



UNIVERSIDADE FEDERAL DO MARANHÃO
Curso de Ciência da Computação

Lucas Reis Abreu

**Mineração de texto aplicada à identificação de
Fake News**

São Luís/MA
2019

Lucas Reis Abreu

Mineração de texto aplicada à identificação de Fake News

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Trabalho aprovado em São Luís - MA, 10 de Dezembro de 2019

Prof. Dr. Ivo José da Cunha Serra
Orientador

Prof. Dra. Simara Vieira da Rocha
Examinadora 1

Prof. Dr. Geraldo Braz Júnior
Examinador 2

São Luís/MA
2019

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Reis Abreu, Lucas.

Mineração de Texto aplicada à identificação de Fake News / Lucas Reis Abreu. - 2019.

50 f.

Orientador(a): Ivo José da Cunha Serra.

Monografia (Graduação) - Curso de Ciência da Computação, Universidade Federal do Maranhão, São Luís, 2019.

1. Aprendizado de Máquina. 2. Fake News. 3. Mineração de Texto. 4. Notícias. I. da Cunha Serra, Ivo José. II. Título.

Agradecimentos

Em primeiro lugar, agradeço à minha família que me acolheu, me deu amor, educação e tudo que foi necessário para o meu caminho até aqui.

Agradeço a minha namorada linda e maravilhosa, Erika Jordana, que nunca me abandonou nas minhas surtadas durante esse último ano, sempre se fez presente e me apoiou nos momentos de desespero. Agradeço por sempre me dar forças e acreditar em mim, mesmo quando eu não acreditei, não sei onde estaria ultimamente sem essa mulher, te amo.

Agradeço também à minha filha de quatro patas, Mera, que embora me acorde sempre de madrugada ainda me faz sorrir todos os dias.

Aos meus poucos amigos do ensino médio que se mantiveram, em especial sei que posso contar com vocês, Edy, Kuro, Coelho e Kayla,.

Ao PETComp, grupo do qual participei por quase toda a minha graduação e influenciou diretamente na minha vida acadêmica e pessoal. Em especial, um agradecimento a turma liderada pelo professor Geraldo, digo com tranquilidade que foi o melhor grupo, essa é pra vocês True Friends.

À professora Simara, que foi minha primeira orientadora e me concedeu conselhos de uma mãe, me dando forças para seguir em frente. E diga-se de passagem a melhor coordenadora do curso.

Ao professor Geraldo, o qual tive o prazer de trabalhar junto durante minha passagem pelo PET, se mostrando um excelente tutor e amigo.

Ao meu orientador Ivo, sempre se mostrando prestativo e atencioso.

Agradeço imensamente aos meus amigos do "Void UpdateCarol", Sky, Rod, Carol, Nenê, Vitu e Vini, so quem viveu sabe. Amo vocês demais, obrigado por sempre me fazerem rir quando estava na pior.

À Winne, uma das melhores amigas que eu poderia fazer nessa universidade. Apesar da nossa relação ser entre tapas e beijos, você sempre me ouviu quando eu precisei desabafar, sempre esteve lá para me apoiar ou criticar, dando conselhos ou me batendo.

À Sky, que também tenho o prazer de chamar de melhor amiga, agradeço por sempre me escutar, rir das minhas idiotices e fazer nossas saídas sempre melhores.

A Bubuca que foi a melhor amizade que cultivei no meu curso anterior, se mantendo presente até hoje, formamos a melhor dupla meu lindo (DC pai).

A Vitu, que carregou toda uma turma no 3º período, agradeço por me aguentar e ser esse grande amigo.

Agradeço a Marvin, pois só ele pra me salvar dizendo que a parte dos agradecimentos não tem quantidade específica.

Quero registrar um agradecimento à Kayla, que embora estejamos longe, ela foi e sempre será uma parte importante da minha vida, sendo a irmã que eu nunca tive. Pra mim tu sempre vai ter 15 anos, e vou sempre querer cuidar de ti.

À Katia Constantino, minha melhor amiga virtual, sempre me deixando a par da cultura japonesa e conversando horas entre idiotices e assuntos sérios.

Deixo um agradecimento também a Rezende, Bruno e Hugo, vocês são feras.

Tenho que agradecer a Pereirão e Eug, a gente quase nunca sai, mas as conversas são sempre muito boas e os conselhos então, nem se falam.

E por fim, ao Sr. França, se mostrou um cara excepcional nos momentos lúdicos em que reservávamos para pensar sobre a vida.

"Part of the journey is the end. What am i even tripping for? Everything's going to work out exactly the way it's supposed to. I LOVE YOU 3000"

(Tony Stark)

Resumo

Na atualidade devido ao avanço nas Tecnologias de Informação e Comunicação que influenciaram principalmente o avanço da internet e de suas ferramentas, como as mídias sociais, por exemplo, o crescimento dos dados gerados e armazenados cresceram de forma exponencial na última década. Este fato acabou possibilitando que os usuários pudessem expressar suas opiniões e compartilhar diversas notícias e informações pela internet, porém algumas vezes tais informações não são verídicas e mesmo assim são espalhadas como se fossem verdade, tal fenômeno é conhecido como *fake news*. Este evento pode ser compreendido como notícias falsas, as quais são criadas e divulgadas com um objetivo específico, seja ele político ou financeiro. Dentro desse contexto existe um cenário que necessita da aplicação de técnicas e métodos para a identificação automática das *fake news*, haja vista a quantidade de informações disponíveis se torna impossível uma classificação de modo manual. Neste trabalho são elucidados abordagens que utilizam a mineração de texto, podendo ser utilizadas como solução automática na identificação e classificação de notícias falsas, assim como é demonstrado nos estudos de casos já desenvolvidos nessa perspectiva, visando mostrar os resultados dessas pesquisas ao discutir sua relevância para a área.

Palavras-chave: Mineração de texto. *Fake news*. Aprendizado de máquina. Notícias.

Abstract

In the present due to the advancement in the Technologies of Information and Communication that they influenced mainly the advancement of the Internet and of his tools, like the social medias, for example, the growth of the produced and stored data they grew in the form exponencial in the last decade. This fact finished making possible that the users could express his opinions and share several news and informations for the Internet, however sometimes such informations are not true and are even so spread as if they were true, such a phenomenon is known how fake news. This event can be understood how false, what news is created and spread with a specific objective, be political or financial he. Inside this context there is a scenery who needs the application of techniques and methods for the automatic identification of the fake news, there is in sight the quantity of available informations impossible compensation a classification of manual way. In this work is elucidated approaches that use the text mining, being able to be used like automatic solution in the identification and classification of false news, as well as it is demonstrated in the studies of cases already developed in this perspective, aiming to show the results of these inquiries while discussing his relevance for the area.

Keywords: Text mining. Fake news. Machine learning. News.

Lista de ilustrações

Figura 1 – Tipos de <i>fake news</i> segundo Wardle (2017). Fonte: (CARVALHO, 2018).	16
Figura 2 – Exemplo de paródia/sátira. Fonte: (MORAES, 2019).	18
Figura 3 – Exemplo de Contexto Enganoso. Fonte: (LOPES, 2018).	18
Figura 4 – Exemplo de Conteúdo Impostor. Fonte: (G1, 2018).	19
Figura 5 – Exemplo de Contexto Fabricado. Fonte: (VEJA, 2018).	20
Figura 6 – Exemplo de Falsa Conexão. Fonte: (SOUZA, 2017).	20
Figura 7 – Exemplo de Falso Contexto. Fonte: (PAIVA, 2018).	21
Figura 8 – Exemplo de Conteúdo Manipulado. Fonte: (MARES, 2019).	21
Figura 9 – Processo da Mineração de texto. Fonte: (BARION; LAGO, 2015).	27
Figura 10 – Exemplo de <i>fake news</i> . Fonte: (MARUMO, 2018).	31
Figura 11 – Exemplo de notícia verídica. Fonte: (MARUMO, 2018).	31
Figura 12 – Exemplo de notícia real sumarizada. Fonte: (MARUMO, 2018).	32
Figura 13 – Colunas da Tabela dos dados coletados do Twitter. Fonte: (LEAL, 2018).	36
Figura 14 – <i>Hashtags</i> utilizadas na aquisição dos dados. Fonte: (LEAL, 2018).	36
Figura 15 – Exemplo de arquivo .ARFF. Fonte: (LEAL, 2018).	37
Figura 16 – Código do autor para geração do arquivo .ARFF referente a sua base de dados. Fonte: (LEAL, 2018).	37
Figura 17 – Exemplo de notícia classificada como fato. Fonte: (LEAL, 2018).	39
Figura 18 – Exemplo de notícia classificada como fake. Fonte: (LEAL, 2018).	39
Figura 19 – Exemplo de notícia classificada como opinião. Fonte: (LEAL, 2018).	40
Figura 20 – Quantidade de registros classificados nas respectivas classes. Fonte: (LEAL, 2018).	40
Figura 21 – Percentual de registros classificados em cada classe. Fonte: (LEAL, 2018).	40
Figura 22 – Percentual de acertos com o parâmetro <i>seed</i> variando de 1 a 10. Fonte: (LEAL, 2018).	41
Figura 23 – Melhor desempenho do algoritmo <i>NaiveBayes</i> . Fonte: (LEAL, 2018).	41
Figura 24 – Melhor desempenho do algoritmo J48. Fonte: (LEAL, 2018).	42
Figura 25 – Melhor resultado do algoritmo <i>JRip</i> . Fonte: (LEAL, 2018).	42
Figura 26 – Melhor desempenho do algoritmo iBK. Fonte: (LEAL, 2018).	42

Lista de tabelas

Tabela 1 – Estatísticas sobre notícias reais e <i>fake news</i> . Fonte: (MARUMO, 2018).	31
Tabela 2 – Quantidade de amostras do domínio. Fonte: (MARUMO, 2018).	34
Tabela 3 – Resultados da classificação com <i>Deep Learning</i> . Fonte: (MARUMO, 2018).	34
Tabela 4 – Resultados da classificação com IA Tradicional. Fonte: (MARUMO, 2018).	34

Lista de abreviaturas e siglas

UFMA	Universidade Federal do Maranhão
UEL	Universidade Estadual de Londrina
EUA	Estados Unidos da América
IA	Inteligência Artificial
RNN	Recurrent Neural Network

Sumário

	Lista de tabelas	9
1	INTRODUÇÃO	12
1.1	Objetivos	13
1.1.1	Objetivos Específicos	13
1.2	Estrutura do Trabalho	13
2	FAKE NEWS	14
2.1	O que são as <i>Fake News</i>?	14
2.2	Como funcionam as <i>Fake News</i>	15
2.3	Tipos de <i>Fake News</i>	16
2.3.1	Sátira ou Paródia	17
2.3.2	Contexto Enganoso	18
2.3.3	Conteúdo Impostor	19
2.3.4	Contexto Fabricado	19
2.3.5	Falsa Conexão	20
2.3.6	Falso Contexto	21
2.3.7	Conteúdo Manipulado	21
2.4	Disseminação	22
2.5	Consequências das <i>Fake News</i> e seus impactos na sociedade	23
2.6	Como combater as <i>Fake News</i>	24
3	MINERAÇÃO DE TEXTO E <i>FAKE NEWS</i>	26
3.1	Mineração de Texto	26
3.1.1	Tarefas da Mineração de Texto	28
3.2	Estudos de Caso: Mineração de Texto e <i>fake news</i>	29
3.2.1	Classificação de <i>fake news</i> com o uso de <i>Deep Learning</i>	29
3.2.2	Aprendizagem de máquina na classificação de <i>fake news</i> obtidas do <i>Twitter</i>	35
3.3	Discussão	43
4	CONCLUSÃO	46
	REFERÊNCIAS	47

1 INTRODUÇÃO

Para conceituar as *fake news* com base nos estudos de Gelfert (2018), o termo pode ser apontado como a apresentação deliberada de notícias falsas ou enganosas, onde segundo o autor, tais informações são tipicamente falsas podendo então serem fabricadas afirmações inverídicas sobre fatos verdadeiros. Nos estudos para a elaboração desta monografia, foi constatado que as *fake news* se tornaram um fenômeno mundial, embora seu conceito não seja novo. Tal expansão se deve principalmente ao uso de seus métodos em campanhas políticas para ganho monetário ou alguma vantagem política, como por exemplo nas eleições presidenciais dos Estados Unidos da América (EUA), onde Donald Trump saiu vitorioso, fato que alguns atribuem a uma significativa influência por parte das *fake news* que seus eleitores divulgaram a respeito da candidata adversária, Hillary Clinton.

Problemas envolvendo *fake news* não são poucos, há relatos desde a disseminação de notícias falsas com a intenção de manchar a reputação de alguém, até casos de óbitos ocasionados por simples boatos espalhados dentro de redes sociais. Com o significativo avanço da internet, a disseminação das *fake news* se tornou um evento difícil de ser contido, uma vez que existem algoritmos, robôs, instituições mal intencionadas e um público desinformado que contribuem para a dispersão dessas notícias inverídicas.

A propagação de notícias falsas pode gerar grandes problemas para democracia de um país, tendo em vista que as *fake news* são amplamente utilizadas em campanhas políticas, como pode ser observado nas recentes eleições presidenciais do Brasil no ano de 2018. Neste dado período foi possível notar muitas notícias inverídicas por parte de alguns partidos políticos, visando claramente obter vantagens sobre seus concorrentes. Tal ato ameaça não só a democracia do Brasil, como a de qualquer outro país, uma vez que um país desestabilizado ou um público mais desinformado se torna mais fácil de ser controlado.

Visando tais consequências descritas acima, pesquisadores, estudantes e até mesmo empresas, perceberam que seria interessante realizar o desenvolvimento de ferramentas automatizadas para identificar informações falsas, uma vez que fazer tal trabalho de maneira manual, perante um numeroso aglomerado de notícias possivelmente inverídicas, pode ser inviável. É neste âmbito que é desenvolvido a aplicação da mineração textual, cuja qual se baseia principalmente no processamento de dados em linguagem natural, como por exemplo, para identificação e classificação das *fake news*.

1.1 Objetivos

O objetivo geral é o estudo do uso da mineração de texto no contexto das *fake news*, através da análise de trabalhos que abordam o desenvolvimento de ferramentas automatizadas para a classificação entre *fake news* e notícias verdadeiras.

1.1.1 Objetivos Específicos

- Evidenciar o advento das *fake news*, mostrando seus conceitos, disseminação e consequências.
- Explicar sobre a mineração de texto, suas classificações, divisões e principais utilizações.
- Apresentar trabalhos que abordem a mineração de texto como forma de automatizar a classificação de *fake news*.
- Discutir acerca dos trabalhos apresentados, expondo suas contribuições e relevância para à área.

1.2 Estrutura do Trabalho

Esse estudo está organizado em quatro capítulos, sendo o primeiro a introdução, que contém os objetivos desta monografia. O capítulo 2 apresenta a fundamentação teórica feita através de pesquisa bibliográfica sobre o tema, com pesquisas realizadas em estudos como os de Braga (2018), Luce (2018), Paula, Silva e Blanco (2018), Gelfert (2018) e entre outros. No capítulo 3 é explanado a fundamentação e a metodologia da mineração de texto, além de apresentar e discutir duas pesquisas, sendo elas Marumo (2018) e Leal (2018). E por fim, o capítulo 4 é constituído pelas considerações finais e sugestões de trabalhos futuros.

2 Fake News

Neste capítulo se apresenta a fundamentação teórica à respeito do tema deste trabalho, as *fake news*, de modo que no decorrer deste capítulo, é discutido sobre o conceito deste fenômeno, assim como sua origem, seu funcionamento e quais os seus principais efeitos/impactos na sociedade.

2.1 O que são as *Fake News*?

Para Paula, Silva e Blanco (2018) o conceito das *fake news*, ou em sua tradução para a língua portuguesa sendo "notícias falsas", são compreendidas como informações chamativas tendo como objetivo alertar seu público alvo para uma dada situação ou demonstrar um ponto de vista à respeito de um acontecimento. Embora seja uma informação noticiada, seu escopo possui parte, ou todo seu conteúdo instituído por informações inverídicas.

Normalmente, as *fake news* simulam informações reais com o objetivo principal de gerar polêmica, ou denegrir a imagem de algum indivíduo ou empresa. Devido ao seu elevado poder viral (disseminação acelerada pela internet), este evento costuma atrair a atenção das grandes massas, que em boa parte é desprovida de senso crítico, devido ao seu baixo grau de escolaridade, apresentando assim uma grande força de persuasão sobre o público.

Tais notícias falsas são muito utilizadas na área da política, devido principalmente à sua grande eficácia em manchar as reputações de candidatos rivais, obtendo-se assim vantagens. Isso se deve em especial porque tais informações tem um poder apelativo emocional muito forte em seus títulos, o que acaba atraindo muito a atenção do público. Segundo Braga (2018), tal potencial político é caracterizado por seus autores se utilizarem principalmente de ambientes politicamente polarizados e com a ausência de tolerância, ou seja, os temas noticiosos são sabiamente escolhidos para confirmar críticas e pré-conceitos de um grupo ou outro.

A palavra *fake news* pode ser considerada atual, porém seu significado e atuação são mais antigos. O hábito das pessoas espalharem boatos ou informações falsas pode ser notado há séculos dentro da sociedade, porém não existem registros ou uma data oficial para quando se deu o início de tais atos.

A mentira sempre existiu, seja em rodas de conversas, nos bares ou até na famosa conversa de pescador. Sempre em um tom descontraído, quase jocoso, o assunto se desenvolve a partir de fantasias e contos com baixa veracidade, a fim de distrair as pessoas que estejam envolvidas

nessa conversa. O problema se inicia quando essas conversas ganham proporções maiores e assumem um papel que não é cabido à elas, o de verdade (LUCÉ, 2018).

É percebido então que as *fake news* não são uma novidade, elas sempre existiram, o que mudou foi sua nomenclatura, seu meio de divulgação e potencial de persuasão, sendo esses dois últimos fortemente instigados pelo avanço rápido e grande poder de alcance da internet nos dias atuais.

O termo em inglês, porém, ganhou força e popularidade no ano de 2016, especificamente durante as eleições presidenciais dos Estados Unidos da América (EUA), as quais tinham como candidatos Donald Trump e Hillary Clinton. Segundo o jornal The Guardian (PARKINSON, 2016), o fato dos apoiadores de ambos os candidatos alimentarem diversos sites e/ou redes sociais com diversos rumores ou notícias corrompidas, podem ter levado ao resultado que já conhecemos, a vitória de Donald Trump, fato que contrariava boa parte dos prognósticos daquelas eleições. Com isso a expressão *fake news* ganhou ainda mais evidência ao redor do mundo.

2.2 Como funcionam as *Fake News*

A produção e distribuição das *Fake News* estabelecem um verdadeiro mercado de rumores. Tal universo é alimentado, em sua grande maioria, por pessoas com grande influência (empresários, políticos, grandes empresas, dentre outras), as quais contratam equipes empenhadas e especializadas neste tipo de conteúdo viral. Tais equipes podem ser compostas por ex-jornalistas, profissionais de marketing, publicitários, profissionais de Tecnologia da Informação (TI) e até mesmo policiais para realizar a segurança das sedes e dos equipamentos.

Alguns desses produtores de conteúdo costumam realizar compras ilegais de números e até e-mails de milhões de pessoas, de preferência com grande poder de influência, como líderes religiosos e de movimentos políticos. Uma vez em posse desses meios, os criadores de *fake news*, começam a divulgar publicações verdadeiras, em um primeiro momento, para garantir que seu público alvo irá compartilhar aquela informação e após se consolidarem, iniciam as postagens com títulos chamativos e inverídicos. Como estão em posse de meios reais (telefones, e-mails, páginas oficiais), a credibilidade daquela informação falsa ganha mais força e seus numerosos seguidores passam a compartilhar aquela mensagem como verdadeira.

De acordo com Braga (2018), o alvo principal das *fake news* seria no âmbito da política, onde as informações falsificadas tomam um partido, quer seja ele mais liberal ou conservador, uma vez que as notícias fabricadas geralmente tendem a fortalecer o pensamento pré-estabelecido do indivíduo quanto ao lado oposto, ganhando assim

argumentos para atacar seus adversários. Por exemplo, em um ambiente de um grupo esquerdista, notícias apresentando relatos de pessoas da direita tratando outros indivíduos de forma ríspida ou com desprezo, reforçaria a natureza egoísta e desumana deles, e assim vice-versa em um grupo de direita.

Infelizmente, dentro deste público descrito, parte das pessoas compõem um grupo desinformado ou que não buscam se aprofundar em checar a veracidade daquela informação, o que significa que tais indivíduos são mais propensos a serem persuadidos pelos títulos apelativos e emotivos das notícias adulteradas. Sendo assim, este coletivo, uma vez cativado, passa a compartilhar mais facilmente as notícias sem verificar suas procedências, dando forças ao processo de expansão das *fake news*.

Existe o investimento em diversas locações, uso de cartões de crédito recarregáveis e até mesmo o uso do CPF das próprias vítimas, visando assim dificultar o rastreamento do real paradeiro de onde partem as *fake news*. Como medida de segurança desses investidores, as equipes costumam mudar regularmente de locação, utilizando-se de policiamento real para realizar a segurança dos equipamentos. Além de todas essas medidas físicas, normalmente os dados são salvos na nuvem, o que dificulta ainda mais o rastreamento de quem realmente divulga as notícias.

2.3 Tipos de Fake News

De acordo com Wardle (2017), para entender melhor o ecossistema da informação em que as *fake news* se encontram, elas foram divididas em sete tipos, como pode ser observado na Figura 1.



Figura 1 – Tipos de *fake news* segundo Wardle (2017). Fonte: (CARVALHO, 2018).

A seguir será melhor detalhado e exemplificado cada uma dessas categorias estabelecidas pelo autor.

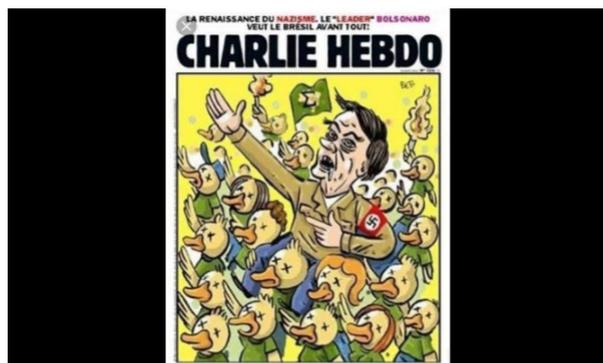
2.3.1 Sátira ou Paródia

As sátiras são definidas como um estilo literário, tendo o objetivo principal de criticar instituições morais, políticas, hábitos e costumes (PORTUGUESA, 2019). Sua característica mais marcante é a forte carga de ironia e sarcasmo, portanto trata-se de uma crítica social feita às pessoas e aos costumes de forma caricata. Embora nem sempre as sátiras tenham por objetivo levar seu público ao riso, seu estilo literário se aproxima bastante da comédia. Por esses motivos, muitas sátiras alvejam pessoas com certa relevância social, como os políticos por exemplo.

Paródias por sua vez são uma imitação de um texto literário, podendo ser a respeito de um personagem ou tema, tendo a finalidade cômica ou irônica. (SIGNIFICADOS, 2016) Por norma, as paródias são como ferramentas de discussão para assuntos polêmicos, de modo que este debate se sucede de maneira mais descontraída e com menos tensão. Uma paródia pode ser feita a partir de um poema, uma música, um filme, uma peça teatral e dentre outras.

Embora as sátiras ou paródias sejam estilos literários que visam ridicularizar, gozar ou criticar comicamente um dado de época, uma instituição ou pessoa, elas serão tratadas, neste trabalho, como um dos tipos de *fake news*. Isso se justifica pelo fato já mostrado na seção 2.2, onde uma vez que boa parte do público alvo possui um menor grau de senso crítico, os mesmos estão menos propensos a diferenciarem uma notícia real de uma sátira ou paródia.

Muitos portais, hoje em dia, tem seu conteúdo centrado em sátiras, como por exemplo o Sensacionalista (2019) e o G17 (2019), que são dois dos principais sites brasileiros com esse enfoque. Apesar de portais como esses serem bastante claros e explícitos sobre a natureza de seus conteúdos, muitos de seus posts (publicações) são compartilhados por usuários das mídias sociais como forma de propagar as *fake news*. Como exemplo recente podemos citar uma suposta charge que estamparia a nova capa da revista satírica francesa Charlie Hebdo. A imagem foi compartilhada no dia 22 de Agosto de 2019, pela página "Jornalistas Livres" no Facebook, com a legenda: "acaba de sair do forno a última edição da publicação francesa". A charge em questão, que pode ser vista na Figura 2, mostra o presidente Jair Bolsonaro (PSL) vestido como Adolf Hitler, ao lado de vários patos de uniforme verde, fazendo o gesto de saudação nazista. Um texto acompanha a ilustração: "O renascimento do nazismo. O 'líder' Bolsonaro quer o Brasil acima de tudo". Por meio de ferramentas de verificação de notícias, usuários do Facebook verificaram que este material se tratava de uma *fake news*.



#Verificamos: É falso que revista Charlie Hebdo publicou capa satirizando Bolsonaro

Figura 2 – Exemplo de paródia/sátira. Fonte: (MORAES, 2019).

2.3.2 Contexto Enganoso

Se dá pelo uso enganoso de uma informação para empregá-la contra um assunto ou uma pessoa, basicamente é a má utilização da informação para moldar um problema acerca de uma situação ou de um indivíduo.

Na política esse tipo de conteúdo é frequentemente utilizado. Um exemplo a ser citado é uma informação, divulgada pelas redes sociais, afirmando que iria acontecer uma turnê entre a cantora Pablllo Vittar e o deputado federal Jean Wyllys, do partido político do PSol - Rio de Janeiro, para mostrar a “diversidade sexual infantil” nas escolas de ensino fundamental. A imagem que teria surgido no Twitter, no dia 19 de Julho de 2017, acompanha um texto do deputado em sua conta oficial da rede juntamente com uma imagem de divulgação da cantora, como pode ser visto na Figura 3.



Figura 3 – Exemplo de Contexto Enganoso. Fonte: (LOPES, 2018).

É constatada a farsa uma vez que, nas contas oficiais do deputado há apenas uma publicação feita no dia 19 de Julho do mesmo ano, e em sua página oficial do Facebook não há menção alguma a respeito da suposta turnê.

2.3.3 Conteúdo Impostor

Ocorre quando fontes verdadeiras têm seus nomes usados porém incorporando dados falsos em seu escopo. Um exemplo foi uma suposta declaração da ex-presidente do Brasil, Dilma Russeff, sobre uma greve de fome em apoio pela liberdade do também ex-presidente Luís Inácio Lula da Silva.

"Gostaria de agradecer a todos os que estão empenhados fazendo greve de fome e tendo que se alimentar escondidos", diz o texto que circula nas redes.



Foto usada em montagem é verdadeira, mas legenda é falsa — Foto: G1

Figura 4 – Exemplo de Conteúdo Impostor. Fonte: (G1, 2018).

A imagem vista na Figura 4, foi registrada por um fotógrafo do jornal O Globo, em Agosto de 2019, durante a comitiva de políticos apoiadores da candidatura de Lula. No discurso de Dilma, o qual foi publicado inteiramente em sua página oficial do Facebook, não houve qualquer menção aos apoiadores de Lula que realizavam uma greve de fome por sua liberdade.

2.3.4 Contexto Fabricado

Nesta classificação o conteúdo é 100% falso, projetado para enganar, desinformar o público e fazer mal. Um caso a ser lembrado sobre esse tema é o da ex-vereadora Marielle Franco, assassinada a tiros no Rio de Janeiro. Algumas horas após o crime, diversas informações falsas estavam disponíveis na internet e se espalharam rapidamente.

Como pode ser observado na Figura 5, a desembargadora Marília Castro Neves do Tribunal de Justiça do Rio de Janeiro (TJ-RJ), afirmou em sua conta oficial do Facebook que a vereadora Marielle Franco, assassinada em Março de 2018, estava envolvida com



Figura 5 – Exemplo de Contexto Fabricado. Fonte: (VEJA, 2018).

bandidos. Ao fazer o post, a desembargadora se baseou sobre as *fake news* que estavam circulando pela web, afirmando que Marielle possuía relações com a facção criminosa do Comando Vermelho.

2.3.5 Falsa Conexão

Caracterizada quando manchetes, imagens ou legendas concedem falsas dicas do que é o conteúdo realmente. São notícias reais, porém apresentam edições e manipulações.

Exemplo disso foi a notícia falsa/manipulada do Jornal O Globo, vista na Figura 6, em que a manchete e a foto nada tinham a ver uma com a outra.



Figura 6 – Exemplo de Falsa Conexão. Fonte: (SOUZA, 2017).

2.3.6 Falso Contexto

Consiste em um conteúdo genuíno compartilhado com um contexto falso, ou seja, quando o verdadeiro conteúdo é compartilhado com informações falsas contextuais.



Figura 7 – Exemplo de Falso Contexto. Fonte: (PAIVA, 2018).

A imagem vista na Figura 7 retrata o incêndio do Reichstag, o prédio onde funciona o parlamento alemão. Um mês após tomar a posição de chanceler da Alemanha, Adolf Hitler viu tal situação como a desculpa perfeita para começar o enrijecimento de seu governo e transformá-lo em uma ditadura. Sua perseguição árdua ao Partido Comunista, a partir de tal informação manipulada a respeito do incêndio no Reichstag, marcou-se como passo fundador do que viria a ser sua ditadura Nazista.

2.3.7 Conteúdo Manipulado

Ocorre quando uma informação genuína ou imagens são manipuladas para enganar o público, como fotos adulteradas.

Exemplo: A ex-presidente Dilma Rousseff, de fato, integrou grupos de combate ao Regime Militar (1964-1985) sendo por isso presa e torturada. No entanto, a foto, vista na Figura 8, amplamente divulgada no Facebook e no Whatsapp mostra a então fundadora do Partido Democrático Trabalhista (PDT) nos anos 80 (já após a anistia) em posse de uma arma. Porém a imagem do armamento é, na verdade, uma montagem.



Figura 8 – Exemplo de Conteúdo Manipulado. Fonte: (MARES, 2019).

2.4 Disseminação

Como já dito anteriormente na seção 2.2, as *fake news* são disseminadas fortemente no meio digital, principalmente através de redes sociais como Facebook, Twitter e Whatsapp; através de jornalistas, seja por viés político dos mesmos, ou contratação/suborno desses profissionais da informação por entidades de grande poderio e influência (políticos, empresários, líderes religiosos, etc.); e por último e em ampla escala, existe a disseminação por *bots*.

Os *bots* são robôs constituídos por algoritmos com a finalidade de automatizar um dado processo. Tais robôs podem ser utilizados tanto para benefícios, como atendimentos virtuais em chats, tanto quanto para malefícios, como por exemplo para compartilhamento automático de notícias propagandiosas e/ou fictícias. Esses robôs são baseados em aprendizado de máquina, de modo que são treinados para se passarem por usuários reais. Uma vez que são mecanismos automáticos, a força de propagação dos bots, em relação a disseminar uma *fake news*, é imensuravelmente maior que força de um usuário comum, levando essas notícias a uma ampla gama de leitores.

Esse método da utilização de *bots* é empregado essencialmente no Twitter, pois a plataforma não oferece nenhum método preventivo contra a criação de contas falsas. Além disso, existem ainda perfis de usuários reais que se aliam com a automação dos *bots*, ficando conhecidos como *cyborgs*. Tal método dificulta e muito a descoberta de perfis falsos, pois os ditos *cyborgs* quebram a previsibilidade caracterizada pelas contas de *bots*, fugindo dos padrões maquinários. Como as postagens das notícias, por esses tipos de conta, contém intervenção humana, tais perfis falsos passam maior credibilidade daquela informação, ampliando o espectro de alcance do público desejado e aumento da complexidade em ser classificado como um perfil fake.

Outro fator que implica fortemente no compartilhamento desse conteúdo inverídico é o seu público atingido. Uma pesquisa americana constatou que, pessoas mais velhas, de antes da era digital, possuem maiores dificuldades em fazer a distinção entre *fake news* e notícias reais. O estudo realizado nas universidades de Princeton e Nova York, constatou que pessoas com mais de 65 anos de idade compartilharam *sete vezes* mais notícias falsas do que entre os jovens (18 a 29 anos). A barreira do aprendizado de novas tecnologias somada com a idade avançada, pode ser observado como um dos fatores para idosos compartilharem um maior número de conteúdo falsificado/enganoso. Como no mundo digital somos bombardeados constantemente com notícias, isso se torna mais um obstáculo para as pessoas de idade avançada, ou seja, as *fake news* se fazem mais aceitas por esse público que passa a compartilhar o conteúdo falso em suas redes sociais, como Facebook e Whatsapp (GUESS; NAGLER; TUCKER, 2019).

2.5 Consequências das *Fake News* e seus impactos na sociedade

Em um primeiro momento pode parecer que propagar notícias distorcidas não tem grandes impactos, mas não pode-se ter uma noção tão leviana. Uma vez que os objetivos das *fake news* são de enganar, manchar/denegrir a imagem de alguém, a fim de se obter algum tipo de ganho (financeiro ou político). O ato de divulgar tais informações falsas, vídeos ou fotos manipuladas, além de publicações duvidosas, os temas noticiosos podem trazer grandes riscos a saúde e segurança pública, onde há casos que resultaram até em mortes.

Adentrando em exemplificações, no ano de 2014 o caso da dona de casa, conhecida como Fabiane Maria de Jesus, que foi linchada por dezenas de moradores da cidade do Guarujá (São Paulo), é uma clara amostra das consequências de uma *fake news* dentro da sociedade. A investida agressiva dos moradores se deu em virtude de uma informação falsa publicada em uma página do Facebook, contendo ainda um retrato falado que por engano era uma foto da vítima, relatando que a mesma praticava o sequestro de crianças para uso em rituais de magia negra. A mulher veio a óbito alguns dias após o ocorrido e segundo a polícia não havia nenhum caso de sequestro de crianças na cidade (ROSSI, 2014).

Nos últimos anos um movimento preocupante que anda ganhando força, devido principalmente a divulgação de informações distorcidas, é o movimento antivacina. O medo das vacinas se espalhou, principalmente nas redes sociais, por causa de uma interpretação errônea a respeito de um médico - que nunca foi partidário do movimento em questão. Tal interpretação se deu quando ele forjou um trabalho científico, a fim de relacionar a vacina triplice viral MMR, que age em combate contra sarampo, rubéola e caxumba, com o autismo. Tal ato, de falsificar sua pesquisa a respeito de uma vacina, ocorreu no ano de 1998, cujo protagonista foi o médico britânico Andrew Wakefield. Embora seu estudo tenha sido publicado em uma revista de renome, anos depois foi descoberto a fraude com todos os dados falsificados. É importante frisar que Wakefield nunca se mostrou contra imunizantes, ele queria apenas enriquecer em cima de sua suposta descoberta e emplacar sua solução como arma definitiva contra o sarampo (PREVIDELLI, 2019).

Mesmo com as conclusões sendo classificadas como adulteração, e suas pesquisas sendo retiradas do periódico, seus estudos conquistaram inúmeros seguidores ao redor do globo, principalmente nos EUA, onde se observou o início de um movimento antivacinação nunca antes visto. Devido a tal compartilhamento destas conclusões falsas, hoje milhares de pessoas estão convencidas de que vacinas, como um todo – e não somente a MMR – são a causa do autismo. Tais dados se mostram preocupantes, pois uma vez que o número de crianças não vacinadas está crescendo, doenças antigas, quase erradicadas, estão reemergindo, se tornando um grave problema de saúde na sociedade.

2.6 Como combater as *Fake News*

De acordo com o que foi elucidado na seção 2.2, os autores das *fake news* adotam algumas medidas de segurança, tanto físicas como virtuais, que acabam dificultando o processo de rastreamento para as fontes reais daquela notícia fabricada.

Porém, ainda assim, existem alguns métodos que podem ajudar para realizar uma primeira identificação e combater a propagação de tais dados corrompidos. Atualmente sabe-se que a maior parte das informações são divulgadas e encontradas na internet, assim sendo, uma forma de combate as *fake news* é procurar a informação de uma determinada notícia em outros sites ou procurar dados sobre o autor daquela publicação, a fim de verificar sua procedência.

Tendo em vista o advento das mídias sociais, é indispensável que os usuários das mesmas, ou de outros portais de informações, tenham ou aprendam a exercer certas condutas. Um desses comportamentos é a checagem da fonte de uma dada notícia. Como uma mesma notícia pode ser distribuída por inúmeros sites, blogs, portais e entre outros, a credibilidade daquela informação vai depender diretamente do seu autor de origem, ou seja, uma notícia registrada por uma fonte confiável e que se tenha conhecimento das suas precedências é bem mais provável publicar uma notícia verdadeira se comparado a um site que tem uma fama de publicar informações sensacionalistas.

Averiguar a fonte da notícia, antes de compartilhar a mesma, é um dos principais métodos de combate as *fake news*, tendo em vista que sua propagação se dá muitas vezes sem que as pessoas se preocupem de onde surgiu determinada informação. Procurar dados sobre o autor, checar credenciais, procurar por notícias antigas daquele site em questão, que realizou a publicação, averiguando se houveram notícias sensacionalistas antes, ou até mesmo realizar uma nova pesquisa na internet com o mesmo título ou tema daquela informação, são algumas maneiras básicas que o usuário tem em mãos para evitar de espalhar informações inverídicas.

Outra questão a ser levantada é a respeito de notícias traduzidas para o idioma local. Não é fora do comum existirem más traduções em uma informação de língua estrangeira. Visando este fato, pessoas mal intencionadas desenvolvem *fake news* baseadas em conteúdos de outros países, quer seja confiando que os leitores não irão checar a veracidade ou construindo um dos tipos de *fake news*, a Falsa Conexão, retratada na seção 2.3. Neste último caso, é comum que alguns autores desenvolvam informações associando uma imagem de uma manchete estrangeira, com um título e escopo que não dizem respeito a imagem. Para evitar o compartilhamento desse tipo de *fake news* é importante que o usuário leia a manchete por completo, não se atendo somente ao título ou imagem da notícia. É importante frisar que, se durante a leitura se perceber erros de escrita ou concordância, é um possível alerta para uma notícia falsa. Além disso, é muito comum

que algumas *fake news* sejam desenvolvidas sobre assuntos antigos, notícias com anos de publicação, então é necessário que a pessoa se atente a esta questão também.

Existem também plataformas online que podem ajudar na identificação de uma notícia compartilhada nas redes. Essa prática é conhecida como checagem de fatos, ou por seu termo em inglês *Fact Checking*. Este método consiste na verificação de confiabilidade das fontes apuradas a respeito de uma informação. Hoje, são inúmeros os sites, iniciativas ou agências que fornecem a checagem compartilhada das notícias, propiciando assim meios para que jornalistas ou usuários comuns verifiquem a veracidade da informação antes de compartilhá-la (DINIZ, 2017). Algumas destas plataformas são: a Lupa (2019), que se hospeda nos domínios da Folha de S.Paulo, Revista Piauí e portal UOL, seus serviços em 2018 se voltaram preferencialmente as questões eleitorais da época, mas seu monitoramento também se faz presente em questões da economia, cidade, educação e entre outros; o Snopes (2019), existente desde o ano de 1994, iniciou com foco em lendas urbanas e se tornou, com o tempo, um dos maiores sites de checagem de fatos da internet; o Full-Fact (2019), sendo uma instituição de caridade sediada em Londres, que busca verificar e corrigir notícias relacionadas à crimes, economia, direito, saúde e entre outros.

Abordando um aspecto mais técnico, existem ferramentas com algoritmos baseadas em mineração de dados, inteligência artificial, aprendizado de máquina, mineração de texto e entre outros, que podem ser voltadas e treinadas especificamente para a detecção automática das *fake news*. Como será demonstrado melhor, no próximo capítulo, diante dos referenciados trabalhos de Marumo (2018) e Leal (2018), é possível estruturar um processo automático com a finalidade de classificar textos reais e *fake news* fazendo uso da mineração de texto.

No capítulo seguinte serão abordados iniciativas que utilizam a técnica da mineração de texto direcionadas à detecção de notícias caracterizadas como *fake news*.

3 Mineração de Texto e *Fake News*

Neste capítulo será exposto o que é, e como se caracteriza a mineração de texto, também serão apresentados e discutidos dois trabalhos, sendo eles Marumo (2018) e Leal (2018), onde o primeiro cria uma ferramenta capaz de classificar um texto em notícia falsa ou não, utilizando abordagens de Deep Learning; o segundo trabalho realiza a validação do aprendizado de máquina, utilizando 04 (quatro) algoritmos distintos, para identificar as *fake news* no âmbito das eleições presidenciais do Brasil em 2018.

Ambos os estudos são utilizados como amostras para exemplificar o uso de ferramentas da mineração de texto, com o intuito de realizar a detecção e classificação automatizada do atual fenômeno das *fake news*.

3.1 Mineração de Texto

Com o grande avanço da internet, e conseqüentemente das mídias sociais, o volume do armazenamento de dados tem crescido exponencialmente. Dentro dessa enorme quantidade de dados textuais, é possível encontrar diversas informações sobre os usuários, como as suas opiniões sobre um dado tema, por exemplo. Em vista disso as empresas começaram a procurar e desenvolver ferramentas que fossem capazes de interpretar tais dados, visando abstrair informações importantes a respeito de seus produtos ou serviços. É justamente neste campo que surgiu a Mineração de Dados ou *Data Mining*.

Por ser considerada uma área multidisciplinar, as definições do termo mineração de dados variam de acordo com o campo de atuação em que seus autores desejam aplicá-las. Em destaque existem três áreas consideradas as de maior expressão dentro da Mineração de Dados:

- *Estatística*
- *Aprendizado de Máquina*
- *Banco de Dados*

Apesar da existência de diversas ferramentas que auxiliam na execução dos algoritmos de mineração, os resultados ainda necessitam de uma análise humana. Embora isso, o método da mineração de dados contribui de forma significativa no processo de descoberta do conhecimento, permitindo aos especialistas concentrarem seus esforços apenas nas partes mais significativa dos dados.

Essas novas ferramentas de monitoramento influenciaram fortemente em uma nova abordagem de acompanhamento do público na coleta de informação. Durante as pesquisas de estudo para este trabalho, foi observado que os consumidores buscam, cada vez mais, aconselhamentos online, disseminando suas opiniões em blogs, fóruns e outros meios sociais na internet, exercendo assim uma influência direta na força de consumo. Dados como estes são muito relevantes para estratégias de marketing, aumento da competitividade empresarial e aplicabilidades políticas.

Observando esse grande e complexo volume de dados textuais, foi necessário a criação de técnicas para efetuar sua interpretação, e é dentro deste âmbito que veio a surgir a **Mineração de Texto**, sendo esta um conjunto de processos focados no processamento da linguagem natural.

Mineração de textos é um conjunto de métodos usados para navegar, organizar, achar e descobrir informação em bases textuais. Pode ser vista como uma extensão da área de Data Mining, focada na análise de textos. Também é chamada de Text Data Mining, Knowledge Discovery in Texts (ARANHA; PASSOS, 2006).

Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996) a transformação de dados digitais em informações, consiste em um processo manual de todas essas informações por especialistas, que produzirão relatórios afim de analisá-los. Com o advento da internet e seu elevado volume de dados, tal processo manual se torna impraticável. Assim sendo, a Descoberta de Conhecimento nas Bases de Dados ou *Knowledge Discovery in Databases* (KDD) é uma tentativa de resolver esse problema causado pela chamada "era da informação".

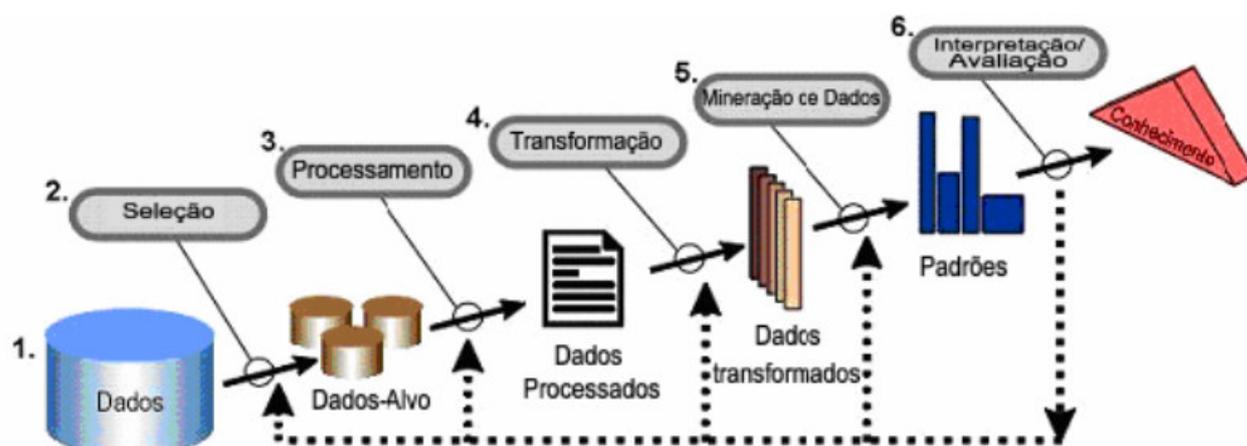


Figura 9 – Processo da Mineração de texto. Fonte: (BARION; LAGO, 2015).

Segundo a autora, o processo de conhecimento/mineração deve ser realizado de maneira iterativa, interativa e dividida em fases Fayyad, Piatetsky-Shapiro e Smyth (1996). Tal representação do KDD pode ser analisada com a Figura 9. A seguir é descrito cada uma dessas etapas deste processo:

- **Dados:** Escolha das informações, ou domínios, que serão adquiridos e analisados para atingir o objetivo do trabalho.
- **Coleta/Seleção de Dados:** Aqui é realizada a obtenção dos dados/textos que serão trabalhados, eles podem ser obtidos de redes sociais, livros, artigos, revistas e dentre outros meios. A coleta de dados pode ser uma etapa trabalhosa e custosa.
- **Pré-processamento:** O objetivo central é transformar e organizar os dados não-estruturados, obtidos na etapa anterior, de preferência utilizando uma relação atributo-valor.
- **Transformação:** Os dados são organizados em grupos, afim de facilitar e ordenar o acesso a informação colhida.
- **Mineração:** Análise dos objetivos da pesquisa para utilização de técnicas específicas da mineração de dados, com intuito de extrair as informações desejadas dos dados coletados. Por exemplo, podem ser aplicadas técnicas como a sumarização, classificação, regressão, clusterização e dentre outros.
- **Interpretação/Avaliação:** Processo de inferência humana para analisar os resultados obtidos até aqui. Documentação e/ou checagem das soluções encontradas através dos processos até o momento, avaliando se está balanceado com as expectativas.

3.1.1 Tarefas da Mineração de Texto

Essencialmente a mineração textual se dá em dois principais nichos: Texto Livre e Texto semiestruturado.

- **Texto Livre:** diz respeito basicamente a interpretação da linguagem natural, da linguagem corrente que podemos encontrar em revistas, artigos, livros, pdfs, e-mails e etc.
- **Texto semiestruturado:** são dados textuais demarcados, que possuem uma estrutura determinada para agrupar e especificar as informações daquela informação, podendo ser exemplificados por tags, arquivos XML, arquivos JSON e entre outros

Além dos nichos especificados acima a mineração textual pode ainda ser classificada pela sua capacidade em realizar determinadas tarefas. Dentre as mais comuns temos:

1. **Classificação:** caracterizada como uma das tarefas mais comuns, a classificação visa a análise das características de um objeto para atribuí-lo à uma classe. A função desta tarefa é fazer a rotulação dos dados, incorporar um determinado aspecto e por fim

atribuir um rótulo a estes dados. Tais classificações são representadas, normalmente, por registros em tabelas de um banco de dados ou em arquivos.

2. **Estimação:** trabalha comumente com dados numéricos, dado um valor de entrada a estimativa atribui um valor para um dado desconhecido. Exemplos: renda, altura, saldo do cartão.
3. **Predição:** bem similar aos dois anteriores, o que difere em sua composição são seus registros que são classificados visando descobrir e/ou estimar um valor futuro para um atributo. Exemplos são algoritmos que trabalham com dados da bolsa de valores, índice de evasão ou previsão do campeão de um jogo.
4. **Agrupamento:** associar dois ou mais dados em determinados grupos com características semelhantes.
5. **Associação:** responsável pela extração de padrões. Sua principal utilização pode ser observada em ferramentas de comércio para análise de itens adquiridos em uma compra, representando então a probabilidade de um item existir em um conjunto ou transação, dado que outro também esteja presente.

3.2 Estudos de Caso: Mineração de Texto e *fake news*

Visto a apresentação das definições a respeito de mineração de texto na seção anterior (3.1) e as características das *fake news* no capítulo 2, serão abordados, então, nesta seção, dois estudos que ilustram a utilização da mineração de texto para a classificação de *fake news*. A primeira pesquisa utiliza diferentes abordagens de *Deep Learning*, fazendo uso do Long Short-Term Memory (LSTM) e do Support Vector Machine (SVM) para a classificação das notícias através do processo de sumarização de texto. O segundo estudo expõe uma proposta similar, porém o autor faz uso de 04 (quatro) algoritmos - sendo eles o *Naive Bayes*, J48, JRip e o iBK - para validar o processo do aprendizado de máquina na identificação e classificação das *fake news*, sendo estas provenientes do Twitter, no âmbito das eleições para presidente do Brasil no ano de 2018.

3.2.1 Classificação de *fake news* com o uso de *Deep Learning*

Neste trabalho, o autor aponta como objetivo central de sua pesquisa realizar a diferença entre notícias reais e *fake news* através da estrutura textual da informação. Para este processo, ele organizou o domínio de seu estudo em um aglomerado estruturado de notícias reais e falsas, abstraído de domínios públicos, como será demonstrando no decorrer desta seção. Além disso, foi empregado o aprendizado de máquina juntamente com o processo do *Deep Learning* - que segundo Deng e Yu (2014) é um campo dentro

do aprendizado de máquina, que utiliza grupos de algoritmos visando aprender diferentes níveis de abstração e empregando várias camadas de processamento - para realizar a classificação das *fake news*.

Conforme a pesquisa conduzida por Marumo (2018), as *fake news* utilizam em seus títulos palavras-chave populares, com a intenção de serem chamativas ou apelativas, para assim alcançarem facilmente um público maior. Para analisar os padrões dessas notícias, o autor utilizou duas técnicas de pré-processamento, sendo elas a *sumarização* e o *word2vec*. Onde, dentre estas técnicas, a primeira consiste, basicamente, em resumir um evento, ou seja, mostrar o fato/informação de forma direta, sem aprofundar-se em detalhes (MARTINS et al., 2001); já a segunda trata de um grupo de modelos que se caracterizam por utilizar Redes Neurais Artificiais (RNA) com o intuito de associar um conjunto de palavras (RONG, 2014) .

Temos então que em o objetivo principal do autor é o desenvolvimento de um classificador de *fake news* baseado em notícias da língua portuguesa. Uma vez que a maior parte da literatura de mineração textual tem seu domínio na língua inglesa, onde muitas vezes para estudos, no Brasil por exemplo, é necessário realizar a tradução de tal domínio, ou efetuar sua criação do zero, o estudo de Marumo (2018) se mostra interessante, pois encontrar trabalhos com bases de dados que atuem no idioma brasileiro é um tanto raro.

As amostras da base de dados desenvolvida pelo autor se dividem em dois tipos, notícias verdadeiras e *fake news*. Para o primeiro tipo de registro, a coleta dos dados se estabeleceu no período de 2017 a 2018 dentro do portal de notícias do G1 ¹ e no site da Agência de notícias da Universidade Estadual de Londrina (UEL) ². Referente ao segundo tipo de amostra, para a coleta de informações caracterizadas como falsas, foram utilizados domínios públicos, tais como o Diário de Pernambuco ³ e o site Sensacionalista ⁴, onde dentre essas informações, existiam algumas notícias datadas com mais de 10 anos desde sua data de publicação oficial, reforçando o que foi elucidado na seção 2.1, de que o conceito das *fake news* não é algo recente dentro da sociedade.

É importante frisar duas questões a respeito da base de dados deste estudo. A primeira, é que as amostras são o corpo textual por completo das notícias, ou seja, são as informações (falsas ou verídicas) em si, como ilustrado nas Figuras 10 e 11. Em segundo vale ressaltar que as temáticas das notícias coletadas são diversificadas, ou seja, abordam conteúdos como saúde, política, justiça, economia e entre outros, todas provenientes de domínios públicos do Brasil. O autor tomou esta decisão com o intuito de obter a maior proximidade possível com um cenário real, contendo informações diárias vistas pela internet.

¹ <https://g1.globo.com/>

² <http://www.uel.br/com/agenciaueldenoticias/>

³ <https://www.diariodepernambuco.com.br/>

⁴ <https://www.sensacionalista.com.br/>

Tipo de Notícia	FN
Conteúdo	A prefeitura do Rio de Janeiro anunciou hoje que as passagens na cidade passarão a custar R4., <i>Comisso, oscariocaspassarãoa</i> ser oficialmente assaltados. O prefeito Marcelo Crivella explicou que como muitos cidadãos saem de casa com medo de ser assaltados, a nova medida vai trazer paz. “Já será certo que assim que você entrar no ônibus você já vai perder a carteira”, disse um assessor.,O novo preço despertou esperança nos passageiros. “Por R4talvezagentetenhacafezinhoechurroscomNutella], disse um deles.

Figura 10 – Exemplo de *fake news*. Fonte: (MARUMO, 2018).

Tipo de Notícia	Notícia real
Conteúdo	O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), em parceria com o Instituto Euvaldo Lodi (IEL), e com o apoio do Fórum Nacional de Gestores de Inovação e Transferência de Tecnologia (FORTEC), divulgam a Chamada Pública para o Programa INOVATEC. A Chamada é fruto de acordo da parceria firmado entre o IEL e o CNPq, cujo objetivo é a implantação do Programa INOVATEC. Tal Programa pretende fomentar a participação de estudantes de graduação em projetos de pesquisa, desenvolvimento e inovação PD&I de interesse do setor empresarial, em parceria com instituições de ensino superior e empresas. O Programa concederá até 200 bolsas de Iniciação Tecnológica e Industrial (ITI) do CNPq para estudantes regularmente matriculados em curso superior ou superior tecnológico, além de auxílios à pesquisa aos professores (coordenadores de projeto). As empresas serão responsáveis pelo custeio integral dos auxílios aos projetos, no valor de R\$ 4.800, para despesas de custeio e capital. As propostas podem ser submetidas até 15 de dezembro deste ano, por meio do endereço http://www.portaldaindustria.com.br/iel/canais/inovatec/ . (Com informações da Coordenação de Comunicação Social do CNPq)

Figura 11 – Exemplo de notícia verdadeira. Fonte: (MARUMO, 2018).

Com os dados textuais obtidos, Marumo (2018), levantou algumas estatísticas interessantes, nas quais é possível observar que as notícias reais possuem um número maior de palavras e sentenças se comparadas as *fake news*. A Tabela 1 ilustra exatamente tais dados levantados por ele.

	Fake News	Real	Média
Média de palavras por notícia	184	288	236
Média de sentenças por notícia	9.74	14.64	12.19

Tabela 1 – Estatísticas sobre notícias reais e *fake news*. Fonte: (MARUMO, 2018).

Segundo o autor para realizar a fase do pré-processamento das informações coletadas foram executadas as seguintes etapas:

1. Primeiro houve a troca de letras maiúsculas por minúsculas, através da biblioteca de string da linguagem de programação Python;
2. Em seguida ocorreu a remoção de palavras vazias (*stop words*) pela biblioteca do *Natural Language Toolkit* (NLTK), e de dígitos numéricos através do framework

Gensim, que consiste em uma biblioteca de código-fonte aberto para a modelagem de abordagens não-supervisionadas e o processamento em linguagem natural. O diferencial deste framework é sua capacidade de lidar com grandes fluxos de dados textuais, utilizando algoritmos online de maneira incremental e não somente visando o processamento em memória como outros programas de aprendizagem de máquina (ŘEHŘEK; SOJKA, 2011).

3. Logo após, foi aplicado o processo da sumarização, que consiste na criação de um título, ou resumo, que capta as ideias principais de um texto, verificando também o contexto, ao invés de limitar-se somente a análise de certas frases (NALLAPATI et al., 2016).
4. Na sequência houve a tarefa de *stemming*, que consiste na redução da palavra ao seu radical, facilitando a automatização e o treinamento do algoritmo proposto pelo autor. Também foi realizada a troca de espaços múltiplos por espaçamento único, além da remoção de palavras com menos de três caracteres utilizando o framework *Gensim* novamente.

O resultado após todo este processo de limpeza é a geração de dois tipos de amostras, sendo elas, textos sumarizados e textos pré-processados. Este último é obtido seguindo todos os passos descritos acima, omitindo somente a parte de sumarização, onde nesta etapa, a ferramenta *Gensim*, não consegue processar textos de apenas uma sentença, nesses casos, foram mantidos os textos originais como textos sumarizados, visando deixar a maior quantidade de amostras. Um exemplo de sumarização aplicada no domínio do autor, pode ser visto na Figura 12, tendo como referência a Figura 11.

Tipo de Notícia	Notícia real
Conteúdo	conselho nacional desenvolvimento científico tecnológico cnpq parceria instituto euvaldo lodi iel apoio fórum nacional gestores inovação transferência tecnologia fortex divulgam chamada pública programa inovatec

Figura 12 – Exemplo de notícia real sumarizada. Fonte: (MARUMO, 2018).

Cada um desses dois tipos de amostras, são utilizadas para a aplicação do algoritmo *Word2Vec* do framework *Gensim*, gerando assim o seu próprio vocabulário. Os parâmetros escolhidos para o *Word2Vec* foram a utilização do modelo *Skip Gram* - que segundo Rong (2014) possui uma funcionalidade melhor para uma pequena quantidade de dados -, dimensionalidade dos vetores que representam as palavras com tamanho 10 (dez) e a escolha de não utilizar palavras com frequência menor que 2 (dois). Nos parâmetros que restaram foram atribuídos valores padrões.

Segundo a descrição de Marumo (2018), na fase da utilização do Aprendizado de máquina, foram seguidos os próximos passos:

1. Balanceamento das amostras, nivelando uma quantidade igual para cada tipo de dado, sendo então uma notícia real para uma *fake news*. Para tal, foi utilizada a técnica de sub-amostragem (*undersampling*) aleatória, que consiste na retirada aleatória de amostras da classe de maior quantidade. Assim o número de notícias para cada classe foi modificado para 286 (duzentos e oitenta e seis).
2. Uso da validação cruzada utilizando *holdout*. As amostras são separadas em dois subconjuntos, sendo um para treinamento (modelagem do problema) e o outro voltado para testes, que serão classificados pelo modelo gerado anteriormente. Deste modo, Marumo (2018) decidiu utilizar 80% da base para treino e os 20% restantes para fase de testes.
3. Definição da arquitetura de *Deep Learning* (DL), utilizando as camadas da rede neural.
4. Definição dos valores dos hiperparâmetros que serão utilizados nos algoritmos.
5. Aplicação dos algoritmos de classificação de IA tradicional (Random Forest [RF], SVM) e DL (composto por RNN e LSTM), utilizando as amostras de treinamento para a criação dos modelos de predição, tomando como parâmetro vinte épocas.
6. Classificação das amostras de teste, utilizando o modelo gerado.

Para a avaliação dos resultados, Marumo (2018) estabeleceu a utilização de três métricas: acurácia, precisão e revocação (*recall*). A primeira visa medir a frequência de acertos; a segunda se responsabiliza por medir o grau de acerto dos conjuntos preditos anteriormente; e por último, a terceira, calcula a frequência de acerto quando a classe do problema é associada corretamente, ou seja, grau de frequência da classificação de uma *fake news* quando o texto é de fato uma *fake news*.

As métricas utilizadas pelo autor se compõem em quatro tipos:

- Verdadeiro positivo: Análise e texto são *fake news*. Representam as notícias dadas como falsas.
- Verdadeiro negativo: Análise e texto não são *fake news*. Representam então as notícias dadas como reais.
- Falso positivo: A análise não é *fake news* e o texto é uma *fake news*.
- Falso negativo: A análise é *fake news* e o texto não é *fake news*.

Houveram casos onde a sumarização não gerou sumários, mas como pode ser visto na Tabela 2, das amostras de sumários criados, 65% delas foram geradas com sucesso para a categoria das *Fake News* e 82% para Notícias Reais (MARUMO, 2018).

	Fake News	Real	Total
Textos Pré-processados	286	513	799
Textos Sumarizados	187	422	609
Sumarização com sucesso	0.65	0.82	0.76

Tabela 2 – Quantidade de amostras do domínio. Fonte: (MARUMO, 2018).

Adentrando nos resultados é possível observar através das Tabelas 3 e 4 que o melhor desempenho foi obtido pelo algoritmo LSTM em textos pré-processados, obtendo uma acurácia de 79.3% seguida pelo RF com 78.2%. Devido a sua característica de possuir um controle de longa memória, o algoritmo LSTM mostra lidar melhor quando o texto possui uma quantidade de dados mais significativa, como ocorre em textos pré-processados, alcançando resultados com um maior equilíbrio na classificação entre *fake news* e informações reais.

	LSTM		RNN	
	Pré-processado	Sumarizado	Pré-processado	Sumarizado
Verdadeiro Positivo (%)	77	59	24	39
Verdadeiro Negativo (%)	81	52	79	55
Falso Positivo (%)	19	48	21	61
Falso Negativo (%)	23	41	74	45
Acurácia (%)	79.3	55.9	52.5	47.3
Precisão	0.805	0.555	0.555	0.468
Revocação	0.772	0.5294	0.258	0.394

Tabela 3 – Resultados da classificação com *Deep Learning*. Fonte: (MARUMO, 2018).

	RF		SVM	
	Pré-processado	Sumarizado	Pré-processado	Sumarizado
Verdadeiro Positivo (%)	74	56	69	67
Verdadeiro Negativo (%)	82	69	66	48
Falso Positivo (%)	18	31	34	52
Falso Negativo (%)	26	44	31	33
Acurácia (%)	78.2	62.8	67.2	57.5
Precisão	0.807	0.648	0.667	0.562
Revocação	0.743	0.563	0.689	0.671

Tabela 4 – Resultados da classificação com IA Tradicional. Fonte: (MARUMO, 2018).

A respeito dos resultados utilizando os algoritmos de inteligência artificial (IA) tradicional, as duas maiores acurácias ocorreram em textos pré-processados, mostrando

que tanto o algoritmo RF quanto o SVM, tiveram melhores desempenhos analisando uma maior quantidade de dados.

Comparando as duas tabelas anteriores (3 e 4), é possível constatar que o algoritmo *Random Forest* (RF) foi o único que mostrou resultados próximos ao do *Long Short Term Memory* (LSTM) em textos pré-processados, porém no âmbito dos textos sumarizados, é notório que o RF obteve a maior acurácia dos resultados, contando com 62.8% seguido do SVM com 56.2%, deixando então o LSTM em terceiro lugar com 55.9% de acurácia.

Em relação aos resultados de classificação com o algoritmo do RNN, a maior taxa de predições corretas é identificada na classe de notícias reais, possuindo uma precisão de 79% em textos pré-processados e 55% em sumarizados, no qual ainda nesta última, é possível perceber que a detecção teve um aumento de 15% na classificação correta de *fake news* (MARUMO, 2018).

3.2.2 Aprendizagem de máquina na classificação de *fake news* obtidas do *Twitter*

Observando o contexto das eleições presidenciais do Brasil no ano de 2018, Leal (2018) propôs em sua pesquisa o desenvolvimento de um protótipo apto a classificar e identificar *fake news* dentro deste âmbito político. Para isto, o autor utilizou técnicas de aprendizagem de máquina, mineração de texto e mineração de dados, onde a coleta dessas informações foi realizada na rede social do Twitter, para obter notícias referentes aos candidatos da corrida presidencial. Além disso, ele utilizou quatro algoritmos para efetuar a comparação estatística a respeito da classificação das *fake news*, sendo eles o *Naive Bayes*, J48, *JRip* e *iBK*. Para a programação do seu sistema, foram aplicadas técnicas de aprendizado de máquina supervisionado, mineração de texto e mineração de dados, sendo que os algoritmos empregados no processo de classificação das notícias inverídicas o autor fez uso do *Naive Bayes*, *RNA*, *Árvore de Decisão*, *Aprendizagem por Regras* e *Aprendizagem Baseada em Instâncias*. Objetivando sua prototipação, Leal (2018) operou com o framework WEKA (*Waikato Environment for Knowledge Analysis*), por este possuir uma coleção de algoritmos de diferentes abordagens na área da IA, já prontos para serem implementados e testados (GARNER et al., 1995).

Ao longo do processo da modelagem de seu domínio, a estrutura da tabela das informações dos *tweets*, obtidos durante a fase de coleta de dados, foi criada de acordo com a resposta de retorno da API oficial do Twitter, a qual é possível observar na Figura 13, onde as mensagens capturadas foram armazenadas em um banco de dados próprio. Devido ao tipo de conta do autor ser *Standart*, a mesma possuía restrições, tendo uma captura limitada dos *tweets* aos últimos sete dias e além disso, só era possível realizar um total de cento e oitenta requisições dentro de uma janela de tempo de quinze minutos. Durante a

Atributo	Descrição
<code>id_tweet</code>	Código sequencial único para cada <i>tweet</i> enviado
<code>message</code>	Mensagem publicada
<code>created_at</code>	Data e hora que o <i>tweet</i> foi enviado
<code>latitude</code>	Latitude do local que o usuário do Twitter estava na hora de enviar o <i>tweet</i>
<code>longitude</code>	Longitude do local que o usuário do Twitter estava na hora de enviar o <i>tweet</i>
<code>favorite_count</code>	Quantidade de vezes que outros usuários marcaram que gostou do <i>tweet</i>
<code>retweet_count</code>	Número de vezes que foi retweetado
<code>id_user</code>	Código sequencial único para cada usuário cadastrado
<code>name_user</code>	O nome do usuário, não sendo necessariamente o nome da pessoa
<code>is_retweet</code>	Campo booleano se foi um <i>retweet</i>
<code>is_retweeted</code>	Campo booleano se foi retweetado
<code>tag</code>	Campo para armazenar a chave utilizada para achar o <i>tweet</i>
<code>class</code>	Classe do tweet (fato, <i>fake</i> , opinião)

Figura 13 – Colunas da Tabela dos dados coletados do Twitter. Fonte: (LEAL, 2018).

coleta dos dados, o autor utilizou algumas *hashtags* como palavras-chave para a captura dos *tweets*, tais quais podem ser vistas na Figura 14, onde a ideia consistiu basicamente em reunir, dentro de uma *Array List*, o nome e sobrenome dos candidatos políticos ou seus nomes associados com o número de seus respectivos partidos representantes.

```
ArrayList<String> tags = new ArrayList();
Collections.addAll(tags, "jair messias bolsonaro", "fernando haddad",
    "vera lucia pereira da silva salgado", "alvaro dias", "cabo daciolo",
    "ciro gomes", "jose maria eymael", "gerald alckmin", "guilherme boulos",
    "henrique meirelles", "joão amoêdo", "joão goulart filho", "marina silva",
    "bolsonaro", "haddad", "daciolo", "ciro", "alckmin", "boulos", "meirelles",
    "amoêdo", "joão goulart", "marina", "bolsonaro17", "haddad13", "ciro12",
    "daciolo51", "alvaro19", "eymael27", "alckmin45", "boulos50", "meirelles15",
    "amoedo30", "joaogoulart54", "marina18", "vera16");
```

Figura 14 – *Hashtags* utilizadas na aquisição dos dados. Fonte: (LEAL, 2018).

Após a fase de captura de dados, realizada no período entre 26 de Setembro de 2018 a 27 de Outubro de 2018, se estabeleceu um armazenamento de 1.780.821 (um milhão setecentos e oitenta mil oitocentos e vinte e um) *tweets*, sendo que deste total, o autor verificou que aproximadamente 500.00 (quinhentos mil) *tweets* não possuíam a categoria do compartilhamento (*retweets*), parâmetro o qual ele utilizou como determinante para uma possível notícia falsa.

Para o uso desta base de dados dentro da ferramenta do WEKA, foi necessário a criação de um arquivo `.ARFF`, o qual é um tipo de texto ASCII que descreve a lista de instâncias que compartilham um conjunto de atributos. O arquivo é dividido em duas

seções, sendo a primeira um cabeçalho com o nome da relação, a lista de atributos e seus tipos. A segunda seção são os dados em suma. Na Figura 15 é possível ver um exemplo deste tipo de arquivo.

```
@relation weather
@attribute outlook {sunny, overcast, rainy}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
```

Figura 15 – Exemplo de arquivo .ARFF. Fonte: (LEAL, 2018).

De acordo com a Figura 16, o autor ilustra como implementou a geração de seu arquivo .ARFF, visando assim poder utilizar seu domínio dentro do ambiente WEKA.

```
public class GeraArquivoArff {
    public static File geraArquivo() {
        File arquivo = new File(System.getProperty("user.dir") + "\\TccIthaloLeal\\src\\fakenewsTCC.arff");
        try {
            Connection con = ColetaAPI.conexaoBD();

            PreparedStatement cmd1 = con.prepareStatement("select message, class from twitter where class is not null");
            ResultSet result = cmd1.executeQuery();

            // Cabeçalho do arquivo Weka
            String exportacao = "@relation fakenewsTCC\n\n";
            exportacao += "@attribute message string\n\n";
            exportacao += "@attribute class {opiniao, fato, fake}\n\n";
            exportacao += "@data\n\n";

            while (result.next()) {
                exportacao += "" + Normalizer.normalize(result.getString("message").replaceAll("\n", " ")).
                    replaceAll("'", ""), Normalizer.Form.NFD).replaceAll("[^\\p{ASCII}]", "") + ",'" +
                    result.getString("class") + "\n\n";
            }

            FileOutputStream f = new FileOutputStream(arquivo);
            f.write(exportacao.getBytes());
            f.close();

        } catch (SQLException | IOException e) {
            e.printStackTrace();
        }

        return arquivo;
    }
}
```

Figura 16 – Código do autor para geração do arquivo .ARFF referente a sua base de dados. Fonte: (LEAL, 2018).

Como é descrito por (LEAL, 2018), para as etapas de pré-processamento, a ferramenta do WEKA disponibiliza vários filtros, onde foram utilizados principalmente o *StringToWordVector* e o *Reorder*, que serão melhor esclarecidos a seguir.

1. StringToWordVector

Tem por objetivo converter a tipagem dos dados de string para atributos numéricos, que representam a frequência das palavras no texto (BOUCKAERT et al., 2008). Dentro deste filtro existem outros algoritmos para auxiliar na conversão dos atributos:

- **Stemmer**

Como já explicado na seção 3.2.1, esta técnica tem o objetivo de reduzir as palavras ao seu radical, eliminando a variação da mesma e reduzindo o tempo de processamento do algoritmo.

Exemplo:

A palavra corr[**endo**] após o processo de *stemming* -> **corr**

- **Stopwords**

São palavras, pronomes, artigos, pontuações muito utilizados em textos, porém não agregam tanto valor semântico, sendo úteis apenas para esclarecimento de sentença.

Exemplo: "A casa é amarela-> "[A] casa [é] amarela". O resultado após o processo seria então "casa amarela".

Sua eliminação não afeta a essência do enunciado, logo sua remoção em grandes quantidades aumenta significativamente o custo e desempenho do algoritmo que será executado.

2. Reorder

Algoritmo responsável por reorganizar os atributos de um dado, sendo então muito útil dentro de trabalhos com o WEKA, pois alguns classificadores necessitam que o atributo classe esteja na última posição e além disso, também é possível remover e duplicar atributos.

Seguindo para a próxima etapa, o autor descreve que para a implementação do protótipo da sua pesquisa, foram escolhidos os seguintes algoritmos, com base na literatura, sendo estes os mais utilizados atualmente:

1. *Naive Bayes*
2. *J48* (Árvore de Decisão)
3. *JRip* (Aprendizagem por Regras)
4. *iBK* (Aprendizagem Baseada em Instâncias)

Observando a divisão da base de dados, Leal (2018) mostra que cerca de 5.000 (cinco mil) tweets foram registrados manualmente para a etapa de treino do seu protótipo. Sua

escolha dos *tweets* foi baseada principalmente no número de compartilhamentos (*retweet*), onde durante seus estudos foi possível observar que as *fake news* são 70% mais propensas à serem compartilhadas.

Observando os passos até então, a classificação das notícias dentro do domínio de Leal (2018) é observada da seguinte maneira:

- **Fato:** registros são considerados como fato quando o *tweet* cita um candidato e o fato realmente aconteceu.

Mensagem	Classe
Marina Silva participa de ato contra Bolsonaro https://bit.ly/2QiYcRo #Eleições2018	fato

Figura 17 – Exemplo de notícia classificada como fato. Fonte: (LEAL, 2018).

Como vários órgãos da imprensa noticiaram este ato⁵ de maneira similar, a notícia foi categorizada como verdadeira.

- **Fake:** registros classificados como fake ocorrem quando a informação cita um candidato e aquele fato não ocorreu, ou o fato foi distorcido na notícia, conforme observado na Figura 18.

Mensagem	Classe
Manifestações contra @jairbolsonaro ocorrem em 66 países https://t.co/rDt6pjj7HP https://t.co/gWPoQxbzWJ	fake

Figura 18 – Exemplo de notícia classificada como fake. Fonte: (LEAL, 2018).

Apesar das manifestações realmente terem ocorrido, o título real da notícia era: "Manifestações contra Jair Bolsonaro ocorrem em 66 cidades pelo mundo"⁶, constatando então que houve uma manipulação da informação.

- **Opinião:** já os dados classificados como opinião, foram registrados que poderiam ou não, citar o nome de um candidato, mas o usuário estava expressando sua própria opinião do assunto, como pode ser visto na Figura 19.

Além dos registros manuais, também foi utilizado um *corpus* público com 3.600 (três mil e seicentas) notícias falsas e 3.600 (três mil e seicentas) notícias verdadeiras já classificadas ⁷.

⁵ <https://politica.estadao.com.br/noticias/eleicoes,ato-contra-bolsonaro-une-adversarios-em-saopaulo,70002525476>

⁶ <https://oglobo.globo.com/brasil/manifestacoes-contra-jair-bolsonaro-ocorrem-em-66-cidades-pelomundo-23113418>

⁷ <https://github.com/roneysco/Fake.br-Corpus>

Mensagem	Classe
Ato da Virada! No Mercado de São Brás! Vem pra democracia! #Haddad13 https://t.co/TMKmXQQ6At	opinião

Figura 19 – Exemplo de notícia classificada como opinião. Fonte:(LEAL, 2018).

Como resultado inicial, foram gerados modelos dos algoritmos utilizados e posteriormente, todos os registros não classificados de forma manual, foram submetidos a classificação dos modelos gerados. Assim, o autor obteve o número e percentual do total de registros classificados em cada uma das três classes: fato, fake ou opinião.

	<i>NaiveBayes</i>	J48	JRip	IBK
opinião	1.342.164	1.390.367	1.696.925	1.348.893
fato	245.825	214.722	48.698	255.551
fake	192.832	175.732	35.198	176.377

Figura 20 – Quantidade de registros classificados nas respectivas classes. Fonte: (LEAL, 2018).

Segundo a Figura 20, é possível observar a quantidade de registros classificados em cada uma das classes, onde é possível observar que os registros da classe fake são superiores através do algoritmo *NaiveBayes*. Na Figura 21 é possível acompanhar o percentual de registros classificados em cada classe, tomado a seguinte fórmula: $P = (qc/qt) \times 100$ (LEAL, 2018). Onde,

- P: Percentual
- qc: Quantidade de registros da classe
- qt: Quantidade de registros total

	<i>NaiveBayes</i>	J48	JRip	IBK
opinião	75,36771%	78,07449%	95,28891%	75,74557%
fato	13,80403%	12,05747%	2,73458%	14,35018%
fake	10,82826%	9,86803%	1,97650%	9,90425%

Figura 21 – Percentual de registros classificados em cada classe. Fonte: (LEAL, 2018).

Segundo o autor, as Figuras 20 e 21 retratam apenas as informações referentes aos registros classificados em cada classe. Então, para avaliar os resultados acerca do seu domínio como um todo, ele utilizou a técnica *K-fold Cross Validation*.

Segundo Kohavi et al. (1995), *K-fold Cross Validation* é uma técnica de reamostragem utilizada para estimar a precisão de um modelo na prática. Tal processo possui um parâmetro "K" que determina o número de grupos em que uma dada amostra deve ser dividida. Além disso, Leal (2018) realizou dez testes variando o parâmetro *Random Seed*, na qual a função deste é decidir quais partes das amostras serão destinadas para a fase de treino e teste. Em vista disto, o autor utilizou um *Random Seed* com variações de 1 a 10 objetivando assim, sempre obter partes diferentes do conjunto total de amostras da sua base de dados.

	NAIVEBAYES	J48	JRIP	IBK
Seed 1	91,20491%	91,59836%	91,68032%	73,71311%
Seed 2	91,20491%	91,72131%	91,67213%	73,72131%
Seed 3	91,21311%	91,68032%	91,89344%	73,93442%
Seed 4	91,22950%	91,56557%	91,93442%	73,58196%
Seed 5	91,21311%	91,37704%	91,86065%	73,86065%
Seed 6	91,19672%	91,34426%	91,87704%	74,06557%
Seed 7	91,21311%	91,63934%	91,71311%	73,85245%
Seed 8	91,21311%	91,45901%	91,68852%	73,50000%
Seed 9	91,22131%	91,23770%	91,89344%	73,89344%
Seed 10	91,21311%	91,85245%	91,80327%	73,96721%
MÉDIA:	91,21229%	91,54754%	91,80163%	73,80901%

Figura 22 – Percentual de acertos com o parâmetro *seed* variando de 1 a 10. Fonte: (LEAL, 2018).

O autor indica que ao todo foram utilizados 12.200 (doze mil e duzentos) registros para a fase de treinamento, onde os percentuais destacados na cor azul representam os melhores resultados de cada algoritmo e na cor laranja o pior, os quais são possíveis visualizar através da Figura 22.

Realizar avaliações quantitativas sobre classificações é feita geralmente através de uma matriz de confusão. Em vista disso, a seguir serão ilustradas as respectivas matrizes para cada algoritmo utilizado pelo autor, mostrando o melhor desempenho de cada um deles.

	opinião	fato	fake
opinião	4.159	0	29
fato	458	3.512	90
fake	416	77	3.459

Figura 23 – Melhor desempenho do algoritmo *NaiveBayes*. Fonte: (LEAL, 2018).

Conforme a Figura 23, é notório que o algoritmo em questão obteve um total de 11.130 acertos contra 1.070 erros. Em percentual pode-se avaliar que a taxa de precisão para identificar a classe das *fake news* é de 96.67%

	opinião	fato	fake
opinião	3.869	150	169
fato	238	3.754	68
fake	295	74	3.583

Figura 24 – Melhor desempenho do algoritmo J48. Fonte: (LEAL, 2018).

A matriz da Figura 24 remete ao algoritmo *J48* o qual apresenta 11.206 acertos contra 994 erros, que comparado ao algoritmo anterior (*NaiveBayes*) percebe-se uma melhora de 76 registros na sua precisão. Ao todo é percebido uma acurácia de 93.79% na classificação das *fake news*.

	opinião	fato	fake
opinião	4.090	25	73
fato	419	3.586	55
fake	384	28	3.540

Figura 25 – Melhor resultado do algoritmo *JRip*. Fonte: (LEAL, 2018).

Dentre todos os algoritmos utilizados pelo autor, a melhor acurácia se encontra no algoritmo do *JRip*, como pode ser comprovado na matriz de confusão da Figura 25, se comparado aos outros dois algoritmos (*NaiveBayes* e *J48*). Ele obteve um total de 11.216 acertos contra 984 erros, totalizando um percentual de acurácia de 96.41% na identificação das *fake news*.

	opinião	fato	fake
opinião	3.981	126	81
fato	545	2.689	826
fake	1.518	68	2.366

Figura 26 – Melhor desempenho do algoritmo *iBK*. Fonte: (LEAL, 2018).

E por fim, o algoritmo *iBK* teve a pior acurácia dentre os demais já mostrados, embora sua precisão se encontre próxima do *NaiveBayes*, *J48* e *JRip*. Os resultados podem ser vistos na Figura 26.

3.3 Discussão

Diante dos trabalhos apresentados na seção anterior (3.2) podemos observar que o fenômeno das *fake news* atrai também a atenção de pesquisadores, afim de que estes desenvolvam programas computacionais capazes de identificar e classificar automaticamente, e com certa precisão, notícias falsas ou manipuladas. Isso de fato se torna uma grande contribuição, pois devido ao elevado número de notícias publicadas com o avanço da internet, mesmo se utilizando de artifícios manuais ou online, sendo estes a verificação da fonte, checagem das credenciais do autor, como descrito na seção 2.6, pode ser inviável uma detecção manual perante um numeroso conjunto de notícias dentro de um domínio.

Considere que um pesquisador tenha por objetivo averiguar uma base de dados com 10.000 (dez mil) notícias cadastradas, afim de categorizar as mesmas como notícias reais e possíveis *fake news*. Realizar tal tarefa de forma manual tomaria muito tempo e esforço, e é em vista de situações como essas que os trabalhos de Leal (2018) e Marumo (2018) se fazem importantes, pois com a automatização da identificação de informações manipuladas, é possível desenvolver programas que realizem todo, ou boa parte do processo manual de classificar uma notícia, sendo ela verídica ou não, em vista de um elevado número de registros.

De acordo com o que foi observado na literatura, existem propostas para automatizar o processo de classificação de uma notícia, como pode ser observado nas pesquisas de SAKURAI (2019), onde este tem por objetivo aliar o Processamento de Linguagem Natural com o Aprendizado de Máquina para conseguir modelar um protótipo que verifique a veracidade de uma notícia, abstraindo as características mais comuns em *fake news*. Também é possível constatar através dos estudos de Andrade (2019) que é viável a proposta de identificação automática de notícias falsas através de mineração de dados e aprendizado de máquina, em seu trabalho o autor trabalha com redes neurais e captação dos dados através do *Data Mining*. Outro trabalho interessante na literatura é o de Vaz, Cassiano e Cordeiro (2018) que através da Análise de Sentimentos mostra uma solução para pesquisas de opinião dentro de campanhas eleitorais, onde cujo trabalho, com algumas modificações, pode ser utilizado para monitorar as informações publicadas durante as eleições, mais especificamente voltando-se para classificar uma possível *fake news*.

Em vista de pesquisas como as relatadas anteriormente, foi destacado o trabalho de Marumo (2018), no qual o autor propôs uma classificação de notícias falsas com enfoque na técnica de *Deep Learning* juntamente com o processo da mineração de texto. Seu domínio de estudo se concentrou em notícias de variados temas, para que se aproximasse o máximo possível de um cenário real, utilizando das técnicas de inteligência artificial para realizar tal classificação. Em seus resultados, ilustrados na subseção 3.2.1, é possível ver que a precisão de algoritmos baseados em IAs apresentam uma precisão aceitável, cerca de 80%, para classificação de *fake news*.

O interessante de seu trabalho se faz pela abordagem diferenciada no processo da mineração de texto, pois além de utilizar técnicas amplamente difundidas como o *stemming*, remoção das *stop words*, limpeza das notícias através da remoção de espaços múltiplos, palavras não alfabéticas, Marumo (2018) utilizou um processo denominado sumarização, onde este por sua vez cria sumários a partir de um texto fonte, levando em consideração assim não somente o valor das palavras separadamente, mas também o seu contexto em sim. Como explicado anteriormente (3.2.1), a sumarização realiza o resumo de um evento/informação, sendo assim algoritmos de *Deep Learning* podem associar as *fake news* com o uso de palavras-chave, que compõem o corpo de um notícia falsa, para isso foi utilizado o método do *Word2Vec*, uma vez que ele consegue associar um conjunto de palavras verificando o contexto onde se encontram, tornando assim a abordagem interessante para outras áreas que também levam em consideração o domínio do contexto de uma informação, por exemplo na Análise de Sentimentos, onde o enfoque principal se trata da opinião do usuário a respeito de um produto, algum evento ou entre outros (LIU, 2012). Como levantado pelo próprio autor, os resultados de acurácia podem ser melhorados utilizando uma arquitetura de *Deep Learning* mais complexa, além de expandir a quantidade de notícias podendo fornecer assim mais dados para o algoritmo do LSTM e outros, descobrirem novos padrões dentro das *fake news* e assim classificá-las.

No âmbito do outro trabalho ressaltado, Leal (2018) tem seu domínio mais focado em questões políticas, onde a ocorrência da divulgação de *fake news* é mais constante. Sua proposta em identificar as notícias falsas se deu através do *framework* WEKA, que já possui diversos algoritmos de aprendizado de máquina pré-programados para realizar testes com a base de dados que o programador desejar. Seu domínio de dados contou com um aglomerado de 1.780.821 (um milhão setecentos e oitenta mil oitocentos e vinte e um) tweets referentes a eleição presidencial do Brasil no ano de 2018. Utilizando algoritmos como *NaiveBayes*, *JRip*, *J48* e *iBK*, o autor conseguiu alcançar, em seus melhores resultados, uma acurácia de cerca de 90% na classificação de *fake news*, evidenciando assim que o desenvolvimento de algoritmos para a detecção automática das *fake news* através do aprendizado de máquina é uma solução viável de identificação de notícias dentro de um armazenamento grande de dados.

Em vista de uma base de dados tão numerosa e se tratando de dados em linguagem natural, Leal (2018) utilizou da mineração de texto principalmente para adequar sua base aos diferentes algoritmos que ele se propôs a realizar um comparativo. Para isso ele utilizou, assim como Marumo (2018), o processo do *stemming*, remoção de *stop words* e uso do filtro *reorder* para agilizar a normalização da base de dados aos diferentes classificadores utilizados na prototipação. Vale levantar que uma possível limitação para esta pesquisa, pode ser em relação a temática sobre suas notícias, tendo em vista que as notícias registradas são situadas somente no âmbito político, sua precisão na classificação pode ser afetada dentro de outros domínios.

Através dos trabalhos presentes na literatura, se torna evidente a colaboração da mineração de texto para o advento de identificar as *fake news* dentro de um cenário onde o fluxo das informações é constante. É possível perceber através dos estudos evidenciado nesta monografia que embora as notícias publicadas na internet sejam dados complexos e se apresentem no formato de linguagem natural, existem abordagens que conseguem capturar tais informações, transformá-las em dados que o computador compreenda e desenvolver, assim, programas que agilizem a classificação de uma possível notícia corrompida.

Finalizada a discussão, o próximo capítulo tratará das considerações finais do trabalho, objetivos alcançados e sugestões de trabalhos futuros.

4 CONCLUSÃO

Na conjuntura atual do compartilhamento de notícias é inquestionável a presença da divulgação de *fake news* dentre notícias reais, e neste contexto se tornam necessários estudos que fundamentam tal advento, explorando seus conceitos, impactos, modo de disseminação e métodos de combate. Nessa perspectiva o uso da mineração de texto é uma das possíveis abordagens que podem ser adotadas como solução para uma classificação automatizada de *fake news* haja vista o elevado armazenamento de informações com o avanço da internet.

Esta monografia, possibilitou realizar uma pesquisa bibliográfica na qual se identificaram estudos a respeito do assunto, tomando por base artigos como Braga (2018), que mostra o aspecto das *fake news* dentro do âmbito político, Luce (2018) que aborda maneiras de lidar com este fenômeno dentro de ambientes virtuais, Paula, Silva e Blanco (2018) e Gelfert (2018) onde seus trabalhos buscam retratar estudos de definição para as notícias falsas. Através destes conceitos, foi possível realizar a apresentação e discussão de duas pesquisas que evidenciaram propostas de automatização através da mineração de texto com o intuito de executar a classificação das *fake news* entre notícias reais.

Com a discussão levantada, através dos estudos de caso, foi possível atingir o objetivo central deste trabalho, explanando casos em que a mineração de texto se mostra um viés viável como forma de automatizar a classificação de *fake news* em vista de um número grande de notícias, onde um processo manual se faria inviável.

Conhecendo os conceitos das *fake news*, e dada a importância do assunto atualmente, se torna interessante que os estudos a respeito desta temática sigam em frente, reforçando e ampliando métodos de combate as divulgações de notícias corrompidas, sendo necessário um estudo mais aprofundando sobre os impactos que o compartilhamento de *fake news* podem acarretar dentro da sociedade. Seguindo esse cenário, como trabalhos futuros é sugerido a avaliação dos trabalhos apresentados nos estudos de caso, efetuando sua reprodução e verificando a veracidade de seus resultados, além de experimentar essas abordagens em contextos sociais específicos.

Referências

- ANDRADE, R. R. Utilização de técnicas de aprendizado de máquina para detecção de fake news. *Ciência da Computação-Tubarão*, 2019. Citado na página 43.
- ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. *Revista Eletrônica de Sistemas de Informação*, v. 5, n. 2, 2006. Citado na página 27.
- BARION, E. C. N.; LAGO, D. Mineração de textos. *Revista de Ciências Exatas e Tecnologia*, v. 3, n. 3, p. 123–140, 2015. Citado 2 vezes nas páginas 8 e 27.
- BOUCKAERT, R. R.; FRANK, E.; HALL, M.; KIRKBY, R.; REUTEMANN, P.; SEEWALD, A.; SCUSE, D. Weka manual for version 3-6-0. *University of Waikato, Hamilton, New Zealand*, p. 1–341, 2008. Citado na página 38.
- BRAGA, R. M. d. C. A indústria das fake news e o discurso de ódio. *Direitos políticos, liberdade de expressão e discurso de ódio: volume I*, Instituto para o Desenvolvimento Democrático, 2018. Citado 4 vezes nas páginas 13, 14, 15 e 46.
- CARVALHO, T. de. *COMO IDENTIFICAR NOTÍCIAS FALSAS? O CASO DE MARIELLE FRANCO*. 2018. Acesso em: 11/12/2019. Disponível em: <<https://www.politize.com.br/como-identificar-noticias-falsas/>>. Citado 2 vezes nas páginas 8 e 16.
- DENG, L.; YU, D. *Deep Learning: Methods and Applications*. [S.l.], 2014. Disponível em: <<https://www.microsoft.com/en-us/research/publication/deep-learning-methods-and-applications/>>. Citado na página 29.
- DINIZ, A. T. de M. Fact-checking no ecossistema jornalístico digital: Práticas, possibilidades e legitimação. *Mediapolis–Revista de Comunicação, Jornalismo e Espaço Público*, n. 5, p. 23–37, 2017. Citado na página 25.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37–37, 1996. Citado na página 27.
- FULL-FACT. *Full Fact is the UK's independent fact checking organisation*. 2019. Acesso em: 03 de Dezembro de 2019. Disponível em: <<https://fullfact.org/>>. Citado na página 25.
- G1, P. *É FAKE frase atribuída a Dilma sobre greve de fome*. 2018. Acesso em: 12 de Novembro de 2019. Disponível em: <<https://g1.globo.com/fato-ou-fake/noticia/2018/08/28/e-fake-frase-atribuida-a-dilma-sobre-greve-de-fome.ghtml>>. Citado 2 vezes nas páginas 8 e 19.
- G17. *G17*. 2019. Acesso em: 11 de Novembro de 2019. Disponível em: <<http://www.g17.com.br/>>. Citado na página 17.
- GARNER, S. R. et al. Weka: The waikato environment for knowledge analysis. In: *Proceedings of the New Zealand computer science research students conference*. [S.l.: s.n.], 1995. p. 57–64. Citado na página 35.

- GELFERT, A. Fake news: A definition. *Informal Logic*, v. 38, n. 1, p. 84–117, 2018. Citado 3 vezes nas páginas 12, 13 e 46.
- GUESS, A.; NAGLER, J.; TUCKER, J. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances*, American Association for the Advancement of Science, v. 5, n. 1, p. eaau4586, 2019. Citado na página 22.
- KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. *Ijcai*. [S.l.], 1995. v. 14, n. 2, p. 1137–1145. Citado na página 41.
- 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. 2018. Citado 14 vezes nas páginas 8, 13, 25, 26, 35, 36, 37, 38, 39, 40, 41, 42, 43 e 44.
- LIU, B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012. Citado na página 44.
- LOPES, G. *Relembre 9 fake news envolvendo o nome do Pablllo Vittar!* 2018. Acesso em: 11 de Novembro de 2019. Disponível em: <<http://www.e-farsas.com/relembre-9-fake-news-envolvendo-o-nome-do-pablllo-vittar.html>>. Citado 2 vezes nas páginas 8 e 18.
- LUCE, B. F. O bibliotecário e as fake news: atuação do profissional da informação na era da pós-verdade. 2018. Citado 3 vezes nas páginas 13, 15 e 46.
- LUPA, A. *LUPA, A PRIMEIRA AGÊNCIA DE FACT-CHECKING DO BRASIL*. 2019. Acesso em: 03 de Dezembro de 2019. Disponível em: <<https://piaui.folha.uol.com.br/lupa/>>. Citado na página 25.
- MARES, C. *Verificamos: É montagem foto de Dilma Rousseff com fuzil automático ao fundo*. 2019. Acesso em: 12 de Novembro de 2019. Disponível em: <<https://piaui.folha.uol.com.br/lupa/2019/08/06/verificamos-dilma-fuzil/>>. Citado 2 vezes nas páginas 8 e 21.
- MARTINS, C. B.; PARDO, T. A. S.; ESPINA, A. P.; RINO, L. H. M. Introdução à sumarização automática. *Relatório Técnico RT-DC*, v. 2, n. 1, p. 35, 2001. Citado na página 30.
- MARUMO, F. S. Deep learning para classificação de fake news por sumarização de texto. 2018. Citado 13 vezes nas páginas 8, 9, 13, 25, 26, 30, 31, 32, 33, 34, 35, 43 e 44.
- MORAES, M. *Verificamos: É falso que revista Charlie Hebdo publicou capa satirizando Bolsonaro*. 2019. Acesso em: 11 de Novembro de 2019. Disponível em: <<https://piaui.folha.uol.com.br/lupa/2019/08/23/verificamos-charlie-hebdo-capa-bolsonaro/>>. Citado 2 vezes nas páginas 8 e 18.
- NALLAPATI, R.; ZHOU, B.; GULCEHRE, C.; XIANG, B. et al. Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*, 2016. Citado na página 32.

- PAIVA, V. *Pega na mentira: 8 fake news que mudaram o curso da história antes da era Trump*. 2018. Acesso em: 12 de Novembro de 2019. Disponível em: <<https://www.hypeness.com.br/2018/08/pega-na-mentira-8-fake-news-que-mudaram-o-curso-da-historia-antes-da-era-trump/>>. Citado 2 vezes nas páginas 8 e 21.
- PARKINSON, H. J. Click and elect: how fake news helped donald trump win a real election. *The Guardian*, v. 14, 2016. Citado na página 15.
- PAULA, L. T. de; SILVA, T. d. R. S. da; BLANCO, Y. A. Pós-verdade e fontes de informação: um estudo sobre fake news. *Revista Conhecimento em Ação*, v. 3, n. 1, p. 93–110, 2018. Citado 3 vezes nas páginas 13, 14 e 46.
- PORTUGUESA. *Definição ou significado de sátira*. 2019. Acesso em: 11 de Novembro de 2019. Disponível em: <<https://www.infopedia.pt/dicionarios/lingua-portuguesa/sátira>>. Citado na página 17.
- PREVIDELLI, F. *ANDREW WAKEFIELD E A FAKE NEWS DAS VACINAS – QUE AINDA É PROPAGADA*. 2019. Acesso em: 12 de Novembro de 2019. Disponível em: <<https://aventurasnahistoria.uol.com.br/noticias/almanaque/andrew-wakefield-e-o-mito-das-vacinas-que-causam-o-autismo.phtml>>. Citado na página 23.
- ŘEHŘEK, R.; SOJKA, P. Gensim—statistical semantics in python. *statistical semantics; gensim; Python; LDA; SVD*, 2011. Citado na página 32.
- RONG, X. word2vec parameter learning explained. *CoRR*, abs/1411.2738, 2014. Disponível em: <<http://arxiv.org/abs/1411.2738>>. Citado 2 vezes nas páginas 30 e 32.
- ROSSI, M. Mulher espancada após boatos em rede social morre em guarujá, sp. *Publicado em*, v. 5, n. 05, 2014. Citado na página 23.
- SAKURAI, G. Y. Processamento de linguagem natural-deteção de fake news. 2019. Citado na página 43.
- SENSACIONALISTA. *Sensacionalista*. 2019. Acesso em: 11 de Novembro de 2019. Disponível em: <<https://www.sensacionalista.com.br/>>. Citado na página 17.
- SIGNIFICADOS. *Significado de Paródia*. 2016. Acesso em: 11 de Novembro de 2019. Disponível em: <<https://www.significados.com.br/parodia/>>. Citado na página 17.
- SNOPEs. *The definitive Internet reference source for urban legends, folklore, myths, rumors, and misinformation*. 2019. Acesso em: 03 de Dezembro de 2019. Disponível em: <<https://www.snopes.com/>>. Citado na página 25.
- SOUZA, A. de. <https://oglobo.globo.com/brasil/janot-denuncia-lula-dilma-mais-seis-por-organizacao-criminosa-21789906>. 2017. Acesso em: 12 de Novembro de 2019. Disponível em: <<https://oglobo.globo.com/brasil/janot-denuncia-lula-dilma-mais-seis-por-organizacao-criminosa-21789906>>. Citado 2 vezes nas páginas 8 e 20.
- VAZ, I. C.; CASSIANO, K. K.; CORDEIRO, D. F. Uma análise da disputa eleitoral presidencial utilizando mineração de dados em textos curtos de redes sociais. *XII SEMINÁRIO NACIONAL DE MÍDIA E CULTURA*, p. 64, 2018. Citado na página 43.

VEJA, R. *Desembargadora diz que Marielle ‘estava engajada com bandidos’*. 2018. Acesso em: 12 de Novembro de 2019. Disponível em: <<https://veja.abril.com.br/brasil/desembargadora-diz-que-marielle-estava-engajada-com-bandidos/>>. Citado 2 vezes nas páginas 8 e 20.

WARDLE, C. Fake news. it’s complicated. *First Draft News*, v. 16, 2017. Citado 2 vezes nas páginas 8 e 16.