{if } \phi(z) \geq 0.5 \\ 0 & \text{caso contrário.} \end{cases} \quad (10)$$

ou equivalente:

$$y = g(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{caso contrário.} \end{cases} \quad (11)$$

Função objetivo – Log-Verossimilhança

Para parametrizar uma modelo de regressão logística, deve-se maximizar a probabilidade

$L(\cdot)$, ou minimizar a função logística de custo.

$$L(\mathbf{w}) = P(\mathbf{y} | \mathbf{x}; \mathbf{w}) = \prod_{i=1}^n P(y^{(i)} | x^{(i)}; \mathbf{w}) = \prod_{i=1}^n \left(\phi(z^{(i)}) \right)^{y^{(i)}} \left(1 - \phi(z^{(i)}) \right)^{1-y^{(i)}}, \quad (12)$$

sob a suposição de que as amostras de treinamento são independentes umas das outras. Porém, na prática, é mais fácil maximizar o logaritmo (natural) desta equação, que é chamada de função de probabilidade logarítmica:

$$l(\mathbf{w}) = \log L(\mathbf{w}) = \sum_{i=1}^n y^{(i)} \log \left(\phi(z^{(i)}) \right) + (1 - y^{(i)}) \log (1 - \phi(z^{(i)})) \quad (13)$$

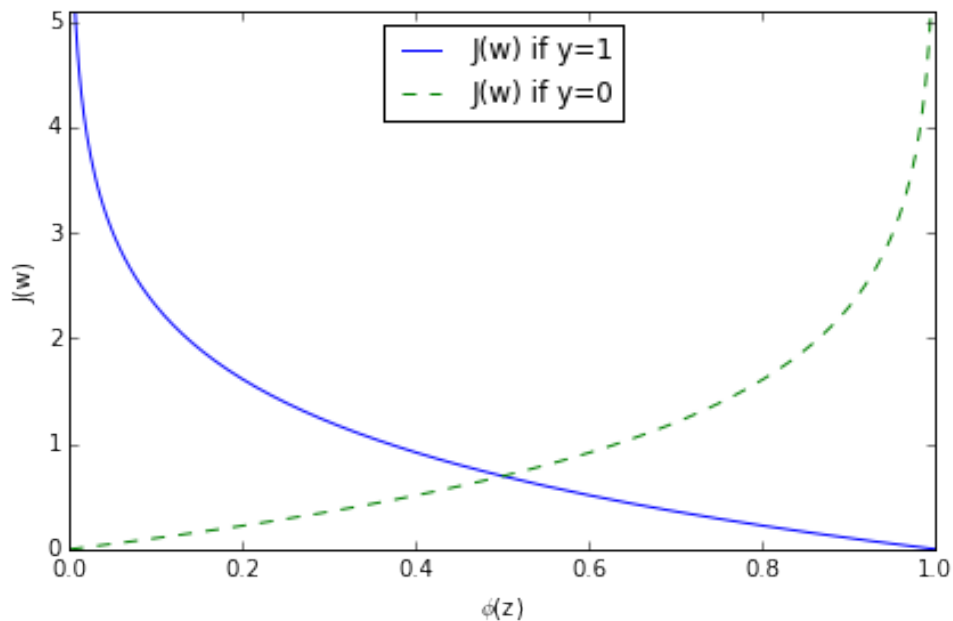
Uma vantagem de obter o log da função é evitar o *underflow* numérico para probabilidades muito pequenas. Outra vantagem é obter a derivada mais facilmente, usando o truque da adição para reescrever o produto dos fatores como um termo de soma, que pode ser maximizado usando algoritmos de otimização como gradiente ascendente.

Função objetivo – Função Custo Logístico

Uma alternativa para maximizar a log-verossimilhança é definir uma função custo $J(\cdot)$ para ser minimizada. Assim, pode-se reescrever a log-verossimilhança como:

$$J(\mathbf{w}) = \sum_{i=1}^m -y^{(i)} \log \left(\phi(z^{(i)}) \right) - (1 - y^{(i)}) \log \left(1 - \phi(z^{(i)}) \right) \quad (14)$$

$$J(\phi(z), y; \mathbf{w}) = \begin{cases} -\log(\phi(z)) & \text{if } y = 1 \\ -\log(1 - \phi(z)) & \text{if } y = 0 \end{cases}$$

Figura 13 – Função Custo.

Fonte: Raschka et al. (2022).

Como apresentado na Figura 13, as previsões erradas são penalizadas com um custo cada vez maior.

Otimização com *Gradient Descent* (GD) e *Stochastic Gradient Descent* (SGD)

Gradient Ascent & log-likelihood

Para aprender o coeficiente de peso de um modelo de regressão logística por meio de otimização baseada em gradiente, calculamos a derivada parcial da função de probabilidade de log da seguinte forma:

$$\frac{\partial}{\partial w_j} l(\mathbf{w}) = \left(y \frac{1}{\phi(z)} - (1 - y) \frac{1}{1 - \phi(z)} \right) \frac{\partial}{\partial w_j} \phi(z) \quad (15)$$

Como passo intermediário, calcula-se a derivada parcial da função sigmoide:

$$\begin{aligned}
 \frac{\partial}{\partial z} \phi(z) &= \frac{\partial}{\partial z} \frac{1}{1 + e^{-z}} \\
 &= \frac{1}{(1 + e^{-z})^2} e^{-z} \\
 &= \frac{1}{1 + e^{-z}} \left(1 - \frac{1}{1 + e^{-z}} \right) \\
 &= \phi(z)(1 - \phi(z))
 \end{aligned} \tag{16}$$

substituindo a Equação 16 na Equação 15 obtém-se:

$$\begin{aligned}
 &\left(y \frac{1}{\phi(z)} - (1 - y) \frac{1}{1 - \phi(z)} \right) \frac{\partial}{\partial w_j} \phi(z) \\
 &= \left(y \frac{1}{\phi(z)} - (1 - y) \frac{1}{1 - \phi(z)} \right) \phi(z)(1 - \phi(z)) \frac{\partial}{\partial w_j} z \\
 &= (y(1 - \phi(z)) - (1 - y)\phi(z)) x_j \\
 &= (y - \phi(z)) x_j
 \end{aligned} \tag{17}$$

Agora, para encontrar os pesos do modelo, dá-se um passo proporcional à direção positiva do gradiente para maximizar o log da verossimilhança (*log-likelihood*). Além disso, adiciona-se um coeficiente, a taxa de aprendizado η para a atualização do peso:

$$\begin{aligned}
 w_j &:= w_j + \eta \frac{\partial}{\partial w_j} l(\mathbf{w}) \\
 w_j &:= w_j + \eta \sum_{i=1}^n (y^{(i)} - \phi(z^{(i)})) x_j^{(i)}
 \end{aligned} \tag{18}$$

Observe que o gradiente (e a atualização de peso) é calculado a partir de todas as amostras no conjunto de treinamento do gradiente de subida/descida em contraste com a subida/descida

do gradiente estocástico.

A Equação 18 mostra a atualização de peso para um único peso j . Na otimização baseada em gradiente, todos os coeficientes de peso são atualizados simultaneamente; a atualização de peso pode ser escrita de forma mais compacta como:

$$\mathbf{w} := \mathbf{w} + \Delta \mathbf{w}, \quad (19)$$

onde

$$\Delta \mathbf{w} = \eta \nabla l(\mathbf{w}) \quad (20)$$

Gradient Descent & função de custo logístico

Na seção anterior, derivou-se o gradiente da função *log-likelihood*, a qual pode ser otimizada por meio da subida do gradiente. Da mesma forma, pode-se obter o gradiente de custo da função de custo logístico $J(\cdot)$ e minimizá-lo via gradiente descendente para aprender o modelo de regressão logística.

A regra de atualização para um único peso é:

$$\begin{aligned} \Delta w_j &= -\eta \frac{\partial J}{\partial w_j} \\ &= -\eta \sum_{i=1}^n (y^{(i)} - \phi(z^{(i)})) x^{(i)} \end{aligned} \quad (21)$$

A atualização de peso simultânea:

$$\mathbf{w} := \mathbf{w} + \Delta \mathbf{w} \quad (22)$$

Onde

$$\Delta \mathbf{w} = -\eta \nabla J(\mathbf{w}). \quad (23)$$

Regularização

Como forma de lidar com o *overfitting*, é possível adicionar um viés adicional ao modelo de regressão logística por meio de termos de regularização. Através do termo de regularização $L2$,

reduzimos a complexidade do modelo penalizando grandes coeficientes de peso:

$$L2 : \frac{\lambda}{2} \|\mathbf{w}\|_2 = \frac{\lambda}{2} \sum_{j=1}^m w_j^2 \quad (24)$$

Para aplicar a regularização, basta adicionar o termo de regularização à função de custo (Equação 14), que foi definida para a regressão logística para diminuir os pesos:

$$J(\mathbf{w}) = \sum_{i=1}^m \left[-y^{(i)} \log(\phi(z^{(i)})) - (1 - y^{(i)}) \log(1 - \phi(z^{(i)})) \right] + \frac{\lambda}{2} \sum_{j=1}^m w_j^2 \quad (25)$$

A regra de atualização para um único peso:

$$\begin{aligned} \Delta w_j &= -\eta \left(\frac{\partial J}{\partial w_j} + \lambda w_j \right) \\ &= -\eta \sum_{i=1}^n (y^{(i)} - \phi(z^{(i)}) x^{(i)}) - \eta \lambda w_j \end{aligned} \quad (26)$$

Atualização simultânea de peso:

$$\mathbf{w} := \mathbf{w} + \Delta \mathbf{w} \quad (27)$$

onde

$$\Delta \mathbf{w} = -\eta (\nabla J(\mathbf{w}) + \lambda \mathbf{w}) \quad (28)$$

3.5 MODELO EPIDEMIOLÓGICO BAYESIANO

Devido ao avanço computacional que contribui para a sua popularização no meio científico, essa abordagem tem ganhado destaque na estatística aplicada, incluindo sua utilização em modelos epidemiológicos. A estatística Bayesiana é uma abordagem para análise de dados e estimativa de parâmetros em um modelo estatístico que recebe uma distribuição de probabilidade conjunta. O fluxo de trabalho utilizando a abordagem Bayesiana consiste em três etapas principais: (1) capturar o conhecimento disponível sobre um determinado parâmetro em um modelo estatístico por meio da distribuição *a priori*, que normalmente é determinada antes da coleta de dados; (2) determinar a função de verossimilhança usando as informações sobre os parâmetros disponíveis nos dados observados; e (3) combinando a distribuição *a priori* com a função de verossimilhança por meio do teorema de Bayes (STIGLER, 1982), determinar a distribuição

posterior, esta última representando o conhecimento atualizado sobre os parâmetros do modelo estatístico (SCHOOT et al., 2021).

Nesta pesquisa, o modelo compartimental é utilizado para modelar a dinâmica de uma doença infecciosa (COVID-19) em uma escala populacional.

A dinâmica do SIRD é governada por um sistema de equações diferenciais ordinárias (ODE):

$$\frac{dS}{dt} = -\beta S \frac{I}{N} \quad (29)$$

$$\frac{dI}{dt} = \beta S \frac{I}{N} - \frac{1}{d_I} I \quad (30)$$

$$\frac{dR}{dt} = \frac{1}{d_I} I (1 - \omega) \quad (31)$$

$$\frac{dT}{dt} = \frac{1}{d_I} I \omega - \frac{1}{d_T} T \quad (32)$$

$$\frac{dD}{dt} = \frac{1}{d_T} T \quad (33)$$

Nas equações 29, 30, 31, 32 e 33, as variáveis S , I , R , T e D representam o número de indivíduos de uma população homogênea de tamanho fixo N em cada um dos respectivos compartimentos no tempo t . As derivadas das variáveis ocorrem em relação ao tempo dt e representam uma variação do número de indivíduos em cada compartimento por unidade de tempo. β é a taxa constante de contatos entre indivíduos por unidade de tempo, que são suficientes para levar à transmissão se um dos indivíduos for infeccioso e o outro for suscetível. ω é a taxa de mortalidade constante de indivíduos recuperados. d_I e d_T são o tempo médio, em dias, em que os indivíduos são infecciosos e o tempo médio, em dias, em que os indivíduos estão em estado terminal, respectivamente.

Indivíduos suscetíveis (estado S) entrarão aleatoriamente em contato com indivíduos infectados (estado I) e, conseqüentemente, a partir desse contato, se infectarão com taxa β (equação 29). Uma vez que o indivíduo suscetível é infectado, ele pode infectar outros indivíduos suscetíveis por encontros aleatórios e permanece infectado/infeccioso por uma média de d_I dias (equação 30). Indivíduos infectados podem se recuperar (estado R) com probabilidade $1 - \omega$ (equação 31) ou ficar terminalmente doente (estado T) com probabilidade ω (equação 32). Na etapa final, os indivíduos com doença terminal acabarão por falecer (estado D) em uma média de d_T dias (equação 33). O modelo também pode ser representado em um gráfico acíclico direcionado (Figura 14).

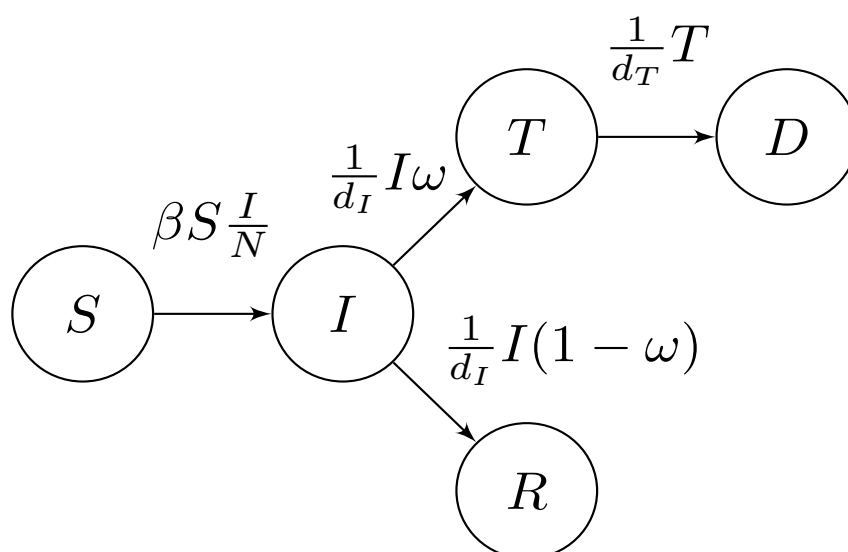


Figura 14 – *SIRTD Modelo Compartimental (STOROPOLI et al., 2021a).*

O modelo SIRTD tem várias suposições. Primeiro, assume que a população N é constante. Em segundo lugar, todo estado é povoado por indivíduos homogêneos, ou seja, sem diferenças demográficas, características sociais ou variáveis relacionadas à saúde. Em terceiro, o modelo assume uma mistura aleatória da população, onde os suscetíveis estão em contato com infectados governados apenas pelo acaso. Quarto, os infectados se tornarão infecciosos (podem espalhar a doença) e se recuperarão ou ficarão doentes terminais. Quinto, os infectados, durante o tempo em que permanecerem infectados, potencialmente infectarão suscetíveis, ou seja, não são tomadas medidas de auto-quarentena ou isolamento.

Este trabalho foi inspirado no modelo Liverpool do CoDatMo (MOORE; PHILLIPS, 2021), que combinou chamadas telefônicas para o serviço público de emergência (111) de pessoas relatando sintomas de SRAG, às autoridades de saúde, com dados semanais de óbitos, em um sofisticado modelo sub-compartimentado. Apesar de utilizarmos, nesta pesquisa, uma versão mais simplificada do modelo proposto por Liverpool, a maior complexidade está em incorporar dados de mídia social em um modelo epidemiológico, principalmente considerando a dimensão continental que o Brasil possui. Além disso, há o grande obstáculo de apenas uma pequena fração dos tweets ser georreferenciada, e alguns deles possuem dados de localização imprecisos (PORCHER; RENAULT, 2021; HUANG et al., 2020). Para superar essas lacunas, desenvolvemos um modelo SIRTD que usa menções de sintomas nas mídias sociais, em Língua Portuguesa do Brasil, para melhor inferir e prever o número de indivíduos infectados (estado I) que dão entrada em serviços de saúde no Brasil. Com isso, nossa intenção é demonstrar como dados de mídias sociais, especialmente menções de sinais e sintomas, podem aprimorar modelos epidemiológicos simples e, por conseguinte, compreender, por meio das menções em redes sociais, quais os principais fatores que levam as pessoas a buscarem atendimento médico em situações de crises sanitárias, como no caso da pandemia da COVID-19.

A fim de testar a hipótese de que as mídias sociais podem contribuir para aprimorar modelos

epidemiológicos, são propostos dois experimentos. O primeiro consiste em um modelo epidemiológico com dados reais de registros de casos graves de COVID-19 pela Vigilância em Saúde do Brasil (Ministério da Saúde, 2021f). No segundo experimento é incorporado, no modelo do experimento, dados de menções realizadas em mídias sociais (Twitter) sobre sinais e sintomas relacionados à COVID-19 (SANTOS et al., 2021). O desenvolvimento dos modelos propostos seguiu o fluxo de trabalho Bayesiano para a modelagem de transmissão de doenças proposto por Grinsztajn et al. (2021): primeiro os modelos foram construídos, depois realizados os ajustes de interesse; na sequência, cada modelo foi criticado e assim o processo se repete. Este ciclo também é semelhante ao fluxo de trabalho Bayesiano proposto por Gelman et al. (2020), que inclui etapas de construção do modelo, inferência e verificação/melhoria do modelo, juntamente com a comparação de diferentes modelos.

Para todos os experimentos, foi utilizada a linguagem de programação probabilística Bayesiana Stan, ideal para especificar modelos estatísticos complexos e realizar inferências usando *Markov Chain Monte Carlo* (MCMC) (CARPENTER et al., 2017). O sistema de equações diferenciais ordinárias (ODE) descrito nas equações 29, 30, 31, 32 e 33, foi implementado e resolvido por um método Runge-Kutta de 4ª/5ª ordem (ISERLES, 2008), usando o algoritmo Dormand-Prince (DORMAND; PRINCE, 1980), com tolerância relativa e tolerância absoluta de $1e-6$ e número máximo de passos $h = 1e4$.

Em relação às especificações, o modelo pode ser da seguinte forma: Primeiro, as distribuições *a priori*. O símbolo $+$ indica que a distribuição é truncada para apenas valores positivos; em outras palavras, somente há suporte positivo para valores acima de 0, \mathbb{R}^+ . O domínio da distribuição se torna $[0, \infty)$. A taxa constante de infecção β é amostrada a partir de uma distribuição normal restrita a valores positivos (equação 34), com média μ_β e desvio padrão σ_β . A taxa de mortalidade constante de indivíduos recuperados ω é amostrada de uma distribuição beta (equação 35), com parâmetros α_ω , representando o número de pessoas que se tornarão doentes terminais e falecidas, e β_ω representando o número de pessoas que se recuperam da doença. O tempo médio durante o qual os indivíduos são infecciosos ou doentes terminais, d_I e d_T , são ambos amostrados a partir de uma distribuição normal restrita a valores positivos (equações 36 e 37), com média μ_{d_I}, μ_{d_T} e desvio padrão $\sigma_{d_I}, \sigma_{d_T}$, respectivamente. A proporção de pessoas infectadas que irão escrever *tweets* diariamente sobre seus sintomas, Proportion Tweets, estando no estado I , é amostrado *a priori* de uma distribuição anterior para proporções como uma distribuição beta (equação 38).

O modelo tem as seguintes especificações de verossimilhança: Ambas as contagens diárias de *tweets* sobre sintomas e contagens cumulativas de mortes possuem distribuição binomial negativa. Para contagens cumulativas de mortes (equação 41), *Deceased*, o parâmetro de localização é o número de indivíduos no estado D (resolvido pelo solucionador numérico de ODEs do Stan), e o parâmetro de precisão ϕ que segue uma distribuição exponencial com parâmetro de taxa λ_ϕ (equação 39). Para as contagens diárias de *Tweets* sobre sinais e sintomas (equação 42), o parâmetro de localização é o número de indivíduos no estado I (também resolvido pelo

solucionador numérico de ODEs do Stan) multiplicado pela proporção de pessoas infectadas que “twittarão” diariamente sobre os seus sinais e sintomas, Proportion Tweets, estando no estado I ; e o parâmetro de precisão ϕ_{tweets} , que segue uma distribuição exponencial com parâmetro de taxa $\lambda_{\phi_{\text{tweets}}}$ (equação 40).

$$\beta \sim \text{Normal}^+(\mu_{\beta}, \sigma_{\beta}) \quad (34)$$

$$\omega \sim \text{Beta}(\alpha_{\omega}, \beta_{\omega}) \quad (35)$$

$$d_I \sim \text{Normal}^+(\mu_{d_I}, \sigma_{d_I}) \quad (36)$$

$$d_T \sim \text{Normal}^+(\mu_{d_T}, \sigma_{d_T}) \quad (37)$$

$$\text{Proportion Tweets} \sim \text{Beta}(1, 1) \quad (38)$$

$$\phi \sim \text{Exponential}(\lambda_{\phi}) \quad (39)$$

$$\phi_{\text{tweets}} \sim \text{Exponential}(\lambda_{\phi_{\text{tweets}}}) \quad (40)$$

$$\text{Deceased} \sim \text{Negative Binomial} \left(\text{state } D, \frac{1}{\phi} \right) \quad (41)$$

$$\text{Tweets} \sim \text{Negative Binomial} \left(\text{state } I \cdot \text{Proportion Tweets}, \frac{1}{\phi_{\text{tweets}}} \right) \quad (42)$$

Em ambos os experimentos, foram definidas as probabilidades *a priori* dos parâmetros do modelo como probabilidades *a priori* semelhantes as que são usadas em alguns modelos epidemiológicos da COVID-19 (MOORE; PHILLIPS, 2021):

$$\beta \sim \text{Normal}^+(2, 1);$$

$$\omega \sim \text{Normal}^+(0.4, 0.5);$$

$$\lambda \sim \text{Beta}(1, 2);$$

$$d_I \sim \text{Normal}^+(7, 2);$$

$$d_T \sim \text{Normal}^+(10, 2);$$

$$\phi \sim \text{Exponencial}(5); \text{ e}$$

$$\phi_{\text{tweets}} \sim \text{Exponencial}(5).$$

Para toda a nossa amostragem, foram utilizadas as configurações padrão de Stan. Isso se traduz em amostragem MCMC usando *Hamiltoniano Monte Carlo* (HMC) (NEAL, 2011) e *No-U-Turn-Sampling* (HOFFMAN; GELMAN, 2011), com quatro cadeias separadas, cada uma com 2.000 iterações, e as primeiras 1.000 (metade do total de iterações) sendo descartadas como *warm-up* (aquecimento), e as últimas 1.000 sendo usadas como amostras da cadeia

de Markov subjacente. Teve-se o cuidado de definir sementes geradoras de números aleatórios específicas para tornar nossos resultados reprodutíveis. Também usamos parâmetros padrão para o amostrador NUTS HMC, o que significa que a taxa de aceitação do valor alvo é 80% (`adapt_delta = 0.8`), e o limite da profundidade das árvores que avalia durante cada iteração é 2^{10} (`max_treedepth = 10`). Por fim, para o ambiente de programação, foram usados o R-Programming version 4.2.0 (TEAM, 2021), Stan version 2.29.1 (CARPENTER et al., 2017), e CmdStanR version 0.5.0 (GABRY; ČEŠNOVAR, 2021).

As previsões considerando as várias ondas da pandemia são normalmente produzidas tendo como suposição que a tendência de crescimento permanece a mesma ao longo do horizonte de previsão, que é plausível para a previsão de uma semana (BHATIA et al., 2021). Corroborando com o argumento anterior, foi observado que, durante a pandemia de COVID-19, as estimativas realizadas com horizonte de previsão de curto prazo (uma semana) produziram resultados melhores do que as realizadas em médio prazo (quatro semanas), e nas tentativas de previsões com um horizonte mais longo os modelos falharam mais em capturar as mudanças abruptas na contagem de casos observados (BHATIA et al., 2021; YOGURTCU et al., 2021). Posto isso, nesta pesquisa, para prever o número de casos graves por COVID-19, foi realizada a predição para a última semana do mês de dezembro.

3.6 LIMITAÇÕES

Embora as mídias sociais ofereçam benefícios para uma estratégia de comunicação em saúde pública mais direcionada, com base nas características individuais de cada usuário do sistema de saúde, é preciso reconhecer que uma parcela significativa da população (quase a metade) não tem acesso à *Internet* (HOOTSUITE, 2021), e grupos específicos, como crianças e idosos, necessitam de cuidados especiais, que ainda não são possíveis de se alcançar utilizando *Internet* e mídias sociais (BERRY, 2007). Outra dificuldade é em relação à informalidade das mídias sociais, fazendo destas mecanismos não regulamentados para coleta, compartilhamento e promoção de informações, de modo que as informações são de qualidade e consistência variadas. No mais, esses problemas são agravados pela natureza interativa da mídia social, que permite que usuários leigos disseminem informações independentemente da qualidade (MORHEAD et al., 2013). Tais situações não foram avaliadas nesta pesquisa.

Esta pesquisa também investiga aspectos culturais, como crenças, religiosidade, características políticas e econômicas das regiões onde as pessoas vivem e trabalham. Também não se aprofunda sobre outras variáveis que podem influenciar o comportamento em saúde, como nível educacional, classe social etc..

Resumo do capítulo

Neste capítulo são apresentados os resultados da pesquisa no que diz respeito a: (a) publicações científicas e produções tecnológicas; (b) estatísticas descritivas da amostra; (c) modelo de Processamento de Linguagem Natural; e (d) modelo epidemiológico Bayesiano.

4.1 PRODUÇÃO CIENTÍFICA E TECNOLÓGICA DA TESE

Esta pesquisa resultou em contribuições científicas e tecnológicas. Em primeiro lugar, destacou-se a publicação do artigo científico dos autores [Storopoli et al. \(2021a\)](#). Nesse trabalho, os autores propõem um algoritmo inovador de simulação para a transmissão de doenças. Esse algoritmo, compatível com o modelo epidemiológico de compartimento padrão para a COVID-19, pode ser adaptado a várias parametrizações, refletindo uma gama de cenários epidêmicos plausíveis. Além disso, o artigo introduz uma maneira de integrar dados provenientes das redes sociais, especificamente menções diárias de sintomas, nos modelos epidemiológicos da COVID-19. Isso representa um avanço significativo na modelagem da doença, pois permite uma compreensão mais detalhada e atualizada de sua disseminação.

Os autores defendem que a simulação proposta pode auxiliar pesquisadores a compreender melhor o comportamento do modelo em um ambiente controlado. Por meio de testes de estresse em diferentes cenários, o modelo pode ser aprimorado antes de ser aplicado a dados reais. Esse processo não apenas promove uma melhor compreensão dos modelos epidemiológicos, mas também potencializa seu desempenho em situações reais, contribuindo, assim, para a eficácia da resposta à pandemia. Vale ressaltar que o artigo publicado foi citado no trabalho de [Moore, Rosato e Maskell \(2022\)](#), onde os autores desenvolveram um novo modelo estatístico para dados de vigilância COVID-19 de várias fontes, ajustando seu hiperparâmetro de suavidade, o que evidencia que frutos desta tese tem auxiliado no desenvolvimento de outras pesquisas.

Outro trabalho de extrema relevância foi a mineração de dados, e posterior publicação, de 13.857.517 de tweets sobre sintomas e sinais de SRAG e COVID-19 ([SANTOS et al., 2021](#)), que em maio de 2023 já possuía 461 *downloads*. Essa estatística mostra a relevância deste trabalho e as contribuições para outras pesquisas.

Outro resultado importante nesta pesquisa veio na publicação no formato pôster, apresentado no 11º Congresso Brasileiro de Epidemiologia. Este pôster detalha a aplicação do Truque de Corrente Linear (LCT, do inglês Linear Chain Trick) em modelos epidemiológicos compartimentais para COVID-19. O LCT modela tempos de espera nas transições de compartimentos utilizando subcompartimentos com uma distribuição Erlang ([STOROPOLI et al., 2021b](#)). Os achados demonstraram uma margem de erro de cerca de ± 85 óbitos semanais ao comparar as

mortes previstas com as reais. Os dados e códigos utilizados neste estudo estão disponíveis para consulta no repositório LabCidades no Github ([LabCidades, 2021](#)). Os autores desta pesquisa também foram convidados a participar como palestrantes e compartilhar parte dos achados desta pesquisa no *webinar* "Cidades e Pandemias: Lições e Desafios para uma Recuperação Resiliente em São Paulo", disponível no canal da Universidade Nove de Julho no YouTube ([UNINOVE, 2022](#)).

Por fim, vale ressaltar que esta pesquisa foi totalmente desenvolvida com ferramentas *open-source*. Assim, todas aplicações desenvolvidas ao longo da pesquisa estão disponíveis no repositório LabCidades ([LABCIDADE, 2021](#)), o que permite que outros pesquisadores consigam reproduzir tanto o modelo epidemiológico proposto, quanto o modelo de NLP ou outras aplicações de apoio desenvolvidas que auxiliaram na manipulação dos dados, para gerar gráficos, mapas e outras estatísticas de interesse. Este vasto conjunto de recursos disponíveis representa uma contribuição valiosa para a comunidade científica e tecnológica, auxiliando outros pesquisadores na compreensão e expansão dos conhecimentos adquiridos neste estudo.

4.2 ESTATÍSTICA DESCRITIVA DA AMOSTRA

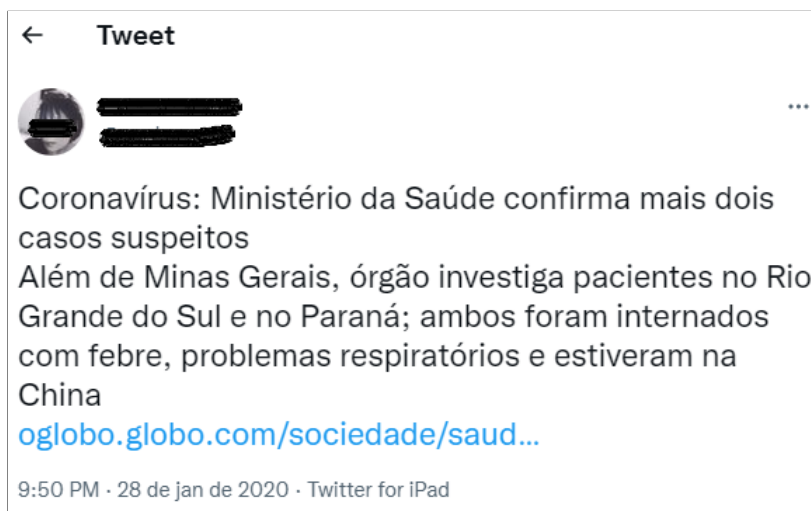
A análise dos resultados iniciam-se com a parte descritiva dos *tweets* coletados para essa pesquisa. Na fase inicial foram coletados, para o ano de 2020, um total de 6.155.844 de *tweets*. Contudo, após aplicar os critérios de exclusão restaram apenas 95.964 *tweets*, o que representa 1,5% do total dos dados coletados. A seguir, na Tabela 3, é possível observar a característica (por mês) da amostra com os 95 mil *tweets* que foram classificados como sinal ou ruído, bem como a amostra aleatória por proporção de cada mês, de 10.004 *tweets*, dos 95 mil que foram utilizados para treinar e validar o algoritmo de processamento de linguagem natural.

Tabela 3 – Amostra com dados rotulados como sinal ou ruído (Etapa 3) – ano 2020.

Mês	Dados coletados		Dados rotulados		
	Tweets	Freq (%)	Ruído	Sinal	Total
Janeiro	3.286	3,42	178 (3,9%)	178 (3,9%)	343
Fevereiro	3.685	3,84	162 (2,9%)	222 (4,9%)	384
Março	17.712	18,46	982 (17,7%)	864 (19,2%)	1.846
Abril	10.778	11,23	592 (10,7%)	532 (11,8%)	1.124
Maiο	9.105	9,49	487 (8,8%)	462 (10,3%)	949
Junho	5.976	6,23	309 (5,5%)	314 (7,0%)	623
Julho	9.658	10,06	611 (11,0%)	396 (8,8%)	1.007
Agosto	8.682	9,05	586 (10,6%)	319 (67,1%)	905
Setembro	6.629	6,91	428 (7,7%)	263 (5,8%)	691
Outubro	6.103	6,36	359 (6,5%)	277 (6,1%)	636
Novembro	6.101	6,36	352 (6,3%)	284 (6,3%)	636
Dezembro	8.249	8,60	493 (8,9%)	367 (8,2%)	860
Total	95.964	100	5.526 (100%)	4.478 (100%)	10.004

Fonte: elaborado pelo autor.

Os resultados da Tabela 3 referem-se à Etapa 3 do processo de tratamento dos dados da pesquisa (Figura 7), onde o objetivo era utilizar o classificador de *machine learning* para identificar quais *tweets* (sinais) seriam incorporados ao modelo Bayesiano, e para identificar os principais temas nas mensagens sobre a pandemia de COVID-19 e associados aos construtos do Modelo de Crença em Saúde. Como pode ser observado na Tabela 3, os primeiros relatos de sintomas de Covid-19 no Brasil aconteceram em janeiro, período em que começaram os primeiros relatos de casos em viajantes com sintomas, como pode ser observado na Figura 15.

Figura 15 – *Tweet relatando casos de COVID-19 em viajantes*

Fonte: dados da pesquisa

Da amostra de *tweets* rotulados como sinal, visando uma melhor compreensão, conforme a teoria HBM, das motivações das pessoas para procurarem serviços de saúde, foram selecionados aleatoriamente quatro mil *tweets*, e observou-se que 60% dos usuários mencionaram que já tinham procurado atendimento médico para COVID-19 e 40% que iriam ou tinham interesse em buscar auxílio médico. Os resultados também mostraram que 51% das pessoas utilizaram as redes sociais para falar dos seus próprios sinais e sintomas e problemas relacionados à COVID-19. Os demais utilizaram para falar de familiares, amigos, colegas e até pessoas desconhecidas que tiveram algum sintoma relacionado à doença (Tabela 4). A princípio, é possível perceber que quanto mais próxima a relação entre a pessoa que escreve a mensagem com a pessoa mencionada com o sinal/sintoma maior é a percepção sobre a gravidade da doença e, consequentemente, maior é a necessidade de buscar auxílio médico.

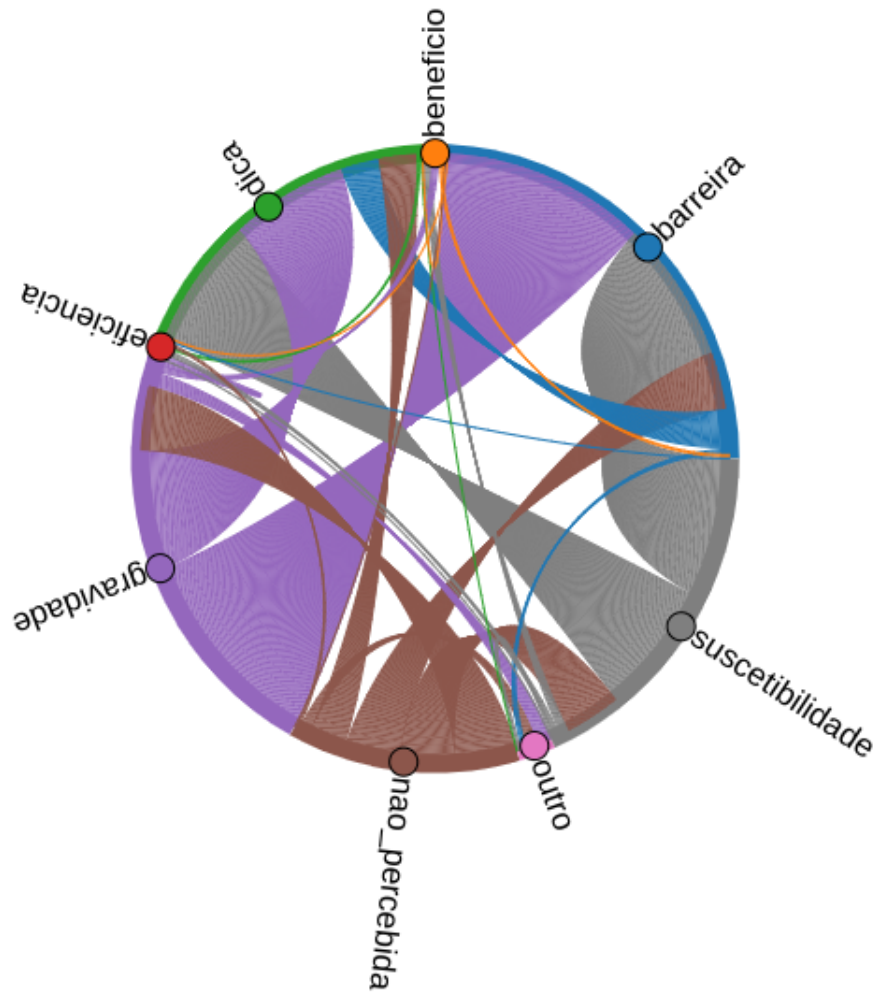
Tabela 4 – Distribuição (%) das pessoas mencionadas nos tweets segundo os construtos do Modelo de Crenças em Saúde. N = 4.000

Pessoas	Frequência (%)	COVID-19	Não Percebida	Suscetibilidade	Gravidade
Própria	51,47	27,48	46,84	58,71	46,86
Rede	8,26	2,25	6,15	17,86	5,85
Outros	7,16	11,85	5,42	5,14	9,26
Família	7,10	9,70	10,31	4,14	5,71
Amigos	5,52	11,64	7,78	5,43	5,23
Filhos	3,70	5,82	3,25	1,86	4,80
Mãe	3,54	1,23	4,70	0,86	4,99
Avós	3,38	7,87	1,99	0,71	5,08
Pai	2,49	4,19	2,89	1,29	3,17
Tios	1,80	3,78	2,17	1,14	1,92
Casal	1,72	4,70	3,44	0,86	2,11
Irmãos	1,66	3,17	2,53	1,00	2,06
Primos	0,86	2,55	1,27	0,14	1,25
Sobrinhos	0,46	1,02	0,18	0,29	0,62
Sogros	0,35	1,23	0,72	0,29	0,38
Cunhados	0,24	0,92	0,36	0,14	0,29
Netos	0,16	0,30	0,00	0,00	0,24
Padrinhos	0,11	0,30	0,00	0,00	0,18
Genros	0,02	0,00	0,00	0,14	0,00
Total (%)	100	100	100	100	100

Nota: Os *tweets* não classificados como "Família" quando são usadas expressões como "um parente meu ficou doente", mas não é possível identificar o grau de parentesco. Em uma mesma mensagem é possível encontrar referências a mais de um construto.

Fonte: elaborado pelo autor.

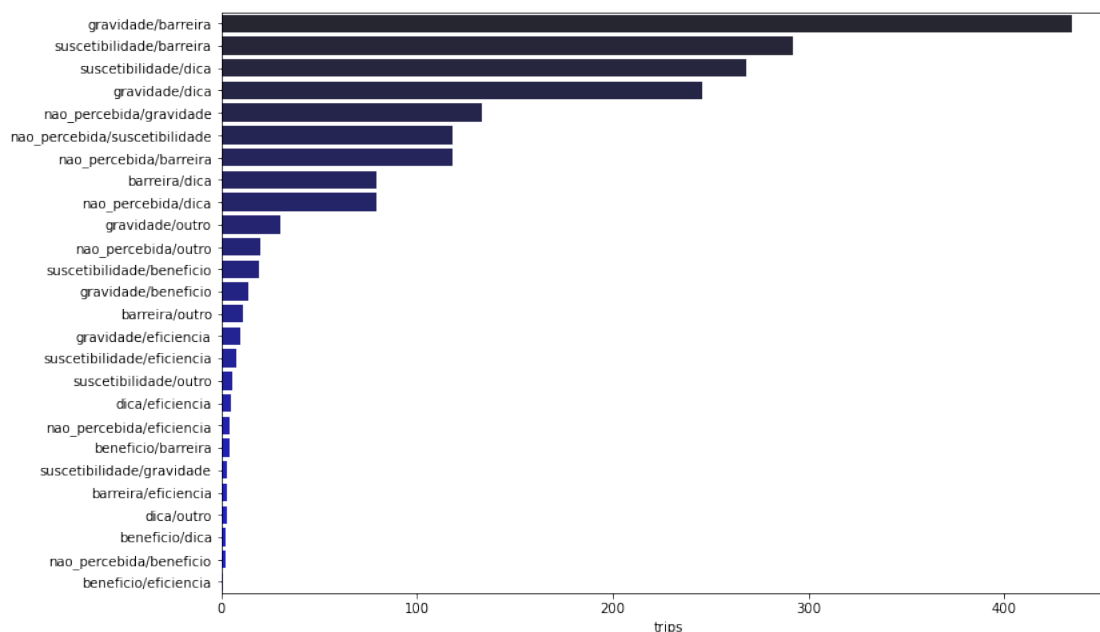
Sob a ótica da HBM, uma característica identificada nos dados analisados é a associação entre os construtos da HBM, os quais geralmente são mencionados juntos. Nas redes sociais, no contexto da pandemia da COVID-19, em uma mesma mensagem podem ser identificados construtos mais fortemente associados a uns ou outros construtos. Por exemplo, na representação gráfica (Figura 16) dos resultados parciais desta pesquisa, podemos observar que o construto Gravidade Percebida (rótulo: gravidade) tende a ser mencionado junto a outros construtos da HBM, entre eles Barreira Percebida (rótulo: barreira) e Dica (rótulo: dica).

Figura 16 – Representação das relações entre os construtos da Health Belief Model

Fonte: dados da pesquisa

Considerando apenas as associações em pares dos construtos da HBM identificados nos dados, a associação mais frequente nas mensagens é entre a "Gravidade Percebida" e "Barreira Percebida" (Figura 17). Uma análise qualitativa do tipo de comunicação presente nos tweets poderá indicar quais são os principais fatores relacionados à associação entre a "Gravidade Percebida" e "Barreira Percebida", assim como para as demais associações.

Figura 17 – *Frequência com que os construtos da Health Belief Model são mencionados (aos pares) em um mesmo tweet. N = 4.000*



Fonte: elaborado pelo autor.

No contexto da pandemia da COVID-19, é possível estender o conceito dos construtos da HBM (Quadro 6). No modelo original proposto por [Rosenstock \(1974a\)](#), os construtos da HBM eram relacionados apenas às crenças sobre uma doença específica. Mas, como pode ser observado no contexto da pandemia do novo coronavírus, representado no Quadro 6, além da doença, outros fatores também devem ser considerados, pois influenciam na percepção que as pessoas têm sobre a gravidade da pandemia.

O HBM tem como objetivo explicar os comportamentos associados ao risco à saúde, e sua proposta central é que as pessoas têm maior probabilidade de tomar medidas preventivas se perceberem que a ameaça do risco à saúde é grave ([MICHIE, 2014](#); [ROSENSTOCK, 1974a](#)). Porém, em um cenário de crise sanitária, como na pandemia de COVID-19, não são apenas os efeitos diretos da doença que podem causar risco à saúde; problemas causados por sedentarismo, como depressão, doença coronariana, diabetes tipo 2, certos tipos de câncer, além de saúde mental ([REZENDE et al., 2014](#)) e vulnerabilidades socioeconômicas críticas durante a pandemia de COVID-19, relacionadas a emprego, habitação, condição financeira e inclusão digital ([SARKER; SHRESTHA; TAMANG, 2022](#); [MIKOLAI; KEENAN; KULU, 2020](#)), também podem influenciar a percepção sobre a gravidade da doença. As afirmações anteriores também foram observadas nos dados da pesquisa, nos relatos feitos pelas pessoas no Twitter, o que evidencia a necessidade de ampliar o conceito dos construtos da HBM no contexto da COVID-19.

Quadro 6 – *Extensão dos Construtos da HBM para a COVID-19, para os tweets da amostra.*

Construtos da HBM	Construtos da HBM para COVID-19
Ameaça não percebida:	Ameaça não percebida:
(a) Sobre a doença	(a) Sobre a doença (b) Medidas de combate à pandemia
Suscetibilidade percebida:	Suscetibilidade percebida:
(a) Sobre a doença	(a) Sobre a doença (b) Medidas de combate à pandemia
Gravidade percebida:	Gravidade percebida:
(a) Sobre a doença	(a) Sobre a doença (b) Medidas de combate à pandemia
Barreira percebida:	Barreira percebida:
(a) Sobre a doença	(a) Sobre a doença (c) Serviços de Saúde (d) Intervenção dos órgãos competentes
Benefícios percebidos:	Benefícios percebidos:
(a) Sobre a doença	(a) Sobre a doença (c) Serviços de Saúde
Dicas para ação:	Dicas para ação:
(a) Sobre a doença	(a) Sobre a doença (c) Assuntos relacionados à pandemia: denúncias; busca por apoio psicológico nas redes sociais; e orientações sobre processos administrativos.

Continua na próxima página.

Quadro 6 – *Extensão dos Construtos da HBM para a COVID-19, para os tweets da amostra.*

Construtos da HBM	Construtos da HBM para COVID-19
Auto-eficácia percebida:	Auto-eficácia percebida:
(a) Sobre a doença	(a) Sobre a doença
Outras variáveis:	Outras variáveis:
(e) Demográficas	(e) Demográficas
(f) Estruturais	(f) Estruturais

Fonte: desenvolvido pelo autor.

Além dos problemas de saúde relacionados diretamente à COVID-19, como um simples resfriado a casos mais graves levando à morte, os resultados mostraram que outros fatores também influenciam a percepção sobre a gravidade da doença, em um cenário de pandemia. Entre eles, as medidas de combate à pandemia que foram encontrados em três níveis diferente de percepção: (i) ameaça não percebida, (ii) suscetibilidade percebida, e (iii) gravidade percebida. Medidas protetivas, como o uso de máscaras, distanciamento social, uso de álcool em gel e medicamentos experimentais ou sem efeito comprovado, como a cloroquina, foram os principais influenciadores para esses três construtos mencionados. O tipo e a qualidade dos serviços de saúde também foram, para muitas pessoas, considerados como uma barreira para controlar e mitigar os efeitos da pandemia. O principal temor das pessoas, em relação às unidades de saúde, é sobre as chances de se infectarem nesses lugares, ou, ainda, de agravar o quadro clínico. Outra barreira sobre os serviços de saúde diz respeito às diferenças de atendimento nas unidades públicas em relação às particulares.

Um benefício percebido sobre os serviços de saúde que foi relatado pelas pessoas nas redes sociais, foi a possibilidade de realizar consultas médicas de forma remota por telemedicina. A telemedicina, durante a pandemia do novo coronavírus, foi uma das linhas de defesa dos profissionais de saúde. Contribuiu para retardar a propagação do coronavírus, mantendo o distanciamento social e prestando serviços por telefone ou videoconferência para cuidados pessoais leves (VIDAL-ALABALL et al., 2020).

As ações dos órgãos competentes divergiu opiniões e, como no caso dos serviços de saúde presenciais, também foram consideradas uma barreira ao combate à pandemia. A falta de uma estratégia de comunicação em saúde a nível nacional eficaz na pandemia foi considerada como uma barreira, e gerou muitas dúvidas na população, nas redes sociais, sobre se as ações tomadas pelo governo para o combate à pandemia seriam capazes de contê-la. Assim, os discursos divergentes entre os governos federal, estaduais e municipais levaram a população a questionar quais diretrizes seriam as melhores e, conseqüentemente, quais a população deveria seguir.

4.3 MODELO DE *Machine Learning*

Na Tabela 5, é possível observar os resultados para cada modelo testado nessa pesquisa. Os resultados dessa tabela referem-se aos 1.001 dados de treino (*support*) que foram reservados para avaliar os modelos em questão, do total de 10.004 dados (*tweets*) rotulados para executar o modelo de classificação (Tabela 3). Esses resultados corroboram com a escolha, nesta pesquisa, pelo modelo LR (*tf-idf*), visto que este teve melhor desempenho em todas as métricas avaliadas.

Tabela 5 – Estatísticas dos modelos de NLP testados na pesquisa

Algoritmos	AUC
LR (<i>tf-idf</i>)	0,9235
LightGBM	0,9135
LR (<i>w2v</i>)	0.8589
NB (<i>tf-idf</i>)	0.9014
ComplementNB	0.9014

Fonte: dados da pesquisa.

Uma métrica importante, utilizada para avaliar modelos de classificação em *machine learning*, é a AUC (do inglês "*Area under the ROC Curve*"). Essa medida, também denominada acurácia do modelo (*Accuracy*), pode ser obtida pela taxa da classe "verdadeiro positivo" (*recall*). No entanto, a AUC é sensível ao desequilíbrio de uma classe; logo, deve ser usada com cautela (CHICCO; JURMAN, 2020), o que se aplica a essa pesquisa, em visto que na coleta inicial foi possível observar um desequilíbrio acentuado para a classe *ruído*. A métrica AUC pode assumir valores entre os limites 0 e 1 (PEDREGOSA et al., 2011); assim, as estatísticas apresentadas na Tabela 6 mostram que, de acordo com a AUC, o algoritmo LR (*tf-idf*) apresentou o melhor resultado, atingindo um desempenho de 92,35% (Tabela 5).

Considerando apenas o classificador LR (*tf-idf*), as estatísticas presentes na Tabela 6 mostram uma pequena diferença nos resultados por classe. Em todas as métricas avaliadas, o classificador teve um desempenho superior para a classe *ruído*, sendo que a maior diferença foi na métrica *precision*, que avalia o número de verdadeiros positivos, ou seja, o quanto dos dados válidos para pesquisa foram classificados corretamente, enquanto o *recall* expressa a capacidade de encontrar todas as instâncias relevantes de uma classe e à medida que *precision* aumenta, o *recall* tende a cair. No entanto, para encontrar uma combinação ideal entre *recall* e *precision*, e assim avaliar o modelo por uma única métrica, pode-se combinar as duas métricas usando o *f1-score*, que é uma das métricas de avaliação de modelos de *machine learning* mais utilizadas e uma das mais importantes (Equação 43).

$$F_1 \text{ score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (43)$$

Quando TP^1 é igual a 0 ($TP = 0$) significa que todas as amostras positivas são classificadas erroneamente, e o quando FN^2 é igual a FP^3 e são iguais a 0 ($FN = FP = 0$) significa que o máximo para uma classificação foi alcançado (CHICCO; JURMAN, 2020). Então, nessa pesquisa, o algoritmo de classificação LR(tf-idf) obteve a maior média harmônica para a precisão na classe ruído, conseguindo classificar de forma correta 85% dos dados dessa classe. Isso é positivo para a pesquisa, pois é melhor que o modelo erre mais excluindo dados "bons"(sinal) do que ele incluir na amostra final, que será analisada, mais dados (tweets) "ruins"(ruído), o que poderia dificultar a interpretação dos resultados. Isso é diferente do que se esperaria se o objetivo fosse diagnosticar pessoas para o tratamento de COVID-19, que nesse caso seria melhor termos na triagem mais pessoas classificadas erroneamente pelo modelo como positivas, do que excluir da triagem possíveis casos graves.

Ademais, outras duas métricas também foram avaliadas no modelo LR(tf-idf): `macro avg` (do inglês, *averaging the unweighted mean per label*) e `weighted avg` (do inglês, *averaging the support-weighted mean per label*). A `macro avg` considera o quanto de cada classe há no modelo, portanto, menos de uma classe significa que ela tem menos impacto no modelo de classificação. Por outro lado, a `weighted avg` refere-se ao número de ocorrências reais da classe no conjunto de dados. Assim, considerando que, para todas as métricas na Tabela 6, as estatísticas apresentadas sobre `macro avg` e `weighted avg` foram iguais, pode-se considerar que a diferença no número observações (tweets) entre as classes não foi significativa para a fase de teste.

Tabela 6 – Estatísticas dos modelo RL com dados de teste. $N = 1.001$

	precision	recall	f1-score	support
Ruído (rótulo 0)	0,86	0,85	0,85	539
Sinal (rótulo 1)	0,83	0,84	0,83	462
Accuracy			0,84	1.001
macro avg	0,84	0,84	0,84	1.001
weighted avg	0,84	0,84	0,84	1.001

Fonte: dados da pesquisa.

Os resultados (`f1-score`) podem ser representados de forma absoluta na matriz de confusão (Tabela 7), onde-se pode observar que dos 1.001 tweets utilizados no teste (`support`) dos rotulados como ruído, 75 foram classificados errados, e dos rotulados como sinal, 82 foram classificados errados pelo modelo LR(tf-idf). Ao analisar os tweets que o modelo não conseguiu prever adequadamente, sinal e ruído, é possível observar uma similaridade nos textos (Figura 18). No geral, esses dois grupos são muito parecidos, são sintomas e outros

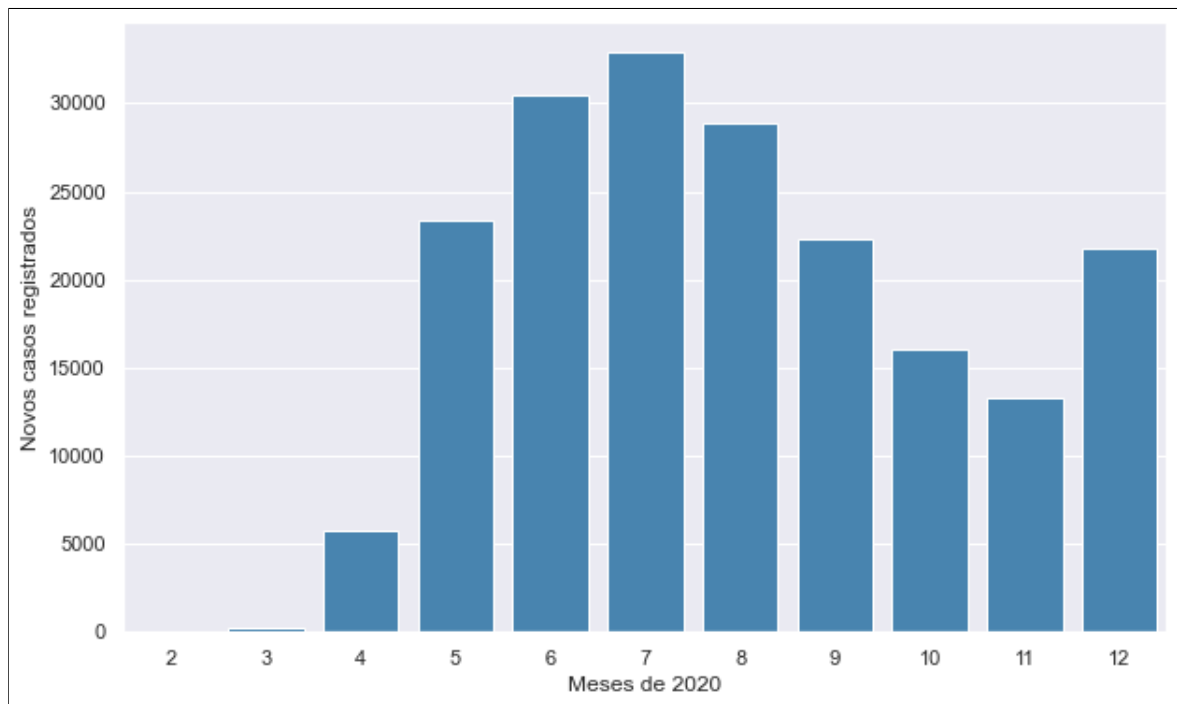
¹TP: True Positive

²FN: False Negative

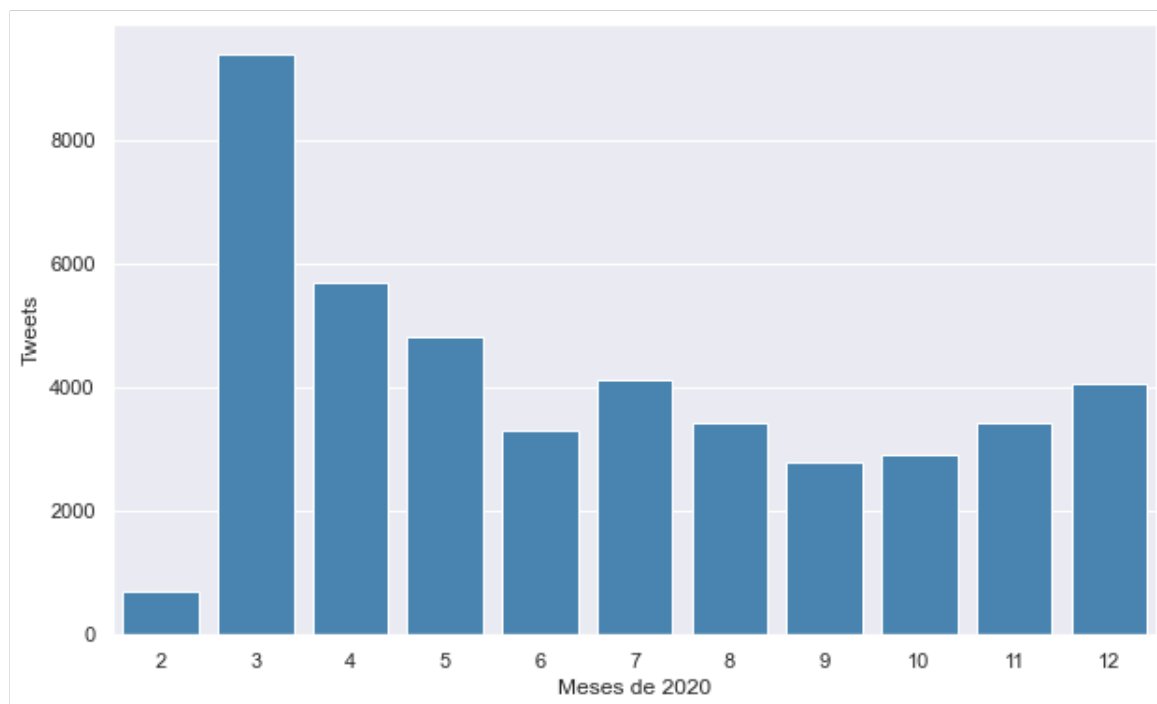
³FP: False Positive

de COVID-19 no Brasil foi registrado no mês de julho, nas redes sociais, conforme os dados coletados nesta pesquisa, o pico foi em março, com um total de 9.413 tweets, e o menor valor foi de 716 tweets em fevereiro. A média mensal foi de 4.073 tweets com uma variação para mais ou para menos de aproximadamente 46%, conforme é demonstrado na Figura 20.

Figura 19 – *Novos casos de SRAG por COVID-19 em 2020.*



Fonte: dados da pesquisa.

Figura 20 – *Tweets (em Língua Portuguesa) sobre sinais e sintomas de COVID-19 em 2020. N = 48.281*

Fonte: dados da pesquisa.

Como pode ser observado nas Figuras 19 e 20, no início de 2020, o número de registro de novos casos de COVID-19 não acompanhou a tendência de menções sobre os sintomas de COVID-19 nas redes sociais. O fato é que, no primeiro semestre de 2020, houve uma campanha nacional muito forte contra a COVID-19 para orientar as pessoas a ficarem em casa (veja alguns exemplos na Figura 21). Isso acabou refletindo em um baixo número de registros de novos casos de COVID-19 em todo território nacional, em relação à média mensal. Contudo, conforme os meses foram passando e a campanha começou a perder força, os perfis de pessoas que antes apenas comentavam nas redes sociais que estavam com sintomas de COVID-19, mas não procuravam atendimento médico, passaram a procurar as unidades de saúde para terem um diagnóstico relacionado à COVID-19. Assim, a partir do mês de junho, é possível observar que a relação entre os tweets, registros de novos casos e também novas mortes por COVID-19 passou a ser fortemente correlacionada (tweets e novos casos) e moderadamente correlacionada (tweets e novas mortes) (Figura 22).

Figura 21 – Campanha nacional de combate à COVID-19 "Fique Em Casa", março de 2020.



Fonte: (SP, 2020; IFBA, 2020; MG, 2020; ES, 2020; PR, 2020).

Figura 22 – Matriz de correlação entre casos, mortes e tweets de COVID-19 em 2020.



Fonte: dados da pesquisa.

A força de associação das variáveis na Figura 22 corrobora com o argumento exposto nesta tese, de que os comentários em redes sociais tendem a refletir o comportamento das pessoas no dia a dia. Assim, a utilização de dados de redes sociais pode ajudar a melhorar modelos epidemiológicos e prever o número de hospitalizações por COVID-19, como pode ser verificado

no resultado do modelo epidemiológico na Tabela 8, onde o modelo que utilizou dados de redes sociais apresentou um melhor resultado preditivo. Todavia, a correlação fraca entre tweets e casos graves de COVID-19 no início da pandemia deve ser melhor investigada em estudos futuros, pois pode prejudicar as análises se o objetivo for utilizar dados de redes sociais em um curto intervalo de tempo (o que foge do escopo desta pesquisa), principalmente na utilização no início de surtos.

Tabela 8 – Comparação entre o modelo sem tweets (1) e o modelo com tweets (2)

Parâmetros	modelo1	modelo2
elpd_diff	-107.499	0.000
se_diff	85.492	0.000
elpd_loo	-126.766	-19.268
se_elpd_loo	85.560	0.068
p_loo	111.014	0.021
se_p_loo	84.386	0.000
looic	253.533	38.536
se_looic	171.119	0.136

Fonte: O autor.

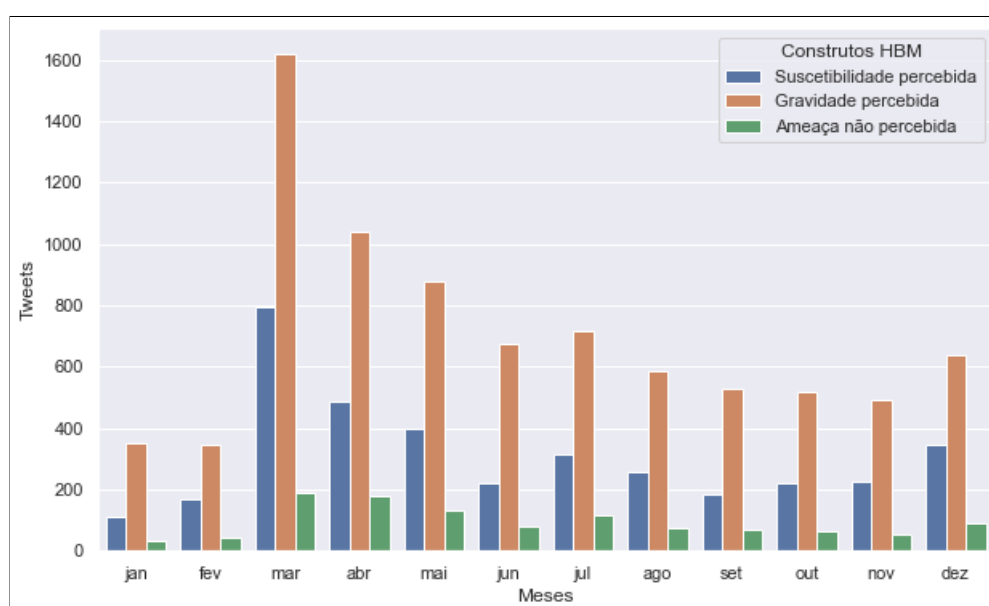
Como resultado final do modelo epidemiológico, considerando a estimativa para a última semana do mês de dezembro de 2020, pode-se concluir que o modelo2, que incorporou dados de redes sociais, teve um desempenho superior que o modelo sem dados de redes sociais. Para a melhor compreensão desse resultado, os valores apresentados na Tabela 8 podem ser interpretados tomando como referência o modelo2, e a partir dele analisar as diferenças entre os dois modelos. O $elpd$ que é a densidade preditiva logarítmica esperada teórica para um novo conjunto de dados. Nesse caso, os dados da última semana de registro de novos casos de COVID-19, que deseja-se estimar, apresentou um resultado favorável para o modelo com tweets (modelo2).

O $elpd_{loo}$ é a estimativa Bayesiana LOO (*Leave-one-out cross-validation*) (VEHTARI; GELMAN; GABRY, 2017) da densidade preditiva logarítmica esperada, e é uma soma de N densidades preditivas logarítmicas pontuais individuais. As densidades de probabilidade podem ser menores ou maiores que 1 e, portanto, as densidades preditivas logarítmicas podem ser negativas ou positivas. O erro padrão do $elpd_{loo}$ ($se_{elpd_{loo}}$) é uma descrição da incerteza sobre o desempenho preditivo para dados futuros desconhecidos. Quando N é pequeno ou há especificação incorreta grave do modelo, a estimativa atual de SE é superotimista, e o SE real pode até ser duas vezes maior. Nessa pesquisa, o resultado referente ao $se_{elpd_{loo}}$ também corrobora com o argumento de que o modelo com tweets teve um desempenho melhor. Contudo, a recomendação ao fazer comparações de modelos é que o SE das diferenças entre componentes (se_{diff}) deve ser utilizado (VEHTARI et al., 2022).

O resultado do p_{loo} refere-se à diferença entre $elpd_{loo}$ e a densidade preditiva posterior de log não validada cruzadamente, e descreve o quanto é mais difícil prever novos dados do que os dados observados. Como o valor de p_{loo} (e o erro padrão $se_{p_{loo}}$) no modelo sem tweets foi muito superior aos resultados do modelo com tweet e muito superior ao número total de parâmetros no modelo, isso indica que o modelo sem dados de redes sociais tem uma capacidade preditiva muito fraca. Já o $looic$ tem a mesma função do AIC (Akaike information criterion), ou seja, é um estimador de erro de previsão e representa, portanto, a qualidade relativa de modelos estatísticos para um determinado conjunto de dados o qual é apenas $-2 * elpd_{loo}$, e quanto menor este valor melhor será o modelo (VEHTARI; GELMAN; GABRY, 2017). Um ponto importante sobre o $looic$ é que deve ser analisado em diferentes modelos considerando o mesmo conjunto de dados, e, como nesta pesquisa tem-se um conjunto de dados com dados de redes sociais e outro sem, nesse caso é recomendado olhar diretamente o $elpd_{loo}$.

A utilização de dados de redes sociais vai além de permitir melhorar a predição dos modelos epidemiológicos. Os comentários realizados nas redes sociais permitem análises contextuais relacionados a um determinado fenômeno e sua evolução, como no caso desta pesquisa sobre a pandemia de COVID-19. Como pode ser observado na Figura 23, utilizando a mesma amostra de tweets incorporada no modelo2, e com base na classificação desses dados considerando os construtos da HBM (suscetibilidade percebida, gravidade percebida e ameaça não percebida), é possível observar que o construto gravidade percebida é o principal construto da HBM responsável por fazer com que as pessoas realizem uma ação ligada à saúde, nesse caso procurar atendimento médico por considerarem a COVID-19 gravíssima ou não irem a uma unidade de saúde por acharem que o risco de irem, se infectarem e morrerem é maior do que ficar em casa.

Figura 23 – Percepção da COVID-19 nas redes sociais de acordo com os construtos da Health Belief Model, 2020.



Fonte: dados da pesquisa.

Resumo do capítulo

Neste capítulo são apresentados os pontos centrais desta dissertação e as considerações finais.

Esta pesquisa teve como objetivo aprimorar um modelo epidemiológico incorporando dados de redes sociais e processamento de linguagem natural para estimar casos graves por COVID-19 no Brasil. Este trabalho foi inspirado no projeto de Liverpool CoDatMo (MOORE; PHILLIPS, 2021), que utilizou modelos epidemiológicos Bayesianos combinados com dados de chamadas telefônicas para o serviço público de emergência de pessoas relatando sinais e sintomas com dados semanais de óbitos.

A principal diferença desta pesquisa para a de Liverpool é que nesta pesquisa, foram utilizados registros de casos graves de COVID-19, no lugar de óbitos, e ao invés de utilizar dados de chamadas telefônicas, foram utilizados dados de redes sociais, especificamente tweets com menções de sinais e sintomas relacionados à COVID-19. A incorporação de dados de redes sociais, ao invés de chamadas telefônicas, trouxe mais complexidade à pesquisa, visto que, diferentemente das chamadas telefônicas que conforme vão acontecendo, vão sendo gradualmente armazenadas e classificadas, no caso dos tweets o processo de coleta e de tratamento de dados teve que ser desenvolvido nesta pesquisa empregando técnicas de inteligência artificial para a mineração e classificação dos dados. Já, a utilização de registros de casos graves por COVID-19, ao invés de óbitos, foi importante pois permite aos tomadores de decisão terem uma melhor percepção do avanço da doença em situações que permitem tratamento.

Para estimar o número de pessoas hospitalizadas com casos graves de COVID-19, primeiro foi realizada a mineração e o tratamento de dados no Twitter, utilizando técnicas de aprendizado de máquina com apoio computacional da linguagem de programação Python (ROSSUM; WARSAW; COGHLAN, 2001). Foram quase 14 milhões de dados coletados entre os anos de 2019 e junho de 2021, dados que estão disponíveis no repositório Zenodo (SANTOS et al., 2021) para consulta e para serem aplicados em pesquisas futuras. Inclusive, sugere-se que em estudos futuros sejam utilizados os dados do ano de 2019, período antes da pandemia, para se realizar uma abordagem comparativa entre o período pré e pós-pandemia. Ainda, considerando que no Brasil, a partir de janeiro de 2021, iniciou-se a campanha de vacinação contra a COVID-19, também é possível estender essa pesquisa e analisar os efeitos que a campanha de vacinação teve na propagação do novo coronavírus.

Finalizada a coleta e o tratamento dos dados de redes sociais, foi gerada outra amostra, desta vez com dados coletados no banco de dados de Síndrome Respiratória Aguda Grave (Ministério da Saúde, 2021e) de casos graves de COVID-19. A fim de responder a questão de pesquisa e identificar como os dados de redes sociais podem aprimorar modelos epidemiológicos, foram desenvolvidos dois modelos, um apenas com registros de casos graves de COVID-19 e outro

incorporando dados de redes sociais. Dessa forma, foi possível observar que o modelo epidemiológico que utilizou dados de redes sociais teve um desempenho superior e conseguiu prever melhor o número de hospitalizações por COVID-19. Contudo, esse estudo limitou-se a considerar apenas os dados agrupados em nível nacional, assim, não foram analisados os efeitos em cada região do Brasil. Nesse aspecto, para analisar o efeito de cada região, o modelo epidemiológico proposto pode ser adaptado para um modelo hierárquico como em [Chatzilena et al. \(2019\)](#), que consideraram um modelo hierárquico, para inferir sobre incidências de gonorreia para idades específicas e regiões na Inglaterra. Em estudos futuros, esses modelos hierárquicos podem ser utilizados para analisar o efeito de diversas variáveis na disseminação da doença, tais como idade, gênero, raça e região, por exemplo.

Os dados de redes sociais sobre sintomas relacionados à COVID-19 também tiveram um papel fundamental para avançar na literatura da comunicação em saúde, especificamente na teoria da *Health Belief Model* ([ROSENSTOCK, 1974a](#)). No modelo original, os construtos da HBM eram utilizados apenas para analisar e classificar as crenças das pessoas sobre uma determinada doença. Porém, nesta pesquisa, observou-se que, no contexto da pandemia, esses construtos se estendem também às crenças que as pessoas possuem sobre medidas de combate à pandemia. Além do mais, relacionando-se os dados classificados com base nos construtos da HBM, e tendo como referência o modelo epidemiológico que incorporou dados de redes sociais, foi possível concluir que a gravidade percebida sobre a pandemia foi responsável por levar as pessoas a tomarem uma ação de saúde, seja para procurarem uma unidade de saúde, ou no início da pandemia para ficarem em casa, por medo de irem ao hospital e serem infectados.

Por fim, o uso de mídia social pode ter inúmeros benefícios para a saúde pública, tanto para o público em geral, como pacientes e profissionais de saúde. As mídias sociais são acessíveis para vários grupos populacionais, independentemente de idade, educação, raça, etnia ou localidade, e, em comparação aos métodos tradicionais de comunicação, as redes sociais são mais democráticas ([MOORHEAD et al., 2013](#)). Agora, o que compete aos órgãos públicos é ampliar o acesso da população à Internet e, conseqüentemente, ampliar o acesso às informações de saúde para alcançar um público ainda maior e permitir compreender melhor o desenvolvimento das epidemias e pandemias e, assim, auxiliar nas diretrizes e tomadas de decisões de enfrentamento, para a melhora da saúde pública.

REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMS, S. A. Blog-based applications and health information: Two case studies that illustrate important questions for consumer health informatics (chi) research. *International Journal of Medical Informatics*, v. 79, n. 6, p. e89–e96, 2010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1386505608001032>>. Citado na pág. 42.
- ADLER, P. S.; KWON, S.-W. Social capital: Prospects for a new concept. *The Academy of Management Review*, v. 27, n. 1, p. 17, 2002. Disponível em: <<https://www.jstor.org/stable/4134367>>. Citado na pág. 49.
- Agência Brasil. *Brazil confirms first case of coronavirus*. 2020. <<https://agenciabrasil.ebc.com.br/en/saude/noticia/2020-02/brazil-confirms-first-case-coronavirus>>. (Accessed on 08/02/2021). Citado na pág. 22.
- AJZEN, I. From intentions to actions: A theory of planned behavior. In: _____. *Action Control*. Springer Berlin Heidelberg, 1985. p. 11–39. ISBN 9783642697487 9783642697463. Disponível em: <http://link.springer.com/10.1007/978-3-642-69746-3_2>. Citado na pág. 163.
- AJZEN, I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, v. 50, n. 2, p. 179–211, 1991. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/074959789190020T>>. Citado na pág. 35, 163.
- AL-DMOUR, H.; MASA'DEH, R.; SALMAN, A.; ABUHASHESH, M.; AL-DMOUR, R. Influence of social media platforms on public health protection against the covid-19 pandemic via the mediating effects of public health awareness and behavioral changes: Integrated model. *Journal of Medical Internet Research*, v. 22, n. 8, p. e19996, Aug 2020. Disponível em: <<http://www.jmir.org/2020/8/e19996/>>. Citado na pág. 29, 31, 32, 38, 39, 47.
- ALFONZO, M. A. To walk or not to walk? the hierarchy of walking needs. *Environment and Behavior*, v. 37, n. 6, p. 808–836, 2005. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0013916504274016>>. Citado na pág. 163.
- ANAND, M.; SAHAY, K. B.; AHMED, M. A.; SULTAN, D.; CHANDAN, R. R.; SINGH, B. Deep learning and natural language processing in computation for offensive language detection in online social networks by feature selection and ensemble classification techniques. *Theoretical Computer Science*, v. 943, p. 203–218, 2023. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0304397522003887>>. Citado na pág. 26.
- ANTUNES, M. N.; SILVA, C. H. d.; GUIMARÃES, M. C. S.; RABAÇO, M. H. L. Monitoramento de informação em mídias sociais: o e-monitor dengue. *Transinformação*, v. 26, n. 1, p. 9–18, 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862014000100002&lng=pt&tlng=pt>. Citado na pág. 24, 31, 40.
- ANVISA. *Orientações sobre medidas de prevenção e controle de influenza nos serviços de saúde*. 2020. <https://www.saude.sc.gov.br/coronavirus/arquivos/NOTA_TECNICA_GVIMS_GGTES_ANVISA_04_2020_Reviso_27.10.2020.pdf>. (Accessed on 08/13/2021). Citado na pág. 53.
- AREA, I.; BATARFI, H.; LOSADA, J.; NIETO, J. J.; SHAMMAKH, W.; TORRES, A. On a fractional order ebola epidemic model. *Advances in Difference Equations*, v. 2015, n. 1, p. 278, 2015. Disponível em: <<http://www.advancesindifferenceequations.com/content/2015/1/278>>.

- BAGOZZI, R. P. The poverty of economic explanations of consumption and an action theory alternative. *Managerial and Decision Economics*, v. 21, n. 3–4, p. 95–109, 2000. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/mde.975>>. Citado na pág. 160.
- BAGOZZI, R. P.; WARSHAW, P. R. An examination of the etiology of the attitude-behavior relation for goal-directed behaviors. *Multivariate Behavioral Research*, v. 27, n. 4, p. 601–634, 1992. Disponível em: <http://www.tandfonline.com/doi/abs/10.1207/s15327906mbr2704_6>. Citado na pág. 161.
- BAILEY, N. T. et al. *The Mathematical Theory of Infectious Diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1457188/>>. Citado na pág. 45.
- BALLI, C.; GUZEL, M. S.; BOSTANCI, E.; MISHRA, A. Sentimental analysis of twitter users from turkish content with natural language processing. v. 2022, p. 1–17, Apr 2022. Disponível em: <<https://www.hindawi.com/journals/cin/2022/2455160/>>. Citado na pág. 65.
- BANDURA, A. Self-efficacy: Toward a unifying theory of behavioral change. *Advances in Behaviour Research and Therapy*, v. 1, n. 4, p. 139–161, 1978. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0146640278900024>>. Citado na pág. 162.
- BANDURA, A. *Social foundations of thought and action: a social cognitive theory*. [S.l.]: Prentice-Hall, 1986. (Prentice-Hall series in social learning theory). ISBN 9780138156145. Citado na pág. 162.
- BEARD, L.; WILSON, K.; MORRA, D.; KEELAN, J. A survey of health-related activities on second life. *Journal of Medical Internet Research*, v. 11, n. 2, p. e17, May 2009. Disponível em: <<http://www.jmir.org/2009/2/e17/>>. Citado na pág. 42.
- BECKER, G. S.; MURPHY, K. M. A theory of rational addiction. *Journal of Political Economy*, v. 96, n. 4, p. 675–700, 1988. Disponível em: <<https://www.journals.uchicago.edu/doi/10.1086/261558>>. Citado na pág. 162.
- BEKKERS, V.; EDWARDS, A.; KOOL, D. de. Social media monitoring: Responsive governance in the shadow of surveillance? *Government Information Quarterly*, v. 30, n. 4, p. 335–342, 2013. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0740624X13000877>>. Citado na pág. 25.
- BELLG, A. J. Maintenance of health behavior change in preventive cardiology: Internalization and self-regulation of new behaviors. *Behavior Modification*, v. 27, n. 1, p. 103–131, 2003. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0145445502238696>>. Citado na pág. 161.
- BENCKE, L.; CECHINEL, C.; MUNOZ, R. Automated classification of social network messages into smart cities dimensions. *Future Generation Computer Systems*, v. 109, p. 218–237, 2020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0167739X19303000>>. Citado na pág. 65.
- BERRY, D. *Health communication theory and practice*. [S.l.]: Open University Press, 2007. Citado na pág. 30, 36, 85.
- BERRY, I.; O'NEILL, M.; STURROCK, S. L.; WRIGHT, J. E.; ACHARYA, K.; BRANKSTON, G.; HARISH, V.; KORNAS, K.; MAANI, N.; NAGANATHAN, T.;

OBRESS, L.; ROSSI, T.; SIMMONS, A. E.; CAMP, M. V.; XIE, X.; TUIITE, A. R.; GREER, A. L.; FISMAN, D. N.; SOUCY, J.-P. R. A sub-national real-time epidemiological and vaccination database for the COVID-19 pandemic in Canada. *Scientific Data*, Springer Science and Business Media LLC, v. 8, n. 1, jul. 2021. Disponível em: <<https://doi.org/10.1038/s41597-021-00955-2>>. Citado na pág. 55.

BHATIA, S.; PARAG, K. V.; WARDLE, J.; IMAI, N.; ELSLAND, S. L. V.; LASSMANN, B.; CUOMO-DANNENBURG, G.; JAUNEIKAITE, E.; UNWIN, H. J. T.; RILEY, S.; FERGUSON, N.; DONNELLY, C. A.; CORI, A.; NOUVELLET, P. Global predictions of short- to medium-term COVID-19 transmission trends: a retrospective assessment. medRxiv, Jul 2021. Disponível em: <<https://www.medrxiv.org/content/10.1101/2021.07.19.21260746v1>>.

Citado na pág. 85.

BIBLIOGRAPHY, A.; BERKOWITZ, A. The social norms approach. 01 2004. Citado na pág. 163.

BILANDI, N.; VERMA, H. K.; DHIR, R. An intelligent and energy-efficient wireless body area network to control coronavirus outbreak. *Arabian Journal for Science and Engineering*, v. 46, n. 9, p. 8203–8222, 2021. Disponível em: <<https://link.springer.com/10.1007/s13369-021-05411-2>>. Citado na pág. 65.

BORLAND, R. *Understanding hard to maintain behaviour change: a dual process approach*. [S.l.]: Wiley-Blackwell, Addiction Press, 2014. ISBN 9781118572917 9781118572924. Citado na pág. 160.

BORLAND, R. Ceos theory: A comprehensive approach to understanding hard to maintain behaviour change. *Applied Psychology: Health and Well-Being*, v. 9, n. 1, p. 3–35, 2017. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/aphw.12083>>. Citado na pág. 160.

BOSSLET, G. T.; TORKE, A. M.; HICKMAN, S. E.; TERRY, C. L.; HELFT, P. R. The patient–doctor relationship and online social networks: Results of a national survey. *Journal of General Internal Medicine*, v. 26, n. 10, p. 1168–1174, 2011. Disponível em: <<http://link.springer.com/10.1007/s11606-011-1761-2>>. Citado na pág. 42.

BOULOS, M. N. K.; WHEELER, S. The emerging web 2.0 social software: an enabling suite of sociable technologies in health and health care education1. *Health Information & Libraries Journal*, v. 24, n. 1, p. 2–23, 2007. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-1842.2007.00701.x>>. Citado na pág. 30.

BRANDÃO, A. Mediatization beyond western democracies: a three-dimensional proposal to measure the influence of the media in Brazil. *Cuadernos.info*, v. 34, p. 153–164, 2014. Disponível em: <<http://ojs.uc.cl/index.php/cdi/article/view/22791>>. Citado na pág. 31.

BRASIL. *LEI No. 6.259, DE 30 DE OUTUBRO DE 1975*. 1975. <https://www.planalto.gov.br/ccivil_03/leis/l6259.htm>. Accessed: 24/06/2023. Citado na pág. 23.

BRASIL. *Constituição da República Federativa do Brasil*. 2016. <https://www2.senado.leg.br/bdsf/bitstream/handle/id/518231/CF88_Livro_EC91_2016.pdf>. (Accessed on 06/11/2022).

Citado na pág. 23.

BRASIL. *Resolução Nº588, DE 12 DE JULHO DE 2018*. 2018. <<http://conselho.saude.gov.br/resolucoes/2018/Reso588.pdf>>. (Accessed on 06/11/2022). Citado na pág. 23.

BRASIL. *Lei Geral de Proteção de Dados Pessoais (LGPD)*. 2019. <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm>. (Accessed on 05/08/2022). Citado na pág. 61.

BRONIATOWSKI, D. A.; JAMISON, A. M.; QI, S.; ALKULAIB, L.; CHEN, T.; BENTON, A.; QUINN, S. C.; DREDZE, M. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American Journal of Public Health*, v. 108, n. 10, p. 1378–1384, 2018. Disponível em: <<https://ajph.aphapublications.org/doi/10.2105/AJPH.2018.304567>>.

Citado na pág. 29.

BRONIATOWSKI, D. A.; PAUL, M. J.; DREDZE, M. National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic. *PLoS ONE*, v. 8, n. 12, p. e83672, Dec 2013. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0083672>>. Citado na

pág. 24.

BURNET, D.; PLAUT, A.; COURTNEY, R.; CHIN, M. H. A practical model for preventing type 2 diabetes in minority youth. *The Diabetes Educator*, v. 28, n. 5, p. 779–795, 2002. Disponível em: <<http://journals.sagepub.com/doi/10.1177/014572170202800519>>. Citado na pág. 160.

CAPEL. *Programa de Suporte à Pós-Graduação de Instituições de Ensino Particulares (PROSUP) — Português (Brasil)*. 2021. <<https://www.gov.br/capes/pt-br/acao-a-informacao/acoes-e-programas/bolsas/bolsas-no-pais/programa-de-suporte-a-pos-graduacao-de-instituicoes-de-ensino-particulares-prosup>>. (Accessed on 07/26/2021). Citado na pág. 48.

CARPENTER, B.; GELMAN, A.; HOFFMAN, M. D.; LEE, D.; GOODRICH, B.; BETANCOURT, M.; BRUBAKER, M.; GUO, J.; LI, P.; RIDDELL, A. Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, American Statistical Association, v. 76, n. 1, 2017. ISSN 1548-7660. Citado na pág. 83, 85.

CARVER, C. S.; SCHEIER, M. F. Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, v. 92, n. 1, p. 111–135, 1982. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.92.1.111>>. Citado na pág. 160.

CATANIA, J. A.; KEGELES, S. M.; COATES, T. J. Towards an understanding of risk behavior: An aids risk reduction model (arm). *Health Education Quarterly*, v. 17, n. 1, p. 53–72, 1990. Disponível em: <<http://journals.sagepub.com/doi/10.1177/109019819001700107>>. Citado na pág. 160.

CHANG, M.-L. E.; CHUA, H. N. Sql and nosql database comparison: From performance perspective in supporting semi-structured data. In: _____. *Advances in Information and Communication Networks*. Springer International Publishing, 2019. v. 886, p. 294–310. ISBN 9783030034016 9783030034023. Disponível em: <http://link.springer.com/10.1007/978-3-030-03402-3_20>. Citado na pág. 51.

CHATZILENA, A.; van Leeuwen, E.; RATMANN, O.; BAGUELIN, M.; DEMIRIS, N. Contemporary statistical inference for infectious disease models using stan. *Epidemics*, v. 29, p. 100367, 2019. ISSN 1755-4365. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1755436519300325>>. Citado na pág. 104.

CHICCO, D.; JURMAN, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, v. 21, n. 1, p. 6, 2020. Disponível em: <<https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7>>. Citado na pág. 95, 96.

CHIPIDZA, W.; AKBARIPOURDIBAZAR, E.; GWANZURA, T.; GATTO, N. M. Topic analysis of traditional and social media news coverage of the early covid-19 pandemic and implications for public health communication. *Disaster Medicine and Public Health Preparedness*, p. 1–8, Mar 2021. Disponível em: <https://www.cambridge.org/core/product/identifier/S1935789321000653/type/journal_article>. Citado na pág. 22, 47.

CHOU, W.-y. S.; HUNT, Y. M.; BECKJORD, E. B.; MOSER, R. P.; HESSE, B. W. Social media use in the united states: Implications for health communication. *Journal of Medical Internet Research*, v. 11, n. 4, p. e48, Nov 2009. Disponível em: <<http://www.jmir.org/2009/4/e48/>>.

Citado na pág. 24, 29, 41.

CIALDINI, R. B. The focus theory of normative conduct. In: _____. *Handbook of Theories of Social Psychology: Volume 2*. SAGE Publications Ltd, 2012. p. 295–312. Disponível em: <https://sk.sagepub.com/reference/hdbk_socialpsychtheories2/n41.xml>. Citado na pág. 160.

CIALDINI, R. B.; KALLGREN, C. A.; RENO, R. R. A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In: _____. *Advances in Experimental Social Psychology*. Elsevier, 1991. v. 24, p. 201–234. ISBN 9780120152247. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0065260108603305>>. Citado na pág. 160.

CODATMO. *Covid Data Model*. 2021. Disponível em: <<https://codatmo.github.io/>>. Citado na pág. 48.

COFEN. *Brasil perde ao menos um profissional de saúde a cada 19 horas para a Covid Conselho Federal de Enfermagem - Brasil*. 2021. <http://www.cofen.gov.br/brasil-perde-ao-menos-um-profissional-de-saude-a-cada-19-horas-para-a-covid_85778.html>. (Accessed on 10/16/2022). Citado na pág. 45.

COLINEAU, N.; PARIS, C. Talking about your health to strangers: understanding the use of online social networks by patients. *New Review of Hypermedia and Multimedia*, v. 16, n. 1–2, p. 141–160, 2010. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/13614568.2010.496131>>. Citado na pág. 42.

DAHER, V. *O que o mundo precisa aprender com a pandemia*. 2020. <<https://cienciahoje.org.br/artigo/o-que-o-mundo-precisa-aprender-com-a-pandemia/>>. Accessed: 24/06/2023. Citado na pág. 27.

DANIEL, E. E.; MASILAMANI, R.; RAHMAN, M. The effect of community-based reproductive health communication interventions on contraceptive use among young married couples in bihar, india. *International Family Planning Perspectives*, v. 34, n. 04, p. 189–197, 2008. Disponível em: <<http://guttmacher.org/pubs/journals/3418908.pdf>>. Citado na pág. 29.

DATASUS. *Painel Coronavírus*. 2021. Disponível em: <<https://covid.saude.gov.br/>>. Citado na pág. 22.

DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, v. 13, n. 3, p. 319, 1989. Disponível em: <<https://www.jstor.org/stable/249008?origin=crossref>>. Citado na pág. 163.

DAWSON, M. R. W. *Minds and Machines Connectionism and Psychological Modeling*. [s.n.], 2008. ISBN 9780470752982. Disponível em: <<https://nbn-resolving.org/urn:nbn:de:101:1-201412163737>>. Citado na pág. 74.

- DECI, E. L.; RYAN, R. M. *Intrinsic motivation and self-determination in human behavior*. [S.l.]: Plenum, 1985. (Perspectives in social psychology). ISBN 9780306420221. Citado na pág. 162.
- DECI, E. L.; RYAN, R. M. Facilitating optimal motivation and psychological well-being across life's domains. *Canadian Psychology/Psychologie canadienne*, v. 49, n. 1, p. 14–23, 2008. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0708-5591.49.1.14>>. Citado na pág. 162.
- DHOLAKIA, U. M.; BAGOZZI, R. P.; PEARO, L. K. A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing*, v. 21, n. 3, p. 241–263, 2004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S016781160400028X>>. Citado na pág. 163.
- DORMAND, J. R.; PRINCE, P. J. A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics*, v. 6, n. 1, p. 19–26, mar. 1980. ISSN 0377-0427. Citado na pág. 83.
- DOWDELL, E. B.; BURGESS, A. W.; FLORES, J. R. Original research: Online social networking patterns among adolescents, young adults, and sexual offenders. *AJN, American Journal of Nursing*, v. 111, n. 7, p. 28–36, 2011. Disponível em: <<https://journals.lww.com/00000446-201107000-00021>>. Citado na pág. 42.
- ENDOMBA, F. T.; BIGNA, J. J.; NOUBIAP, J. J. The impact of social networking services on the coronavirus disease 2019 (covid-19) pandemic in sub-saharan africa. *The Pan African Medical Journal*, v. 35, n. Supp 2, Jun 2020. Disponível em: <<https://www.panafrican-med-journal.com/content/series/35/2/67/full>>. Citado na pág. 31.
- ES, G. *COVID-19 - Informativos Coronavírus*. 2020. <<https://coronavirus.es.gov.br/informativos-coronavirus>>. (Accessed on 11/12/2022). Citado na pág. 100.
- EWART, C. K. Social action theory for a public health psychology. *American Psychologist*, v. 46, n. 9, p. 931–946, 1991. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.46.9.931>>. Citado na pág. 162.
- FAN, W.; GORDON, M. D. The power of social media analytics. *Communications of the ACM*, v. 57, n. 6, p. 74–81, 2014. Disponível em: <<https://dl.acm.org/doi/10.1145/2602574>>. Citado na pág. 24, 40.
- FISHBEIN, M. The role of theory in hiv prevention. *AIDS Care*, v. 12, n. 3, p. 273–278, 2000. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/09540120050042918>>. Citado na pág. 161.
- FISHER, J. D.; FISHER, W. A. Changing aids-risk behavior. *Psychological Bulletin*, v. 111, n. 3, p. 455–474, 1992. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.111.3.455>>. Citado na pág. 161.
- FISHER, J. D.; FISHER, W. A.; AMICO, K. R.; HARMAN, J. J. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. *Health Psychology*, v. 25, n. 4, p. 462–473, 2006. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-6133.25.4.462>>. Citado na pág. 161.
- FLAY, B. *On improving the chances of mass media health promotion programs causing meaningful changes in behaviour*. [S.l.]: Health Education by Television and Radio, 1981. Citado na pág. 161.

FLAY, B.; D'AVERNAS, J.; BEST, A.; KERSELL, M.; RYAN, K. Cigarette smoking: Why young people do it and ways of preventing it. In: _____. [S.l.: s.n.], 1983. p. 132–183. Citado na pág. 161.

FLAY, B.; PETRAITIS, J. The theory of triadic influence: A new theory of health behavior with implications for preventive interventions. *Advances in Medical Sociology*, v. 4, p. 19–44, 01 1994. Disponível em: <https://www.researchgate.net/publication/224942196_The_Theory_of_Triadic_Influence_A_New_Theory_of_Health_Behavior_With_Implications_for_Preventive_Interventions>. Citado na pág. 163.

FLAY, B. R.; DITECCO, D.; SCHLEGEL, R. P. Mass media in health promotion: An analysis using an extended information-processing model. *Health Education Quarterly*, v. 7, n. 2, p. 127–147, 1980. Disponível em: <<http://journals.sagepub.com/doi/10.1177/109019818000700203>>. Citado na pág. 160.

FLEISHER, L.; BUZAGLO, J.; COLLINS, M.; MILLARD, J.; MILLER, S. M.; EGLESTON, B. L.; SOLARINO, N.; TRINASTIC, J.; CEGALA, D. J.; BENSON, A. B.; AL. et. Using health communication best practices to develop a web-based provider–patient communication aid: The connectm study. *Patient Education and Counseling*, v. 71, n. 3, p. 378–387, 2008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0738399108001274>>. Citado na pág. 29.

FUTURE AGENDA. *Global Pandemics - Future Agenda*. 2021. Disponível em: <<https://www.futureagenda.org/foresights/global-pandemics/>>. Citado na pág. 22, 47.

GABRY, J.; ČEŠNOVAR, R. *cmdstanr: R Interface to 'CmdStan'*. [S.l.], 2021. <https://mc-stan.org/cmdstanr>, <https://discourse.mc-stan.org>. Citado na pág. 85, 157.

GAJARIA, A.; YEUNG, E.; GOODALE, T.; CHARACH, A. Beliefs about attention-deficit/hyperactivity disorder and response to stereotypes: Youth postings in facebook groups. *Journal of Adolescent Health*, v. 49, n. 1, p. 15–20, 2011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1054139X10004325>>. Citado na pág. 42.

GELMAN, A.; VEHTARI, A.; SIMPSON, D.; MARGOSSIAN, C. C.; CARPENTER, B.; YAO, Y.; KENNEDY, L.; GABRY, J.; BÜRKNER, P.-C.; MODRÁK, M. Bayesian Workflow. *arXiv:2011.01808 [stat]*, nov. 2020. Citado na pág. 83.

GENDRON, Y. Rethinking the concept of research contribution. *Comptabilité Contrôle Audit*, Tome 19, n. 2, p. 133–155, Jul 2013. Disponível em: <<https://www.cairn.info/revue-comptabilite-contrrole-audit-2013-2-page-133.htm?ref=doi>>. Citado na pág. 48.

GERRARD, M.; GIBBONS, F. X.; HOULIHAN, A. E.; STOCK, M. L.; POMERY, E. A. A dual-process approach to health risk decision making: The prototype willingness model. *Developmental Review*, v. 28, n. 1, p. 29–61, 2008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0273229707000512>>. Citado na pág. 162.

GOLDENBERG, J. L.; ARNDT, J. The implications of death for health: A terror management health model for behavioral health promotion. *Psychological Review*, v. 115, n. 4, p. 1032–1053, 2008. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/a0013326>>.

Citado na pág. 163.

GONZALEZ, G. An integrated theoretical model for alcohol and other drug abuse prevention on the college campus. v. 30, n. 6, p. 492–503, 1989. Disponível em: <<https://psycnet.apa.org/record/1990-10896-001>>. Citado na pág. 161.

GOSTIN, L. O.; WILEY, L. F. . *Public health law and ethics: a reader*. Third edition. Oakland, California: University of California Press, 2018. ISBN 9780520967731. Citado na pág. 23.

GOTTFREDSON, M. R.; HIRSCHI, T. *A general theory of crime*. [S.l.]: Stanford University Press, 1990. ISBN 9780804717731 9780804717748. Citado na pág. 160.

GOUNANE, S.; BARKOUCH, Y.; ATLAS, A.; BENDAHMANE, M.; KARAMI, F.; MESKINE, D. An adaptive social distancing sir model for covid-19 disease spreading and forecasting. *Epidemiologic Methods*, v. 10, n. s1, p. 20200044, Mar 2021. Disponível em: <<https://www.degruyter.com/document/doi/10.1515/em-2020-0044/html>>. Citado na pág. 43.

GREENBERG, J.; PYSZCZYNSKI, T.; SOLOMON, S. The causes and consequences of a need for self-esteem: A terror management theory. In: _____. *Public Self and Private Self*. Springer New York, 1986. p. 189–212. ISBN 9781461395669 9781461395645. Disponível em: <http://link.springer.com/10.1007/978-1-4613-9564-5_10>. Citado na pág. 163.

GREENHOW, C.; ROBELIA, B.; HUGHES, J. E. Learning, teaching, and scholarship in a digital age: Web 2.0 and classroom research: What path should we take now? *Educational Researcher*, v. 38, n. 4, p. 246–259, 2009. Disponível em: <<http://journals.sagepub.com/doi/10.3102/0013189X09336671>>. Citado na pág. 30.

GRINSZTAJN, L.; SEMENOVA, E.; MARGOSSIAN, C. C.; RIOU, J. Bayesian workflow for disease transmission modeling in Stan. *arXiv:2006.02985 [q-bio, stat]*, fev. 2021. Citado na pág. 83.

GULJAŠ, S.; BOSNIĆ, Z.; SALHA, T.; BERECKI, M.; DUPAN, Z. K.; RUDAN, S.; TRTICA, L. M. Lack of informations about covid-19 vaccine: From implications to intervention for supporting public health communications in covid-19 pandemic. *International Journal of Environmental Research and Public Health*, v. 18, n. 11, p. 6141, Jun 2021. Disponível em: <<https://www.mdpi.com/1660-4601/18/11/6141>>. Citado na pág. 23, 35, 38.

GÓMEZ-CARRASCO, P.; GUILLAMÓN-SAORÍN, E.; OSMA, B. G. Stakeholders versus firm communication in social media: The case of twitter and corporate social responsibility information. *European Accounting Review*, v. 30, n. 1, p. 31–62, Jan 2021. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/09638180.2019.1708428>>. Citado na pág. 26, 27, 40.

HADI, T. A.; FLESHLER, K. Integrating social media monitoring into public health emergency response operations. *Disaster Medicine and Public Health Preparedness*, v. 10, n. 5, p. 775–780, 2016. Disponível em: <https://www.cambridge.org/core/product/identifier/S1935789316000392/type/journal_article>. Citado na pág. 24, 25, 31.

HAGGER, M. S.; CHATZISARANTIS, N. L. D. Transferring motivation from educational to extramural contexts: a review of the trans-contextual model. *European Journal of Psychology of Education*, v. 27, n. 2, p. 195–212, 2012. Disponível em: <<http://link.springer.com/10.1007/s10212-011-0082-5>>. Citado na pág. 163.

HAGGER, M. S.; CHATZISARANTIS, N. L. D.; CULVERHOUSE, T.; BIDDLE, S. J. H. The processes by which perceived autonomy support in physical education promotes leisure-time physical activity intentions and behavior: A trans-contextual model. *Journal of Educational Psychology*, v. 95, n. 4, p. 784–795, 2003. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0663.95.4.784>>. Citado na pág. 163.

HALL, P. A.; FONG, G. T. Temporal self-regulation theory: A model for individual health behavior. *Health Psychology Review*, v. 1, n. 1, p. 6–52, 2007. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/17437190701492437>>. Citado na pág. 163.

HAMDI, A.; SHABAN, K.; ZAINAL, A. Clasenti: A class-specific sentiment analysis framework. *ACM Transactions on Asian and Low-Resource Language Information Processing*, v. 17, n. 4, p. 1–28, Dec 2018. Disponível em: <<https://dl.acm.org/doi/10.1145/3209885>>. Citado na pág. 66.

HAWKINS, J. D.; WEIS, J. G. The social development model: An integrated approach to delinquency prevention. *The Journal of Primary Prevention*, v. 6, n. 2, p. 73–97, 1985. Disponível em: <<http://link.springer.com/10.1007/BF01325432>>. Citado na pág. 163.

HAZEL, A.; DAVIDSON, M. C.; ROGERS, A.; BARRIE, M. B.; FREEMAN, A.; MBAYOH, M.; KAMARA, M.; BLUMBERG, S.; LIETMAN, T. M.; RUTHERFORD, G. W.; JONES, J. H.; PORCO, T. C.; RICHARDSON, E. T.; KELLY, J. D. Social network analysis of ebola virus disease during the 2014 outbreak in sukudu, sierra leone. *Open Forum Infectious Diseases*, v. 9, n. 11, p. ofac593, Nov 2022. Disponível em: <<https://academic.oup.com/ofid/article/doi/10.1093/ofid/ofac593/6795257>>. Citado na pág. 31.

HERNANDEZ-CERON, N.; FENG, Z.; CASTILLO-CHAVEZ, C. Discrete epidemic models with arbitrary stage distributions and applications to disease control. *Bulletin of Mathematical Biology*, v. 75, n. 10, p. 1716–1746, 2013. Disponível em: <<http://link.springer.com/10.1007/s11538-013-9866-x>>. Citado na pág. 43.

HIGGINS, E. T. Making a good decision: Value from fit. *American Psychologist*, v. 55, n. 11, p. 1217–1230, 2000. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.55.11.1217>>. Citado na pág. 162.

HIGGINS, E. T. Value from regulatory fit. *Current Directions in Psychological Science*, v. 14, n. 4, p. 209–213, 2005. Disponível em: <<http://journals.sagepub.com/doi/10.1111/j.0963-7214.2005.00366.x>>. Citado na pág. 162.

HOFFMAN, M. D.; GELMAN, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, v. 15, n. 1, p. 1593–1623, nov. 2011. Citado na pág. 84.

HOOTSUITE. *Digital 2021: Brazil*. 2021. <<https://datareportal.com/reports/digital-2021-brazil>>. (Accessed on 09/08/2021). Citado na pág. 31, 85.

HOVELL, M. F.; HILLMAN, E. R.; BLUMBERG, E.; SIPAN, C.; ATKINS, C.; HOFSTETTER, C. R.; MYERS, C. A. A behavioral-ecological model of adolescent sexual development: A template for aids prevention. *Journal of Sex Research*, v. 31, n. 4, p. 267–281, 1994. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/00224499409551762>>.

Citado na pág. 160.

HU, Y.; SUNDAR, S. S. Effects of online health sources on credibility and behavioral intentions. *Communication Research*, v. 37, n. 1, p. 105–132, 2010. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0093650209351512>>. Citado na pág. 41, 42.

HUANG, X.; LI, Z.; JIANG, Y.; LI, X.; PORTER, D. *Twitter, human mobility, and COVID-19*. 2020. Citado na pág. 82.

IFBA. *Campanha Fica em Casa Rede Federal — IFBA - Instituto Federal de Educação, Ciência e Tecnologia da Bahia Instituto Federal da Bahia*. 2020. <<https://portal.ifba.edu.br/barreiras/noticias-2/campanha-fica-em-casa-rede-federal>>. (Accessed on 11/12/2022). Citado na pág. 100.

INÁCIO, A.; SAVARIS, A.; ALVES, J.; WANGENHEIM, A. von. Sistema integrado de análise e visualização epidemiológica georreferenciada online para telemedicina. In: *Anais do XVI Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2016)*. Sociedade Brasileira de Computação - SBC, 2016. p. 2547–2556. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/9901>>. Citado na pág. 51.

ISERLES, A. *A First Course in the Numerical Analysis of Differential Equations*. Second. USA: Cambridge University Press, 2008. ISBN 0-521-73490-8. Citado na pág. 83.

JAJOSKY, R. A.; GROSECLOSE, S. L. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health*, v. 4, n. 1, p. 29, Jul 2004. Disponível em: <<https://doi.org/10.1186/1471-2458-4-29>>. Citado na pág. 23.

JESSOR, R. Problem-behavior theory, psychosocial development, and adolescent problem drinking. *Addiction*, v. 82, n. 4, p. 331–342, 1987. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1360-0443.1987.tb01490.x>>. Citado na pág. 162.

JESSOR, R. Risk behavior in adolescence: A psychosocial framework for understanding and action. *Journal of Adolescent Health*, v. 12, n. 8, p. 597–605, 1991. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/1054139X9190007K>>. Citado na pág. 162.

JOHNSON, N. P. A. S.; MUELLER, J. Updating the accounts: Global mortality of the 1918-1920 "spanish"; influenza pandemic. *Bulletin of the History of Medicine*, v. 76, n. 1, p. 105–115, 2002. Disponível em: <http://muse.jhu.edu/content/crossref/journals/bulletin_of_the_history_of_medicine/v076/76.1johnson.html>. Citado na pág. 22.

KAHNEMAN, D.; TVERSKY, A. Prospect theory: An analysis of decision under risk. *Econometrica*, v. 47, n. 2, p. 263, 1979. Disponível em: <<https://www.jstor.org/stable/1914185?origin=crossref>>. Citado na pág. 162.

KANFER, F. H.; GAELICK, L. Self-management methods. In: _____. *Helping people change: a textbook of methods*. [S.l.]: Pergamon Press, 1991. (Pergamon general psychology series, 52), p. 305–360. ISBN 9780080182728 9780080182711. Citado na pág. 162.

KAPLAN, A. M.; HAENLEIN, M. Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, v. 53, n. 1, p. 59–68, 2010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0007681309001232>>. Citado na pág. 25.

KAPLAN, H. B. Toward a general theory of psychosocial deviance: The case of aggressive behavior. *Social Science & Medicine (1967)*, v. 6, n. 5, p. 593–617, 1972. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0037785672900753>>. Citado na pág. 161.

KAPLAN, H. B. Self-attitudes and deviant behavior: New directions for theory and research. *Youth & Society*, v. 14, n. 2, p. 185–211, 1982. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0044118X82014002004>>. Citado na pág. 161.

- KAPLAN, H. B.; MARTIN, S. S.; ROBBINS, C. Application of a general theory of deviant behavior: Self-derogation and adolescent drug use. *Journal of Health and Social Behavior*, v. 23, n. 4, p. 274, 1982. Disponível em: <<http://www.jstor.org/stable/2136487?origin=crossref>>. Citado na pág. 161.
- KATZ, D. L. Behavior modification in primary care: The pressure system model. *Preventive Medicine*, v. 32, n. 1, p. 66–72, 2001. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0091743500907734>>. Citado na pág. 162.
- KEELING, M. J.; EAMES, K. T. Networks and epidemic models. *Journal of The Royal Society Interface*, v. 2, n. 4, p. 295–307, Sep 2005. Disponível em: <<https://royalsocietypublishing.org/doi/10.1098/rsif.2005.0051>>. Citado na pág. 44.
- KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, The Royal Society London, v. 115, n. 772, p. 700–721, 1927. Citado na pág. 44.
- KERSELL, M. W.; MILSUM, J. H. A systems model of health behavior change. *Behavioral Science*, v. 30, n. 3, p. 119–126, 1985. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/bs.3830300302>>. Citado na pág. 163.
- KICKBUSCH, I.; PELIKAN, J. M.; APFEL, F.; TSOUROS, A. D. *Health literacy: the solid facts*. [S.l.]: World Health Organization Regional Office for Europe, 2013. (The solid facts). ISBN 9789289000154. Citado na pág. 23.
- KIETZMANN, J. H.; HERMKENS, K.; MCCARTHY, I. P.; SILVESTRE, B. S. Social media? get serious! understanding the functional building blocks of social media. *Business Horizons*, v. 54, n. 3, p. 241–251, 2011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0007681311000061>>. Citado na pág. 25.
- KIRSCHBAUM, C. Network analysis: emergence, criticism and recent trends. *RAUSP Management Journal*, v. 54, n. 4, p. 533–547, Oct 2019. Disponível em: <<https://www.emerald.com/insight/content/doi/10.1108/RAUSP-05-2019-0074/full/html>>. Citado na pág. 49.
- KLUGER, A. N.; DENISI, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. p. 254–284, 1996. Citado na pág. 160.
- KOLLMUSS, A.; AGYEMAN, J. Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, v. 8, n. 3, p. 239–260, 2002. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/13504620220145401>>. Citado na pág. 161.
- KONTOS, E. Z.; EMMONS, K. M.; PULEO, E.; VISWANATH, K. Communication inequalities and public health implications of adult social networking site use in the united states. *Journal of Health Communication*, v. 15, n. sup3, p. 216–235, Dec 2010. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/10810730.2010.522689>>. Citado na pág. 41.
- KRAHN, G. L. A call for better data on prevalence and health surveillance of people with intellectual and developmental disabilities. *Intellectual and Developmental Disabilities*, v. 57, n. 5, p. 357–375, Oct 2019. Disponível em: <<https://meridian.allenpress.com/idd/article/57/5/357/364934/A-Call-for-Better-Data-on-Prevalence-and-Health>>. Citado na pág. 23.

LABCIDADE. *COVID-Classifier: Classifier for Brazilian Twitter Data*. 2021. <<https://github.com/LabCidades/COVID-Classifier>>. (Accessed on 11/09/2021). Citado na pág. 54, 87.

LabCidades. *Epi-Subcompartimentos Repository*. 2021. <<https://github.com/LabCidades/Epi-Subcompartimentos>>. Acessado em 15 de Maio de 2023. Citado na pág. 87.

LABCIDADES. *Twitter-Model*. 2023. GitHub repository. Disponível em: <<https://github.com/LabCidades/Twitter-Model/tree/main/data>>. Citado na pág. 156.

LABCIDADES. *Twitter-Model*. 2023. GitHub repository. Disponível em: <<https://github.com/LabCidades/Twitter-Model>>. Citado na pág. 157.

LAGU, T.; HANNON, N. S.; ROTHBERG, M. B.; LINDENAUER, P. K. Patients' evaluations of health care providers in the era of social networking: An analysis of physician-rating websites. *Journal of General Internal Medicine*, v. 25, n. 9, p. 942–946, 2010. Disponível em: <<http://link.springer.com/10.1007/s11606-010-1383-0>>. Citado na pág. 42.

LARANJO, L. Social media and health behavior change. In: _____. *Participatory Health Through Social Media*. Elsevier, 2016. p. 83–111. ISBN 9780128092699. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/B9780128092699000062>>. Citado na pág. 34, 35, 36, 38.

LARANJO, L.; ARGUEL, A.; NEVES, A. L.; GALLAGHER, A. M.; KAPLAN, R.; MORTIMER, N.; MENDES, G. A.; LAU, A. Y. S. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. *Journal of the American Medical Informatics Association*, v. 22, n. 1, p. 243–256, Jan 2015. Disponível em: <<https://academic.oup.com/jamia/article/22/1/243/833940>>. Citado na pág. 31, 35.

LAWRENCE, T.; GATICA, J.; LOYOLA, I.; FICA, A.; LAWRENCE, T.; GATICA, J.; LOYOLA, I.; FICA, A. Impact on social networks after announcement of a case of hansen's disease in an immigrant in chile. *Revista chilena de infectología*, v. 35, n. 6, p. 689–694, 2018. Disponível em: <http://www.scielo.cl/scielo.php?script=sci_abstract&pid=S0716-10182018000600689&lng=en&nrm=iso&tlng=en>. Citado na pág. 31.

LEMAÎTRE, G.; NOGUEIRA, F.; ARIDAS, C. K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, v. 18, n. 17, p. 1–5, 2017. Disponível em: <<http://jmlr.org/papers/v18/16-365.html>>.

Citado na pág. 68.

LEWIN, K. Frontiers in group dynamics: Concept, method and reality in social science; social equilibria and social change. *Human Relations*, v. 1, n. 1, p. 5–41, 1947. Disponível em: <<http://journals.sagepub.com/doi/10.1177/001872674700100103>>. Citado na pág. 160.

LEWIN, K. *Field theory in social science : selected theoretical papers*. London: Tavistock Publications, 1952. Citado na pág. 160.

LEWIN, K. *Group Decision and Social Change*. In E.E. Maccoby, T.M. Newcomb & E.L. Hartley. (Eds.). [S.l.]: Hartley, Holt, Rinehart and Winston, 1958. 197–211 p. Citado na pág. 160.

LIANG, B.; SCAMMON, D. L. E-word-of-mouth on health social networking sites: An opportunity for tailored health communication: ewom on health social networking sites. *Journal of Consumer Behaviour*, v. 10, n. 6, p. 322–331, 2011. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/cb.378>>. Citado na pág. 41, 42.

- LINDENBERG, S.; STEG, L. Normative, gain and hedonic goal frames guiding environmental behavior. *Journal of Social Issues*, v. 63, n. 1, p. 117–137, 2007. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1540-4560.2007.00499.x>>. Citado na pág. 161.
- LIU, Z.; MAGAL, P.; SEYDI, O.; WEBB, G. A covid-19 epidemic model with latency period. *Infectious Disease Modelling*, v. 5, p. 323–337, Jan 2020. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2468042720300099>>. Citado na pág. 45.
- LIU, Z.; MAGAL, P.; SEYDI, O.; WEBB, G. A covid-19 epidemic model with latency period. *Infectious Disease Modelling*, v. 5, p. 323–337, 2020. ISSN 2468-0427. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2468042720300099>>. Citado na pág. 45.
- LO, A. S.; ESSER, M. J.; GORDON, K. E. Youtube: A gauge of public perception and awareness surrounding epilepsy. *Epilepsy & Behavior*, v. 17, n. 4, p. 541–545, 2010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1525505010002088>>. Citado na pág. 42.
- LOCKE, E. A. Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, v. 3, n. 2, p. 157–189, 1968. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0030507368900044>>. Citado na pág. 161.
- LOCKE, E. A.; LATHAM, G. P. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, v. 57, n. 9, p. 705–717, 2002. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.57.9.705>>. Citado na pág. 161.
- LOEWENSTEIN, G. F.; WEBER, E. U.; HSEE, C. K.; WELCH, N. Risk as feelings. *Psychological Bulletin*, v. 127, n. 2, p. 267–286, 2001. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.127.2.267>>. Citado na pág. 162.
- LOUKIS, E.; CHARALABIDIS, Y.; ANDROUTSOPOULOU, A. Promoting open innovation in the public sector through social media monitoring. *Government Information Quarterly*, v. 34, n. 1, p. 99–109, 2017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0740624X16301770>>. Citado na pág. 25.
- MAES, S.; GEBHARDT, W. Self-regulation and health behavior. In: _____. *Handbook of Self-Regulation*. Elsevier, 2000. p. 343–368. ISBN 9780121098902. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/B9780121098902500408>>. Citado na pág. 161.
- MANRIQUE, A. P. P. *MODELOS EPIDEMIOLÓGICOS EM REDES*. 2016. <[https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/18662/20_02%20\(1\).pdf;jsessionid=DCFFB5CE3646F06462B720CAD74A6315?sequence=1](https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/18662/20_02%20(1).pdf;jsessionid=DCFFB5CE3646F06462B720CAD74A6315?sequence=1)>. Accessed: 24/06/2023. Citado na pág. 27.
- MARLATT, G. A.; GEORGE, W. H. Relapse prevention: Introduction and overview of the model. *Addiction*, v. 79, n. 3, p. 261–273, 1984. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1360-0443.1984.tb00274.x>>. Citado na pág. 162.
- MARQUEZ, J. L. J.; CARRASCO, I. G.; CUADRADO, J. L. L. Challenges and opportunities in analytic-predictive environments of big data and natural language processing for social network rating systems. *IEEE Latin America Transactions*, v. 16, n. 2, p. 592–597, 2018. Disponível em: <<https://ieeexplore.ieee.org/document/8327417/>>. Citado na pág. 26.
- MARTINS, M. H. d. M.; SPINK, M. J. P. Comunicação em saúde nas campanhas de prevenção à leptospirose humana em maceió, alagoas, brasil. *Interface - Comunicação, Saúde, Educação*,

v. 23, p. e180709, 2019. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-32832019000100281&tlng=pt>. Citado na pág. 29.

MELO, T. de; FIGUEIREDO, C. M. S. Comparing news articles and tweets about COVID-19 in Brazil: Sentiment analysis and topic modeling approach. *JMIR Public Health and Surveillance*, JMIR Publications Inc., v. 7, n. 2, p. e24585, fev. 2021. Disponível em: <<https://doi.org/10.2196/24585>>. Citado na pág. 47.

MG, G. *Prefeitura Municipal de Santa Efigênia de Minas*. 2020. <<https://www.santaefigenia.mg.gov.br/noticias/campanha-fica-em-casa>>. (Accessed on 11/12/2022). Citado na pág. 100.

MICHIE, S. *ABC of behaviour change theories: an essential resource for researchers, policy makers and practitioners; 83 theories*. Bream: Silverback Publishing, 2014. ISBN 978-1-291-88667-2. Citado na pág. 92, 160.

MICHIE, S.; STRALEN, M. M. van; WEST, R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, v. 6, n. 1, p. 42, 2011. Disponível em: <<http://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-6-42>>. Citado na pág. 160.

MIELE, A.; WANG, T.; BASAPUR, V. Primal and dual formulations of sequential gradient-restoration algorithms for trajectory optimization problems. *Acta Astronautica*, v. 13, n. 8, p. 491–505, 1986. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0094576586900287>>. Citado na pág. 67.

MIKOLAI, J.; KEENAN, K.; KULU, H. Intersecting household-level health and socio-economic vulnerabilities and the COVID-19 crisis: An analysis from the UK. *SSM - Population Health*, v. 12, p. 100628, 2020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2352827320302652>>. Citado na pág. 92.

MILLER, N. E.; DOLLARD, J. *Social Learning and Imitation*. [S.l.]: Yale U.P, 1962. Citado na pág. 163.

Ministério da Saúde. *2020_03_13_Boletim-Epidemiologico-05.pdf*. 2020. <http://maismedicos.gov.br/images/PDF/2020_03_13_Boletim-Epidemiologico-05.pdf>. (Accessed on 08/10/2021). Citado na pág. 57, 58.

Ministério da Saúde. *Coronavírus: Brasil confirma primeiro caso da doença - Notícia - UNA-SUS*. 2020. <<https://www.unasus.gov.br/noticia/coronavirus-brasil-confirma-primeiro-caso-da-doenca>>. (Accessed on 07/26/2021). Citado na pág. 51.

Ministério da Saúde. *Guia de Vigilância Epidemiológica*. 2020. <https://portalarquivos.saude.gov.br/images/af_gvs_coronavirus_6ago20_ajustes-finais-2.pdf>. (Accessed on 08/10/2021). Citado na pág. 22.

Ministério da Saúde. *Coronavirus Sintomas — Português (Brasil)*. 2021. <<https://www.gov.br/saude/pt-br/coronavirus/sintomas>>. (Accessed on 08/08/2021). Citado na pág. 53.

Ministério da Saúde. *Covid-19 Casos e Óbitos*. 2021. <https://qsprod.saude.gov.br/extensions/covid-19_html/covid-19_html.html>. (Accessed on 07/26/2021). Citado na pág. 51.

Ministério da Saúde. *Plataforma Integrada de Vigilância em Saúde - Secretaria de Vigilância em Saúde*. 2021. <<http://plataforma.saude.gov.br/>>. (Accessed on 08/10/2021). Citado na pág. 56.

Ministério da Saúde. *SRAG 2020 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 - Conjuntos de dados - Open Data*. 2021. <<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>>. (Accessed on 08/07/2021). Citado na pág. 57.

Ministério da Saúde. *SRAG 2021 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 - SRAG 02/08/2021 - Open Data*. 2021. <<https://opendatasus.saude.gov.br/dataset/bd-srag-2021/resource/42bd5e0e-d61a-4359-942e-ebc83391a137>>. (Accessed on 08/10/2021). Citado na pág. 103.

Ministério da Saúde. *SRAG 2021 - Severe Acute Respiratory Syndrome Database - including COVID-19 data - Datasets - Open Data*. 2021. <<https://opendatasus.saude.gov.br/dataset/bd-srag-2021>>. (Accessed on 08/10/2021). Citado na pág. 57, 83.

Ministério da Saúde. *Painel COVID-19*. 2023. <https://infoms.saude.gov.br/extensions/covid-19_html/covid-19_html.html>. Accessed on June 27, 2023. Citado na pág. 51, 52.

Ministério da Saúde, Secretaria de Vigilância em Saúde. *SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19*. 2023. Disponível em: <<https://opendatasus.saude.gov.br/dataset/srag-2021-a-2023>>. Citado na pág. 52, 156.

MOORE, R.; PHILLIPS, A. *Liverpool Covid Model*. [S.l.]: GitHub, 2021. <<https://github.com/codatmo/Liverpool>>. Citado na pág. 82, 84, 103.

MOORE, R. E.; ROSATO, C.; MASKELL, S. Refining epidemiological forecasts with simple scoring rules. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 380, n. 2233, p. 20210305, Oct 2022. Disponível em: <<https://royalsocietypublishing.org/doi/10.1098/rsta.2021.0305>>. Citado na pág. 86.

MOORHEAD, S. A.; HAZLETT, D. E.; HARRISON, L.; CARROLL, J. K.; IRWIN, A.; HOVING, C. A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*, v. 15, n. 4, p. e85, Apr 2013. Disponível em: <<http://www.jmir.org/2013/4/e85/>>. Citado na pág. 24, 31, 32, 41, 42, 43, 85, 104.

MURRAY, C. J.; FRENK, J. Ranking 37th — measuring the performance of the u.s. health care system. *New England Journal of Medicine*, v. 362, n. 2, p. 98–99, Jan 2010. Disponível em: <<http://www.nejm.org/doi/abs/10.1056/NEJMp0910064>>. Citado na pág. 29.

MURUGANANTHAM, A.; GANDHI, G. M. Framework for social media analytics based on multi-criteria decision making (mcdm) model. *Multimedia Tools and Applications*, v. 79, n. 5–6, p. 3913–3927, 2020. Disponível em: <<http://link.springer.com/10.1007/s11042-019-7470-2>>. Citado na pág. 26, 40.

NAGPAL, A. *Métodos de Regularização L1 e L2*. 2017. <<https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c>>. (Accessed on 05/11/2022). Citado na pág. 67.

NEAL, R. M. MCMC using Hamiltonian dynamics. In: BROOKS, S.; GELMAN, A.; JONES, G. L.; MENG, X.-L. (Ed.). *Handbook of Markov Chain Monte Carlo*. [S.l.]: Chapman and Hall/CRC, 2011. Citado na pág. 84.

NETO, O. P.; KENNEDY, D. M.; REIS, J. C.; WANG, Y.; BRIZZI, A. C. B.; ZAMBRANO, G. J.; de Souza, J. M.; PEDROSO, W.; de Mello Pedreiro, R. C.; de Matos Brizzi, B.; ABINADER, E. O.; ZÂNGARO, R. A. Mathematical model of COVID-19 intervention scenarios for São Paulo—Brazil. *Nature Communications*, Nature Publishing Group, v. 12, n. 1, p. 418, jan. 2021. ISSN 2041-1723. Citado na pág. 45.

NETTO, G. F.; VILLARDI, J. W. R.; MACHADO, J. M. H.; SOUZA, M. d. S. d.; BRITO, I. F.; SANTORUM, J. A.; OCKÉ-REIS, C. O.; FENNER, A. L. D. Vigilância em saúde brasileira: reflexões e contribuição ao debate da 1ª conferência nacional de vigilância em saúde. *Ciência & Saúde Coletiva*, v. 22, n. 10, p. 3137–3148, 2017. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-81232017021003137&lng=pt&tlng=pt>. Citado na pág. 23.

NEWMAN, M. E. J. *Networks*. Second edition. [S.l.]: Oxford University Press, 2018. ISBN 9780198805090. Citado na pág. 59.

NIELSEN, R. C.; LUENGO-OROZ, M.; MELLO, M. B.; PAZ, J.; PANTIN, C.; ERKKOLA, T. Social media monitoring of discrimination and hiv testing in brazil, 2014–2015. *AIDS and Behavior*, v. 21, n. S1, p. 114–120, 2017. Disponível em: <<http://link.springer.com/10.1007/s10461-017-1753-2>>. Citado na pág. 24, 29, 31.

NIEUWENHUIJSEN, E. R.; ZEMPER, E.; MINER, K. R.; EPSTEIN, M. Health behavior change models and theories: contributions to rehabilitation. *Disability and Rehabilitation*, v. 28, n. 5, p. 245–256, 2006. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/09638280500197743>>. Citado na pág. 35, 36, 37, 38.

NSF. *NSF Award Search: Award # 2055251 - RAPID: Flexible, Efficient, and Available Bayesian Computation for Epidemic Models*. 2020. <https://www.nsf.gov/awardsearch/showAward?AWD_ID=2055251>. (Accessed on 07/26/2021). Citado na pág. 48.

ÖLANDER, F.; THØGERSEN, J. Understanding of consumer behaviour as a prerequisite for environmental protection. *Journal of Consumer Policy*, v. 18, n. 4, p. 345–385, 1995. Disponível em: <<http://link.springer.com/10.1007/BF01024160>>. Citado na pág. 162.

OMS. *WHO guidelines on ethical issues in public health surveillance*. Geneva: World Health Organization, 2017. ISBN 9789241512657. Disponível em: <<https://apps.who.int/iris/handle/10665/255721>>. Citado na pág. 23.

OMS. *The best time to prevent the next pandemic is now: countries join voices for better emergency preparedness*. 2020. Disponível em: <<https://www.who.int/news/item/01-10-2020-the-best-time-to-prevent-the-next-pandemic-is-now-countries-join-voices-for-better-emergency-preparedness>>. Citado na pág. 22.

OMS. *Coronavirus disease (COVID-19)*. 2023. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019?adgroupsurvey={adgroupsurvey}&gclid=Cj0KCQjw1vSZBhDuARIsAKZlijR_H0d0rIX3jAHT1qPjbv7QNtVTsqr481nuiJ-fiUHRrUEhcjL44dUaAvtMEALw_wcB>. (Accessed on 24/06/2023). Citado na pág. 22.

OMS. *Coronavirus disease (COVID-19): What are the symptoms of COVID-19*. 2023. <<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>>. (Accessed on 08/13/2021). Citado na pág. 53.

PAIM, J. S. *O que é o SUS*. Rio de Janeiro: Fiocruz, 2020. (Coleção Temas em Saúde). ISBN 9788575411858. Citado na pág. 23.

PALMEIRA, A. L.; TEIXEIRA, P. J.; BRANCO, T. L.; MARTINS, S. S.; MINDERICO, C. S.; BARATA, J. T.; SERPA, S. O.; SARDINHA, L. B. Predicting short-term weight loss using four leading health behavior change theories. *International Journal of Behavioral Nutrition and Physical Activity*, v. 4, n. 1, p. 14, 2007. Disponível em: <<http://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-4-14>>. Citado na pág. 35, 36.

Pan American Health Organization. *Folha informativa sobre COVID-19*. 2021. <<https://www.paho.org/pt/covid19>>. (Accessed on 08/13/2021). Citado na pág. 53.

Pandas Development Team. *pandas documentation — pandas 1.3.1 documentation*. 2021. <<https://pandas.pydata.org/docs/>>. (Accessed on 08/13/2021). Citado na pág. 54, 62.

PANTER-BRICK, C.; CLARKE, S. E.; LOMAS, H.; PINDER, M.; LINDSAY, S. W. Culturally compelling strategies for behaviour change: A social ecology model and case study in malaria prevention. *Social Science & Medicine*, v. 62, n. 11, p. 2810–2825, 2006. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0277953605005356>>. Citado na pág. 163.

PAVLOV, I. Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences*, v. 17, n. 3, p. 136–141, Jul 1927. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4116985/>>. Citado na pág. 160.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na pág. 65, 66, 67, 68, 95.

PENDER, N. J.; WALKER, S. N.; SECHRIST, K. R.; STROMBORG, M. F. Development and testing of the health promotion model. *Cardio-vascular nursing*, v. 24, n. 6, p. 41–43, 1988. Cited By :25. Disponível em: <www.scopus.com>. Citado na pág. 161.

POLDI, F.; ZACHARIAS, C.; LEVIN, M.; SHABASHEV, F.; URBANSKI, A.; ROBERTSON, C.; LUNN, L. *twintproject/twint: An advanced Twitter scraping & OSINT tool written in Python that doesn't use Twitter's API, allowing you to scrape a user's followers, following, Tweets and more while evading most API limitations*. 2022. <<https://github.com/twintproject/twint>>. (Accessed on 05/08/2022). Citado na pág. 61.

PORCHER, S.; RENAULT, T. Social distancing beliefs and human mobility: Evidence from twitter. *PLOS ONE*, Public Library of Science (PLoS), v. 16, n. 3, p. e0246949, Mar 2021. ISSN 1932-6203. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0246949>>. Citado na pág. 82.

PR, G. *Prefeitura lança campanha #FicaEmCasaFoz - Clickfoz*. 2020. <<https://www.clickfozdoiguacu.com.br/prefeitura-lanca-campanha-ficaemcasafoz/>>. (Accessed on 11/12/2022). Citado na pág. 100.

PROCHASKA, J. O.; DICLEMENTE, C. C. Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy: Theory, Research & Practice*, v. 19, n. 3, p. 276–288, 1982. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/h0088437>>. Citado na pág. 163.

PROCHASKA, J. O.; DICLEMENTE, C. C.; NORCROSS, J. C. In search of how people change: Applications to addictive behaviors. *American Psychologist*, v. 47, n. 9, p. 1102–1114, 1992. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.47.9.1102>>.

Citado na pág. 163.

PROCHASKA, J. O.; REDDING, C. A.; HARLOW, L. L.; ROSSI, J. S.; VELICER, W. F. The transtheoretical model of change and hiv prevention: A review. *Health Education Quarterly*, v. 21, n. 4, p. 471–486, 1994. Disponível em: <<http://journals.sagepub.com/doi/10.1177/109019819402100410>>. Citado na pág. 163.

PROCHASKA, J. O.; VELICER, W. F. The transtheoretical model of health behavior change. *American Journal of Health Promotion*, v. 12, n. 1, p. 38–48, 1997. Disponível em: <<http://journals.sagepub.com/doi/10.4278/0890-1171-12.1.38>>. Citado na pág. 35, 36.

RASCHKA, S.; BOURBEAU, J.; NAKANO, R.; MHASAKAR, M.; MORI, K. *LogisticRegression: A binary classifier - mlxtend*. 2022. <http://rasbt.github.io/mlxtend/user_guide/classifier/LogisticRegression/>. (Accessed on 05/11/2022). Citado na pág. 65, 73, 75, 77.

RAVESLOOT, C.; RUGGIERO, C.; IPSEN, C.; TRACI, M.; SEEKINS, T.; BOEHM, T.; WARE-BACKS, D.; RIGLES, B. Disability and health behavior change. *Disability and Health Journal*, v. 4, n. 1, p. 19–23, 2011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1936657410000488>>. Citado na pág. 35, 38.

RECKLESS, W. *A New Theory of Delinquency and Crime*. [s.n.], 1961. ISBN 9781315087900 9781351548502. Disponível em: <<https://www.taylorfrancis.com/chapters/mono/10.4324/9781315087900-12/walter-reckless-1961-new-theory-delinquency-crime-federal-probation-25-pp-42-6-edward-wells>>. Citado na pág. 160.

REZENDE, L. F. M. d.; LOPES, M. R.; REY-LÓPEZ, J. P.; MATSUDO, V. K. R.; LUIZ, O. d. C. Sedentary behavior and health outcomes: An overview of systematic reviews. *PLoS ONE*, v. 9, n. 8, p. e105620, Aug 2014. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0105620>>. Citado na pág. 92.

RIMAL, R. N.; REAL, K. How behaviors are influenced by perceived norms: A test of the theory of normative social behavior. *Communication Research*, v. 32, n. 3, p. 389–414, 2005. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0093650205275385>>. Citado na pág. 163.

RIZK, A.; ELRAGAL, A. Data science: developing theoretical contributions in information systems via text analytics. *Journal of Big Data*, Springer Science and Business Media LLC, v. 7, n. 1, jan. 2020. Disponível em: <<https://doi.org/10.1186/s40537-019-0280-6>>. Citado na pág. 47, 48, 49.

ROGERS, E. M. *Diffusion of Innovations*. [S.l.]: Free Press, 1983. Citado na pág. 160, 162.

ROGERS, E. M. *Diffusion of innovations*. 5th ed. ed. [S.l.]: Free Press, 2003. ISBN 9780743222099. Citado na pág. 160.

ROGERS, E. M.; SINGHAL, A.; QUINLAN, M. M. Diffusion of innovations 1. In: _____. *An Integrated Approach to Communication Theory and Research*. 3. ed. Routledge, 2019. p. 415–434. ISBN 9780203710753. Disponível em: <<https://www.taylorfrancis.com/books/9781351358712/chapters/10.4324/9780203710753-35>>. Citado na pág. 160.

- ROGERS, R. W. A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology*, v. 91, n. 1, p. 93–114, 1975. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00223980.1975.9915803>>. Citado na pág. 162.
- ROMER, D.; HORNIK, R. Hiv education for youth: The importance of social consensus in behaviour change. *AIDS Care*, v. 4, n. 3, p. 285–303, 1992. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/09540129208253100>>. Citado na pág. 162.
- RONCATO, C. L.; ANDRADE, T. S. *Relação entre Taxa de Mortalidade da Covid-19 e a Desigualdade Tributária no Brasil*. 2021. <<https://cadernopaic.fae.edu/cadernopaic/article/viewFile/454/401>>. (Accessed on 10/12/2022). Citado na pág. 45.
- ROSENSTOCK, I. M. The health belief model and preventive health behavior. *Health Education Monographs*, v. 2, n. 4, p. 354–386, 1974. Disponível em: <<http://journals.sagepub.com/doi/10.1177/109019817400200405>>. Citado na pág. 36, 49, 50, 69, 92, 104.
- ROSENSTOCK, I. M. Historical origins of the health belief model. *Health Education Monographs*, v. 2, n. 4, p. 328–335, 1974. Disponível em: <<http://journals.sagepub.com/doi/10.1177/109019817400200403>>. Citado na pág. 36, 69, 161.
- ROSENSTOCK, I. M.; STRECHER, V. J.; BECKER, M. H. Social learning theory and the health belief model. *Health Education Quarterly*, v. 15, n. 2, p. 175–183, 1988. Disponível em: <<http://journals.sagepub.com/doi/10.1177/109019818801500203>>. Citado na pág. 36, 69, 161.
- ROSSUM, G. van; WARSAW, B.; COGHLAN, N. *Style Guide for Python Code*. [S.l.], 2001. Disponível em: <<https://www.python.org/dev/peps/pep-0008/>>. Citado na pág. 54, 61, 62, 103.
- RYAN, P. Integrated theory of health behavior change: Background and intervention development. *Clinical Nurse Specialist*, v. 23, n. 3, p. 161–170, 2009. Disponível em: <<https://journals.lww.com/00002800-200905000-00011>>. Citado na pág. 161, 162.
- RYAN, R. M.; DECI, E. L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, v. 55, n. 1, p. 68–78, 2000. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.55.1.68>>. Citado na pág. 35.
- SALATHÉ, M.; FREIFELD, C. C.; MEKARU, S. R.; TOMASULO, A. F.; BROWNSTEIN, J. S. Influenza a (h7n9) and the importance of digital epidemiology. *New England Journal of Medicine*, v. 369, n. 5, p. 401–404, 2013. Disponível em: <<http://www.nejm.org/doi/10.1056/NEJMp1307752>>. Citado na pág. 24.
- SANFORD, A. A. “i can air my feelings instead of eating them”: Blogging as social support for the morbidly obese. *Communication Studies*, v. 61, n. 5, p. 567–584, 2010. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/10510974.2010.514676>>. Citado na pág. 41, 42.
- SANTOS, A. L. M. F. D.; PELLINI, A. C. G.; NODA, E. S.; NÓIA, J. V. V.; FRAGA, P.; BRICHTA, C.; SANTOS, L. R. D.; Junior De Sousa Silva; STOROPOLI, J. E. *Brazilian Portuguese COVID-19 Tweets*. Zenodo, 2021. Disponível em: <<https://zenodo.org/record/5073680>>. Citado na pág. 54, 83, 86, 103, 156.
- SANTOS, A. L. M. F. dos. *COVID-19 Datasets*. 2023. <https://github.com/andreimfsantos/Tese_COVID19/tree/main/datasets>. Accessed on May 18, 2023. Citado na pág. 157.

SANTOS, A. L. M. F. dos. *Programa para ler o Banco de Dados de Síndrome Respiratória Aguda Grave (SRAG)*. 2023. GitHub repository. Disponível em: <https://github.com/andreImfsantos/Tese_COVID19/blob/main/Tese_Jupyter_Notebook/openDataSUS.ipynb>.

Citado na pág. 156.

SANTOS, A. L. M. F. dos. *Programa para minerar dados no Twitter sem API*. 2023. GitHub repository. Disponível em: <https://github.com/andreImfsantos/Tese_COVID19/blob/main/Tese_Jupyter_Notebook/TwitterCOVID_Symptoms01_ScrapingTwint.ipynb>. Citado na pág. 156.

SANTOS, A. L. M. F. dos. *Time Series Analysis of COVID-19 Data*. 2023. <https://github.com/andreImfsantos/Tese_COVID19/blob/main/Tese_Jupyter_Notebook/HBM_TimeSeries.ipynb>. Accessed on May 18, 2023. Citado na pág. 157.

SANTOS, A. L. M. F. dos. *Twitter COVID-19 Symptoms Amostragem*. 2023. <https://github.com/andreImfsantos/Tese_COVID19/blob/main/Tese_Jupyter_Notebook/TwitterCOVID_Symptoms04_Amostragem.ipynb>. Accessed on May 18, 2023. Citado na pág. 157.

SANTOS, A. L. M. F. dos. *COVID-19 Classifier Using Machine Learning – Logit Model*. 2023. <https://github.com/andreImfsantos/Tese_COVID19/blob/main/Tese_Jupyter_Notebook/HBM_COVID19_Classifier.ipynb>. Accessed on May 18, 2023. Citado na pág. 157.

SARKER, D.; SHRESTHA, S.; TAMANG, S. K. B. “we’ll starve to death”: The consequences of covid-19 over the lives of poor people with disabilities in rural nepal. *Asian Social Work and Policy Review*, p. aswp.12250, Feb 2022. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/aswp.12250>>. Citado na pág. 92.

SCANFELD, D.; SCANFELD, V.; LARSON, E. L. Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, v. 38, n. 3, p. 182–188, 2010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0196655310000349>>. Citado na pág. 41.

SCHIAVO, R. *Health communication: from theory to practice*. Second edition. [S.l.]: Jossey-Bass, 2014. (Jossey-bass public health). ISBN 9781118416396 9781118419120. Citado na pág. 24, 36.

SCHOOT, R. van de; DEPAOLI, S.; KING, R.; KRAMER, B.; MÄRTENS, K.; TADESSE, M. G.; VANNUCCI, M.; GELMAN, A.; VEEN, D.; WILLEMSSEN, J.; YAU, C. Bayesian statistics and modelling. *Nature Reviews Methods Primers*, v. 1, n. 1, p. 1, 2021. Disponível em: <<http://www.nature.com/articles/s43586-020-00001-2>>. Citado na pág. 81.

SCHWARTZ, S. The justice of need and the activation of humanitarian norms. *Journal of Social Issues*, v. 31, n. 3, p. 111–136, 1975. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1540-4560.1975.tb00999.x>>. Citado na pág. 162.

SCHWARZER, R. Self-efficacy in the adoption and maintenance of health behaviors: Theoretical approaches and a new model. *Self-efficacy: Thought control of action*, 01 1992. Citado na pág. 161.

SCHWARZER, R. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology*, v. 57, n. 1, p. 1–29, 2008. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1464-0597.2007.00325.x>>. Citado na pág. 161.

SCHWARZER, R. *Self-efficacy in the adoption and maintenance of health behaviors: Theoretical approaches and a new model*. In R. Schwarzer (Ed.), *Self-Efficacy: Thought Control Of Action*. 1a edição. ed. [S.l.]: Taylor & Francis, 2014. Citado na pág. 161.

SEKALALA, S.; DAGRON, S.; FORMAN, L.; MEIER, B. M. Analyzing the human rights impact of increased digital public health surveillance during the covid-19 crisis. *Health and Human Rights*, v. 22, n. 2, p. 7–20, Dec 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7762901/>. Citado na pág. 24.

SKINNER, B. F. *Science And Human Behavior*. [S.l.]: Simon and Schuster, 1953. ISBN 9781476716152. Citado na pág. 162.

SNYDER, L. B.; HAMILTON, M. A.; MITCHELL, E. W.; KIWANUKA-TONDO, J.; FLEMING-MILICI, F.; PROCTOR, D. A meta-analysis of the effect of mediated health communication campaigns on behavior change in the united states. *Journal of Health Communication*, v. 9, n. sup1, p. 71–96, 2004. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/10810730490271548>. Citado na pág. 29.

SP, G. (70) *Campanha de Utilidade Pública: Fique em Casa - YouTube*. 2020. <https://www.youtube.com/watch?v=D2zESBXbauA>. (Accessed on 11/12/2022). Citado na pág. 100.

SPAARGAREN, G. Ecological modernization theory and domestic consumption. *Journal of Environmental Policy & Planning*, v. 2, n. 4, p. 323–335, 2000. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/714038564>. Citado na pág. 160.

SPAARGAREN, G.; VLIET, B. V. Lifestyles, consumption and the environment: The ecological modernization of domestic consumption. *Environmental Politics*, v. 9, n. 1, p. 50–76, 2000. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/09644010008414512>.

Citado na pág. 160.

SRISATHAN, W. A.; NARUETHARADHOL, P. A covid-19 disruption: The great acceleration of digitally planned and transformed behaviors in thailand. *Technology in Society*, v. 68, p. 101912, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0160791X22000537>. Citado na pág. 25.

STATISTA. *Number of social media users 2025*. 2021. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. (Accessed on 09/04/2021). Citado na pág. 24.

STEIJVERS, L. C. J.; BRINKHUES, S.; HOEBE, C. J. P. A.; TILBURG, T. G. V.; CLAESSEN, V.; BOUWMEESTER-VINCKEN, N.; HAMERS, F.; VRANKEN, P.; DUKERS-MUIJRS, N. H. T. M. Social networks and infectious diseases prevention behavior: A cross-sectional study in people aged 40 years and older. *PLOS ONE*, v. 16, n. 5, p. e0251862, May 2021. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0251862>. Citado na pág. 31.

STERN, P. C.; DIETZ, T.; ABEL, T. D.; GUAGNANO, G. A.; KALOF, L. A value-belief-norm theory of support for social movements: The case of environmentalism. *Human Ecology Review*, v. 6, p. 81–97, 1999. Disponível em: <https://humanecologyreview.org/pastissues/her62/62sternetal.pdf>. Citado na pág. 163.

STIEGLITZ, S.; MIRBABAIE, M.; ROSS, B.; NEUBERGER, C. Social media analytics – challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, v. 39, p. 156–168, 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0268401217308526>. Citado na pág. 26, 40.

- STIGLER, S. M. Thomas bayes's bayesian inference. *Journal of the Royal Statistical Society. Series A (General)*, v. 145, n. 2, p. 250, 1982. Disponível em: <<https://www.jstor.org/stable/2981538?origin=crossref>>. Citado na pág. 80.
- STOK, F. M.; RIDDER, D. T. D. de. The focus theory of normative conduct. In: _____. *Social Psychology in Action*. Springer International Publishing, 2019. p. 95–110. ISBN 9783030137878 9783030137885. Disponível em: <http://link.springer.com/10.1007/978-3-030-13788-5_7>. Citado na pág. 160.
- STOROPOLI, J.; SANTOS, A. L. M. F. d.; PELLINI, A. C. G.; BALDWIN, B. Simulation-driven covid-19 epidemiological modeling with social media. Jun 2021. Disponível em: <<https://arxiv.org/abs/2106.11686v1>>. Citado na pág. 12, 24, 50, 82, 86.
- STOROPOLI, J. E.; SANTOS, A. L. M. F. D.; NÓIA, J. V. V.; NODA, E.; PELLINI, A. Modelo epidemiológico compartimental usando o truque de corrente linear para covid-19. 2021. Disponível em: <<https://rgdoi.net/10.13140/RG.2.2.27786.64961>>. Citado na pág. 86.
- STRACK, F.; DEUTSCH, R. Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, v. 8, n. 3, p. 220–247, 2004. Disponível em: <http://journals.sagepub.com/doi/10.1207/s15327957pspr0803_1>. Citado na pág. 162.
- SU, Y.; XIAO, X. Beyond the differential gains model: the effects of authoritarian orientation, social media use, and political discussion on political participation in taiwan and south korea. *Atlantic Journal of Communication*, v. 28, n. 4, p. 241–256, Aug 2020. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/15456870.2019.1618304>>. Citado na pág. 31.
- SUTHERLAND, E. H. *Principles of criminology*. 4th edition. ed. [S.l.]: J.B. Lippincott Co, 1947. Citado na pág. 160.
- SUTHERLAND, E. H.; CRESSEY, D. R.; LUCKENBILL, D. F. *Principles of criminology*. 11. ed. ed. [S.l.]: General Hall, 1992. (The Reynolds Series in sociology). ISBN 9780930390693 9780930390709. Citado na pág. 160.
- TAJFEL, H.; TURNER, J. C. Social identity theory of intergroup behavior. In: _____. *Political Psychology*. 0. ed. Psychology Press, 1974. p. 276–293. ISBN 9780203505984. Disponível em: <<https://journals.sagepub.com/doi/10.1177/053901847401300204>>. Citado na pág. 163.
- TEAM, R. C. R. *A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Citado na pág. 85, 157.
- TEIXEIRA, M. G.; COSTA, M. d. C. N.; CARMO, E. H.; OLIVEIRA, W. K. d.; PENNA, G. O. Vigilância em saúde no sus - construção, efeitos e perspectivas. *Ciência & Saúde Coletiva*, v. 23, n. 6, p. 1811–1818, 2018. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-81232018000601811&lng=pt&tlng=pt>. Citado na pág. 23.
- TENNENHOLTZ, M. On computational social laws for dynamic non-homogeneous social structures. *Journal of Experimental and Theoretical Artificial Intelligence*, v. 7, p. 379–390, 1994. Citado na pág. 49.
- TESSER, C. D.; NORMAN, A. H. Prevenção quaternária e medicalização: conceitos inseparáveis. *Interface - Comunicação, Saúde, Educação*, v. 25, p. e210101, Aug 2021. Disponível em: <<https://www.scielo.br/j/icse/a/7mkSdVRspSG34PnChDr8B9L/?lang=pt>>.

THOMPSON, B.; KINNE, S. Social change theory: Applications to community health. In: _____. *Health Promotion at the Community Level: New Advance*. SAGE Publications, Inc., 1999. p. 29–46. ISBN 9780761913047 9781452204789. Disponível em: <<http://sk.sagepub.com/books/health-promotion-at-the-community-level-2e/n2.xml>>. Citado na pág. 162.

THUNSTRÖM, L.; NEWBOLD, S. C.; FINNOFF, D.; ASHWORTH, M.; SHOGREN, J. F. The benefits and costs of using social distancing to flatten the curve for covid-19. *Journal of Benefit-Cost Analysis*, v. 11, n. 2, p. 179–195, 2020. Disponível em: <https://www.cambridge.org/core/product/identifier/S2194588820000123/type/journal_article>. Citado na pág. 29.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 58, n. 1, p. 267–288, 1996. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1996.tb02080.x>>. Citado na pág. 67.

TIBSHIRANI, R. The lasso method for variable selection in the cox model. *Statistics in Medicine*, v. 16, n. 4, p. 385–395, Feb 1997. Disponível em: <[https://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3.0.CO;2-3](https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3)>. Citado na pág. 67.

TORTORELLA, G. L.; FOGLIATTO, F. S.; SAURIN, T. A.; TONETTO, L. M.; MCFARLANE, D. Contributions of healthcare 4.0 digital applications to the resilience of healthcare organizations during the covid-19 outbreak. *Technovation*, v. 111, p. 102379, 2022. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0166497221001607>>. Citado na pág. 24.

TRIANDIS, H. C. *Interpersonal behavior*. [S.l.]: Brooks/Cole Pub. Co, 1977. ISBN 9780818501883. Citado na pág. 163.

TUTTE, W. T.; TUTTE, W. T.; NASH-WILLIAMS, C. S. J. A. *Graph theory*. Transferred to digital print. [S.l.]: Cambridge Univ. Press, 2005. (Encyclopedia of mathematics and its applications). ISBN 9780521794893. Citado na pág. 59.

TWITTER. *Q2 2020 Letter to Shareholders*. 2020. Disponível em: <https://s22.q4cdn.com/826641620/files/doc_financials/2020/q2/Q2-2020-Shareholder-Letter.pdf>. Citado na pág. 82.

UNINOVE. *Cidades e Pandemias: Lições e Desafios para uma Recuperação Resiliente em São Paulo*. 2022. <<https://www.youtube.com/watch?v=3NWhWEZyxA8>>. Acessado em 15 de Maio de 2023. Citado na pág. 87.

USHER, W. Types of social media (web 2.0) used by australian allied health professionals to deliver early twenty-first-century practice promotion and health care. *Social Work in Health Care*, v. 50, n. 4, p. 305–329, Apr 2011. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00981389.2010.534317>>. Citado na pág. 42.

VAUGHAN, P. W.; ROGERS, E. M. A staged model of communication effects: Evidence from an entertainment-education radio soap opera in tanzania. *Journal of Health Communication*, v. 5, n. 3, p. 203–227, 2000. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/10810730050131398>>. Citado na pág. 162.

VEHTARI, A.; GABRY, J.; MAGNUSSON, M.; YAO, Y.; BÜRKNER, P.-C.; PAANANEN, T.; GELMAN, A. *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*. 2022. R package version 2.5.0. Disponível em: <<https://mc-stan.org/loo/>>. Citado na pág. 101.

- VEHTARI, A.; GELMAN, A.; GABRY, J. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, v. 27, n. 5, p. 1413–1432, 2017. Disponível em: <<http://link.springer.com/10.1007/s11222-016-9696-4>>. Citado na pág. 101, 102.
- VENKATESH, V.; BALA, H. Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, v. 39, n. 2, p. 273–315, 2008. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1540-5915.2008.00192.x>>. Citado na pág. 163.
- VENKATESH, V.; DAVIS, F. D. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, v. 46, n. 2, p. 186–204, 2000. Disponível em: <<http://pubsonline.informs.org/doi/abs/10.1287/mnsc.46.2.186.11926>>. Citado na pág. 163.
- VERBEKE JACQUES VIAENE, O. W. Health communication and consumer behavior on meat in belgium: From bse until dioxin. *Journal of Health Communication*, v. 4, n. 4, p. 345–357, 1999. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/108107399126869>>. Citado na pág. 29.
- VIDAL-ALABALL, J.; ACOSTA-ROJA, R.; HERNÁNDEZ, N. P.; LUQUE, U. S.; MORRISON, D.; PÉREZ, S. N.; PEREZ-LLANO, J.; VÈRGES, A. S.; SEGUÍ, F. L. Telemedicine in the face of the covid-19 pandemic. *Atención Primaria*, v. 52, n. 6, p. 418–422, 2020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0212656720301268>>. Citado na pág. 94.
- VLEK, C. Household consumption, quality-of-life and environmental impacts: a psychological perspective and empirical study. In: _____. *Green households?* [S.l.]: Earthscan Publications, 1998. p. 141 – 183. Citado na pág. 162.
- VRIES, D.; MUDDE, A.; DIJKSTRA, A. *The attitude–social influence–efficacy model applied to the prediction of motivational transitions in the process of smoking cessation*. Psychology Press, 2013. ISBN 9781315080055. Disponível em: <<https://www.taylorfrancis.com/chapters/edit/10.4324/9781315080055-19/attitude-social-influence-efficacy-model-applied-prediction-motivational-transitions-process-smoking-cessation>>. Citado na pág. 161.
- VRIES, H. d.; MESTERS, I.; STEEG, H. v. d.; HONING, C. The general public’s information needs and perceptions regarding hereditary cancer: an application of the integrated change model. *Patient Education and Counseling*, v. 56, n. 2, p. 154–165, 2005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0738399104000758>>. Citado na pág. 161.
- VRIES, H. de; DIJKSTRA, M.; KUHLMAN, P. Self-efficacy: the third factor besides attitude and subjective norm as a predictor of behavioural intentions. *Health Education Research*, v. 3, n. 3, p. 273–282, 1988. Disponível em: <<https://academic.oup.com/her/article-lookup/doi/10.1093/her/3.3.273>>. Citado na pág. 161.
- WAGENAAR, A. C.; PERRY, C. L. Community strategies for the reduction of youth drinking: Theory and application. *Journal of Research on Adolescence*, v. 4, n. 2, p. 319–345, Apr 1994. Disponível em: <https://www.tandfonline.com/doi/abs/10.1207/s15327795jra0402_8>. Citado na pág. 161.
- WALSH, J. C.; GROARKE, J. M. Integrating behavioral science with mobile (mhealth) technology to optimize health behavior change interventions. *European Psychologist*, v. 24,

n. 1, p. 38–48, Jan 2019. Disponível em: <<https://econtent.hogrefe.com/doi/10.1027/1016-9040/a000351>>. Citado na pág. 35.

WEBER, M.; ROTH, G.; WITTICH, C. *Economy and society: an outline of interpretive sociology*. [S.l.]: University of California Press, 1978. ISBN 9780520028241 9780520035003.

Citado na pág. 162.

WEBER, M.; TRIBE, K. *Economy and society: a new translation*. [S.l.]: Harvard University Press, 2019. ISBN 9780674916548. Citado na pág. 162.

WEINSTEIN, N.; SANDMAN, P.; BLALOCK, S. The precaution adoption process model. In: _____. *Health Behaviour and Health Education: Theory, Research and Practice*. [S.l.]: John Wiley & Sons, Inc., 2008. p. 123 – 148. Citado na pág. 162.

WEINSTEIN, N. D.; SANDMAN, P. M. A model of the precaution adoption process: Evidence from home radon testing. *Health Psychology*, v. 11, n. 3, p. 170–180, 1992. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-6133.11.3.170>>. Citado na pág. 162.

WEISS, H.; CROPANZANO, R. *Affective events theory: A theoretical discussion of the structure, causes, and consequences of affective experiences at work*. *Research in organizational behavior*. [S.l.]: JAI Press, 1996. v. 18. ISBN 9781559389389. Citado na pág. 160.

WEST, R.; BROWN, J. *Theory of addiction*. Second edition. [S.l.]: Wiley Blackwell/Addiction Press, 2014. ISBN 9781118484913 9781118484906 9781118484920. Citado na pág. 162.

WEST, R.; GODINHO, C. A.; BOHLEN, L. C.; CAREY, R. N.; HASTINGS, J.; LEFEVRE, C. E.; MICHIE, S. Development of a formal system for representing behaviour-change theories. *Nature Human Behaviour*, v. 3, n. 5, p. 526–536, 2019. Disponível em: <<http://www.nature.com/articles/s41562-019-0561-2>>. Citado na pág. 36.

WITTE, K. Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs*, v. 59, n. 4, p. 329–349, 1992. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/03637759209376276>>. Citado na pág. 160.

W.O.KERMACK; MCKENDRICK, A. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, v. 115, n. 772, p. 700–721, 1927. Disponível em: <<https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118>>. Citado na pág. 44.

WWF. *A Call to Stop the Next Pandemic | Stories | WWF*. 2021. Disponível em: <<https://www.worldwildlife.org/stories/a-call-to-stop-the-next-pandemic>>. Citado na pág. 22.

XUE, Y.; HASSAN, S.; BAHADER, S.; HABIB, S. Investigating the impact of fear related to covid-19 disease on situational humor via social networking sites. *Frontiers in Psychology*, v. 12, p. 721304, Sep 2021. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpsyg.2021.721304/full>>. Citado na pág. 31.

YADAV, S. K.; AKHTER, Y. Statistical modeling for the prediction of infectious disease dissemination with special reference to covid-19 spread. *Frontiers in Public Health*, v. 9, p. 645405, Jun 2021. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpubh.2021.645405/full>>. Citado na pág. 26.

- YIN, F.; XIA, X.; SONG, N.; ZHU, L.; WU, J. Quantify the role of superspreaders -opinion leaders- on covid-19 information propagation in the chinese sina-microblog. *PLOS ONE*, v. 15, n. 6, p. e0234023, Jun 2020. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0234023>>. Citado na pág. 47.
- YOGURTCU, O. N.; MESSAN, M. R.; GERKIN, R. C.; BELOV, A. A.; YANG, H.; FORSHÉE, R. A.; CHOW, C. C. A quantitative evaluation of covid-19 epidemiological models. *medRxiv*, p. 2021.02.06.21251276, Feb 2021. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7872378/>>. Citado na pág. 85.
- YOUNG, S. D.; RICE, E. Online social networking technologies, hiv knowledge, and sexual risk and testing behaviors among homeless youth. *AIDS and Behavior*, v. 15, n. 2, p. 253–260, 2011. Disponível em: <<http://link.springer.com/10.1007/s10461-010-9810-0>>. Citado na pág. 42.
- ZHANG, E. X.; YANG, Y.; SHANG, R. D.; SIMONS, J. J. P.; QUEK, B. K.; YIN, X. F.; SEE, W.; OH, O. S. H.; NANDAR, K. S. T.; LING, V. R. Y.; CHAN, P. P.; WANG, Z.; GOH, R. S. M.; JAMES, L.; TEY, J. S. H. Leveraging social networking sites for disease surveillance and public sensing: the case of the 2013 avian influenza a(h7n9) outbreak in china. *Western Pacific Surveillance and Response Journal*, v. 6, n. 2, p. 66–72, Jun 2015. Disponível em: <<http://ojs.wpro.who.int/ojs/index.php/wpsar/article/view/307/487>>. Citado na pág. 24.

Apêndices

DICIONÁRIO DE EXPRESSÕES DE TWEETS

Dicionário de expressões para o modelo de classificação de tweets.

1. abril: abrilmes
2. adv: advogado
3. aera: aereas
4. aglomerabrasil: aglomera brasil
5. ainsa: ainda
6. alemaoferando: alemao fernando
7. art: arte
8. azi: azitromicima
9. azidroxina: azitromicima
10. azimitrina: azitromicima
11. azitro: azitromicima
12. azitrom: azitromicima
13. azitromessias: azitromicima
14. azitromicima: azitromicima
15. azitromincina: azitromicima
16. azm: azitromicima
17. azt: azitromicima
18. aztmcn: azitromicima
19. bisavó: avó
20. bisavô: avó
21. bolso: bolsonaro
22. bolsomato: bolsonaro
23. bolsomico: bolsonaro
24. bolsominion: bolsonaro
25. bolsomion: bolsonaro
26. bolsomito: bolsonaro
27. bolsonar: bolsonaro
28. bolsonariana: bolsonaro
29. bolsonariano: bolsonaro
30. bolsonarista: bolsonaro
31. bolsonaristas: bolsonaro
32. bolsonaro: bolsonaro
33. bolsonaroangelo: bolsonaro
34. bolsonaroenlouqueceu: bolsonaro enlouqueceu
35. bolsonaroestavacerto: bolsonaro estava certo
36. bolsonarofiel: bolsonaro fiel
37. bolsonarogenocida: bolsonaro genocida
38. bolsonaroorgulhaobrasil: bolsonaro orgulha o brasil
39. bolsonaroreeleito: bolsonaro reeleito
40. bolsonarosempreterazao: bolsonaro estava certo
41. bolsonarosp: bolsonaro
42. bolsonarotemrazao: bolsonaro estava certo

- | | |
|--|--|
| 43. bolsonarotemrazão: bolsonaro estava certo | 67. cansaço: cansaço |
| 44. bolsonarotemrazaosim: bolsonaro estava certo | 68. cansada: cansaço |
| 45. bolsonarovirus: bolsonaro virus | 69. cansado: cansaço |
| 46. bolsopil: bolsonaro | 70. cansados: cansaço |
| 47. bolsopill: bolsonaro | 71. cansadossimvencidosnunca: cansaço |
| 48. bolsoquina: bolsonaro cloroquina | 72. cansando: cansaço |
| 49. bolsoregrets: bolsonaro | 73. cansativa: cansaço |
| 50. bolsoridr: bolsonaro | 74. cansativo: cansaço |
| 51. bolsoshower: bolsonaro | 75. cansava: cansaço |
| 52. bolsotroopers: bolsonaro | 76. cansei: cansaço |
| 53. br: brasil | 77. canso: cansaço |
| 54. bra: brasil | 78. caosedevastacao: caos e devastacao |
| 55. brasi: brasil | 79. caossocial: caos social |
| 56. brasile: brasil | 80. caraaaa: cara |
| 57. brasilsemcorrupção: brasil sem corrupção | 81. carai: caramba |
| 58. butantã: butantã | 82. caraí: caramba |
| 59. butantan: butantã | 83. caraio: caramba |
| 60. butantanoficial: butantã | 84. caralhada: caramba |
| 61. cancer: cancer | 85. caralho: caramba |
| 62. cancêr: cancer | 86. caramba: caramba |
| 63. câncer: cancer | 87. carambola: caramba |
| 64. cancerjack: cancer | 88. cardi: cardio |
| 65. cancerpain: cancer | 89. cardiaca: cardio |
| 66. cansaco: cansaço | 90. cardíaca: cardio |
| | 91. cardiacar: cardio |
| | 92. cardíacas: cardio |

93. cardiaco: cardio
94. cardíaco: cardio
95. cardiacos: cardio
96. cardíacos: cardio
97. cardio: cardio
98. cardiogênico: cardio
99. cardiologista: cardio
100. cardiopata: cardio
101. cardiopatia: cardio
102. cardiorespiratoria: cardio
103. cardiorrespiratória: cardio
104. cardiovascular: cardio
105. cardiovasculares: cardio
106. carênciavirus: carência virus
107. carentevirus: carência virus
108. casossuspeitos: casos suspeitos
109. catastrophe: catastrophe
110. catástrofe: catastrophe
111. catástrofes: catastrophe
112. catastrófico: catastrophe
113. cbç: cabeça
114. cbn: cbn
115. cbnbrasil: cbn
116. cbnc: cbn
117. cbndiario: cbn
118. cbnjoinville: cbn
119. cbnoficial: cbn
120. ccabeçaa: cabeça
121. ccambara: caramba
122. cega: cegueira
123. cegaa: cegueira
124. cego: cegueira
125. cegou: cegueira
126. cegueira: cegueira
127. celula: celula
128. célula: celula
129. celular: celula
130. censurados: censura
131. censuradostf: censura
132. centrooeste: centro oeste
133. chloroquine: cloroquina
134. ciência: ciencia
135. ciências: ciencia
136. ciente: ciencia
137. cientifica: ciencia
138. científica: ciencia
139. cientificamente: ciencia
140. científico: ciencia
141. científicos: ciencia
142. cientista: ciencia
143. cientistas: ciencia
144. civilizedngr: civil

145. civis: civil
146. clo: cloroquina
147. cloqna: cloroquina
148. cloriquina: cloroquina
149. cloro: cloroquina
150. cloroaquina: cloroquina
151. cloroclina: cloroquina
152. cloroqui: cloroquina
153. cloroquina: cloroquina
154. cloroquinacura: cloroquina cura
155. cloroquinada: cloroquina
156. cloroquinasalvavidas: cloroquina salva
vidas
157. cloroquinions: cloroquina
158. cnn: cnn
159. cnnada: cnn
160. cnnbarsil: cnn
161. cnnbrasil: cnn
162. cnnbrasillixo: cnn
163. cnnnovodia: cnn
164. codvid: covid
165. conona: covid
166. conoravirus: covid
167. convd: covid
168. convid: covid
169. coroanavírus: covid
170. corona: covid
171. coronaravirus: covid
172. coronariana: covid
173. coronaravirus: covid
174. coronas: covid
175. coronasus: covid sus
176. coronavairus: covid
177. coronavgd: covid
178. coronavirus: covid
179. coronavírus: covid
180. CORONAVÍRUS: covid
181. coronavirusbra: covid brasil
182. coronavirusbrasil: covid brasil
183. coronavirusbrazil: covid brasil
184. coronavirusnobrasil: covid brasil
185. coronavirusoutbreak: covid
186. coronavirusplantão: covid
187. coronavirustime: covid
188. coronga: covid
189. corongada: covid
190. corongavirus: covid
191. coroninha: covid
192. corony: covid
193. coroquina: cloroquina
194. corouna: covid
195. corovairis: covid
196. cov: covid

197. covi: covid
198. covic: covid
199. covid: covid
200. covid-: covid
201. covida: covid
202. covidao: covid
203. covidão: covid
204. covidi: covid
205. covidmata: covid mata
206. covit: covid
207. crianccasyanomami: crianccas yanomami
208. cv: covid
209. cvd: covid
210. cvk: covid
211. cvs: covid
212. diabete: diabete
213. diabetes: diabete
214. diabetica: diabete
215. diabética: diabete
216. diabetico: diabete
217. diabético: diabete
218. diabeticos: diabete
219. diabéticos: diabete
220. dnv: novamente
221. dorcronica: dor cronica
222. dorde: dor de
223. dordecabeca: dor de cabeça
224. dordecabeça: dor de cabeça
225. dores: dor
226. doresssss: dor
227. doria: doria
228. dória: doria
229. dorian: doria
230. dórias: doria
231. doriavaicair: doria vai cair
232. dorm1: dormir
233. dorn: dor
234. dorzinha: dor
235. dr: doutor
236. dra: doutor
237. dr^a: doutor
238. drauzio: drauzio
239. dráuzio: drauzio
240. drauziovarella: drauzio
241. einsten: einstein
242. epidemia: epidemiologia
243. epidemias: epidemiologia
244. epidêmico: epidemiologia
245. epidemiologia: epidemiologia
246. epidemiológica: epidemiologia
247. epidemiológicas: epidemiologia
248. epidemiológico: epidemiologia

249. epidemiológicos: epidemiologia
250. epidemiologista: epidemiologia
251. eua: estadounidos
252. euamobrasil: estadounidos brasil
253. evolua: evoluir
254. evolucao: evoluir
255. evolução: evoluir
256. evolue: evoluir
257. evoluem: evoluir
258. evolui: evoluir
259. evoluído: evoluir
260. evoluindo: evoluir
261. evoluir: evoluir
262. evoluíram: evoluir
263. evoluíram: evoluir
264. evoluiu: evoluir
265. evoluo: evoluir
266. febra: febre
267. febrão: febre
268. febre: febre
269. fébre: febre
270. febreaftosa: febre
271. febres: febre
272. febreteen: febre
273. febrezinha: febre
274. febril: febre
275. febris: febre
276. filhos: filho
277. filha: filho
278. filhaadomar: filho
279. filhadaputa: filho
280. filhadoshouis: filho
281. filhas: filho
282. filhinha: filho
283. filhinhos: filho
284. filho: filho
285. filhos: filho
286. filhota: filho
287. filhote: filho
288. filmesdochico: filmes do chico
289. filminho: filme
290. fiqueemcasa: fique em casa
291. fiquei: fique
292. fiquem: fique
293. fiquememcasa: fique em casa
294. fiquemos: fique
295. gov: governo
296. govbr: governo
297. governa: governo
298. governaderes: governo
299. governado: governo
300. governador: governo

301. governadora: governo
302. governadores: governo
303. governante: governo
304. governar: governo
305. governo: governo
306. governoes: governo
307. governoma: governo
308. governopara: governo
309. governos: governo
310. governosc: governo
311. governosp: governo
312. govpt: governo
313. govrj: governo
314. govvs: governo
315. govsc: governo
316. hidrocloroquina: cloroquina
317. hidrox: cloroquina
318. hidrox: cloroquina
319. hidroxicloquina: cloroquina
320. hidroxicloraquina: cloroquina
321. hidroxicloriquina: cloroquina
322. hidroxicloroq: cloroquina
323. hidroxicloroquina: cloroquina
324. hidróxido: cloroquina
325. hidróxidocloroquina: cloroquina
326. hidroxizina: cloroquina
327. hidroxocloroquina: cloroquina
328. hidroxycloroquina: cloroquina
329. hidroxycloroquinasalva: cloroquina
330. hospitais: hospital
331. hospital: hospital
332. hospitala: hospital
333. hospitalalvarenga: hospital
334. hospitalar: hospital
335. hospitalares: hospital
336. hospitalarios: hospital
337. hospitaldebase: hospital
338. hospitaldeicarai: hospital
339. hospitaleiro: hospital
340. hospitalgeraldeguaianazes: hospital
341. hospitalhcfmusp: hospital
342. hospitalidade: hospital
343. hospitaliza: hospital
344. hospitalizacao: hospital
345. hospitalização: hospital
346. hospitalizacoes: hospital
347. hospitalizações: hospital
348. hospitalizada: hospital
349. hospitalizadah: hospital
350. hospitalizadas: hospital
351. hospitalizado: hospital
352. hospitalizados: hospital

353. hospitalizar: hospital
354. hospitalizations: hospital
355. hospitalized: hospital
356. hospitalizem: hospital
357. hospitalizou: hospital
358. hospitalmaedeus: hospital
359. hospitalveterinario: hospital
360. hospitalzinho: hospital
361. hospsaoluiz: hospital
362. hosputal: hospital
363. ideia: ideologia
364. idéia: ideologia
365. ideias: ideologia
366. idéias: ideologia
367. ideologia: ideologia
368. ideologica: ideologia
369. ideológico: ideologia
370. ideológicos: ideologia
371. idosa: idoso
372. idosas: idoso
373. idoso: idoso
374. idosos: idoso
375. imune: imunidade
376. imunidade: imunidade
377. imunizado: imunidade
378. imunitários: imunidade
379. imunizacaourgente: imunidade
380. imunizações: imunidade
381. imunizada: imunidade
382. imunizado: imunidade
383. imunizados: imunidade
384. imunizar: imunidade
385. imunizaram: imunidade
386. imunizem: imunidade
387. imuno: imunidade
388. imunológica: imunidade
389. imunologico: imunidade
390. imunológico: imunidade
391. imunológicos: imunidade
392. imunomodulador: imunidade
393. imunossuprimidos: imunidade
394. imunosuprimida: imunidade
395. infecção: infecção
396. infecçao: infecção
397. infecção: infecção
398. infeccções: infecção
399. infecciona: infecção
400. infeccionada: infecção
401. infeccionado: infecção
402. infeccionar: infecção
403. infeccionou: infecção
404. infecciosa: infecção

405. infeccioso: infecção
406. infecciosos: infecção
407. infeccoes: infecção
408. infeccões: infecção
409. infecções: infecção
410. infeccionada: infecção
411. infecsioso: infecção
412. infecta: infecção
413. infectada: infecção
414. infectadas: infecção
415. infectado: infecção
416. infectados: infecção
417. infectam: infecção
418. infectando: infecção
419. infectar: infecção
420. infectaram: infecção
421. infectasse: infecção
422. infectcontagiosa: infecção
423. infectibilidade: infecção
424. infections: infecção
425. infecto: infecção
426. infectocontagiosa: infecção
427. infectologista: infecção
428. infectologistas: infecção
429. infectou: infecção
430. inflama: inflamação
431. inflamação: inflamação
432. inflamação: inflamação
433. inflamações: inflamação
434. inflamada: inflamação
435. inflamadas: inflamação
436. inflamado: inflamação
437. inflamados: inflamação
438. inflamar: inflamação
439. inflamatória: inflamação
440. inflamatória: inflamação
441. inflamatórias: inflamação
442. inflamatorio: inflamação
443. inflamatório: inflamação
444. inflamatórios: inflamação
445. inflamção: inflamação
446. inflammatories: inflamação
447. inflamou: inflamação
448. internac: internação
449. internacao: internação
450. internação: internação
451. internação: internação
452. internacionais: internação
453. internacional: internação
454. internacionalista: internação
455. internacionalização: internação
456. internacionalmente: internação

- | | |
|--------------------------------|---------------------------------------|
| 457. internacionasi: interação | 482. kkkkkk: e |
| 458. internacoes: interação | 483. kkkkkkk: e |
| 459. interações: interação | 484. kkkkkkkk: e |
| 460. internada: interação | 485. kkkkkkkkk: e |
| 461. internadacom: interação | 486. kkkkkkkkkk: e |
| 462. internadado: interação | 487. kkkkkkkkkkk: e |
| 463. internadas: interação | 488. kkkkkkkkkkkk: e |
| 464. internado: interação | 489. kkkkkkkkkkkkj: e |
| 465. internados: interação | 490. kkkkkkkkkkkkk: e |
| 466. internamos: interação | 491. kkkkkkkkkkkkkk: e |
| 467. internando: interação | 492. kkkkkkkkkkkkkkk: e |
| 468. internar: interação | 493. kkkkkkkkkkkkkkkk: e |
| 469. internaram: interação | 494. kkkkkkkkkkkkkkkkk: e |
| 470. internarem: interação | 495. kkkkkkkkkkkkkkkkkk: e |
| 471. internaria: interação | 496. kkkkkkkkkkkkkkkkkkk: e |
| 472. internedo: interação | 497. kkkkkkkkvintao: e |
| 473. internei: interação | 498. kkkkkklkkkkkkkkk: e |
| 474. internem: interação | 499. kkkkkkmkkkkkkkkkk: e |
| 475. kk: e | 500. kkkkkm: e |
| 476. kkk: e | 501. kkkkkk: e |
| 477. kkkjj: e | 502. kkmorais: e |
| 478. kkkk: e | 503. kkqx: e |
| 479. kkkkk: e | 504. kksk: e |
| 480. kkkkkj: e | 505. kkskkskajskakskakakakajskaksk: e |
| 481. kkkkkjj: e | 506. kkzinha: e |
| | 507. kkzn: e |
| | 508. kl: e |

561. nãoo: nao
562. nãooo: nao
563. nãoooo: nao
564. naosalvo: nao
565. naosomosfasdecanalhas: nao
566. naot: nao
567. ãotomehidroxicloroquina: não tome
cloroquina
568. narigudo: nariz
569. narinas: nariz
570. nasais: nasal
571. nausea: náusea
572. náusea: náusea
573. náuseas: náusea
574. ne: não é
575. né: não é
576. necarios: necessário
577. necess: necessário
578. necessária: necessário
579. necessariamente: necessário
580. necessárias: necessário
581. necessario: necessário
582. necessário: necessário
583. necessários: necessário
584. necessidade: necessário
585. necessidades: necessário
586. necessita: necessário
587. necessitada: necessário
588. necessitam: necessário
589. necessitando: necessário
590. necessitar: necessário
591. necessitarem: necessário
592. necessitaria: necessário
593. necessite: necessário
594. necessitou: necessário
595. neg: negar
596. nega: negar
597. negacionismo: negar
598. negacionista: negar
599. negacionistas: negar
600. negada: negar
601. negado: negar
602. negah: negar
603. negam: negar
604. negamos: negar
605. negando: negar
606. negar: negar
607. negaram: negar
608. negarão: negar
609. negaria: negar
610. negat: negar
611. negativado: negativo
612. negativamente: negativo

613. negativar: negativo
614. negativaram: negativo
615. negativo: negativo
616. negativos: negativo
617. negativou: negativo
618. negava: negar
619. negligencia: negligência
620. negligência: negligência
621. negligenciada: negligência
622. negligenciado: negligência
623. negligenciar: negligência
624. negligenciou: negligência
625. negligênciou: negligência
626. negligentes: negligência
627. negou: negar
628. netflixbrasil: netflix
629. netinho: neto
630. neto: neto
631. netoemerson: neto
632. netolab: neto
633. netos: neto
634. ng: ninguém
635. ngm: ninguém
636. ngn: ninguém
637. noticia: notícia
638. notícia: notícia
639. notíciaatual: notícia
640. noticiadas: notícia
641. noticiado: notícia
642. noticiados: notícia
643. noticiário: notícia
644. noticias: notícia
645. notícias: notícia
646. noticiasdatv: notícia
647. noticiasdojapao: notícia
648. noticiasdopara: notícia
649. noticiou: notícia
650. notifica: notificação
651. notificação: notificação
652. notificações: notificação
653. notificada: notificação
654. notificado: notificação
655. notificados: notificação
656. notificam: notificação
657. notificando: notificação
658. notificar: notificação
659. novos: novo
660. num: numeros
661. núm: numeros
662. numéricas: numeros
663. numero: numeros
664. número: numeros

665. numeros: numeros
666. números: numeros
667. nuncaaa: nunca
668. nutri: nutrição
669. nutrição: nutrição
670. nutricionista: nutrição
671. nutrido: nutrição
672. nutriente: nutrição
673. obito: óbito
674. óbito: óbito
675. obitos: óbito
676. óbitos: óbito
677. objetiva: objetivo
678. objetivo: objetivo
679. objetivos: objetivo
680. obras: obra
681. obriga: obrigatório
682. obrigacao: obrigatório
683. obrigação: obrigatório
684. obrigadabolsonaro: obrigada bolsonaro
685. obrigaria: obrigatório
686. obrigatoria: obrigatório
687. obrigatória: obrigatório
688. obrigatoriamente: obrigatório
689. obrigatórias: obrigatório
690. obrigatoriedade: obrigatório
691. obrigatoriedades: obrigatório
692. obrigatório: obrigatório
693. obrigava: obrigatório
694. obrigou: obrigatório
695. observa: observar
696. observação: observar
697. observação: observar
698. observacional: observar
699. observada: observar
700. observador: observar
701. observados: observar
702. observando: observar
703. observar: observar
704. observatório: observar
705. obtos: óbito
706. obv: obvio
707. obvio: obvio
708. óbvio: obvio
709. óbvios: obvio
710. onibus: ônibus
711. oocportugal: ooc portugal
712. outbrea: outbreak
713. pandêmia: pandemia
714. pandemias: pandemia
715. pandemic: pandemia
716. policia: polícia

717. polícia: polícia
718. policiafederal: polícia
719. policiais: polícia
720. policial: polícia
721. policlinica: hospital
722. política: política
723. política: política
724. politicagem: política
725. politicamente: política
726. políticas: política
727. político: política
728. politicos: política
729. políticos: política
730. politiqueiro: política
731. politiza: política
732. politizadas: política
733. politizado: política
734. politizando: política
735. politizaram: política
736. populaç: população
737. populacao: população
738. população: população
739. população: população
740. populacional: população
741. popular: população
742. populistas: política
743. populoso: população
744. portugall: portugal
745. português: portugal
746. portugueses: portugal
747. positiva: positivo
748. positivadas: positivo
749. positivar: positivo
750. positivaram: positivo
751. positivas: positivo
752. positive: positivo
753. positivo: positivo
754. positivos: positivo
755. positivou: positivo
756. postinhokkkkkkkkk: postinho
757. postinhos: postinho
758. posto: postinho
759. postos: postinho
760. povinho: população
761. povo: população
762. povos: população
763. prescreve: prescreve
764. prescrevendo: prescreve
765. prescrever: prescreve
766. prescreveram: prescreve
767. prescreverem: prescreve
768. prescreveu: prescreve

769. prescricao: prescreve
770. prescrição: prescreve
771. prescrita: prescreve
772. prescritas: prescreve
773. prescrito: prescreve
774. prescritos: prescreve
775. psico: psicológico
776. psicóloga: psicológico
777. psicologia: psicológico
778. psicológica: psicológico
779. psicológica: psicológico
780. psicologicamente: psicológico
781. psicológicas: psicológico
782. psicologico: psicológico
783. psicológico: psicológico
784. psicológicos: psicológico
785. psicólogo: psicológico
786. psicólogo: psicológico
787. psicopata: psicológico
788. psicóticos: psicológico
789. ptbrasil: pt brasil
790. publica: pública
791. pública: pública
792. publicação: publicar
793. publicada: publicar
794. publicado: publicar
795. publicar: publicar
796. públicas: pública
797. publico: público
798. público: público
799. publicos: público
800. públicos: público
801. publicou: publicar
802. publiquei: publicar
803. puc: puc
804. puccampinas: puc
805. pucrio: puc
806. pulmao: pulmão
807. pulmão: pulmão
808. pulmoes: pulmão
809. pulmões: pulmão
810. pulmonar: pulmão
811. pulmonares: pulmão
812. racismo: racismo
813. racista: racismo
814. racistas: racismo
815. radiografia: radiografia
816. radiografias: radiografia
817. radioitatiaia: radiografia
818. radiologia: radiografia
819. radiológica: radiografia
820. radiologista: radiografia

821. radiology: radiografia
822. raiox: radiografia
823. raioxdosafetos: radiografia
824. remdesvir: remdesivir
825. remdio: remédio
826. remedio: remédio
827. remedios: remédio
828. remédios: remédio
829. reportagem: reportagem
830. reportagens: reportagem
831. repórter: reportagem
832. semaaaanas: semana
833. semana: semana
834. semanais: semana
835. semanal: semana
836. semanas: semana
837. semaninhas: semana
838. sensibilidade: sensível
839. sensíveis: sensível
840. sensivel: sensível
841. sensível: sensível
842. Sinto: sinto
843. sintoma: sintomas
844. sintomas: sintomas
845. síntomas: síntomas
846. síntomas: síntomas
847. sintomassãoidênticos: sintomas são idênticos
848. sintomática: sintomático
849. sintomáticas: sintomático
850. sintomático: sintomático
851. sintomaticos: sintomático
852. sintomáticos: sintomático
853. sintomatologias: síntomas
854. sintominhas: síntomas
855. sintomo: síntomas
856. sinus: sinusite
857. sinusite: sinusite
858. sinuzite: sinusite
859. sitoma: síntomas
860. sitomas: síntomas
861. socorremos: socorro
862. socorrendo: socorro
863. socorrer: socorro
864. socorrida: socorro
865. socorro: socorro
866. socorros: socorro
867. socorrrooo: socorro
868. tose: tosse
869. tosse: tosse
870. tosses: tosse

871. tossezinha: tosse
872. tossi: tosse
873. tossia: tosse
874. tossiamos: tosse
875. tossida: tosse
876. tossidas: tosse
877. tossindo: tosse
878. tossinha: tosse
879. tossir: tosse
880. tossiu: tosse
881. trab: trabalhar
882. trabalha: trabalhar
883. trabalhado: trabalhar
884. trabalhador: trabalhar
885. trabalhadores: trabalhar
886. trabalham: trabalhar
887. trabalhando: trabalhar
888. trabalhar: trabalhar
889. trabalharam: trabalhar
890. trabalharemos: trabalhar
891. trabalhasse: trabalhar
892. trabalhava: trabalhar
893. trabalhe: trabalhar
894. trabalhei: trabalhar
895. trabalhistas: trabalhar
896. trabalho: trabalhar
897. trabalhoou: trabalhar
898. trabalhou: trabalhar
899. trabalhos: trabalhar
900. trabalhou: trabalhar
901. trabalnar: trabalhar
902. tweet: tweet
903. tweetando: tweet
904. tweetar: tweet
905. tweetninguem: tweet
906. tweets: tweet
907. twekamxcw: tweet
908. twice: tweet
909. twit: tweet
910. twitre: tweet
911. twitter: tweet
912. twitterbrasil: tweet
913. twittersafety: tweet
914. twitterseguro: tweet
915. twittter: tweet
916. urgencia: urgência
917. urgência: urgência
918. urgências: urgência
919. urgente: urgência
920. urgentemente: urgência
921. urgentes: urgência
922. urina: urinária

















































923. urinaria: urinária
924. urinária: urinária
925. urinario: urinária
926. urologista: urinária
927. vet: veterinário
928. veterinária: veterinário
929. veterinario: veterinário
930. veterinário: veterinário
931. veterinários: veterinário
932. vicina: vacina
933. vó: avó
934. vô: avô
935. voce: você
936. você: você
937. voçê: você
938. vocêeee: você
939. voceis: você
940. voces: você
941. vocês: você
942. vomita: vomitar
943. vomitada: vomitar
944. vomitando: vomitar
945. vomitar: vomitar
946. vomitava: vomitar
947. vomitei: vomitar
948. vomito: vomitar
949. vômito: vomitar
950. vômito: vomitar
951. vomitos: vomitar
952. vômitos: vomitar
953. vômitos: vomitar
954. vomitou: vomitar
955. vomitp: vomitar
956. vovó: avó
957. vovô: avô

EMOJIS PRESENTES NOS DADOS DA PESQUISA

Emoji	Significado	Emoji	Significado
	mãos cruzadas		piscando o rosto com a língua
	cara de choro alto		cara de chapéu de cowboy
	rosto pensativo		peessoa correndo
	Pele clara		cara de lua nova
	peessoa facepalming		exclamação do coração
	coração vermelho		balão de pensamento
♀️	signo feminino		deslumbrado
	tom de pele médio-claro		hospital
	rosto com lágrimas de alegria		pilha de cocô
	rosto suplicante		coração preto
	cara de palhaço		rosto sem expressão
	rosto pensante		Coração azul
	rosto com olhos rolando		caveira e ossos cruzados
	cara de choro	™	marca comercial
	levantando as mãos		lar
♂️	signo masculino		rosto ligeiramente carrancudo
	coração partido	?	ponto de interrogação branco
	tom de pele médio	⏏	botão p
	peessoa dando de ombros		arco Iris
	cara de sono		fontes termiais
	notas musicais		teia de aranha
	rosto sorridente com olhos de coração		bandeira branca
	rosto abatido com suor		rosto sorridente com chifres
	pele morena escura		casal com coração
	rosto perseverante		copo de vidro
	afirmativo		ciclone
	rosto desapontado		folha caída
	índice de backhand apontando para baixo		erva
	rosto sorridente com corações		folha de Carvalho
	rosto cansado		tornado
	cara espirrando		folha flutuando ao vento
	rosto triste mas aliviado		árvore de folhas secas
	rolando de rir no chão		cara de vento
	cara de beicinho	♣️	trevo
	terno de coração		pegadas
✌️	mão da vitória		selfie

Emoji	Significado	Emoji	Significado
😬	rosto sem graça	👦	Garoto
👁️	cara piscando	🎬	projektor de filme
😓	rosto sorridente com suor	📺	presente embrulhado
😵	rosto tonto	👽	estrangeiro
🤯	rosto com termômetro	🏆	troféu
😫	rosto cansado	💵	nota de dólar
😱	cara gritando de medo	🏆	medalha militar
😬	cara carrancuda	💰	dinheiro com asas
😜	cara doida	🍸	copo de coquetel
😵	rosto confuso	🚬	cigarro
🗨️	rosto com símbolos na boca	📌	alfinete redondo
😘	cara mandando um beijo	📷	câmera com flash
⚠️	aviso	👉	dorso da mão com dedo indicador apontando para a esquerda
👨‍⚕️	rosto com máscara médica	🔊	alto-falante alto volume
👏	mãos de aplauso	❤️	coração laranja
✨	brilhos	🐮	boi
😬	Cara corada	📻	rádio
😡	cara de bravo	↗️	seta para cima e para direita
🤮	cara vomitando	🎤	microfone de estúdio
✅	botão de marca de seleção	🍪	panela rasa de comida
👤	cara	🐱	cara zangada com chifres
👩	mulher	🍚	molho de arroz
🔥	incêndio	🐮	cara de vaca
👌	OK mão	🐶	cara de cachorro
🙄	cara de cabeça para baixo	👑	Princesa
😓	rosto ansioso com suor	👑	Príncipe
💜	coração roxo	🌡️	termômetro
🦠	micróbio	🐊	crocodilo
🤔	rosto enjoado	🦵	perna
😐	rosto neutro	🐒	macaco
😊	rosto sorridente	😊	rosto sorridente com olhos grandes
😊	rosto sorridente	💻	computador portátil
😊	rosto levemente sorridente	💰	cara de dinheiro
💕	dois corações	👩	peessoa levantando a mão
💃	mulher dançando	👆	indicador apontando para cima
👉	índice de backhand apontando para a direita	👋	mão levantada

Emoji	Significado	Emoji	Significado
	cabeça falante		beijando o rosto com os olhos fechados
	rosto aliviado		bota de caminhada
	rosto com vapor do nariz		esmalte de unha
	rosto sorridente com olhos sorridentes	—	menos
	rosto radiante com olhos sorridentes	+	mais
	rosto confuso		mulher grávida
	olhos		robô
	corações girando		melancia
	rosto abraçado		mão aberta com os dedos separados
	coração amarelo		pepino
→	seta direita		gotas de suor
	punho levantado		lua crescente
	rosto sorridente com óculos de sol		gato cansado
	cabeça explodindo		baleia jorrando
	Câmera		rosto angustiado
	luz do carro da polícia	✓	marca de seleção
	coração verde		LOGO seta
	rosto com bandagem		dorso da mão levantada
	careta		cara de zíper
	trevo-de-quatro-folhas	!?	pontos de exclamação e interrogação
	pedra preciosa		alvo
	coroa		bíceps
©	direito autoral		pote de mel
	fone de ouvido		limão
	rosto com a mão na boca	!!	ponto de exclamação
	cara quente		jornal enrolado
	tom de pele escuro		cara fria
	cara de beijo		Cara de cavalo
	tulipa		globo mostrando Ásia-Austrália
✗	marca de cruz		círculo vermelho
	cara dormindo		cara saboreando comida
▶	botão play		crânio
	cara de festa		dedos cruzados
	punho que se aproxima		megafone
	rosto com sobrelha levantada		cara de mentira
	coração crescendo		rosto com boca aberta

Emoji	Significado	Emoji	Significado
	celular com seta		microfone
	correndo para longe		alfinete
	rosto sorridente com halo		beijando gato
	coração batendo		bode
	coração cintilante		cem pontos
	marca de beijo		Artes performáticas
	rosa		globo mostrando as Américas
	ponto de interrogação vermelho		seta para baixo
	símbolo médico		anel
	pêssego		yin yang
	cara de medo		ogro
	nota musical		cara babando
	canecas de cerveja tinindo		borboleta
	pop de festa		pontuação musical
	cara com lingua		seringa
	comprimido		tartaruga
	girassol		rosto preocupado
	ambulância		sol
	beijando o rosto com olhos sorridentes		gato chorando
	pessoa inclinando a mão		cara de nocaute
	Símbolo japonês para iniciante		urso
	pessoa se curvando		gesto de amor
	lâmpada elétrica		Botão SOS
	botão de descida rápida		mãos abertas

REPRODUÇÃO DO *Framework* DA PESQUISA

O objetivo desta seção é apresentar as principais etapas para implementar os modelos propostos nesta pesquisa, seguindo o *Framework* exibido na Figura 3, que pode ser dividido nas seguintes etapas:

1. Mineração de dados (mensagens) sobre sinais e sintomas de COVID-19 no Twitter.
2. Extração de dados (casos de COVID-19) no SIVEP-Gripe.
3. Desenvolvimento do Modelo de *Machine Learning* (NLP).
4. Classificação dos *tweets*, com base nos construtos do Modelo de Crença em Saúde.
5. Desenvolvimento do Modelo Epidemiológico Bayesiano.

Para reproduzir esta pesquisa serão necessários dois conjuntos de dados, um conjunto de dados de mensagens de redes sociais (dados do Twitter) sobre sintomas e sinais sobre COVID-19 e outro conjunto de dados de casos de COVID-19 (dados do SIVEP-Gripe). Caso seja de interesse fazer uma nova mineração de dados no Twitter, isso poderá ser feito utilizando o código em Python disponibilizado no repositório desta pesquisa (SANTOS, 2023c), ou utilizando qualquer outro tipo de biblioteca em Python para *web scrapping* ou outra linguagem de programação de interesse; todavia isso não será necessário, visto que os dados coletados estão disponíveis publicamente no repositório Zenodo (SANTOS et al., 2021), inclusive é possível ampliar a pesquisa incorporando os dados já coletados do ano de 2021.

Coletado os dados de redes sociais, o próximo passo é fazer a coleta e tratamento dos dados de casos de COVID-19. Assim, como nos dados do Twitter, para os dados do SIVEP-Gripe os dados também estão tratados e disponíveis publicamente no repositório *Twitter Model Data* no Git-hub (LABCIDADES, 2023a). Todavia, se o interesse for fazer a coleta dos dados de casos de COVID-19 e posterior tratamento, então, será necessário acessar o site openDataSUS (Ministério da Saúde, Secretaria de Vigilância em Saúde, 2023) e baixar os arquivos necessários. Todo o processo de tratamento e limpeza dos dados de casos de COVID-19 utilizado nesta pesquisa foi realizado utilizando o código desenvolvido em Python e está disponível para consulta e utilização (SANTOS, 2023b).

Realizadas as coletas dos dados, a próxima etapa é começar o desenvolvimento dos modelos. Iniciando pelo modelo de processamento de linguagem natural (NLP), para desenvolver esse modelo, será necessário fazer um pré-tratamento nos dados coletados no Twitter. Esse tratamento tem como objetivo filtrar as mensagens expressões que indicam que as pessoas foram ou pretendiam buscar serviços de saúde (ver lista das expressões relacionadas no Quadro 4), para atendimento sobre a COVID-19. Na sequência, foi aplicado um novo tratamento nos dados para normalizar as expressões e emojis de acordo com o "Dicionário de Expressões de Tweets"

desenvolvido nessa pesquisa (ver Apêndices A e B). Essa última etapa não era necessária, contudo foi possível observar que esse tratamento melhorou a acurácia do modelo em um ponto percentual.

Finalizado o pré-tratamento, nesta etapa é necessário recorrer mais uma vez à alguma linguagem de programação, só que desta vez o objetivo é gerar uma amostra aleatória de dados do Twitter, o modelo do código usado para gerar a amostra também encontra-se no Github ([SANTOS, 2023e](#)). Essa amostra será importante para realizar uma classificação manual, onde um pesquisador deverá atribuir valor um se a mensagem será incluída na pesquisa ou zero, caso contrário. No total foram dez mil mensagens classificadas como um ou zero que podem ser acessadas na repositório de dados desta pesquisa ([SANTOS, 2023a](#)).

De posse desses dados classificados em um ou zero, chegou o momento de rodar o modelo de processamento de linguagem natural. Nesta pesquisa foram testados vários modelos, contudo o melhor modelo foi o modelo logístico. O código para reprodução do modelo está disponível em [Santos \(2023f\)](#). Importante mencionar que para executar o código, além do arquivo com os dez mil *tweets*, utilizado para treinar o modelo, será necessário ter a base completa com os noventa e seis mil *tweets*, contidos no arquivo zipado de nome "Tweets96978_com_Descriptions.zip" ([SANTOS, 2023a](#)). Como saída do modelo será gerado um arquivo que irá classificar toda a base de dados de *tweets* em um e zero.

Os dados classificados como um serão utilizados para incorporar o modelo epidemiológico Bayesiano. Porém, antes de incorporar os dados no modelo é necessário transformar os dados em *time series*, para isso, basta utilizar o código disponível em [Santos \(2023d\)](#). Agora sim, todos os recursos necessários para executar o modelo epidemiológico estão disponíveis e preparados. As bases de dados estão prontas e o código do modelo está pronto e disponível para download em [LabCidades \(2023b\)](#). É importante ressaltar que até o momento todas as outras aplicações foram desenvolvidas utilizando o Python; contudo, o modelo epidemiológico é a única aplicação desenvolvida na linguagem *Stan* ([GABRY; ČEŠNOVAR, 2021](#)) e compilada no ambiente do *R-programming* ([TEAM, 2021](#)).

Até o presente momento, tudo que foi executado foi para fazer a predição de casos de COVID-19 e agora, na parte final, será mais uma vez utilizada a base de *tweets*, classificadas inicialmente como sinal ($= 1$); porém o objetivo dessa vez será analisar as percepções das pessoas sobre a pandemia de COVID-19. Então, recorrendo novamente para a base com dez mil *tweets* deve-se analisar quais os construtos do Modelo de Crença em Saúde estão presentes em cada mensagem, lembrando que em uma única mensagem é possível encontrar mais de um construto, logo a classificação pode ser múltipla. Esta é uma etapa mais subjetiva e dependendo de quem classificar pode gerar resultados diferentes, mas reduzir a subjetividade nesta pesquisa foram criados alguns critérios de classificação:

- **Ameaça não percebida:** Quando a ameaça não é percebida. Pode ser em relação à doença ou tratamento/medicação; por exemplo, quando não percebe a ineficácia do "kit covid". Não há necessidade de ser o doente, por exemplo a pessoa pode ter todos os sintomas, mas

o médico achar que não é COVID-19 (nesse caso o médico não percebeu). Na ausência de um sintoma a pessoa pode considerar improvável ter a doença. Quando a pessoa não percebe que outros podem ter dificuldade em seguir uma recomendação, por exemplo, usar máscara por ser claustrofóbica (não confundir com eficácia percebida que é quando a pessoa declara que não consegue seguir um tratamento, por exemplo, por ser claustrofóbica). Quando a pessoa menciona que deve ter pego COVID-19 em um período anterior ao primeiro caso registrado, por exemplo, quando fala que deve ter pegado na infância, pois os sintomas eram parecidos; isso indica que a pessoa não entendeu o que é o novo coronavírus.

- **Suscetibilidade percebida:** Quando há na mensagem algum sinal ou sintoma de COVID-19 e por causa disso a pessoa (ou terceiro) indica que vai tomar uma ação preventiva, ou, por exemplo, o teste deu positivo. Contudo, nada na mensagem indica que a pessoa (ou terceiros) acha ser grave.
- **Gravidade percebida:** Quando a mensagem indica que a pessoa (a mesma ou terceiros) acredita que a COVID-19 é grave. Se a pessoa pegou a COVID-19 e acha que pode evoluir e se tornar grave. Analisar também emojis de sentimentos de gravidade alta e se no texto tem expressões como: "corre por hospital"; "Jesus...". Se a pessoa relata que foi várias vezes no médico considere que ela acredita que é grave e que possivelmente os sintomas estão piorando. Se a pessoa relata morte por COVID-19. Se a pessoa fala que apesar de não ter ficado internada passou o dia todo ou a noite toda no hospital, em observação etc..
- **Benefícios percebidos:** Quando a pessoa acredita que a intervenção médica que ela (ou terceiro) passou (ou pode passar) foi positiva.
- **Barreiras percebidas:** Quando a pessoa acredita que a intervenção médica tem mais aspectos negativos do que positivo. Ou ainda, quando não tomou uma ação por fatores externos, por exemplo, hospital que só atende emergência. Se não consegue um diagnóstico para saber se teve COVID-19. Ou quando comenta que o fato das pessoas não respeitarem os protocolos pode prejudicar. Ou quando a pessoa manifesta que há problema em falar para outros que está com COVID-19 (medo do preconceito da doença). Quando não é possível identificar o vírus, por exemplo quando a pessoa tem pneumonia (que segundo os médicos pode ser de ocorrência do coronavírus), mas o exame deu negativo. Ter a doença e não poder ver os entes queridos. Quando a pessoa indica que foi varias vezes no médico e não adiantou etc..
- **Dicas para a ação:** Quando a mensagem é sobre conselhos, sugestões, recomendações de tratamento (ir ao médico) e orientações relacionadas à COVID-19. Se a pessoa diz que conversou com outras pessoas para saber mais sobre a doença ou como agir na pandemia. Ou quando a pessoa coloca de forma explícita seu caso para servir de exemplo e

orientação para pessoas na mesma situação. *Auto-eficácia*: Crença de uma pessoa sobre se ela mesmo é ou não capaz de tomar a ação preventiva. Aqui deve ser considerado se a pessoa manifestar que tomou todos os cuidados recomendados (isolamento social, não aglomerar, etc) ou se não conseguiria seguir as recomendações (se aplica a terceiros). Também deve ser considerado se a pessoa mencionou um tratamento que deve seguir (independente de mencionar a chance de seguir ou não - isso será analisado posteriormente). *Outras variáveis*: Influências por variáveis externas relacionadas à COVID-19: política, crenças religiosas, outros aspectos culturais etc..

- **COVID-19**: se a pessoa falar que o exame deu positivo para COVID-19 ou se pelo exame clínico o médico falar que está com COVID-19. Caso não esteja explícito no texto que está com a doença, mas se falar que o médico mandou ficar em quarenta, então deve se admitir que está com COVID.
- **Ruído**: *Tweets* não relacionados a pesquisa. Exemplo: quando alguém menciona sintomas de animais. Ou menciona um evento que aconteceu antes da pandemia (2019, por exemplo).

Utilizando esse guia para classificar as mensagens será possível identificar quais construtos do Modelo de Crença em Saúde foram predominantes durante a pandemia de COVID-19, e, posteriormente, identificar os fatores que motivam as pessoas a procurarem os serviços de saúde. Desta forma, seguindo todas as etapas descritas chega-se a etapa final do *Framework* da pesquisa e todas as principais etapas e respectivos resultados serão reproduzidos.

Oitenta e três teorias de mudança de comportamento de acordo com Michie (2014):

1. Modelo de Teoria da Ação de Consumo (BAGOZZI, 2000).
2. Teoria dos Eventos Afetivos (WEISS; CROPANZANO, 1996).
3. Modelo de Redução de Risco de AIDS (CATANIA; KEGELES; COATES, 1990).
4. Modelo Ecológico-Comportamental de Prevenção da Aids em Adolescentes (HOVELL et al., 1994).
5. Teoria dos CEOS (BORLAND, 2014; BORLAND, 2017).
6. Teoria da Mudança (LEWIN, 1947; LEWIN, 1952; LEWIN, 1958).
7. Condicionamento Clássico (PAVLOV, 1927).
8. Sistema COM-B (MICHIE; STRALEN; WEST, 2011).
9. Consumo como Práticas Sociais (SPAARGAREN; VLIET, 2000; SPAARGAREN, 2000).
10. Teoria de Contenção (RECKLESS, 1961).
11. Teoria de Controle (CARVER; SCHEIER, 1982).
12. Teoria da Associação Diferencial (SUTHERLAND, 1947; SUTHERLAND; CRESSEY; LUCKENBILL, 1992).
13. Difusão de Inovações (ROGERS, 1983; ROGERS, 2003; ROGERS; SINGHAL; QUINLAN, 2019).
14. Modelo Ecológico para a Prevenção do Diabetes Tipo 2 em Jovens Minoritários (BURNET et al., 2002).
15. Modelo de Processamento de Informação Estendida (FLAY; DITECCO; SCHLEGEL, 1980).
16. Modelo de Processamento Paralelo Estendido (WITTE, 1992).
17. Teoria de Intervenção de Feedback (KLUGER; DENISI, 1996).
18. Teoria do Foco de Conduta Normativa (CIALDINI; KALLGREN; RENO, 1991; CIALDINI, 2012; STOK; RIDDER, 2019).
19. Teoria Geral do Crime (GOTTFREDSON; HIRSCHI, 1990).

20. Teoria Geral do Comportamento Desviante (KAPLAN, 1972; KAPLAN, 1982; KAPLAN; MARTIN; ROBBINS, 1982).
21. Teoria do Objetivo Dirigido (BAGOZZI; WARSHAW, 1992).
22. Teoria de Definição de Metas (LINDENBERG; STEG, 2007).
23. Teoria do Estabelecimento de Metas (LOCKE, 1968; LOCKE; LATHAM, 2002).
24. Abordagem do Processo de Ação em Saúde (SCHWARZER, 1992; SCHWARZER, 2008; SCHWARZER, 2014).
25. Modelo de Metas de Comportamento de Saúde (MAES; GEBHARDT, 2000).
26. Modelo de Internalização de Comportamento de Saúde (BELLG, 2003).
27. Modelo de Crenças de Saúde (ROSENSTOCK, 1974b; ROSENSTOCK; STRECHER; BECKER, 1988).
28. Modelo de Promoção da Saúde (PENDER et al., 1988).
29. Modelo de Mudança I (VRIES; DIJKSTRA; KUHLMAN, 1988; VRIES et al., 2005; VRIES; MUDDE; DIJKSTRA, 2013)
30. Modelo de Habilidades de Informação-Motivação-Comportamento (FISHER; FISHER, 1992).
31. Modelo de adesão informação-motivação-habilidades comportamentais (FISHER et al., 2006).
32. Modelo Teórico Integrado para a Prevenção do Abuso de Álcool e Outras Drogas (GONZALEZ, 1989).
33. Teoria Integrada do Comportamento de Beber (WAGENAAR; PERRY, 1994).
34. Teoria Integrada de Mudança de Comportamento em Saúde (RYAN, 2009).
35. Modelo Integrativo de Predição Comportamental (FISHBEIN, 2000).
36. Modelo Integrativo de Fatores que Influenciam os Comportamentos de Fumar (FLAY et al., 1983).
37. Modelo Integrativo de Mudança de Atitude e Comportamento em Saúde (FLAY, 1981).
38. Modelo Integrativo de Fatores que Influenciam o Tabagismo e a Mudança de Atitude e Comportamento de Saúde (FLAY et al., 1983).
39. Modelo de Comportamento Pró-Ambiental (KOLLMUSS; AGYEMAN, 2002).

40. Modelo de Motivação-Oportunidades-Habilidades (ÖLANDER; THØGERSEN, 1995).
41. Modelo de Necessidades-Oportunidades-Habilidades (VLEK, 1998).
42. Teoria de Ativação de Normas (SCHWARTZ, 1975).
43. Teoria de Aprendizagem Operante (SKINNER, 1953).
44. Modelo de Processo de Adoção de Precaução (WEINSTEIN; SANDMAN, 1992; WEINSTEIN; SANDMAN; BLALOCK, 2008).
45. Modelo de Sistema de Pressão (KATZ, 2001).
46. Teoria PRIME (WEST; BROWN, 2014).
47. Teoria do Comportamento Problemático (JESSOR, 1987; JESSOR, 1991).
48. Teoria da Perspectiva (KAHNEMAN; TVERSKY, 1979).
49. Teoria da Motivação de Proteção (ROGERS, 1975; ROGERS, 1983).
50. Modelo de Protótipo de Vontade (GERRARD et al., 2008).
51. Modelo Racional de Vício (BECKER; MURPHY, 1988).
52. Modelo Impulsivo Reflexivo (STRACK; DEUTSCH, 2004).
53. Teoria de Ajuste Regulatório (HIGGINS, 2000; HIGGINS, 2005).
54. Modelo de Prevenção de Recaída (MARLATT; GEORGE, 1984).
55. Risco como Teoria de Sentimentos (LOEWENSTEIN et al., 2001).
56. Teoria de Autodeterminação (DECI; RYAN, 1985; DECI; RYAN, 2008; RYAN, 2009).
57. Teoria da Autoeficácia (BANDURA, 1978).
58. Teoria da Autorregulação (KANFER; GAELICK, 1991).
59. Seis Modelos de Efeitos de Comunicação (VAUGHAN; ROGERS, 2000).
60. Teoria da Ação Social (EWART, 1991).
61. Teoria da Ação Social (WEBER; ROTH; WITTICH, 1978; WEBER; TRIBE, 2019).
62. Teoria da Mudança Social (THOMPSON; KINNE, 1999).
63. Teoria Social Cognitiva (BANDURA, 1986).
64. Modelo de Consenso Social de Educação em Saúde (ROMER; HORNIK, 1992).

65. Modelo de Desenvolvimento Social (HAWKINS; WEIS, 1985).
66. Modelo Social Ecológico de Mudança de Comportamento (PANTER-BRICK et al., 2006).
67. Modelo Social Ecológico de Caminhada (ALFONZO, 2005).
68. Teoria da Identidade Social (TAJFEL; TURNER, 1974).
69. Modelo de Influência Social de Participação do Consumidor (DHOLAKIA; BAGOZZI; PEARO, 2004).
70. Teoria da Aprendizagem Social (MILLER; DOLLARD, 1962).
71. Teoria das Normas Sociais (BIBLIOGRAPHY; BERKOWITZ, 2004).
72. Modelo de sistemas de Mudança de Comportamento em Saúde (KERSELL; MILSUM, 1985).
73. Modelo de Aceitação de Tecnologia 1, 2 e 3 (DAVIS, 1989; VENKATESH; DAVIS, 2000; VENKATESH; BALA, 2008).
74. Teoria da Autorregulação Temporal (HALL; FONG, 2007).
75. Teoria de Gestão do Terror (GREENBERG; PYSZCZYNSKI; SOLOMON, 1986).
76. Modelo de Saúde de Gestão do Terrorismo (GOLDENBERG; ARNDT, 2008).
77. Teoria do Comportamento Interpessoal (TRIANDIS, 1977).
78. Teoria do comportamento Social Normativo (RIMAL; REAL, 2005).
79. Teoria do Comportamento Planejado (AJZEN, 1985; AJZEN, 1991).
80. Teoria da Influência Triádica (FLAY; PETRAITIS, 1994).
81. Modelo Transcontextual de Motivação (HAGGER et al., 2003; HAGGER; CHATZISARANTIS, 2012).
82. Modelo Transteórico de Mudança de Comportamento (PROCHASKA; DICLEMENTE, 1982; PROCHASKA; DICLEMENTE; NORCROSS, 1992; PROCHASKA et al., 1994).
83. Teoria da Norma da Crença de Valor (STERN et al., 1999).