MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO

FLÁVIO ROBERTO MATIAS DA SILVA

# FAKENEWSSETGEN: UM PROCESSO PARA CONSTRUÇÃO DE *DATASETS* QUE VIABILIZEM A COMPARAÇÃO ENTRE MÉTODOS DE DETECÇÃO DE *FAKE NEWS* BASEADOS EM DIFERENTES DEMANDAS DE INFORMAÇÃO.

RIO DE JANEIRO
2020

FLÁVIO ROBERTO MATIAS DA SILVA

# FAKENEWSSETGEN: UM PROCESSO PARA CONSTRUÇÃO DE *DATASETS* QUE VIABILIZEM A COMPARAÇÃO ENTRE MÉTODOS DE DETECÇÃO DE *FAKE NEWS* BASEADOS EM DIFERENTES DEMANDAS DE INFORMAÇÃO.

Dissertação apresentada ao Programa de Pós-graduação em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências em Sistemas e Computação.

Orientador(es): Ronaldo Ribeiro Goldschmidt, D.Sc.

Rio de Janeiro

2020

Silva, Flávio Roberto Matias da.
FakeNewsSetGen: um processo para construção de *datasets* que viabilizem a comparação entre métodos de detecção de *Fake News* baseados em diferentes demandas de informação. / Flávio Roberto Matias da Silva. – Rio de Janeiro, 2020.
80 f.

Orientador(es): Ronaldo Ribeiro Goldschmidt.

Dissertação (mestrado) – Instituto Militar de Engenharia, Sistemas e Computação, 2020.

1. fake news. 2. aprendizado de máquina. 3. dataset. 4. métodos de detecção. 5. redes sociais. i. Goldschmidt, Ronaldo Ribeiro (orient.) ii. Título

**FLÁVIO ROBERTO MATIAS DA SILVA**

# FakeNewsSetGen: um processo para construção de *datasets* que viabilizem a comparação entre métodos de detecção de *Fake News* baseados em diferentes demandas de informação.

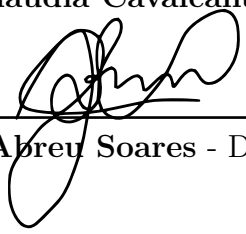
Dissertação apresentada ao Programa de Pós-graduação em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências em Sistemas e Computação.

Orientador(es): Ronaldo Ribeiro Goldschmidt.

Aprovado em Rio de Janeiro, 04 de dezembro de 2020, pela seguinte banca examinadora:

Prof. **Ronaldo Ribeiro Goldschmidt** - D.Sc. do IME - Presidente

Prof. **Maria Claudia Cavalcanti** – D.Sc. do IME



Prof. **Jorge de Abreu Soares** - D.Sc. do CEFET-RJ

Rio de Janeiro

2020

*À Marinha do Brasil pela formação e pela oportunidade de crescimento pessoal e profissional.*

# AGRADECIMENTOS

Agradeço a Deus por estar presente em todos os momentos da minha vida, a minha querida família e a cada um dos companheiros deste curso, que trilharam esta jornada comigo.

Agradeço ainda aos professores que me incentivaram, em especial ao meu orientador Ronaldo Ribeiro Goldschmidt e ao professor Paulo Márcio Souza Freire, pela imensa paciência, compreensão, apoio e atenção que dedicaram a mim.

Por fim, agradeço ao Instituto Militar de Engenharia pelo apoio proporcionado.

*"Pois o Senhor é quem dá a sabedoria;*
*de sua boca procedem o conhecimento e o discernimento.*
*(Bíblia Sagrada, Provérbios 2, 6)*

# RESUMO

Devido ao fácil acesso e ao baixo custo, o consumo de notícias *on-line* em redes sociais aumentou significativamente na última década. Apesar de seus benefícios, algumas redes sociais permitem que qualquer pessoa divulgue notícias com intenso poder de difusão, o que amplia um problema antigo: a disseminação do *fake news* (i.e., notícias falsas veiculadas de forma intencional). A proliferação de *fake news*, geralmente, afeta não apenas a integridade jornalística, mas também perturba as áreas social, política, econômica, cultural, assim como da saúde e segurança. Diante desse cenário, foram propostos vários métodos baseados em aprendizado de máquina para detectar automaticamente *fake news* (*machine learning-based methods to automatically detect fake news*- MLFN). Esses métodos necessitam de *datasets* para treinar e avaliar seus modelos de detecção. Embora os MLFN recentes tenham sido projetados para considerar dados sobre a propagação de notícias em redes sociais, poucos dos *datasets* disponíveis contêm esses dados. Assim, a comparação de desempenho entre MLFN está restrita à utilização de um número limitado de *datasets*. Além disso, os *datasets* existentes com dados de propagação não contêm notícias em português, o que prejudica a avaliação do MLFN nesse idioma. Portanto, este trabalho propõe o *FakeNewsSetGen*, um processo de construção de *datasets* para o estudo de *fake news* que contenham dados de propagação de notícias e viabilizem a comparação entre MLFN. O processo de engenharia de software do *FakeNewsSetGen* foi orientado para incluir todos os tipos de dados exigidos pelos MLFN existentes. Para ilustrar a viabilidade e adequação do *FakeNewsSetGen*, foi realizado um estudo de caso que abrange a implementação de um protótipo do *FakeNewsSetGen* e a aplicação desse protótipo para criar uma instância de *dataset* denominada *FakeNewsSet*, composta de notícias em português. Dez MLFN com diferentes tipos de requisitos de dados (sete deles exigindo dados de propagação de notícias) foram aplicados ao *FakeNewsSet* e comparados, demonstrando o potencial de utilização do processo proposto e do *dataset* criado.

**Palavras-chave**: fake news. aprendizado de máquina. dataset. métodos de detecção. redes sociais.

# ABSTRACT

Due to easy access and low cost, social media online news consumption has increased significantly for the last decade. Despite their benefits, some social media allow anyone to post news with intense spreading power, which amplifies an old problem: the dissemination of fake news (ie., false information that is spread deliberately to deceive). The proliferation of fake news generally affects not only journalistic integrity, but also disrupts social, political, economic, cultural, as well as health and safety. In the face of this scenario, several machine learning-based methods to automatically detect fake news (MLFN) have been proposed. All of them require fake news datasets to train and evaluate their detection models. Although recent MLFN were designed to consider data regarding the news propagation on social media, most of the few available fake news datasets do not contain this kind of data. Hence, comparing the performances amid those recent MLFN and the others is restricted to a very limited number of datasets. Moreover, all existing datasets with propagation data do not contain news in Portuguese, which impairs the evaluation of the MLFN in this language. Thus, this work proposes *FakeNewsSetGen*, a process that builds fake news datasets that contain news propagation data and support comparison amid the state-of-the-art MLFN. *FakeNewsSetGen*'s software engineering process was guided to include all kind of data required by the existing MLFN. In order to illustrate *FakeNewsSetGen*'s viability and adequacy, a case study was carried out. It encompassed the implementation of a *FakeNewsSetGen* prototype and the application of this prototype to create a dataset called *FakeNewsSet*, with news in Portuguese. Ten MLFN with different kind of data requirements (seven of them demanding news propagation data) were applied to *FakeNewsSet* and compared, demonstrating the potential use of both the proposed process and the created dataset.

**Keywords**: fake news. machine learning. dataset. detection methods. social media.

# LISTA DE ILUSTRAÇÕES

# LISTA DE QUADROS

# LISTA DE TABELAS

# LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| MLFN | **M**achine **L**earning-based methods to automatically detect **F**ake **N**ews |
| MDDN | **M**eios **D**igitais de **D**ivulgação de **N**otícias |
| RSV | **R**edes **S**ociais **V**irtuais |
| COVID | **CO**rona **V**irus **D**isease |
| AM | **A**prendizado de **M**áquina |
| ICS | **I**mplicit **C**rowd **S**ignals |
| HCS | **H**ybrid **C**rowd **S**ignals |
| POS | **P**art **O**f **S**peech |
| 3PFC | **Third-P**arty **F**act-**C**hecking **P**rogram |
| RI | **R**ecuperação de **I**nformação |
| QE | **Q**uery **E**xpansion |
| UML | **U**nified **M**odeling **L**anguage |
| URL | **U**niform **R**esource **L**ocator |
| RSS | **R**DF **S**ite Summary or Really **S**imple Syndication |
| API | **A**pplication **P**rogramming **I**nterface |
| SVM | **S**upport **V**ector **M**achine |
| JED | **J**ogos **E**ducacionais **D**igitais |

# LISTA DE SÍMBOLOS

| | |
|---|---|
| $S$ | Conjunto de fontes fidedignas a serem consultadas em busca de notícias fake e não fake previamente rotuladas |
| $s$ | Uma determinada fonte pertencente a $S$ |
| $s.l$ | Atributo do endereço do *feed* de notícias de $s$ |
| $URLs$ | Conjunto de endereços de notícias previamente rotuladas por $s$ |
| $N^R$ | Conjunto de notícias rotuladas |
| $n^r$ | Uma determinada notícia pertencente a $N^R$ |
| $n^r.text$ | Atributo que representa o texto de $n^r$ |
| $t$ | Intervalo de tempo a ser utilizado para restringir a busca por divulgações |
| $query$ | Conjunto de palavras-chave utilizado, juntamente com $t$, para identificar divulgações de $n^r$ na rede social |
| $ND$ | Conjunto de associações entre os identificadores de notícias e de divulgações |
| $nd$ | Uma determinada associação de $ND$ |
| $nd.urlDiv$ | Endereço da divulgação de $n^r$ |
| $D$ | Conjunto de divulgações relacionadas a cada $nd$ de $ND$ |
| $d$ | Uma determinada divulgação pertencente a $D$ |
| $d.mddn$ | Atributo que indica o origem da divulgação |
| $rsv$ | Rede Social Virtual |
| $mv$ | Mídia Virtual |
| $U$ | Conjunto de usuários divulgadores de $ND$, incluindo os seus respectivos seguidores e seguidos |

# SUMÁRIO

# 1 INTRODUÇÃO

## 1.1 Contextualização e Motivação

Historicamente, a publicação de notícias se restringia à mídia tradicional, como rádio, TV, revistas e jornais impressos. Todavia, devido ao fácil acesso e baixo custo, as pessoas vêm, a cada dia, aumentando o consumo de notícias on-line [VOSOUGHI; MOHSENVAND; ROY, 2017]. Este consumo é realizado por intermédio dos meios digitais de divulgação de notícias (MDDN), compostos, basicamente, pelas mídias virtuais (e.g.: jornais e revistas *on-line*) e, em especial, pelas redes sociais virtuais (RSV).

Na Figura 1 podemos observar que cerca de 13,8% de uma amostra da população americana se informou sobre as eleições presidenciais por meio de redes sociais.

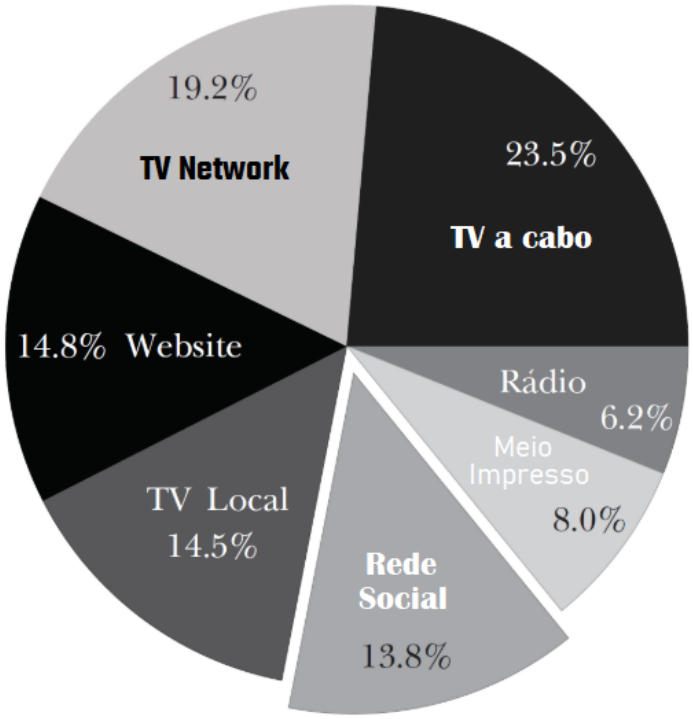



Figura 1 – Principais fontes das eleições presidenciais americanas em 2016. Fonte: Allcott e Gentzkow [2017]

Apesar de seus benefícios, algumas RSV permitem que qualquer pessoa, independentemente de sua credibilidade, divulgue (publique/propague) notícias com intenso poder de espalhamento [SHU et al., 2017; WANG et al., 2018]. Tal permissividade amplificou um problema antigo: a disseminação de notícias falsas [CONROY; RUBIN; CHEN, 2015; ZHANG et al., 2018]. Esse problema pode se tornar ainda mais prejudicial quando essas notícias são divulgadas intencionalmente, pois uma inverdade deliberada tende a ser melhor

elaborada e, portanto, mais eficaz em seu objetivo principal, que é influenciar na mudança de opinião [FREIRE; GOLDSCHMIDT, 2019b]. Esse tipo particular de notícia falsa, cuja divulgação acontece de forma proposital, é denominado de *fake news* [SHU et al., 2017].

Abaixo, seguem exemplos de casos significativos que ilustram o impacto que têm as *fake news* e a importância de se combater a sua disseminação:

- Nos três meses finais das eleições presidenciais americanas de 2016, as notícias falsas publicadas na rede social *Facebook*, que favoreceram qualquer um dos dois candidatos, foram compartilhadas 37 milhões de vezes [FARAJTABAR et al., 2017]; Uma análise do *Buzzfeed News*[1] mostra que, a partir das 20 principais notícias falsas sobre essas eleições, criadas por sites fraudulentos, foram geradas quase 1,5 milhões de atividades de engajamento (i.e.: interações de usuários a partir de uma determinada notícia) no *Facebook*;

- Um homem, carregando um rifle AR-15, aterrorizou os frequentadores de uma pizzaria em *Washington*, porque ele havia lido uma notícia falsa *on-line*, afirmando que o referido estabelecimento usava crianças como escravas sexuais [FARAJTABAR et al., 2017];

- Segundo o jornal *on-line G1*[2], uma dona de casa morreu dois dias após ter sido espancada por dezenas de moradores do município de Guarujá em São Paulo. De acordo com o jornal, ela foi agredida a partir de uma notícia falsa, divulgada por uma rede social, que atribuía à dona de casa o sequestro de crianças para utilizá-las em rituais de magia negra;

- As inúmeras noticias falsas sobre a *COVID-19* têm dificultado de forma significativa o esclarecimento da população brasileira sobre a disseminação da pandemia e a respeito das devidas medidas de enfrentamento da doença [MEJOVA; KALIMERI, 2020].

De acordo com os exemplos acima apresentados, casos relacionados às *fake news* não se limitam aos EUA. Dessa forma, surgem diferentes iniciativas da industria para combater esse tipo de notícia. Por exemplo, conforme divulgado pela *BBC News*[3], após notícias falsas terem, supostamente, levado a linchamentos em 2018 na Índia, o *WhatsApp* anunciou um limitador para a quantidade de encaminhamentos de mensagem.

[1] https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

[2] http://g1.globo.com/sp/santos-regiao/noticia/2014/05/mulher-espancada-apos-boatos-em-rede-social-morre-em-guaruja-sp.html

[3] <https://www.bbc.com/portuguese/salasocial-44897990#:~:text=WhatsApp%20limita%20mensagens%20na%20%C3%8Dndia%20ap%C3%B3s%20not%C3%ADcias%20falsas%20levarem%20a%20linchamentos,-20%20julho%202018&text=O%20WhatsApp%20anunciou%20que%20vai,informa%C3%A7%C3%B5es%20falsas%20em%20sua%20plataforma>

Infelizmente, o Brasil também apresenta acontecimentos vinculados às *fake news* e esses eventos estão se tornando cada vez mais frequentes, especialmente quando se trata de assuntos relacionados à política. Com essa crescente incidência desse tipo de notícia no país, o interesse da sociedade brasileira pelo assunto também aumenta. De acordo com extrato obtido por meio do *Google Trends*[4], conforme apresentado nas Figuras 2 e 3, as cidades brasileiras estão entre as primeiras posições do *ranking* de interesse pelo tópico *fake news*.


Figura 2 – Análise por região. Fonte: https://trends.google.com.br/trends

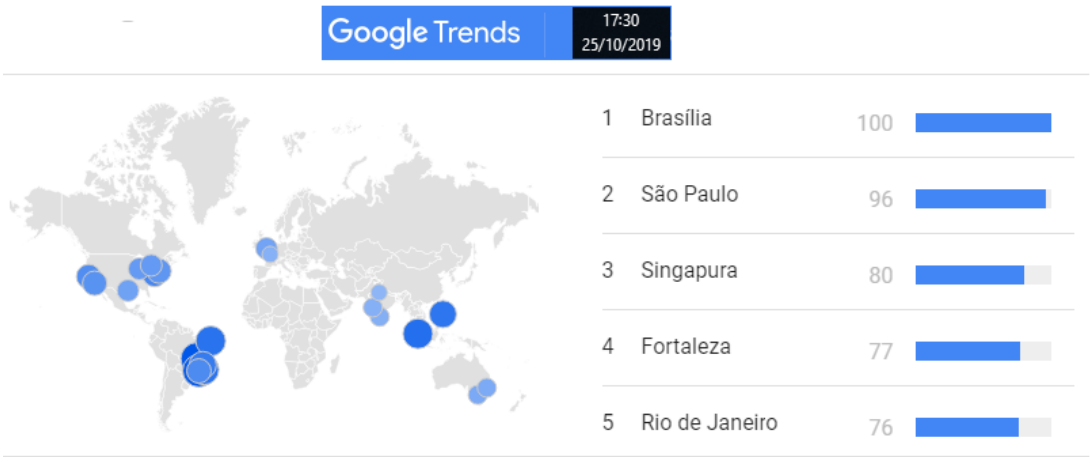



Figura 3 – Análise por cidade. Fonte: https://trends.google.com.br/trends

## 1.2 Problema de Pesquisa

Diante desse cenário, tanto a academia quanto a indústria têm pesquisado como combater *fake news* disponibilizadas em MDDN, em especial nas redes sociais [FREIRE; GOLDSCHMIDT, 2019b]. Contudo, esse problema apresenta-se como não trivial, tanto

[4] Ferramenta que viabiliza a analise da popularidade das principais consultas no Google.

pelo volume de publicações quanto pela velocidade das suas respectivas propagações. Assim, o emprego de abordagens computacionais, devido à sua maior velocidade de atuação, vem se destacando no combate às *fake news* [RUCHANSKY; SEO; LIU, 2017].

Grande parte dos métodos pertencentes a essas abordagens utiliza aprendizado de máquina para a classificação de notícias em *fake* e não *fake* [FREIRE; GOLDSCHMIDT, 2019b]. Esses métodos baseados em aprendizado de máquina para detectar automaticamente *fake news* (*machine learning-based methods to automatically detect fake news* - MLFN) precisam treinar, validar e testar modelos a partir de *datasets*.

Considerável parcela de trabalhos envolvendo MLFN utiliza *datasets* que são construídos para atender a um conjunto específico de métodos. Pesquisas envolvendo MLFN relacionados à análise do texto de notícias, por exemplo, tendem a utilizar *datasets* que contenham apenas informações textuais. Embora os métodos recentes tenham sido projetados para considerar dados sobre a propagação de notícias em redes sociais, poucos dos *datasets* disponíveis contêm esses dados [FREIRE; GOLDSCHMIDT, 2019b].

Assim, a comparação de desempenho entre esses MLFN está restrita à utilização de um número limitado de *datasets*. Essa comparação torna-se ainda mais difícil caso seja realizada entre métodos de diferentes abordagens de detecção, os quais demandam distintos tipos de dados. Além disso, até onde foi observado, os *datasets* existentes com dados de propagação não contêm notícias em português, o que prejudica a avaliação do MLFN nesse idioma.

## 1.3 Objetivo

Diante do exposto, o presente trabalho tem como objetivo propor um processo de construção de *datasets* com notícias na língua portuguesa que contenham tanto dados sobre o conteúdo das notícias quanto dados referentes às suas propagações. Desse modo, *datasets* são criados a partir do processo proposto de forma a viabilizar a aplicação e a comparação de métodos do estado da arte na detecção de *fake news* em MDDN com diferentes demandas de dados (e.g. conteúdo e/ou propagação).

## 1.4 Metodologia

Esse processo foi desenvolvido com base em requisitos que foram levantados a partir da demanda de dados dos MLFN do estado da arte. Por meio do levantamento desses requisitos foi possível identificar atributos através da observação (i.e.: engenharia reversa) dos *datasets* utilizados por esses métodos. O conjunto de atributos resultante da engenharia reversa aplicada e da análise das informações disponibilizadas por MDDN foi, então, estruturado e abstraído em um modelo de dados conceitual.

Ademais, por meio da observação dos trabalhos relacionados que descrevem processos de construção de *datasets* sobre *fake news*, o levantamento de requisitos ainda possibilitou a identificação das principais funcionalidades a serem incorporadas ao processo. Com base no modelo de dados conceitual desenvolvido e nas funcionalidades identificadas, foi criado um modelo macro-funcional do processo que favorece o entendimento de como e quando ocorrem a coleta (recuperação de informação em diferentes MDDN), o processamento (e.g., preparação da busca por similaridade) e o armazenamento dos dados referentes a cada conjunto de atributos do modelo de dados conceitual.

Cada etapa do modelo macro-funcional foi detalhada por meio de pseudocódigos e os passos desses algoritmos foram detalhados formalmente a fim de viabilizar uma implementação básica do processo.

Como forma de apresentar a viabilidade e adequação do processo proposto, foi realizado um estudo de caso que envolveu a implementação de um protótipo customizado para utilizar a rede social *Twitter*. A partir do *dataset* criado nesse estudo de caso, foram aplicados e comparados dez diferentes métodos de detecção de *fake news* com diferentes demandas de dados (sendo sete com demanda por dados de propagação) a fim de demonstrar o potencial de utilização tanto do processo quanto do *dataset* criado para viabilizar a comparação de diferentes MLFN.

## 1.5 Contribuições Esperadas

Em suma, as principais contribuições esperadas desta dissertação são:

(i) a especificação do processo de construção de *datasets* proposto;

(ii) a implementação de um protótipo desse processo, capaz de gerar instâncias de *datasets* que viabilizam a comparação entre MLFN com diferentes necessidades de informação; e

(iii) a instância do *dataset* gerada para o estudo de caso.

## 1.6 Organização do Texto

Esta dissertação está organizada em mais cinco capítulos. O Capítulo 2 introduz os conceitos básicos para o entendimento deste trabalho. O Capítulo 3 descreve os trabalhos relacionados, por intermédio da proposta e da aplicação de um modelo comparativo. O Capítulo 4, por sua vez, apresenta o processo de construção de *datasets* proposto. O estudo de caso é apresentado no Capítulo 5, por meio do detalhamento do protótipo, da descrição das peculiaridades do *dataset* gerado e da avaliação da aplicação dos métodos de

detecção de *fake news*. Por fim, o Capítulo 6 apresenta as considerações finais, destacando as contribuições e os principais resultados obtidos pelo presente trabalho.

# 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda as definições e os conceitos relevantes ao entendimento do presente trabalho. Nele, são apresentados fundamentos de aprendizado de máquina (Seção 2.1), *fake news* (Seção 2.2) e recuperação de informação (Seção 2.3).

## 2.1 Aprendizado de Máquina

Ultimamente, com a crescente demanda de complexos problemas a serem tratados computacionalmente devido ao enorme volume de dados existente, tornou-se clara a necessidade de ferramentas computacionais mais sofisticadas, que fossem mais autônomas, reduzindo a necessidade de intervenção humana e a dependência de especialistas. Uma ideia seria então desenvolver técnicas que fossem capazes de propor uma solução para o problema, a partir do processamento adequado de registros anteriores ao problema, que seria então apresentada, por exemplo, por meio da geração de uma função aproximada ou formulação de uma hipótese. A esse processo de indução de uma hipótese a partir da experiência passada dá-se o nome de Aprendizado de Máquina (AM) [FACELI et al., 2011].

Entre os problemas que são tratados atualmente com técnicas de AM podemos citar:

- Reconhecimento de palavras faladas;
- Detecção do uso fraudulento de cartões de crédito;
- Programas capazes de jogar xadrez como os melhores jogadores existentes;
- Auxílio no diagnóstico de doenças;
- Detecção de *Fake News*.

Os métodos utilizados pelo AM para encontrar regularidades, padrões e conceitos em conjuntos de dados podem ser organizados de diferentes formas, como, por exemplo, de acordo com o paradigma de aprendizado adotado para executar a tarefa a que se destina [FACELI et al., 2011], conforme descrito a seguir:

- Tarefas preditivas (aprendizado supervisionado): compreende a definição de uma função que associe os dados a um rótulo ou valor que os caracterize; conhecido

como paradigma de aprendizado supervisionado por considerar a simulação de um "supervisor externo" responsável por fazer algumas definições;

- Tarefas descritivas (aprendizado não supervisionado): procura descrever o conjunto de dados; conhecido como paradigma de aprendizado não supervisionado por não necessitar de informações externas ao conjunto de dados;

No paradigma de aprendizado não supervisionado, como não existe a informação de saída desejada, o aprendizado tenta identificar regularidade entre os dados, em busca de similaridades [GOLDSCHMIDT; PASSOS; BEZERRA, 2015]. Os principais tipos de aprendizado não supervisionado são:

- Agrupamento: consiste em encontrar e agrupar conjuntos de objetos com características similares (e.g., identificar grupos de clientes com características semelhantes);
- Associação: encontrar padrões frequentes de associações entre os atributos do conjunto de dados. Por exemplo, descobrir no conjunto de registros de compras de uma loja quais os produtos que estão associados;
- Detecção de desvios: encontrar dados que não obedecem ao comportamento dos demais dados existentes. Por exemplo, encontrar as categorias profissionais que apresentem salários acima da média nacional;
- Sumarização: encontrar uma descrição simples e compacta para um conjunto de dados. Por exemplo, definir o perfil dos clientes de um banco de acordo com faixa etária, faixa de renda, local da residência, tipos e valores de aplicações.

As subseções a seguir apresentam o aprendizado supervisionado e alguns conceitos a respeito dos *datasets* nos quais os registros anteriores aos problemas tratados com técnicas de AM são armazenados.

### 2.1.1 Aprendizado Supervisionado

O aprendizado supervisionado compreende a abstração de um modelo de conhecimento a partir de uma entrada e uma respectiva saída desejada [GOLDSCHMIDT; PASSOS; BEZERRA, 2015]. Assim, essa forma de aprendizado compreende estimar uma função aproximada a partir dos dados de entrada, descritos com base em seus atributos (ou características) que leve a um rótulo, conhecido como classe, ou um valor. Ou seja, considerando $H = (x_i, f(x_i)), i = 1, ..., n$, onde $x_i$ é o conjunto de atributos do elemento $i$ do conjunto de dados e $f(x_i) = y_i$ um valor conhecido associado ao elemento $x_i$, estimar a função $\widehat{f}$, isso é, uma função $\widehat{f} \cong f$ que aproxime a função original $f$ com um erro mínimo nessa aproximação.

Quando $y_i = f(x_i) \in c_1, c_2, ..., c_n$, isso é, são valores discretos, o aprendizado é denominado "classificação", enquanto que quando $y_i = f(x_i) \in R$, isso é, são valores contínuos, o aprendizado é denominado "regressão". Como exemplo de classificação temos, a partir de dados históricos de pacientes, definir uma função a ser aplicada aos dados de um novo paciente que indique se ele tem propensão à doença em questão. E como exemplo de regressão temos, a partir de tuplas de dados, aproximar uma curva que passe por esses dados, sendo a regressão caracterizada pela curva utilizada, podendo ser linear ou polinomial entre outros tipos.

### 2.1.2 Dataset

Um enorme volume de dados é gerado a cada dia por meio de transações financeiras, monitoramento ambiental, obtenção de dados clínicos e genéticos, captura de imagens, navegação na internet, dentre outras atividades. Os dados podem ainda assumir diferentes formatos, como série temporais, *itemsets*, transações, grafos ou redes sociais, textos, páginas web, imagens (vídeos) e áudios (músicas). Esses dados formam um conjunto, simplesmente denominado conjunto de dados ou *dataset*.

*Datasets* são formados por objetos que podem representar um objeto físico, como uma cadeira, ou uma noção abstrata, como os sintomas apresentados por um paciente que se dirige a um hospital. Em geral, cada objeto é descrito por um conjunto de atributos ou vetor de características. Cada objeto corresponde a uma ocorrência de dados e cada atributo está associada a uma propriedade do objeto [FACELI et al., 2011].

Formalmente, os dados podem ser representados por uma matriz de objetos $X_{n \times d}$, em que $n$ é o número de objetos e $d$ é o número de atributos de cada objeto. O valor de $d$ define a dimensionalidade dos objetos ou do espaço de objetos (também chamado de espaço de entradas ou espaço de atributos). Cada elemento dessa matriz. $x_i^j$, contém o valor da j-ésima característica para o i-ésimo objeto. Os $d$ atributos também podem ser vistos como um conjunto de eixos ortogonais e os objetos, como pontos no espaço de dimensão $d$, chamado espaço de objetos [FACELI et al., 2011].

A Figura 4 ilustra um exemplo de espaço de objetos. Nesse espaço, a posição de cada objeto é definida pelos valores de dois atributos ($d = 2$), nesse caso, os resultados de dois exames clínicos. O atributo de saída é representado pelo formato do objeto na figura: círculo para pacientes doentes e triângulo para pacientes saudáveis.

Os MLFN precisam treinar, validar e testar modelos a partir de *datasets*, compostos de diferentes atributos de diversos tipos, tais como os relacionados ao conteúdo das notícias e/ou aqueles referentes à propagação dessas notícias. Dessa forma, *datasets* que possuam uma variedade de atributos são importantes para o desenvolvimento, a aplicação e comparação desses diferentes métodos.

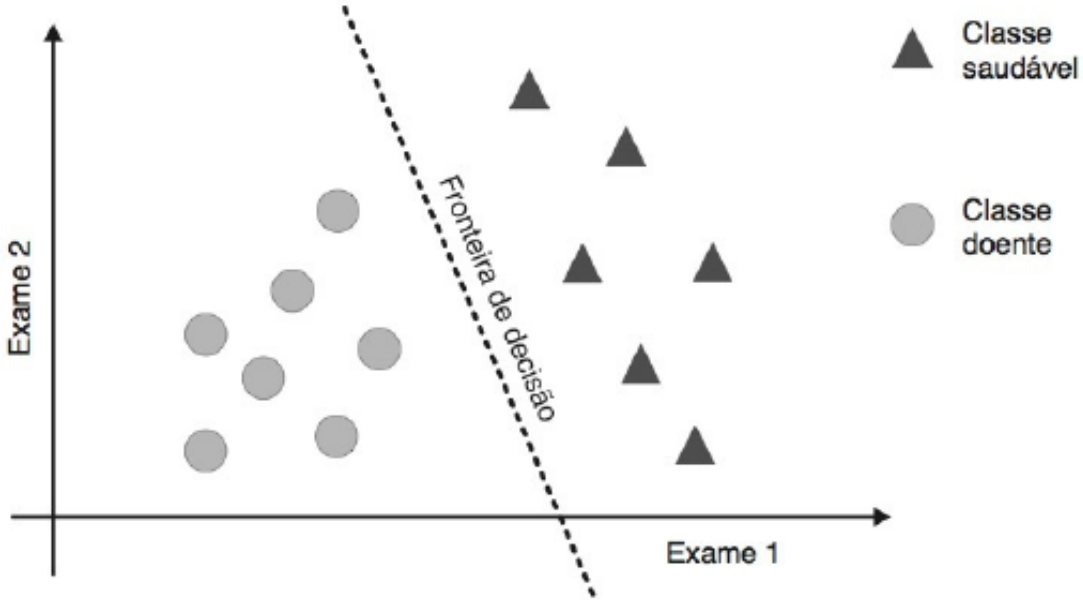



Figura 4 – Espaço de objetos. Fonte: Faceli et al. [2011].

## 2.2 *Fake News*

### 2.2.1 Conceituações de Verdade

Para facilitar o entendimento de como as notícias *fake* são rotuladas quanto a sua falsidade, torna-se importante descrever alguns conceitos consolidados sobre constatação de verdade, visto que a definição de *fake news* adotada por este trabalho é uma notícia **falsa**, publicada de forma intencional. Dessa forma, diferentes conceituações sobre verdade são abordadas a seguir.

Há diferentes concepções filosóficas sobre a natureza do conhecimento verdadeiro. Inclusive, visando evidenciar tal amplitude, pode-se enfatizar o ceticismo e o relativismo. A primeira concepção, na sua forma clássica (ou ceticismo pirrônico), defende uma postura suspensiva, haja vista que a multiplicidade de explicações acerca de uma mesma questão constitui, por si só, razão suficiente para nada se afirmar de forma absoluta [EMPIRICO, 1997]. Já a segunda, baseada no relativismo absoluto, aceita diferentes verdades para uma mesma questão [RODRIGUES, 2013]. Inclusive, é possível ressaltar a refutação tanto do ceticismo quanto do relativismo, pois é difícil para alguém declarar-se ceticista ou relativista sem se colocar fora ou acima de tal declaração. Isso acontece porque, se uma pessoa declara que "não há verdade absoluta"ou que "todas as verdades são relativas", aparece a dúvida se tal afirmação é ou não, respectivamente, cética ou relativa [FREIRE; GOLDSCHMIDT, 2019b].

Em outra perspectiva, segundo SPONHOLZ [2009], a veracidade de uma notícia pode ser constatada por meio de declarações descritivas que são baseadas em verificação dos fatos e/ou por intermédio da emissão de juízo de valor que se baseia em argumentações

(i.e.: justificativas de opiniões). Assim, uma notícia pode ter a verdade constatada tanto pela verificação dos fatos quanto dos argumentos. Este trabalho parte da premissa de que uma verdade deve ser constatada por meio de verificação de fatos (i.e.: utilizando as declarações descritivas).

### 2.2.2 Combate às *Fake News*

As abordagens automáticas de combate às *fake news* podem, basicamente, possuir três funcionalidades: Coleta, Detecção e Intervenção [FREIRE; GOLDSCHMIDT, 2019b]. Independente da abordagem adotada, a coleta dos dados inerentes à divulgação da notícia se faz necessária. Tanto a coleta de dados na rede social quanto as tarefas de detecção e intervenção são fases iterativas do processo de combate, conforme ilustra a Figura 5. Cabe ressaltar que quanto mais cedo ocorrer a detecção e a intervenção da *fake news*, menores serão os impactos negativos desta notícia.

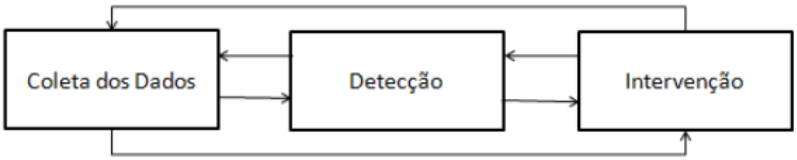



Figura 5 – Fluxo do processo de combate às *fake news* Fonte: Freire e Goldschmidt [2019b]

A *Detecção* automática da *fake news*, foco deste trabalho, pode ser um problema de classificação binária, onde dada uma rede social $\mathcal{G}$, uma notícia $a$ e um conjunto de postagens (publicações/propagações) $\mathcal{P}$, relacionadas à $a$, são espalhadas através da $\mathcal{G}$ por um conjunto de usuários $U$ em um intervalo de tempo $t$. Assim o referido classificador binário $\mathcal{F}$ deve, aprendendo a partir dos dados, prever se $a$ é uma *fake news* ou não, como formalmente indicado na equação 2.1. Uma outra forma é a utilização de técnicas mais subjetivas que definam a probabilidade, peso ou pertinência de uma notícia $a$ ser *fake* [FREIRE; GOLDSCHMIDT, 2019b].

$$\mathcal{F}(\mathcal{G}, a, \mathcal{P}, U, t) = \begin{cases} 1, & \text{se } a \text{ é uma } \textit{fake news}; \\ 0, & \text{caso contrário.} \end{cases} \tag{2.1}$$

O combate às *fake news* em redes sociais possui uma variedade de aspectos que podem ser considerados, tais aspectos são categorizados na figura 6 e, logo após, o aspecto referente aos dados, que representa o foco deste trabalho, será detalhado.

Segundo Freire e Goldschmidt [2019b], o aspecto relacionado aos dados que podem ser utilizados pelas abordagens computacionais de combate às *fake news* subdivide-se em dados obtidos a partir da publicação da notícia e aqueles associados com a sua propagação,

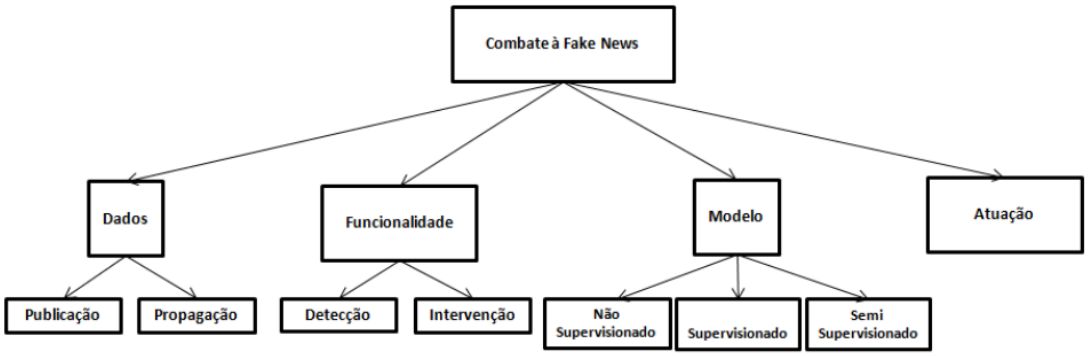



Figura 6 – Aspectos de Abordagens de Combate. Fonte: Freire e Goldschmidt [2019b]

conforme ilustrado na Figura 7. Os dados de publicação representam as informações inerentes ao surgimento da notícia na rede social. Estes dados podem ser classificados em notícia, usuário, assunto e temporalidade. No que diz respeito à notícia, a abordagem pode ser capaz de analisar dados oriundos da publicação a partir de diferentes tipos de mídia (texto, áudio, imagem e vídeo). A análise do conteúdo pode ser realizada de forma léxica, sintática, semântica e legibilidade. Com relação ao usuário publicador, a abordagem pode identificar diferentes tipos, tais como: humano, bot ou cyborg. Pode-se analisar também dados referentes ao Perfil do usuário na rede social, tais como: identificação e idade. Outro aspecto relevante está relacionado à Reputação do publicador, que pode estar vinculada à sua capacidade em identificar ou publicar *fake news*. A abordagem pode também utilizar o Assunto abordado no momento da publicação. Assim, seria possível tratar especificidades, tais como: relacionamento entre assuntos, assuntos controversos ou análise de tópicos. Outro aspecto leva em consideração a relevância do assunto publicado, haja vista que assuntos em voga motivam a criação de *fake news*. A variação das características de uma notícia com o passar do tempo torna a temporalidade mais um relevante recurso para a identificação de *fake news*.

Os dados de Propagação representam as informações obtidas após a publicação, consequentemente, aquelas inerentes às contribuições devido ao espalhamento da notícia na rede social (reações tais como curtida/like, comentário/reply ou compartilhar/retweet). Portanto, estes dados podem ser classificados em contribuição, usuário, assunto, temporalidade e rede. No que diz respeito à contribuição, usuário, assunto e temporalidade, a abordagem pode ser capaz de analisar os dados oriundos da propagação a partir dos mesmos aspectos anteriormente citados referentes aos dados de publicação. Ademais, as informações relacionadas à rede criada, com base na propagação da notícia, possibilitam não só a identificação de uma *fake news* como uma possível atuação contra a mesma (intervenção).


Figura 7 – Dados das abordagens computacionais [FREIRE; GOLDSCHMIDT, 2019b].

### 2.2.3 Métodos de detecção

O desenvolvimento e a avaliação de diferentes MLFN permaneceram isolados por algum tempo. Mais recentemente, esforços de integração metodológica e abordagens híbridas produziram resultados promissores [RUBIN; CONROY, 2015]. A variedade de práticas jornalísticas e as fontes de notícias disponíveis exigem a consideração dos vários métodos existentes, uma vez que uma abordagem geralmente atua em pontos fracos conhecidos em outra.

Segundo Conroy, Rubin e Chen [2015], os métodos de detecção podem ser enquadrados em três abordagens:

- Abordagem linguística - o conteúdo de mensagens *fake* é extraído e analisado para associar a padrões de linguagem;
- Abordagem de rede - informações de rede, como metadados de mensagens ou consultas estruturadas em rede de conhecimento, podem ser aproveitadas para fornecer medidas agregadas.
- Abordagem híbrida - métodos mais efetivos de detecção que agrupam características das outras abordagens.

Essas abordagens normalmente incorporam técnicas de aprendizado de máquina a fim de gerar modelos a partir do treinamento de classificadores.

Uma breve descrição de alguns MLFN é apresentada a seguir, detalhes a respeito de suas aplicações podem ser consultados por intermédio das respectivas referências.

**Detective Ótimo** [TSCHIATSCHEK et al., 2018]: Esse método usa inferência Bayesiana para detectar *fake news* a partir de *Crowd Signals*. Esse *Crowd* é formado pela opinião dos usuários sobre a notícia, juntamente com a sua capacidade histórica de acertar ou errar as suas opiniões. O objetivo é detectar, de forma antecipada, a *fake news* e bloqueá-la;

***Implicit Crowd Signals*** **(*ICS*)** [FREIRE; GOLDSCHMIDT, 2019a]: Segundo esse método, em uma rede social $\mathcal{G}$, a opinião do usuário $u$ sobre uma notícia $n$ é obtida, de forma implícita, a partir da divulgação de $n$ por $u$ em $\mathcal{G}$. Assim, objetivando a detecção de $n$, o $ICS$ se inicia selecionando os usuários divulgadores (publicador(es)/propagador(es)) de $n$ como os membros que irão compor o $Crowd$. Em seguida, esse método afere as reputações desses membros a partir das suas respectivas divulgações de notícias, $fake$ e não $fake$, já realizadas. Por fim, o $ICS$ classifica $n$ como $fake$, ou não, associando as reputações dos membros do $Crowd$ com as suas respectivas opiniões ($Signals$) implícitas sobre $n$.

***Hybrid Crowd Signals (*****$HCS$*****)*** [FREIRE; GOLDSCHMIDT, 2020]: Esse método disponibiliza a formação de um *Crowd* híbrido, a partir do momento em que os membros do *Crowd* podem ser tanto usuários divulgadores do meio digital quanto máquinas (modelos de classificação de notícias) disponibilizadas para uso da *HCS*.

Por meio de usuários divulgadores as opiniões implícitas são inferidas a partir dos padrões de comportamento desses usuários ao divulgar (publicar/propagar) notícias. Sendo assim, o fato de o usuário $u$ divulgar a notícia $n$ é um sinal implícito de que, na opinião de $u$, $n$ não é fake. Tal princípio se baseia na citação do filósofo Habermas, segundo a qual toda ação comunicativa traz consigo uma, inevitável, pretensão à verdade [HABERMAS, 1982].

As máquinas do *Crowd* representam a implementação de métodos de classificação para detecção de Fake News já existentes na literatura. Inclusive, essa utilização de máquinas permite a integração com outros trabalhos de combate às Fake News. Assim sendo, quando uma máquina $m$ classifica uma notícia $n$ como fake (resp. não fake), a opinião explícita deste membro $m$ do *Crowd* é que $n$ é fake (resp. não fake). Desta forma, a reputação das máquinas, utilizada para ponderar as suas respectivas opiniões, é obtida por meio do seus acertos e erros ao opinar de forma explícita sobre as notícias já classificadas.

Dessa forma, diferentemente da abordagem baseada em *Crowd Signals* para detecção de Fake News, a *HCS* não precisa que o meio digital disponibilize uma funcionalidade para o usuário expressar a sua opinião sobre a notícia, ressaltando que, mesmo com essa funcionalidade, ainda é necessário contar com a boa vontade do usuário em opinar.

***Fake News Detection through Textual Patterns*** [MORAES et al., 2019] : Esse método utiliza a estrutura linguística de textos de notícias em português para classificá-las em *fake* ou não. Dessa forma, indexa os dados de cada notícia das diferentes fontes na base única *elasticsearch* e realiza alguns cálculos para enriquecer essa base com novos atributos, como o *POS (Part Of Speech) Tagging* dos textos, onde cada *token* é classificado em substantivo, adjetivo, verbo, verbo auxiliar, advérbio, etc. Também são utilizados atributos obtidos por meio do cálculo dos percentuais de cada classe gramatical em relação ao total de palavras, quantidade e percentual de caracteres maiúsculos, pontos

de exclamação e sentimento no texto. Por fim, a polarização léxica do texto é verificada e algoritmos conhecidos para a classificação de texto, como o *Multinomial Naive Bayes* e o *SVM*, são aplicados à essa base única enriquecida.

**SL_XGBOOST|RF|SVM** [REIS et al., 2019]: O estudo que apresentou esse método considera que a detecção pode se basear no conteúdo de notícias (e.g., técnicas de processamento de linguagem natural), na reputação da fonte de notícias ou no ambiente (e.g., estrutura de rede social). Dessa forma, o método combina a aplicação de classificadores existentes que atuam em ao menos uma dessas três áreas a fim de buscar resultados melhores do que os obtidos com a aplicação desses classificadores de forma isolada.

### 2.2.4 Agência de Checagem de Fatos

Notícias on-line podem ser coletadas de diferentes fontes, como *homepages* de agências de notícias, mecanismos de busca e de RSV. No entanto, determinar manualmente a veracidade dessas notícias é uma tarefa desafiadora, geralmente exigindo experientes especialistas de domínio e uma análise criteriosa das evidências de falsidade, do contexto e de relatórios adicionais de fontes autorizadas.

Diante desse cenário, pesquisas relacionadas ao estudo de *fake news* buscam construir seus *datasets* a partir de notícias previamente rotuladas por jornalistas especialistas, agências de checagem de fatos, ferramentas detectoras ou *Crowd-sourced workers* [SHU et al., 2017].

Grande parte dos MLFN constata se uma notícia é falsa por meio de Agências de Checagem de Fatos. Embora essas organizações sejam baseadas no "jornalismo de verificação de fatos", que depende da verificação humana, soluções tecnológicas automatizadas podem contribuir com o processo, atribuindo pontuação de credibilidade ao conteúdo da Web por meio de inteligência artificial. O propósito dessas agências é analisar uma reclamação e rotular uma determinada notícia, principalmente, como *fake* ou não. As Agências de Checagem de Fatos desempenham um papel muito importante para a tarefa de detecção de *fake news*, visto que grande parte dos métodos do estado da arte utiliza modelos supervisionados de aprendizado.

Existem 225 agências ativas no mundo, de acordo com o site da Duke Reporters LAB[1]. O Brasil, por sua vez, possui 10 dessas iniciativas:

- **AFP fact check**[2] - A Agência brasileira do serviço de notícias "Agence France-Presse" mantém o blog do AFP Checamos que foi criado em junho de 2018 e faz parte do programa de verificação de fatos do *Facebook*. A equipe dessa agência, alocada em um escritório no Rio de Janeiro, conta atualmente com duas jornalistas

[1] <https://reporterslab.org/fact-checking/>
[2] <https://checamos.afp.com/afp-brasil>

dedicadas à verificação de imagens duvidosas, vídeos, declarações oficiais e outras informações falsas que surgem nas redes sociais. As jornalistas são supervisionadas por um chefe de redação e pela diretoria do escritório. Cada verificação é editada por ao menos dois integrantes da equipe de coordenação do projeto.

- **Agência Lupa**[3] - A Lupa, fundada em 1º de novembro de 2015, é a primeira agência de notícias do Brasil a se especializar na técnica jornalística mundialmente conhecida como *fact-checking*. A agência é dividida em três conselhos: editorial, de ética e de negócios, cada um com cerca de três profissionais. Além disso, a Lupa também possui parceria com a principal iniciativa de combate à desinformação do *Facebook*. Desde maio de 2018, a agência integra o *Third-Party Fact-Checking Program* (3PFC), pelo qual confere informações denunciadas pelos usuários do *Facebook* como possivelmente falsas.

- **Aos fatos**[4] - Com bases no Rio de Janeiro e em São Paulo, a agência "Aos Fatos" possui uma equipe composta de dezesseis profissionais multidisciplinares que, diariamente, acompanha declarações de políticos e autoridades de expressão nacional, de diversas colorações partidárias, de modo a verificar a veracidade dessas declarações. Atualmente, essa agência mantém parceria remunerada com o *Facebook* por meio do programa de checadores independentes. De forma não remunerada, a TV Cultura também é parceira de republicação do Aos Fatos.

- **Boatos**[5] - Fundada em junho de 2013, essa agência independente é suportada pela divulgação de anúncios de combate a *fake news* e possui uma equipe composta de quatro jornalistas.

- **Comprova**[6] - Essa agência é coordenada pela Associação Brasileira de Jornalismo Investigativo com a *First Draft*, uma organização independente que patrocinou parcerias semelhantes em vários países. Os parceiros da Comprova coordenam seu trabalho usando a metodologia de verificação cruzada do *First Draft*. A *Google News Initiative* e o *Facebook Journalism Project* fornecem suporte financeiro e técnico. A PROJOR, uma organização sem fins lucrativos focada em questões relacionadas à mídia brasileira, também foi uma das primeiras apoiadoras dessa agência.

- **É isso mesmo?**[7] - Essa agência é composta por uma equipe de jornalistas do jornal "O Globo"que analisam declarações de políticos e informações divulgadas nas redes sociais.

[3] <https://piaui.folha.uol.com.br/lupa/>
[4] <http://aosfatos.org/>
[5] <https://boatos.org>
[6] <https://projetocomprova.com.br/>
[7] <https://blogs.oglobo.globo.com/eissomesmo/>

- **E-farsas**[8] - Agência independente que verifica e desmascara o conteúdo *fake* online. É mantida desde 2012 com receita de publicidade da empresa brasileira de mídia "Portal R7";

- **Estadão Verifica**[9] - Essa agência que pertence ao "Grupo Estado" é uma organização de mídia mantida com receita publicitária e assinaturas de leitores e possui uma equipe composta de quatro jornalistas. O conteúdo analisado inclui notícias de relevância nacional e internacional.

- **Portal EBC's Hoax reports**[10] - Agência do serviço público de radiodifusão e notícias do Brasil que disponibiliza uma série contínua de *fake news*.

- **UOL Confere**[11] - As checagens e explicações dos fatos dessa agência se concentram, principalmente, em declarações de funcionários públicos do Brasil. Essas checagens são disponibilizadas no portal "UOL web" e o serviço é suportado por rendimentos de anúncios.

## 2.3 Recuperação de Informação

O significado desse termo pode ser muito amplo. Por exemplo, uma simples recuperação da informação de um cartão de crédito em uma carteira a fim de viabilizar o registro do número desse cartão durante uma compra online. No entanto, como campo de estudo acadêmico, Salton [1968] define Recuperação de Informação (RI) como a ciência que lida com a representação, o armazenamento, a organização e o acesso a itens de informação para atender a requisitos de usuários, referentes a essas informações. Manning, Raghavan e Schütze [2008], por sua vez, definem RI como a busca por material (geralmente documentos) de natureza não estruturada (geralmente texto) que satisfaça uma demanda por informações dentro de grandes coleções de dados (geralmente armazenadas em computadores).

A RI não começou na Web, mas sim a partir de demandas por resgate de informações em publicações científicas e em registros de bibliotecas [MANNING; RAGHAVAN; SCHÜTZE, 2008]. Em resposta a vários desafios de se fornecer acesso à informação, o campo da RI evoluiu para oferecer abordagens científicas que viabilizassem pesquisas em várias formas de conteúdo. A RI está se tornando rapidamente a forma dominante de acesso às informações, superando a pesquisa de dados tradicional que ocorre, por exemplo, quando um funcionário diz: "Sinto muito, só posso procurar seu pedido se você puder me fornecer seu ID do pedido".

[8] <https://www.e-farsas.com/>
[9] <https://politica.estadao.com.br/blogs/estadao-verifica/>
[10] <http://www.ebc.com.br/hoax>
[11] <https://noticias.uol.com.br/confere>

Um sistema de RI é um domínio muito bem estabelecido, avaliado em sua precisão e capacidade de recuperar informações ou documentos com qualidade, ou seja, quanto mais as respostas corresponderem às expectativas dos usuários, melhor é o sistema [BOUADJENEK; HACID; BOUZEGHOUB, 2016].

Dentre as tarefas de um sistema de RI, destaca-se a recuperação de informações textuais em redes sociais, pois possui notória demanda, já que o conhecimento humano é armazenado de forma textual e, depois da fala, o texto é a principal forma de transmissão de informação [BOUADJENEK; HACID; BOUZEGHOUB, 2016].

### 2.3.1 Abordagens de RI em redes sociais

Segundo Bouadjenek, Hacid e Bouzeghoub [2016], diferentes modelos de redes sociais têm sido utilizados por métodos de RI de formas distintas. Assim, várias abordagens de RI podem ser categorizadas de acordo com os tipos de informações de redes sociais utilizados. A figura 8 propõe uma taxonomia que agrupa diferentes iniciativas e viabiliza uma compreensão comum desse domínio.

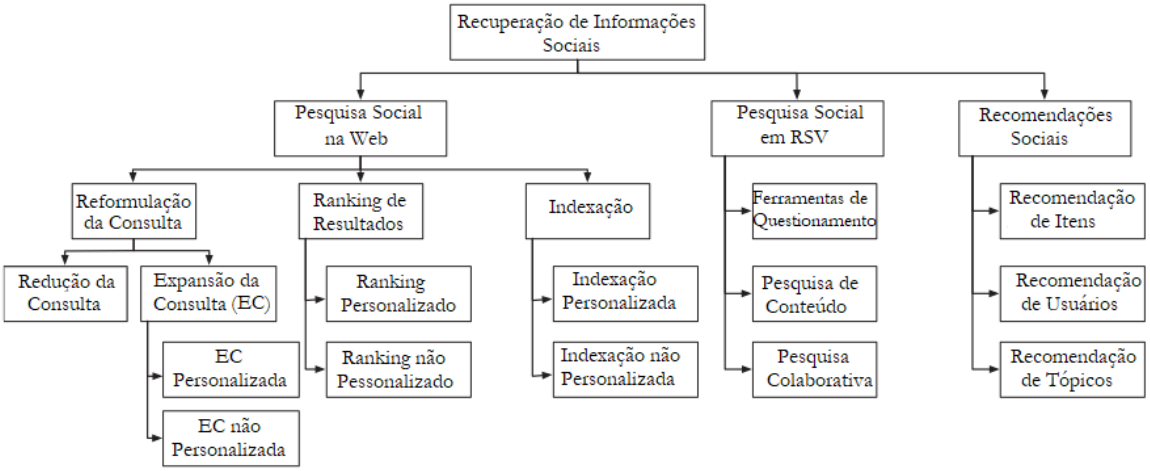



Figura 8 – Taxonomia para um modelo de RI. Fonte: [BOUADJENEK; HACID; BOUZEGHOUB, 2016]

Há três categorias principais, segundo a taxonomia proposta por Bouadjenek, Hacid e Bouzeghoub [2016]:

- *Pesquisa Social na Web* - Nessa categoria, as informações sociais são utilizadas para melhorar o processo clássico de RI. Os sistemas de RI geralmente utilizam índices e ontologias para interpretar e processar documentos, mas os resultados recuperados não são necessariamente relevantes da perspectiva do usuário final, apesar de a classificação ter sido realizada por um mecanismo de pesquisa. Para

expandir o processo clássico de RI e reduzir a quantidade de documentos irrelevantes, há principalmente três possibilidades de melhoria:

  - *Reformulação da Consulta* - É o processo que consiste em transformar uma consulta inicial Q em outra consulta Q'. Essa transformação pode ser uma redução ou uma expansão da consulta (EC). A primeira reduz a consulta de forma que as informações supérfluas sejam removidas [KUMARAN; CARVALHO, 2009], enquanto que a expansão aprimora a consulta com informações adicionais que provavelmente resultarão em documentos relevantes [EFTHIMIADIS, 2000]. Caso uma consulta de um determinado usuário consista de n termos $Q = \{t_1, t_2, ..., t_i, t_{i+1}, ..., t_n\}$, a EC pode ter dois componentes: adição de novos termos $T' = \{t'_1, t'_2, ..., t'_m\}$ do(s) *data source(s) D* e remoção de *stop words* $T'' = \{t_{i+1}, t_{i+2}, ..., t_{i+n}\}$. Assim, essa consulta pode ser representada como:

$$Q_{exp} = (Q - T'') \cup T' = \{t_1, t_2, ..., t_i, t'_1, t'_2, ..., t'_m\}$$

    Na definição acima, o principal aspecto da EC é o $T'$: conjunto de novos termos significativos adicionados à consulta original do usuário para recuperar documentos mais relevantes e reduzir a ambiguidade.

  - *Ranking de Resultados* - Consiste na definição de uma função que permite quantificar as semelhanças entre documentos e consultas. Esse processo pode ser realizado de forma personalizada (utilizando informações sociais para personalizar os resultados da pesquisa) ou não personalizada (usando informações sociais para adicionar uma relevância social ao processo de classificação)

  - *Indexação* - Melhoria do modelo de RI por meio de contextos sociais, como anotações, comentários e menções. Esses contextos são frequentemente associados a páginas da web e são capazes de enriquecer consideravelmente o conteúdo dessas páginas.

- *Pesquisa Social em RSV* - Processo de pesquisa de informações que considera apenas entidades, interações e contribuições das redes sociais [BOUADJENEK; HACID; BOUZEGHOUB, 2016]. Plataformas sociais como *Twitter* e *Facebook* não só permitem o compartilhamento e a publicação de informações, como também viabilizam a utilização de consultas muito precisas e altamente contextualizadas, como o uso de uma *hashtag* no *Twitter* a fim de se obter informações a respeito de uma determinada conferência a ser realizada. Esse processo pode ser dividido em três categorias principais:

  - *Ferramentas de Questionamento* - Ferramentas de perguntas e respostas. Basicamente, essas ferramentas fornecem meios para responder a vários tipos de perguntas, como recomendações, busca de opinião, conhecimento factual, solução de problemas, entre outros.

  - *Pesquisa de Conteúdo* - Esse é o tipo de pesquisa utilizado no processo proposto pelo presente trabalho. As redes sociais permitem que os usuários forneçam, publiquem e compartilhem informações. Nesse contexto, uma enorme quantidade de informações é criada nessas redes, o que representa uma fonte valiosa de informações relevantes. Portanto, sistemas de pesquisa de conteúdo social são um meio de indexar conteúdo criado explicitamente por usuários nas redes sociais e fornecer um suporte de pesquisa em tempo real [JANSEN; CHOWDURY; COOK, 2010].
  - *Pesquisa Colaborativa* - Uma das deficiências dos mecanismos de pesquisa disponíveis atualmente é o fato de que eles foram projetados para um único usuário que pesquisa sozinho. Essa categoria busca suprir essa deficiência, viabilizando o emprego de ferramentas colaborativas de pesquisa.

- *Recomendações Sociais* - É um conjunto de técnicas que tentam sugerir itens (por exemplo, filmes, músicas, livros, notícias, páginas da web), entidades sociais (por exemplo, pessoas, eventos, grupos) ou tópicos de interesse (por exemplo, esporte, cultura e culinária) que provavelmente sejam do interesse do usuário, por meio do uso de informações sociais [BOUADJENEK; HACID; BOUZEGHOUB, 2016].

# 3 TRABALHOS RELACIONADOS

Este estudo considera trabalhos que de alguma forma constroem *datasets*, mesmo que essa construção não seja o objetivo principal de alguns desses trabalhos, já que vários deles destinam-se, principalmente, à criação de métodos de detecção automática de *fake news*.

Cada *dataset* utilizado nos trabalhos relacionados pode ser representado por uma matriz de objetos $X_{n \times d}$, em que $n$ é o número de objetos e $d$ é o número de atributos de entrada de cada objeto, conforme detalhado na Subseção 2.1.2.

A fim de facilitar a análise dos trabalhos relacionados, um modelo comparativo é proposto e aplicado. Para tanto, alguns critérios de comparação foram estabelecidos a partir de características comuns entre objetos e atributos de *datasets*, pertencentes aos trabalhos do estado da arte pesquisados.

## 3.1 Proposta de Modelo Comparativo

De uma forma geral, os trabalhos baseados em aprendizado de máquina para detecção de *Fake News* desenvolveram modelos de classificação binária como o representado conceitualmente pela Equação 3.1. Nesta equação, $a$ é uma notícia arbitrária descrita a partir de um registro de dados $(a_1, a_2, ..., a_n)$ onde cada $a_i$, $i = 1, ..., n$, corresponde ao valor de um atributo $A_i$ que, por sua vez, representa alguma característica associada a $a$ e pertence a um conjunto ordenado de atributos $E = \{A_1, A_2, ..., A_n\}$. São exemplos de elementos de $E$: o assunto de que trata uma notícia, o texto e/ou a imagem e/ou áudio e/ou vídeo veiculados nela, a data e a hora em que a notícia foi divulgada, os meios de divulgação, a identificação dos responsáveis pela propagação da notícia nas redes sociais, os comentários e compartilhamentos associados a ela, entre outros. No contexto desta dissertação, qualquer subconjunto não vazio de $E$ será denominado *esquema de representação de notícias*, ou simplesmente, *esquema de representação*.

$$\mathcal{F}(a) = \begin{cases} 1, & \text{se } a \text{ é uma } \textit{fake news}; \\ 0, & \text{caso contrário.} \end{cases} \tag{3.1}$$

Em linhas gerais, *datasets* utilizados por MLFN podem ser agrupados em função do esquema de representação utilizado. Para tanto, é importante, primeiramente, distinguir dois tipos de eventos relacionados à divulgação de notícias que podem conter informações relevantes para a identificação de fake news: a *publicação* de uma notícia e a *propagação* dessa notícia. A publicação de uma notícia corresponde ao evento em que tal notícia é disponibilizada em algum MDDN pela primeira vez. Por outro lado, toda e qualquer

reação ocorrida em uma rede social (tais como curtida/*like*, comentário/*reply* ou compartilhamento/*retweet*) com relação a uma notícia previamente publicada corresponde a um evento de engajamento e, consequentemente, de propagação dessa notícia.

Além da distinção entre publicação e propagação, o agrupamento dos trabalhos relacionados em esquemas de representação requer que cada atributo utilizado nesses trabalhos seja classificado em um dentre quatro tipos. No primeiro tipo estão os atributos relacionados aos conteúdos das notícias, como, por exemplo, textos e imagens divulgados. O segundo engloba os atributos associados aos responsáveis pelas divulgações das notícias (e.g.: conjuntos de seguidores e de seguidos nas redes sociais). Os atributos relacionados à dimensão temporal, tais como data e hora de uma publicação ou de uma propagação, se enquadram no terceiro tipo. Finalmente, o quarto e último tipo abrange os atributos relacionados às redes de propagação das notícias ao longo do tempo, como, por exemplo, sequências de compartilhamentos, comentários e curtidas, entre outros. Na Tabela 1 é apresentada uma taxonomia que sumariza os esquemas de representação em função do cruzamento entre os tipos de evento e os tipos de atributo considerados pelos *datasets* utilizados nos trabalhos relacionados.

| Esquema de Representação | Tipos de Atributo | | | | | | |
|---|---|---|---|---|---|---|---|
| | Publicação | | | Propagação | | | |
| | Conteúdo | Responsável | Temporalidade | Conteúdo | Responsável | Temporalidade | Rede |
| $E_1$ | X | | | | | | |
| $E_2$ | X | X | | | | | |
| $E_3$ | X | X | X | X | | | |
| $E_4$ | X | X | X | X | X | X | X |
| $E_5$ | X | | | X | | | |
| $E_6$ | X | | | X | | | X |
| $E_7$ | | X | X | | X | X | X |
| $E_8$ | | X | | | X | | X |
| $E_9$ | | | | | | | X |

Tabela 1 – Taxonomia de esquemas de representação utilizados pelos trabalhos relacionados

Em termos formais, a relação entre os conceitos "métodos de detecção de *fake news*" e "esquema de representação" pode ser definida da seguinte maneira: *diz-se que um método de detecção $M$ é baseado em um esquema de representação $E$, se, e somente se, cada tipo de atributo necessário ao funcionamento de $M$ pertencer a $E$*. Logo, um esquema $E$ de um *dataset* qualquer viabiliza a aplicação de $M$ baseado em um esquema de representação $E_M$, se, e somente se, $E_M \subseteq E$.

O presente trabalho adotou o esquema de representação em que o *dataset* se baseia em conjunto com outros critérios para compor o modelo comparativo dos trabalhos relacionados. Logo, uma breve descrição de todos os critérios adotados é apresentada a seguir:

- **Esquema de Representação (C1)** - Esse critério indica em qual esquema o *dataset* pode ser representado, de acordo com a taxonomia descrita por meio da Tabela 1. Dependendo do esquema de representação utilizado, diferentes métodos de detecção de *fake news* têm sido investigados. Por exemplo, o esquema $E_1$ tornou-se um dos mais populares dentre os MLFN que utilizam atributos relacionados ao conteúdo das publicações por permitir que alguns trabalhos avaliassem a natureza do assunto tratado [YANG et al., 2019], assim como a relevância e a controvérsia potencialmente envolvida [WOLOSZYN; NEJDL, 2018], tanto em assuntos considerados de forma isolada [Helmstetter; Paulheim, 2018] quanto em assuntos relacionados entre si [WANG, 2017].

- **Idioma Português (C2)** - Esse critério indica se o *dataset* possui notícias no idioma português. Métodos de detecção de fake news, principalmente os que utilizam abordagens voltadas para análise de texto, em grande parte, são aplicados especificamente a *datasets* com notícias em um único idioma, em decorrência das especificidades de cada língua [CORDEIRO; PINHEIRO, 2019; FREIRE; GOLDSCHMIDT, 2019b].

- **Modelo de Dados Conceitual (C3)** - Esse critério indica se o trabalho analisado possui uma abstração do *dataset* utilizado. Essa abstração pode ser expressa por meio de uma modelagem conceitual que pode ser usada para representar níveis mais altos de abstração dos dados, em prol da facilitação de um maior entendimento de aspectos essenciais à descrição do problema. Dessa forma, um modelo de dados conceitual facilitaria a compreensão do *dataset* utilizado em cada trabalho.

- **Politemático (C4)** - Esse critério indica se o *dataset* contém notícias sobre mais de um tema. Apesar de o debate midiático a respeito de *fake news* abordar demasiadamente a área política, essas notícias não se restringem a um único tema. Entretanto, uma parcela dos *datasets* utilizados para detecção de *fake news* contém apenas notícias monotemáticas [CASTELO et al., 2019]. Apesar disso, a pluralidade de temas das notícias nos *datasets* é importante para que a aplicação de métodos de detecção *fake news* não se restrinja a um tema específico [MONTEIRO et al., 2018].

- **Atualizável (C5)** - Esse critério indica se existe um processo que viabilize a atualização do *dataset* com novas notícias. Com o passar do tempo, *fake news* vêm sendo elaboradas de forma cada vez mais sofisticada a fim de enganar métodos de detecção [SHU et al., 2017]. Logo, espera-se que *datasets* possam de alguma forma ser atualizados com notícias recentes para mitigar esse problema.

- **Agência de Checagem de Fatos (C6)** - Esse critério indica se o *dataset* utiliza notícias previamente rotuladas em *fake* ou não por agências especializadas em checagem de fatos. Atualmente, existem 225 dessas agências ativas no mundo, de

acordo com o site da Duke Reporters LAB[1]. Cada uma dessas agências segue o código de conduta e princípios éticos[2] da *International Fact-Checking Network* (IFCN), sediada no *Poynter Institute*, nos Estados Unidos. Dessa forma, a utilização de notícias coletadas dessas agências fornece credibilidade ao processo de rotulação do *dataset*.

- **Processo de Construção Definido (C7)** - Esse critério revela se o trabalho definiu e demonstrou o processo utilizado para a construção do *dataset*. O detalhamento das etapas metodológicas de construção podem viabilizar a implementação de diferentes protótipos e, consequentemente, a criação de diferentes instâncias de *datasets* compatíveis com diferentes métodos de detecção [SHU et al., 2020].

[1] https://reporterslab.org/fact-checking/
[2] https://www.poynter.org/international-fact-checking-network-fact-checkers-code-principles

## 3.2 Aplicação do Modelo Proposto

Esta pesquisa, após analisar os resultados recuperados a partir da *string* de busca *fake news* desde o ano 2015, identificou trinta e cinco trabalhos relacionados à utilização de *datasets* destinados à aplicação de MLFN. Esses *datasets* foram classificados de acordo com os critérios acima detalhados e podem ser observados na Tabela 2.

| ***Dataset*** | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** | **C7** |
|---|---|---|---|---|---|---|---|
| **Bracis2019FakeNews** [Pedro Faustini, 2019] | $E_5$ | x | | x | | x | x |
| BS Detector [SHU et al., 2017] | $E_4$ | | | x | | | |
| BuzzFace [SANTIA; WILLIAMS, 2018] | $E_4$ | | | x | | | |
| BuzzFeedNews [JANZE; RISIUS, 2017] | $E_6$ | | | x | | | |
| BuzzFeedNews [SHU et al., 2017] | $E_4$ | | | x | | | |
| Celebrity [PéREZ-ROSAS et al., 2018] | $E_2$ | | | | | | |
| CredBank [SHU et al., 2017] | $E_2$ | | | x | | | |
| DataSet Emergent [ZHANG et al., 2018] | $E_1$ | | | x | | | |
| DistrustRank [WOLOSZYN; NEJDL, 2018] | $E_2$ | | | x | | | |
| Facebook para Detective [TSCHIATSCHEK et al., 2018] | $E_7$ | | | | | | |
| **Factck.Br Corpus** [MORENO; BRESSAN, 2019] | $E_1$ | x | | x | x | x | x |
| **FakeBr Corpus** [MONTEIRO et al., 2018] | $E_2$ | x | | x | | | x |
| FakeNews vs Satire [GOLBECK et al., 2018] | $E_1$ | | | x | | | |
| FakeNewsAMT [PéREZ-ROSAS et al., 2018] | $E_1$ | | | x | | | |
| FakeNewsData1 [HORNE; ADALI, 2017] | $E_1$ | | | | | | |
| **FakeNewsNet1** [SHU et al., 2017] | $E_4$ | | | x | | x | x |
| **FakeNewsNet2** [SHU et al., 2020] | $E_4$ | | | x | | x | x |
| **FakeTweet.Br Corpus** [CORDEIRO; PINHEIRO, 2019] | $E_5$ | x | | x | | | x |
| KV [DONG et al., 2014] | $E_3$ | | | x | | | |
| Large-scale Training Dataset [Helmstetter; Paulheim, 2018] | $E_3$ | | | | | | |
| **LIAR** [WANG, 2017] | $E_2$ | | | | | x | x |
| PoliticalNews [CASTELO et al., 2019] | $E_1$ | | | | | x | |
| PolitiFact para XFake [YANG et al., 2019] | $E_2$ | | | | | x | |
| RST-SVM Dataset [RUBIN; CONROY, 2015] | $E_1$ | | | x | | | |
| Signal Media para Evaluating M. L. Algorithms [Gilda, 2017] | $E_1$ | | | x | | | |
| Soc-LiveJournal [SRIVASTAVA et al., 2018] | $E_9$ | | | | | | |
| Twitter e Sina Weibo para CSI [RUCHANSKY; SEO; LIU, 2017] | $E_4$ | | | x | | | |
| Twitter e Sina Weibo para EANN [RUCHANSKY; SEO; LIU, 2017] | $E_4$ | | | x | | | |
| Twitter e Sina Weibo para Early Detection [LIU; BROOKWU, 2018] | $E_4$ | | | x | | | |
| Twitter e Sina Weibo para TCNN-URG [QIAN et al., 2018] | $E_5$ | | | x | | | |
| Twitter Auto Identifying fake news [BUNTAIN; GOLBECK, 2017] | $E_4$ | | | x | | x | |
| Twitter para Content Polluters [NASIM et al., 2018] | $E_4$ | | | x | | | |
| Twitter para Mitigation via Point Process [FARAJTABAR et al., 2017] | $E_9$ | | | x | | | |
| Twitter para TraceMiner [WU; LIU, 2018] | $E_8$ | | | x | | x | |
| Twitter Trec [SRIVASTAVA et al., 2018] | $E_9$ | | | x | | | |
| **LEGENDA:**<br>**C1**-Esquema de Representação —— **C2**-Idioma Português —— **C3**-Modelo de Dados Conceitual<br>**C4**-Politemático ———————— **C5**-Atualizável<br>**C6**-Agência de Checagem ————— **C7**-Processo de Construção Definido | | | | | | | |

Tabela 2 – Aplicação do Modelo Comparativo

### 3.2.1 Descrição dos Trabalhos

Até onde foi possível observar, apenas sete trabalhos explicitam o processo de construção do *dataset*. Os nomes dos *datasets* desses trabalhos estão destacados em negrito

na Tabela 2. Nesta subseção, esses destaques são brevemente descritos e seus detalhes podem ser consultados por intermédio das respectivas referências:

- **Bracis2019FakeNews** [Pedro Faustini, 2019] - Dois *datasets* foram construídos nesse trabalho. O primeiro é baseado em textos de 115 notícias falsas e 12 verdadeiras divulgadas pelo *WhatsApp*, já o outro é composto por 87 identificadores de postagens falsas e por 21 de postagens verdadeiras, coletadas do *Twitter*. O texto das mensagens do *WhatsApp* e as postagens relacionadas aos identificadores do *Twitter* estão no idioma português. Esses *datasets* contém dados de publicação da notícia, mas não armazenam dados referentes à sua propagação. O acesso a esse trabalho pode ser realizado por meio do repositório *github* do projeto[3].

- **Factck.Br Corpus** [MORENO; BRESSAN, 2019] - Esse *dataset* é composto por textos de 1313 notícias no idioma português que foram verificadas individualmente pelas agências de checagem de fatos brasileiras 'Lupa', 'Aos Fatos' e 'Truco'. Possui como principal característica o fato de ser atualizável, ou seja, há uma API que possibilita a atualização do *dataset* com novas notícias analisadas por agências de checagem. As notícias são rotuladas em 'Ainda é cedo para dizer', 'de olho', 'discutível', 'distorcido', 'exagerado', 'impossível provar', 'impreciso', 'insustentável', 'verdadeiro', 'falso' e 'outros'. Esse *dataset* contém dados de publicação da notícia, mas não armazena dados referentes à sua propagação. Todo o projeto pode ser acessado por meio de seu respectivo repositório *github*[4].

- **FakeBr Corpus** [MONTEIRO et al., 2018] - Esse é um dos primeiros *datasets* disponibilizados para viabilizar a construção de modelos de detecção de *Fake news* no idioma português. Possui 7200 notícias, das quais 3600 são consideradas verdadeiras, obtidas em versões digitais de mídias virtuais tradicionais e as outras 3600, obtidas em sites considerados propagadores de notícias falsas, são rotuladas como 'fake'. Esse *dataset* contém dados de publicação da notícia, mas não armazena dados referentes à sua propagação. O acesso a esse *dataset* pode ser realizado por meio do repositório *github* do projeto[5].

- **FakeNewsNet1** [SHU et al., 2017] - Esse *dataset* foi coletado do *Twitter* e contém 211 notícias rotuladas como *'Fake'* e outras 211 como *'Real'*, obtidas a partir das agências de checagem de fatos americanas *BuzzFeed* e *PolitiFact*. Ele contempla tanto dados de publicação quanto dados de propagação de notícias no idioma inglês. Todo esse trabalho pode ser acessado por meio seu respectivo repositório *github*[6].

[3] https://github.com/phfaustini/BRACIS2019_FAKENEWS
[4] https://github.com/jghm-f/FACTCK.BR
[5] https://github.com/roneysco/Fake.br-Corpus
[6] https://github.com/KaiDMML/FakeNewsNet/tree/old-version

- **FakeNewsNet2** [SHU et al., 2020] - Esse trabalho, além de disponibilizar um *dataset*, propõe um processo de construção de *datasets* multidimensionais que viabilize diferentes pesquisas relacionadas à detecção de *fake news*. O *dataset* disponibilizado é composto por 5.808 notícias rotuladas como *'Fake'* e outras 16.987 como *'Real'*, coletadas a partir das agências de checagem de fatos americanas *GossipCop* e *PolitiFact*. Esse *dataset* contempla tanto dados de publicação quanto dados de propagação de notícias no idioma inglês. Uma implementação do processo e o *dataset* podem ser acessados pelo repositório *github* do projeto[7].

- **FakeTweet.Br Corpus** [CORDEIRO; PINHEIRO, 2019] - Esse *dataset* foi construído para detecção *fake news* escritas em língua portuguesa e é composto por 206 notícias falsas e 94 verdadeiras. Apesar dos dados terem sido coletados a partir do *Twitter*, ele só contém o texto de publicação das notícias. O acesso ao *dataset* pode ser realizado através do seu respectivo repositório *github*[8].

- **LIAR** [WANG, 2017] - Esse *dataset* foi coletado da agência de checagem de fatos *PolitiFact* e, dessa forma, é composto por notícias no idioma inglês. Ele inclui 12.836 notícias rotuladas manualmente como 'Pants-fire', 'False', 'Barely-true', 'Half-true', 'Mostly true' e 'True'. Tanto dados de publicação quanto de propagação de notícias estão presentes no *dataset*, mas cabe salientar que os dados referentes ao usuário se resumem ao nome do autor da postagem. O *dataset* pode ser obtido pelo link da *University of California Santa Barbara (UCSB)*[9].

### 3.2.2 Análise

No inicio, as notícias *fake* e não *fake* possuíam formas muito distintas, mas, com o passar do tempo, a variação das características dessas notícias se torna cada vez mais sutil, de forma a dificultar a detecção por meio do próprio conteúdo, sendo necessário, dessa forma, efetuar a exploração além do conteúdo de notícias [SHU et al., 2020], como por meio do envolvimento de usuários e dos dados de propagação.

No entanto, o presente estudo identificou que poucos trabalhos disponibilizam *datasets* com esses dados (i.e., *datasets* cujo esquema de representação contemple atributos relacionados à publicação e à propagação de notícias), o que dificulta consideravelmente a realização de estudos comparativos entre métodos de detecção de diferentes abordagens. Isso decorre do fato de esses métodos normalmente utilizarem *datasets* de características distintas, diante da carência de conjuntos de dados mais robustos e completos que contemplem essas diferentes demandas de atributos.

[7] https://github.com/KaiDMML/FakeNewsNet
[8] https://github.com/prc992/FakeTweet.Br
[9] https://www.cs.ucsb.edu/william/data/liardataset.zip

Outro importante fato observado refere-se ao idioma. Grande parte dos métodos de detecção de *fake news* são dependentes de um idioma específico, devido não somente às variações linguísticas da notícia, mas também à grande quantidade de variáveis relacionadas às características dialéticas regionais que complementam a notícia, como, por exemplo, o uso de diferentes gírias e jargões na composição de comentários e *hashtags* [Zhou; Zafarani, 2018; CONROY; RUBIN; CHEN, 2015].

Assim, pesquisas sobre aplicação de métodos específicos para o idioma português são necessárias. Entretanto, até onde foi observado por esta pesquisa, há apenas quatro iniciativas de construção de *datasets* com notícias no idioma português [MONTEIRO et al., 2018; MORENO; BRESSAN, 2019; Pedro Faustini, 2019; CORDEIRO; PINHEIRO, 2019]. Além disso, nenhuma dessas iniciativas disponibiliza *datasets* que contenham dados de propagação da notícia.

Destaca-se, também, em todos os trabalhos analisados, a ausência de modelos conceituais que favorecessem o entendimento dos aspectos essenciais por meio de níveis mais altos de abstração.

Diante do exposto, identificou-se que, dentre os trabalhos referentes à construção de *datasets* destinados a aplicação de MLFN, poucos deles disponibilizam dados de propagação e esses poucos não possuem notícias no idioma português, o que inviabiliza a aplicação e, consequentemente, o estudo comparativo de métodos que demandem dados de propagação de notícias nesse idioma.

# 4 PROCESSO DE CONSTRUÇÃO DE *DATASETS*

Este capítulo apresenta a metodologia adotada no desenvolvimento do *FakeNewsSetGen*, processo de construção de *datasets* proposto por este trabalho. Além disso, descreve ainda os dois principais modelos resultantes desse desenvolvimento: o modelo de dados conceitual e o modelo macro-funcional, bem como a implementação base do processo proposto.

## 4.1 Metodologia de Desenvolvimento

O levantamento de requisitos[1] foi a base para a especificação do *FakeNewsSetGen*. A partir desse levantamento foi possível identificar atributos por meio da observação (i.e.: engenharia reversa) dos *datasets* utilizados pelos métodos do estado da arte de detecção de *fake news*, referenciados na Tabela 2.

Além da identificação desses atributos que, por sua vez, já viabiliza a representação da demanda de dados pelos métodos existentes, informações disponíveis em diferentes MDDN foram analisadas e, dessa forma, outros atributos foram identificados.

O conjunto de atributos resultante da engenharia reversa aplicada e da análise das informações disponibilizadas por MDDN foi, então, estruturado e abstraído em um modelo de dados conceitual. Esse modelo foi concebido de forma que *datasets* gerados, segundo sua estrutura, possam ser utilizados por diferentes métodos de detecção de *fake news*, bem como possam ser compatíveis com diferentes MDDN.

Ademais, por intermédio da observação dos trabalhos relacionados que descrevem processos de construção de *datasets* sobre *fake news*, o levantamento de requisitos ainda possibilitou a identificação das principais funcionalidades a serem incorporadas ao *FakeNewsSetGen*.

Com base no modelo de dados conceitual desenvolvido e nas funcionalidades identificadas, um modelo macro-funcional foi criado para descrever conceitualmente o funcionamento do *FakeNewsSetGen*. Esse modelo funcional favorece o entendimento de como e quando ocorrem a coleta, o processamento e o armazenamento dos dados referentes a cada conjunto de atributos do modelo de dados conceitual.

O presente trabalho desenvolveu ainda uma implementação base compatível com o processo proposto. Disponibilizada no repositório *github* do projeto[2], essa implementação foi desenvolvida na linguagem *Python* e possui funcionalidades que possibilitam o resgate

[1] Fase da engenharia de requisitos que busca levantar informações sobre o que o sistema deve fazer [PRESSMAN; MAXIM, 2016].

[2] https://github.com/kamplus/FakeNewsSetGen

dos dados de divulgação de notícias no *Twitter*. A implementação desenvolvida pode ser adaptada para atender a diferentes cenários tanto por meio de customizações (i.e.: ajustes no código da solução implementada) quanto por intermédio de parametrizações (i.e.: ajustes de parâmetros de entrada).

Algumas customizações na implementação base podem, por exemplo, permitir a interação com diferentes redes sociais. Já a escolha de outras entidades externas, tais como mídias virtuais e agências de checagem de fatos com as quais o processo vai interagir em cada aplicação, pode ser realizada simplesmente por meio de parametrizações. As instruções para a configuração desses parâmetros de entrada podem ser obtidas no repositório do projeto.

A Seção 5.1 contém um exemplo de protótipo customizado a partir da implementação base.

## 4.2 Modelo de Dados Conceitual

Ilustrado na Figura 9, o modelo de dados conceitual explicita os grupos de informação e as relações entre eles que são relevantes para um ou mais métodos de detecção de *fake news* do estado da arte. Para apresentação desse modelo foi utilizada a notação de Diagramas de Classes de Análise da UML [BEZERRA, 2007]. Abaixo segue uma breve descrição das classes do modelo desenvolvido.

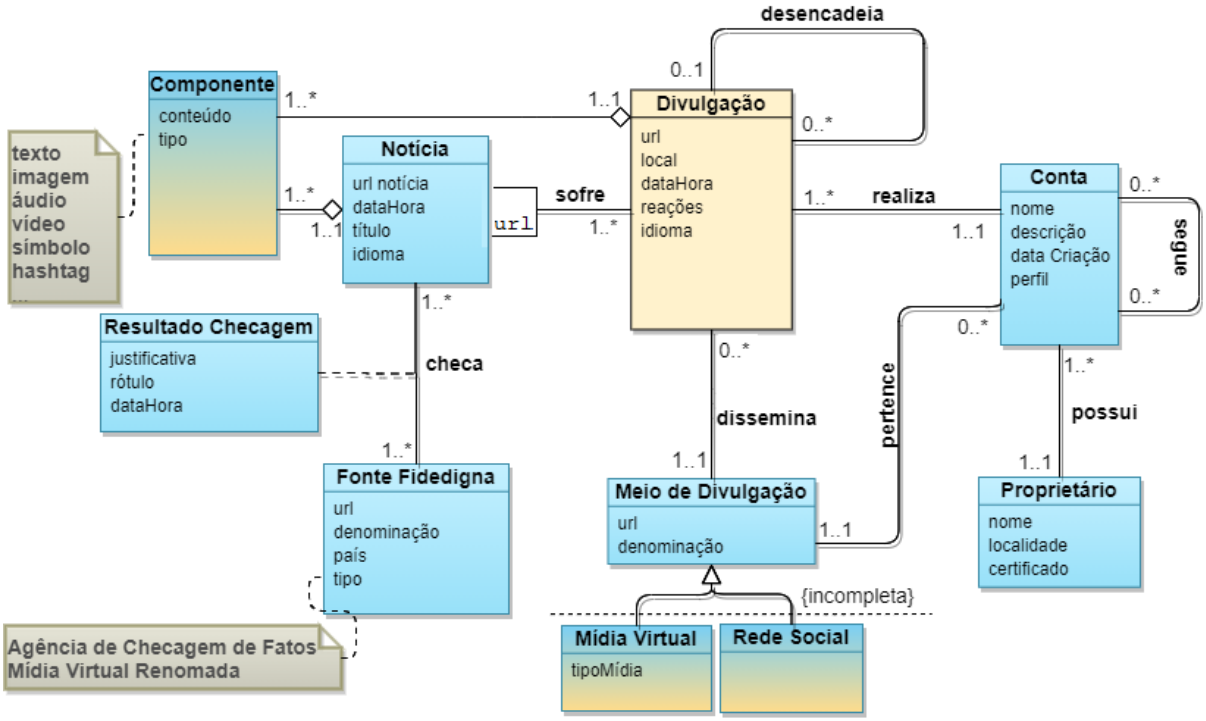



Figura 9 – *Modelo de Dados Conceitual.*

- **Meio de Divulgação -** Compreende todos os canais digitais por onde as notícias podem ser divulgadas (i.e., os *MDDN*) e representa uma generalização das classes Mídia Virtual e Rede Social;
- **Rede Social -** Representa estruturas formadas por pessoas e/ou organizações que se conectam a partir de interesses ou valores comuns (e.g.: *Facebook, Instagram e Twitter*);
- **Mídia Virtual -** Caso particular da classe Meio de Divulgação que se refere às mídias digitais como os jornais e revistas online, por exemplo;
- **Fonte Fidedigna -** Corresponde a uma agência de checagem de fatos (e.g.: Lupa, Aos Fatos e AFP) ou a uma mídia virtual renomada (e.g.: portais de notícias G1 e R7) que fornecem credibilidade ao processo de rotulação de notícias;
- **Resultado da Checagem -** Armazena o rótulo (fake e não fake) de cada notícia, obtido a partir de fontes fidedignas, bem como outros atributos complementares, como a justificativa do rótulo e o momento em que a checagem foi realizada;
- **Notícia -** Engloba características específicas da notícia checada por uma fonte fidedigna e pode conter diferentes componentes (e.g.: texto, imagem e áudio). A fim de viabilizar a comparação entre MLFN aplicados a diferentes classes desse modelo, somente são consideradas notícias com pelo menos uma divulgação na RSV;
- **Divulgação -** Representa todas as divulgações sofridas por uma determinada notícia ou desencadeadas a partir de outra divulgação (e.g. *tweets* e *retweets*). Devido ao dinamismo da propagação nas RSV, pode haver considerável diferença entre os quantitativos de divulgações de instâncias *FakeNewsSet*, pois quanto mais atual for a instância maior será a quantidade de divulgações capturadas;
- **Componente -** Engloba atributos de diferentes tipos (e.g.: texto, imagem e áudio) que podem estar presentes em notícias e em divulgações. O texto é o componente comumente utilizado pelos MLFN e normalmente está contido tanto na notícia quanto na divulgação;
- **Conta -** Contém dados das contas de usuários de meios de divulgação (e.g. idade da conta, contas seguidoras, contas seguidas). Não pertencem necessariamente a humanos, *bots*, por exemplo, podem controlar essas contas para realizar ações legítimas ou de cunho malicioso; e
- **Proprietário -** Compreende os dados do responsável por cada conta de usuário, sempre que esse responsável for identificado. Essa informação, ainda que disponível, não é necessariamente confiável, visto que não há um controle efetivo nas informações de cadastro de usuários nas RSV.

As associações (i.e. relacionamentos) entre as classes do modelo conceitual foram modeladas com o objetivo de atender à demanda dos MLFN. Assim, torna-se necessário detalhar algumas delas a fim de esclarecer as razões pelas quais esses relacionamentos foram adotados.

Dessa forma, como grande parte dos MLFN existentes utilizam algoritmos supervisionados de aprendizado de máquina, o escopo de notícias do modelo está restrito àquelas previamente rotuladas. Ou seja, todas as notícias previstas pelo modelo são obrigatoriamente checadas por, pelo menos, uma fonte fidedigna e o resultado dessa checagem é composto, basicamente, por um rótulo *fake* ou não *fake* e pelos motivos que justificam o rótulo atribuído.

Outro ponto a ser destacado refere-se às auto-associações das classes *Divulgação* e *Conta*, pois foram modeladas a partir de características comuns dos MDDN. Apesar das especificidades de cada meio (e.g., tamanho das notícias, características de propagação), grande parte dos MDDN possui funcionalidades que viabilizam a retransmissão de postagens e a composição grupos de seguidores de contas de usuários. Assim, o modelo prevê a possibilidade de uma *Divulgação* desencadear outras e de ser desencadeada por outra, bem como de uma *Conta* seguir ou ser seguida por outras.

Além das associações entre as classes, também é importante destacar o motivo da caracterização das mídias virtuais como fontes fidedignas. Esse motivo decorre do fato de que grande parte das notícias provenientes das agências de checagem de fatos são rotuladas *fake* [SHU et al., 2020], pois, diante das restrições de pessoal e de estrutura, essas agências tendem à direcionar a força de trabalho para a análise de notícias com maior probabilidade de serem falsas.

Com isso, *datasets* construídos a partir de notícias obtidas apenas dessas agências seriam desbalanceados, o que poderia prejudicar a aplicação dos MLFN, visto que os resultados dos modelos gerados por esses métodos seriam enviesados, ou seja, tenderiam a classificar novos dados como sendo da classe que possui mais exemplos. Assim, notícias divulgadas por mídias virtuais de renome são rotuladas não *fake* a fim de viabilizar a construção de *datasets* menos desbalanceados.

Logo, *datasets* criados com base no modelo de dados conceitual podem pertencer ao esquema de representação $E_4$, pois, caso o processo seja executado de forma completa, esses *datasets* seriam compostos por todos os atributos (i.e. conteúdo, responsável, temporalidade e rede) relacionados à publicação e à propagação de notícias.

## 4.3 Modelo Macro-Funcional

A Figura 10 ilustra graficamente o modelo macro-funcional do processo *FakeNewsSetGen* por meio de três divisões tradicionais dos sistemas de informação: entrada, processamento e saída. Nela, os retângulos de cantos arredondados representam as etapas do processo; os retângulos com cantos afiados representam entidades externas ao processo; a seta contínua indica uma relação de generalização entre as entidades externas; e as setas tracejadas indicam fluxos de dados.

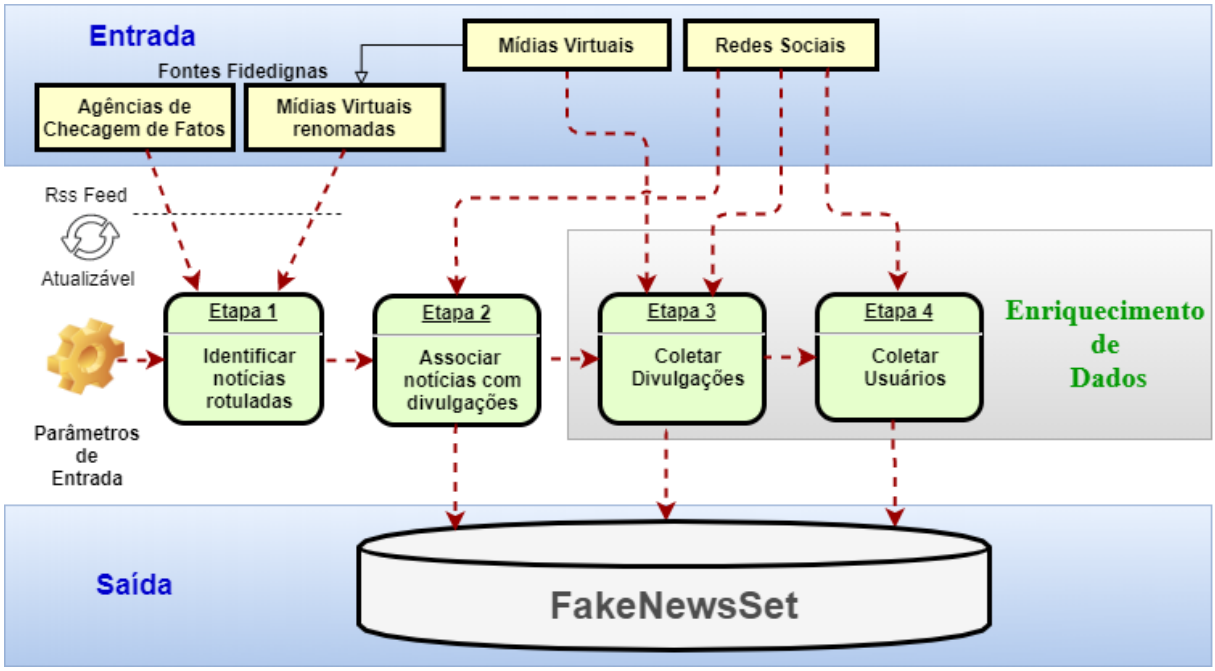



Figura 10 – Modelo Macro-Funcional do *FakeNewsSetGen*

Sempre que executado, o *FakeNewsSetGen* permite gerar uma instância de *dataset* denominada *FakeNewsSet*, compatível com o modelo de dados conceitual apresentado na seção anterior. Para tanto, antes da execução da primeira etapa do processo, é preciso que sejam realizadas configurações nos parâmetros de entrada (e.g.: URL do *feed* de notícias das fontes fidedignas, o período a ser considerado para fins busca de divulgações nas redes sociais, o idioma do conteúdo das notícias) a fim de se adequar o processo às características esperadas do *dataset* que será construído. A instância gerada pode ser atualizada de forma parcial ou completa por meio de iterativas execuções das quatro etapas do *FakeNewsSetGen* descritas a seguir.

### 4.3.1 Identificar Notícias Rotuladas

Resumida pelo Algoritmo 1, a primeira etapa do processo (identificação de notícias rotuladas) recebe como parâmetros de entrada o conjunto $S$ das fontes fidedignas a serem consultadas em busca de notícias fake e não fake previamente rotuladas e o conjunto notícias rotuladas a ser atualizado, o qual opcionalmente pode ser vazio. Cada fonte $s \in S$ possui um atributo $s.l$ que contém o endereço do *feed* de notícias de $s$. Assim, para cada $s \in S$, esta etapa recupera, a partir do feed $s.l$, o conjunto $URLs$ de endereços de notícias previamente rotuladas por $s$. Em seguida, para cada $url \in URLs$, a notícia rotulada localizada em $url$ é recuperada e armazenada em um conjunto de notícias rotuladas $N^R$, caso essa notícia ainda não exista nesse conjunto. Cada $n^r \in N^R$ possui, dentre outros, os atributos $n^r.id$ e $n^r.text$ que contêm, respectivamente, o identificador e o texto da notícia $n^r$.

**Algoritmo 1:** Identificar notícias rotuladas

**Entrada:**
(1) Fontes fidedignas selecionadas, $S$
(2) Conjunto de notícias rotuladas, $N^R$
**Saída:** Conjunto de notícias rotuladas atualizado, $N^R$

```
1  para cada s ∈ S faça
2      URLs ← ColetarURLsFeedRSS(s.l);
3      para cada url ∈ URLs faça
4          noticia ← lerNoticiaRotulada(url);
5          se noticia ∉ N^R então
6              N^R ← N^R ∪ noticia
7          fim-se
8      fim
9  fim
10 retorna N^R;
```

Um exemplo de registro de saída do conjunto de notícias rotuladas dessa etapa é ilustrado no Quadro 1.

Quadro 1 – Exemplo de registro de saída do conjunto de notícias rotuladas

**Identificador:** 1234
**Notícia:** *'Homem que morreu atropelado em São Paulo tem atestado de óbito como morte por Covid-19.'*
**Data:** 02-04-2020
**Fonte:** `MidiaFive`
**url:** <https://www.midiafive.com/2020/03/homem-que-morreu-atropelado-em-sao.html>
**Fonte Fidedigna:** `Aos Fatos` *(Agência de Checagem de Fatos)*
**Resultado da Checagem:** *'O acidente citado na peça de desinformação ocorreu, na verdade, em ...*
**Rótulo:** `FAKE`

### 4.3.2 Associar Notícias com Divulgações

Posteriormente, ocorre a segunda etapa do processo que se refere à identificação dos eventos de divulgação na rede social e nas mídias virtuais de cada notícia rotulada de $N^R$, conforme retratado no Algoritmo 2. Esta etapa realiza iterações sobre o conjunto $N^R$, de modo que, para cada $n^r \in N^R$, $n^r.text$ é preprocessado[3] a fim de se obter um conjunto de palavras chave *query* que é utilizado, juntamente com o intervalo de tempo de restrição da busca $t$, para identificar divulgações de $n^r$ na rede social, por intermédio de RI do tipo pesquisa de conteúdo, conforme apresentado na Subseção 2.3.1.

**Algoritmo 2:** Associar notícias com divulgações

**Entrada:**
(1) Notícias rotuladas, $N^R$
(2) Intervalo de tempo a ser utilizado para restringir a busca por divulgações, $t$
**Saída:** Conjunto de associações entre os identificadores de notícias e de divulgações, $ND$

```
1  para cada n^r ∈ N^R faça
2      query, URLsDIV ← ∅;
3      query ← PreProcessar(n^r.text);
4      URLsDIV ← PesquisarDivulgacoes(query,t);
5      se URLsDIV ≠ ∅ então
6          para cada urlDiv ∈ URLsDIV faça
7              ND ← ND ∪ {(n^r.id, urlDiv)}
8          fim
9      fim-se
10 fim
11 Armazenar(N^R);
12 Armazenar(ND);
13 retorna ND;
```

Essa identificação (i.e. *matching* entre os textos da *query* e das respectivas divulgações) é realizada por intermédio do acesso a serviços de consulta da RSV, pois esses serviços são um meio de indexar o conteúdo criado explicitamente pelos usuários dessas redes e fornecer, dessa forma, um suporte de pesquisa em tempo real.

Logo, esses serviços recebem *query* e $t$ como parâmetros de entrada e retornam *urls* de divulgações que são armazenadas no conjunto $URLsDIV$. Diante do exposto, torna-se importante mencionar que todas as *urls* retornadas pelos serviços de consulta da RSV são consideradas similares à respectiva notícia.

[3] O pré-processamento envolve as seguintes operações sobre *strings* textuais: *tokenização* [MULLEN et al., 2018] e a remoção de *stop words* [SHARMA; JAIN, 2015].

O processo de tokenização divide o texto em partes menores que são chamadas de *tokens* e remove dos dados textuais todas as pontuações e todos os ruídos (i.e.: palavras específicas dos MDDN que não fazem parte da notícia). *Stop words*, por sua vez, são irrelevantes para o processo de pesquisa na RSV (e.g., conjunções, pronomes e preposições).

Em seguida, para cada $urlDiv \in URLsDIV$, um par ordenado ($n^r.id$, $urlDiv$) é criado e armazenado no conjunto $ND$. Dessa forma, ao final da etapa 2, cada $nd \in ND$ possui os atributos $nd.idNoticia$ e $nd.urlDiv$ que contêm, respectivamente, o identificador da notícia $n^r$ e o endereço da divulgação de $n^r$ ocorrida na rede social ou na mídia virtual.

Por fim, os conjuntos $N^R$ e $ND$ são armazenados no *FakeNewsSet*. O Quadro 2 apresenta um exemplo de subconjunto de saída de associações entre identificadores de notícias e de divulgações dessa etapa do processo.

Quadro 2 – Exemplo de subconjunto de saída de associações entre identificadores

| **Notícia Rotulada** | **Divulgação em MDDN** |
|---|---|
| 1234 | https://twitter.com/xpto974876/status/974876974876 |
| 1234 | https://twitter.com/xpto6777670/status/67776700900 |
| 1234 | https://twitter.com/xpto5667435/status/56674359876 |
| 5678 | https://twitter.com/xpto9766789/status/98659788765 |
| 5678 | https://twitter.com/xpto7654567/status/23456754775 |
| 5678 | https://www1.folha.uol.com.br/poder/2020/11/99786 |

### 4.3.3 Coletar Divulgações

Em seguida, ocorre a terceira etapa do processo (resumida no Algoritmo 3) que se refere à coleta (i.e. recuperação) dos detalhes das divulgações de notícia provenientes das mídias virtuais e/ou das redes sociais. Essa etapa realiza iterações sobre o conjunto $ND$, de modo que para cada $nd \in ND$, um método recupera os dados da divulgação a partir de $nd.urlDiv$ sobre a notícia $nd.idNoticia$. Esses dados são então armazenados em um conjunto de divulgações $D$, onde cada $d \in D$ possui, entre outros, o atributo $d.mddn$ que indica se a divulgação provém de uma rede social $rsv$ ou de uma mídia virtual $mv$.

**Algoritmo 3:** Coletar divulgações

**Entrada:** Conjunto de associações entre os identificadores de notícias e de divulgações, $ND$
**Saída:** Conjunto de divulgações de notícia, $D$

```
1  para cada nd ∈ ND faça
2  |  D ← D ∪ Obter(nd.urlDiv);
3  fim
4  para cada d ∈ D faça
5  |  se d.mddn = rsv então
6  |  |  R ← R ∪ PesquisarRetransmissoes(d.id);
7  |  fim-se
8  fim
9  D ← D ∪ R;
10 Armazenar(D);
11 retorna D;
```

Logo após, para cada $d \in D$, caso $d.mddn$ seja do tipo $rsv$ e haja outras divulgações que tenham sido desencadeadas por $d$, essas são coletadas e armazenadas no conjunto provisório de divulgações desencadeadas $R$. Por fim, o conjunto de divulgações $D$ é acrescido de $R$ e armazenado no *FakeNewsSet*.

O Quadro 3 refere-se a um extrato da saída dessa etapa e, dessa forma, apresenta um exemplo de divulgação de notícia, destacando apenas alguns dos atributos mais importantes, a fim de facilitar a compreensão do processo.

Quadro 3 – Exemplo de divulgação de notícia

**Criada em:** "Wed Nov 09 14:12:32 +0000 2011"
**id:** 134272162902183936
**texto:** Pedestre morre atropelado na Régis Bittencourt:Um homem, de 59 anos, morreu atropelado por um...
**truncado:** falso
**componentes:** hashtags-#foraXPTO, símbolos-23443222, comentários-..., imagens-http://t.co/MqGeN3Kw,...
**fonte:** http://twitterfeed.com...
**em resposta ao usuário cujo identificador é:** nulo
**identificador do usuário:** 296277943
**estatísticas:** 529 (retransmissões), 121 (curtidas), 1940 (comentários)
**favorito:** falso
**retransmissão:** falso
**possivelmente sensível:** falso
**idioma:** pt

### 4.3.4 Coletar Usuários

Denominada *Coletar Usuários*, a quarta e última etapa do processo é responsável por recuperar os dados dos usuários envolvidos em divulgações de notícias na rede social. Apresentada no Algoritmo 4, esta etapa realiza iterações sobre o conjunto $D$, de modo que para cada divulgação $d \in D$ onde $d.mddn = rsv$, um método coleta os dados do usuário $u$ divulgador de $d$, assim como os usuários seguidores e seguidos associados a $u$ na rede social correspondente. Esses dados coletados são então armazenados inicialmente em um conjunto de usuários $U$ e, posteriormente, no *FakeNewsSet*.

**Algoritmo 4:** Coletar usuários

**Entrada:** Conjunto de divulgações da notícia, $D$
**Saída:** Conjunto de usuários divulgadores, incluindo os seus respectivos seguidores e seguidos, $U$

```
1 para cada d ∈ D faça
2 |   se d.mddn = rsv então
3 |   |   U ← U ∪ Obter(d);
4 |   fim-se
5 fim
6 Armazenar(U);
7 retorna U;
```

Um exemplo de registro de saída do conjunto de usuários divulgadores dessa etapa é ilustrado no Quadro 4.

Quadro 4 – Exemplo resumido de registro de um usuário divulgador

| |
|---|
| data de criação: 10/10/2000<br>identificador: 296277908<br>nome: Vitor da Silva<br>descrição: "santista, sagitariano e o mais importante amigo para todas as horas..."<br>website: nulo<br>imagem de perfil: link para imagem<br>quantidade de divulgações: 503<br>quantidade de seguidores: 10<br>quantidade de seguidos: 20<br>verificado: falso<br>lista de identificadores dos seguidores: (23454356, 478356832, ...)<br>lista de identificadores dos seguidos: (94383475, 124843924, ...)<br>localidade: Brasil |

É importante enfatizar que uma vez gerado, o *FakeNewsSet* pode ser atualizado por meio de reiteradas execuções do *FakeNewsSetGen*, dando origem a novas instâncias do *dataset*. Assim, novas notícias rotuladas provenientes de fontes fidedignas, bem como seus respectivos dados de propagação podem ser adicionados ao *FakeNewsSet* de forma incremental.

## 4.4 Implementação Base

A implementação base do *FakeNewsSetGen* foi desenvolvida na linguagem *Python* de forma a viabilizar customizações (i.e.: ajustes no código da solução implementada). Para tanto, o código foi dividido em etapas desacopladas, de acordo com o modelo macro-funcional apresentado. As interações entre essas etapas são realizadas por meio de leitura e escrita em arquivos texto, no formato *csv*. Dessa forma, caso seja necessário buscar associações entre notícias rotuladas em outra rede social, no *facebook* por exemplo, uma customização pode ser realizada através de ajustes específicos no código que implementa a etapa 2 do processo.

Além das adaptações no código por meio de customização, também há a possibilidade de se efetuar mudanças mais simples, no tocante ao comportamento do processo, por intermédio ajustes em parâmetros de entrada. Alguns dos parâmetros que podem ser configurados nessa implementação base são apresentados a seguir.

- **agency_collection_url** - Esse parâmetro destina-se à inclusão de *URLs* dos espaços destinados à divulgação de *feeds* de notícias por agências de checagem de fatos. Essas informações são utilizadas na etapa de coleta de notícias rotuladas.
- **virtual_media_collection_url** - Assim como as agências de checagem, as mídias virtuais renomadas também disponibilizam notícias por meio de *RSS-feed*. Logo, esse

parâmetro destina-se à inclusão das URLs referentes a essas mídias e essa informação também é utilizada pela primeira etapa do processo.

- **stopwords_lang** - Parâmetro destinado ao idioma dos textos das notícias. Utilizado principalmente pelo pré-processamento da etapa de associação.

- **stopwords_add_collection** - Devido às especificidades de cada agência de checagem, por vezes no texto da notícia rotulada há marcadores que não fazem parte da notícia (e.g, "LUPA confere", "G1 - Economia"). Todos esses marcadores devem ser incluídos neste parâmetro a fim de que esses termos sejam desconsiderados na consulta da etapa de associação.

- **query_since_date** - Esse parâmero também é utilizado na etapa de associação. Deve ser preenchido com a data inicial que vai compor a *query* da rede social.

- **query_until_date** - Esse outro parâmetro complementa o anterior, estabelecendo uma data final para a *query*.

- **max_results** - Parâmetro que indica o número máximo de divulgações que podem ser coletadas para cada notícia.

- **fake_name_collection** - Infelizmente, ainda não existe uma padronização dos rótulos utilizados pelas agências de checagem de fatos. Logo, este parâmetro deve ser preenchido com todos os rótulos que se deseja utilizar para caracterizar uma notícia como *fake* (e.g., "Falso", "F", "FALSE", "NOT TRUE"). Esse preenchimento dependerá de uma análise prévia às nomenclaturas utilizadas por cada agência de checagem considerada.

- **notFake_name_collection** - Seu emprego é similar ao parâmetro anterior e deve ser preenchido com os rótulos relacionados às notícias não *fake*.

- **dataset_dir** - Diretório que contém o arquivo com dados históricos, visto que a solução possibilita o incremento do *dataset* com novas notícias.

- **data_features_to_collect** - Esse parâmetro permite a execução parcial das etapas 3 e 4. Ou seja, por meio dele é possível especificar exatamente o que será coletado: publicações, propagações, dados de usuários, usuários seguidores e/ou usuários seguidos.

- **social_media_keys_file** - Este é um arquivo destinado a armazenar chaves individuais necessárias para a coleta de dados por meio de APIs das redes sociais. Essas chaves precisam ser obtidas através de solicitações formais à rede social. Contudo, normalmente, há uma rígida política de segurança, especialmente na disponibilização

de dados de usuários em decorrência da obrigatoriedade da anonimização dos usuários, conforme previsto na Lei Geral de Proteção de Dados Pessoais (LGPD[4]). O *Twitter*, por exemplo, só permite o acesso à sua API se o solicitante concordar com sua política, assinar uma série de termos legais e responder detalhadamente sobre os motivos pelos quais pretende utilizá-la, bem como quais são os atributos pretendidos. Detalhes sobre a política de uso da API do *Twitter* podem ser verificados em seu ambiente de desenvolvimento web [5].

[4] http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm
[5] https://developer.twitter.com/en/docs/basics/authentication/guides/access-tokens.html

# 5 ESTUDO DE CASO

## 5.1 Protótipo

### 5.1.1 Customização

A fim de demostrar a viabilidade de aplicação do processo proposto, um protótipo foi customizado a partir de uma implementação base do *FakeNewsSetGen* de forma a coletar notícias em língua portuguesa disponíveis em três agências de checagem de fatos brasileiras (*Lupa*[1], *Aos Fatos*[2] e *AFP*[3]) e em duas mídias virtuais nacionais de renome (*G1*[4] e *R7*[5]), além de obter divulgações dessas notícias ocorridas na rede social do *Twitter*.

A escolha dessas agências e mídias de renome foi motivada tanto pelas suas credibilidades quanto pela regularidade de postagens diárias nos respectivos *feeds* de notícias checadas. A opção pelo *Twitter*, por sua vez, decorreu da grande utilização dessa rede como canal de divulgação de notícias, bem como da possibilidade de obtenção de informações de divulgação de forma ampla, contemplando até mesmo informações de usuários, por meio da API disponibilizada pela plataforma.

Uma das principais customizações implementadas foi decorrente da falta de padronização dos rótulos utilizados pelas agências de checagens para classificar notícias como *fake* e não *fake*. A agência *Lupa*, por exemplo, utiliza os rótulos 'VERDADEIRO', 'VERDADEIRO MAS', 'AINDA É CEDO PARA DIZER', 'EXAGERADO', 'CONTRADITÓRIO', 'SUBESTIMADO', 'INSUSTENTÁVEL' e 'FALSO'. Neste caso, o protótipo foi implementado de forma a selecionar apenas notícias com os rótulos 'FALSO' e 'VERDADEIRO', desprezando as demais. Com relação à rede social, o protótipo foi ajustado para utilizar a interface do *Twitter* e se adequar às características e restrições tecnológicas (e.g., uso de credenciais de acesso e limite no número de consultas permitido por sessão de acesso) impostas pelo uso da API dessa rede social.

A partir de percepções obtidas no decorrer dos experimentos, outra importante customização foi realizada com o propósito de evitar que palavras chave adicionadas às notícias pelas agências de checagem fossem consideradas na construção da *query* da etapa de associação. Dessa forma, algumas *strings* (e.g., LUPA, #FAKE, AOS FATOS, FALSO)

[1] https://piaui.folha.uol.com.br/lupa/feed
[2] https://aosfatos.org/noticias/feed
[3] https://checamos.afp.com/rss.xml
[4] https://g1.globo.com/rss/g1
[5] https://noticias.r7.com/feed.xml

foram adicionadas ao conjunto de *stop words* e, consequentemente, desconsideradas da fase de consulta à rede social.

Todas as customizações realizadas no protótipo podem ser consultadas no repositório do projeto, exceto o conteúdo dos parâmetros que armazenam as credenciais do Twitter (API key, API secret key, Access token e Access token secret). Essas credenciais podem ser solicitadas ao Twitter por meio do seu ambiente de desenvolvimento[6].

O protótipo desenvolvido foi executado de forma modularizada, de acordo com as etapas do processo descritas na Seção 4.3.

### 5.1.2 Execução

Inicialmente, a etapa *"Identificar notícias rotuladas"* coletou dados históricos das fontes que disponibilizavam seus arquivos de checagem de anos anteriores. Logo após, cerca de setenta notícias provenientes de atualizações diárias das fontes indicadas foram acrescentadas ao *dataset* durante 60 dias. Após a coleta, verificou-se quais dessas notícias haviam sido divulgadas no *Twitter* e somente essas foram consideradas por este protótipo.

A fim de balancear a quantidade de notícias *fake* e não *fake*, uma amostra estratificada com trezentas notícias de cada classe foi selecionada, sendo suas notícias associadas aos seus respectivos *tweets* na rede social. Em seguida, as informações relacionadas aos *tweets* identificados, aos *retweets* desencadeados por esses *tweets*, aos usuários divulgadores, seguidores e seguidos foram coletadas.

Os dados obtidos por cada módulo executado foram persistidos e assim foi construída uma instância *FakeNewsSet*, composta de notícias publicadas entre março de 2017 e maio de 2020 (as notícias mais antigas foram capturadas a partir de agências que disponibilizavam arquivos históricos).

## 5.2 Dataset Gerado

A Tabela 3 apresenta um resumo estatístico da instância *FakeNewsSet* gerada e a Tabela 4 ilustra a distribuição das notícias do *dataset* por temática. A classificação das notícias por tema foi realizada de forma manual, sendo cada notícia classificada apenas com o tema mais evidente. Não foram considerados, portanto, mais de uma categoria para uma mesma notícia a fim de simplificar o processo, apesar de algumas notícias se referirem a mais de um tema (e.g., atualmente notícias sobre COVID referem-se, em grande parte, à política e também à saúde pública).

Dentre as estatísticas apresentadas, destaca-se o elevado desbalanceamento entre os totais de divulgações (*tweets e retweets*) de notícias *fake* e *not Fake*. Esse desbalanceamento

[6] https://developer.twitter.com/en/docs/authentication/overview

| **Recursos** | $Fake$ | $notFake$ |
|---|---|---|
| Notícias | 300 | 300 |
| Divulgações (tweets) | 20.498 | 6.506 |
| Divulgações (retweets) | 17.927 | 4.816 |
| Média de caracteres por notícia | 144,55 | 81,37 |
| Média de palavras por notícia | 24,2 | 14,05 |
| Usuários divulgadores | 15.983 | |
| Média de seguidores por usuário | 96.239 | |
| Média de seguidos por usuário | 6.120 | |
| Média de notícias por usuário | 1,53 | |
| Média de usuários por notícia | 40,7 | |

Tabela 3 – Estatísticas da instância *FakeNewsSet* gerada

| | Política | Economia | Saúde | Educação | Tecnologia | Segurança | Outros |
|---|---|---|---|---|---|---|---|
| ***Fake*** | 155 | 27 | 46 | 7 | 3 | 37 | 25 |
| ***não Fake*** | 18 | 119 | 29 | 58 | 34 | 8 | 34 |

Tabela 4 – Distribuição de notícias por tema

também foi observado no estudo de Shu, Mahudeswaran e Liu [2019], corroborando, dessa forma, a tendência que as *fake news* venham a atingir um maior volume de propagação do que as notícias *not Fake* [RUCHANSKY; SEO; LIU, 2017].

O meio mais direto para se detectar *fake news* é por intermédio de seu conteúdo textual, pois, ainda que esse tipo de notícia possua outros componentes, como imagem e áudio, o texto geralmente é falso [SHU et al., 2017]. No entanto, apesar de os textos das notícias serem ligeiramente diferentes, é difícil detectar *fake news* com base apenas em tópicos de conteúdo textual [SHU et al., 2018].

As Figuras 11 e 12 apresentam, respectivamente, os tópicos que se destacam nos textos de notícias *fake* e *não fake* da instância *FakeNewsSet*, utilizada neste Estudo de Caso. Nas notícias *fake*, palavras relacionadas a personagens políticos se destacam. Por outro lado, os tópicos de destaque das notícias *not fake* estão relacionados à pandemia da COVID-19.


Figura 11 – *Nuvem de palavras do conjunto de notícias* ***fake*** *da instância FakeNewsSet*


Figura 12 – *Nuvem de palavras do conjunto de notícias* ***não fake*** *da instância FakeNewsSet*

É importante destacar que o *dataset* gerado contém informações compatíveis com a estrutura prevista no modelo de dados conceitual exposto na Seção 4.2. Com o intuito de brevemente ilustrar o conteúdo da instância *FakeNewsSet* criada, o Quadro 5 apresenta um exemplo simplificado de registro de notícia rotulada como '*fake*' e a Figura 13 um exemplo de imagem pertencente a notícia também rotulada como '*fake*' por uma agência

de checagem.

Quadro 5 – Exemplo de *fake news* do *FakeNewsSet*

**Notícia:** *'Homem que morreu atropelado em São Paulo tem atestado de óbito como morte por Covid-19.'*
**Data:** 02-04-2020
**Fonte:** `MidiaFive`
**url:** <https://www.midiafive.com/2020/03/homem-que-morreu-atropelado-em-sao.html>
**Fonte Fidedigna:** `Aos Fatos` *(Agência de Checagem de Fatos)*
**Resultado da Checagem:** *'O acidente citado na peça de desinformação ocorreu, na verdade, em 1 de março no trecho da estrada em Mogi Guaçu, município que ainda não registrou casos de coronavírus. Além disso, o atestado exibido nas postagens checadas mostra sobrenome "de Carvalho", quando o da vítima do atropelamento é "de Cruz". O site MidiaFive, o primeiro a difundir a desinformação, copiou trechos de uma notícia publicada no "Mogi Guaçu Acontece" em 1º de março e acrescentou acusações de que o governador de São Paulo, João Doria, estaria manipulando as estatísticas de letalidade de Covid-19. Essa noticia já foi removida pelo site MidiaFive, mas a informação enganosa permanece em posts no Facebook que reúnem ao menos 10.500 compartilhamentos até a tarde desta quinta-feira.'*
**Quantidade de divulgações no Twitter:** 30
**Exemplo de divulgação no Twitter:** *'Seria muito engraçado se não fosse trágico e criminoso!!! que Homem morreu atropelado tem atestado de morte por Covid-19.'*
**Quantidade de propagações dessa divulgação**: 4
**Características da conta divulgadora**: *Segue 15114 usuários e é seguido por outros 11524.*
**Data de criação da conta:** 10-12-2009.

Figura 13 – *Foto fake de Che Guevara executando mulheres.* Fonte: Aos Fatos

## 5.3 Avaliação

Esta seção tem como objetivo mostrar que a instância do *FakeNewsSet* gerada pelo protótipo do *FakeNewsSetGen* permite realizar experimentos tanto com MLFN baseados no conteúdo da notícia quanto com métodos que se baseiam em dados de propagação, viabilizando uma comparação direta entre os desempenhos desses métodos na detecção de *fake news*.

A Tabela 5 apresenta o conjunto de métodos utilizados neste estudo de caso, evidenciando as principais informações utilizadas por cada um deles, assim como as principais classes do modelo de dados conceitual que contém essas informações. Esses métodos foram selecionados em decorrência de suas demandas distintas por dados, de seus processos de construção e/ou suas implementações estarem disponíveis, bem como dos bons resultados obtidos nos experimentos realizados nos trabalhos originais.

| **Método** | **Principais informações utilizadas** | **Principais classes** |
|---|---|---|
| ICS [FREIRE; GOLDSCHMIDT, 2019a] | Conta usuário publicador/propagador | Notícia \| Conta |
| Detective[7] [TSCHIATSCHEK et al., 2018] | Conta usuário publicador/propagador | Notícia \| Conta |
| Naive Bayes [MORAES et al., 2019] | Texto da notícia publicada | Notícia \| Componente |
| Ada Boost [MORAES et al., 2019] | Texto da notícia publicada | Notícia \| Componente |
| SVM [MORAES et al., 2019] | Texto da notícia publicada | Notícia \| Componente |
| HCS-I [FREIRE; GOLDSCHMIDT, 2020] | Conta usuário publicador/propagador | Notícia \| Conta |
| HCS-F [FREIRE; GOLDSCHMIDT, 2020] | Conta usuário / Texto notícia / Divulgação | Notícia \| Conta \| Divulgação |
| SL_RF [REIS et al., 2019] | Texto da notícia / Divulgação | Notícia \| Divulgação |
| SL_XGBOOST [REIS et al., 2019] | Texto da notícia / Divulgação | Notícia \| Divulgação |
| SL_SVM [REIS et al., 2019] | Texto da notícia / Divulgação | Notícia \| Divulgação |

Tabela 5 – Demanda dos métodos de detecção de *fake news*

Na Subseção 2.2.3 é apresentada uma breve descrição do conjunto de MLFN escolhido para este estudo. Maiores detalhes sobre o funcionamento de cada um desses métodos podem ser obtidos por meio de consulta às respectivas referências. Cabe enfatizar que, de fato, no conjunto de métodos utilizados estão exemplares que somente demandam informações sobre o conteúdo das notícias e exemplares que requerem informações sobre as propagações dessas notícias, viabilizando, portanto, a comparação entre métodos com diferentes necessidades de informação.

Para experimentação e validação do *FakeNewsSetGen*, os métodos selecionados foram aplicados ao *FakeNewsSet* através de validação cruzada [KOHAVI et al., 1995] com 10 conjuntos. Dessa forma, foram utilizadas as médias das métricas de avaliação de desempenho acurácia e F1, juntamente com as respectivas interseções entre intervalos (desvio padrão). Mesmo sem ter sido utilizada em parte dos estudos originais dos métodos em questão, a métrica F1 foi calculada a fim de analisar um possível impacto do desbalanceamento entre as quantidades de divulgações de notícias *fake* e *not fake* nos resultados da aplicação dos métodos no *FakeNewsSet*.

Em seguida, os resultados obtidos foram comparados com os originalmente publicados pelos trabalhos que propuseram tais métodos. A Tabela 6 permite comparar tais resultados.

De uma forma geral, pode-se perceber que todos os métodos obtiveram bons resultados, tanto nos *datasets* originais, quanto no *FakeNewsSet*. Além disso, percebe-se que os resultados obtidos pelos métodos a partir do *FakeNewsSet* foram comparáveis com os obtidos pelos trabalhos correspondentes nos seus respectivos *datasets*, com exceção do Naive Bayes e do SVM. Essa exceção pode ser justificada pela diferença do tamanho

| | Dataset original | | FakeNewsSet | |
|---|---|---|---|---|
| Método | Acurácia | F1 | Acurácia | F1 |
| ICS [FREIRE; GOLDSCHMIDT, 2019a] | 0,9402 | - | 0,9591 ± 0.0397 | 0.9594 ± 0.0362 |
| Detective [TSCHIATSCHEK et al., 2018] | 0,9791 ± 0.0169 | 0.9804 ± 0.0158 | 0,9793 ± 0.0264 | 0,9793 ± 0.0327 |
| Naive Bayes [MORAES et al., 2019] | 0,8400 | - | 0,7817 ± 0.0028 | 0,7783 ± 0.0028 |
| Ada Boost [MORAES et al., 2019] | 0,9300 | - | 0,9017 ± 0.0028 | 0,9016 ± 0.0028 |
| SVM [MORAES et al., 2019] | 0,9400 | - | 0,8433 ± 0.0028 | 0,8429 ± 0.0028 |
| HCS-I [FREIRE; GOLDSCHMIDT, 2020] | 0.9389 ± 0.0094 | 0.9345 ± 0.0106 | 0.9179 ± 0.0859 | 0.9222 ± 0.0968 |
| HCS-F [FREIRE; GOLDSCHMIDT, 2020] | 0,9671 ± 0.0094 | 0.9657 ± 0.0100 | 0.9639 ± 0.0281 | 0.9619 ± 0.0378 |
| SL_RF [REIS et al., 2019] | 0,8500 ± 0.0060 | 0,8100 ± 0.0080 | 0.8389 ± 0.0631 | 0.8311 ± 0.0829 |
| SL_XGBOOST [REIS et al., 2019] | 0,8600 ± 0.0070 | 0,8100 ± 0.0190 | 0.8117 ± 0.0864 | 0.8166 ± 0.0954 |
| SL_SVM [REIS et al., 2019] | 0,7900 ± 0.0300 | 0,7600 ± 0.0111 | 0.6295 ± 0.0818 | 0.7166 ± 0.0874 |

Tabela 6 – Comparação dos resultados obtidos

médio das notícias dos *datasets*, o que impacta consideravelmente nos métodos baseados em análise de texto.

Apesar de os métodos adotados terem apresentado bons resultados, há uma perceptível diferença de desempenho entre os métodos que utilizam os dados de propagação para obter a reputação do usuário divulgador e aqueles que são baseados apenas na análise do texto da notícia. Esses últimos são prejudicados não só pela tendência de utilização de textos curtos em MDDN, mas também em decorrência do contínuo aprimoramento do processo de fabricação de *fake news*, de forma que essas notícias vêm se tornando, cada vez mais, similares às notícias não *fake*.

Esses bons resultados obtidos, comparáveis com os obtidos nos trabalhos originais, bem como a aplicação e a comparação desses métodos de diferentes abordagens em uma mesma instância de *dataset* são indicadores positivos quanto à corretude dos dados coletados e, consequentemente, quanto à adequação do processo de construção de *datasets* proposto.

## 5.4 Exemplos de utilização do *FakeNewsSet*

A disseminação de *fake news* em MDDN provocou o desenvolvimento de abordagens para detecção desse tipo de notícia em várias frentes de atuação, de forma a abranger diferentes aspectos para solução do problema. Dentre essas abordagens, destacam-se as linguísticas que utilizam informações que podem ser extraídas diretamente do texto da notícia.

Baseados nessas abordagens, foram desenvolvidos métodos que utilizam a classificação gramatical e a análise de sentimentos presentes na escrita das notícias em língua portuguesa. Entretanto, a análise de sentimentos considerada nesses trabalhos se limita à utilização da polaridade.

Tendo como base essa limitação, foi desenvolvido um método estendido que classifica emoções humanas e transforma essa informação em elementos que possam ser processados

por algoritmos. Esse método utiliza instâncias *FakeNewsSet* para o treinamento dos modelos e o Apêndice A apresenta um resumo do artigo [SOUZA et al., 2020] que detalha esse trabalho.

Apesar de os *datasets* gerados pelo processo proposto terem sido originalmente destinados à aplicação de métodos de detecção de *fake news*, eles podem ser empregados para atender a diferentes demandas. Diante desse cenário, a instância *FakeNewsSet*, gerada pelo protótipo do presente trabalho, também foi utilizada como base de perguntas de um jogo didático online com a finalidade de viabilizar a expansão do senso crítico de jovens quanto à identificação de *fake news*.

Embora existam iniciativas como essa que utilizam jogos educacionais digitais (JED), os JED utilizados não dispõem de notícias escritas em Língua Portuguesa. Para suprir esta lacuna, um estudo de caso com quarenta e três alunos de ensino médio foi realizado com base em uma versão do "Jogo da Trilha" contendo notícias *fake* e não *fake*, extraídas da instância *FakeNewsSet*. Esse estudo revelou evidências de que o JED proposto contribuiu para a capacitação discente na identificação de *fake news*. O Apêndice B refere-se a um resumo do artigo que apresenta esse estudo de caso [PASSOS et al., 2020].

# 6 CONSIDERAÇÕES FINAIS

O consumo de notícias por intermédio de MDDN aumentou significativamente na última década. Todavia, infelizmente, as *fake news* também passaram a ser divulgadas nesses meios com muita frequência. Essas notícias falsas são divulgadas de forma intencional e, portanto, detectá-las não é uma tarefa trivial. Logo, abordagens computacionais vêm sendo utilizadas, com destaque para os MLFN, os quais necessitam de *datasets* para treinar e avaliar seus modelos de detecção.

Com o objetivo de analisar os *datasets* utilizados pelos MLFN existentes, este estudo se propôs a identificar os trabalhos do estado da arte relacionados à construção desses *datasets* por meio de um modelo comparativo que foi aplicado com base em alguns critérios estabelecidos, a partir de características comuns entre objetos e atributos dos *datasets* existentes.

Dessa forma, percebeu-se que, embora os MLFN recentes tenham sido projetados para considerar dados sobre a propagação de notícias em redes sociais, poucos dos *datasets* disponíveis contêm esses dados, dificultando, assim, a comparação de desempenho entre MLFN. Além disso, também foi identificado que os *datasets* existentes com dados de propagação não contêm notícias em português, o que prejudica a avaliação do MLFN nesse idioma.

Uma das contribuições obtidas por este trabalho foi a construção e aplicação de um modelo que viabiliza o estudo comparativo dos trabalhos relacionados por meio da indicação de critérios pré-estabelecidos. Dentre esses critérios do modelo proposto, destaca-se o esquema de representação do *dataset*, pois foi criada uma taxonomia de esquemas relacionados aos conjuntos de atributos considerados nos *datasets* para o estudo de *fake news* existentes.

Outra importante contribuição do presente trabalho foi o desenvolvimento do modelo de dados conceitual e do modelo macro-funcional. Com base nesses modelos, este trabalho propôs o processo *FakeNewsSetGen*, capaz de construir *datasets* compostos de dados sobre o conteúdo de notícias e de dados referentes às suas propagações, a fim de viabilizar a comparação entre métodos de detecção de *fake news*, baseados em diferentes demandas de informação.

Para ilustrar a viabilidade e adequação do *FakeNewsSetGen*, foi realizado um estudo de caso que abrange a criação de outras duas contribuições do presente trabalho: (i) um protótipo customizado a partir da implementação base do processo e, por meio da aplicação desse protótipo, (ii) um *dataset*, denominado *FakeNewsSet*, que contempla os dados de propagação de notícias no idioma português, sendo representado pelo esquema

$E_4$.

Em seguida, foram aplicados ao *FakeNewsSet* dez métodos de detecção de *fake news* com diferentes tipos de demandas de dados (sete deles exigindo dados de propagação). Os resultados obtidos pelos métodos a partir do *FakeNewsSet* foram comparáveis com os resultados obtidos pelos trabalhos correspondentes nos seus respectivos *datasets*, demonstrando o potencial de utilização do processo proposto e do *dataset* por ele construído.

Contudo, como o processo proposto se baseou na demanda de dados de MLFN existentes, o *FakeNewsSeGen* possivelmente demandará ajustes em virtude de novas demandas por atributos que podem ser geradas por novos MLFN criados. Além disso, futuras inovações funcionais de RSV podem também provocar a necessidade de adequação do processo de construção, demandando a inclusão de outros tipos de dados e/ou alterações nos atributos/associações existentes no modelo de dados conceitual.

Baseadas nas limitações da solução proposta, as perspectivas de trabalho futuro incluem: implementar interfaces com outras redes sociais e com outras fontes fidedignas; estender o processo para gerar *fake news* a partir de notícias verdadeiras, com base em técnicas de fabricação identificadas por trabalhos do estado da arte; aperfeiçoar o processo de associação entre notícias e suas divulgações nas redes sociais; criar um processo de deduplicação[1] de notícias para remover falhas de importação; registrar a similaridade entre notícias e divulgações; criar um estudo de caso para gerar uma instância *FakeNewsSet* em inglês, bem como aplicar MLFN de diferentes abordagens nessa instância; expandir a recuperação dos caminhos de propagação de cada notícia; enriquecer o processo com a identificação das contas utilizadas por *bots* sociais; e atender os princípios orientadores para a gestão de dados científica FAIR[2] [WILKINSON et al., 2016].

[1] Processo responsável por analisar, identificar e remover duplicidade nos dados

[2] Dados FAIR são dados que atendem a padrões de encontrabilidade, acessibilidade, interoperabilidade e reusabilidade.

# REFERÊNCIAS

AJAO, O.; BHOWMIK, D.; ZARGARI, S. Sentiment Aware Fake News Detection on Online Social Networks. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2019. v. 2019-May, p. 2507–2511. ISBN 9781479981311. ISSN 15206149.

ALLCOTT, H.; GENTZKOW, M. Social media and fake news in the 2016 election. Journal of Economic Perspectives, v. 31, n. 2, p. 211–36, May 2017. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jep.31.2.211>.

BEZERRA, E. Princípios de Análise e Projeto de Sistemas com UML. Rio de Janeiro, RJ, Brazil: Elsevier, 2007. v. 2.

BONDIELLI, A.; MARCELLONI, F. A survey on fake news and rumour detection techniques. Information Sciences, Elsevier Inc., v. 497, p. 38–55, 2019. ISSN 00200255.

BOUADJENEK, M. R.; HACID, H.; BOUZEGHOUB, M. Social networks and information retrieval, how are they converging? a survey, a taxonomy and an analysis of social information retrieval approaches and platforms. Information Systems, Elsevier, v. 56, p. 1–18, 2016.

BUNTAIN, C.; GOLBECK, J. Automatically identifying fake news in popular twitter threads. In: 2017 IEEE International Con on Smart Cloud (SmartCloud). New York, NY, USA: IEEE, 2017. p. 208–215.

CASTELO, S. et al. A topic-agnostic approach for identifying fake news pages. In: Companion Proceedings of The 2019 World Wide Web Conference. New York, NY, USA: ACM, 2019. (WWW '19), p. 975–980. ISBN 978-1-4503-6675-5.

CONROY, N.; RUBIN, V.; CHEN, Y. Automatic deception detection: Methods for finding fake news. Association for Information Science and Technology, v. 52, p. 1–4, November 2015. ISSN 2330-1643.

CORDEIRO, P. R.; PINHEIRO, V. Um Corpus de Notícias Falsas do Twitter e Verificação Automática de Rumores em Língua Portuguesa. In: STIL - Brazilian Symposium in Information and Human Language Technology. Salvaldor, BA, Brazil: IEEE, 2019. p. 220–228.

DONG, X. et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In: ACM SIGKDD international Con on Knowledge discovery and data mining. New York, NY, USA: ACM, 2014. p. 601–610. ISBN 978-1-4503-2956-9.

EFTHIMIADIS, E. N. Interactive query expansion: A user-based evaluation in a relevance feedback environment. Journal of the American Society for Information Science, Wiley Online Library, v. 51, n. 11, p. 989–1003, 2000.

EMPIRICO, S. Hipotiposes pirrônicas livro i (tradução de danilo marcondes). O que nos faz pensar, v. 9, n. 12, p. 115–122, 1997.

FACELI, K.; LORENA, A.; GAMA, J.; CARVALHO, A. D. Inteligência Artificial: uma Abordagem de Aprendizado de Máquina. LTC, 2011. ISBN 9788521618805. Disponível em: <https://books.google.com.br/books?id=4DwelAEACAAJ>.

FARAJTABAR, M. et al. Fake news mitigation via point process based intervention. In: Proceedings of the 34th International Con on Machine Learning - Volume 70. Sydney, NSW, Australia: JMLR.org, 2017. (ICML'17), p. 1097–1106.

FREIRE, P. M. S.; GOLDSCHMIDT, R. R. Fake news detection on social media via implicit crowd signals. In: Proceedings of the 25th Brazilian Symposium on Multimedia and the Web. New York, NY, USA: ACM, 2019. (WebMedia '19), p. 521–524. ISBN 978-1-4503-6763-9. Disponível em: <http://doi.acm.org/10.1145/3323503.3360626>.

FREIRE, P. M. S.; GOLDSCHMIDT, R. R. Uma introdução ao combate automático às fake news em redes sociais virtuais. In: Tópicos de Gerenciamento de Dados e Informação. Fortaleza, CE, Brazil: SBC, 2019. (34th SBBD), p. 38–67. ISSN 2016-5170.

FREIRE, P. M. S.; GOLDSCHMIDT, R. R. Detecção de fake news baseada em sinais explícitos e implícitos de um crowd híbrido: uma abordagem inspirada em meta-aprendizagem. In: Tópicos de Gerenciamento de Dados e Informação. Fortaleza, CE, Brazil: SBC, 2020. (34th SBBD), p. 38–67. ISSN 2016-5170.

Gilda, S. Evaluating machine learning algorithms for fake news detection. In: 2017 IEEE 15th Student Con on Research and Development (SCOReD). Putrajaya, Malaysia: IEEE, 2017. p. 110–115.

GOLBECK, J. et al. Fake news vs satire: A dataset and analysis. In: WebSci 2018 - Proceedings of the 10th ACM Conference on Web Science. New York, NY, USA: ACM, 2018. p. 17–21.

GOLDSCHMIDT, R. R.; PASSOS, E.; BEZERRA, E. Data Mining. Conceitos, técnicas, algoritmos, orientações e aplicações. [S.l.]: Elsevier, 2015. ISBN 9788535278224.

HABERMAS, J. Teoria de la acción comunicativa: complementos y estudios prévios. [S.l.]: Madrid: Cátedra, 1982.

Helmstetter, S.; Paulheim, H. Weakly supervised learning for fake news detection on twitter. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Barcelona, Spain: IEEE, 2018. p. 274–277. ISSN 2473-991X.

HORNE, B. D.; ADALI, S. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. In: Association for the Advancement of Artificial Inteligence. New York, USA: The 2nd International Workshop on News, 2017. p. 163–173.

JANSEN, B. J.; CHOWDURY, A.; COOK, G. The ubiquitous and increasingly significant status message. interactions, ACM New York, NY, USA, v. 17, n. 3, p. 15–17, 2010.

JANZE, C.; RISIUS, M. Automatic Detection of Fake News on Social Media Platforms. In: PACIS 2017. Frankfurt, Germany: AISeL, 2017. p. 261.

KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. Ijcai. Montreal, Canada: researchgate, 1995. v. 14, p. 1137–1145.

KUMARAN, G.; CARVALHO, V. R. Reducing long queries using query quality predictors. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. [S.l.: s.n.], 2009. p. 564–571.

LIU, Y.; BROOKWU, Y. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: AAAI Con on Artificial Intelligence. Newark, New Jersey, USA: AAAI Publications, 2018. p. 354–361.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. Introduction to information retrieval. [S.l.]: Cambridge university press, 2008.

MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, Faculty of Engineering, Ain Shams University, v. 5, n. 4, p. 1093–1113, 2014. ISSN 20904479. Disponível em: <http://dx.doi.org/10.1016/j.asej.2014.04.011>.

MEJOVA, Y.; KALIMERI, K. Advertisers jump on coronavirus bandwagon: Politics, news, and business. ArXiv, abs/2003.00923, p. 90–101, 2020.

MONTEIRO, R. A. et al. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In: Computational Processing of the Portuguese Language. Cham: Springer International, 2018. p. 324–334. ISBN 978-3-319-99722-3.

MORAES et al. Data mining applied in fake news classification through textual patterns. In: Proceedings of the 25th Brazillian Symposium on Multimedia and the Web. New York, NY, USA: ACM, 2019. (WebMedia '19), p. 321–324. ISBN 9781450367639.

MORENO, J. a.; BRESSAN, G. Factck.br: A new dataset to study fake news. In: Proceedings of the 25th Brazillian Symposium on Multimedia and the Web. New York, NY, USA: ACM, 2019. (WebMedia '19), p. 525–527. ISBN 9781450367639.

MULLEN, L. A.; BENOIT, K.; KEYES, O.; SELIVANOV, D.; ARNOLD, J. Fast, consistent tokenization of natural language text. Journal of Open Source Software, v. 3, n. 23, p. 655, 2018.

NASIM, M. et al. Real-time detection of content polluters in partially observable twitter networks. In: Companion Proceedings of the The Web Con 2018. Republic and Canton of Geneva, Switzerland: International World Wide Web Con Steering Committee, 2018. (WWW '18), p. 1331–1339. ISBN 978-1-4503-5640-4.

NEWMAN, M. L.; PENNEBAKER, J. W.; BERRY, D. S.; RICHARDS, J. M. Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin, v. 29, n. 5, p. 665–675, 2003. PMID: 15272998. Disponível em: <https://doi.org/10.1177/0146167203029005010>.

PASSOS, C. et al. Digital educational games as tools to support student training in identifying fake news written in portuguese: A case study. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). [S.l.: s.n.], 2020. v. 1, n. 1.

Pedro Faustini, T. C. Fake News Detection Using One-Class Classification. In: 2019 8th Brazilian Conference on Intelligent Systems (BRACIS). Salvador, Brazil: IEEE, 2019. p. 592–597.

PRESSMAN, R.; MAXIM, B. Engenharia de Software-8ª Edição. Connecticut, EUA: McGraw Hill Brasil, 2016.

PéREZ-ROSAS, V.; KLEINBERG, B.; LEFEVRE, A.; MIHALCEA, R. Automatic detection of fake news. In: International Conference on Computational Linguistics. Santa Fe, New Mexico, USA: arXiv.org, 2018. p. 3391–3401.

QIAN, F. et al. Neural user response generator: Fake news detection with collective user intelligence. IJCAI International Joint Conference on Artificial Intelligence, v. 2018-July, p. 3834–3840, 2018. ISSN 10450823.

REIS, J. C. S.; CORREIA, A.; MURAI, F.; VELOSO, A.; BENEVENUTO, F. Supervised learning for fake news detection. IEEE Intelligent Systems, v. 34, n. 2, p. 76–81, March 2019. ISSN 1541-1672.

RODRIGUES, J. G. O relativismo acerca da verdade refuta-se a si mesmo? Revista Portuguesa de Filosofia, Revista Portuguesa de Filosofia, v. 69, n. 3/4, p. 777–806, 2013. ISSN 08705283, 2183461X. Disponível em: <http://www.jstor.org/stable/23785891>.

RUBIN, V. L.; CONROY, N. J. Towards news verification: Deception detection methods for news discourse. In: Hawaii International Conference on System Sciences. Ontario, CANADA: WebSci 2018 - Proceedings of the 10th ACM Conference on Web Science, 2015.

RUCHANSKY, N.; SEO, S.; LIU, Y. Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Con on Information and Knowledge Management. New York, NY, USA: ACM, 2017. (CIKM '17), p. 797–806. ISBN 978-1-4503-4918-5. Disponível em: <http://doi.acm.org/10.1145/3132847.3132877>.

SALTON, G. Automatic information organization and retrieval. [S.l.]: McGraw Hill Text, 1968.

SANTIA, G. C.; WILLIAMS, J. R. Buzz face: A news veracity dataset with facebook user commentary and egos. In: 12th International AAAI Conference on Web and Social Media. Pennsylvania, USA: AAAI, 2018. p. 531–540. ISBN 9781577357988.

SHARMA, D.; JAIN, S. Evaluation of stemming and stop word techniques on text classification problem. International Journal of Scientific Research in Computer Science and Engineering, v. 3, n. 2, p. 1–4, 2015.

SHU, K.; MAHUDESWARAN, D.; LIU, H. Fakenewstracker: a tool for fake news collection, detection, and visualization. Computational and Mathematical Organization Theory, v. 25, n. 1, p. 60–71, Mar 2019. ISSN 1572-9346. Disponível em: <https://doi.org/10.1007/s10588-018-09280-3>.

SHU, K.; MAHUDESWARAN, D.; WANG, S.; LEE, D.; LIU, H. Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. September 2018.

SHU, K. et al. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data, v. 8, n. 3, p. 171–188, 2020. PMID: 32491943. Disponível em: <https://doi.org/10.1089/big.2020.0062>.

SHU, K.; SLIVA, A.; WANG, S.; TANG, J.; LIU, H. Fake news detection on social media: A data mining perspective. SIGKDD Explor. Newsl., ACM, New York, NY, USA, v. 19, n. 1, p. 22–36, set. 2017. ISSN 1931-0145. Disponível em: <http://doi.acm.org/10.1145/3137597.3137600>.

SOUZA, M.; SILVA, F.; FREIRE, P.; GOLDSCHMIDT, R. A linguistic-based method that combines polarity, emotion and gramatical characteristics to detect fake news in portuguese. In: Proceedings of the 25th Brazillian Symposium on Multimedia and the Web. New York, NY, USA: ACM, 2020. (WebMedia '20), p. 241–248.

SPONHOLZ, L. O que é mesmo um fato? conceitos e suas conseqüências para o jornalismo. In: Revista Galáxia. [S.l.]: PUC-SP, 2009. p. 56–69. ISSN 1982-2553.

SRIVASTAVA, A. et al. Factcheck: Keeping activation of fake news at check. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2018. (AAMAS '18), p. 2079–2081.

TAUSCZIK, Y. R.; PENNEBAKER, J. W. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology, v. 29, n. 1, p. 24–54, 2010. Disponível em: <http://jls.sagepub.com>.

TSCHIATSCHEK, S. et al. Fake news detection in social networks via crowd signals. In: Companion Proceedings of the The Web Con 2018. Republic and Canton of Geneva, Switzerland: International World Wide Web Con Steering Committee, 2018. (WWW '18), p. 517–524. ISBN 978-1-4503-5640-4. Disponível em: <https://doi.org/10.1145/3184558.3188722>.

VOSOUGHI, S.; MOHSENVAND, M. N.; ROY, D. Rumor gauge: Predicting the veracity of rumors on twitter. ACM Trans. Knowl. Discov. Data, ACM, New York, NY, USA, v. 11, n. 4, p. 50:1–50:36, jul. 2017. ISSN 1556-4681. Disponível em: <http://doi.acm.org/10.1145/3070644>.

WANG, P. et al. Is this the era of misinformation yet: Combining social bots and fake news to deceive the masses. In: Companion Proceedings of the The Web Con 2018. Republic and Canton of Geneva, Switzerland: International World Wide Web Con Steering Committee, 2018. (WWW '18), p. 1557–1561. ISBN 978-1-4503-5640-4.

WANG, W. Y. "liar, liar pants on fire": A new benchmark dataset for fake news detection. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017. p. 422–426. Disponível em: <https://www.aclweb.org/anthology/P17-2067>.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.-W.; SANTOS, L. B. da S.; BOURNE, P. E. et al. The fair guiding principles for scientific data management and stewardship. Scientific data, Nature Publishing Group, v. 3, n. 1, p. 1–9, 2016.

WOLOSZYN, V.; NEJDL, W. Distrustrank: Spotting false news domains. In: Proceedings of the 10th ACM Con on Web Science. New York, NY, USA: ACM, 2018. (WebSci '18), p. 221–228. ISBN 978-1-4503-5563-6. Disponível em: <http://doi.acm.org/10.1145/3201064.3201083>.

WU, L.; LIU, H. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In: Proceedings of the Eleventh ACM International Con on Web Search and Data Mining. New York, NY, USA: ACM, 2018. (WSDM '18), p. 637–645. ISBN 978-1-4503-5581-0. Disponível em: <http://doi.acm.org/10.1145/3159652.3159677>.

YANG, F. et al. Xfake: Explainable fake news detector with visualizations. In: The World Wide Web Conference. New York, NY, USA: ACM, 2019. (WWW '19), p. 3600–3604. ISBN 978-1-4503-6674-8. Disponível em: <http://doi.acm.org/10.1145/3308558.3314119>.

ZHANG, Q. et al. Ranking-based method for news stance detection. In: Companion Proceedings The Web Con 2018. Rep. Canton of Geneva, Switzerland: International World Wide Web Con Steering Committee, 2018. (WWW '18), p. 41–42. ISBN 978-1-4503-5640-4.

Zhou, X.; Zafarani, R. Fake News: A Survey of Research, Detection Methods, and Opportunities. 2018. arXiv:1812.00315 p.

# APÊNDICE A – MÉTODO DE DETECÇÃO DE *FAKE NEWS* QUE UTILIZA INSTÂNCIA *FAKENEWSSET*

Dentre as abordagens para a detecção automatizada de *fake news* destaca-se a abordagem linguística, que se caracteriza pela utilização de informações extraídas diretamente dos textos das notícias [BONDIELLI; MARCELLONI, 2019] [CONROY; RUBIN; CHEN, 2015]. Baseados neste tipo de abordagem, foram criados métodos promissores que utilizam a classificação gramatical associada à análise de sentimentos [AJAO; BHOWMIK; ZARGARI, 2019], inclusive para notícias escritas em língua portuguesa [MORAES et al., 2019].

Apesar de a análise de sentimentos ser caracterizada pela aplicação de técnicas computacionais que identifiquem a polaridade (opinião) e/ou a emoção presente em um texto [MEDHAT; HASSAN; KORASHY, 2014], até onde foi possível observar, os métodos de detecção de *fake news* desenvolvidos até o momento concentram-se exclusivamente na determinação da polaridade, ou seja, se o texto da notícia em análise expressa opinião positiva, negativa ou neutra.

A utilização da análise de sentimentos para identificar *fake news* restrita apenas à polaridade se apresenta como uma limitação devido, basicamente, a dois fatores. O primeiro é que os resultados dos próprios trabalhos de detecção de *fake news* baseados em polaridade apontam que há pouca capacidade de separação das notícias em *fake* e não *fake* na tarefa de classificação dos textos [MORAES et al., 2019]. O segundo fator limitante pode ser percebido a partir dos estudos que indicam que emoções contidas na escrita dos textos são pontos reveladores de características psicológicas ou sociais do contexto onde seu autor está inserido [TAUSCZIK; PENNEBAKER, 2010] [NEWMAN et al., 2003], podendo, inclusive, identificar se a narrativa possui indícios de não ser verdadeira.

Diante da limitação dos métodos de detecção de notícias *fake* baseados em análise de sentimento em se restringirem à identificação da polaridade dos textos, este estudo levanta a hipótese de que a ampliação do uso de técnicas de análise de sentimentos, em particular com a inclusão da classificação de emoções, associadas a classificação gramatical dos textos das notícias pode viabilizar a construção de modelos de detecção de *fake news* em língua portuguesa, mais robustos que os existentes na literatura.

Denominado FNE, o método proposto pelo presente trabalho constrói modelos de aprendizado de máquina voltados à detecção de *Fake News*. Para tanto, baseia-se na combinação de informações linguísticas, entre elas classificações gramaticais, polaridade e emoções extraídas de textos escritos de notícias previamente rotuladas como *fake* e *não*

*fake*. A Figura 14 apresenta uma visão macro-funcional do método proposto.

O FNE recebe como entrada um conjunto de notícias N, onde cada notícia $n \in N$ possui dois atributos: $n.t$ que contém o texto divulgado em n e $n.r$ o rótulo indicando se $n$ é *fake* ou *não fake*. As próximas seções detalham as etapas do método proposto.

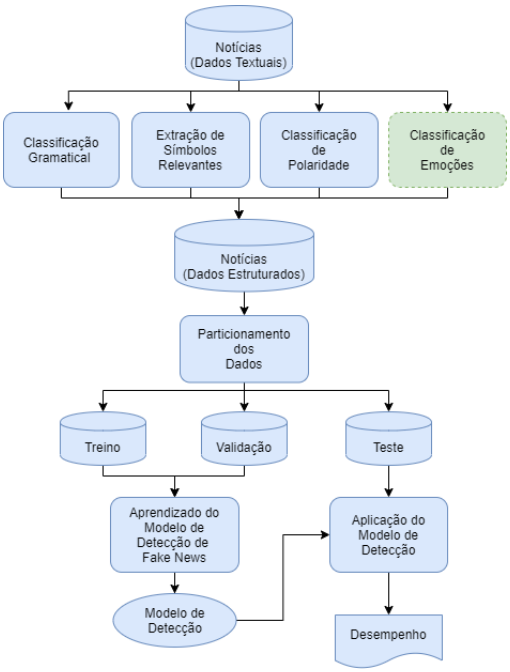



Figura 14 – Modelo Conceitual

## A.1 Classificação Gramatical

Para cada notícia $n \in N$, esta etapa tem como objetivo identificar todas as palavras ou símbolos de pontuação, também chamados de *tokens*, existentes em n.t, gerando um conjunto ordenado $TK_{n.t} = \{p_1, p_2, ..., p_k\}$, onde cada $p_i$ é um *token*. Assim, para cada token $p_i \in TK_{n.t}$, esta etapa aplica um processo de etiquetagem $\alpha(p_i)$ cujo objetivo é identificar a classe gramatical $cg \in CG$, a qual $p_i$ pertence , onde $CG = \{cg_1, cg_2, ..., cg_{|CG|}\}$ é o conjunto de classes gramaticais consideradas. É importante observar, neste ponto, que o FNE é configurável, cabendo ao analista de dados escolher a implementação de $\alpha$ a ser adotada e, consequentemente, o conjunto $CG$ de classes gramaticais a ser utilizado.

Para otimizar a análise pelos algoritmos de aprendizado de máquina, a frequência de *tokens* em cada classe é contabilizada, sendo os valores resultantes organizados em uma matriz linha $T_{n.t}$ de forma que cada coluna indica a quantidade de *tokens* etiquetados em uma das $|CG|$ classes de $CG$, conforme mostra a Equação (A.1). Nela $|n.t_{cg_j}|$ corresponde à quantidade de *tokens* de $n.t$ cuja classe gramatical é $cg_j$. Cabe destacar que as referidas quantidades são normalizadas pelo total de *tokens* em $TK_{n.t}$ (ie., k).

$$T_{n.t} = \frac{1}{k}\Big[\mid n.t_{cg_1} \mid, \mid n.t_{cg_2} \mid, ..., \mid n.t_{cg_{|CG|}} \mid\Big] \tag{A.1}$$

## A.2 Identificação de Símbolos Relevantes

Esta etapa analisa cada símbolo $s$ em $n.t$ de forma a contabilizar símbolos considerados relevantes para o processo de detecção de *Fake News*, como, por exemplo, pontos de exclamação, aspas, caracteres em caixa alta, entre outros. Segundo Moraes et al. [2019], quanto mais frequentes forem esses símbolos em um texto, maior a possibilidade de que esse texto seja *fake*. A fim de flexibilizar a escolha de quais símbolos especiais devem ser considerados, o FNE permite que o analista de dados especifique o conjunto de símbolos relevantes $CSR = \{s_1, s_2, ..., s_{|CSR|}\}$, onde, conforme o nome sugere, cada $s_j$ é um símbolo relevante.

Assim, esta etapa percorre a cadeia $n.t$ de forma a contabilizar, para cada $s_j \in CSR$, $|n.t_{s_j}|$, i.e., o número de vezes que o símbolo $s_j$ ocorreu em $n.t$. Após contabilizar a frequência de todos os termos de $CSR$, uma matriz linha $C_{n.t}$ é construída, como mostra a Equação A.2, onde $z$ é a quantidade total de caracteres em $n.t$.

$$C_{n.t} = \frac{1}{z}\Big[\mid n.t_{s_1} \mid, \mid n.t_{s_2} \mid, ..., \mid n.t_{s_{|CSR|}} \mid\Big] \tag{A.2}$$

## A.3 Classificação de Polaridade

Para cada *token* $p_i \in TK_{n.t}$, esta etapa aplica a função parcial $P : L \rightarrow \{-1, 0, 1\}$ que procura se $p_i$ pertence ao léxico $L$ a fim de recuperar o valor de polaridade associado a $p_i$. Os valores de polaridade recuperados são somados à polaridade total de $n.t$, conforme indicado na Equação (A.3). Para estabelecer uma independência em relação aos diferentes tamanhos de texto, os valores de polaridade $P_{n.t}$ são normalizados (i.e., processados de forma a assumir valores no intervalo $[0, 1]$), considerando os resultados de polaridade obtidos para todas as notícias do *dataset*. É importante destacar, neste ponto, que cabe ao analista de dados configurar o FNE com o léxico de polaridade desejado, incluindo a função $P$ a ser utilizada.

$$P_{n.t} = \sum_{i=1}^{k} P(p_i) \tag{A.3}$$

## A.4 Classificação de Emoção

Usando léxicos afetivos (i.e., dicionários que relacionam emoções às palavras de um texto), é possível extrair atributos que permitem identificar a presença de emoções em textos [TAUSCZIK; PENNEBAKER, 2010].

Cada palavra $p_i \in n.t$, é classificada inicialmente segundo uma função binária $affect(p_i)$, cujo resultado é 1, caso $p_i$ pertença ao léxico afetivo escolhido pelo analista, ou 0, caso contrário. Então, cada $p_i$ em que $affect(p_i) = 1$ é submetida às funções binárias mutuamente exclusivas $posemo(p_i)$ ou $negemo(p_i)$. Caso a emoção relacionada à $p_i$ seja positiva no léxico afetivo, então $posemo(p_i) = 1$ e $negemo(p_i) = 0$. Caso seja negativa, então $posemo(p_i) = 0$ e $negemo(p_i) = 1$. Cada $p_i$ com $negemo(p_i) = 1$ pode ainda ser subclassificada em uma de três emoções negativas: raiva, ansiedade ou tristeza. Para tanto, são utilizadas, respectivamente, as funções binárias $anger(p_i), anx(p_i)$ e $sad(p_i)$, que retornam valor 1, caso $p_i$ se enquadre na referida emoção, e 0, caso contrário. Como resultado deste processo, tem-se uma matriz linha $E_{n.t}$ representada na Equação (A.4), onde cada coluna indica o somatório de ocorrências de *tokens* em cada uma das seis categorias afetivas apresentadas acima.

$$E_{n.t} = \left[\sum_{i=1}^{k} affect(p_i), \sum_{i=1}^{k} posemo(p_i), \sum_{i=1}^{k} negemo(p_i), \sum_{i=1}^{k} anger(p_i), \sum_{i=1}^{k} anx(p_i), \sum_{i=1}^{k} sad(p_i)\right] \tag{A.4}$$

## A.5 Experimentos e Resultados

De forma a obter evidências experimentais que apontem para a validade da hipótese levantada, foi criado um método estendido que, além da classificação gramatical e da análise de sentimentos por polaridade, identifica e utiliza a emoção presente nos textos das notícias para detectar *fake news* escritas no idioma português. Nos experimentos realizados, esse método apresentou resultados de acurácia superior a 92% na identificação de *fake news*, valor em média 1,4 p.p. superior quando comparado aos métodos que não utilizam a classificação de emoções.

# APÊNDICE B – INSTÂNCIA *FAKENEWSSET* UTILIZADA POR JOGO EDUCACIONAL DIGITAL

O problema das *fake news* tem preocupado os mais diversos segmentos sociais. Uma das estratégias para combater essas notícias é capacitar pessoas para identificá-las. Embora existam iniciativas em que tal capacitação é apoiada por jogos educacionais digitais (JED), os JED utilizados não dispõem de notícias escritas em Língua Portuguesa. Para suprir esta lacuna, a instância *FakeNewsSet*, gerada no estudo de caso do presente trabalho, foi utilizada como base de perguntas para um JED denominado "Jogo da Trilha" que exercita a identificação de *fake news* escritas em Português, além de oferecer suporte para mineração de dados sobre o desempenho de seus jogadores. Um estudo de caso com quarenta e três alunos de ensino médio revelou evidências de que esse jogo contribuiu para a capacitação discente na identificação de *fake news*.

O "Jogo da Trilha" propõe uma competição saudável entre os participantes e pode ser jogado individualmente ou em grupos de, no máximo, seis jogadores. Trata-se de um JED de tabuleiro que utiliza o mouse e o teclado do computador para que os jogadores acionem o dado e respondam às perguntas propostas, selecionadas aleatoriamente e sem reposição a partir de uma base de dados previamente configurada. As perguntas podem ser acompanhadas de imagens e são apresentadas na forma de múltipla escolha com até quatro opções de respostas.

Além disso, esse jogo pode ter seu conteúdo configurado pelo professor, além de ser capaz de coletar dados sobre o comportamento dos alunos durante as partidas ao longo do tempo. Os dados coletados podem ser posteriormente analisados pelo professor por meio de recursos de mineração de dados, a fim de apoiar o docente na avaliação discente e na identificação de lacunas de aprendizado.

No estudo de caso o "Jogo da Trilha" foi configurado de forma a desafiar os jogadores a identificar *fake news*. Assim, cada pergunta corresponde a uma afirmação cujas respostas possíveis são *fake* ou não *fake*. As perguntas são obtidas a partir das notícias contidas na instância *FakeNewsSet*. Além das perguntas, o jogo também utiliza os atributos data de publicação da notícia, a resposta correta e, se for o caso, alguma imagem associada.

A Figura 15 mostra exemplos de algumas telas do "Jogo da Trilha" configurado para o desafio de identificação de notícias *fake*. No item (a) está a interface onde os alunos escolhem o avatar desejado, informam sua idade e uma autoavaliação quanto à sua capacidade de identificar *fake news*. Para a autoavaliação, o aluno deve escolher um dentre os seguintes cinco níveis, organizados em ordem crescente de aptidão: *noob* (i.e., *newbie*, novato em inglês), iniciante, casual, avançado e *proplayer*. Cabe destacar que a

nomenclatura adotada na denominação desses níveis procurou espelhar uma terminologia que fosse familiar aos alunos no universo dos jogos digitais modernos. Nos itens (b) e (c) encontram-se ilustradas, respectivamente, a interface principal do jogo e a tela apresentada ao final de cada partida, contendo o *ranking* dos jogadores com melhor desempenho até então.


Figura 15 – Telas do JED: (a) Configuração; (b) Principal; (c) Fim de partida

A fim de buscar evidências experimentais que indiquem que JED podem ser utilizados como ferramentas de apoio ao combate às *fake news* escritas na Língua Portuguesa, foi realizado um estudo de caso com o "Jogo da Trilha". O referido estudo foi desenvolvido junto a quarenta e três alunos na faixa de 14 a 21 anos (idade média de 15,3 anos e desvio padrão 1,1) do ensino médio de uma escola da rede privada de educação do município de Foz do Iguaçu no Paraná.

Diversos resultados relacionados ao desempenho discente puderam ser apurados após a utilização do JED. A Tabela 7 apresenta um resumo estatístico global desses resultados. Por meio dela, é possível perceber que todos os estudantes jogaram exatamente duas partidas, se limitando ao mínimo de partidas solicitado durante as orientações. Acredita-se que possa não ter ficado claro para os alunos que eles poderiam ter jogado mais partidas além do mínimo mencionado. Outro aspecto interessante é quanto ao percentual médio de acertos por partida. Pelo desvio padrão elevado, percebe-se uma heterogeneidade quanto à capacidade dos discentes em discernir entre notícias *fake* e não *fake*. Similarmente, é possível identificar uma alta dispersão no tempo médio por partida.

| **Aspecto analisado** | **Média apurada ± desvio padrão** |
|---|---|
| Número de partidas por aluno | $2,00 \pm 0,00$ |
| Tempo por partida (segundos) | $341,11 \pm 230.57$ |
| Tempo de resposta por pergunta (segundos) | $9,97 \pm 4,70$ |
| Percentual de acertos por partida | $69,50 \pm 29,00$ |

Tabela 7 – Resumo dos resultados obtidos

Um gráfico em barras contendo quantidades de erros e de acertos dos jogadores, distribuídos de acordo com a temática das perguntas, é apresentado na Figura 16(a). Por meio da leitura desse gráfico, é possível deduzir que notícias *fake*, relacionadas à tecnologia, são mais difíceis de ser identificadas.

A Figura 16(b) apresenta uma comparação entre a autoavaliação dos alunos em relação à sua capacidade em identificar *fake news* e os resultados efetivamente obtidos durante o jogo[1]. Percebe-se uma clara divergência entre as impressões dos discentes quanto à sua capacidade de acerto e o seu desempenho apurado por meio do jogo.

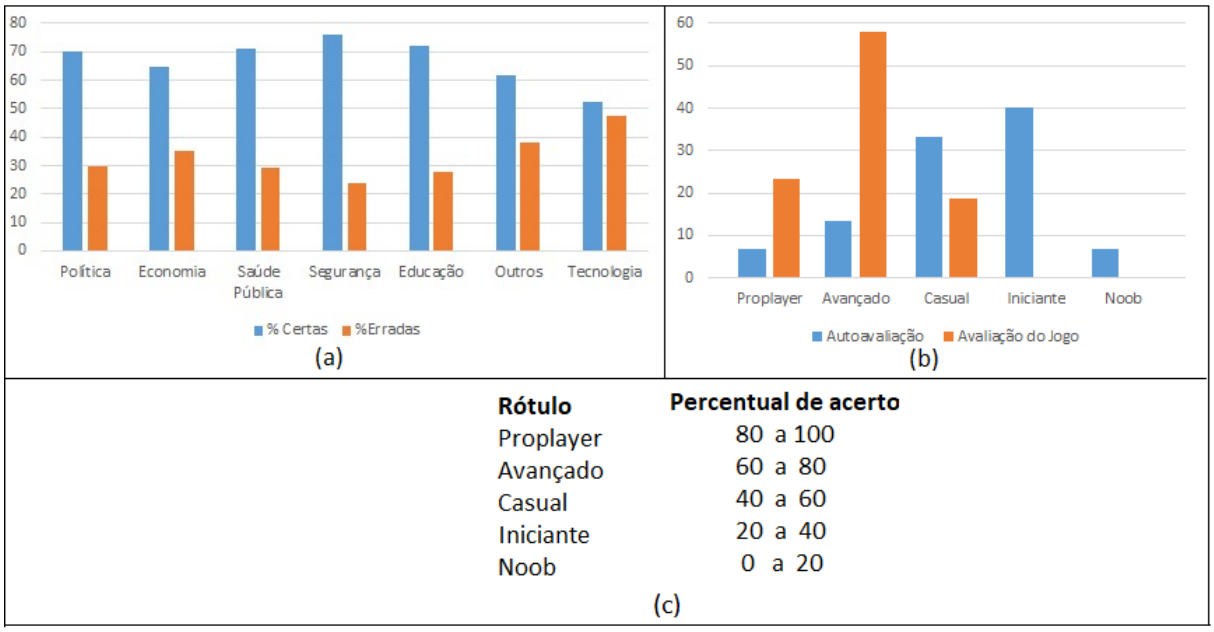
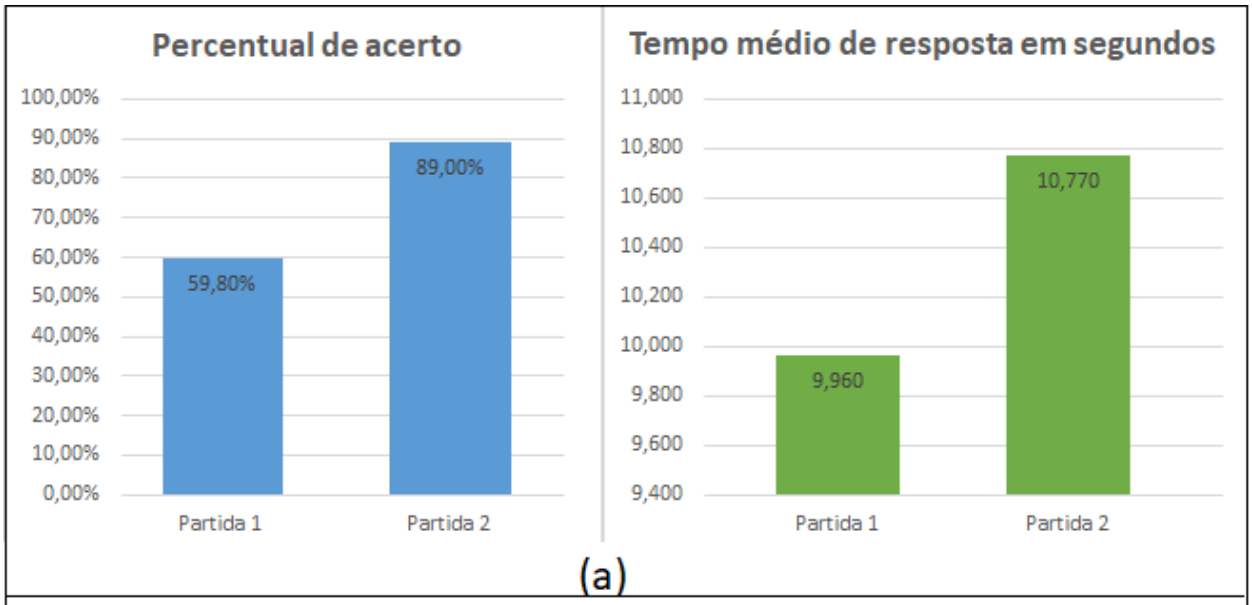



Figura 16 – (a) Acerto/erro por tema (b) Autoavaliação e desempenho real (c) Faixas

A Figura 17(a) ilustra o desempenho dos alunos nas duas partidas. Neste caso, percebe-se um aumento no percentual de acertos e no tempo médio de resposta, o que denota tanto uma melhoria na capacidade de identificação de *fake news* quanto uma análise mais atenta e criteriosa das notícias. Algumas regras de associação foram identificadas com a aplicação do algoritmo *Apriori* e três exemplos delas estão expostos na Figura 17(b). O termo consequente das três regras apresentadas (i.e.: capacidade crítica aumentou) refere-se à evolução da capacidade discente em diferenciar notícias *fake* das não *fake* de uma partida para outra. Essas regras indicam, com elevada confiança, que na primeira partida, os alunos classificados no nível avançado pelo jogo, bem como discentes que se autoavaliam casuais ou iniciantes, tendem a aumentar suas capacidades críticas.

Outra observação que merece destaque foi a capacidade dos alunos em perceber como *fake* algumas notícias cujo teor reflete situações absurdas, como a retratada na notícia *"Publicado em 11/09/2018. Jovem se esfaqueia para provar que facada de Bolsonaro foi falsa e acaba morrendo."*. Dos vinte e oito alunos que foram confrontados com esta notícia apenas um deles (3%) errou, ao indicá-la como não *fake*.

[1] A classificação dos resultados pelo jogo foi obtida a partir da discretização do número médio de acertos de cada aluno conforme as faixas de valores indicadas na Figura 16(c).




| Exemplos de regras de associação identificadas | Suporte | Confiança |
|---|---|---|
| {Avaliação do jogo (1ª partida) = avançado} => {Capacidade crítica aumentou} | 0.5833333 | 0.7777778 |
| {Autoavaliação (1ª partida) = casual} => {Capacidade crítica aumentou} | 0.1666667 | 1.0000000 |
| {Autoavaliação (1ª partida) = iniciante} => {Capacidade crítica aumentou} | 0.3333333 | 0.8000000 |

(b)

Figura 17 – (a) Desempenho dos alunos (b) Exemplos de regras de associação.