

A IMPORTÂNCIA DO DATA SCIENCE PARA PREVENÇÃO DE FAKE NEWS THE IMPORTANCE OF DATA SCIENCE FOR PREVENTING FAKE NEWS

Stefany Silva de Oliveira¹
Rebecca Bignardi Arambasic da Silva²

RESUMO:

O avanço tecnológico é uma realidade mundial, essencialmente no ramo da comunicação, de maneira rápida, uma informação pode atravessar longas distâncias em poucos minutos. Durante esse percurso, pode ocorrer alteração na mensagem transmitida, muitas vezes para tornar algo surpreendente ou até mesmo expor opiniões pessoais como uma fonte verídica. A transmissão de mensagens falsas para causar impacto ou ganhar reconhecimento pessoal se tornou comum atualmente, visto que a comunicação facilitada permite que qualquer indivíduo possa não só alterar, mas também criar notícias. Essas ações causam impacto direto na sociedade, gerando insegurança e desinformação. Neste projeto será realizada a revisão literária de artigos que abordam Data Science e seus derivados para prevenção de notícias falsas. Para este estudo temos como hipótese que é possível utilizar este recurso através do machine learning, exercitando e aprimorando suas técnicas analíticas de texto que podem contribuir na redução da propagação de notícias falsas. O objetivo principal deste estudo está envolvido na análise de impactos das notícias falsas na ciência de dados. Os objetivos específicos traçados foram divididos em três, a saber: (i) verificar os efeitos e impactos que as notícias falsas causam na sociedade; (ii) descrever as técnicas do data science; (iii) discutir como o data science pode auxiliar na mitigação de notícias falsas. A escolha para este tema, foram acontecimentos recentes que influenciaram a população na tomada de decisões duvidosas causada pela insegurança transmitida através de conteúdos duvidosos trazendo à questão como a tecnologia e a ciência de dados poderiam contribuir para esta precaução. O arcabouço teórico que sustenta este trabalho são os estudos sobre a utilização de Data Science para detecção de fake news são artigos dos autores Shoemaker 2019, Wazib Ansar, Saptarsi Goswami 2021, Renny Thomas 2021, e Ahmed, Aljarbouh, Donepudi e Choi 2021. A metodologia utilizada é a revisão sistemática da literatura, como método aplicado nos estudos, o treinamento de classificadores capazes de identificar através de sites que possuem compilado de informações verdadeiras e falsas, analisando a notícia de maneira textual com contagem e identificação de palavras semelhantes. Os resultados alcançados confirmam que este

¹ Graduando do curso Análise e Desenvolvimento de Sistemas da Faculdade de Tecnologia da Zona Sul. stefany.oliveira3@fatec.sp.gov.br

² Rebecca Bignardi Arambasic da Silva da Faculdade de Tecnologia da Zona Sul. rebecca.silva3@fatec.sp.gov.br

método é eficaz, porém requer um treinamento mais acentuado com uma necessidade mais abrangente de dados rotulados para tornar a machine learning mais independentes. Dessa forma, evidenciamos que é possível utilizar machine learning para mitigar as fake news, porém é um trabalho que necessita de aperfeiçoamento gradual e também a necessidade contínua de atualização e constantes cargas de treinamento para aperfeiçoamento do machine learning na adaptação para do acompanhamento da evolução das técnicas de criação das notícias falsas, mostrando que o data science é uma das ferramentas para prevenção, mas não a solução principal.

Palavras-chave: Notícias, Machine-learning. Tecnologia. Data Science. Falsas.

ABSTRACT

Technological advancement is a global reality, essentially in the field of communication. Information can travel long distances in a few minutes. During this journey, changes may occur in the message transmitted, often to make something surprising or even expose personal opinions as a truthful source.

The transmission of false messages to make an impact or gain personal recognition has become common today, as facilitated communication allows any individual to not only change, but also create news. These actions have a direct impact on society, generating insecurity and misinformation. In this project, a literary review of articles that address Data Science and its derivatives will be carried out to prevent fake news. For this study, we hypothesize that it is possible to use this resource through machine learning, exercising and improving text analytical techniques that can contribute to reducing the spread of fake news. The main objective of this study is involved in analyzing the impacts of fake news on data science. The specific objectives outlined were divided into three, namely: (i) verify the effects and impacts that fake news causes in society; (ii) describe data science techniques (iii) discuss how data science can help in the spread of fake news. The main objective of this study is involved in analyzing the impacts of fake news on data science. The specific objectives outlined were divided into three, namely: (i) verify the effects and impacts that fake news causes in society; (ii) Describe data science techniques (iii) discuss how data science can help in the spread of fake news. The choice for this topic were recent events that influenced the population in making dubious decisions caused by insecurity transmitted through dubious content, raising the question of how technology and data science could contribute to this precaution. The theoretical framework that supports this work are studies on the use of Data Science to detect fake news are articles by the authors Shoemaker 2019, Wazib Ansar, Saptarsi Goswami 2021, Renny Thomas 2021, and Ahmed, Aljarbouh,

Donepudi and Choi 2021. The methodology used is the systematic literature review, as a method applied in the studies, the training of classifiers capable of identifying through websites that have compiled true and false information, analyzing the news in a textual way with counting and identification of similar words. The results confirm that this method is effective but requires more extensive training with a more comprehensive need for labeled data to make machine learning more independent.

In this way, we demonstrate that it is possible to use machine learning to mitigate fake news, but it is a work that requires gradual improvement and also the continuous need for updating and constant training loads to improve machine learning in adapting to monitoring the evolution of techniques for creating fake news, showing that data science is one of the tools to use for prevention, but not the main solution.

Keywords: Fake-. Machine Learning. Technology. Data Science- News.

1.0 INTRODUÇÃO

A utilização da tecnologia é um grande aliado para se obter e transmitir informações, na atualidade em qual se vive, uma notícia pode percorrer longas distâncias em segundos com apenas um clique.

Durante este percurso, a informação pode ser modificada propositalmente gerando a quem recebe.

As chamadas fake news ou nacionalmente conhecidas como notícias falsas, estão ganhando cada vez mais popularidade ocasionando insegurança aos usuários, podendo ocasionar a perda de credibilidade na ciência e em instituições democráticas.

Recentemente áreas como política e saúde foram as mais atingidas, visto que o termo Fake News ganhou destaque nas eleições de 2016 nos Estados Unidos, sendo utilizada como ferramenta imprópria para disseminar e manipular opiniões públicas.

Em um contexto geral, situações como estas onde notícias falsas se espalham de uma maneira irrefreável podem gerar um impacto à população e um grande desafio para a ciência.

A problemática para este, foram algumas consequências dessas notícias que geram impacto negativo na população brasileira, iniciando com casos como utilização de medicamentos impróprios, difamação política, compartilhamento de curas incertas sobre o covid-19 e recentemente até em assassinato.

A justificativa é que através do ramo diversificado, todas as áreas da tecnologia têm o mesmo objetivo de promover o avanço do conhecimento, além de facilitar trabalhos manuais e complexos, porém isso se torna repulsivo quando utilizam deste recurso para transmitir conhecimento hostis, causando malefícios ou incentivando não só uma pessoa, mas também uma população em decisões e até ações errôneas.

E a partir deste que tende-se a necessidade do desenvolvimento ou aprimoramento tecnológico de uma ferramenta capaz de contestar conteúdos enganosos.

A hipótese é demonstrar que a machine learning é uma ferramenta capaz de realizar o trabalho de identificação de conteúdos enganosos alertando o usuário, sem a necessidade de outros objetos.

O objetivo principal deste trabalho, é demonstrar que o data science é capaz de auxiliar na propagação das fake news, utilizando o machine learning. Como objetivo específico, temos a verificação dos efeitos e impactos que a desinformação causada pelas notícias falsas causam, explicar as técnicas da ciência de dados e por fim abordar o papel da ciência de dados na disseminação das notícias falsas.

A metodologia utilizada para este trabalho, foi feita a revisão sistêmica literária qualitativa, além da aplicação de um formulário online, voltado a público livre compondo testes A e B mostrando notícias reais e informações falsas, e quais os fatores para distinção das notícias ali aplicadas. O grupo escolhido tem idade entre 18 e 66 anos, para uma maior abrangência e diversidade para pesquisa, levando em consideração a faixa etária e sexo que impacta indiretamente no projeto.

2.0 FUNDAMENTAÇÃO TEÓRICA

Ao citarmos fonte de informações, todas se originam de dados, trazendo isso para o lado tecnológico, podemos citar a ciência de dados, designada a partir de grande volumetria de dados utilizando da estatística para traçar tendências preditivas auxiliando no processo de tomada de decisões.

Segundo Morettin e Singer (2020), este ramos estatístico é utilizado desde 1980, tendo ênfase durante uma análise exploratória feita por Tukey (1962, 1977) relatando a importância de uma inferência estatística antes de uma análise confirmatória das informações com base em dados, partindo deste, para facilitar este processo seria importante utilização de gráficos e tabelas, para análises preliminares. Com o passar dos anos foi se aprimorando até chegar nas suas ramificações e aprendizados.

Os autores citam que Blei e Smyth (2017), relatam que a ciência de dados é uma junção estatística, computacional e humana, desenvolvida para análises mais complexas com grandes capacidades de armazenamento.

A estatística desempenha papel fundamental na ciência de dados, orientando uma coleta de dados complexos, enquanto a computação com algoritmos e processadores proporciona rapidez e capacidade de armazenamento mais eficiente, já a ação humana tende-se escolher quais métodos adequados deve ser explorado para determinadas situações, como visualizar os resultados e compreendê-los de forma que auxilie na resolução ou decisão. A ciência de dados possui suas origens,

partindo delas o Aprendizado com Estatística e o Aprendizado de Máquina, que segundo Morettin e Singer (2020), carregam a essência da ciência de dados utilizando modelos estatísticos e ação computacional para extrair informações de conjuntos de dados que geralmente envolvem diversas unidades amostras e variáveis. O aprendizado com estatística possui duas principais abordagens, sendo a inferência estatística e a previsão. A inferência busca compreender como o resultado é atingido pela variação de seus preditores ou localizar uma relação relevante entre eles, já a previsão busca identificar o melhor caminho para antecipar valores. No contexto de previsão, o aprendizado de estatística pode ser categorizado como supervisionado, onde o intuito é prever o resultado de uma variável resposta de acordo com as variáveis preditoras. Essa variável pode ser numérica, onde é comum a aplicação do modelo de regressão, já em casos de respostas categóricas é utilizado a logística de regressão para análise. Em um modelo não supervisionado o intuito é apenas identificar a padronização entre as variáveis.

2.1 DATA SCIENCE NA TECNOLOGIA

Trazendo para a tecnologia, o aprendizado de máquina ou chamado machine learning faz parte da subárea da inteligência artificial, entretanto com foco no desenvolvimento de algoritmos que identificam padrões dentre os dados sem necessariamente precisar da ação humana. Segundo Morettin e Singer (2020), Turing (1950) foi o pioneiro em considerar que as máquinas seriam capazes de pensar, se tornando inteligentes. Automatizar processos e técnicas faz parte da evolução tecnológica, sendo a inteligência artificial um método de automatizar demandas intelectuais humanas. O processo de aprendizagem da máquina pode ser utilizado para executar diversas atividades, entretanto, há uma certa limitação do algoritmo, mostrando a necessidade de haver modelos que auxiliam o aprendizado para problemas específicos (Monard; Baranauskas,2017). Semelhante ao aprendizado de estatística, o aprendizado de máquina também compartilha de métodos estatísticos para realização de demandas, entretanto é utilizada a indução, como forma de inferência lógica para gerar resultados a partir dos modelos aplicados, esta metodologia é a mais utilizada para o aprendizado, entretanto é necessário a utilização com precaução, visto que se mal aplicada os resultados

podem ser errôneos ou de pouco valor. O aprendizado de máquina também compartilha do aprendizado indutivo, onde é categorizado em supervisionado e não supervisionado.

Segundo Monard e Baranauska (2020), o modelo supervisionado tem como intuito o desenvolvimento de um classificador que consegue identificar e classificar o conjunto de dados aplicados e neste processo envolve a utilização de um indutor, algoritmo de aprendizado, que extrai um classificador a partir de exemplos rotulados. Cada exemplo consiste em um conjunto de valores de atributos que descrevem um objeto de interesse, junto com uma classe associada. Os atributos podem ser nominais que são sem ordem ou contínuos que contém uma ordem linear. As classes podem ser discretas denominadas a nominais ou valores reais, regressão.

Na prática de um aprendizado supervisionado, a máquina possui a fase de treinamento para ensinar o modelo que deve ser seguido, e logo após temos a fase teste, utilizada para avaliar seu desempenho. Em uma classificação, temos os ruídos, denominados como erros de rotulação ou outliers, que podem impactar a qualidade do modelo, problemas de underfitting que acontecem quando o modelo é muito simples. Já o overfitting ocorre quando é excessivamente específico para o conjunto de treinamento, por isso a necessidade de encontrar o equilíbrio entre eles é essencial. A avaliação do desempenho do classificador muitas vezes envolve o uso de uma matriz de confusão, que detalha classificações corretas e incorretas para cada classe (Monard; Baranauskas, 2017).

A escolha adequada de atributos, o tratamento de ruídos e o gerenciamento de overfitting e underfitting são fatores críticos para a conclusão do modelo no aprendizado supervisionado. A compreensão desses conceitos fundamentais é essencial para desenvolver classificadores eficazes.

Já o modelo não supervisionado realiza análises de dados sem nenhum tipo de padronização, buscando exemplos ou não rotulados, onde o indutor procura identificar padrões e agrupamentos nos dados. A função principal é determinar se alguns exemplos podem ser agrupados de alguma maneira, formando clusters.

Após a identificação desses agrupamentos, uma análise adicional é geralmente necessária para compreender o significado de cada conjunto no contexto do problema sendo analisado. Os dois modelos têm funções diferentes para a análise de dados, visto que o supervisionado é orientado a previsão e classificação, enquanto o não supervisionado tem o enfoque na descoberta de padronizações e estruturas dos dados necessitando de uma análise adicional. (Monard; Baranauskas,2017).

Tende-se então o aprendizado de máquina utilizando de métodos estatísticos e computacionais, semelhantes ao aprendizado estatístico, entretanto tendo como aliado a inteligência artificial, para extração automática de resultados a partir dos dados aplicados.

De acordo com Morettin e Singer (2020), a evolução do aprendizado estatístico e a evolução computacional estão correlacionados, acompanhando mudanças benéficas, onde em 1960 eram necessários máquinas manuais ou elétricas para realizar cálculos, após duas décadas com o surgimento do computador, os algoritmos permitiram melhor capacidade e funcionalidade, o método estatístico incluem diversos métodos como citados acima permitindo, melhores resultado e precisões.

2.2 USANDO O DATA SCIENCE NA MITIGAÇÃO DE FAKE NEWS

Sabemos que o aprendizado de máquina e o aprendizado de estatística são grandes aliados, buscando na revisão bibliográfica há estudos que utilizaram destes para identificação de notícias falsas. Shoemaker (2019), realizou um estudo que consistia na automação da detecção de notícias falsas, através da padronização do conjunto utilizado, fazendo com que o aprendizado de máquina extraísse características dos textos, realizasse contagem de palavras, utiliza se também a frequência dos termos abordados, e a lematização que consiste em diminuir palavras para uma melhor análise, muito utilizado em processo de linguagem natural (PLN), além da análise de sentimento. Neste foram utilizados os classificadores Naive Bayes e Floresta Aleatórias (Random Forest), para concluir o processo de aprendizado a autora utilizou três conjuntos de dados, sendo um site

que fornecia informações falsas, um arquivo que continham informações verdadeiras e outro conjunto apenas com fake news. Os resultados mostraram que durante os treinamentos e testes, que o método Naive Bayes classificou com mais precisão se comparado à Floresta Aleatória, notou-se também que características como a contagem de palavras e a frequência de termos se mostraram mais eficazes para classificação se comparado a lematização e análise de sentimento.

Já Ahmed (2021), realizou uma revisão de estudo que consistia em três questões, por que é necessário o aprendizado de máquina para detecção das fake news, quais os classificadores supervisionados poderiam ser utilizados para tal e como os classificadores eram treinados para detecção.

Os resultados obtidos neste estudo mostraram que o crescimento da utilização da internet e redes sociais facilitou a divulgação das notícias falsas, prejudicando não só pessoas como indivíduos, mas organizações e o ramo corporativo. Viram o aprendizado de máquina como uma maneira de solucionar a verificação de notícias falsas, visto que essa tarefa para ação humana demandaria muito tempo e dificuldade. Partindo deste, foram utilizados alguns classificadores como método Naive Bayes, florestas aleatórias, máquina de vetores de suporte e também redes neurais. O treinamento dos classificadores consiste na identificação de características semelhantes entre o conjunto analisado, como a frequência de termos utilizados, termos diferenciados e também a desordem das palavras. Esses modelos de identificação são nomeados em Term Frequency-Inverse Document Frequency, Inverse Document Frequency, N-Gram e Bag of Words, que convertem o texto para o entendimento da máquina, ensinando-a na classificação.

Deu-se a conclusão que o classificador Naive Bayes, que consiste em um aprendizado acessível, que utiliza o método probabilístico Bayes, com resultados mais robustos, se mostrou competente para detecção das fakes news. Entretanto, a máquina de vetores de suporte apresentou maior precisão nos resultados.

Em 2021 Wazib Ansar e Saptarsi Goswami, realizaram um estudo propondo a mitigação das notícias falsas utilizando o aprendizado de máquina junto a estatística, os autores relatam também sobre os malefícios causados à sociedade, em período de pandemia que ocorreu em 2019, para identificar quais princípios por

trás da desinformação. Os autores concluem, a necessidade de mais estudo para a utilização do data science para redução das notícias falsas.

2.3 FAKE NEWS

Desde os primórdios, os seres humanos demonstraram a necessidade de estabelecer comportamentos para compreensão, expressão para aprimoração da relação entre eles. Segundo Recuero (2000), a comunicação foi desenvolvida através da luta pela sobrevivência, em busca de transmissão de mensagens, ideias e até sentimentos. De acordo com Bragança (2000), a comunicação é uma ferramenta fundamental para a humanidade, responsável por diversos desenvolvimentos sociais ao longo dos anos, seguindo Eras. Após passar por constantes evoluções e aprimoramento, chegamos a Era da computação, atual período em qual vivemos, com a popularização dos computadores seguindo para uma fase mais digital, possibilitando melhoria nos meios de comunicação.

Nesta era digital há uma imensa facilidade de se comunicar, a tecnologia nos permite uma comunicação mais ampla, com vídeos, imagens, sons, seja a curta ou a longa distâncias, além da agilidade que uma mensagem pode percorrer. Entretanto, com inúmeros recursos, há também possibilidades diversas que podem ser utilizadas para desenvolvimento de uma comunicação benéfica, mas também uma comunicação maléfica, através da desconstrução e distorcimento de uma mensagem.

De acordo Riemsdijk et al. (2020), uma notícia é denominada a um produto valioso e específico no nicho da comunicação, podendo despertar interesse ao público de acordo com sua escrita, exemplificando como em casos de manchetes sensacionalistas que auxiliam na popularização de acessos a sites e blogs que muitas vezes não tem ligação nenhuma com a mensagem transmitida, visando apenas o alcance do público alvo e lucros por trás de cada acesso a página.

Os autores ainda ressaltam a diferenciação de paródia, propagandas, sátiras, manipulação e fabricação das fakes news, citando que a utilização da mínima credibilidade pode ser ocultada na elaboração da escrita.

A partir dessas diferenciações as informações devem ser bem averiguadas, visto que o chamado Distúrbio da Informação que categorizam a informação incorreta, quando não há intenção de prejudicar, a desinformação, denominada a manipulações, teorias conspiratórias e boatos com intuito de causar danos e a má informação caracterizada por informações confidenciais divulgada sem autorização para fins prejudiciais (Derakhshan;Wardle, 2017).

Para prevenção e combate às distorções e o entendimento inicial do processo de criação das fakes news, desde a pessoa que elabora a notícia a importância que a mensagem passa, até chegar ao receptor que é impactado e a partir daí decidir suas decisões. Segundo o Derakhshan e Wardle(2017), o processo de recriação e compartilhamento de uma mensagem falsa tem um efeito mais aumentativo sobre a sociedade, aumentando o alcance das pessoas gerando cada vez mais visualizações, acessos e popularização e mais uma vez gerando lucros.

De acordo com Allcot e Gentzkow (2017), a motivação para a criação de notícias falsas é o retorno financeiro que ela pode causar e também a ideologia ligada diretamente a valores e crenças da população.

Derakhshan e Wardle (2017), confirmam tais motivações, complementando com mais duas motivações, sendo política com o intuito de impactar a opinião pública durante períodos de eleição e também questões psicológicas como forma de afirmação de demais pessoas ligadas a determinados assuntos.

A propagação das fake news ganharam notoriedade em movimentos políticos em 2016, segundo Gonçalves (2019) partindo de uma decisão da formação da União Europeia, defendendo a permanência do Reino Unido no grupo e em seguida a eleição presidencial norte americana, onde as notícias falsas foram utilizadas como métodos de derrubar os oponentes e também a imprensa, gerando desconfiança nas informações transmitidas pelo jornalismo.

No Brasil a disseminação das notícias seguiu os passos políticos internacionais, entretanto com um período eleitoral reduzido se comparado ao padrão, além de diversas restrições aos candidatos para promoção de suas campanhas (Gonçalves, 2019). Partindo destes, a corrida presidencial foi marcada pela propagação de notícias falsas nas redes sociais, causando insegurança nas decisões políticas da

população, mas impactando positivamente o retorno financeiro aos criadores de conteúdos enganosos (Sousa Júnior, Petroll e Rocha (2019)).

Durante o período de pandemia, no ano de 2019 não só o país, mas o mundo se viu de frente a propagação irrefreável de desinformações, um período caracterizado pelas incertezas e medo de um vírus desconhecido. De acordo com Raasch et al (2020), foi um grande desafio rebater todas as falsas notícias que surgiram neste tempo, o que gerou um desafio não só para jornalistas e a imprensa, mas também cientistas na área da saúde, que corriam contra o tempo para a busca da prevenção da doença.

O compartilhamento de notícias sobre o COVID 19 nas redes sociais, cresceu expressivamente, mesmo sem saber a veracidade do conteúdo, a população reproduzia as mensagens como método de alerta outras pessoas, muitas dessas desinformações repercutiu negativamente na aversão da vacina como método de prevenção e também na conscientização sobre a doença (Raasch et al, 2020), vale ressaltar que no Brasil até o ano de 2023 houve 706 mil mortes, segundo o portal oficial do governo do Brasil.

2.4 CASOS DE FAKE NEWS

Além da política e saúde, no Brasil alguns casos regionais foram expostos a público de como as fake news são prejudiciais, em maio 2014, Fabiane Maria de Jesus (1981,2014) foi a primeira vítima da desinformação após ser confundida com uma suposta sequestradora de crianças para rituais religiosos, a população local da região do litoral norte paulista, tomaram uma atitude drástica de espancar a mulher até a morte. O caso ganhou repercussão internacional, por ser o primeiro óbito por conta de notícias falsas. Outro caso surgiu na mesma região após 8 anos, como o do Osil Vicente Guedes (1974,2023) que foi vítima de uma informação falsa pelo furto de uma motocicleta, o que gerou revolta nos moradores espancando o homem, que veio a óbito horas depois.

Já no norte do país, em Belém, Messias Silva de Almeida teve sua imagem denegrada por ser um motorista de aplicativo que furtava os pertences dos

passageiros, após a sinalização de outras pessoas a Messias, o mesmo foi até a delegacia informar sobre a desinformação gerada contra ele (g1.globo, 2023).

No México, outra vítima da desinformação pelas redes sociais, foi linchado e queimado vivo por ser um suposto ladrão de crianças, Daniel Picazo (1991,2022) morreu após moradores atearam fogo no corpo do turista, que estava de passagem pela região (g1.globo, 2023).

É notório as consequências causadas pela desinformação, seja em busca de reconhecimento, lucros, ou questões psicológicas e ideológicas como citado por Derakhshan e Wardle (2017).

Com diversos acontecimentos, o governo do brasileiro desenvolveu o programa “Brasil Contra Fake”, portal que auxilia e educa digitalmente a população sobre a distinção de notícias verdadeiras e falsas envolvendo o governo federal, além de auxiliar em como deve ser denunciado notícias falsas nas redes sociais. Neste programa o caso da Fabiane foi reconhecido e utilizado para conscientizar a população sobre as consequências e a destruição que uma mentira digital pode ocasionar.

Parte das mídias digitais também se unificaram ao combate a fake news, não só a opção de denúncia dentro de plataformas como Instagram, Facebook, YouTube entre outros. Há também agências de *fact checking*, denominadas portais online que comprovam a veracidade ou não informações que viralizam na internet, ressaltando que são agências voluntárias, sem possuir fins lucrativos.

3. METODOLOGIA

Além da revisão bibliográfica qualitativa, foi aplicado um formulário desenvolvido pela ferramenta Google Forms e foi disponibilizado ao público por um período de 10 dias, sendo totalmente online pelo período do dia 17 ao dia 27 de setembro de 2023. O link foi compartilhado através das mídias sociais. O questionário continha 8 imagens noticiadas com fonte, imagem e escrita. Dentro deste tinha-se a mistura de informações verdadeiras e falsas e também opções de como foram realizadas as distinções das notícias, para explorar o lado cognitivo da pessoa que preenchia o

formulário, além de sexo e idade para uma melhor granularidade dos dados. Cada notícia continha 5 questões totalizando 40 questões, dentre os temas abordados, foram adicionados temas voltados à política, saúde, informações gerais e entretenimento.

4.DIUSSÃO E RESULTADOS

O intuito deste formulário foi explorar o conhecimento e também o senso analítico das pessoas que foram questionadas, notando quais fatores eram utilizados para distinção das informações ali aplicadas e de forma qualitativa o que tornava aquela notícia verdadeira ou falsa.

Os resultados foram cento e cinco (105) pessoas, dentre elas 62 mulheres e 43 homens. Por uma divisão etária , 33 pessoas possuíam 18 a 25 anos, 55 pessoas que continham 26 a 45 anos e com 46 a 65 anos teve-se 16 pessoas e apenas uma pessoa com mais de 66 anos.

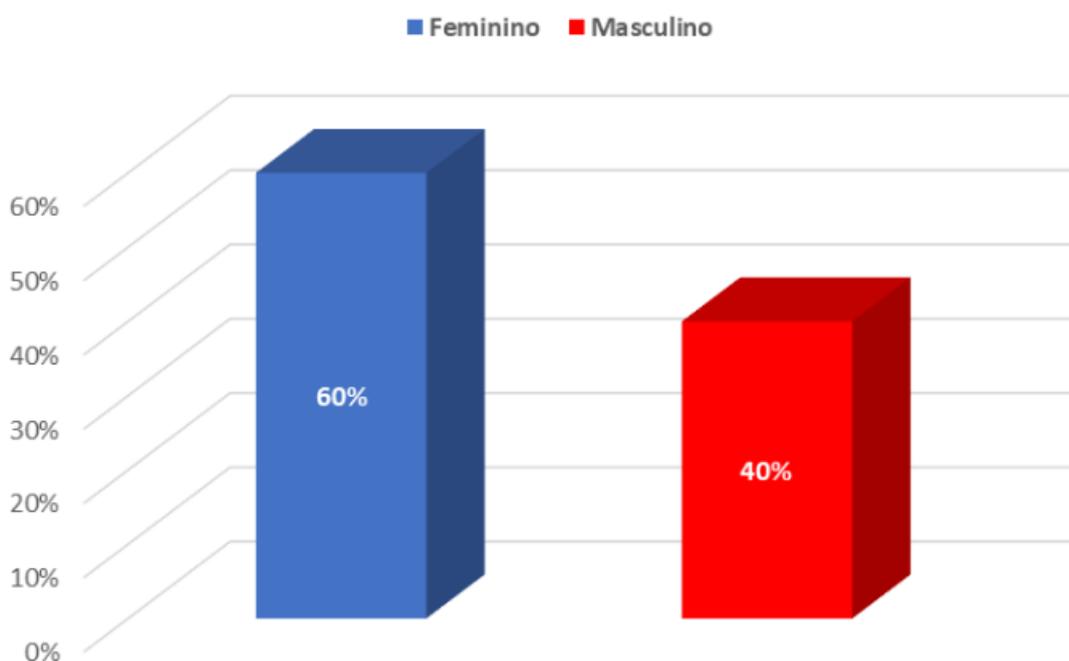


FIGURA 1: NOTÍCIA COM ALTO ÍNDICE DE ERRO: RESULTADO POR SEXO

Fonte: próprio autor.

Dentre as notícias falsas, a informativa sobre a abertura do concurso público da linha ferroviária do estado de São Paulo apresentou o maior resultado de pessoas

que acreditaram, sendo sua grande parte, pessoas com faixa etária entre 26 e 46 anos sendo predominante o sexo feminino.

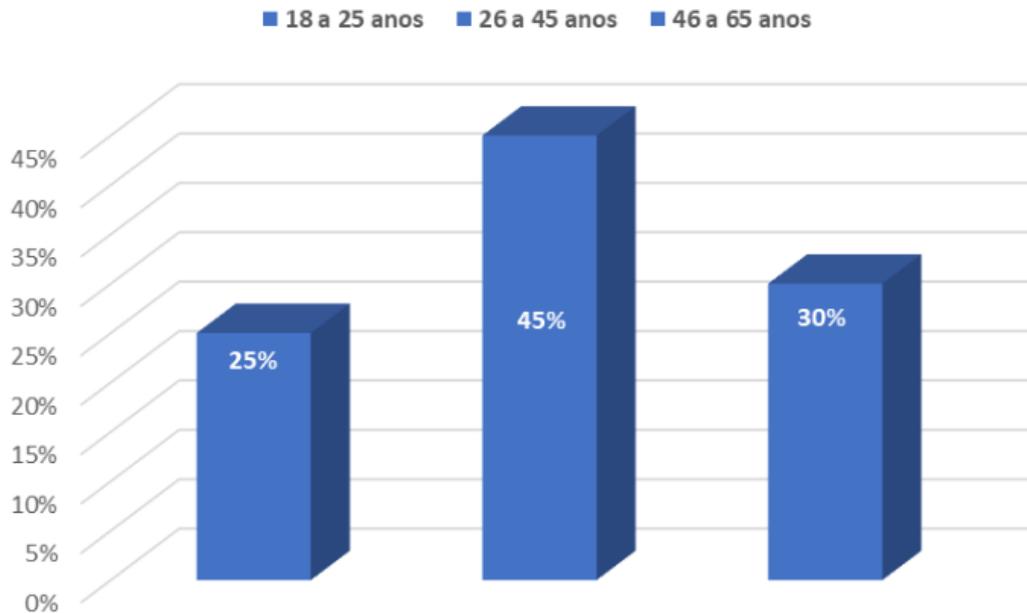


FIGURA 2: NOTÍCIA COM ALTO ÍNDICE DE ERRO: RESULTADO POR IDADE

Fonte: próprio autor.

A imagem continha uma escrita correta, despertando confiança no indivíduo para acreditar na informação transmitida, no formulário contemplava a justificativa da escolha onde grande maioria informou que a distinção fora justamente pelo método de formatação contextual da notícia. Isso demonstra que os homens entre 26 e 46 anos souberam melhor distinguir a informação e como distinção utilizaram a fonte, a origem da notícia.

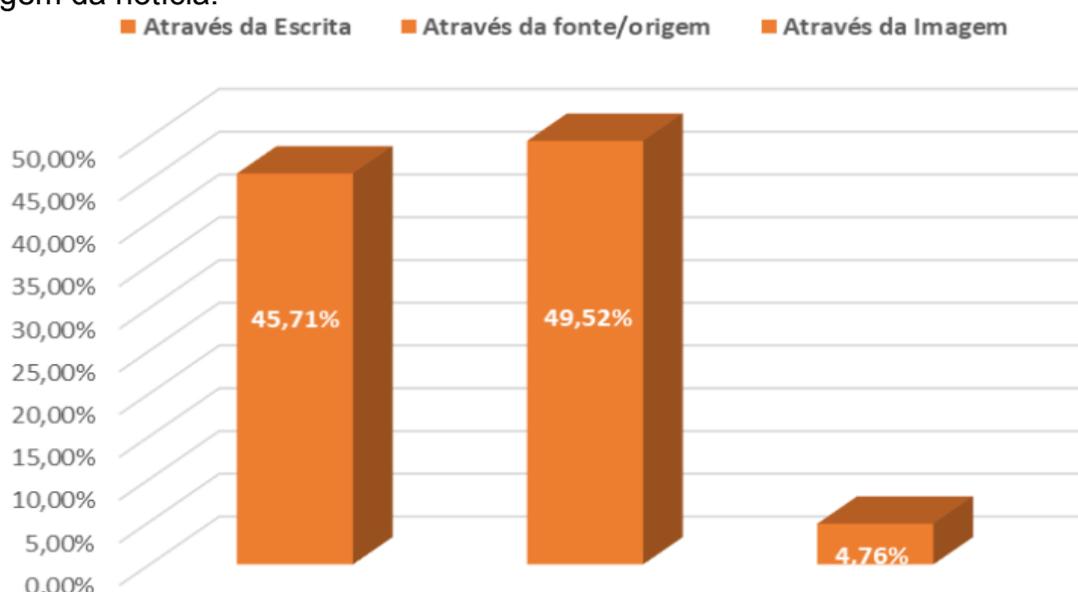


FIGURA 3: NOTÍCIA COM ALTO ÍNDICE DE ERRO: MÉTODO DE DISTINÇÃO

Fonte: próprio autor.

Já o informativo que houve grande número de acertos se trata de uma notícia falsa, nesta a escrita se tornava duvidosa pelo método sensacionalista alertando a pessoas que tomaram a vacina contra pólio contraíram câncer e se tornaram inférteis. A notícia contemplava a imagem de um médico em um evento do CDC (centro de controle de doenças), e a publicação tinha como fonte o site Tribuna Nacional, pouco conhecido no mundo digital. Nas respostas obtidas, a fonte e a escrita da notícia denunciaram sua veracidade, com predominância da escrita, e com grande maioria do gênero feminino entre as idades 26 e 46, totalizando 55 respostas. Vale ressaltar a predominância de mulheres, devido a sua participação em sua grande maioria no questionário.

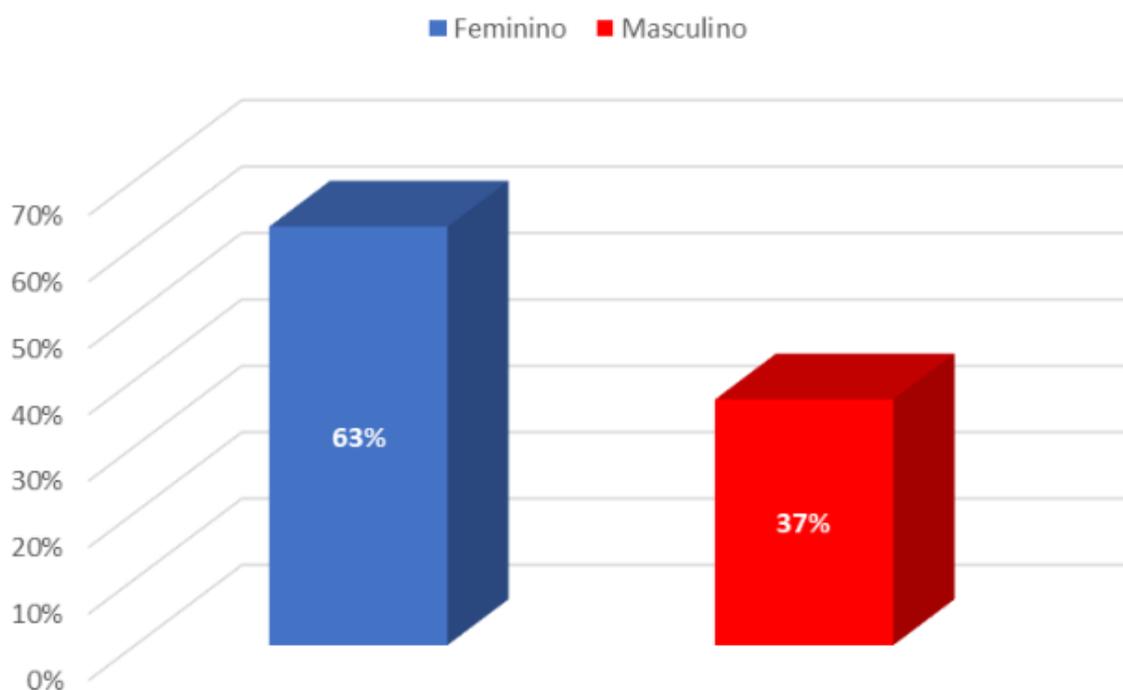


FIGURA 4:NOTÍCIA COM ALTO ÍNDICE DE ACERTO: RESULTADO POR SEXO

Fonte: próprio autor.

Em uma análise assertiva de notícias verdadeiras, a faixa etária que soube melhor distinguir as informações foram 26 e 46 anos, já entre gêneros, não houve diferenciação exata havendo diversas respostas mostrando que o discernimento não tem relação direta com o sexo.

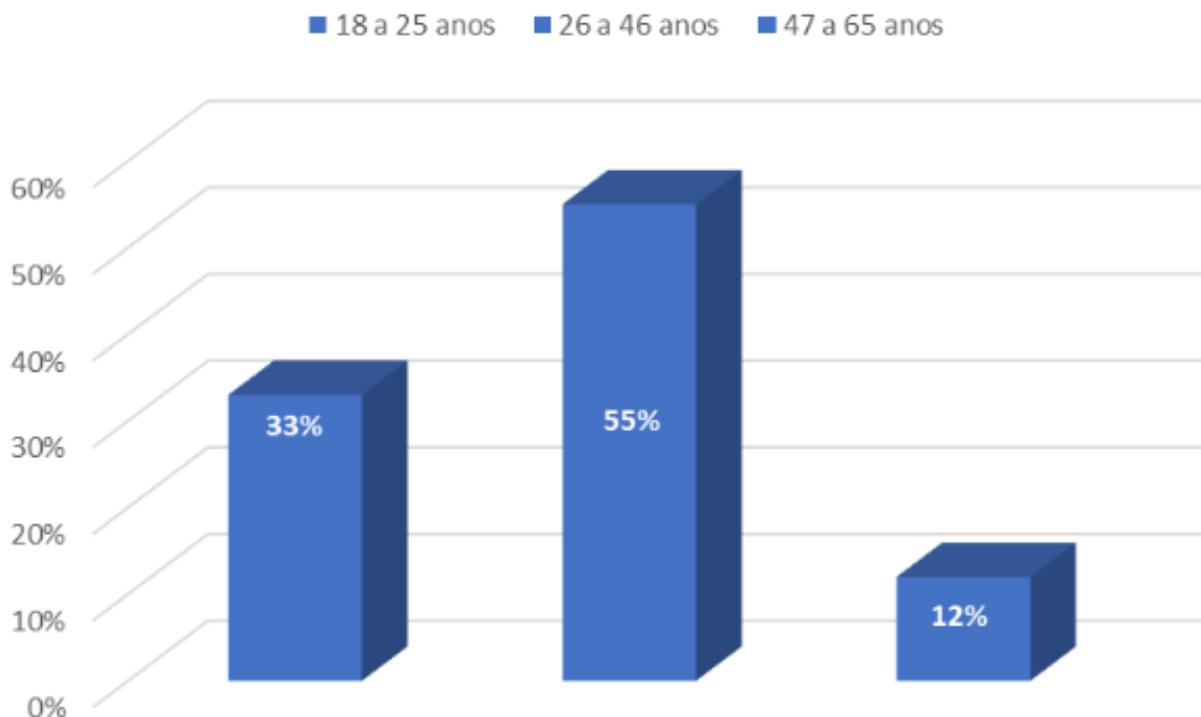


FIGURA 5: NOTÍCIA COM ALTO ÍNDICE DE ACERTO: RESULTADO POR IDADE
 Fonte: próprio autor.

Já os resultados por crença, em uma divisão por gênero conclui -se tanto o sexo feminino quanto ao masculino, informaram não acreditar nas notícias com base na imagem apresentada, relatando que a escrita e fonte são pontos essenciais para distinção das informações verdadeiras de falsas, pois é um fator que gera uma desconfiança aguçando o lado analítico. Em todas as faixas etárias demonstraram grande parte do público, partem deste mesmo método para identificar informações falsas.

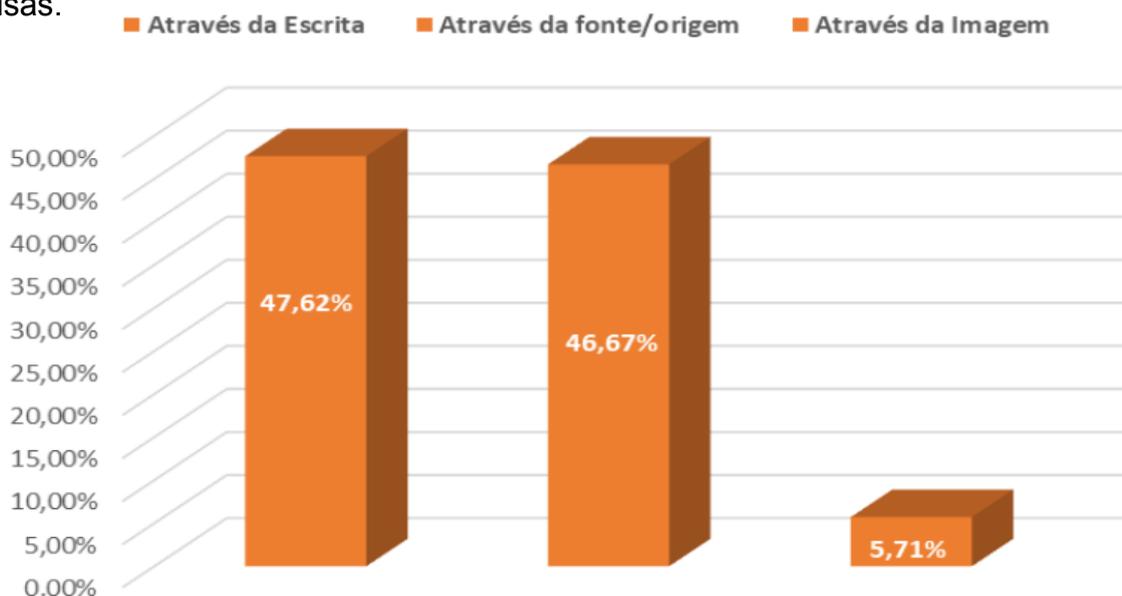


FIGURA 6: NOTÍCIA COM ALTO ÍNDICE DE ACERTO: MÉTODO DE DISTINÇÃO
 Fonte: próprio autor.

5. CONSIDERAÇÕES FINAIS

A evolução da comunicação foi um grande passo para a humanidade, surgindo na era primitiva como método de sobrevivência, passando por diversas modificações e eras, se aprimorando e tornando essencial para evolução humana. Com a comunicação se desenvolveu a informação além do surgimento de jornais e aparelhos eletrônicos, a o início da era digital, a comunicação se tornou abrangente alcançando diversos lugares.

A evolução digital por sua vez, possibilitou o acesso a internet que permitiu que as informações percorrem de uma maneira mais rápida a alcances mais longos, além de uma fácil acessibilidade. A maneira facilitada de se comunicar modificou a maneira de recebermos e encaminharmos informações, e neste caminho a mensagem pode ser modificada propositalmente, causando dúvidas e desinformação a quem recebe. Este tipo de desinformação foi denominado fake news e se propagou em períodos de grandes decisões sociais, como em período de eleição e durante a pandemia. Essas informações falsas ganharam investidores que acreditam na ideologia que está ligada a valores e crenças da sociedade. Entretanto, esse tipo de desinformações ocasionou também mortes pelo país e no mundo, mostrando-se totalmente prejudicial, não se tratando apenas de opiniões a partidos políticos, mas de vidas de pessoas.

A ciência de dados é uma opção de como pode ser mitigado as notícias falsas, com o aprendizado de máquina e estatística, ramos que sustentam o data science. A aplicação de modelos probabilísticos como o Bayes e o suporte vetorial de máquina que se mostraram eficientes nos estudos apresentados.

Os autores citados como Ahmed (2021) e Shoemaker (2019), encerram os estudos concluindo com resultados semelhantes, ambos demonstram o método Naive Bayes eficaz, e demonstra também a necessidade de mais estudos para utilização da ciência de dados para mitigação. Partindo destes, conclui-se que é possível a utilização do data science para mitigação das fake news, entretanto há uma necessidade contínua do aprendizado das máquinas, visto que o mundo passa por

constantes acontecimentos, e grande parte são noticiados e com uma comunicação de acesso facilitado através da internet, permite não só a modificação deste, mas também o compartilhamento gerando insegurança e desinformação social.

De acordo com o questionário aplicado a público, homens e mulheres demonstraram confiar na qualidade da escrita e na fonte da notícia para distinção de informações verdadeiras e falsas, entretanto gera atenção neste dados, visto que Riemsdijk et al. (2020), relataram, que o método mais utilizado para despertar interesse é a escrita utilizando da mínima credibilidade para a fabricação de notícias falsas.

Partindo disto mostra-se a necessidade de uma educação digital robusta não apenas em sites governamentais e agências de *fact checking*, mas uma possível força tarefa nacional utilizando uma colaboração do governo com as principais redes sociais utilizadas no país, para melhores distinções de informações voltada a todo o público, independentemente da faixa etária ou gênero, para melhor distinção das informações recebidas, considerada a minha pesquisa ressaltou a importância de avaliar uma informação por um todo não apenas pela escrita e fonte, mas também os outros elementos que podem conter na notícia.

Concluo que neste trabalho agregamos conhecimento ao entendermos os fundamentos da ciência de dados, originadas da estatística e aprendizado de máquina, a revisão de estudos realizados para identificação de notícias falsas, que mostraram eficiência entretanto necessitando de constantes aprendizados, além das fases de comunicação e o surgimento das fake news e o impacto delas na sociedade. Por fim, uma perspectiva do público avaliado no questionário aplicado, mostrando atenção nos fatores de distinção utilizados pelos entrevistados.

O objetivo deste trabalho foi alcançado confirmando que a ciência de dados pode ser uma das ferramentas para mitigação, mas não a solução principal, entretanto a hipótese não se tornou válida, a machine learning necessita de constantes aprendizados, demonstrando que combater as notícias falsas, se torna um desafio tecnológico e social reforçando a necessidade de uma educação digital destinada à população para melhor distinção das informações recebidas.

Aos meus colegas acadêmicos e a instituição de ensino, deixo este trabalho sugerindo para que explorem e continuem o estudo deste tema para próximos anos, para demonstrar que a tecnologia é capaz de transformar cada vez mais o ambiente digital em uma fonte segura e confiável de informações.

REFERÊNCIAS

Ansar, A.; Goswami.S.Combating the menace: A survey on characterization and detection of fake news from a data science perspective. Disponível em: elsevier.com/locate/jjime.

BARROS, Álvaro Gonçalves de; SOUZA, Carlos Henrique Medeiros de; TEIXEIRA, Risiberg. Evolução das Comunicações até a Internet das Coisas: A Passagem para uma Nova Era da Comunicação Humana. Instituto Federal do Rio de Janeiro – Campus Arraial do Cabo; Universidade Estadual do Norte Fluminense, 2020. Disponível em: <https://portalespiral.cp2.g12.br/index.php/cadernos/article/view/3065>. Acesso em: 3 de novembro de 2023

Bragança, I.Evolução Da Comunicação Humana Podemos Explicar A História Da Existência Humana Através Das Etapas Dos Meios de Comunicação. Disponível em:<https://pt.scribd.com/doc/16088693/Evolucao-da-comunicacao-humana-Podemos-explicar-a-historia-da-existencia-humana-atraves-das-etapas-do-desenvolvimento-da-comunicacao>. Acesso em: 2 de novembro de 2023.

Dacorso, S.T.M; Neto, T.G. O BOOM DAS FAKE NEWS: O COMEÇO DA ERA DAS MENTIRAS.CADERNOS DE PSICOLOGIA – CESJF - jun.2019 v.1 n.1 p.589-608. jun-2021.

Fagundes V.O; et al.Jovens e sua percepção sobre fake news na ciência.Bol. Mus. Para. Emílio Goeldi. Cienc. Hum., Belém, v. 16, n. 1, 2021.

Gonçalves, A; Sousa, C.H.M; Teixeira, R. Evolução das Comunicações até a internet das coisas: a passagem para uma nova era da comunicação humana.

<https://covid.saude.gov.br/>. Acesso em [02 de novembro de 2023]

<https://g1.globo.com/sp/santos-regiao/noticia/2022/06/15/oito-anos-apos-mulher-ser-espancada-ate-a-morte-em-sp-fake-news-segue-fazendo-vitimas-como-o-turista-que-imado-vivo-no-mexico.ghtml> Acesso em [03 de novembro de 2023]

<https://g1.globo.com/sp/santos-regiao/noticia/2023/05/10/o-que-falta-saber-sobre-a-morte-do-homem-brutalmente-agredido-e-vitima-de-acusacao-falsa-de-roubo-de-mo-to.ghtml>. Acesso em [03 de novembro de 2023]

Morettin,A.P. Singer, J.M.Introdução à Ciência de Dados Fundamentos e Aplicações.Junho, 2020. Disponível em <https://www.ime.usp.br/~jmsinger/MAE0217/cdados2020jun03.pdf> Acesso em: 15 nov.2023

Monard, M.C; Baranauskas J.A.Conceitos sobre Aprendizado de Máquina. Disponível em: <https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>. Acesso em 8 de nov.2023

Riemsdijk.G et al. O fator Fake News na Atualidade, na mira da psicologia. International Journal of Developmental and Education Psychology. InFAD Revista de Psicologia, Nº1- Volumen 1,2020. ISSN. pp:255-262.

Ripoll, L., & Canto, F. L. do. (2019). Fake news e "viralização": responsabilidade legal na disseminação de desinformação. Revista Brasileira De Biblioteconomia E Documentação, 15, 143–156.

SABBATINI, Renato M.E., PhD. A Evolução da Inteligência - Parte 5: Linguagem e Evolução. Revista Cérebro & Mente, Fevereiro/Abril 2001. Disponível em: https://cerebromente.org.br/n12/mente/evolution/evolution05_p.html. Acesso em: [17 de novembro de 2023].

SHOEMAKER, Eliza. Using Data Science to Detect Fake News. James Madison University, Senior Honors Projects, Spring 2019. Disponível em: <https://commons.lib.jmu.edu/cgi/viewcontent.cgi?article=1754&context=honors201019>. Acesso em: [17 de novembro de 2023].