CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
CAMPUS TIMÓTEO

Athos Caetano de Assis

A EFICIÊNCIA DOS MÉTODOS PARA DETECÇÃO AUTOMÁTICA DE *FAKE NEWS* NA INTERNET: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Timóteo

2020

Athos Caetano de Assis

A EFICIÊNCIA DOS MÉTODOS PARA DETECÇÃO AUTOMÁTICA DE *FAKE NEWS* NA INTERNET: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Monografia apresentada à Coordenação de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, campus Timóteo, para obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Maurílio Alves Martins da Costa
Coorientador: Deisymar Botega Tavares

Timóteo

2020

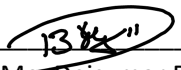
Athos Caetano de Assis

A EFICIÊNCIA DOS MÉTODOS PARA DETECÇÃO AUTOMÁTICA DE FAKE NEWS NA INTERNET: UMA REVISÃO SISTEMÁTICA DA LITERATURA

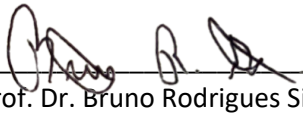
Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, campus Timóteo, como requisito parcial para obtenção do título de Engenheiro de Computação.

Trabalho aprovado. Timóteo, 11 de dezembro de 2020:

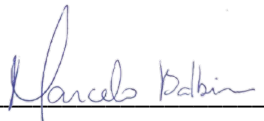
Prof. Dr. Maurilio Alves Martins da Costa
Orientador



Prof. Me. Deisymar Botega Tavares
Coorientadora



Prof. Dr. Bruno Rodrigues Silva
Professor Convidado



Prof. Me. Marcelo de Sousa Balbino
Professor Convidado

Timóteo
2020

.

.

.

Primeiramente, toda honra e toda glória a Deus.
Dedico a todos aqueles que, de alguma forma,
me sustentaram nesta jornada, árdua e prazerosa,
de busca pelo conhecimento.
Este trabalho é o nosso resultado.

Agradecimentos

O resultado deste trabalho é constituído por muitas pessoas, muitos momentos e muitas emoções. Devido a isso, muitos agradecimentos precisam ser realizados.

Agradeço à Deus, pela seu eterno amor e pela sua infinita misericódia, que se renova a cada manhã. Por Ele conceder a salvação e sempre querer o melhor para os seus servos, mesmo não sendo merecedores. Pelo sustento diário ao longo desta jornada, presenteando com saúde, ânimo, paciência, inteligência e sabedoria.

Agradeço aos meus pais, Adélio Carlos de Oliveira Assis e Magna Morais Caetano de Assis, por todos os conselhos, sempre mostrando os melhores caminhos sob seus pontos de vista. Por investirem tempo de vida para com seu primogênito, dando amor, carinho e cuidado. Espero poder retribuir em vida parte deste esforço dispensado.

Agradeço ao meu irmão, Daniel Caetano de Assis, pelo constante companheirismo. Por deixar meus dias mais leves e felizes com seu amor, humor e amizade. Saiba que estou disposto a ajudá-lo em tudo que for necessário.

Agradeço à minha namorada, que mesmo chegando em sua reta final tornou-se fundamental para o alcance desta conquista. Obrigado pela paciência, pelas doces palavras de incentivo e confiança. Que possamos compartilhar muitas conquistas no decorrer desta vida.

Agradeço aos meus amigos e amigas, aos antigos e recentes, com os quais tive a oportunidade de dividir momentos únicos e incríveis ao lado. Vocês sempre serão lembrados.

Agradeço a todos os profissionais da instituição CEFET-MG, sem nenhuma exceção. Em especial, quero deixar exposto minha admiração pelos professores, os verdadeiros heróis desta sociedade. Obrigado por divirem seus conhecimentos, tanto acadêmicos quanto de vida, para com seus alunos.

Agradeço ao meu primeiro chefe, Rodrigo Muzzi, por oferecer minha primeira oportunidade no mercado de trabalho, por acreditar em meu potencial. Espero ter contribuído para o sucesso da empresa.

Por fim, sou grato a todos aqueles que, de alguma forma, me sustentaram nesta jornada, árdua e prazerosa, de busca pelo conhecimento.

“Os grandes navegadores
devem sua reputação aos temporais e tempestades”.
Epicuro

Resumo

A internet transformou a comunicação, potencializou o alcance e a velocidade de propagação das informações e modificou o comportamento das pessoas na busca destas informações. Em reflexo, os meios tradicionais de notícias, como rádio e televisão, foram substituídos por plataformas onlines, principalmente mensageiros instantâneos, blogs e redes sociais. O problema desta mudança de comportamento é que o ambiente virtual não apresenta nenhuma restrição ou verificação de veracidade das informações que são criadas e compartilhadas, permitindo que qualquer tipo de notícia se espalhe, incluindo *fake news*. A disseminação de *fake news* se tornou um dos maiores problemas enfrentados pela nossa sociedade, atingiu tal proporção que influenciou o resultado das eleições presidenciais dos Estados Unidos, a maior economia do mundo. Para alcançar seus objetivos, os autores de *fake news* manipulam as pessoas e colocam em risco o equilíbrio de autenticidade do ecossistema de notícias. Diante do cenário exposto, o presente trabalho realizou um levantamento bibliográfico em algumas das principais bases científicas da área de ciências exatas, *ACM Digital Library*, *Google Scholar*, *IEEE Xplore* e *ScienceDirect*, acerca do grau de eficiência dos métodos descritos em artigos sobre detecção automática de *fake news* na internet. Nossa contribuição para a academia pode ser separada em duas partes. Na primeira, identificamos os cinco métodos prevalentes divididos em duas linhas de pesquisa, baseada em conteúdo e baseada em contexto, e relatamos sobre as estratégias utilizadas por cada. Na segunda, listamos mais de 60 ferramentas para detecção automática de *fake news* advindos dos artigos relevantes, propomos uma classificação destas ferramentas mediante quatro critérios (linha de pesquisa, método, campo de conhecimento e eficácia) e analisamos os resultados. Logo, este trabalho pode ser um material fundamental tanto para quem necessita entender o fenômeno das *fake news* quanto para quem deseja conhecer ou desenvolver métodos e ferramentas de combate.

Palavras-chave: fake news, notícias falsas, métodos, detecção automática, levantamento bibliográfico, revisão sistemática.

Abstract

The internet was capable to change and optimize the reach and the speed of spread of information and this changed the people's behavior at search that information. Accordingly, the traditional media of information, such as radio and television, has been replacing by online platforms, mainly instant messengers, blogs and social medias. The problem of this behavior changing is that the virtual environment does not have restriction or veracity check of the information that is being created and shared, allowing any type of news to be spread, including fake news. The spread of fake news has become one of the biggest challenges to our society and has even reached the presidential elections of the country with largest economy in the world, the United States. To achieve your goals, the fake news authors manipulate people and put at risk the balance of authenticity of the news ecosystem. Therefore, this document carried out a bibliographic survey on some of the main scientific bases in IT Field (Information Technology), ACM Digital Library, Google Scholar, IEEE Xplore e ScienceDirect, about the grade of efficiency of the described methods in articles regarding automatic detection of fake news on the internet. Our contribution to the academy can be divide in two parts. First, we identified the five main automated methods divided in two research lines, content-based and context-based, and we report about the strategies used in each one. Second, we listed more than 60 mechanisms for automatic detection of fake news from relevant articles, we put forward a classification to each mechanisms through four standards (research line, method, field of knowledge and efficiency). In brief, this document can be a fundamental material for those who need to understand the fake news phenomenon and for who want to know or develop methods and mechanisms against this new phenomenon.

Keywords: fake news, methods, automatic detection, bibliographic survey, systematic review.

Lista de ilustrações

Lista de tabelas

Sumário

1 Introdução

“A resposta certa, não importa nada:
o essencial é que as perguntas estejam certas”.
Mário Quintana

A rede mundial de computadores revolucionou a maneira como as informações são disseminadas, possibilitando a cada indivíduo procurar, consumir e gerar conteúdos. Em decorrência disto, as organizações tradicionais de informação, como rádio, jornal e televisão, perderam espaço e audiência na busca por notícias (SHU et al., 2017). De acordo com Levy et al. (2017), a internet ultrapassou a televisão como sendo a principal fonte de notícias para os jovens entre 18 e 24 anos, ilustrando assim, o avanço desta tendência.

A mudança no comportamento na busca de notícias pode ser esclarecida, em parte, pela popularização dos dispositivos móveis (JUNIOR, 2016), visto que simplificaram o acesso a internet e, como resultado, permitiram com que as pessoas tivessem contato com plataformas online consideradas mais atrativas. De modo sucinto, plataformas online são serviços disponíveis na internet que facilitam a comunicação entre dois ou mais conjuntos distintos de usuários simultaneamente (OECD, 2019). Segundo Shu et al. (2017), além destas plataformas serem mais acessíveis que as organizações tradicionais, elas facilitam o compartilhamento de notícias entre as pessoas e promovem um espaço mais oportuno para comentários e debates.

Dentre as diversas plataformas online existentes, como mecanismos de buscas, mensageiros instantâneos e *streamings*, as redes sociais virtuais apresentam grande relevância, se destacam tanto em poder de comunicação, quanto em quantidade de usuários (JIN et al., 2013). As redes sociais virtuais são serviços nos quais um indivíduo consegue criar perfis públicos ou semipúblicos em um sistema e os vincular a uma lista de outros usuários com os quais deseja distribuir informações (BOYD; ELLISON, 2007). O foco principal das redes sociais é estabelecer o relacionamento entre seus usuários e, para tanto, permitem entre eles uma interação virtual sobre diversos assuntos e o acesso a diferentes conteúdos.

Durante a navegação nas redes sociais virtuais um indivíduo recebe muitas informações, sendo alcançado por notícias livremente criadas e compartilhadas. Deste modo, as chances dele se deparar com notícias falsas produzidas intencionalmente, denominadas *fake news* (ALLCOTT; GENTZKOW, 2017), são grandes. Conforme Shu et al. (2017), *fake news* são elaboradas com o propósito de confundir seus receptores. Para alcançar este objetivo elas exploram a diversificação das formas de apresentação de seu conteúdo, utilizam de variações linguísticas e, ao mesmo tempo, desvirtuam notícias verdadeiras.

Na prática, as estratégias utilizadas para confundir os receptores de *fake news* se apresentam eficazes e cada vez mais as pessoas são enganadas. De acordo com Figueira e Oliveira (2017), até mesmo os mais jovens, que normalmente possuem maior capacidade e conhecimento tecnológico, parecem confusos tanto como o restante da sociedade para identi-

ficar *fake news*. Uma pesquisa realizada pela organização *Common Sense Media* revelou que pelo menos um em cada quatro adolescentes, de 10 a 18 anos, já compartilhou *fake news* na internet, comprovando que as vítimas desta manipulação também não se limitam a idade.

Diante do exposto, questionamentos são feitos sobre a capacidade das pessoas em diferenciar o que é verdadeiro e o que é falso no mundo virtual (FIGUEIRA; OLIVEIRA, 2017). De acordo com Vosoughi, Roy e Aral (2018), *fake news* se espalham em maior quantidade, velocidade e profundidade, quando comparado com notícias verdadeiras, porque os usuários das redes sociais virtuais apresentam maior probabilidade de compartilharem as mesmas em seus perfis, evidenciando assim, a complexidade presente em torno da identificação de *fake news* e a ausência de habilidade nesta atividade por parte dos seres humanos.

A alta propagação de *fake news* pode romper o equilíbrio de autenticidade do ecossistema de notícias, manipular pessoas para aceitarem crenças tendenciosas, e até mesmo modificar o jeito como elas interpretam e respondem a notícias verdadeiras (SHU et al., 2017). Em virtude disso, a disseminação deliberada de *fake news* foi enumerada pelo Fórum Econômico Mundial como uma das principais ameaças à nossa sociedade (HOWELL et al., 2013). Diante desta situação o combate às *fake news* tem sido ampliado, propagandas educativas e campanhas de conscientização se tornaram mais recorrentes e ferramentas para detecção automática estão sendo desenvolvidas por pesquisadores e cientistas (CONROY; RUBIN; CHEN, 2015).

A detecção automática pode ser o caminho mais eficiente para diminuir os impactos causados pelas *fake news* (LAZER et al., 2018), entretanto criar ferramentas que desempenham esta tarefa é desafiador. Dentre os motivos para a afirmação anterior estão as várias etapas necessárias para analisar a veracidade de uma determinada notícia, a inexistência de uma governança para controlar o que as pessoas podem ler e qual domínio elas acessam, como também a exigência de uma classificação rápida e precisa (PARIKH; ATREY, 2018). Todos esses obstáculos abrem margem para várias abordagens durante o desenvolvimento destas ferramentas, possibilitando a aplicação de diferentes métodos em torno de uma possível solução.

Sob tal enfoque, este trabalho teve como objetivo responder a seguinte pergunta: qual é o grau de eficiência dos métodos descritos em artigos sobre detecção automática de *fake news* na internet? Partindo deste propósito, surgiu então a necessidade da realização de um levantamento bibliográfico para descobrir e descrever os fundamentos contidos nestes métodos e seus desempenhos atuais.

1.1 Objetivos

O presente trabalho teve como objetivo geral realizar um levantamento bibliográfico sistemático sobre o grau de eficiência dos métodos descritos em artigos sobre detecção automática de *fake news* na internet.

Para atingir o objetivo geral citado acima este trabalho buscou ainda:

1. Levantar os métodos prevalentes nos artigos encontrados;
2. Relatar a estratégia e os fundamentos destes métodos;
3. Classificar os artigos encontrados conforme método e grau de eficiência descritos.

Sendo, portanto, os objetivos específicos deste trabalho.

1.2 Justificativa

A disseminação de *fake news* atingiu um novo patamar na era digital, a internet potencializou o alcance e a velocidade de propagação desta forma de manipulação. As plataformas online, principalmente as redes sociais virtuais, permitiram com que os mais variados tipos de conteúdo, inclusive notícias, fossem compartilhados entre as pessoas sem nenhum filtro, investigação de fatos ou julgamento editorial, admitindo assim, a reprodução de *fake news*. De fato, no decorrer deste novo mundo conectado, *fake news* se tornaram cada vez mais presentes em nosso cotidiano.

Diversas são as motivações por detrás da elaboração e disseminação de *fake news*, entre as predominantes estão as ideológicas e financeiras. Segundo Allcott e Gentzkow (2017), esta atividade faz parte de um mercado lucrativo com receitas de publicidade e manipulação de pessoas. Para alcançar seus objetivos os divulgadores de *fake news* utilizam diferentes ferramentas, *bots* e *clickbaits* aumentam a propagação de informações distorcidas e imprecisas circulando na internet, adquirindo grande potencial para causar impactos reais.

Os impactos prejudiciais causados pelo consumo de *fake news*, como a fomentação de boatos e o aumento do extremismo, também tomaram novas proporções, sendo capazes de fazer um número considerável de vítimas. Um conjunto complexo de aspectos sociais, psicológicos e cognitivos contribuem para a vulnerabilidade dos seres humanos diante desta manipulação. De acordo Roets et al. (2017), dependendo do nível de habilidade cognitiva de um indivíduo, a influência causada pelas fake news não pode ser desfeita. Em geral, mesmo após a verdade ser exposta, as pessoas continuam confiando pelo menos parcialmente nas informações que elas sabem serem falsas.

Diante disto, as *fake news* tornaram-se um tema de discussão importante para toda a sociedade, procurar entender este fenômeno e combatê-lo é fundamental. Uma intervenção apresentada como possível solução para este problema é o uso da tecnologia para o desenvolvimento de ferramentas que consigam detectar automaticamente as *fake news*, acabando com seu fluxo de disseminação e a manipulação de pessoas. Estas ferramentas deveriam contribuir para a prevenção em primeira instância, evitando a exposição de indivíduos a *fake news*.

Neste cenário, ferramentas para detecção automática de *fake news* na internet vêm sendo elaboradas por pesquisadores e cientistas mediante a aplicação de diferentes métodos (TACCHINI et al., 2017). Geralmente, estes métodos são documentados em algum tipo de trabalho acadêmico, como artigos, dissertações, teses, entre outros, servindo como material de

consulta. Contudo, os documentos encontrados sobre este tema exploram, em sua maioria, um único método e uma única linha de pensamento, limitando em amplitude o conhecimento dos leitores e sendo necessário que os mesmos procurem em diversos lugares e fontes quando necessitam de maiores informações.

Finalmente, por se tratar de assunto de interesse de toda a comunidade e de suma importância, produzir um documento abrangente acerca do grau de eficiência dos métodos descritos em artigos sobre detecção automática de *fake news* na internet, além de ser um modo de descoberta de conhecimento e de enriquecimento da literatura acadêmica nacional, pode servir como fonte de auxílio para aquelas pessoas que desejam aprender sobre o tema ou buscam direcionamento na elaboração de futuros projetos.

1.3 Estrutura da monografia

Esta monografia foi estruturada em sete capítulos, ordenados pelo momento em que foram concluídos dentro do ciclo de vida deste trabalho, e anexos, a saber:

- As bases teóricas são abordadas no capítulo 2. Para auxiliar o entendimento do leitor, conceitos importantes sobre *fake news* e revisão sistemática da literatura são mostrados.
- O capítulo 3 abrange os procedimentos metodológicos através dos quais este trabalho se desenvolve, expondo com detalhes a elaboração de todas as etapas da revisão sistemática da literatura efetuada, desde o processo de pesquisa até a apresentação dos dados coletados.
- Os principais métodos descritos em artigos sobre detecção automática de *fake news* na internet, advindos da revisão sistemática, são definidos ao decorrer do capítulo 4. Os fundamentos contidos em cada um deles são detalhados e, alguns, representados visualmente.
- O desenvolvimento do trabalho é abordado no capítulo 5, todo processo de aplicação dos procedimentos metodológicos foi narrado, os dados coletados apresentados e os resultados da revisão sistemática da literatura são exibidos.
- No capítulo 6 uma análise embasada nos resultados é realizada, dando destaque sobretudo ao grau de eficiência dos métodos descritos nos artigos. Nela, possíveis relações também são apontadas.
- Finalmente, no capítulo 7 são feitas as conclusões e considerações finais. Nele são relatadas as principais contribuições, limitações deste trabalho e indicadas algumas direções para trabalhos futuros.
- O trabalho inclui parte dos seus resultados e produtos em anexos, tendo em vista que sua extensão poderia comprometer a legibilidade do texto e a compreensão do leitor. Os artigos bases não são disponibilizados entre os anexos pois isso violaria alguns direitos de propriedade intelectual dos seus autores, entretanto, outros meios para acessá-los são indicados nas referências bibliográficas.

2 Fundamentação teórica

“O período de maior ganho em conhecimento e experiência
é o período mais difícil da vida”.
Dalai Lama

Neste capítulo são abordados os conceitos fundamentais para o desenvolvimento deste trabalho com o intuito de fornecer ao leitor suporte necessário para uma melhor compreensão. Primeiramente, diversas definições de *fake news* encontradas na literatura são apresentadas, seguidas pelo seu contexto histórico, aspectos gerais e impactos. Finalmente, uma dissertação sobre revisão sistemática complementa a seção.

2.1 *Fake news*

O termo *fake news* é considerado recente na literatura, diversas são as definições propostas para o mesmo, contudo, não existe um acordo estabelecido (SHU et al., 2017). Tal constatação pode ser feita tendo em vista uma revisão sobre o termo em publicações acadêmicas realizada por Jr, Lim e Ling (2018), na qual são identificadas seis maneiras diferentes pelas quais os autores destas obras buscam estabelecer e caracterizar as *fake news*.

Enquanto algumas definições são bem amplas, a exemplo de Marchi (2012) que também considera as notícias em forma de sátira e paródia transmitidas pela TV em programas de entretenimento como sendo *fake news*, outras são bastante restritas, levando em conta de forma rigorosa o ambiente, conteúdo e intenção. Lazer et al. (2018) conceitua *fake news*, de forma específica, como informações produzidas para imitar as notícias das organizações tradicionais, como rádio e televisão, em seu conteúdo, mas não em sua intenção ou processo de desenvolvimento.

Em relação às definições restritas, uma característica comum evidenciada nas mesmas é como as *fake news* tentam se assemelhar com notícias verdadeiras, procurando apresentar uma suposta imagem de legitimidade e credibilidade através do modo como são exibidas e escritas. Entretanto, quando ponderamos a intenção por detrás das *fake news*, a semelhança termina (LAZER et al., 2018).

A propósito desta característica, pelo contexto deste trabalho e pretendendo eliminar as ambiguidades entre *fake news* e conceitos relacionados, optou-se pela definição de Allcott e Gentzkow (2017), amplamente adotada em estudos recentes e pautado em duas ideias: autenticidade e finalidade. Segundo os autores, *fake news* são notícias, comprovadamente falsas, criadas de forma intencional para enganar seus leitores.

2.1.1 Contexto histórico

Apesar do termo recente, *fake news* são antigas, possuem quase a mesma quantidade de tempo que a imprensa, inventada na idade média (SHU et al., 2017). Por meio desta invenção, as notícias, tanto verdadeiras quanto falsas, começaram a circular amplamente e a modificar o cotidiano das pessoas. Os *canards* ilustram bem este novo cenário, consistiam em folhetos vendidos nas esquinas de cidades francesas no século XI com diversos conteúdos, inclusive notícias enganosas. Em um deles era informado que um monstro capturado no Chile estaria sendo enviado para a França, o que causou pequeno alvoroço na região (BURKHARDT, 2017).

Considerando a evolução dos meios de comunicação desde os *canards*, é perceptível que o ambiente midiático das *fake news* foi ampliado e alterado com o tempo, transitou-se do papel para o rádio, televisão e, mais recentemente, a internet (SHU et al., 2017).

Ao longo desta transição, normas jornalísticas de equilíbrio e objetividade foram instauradas para que *fake news*, como a da descoberta de vida na lua publicada pelo famoso jornal estadunidense *New York Sun*, não fossem divulgadas (ALLCOTT; GENTZKOW, 2017). De acordo com Lazer et al. (2018), os oligopólios criados pelos dominantes das tecnologias de distribuição de notícias, ou seja, as organizações tradicionais de informações, conseguiram sustentar essas normas até a popularização da rede mundial de computadores. No entanto, ainda segundo o autor, a internet permitiu a entrada de novos concorrentes no mercado de notícias, muito dos quais rejeitaram as normas antes adotadas, e abalou o modelo de negócio das organizações tradicionais que desfrutavam de grande confiança e credibilidade do público.

Compreende-se, assim, como as *fake news* se tornaram um problema em escala mundial, a ponto de influenciar as eleições presidenciais dos Estados Unidos em 2016, onde as mesmas foram usadas para ajudar um ou outro candidato (ALLCOTT; GENTZKOW, 2017). Principalmente durante e após este fato, *fake news* se tornou um assunto destacado frequentemente pela mídia e muito popular entre a sociedade, tanto que, no mesmo ano das eleições, o termo foi nomeado a palavra do ano pelo dicionário *Macquarie* (SHU et al., 2017).

2.1.2 Propagação e motivação

O processo de geração, busca e consumo de notícias modificou-se com a *world wide web*, a quantidade de pessoas que procuram se informar através da internet, em substituição às organizações tradicionais, cresce cada vez mais. Entretanto, este ambiente não possui mediador e é considerado desafiador ao jornalismo, visto que todo indivíduo com acesso a internet passou a ser capaz de desempenhar o papel de jornalista, produzindo e publicando notícias em plataformas online (JR; LIM; LING, 2018).

Por sua vez, as diversas plataformas de divulgação de notícias existentes na internet, como blogs, portais e principalmente as redes sociais, fornecem mecanismos para atingir um grande público de maneira acessível, com baixo custo e sem nenhuma regulamentação para garantir a precisão e credibilidade das informações (LAZER et al., 2018). Assim, qualquer tipo de notícia, criada por qualquer indivíduo, pode se tornar viral e ser propagada rapidamente,

incluindo *fake news*.

A proliferação de *fake news* também se deve ao engajamento por parte de seus consumidores, usuários que as curtem e compartilham. Dentre os motivos para este engajamento, podemos citar a má índole por parte de algumas pessoas, o fato de que a maioria raramente verifica as informações que compartilha (JR; LIM; LING, 2018), e o principal, muitas não conseguem identificar o teor implícito das notícias, fenômeno que pode ser explicado por aspectos comportamentais e psicológicos.

Conforme Shu et al. (2017), os dois principais fatores que tornam os consumidores de informação vulneráveis às *fake news* são: o realismo direto e o viés de confirmação. Enquanto o primeiro afirma que os consumidores tendem a acreditar que sua percepção da realidade é absoluta e aqueles que discordam da mesma são desinformados e irracionais, o segundo declara que estes mesmos consumidores possuem maior aceitação por notícias que confirmem seus pontos de vista. Devido a esses vieses cognitivos, ou tendências de pensamento, pertencentes à natureza humana, as *fake news* são compreendidas como verdadeiras pelos consumidores.

Além da falta de habilidade das pessoas em diferenciar notícias reais e falsas, os *bots* sociais, presentes de forma considerável nas redes sociais, também podem ampliar a disseminação de *fake news* nos momentos em que interagem com as mesmas. Segundo Shu et al. (2017), *bots* referem-se a contas nas redes sociais controlados por algoritmo de computador, muitas vezes de difícil identificação, para interagir com humanos e produzir conteúdo automaticamente. Sendo assim, *bots* podem ser desenvolvidos com o objetivo de manipular, enganar e causar danos com a disseminação de *fake news* como nas eleições presidenciais de 2016 dos Estados Unidos, onde os mesmos distorceram discussões online na semana que antecedeu o pleito (LAZER et al., 2018).

As motivações para propagar *fake news* principais são a tentativa de ganhar receita com publicidades geradas pela mesma, através de visualizações ou cliques nos sites em que a *fake news* é publicada. Outra motivação é a ideológica, onde a propagação busca impor, convencer ou manipular o consumidor sobre uma ideia ou pensamento (ALLCOTT; GENTZKOW, 2017). Em casos menos comuns, os produtores de *fake news* tem como motivação a diversão, apresentam prazer em ludibriar outras pessoas.

2.1.3 Impactos

Shu et al. (2017) aponta o rompimento do equilíbrio de autenticidade do ecossistema de notícias como um dos impactos causados pela proliferação das *fake news*, e, pesquisas atuais, comprovam a declaração do autor. Estudos recentes mostraram que as *fake news*, devido ao sensacionalismo presente em suas manchetes ou textos, possuem maiores chances de serem espalhadas pela rede, sendo compartilhadas mais rapidamente e por mais pessoas do que as notícias verdadeiras (LAZER et al., 2018). Considerando que, em uma rede, as notícias ganham mais credibilidade à medida que são distribuídas, chega um momento em que um limiar é cruzado e acredita-se que uma *fake news* é verdadeira por parte daquela comunidade (FIGUEIRA; OLIVEIRA, 2017).

Nesse processo, saber quantos indivíduos se depararam ou compartilharam uma *fake news* não é o mesmo que saber quantos visualizaram seu conteúdo e foram afetados pelo mesmo (LAZER et al., 2018). Quando uma *fake news* passa a ser assimilada como verdadeira é muito difícil corrigir este equívoco, em alguns casos a correção de informações falsas pela apresentação de informações verdadeiras e factuais não é apenas inútil para reduzir as percepções errôneas, mas às vezes pode até aumentá-las, especialmente entre grupos ideológicos. Desta maneira, *fake news* podem ser usadas para manipular pessoas a aceitarem crenças tendenciosas e até mesmo modificar o jeito como elas interpretam e respondem a notícias reais (SHU et al., 2017).

Além disso, de acordo com Allcott e Gentzkow (2017), o aumento do cinismo, da apatia e do extremismo provocado pela propagação de *fake news* podem provocar diversos impactos, inclusive sociais. A diminuição dos índices de engajamento em campanhas de vacinação e o declínio da confiança nos principais meios de comunicação, caindo para mínimos históricos nos últimos anos, são alguns dos exemplos.

2.2 Revisão sistemática da literatura

Uma das maneiras de conseguir maior precisão e melhores níveis de credibilidade em uma revisão bibliográfica é adotar uma abordagem sistemática, isto é, utilizar uma estratégia e um método sistematizado para realizar as buscas e analisar os resultados. A condução rigorosa desta abordagem permite ao pesquisador responsável pela revisão bibliográfica aprimorar suas hipóteses, definir a condução adequada para buscar e selecionar dados e encontrar direções para futuras pesquisas. Em decorrência destes benefícios, outros pesquisadores podem desfrutar dos resultados obtidos com maior confiabilidade, seja reutilizando em novos estudos ou aplicando em projetos (CONFORTO; AMARAL; SILVA, 2011).

De acordo com Levy e Ellis (2006), revisão sistemática da literatura é um processo que vai desde a coleta até a avaliação de um conjuntos de artigos científicos com o propósito de um criar um embasamento sobre determinado assunto. Kitchenham (2004) segue a mesma linha de pensamento, para o autor a revisão sistemática é um meio de conhecer e verificar toda pesquisa disponível relevante sobre uma área de interesse. Rother (2007) complementa as definições anteriores afirmando que se trata de uma revisão planejada para responder uma pergunta específica empregando criticamente métodos explícitos e sistemáticos ao longo de todos os passos.

É oportuno lembrar que, a realização de uma revisão sistemática da literatura é somente possível após a publicação de muitos estudos sobre a área ou assunto a ser investigado, intitulados estudos primários, e que, mesmo consumindo bastante recurso e tempo, ela ainda é mais rápida do que iniciar um novo estudo completo (CONFORTO; AMARAL; SILVA, 2011). O resultado de uma revisão sistemática, ou estudo secundário, é dependente da qualidade de suas fontes, deve constituir o estado da arte e gerar conhecimento além do existente, isto é, conhecimento complementar aos estudos primários (LEVY; ELLIS, 2006).

Algumas revisões sistemáticas de literatura incluem uma síntese estatística dos re-

sultados dos estudos, conhecida como metanálise. De acordo com Higgins et al. (2019), a metanálise é um método estatístico onde resultados de estudos independentes são combinados e sintetizados, fornecendo estimativas mais precisas, facilitando as investigações de consistência e diferenças entre estudos. Ainda segundo os autores, como a metanálise não é obrigatória na composição de uma revisão sistemática, parte da comunidade científica designa revisões sistemáticas que possuam este recurso com outra nomenclatura.

3 Procedimentos metodológicos

“Um bom começo é a metade”.
Aristóteles

Este trabalho consiste em uma pesquisa exploratória, estabelecida por Raupp e Beuren (2006), realizada em bases de dados científicas sobre o tema proposto. Pesquisas desta natureza buscam esclarecer questões superficialmente abordadas sobre determinado assunto, gerando um conhecimento mais profundo e compreensível. Por meio de uma revisão sistemática da literatura, baseada no método desenvolvido por Kitchenham et al. (2009) e com adaptações de Teixeira (2016), almejou-se identificar o grau de eficiência dos métodos descritos em artigos sobre detecção automática de *fake news* na internet. O procedimento utilizado e suas etapas são detalhadas a seguir.

3.1 Processo de pesquisa

A busca de estudos primários foi efetuada em `julho de 2019` nas bases de dados científicas relacionadas abaixo, na tabela 1.

Tabela 1 – Relação das bases de dados científicas consultadas.

Bases	**Acrônimos**
ACM Digital Library	ACM
Google Scholar	GS
IEEE Xplore	IEEE
ScienceDirect	SD

Fonte: elaborada pelo autor.

As bases *ACM Digital Library*, *IEEE Xplore* e *ScienceDirect* foram escolhidas por serem consideradas referência em repositórios de obras das áreas de ciências exatas, possuindo uma coleção abrangente de publicações relacionadas à tecnologia. O *Google Scholar*, por sua vez, foi escolhida por ser um dos maiores mecanismos de pesquisa acadêmica, possibilitar acesso simples a publicações da língua portuguesa, manter elevado nível de transparência e capacidade de atualização (HADDAWAY et al., 2015).

As palavras chaves foram elaboradas a partir de artigos suportes encontrados nas bases de dados científicas, para tanto, utilizou-se nas buscas os termos *fake news detection* e *fake news identification*. Os artigos suportes foram determinados após estudo e pré-análise com o intuito de confirmar que atendiam a abordagem desta pesquisa exploratória. A tabela 2 apresenta as *strings* definidas após aplicação desta estratégia e sua ordem de prioridade nos motores de busca.

Tabela 2 – *Strings* para busca em português.

Ordem	***Strings* (PT)**
1	(notícias falsas detecção) *AND* (automática *OR* automatizada)
2	(notícias falsas detecção) *AND* método
3	(notícias falsas detecção) *AND* (classificação *OR* verificação)
4	(notícias falsas detecção) *AND* (online *OR* internet)
5	(notícias falsas detecção) *AND* (mídia social *OR* rede social)
6	notícias falsas detecção
7	(notícias falsas) *AND* método
8	(notícias falsas) *AND* (classificação *OR* verificação)
9	(desinformação) *AND* método
10	(desinformação) *AND* (classificação *OR* verificação)

Fonte: elaborada pelo autor.

Em alguns casos, como os mecanismos de processamento dos motores de busca das bases consultadas possuem especificações diferentes, tornou-se necessária algumas adaptações nas *strings*. Um exemplo a ser citado é a tradução das palavras chaves, a princípio elaboradas em português, pelo fato de algumas bases serem desenvolvidas em inglês. Assim, as *strings* definidas para bases com esta característica está exposta na tabela 3.

Tabela 3 – *Strings* para busca em inglês.

Ordem	***Strings* (EN)**
1	*(fake news detection) AND (automatic OR automated)*
2	*(fake news detection) AND method*
3	*(fake news detection) AND (classification OR verification)*
4	*(fake news detection) AND (online OR internet)*
5	*(fake news detection) AND (social media OR social network)*
6	*fake news detection*
7	*(fake news) AND method*
8	*(fake news) AND (classification OR verification)*
9	*(misinformation) AND method*
10	*(misinformation) AND (classification OR verification)*

Fonte: elaborada pelo autor.

3.2 Critérios de inclusão e exclusão

Após a busca inicial para encontrar publicações potencialmente relevantes, também chamadas de estudos primários, parâmetros de seleção foram estabelecidos com a finalidade de obter uma melhoria nos resultados e diminuir as chances de insucesso. Nesta etapa dois passos foram realizados, o primeiro deles visa examinar as características externas das publicações através dos critérios de inclusão e exclusão definidos e relatados abaixo.

Inclusão

- Publicações em formato de artigo científico;
- Publicações veiculadas entre `janeiro de 2014` e `julho de 2019`;
- Publicações escritas no idioma inglês ou português;
- Publicações disponíveis em sua íntegra.

Exclusão

- Publicações que não tenham relevância com tema proposto, ou seja, não dispõem de informações sobre métodos para detecção automática de *fake news* na internet;
- Publicações desprovidos de resumo, introdução ou conclusão;
- Publicações veiculadas em órgãos não reconhecidos;
- Publicações não disponibilizadas gratuitamente ou não acessíveis via assinatura do portal Periódicos da Capes feita pela instituição CEFET-MG;
- Publicações repetidas (quando vários relatos de um mesmo estudo foram encontrados, apenas a versão mais completa foi incluída na revisão).

O segundo passo examina certas partes internas dos artigos de forma manual, uma triagem, com três filtros de pesquisa, foi realizada para verificar quais seriam os estudos primários com maior probabilidade de contribuir com respostas para a questão de pesquisa levantada. Neste caso, quando um artigo atendia os requisitos de todos os filtros ele era incluído na seleção, no contrário, excluído. Os filtros de pesquisa aplicados são:

1. Ler o título de cada artigo incluído na etapa anterior e verificar sua aderência ao tema deste trabalho;
2. Ler o resumo de cada artigo incluído no filtro anterior e verificar sua aderência ao tema deste trabalho;
3. Ler a introdução e conclusão de cada artigo incluído no filtro anterior e verificar sua aderência ao tema deste trabalho.

3.3 Avaliação de qualidade

Ainda preocupado com a qualidade dos artigos relevantes selecionados na etapa anterior optou-se por fazer uma avaliação de qualidade dos mesmos utilizando a estratégia elaborada por Teixeira (2016). Todos os artigos relevantes foram integralmente lidos e avaliados conforme sua importância para com esta pesquisa exploratória. Mediante a atribuição de notas em uma escala crescente de 1 a 5, como mostrado no tabela 4, pretendeu-se gerar um índice de qualidade para cada artigo e assim aprimorar os resultados.

Tabela 4 – Parâmetros para avaliação de qualidade dos artigos relevantes.

Notas	**Parâmetros**
1	Apresenta pouco conteúdo relevante sobre os métodos para detecção automática de *fake news* na internet.
2	Apresenta uma ferramenta de detecção automática de *fake news* com poucos detalhes e pouco conteúdo relevante sobre o método aplicado.
3	Apresenta uma ferramenta de detecção automática de *fake news* com muitos detalhes e pouco conteúdo relevante sobre o método aplicado.
4	Apresentam muito conteúdo relevante sobre os métodos para detecção automática de *fake news* na internet.
5	Apresenta uma ferramenta de detecção automática de *fake news* com muitos detalhes e muito conteúdo relevante sobre o método aplicado.

Fonte: elaborada pelo autor.

Artigos relevantes que obtiveram notas abaixo de 3 não integraram os estudos primários deste trabalho, a relação de nome e demais informações desses artigos estão disponíveis no Apêndice A. Todavia, por serem relevantes, seus dados também foram coletados e apresentados conforme exposto na seção 3.4.

3.4 Coleta e apresentação dos dados

Todos os artigos considerados relevantes, mesmo aqueles com mínima relevância, tiveram alguns dados específicos coletados e registrados com precisão, sendo importantes para entendimento do contexto no qual este trabalho foi desenvolvido. De cada artigo extraiu-se os seguintes dados: sua obra completa, sua referência completa, a base de dados científica onde foi publicado e o seu número de citações.

A obra completa, além de ser extraída para leitura e desenvolvimento do trabalho, possibilitou classificar cada artigo com um índice de qualidade e encontrar aqueles com maior relevância. A referência completa, mesmo sendo parte obrigatória na composição do trabalho, foi necessária para visualização do número de artigos publicados durante cada ano, permitindo notar a variação de interesse pelo tema da detecção automática de *fake news* no decorrer do tempo. A extração da base de dados científica facilitou identificar em qual era possível encontrar mais documentos sobre o tema. E por último, o número de citações proporcionou

mensurar se o tema era recente ou pouco explorado na literatura.

Baseado no dados mencionados anteriormente, uma apresentação foi elaborada na tentativa de fazer uma comparação e um resumo de maneira qualitativa e quantitativa. Nesta apresentação temos um conjunto de gráficos originado de cada dado extraído. A apresentação dos dados e a aplicação da etapa de coleta destes dados, bem como das etapas precedentes, podem ser vistas com mais clareza no capítulo 5.

4 Métodos para detecção de fake news

“Não se possui o que não se compreende”.
Johann Goethe

De acordo com Elhadad, Li e Gebali (2019), a detecção de *fake news* pode ser definida como o processo de estimar se uma notícia específica, de qualquer natureza, é intencionalmente ou involuntariamente enganosa. Mahid, Manickam e Karuppayah (2018) complementam a definição anterior estabelecendo o conceito de ferramentas que realizam esta tarefa de maneira automática: são técnicas ou sistemas que auxiliam as pessoas com recursos e funções na previsão de conteúdo de notícias enganosas. Ainda segundo os autores, essas ferramentas exigem um complexo planejamento e elaboração.

Em nossa revisão sistemática da literatura foram encontrados diferentes métodos descritos nos artigos a respeito de detecção automática de *fake news* no ambiente online, entre eles estão os métodos de verificação de fatos, verificação de adulteração visual, verificação de credibilidade e outros. Entretanto, os diferentes métodos descritos podem ser classificados em duas linhas de pesquisa, a baseada no conteúdo de uma notícia e a baseada no contexto social de uma notícia (SHU et al., 2017). Basicamente, a classificação de um método quanto a sua linha de pesquisa é vinculado ao conjunto de informações necessário pelo mesmo para realizar a tarefa de detecção, informações provenientes do conteúdo ou do contexto. Desse modo, neste capítulo optou-se por dividir a apresentação dos métodos conforme sua linha de pesquisa.

4.1 Baseado no conteúdo

A principal linha de pesquisa para detecção automática de *fake news* na internet é a baseada em conteúdo, utilizada na maioria das ferramentas desenvolvidas para esse objetivo (MONTI et al., 2019). Isto porque, a maior parte das notícias que circulam na internet apresentam apenas texto em seu corpo e dele é possível retirar muitos recursos que fornecem pistas para detectar *fake news*. Historicamente, essa linha de pesquisa é também usada para a detecção de *spam* em mensagens de e-mail e páginas da *web*, que também possuem como essência o texto (VEDOVA et al., 2018).

Um método que segue uma linha de pesquisa baseada em conteúdo é aquele que pretende classificar uma notícia como falsa apenas com os recursos contidos no corpo da mesma, sendo que a maior parte desses recursos são textuais. Uma notícia é composta por diferentes elementos como fonte, título, subtítulo, lead, corpo, imagens e vídeo anexados, as informações contidas nestes elementos podem ser fundamentais na atribuição do valor de verdade, classificando-a como falsa ou verdadeira (SHU et al., 2017). A linha de pesquisa baseada em conteúdo contempla três métodos, verificação de fatos, verificação do estilo de escrita e verificação de adulteração visual.

4.1.1 Verificação de fatos

Um dos métodos descritos nos artigos sobre detecção automática de *fake news* na internet que segue uma linha de pesquisa baseada em conteúdo é a verificação de fatos. Fundamentado em uma estratégia desenvolvida pelo jornalismo, este método visa avaliar a autenticidade de uma notícia através da investigação do conhecimento existente em seus elementos textuais. Basicamente, todo conhecimento extraído de uma notícia, por exemplo, suas afirmações e declarações, é comparado com fatos verídicos conhecidos, denominados 'conhecimento verdadeiro', e uma classificação quanto sua autenticidade é efetuada (STAHL, 2018).

A verificação de fatos pode ser realizada de forma manual ou automática, entretanto, neste trabalho estamos interessados somente na segunda perspectiva pois ela consegue se adequar ao grande volume de notícias propagadas no ambiente virtual, principalmente nas redes sociais. Para lidar com a questão de escalabilidade o método da verificação de fatos se utiliza principalmente das técnicas de rede neural, processamento de linguagem natural, recuperação de informação e gráficos de conhecimento.

De maneira geral, a verificação de fatos automática possui apenas duas etapas: a construção da base de 'conhecimento verdadeiro' e a sua comparação com o conhecimento existente na notícia. Todo processo é representado na figura 1.

Figura 1 – Processo do método verificação de fatos.

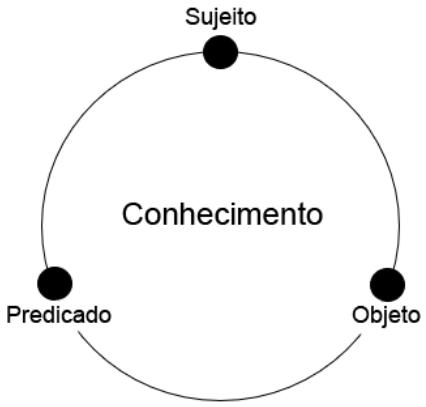
Fonte: elaborada pelo autor.

Antes de descrever cada etapa, é preciso apresentar a definição de conhecimento. Um conceito padrão amplamente adotado em estudos relacionados de conhecimento que pode ser assimilado automaticamente por máquinas é "um conjunto de triplica (Sujeito, Predicado, Objeto) extraído das informações fornecidas que representam satisfatoriamente as mesmas" (ZHOU; ZAFARANI, 2018). Por exemplo, na frase "Jair Bolsonaro é presidente do Brasil" podemos obter o seguinte conhecimento (Jair Bolsonaro, Profissão, Presidente), conforme ideia exposta na figura 2.

Na primeira etapa acontece a extração de conhecimento originário da internet para a construção da base de 'conhecimento verdadeiro (STAHL, 2018). A extração de conhecimento, também conhecida como extração de relação, pode ocorrer mediante duas maneiras nesta etapa, extração de fonte única ou extração de fonte múltipla. A extração de fonte única procura adquirir conhecimento a partir de um único site enciclopédico confiável, como o *Wikipédia*, porém, apesar de ser relativamente eficiente e mais rápida pode levar a um conhecimento

incompleto. A extração de fonte múltipla, por sua vez, incorpora vários sites enciclopédicos confiáveis, gerando um conhecimento mais completo mas menos eficiente.

Figura 2 – Visualização da definição de conhecimento.



Fonte: elaborada pelo autor.

O conhecimento extraído dos sites enciclopédicos são derivados, em sua maioria, de documentos online com informações não estruturadas e são caracterizados por serem redundantes, inválidos, conflitantes, incompletos e não confiáveis, o que poderia resultar em problemas na formação da base de 'conhecimento verdadeiro'. Sendo assim, é necessário que o conhecimento extraído da internet seja limpo através das cinco tarefas listadas (ZHOU; ZAFARANI, 2018):

1. resolução da entidade, para reduzir redundâncias;
2. registro de tempos, para remover conhecimento desatualizado;
3. fusão de conhecimentos, para lidar com conhecimentos conflitantes;
4. avaliação de credibilidade, para melhorar a credibilidade no conhecimento;
5. vínculo de previsão, para inferir novos fatos.

Uma vez realizada a limpeza temos a formação de uma base de 'conhecimento verdadeiro' formada apenas por conhecimento 'limpo' baseado no conjunto de triplos, sujeito, predicado e objeto. Uma forma de representação desta base é por intermédio de gráficos de conhecimento, nele as entidades, isto é, os sujeitos e objetos, são retratados como nós, e os relacionamentos, predicados, são retratados como arestas (STAHL, 2018). Os gráficos de conhecimento são eficientes para fornecer informações básicas para estudo de *fake news*, ou seja, triplas existentes em uma base de 'conhecimento verdadeiros' representam fatos realmente verdadeiros e triplas inexistentes indicam fatos falsos ou desconhecidos.

Na segunda, e última, etapa para definir a autenticidade de uma determinada notícia é necessário extrair todo conhecimento presente em seu conteúdo e compará-lo com os fatos armazenados na base de 'conhecimento verdadeiro' construída na etapa anterior. Geralmente,

para verificar se um fato relatado no conteúdo de uma notícia é verdadeiro é preciso que o mesmo exista no gráfico de conhecimento representante da base de "conhecimento verdadeiro", digo, exista uma aresta denominada Predicado ligando os nós Sujeito e Objeto. A partir desta verificação pode-se classificar se uma notícia é considerada uma *fake news*.

4.1.2 Verificação do estilo de escrita

Outro método método descrito dos artigos sobre detecção automática de *fake news* na internet que segue uma linha de pesquisa baseada em conteúdo é a verificação do estilo de escrita. Este método busca classificar uma determinada notícia com base na intenção de quem escreveu, se o seu autor teve, ou não, como objetivo enganar os consumidores daquela notícia. Normalmente, autores de *fake news* tem a intenção maliciosa de influenciar uma grande quantidade de pessoas por meio da disseminação de informações enganosas e distorcidas, para persuadir as pessoas e atrair sua atenção eles utilizam um estilo de escrita que não é encontrado em notícias verdadeiras (SHU et al., 2017).

Em relação a validade da diferença do estilo de escrita entre *fake news* e notícias verdadeiras, o método em questão está pautado em teorias do estilo do engano, entre as quais podemos citar a teoria do engano interpessoal e a hipótese de *Undeutsch* (SHU et al., 2017; ZHOU; ZAFARANI, 2018). Ademais, em alguns estudos psicológicos forenses foram comprovados que declarações decorrentes de fatos reais diferem em conteúdo e em qualidade das declarações fundamentadas na fantasia (STAHL, 2018). Sendo assim, seria de fato possível caracterizar fraudes no estilo de escrita.

De acordo com Zhou e Zafarani (2018), o estilo de uma *fake news* pode ser definido como sendo "um conjunto de características quantificáveis que podem as representar e as diferenciar de notícias verdadeiras". Em geral, características quantificáveis são recursos de aprendizado de máquina que podem variar desde o número de caracteres existentes na notícia até frases contidas na mesma, verificar tabela 5. As características quantificáveis são extraídas do conteúdo das notícias e analisadas de diferentes maneiras.

As principais formas de investigação do estilo de escrita de uma notícia são as análises léxica, sintática, semântica e de discurso (STAHL, 2018; PARIKH; ATREY, 2018). Na análise léxica é avaliado a frequência estatística de letras, palavras e outros dados por meio de n-gramas. Na análise sintática, mediante tarefas executadas por gramáticas probabilísticas livres de contexto e árvores de decisão, é averiguado a disposição de palavras nas frases, as frases e a relação lógica entre as mesmas. Na análise semântica é examinado os significados das palavras, frases e textos contidos na notícia, para isso procedimentos de consulta linguística e contagem de palavras são usadas. Por último temos a análise de discurso, que captura como ocorrem as construções ideológicas no corpo da notícia pela teoria da estrutura retórica.

No que diz respeito às análises, são utilizados um vetor de características representando o estilo de conteúdo das informações fornecidas em uma estrutura de aprendizado de máquina para classificar uma notícia como enganosa ou não, e em caso positivo, quanto enganosa é a mesma. A maioria das ferramentas para detecção automática de *fake news* que aplicam este método utilizam o aprendizado supervisionado, onde é necessário dados de trei-

Tabela 5 – Características quantificáveis do conteúdo de uma notícia.

Categorias	**Exemplos**
Legibilidade	Índice de *Flesch*
	Índice de *Coleman-Liau*
	Índice de *Dale-Chall*
	Índice de nevoeiro de *Gunning*
Dimensões linguísticas	Contagem de sílabas
	Contagem de palavras
	Contagem de sentenças
	Porcentagem de pronomes
	Porcentagem de artigos
	Porcentagem de preposições
	Porcentagem de verbos
	Porcentagem de advérbios
	Porcentagem de conjunções
Dicas somativas	Pensamento analítico
	Influência
	Tom emocional
Pistas afetivas	Realização
	Ansiedade
	Raiva
	Tristeza
	Emoções positivas e negativas
Pistas cognitivas	Discernimento
	Casualidade
	Discrepância
	Certeza
	Diferenciações
Pistas de informalidade	Palavrões
	Gírias da internet
	Palavras sexuais
Dicas de orientação temporal	Foco no passado
	Foco na atualidade
	Foco no futuro
Dicas de pontuação	Contagem de pontos finais
	Contagem de pontos de interrogação
	Contagem de pontos de exclamação
	Contagem de vírgulas
	Contagem de apóstrofos
	Contagem de parênteses

Fonte: elaborada pelo autor com base em Fernandez e Devaraj (2019).

namento rotulados, ou seja, conjunto de vetores com seus rótulos correspondentes (OSHIKAWA; QIAN; WANG, 2018). Porém, o aprendizado semi-supervisionado também pode ser de grande valia visto que o conjunto de vetores rotulados de artigos de notícias são limitados e a

grande construção de um conjunto com excelência é complexo.

Independentemente da estrutura de aprendizado usada, semi-supervisionada ou supervisionada, o método de verificação do estilo de escrita é um complemento do método de verificação de fatos e pode contribuir muito para a classificação da autenticidade de uma notícia.

4.1.3 Verificação de adulteração visual

Outro método descrito nos artigos sobre detecção automática de *fake news* na internet que segue uma linha de pesquisa baseada em conteúdo é a verificação de adulteração visual. A ideia deste método pode ser adequadamente resumida com o ditado 'uma imagem vale mais do que mil palavras' e sua estratégia é avaliar a veracidade de uma notícia através da examinação das imagens anexadas em seu corpo, reconhecendo se elas são falsas, foram manipuladas ou pertencem a uma conjuntura diferente da relatada pela notícia (HUH et al., 2018; ELKASRAWI et al., 2016). Todavia, o avanço recente das ferramentas de edição e manipulação de fotos torna a aplicação deste método desafiadora.

O anexo de imagens em uma notícia, além de fornecer mais crédito, a torna mais atraente para seus leitores, por esses motivos os autores de *fake news* as usam para reforçar seus pensamentos e relatos mentirosos (SHU et al., 2017). Para identificar a adulteração de uma imagem as estratégias existentes se reduzem a sua comparação de paridade ou conjuntura com outras imagens semelhantes encontradas na internet e a busca por vestígios que indiquem o uso de alguma ferramenta de edição. Em ambas as estratégias, geralmente, são utilizadas diversos algoritmos de processamento de imagem.

Encontrar a versão original de uma imagem anexada em uma notícia ou versões muito semelhantes postadas por outras fontes pode ser decisivo para avaliar tanto sobre a autenticidade da imagem quanto da notícia. A suposição de Pasquini et al. (2015) e de Elkasrawi et al. (2016) é que a versão original de uma imagem manipulada quase sempre pode ser encontrada em algum lugar da rede mundial de computadores. Uma forma para descobrir outras versões é recorrer a pesquisa reversa, uma técnica de recuperação que retorna resultados com base no conteúdo da imagem de consulta. O *Google Image Search*, por exemplo, pode ser citado como um exemplo de serviço e tecnologia que emprega essa técnica e pode vir a auxiliar nesta atividade (ELKASRAWI et al., 2016).

Uma vez descoberta a versão original ou similares da imagem anexada em uma notícia torna-se possível as comparar e identificar possíveis adulterações. Neste momento, técnicas de processamento de imagem, como a interpolação *Color Filter Array* e a detecção de borda, podem ser usadas (HUH et al., 2018). Além disso, se alguma versão estiver vinculada a notícia originária de outra fonte também existe a possibilidade de analisar informações como a data de criação ou publicação. Segundo Elkasrawi et al. (2016), excluindo eventos específicos recorrentes, tal como shows e conferências, se a data de publicação da notícia originária de outra fonte for muito antiga as chances da imagem analisada não ser verdadeira são maiores e, logo, a notícia em evidência provavelmente seria uma *fake news*.

Caso nenhuma outra versão da imagem anexada em uma notícia esteja disponível, a estratégia restante para verificar adulterações é investigar o conjunto de informações fornecidas pela própria imagem. Os metadados de uma imagem podem indicar uma série de diferenças entre sua pipeline e a de outra qualquer, isso porque eles se modificam de acordo com o equipamento de registro, a distância focal, as configurações de qualidade e demais características. Uma das maneiras de verificar adulterações a partir dos metadados é verificar se as diferentes partes da mesma imagem poderiam ter sido produzidas por um único pipeline, localizando regiões de emenda (HUH et al., 2018). Ademais, as análises forenses também são opções relevantes.

4.2 Baseado no contexto

A outra linha de pesquisa para detecção automática de *fake news* na internet é a baseada em contexto, um método que segue esta linha de pesquisa é aquele que pretende classificar uma notícia como falsa apenas com elementos adicionais ao seu conteúdo. Como exemplo de elementos adicionais de uma notícia podemos citar o engajamento de seus consumidores e o seu caminho de propagação na rede mundial de computadores, as informações contidas nestes elementos podem ser fundamentais na atribuição do valor de verdade, classificando-a como falsa ou verdadeira (SHU et al., 2017). A linha de pesquisa baseada em contexto contempla dois métodos, verificação de propagação e verificação de credibilidade.

4.2.1 Verificação de propagação

Um dos métodos descritos nos artigos sobre detecção automática de *fake news* na internet que segue uma linha de pesquisa baseada em contexto é a verificação de propagação, também denominada verificação de rede por alguns autores. Este método se baseia nas informações referentes à disseminação de uma notícia no ambiente online, formato de propagação e como ela é difundida por seus consumidores, para avaliar a sua autenticidade (SHU et al., 2017; JANG et al., 2018). Para alcançar o resultado desejado este método se resume principalmente ao estudo e investigação de *fake news cascade*.

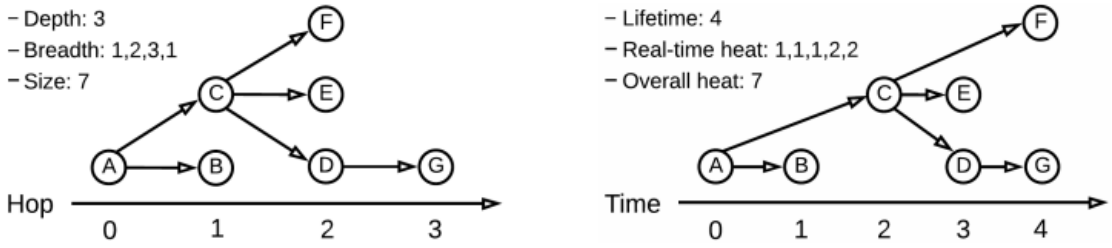
Primeiramente, é importante esclarecer o significado de *fake news cascade*. Uma definição estabelecida por Zhou e Zafarani (2018), pautada em estudos relacionados, diz que *fake news cascade* “é uma árvore ou estrutura semelhante a uma árvore que representa a propagação de um determinada *fake news* em uma rede social”. Ainda segundo Zhou e Zafarani (2018), uma *fake news cascade* pode ser representada de duas maneiras: com base em etapas e com base no tempo. Enquanto a primeira retrata os compartilhamentos que uma *fake news* obteve por meio de etapas, a segunda retrata os compartilhamentos em concordância com seus respectivos horários.

Parece oportuno exemplificar as duas representações de uma *fake news cascade* para melhor visualização e esclarecimento, portanto, na figura 3 é possível visualizar os dois tipos de representações. O nó de origem, ou nó raiz, da árvore simboliza o autor da *fake news*, o primeiro usuário a publicá-la. Os demais nós simbolizam outros usuários que, em sequência,

replicaram a *fake news* depois de seus nós pais, sendo que a relação entre eles são indicadas por arestas.

Algumas observações relevantes devem ser salientadas após o exemplo, a primeira é que cada nó representando um usuário é composto por uma série de atributos e dados deste usuário, como por exemplo, detalhes do seu perfil, postagens antigas, atividades recentes e sua tendência a compartilhar *fake news*. A segunda observação que deve ser salientada é que uma *fake news* pode gerar várias *fake news cascades* simultâneas, caso seja criada por vários usuários. Assim, a propagação de *fake news* pode ser analisada de diferentes maneiras.

Figura 3 – Representações de uma *fake news cascade*.



Fonte: (ZHOU; ZAFARANI, 2018).

As *fake news cascades* permitem analisar a propagação de duas formas, qualitativa e quantitativa. Uma análise qualitativa tem como objetivo entender o comportamento do objeto de estudo e assim descobrir tendências, padrões. Esse tipo de análise não apresenta resultados em número exatos, diferentemente da análise quantitativa. Nesta segunda forma de analisar, o objetivo é mensurar um problema por meio da geração de dados numéricos ou dados que podem ser transformados em estatísticas utilizáveis (ZHOU; ZAFARANI, 2018).

No que diz respeito a análise qualitativa, conseguimos categorizar tanto padrões que descrevem somente a propagação de *fake news* quanto padrões derivados da comparação entre a propagação de *fake news* e a de notícias verdadeiras (JANG et al., 2018). O estudo desses padrões trazem benefícios específicos entre os quais podemos citar: entender como as propagações variam em diferentes domínios, idiomas e sites, identificar conteúdo digno de verificação e acelerar a detecção e bloqueio de *fake news*.

Abaixo são listados alguns padrões de propagação de *fake news* descobertos por estudos recentes por meio da análise quantitativa em *fake news cascades*:

- *Fake news* são disseminadas mais rapidamente, mais amplamente e mais longe do que notícias verdadeiras, isto porque a largura máxima, o tamanho e a profundidade de uma *fake news cascade* são maiores se comparados com cascatas de notícias verdadeira.
- *Fake news* sobre política também são disseminadas mais rapidamente, mais amplamente e mais longe do que *fake news* sobre outros assuntos.

Por outro lado, na análise quantitativa é feita uma inserção de modelos matemáticos sobre a propagação de *fake news* buscando caracterizá-la. Determinar um modelo preciso

pode ser fundamental para descrever, quantificar e prever de forma realista as *fake news*. Uma das abordagens mais utilizadas para modelar a propagação de *fake news*, levando em consideração a série temporal das *fake news cascades*, é a análise de regressão. Outros modelos candidatos a mensurar a propagação de *fake news* são os modelos clássicos de pandemia e de economia pelo fato de compartilharem muitas semelhanças, dentre eles o formato de disseminação e a tomada de decisão de seus consumidores.

Após discorrer sobre a definição de *fake news cascades*, e sobre possíveis análises a serem feitas sobre ela, é possível apresentar as estratégias aplicadas por este método nas ferramentas de detecção automática. Através das *fake news cascades* é possível detectar *fake news* comparando a sua cascata com a cascatas de outras *fake news* verdadeiras ou representar corretamente sua cascata usando um representação informativa que facilite sua distinção (JANG et al., 2018).

Quando a estratégia utilizada é a comparação de *fake news cascades* então, para calcular a semelhança existentes entre elas, são empregues os gráficos de *Kernel*, que podem ser entendidos como funções que medem a semelhança de pares de gráficos. Com as semelhanças encontradas nas *fake news cascades*, as mesmas são utilizadas como recursos em uma estrutura de aprendizado supervisionado para detectar *fake news* a partir da estratégia de representação informativa. Como exemplo citamos o trabalho de Wu, Yang e Zhu (2015).

4.2.2 Verificação de credibilidade

Outro método descritos nos artigos sobre detecção automática de *fake news* na internet que segue uma linha de pesquisa baseada em contexto é a verificação de credibilidade, nele a veracidade da notícia é avaliada mediante suas informações sociais. Intuitivamente, uma notícia publicada em um site não confiável sendo compartilhada por usuários não confiáveis possui mais chance de ser *fake news* do que uma notícia em cenário oposto. Partindo deste princípio somos capazes de dizer que as informações sociais podem ter um papel de destaque para detectar *fake news*. As informações sociais avaliadas quanto à verificação de credibilidade nesta seção são: as manchetes, as fontes, os comentários e os divulgadores de uma notícia.

De acordo com Zhou e Zafarani (2018) a avaliação de credibilidade de uma manchete de notícia é reduzida à descoberta de *clickbaits*, ou seja, conteúdo na internet destinado à geração de receita de publicidade através do acesso. O principal objetivo de um *clickbait* é atrair a atenção de internautas para clicar em um *link* direcionado a uma determinada página da *web*, para isso suas manchetes são curtas o suficiente para não satisfazer a curiosidade do leitor (CHEN; CONROY; RUBIN, 2015).

As táticas para atrair cliques não cessam apenas no tamanho da manchete, em geral elas também são sensacionalistas, chamativas e contém informações de precisão e qualidade baixa. Segundo Wu, Yang e Zhu (2015), *clickbaits* são frequentemente combinados com *fake news*, isto porque ao fazer essa associação em suas manchetes eles, além de obterem a confiança dos internautas, alcançam altas taxas de cliques. Portanto, embora alguns *clickbaits* sejam "bons"para publicidade ou marketing de produtos, poucos devem ser permitidos em

artigos de notícias.

Diantes deste cenário, a credibilidade de uma manchete pode revelar muito sobre a credibilidade da sua notícia. Atualmente a avaliação de credibilidade de uma manchete é baseada na avaliação de seus recursos não linguísticos, como url de *links*, e linguísticos, como palavras, dentro de estruturas de aprendizado supervisionado e de aprendizado profundo. Um processo para atribuição de valor muito parecido com a do método de verificação do estilo de escrita.

Em adendo, também temos a avaliação de credibilidade das fontes. Um estudo de Silverman (2016) evidenciou que grande parte das *fake news* são originárias de sites hiper partidários que se apresentam como 'donos da verdade' ou de sites de notícias que publicam apenas boatos. Assim sendo, fatores que indicam o viés político, a qualidade e confiança do site de origem de uma notícia podem ser determinantes para avaliar sua credibilidade e a categorizá-la como *fake news*.

A análise de credibilidade na internet é um assunto de pesquisa bastante ativo, muitos algoritmos classificadores foram desenvolvidos nas últimas décadas, entre eles os tradicionais *PageRank* e *HITS* (ZHOU; ZAFARANI, 2018). Esses algoritmos tradicionais analisam a credibilidade de sites espalhados por toda rede de computadores objetivando melhorar a resposta dos mecanismos de busca, como o *Google*, às pesquisas de seus usuários. Contudo, alguns pontos fracos destes algoritmos tradicionais permitem a manipulação desta análise, dando visibilidade a páginas de pouca credibilidade, as chamadas páginas spam, e contribuindo para o aumento da disseminação de *fake news*.

Apesar das falhas apresentadas pelos principais algoritmos de classificação de sites, este parece ser um caminho promissor no combate às *fake news* visto que a credibilidade da fonte de uma notícia também pode apontar sua veracidade. Portanto, a avaliação de credibilidade da fonte de uma notícia atualmente é pautada na detecção precisa de spam. Os principais algoritmos utilizados nesta detecção são aqueles baseados no conteúdo, em *links*, no fluxo de cliques e no comportamento do usuário.

A terceira avaliação que pode ser feita para verificar a credibilidade de uma notícia é a de seus comentários ou respostas. Os comentários publicados por consumidores de uma notícia indicam as posturas e as opiniões deles sobre aquela notícia e podem trazer informações valiosas para detectar *fake news*. Contudo, segundo Zhou e Zafarani (2018), a maioria das ferramentas de detecção automática de *fake news* simplesmente ignoram os comentários ou, aquelas que os levam em consideração, prestam pouca atenção à credibilidade dos mesmos.

O estudo da credibilidade de comentários é importante para garantir a confiabilidade das informações obtidas, por exemplo, uma notícia sobre um fato político pode ser ao mesmo tempo atacada maliciosamente e receber muitos elogios das pessoas pelo simples fato de seu conteúdo ser contrário ou conivente, respectivamente, com a posição política dessas pessoas. Neste cenário, métodos que avaliem a credibilidade de cada comentário, identificando os não confiáveis e spams, possuem sua parcela de ajuda na detecção de *fake news* (SHU et al., 2017).

Os modelos de avaliação de credibilidade de comentários encontrados nos artigos consultados neste trabalho podem ser divididos em três categorias. A primeira é a baseada em conteúdo, nela vários recursos de linguagem são extraídos do comentário e uma estratégia semelhante a verificação de estilo de escrita é aplicada. A segunda é baseada em comportamento, esta categoria 'geralmente utiliza recursos indicativos de comentários não confiáveis extraídos dos metadados associados ao comportamento do usuário'. A terceira, e última categoria, é a baseada em gráficos, ela adota técnicas de decomposição de matriz e algoritmos probabilísticos para encontrar possíveis comentários spam.

A última forma de avaliação analisada nesta seção é a avaliação da credibilidade dos usuários que compartilham uma notícia. Os usuários são os grandes responsáveis pela disseminação das *fake news*, eles podem interagir e espalhar *fake news* de diversas formas, seja compartilhando, encaminhando ou apenas curtindo. Uma notícia compartilhada em sua maioria por usuários com pouca credibilidade também possui pouca credibilidade, o que pode indicar que a notícia seja uma *fake news*.

A definição da credibilidade de um usuário é dada a partir da examinação de seu histórico de atividades, incluindo seus *posts* e interconexões. Segundo Zhou e Zafarani (2018), todo usuário pertence a um destes dois grupos: usuários mal-intencionados com baixa credibilidade e usuários normais com alta credibilidade, entretanto Zhou e Zafarani (2018) adverte para a vulnerabilidade do usuário uma vez que "(...) o limite entre aquele que seja mal-intencionado e normal torna-se incerto". Para ele, um melhor agrupamento de usuários seria de participantes e não participantes, sendo que o primeiro seria subdividido em usuário mal-intencionado e usuário ingênuo uma vez que usuários normais podem com frequência, mesmo sem a intenção, participar de atividades relacionadas com *fake news*.

Usuários maliciosos espalham *fake news* de forma intencional, seja para ganhar alguma vantagem financeira ou outro tipo de benefício, como poder ou popularidade. Eles podem ser de fato humanos ou não, alguns se tratam de programas de *software* que executam suas atividades online por meio de tarefas automatizadas denominados *bots* ou *cyborgs*.

5 Desenvolvimento

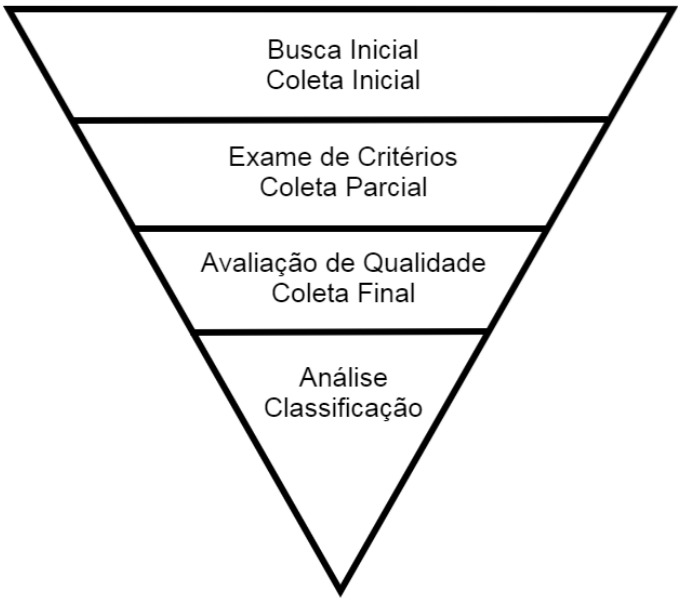
“Há apenas um bem, o saber;
e apenas um mal, a ignorância”.
Sócrates

Neste capítulo é apresentado a maneira como a revisão sistemática da literatura foi realizada, o processo de pesquisa e seleção dos artigos relevantes e dos estudos primários, os dados provenientes destes artigos e a classificação das ferramentas para detecção automática de *fake news* procedentes. A primeira seção descreve o desenvolvimento de todas as etapas do procedimento metodológico, partindo da definição da estratégia de pesquisa nas bases de consulta até a escolha dos artigos que constituem nosso estudo secundário. Em seguida, dados extraídos e padrões observados durante a pesquisa são expostos, como, por exemplo, o aumento recente da produção científica direcionada para a detecção automática de *fake news*. Por último, classificamos as ferramentas de detecção descritas nos estudos primários com base em quatro critérios: linha de pesquisa, método, campo de conhecimento e acurácia.

5.1 Aplicação dos métodos

Para uma melhor compreensão relatamos o desenvolvimento da revisão sistemática em quatro etapas, elas funcionam como filtros e podem ser representadas pela imagem de um funil, conforme a figura 4. A quantidade de documentos e/ou artigos restantes diminuem a cada etapa.

Figura 4 – Etapas da revisão sistemática da literatura.

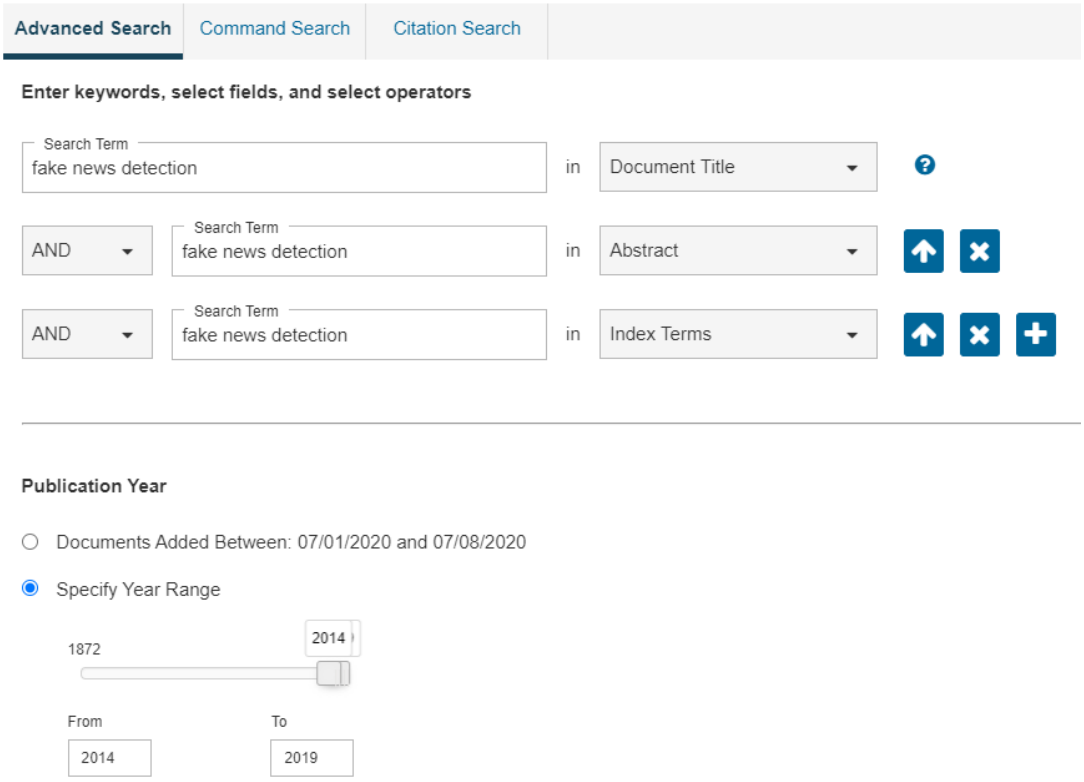


Fonte: elaborada pelo autor.

Uma vez definidos os procedimentos metodológicos, começamos a efetuar todas as etapas de maneira sequencial. A primeira etapa foi composta pela busca e coleta inicial de documentos nas bases de dados científicas predefinidas. Respeitamos a seguinte ordem: *IEEE Xplore*, *ACM Digital Library*, *Science Direct* e, por último, o *Google Scholar*. Optamos por essa ordem para evitar a coleta duplicada, como o *Google Scholar* realiza em sua plataforma a indexação de documentos das mais variadas fontes provavelmente, utilizando uma ordem de busca diferente, seria necessário um retrabalho para garantir a unicidade dos documentos coletados.

Todas as bases de dados científicas dispunham de uma funcionalidade, denominada pesquisa avançada, que possibilitava encontrar uma determinada palavra, ou conjuntos de palavras, em alguma parte específica dos documentos disponíveis. Para uma melhor visualização, na figura 5 é apresentada a interface e parte dos recursos presentes na pesquisa avançada da *IEEE Xplore*. Por meio desta funcionalidade foram efetuadas as buscas de documentos considerando as *strings* referentes às tabelas 2 e 3, encontradas na seção 3.1. Limitamos os resultados a mostrarem somente documentos nos quais o título, o resumo ou as palavras-chaves eram constituídos pelas *strings* procuradas, a única exceção foi o *Google Scholar* que possibilitava encontrar conjuntos de palavras somente no título dos documentos.

Figura 5 – Exemplo da funcionalidade de pesquisa avançada: *IEEE Xplore*.



Fonte: (IEEE Xplore, 2020).

A pesquisa avançada também possuía outro recurso interessante, a possibilidade de limitar os resultados de acordo com um intervalo de tempo, geralmente anual. Durante a busca inicial nas bases de dados científicas aproveitamos para validar a data de publicação dos do-

cumentos, um dos critérios de inclusão que seria examinado na próxima etapa. O intervalo entre 2014 e 2019 foi aplicado nos filtros inteligentes disponíveis em cada base de dados, diminuindo ainda mais a quantidade de resultados e permitindo melhor examinação dos mesmos pelo autor deste trabalho.

Em nossa busca inicial, levando em conta todos os resultados provenientes das bases de dados científicas, centenas de documentos foram considerados. Para acelerar a coleta inicial um critério de exclusão também foi validado, somente os documentos em que o título tinha alguma aderência ao tema deste trabalho foram coletados. A aplicação deste critério, neste momento, foi crucial para a velocidade de desenvolvimento do trabalho, reduzindo muito o número de documentos que deveriam ser coletados. Quando o título era concernente com o trabalho a versão original do documento era armazenada em um serviço de nuvem, diminuindo a possibilidade de perda do mesmo.

Finalizamos a primeira etapa com a consumação da coleta inicial e avançamos para a segunda, composta pela examinação dos critérios e coleta parcial. Todos os documentos armazenados foram fiscalizados conforme demais critérios de inclusão e exclusão descritos na subseção 3.2. Durante a examinação das características externas encontramos muita literatura cinzenta, documentos que não estavam disponíveis em sua totalidade, que eram desprovidos de resumo, introdução, conclusão ou não tinham formato permitido. Também encontramos, em um número bem menor, documentos escritos em outras línguas além do inglês e português. Documentos que não preenchiam a lista de critérios foram descartados da pesquisa, deste modo, após o exame externo restaram somente artigos científicos.

A examinação das partes internas foi decisiva para determinar os artigos relevantes deste trabalho, através da leitura do resumo e da introdução de mais de uma centena de artigos remanescentes foi possível verificar quais poderiam acrescentar em algo neste trabalho e contribuir com informações sobre os métodos para detecção automática de *fake news*. Sempre que um artigo tinha sua leitura concluída ou ele era descartado pela falta de aderência, ou tinha sua referência bibliográfica e demais dados registrados por ser considerado relevante. Todos os registros foram feitos pela ferramenta *Mendeley*, um software gratuito que auxilia nos trabalhos acadêmicos e tem a finalidade de gerenciar arquivos eletrônicos, além de ajudar na normalização de citações e referências geradas automaticamente. Ao final da segunda etapa, um total de 92 artigos aderentes ao tema deste trabalho foram levantados e constituíam a nossa coleta parcial.

Em seguida, partimos para a etapa mais longa deste trabalho, a leitura completa de todos os artigos considerados relevantes e a avaliação de qualidade dos mesmos. Nesta terceira etapa optou-se por uma leitura dinâmica com o objetivo de aumentar a velocidade da leitura mantendo o entendimento e a retenção de informações. Primeiramente, seguindo as diretrizes apresentadas na seção 3.3 e para termos uma base de parâmetro, encontramos um artigo considerado nota 5 e outro considerado nota 1. A partir deles demos as notas aos demais artigos. O artigo era lido e logo depois avaliado, de forma a garantir a coerência das avaliações.

Após avaliação dos artigos percebemos que tínhamos um bom número de artigos com

nota igual ou superior a 3 e decidimos que, por questão de tempo e para uma melhor qualidade da análise que seria realizada, eles deveriam constituir a nossa coleta final e fazer parte de nosso estudo secundário.

Chegamos na quarta, e última, parte deste trabalho com mais de 70 artigos. Para facilitar a análise destes artigos, preferimos estabelecer algumas perguntas que deveriam ter suas respostas retiradas dos artigos conforme realização da leitura analítica. Segundo Adler e Doren (2010), leitura analítica consiste em uma leitura compassada, lenta se necessário, que têm por objetivo a absorção total do conteúdo, ou seja, das proposições e argumentos. Abaixo são listadas as perguntas estabelecidas que posteriormente nortearam a nossa análise.

- Em que se baseava o método descrito?
- Qual subárea de conhecimento da ciência foi mais explorada?
- Quais algoritmos foram usados?
- Para que foram usados estes algoritmos?
- Qual a taxa de sucesso de detecção?
- Quais são os pontos fortes deste método?
- Quais são os pontos fracos deste método?

Durante desenvolvimento do trabalho dados referentes a cada etapa foram captados e serão mostrados na próxima subseção. É importante ressaltar que os artigos relevantes com notas igual ou inferior a 3 também foram considerados na apresentação dos dados visto que eles poderiam agregar para o entendimento do contexto geral e porque, de fato, representam a produção acadêmica sobre a detecção automática de *fake news*. Mesmo sendo descartados da coleta final, as referências destes artigos podem ser visualizadas no apêndice A. Para a tabulação dos dados foi usado o *Google Docs*, especialmente a parte de planilhas, que permite melhor disposição e geração de diferentes estilos de gráficos.

5.2 Apresentação dos dados

Durante todo o procedimento metodológico diversos dados foram coletados ao longo das etapas do processo. O registro e o entendimento destes dados poderiam revelar alguns conhecimentos como, por exemplo, tendências e padrões de curto ou longo prazo. Além disso, o levantamento de dados é importante em uma revisão sistemática pois enriquece e ajuda a fornecer estimativas mais precisas ao leitor. A seguir, todos os dados coletados entre a busca inicial de documentos até a coleta final dos artigos são apresentados.

A quantidade de resultados advindos da busca inicial com as *strings* definidas na subseção 3.1 e delimitados pela data de publicação foi diversificado entre as bases de dados científicas. A *Science Direct* apresentou o menor número de resultados, sendo um pouco menos de duas centenas de documentos. Logo depois veio a *ACM Digital Library*, com um pouco

mais de duas centenas de documentos. O *Google Scholar* e o *IEEE Xplore* foram além, com a quantidade variando entre a casa de quatro centenas de documentos. Sendo assim, o número total de documentos examinados de alguma forma neste trabalho foi de 1339 e são explicitados abaixo, na tabela 6.

Tabela 6 – Quantidade de resultados provenientes da busca inicial.

String	***ACM Digital Library***	***Google Scholar***	***IEEE Xplore***	***ScienceDirect***
1	6	22	20	4
2	9	2	33	6
3	14	8	30	4
4	11	11	33	6
5	30	36	82	12
6	33	279	107	16
7	32	6	69	15
8	48	44	49	11
9	32	7	43	101
10	15	5	26	19
Total	**230**	**423**	**492**	**194**

Fonte: elaborada pelo autor.

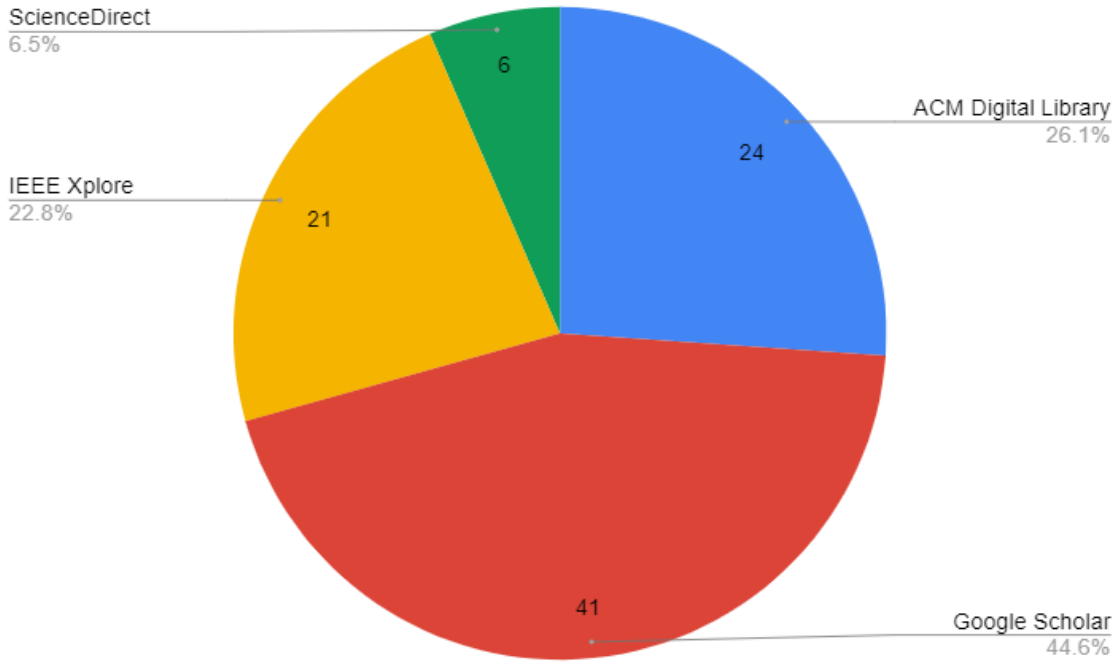
Os 1339 documentos resultantes da busca inicial tiveram como primeira examinação a aderência de seu título com o tema e objetivo deste trabalho, mesmo antes da coleta inicial. O percentual de artigos em que o título era divergente foi alto, aproximadamente 87,38%. Deste modo, em nossa coleta inicial foram reunidos 164 documentos compatíveis, correspondendo a aproximadamente 12,62% do total de documentos resultantes da busca inicial. Apesar do grande número de documentos descartados, a quantidade coletada foi considerável, indicando inicialmente que poderia existir muitos estudos primários de nosso interesse.

Para encontrar os artigos relevantes, digo, artigos que poderiam contribuir com este trabalho, os 164 documentos que compunham a coleta inicial passaram pela examinação de diversos critérios listados e esclarecidos na subseção 3.2. Neste ponto também tivemos um grande descarte em porcentagem de documentos, 44% dos artigos não contemplavam os critérios de inclusão ou apresentavam algum critério de exclusão. Portanto, o número de artigos na coleta parcial foi de 92, equivalente a 56% do total de documentos resultantes da coleta inicial. Estes são os artigos que poderiam ser relevantes de alguma maneira com este trabalho.

Os artigos relevantes estão distribuídos entre as bases de consulta de acordo com a figura 6. Alguns aspectos interessantes puderam ser observados, primeiramente, mesmo sendo a última base de dados científica a ser consultada o *Google Scholar* forneceu o maior número de artigos relevantes, muito devido ao grande porte da plataforma, por unir diversas bases de dados científicas e possibilitar o encontro de artigos em diferentes linguagens. Segundamente, a base da *ACM Digital Library*, mesmo apresentando metade da quantidade de resultados na

busca inicial, contribuiu com mais artigos relevantes do que da *IEEE Xplore*. Por fim, a *ScienceDirect* pouco contribuiu para o trabalho quando comparada com as demais bases de dados científicas, fruto do pouco acervo de artigos acerca da detecção automática de *fake news*.

Figura 6 – Número de artigos relevantes por base de consulta.



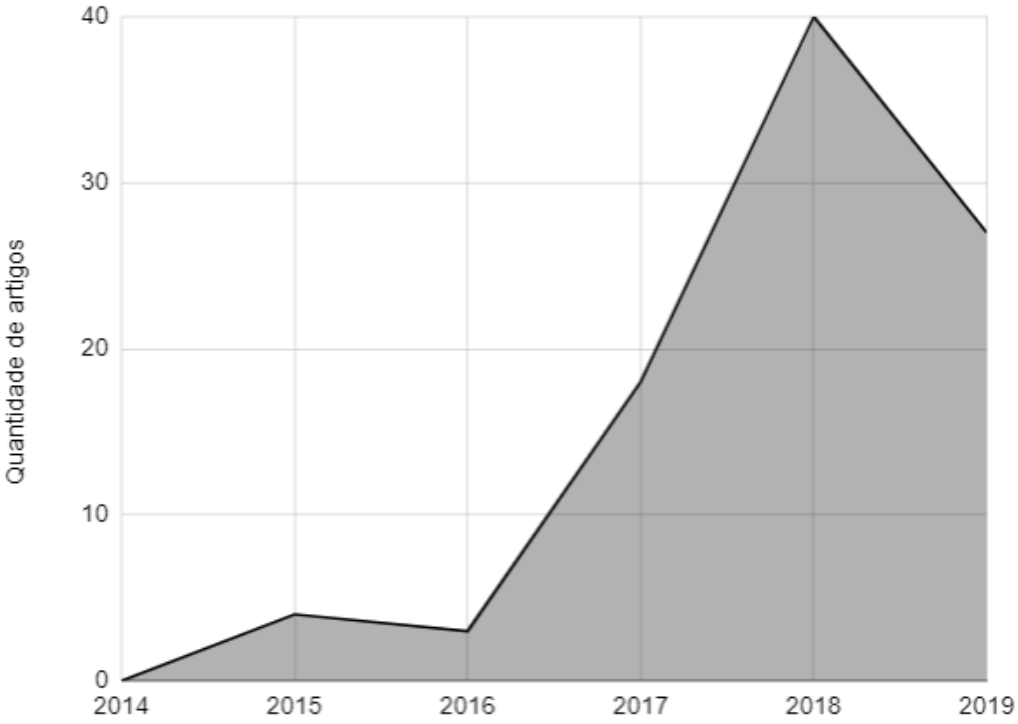
Fonte: elaborada pelo autor.

Quando observamos a data de publicação dos artigos relevantes, conforme figura 7, encontramos um resultado que já era esperado, o número de artigos relevantes sobre detecção automática de *fake news* apresenta um aumento considerável depois de 2016, ano que muitos autores relatam a influência das *fake news* nas eleições presidenciais de um dos maiores países do mundo, os Estados Unidos. Dos 92 artigos relevantes, ao menos 85 deles foram publicados após 2016, representando cerca de 92,39% do total. Levando em consideração que a pesquisa levantou artigos publicados até `julho de 2019`, provavelmente este mesmo ano superaria 2018, indicando o aumento contínuo de interesse sobre o tema deste trabalho e importância para a sociedade como um todo.

Quando levamos em consideração o número de citações de cada artigo relevante, a maior parte dos artigos possuem menos de 100 citações. Este dado pode indicar que a detecção automática de *fake news* é um assunto atual e emergente, que foi abordado na literatura acadêmica recentemente e que ainda não existem artigos considerados base. Também pode ser reflexo das várias abordagens de detecção existentes, impedindo com que as citações se concentrem em uma gama de artigos de mesmo viés. Mais detalhes sobre o número de citações dos artigos relevantes podem ser visualizadas na figura 8.

A pequena quantidade de artigos escritos em português, juntamente com o pequeno número de citações concedidas, refletem que este assunto é ainda pouco abordado nessa língua, inclusive em nosso país. Dentre os 10 artigos escritos em português que constituíam

Figura 7 – Número de artigos relevantes por ano de publicação.



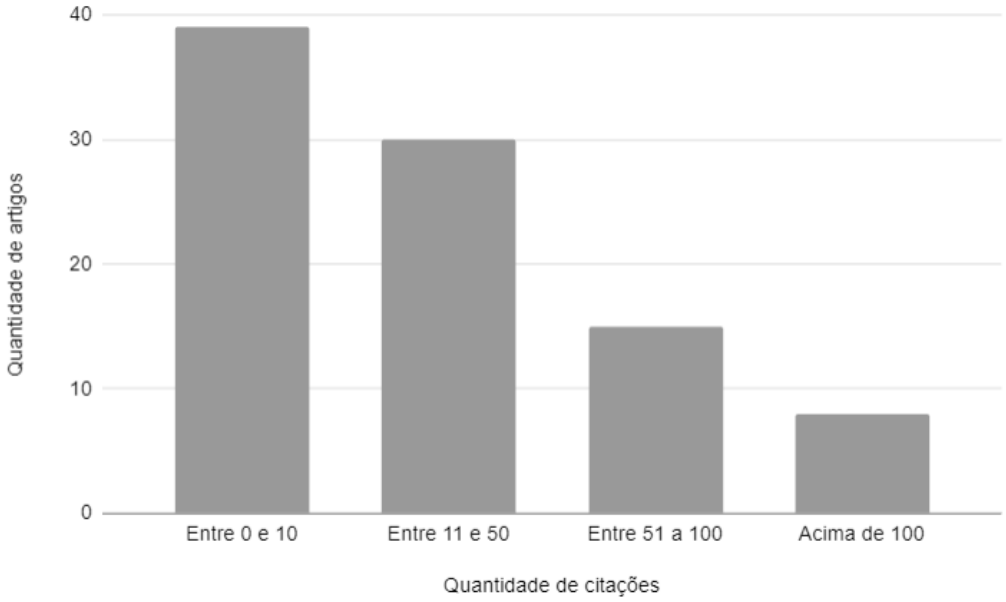
Fonte: elaborada pelo autor.

parte de nossa coleta parcial apenas um deles tinha alguma citação, sendo um total de 6 citações. Podemos perceber nesta etapa, mas também desde a busca inicial, que produção científica e acadêmica do Brasil sobre detecção automática de *fake news* era muito baixo, pelo menos em nossa língua.

Posteriormente ao encontro dos artigos relevantes sucedeu-se a última etapa antes da coleta final dos estudos primários e análise dos mesmos, a avaliação de qualidade. Todos os 92 artigos artigos relevantes foram avaliados com notas dentre intervalo de 1 e 5. Quanto menor a nota de um determinado artigo, menor seria a sua contribuição para com este trabalho, e vice-versa, quanto maior a sua nota, mais o artigo poderia contribuir com informações. Na tabela 7 é possível visualizar a quantidade de artigos relevantes por nota de acordo com a base de dados científica consultada.

Observando a tabela 7 pode-se notar que em geral os artigos relevantes obtiveram pontuações bem diversificadas. É importante ressaltar que os artigos que obtiveram notas inferiores a 3 não podem ser considerados de baixa qualidade, apenas não apresentaram um bom conteúdo para o propósito deste trabalho. Quanto aos demais artigos relevantes que alcançaram notas igual ou superior a 3 foram efetivamente utilizados nesta pesquisa. Concluindo assim todo o processo de seleção dos estudos primários, chegamos em nossa última etapa, a coleta final, com um total de 77 artigos.

Figura 8 – Número de artigos relevantes por quantidade de citações.



Fonte: elaborada pelo autor.

Tabela 7 – Quantidade de artigos relevantes por nota conforme base de dados científica.

Base de dados científica	**Nota 1**	**Nota 2**	**Nota 3**	**Nota 4**	**Nota 5**
ACM Digital Library	2	1	17	4	0
Google Scholar	6	1	20	4	10
IEEE Xplore	2	1	14	2	2
ScienceDirect	1	1	2	2	0
Total de Artigos	**11**	**4**	**53**	**12**	**12**

Fonte: elaborada pelo autor.

5.3 Resultados

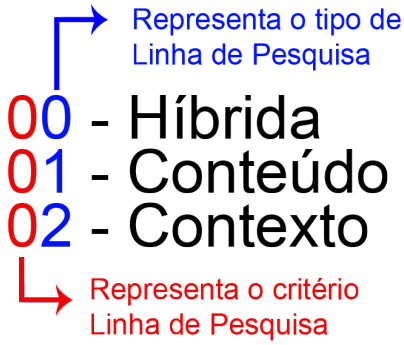
A maior parte dos estudos primários apresentavam com detalhes a elaboração e o desenvolvimento de alguma ferramenta para detecção automática de *fake news*. Sob a influência de Vedova et al. (2018) e Marra et al. (2018), que comparam a eficácia de suas ferramentas com outras existentes na literatura, e Hardalov, Koychev e Nakov (2016), que comparam a variação da eficácia da sua ferramenta em relação ao conjuntos de informações de entrada, propomos uma classificação das ferramentas de detecção encontradas na nossa pesquisa considerando quatro critérios: linha de pesquisa, método, campo de conhecimento e eficácia.

A linha de pesquisa se refere ao conjunto de informações de entrada do método aplicado a ferramenta. Como visto ao longo deste trabalho um método pode utilizar de diversas informações relacionadas a uma notícia para classificar a mesma como sendo ou não *fake news*. Essas informações estão divididas em dois grupos: de conteúdo e de contexto. O pri-

meiro grupo é vinculado ao corpo de uma notícia e o segundo grupo a credibilidade da fonte e a forma de propagação.

Em nossa classificação a linha de pesquisa é o primeiro critério, ele é representado por dois números sendo que o primeiro deles é obrigatoriamente o número zero e possui três ramificações: Híbrida (ferramenta na qual seu conjunto de informações de entrada é constituído por informações de conteúdo e contexto), Conteúdo (ferramenta na qual seu conjunto de informações de entrada é constituído somente por informações de conteúdo) e Contexto (ferramenta na qual seu conjunto de informações de entrada é constituído somente por informações de contexto). A figura 9 exemplifica o critério da linha de pesquisa de forma visual.

Figura 9 – Exemplo de classificação: linha de pesquisa.



Fonte: elaborada pelo autor.

O método se refere a estratégia aplicada pela ferramenta para determinar se uma notícia é considerada *fake news*. Conforme visto detalhadamente na seção 4, diferentes métodos podem ser aplicados em uma ferramenta com base no conjunto de informações de entrada utilizado pela mesma. Se o conjunto de entrada for constituído por informações de conteúdo então a autenticidade da notícia pode ser definida pela comparação de seus fatos com fontes consideradas confiáveis, pelo estilo de escrita e pelas possíveis adulterações feitas nas imagens anexadas ao seu corpo. Se o conjunto de entrada for constituído por informações de contexto, então a ferramenta pode definir a autenticidade de uma notícia pela maneira como ela se propaga na internet ou verificando a credibilidade de quem a compartilhou ou escreveu.

Em nossa classificação o método aplicado é o segundo critério, ele é representado por dois números, sendo que o primeiro deles é obrigatoriamente o número um e possui seis ramificações: Múltiplos (ferramenta que aplicou mais de um método para detecção), Fatos (ferramenta que aplicou a verificação de fatos para detecção), Estilo (ferramenta que aplicou a verificação do estilo de escrita para detecção), Adulteração visual (ferramenta que aplicou a verificação de adulteração visual para detecção), Propagação (ferramenta que aplicou a verificação de propagação para detecção) e Credibilidade (ferramenta que aplicou a verificação de credibilidade para detecção). A figura 10 exemplifica o critério do método de forma visual.

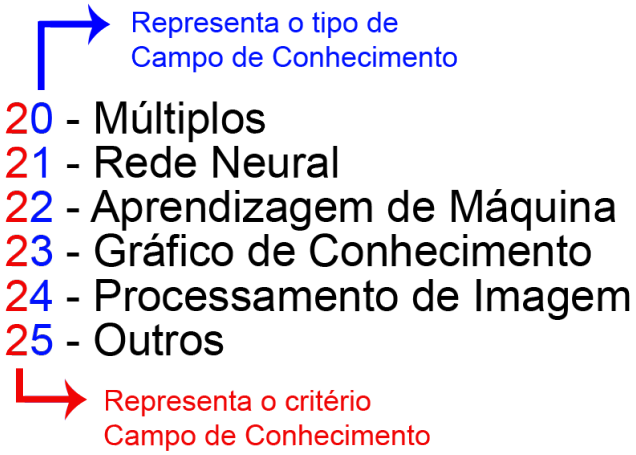
O campo de conhecimento é referente à subárea da ciência da computação na qual a ferramenta está fundamentada ou explora. Para definir as subáreas da ciência de computação pautamos Bigonha (1997), que em seu artigo propõe uma definição amplamente aceita pela

Figura 10 – Exemplo de classificação: método.

Fonte: elaborada pelo autor.

comunidade acadêmica. De modo geral, a maioria dos ferramentas encontradas são fundamentadas no Processamento de Linguagem Natural, uma subárea que estuda os problemas da geração e compreensão automática de línguas humanas naturais. Entretanto, como os algoritmos de Processamento de Linguagem Natural empregados nas ferramentas se baseiam nas Redes Neurais e na Aprendizagem de Máquina optamos por usá-las como ramificações deste critério.

Figura 11 – Exemplo de classificação: campo de conhecimento.



Fonte: elaborada pelo autor.

Em nossa classificação o campo de conhecimento é o terceiro critério, representado por dois números, sendo que o primeiro deles é obrigatoriamente o número dois e possui seis ramificações: Múltiplos (ferramenta que para ser entendida e reproduzida é necessário conhecer sobre mais de uma subárea da ciência), Rede Neural, Aprendizagem de Máquina, Gráfico de Conhecimento, Processamento de Imagem e Outros (ferramenta que para ser entendida e

reproduzida é necessário conhecer sobre subárea da ciência não mencionada anteriormente). A figura 11 exemplifica o critério do campo de conhecimento de forma visual.

A eficácia corresponde a faixa de acurácia relatada pelo ferramenta. Acurácia é a proximidade entre o valor obtido experimentalmente e o valor requerido, calculada dividindo o número de notícias detectadas corretamente pelo número total de notícias (MOHTARAMI et al., 2018). Esta é uma medida de avaliação comum usada para calcular o desempenho das ferramentas, quase todos os artigos que continham alguma ferramenta apresentavam a acurácia da mesma.

Em nossa classificação a eficácia é o quarto critério, ele é representado por dois números, sendo que o primeiro deles é obrigatoriamente o número três e possui quatro ramificações: Indisponível (ferramentas que não exibiram seus resultados ou exibiram resultados sem o cálculo da acurácia), até 60%, entre 60% e 80%, acima de 80% (ferramentas que apresentaram esta faixa de acurácia em seus resultados). A figura 12 exemplifica o critério da eficácia de forma visual.

Figura 12 – Exemplo de classificação: eficácia.

Fonte: elaborada pelo autor.

Segue abaixo a classificação de todos os artigos que eram compostos por alguma ferramenta para detecção automática de *fake news* na internet. Ao todo, 65 artigos foram classificados em conformidade com os quatro critérios detalhados anteriormente. Para melhor compreensão do leitor optamos por dividir a lista de artigos em três tabelas com base no primeiro critério de classificação, a linha de pesquisa. Além disso, optamos também por listar em ordem crescente quanto ao código classificador.

Tabela 8 – Classificação de artigos que seguem uma linha de pesquisa híbrida.

Artigos	Códigos
(HELMSTETTER; PAULHEIM, 2018)	00 10 20 30
(ZHOU et al., 2015)	00 10 20 32
(YANG et al., 2019)	00 10 21 30
(SHU; MAHUDESWARAN; LIU, 2019)	00 10 21 32
(SHU et al., 2019)	00 10 21 33
(QIAN et al., 2018)	00 10 21 33
(LONG, 2017)	00 10 21 31
(CASTELO et al., 2019)	00 10 22 32
(BUNTAIN; GOLBECK, 2017)	00 10 22 32
(GUPTA et al., 2018)	00 10 22 32
(RASOOL et al., 2019)	00 10 22 32
(JANZE; RISIUS, 2017)	00 10 22 33
(VEDOVA et al., 2018)	00 10 22 33
(SHU; WANG; LIU, 2019)	00 10 22 33
(REIS et al., 2019)	00 10 22 33
(KOTTETI et al., 2018)	00 10 22 33

Fonte: elaborada pelo autor.

Tabela 9 – Classificação de artigos que seguem uma linha de pesquisa em conteúdo.

Artigos	Códigos
(DEY et al., 2018)	01 10 20 32
(KARIMI; TANG, 2019)	01 10 20 33
(RUBIN et al., 2016)	01 10 20 33
(PÉREZ-ROSAS et al., 2017)	01 10 22 32
(HARDALOV; KOYCHEV; NAKOV, 2016)	01 10 22 33
(ATODIRESEI; TĂNĂSELEA; IFTENE, 2018)	01 10 25 30
(MARUMO, 2018)	01 11 20 32
(KALIYAR, 2018)	01 11 20 33
(SADIQ et al., 2019)	01 11 20 33

Continua na próxima página

Artigos	Códigos
(LEAL, 2018)	01 11 20 33
(CHOPRA; JAIN; SHOLAR, 2017)	01 11 20 33
(PFOHL; LEGROS, 2017)	01 11 21 30
(KHATTAR et al., 2019)	01 11 21 32
(RIEDEL et al., 2017)	01 11 21 33
(MOHTARAMI et al., 2018)	01 11 21 33
(BHATT et al., 2018)	01 11 21 33
(SU; MACDONALD; OUNIS, 2019)	01 11 21 33
(THOTA et al., 2018)	01 11 21 33
(BAHAD; SAXENA; KAMAL, 2019)	01 11 21 33
(AJAO; BHOWMIK; ZARGARI, 2018)	01 11 21 33
(DAVIS; PROCTOR, 2017)	01 11 21 33
(SEO; SEO; JEONG, 2018)	01 11 21 33
(KARIMI et al., 2018)	01 11 21 31
(TRAYLOR et al., 2019)	01 11 22 32
(GRANIK; MESYURA, 2017)	01 11 22 32
(PRATIWI; ASMARA; RAHUTOMO, 2017)	01 11 22 32
(Al Asaad; Erascu, 2018)	01 11 22 33
(FONTANA et al.,)	01 11 22 33
(MONTEIRO; NOGUEIRA; MOSER,)	01 11 22 33
(AHMED; TRAORE; SAAD, 2017)	01 11 22 33
(JAIN; KASBE, 2018)	01 11 22 33
(AL-ASH; WIBOWO, 2018)	01 11 22 33
(BOURGONJE; SCHNEIDER; REHM, 2017)	01 11 22 33
(SAKURAI,)	01 11 22 31
(PAN et al., 2018)	01 11 24 33
(AJAO; BHOWMIK; ZARGARI, 2019)	01 12 20 33
(WANG et al., 2018)	01 12 21 32
(GIACHANOU; ROSSO; CRESTANI, 2019)	01 12 21 32
(TARMIZI et al., 2019)	01 12 21 33
(YANG et al., 2018)	01 12 21 33
(FERNANDEZ; DEVARAJ, 2019)	01 12 22 33
(SAMONTE, 2018)	01 12 22 31
(HUH et al., 2018)	01 13 20 30
(ELKASRAWI et al., 2016)	01 13 24 32

Continua na próxima página

Artigos	**Códigos**

Fonte: elaborada pelo autor.

Tabela 10 – Classificação de artigos que seguem uma linha de pesquisa em contexto.

Artigos	**Códigos**
(RUCHANSKY; SEO; LIU, 2017)	02 10 21 33
(WU; LIU, 2018)	02 10 21 33
(MONTI et al., 2019)	02 14 20 33
(LIU; WU, 2018)	02 14 21 33
(TACCHINI et al., 2017)	02 15 22 33

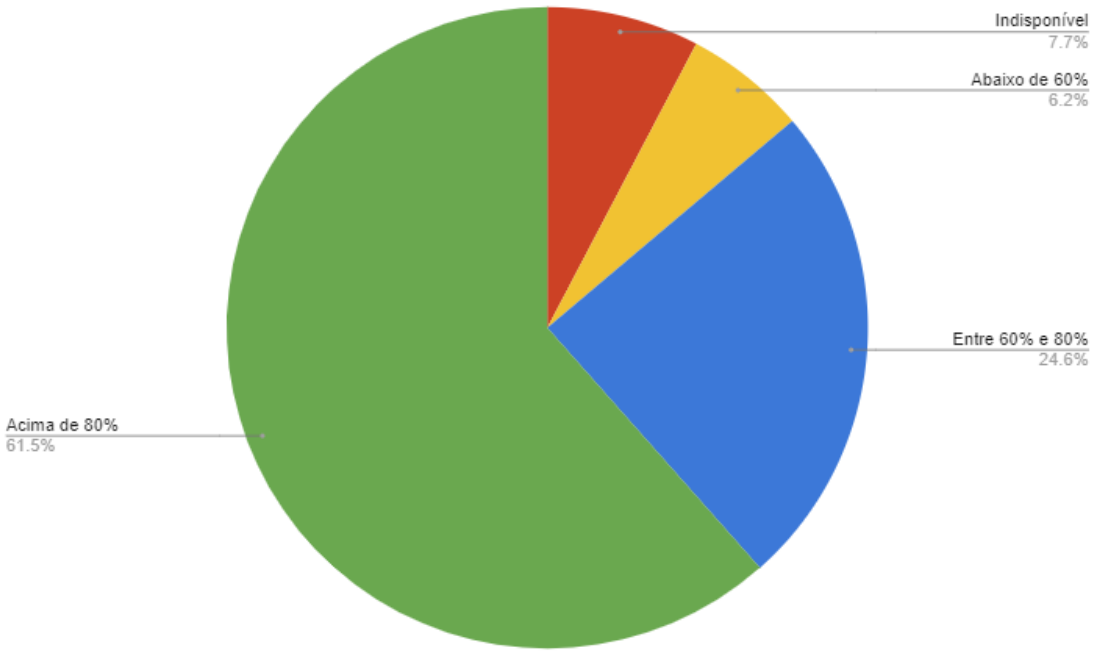
Fonte: elaborada pelo autor.

6 Análise da pesquisa

“No tempo certo, tudo dará certo”.
Desconhecido

Os resultados expostos no capítulo precedente podem estar repletos de informações e dados pertinentes sobre a detecção automática de *fake news*, principalmente em relação a sua eficácia. Estes resultados possibilitam com que análises sobre diferentes pontos de vista sejam realizadas e alguns indícios e relações sejam apontados. Optamos por dividir nossa análise em três níveis: cenário global, cenário local e cenário associativo. O nível de cenário global pretende abordar a eficácia das ferramentas de detecção de forma geral, analisando o conjunto de elementos superficialmente, ignorando associações específicas entre eficácia e demais critérios. O nível de cenário local, ao contrário do anterior, destaca as associações entre eficácia e demais critérios, porém a análise é reduzida a relações únicas e diretas. Por último, o nível de cenário associativo enfatiza relações específicas entre eficácia e no mínimo outros dois critérios. Destaca-se que ferramentas de detecção com desempenho indisponível não integraram a nossa análise.

Figura 13 – A eficácia dos artigos em um cenário global.



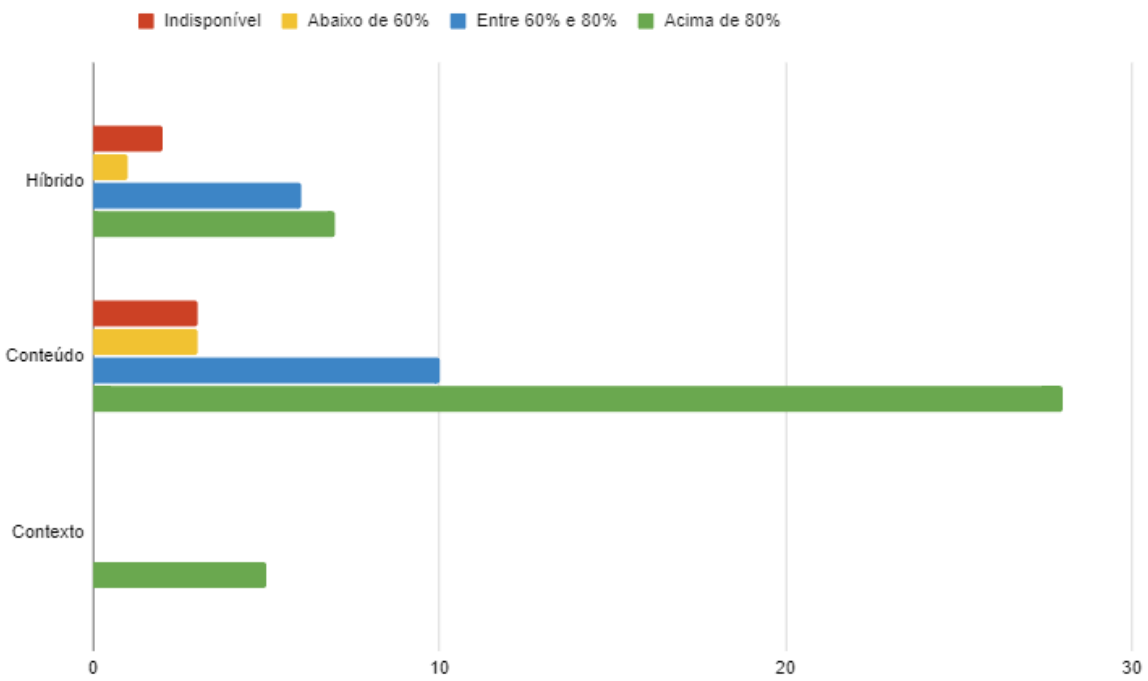
Fonte: elaborada pelo autor.

Em um cenário global, exibido pela figura 13, mais da metade das ferramentas de detecção levantadas na revisão sistemática apresentaram uma eficácia acima de 80%, isto significa que, a cada 10 notícias avaliadas ao menos 8 foram classificadas corretamente quanto a

sua autenticidade. A propósito, algumas ferramentas alcançaram índices acima de 90%. Estes bons números de eficiência não se concentram em uma única linha de pesquisa, único método ou único campo de conhecimento, reafirmando as diversas possibilidades em busca de uma possível solução para a detecção automática de *fake news*. A independência de desempenho pode nos deixar otimistas pois indica que seja viável executar esta tarefa independente do formato, estilo e aparência de uma *fake news*. Contudo, apesar do exposto, alguns pontos devem ser levados em consideração.

O principal ponto a ser considerado é que nenhuma das ferramentas de detecção atingiu a eficácia julgada ideal, que neste caso tende a ser a máxima possível. Outro ponto é referente a validação da eficácia apresentada por cada ferramenta, é importante salientar que o conjunto de notícias utilizado para cálculo da eficácia era condizente com o método aplicado, ainda assim ferramentas que aplicaram métodos iguais também utilizaram conjunto de notícias heterogêneas. Por exemplo, as ferramentas de Jain e Kasbe (2018) e Ahmed, Traore e Saad (2017) possuem a mesma classificação porém o desempenho das mesmas foram calculados mediante conjunto de notícias distintas. Por último, mesmo sendo intuitivo, é necessário relatar que algumas ferramentas de detecção são limitadas ao seu propósito. Enquanto ferramentas que aplicam o método de verificação de adulteração visual seriam ineficientes em notícias que não possuem imagens em anexo, ferramentas que seguem a linha de pesquisa de contexto tendem a ter mais eficácia em ambientes como as redes sociais.

Figura 14 – A eficácia dos artigos em um cenário local: linha de pesquisa.

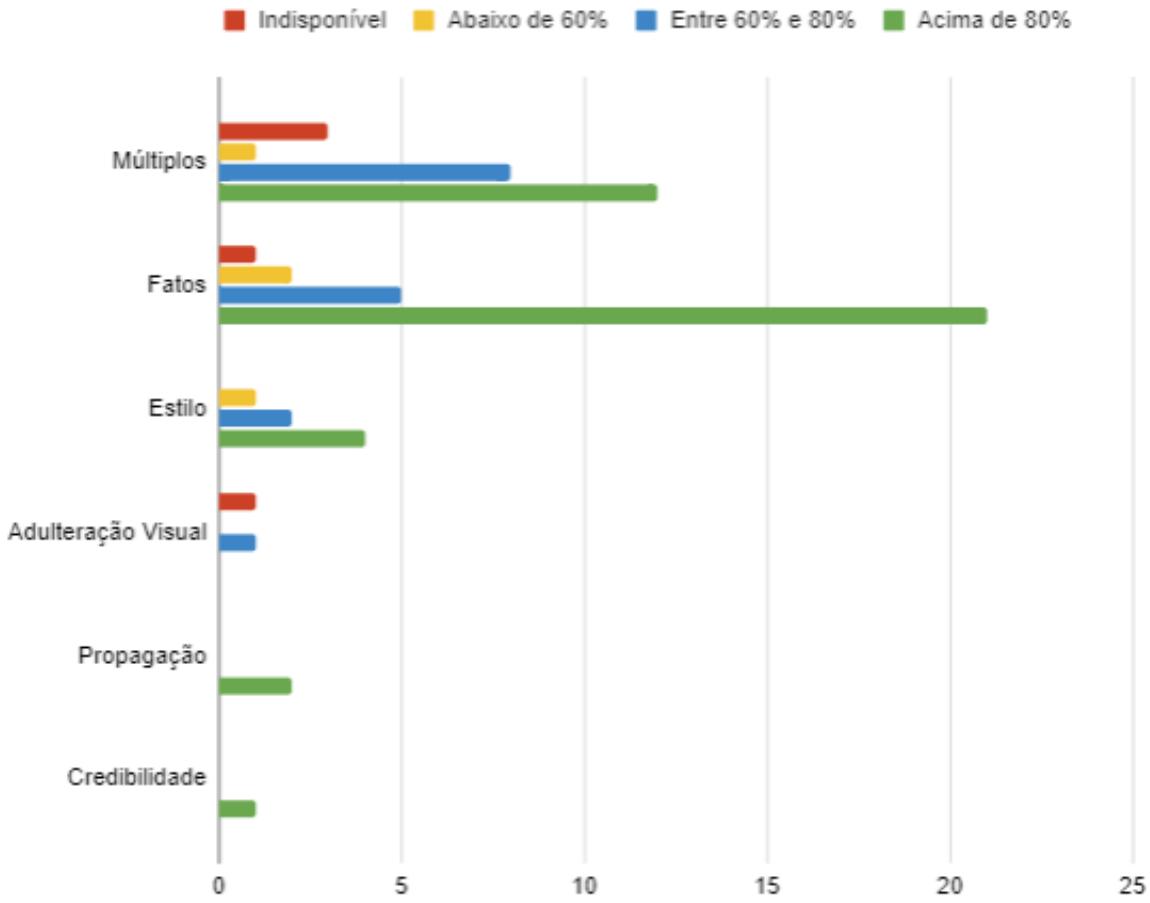


Fonte: elaborada pelo autor.

Em um cenário local, quando o critério da linha de pesquisa é observado, o que nos ressalta a atenção é que todas as ferramentas de detecção que manipularam informações de

contexto apresentaram eficácia acima de 80%, conforme figura 14. Todavia, este grupo de ferramentas foi o menor, com apenas 5 exemplares. Com certeza seria interessante o desenvolvimento de novas ferramentas que manipulam informações de contexto visto que, inicialmente, parece ser um caminho bastante promissor e ainda pouco explorado. Das ferramentas de detecção que manipularam informações de conteúdo mais da metade apresentaram eficácia acima de 80%, um dado mais assertivo devido a grande quantidade de exemplares deste grupo de ferramentas, o mais investigado por cientistas, pesquisadores e pela literatura. Quanto ao grupo de ferramentas de detecção que manipularam ambos os tipos de informação, em comparação com os grupos anteriores, foi o que apresentou menor eficácia, apontando que talvez seja complicado desenvolver ferramentas combinando conceitos e informações.

Figura 15 – A eficácia dos artigos em um cenário local: método.



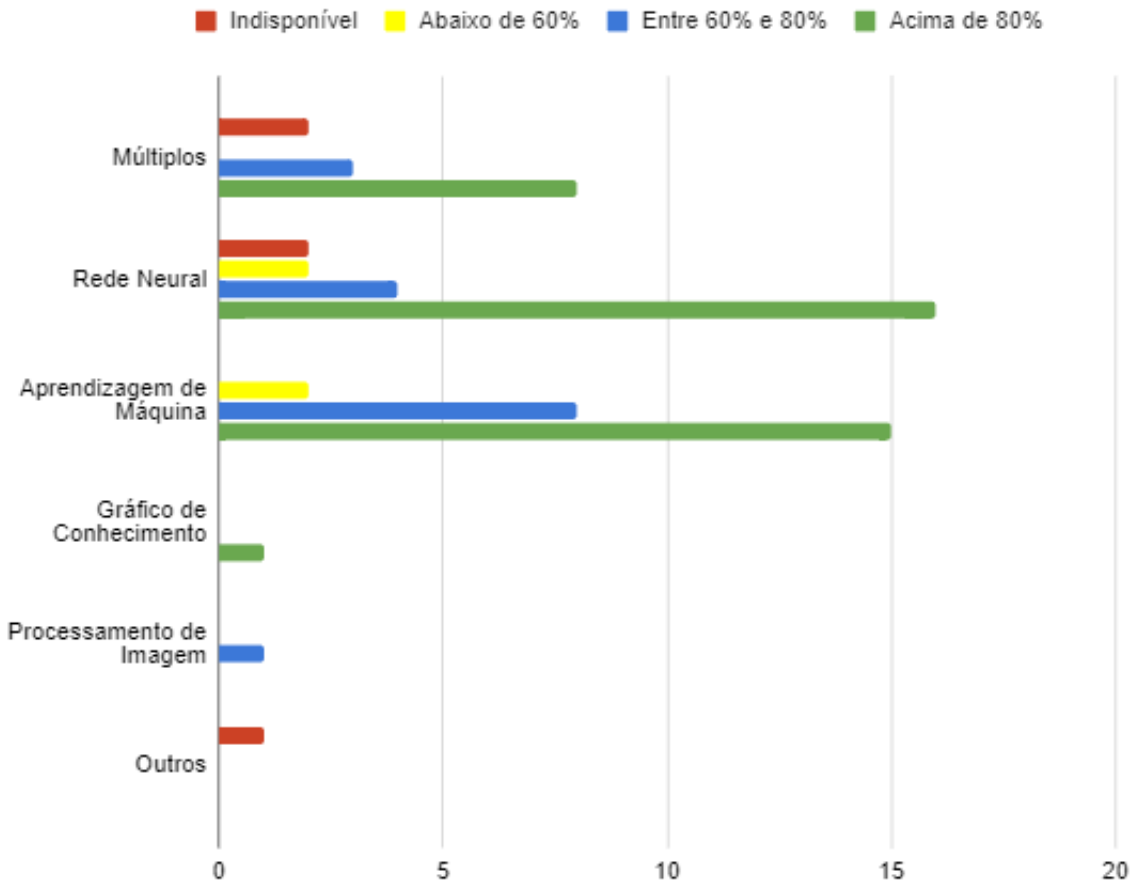
Fonte: elaborada pelo autor.

Quando o critério do método é observado, isoladamente, pouco se pode falar sobre os métodos de verificação de propagação, verificação de credibilidade e verificação de adulteração visual em função da quantidade mínima de ferramentas de detecção que os aplicaram de forma única, conforme figura 15. Mesmo sendo necessário mais estudos sobre estes métodos destaca-se que, as ferramentas que aplicaram os dois primeiros apresentaram desempenho interessante. Em relação aos dois métodos mais aplicados de forma única, a verificação de

fatos e a verificação do estilo de escrita contribuíram para que a maioria das ferramentas de detecção onde foram empregues alcançassem eficácia acima de 80%. Possivelmente em virtude do grande avanço nos últimos anos do Processamento de Linguagem Natural e por serem os métodos mais explorados atualmente. Sobre ferramentas que aplicaram mais de um método também é possível falar pouco por causa da grande volatilidade dos índices de eficácia.

Os últimos apontamentos a serem feitos levando em consideração um cenário local são referentes ao critério campo de conhecimento. As subáreas da ciência prevalecentes entre as ferramentas de detecção foram, quase em sua totalidade, as redes neurais e a aprendizagem de máquina, inclusive de maneira simultânea em muitas delas. Acerca da eficácia, as ferramentas que exploraram somente as redes neurais apresentaram índice melhor quando comparado com aquelas que exploraram somente a aprendizagem de máquina, conforme figura 16. Quando ambas foram exploradas de maneira conjunta pelas ferramentas houve grande volatilidade dos índices de eficácia, prejudicando qualquer tipo de apontamento. Quanto às outras áreas de conhecimento, como gráficos de conhecimento e processamento de imagem, parecem caminhos destoantes ou apostas arriscadas por conta dos poucos exemplares, carece de mais estudo.

Figura 16 – A eficácia dos artigos em um cenário local: campo de conhecimento.



Fonte: elaborada pelo autor.

Em um cenário associativo, dois indicadores, subentende-se evidências, podem ser levantados relacionando a eficácia com os critérios linha de pesquisa, método e campo de conhecimento. Primeiro, ferramentas de detecção que seguiram uma linha de pesquisa híbrida apresentaram eficácia melhor e menos volátil quando o campo de conhecimento explorado foi a aprendizagem de máquina. Segundo, ferramentas que seguiram uma linha de pesquisa em conteúdo tiveram eficácia melhor ao aplicar o método de verificação de fatos explorando, exclusivamente ou simultaneamente, o campo de conhecimento de redes neurais. Demais relações são difíceis de serem observadas porque existe uma escassez de exemplares com determinados códigos de classificação.

Resumidamente, a análise dos cenários confirmam o quão desafiador e complexo é a tarefa da detecção automática de *fake news*. Mesmo após anos de pesquisas, estratégias elaboradas e experimentos realizados pela comunidade científica e acadêmica ainda é necessário uma melhoria contínua. Fato consequente sobretudo do constante poder de transformação contido no fenômeno das *fake news*. Contudo, apoiado nos resultados deste trabalho, é visto que a sociedade tem dado passos, mesmo que pequenos, em direção ao objetivo e que a evolução da tecnologia permite acreditar que em algum momento do futuro os problemas em torno das *fake news* serão solucionados.

7 Considerações Finais

"As palavras fogem quando precisamos delas e sobram quando não pretendemos usá-las."
Carlos Drummond de Andrade

A grande motivação para a realização deste trabalho foi a busca de conhecimento autêntico sobre o fenômeno das *fake news* e a transmissão deste conhecimento para todos os interessados, ou seja, a sociedade em geral. Entender como o fenômeno das *fake news* acontece, divulgar suas consequências e formas de combate podem ser elementos determinantes para alcançar a solução deste desafiador problema originado, infelizmente, pelos próprios seres humanos.

Em razão disso, o presente trabalho apresenta um amplo levantamento bibliográfico sistemático sobre o grau de eficiência dos principais métodos descritos em artigos sobre detecção automática de *fake news* na internet. As informações do levantamento foram extraídas de artigos publicados em bases científicas consideradas referência em repositórios de obras das áreas de ciências exatas e selecionados depois de cumprirem uma série de requisitos referentes a sua qualidade.

De maneira geral, com base nos artigos encontrados, foram mapeados cinco métodos para a detecção automática de *fake news* divididos em duas linhas de pesquisa conforme os recursos de notícia necessários para a concepção dos mesmos. Os conceitos e as estratégias dos métodos mapeados foram expostos de maneira detalhada ao longo deste documento. Além disso, como a maior parte dos estudos primários eram compostos por ferramentas de detecção que aplicavam pelo menos um dos métodos, uma classificação destas foi proposta considerando quatro critérios: linha de pesquisa, campo de conhecimento, acurácia e, obviamente, o método.

Dito isso, podemos afirmar que nosso objetivo geral e objetivos específicos foram alcançados. Através da classificação proposta conseguimos analisar a eficiência dos métodos e apontar algumas relações. Os métodos de verificação de fatos e verificação do estilo de escrita, amplamente mais populares, contribuiram para que a maioria de suas ferramentas alcançassem mais de 80% de eficácia. Os métodos de verificação de propagação e verificação de credibilidade também contribuíram para que suas ferramentas alcançassem uma eficácia parecida, contudo a quantidade de ferramentas que os aplicaram era muito baixa.

Algumas das limitações deste trabalho foram que o número de estudos primários que discorriam de métodos para detecção de *fake news* em formato de imagem eram poucos e para formatos de áudio e vídeo nenhum foi encontrado. Inclusive, as limitações citadas anteriormente podem ser sugestões de trabalhos futuros visto que os criadores de *fake news* cada vez mais utilizam de novos formatos para espalhar suas mentiras. Outros possíveis trabalhos futuros são levantamentos bibliográficos sobre a detecção precoce e a explicabilidade de *fake*

news e o desenvolvimento de ferramentas que auxiliem na correção de equívocos causados em pessoas expostas a *fake news*.

Em suma, os métodos para a detecção automatizada de *fake news* evoluíram muito nos últimos anos e atualmente existem ferramentas que apresentam resultados promissores para esta tarefa, porém ainda não ideais. A expectativa é que o documento em evidência possa contribuir com a literatura acadêmica nacional, e principalmente apoiar projetos que busquem desenvolver e melhorar algum método ou ferramenta a respeito deste importante tema para a sociedade.

Referências

ADLER, M. J.; DOREN, C. V. *Como ler livros: o guia clássico para a leitura inteligente*. [S.l.]: É Realizações, 2010. Citado na página 39.

AHMED, H.; TRAORE, I.; SAAD, S. Detection of online fake news using n-gram analysis and machine learning techniques. In: SPRINGER. *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*. [S.l.], 2017. p. 127–138. Citado nas páginas 48 e 51.

AJAO, O.; BHOWMIK, D.; ZARGARI, S. Fake news identification on twitter with hybrid cnn and rnn models. In: *Proceedings of the 9th international conference on social media and society*. [S.l.: s.n.], 2018. p. 226–230. Citado na página 48.

AJAO, O.; BHOWMIK, D.; ZARGARI, S. Sentiment aware fake news detection on online social networks. In: IEEE. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.], 2019. p. 2507–2511. Citado na página 48.

Al Asaad, B.; Erascu, M. A tool for fake news detection. In: *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. [S.l.: s.n.], 2018. p. 379–386. Citado na página 48.

AL-ASH, H. S.; WIBOWO, W. C. Fake news identification characteristics using named entity recognition and phrase detection. In: IEEE. *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*. [S.l.], 2018. p. 12–17. Citado na página 48.

ALLCOTT, H.; GENTZKOW, M. Social media and fake news in the 2016 election. *Journal of economic perspectives*, v. 31, n. 2, p. 211–36, 2017. Citado nas páginas 11, 13, 15, 16, 17 e 18.

ATODIRESEI, C.-S.; TĂNĂSELEA, A.; IFTENE, A. Identifying fake news and fake users on twitter. *Procedia Computer Science*, Elsevier, v. 126, p. 451–461, 2018. Citado na página 47.

BAHAD, P.; SAXENA, P.; KAMAL, R. Fake news detection using bi-directional lstm-recurrent neural network. *Procedia Computer Science*, Elsevier, v. 165, p. 74–82, 2019. Citado na página 48.

BHATT, G. et al. Combining neural, statistical and external features for fake news stance identification. In: *Companion Proceedings of the The Web Conference 2018*. [S.l.: s.n.], 2018. p. 1353–1357. Citado na página 48.

BOURGONJE, P.; SCHNEIDER, J. M.; REHM, G. From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In: *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*. [S.l.: s.n.], 2017. p. 84–89. Citado na página 48.

BOYD, D. M.; ELLISON, N. B. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated communication*, Oxford University Press Oxford, UK, v. 13, n. 1, p. 210–230, 2007. Citado na página 11.

BUNTAIN, C.; GOLBECK, J. Automatically identifying fake news in popular twitter threads. In: IEEE. *2017 IEEE International Conference on Smart Cloud (SmartCloud)*. [S.l.], 2017. p. 208–215. Citado na página 47.

BURKHARDT, J. M. History of fake news. *Library Technology Reports*, v. 53, n. 8, p. 5–9, 2017. Citado na página 16.

CASTELO, S. et al. A topic-agnostic approach for identifying fake news pages. In: *Companion Proceedings of The 2019 World Wide Web Conference*. [S.l.: s.n.], 2019. p. 975–980. Citado na página 47.

CHEN, Y.; CONROY, N. J.; RUBIN, V. L. Misleading online content: recognizing clickbait as"false news". In: *Proceedings of the 2015 ACM on workshop on multimodal deception detection*. [S.l.: s.n.], 2015. p. 15–19. Citado na página 33.

CHOPRA, S.; JAIN, S.; SHOLAR, J. M. Towards automatic identification of fake news: Headline-article stance detection with lstm attention models. In: *Stanford CS224d Deep Learning for NLP final project*. [S.l.: s.n.], 2017. Citado na página 48.

CONFORTO, E. C.; AMARAL, D. C.; SILVA, S. d. Roteiro para revisão bibliográfica sistemática: aplicação no desenvolvimento de produtos e gerenciamento de projetos. *Trabalho apresentado*, v. 8, 2011. Citado na página 18.

CONROY, N. K.; RUBIN, V. L.; CHEN, Y. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, Wiley Online Library, v. 52, n. 1, p. 1–4, 2015. Citado na página 12.

DAVIS, R.; PROCTOR, C. *Fake news, real consequences: Recruiting neural networks for the fight against fake news*. [S.l.]: Stanford CS224d Deep Learning for NLP final project, 2017. Citado na página 48.

DEY, A. et al. Fake news pattern recognition using linguistic analysis. In: IEEE. *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. [S.l.], 2018. p. 305–309. Citado na página 47.

ELHADAD, M. K.; LI, K. F.; GEBALI, F. Fake news detection on social media: a systematic survey. In: IEEE. *2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*. [S.l.], 2019. p. 1–8. Citado na página 25.

ELKASRAWI, S. et al. What you see is what you get? automatic image verification for online news content. In: IEEE. *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*. [S.l.], 2016. p. 114–119. Citado nas páginas 30 e 48.

FERNANDEZ, A. C. T.; DEVARAJ, M. Computing the linguistic-based cues of fake news in the philippines towards its detection. In: *Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics*. [S.l.: s.n.], 2019. p. 1–9. Citado nas páginas 29 e 48.

FIGUEIRA, Á.; OLIVEIRA, L. The current state of fake news: challenges and opportunities. *Procedia Computer Science*, Elsevier, v. 121, p. 817–825, 2017. Citado nas páginas 11, 12 e 17.

FONTANA, R. M. et al. Aleteia: sistema de classificação de notícias. Citado na página 48.

GIACHANOU, A.; ROSSO, P.; CRESTANI, F. Leveraging emotional signals for credibility detection. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. [S.l.: s.n.], 2019. p. 877–880. Citado na página 48.

GRANIK, M.; MESYURA, V. Fake news detection using naive bayes classifier. In: IEEE. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*. [S.l.], 2017. p. 900–903. Citado na página 48.

GUPTA, S. et al. Cimtdetect: a community infused matrix-tensor coupled factorization based method for fake news detection. In: IEEE. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. [S.l.], 2018. p. 278–281. Citado na página 47.

HADDAWAY, N. R. et al. The role of google scholar in evidence reviews and its applicability to grey literature searching. *PloS one*, Public Library of Science, v. 10, n. 9, p. e0138237, 2015. Citado na página 20.

HARDALOV, M.; KOYCHEV, I.; NAKOV, P. In search of credible news. In: SPRINGER. *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*. [S.l.], 2016. p. 172–180. Citado nas páginas 43 e 47.

HELMSTETTER, S.; PAULHEIM, H. Weakly supervised learning for fake news detection on twitter. In: IEEE. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. [S.l.], 2018. p. 274–277. Citado na página 47.

HIGGINS, J. P. et al. *Cochrane handbook for systematic reviews of interventions*. [S.l.]: John Wiley & Sons, 2019. Citado na página 19.

HOWELL, L. et al. Digital wildfires in a hyperconnected world. *WEF report*, v. 3, n. 2013, p. 15–94, 2013. Citado na página 12.

HUH, M. et al. Fighting fake news: Image splice detection via learned self-consistency. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. [S.l.: s.n.], 2018. p. 101–117. Citado nas páginas 30, 31 e 48.

JAIN, A.; KASBE, A. Fake news detection. In: IEEE. *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. [S.l.], 2018. p. 1–5. Citado nas páginas 48 e 51.

JANG, S. M. et al. A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behavior*, Elsevier, v. 84, p. 103–113, 2018. Citado nas páginas 31, 32 e 33.

JANZE, C.; RISIUS, M. Automatic detection of fake news on social media platforms. In: *PACIS*. [S.l.: s.n.], 2017. p. 261. Citado na página 47.

JIN, L. et al. Understanding user behavior in online social networks: A survey. *IEEE Communications Magazine*, IEEE, v. 51, n. 9, p. 144–150, 2013. Citado na página 11.

JR, E. C. T.; LIM, Z. W.; LING, R. Defining "fake news" a typology of scholarly definitions. *Digital journalism*, Taylor & Francis, v. 6, n. 2, p. 137–153, 2018. Citado nas páginas 15, 16 e 17.

JUNIOR, W. T. L. Mídia social conectada: produção colaborativa de informação de relevância social em ambiente tecnológico digital. *Líbero*, n. 24, p. 95–106, 2016. Citado na página 11.

KALIYAR, R. K. Fake news detection using a deep neural network. In: IEEE. *2018 4th International Conference on Computing Communication and Automation (ICCCA)*. [S.l.], 2018. p. 1–7. Citado na página 47.

KARIMI, H. et al. Multi-source multi-class fake news detection. In: *Proceedings of the 27th International Conference on Computational Linguistics*. [S.l.: s.n.], 2018. p. 1546–1557. Citado na página 48.

KARIMI, H.; TANG, J. Learning hierarchical discourse-level structure for fake news detection. *arXiv preprint arXiv:1903.07389*, 2019. Citado na página 47.

KHATTAR, D. et al. Mvae: Multimodal variational autoencoder for fake news detection. In: *The World Wide Web Conference*. [S.l.: s.n.], 2019. p. 2915–2921. Citado na página 48.

KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele University*, v. 33, n. 2004, p. 1–26, 2004. Citado na página 18.

KITCHENHAM, B. et al. Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, Elsevier, v. 51, n. 1, p. 7–15, 2009. Citado na página 20.

KOTTETI, C. M. M. et al. Fake news detection enhancement with data imputation. In: IEEE. *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. [S.l.], 2018. p. 187–192. Citado na página 47.

LAZER, D. M. et al. The science of fake news. *Science*, American Association for the Advancement of Science, v. 359, n. 6380, p. 1094–1096, 2018. Citado nas páginas 12, 15, 16, 17 e 18.

LEAL, I. H. D. S. O uso de aprendizagem de mÁquina para identificaÇÃo e classificaÇÃo de fake news no twitter referentes a eleiÇÃo presidencial de 2018. dec 2018. Citado na página 48.

LEVY, D. et al. *Reuters Institute digital news report 2019*. [S.l.]: Reuters Institute for the Study of Journalism, 2017. v. 2017. Citado na página 11.

LEVY, Y.; ELLIS, T. J. A systems approach to conduct an effective literature review in support of information systems research. *Informing Science*, v. 9, 2006. Citado na página 18.

LIU, Y.; WU, Y.-F. B. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: *Thirty-second AAAI conference on artificial intelligence*. [S.l.: s.n.], 2018. Citado na página 49.

LONG, Y. Fake news detection through multi-perspective speaker profiles. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. [S.l.], 2017. Citado na página 47.

MAHID, Z. I.; MANICKAM, S.; KARUPPAYAH, S. Fake news on social media: Brief review on detection techniques. In: IEEE. *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*. [S.l.], 2018. p. 1–5. Citado na página 25.

MARCHI, R. With facebook, blogs, and fake news, teens reject journalistic “objectivity”. *Journal of communication inquiry*, SAGE Publications Sage CA: Los Angeles, CA, v. 36, n. 3, p. 246–262, 2012. Citado na página 15.

MARUMO, F. S. *Deep Learning para classificação de Fake News por sumarização de texto*. [S.l.]: Londrina, 2018. Citado na página 47.

MOHTARAMI, M. et al. Automatic stance detection using end-to-end memory networks. *arXiv preprint arXiv:1804.07581*, 2018. Citado nas páginas 46 e 48.

MONTEIRO, R. O.; NOGUEIRA, R. R.; MOSER, G. Desenvolvimento de um sistema para a classificação de fakenews com textos de notícias em língua portuguesa. Citado na página 48.

MONTI, F. et al. Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673*, 2019. Citado nas páginas 25 e 49.

OECD. *An Introduction to Online Platforms and Their Role in the Digital Transformation*. [s.n.], 2019. 216 p. Disponível em: <https://www.oecd-ilibrary.org/content/publication/53e5f593-en>. Citado na página 11.

OSHIKAWA, R.; QIAN, J.; WANG, W. Y. A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*, 2018. Citado na página 29.

PAN, J. Z. et al. Content based fake news detection using knowledge graphs. In: SPRINGER. *International semantic web conference*. [S.l.], 2018. p. 669–683. Citado na página 48.

PARIKH, S. B.; ATREY, P. K. Media-rich fake news detection: A survey. In: IEEE. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. [S.l.], 2018. p. 436–441. Citado nas páginas 12 e 28.

PASQUINI, C. et al. Towards the verification of image integrity in online news. In: IEEE. *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. [S.l.], 2015. p. 1–6. Citado na página 30.

PÉREZ-ROSAS, V. et al. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*, 2017. Citado na página 47.

PFOHL, O. T.; LEGROS, F. Stance detection for the fake news challenge with attention and conditional encoding. *CS224n: Natural Language Processing with Deep Learning*, 2017. Citado na página 48.

PRATIWI, I. Y. R.; ASMARA, R. A.; RAHUTOMO, F. Study of hoax news detection using naïve bayes classifier in indonesian language. In: IEEE. *2017 11th International Conference on Information & Communication Technology and System (ICTS)*. [S.l.], 2017. p. 73–78. Citado na página 48.

QIAN, F. et al. Neural user response generator: Fake news detection with collective user intelligence. In: *IJCAI*. [S.l.: s.n.], 2018. v. 18, p. 3834–3840. Citado na página 47.

RASOOL, T. et al. Multi-label fake news detection using multi-layered supervised learning. In: *Proceedings of the 2019 11th International Conference on Computer and Automation Engineering*. [S.l.: s.n.], 2019. p. 73–77. Citado na página 47.

RAUPP, F. M.; BEUREN, I. M. Metodologia da pesquisa aplicável às ciências. *Como elaborar trabalhos monográficos em contabilidade: teoria e prática. São Paulo: Atlas*, p. 76–97, 2006. Citado na página 20.

REIS, J. C. et al. Explainable machine learning for fake news detection. In: *Proceedings of the 10th ACM Conference on Web Science*. [S.l.: s.n.], 2019. p. 17–26. Citado na página 47.

RIEDEL, B. et al. A simple but tough-to-beat baseline for the fake news challenge stance detection task. *arXiv preprint arXiv:1707.03264*, 2017. Citado na página 48.

ROETS, A. et al. 'fake news': Incorrect, but hard to correct. the role of cognitive ability on the impact of false information on social impressions. *Intelligence*, Elsevier, v. 65, p. 107–110, 2017. Citado na página 13.

ROTHER, E. T. Revisão sistemática x revisão narrativa. *Acta paulista de enfermagem*, Universidade Federal de São Paulo, v. 20, n. 2, p. v–vi, 2007. Citado na página 18.

RUBIN, V. L. et al. Fake news or truth? using satirical cues to detect potentially misleading news. In: *Proceedings of the second workshop on computational approaches to deception detection*. [S.l.: s.n.], 2016. p. 7–17. Citado na página 47.

RUCHANSKY, N.; SEO, S.; LIU, Y. Csi: A hybrid deep model for fake news detection. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. [S.l.: s.n.], 2017. p. 797–806. Citado na página 49.

SADIQ, S. et al. High dimensional latent space variational autoencoders for fake news detection. In: IEEE. *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. [S.l.], 2019. p. 437–442. Citado na página 47.

SAKURAI, G. Y. Processamento de linguagem natural-detecção de fake news. Citado na página 48.

SAMONTE, M. J. C. Polarity analysis of editorial articles towards fake news detection. In: *Proceedings of the 2018 International Conference on Internet and e-Business*. [S.l.: s.n.], 2018. p. 108–112. Citado na página 48.

SEO, Y.; SEO, D.; JEONG, C.-S. Fander: fake news detection model using media reliability. In: IEEE. *TENCON 2018-2018 IEEE Region 10 Conference*. [S.l.], 2018. p. 1834–1838. Citado na página 48.

SHU, K. et al. defend: Explainable fake news detection. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. [S.l.: s.n.], 2019. p. 395–405. Citado na página 47.

SHU, K.; MAHUDESWARAN, D.; LIU, H. Fakenewstracker: a tool for fake news collection, detection, and visualization. *Computational and Mathematical Organization Theory*, Springer, v. 25, n. 1, p. 60–71, 2019. Citado na página 47.

SHU, K. et al. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, ACM New York, NY, USA, v. 19, n. 1, p. 22–36, 2017. Citado nas páginas 11, 12, 15, 16, 17, 18, 25, 28, 30, 31 e 34.

SHU, K.; WANG, S.; LIU, H. Beyond news contents: The role of social context for fake news detection. In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. [S.l.: s.n.], 2019. p. 312–320. Citado na página 47.

SILVERMAN, C. This analysis shows how viral fake election news stories outperformed real news on facebook. *BuzzFeed news*, v. 16, 2016. Citado na página 34.

STAHL, K. Fake news detection in social media. *California State University Stanislaus*, v. 6, 2018. Citado nas páginas 26, 27 e 28.

SU, T.; MACDONALD, C.; OUNIS, I. Ensembles of recurrent networks for classifying the relationship of fake news titles. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. [S.l.: s.n.], 2019. p. 893–896. Citado na página 48.

TACCHINI, E. et al. Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506*, 2017. Citado nas páginas 13 e 49.

TARMIZI, F. A. A. et al. Online news veracity assessment using emotional weight. In: *Proceedings of the 2019 2nd International Conference on Information Science and Systems*. [S.l.: s.n.], 2019. p. 60–64. Citado na página 48.

TEIXEIRA, R. B. AutomatizaÇÃo de teste unitÁrio de software: Levantamento bibliogrÁfico. dec 2016. Citado nas páginas 20 e 23.

THOTA, A. et al. Fake news detection: a deep learning approach. *SMU Data Science Review*, v. 1, n. 3, p. 10, 2018. Citado na página 48.

TRAYLOR, T. et al. Classifying fake news articles using natural language processing to identify in-article attribution as a supervised learning estimator. In: IEEE. *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*. [S.l.], 2019. p. 445–449. Citado na página 48.

VEDOVA, M. L. D. et al. Automatic online fake news detection combining content and social signals. In: IEEE. *2018 22nd Conference of Open Innovations Association (FRUCT)*. [S.l.], 2018. p. 272–279. Citado nas páginas 25, 43 e 47.

VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. *Science*, American Association for the Advancement of Science, v. 359, n. 6380, p. 1146–1151, 2018. Citado na página 12.

WANG, Y. et al. Eann: Event adversarial neural networks for multi-modal fake news detection. In: *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*. [S.l.: s.n.], 2018. p. 849–857. Citado na página 48.

WU, K.; YANG, S.; ZHU, K. Q. False rumors detection on sina weibo by propagation structures. In: IEEE. *2015 IEEE 31st international conference on data engineering*. [S.l.], 2015. p. 651–662. Citado na página 33.

WU, L.; LIU, H. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In: *Proceedings of the eleventh ACM international conference on Web Search and Data Mining*. [S.l.: s.n.], 2018. p. 637–645. Citado na página 49.

YANG, F. et al. Xfake: explainable fake news detector with visualizations. In: *The World Wide Web Conference*. [S.l.: s.n.], 2019. p. 3600–3604. Citado na página 47.

YANG, Y. et al. Ti-cnn: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*, 2018. Citado na página 48.

ZHOU, X. et al. Real-time news cer tification system on sina weibo. In: *Proceedings of the 24th International Conference on World Wide Web*. [S.l.: s.n.], 2015. p. 983–988. Citado na página 47.

ZHOU, X.; ZAFARANI, R. Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*, v. 2, 2018. Citado nas páginas 26, 27, 28, 31, 32, 33, 34 e 35.

Apêndices

APÊNDICE A – Artigos descartados

Estas são as referêncidas dos artigos...

- ALDWAIRI, M.; ALWAHEDI, A. Detecting fake news in social media networks. Procedia Computer Science, Elsevier, v. 141, p. 215–222, 2018.
- ARAUJO, Y. de L.; CHARES, A. C.; SAMPAIO, J. de O. Identificação de fake news: uma abordagem utilizando métodos de busca e chatbots. In: SBC. Anais do VII Brazilian Workshop on Social Network Analysis and Mining. [S.l.], 2018.
- BHATTACHARJEE, S. D.; TALUKDER, A.; BALANTRAPU, B. V. Active learning based news veracity detection with feature weighting and deep-shallow fusion. In: IEEE. 2017 IEEE International Conference on Big Data (Big Data). [S.l.], 2017. p. 556–565.
- GIRGIS, S.; AMER, E.; GADALLAH, M. Deep learning algorithms for detecting fake news in online text. In: IEEE. 2018 13th International Conference on Computer Engineering and Systems (ICCES). [S.l.], 2018. p. 93–97.
- KO, H. et al. Human-machine interaction: A case study on fake news detection using a backtracking based on a cognitive system. Cognitive Systems Research, Elsevier, v. 55, p. 77–81, 2019.
- LIMA, P. de A.; AMARAL, É. Existem ferramentas digitais capazes de reduzir a disseminação das fake news? Anais do Salão Internacional de Ensino, Pesquisa e Extensão, v. 10, n. 2, 2019.
- MARRA, F. et al. Detection of gan-generated fake images over social networks. In: IEEE. 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). [S.l.], 2018. p. 384–389.
- OKORO, E. et al. A hybrid approach to fake news detection on social media. Nigerian Journal of Technology, Faculty of Engineering University of Nigeria Nsukka, v. 37, n. 2, p. 454–462, 2018.
- PINHEIRO, A.; CAPPELLI, C.; MACIEL, C. Adoção da auditabilidade como proposta para identificar informações falsas em redes sociais. In: SBC. Anais do VII Workshop sobre Aspectos da Interação Humano-Computador para a Web Social. [S.l.], 2017. p. 65–71.
- SHABANI, S.; SOKHN, M. Hybrid machine-crowd approach for fake news detection. In: IEEE. 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC). [S.l.], 2018. p. 299–306.
- SILVA, F. R. d. L. Análise de fontes de informação como critério no combate à desinformação e Fake News. 2019. Dissertação (B.S. thesis) — Universidade Federal do Rio Grande do Norte, 2019.

- SIRAJUDEEN, S. M.; AZMI, N. F. A.; ABUBAKAR, A. I. Online fake news detection algorithm. Journal of Theoretical Applied Information Technology, v. 95, n. 17, 2017.
- VO, N.; LEE, K. Learning from fact-checkers: Analysis and generation of fact-checking language. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. [S.l.: s.n.], 2019. p. 335–344.
- ZHANG, Q.; YILMAZ, E.; LIANG, S. Ranking-based method for news stance detection. In: Companion Proceedings of the The Web Conference 2018. [S.l.: s.n.], 2018. p. 41–42.
- ZHOU, X.; ZAFARANI, R. Fake news detection: An interdisciplinary research. In: Companion Proceedings of The 2019 World Wide Web Conference. [S.l.: s.n.], 2019. p. 1292–1292.