UNIVERSIDADE PRESBITERIANA MACKENZIE
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E COMPUTAÇÃO

Vitor da Silva Souza

SENTIMENTUM: UM MÉTODO DE DETECÇÃO DE DISCURSOS ENGANOSOS

São Paulo - SP
2023

Vitor da Silva Souza

SENTIMENTUM: UM MÉTODO DE DETECÇÃO DE DISCURSOS ENGANOSOS

Projeto de Dissertação apresentado ao Programa de Pós-Graduação em Engenharia Elétrica e Computação da Universidade Presbiteriana Mackenzie, como parte dos requisitos para a obtenção do título de Mestre em Engenharia Elétrica e Computação.

Orientador: Prof. Dr. Leandro Augusto da Silva

São Paulo -SP

2023

Elaborado pelo Sistema de Geração Automática de Ficha Catalográfica da Mackenzie
com os dados fornecidos pelo(a) autor(a)

S719s Souza, Vitor Da Silva.

Sentimentum: um método de detecção de discursos enganosos : [recurso eletrônico] / Vitor da Silva Souza.

797 KB ; il.

Dissertação (Mestrado em Engenharia Elétrica e Computação) - Universidade Presbiteriana Mackenzie, São Paulo, 2023.

Orientador(a): Prof(a). Dr(a). Leandro Augusto Da Silva.

Referências Bibliográficas: f. 42-44.

1. Processamento De Língua Natural. 2. Fake News. 3. Mídias Sociais. 4. Inteligência Artificial. I. Da Silva, Leandro Augusto, *orientador(a)*. II. Título.

Bibliotecário(a) Responsável: Maria Gabriela Brandi Teixeira - CRB 8/6339

VITOR DA SILVA SOUZA

SENTIMENTUM: UM MÉTODO DE IDENTIFICAÇÃO DE DISCURSOS ENGANOSOS

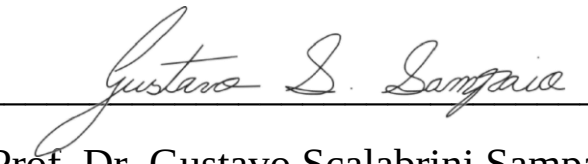
Projeto de Pesquisa apresentado ao Programa de Pós-Graduação em Engenharia Elétrica e Computação da Universidade Presbiteriana Mackenzie, como requisito para a obtendo título de Mestre em Engenharia Elétrica e Computação.

Aprovado em 14 de agosto de 2023.

BANCA EXAMINADORA

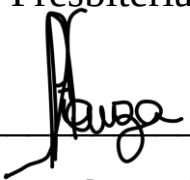
Prof. Dr. Leandro Augusto da Silva

Universidade Presbiteriana Mackenzie



Prof. Dr. Gustavo Scalabrini Sampaio

Universidade Presbiteriana Mackenzie



Profa. Dra. Alexandra Souza

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo

AGRADECIMENTOS

Agradeço a meus pais Adolfo e Lilia, por todo amor, incentivo e paciência, sem eles essa jornada não seria possível.

Ao meu orientador Prof. Dr. Leandro Augusto da Silva, por toda dedicação, paciência e por aceitar o desafio de uma orientação repentina.

À minha companheira, Maria Júlia, por toda a paciência, carinho e incentivo, mesmo nos momentos mais difíceis, sempre ao meu lado me apoiando com os estudos.

A dor passa, a alegria quer eternidade, quer profunda eternidade (Friedrich Nietzsche).

RESUMO

A popularização da internet e das mídias sociais ampliaram as possibilidades de comunicação, permitindo uma comunicação de muitos para muitos, em que conteúdos são publicados e compartilhados para milhares de pessoas em poucos minutos. Por outro lado, essa dinâmica de comunicação permite que conteúdos falsos, conhecidos como *Fake News*, sejam publicados na mesma velocidade para milhares de pessoas. A facilidade de produção e sua velocidade de propagação, fazem com que as *Fake News* se tornem um problema para a sociedade, podendo influenciar os ambientes políticos, econômicos e sociais. Esta pesquisa propõe o desenvolvimento de um método de detecção automática de *Fake News*, utilizando técnicas de processamento de linguagem natural, como "bag of words", análise de sentimentos, *Linguistic Inquiry and Word Count* (LIWC) e algoritmos de aprendizagem de máquina como *Support Vector Machine* (SVM) e Árvores de decisão. O método apresentado nesta dissertação foi adaptado de uma proposta de discursos enganosos propagados por diretores executivos e diretores financeiros nas divulgações de resultados trimestrais de empresas. A adaptação realizada consistiu em utilizar as técnicas de detecção de discursos enganosos em notícias falsas publicadas em jornais de forma escrita. O método foi adaptado ao contexto de *Fake News* e testado em bases de dados textuais da literatura no idioma inglês e seu resultado comparado com outras abordagens da literatura. Após a aplicação dos algoritmos verificou-se resultados positivos da utilização do LIWC para a detecção de *Fake News* com uma acurácia de 99,6% para o algoritmo SVM o que significa um resultado positivo quando comparado com demais resultados encontrados na literatura. Também foi possível identificar categorias do LIWC que frequentemente estão mais associadas com F*ake News* como as categorias *negate* e *I* (pronome pessoal).

Palavras-chave: Processamento de Língua Natural; Fake News; Mídias Sociais; Inteligência Artificial

ABSTRACT

With the popularization of the Internet and social media, communication from many to many has become possible, which allows, among other things, the propagation of false information, known as Fake News. The ease of production and its propagation speed make them a problem for society, which can influence the political, economic, and social environments. This research proposes the development of an automatic Fake News Detection method, using natural language processing (NLP) techniques, such as bag of words, sentiment analysis through LIWC and machine learning such as Support Vector Machine (SVM) and Decision trees. The method presented was adapted from the article called Detecting Deceptive Discussions in Conference Calls, which identifies misleading speeches propagated by CEO and CFO in company quarterly earnings releases. The adapted method, in the context of Fake News, was tested in textual databases of literature in the English language and its result compared with other literature approaches. After applying the algorithms, there were positive results from the use of LIWC for the detection of Fake news with an accuracy of 99.6% for the best SVM algorithm when compared to other results found in the literature. It was also possible to identify LIWC categories that often more associated with fake news, such as *negate* and *I* (personal pronoun) categories.

Keywords: Natural Language Processing, Fake News; Social Media, Artificial Intelligence.

SUMÁRIO

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE QUADROS

1 INTRODUÇÃO

As mídias sociais fazem parte do cotidiano de bilhões de pessoas ao redor do mundo e possuem grande relevância política, econômica e social. Segundo o último levantamento feito em janeiro de 2023, 4,76 bilhões de pessoas utilizam as mídias sociais ativamente, representando um aumento de 140 milhões de usuários quando comparado com o mesmo período de 2022 (Meltwater, 2023). No Brasil, 152 milhões de pessoas utilizam as mídias sociais ativamente, o que corresponde a 70% da população. O número de usuários de mídias sociais teve um crescimento de 2,4 milhões de usuários em comparação com o mesmo período de janeiro de 2022 (Meltwater, 2023).

Com este crescimento no uso das mídias sociais, houve um aumento na disseminação de notícias falsas nas mídias sociais, conhecidas também como *Fake News*. Apesar das notícias falsas serem um problema antigo da humanidade, esse problema se potencializou com o advento da internet e das notícias online, onde é possível publicar, compartilhar, encaminhar e receber notícias de forma muito mais rápida (CARDOSO DURIER DA SILVA, 2019).

O espalhamento de notícias falsas nas mídias sociais tem gerado efeitos negativos na sociedade e, por consequência, vem despertando o interesse de pesquisas sobre o tema em todo o mundo (RECUERO & SOARES, 2021). Como exemplo que marca este advento de *Fake News* destaca-se às notícias falsas que foram compartilhadas nas eleições de 2016 para a presidência dos Estados Unidos e que circularam durante o referendo do *Brexit* (BASTOS, 2019).

Não existe na literatura uma definição universal para *Fake News*. O que se encontram são co3nceitos relacionados quando o assunto é *Fake News*, que, embora imprecisas, ajudam a entender o tema e problemas de pesquisa que estão relacionados ao mesmo. ZHOU e ZAFARANI (2020) distinguem notícias de notícias falsas, notícias com viés e *Fake News*, àquelas que possuem atributos que podem ser encontrados em todas as notícias verdadeiras com informações inverídicas, vieses, manipulação entre outras características. Porém, o que distingue as *Fake News* é a intencionalidade de obter algum tipo de vantagem, seja econômica ou política com a disseminação de notícia falsa, além de apresentarem um espalhamento muito rápido na rede, muitas vezes associado ao uso de *bots* (ZHOU & ZAFARANI, 2020).

Assim, situa-se o contexto que será tratado na pesquisa, as *Fake News*, notícias falsas que são propositalmente disseminadas com o objetivo de beneficiar interesses particulares e não possuem o compromisso com a verdade. Compreender a circulação de *Fake News* é

fundamental para elaborar métodos de combate a esses fenômenos. Isso é particularmente importante no Brasil, país que possui grande polarização política, desde as eleições de 2018, e que possui compartilhamento de notícias falsas acentuado em épocas de eleições (SANTOS, 2021).

Ao mesmo tempo em que as *Fake News* são espalhadas, são propostos métodos diversos com o objetivo de identificá-las automaticamente (OSHIKAWA, QIAN, & WANG, 2018, PARIKH & ATREY, 2018, ZHOU & ZAFARANI, 2020, LILLIE & MIDDELBOE, 2019, CARDOSO DURIER DA SILVA, VIEIRA, & GARCIA, 2019 e SHU, 2017). Essa dissertação trata especificamente de um método de detecção de *Fake News* baseado na análise de mensagens textuais. Um método originalmente proposto para a identificação de discursos enganosos (LARCKER & ZAKOLYUKINA, 2012) que foi adaptado e adotado com o objetivo de identificar *Fake News*, principalmente nas mídias sociais (SOUZA, 2022).

1.1 Objetivos

Este trabalho tem como objetivo geral adaptar o método de identificação de discursos enganosos proposto em (LARCKER & ZAKOLYUKINA, 2012) no contexto de detecção de *Fake News*. Como objetivos específicos essa dissertação pretende:

- Realizar um estudo conceitual sobre as *Fake News*;
- Realizar um estudo sobre técnicas de identificação de *Fake News* baseadas em texto e aprendizagem de máquina;
- Adaptar um método de identificação de discursos enganosos para a detecção de *Fake News*.

1.2 Estrutura

Este trabalho está organizado em cinco capítulos da seguinte forma:

- Capítulo 1 – Introdução e Contextualização da Pesquisa.
- Capítulo 2 – Fundamentação Teórica sobre *Fake News*, análise de textos e trabalhos relacionados.
- Capítulo 3 – Metodologia: Caracterização da pesquisa e apresentação dos resultados obtidos na pesquisa.
- Capítulo 4 – Conclusão e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos fundamentais que envolvem a pesquisa, mais especificamente identificação de *Fake News* e Análise de Textos. A escolha desses tópicos visa abordar os principais conceitos que cercam o projeto de pesquisa.

2.1 *Fake News* e Mídias Sociais

As *Fake News* podem ser entendidas como a distribuição proposital de notícias falsas, desinformações que possuem como meio jornais, televisão, rádio e mídias sociais, que buscam obter ganhos financeiros, sociais ou políticos (KAPLAN, 2020). Essa definição baseia-se em dois conceitos: *autenticidade* e *intencionalidade*. A autenticidade consiste nas informações não verdadeiras que estão contidas na notícia falsa, enquanto a intencionalidade se refere ao objetivo explícito de enganar o leitor para obter algum tipo de benefício (MEDEIROS & BRAGA, 2020).

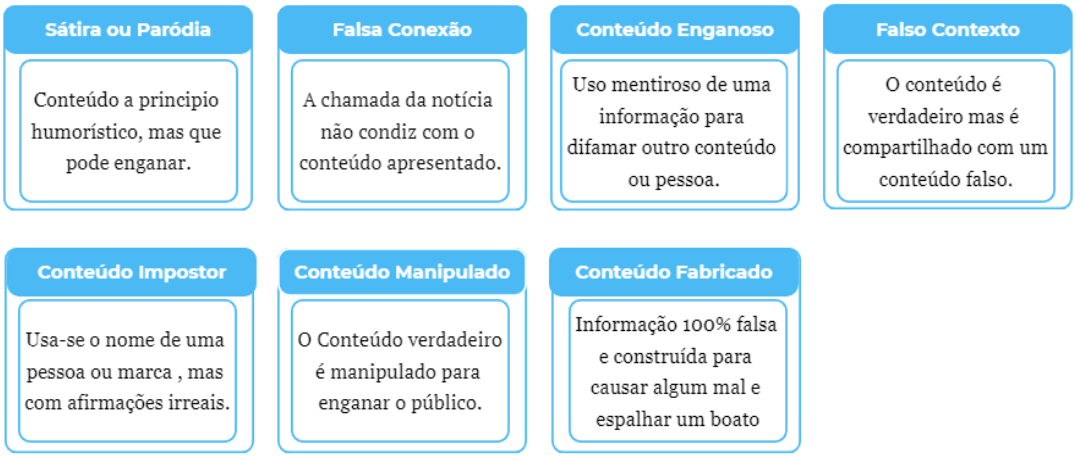
As *Fake News* podem ser classificadas de acordo com sete tipos: sátira ou paródia; falsa conexão; conteúdo enganoso; falso contexto; conteúdo impostor; conteúdo manipulado e conteúdo fabricado. A Figura 1 descreve resumidamente cada um desses sete tipos (WARDLE & DERAKHSHAN, 2017), a seguir detalhamos cada um destes sete tipos:

- Sátira ou paródia: Wardle e Derakhshan (2017) argumentam que mesmo que os receptores da informação compreendam o que são sátiras ou paródias, à medida que este conteúdo é compartilhado nas mídias sociais mais pessoas perdem a conexão com o mensageiro original e não conseguem distinguir a informação como tal. Os autores citam como exemplo um caso escrito no jornal Le Monde, por Adrien Sénécat que mostra o passo a passo para quem quer utilizar uma sátira com o objetivo de desinformar. 1 – Le Gorafi, um site satírico, “informou” que o candidato presidencial Emanuel Macron se sente sujo depois de tocar nas mãos de pessoas pobres. Isso funcionou como um ataque a Macron, por ser considerado elitista. 2 – Páginas do Facebook usaram essa reivindicação para criar conteúdos incluindo imagens de Macron visitando uma fábrica e limpando as mãos durante a visita. 3 – Os vídeos tornaram-se virais e um trabalhador de outra fábrica desafiou Macron a apertar suas mãos “sujas da classe trabalhadora”. O ciclo de notícias continuou.

- Falsa Conexão: Wardle e Derakhshan (2017) explicam que estes tipos de conteúdos são os populares "*clickbaits*" chamadas de notícias que não correspondem ao conteúdo efetivo da notícia e que podem induzir o leitor ao erro. Como exemplo, podemos citar uma chamada de notícia veiculada em um dos maiores jornais do país em época de eleição: "Confrontado, candidato corrige patrimônio após omitir conta bancária em declaração de bens", porém após clicar na matéria o leitor fica sabendo que o candidato, sem embate algum corrigiu os dados após ser questionado – uma conta no banco com saldo de R$ 579,53 (FOLHA, 2023).
- Conteúdo Enganoso: Esse tipo de conteúdo caracteriza-se por utilizar informações enganosas, manipular estatísticas e decidir não cobrir todos os dados, porque pode prejudicar um argumento. Como exemplo, podemos citar cortes em gráficos em um período relevante para a análise, ou até mesmo imagens com proporção distorcida, para acentuar um ponto de vista.
- Falso Contexto: Conteúdos verdadeiros são tirados de contexto para enganar o receptor, podemos citar como exemplo fotos tiradas de contexto dos refugiados Sírios na Gracia que foram utilizadas para reforçar discursos contra a imigração vindos de países da américa central aos Estados Unidos (WARDLE & DERAKHSHAN, 2017).
- Conteúdo Impostor: Este tipo de conteúdo utiliza-se da credibilidade de uma marca ou logotipo conhecido para enganar o leitor, como exemplo, utilizar a logomarca de um grupo jornalístico conhecido para dar uma notícia que não foi o próprio grupo que produziu.
- Conteúdo Manipulado: Este caso ocorre quando um aspecto genuíno é alterado, normalmente esta manipulação está mais associada a imagens e vídeos. Podemos citar como exemplo fotos de manifestações políticas manipuladas que são alteradas para das a impressão de que mais pessoas participaram.
- Conteúdo Fabricado: O conteúdo fabricado é aquele 100% falso, por exemplo, uma alegação falsa de que o Papa Francisco havia endossado a campanha de Donald Trump circulou nas eleições presidenciais dos Estados Unidos em 2016. A manchete apareceu em um site falso chamado WTOE5, que divulgou vários boatos falsos antes das eleições.

A distinção que os autores apresentam auxilia no entendimento do assunto que, por muitas vezes, é tratado apenas como um tipo de conteúdo enganoso. A diversidade de tipos de *Fake News* podem fazer com que algumas categorias não sejam reconhecidas facilmente como, por exemplo, as sátiras (WARDLE & DERAKHSHAN, 2017).

Figura 1 - Taxonomia de Fake News



Fonte: adaptado de (WARDLE & DERAKHSHAN, 2017).

As Fake News são um problema antigo das sociedades e possuem relatos ao longo da história, podemos citar como exemplo a causa da peste negra, que arrasou a Europa entre os anos de 1348-1350. A doença é transmitida pela bactéria *Yersinia Pestis*, porém devido ao conhecimento e tratamento a grupos marginalizados da época a doença foi atribuída indevidamente aos leprosos e judeus (FOLLADOR, 2017).

O ambiente onde encontra-se o maior número de *Fake News* atualmente são as mídias sociais, a popularização da internet e dos *smartphones* potencializou o número de usuários e o compartilhamento de conteúdo, por outro lado potencializou a disseminação de *Fake News* (KAPLAN, 2020). No início dos anos 2010 esperava-se que as mídias sociais trouxessem mais voz aos cidadãos, pois a informação agora poderia ser rapidamente e popularmente disseminada por todos através de mídias como o Facebook ou o Twitter (KAPLAN, 2020). Nessas plataformas a democracia seria exercida por todos. Como exemplo, na primavera árabe – uma série de protestos contra governos opressores que ocorreu ao Norte da África e no oriente Médio – o uso das mídias sociais foi determinante para facilitar a comunicação e o engajamento dos participantes desses protestos (KAPLAN, 2020).

Uma década após o fim da primavera árabe as mídias sociais revelaram ser uma ameaça para a democracia, ao invés de ser um instrumento de ampliação da comunicação e participação dos indivíduos nas sociedades. Para exemplificar essa ameaça, Kaplan (2020) cita o caso da consultoria Cambridge Analytica que utilizou os dados de milhões de usuários do Facebook em diversos eventos políticos, como as eleições presidenciais dos Estados Unidos de 2016 e o referendo do *Brexit* do Reino Unido de 2018. As informações obtidas pela Cambridge Analytica permitiam às empresas desenvolver algoritmos para direcionar mensagens de forma personalizada, adequados ao interesse de cada um.

O processo de eleições presidenciais no Brasil em 2018 também foi marcado por uma forte polarização entre os políticos de esquerda e de direita no país, resultando em uma grande quantidade de *Fake News* disseminadas. Dentre essas *Fake News* a que obteve maior circulação foi a que Fernando Haddad quando foi ministro da educação havia distribuído o "kit gay" nas escolas (SANTOS, 2021). Em todos os exemplos de *Fake News* apresentados verifica-se algum ganho seja econômico, político ou social para o emissor das informações falsas.

2.2 Detecção de *Fake News*

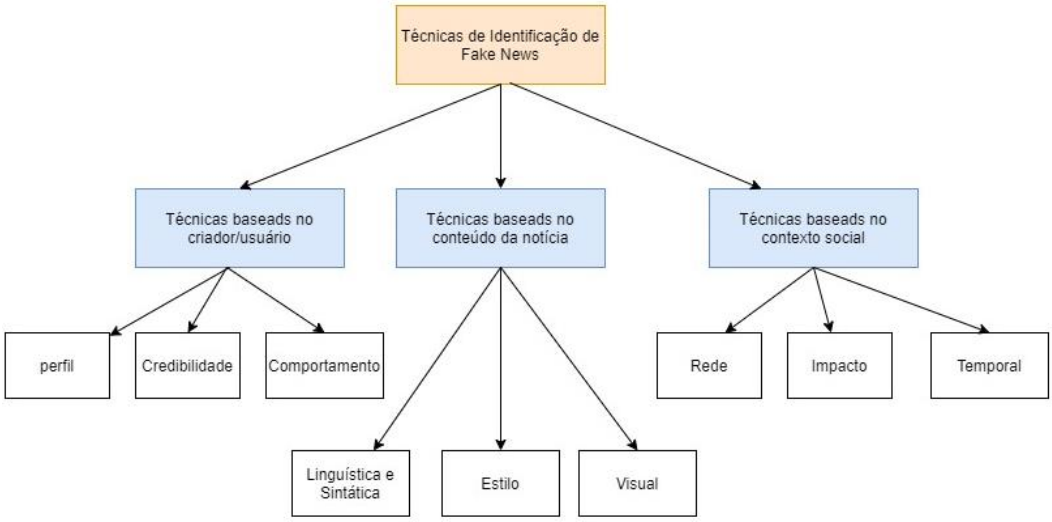
A detecção automática de *Fake News* pode ser definida como uma tarefa de avaliar as afirmações em uma notícia, classificando-as em verdadeiras, em inglês *true News,* ou falsas, em inglês Fake News (OSHIKAWA & WANG, 2018). Com a disseminação de *Fake News* nas mídias sociais, técnicas tradicionais de estruturação de documentos para Processamento de Língua Natural (PLN), como *bag-of-words* e *n-gram*, são utilizadas, porém apresentam as seguintes limitações (LARCKER & ZAKOLYUKINA, 2012).

- Por serem baseadas em contagem de palavras não levam em consideração o contexto que a palavra foi utilizada;
- No caso dos *n-grams* o processo pode apresentar um alto custo computacional, quanto maior for o valor de *n*;

As *Fake News* possuem quatro componentes principais: o *criador*, o *alvo*, o *conteúdo* e o *contexto social*. As técnicas de detecção de *Fake News* podem ser divididas considerando esses quatro componentes (ZHOU & ZAFARANI, 2020). Zang e Ghorbani (2020) realizaram um levantamento com as principais técnicas utilizadas atualmente na detecção de *Fake News*, um resumo dessas técnicas pode ser verificado na figura 2, que separa as técnicas em três grandes grupos: técnicas baseadas no criador/usuário; técnicas baseadas no conteúdo da notícia;

e técnicas baseadas no contexto social. Cada uma dessas técnicas, por sua vez, possuem aspectos que devem ser considerados para avaliação ou não de *Fake News*, que encontram-se representados abaixo das técnicas sinalizadas pelas setas pelas abaixo das técnicas em azul. As técnicas separadas na figura 2, apresentam a diversidade de aspectos que podem ser avaliados para a detecção de *Fake News*, bem como os instrumentos que são verificados por essas técnicas, por exemplo técnicas baseadas no criador/usuário que publicou a notícia, avaliam o perfil que publicou a notícia para classificar se o texto é ou não *Fake News*.

Figura 2 - Técnicas de detecção de Fake News



Fonte: Adaptado de (ZHANG & GHORBANI, 2020)

Além de usuários convencionais as mídias sociais também possuem uma parcela de usuários falsos, que imitam o comportamento humanos e muitas vezes são automatizados, conhecido também como *"bots"*. Esses usuários manipulam discussões, alteram a popularidade de um usuário, poluem conteúdos e disseminam desinformação.

A detecção baseada no usuário/criador é implementada por técnicas de identificação de *Fake News* por meio da detecção de usuários suspeitos. A detecção é feita comparando comportamentos de usuários convencionais com comportamentos de usuários falsos. A detecção baseada no usuário divide-se em análise de usuário, análise do comportamento de postagens ao longo do tempo, análise de sentimento e credibilidade da informação (ZHANG & GHORBANI, 2020):

1) **Análise de usuário**: são analisadas informações básicas do usuário como linguagem utilizada, localização geográfica da conta, data de criação da conta, quantidade de publicações, e se a conta possui verificação ou não;

2) **Análise do comportamento de postagens ao longo do tempo**: O comportamento temporal das postagens precisa tem como atributos de análise o tempo médio entre uma postagem e outra, a frequência de replicação e os compartilhamentos são aspectos analisados;

3) **Análise de sentimento**: A análise de sentimento pode ser utilizada como uma técnica de detecção de *Fake News*. Por ser o gatilho de uma resposta emocional anômala, contas maliciosas podem exagerar nos fatos e enganar usuários legítimos. Várias abordagens são utilizadas para extração de sentimentos, como *arousal-valence-domination score, happiness score, emotion score* e *polarization and strength* (ZHANG & GHORBANI, 2020).

4) **Credibilidade da informação**: O número de amigos e seguidores contribui na diferenciação de uma conta maliciosa de usuários legítimos. Normalmente contas maliciosas possuem muito mais amigos do que seguidores. Mislov et al. 2007, partem dessa premissa para definir uma equação que quantifica a reputação de uma conta com base na quantidade de amigos e seguidores. A Equação (1) parte do pressuposto de que contas maliciosas possuem muitos amigos e pouco seguidores, dessa forma, para contas com essa característica a Account_Reputation será próxima de zero e para contas de pessoas famosas será próxima de 1 (ZHANG & GHORBANI, 2020).

$$Account\_Reputation = \frac{followers}{followers+friends} \quad (1)$$

O contexto social, refere-se a todo sistema ativo e social em que a disseminação de notícias acontece, ele envolve como o dado social é distribuído e como os usuários interagem uns com os outros. Atualmente os meios de compartilhar e disseminar informação são dominados por meios interativos e tecnológicos nas mídias sociais (ZHANG & GHORBANI, 2020). As técnicas de detecção baseadas no contexto social são desenvolvidas buscando refletir o padrão de distribuição das notícias online e a interação entre os usuários. Essas técnicas dividem-se em três tipos: técnicas baseadas em rede, técnicas baseadas na distribuição e técnicas baseadas no contexto temporal (ZHANG & GHORBANI, 2020).

1) **Técnicas baseadas em redes**: São técnicas que tem por objetivo analisar características da rede, como grupos de usuários, localização, educação, bagagem, hábitos dentre outras. Essas características extraídas são utilizadas para estabelecer similaridades e dissimilaridades entre as diferentes contas de usuários. Técnicas como: posição de rede, rede de coocorrência, rede de amizade e rede de difusão fazem parte do contexto das técnicas baseadas em redes.

2) **Técnicas baseadas na distribuição**: São técnicas que tem por objetivo analisar características da difusão de informação feita pelos usuários. Usualmente uma árvore de propagação pode ser construída para facilitar a identificação da natureza de distribuição das notícias. Essa árvore possui atributos que incluem: grau da raiz na propagação, número máximo de filhos, profundidade da árvore dentre outras. Outros atributos também são analisados como o número de retuite, na entrada e saída de cada nó da árvore, essa informação pode ser usada para mensurar o grau de atividade suspeita de uma conta disseminando *fakes news*.

3) **Técnicas baseadas no contexto temporal**: São técnicas que tem por objetivo analisar características temporais dos usuários como: Intervalo entre postagens, frequência de postagens, compartilhamento e comentários das postagens, o horário da postagem e o dia da semana que foi publicada a notícia.

2.3 Análise de Textos

O campo de *Processamento de Língua Natural* (do inglês *Natural Language Processing*, NLP) é a área da *Ciência da Computação* que envolve a aplicação de métodos computacionais para a análise e síntese da língua humana, e pode ser dividido em duas subáreas (OTTER, MEDINA, & KALITA, 2020): *centrais* e *de aplicação*. As áreas centrais lidam com problemas de modelagem da língua, ocorrência natural de palavras, processamento morfológico, segmentação de componentes, identificação das partes do discurso das palavras, processamento sintático, dentre outras. As áreas de aplicação envolvem tópicos como extração de informação útil, tradução de texto entre línguas, sumarização da escrita, resposta automática de questões por inferência, classificação, agrupamento, dentre outras.

As técnicas de NLP podem ser utilizadas para resolução de problemas de análise de textos, tendo como ponto de partida o *corpus*, ou seja, um conjunto de materiais linguísticos a serem analisados (FERREIRA & LOPES, 2021). *Corpus* é um termo latino, cujo plural, *corpora*, significa corpo (GLOSBE, 2023). Para que um *corpus* possa ser analisado e

processado por um programa de computador é necessário efetuar um tratamento, ou preparação, no texto. Um *corpus* possui as seguintes propriedades (FERREIRA & LOPES, 2021):

1. O *corpus* é um produto já realizado e não em processo de realização.
2. Qualquer que seja sua extensão, o *corpus* é sempre finito.
3. Por mais representativo que possa ser, o corpus corresponde a apenas parte das possibilidades realizáveis da língua e do discurso.

Um corpus é dito anotado se, paralelamente aos dados, contém marcações (*tags*) feitas sobre os dados que sirvam para classificá-los de alguma forma (FERREIRA & LOPES, 2021). Anotações comuns nesse processo são as que compõem parte da oração, como verbo, substantivo etc. Essas anotações podem ser realizadas manualmente ou através de programas, que são conhecidos como *etiquetadores* (FERREIRA & LOPES, 2021). Um etiquetador, do inglês *tagger*, é um sistema que tem como meta identificar a categoria gramatical de cada léxico do texto analisado (RAJAN, SALGAONKAR, & JOSHI, 2020).

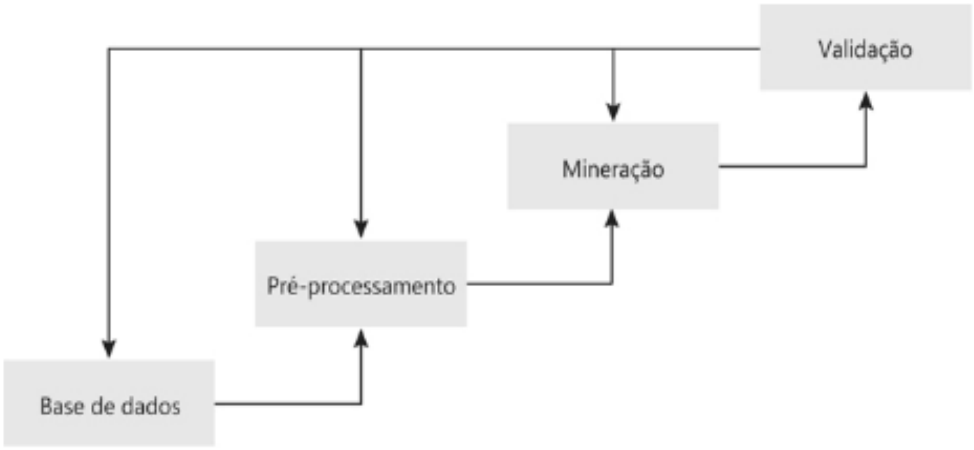
Na análise de textos, um passo importante para a utilização do *corpus* é a limpeza. Isso devido a necessidade de se avaliar as unidades do texto, como sílabas, palavras, frases e não necessariamente todos os caracteres do texto, como artigos e pontuações. No processo de limpeza é comum a criação de uma lista de palavras que não serão utilizadas na análise de texto, além do texto ser convertido todo em minúsculo para facilitar a comparação de palavras como “casa” e “Casa”. A contagem de palavras em análise computacional refere-se a análise quantitativa das unidades contidas no corpus (FERREIRA & LOPES, 2021).

A quantidade de dados trafegada nas mídias atualmente torna inviável a análise manual dos dados e gera a necessidade de métodos e ferramentas computacionais para análise automática dos dados. O processo de *mineração de dados* (MD) pode ser entendido por meio de uma alusão à mineração de ouro e pedras preciosas, onde uma base de dados (mina) é minerada usando algoritmos (ferramentas) adequados para obter conhecimento (minerais preciosos) (DE CASTRO, 2016).

A mineração de dados possui um fluxo que é conhecido por *Descoberta de Conhecimento em Bases de Dados* (do inglês *Knowledge Discovery in Databases*, KDD). O processo consiste inicialmente na definição da base a ser utilizada e, posteriormente, na implementação de pré-processamento, mineração e validação (DE CASTRO, 2016). No pré-processamento algumas técnicas de tratamento dos dados podem ser aplicadas como: imputação

de valores ausentes, deleção de objetos e atributos, conversão e normalização dos dados, tratamento de anomalias, seleção de atributos etc. Uma vez que os dados foram tratados com o objetivo de extração de conhecimento, são aplicados algoritmos de aprendizagem de máquina, como algoritmos de estimação, classificação, regras de associação e agrupamento. Por fim, os resultados obtidos passam por uma etapa de validação. O processo de descoberta de conhecimento é iterativo e interativo, onde cada etapa do processo pode ser reimplementada para a obtenção de melhores resultados (DE CASTRO, 2016). O fluxograma abaixo resume o processo de KDD (Figura 3).

Figura 3 - Processo de descoberta de conhecimento em base de dados.



Fonte: (DE CASTRO & FERRARI, 2016).

Grandes quantidades de dados são geradas nas mídias sociais diariamente, devido ao seu grande número de usuários. Por exemplo, atualmente as mídias sociais possuem 4,76 bilhões de usuários ativos que podem postar notícias, vídeos e realizar comentários na rede. Para exemplificar esses números 350 milhões de fotos são compartilhadas diariamente no facebook, 500 horas de vídeo são enviadas a cada minuto para o Youtube e mais de 500 milhões de tweets diários são enviados no Twitter (Meltwater, 2023). Essa grande quantidade de dados trafegados traz desafios para a filtragem de informações relevantes e obtenção de conhecimento desses dados brutos.

A difusão de informações nas redes sociais ocorre de forma muito acelerada e possui um grande potencial econômico envolvido. A *mineração de dados de mídias sociais* apresenta conceitos e algoritmos para analisar a grande quantidade de dados disponíveis nas mídias sociais. Este é um campo recente e interdisciplinar que abrange áreas do conhecimento como

ciência da computação, sociologia, etnografia, estatística, otimização e matemática (ZAFARANI, ABBASI, & LIU, 2014).

A difusão de informação envolve os seguintes agentes (ZAFARANI, ABBASI, & LIU, 2014):

1. **Emissor**: Um ou mais emissores que iniciam o processo de difusão de informação.
2. **Receptor**: Um ou mais receptores que recebem a informação difundida. Geralmente o conjunto de receptores é muito maior que o conjunto de emissores e pode sobrepor os emissores.
3. **Meio**: É o meio por onde a difusão toma espaço, por exemplo, quando um boato é espalhado o meio seria a comunicação pessoal entre os indivíduos.

A mineração de textos (do inglês *Text Mining*) é um processo de descoberta de conhecimento que utiliza técnicas de análise e extração de dados a partir de textos, frases ou apenas palavras, permitindo uma melhor compreensão do conteúdo de documentos textuais (MORAIS & AMBRÓSIO, 2007). A mineração de textos também pode ser entendida como a aplicação de técnicas de KDD sobre dados extraídos de textos. Contudo, a mineração de textos não inclui somente as técnicas de KDD, mas também técnicas que possam ser aplicadas no sentido de encontrar conhecimento em qualquer tipo de texto (WIVES, 2002):

1) **Tokenização**: É utilizada para separar o texto em palavras, chamadas de *tokens*, e é também conhecida como segmentação de palavras;
2) **Remoção de *stopwords***: As *stopwords* são palavras que podem ser consideradas pouco relevantes para o entendimento do sentido do texto, como artigos, preposições, alguns verbos etc.
3) **Stemming**: É o processo de extração da raiz da palavra, removendo plural, gênero e outros sufixos ou prefixos que não agregam significado contextual. O processo de *stemming* permite uma redução na quantidade de tokens que comporão a matriz de dados final;
4) **Representação dos documentos**: Essa abordagem também conhecida como *Bag of Words* ignora a ordem em que as palavras aparecem no texto, assim como qualquer informação de pontuação, mas retém o número de vezes que cada palavra ocorre no texto.

Ao final do pré-processamento uma matriz de dados é gerada, onde cada linha representa um texto (ou documento) e cada coluna um termo (palavra ou token). Feito isso, é necessário determinar os coeficientes dessa matriz, ou seja, os pesos de cada palavra (*token*) nos documentos.

O TF-IDF (TF do inglês *term frequence* e IDF *inverse document frequency*) é um método de cálculo do peso das palavras no texto baseado na frequência de ocorrências no texto, buscando medir a relevância de um termo em um documento. O TF-IDF mostra a frequência de ocorrência dos termos considerando o equilíbrio entre o termo local e seu significado no contexto da coleção completa de documentos (KACZMAREK, 2021):

$$TF(i,j) = \frac{frequência\ do\ termo\ i\ no\ documento\ j}{número\ total\ de\ palavras\ no\ documento\ j} \qquad (2)$$

$$IDF(i) = \log_{10}\left(\frac{número\ total\ de\ documentos}{documentos\ com\ o\ termo\ i}\right) \qquad (3)$$

$$TF - IDF = TF(i,j) \times IDF(i) \qquad (4)$$

A forma como o ser humano se comunica carrega uma rica informação sobre suas crenças, medos, padrões de pensamento, personalidade e padrões sociais. Após os estudos de Freud, os psicólogos perceberam que as palavras possuem grande valor para a psicologia e podem ser analisadas de forma sistemática (PENNEBAKER, 2015). A popularização dos computadores pessoais, a expansão da internet e a grande quantidade de documentos que passou a ser disponibilizada em formato digital, possibilitou a análise dos fatores psicológicos das palavras usando métodos computacionais (PENNEBAKER, 2015).

O LIWC (*Linguistic Inquiry and word count*) é um software proposto por Pennebaker (2015) para analisar os componentes emocionais, cognitivos e estruturais do texto usando um dicionário de palavras em inglês. O LIWC15 categoriza cada palavra de um corpus em um conjunto de 60 categorias, a lista completa com todas as categorias encontra-se no Apêndice I.

A versão do LIWC15 possui um total de 6.400 palavras organizadas em categorias. O dicionário possui palavras pertencentes a mais de uma categoria que refletem processos psicológicos e linguísticos relacionados a diversos assuntos, como emoções positivas (*posemo*), processos sociais (*social*), pronomes (*pronoun*), dentre outras.

As categorias do LIWC15 possuem subdivisão, por exemplo, a categoria dos pronomes pessoais (*pron*) é dividida em cinco subcategorias: primeira pessoa do singular (*I*), primeira pessoa do plural (*we*), segunda pessoa (*you*), terceira pessoa do singular (*she/he*) e terceira pessoa do plural (*they*)

O LIWC15 possui também uma versão de seu dicionário em português do Brasil que foi traduzida por pesquisadores brasileiros (BALAGE FILHO, 2013) essa versão encontra-se disponibilizada em uma página web (PortLex, 2022), sob licença aberta.

A tarefa de estimação busca predizer uma saída contínua com base no histórico de registros passados. Um exemplo de um modelo de estimação é o de uma empresa de crédito que busca informar qual é o valor do limite disponível para um determinado cliente (DE CASTRO, 2016). A estimação pode ser realizada usando algoritmos como árvores de decisão e redes neurais artificiais.

O desenvolvimento de um modelo preditivo possui duas etapas: *treinamento* e *teste*. No treinamento o preditor é gerado para que seja capaz de distinguir um conjunto determinado de classes ou valores. O preditor é gerado utilizando um conjunto de dados rotulados, ou seja, para cada vetor de entrada existe uma ou mais saídas desejadas, que pode ser, por exemplo, a classe à qual um objeto pertence (DE CASTRO, 2016). No teste, após o preditor ser treinado, é preciso avaliar seu desempenho quando aplicado a dados que não foram utilizados no processo de treinamento. O desempenho do preditor quando aplicado a dados de teste oferece uma estimativa de sua capacidade de responder corretamente a dados não usados no processo de treinamento, ou seja, sua *capacidade de generalização* (DE CASTRO, 2016).

2.4 Trabalhos Relacionados

Para o desenvolvimento da pesquisa foi feita uma revisão bibliográfica preliminar com o objetivo de compreender melhor e conhecer o estado da arte do tema da pesquisa. Para identificação de trabalhos relacionados ao tema utilizou-se a ferramenta Google Acadêmico, pesquisando pelo tema: “Detecção de *Fake News*” A pesquisa trouxe aproximadamente 6.370 resultados, para filtrar a pesquisa escolhemos apenas artigos publicados desde 2018 e artigos de revisão sobre o tema, desta forma obtivemos 53 resultados de pesquisa. Alguns destes trabalhos que apresentaram maior relevância para essa pesquisa serão citados a seguir.

No trabalho realizado por Gusmão, Figueredo e Brito (2021), efetua-se uma avaliação e classificação do disque denúncia da cidade do rio de janeiro, com o objetivo de automatizar e agilizar o trabalho realizado pela polícia, levando em consideração textos, com muitos erros de ortografia. Ambos os trabalhos irão analisar o texto e realizar uma classificação, no caso deste trabalho se é ou não é *Fake News*, já no trabalho realizado por Gusmão, Figueredo e Brito (2021) utilizou-se uma classificação para um tipo de ocorrência policial (GUSMÃO, FIGUEIREDO, & BRITO, 2021).

No trabalho realizado por Recuero e Soares (2021), os autores analisaram 57.295 tweets, com o objetivo de compreender como se deu a circulação de desinformação a respeito de uma possível "cura" para o corona vírus com foco na discussão: (1) dos modos através dos quais os discursos relacionados à desinformação sobre supostas curas da pandemia foram espalhadas, (2) os diferentes tipos de discurso desinformativo e sua prevalência, e (3) os modos de disputa contra os discursos que desmentem a desinformação. Um aspecto relevante verificado neste trabalho refere-se as diferentes conceituações de discurso enganoso, após a análise de diversos autores, Recuero e Soares compreendem o conceito de discurso enganoso em três aspectos:

1) **Informação fabricada**: Informação completamente falsa, fabricada ou sem nenhuma relevância com, por exemplo, teorias da conspiração
2) **Informação com enquadramento enganoso**: Informações verdadeiras utilizadas para criar um sentido falso devido à forma como são apresentados os tipos de conexões que são realizadas a partir delas.
3) **Informações manipuladas**: Informações parcialmente verdadeiras manipuladas para construir um falso sentido. Por exemplo, imagens verdadeiras manipuladas de modo a acrescentar ou retirar uma informação essencial.

Ao fim do trabalho os autores identificam que os líderes de opinião, mesmo quando apresentam um discurso enganoso possuem um grande poder de disseminação dessa informação nas mídias sociais. Os autores baseiam-se na análise de dados que possuíam força no combate as desinformações relacionadas a "cura" da COVID-19 através do uso de cloroquina, porém após uma declaração do presidente da república, um líder de opinião verificou-se um aumento da desinformação da "cura" da COVID-19 no *Twitter* engajado principalmente por líderes de opinião (RECUERO & SOARES, 2021).

No trabalho realizado por (MEDEIROS & BRAGA, 2020) os autores realizam uma revisão sistemática de literatura através da metodologia prisma. Os autores realizam uma

revisão dos algoritmos aplicados para a detecção de *Fake News*, elaborando uma tabela com os principais algoritmos utilizados e acurácia obtida por algoritmo, conforme apresentado na tabela 1.

Tabela 1 - Acurácia de algoritmos na detecção de Fake News.

Acurácia	**Algoritmo**
0,98	CNN+GRU
0,961	Decision Tree
0,956	Self-Att-GRU+BERT
0,953	LSTM
0,948	Att-Based CNN
0,948	LSTM
0,944	Att-Based LSTM
0,943	Bi-LSTM+HAN
0,938	LR+HBLC18
0,936	LR
0,933	CNN
0,921	RNN+CNN
0,92	SVM
0,912	LSTM
0,91	LR
0,906	LR17
0,902	LPT16+SVM
0,896	SVM
0,89	LSTM+HAN+SVM
0,889	Deep Bi-GRU
0,878	NMF
0,84	LSTM
0,827	CNN
0,81	XGBoost
0,809	Bi-LSTM+MLP
0,805	2-Layer LSTM
0,788	Att-Based LTSM
0,78	LSTM
0,75	SVM
0,742	RNN
0,737	Tree-structured RNN
0,691	CRF20+MaxEnt

Fonte: (MEDEIROS & BRAGA, 2020).

No trabalho realizado por (FREIRE, DA SILVA, & GOLDSCHMIDT, 2021) os autores utilizam técnicas de sinal de multidão híbrido, do inglês *Hybrid Crowd Signals* (HCS), para a detecção de *Fake News*.

Essa técnica utiliza sinais implícitos e explícitos de usuários em ambientes de mídias digitais para concluir se são *true news* ou *Fake News*. Os sinais implícitos são obtidos através de inferências utilizando o comportamento de um usuário no momento da disseminação de como, por exemplo, o caminho de propagação da notícia: se um usuário compartilha notícias que são compartilhados com baixa reputação, sua reputação tende a diminuir.

Já os sinais explícitos são, por exemplo, obtidos através de comentários de usuários em notícias. Os autores utilizam esse conjunto de características em 5 bases de dados para classificar uma notícia em *true news of Fake News* os autores identificaram resultados promissores através de algoritmos de aprendizagem de máquina como SVM e *Random Forest*.

3 METODOLOGIA EXPERIMENTAL

Este capítulo tem por objetivo apresentar o método *Detecting Deceptive Discussions in Conference Calls* (D3C2), proposto por (LARCKER & ZAKOLYUKINA, 2012), que originalmente foi aplicado na detecção de discursos enganosos em conferências de demonstração do balanço trimestral de empresas. Após apresentar o método, apresenta-se a adapatação prposta neste trabalho para a detecção de *Fake News* em textos de mídias sociais.

No método proposto no artigo D3C2, os autores utilizam categorias linguísticas do dicionário LIWC para classificar ligações de conferência em enganosas e não enganosas com um resultado de 6%-16% melhor do que um palpite aleatório. A adaptação proposta consite em utilizar o mesmo método de detecção de discursos enganosos no contexto das *Fake News*, isto é, propõe-se realizar um pré-processamento no texto extraindo categorias da psicologia encontradas no dicionário LIWC, baseados nas premissas do artigo D3C2, que serão detalhadas no tópico 3.1 Método proposto em D3C2.

Este capítulo está organizado da seguinte forma. No tópico 3.1 iremos apresentar em detalhes o método D3C2, no tópico 3.2 iremos explicar como adaptamos o método do artigo para a identificação *Fake News* em textos originados de mídias sociais, no tópico 3.3 iremos apresentar a base de dados utilizada em nosso método e por fim, no tópico 3.4 iremos aplicar algoritmos de classificação para a detecção de *Fake News* ou *True News* na base de dados.

3.1 Método proposto em D3C2

O método proposto pelos autores (LARCKER & ZAKOLYUKINA, 2012) realiza uma análise linguística e sintática de textos extraídos de conferências de fechamento das demonstrações contábeis do trimestre de empresas. O conjunto de ligações dessas conferências foram transcritas em textos e serviram de base para construção de um modelo de previsão da probabilidade de um erro na divulgação dos relatórios trimestrais. O conjunto de conferências analisadas compreendem o período de setembro de 2003 a maio de 2007 (LARCKER & ZAKOLYUKINA, 2012).

O objetivo desse método foi identificar discursos enganosos propagados pelos diretores executivos (sigla em inglês, CEO – Chief Executive Officer) e os diretores financeiros (sigla em inglês, CFO - Chief Financial Officer) CEO e CFO dessas empresas em conferências de demonstração de resultados trimestrais. Os autores argumentam que muitas vezes os CEO e

CFO possuem conhecimento real dos dados, porém por motivos econômicos, podem apresentar informações falsas. Este tipo de análise é de interesse para pesquisadores, investidores, credores e órgãos reguladores de mercado financeiro (LARCKER & ZAKOLYUKINA, 2012) por conseguir capturar de forma mais precisa divulgações enganosas.

Para realizar a análise linguística e sintática os autores fundamentam-se na revisão de literatura baseada em (VRIJ, 2008), que possui como premissa quatro perspectivas da psicologia: emoções, esforço cognitivo, tentativa de controle e falta de acolhimento. A seguir serão apresentadas de forma resumida essas quatro teorias utilizadas no método proposto em D3C2.

A **perspectiva de emoção**, parte da premissa que mentirosos, sentem culpa e têm medo de serem descobertos mentindo. Desta forma, eles podem utilizar emoções negativas que são manifestadas tanto em comentários quanto em afeto negativo (LARCKER & ZAKOLYUKINA, 2012).

Autores que trabalham com a **perspectiva do esforço** cognitivo argumentam que fabricar uma mentira é difícil, se o mentiroso não tiver a oportunidade de cuidadosamente preparar a mentira. Isso se deve ao fato de que as suas declarações possuem grandes chances de carecer de detalhes específicos e em vez disso, incluirão termos mais gerais e com poucas menções a experiências pessoais. Consequentemente, essa perspectiva implica em pouca autorreferência e declarações mais curtas (LARCKER & ZAKOLYUKINA, 2012).

Teóricos da **perspectiva do controle** argumentam que os mentirosos evitam produzir declarações que os autoincriminam, desta forma, o conteúdo de declarações enganosas é controlado para que os ouvintes não percebam facilmente que as declarações são enganosas. Essa perspectiva implica em declarações impessoais, com menor quantidade de autorreferencias, declarações com poucos detalhes e informações irrelevantes como substituições para que o enganador não apresente fatos que o autoincrimine (LARCKER & ZAKOLYUKINA, 2012).

Por último, os defensores da **perspectiva da falta de acolhimento** argumentam que mentirosos parecem não ter convicção do que estão dizendo, porque, se sentem desconfortáveis quando mentem ou porque não experimentaram pessoalmente as supostas alegações, acarretando imprecisões. Essa perspectiva implica que os mentirosos utilizam mais termos gerais, menos autorreferências e mais respostas curtas (LARCKER & ZAKOLYUKINA, 2012).

Verifica-se que as quatro teorias psicológicas sugerem que os mentirosos são mais negativos e utilizam menos autorreferências, porém dependendo da teoria associações entre categorias linguísticas podem ser ambíguas (LARCKER & ZAKOLYUKINA, 2012).

Com o entendimento do funcionamento das quatro perspectivas da psicologia, os autores utilizam a ferramenta LIWC, extraindo do texto palavras associadas às categorias de seu dicionário interno, utilizando como premissa que essas categorias são as que melhor se encaixam na detecção de discursos enganosos (LARCKER & ZAKOLYUKINA, 2012).

O software LIWC lê o texto e compara cada palavra com a lista de palavras do seu dicionário interno e calcula a porcentagem do total de palavras no texto que correspondem a cada uma das categorias do dicionário. Internamente o LIWC aplica o modelo "*bag of words*" que representa o texto através de um vetor de palavras, contando quantas vezes uma determinada palavra aparece no texto, a diferença do LIWC para o modelo "*bag of words*" é que com o LIWC são contadas as palavras que se encontram dentro do dicionário interno de categorias, também chamado LIWC. Este dicionário possui categorias especificas que estão associadas a psicologia, desta forma, o dicionário conta a quantidade de palavras que ocorre para cada categoria do LIWC (LARCKER & ZAKOLYUKINA, 2012).

No método proposto em D3C2, após a representação dos textos, em um "bag of words" utilizando o LIWC e técnicas de pré-processamento como a normalização dos atributos identificados, remoção de ruídos, dentre outros, os autores aplicam um algoritmo de regressão logística para classificação dos discursos em enganosos ou verdadeiros.

3.2 Adaptação realizada no método proposto D3C2

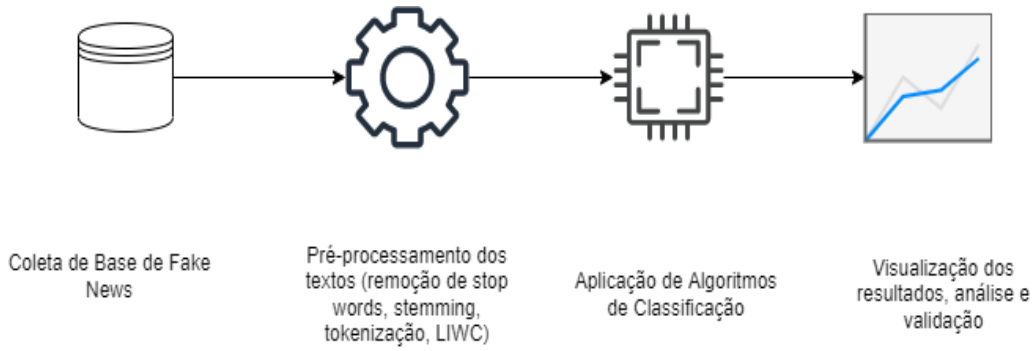
Para adaptação neste trabalho , utilizou-se como base as categorias do LIWC que foram utilizadas do artigo D3C2 e categorias que se encaixem nas premissas elencadas pelos autores, isto é, as quatro perspectivas da psicologia: emoção, esforço cognitivo, tentativa de controle e falta de acolhimento. Desta forma para a aplicação do método foram selecionadas as categorias listadas no Quadro 1, que será exibido a seguir no trabalho.

O processo de construção do método divide-se em quatro etapas, conforme ilustrado na figura 4.

- Coleta de Base de dados *Fake News*;
- Pré-processamento do texto – Aplicação de técnicas de processamento de língua natural no texto como: *tokenização*, *bag of words*, *stemming* e remoção de *stopwords*;

- Aplicação de algoritmos de classificação – Aplicação dos algoritmos Máquina de Vetores de Suporte (*Support Vector Machine*, SVM) e Árvore de Decisão (*Decision Tree*);
- Visualização dos resultados obtidos – Representação gráfica dos resultados obtidos;

Figura 4 - Processo do método proposto.



Fonte: Próprio Autor.

3.3 Base de dados

Para a realização do projeto foi utilizada uma base de dados aberta, disponibilizada pelo site Kaggle *Fake News - Build a system to identify unreliable news articles*" que foi elaborada por estudantes da universidade do Tennessee (KAGGLE, 2021). A base de dados possui 20.800 notícias, no idioma inglês organizadas em cinco atributos: *id*, *title*, *author*, *text* e *label*. A seguir será apresentada uma breve descrição de cada um destes atributos.

O atributo *"id"* representa um identificador único, o atributo *"title"* representa o título do texto, o atributo *"author"* contém o nome do autor da notícia e o atributo *"label"* representa a classificação da notícia, em que (zero (0) significa que temos uma notícia verdadeira, *"true news"* significa que temos uma notícia verdadeira, *"true news"* e um (1) significa que se trata de uma notícia false, uma "*Fake News*" (KAGGLE, 2021). A base de dados possui uma distribuição aleatória de 50% de "*Fake News"* e 50% de "*true news"* e os textos do atributo *text* estão no idioma inglês.

3.4 Aplicação do método sentimentum

Este tópico apresenta o Sentimentum, o método de detecção de discursos enganosos, elaborado neste trabalho. Conforme ilustrado anteriormente na figura 4. Após a coleta da base de dados, os dados foram pré-processados e em seguida aplicamos algoritmos de aprendizagem

de máquina e por fim foi apresentada uma representação visual e análise dos resultados obtidos.

A primeira etapa realizada foi utilizar o software LIWC na base de dados de treino (KAGGLE, 2021), no atributo *text*, que possui o texto de uma notícia em inglês. O LIWC calcula o grau de uso de diferentes categorias de palavras, através do seu dicionário interno e classifica palavras em diversas categorias (ansiedade, raiva, afetividade, positivo, negativo, dentre outras) todas as categorias do LIWC estão listadas no apêndice, na tabela 4 - Lista de Atributos do LIWC). O software realiza esse processo realizando a tokenização, steamming e remoção de stop words para posterior contagem de palavras que estão associadas com o seu dicionário interno, também chamado LIWC.

O software realiza a contagem de palavras dentro do texto que encontra em seu dicionário, em seguida calcula o percentual de palavras que pertencem a cada categoria. A tabela 2 apresenta uma amostra da base de dados após o pré-processamento realizado com o LIWC, essa amostra possui apenas 5 linhas de um total de 20.800 registros e 10 atributos de um total de 28, considerando o atributo *label*, nosso atributo alvo. Na base podemos visualizar o percentual de palavras que pertencem a cada atributo que escolhemos previamente no LIWC.

Na linha 1 verificamos que esse texto possui 2,41% de suas palavras com a categoria *negate* esse valor é maior em comparação com os demais textos que aparecem na tabela 1, de acordo com a perspectiva da emoção, mentirosos sentem culpa e têm medo de serem descobertos mentindo, tendendo a utilizar mais emoções negativas no texto (LARCKER & ZAKOLYUKINA, 2012), desta forma, neste trabalho poderemos verificar se para textos de mídias sociais essa categoria *negate* pode realizar alguma influência no fato do texto ser *true news* ou *Fake News*.

Após a aplicação dos algoritmos de classificação, podemos comparar as diferenças apresentadas em falas, para a detecção de *Fake News*, onde o autor em tese possui maior tempo para elaborar uma mentira em comparação com uma conversa e poderemos verificar quais são os atributos que mais influenciam nesse tipo de discurso.

Tabela 2 - Amostra base de dados.

	label	**pronoum**	**ppron**	**i**	**we**	**prep**	**negate**	**affect**	**posemo**	**negemo**
0	1	19.75	13.03	4.50	1.87	9.71	0.93	5.31	3.78	1.28
1	0	19.40	11.80	5.65	2.21	12.13	2.41	5.11	4.09	0.97
2	1	10.39	5.44	1.65	1.17	13.77	1.68	4.89	3.15	1.66
3	1	13.23	7.56	2.00	1.03	12.73	1.82	4.60	2.71	1.82
4	1	12.56	7.01	2.03	1.01	13.79	1.55	4.83	2.99	1.82

Fonte: Próprio autor.

Após este pré-processamento da base de dados, obtivemos uma nova base de dados já com a contagem de atributos contidos no LIWC. O Software LIWC foi executado contando apenas categorias que estão relacionadas a premissa do artigo D3C2, isto é, estão baseadas em uma extensa revisão e síntese de literatura baseada no comportamento não verbal (VRIJ, 2008) que se fundamenta em teorias da psicologia: emoções, esforço cognitivo, tentativa de controle e falta de acolhimento. Após essa filtragem chega-se a 27 atributos encontrados no LIWC que atendem os critérios utilizados pelos autores do artigo D3C2, que estão listados no quadro 1.

Quadro 1 - Atributos do LIWC utilizados no pré-processamento.

Categoria	Descrição	Tipo
label	Assume 0 para true news e 1 para *Fake News*	Binário
pronoun	Pronome	Numérico
ppron	Pronome pessoal	Numérico
i	Primeira pessoa do singular	Numérico
we	Primeira pessoa do plural	Numérico
prep	Preposição	Numérico
negate	Negativas	Numérico
affect	Palavras afetivas	Numérico
posemo	Emoção positiva	Numérico
negemo	Emoções negativas	Numérico
anx	Ansiedade	Numérico
anger	Raiva	Numérico
sad	Tristeza	Numérico
social	Social	Numérico
family	Família	Numérico
friend	Amigos	Numérico
cause	Causa	Numérico

certain	Certeza	Numérico
feel	Sentimento	Numérico
power	Poder	Numérico
risk	Risco	Numérico
relativ	Relativo	Numérico
money	Dinheiro	Numérico
relig	Religião	Numérico
death	Morte	Numérico
informal	Informal	Numérico
swear	Xingamento	Numérico
assent	Consentimento	Numérico

Com os 27 atributos selecionados, realiza-se uma análise exploratória da base, com o objetivo de compreender o domínio das variáveis e pré-processamentos necessários, como a limpeza de dados, tratamento de valores ausentes e tratamento de dados ruidosos (DE CASTRO, 2016). Na tabela 3 foi realizada uma análise estatística para compreender melhor a base de dados, observando aspectos como medidas de tendência central e de frequência.

Tabela 3 - Atributos LIWC.

Atributo	Quantidade de Textos	Média	Desvio Padrão	Min.	25%	Mediana	75%	Max.
label	20800.0	0.500625	0.500012	0.0	0.00	1.000	1.00	1.00
pronoun	20800.0	8.428.563	4.000.604	0.0	6.08	8.110	10.51	50.00
ppron	20800.0	4.297.357	2.949.731	0.0	2.30	3.840	5.82	50.00
i	20800.0	0.654374	1.238.814	0.0	0.00	0.210	0.78	25.00
we	20800.0	0.690286	1.069.703	0.0	0.00	0.370	0.91	20.00
prep	20800.0	13.972.905	3.457.570	0.0	13.22	14.490	15.62	50.00
negate	20800.0	1.015.364	1.242.512	0.0	0.45	0.880	1.35	100.00
affect	20800.0	4.347.037	3.200.876	0.0	3.01	4.120	5.35	100.00
posemo	20800.0	2.312.049	2.621.934	0.0	1.33	2.070	2.92	100.00
negemo	20800.0	1.974.989	1.959.070	0.0	0.92	1.700	2.67	100.00
anx	20800.0	0.330682	0.646362	0.0	0.00	0.200	0.46	50.00
anger	20800.0	0.811285	1.465.824	0.0	0.10	0.500	1.12	100.00
sad	20800.0	0.281668	0.553384	0.0	0.00	0.170	0.39	33.33
social	20800.0	8.706.360	4.387.272	0.0	6.17	8.450	10.99	100.00

family	20800.0	0.242990	0.609540	0.0	0.00	0.000	0.25	25.00
friend	20800.0	0.196187	1.546.212	0.0	0.00	0.000	0.23	100.00
cause	20800.0	1.536.754	1.347.210	0.0	0.89	1.395	1.98	25.00
certain	20800.0	1.009.898	1.070.481	0.0	0.45	0.850	1.36	33.33
feel	20800.0	0.305711	0.925202	0.0	0.00	0.180	0.40	100.00
power	20800.0	4.095.402	2.200.585	0.0	2.64	3.980	5.38	33.33
risk	20800.0	0.668385	0.791222	0.0	0.18	0.530	0.93	33.33
relativ	20800.0	13.318.793	4.232.404	0.0	11.43	13.370	15.49	60.00
money	20800.0	1.090.219	1.581.851	0.0	0.15	0.560	1.33	25.00
relig	20800.0	0.395912	1.036.999	0.0	0.00	0.000	0.33	27.27
death	20800.0	0.290021	0.620873	0.0	0.00	0.000	0.32	16.67
informal	20800.0	0.553743	1.524.800	0.0	0.00	0.270	0.61	100.00
swear	20800.0	0.070294	0.863595	0.0	0.00	0.000	0.00	100.00
assent	20800.0	0.079201	0.793482	0.0	0.00	0.000	0.06	100.00

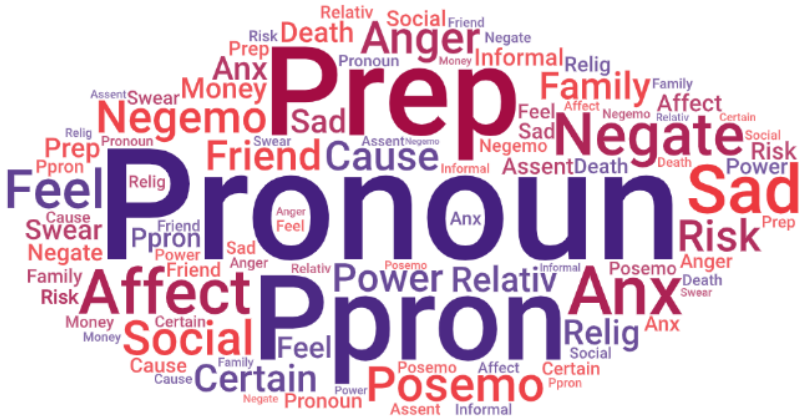
Fonte: Próprio Autor.

A partir da tabela 3 se inicia uma análise de existencia de textos em que todos os atributos apresentaram valor zero, isto é, após a aplicação do LIWC não foi obtida qualquer informação de nenhum atributo associado ao dicionário interno do LIWC esses textos com valores ausentes foram removidos da base de dados. Com a visualização fornecida pela tabela 2, também se identificou que alguns atributos apresentam outliers com 100% do texto em apenas um único atributo. Um segundo tratamento para remover outliers foi necessário para verificar casos em que a ocorrência era maior de 20% do texto em um único atributo. Após o pré-processamento, a base de dados passou de 20.800 textos para 20.552 textos.

Após o tratamento da base de dados com a remoção de outliers foi gerada uma nuvem de palavras, Figura 5, com os atributos do LIWC. Considerou-se para a criação do gráfico a média dos atributos distribuídos nos 20.552 textos.

A nuvem de palavras obtida na figura 5 permite uma visualização rápida dos atributos que mais aparecem nos textos, como esperado: preposições, pronomes pessoais e pronomes são os atributos que mais aparecem nos textos analisados, e atributos como death, Money aparecem em menor quantidade nos textos.

Figura 5 - Nuvem de palavras frequência dos atributos LIWC.



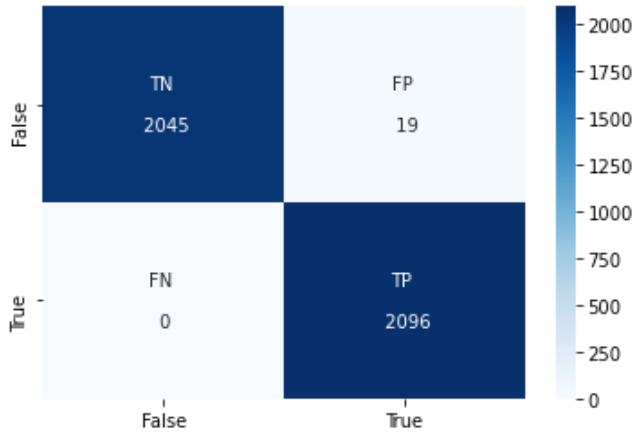
Fonte: Próprio autor.

Após o pré-processamento da base de dados foi possível prosseguir com a aplicação dos algoritmos de aprendizagem de máquina na base de dados. Inicialmente explorou-se o algortimo SVM (Support Vector Machinne), usano o *label com* valores de zero (0) quanto o texto é *True News* ou um (1) quando o texto é *Fake News*.

O algoritmo foi implementando através da linguagem de programação python, utilizando a biblioteca sklearn.svm (SCIKIT-LEARN, 2022). A parametrização escolhida como kernel do algoritmo foi Radial Basis Function (RBF). A base de dados de *Fake News* já com o seu pré-processamento foi dividida em treino e teste. Num primeiro experimento a base foi dividida em 80% de treino e 20% como teste. A acurácia alcançada foi de 0,996.

Em outro experimento, com o objetivo de validar os resultados de classificação aplicou-se o método de validação cruzada, com quantidade de pastas igual a dez. O resultado obtido foi de 0,999, confirmando a generalização do algoritmo. Na figura 6 apresenta-se a matriz de confusão e resultante da validação cruzada. No resultado obtido pela matriz de confusão, pode-se verificar que existe um equilíbrio na base de dados e no algoritmo de classificação, porque a classificação em falso negativo e verdadeiro positivo apresentaram valores muito próximos, isto é, o algoritmo não apresentou um viés de classificação enquadrando maior porte dos valores em um único quadrante.

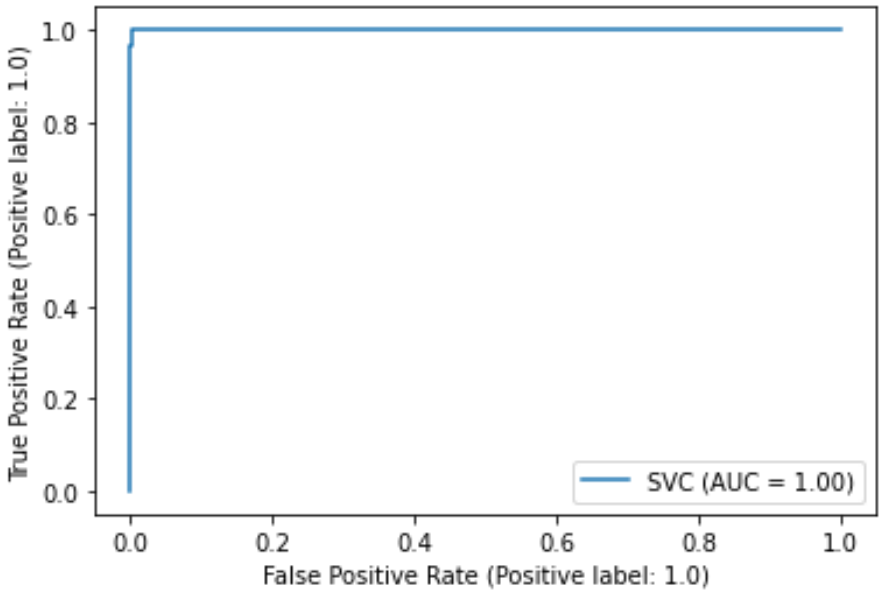
Figura 6 - Matriz de Confusão SVM.



Fonte: Próprio autor.

Além da matriz de confusão, o resultado também será apresentado usando a curva ROC (Figura 7) , para auxiliar na visualização do desempenho de um algoritmo de classificação. No eixo Y são exibidos os valores da taxa de verdadeiro positivo e no eixo X a taxa de falso positivo. A interpretação é que quanto mais a curva se aproxima de 1 melhor é a capacidade preditiva do modelo e valores menores do 0,5 significam que o modelo possui baixa capacidade preditiva. Com 50% se tem a mesma capacidade preditiva do que um palpite aleatório. Na figura 7 verifica-se que o algoritmo apresenta um resultado preditivo satisfatório, porque possui o valor AUC igual a 1. Um ponto que merece atenção neste resultado é verificar se o modelo não possui um *overfiting* e pode apresentar problema com outras bases de dados.

Figura 7 - Curva Roc

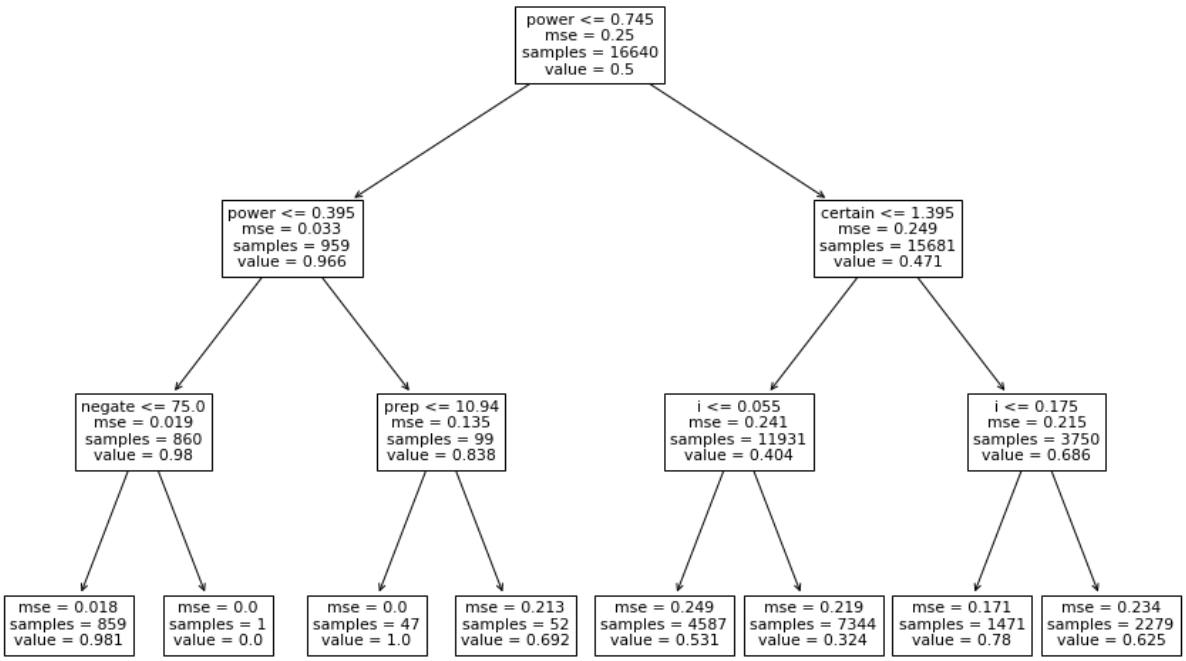


Fonte: Próprio Autor

Outro experimento de classificação foi realizado com o algoritmo Árvore de Decisão. O algoritmo foi implementado através da biblioteca scikitlearn na linguagem de programação python, utilizando a mesma proporção do experimento anterior, ou seja, 80% da base como treinoe 20% da base com teste.

Na figura 8 apresenta-se o resultado da aplicação do algoritmo, como o parâmetro max_depth, que representa a profundidade máxima da árvore que será gerada igual a três. Com o aumento na profundidade fica cada vez mais difícil enxergar a árvore gerada.

Figura 8 - Árvore de Decisão Fake News.



Fonte: Próprio Autor

O Algoritmo de árvore de decisão foi aplicado devido a sua visualização que permite identificarquais atributos exercem maior influência na classificação em *True News* ou *Fake News*. A partir da visualização gerada na figura 8, verifica-se que para o algoritmo de árvore de decisão, o atributo *power* é o fator de maior influência para a detecção de uma *Fake News* dada a amostra que foi utilizada. Nota-se que das 27 categorias do LIWC utilizadas 4 possuem maior relevância para a classificação de uma *Fake News*, são eles: *power*, *certain*, *negate*, *prep* e *i* (pronome pessoal).

Este resultado corrobora as quatro perspectivas seguidas como base no estudo D3C2, isto é, mentirosos tendem a ser mais negativos e procuram retirar a primeira pessoa por trazer detalhes que podem comprometer a veracidade do estudo e tendem a não ter convicção por não terem vivenciado o fato que estão narrando, mesmo em notícias, onde existe um tempo maior para preparar a mentira.

O primeiro atributo que possui grande influência em determinar se temos uma *Fake News* é o atributo *power*, o modelo identifica que para valores menores do que 0,745 existe um conjunto de 959 amostras que possui uma grande probabilidade de serem *Fake News*. A segunda categoria do LIWC que apresenta relevância é a *certain*, onde para valores maiores que 1,395 exista uma maior probabilidade de o texto ser verdadeiro em comparação com valores menores que 1,395 existe uma maior probabilidade de estarmos tratando de uma notícia falsa. Um atributo que não foi identificado com grande relevância no artigo D3C2, neste artigo apresenta maior relevância é o atributo power, isso pode ter ocorrido devido a diferença ambientes que foram realizados os estudos o artigo D3C2 utilizou conversas de conferências enquanto este artigo utilizou notícias em mídias sociais.

4 CONCLUSÃO E TRABALHOS FUTUROS

A disseminação de discursos enganosos, intensificou-se nos últimos anos conforme verificou-se em eventos como o Brexit e as eleições para presidente dos Estados Unidos de Donald Trump em 2016 (BASTOS, 2019). O estudo de técnicas de detecção de discursos enganosos é fundamental para identificar e combater a desinformação que está presente em diversas mídias, principalmente nas mídias sociais (BASTOS & MERCEA, 2019).

Este trabalho propôs adaptar o método de (LARCKER & ZAKOLYUKINA, 2012) ao contexto de identificação de *Fake News*. Essa adaptação foi realizada por técnicas de mineração de textos e análise de sentimentos para a detecção de *Fake News*. Esse processo foi realizado através da utilização do LIWC em conjunto com algoritmos de aprendizado de máquina que resultaram no método que chamamos de *sentimentum*.

O *sentimentum* é um método de identificação de *Fake News*, que realiza a contagem de palavras associadas as categorias previamente selecionadas do dicionário interno do LIWC. As categorias selecionadas do LIWC fundamentam-se na revisão de literatura baseada em (Vrij, 2008). Após a contagem de palavras o método utiliza algoritmos de classificação como o SVM e Árvore de Decisão para classificar se um texto é *Fake News* ou *true news*.

A pesquisa apresentou resultados uma acurácia satisfatória no contexto de detecção de *Fake News*, um exemplo que justifica essa afirmação é a revisão de literatura sistemática elaborada por (MEDEIROS & BRAGA, 2020). Neste estudo os autores realizaram uma revisão sistemática de literatura com artigos de detecção de *Fake News*. Um dos resultados apresentado pelos autores é a compilação dos diferentes métodos de detecção de discursos enganosos identificados na literatura.

De acordo com essa compilação o resultado obtido neste estudo apresenta um resultado satisfatório porque comparado com outros estudos de detecção de *Fake News* o melhor valor para acurácia que foi identificado era de 0,920, neste estudo atingimos uma acurácia de 0,996 para o algoritmo SVM, dentro do contexto de detecção de *Fake News*.

Um segundo aspecto da pesquisa que vale ser ressaltado é a relevância dos atributos identificados no LIWC em comparação com as premissas utilizadas pelos autores do artigo D3C2 para seleção dos atributos do LIWC. Para os resultados obtidos na pesquisa através da Árvore de decisão verificamos que as premissas apresentadas em D3C2 são observadas no contexto de identificação de *Fake News*, isto é, os atributos *Negate* e *I* (pronome pessoal “Eu”,

em inglês). O atributo *Negate* representa palavras negativas no texto e o atributo *I* que representa o uso de palavras do pronome pessoal, são atributos extremamente relevantes para a classificação de *Fake News* e *True News*.

Para a técnica proposta foi necessário realizar um descarte de 248 notícias, após o pré-processamento da base de dados este descarte ocorreu devido a estes textos não terem sido enquadrados em categorias do LIWC ou apresentarem outliers com apenas uma categoria do LIWC identificada em todo o texto. Para trabalhos futuros seria possível ampliar a quantidade de categorias utilizadas pelo LIWC ampliando os estudo sobre categorias efetivas para a identificação de *fake news*.

Uma sugestão de trabalhos futuros seria aplicar o método *Sentimentum* utilizando uma base de dados em português e o dicionário LIWC em português disponibilizado de forma gratuita por (Aluisio, Checchia, & Chishman, 2022). Este processo envolveria realizar o método de tokenização, lematização e contagem de palavras que são realizados pelo software LIWC para o dicionário LIWC em português, além de testar o método com outros algoritmos ao invés de Árvores de Decisão e SVM, que foram utilizadas como algoritmos de aprendizagem profunda como redes neurais convolucionais e redes neurais recorrentes.

Outra sugestão de trabalhos futuros, seria utilizar técnicas que captam parte do discurso como POS (*Part of Speech*), isso porque um dos pontos fracos da utilização da técnica "*bag of words*" é que ela não realiza a análise semântica do texto, podendo trazer imprecisões nas análises que são feitas individualmente, palavra a palavra.

Por fim, uma última sugestão para trabalhos futuros seria a aplicação do modelo para uma maior quantidade de base de dados de textos de fake news, apesar da acurácia apresentar um resultado satisfatório neste modelo, seria importante verificarmos a capacidade de generalização do modelo aplicado a diversas base de dados, realizando uma média para verificar se existe alguma alteração nos atributos que possuem maior relevância na detecção de *fake news*.

REFERENCIAS BIBLIOGRÁFICAS

Aluisio, S., Checchia, R., & Chishman, R. (04 de Abril de 2022). *PortLex.* Fonte: LIWC: http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc

BALAGE FILHO, P. P. (2013). Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology.

BASTOS, M. T. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social science computer review*.

CARDOSO DURIER DA SILVA, F. V. (2019). Can machines learn to detect fake news? a survey focused on social media. *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

DE CASTRO, L. N. (2016). Introdução a mineração de dados (1ª edição ed.). São Paulo, SP, Brasil: Saraiva Educação SA.

FERREIRA, M., & LOPES, M. (2021). *Para conhecer linguística computacional, (1ª ed., Vol. I). (J. Pinsky, Ed.).* São Paulo: J. Contexto.

FOLHA. (11 de Junho de 2023). *Folha de São Paulo*. Fonte: https://www1.folha.uol.com.br/colunas/flavia-lima-ombudsman/2020/10/o-patrimonio-de-r-579-de-boulos.shtml

FOLLADOR, K. J. (2017). A relação entre a peste negra e os judeus. *Vértices, n. 20, p. 26-46*.

FREIRE, P. M., DA SILVA, F. R., & GOLDSCHMIDT, R. R. (30 de Novembro de 2021). Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. *Expert Systems with Applications, p. 115414.*

GLOSBE. (17 de 06 de 2023). Fonte: Latim-Português Dicionário: https://pt.glosbe.com/la/pt/corpus

GUSMÃO, C., FIGUEIREDO, K., & BRITO, W. A. (2021). Técnicas de Processamento de Linguagem Natural em Denúncias Criminais: Automatização e Classificação de Texto em português Coloquial. *Anais do XLVIII Seminário Integrado de Software e Hardware (pp. 172-182)*. Rio de Janeiro, RJ: SBC.

KACZMAREK, I. E. (2021). A machine learning approach for integration of spatial development plans based on natural language processing. *Sustainable Cities and Society, 103479.*

KAGGLE, B. a. (04 de Novembro de 2021). *Build a system to identify unreliable news articles*. Fonte: KAGGLE: https://www.kaggle.com/c/fake-news/data

KAPLAN, A. (2020). Artificial intelligence, social media, and fake news: Is this the end of democracy? MEDIA & SOCIETY, 149.

LARCKER, D. F., & ZAKOLYUKINA, A. A. (2012). Detecting deceptive discussions in

conference calls. *Journal of Accounting Research*, pp. 495-540.

MEDEIROS, F. D., & BRAGA, R. B. (2020). Fake News Detection in Social Media: A Systematic Review. A Systematic Review. *XVI Brazilian Symposium on Information Systems*, pp. 1-8.

Meltwater. (09 de Junho de 2023). *datareportal*. Fonte: We are Social: https://datareportal.com/reports/digital-2023-global-overview-report

MORAIS, E. A., & AMBRÓSIO, A. P. (2007). Mineração de textos. p. 30.

OSHIKAWA, R., & WANG, W. Y. (2018). A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

OTTER, D. W., MEDINA, J. R., & KALITA, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 604-624.

PENNEBAKER, J. W. (2015). The development and psychometric properties of LIWC2015. 22.

RAJAN, A., SALGAONKAR, A., & JOSHI, R. (2020). A survey of Konkani NLP resources. Computer Science Review, 38, 100299.

RECUERO, R., & SOARES, F. (2021). O Discurso Desinformativo sobre a Cura do COVID-19 no Twitter: Estudo de caso.

SANTOS, G. F. (2021). Social media, disinformation, and regulation of the electoral process: a study based on 2018 Brazilian election experience. *Revista de Investigações Constitucionais*, pp. 429-449.

SCIKIT-LEARN. (04 de Julho de 2022). *scikit-learn*. Fonte: scikit-learn: https://scikit-learn.org/stable/

SOUZA, V. d. (2022). Sentimentum: A Method of Detecting Fake News. *3rd International Conference On Computational Intelligence - ICCI 2022*.

VRIJ, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities.* John Wiley & Sons.

WARDLE, C., & DERAKHSHAN, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe.*

WIVES, L. K. (2002). Tecnologias de Descoberta de Conhecimento em Textos Aplicadas à Inteligência Competitiva. *Tese de Doutorado.*

ZAFARANI, R., ABBASI, M., & LIU, H. (2014). *Social media mining: an introduction (1ª ed., Vol. I). (L. Cowles, Ed.).* New York: Cambridge University Press.

ZHANG, X., & GHORBANI, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management.*

ZHOU, X., & ZAFARANI, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. pp. 1-40.

APÊNDICE

Tabela 4 - Lista de Atributos do LIWC

Category Linguistic Process	Abbrev	Examples	Words in category	Validity (judges)	Alpha: Binary/raw
Word count	wc				
words/sentence	wps				
Dictionary words	dic				
Words>6 letters	sixltr				
Total function words	funct		464		.97/.40
Total pronouns	pronoun	I, them, itself	116		.91/.38
Personal pronouns	ppron	I, them, her	70		.88/.20
1st pers singular	i	I, me, mine	12	.52	.62/.44
1st pers plural	we	We, us, our	12		.66/.47
2nd person	you	You, your, thou	20		.73/.34
3rd pers singular	shehe	She, her, him	17		.75/.52
3rd pers plural	they	They, their, they'd	10		.50/.36
Impersonal pronouns	ipron	It, it's, those	46		.78/.46
Articles	article	A, an, the	3		.14/.14
[Common verbs]	verb	Walk, went, see	383		.97/.42
Auxiliary verbs	auxverb	Am, will, have	144		.91/.23
Past tense a	past	Went, ran, had	145	.79	.94/.75
Present tense a	present	Is, does, hear	169		.91/.74
Future tense a	future	Will, gonna	48		.75/.02
Adverbs	adverb	Very, really, quickly	69		.84/.48
Prepositions	prep	To, with, above	60		.88/.35
Conjunctions	conj	And, but, whereas	28		.70/.21
Negations	negate	No, not, never	57		.80/.28
Quantifiers	quant	Few, many, much	89		.88/.12
Numbers	number	Second, thousand	34		.87/.61
Swear words	swear	Damn, piss, fuck	53		.65/.48
Psychological Processes					
Social processesb	social	Mate, talk, they, child	455		.97/.59
Family	family	Daughter, husband, aunt	64	.87	.81/.65
Friends	friend	Buddy, friend, neighbor	37	.70	.53/.12

Humans	human	Adult, baby, boy	61		.86/.26
Affective processes	affect	Happy, cried, abandon	915		.97/.36
Positive emotion	posemo	Love, nice, sweet	406	.41	.97/.40
Negative emotion	negemo	Hurt, ugly, nasty	499	.31	.97/.61
Anxiety	anx	Worried, fearful, nervous	91	.38	.89/.33
Anger	anger	Hate, kill, annoyed	184	.22	.92/.55
Sadness	sad	Crying, grief, sad	101	.07	.91/.45
Cognitive processes	cogmech	cause, know, ought	730		.97/.37
Insight	insight	think, know, consider	195		.94/.51
Causation	cause	because, effect, hence	108	.44	.88/.26
Discrepancy	discrep	should, would, could	76	.21	.80/.28
Tentative	tentat	maybe, perhaps, guess	155		.87/.13
Certainty	certain	always, never	83		.85/.29
Inhibition	inhib	block, constrain, stop	111		.91/.20
Inclusive	incl	And, with, include	18		.66/.32
Category	Abbrev				
Examples	Words in category	Validity			
(judges)	Alpha:				
Binary/raw					
Exclusive	excl	But, without, exclude	17		.67/.47
Perceptual processes	percept	Observing, heard, feeling	273		.96/.43
See	see	View, saw, seen	72		.90/.43
Hear	hear	Listen, hearing	51		.89/.37
Feel	feel	Feels, touch	75		.88/.26
Biological processes	bio	Eat, blood, pain	567	.53	.95/.53
Body	body	Cheek, hands, spit	180		.93/.45
Health	health	Clinic, flu, pill	236		.85/.38
Sexual	sexual	Horny, love, incest	96		.69/.34
Ingestion	ingest	Dish, eat, pizza	111		.86/.68
Relativity	relativ	Area, bend, exit, stop	638		.98/.51
Motion	motion	Arrive, car, go	168		.96/.41

Space	space	Down, in, thin	220	.96/.44
Time	time	End, until, season	239	.94/.58
Personal Concerns				
Work	work	Job, majors, xerox	327	.91/.69
Achievement	achieve	Earn, hero, win	186	.93/.37
Leisure	leisure	Cook, chat, movie	229	.88/.50
Home	home	Apartment, kitchen, family	93	.81/.57
Money	money	Audit, cash, owe	173	.90/.53
Religion	relig	Altar, church, mosque	159	.91/.53
Death	death	Bury, coffin, kill	62	.86/.40
Spoken categories				
Assent	assent	Agree, OK, yes	30	.59/.41
Nonfluencies	nonflu	Er, hm, umm	8	.28/.23
Fillers	filler	Blah, Imean, youknow	9	.63/.18

Fonte: **(PENNEBAKER, FRANCIS, & BOOTH, 2001)**