

Uso de Python para detecção de *fake news* sobre a covid-19: desafios e possibilidades

Using Python to detect fake news about covid-19: challenges and possibilities

Usar Python para detectar notícias falsas sobre covid-19: desafios y posibilidades

Fernanda Vasques Ferreira^{1,2,a}

fernanda.jornalista82@gmail.com | <https://orcid.org/0000-0003-4242-0057>

Rafiza Varão^{3,a}

rafiza@gmail.com | <https://orcid.org/0000-0003-0383-5524>

Marco Aurelio Boselli^{4,b}

maboselli@gmail.com | <https://orcid.org/0000-0002-1030-1144>

Leandro Brito Santos^{5,c}

lbsantos@ufob.edu.br | <https://orcid.org/0000-0003-0132-4712>

Marcelo A. Moret^{1,6,d}

mamoret@gmail.com | <https://orcid.org/0000-0003-2624-5754>

¹ Centro Universitário Senai Cimatec, Faculdade de Tecnologia. Salvador, BA, Brasil.

² Universidade Federal do Oeste da Bahia, Centro Multidisciplinar de Santa Maria da Vitória. Santa Maria da Vitória, BA, Brasil.

³ Universidade de Brasília, Faculdade de Comunicação. Brasília, DF, Brasil.

⁴ Universidade Federal de Uberlândia, Instituto de Física. Uberlândia, MG, Brasil.

⁵ Universidade Federal do Oeste da Bahia, Centro Multidisciplinar de Bom Jesus da Lapa. Bom Jesus da Lapa, BA, Brasil.

⁶ Universidade do Estado da Bahia. Salvador, BA, Brasil.

^a Doutorado em Comunicação pela Universidade de Brasília.

^b Doutorado em Física pela Universidade Estadual de Campinas.

^c Doutorado em Modelagem Computacional pelo Senai Cimatec.

^d Doutorado em Ciências Biológicas (Biofísica) pela Universidade Federal do Rio de Janeiro.

RESUMO

Este trabalho tem como objetivo relatar estratégias para coleta de um conjunto de dados em português para treinamento de modelos de Inteligência Artificial com vistas a identificar de forma automática *fake news* sobre covid-19 disseminadas durante a pandemia, a partir de código Python. Analisamos um método de detecção de *fake news* baseado em uma Rede Neural Recorrente e de aprendizagem supervisionada. Selecionamos um corpus com 7,2 mil textos coletados em *websites* e agências de notícias por Monteiro *et al.* (2018) com cada um previamente catalogado como verdadeiro ou falso como conjunto de dados de treino e validação. O modelo foi usado para detecção de *fake news* sobre covid-19 em um conjunto de notícias coletadas e classificadas pelos autores deste trabalho. O índice de acerto foi de 70%, ou seja, essa foi a taxa de sucesso da detecção dos itens catalogados.

Palavras-chave: *Fake news*; Inteligência artificial; Covid-19; Python; Desinformação.

ABSTRACT

This work aims to report strategies for collecting a dataset in Portuguese for training Artificial Intelligence models to automatically identify fake news about covid-19 disseminated during the pandemic, using Python code. We analyze a fake news detection method based on a Recurrent Neural Network and supervised learning. We selected a corpus with 7,200 texts collected on websites and news agencies by Monteiro *et al.* (2018), each one of them previously cataloged as true or false as a training and validation dataset. This model was used to detect fake news about covid-19 in a set of news collected and classified by the authors of this work. The hit rate was 70%.

Keywords: Fake news; Artificial Intelligence; Covid-19; Python; Misinformation.

RESUMEN

Este trabajo tiene como objetivo informar estrategias para recopilar un conjunto de datos en portugués para entrenar modelos de Inteligencia Artificial para identificar automáticamente noticias falsas sobre covid-19 difundidas durante la pandemia, utilizando el código Python. Analizamos un método de detección de noticias falsas basado en una Red Neuronal Recurrente y de aprendizaje supervisado. Seleccionamos un corpus de 7.200 textos recogidos en webs y agencias de noticias por Monteiro *et al.* (2018) con cada uno catalogado previamente como verdadero o falso como un conjunto de datos de entrenamiento y validación. El modelo se utilizó para detectar noticias falsas sobre covid-19 en un conjunto de noticias recopiladas y clasificadas por los autores de este trabajo. La tasa de acierto fue del 70%, es decir, esta fue la tasa de éxito de detección de los artículos catalogados.

Palabras clave: Noticias falsas; Inteligencia artificial; Covid-19; Python; Desinformación.

INFORMAÇÕES DO ARTIGO

Este artigo compõe o dossiê *Perspectivas multidisciplinares sobre desinformação em ciência e saúde.*

Contribuição dos autores:

Concepção e desenho do estudo: Fernanda Vasques Ferreira, Leandro Brito Santos, Rafiza Varão, Marcelo Moret e Marco Aurélio Boselli.

Aquisição, análise ou interpretação dos dados: Fernanda Vasques Ferreira e Marco Aurélio Boselli.

Redação do manuscrito: Fernanda Vasques Ferreira, Leandro Brito Santos, Rafiza Varão, Marcelo Moret e Marco Aurélio Boselli.

Revisão crítica do conteúdo intelectual: Marcelo Moret, Rafiza Varão, Fernanda Vasques Ferreira, Marco Aurélio Boselli.

Declaração de conflito de interesses: não há.

Fontes de financiamento: não houve.

Considerações éticas: não há.

Agradecimentos/Contribuições adicionais: não há.

Histórico do artigo: submetido: 03 mar. 2022 | aceito: 02 jun. 2022 | publicado: 30 jun. 2022.

Apresentação anterior: não houve.

Licença CC BY-NC atribuição não comercial. Com essa licença é permitido acessar, baixar (*download*), copiar, imprimir, compartilhar, reutilizar e distribuir os artigos, desde que para uso não comercial e com a citação da fonte, conferindo os devidos créditos de autoria e menção à Reciis. Nesses casos, nenhuma permissão é necessária por parte dos autores ou dos editores.

INTRODUÇÃO

A disseminação de *fake news* nas mídias sociais, bem como o fenômeno mais amplo da desinformação, se tornou uma constante desde o início da segunda década do século XXI, ao apresentar desafios e dilemas importantes que extrapolam as questões de âmbito comunicacional e interferem na dinâmica social. Segundo o Dicionário Collins (2022), *fake news* são informações falsas, geralmente sensacionalistas, difundidas sob o disfarce de notícias. Essa definição, entretanto, é insuficiente para tratar da temática. Segundo Volkoff (2000), as *fake news* se valem de técnicas que permitem produzir e disseminar informações geralmente enganosas que tendem a organizar ações coletivas ou difundir juízos desejados pelos desinformados. A noção de falseamento, ou deturpação da realidade, é importante nesse contexto.

No Brasil, durante o primeiro ano da pandemia da covid-19, uma das mais impactantes notícias falsas se referia a caixões vazios em Manaus, capital do Amazonas, um dos maiores polos de contágio da doença, que vivenciou duas ondas de crescimento: no início e no final de 2020. A informação falsa disseminada tinha como objetivo desmerecer a importância da pandemia e negar as milhares de mortes naquele estado (PENNAFORT, 2020).

De acordo com os dados do relatório Digital 2021 (Global Overview Report), fruto de uma parceria entre as organizações We Are Social e Hootsuite (AMBER ENERGIA HUMANA, 2022), 75% da população brasileira usa internet, o que corresponde a mais de 158 milhões de brasileiros conectados à rede em 2019 e, nesse mesmo ano, o tempo médio dos brasileiros na internet passava de 9 horas por dia. O mesmo relatório apontou que, em 2021, o tempo médio de permanência diária passou de 10 horas. Os brasileiros passaram, em média, 3 horas 42 minutos por dia em mídias sociais, sendo o Brasil o país que ocupa o terceiro lugar no *ranking* de tempo dedicado a elas. A pesquisa também indicou que a busca por informações é a principal razão para as pessoas acessarem a internet e que, nesse cenário, as *fake news* preocupam bastante os brasileiros.

Sobre esse tema, uma pesquisa feita pela Avaaz, realizada em 2020, apontou que a cada dez brasileiros, sete já acreditaram em, pelo menos, uma *fake news* sobre a covid-19. Uma evidência disso é a pesquisa Ibope, realizada em 2020, que revelou a probabilidade de que 25% da população não se vacinasse contra a covid-19. No percentual dos que disseram que não iriam tomar a vacina de ‘jeito nenhum’ e os que não tinham certeza, 34% declararam pelo menos um motivo relacionado à desinformação.

Se fora da situação de emergência sanitária dos últimos anos as *fake news* são perniciosas para as práticas de prevenção e difusão da informação sobre saúde, em um cenário pandêmico, em que uma série de medidas de combate ao coronavírus deve ser compreendida como informação de interesse público, a situação se agrava ainda mais. Indivíduos fisicamente isolados, impedidos do convívio social e, por vezes, com laços fraternos fragilizados tendem a ser alvo de conteúdos maliciosos que visam desinformar, confundir, embaralhar e distorcer a realidade.

É nesse sentido que algumas teorias da comunicação – consideradas ultrapassadas – passaram a ser reconsideradas por pesquisadores, tais como: a teoria hipodérmica e a abordagem empírico-experimental. A primeira trabalha com a ideia de manipulação, a partir da psicologia behaviorista segundo a qual, por meio da repetição de mensagens, é possível manipular indivíduos isolados e com laços fraternos fragilizados; a segunda, por sua vez, trata da percepção seletiva, nas quais os indivíduos elegem para si temas e perspectivas que são próximos de suas realidades, e, desse modo, as mensagens passam a exercer um efeito persuasivo.

Em meados da década de 1950, o sociólogo Paul Lazarsfeld afirmava que quando em contato com os meios de comunicação a audiência busca consumir aquilo que reforça opiniões, rejeitando o contraditório. Nas mídias sociais, com a livre publicação, essa atitude é reforçada não pela simples rejeição a conteúdos, mas pela formulação de novos, que defendem a posição de quem emite a mensagem – o que anteriormente

cabia mais amiúde aos jornalistas de opinião e, num aspecto macro, às linhas editoriais dos veículos (VARÃO, 2017).

De acordo com Gomes (2020), as *fake news* são usadas para descredibilizar a mídia de referência e se proliferam, exponencialmente, em razão de processos de dissonância cognitiva, ou seja, de uma aversão a entrar em contato com informações que não têm consonância com crenças e valores prévios. Em trabalho publicado em 2019, Pennycook e Rand (2019) identificaram que há mais suscetibilidade às notícias falsas por conta de um comportamento preguiçoso do que por preconceito partidário em si. Associado a essa descoberta, o trabalho evidencia o que Gomes (2020) nos explica: quando as notícias falsas vão ao encontro da ideologia política ou das crenças dos indivíduos, há um terreno fértil para a circulação, a disseminação e o consumo de informações duvidosas. Acrescenta-se a isso o que Sousa Júnior, Petroll e Rocha (2019), assim como Mello (2020), indicaram sobre o aspecto financeiro que envolve as notícias falsas. Há um modelo de negócio rentável, a partir do sensacionalismo que estimula o ódio nas redes digitais e impulsiona as *fake news*.

Diante desse contexto social complexo inerente à desinformação em saúde, principalmente, com a emergência sanitária acarretada pela covid-19, este trabalho tem como objetivo apresentar estratégias para detecção de *fake news*, com vistas a indicar caminhos e possibilidades para tratar o problema de forma mais sistemática em textos publicados em língua portuguesa (circunscritos ao Brasil), levando em consideração o uso de Inteligência Artificial a partir de codificação em Python.¹ Para isso, nosso trabalho relata as estratégias para coleta de um conjunto de dados em português para treinamento de modelo com uso de Inteligência Artificial para a detecção de *fake news* sobre a covid-19. Desse modo, a pergunta que norteia este artigo é: de que modo a Inteligência Artificial e o uso da modelagem computacional podem contribuir para a detecção de *fake news* sobre a covid-19 no Brasil?

Para realizar o trabalho, analisamos um método de detecção de *fake news* baseado em uma Rede Neural Recorrente (RNR), cujo nome faz referência ao fato de essa rede guardar informações dos passos anteriores ao passo em execução. O treinamento do modelo foi feito por aprendizagem supervisionada, ocasião em que os parâmetros são ajustados a um conjunto de dados tabulados. Usamos para esse conjunto um *corpus* com 7,2 mil textos² com cada um previamente catalogado como verdadeiro ou falso, produzido por Monteiro *et al.* (2018) como conjunto de dados de treino e validação. Esse conjunto de textos foi coletado a partir dos websites Diário do Brasil, Folha do Brasil, The Jornal Brasil³ e Top Five TV como fontes de notícias falsas, e dos sites noticiosos G1, Folha de S.Paulo e O Estado de S. Paulo como fontes de notícias verdadeiras. A estratégia para montar esse *corpus* foi a coleta automática de textos de jornais de referência, como G1, Folha de S.Paulo e O Estado de S. Paulo, cujos textos foram considerados verdadeiros. A partir dessa coleta, foram selecionadas palavras-chave desses textos, e com elas foi feita uma busca – também por *web scraping* – nos sites propagadores de *fake news* ou sites com baixa reputação, como Diário do Brasil, Folha do Brasil, The Jornal Brasil e Top Five TV, contendo as mesmas palavras-chaves. Com isso, temos nesse *corpus* uma contraposição grande de textos verdadeiros *versus* textos falsos sobre o mesmo assunto.

O modelo foi usado para detecção de *fake news* em um conjunto de notícias coletadas e classificadas pelos autores deste artigo. O índice de acerto foi de 70%, ou seja, essa foi a taxa de sucesso da detecção dos itens catalogados e identificados como falsos. Por índice de acerto entendemos a capacidade de o modelo

1 Python é uma linguagem de programação criada em 1991, interpretada por *script* e acessível, com sintaxe clara e concisa que favorece a legibilidade de código-fonte, tornando a linguagem mais produtiva. É a linguagem de programação mais utilizada atualmente no mundo. De código aberto, Python é um *software* que permite automatizar tarefas.

2 Esse *corpus* se refere a conteúdos verdadeiros e falsos de diferentes temáticas e não relativos à Covid-19.

3 Site ativo à época do trabalho de Moreira *et al.* (2020), *link* indisponível em 2022.

computacional desenvolvido em Python reconhecer os conteúdos verdadeiros como verdadeiros e os falsos como falsos. O percurso teórico-metodológico será detalhado ao longo deste artigo.

Desse modo, o trabalho busca ampliar a capacidade de detecção de *fake news*, uma vez que observamos que o conjunto de bibliotecas e programas para processamento simbólico e estatístico da linguagem natural já está bastante desenvolvido para a língua inglesa, mas exige um esforço de pesquisa maior para a língua portuguesa (Brasil), apresentando, assim, uma lacuna que deve ser objeto de investigação.

No geral, os conjuntos de dados em inglês têm mais conteúdo do que os portugueses, com sentenças consideravelmente mais longas. Este fato indica que os conjuntos de dados portugueses podem ser menos beneficiados por modelos mais complexos do que os conjuntos de dados em inglês.⁴ (SOUZA; SOUZA FILHO, 2021, p. 3, tradução nossa)

Essa dificuldade é apontada em trabalhos como *Sentiment analysis on Brazilian Portuguese user reviews* (SOUZA; SOUZA FILHO, 2021) e não é uma peculiaridade da Língua portuguesa, mas uma restrição comum às línguas não tão centrais como o inglês. Rafał Wójcik (2019), em artigo no *Towards Data Science*, um *site* voltado para programadores, aponta dificuldade semelhante na língua polonesa.

Apesar dos avanços recentes na NLP [Processamento de Linguagem Natural], a língua portuguesa dispõe de recursos linguísticos limitados. Como aponta o artigo de Pereira (2020), vários pesquisadores vêm coletando conjuntos de dados que descrevem uma série de problemas de classificação de textos em português brasileiro. No entanto, trabalhos anteriores que visam à classificação de textos avaliam uma ampla variedade de conjuntos de dados ou consideram diferentes alvos a serem previstos. Além disso, mesmo nos casos em que a maioria desses aspectos coincide, os subconjuntos considerados para avaliação do modelo muitas vezes diferem, tornando injusta uma comparação direta dos resultados correspondentes.⁵ (SOUZA; SOUZA FILHO, 2021, p. 2, tradução nossa)

Por meio do estudo de métodos de detecção de *fake news*, e também construindo um conjunto de dados classificados como ‘verdade’ ou ‘falso’, pretende-se indicar possibilidades efetivas para que grandes corporações de mídias, redes sociais e organizações engajadas na luta contra a desinformação possam atuar com capacidade resolutiva sobre o fenômeno.

FAKE NEWS COMO PROBLEMA SOCIAL E COMPLEXO

De acordo com a pesquisa *Desinformação, mídia social e covid-19 no Brasil: relatório, resultados e estratégias de combate* (RECUERO *et al.*, 2021), o fenômeno complexo das *fake news* no Brasil não pode ser analisado sem se considerar o contexto político brasileiro. Assim, o relatório aponta que a pandemia é percebida nas mídias sociais como um assunto político-partidário, mais do que como um assunto de saúde pública. Sendo assim, a discussão sobre como mitigar os efeitos da pandemia se polariza e é confundida com filiação partidária, o que é uma visão equivocada. A pesquisa identificou ainda que a desinformação contribui para sustentar discursos políticos, principalmente os de apoio ao atual presidente da República, Jair Bolsonaro.

Em certa medida, esse contexto de desinformação nas redes sociais se dá pela formação de ‘bolhas’ informacionais alimentadas e com papel central desempenhado por “líderes políticos que produzem

4 Overall, the English datasets have richer content than the Portuguese ones, with considerably longer sentences. This fact indicates that the Portuguese datasets may be less benefited from more complex models than English. (SOUZA; SOUZA FILHO, 2021, p. 3)

5 Despite the recent advances in NLP, the Portuguese language disposes of only limited linguistic resources. As Pereira’s paper points out, several researchers have been collecting datasets describing a range of text classification problems in Brazilian Portuguese. However, previous works envisaging text classification evaluate a wide variety of datasets or considers different targets to be predicted. Moreover, even in the cases where most of these aspects match, the subsets considered for model evaluation often differ, making a straight comparison of the corresponding results unfair. (SOUZA; SOUZA FILHO, 2021, p. 2). O texto mencionado pelos autores é Pereira, 2020.

conteúdo sobre a pandemia a partir de um enquadramento ideológico, e ativistas e outros usuários muito ativos, que compartilham apenas conteúdo alinhado a seu posicionamento político” (RECUERO; SOARES; GRUZD, 2020, p. 13).

O mesmo relatório aponta que, em comparação, a checagem de informação circula menos e tem dificuldade para entrar nas ‘bolhas’ de desinformação, já que “grupos politicamente radicais ou altamente ideologizados tendem a proteger as suas crenças através da negação de fatos ou de evidências contrárias” (RECUERO; SOARES; GRUZD, 2020, p. 17). Além disso, os indivíduos que integram as ‘bolhas’ de desinformação são mais ativos e engajados no compartilhamento da desinformação. Esse último aspecto suscita três questões fundamentais: 1. aspectos econômicos que financiam a desinformação; 2. atores mais engajados e envolvidos na propagação da desinformação; e 3. o uso de *bots* para escalonar essa propagação. Essa última, em especial, é uma importante constatação que interfere em nossa pesquisa.

É importante retomar o que Gomes (2020)⁶ nos explica acerca do fenômeno das *fake news*. Segundo o pesquisador, o fenômeno tal qual conhecemos desde 2016 tem duas causas fundamentais: o avanço sem precedentes da direita conservadora digital e a hiperpolitização. Segundo Silva e Araújo (2019), as mídias sociais contribuíram para o crescimento exponencial da hiperpolítica, e, de acordo com Gomes (2020), da mesma forma, as *fake news* também cresceram nesse ambiente de forma exponencial. A hiperpolítica, conceito cunhado pelo filósofo alemão Peter Sloterdijk (1999) sugere que esta é, na verdade, a aversão à política, à capacidade de diálogo e de convivência. Conforme Silva e Araújo (2019, p. 140), a hiperpolítica corresponde a “uma dificuldade inédita e imensa de se conviver com o outro e de se aceitar o diferente”. Nesse contexto:

O sucesso dos algoritmos, e consequentemente das plataformas, se dá através de um “encantamento de fontes duvidosas”. As fontes de informação não são investigadas, não são rebatidas. Os usuários, em um processo natural, replicam as informações recebidas porque fazem sentido a eles, porque se assemelham a sua forma de enxergar o mundo. São informações calculadas com o auxílio dos algoritmos, que por chegarem de forma sedutora aos usuários adquirem um grande potencial de viralização. Num cenário assim, viver juntos se torna compartilhar fontes duvidosas, receber aprovação e se unir em hordas. (SILVA; ARAÚJO, 2019, p. 144)

A reflexão proposta por Silva e Araújo (2019) parece evidenciar o aspecto fraudulento desse ambiente fértil de produção, disseminação e circulação de *fake news*. Para soar mais crível, Gomes (2020) nos explica que as *fake news* forjam o jornalismo de referência, dizem respeito ao fato inventado, ao fato que aconteceu, mas foi alterado e, portanto, constitui-se como uma falsificação, podendo, em algumas ocasiões, ser fenômeno do mau jornalismo.

Esse fenômeno não se restringe ao campo político, mas atinge todas as esferas da vida social e, a partir de 2020, teve grande impacto nas questões relativas à saúde pública, preocupando autoridades.⁷ Desse modo, para solucionar o impacto das *fake news* como problema social complexo – ou ao menos para minimizar tais impactos na saúde pública –, é preciso haver uma ação conjunta de diferentes setores do Estado e da sociedade. Combatê-las exige que diferentes esferas do poder atuem de forma dialogada e multidisciplinar, conforme esclarece Ferreira (2018). As *fake news* constituem um tema que pode fornecer compreensões a partir de múltiplos campos do conhecimento e indicar a necessidade de uma atuação igualmente complexa para a redução de danos ou a possível solução do problema.

Diante da complexidade do fenômeno e do contexto de difícil solução em que se circunscrevem as *fake news* – bem como da multiplicidade de fatores intervenientes na disseminação desse tipo de conteúdo,

6 Fala do professor Wilson Gomes durante o curso Entendendo as *fake news*, março de 2020.

7 A Comissão Parlamentar de Inquérito da Pandemia, instaurada no Senado brasileiro, é um exemplo de iniciativa que também teve as *fake news* como objeto de preocupação de um dos grupos de trabalho.

incluindo aspectos comportamentais, econômicos, político-partidários e computacionais que envolvem a utilização de algoritmos e *bots* a partir da implementação de conhecimentos em Inteligência Artificial –, este trabalho percorre uma via que alinha os conhecimentos em Inteligência Artificial e aprendizagem de máquina para detecção de *fake news* sem, com isso, eximir-se de ponderações cautelosas sobre esse problema social complexo (BRUCK, 2015) que envolvem questões que vão além dos conhecimentos em computação.

DETECÇÃO DE FAKE NEWS: UM CAMINHO POSSÍVEL

A rapidez e a velocidade de produção, bem como a disseminação e o alcance por conta do fluxo das redes digitais, a capilaridade, fizeram com que as notícias falsas alcançassem lugares onde a informação profissional, verificada e checada tem dificuldade de chegar. As *fake news* foram distribuídas, expondo inadvertidamente o público com a redução ou a ausência de filtros e, por fim, fazendo uso de Inteligência Artificial (HASSON, 2021; GOMES, 2020). A automatização e a criação de *bots* para disseminação, reprodução e compartilhamento de *fake news*, sobretudo na área de saúde – em especial na pandemia da covid-19 –, é motivo de preocupação. Elas podem gerar desinformação, manipulação e confusão, impactando a realidade social, ao distorcer a compreensão sobre a covid-19 e sobre as medidas preventivas. Por conta disso, coloca-se em risco a vida e a saúde dos cidadãos, principalmente quando os *bots* que disseminam desinformação se somam aos algoritmos das mídias sociais que formam as ‘bolhas informacionais’.

Pesquisa apresentada pelo Imperva – Bad bot report 2021: the pandemic of the internet – (HASSAN, 2021) apontou que, em 2021, houve um aumento de mais de 6% no tráfego de *bots* de desinformação na internet em todo o mundo em relação a 2020. Isso significa que, de toda a informação disponível na internet, 1/4 é de responsabilidade dos *bad bots*, que disseminam informações falsas. A junção entre os *bots* que disseminam desinformação e os algoritmos que identificam e formam ‘bolhas informacionais’ potencializa os impactos das *fake news*, principalmente no campo da saúde, afetando a vida das pessoas.

Inteligência Artificial é um nome amplo e mais difundido popularmente para definir um conceito de programação que nasceu em 1950 e que coloca a questão sobre em quais situações os computadores poderiam ‘pensar’ (CHOLLET, 2018). Nessa definição ampla, de maneira mais contemporânea, trabalha-se com conceitos mais restritos, como aprendizagem de máquina e aprendizagem profunda.

Podemos definir Inteligência Artificial como a capacidade de um sistema realizar tarefas características de seres humanos, como a aprendizagem e a tomada de decisões. A aprendizagem de máquina se dá quando o uso de algoritmos e de grandes conjuntos de dados treina modelos computacionais para reconhecer padrões que não foram definidos anteriormente. Kertysova (2018) esclarece que os sistemas alimentados por Inteligência Artificial representam novos riscos para os direitos humanos e para a democracia. Entre as preocupações: a falta de legislação que inclua a política de utilização de algoritmos; a personalização de conteúdo com resultados parciais, provocando cegueira por meio das ‘bolhas informacionais’; a violação da privacidade do usuário por meio de manipulação de áudio e vídeo sem o consentimento individual etc. Kertysova (2018) adverte que:

[...] soluções de Inteligência Artificial têm sido particularmente eficazes na detecção e remoção de conteúdo *on-line* ilegal, duvidoso e indesejável. As técnicas de Inteligência Artificial também foram bem-sucedidas na triagem e identificação de contas de *bot* falsas – técnicas conhecidas como detecção de *bots* e rotulagem de *bots*. Ao rotular contas identificadas como *bots*, as empresas de mídia social estão permitindo que os usuários entendam melhor o conteúdo com o qual estão se envolvendo e julguem sua veracidade por si mesmos. No que diz respeito à sua precisão, no entanto, os algoritmos de detecção

precisam ser desenvolvidos para serem comparáveis à tecnologia de filtro de *spam* de *e-mail*.⁸ (KERTYSOVA, 2018, p. 59, tradução e grifo nossos)

Assim, consideramos a aprendizagem de máquina o processo pelo qual um conjunto de parâmetros é ajustado a partir de um grupo de dados, chamado conjunto de treinamento. A esse ajuste está associado o conceito de aprendizagem, uma vez que o objetivo final é poder fazer previsões com esses parâmetros. Outros nomes que aparecem são rede neural e neurons, e cabe aqui definir brevemente o papel deles no processo de aprendizagem. Um único parâmetro a ser ajustado é associado, nesse tipo de processo, à unidade básica, que é o neurônio.

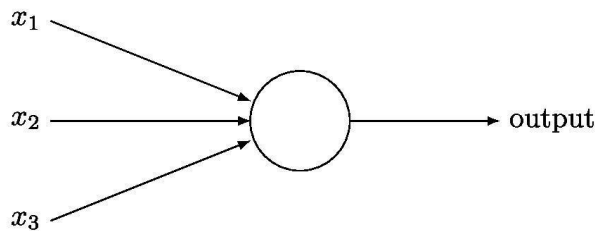


Figura 1 – Esquema de um neurônio, com três *inputs* e um *output*, a ‘sinapse’

Fonte: elaboração dos autores no *software* Tikz LaTeX.

A estrutura mais simples para a sinapse é dada pela expressão na Figura 2.

$$\text{output} = \sigma = \begin{cases} 0 & \text{se } \sum_j w_j x_j + b_j \leq 0 \\ 1 & \text{se } \sum_j w_j x_j + b_j \geq 0 \end{cases}$$

Figura 2 – Expressão da sinapse

Fonte: elaboração dos autores no *software* Tikz LaTeX.

Conforme a Figura 2, os w e os b são os parâmetros a serem ajustados, e x representa os dados. Essa representação de σ é mais didática, em termos práticos normalmente se usam outras formas para a função de ativação: “De fato, é a suavidade da função σ que é o fato crucial, não sua forma detalhada. A suavidade de σ significa que pequenas mudanças Δw_j nos pesos e Δb no viés produzirão uma pequena mudança $\Delta \text{saída}$ na saída do neurônio” (NIELSEN, 2015, tradução nossa).^{9,10}

Um modelo de aprendizagem de máquina é composto por uma rede neural e é caracterizado por:
1. Estrutura da rede neural; 2. Função de ativação; 3. Função de custo; 4. Método de otimização; e 5. Método de aprendizagem.

8 AI solutions have been particularly effective in detecting and removing illegal, dubious, and undesirable content online. AI techniques have also been successful in screening for and identifying fake bot accounts – techniques known as bot-spotting and bot-labelling. By labelling accounts identified as bots, social media firms are enabling users to better understand the content they are engaging with and judge its veracity for themselves. As regards their accuracy, however, detection algorithms need to be further developed in order to be comparable to the e-mail spam filter technology (KERTYSOVA, 2018, p. 59).

9 Sagar Sharma (2017) fornece uma breve explicação sobre as funções de ativação e uma lista das funções mais usadas.

10 Indeed, it's the smoothness of the σ function that is the crucial fact, not its detailed form. The smoothness of σ means that small changes Δw_j in the weights and Δb in the bias will produce a small change Δoutput in the output from the neuron (NIELSEN, 2019).

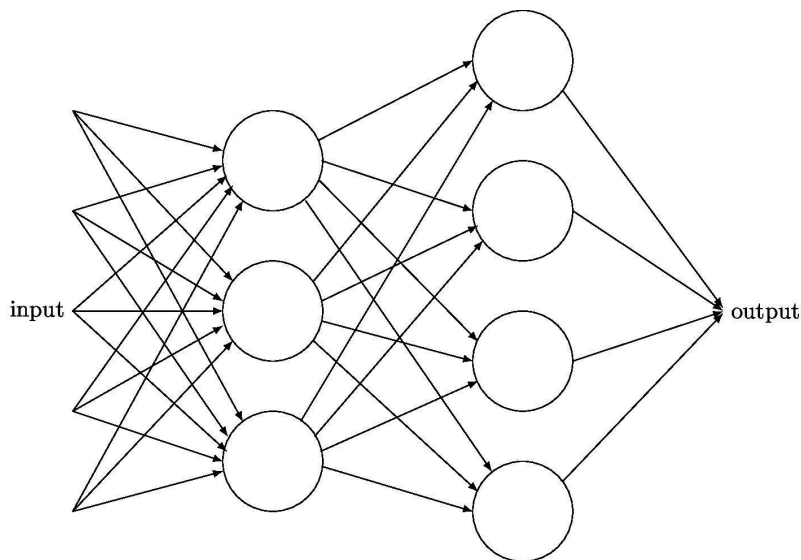


Figura 3 – Esquema geral de uma rede neural em duas camadas
Fonte: elaboração dos autores no *software* Tikz LaTeX.

Estrutura da rede neural é a definição do número de neurons em cada camada e o número total de camadas, montando assim a arquitetura da rede. Quanto maior o número de camadas maior o número de parâmetros w e b a serem ajustados, e espera-se um resultado melhor. Contudo, uma rede demasiada grande levará mais tempo de processamento e poderá levar ao fenômeno de *overfitting*, quando o ajuste dos parâmetros é tão específico ao conjunto de dados de entrada que não é aplicável a qualquer outro conjunto de dados – a consequência é fazer previsões ruins. Uma rede demasiada pequena levará ao fenômeno contrário, o *underfitting*, e também não faz boas previsões. Não há uma receita fundamental a ser seguida para definir essa arquitetura, que, por sua vez, necessariamente, deve ser menor que o número de dados disponíveis. Em nosso trabalho, o número de notícias no conjunto de treino e os diferentes tamanhos de redes são testados. A análise dos resultados leva à escolha da melhor arquitetura.

A função de ativação σ toma conta das decisões no neurônio, se acontece ou não a sinapse numérica. A função custo mede a distância das previsões baseadas nos parâmetros calculados e o conjunto de dados de treino. O processo de otimização visa minimizar a função custo de forma a deixar mais próximo possível da realidade as previsões e os dados reais.

Assim, o método de aprendizagem de máquina a ser usado em nosso trabalho é o de aprendizagem assistida, onde um conjunto de treino dirige o processo de otimização dos parâmetros. O papel do conjunto de validação consiste em estimar a qualidade das previsões para um conjunto diferente do conjunto de treino. “Para monitorar durante o treinamento a precisão do modelo em dados que nunca foram vistos antes, você criará um conjunto de validação separando 10 mil amostras dos dados de treinamento originais”¹¹ (CHOLLET, 2018, p. 75, tradução nossa).

A utilização dos algoritmos tem como objetivo classificar textos para detecção e rastreamento de *fake news* sobre a covid-19. Ressaltamos que na linha de processamento de textos, modelos computacionais já foram desenvolvidos e conseguem classificar críticas de filmes, emoções em redes digitais, fazer traduções, entre outros temas e assuntos (CHOLLET, 2018). Isso porque, embora existam inúmeras iniciativas em todo o mundo para a checagem de desinformação, esse trabalho ainda é realizado manualmente e, por conseguinte, ainda é insuficiente, na medida em que assistimos ao crescimento do número, dos tipos, dos formatos e

¹¹ In order to monitor during training the accuracy of the model on data it has never seen before, you'll create a validation set by setting apart 10.000 samples from the original training data (CHOLLET, 2018, p. 75).

das técnicas de produção de *fake news*, bem como ao crescimento de seus impactos na realidade social, sobretudo, na situação de emergência sanitária imposta pela pandemia da covid-19.

METODOLOGIA

Nosso trabalho foi dividido em cinco etapas fundamentais. A primeira consistiu no levantamento bibliográfico para dar conta da extensão do problema de pesquisa. Essa etapa é imprescindível para a compreensão do fenômeno, conforme a literatura já disponível sobre o assunto (RECUERO; SOARES; GRUZD, 2020; GOMES, 2020; PENNYCOOK; RAND, 2019; KERTYSOVA, 2018; CHOLLET, 2018; MONTEIRO *et al.* 2018). A segunda etapa do trabalho definiu um número grande de notícias falsas e um número também grande de notícias verdadeiras. Essa etapa consistiu na busca por bases de dados, tendo como resultado a base de dados da Agência Lupa, e contou com um levantamento manual de notícias falsas e verdadeiras, sendo que para cada notícia identificada como desinformação havia a correspondência de outras três notícias verdadeiras.

A terceira etapa consistiu em tabular a planilha e atribuir o caráter ‘verdadeiro’ ou ‘falso’ para cada unidade de análise com o objetivo de organizar a base de dados para treinamento e aprendizagem de máquina para reconhecimento de textos verdadeiros e falsos. Esse conjunto de textos precisa passar por um pré-processamento para eliminar termos de pouca informação no conjunto de textos, e armazenar apenas as palavras significativas. Esse conjunto de dados constituiu o *corpus* final da pesquisa. A partir disso, separamos o conjunto de dados em dois conjuntos: um para treino e outro para validação, feito de forma aleatória. Feito isso, elaboramos o código Python (BOSELLI, 2022) para treinar um modelo com o conjunto de dados e fazer a validação para testar diferentes modelos e estratégias de treino e verificar qual leva ao melhor resultado na validação. Na quarta etapa, usamos o modelo mais promissor do mundo real explorando notícias que são propagadas nas redes sociais para, assim, analisarmos os resultados e definirmos a porcentagem de acertos.

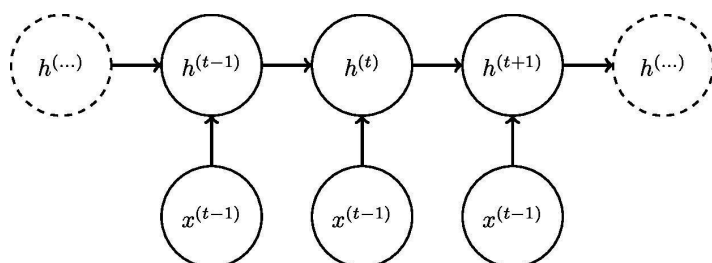


Figura 4 – Esquema de uma Rede Neural Recorrente (RNR)
Fonte: elaboração dos autores no *software* Tikz LaTeX.

Os trabalhos já realizados em classificação de textos apontam que a arquitetura de Redes Neurais Recorrentes (RNR) é eficiente para tratar sequências.

Esse é um tipo de rede neural que utiliza, durante o processo de treinamento, informações do passo anterior ($t-1$) no passo seguinte t (CHOLLET, 2018). As RNR foram desenvolvidas para fazer o processamento de sequência de dados (GOODFELLOW; BENGIO; COURVILLE, 2016), como frases em alguma língua, por exemplo, na qual o conjunto de palavras forma uma sequência de dados. “Usaremos LSTMs [long short-term memory] para construir nosso modelo. Os LSTMs têm a vantagem da memória, o que é muito útil quando trabalhamos com dados sequenciais como: Fala, Texto e Preços de Ações. Basicamente, temos dados sequenciais quando o passado importa para a previsão!¹²” (MAYERS, 2020, tradução nossa).

¹² We'll use LSTM's to build our Model. LSTM's have the advantage of memory, this is very useful when we're working with sequential data like: Speech, Text and Stock Prices. Basically, we have sequential data when the past matters to the prediction! (MAYERS, 2020).

O processo de montagem do banco de dados pode ser feito com a ajuda de códigos Python de *web scraping*¹³ para capturar e tabular os textos e as classificações. Se um texto vem de uma agência de checagem, ele já tem a informação-chave de verdadeiro ou falso, e a tabulação completa é automática. Porém, esse processo de captura de textos pode ser usado também para outras fontes de informação como jornais, redes sociais etc. Nesse caso, a classificação da informação como verdadeira ou falsa precisa ser feita manualmente. Isso finaliza a terceira etapa do processo, que ainda pode exigir interferência manual.

Entre a coleta de textos e a sua classificação, é necessário realizarmos o pré-processamento, processo que é bastante facilitado pelas ferramentas das bibliotecas NLTK¹⁴ (do inglês Natural Language Toolkit) e Regex¹⁵ (do inglês Regular Expression) do Python voltadas para tratamento de textos. “Quase todas as tarefas de Processamento de Linguagem Natural (NLP) exigem que o texto seja pré-processado antes de treinar um modelo. Modelos de aprendizado profundo não podem usar texto bruto diretamente, então cabe a nós, pesquisadores, limpar o texto por conta própria.”¹⁶ (CHENG, 2020, tradução nossa).

É justamente nesse ponto que foi encontrado um dos gargalos para o processamento de textos em português (Brasil). O NLTK conta com a sua funcionalidade para a língua inglesa num estágio bastante avançado, e o pré-processamento acontece de forma natural. Já para o português há algumas deficiências. Por exemplo, para este trabalho foi necessário acrescentar novas palavras no conjunto das *stopwords* no intuito de conseguir um *corpus* limpo e funcional para uso no treino do aprendizado de máquina. Uma segunda filtragem de símbolos e de outros termos sem interesse é necessária, e pode ser feita a partir da biblioteca Regex com a construção de filtros dedicados a cada conjunto de textos.

A última etapa consiste em usar esses textos tabulados como conjunto de treino e validação. Treino é o termo usado para o processo de ajuste dos parâmetros do modelo. A cada passo o modelo é confrontado com o conjunto de validação. Com um modelo treinado, podemos usar o código para classificação de notícias e postagens em redes sociais. Novamente com a ajuda dos códigos de *web scraping*, agora acoplados com o código de classificação/detecção, é possível coletar os textos da rede e ter uma avaliação de verdadeiro ou falso. Essas últimas duas etapas são completamente automáticas.

RESULTADOS PRELIMINARES

Na segunda etapa, que consistia em definir um número grande de notícias como falsas e um número também grande de notícias como verdadeiras, elaboramos manualmente uma base de dados em que cada notícia ou conteúdo identificado como desinformação estava relacionado a três notícias verdadeiras. Somou-se a esse esforço a utilização da base de dados da Agência Lupa, de *fact-checking*. Na coleta de *fake news* realizada manualmente, constatamos um número maior de conteúdos manipulados ou fabricados com viés político – com a finalidade de reproduzir informações falsas ou tendenciosas para difamar a imagem de grupos ou políticos. Em geral, as quatro categorias com maior número de *fake news*, durante a coleta manual foram: 1. informação incompleta, tendenciosa ou simplesmente falsa; e 2. comentários orientados – classificados por Volkoff (2000). Sobre as atuações das estratégias de desinformação, o autor explica: “[...] não difundindo uma informação, ou difundindo uma informação incompleta, tendenciosa ou simplesmente falsa, ou saturando a atenção por meio de uma sobreinformação que faz perder o sentido do

13 Coleta direta de dados de páginas *web* em HTML. Usamos a biblioteca BS4. Disponível em Richardson (c2020).

14 **NLTK** é uma biblioteca Python para processamento de linguagem natural. Ela permite trabalhar com textos fazendo análise léxica, sintática e semântica de forma automática.

15 **Regex** ou RE é uma biblioteca básica do Python que permite a busca, a comparação e a troca de expressões em um texto. É muito útil na construção de filtros para eliminar termos, espaços e caracteres sem interesse em um texto.

16 *Almost every Natural Language Processing (NLP) task requires text to be preprocessed before training a model. Deep learning models cannot use raw text directly, so it is up to us researchers to clean the text ourselves* (CHENG, 2020).

que é e não é importante, ou através de comentários orientados” (VOLKOFF, 2000, p. 132). A categoria ‘comentários orientados’ se refere à utilização de declarações e falas de autoridades em determinado assunto para validar um posicionamento.

Ao analisar o material, consideramos também as categorias de Wardle (2017) e a maior parte do conteúdo entendido como falso classifica-se como: 1. conteúdo manipulado; e 2. conteúdo fabricado. A autora explica que ‘conteúdo manipulado’ se refere a imagens e vídeos deliberadamente manipulados por meio de programas como o Photoshop e são uma grande parte do ecossistema de notícias falsas, uma vez que podem ser criados facilmente por fraudadores e são frequentemente compartilhados pelos mesmos com intuítos maliciosos. Já o ‘conteúdo fabricado’ é um novo conteúdo inteiramente falso, projetado para enganar e causar danos a determinado alvo.

Essas classificações são importantes na medida em que nos oferecem a dimensão de conteúdo e como são produzidas as *fake news* – que elementos as compõem –, de modo a permitir criar relações semânticas com os conteúdos da desinformação sobre a covid-19 para o processo de treino e aprendizagem de máquina que será realizado.

Para confrontar a checagem manual com uma detecção automática de notícias falsas, construímos um modelo de aprendizagem de máquina usando as bibliotecas [Keras/TensorFlow](#), do Python. O modelo consistiu de: 1) uma camada de *word embedding* com dimensão 500, responsável por organizar as palavras por proximidade de seu significado semântico. Esse procedimento melhora muito o desempenho geral do modelo, uma vez que promove um armazenamento denso do conjunto de dados, e os modelos de aprendizagem de máquina convergem mais rapidamente com vetores densos, quando comparados a vetores esparsos (VIRAHONDA, 2020). A dimensão de 500 significa que cada elemento do conjunto tem um tamanho fixo de 500 palavras. Textos com mais de 500 palavras retornam com resultados truncados; 2) uma camada LSTM com dimensão 64. Essa é a parte do código que faz o processamento recorrente mantendo a informação e a memória dos passos anteriores. Esse modelo apresenta os melhores resultados em processamento de textos por Inteligência Artificial (KHAN *et al.*, 2019). A dimensão 64 significa que são 64 neurons artificiais; e, por fim, 3) uma camada de saída com conexão densa e ativação sigmoideal, uma função que retorna de forma binária aos valores de 0 ou 1, formando a classificação final 0 para falso e 1 para verdadeiro. Para treinar esse modelo, usamos o *corpus* preparado por Monteiro *et al.* (2018). A validação do processo indicou um índice de acerto de 95% durante o treino. O último passo foi utilizar o modelo treinado no conjunto de dados coletados neste trabalho, um conjunto de dados completamente independente do *corpus* usado no treino.

Nesse caso, o índice de acerto foi de 70%. Para entender esse número é apropriado recorrer ao trabalho feito por Khan *et al.* (2019), que testaram vários tipos de modelos de aprendizagem de máquina para detecção de *fake news*, incluindo os modernos LSTM que podem ser considerados o estado da arte em processamento de textos por modelos de aprendizagem de máquina. Esses autores relatam resultados de detecção que vão de 70% a 95% de acerto, dependendo do modelo, para um *corpus* de 80 mil textos. Considerando que estávamos limitados a um *corpus* de 7,2 mil textos, entendemos esse índice de acerto como um resultado positivo.

CONCLUSÃO

À guisa de concluir nosso trabalho, entendemos que estamos vivendo, sobretudo no Brasil, após as eleições de 2018, um fenômeno complexo, sintomático e sistêmico de esgarçamento da racionalidade e dos deveres cívicos em sua essência. Nos cercamos, hoje, da ação de compartilhamentos irrefletidos em redes digitais acerca de conteúdos de fontes duvidosas que chegam a grupos pré-selecionados por algoritmos. Esses, por sua vez, formam as ‘bolhas’ de desinformação, separando e confundindo os cidadãos,

desestabilizando as instituições, cedendo lugar ao convencimento calcado em comoção e convicções que corroboram com a ausência de cultura política.

Ao relatar as estratégias para coleta de conjunto de dados em português para treinamento de modelos de Inteligência Artificial com vistas a identificar as *fake news* sobre a covid-19 disseminadas durante a pandemia, identificamos a necessidade de contextualizar o fenômeno das *fake news* dentro da complexidade em que se inserem. Isso significa demonstrar que, para além dos aspectos computacionais e da automação, o problema da desinformação é político, econômico e envolve mudanças culturais recentes que solaparam nossa forma de ser e estar, ler e compreender no/o mundo. As *fake news* sobre a covid-19 não são apenas um problema do campo da comunicação ou da saúde pública e coletiva: elas engendram aspectos que envolvem a automação, o uso de algoritmos, de *bots*, mas têm como pano de fundo aspectos relativos ao crescimento das redes digitais e da hiperpolitização. Sendo assim, entendemos que, numa ambiência em que a opinião é calculada e formatada de acordo com interesses comerciais e políticos (SILVA; ARAÚJO, 2019), não podemos afirmar que vivemos na ‘Era da Informação’, mas na ‘Era da Desinformação’, em que a construção de consensos a partir de debates críticos é algo cada vez mais distante de nossa realidade.

As *fake news* relacionadas à covid-19 têm impactos na saúde, na vida das pessoas, na maneira como compreendem e se previnem. É um problema que exige políticas públicas para o adequado enfrentamento, e quiçá essas políticas passem pela necessidade de utilização das mesmas ‘armas’ usadas para a disseminação: a Inteligência Artificial e a automação. O potencial prejudicial das *fake news* é amplificado pela capacidade de disseminação delas em um contexto cultural e político marcado por uma guerra ideológica que divide a sociedade, polarizando-a. Conforme descrito, esse contexto se torna um terreno fértil para a disseminação do medo e da desconfiança nas instituições políticas e midiáticas, uma condição catalisadora de discursos de ódio, teorias da conspiração e campanhas difamatórias de toda ordem (ALVES; MACIEL, 2020).

Desse modo, deve-se estimular pesquisas que auxiliem de forma mais eficaz na automação da detecção de *fake news*, conteúdos inverídicos que buscam se passar por fatos com certa objetividade, e processos de desinformação. Assim, nosso trabalho contribuiu para o enfrentamento e o combate à infodemia, a partir de uma perspectiva multidisciplinar entre os campos da comunicação, da física e da computação, de forma a evidenciar que, se de um lado é possível automatizar a detecção de *fake news* sobre a covid-19 – a partir de nossa pesquisa –, de outro as estratégias para detecção de *fake news* apresentadas em nosso trabalho assinalam possibilidades efetivas para que grandes corporações midiáticas, redes sociais e organizações engajadas na luta contra a desinformação possam atuar com capacidade resolutiva sobre o fenômeno – compreendido na profundidade e na extensão da complexidade do problema social tal qual se apresenta.

Ressaltamos ainda que, nesse momento, é fundamental que a deficiência de conjuntos de dados existentes para a língua portuguesa seja suprida, tanto no âmbito da NLTK como, principalmente, na implementação e no compartilhamento de novos *corpora* dedicados à detecção de *fake news*. Aqui demonstramos que um modelo bem conhecido dentro das RNR e um *corpus* restrito em tamanho foram capazes de realizar um bom trabalho. A ampliação dos dados hoje disponíveis para uso em sistemas de Inteligência Artificial levará a resultados ainda melhores.

REFERÊNCIAS

ALVES, Marco Antônio Sousa; MACIEL, Emanuella Ribeiro Halfeld. O fenômeno das *fake news*: definição, combate e contexto. **Internet & Sociedade**, Rio de Janeiro, v. 1, n. 1, p. 144-171, 2020. Disponível em: <https://revista.internetlab.org.br/o-fenomeno-das-fake-news-definicao-combate-e-contexto/>. Acesso em: 15 nov. 2021.

AMPER ENERGIA HUMANA. We Are Social e HootSuite: Digital 2021 [resumo e relatório completo]. In: AMPER ENERGIA HUMANA. **Amper: marketing e comunicação**. São Paulo, 03 maio 2022. Disponível em: <https://www.amper.ag/post/we-are-social-e-hootsuite-digital-2021-resumo-e-relatorio-completo>. Acesso em: 16 maio 2021.

AVAAZ. **IBOPE**: 1 em cada 4 brasileiros pode não se vacinar contra a covid-19. [S. l.]: Avaaz, 07 set. 2020. Disponível em: https://secure.avaaz.org/campaign/po/brasileiros_ nao_vacinar_covid/. Acesso em: 14 nov. 2021.

AVAAZ. **O Brasil está sofrendo uma infodemia de covid-19**. [S. l.]: Avaaz, 04 maio 2020. Disponível em: https://avaazimages.avaaz.org/brasil_infodemia_coronavirus.pdf. Acesso em: 25 maio 2022.

BOSELLI, Marco Aurélio. **Reciis**. Uberlândia, 25 maio 2022. Código Python. Disponível em: <https://github.com/maboselli/reciis>. Acesso em: 14 jun. 2022.

BRUCK, Mozahir Salomão. **O jornalismo diante de novos cenários sociais**: a imprensa e o surgimento da aids e do *crack*. São Paulo: Intermeios, 2015.

CHENG, Raymond. Text preprocessing with NLTK. *In*: TDS Editors; HUBERMAN, Ben; Kindig Caitlin (ed.). **Towards Data Science**, [s. l.], 29 jun. 2020. Disponível em: <https://towardsdatascience.com/nlp-preprocessing-with-nltk-3c04ee00edc0>. Acesso em: 28 nov. 2021.

CHOLLET, François. **Deep learning with Python**. New York: Manning Publications, 2018.

COLLINS COBUILD. Fake news. *In*: COLLINS. Collins Dictionary. Nova York: Harper Collins, c2022. Disponível em: <https://www.collinsdictionary.com/dictionary/english/fake-news>. Acesso em: 01 fev. 2022.

FERREIRA, Fernanda Vasques. **O papel do factual nos processos de agendamento e de enquadramento no telejornalismo**. 2018. 438 f., il. Tese (Doutorado em Comunicação) – Universidade de Brasília, Brasília, 2018. Disponível em: <https://repositorio.unb.br/handle/10482/33073>. Acesso em: 16 maio 2021.

GOMES, Wilson. **O que são fake news?**. Brasília, DF: INCT, 2020. 1 vídeo (38 min). Publicado pelo canal INCT em Democracia Digital. Disponível em: <https://www.youtube.com/watch?v=8tvJ4cMt YXY>. Acesso em: 16 maio 2021.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. Massachusetts: The MIT Press, Cambridge, 2016.

HASSON, Erez. Bad bot report 2021: the pandemic of the internet. *In*: IMPERVA, 13 abr. 2021. Disponível em: <https://www.imperva.com/blog/bad-bot-report-2021-the-pandemic-of-the-internet/>. Acesso em: 23 jan. 2022.

KERTYSOVA, Katarina. Artificial Intelligence and Disinformation: how AI changes the way disinformation is produced, disseminated, and can be countered. **Security and Human Rights**, Leiden, v. 29, p. 55-81, 2018. DOI: <http://dx.doi.org/10.1163/18750230-02901005>. Disponível em: https://brill.com/view/journals/shrs/29/1-4/article-p55_55.xml. Acesso em: 14 jun. 2022.

KHAN, Junaed Younus *et al.* A benchmark study on Machine Learning Models for online fake news detection. **arXiv**, Ithaca, 12 maio 2019. Disponível em: <https://arxiv.org/abs/1905.04749>. Acesso em: 06 jun. 2022.

MAYERS, Gabriel. Detecting fake news using Machine Learning. *In*: TEAM AV (ed.). **Analytics Vidhya**, [s. l.], 19 jun. 2020. Disponível em: <https://medium.com/analytics-vidhya/detecting-fake-news-using-machine-learning-95efefab08e4>. Acesso em: 07 mar. 2022.

MELLO, Patrícia Campos. **A máquina do ódio**: notas de uma repórter sobre *fake news* e violência digital. São Paulo: Companhia das Letras, 2020.

MONTEIRO, Rafael A. *et al.* Contributions to the study of fake news in Portuguese: new corpus and Automatic Detection Results. *In*: COMPUTATIONAL FOR PORTUGUESE LANGUAGE, 13., 24-26 set. 2018, Canela. **Proceedings** [...]. Cham: Springer, 2018. p. 324-334. (Lecture Notes in Computer Science, v. 11122). DOI: http://dx.doi.org/10.1007/978-3-319-99722-3_33. Disponível em: <https://github.com/roneysco/Fake.br-Corpus>. Acesso em: 27 fev. 2022.

NIELSEN, Michael A. **Neural networks and deep learning**. [S. l.], Determination Press, 2015. Disponível em: <http://neuralnetworksanddeeplearning.com/index.html>. Acesso em: 27 fev. 2022.

PENNAFORT, Roberta. É #FAKE que foto mostre caixão enterrado vazio para inflar dados de mortos por coronavírus em Manaus. **G1**, [s. l.], 30 abr. 2020. Fato ou Fake. Disponível em: <https://g1.globo.com/fato-ou-fake/coronavirus/noticia/2020/04/30/e-fake-que-foto-mostre-caixao-enterrado-vazio-para-inflar-dados-de-mortos-por-coronavirus-em-manaus.ghtml>. Acesso em: 29 maio 2022.

PENNYCOOK, Gordon; RAND, David G. Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. **Cognition**, [s. l.], v. 188, p. 39-50, 2019. DOI: <https://doi.org/10.1016/j.cognition.2018.06.011>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S001002771830163X>. Acesso em: 16 maio 2021.

PEREIRA, Denilson Alves. A survey of sentiment analysis in the portuguese language. **Artificial Intelligence Review**, [s. l.], v. 54, n. 2, p. 1087-1115, 2020. DOI: <https://doi.org/10.1007/s10462-020-09870-1>. Disponível em: <https://link.springer.com/article/10.1007/s10462-020-09870-1>. Acesso em: 25 maio 2022.

RECUERO, Raquel *et al.* **Desinformação, mídia social e covid-19 no Brasil**: relatório, resultados e estratégias de combate. Pelotas: MIDIARS – Grupo de Pesquisa em Mídia Discurso e Análise de Redes Sociais, 2021. Relatório de pesquisa. Disponível em: <https://wp.ufpel.edu.br/midiars/files/2021/05/Desinformac%CC%A7a%CC%83o-covid-midiars-2021-1.pdf>. Acesso em: 14 jun. 2022.

RECUERO, Raquel; SOARES, Felipe Bonow; GRUZD, Anatolii. Hyperpartisanship, Disinformation and Political Conversations on Twitter: The Brazilian Presidential Election of 2018. In: INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA, 14., 2020, Atlanta. **Proceedings** [...]. Atlanta: AAAI Digital Library, 2020. Disponível em: <https://ojs.aaai.org/index.php/ICWSM/article/view/7324>. Acesso em: 14 jun. 2022.

RICHARDSON, Leonard. **Beautiful Soup Documentation**. [S. l.]: Leonard Richardson, c2020. Disponível em: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. Acesso em: 06 jun. 2022.

SHARMA, Sagar. Activation functions in neural networks: Sigmoid, tanh, Softmax, ReLU, Leaky ReLU explained!!! In: TDS Editors; HUBERMAN, Ben; Kindig Caitlin (ed.). **Towards Data Science**, [s. l.], 6 set. 2017. Disponível em: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>. Acesso em: 06 jun. 2022.

SILVA, Igor Fediczko; ARAÚJO, Rafael de Paula Aguiar. Campanhas políticas em tempos de hiperpolítica: um ensaio sobre Peter Sloterdijk e a campanha de 2018. **Ponto e Vírgula**, São Paulo, n. 26, p. 138-145, 2019. Disponível em: <https://revistas.pucsp.br/index.php/pontoevirgula/article/view/51519/34074>. Acesso em: 28 set. 2021.

SLOTERDIJK, Peter. **No mesmo barco**: ensaio sobre a hiperpolítica. São Paulo: Estação Liberdade, 1999.

SOUSA JUNIOR, João Henrique de; PETROLL, Martin de La Martinière; ROCHA, Rudimar Antunes da. *Fake news* e o comportamento *on-line* dos eleitores nas redes sociais durante a campanha presidencial brasileira de 2018. In: SEMINÁRIOS EM ADMINISTRAÇÃO, 22., 6-8 nov, São Paulo. **Anais** [...], São Paulo: SemeAd, 2019. Disponível em: https://login.semead.com.br/22semead/anais/resumo.php?cod_trabalho=501. Acesso em: 20 dez. 2021.

SOUZA, Frederico Dias; SOUZA FILHO, João Batista de Oliveira. Sentiment analysis on Brazilian Portuguese user reviews. **arXiv**, Ithaca, 10 dez. 2021. Disponível em: <https://arxiv.org/pdf/2112.05459.pdf>. Acesso em: 20 dez. 2021.

VARÃO, Rafiza. Há alguma novidade na ideia de *fake news*? In: SOS IMPRENSA. **Blog SOS Imprensa**. Brasília, DF, 18 out. 2017. Disponível em: <https://sosimprensa.wordpress.com/2017/10/18/ha-alguma-novidade-na-ideia-de-fake-news/>. Acesso em: 03 jan. 2022.

VIRAHONDA, Sergio. An easy tutorial about sentiment analysis with deep learning and Keras. In: TDS Editors; HUBERMAN, Ben; Kindig Caitlin (ed.). **Towards Data Science**, [s. l.], 08 out. 2020. Disponível em: <https://towardsdatascience.com/an-easy-tutorial-about-sentiment-analysis-with-deep-learning-and-keras-2bf52b9cba91>. Acesso em: 03 jan. 2022.

VOLKOFF, Vladimir. **Pequena história da desinformação**: do cavalo de Troia à internet. Lisboa: Editorial Notícias, 2000.

WARDLE, Claire. Fake News. It's complicated. In: FIRST DRAFT. **First Draft Footnotes**, [s. l.], 16 fev. 2017. Disponível em: <https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79>. Acesso em: 14 maio 2021.

WÓJCIK, Rafał. Unsupervised sentiment analysis. In: TDS Editors; HUBERMAN, Ben; Kindig Caitlin (ed.). **Towards Data Science**, [s. l.], 26 nov. 2019. Disponível em: <https://towardsdatascience.com/unsupervised-sentiment-analysis-a38bf1906483>. Acesso em: 25 de maio 2022.