

UNIVERSIDADE FEDERAL DO PAMPA

PABLO DE ANDRADES LIMA

**CFN-AI: UM CLASSIFICADOR DE
NOTÍCIAS FALSAS BASEADO EM
INTELIGÊNCIA ARTIFICIAL**

**Bagé
2021**

PABLO DE ANDRADES LIMA

**CFN-AI: UM CLASSIFICADOR DE
NOTÍCIAS FALSAS BASEADO EM
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Engenharia de Computação como requisito parcial para a obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Érico Marcelo Hoff do Amaral
Coorientador: Sandro da Silva Camargo

**Bagé
2021**

Ficha catalográfica elaborada automaticamente com os dados fornecidos pelo(a) autor(a) através do Módulo de Biblioteca do Sistema GURI (Gestão Unificada de Recursos Institucionais).

-- Lima, Pablo de Andrades

CFN-AI: Um classificador de Notícias falsas baseado em Inteligência Artificial / Pablo de Andrades Lima.

63 p.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal do Pampa, ENGENHARIA DE COMPUTAÇÃO, 2021.

"Orientação: Érico Marcelo Hoff do Amaral; Coorientação: Sandro da Silva Camargo".

1. Aprendizado profundo. 2. Processamento linguagem natural. 3. Redes neurais convolucionais. I. Título.



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
Universidade Federal do Pampa

PABLO DE ANDRADES LIMA

**CFN-AI: UM CLASSIFICADOR DE
NOTÍCIAS FALSAS BASEADO EM
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Computação.

Trabalho de Conclusão de Curso defendido e aprovado em: 2 de outubro de 2021.

Banca examinadora:

Prof. Dr. Érico Marcelo Hoff do Amaral
Orientador
Unipampa

Prof^a. Dra. Sandra Dutra Piovesan
Unipampa

Prof. Dr. Milton Roberto Heinen
Unipampa



Assinado eletronicamente por **SANDRA DUTRA PIOVESAN, PROFESSOR DO MAGISTERIO SUPERIOR**, em 21/09/2022, às 10:55, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **MILTON ROBERTO HEINEN, PROFESSOR DO MAGISTERIO SUPERIOR**, em 21/09/2022, às 20:57, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **ERICO MARCELO HOFF DO AMARAL, PROFESSOR DO MAGISTERIO SUPERIOR**, em 21/09/2022, às 21:11, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



A autenticidade deste documento pode ser conferida no site https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0932244** e o código CRC **8F402EEF**.

Referência: Processo nº 23100.017451/2021-96 SEI nº 0932244

RESUMO

O perfil democrático e livre transformou a *Internet* em um local usual para propagação de informações, tornando este um meio comumente utilizado para o compartilhamento de falsas notícias, as ditas *fake news*. Essas notícias possuem um método de escrita diferenciado de conteúdos jornalísticos focados somente no fato em si. Desta forma, esta pesquisa busca afirmar se é possível minimizar o impacto das notícias falsas através de uma análise textual, com uso de Inteligência Artificial (IA), analisando algoritmos de IA já implementados para classificação de notícias falsas. Busca também desenvolver um modelo que classifique uma notícia como verdadeira ou falsa com uso de Inteligência Artificial, comparando o modelo desenvolvido com outros trabalhos correlatos. A pesquisa para composição dos trabalhos correlatos se deu através de uma revisão sistemática da literatura. O modelo foi desenvolvido em *Python*, no *Google Colab*. Após a importação do *dataset*, foi realizado o pré-processamento dos dados com a tokenização usando *tensorflow datasets*, construção do modelo usando a biblioteca *Keras*, análise e definição dos hiperparâmetros, treinamento do modelo, avaliação com uso de validação cruzada e testes experimentais de classificação com notícias que não estão no conjunto de dados usados para treinamento. A precisão do modelo alcançou um percentual de 95,76%, mas quando submetido a outras notícias externas a base de dados usada para treino, o valor foi de 80%. Nos resultados, é possível verificar, que o modelo proposto através Redes Neurais Convolucionais apresentou bons resultados preliminares, quando comparado às outras pesquisas para classificação de notícias. Porém há de se observar algumas inconformidades quando os algoritmos são submetidos a situações reais pós o treino.

Palavras-chave: Aprendizado profundo. Processamento linguagem natural. Redes neurais convolucionais.

ABSTRACT

The democratic and free profile turned the Internet into a common place for propagation information, making this a commonly used medium for sharing of false news, the so-called fake news. These news have a writing method differentiated from journalistic content focused only on the fact itself. Thus, this research seeks to affirm whether it is possible to minimize the impact of false news through a textual analysis, using Artificial Intelligence (AI), analyzing already implemented AI algorithms for ranking fake news. Search also develop a model that classifies news as true or false with usage of Artificial Intelligence, comparing the developed model with other works correlated. The research for the composition of the related works took place through a systematic review of the literature. The model was developed in Python, on Google Collaboration After importing the dataset, the pre-processing of the data was performed with tokenization using tensorflow datasets, model building using the Keras library, analysis and definition of the hyperparameters, model training, evaluation using cross-validation and experimental ranking tests with news not in the dataset used for training. The model's accuracy reached a percentage of 95.76%, but when submitted to other news outside the database used for training, the value was 80%. In the results, it is possible to verify that the model proposed through Networks Convolutional Neural showed good preliminary results when compared to other searches for news classification. But there are some to be observed nonconformities when algorithms are subjected to real situations after training.

Keywords: *Deep Learning, Natural Language Processing, Convolutional Neural Networks.*

LISTA DE FIGURAS

Figura 1	Diagrama Conceitual	25
Figura 2	Tipos de aprendizado	27
Figura 3	Exemplo de Rede Neural Artificial.....	28
Figura 4	Processo de seleção dos dados na camada de convolução.....	30
Figura 5	Representação do processo inicial da CNN.....	31
Figura 6	Validação Cruzada	35
Figura 7	Atividades de Metodologia.....	37
Figura 8	Modelo CRISP-DM.....	41
Figura 9	Proposta de modelo	42
Figura 10	Fases de desenvolvimento.....	43
Figura 11	Pré processamento do texto	43
Figura 12	Codificação do modelo proposto	44
Figura 13	Treinamento do modelo com validação cruzada	45
Figura 14	Submissão de novos textos	46
Figura 15	Matriz de Confusão com dados da Fiocruz	47
Figura 16	Fases de avaliação dos resultados	48
Figura 17	Matriz de Confusão	50
Figura 18	Função de perda.....	51
Figura 19	Acurácia durante do treinamento.....	52
Figura 20	Comparação de acurácia de outros trabalhos.....	53
Figura 21	Comparação de acurácia de outros trabalhos com uso de CNN.....	54
Figura 22	Logomarca do projeto.....	57

LISTA DE TABELAS

Tabela 1	Média de citações no período do trabalhos correlatos.....	19
Tabela 2	Trabalhos Correlatos finais com suas características.....	20
Tabela 3	Projetos e ações de combate a notícias falsas.....	21
Tabela 4	Classificação das <i>fake news</i>	24
Tabela 5	Modelo de Matriz de confusão	33
Tabela 6	Variação de parâmetros.....	49
Tabela 7	Variação do tamanho de lote.....	49
Tabela 8	Métricas	51

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
Braxit	Reino Unido da União Europeia
BNCC	Base Nacional Comum Curricular
CNN	<i>Convolutional Neural Network</i>
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>
DAPP	Departamento de Análise de Políticas Públicas
DL	<i>Deep Learning</i>
Glove	<i>Global Vectors for Word Representation</i>
GPOPAI	Grupo de Pesquisa em Políticas Públicas para Acesso à Informação
GPU	<i>Graphics Processing Unit</i>
HAN	<i>Hierarchical Attention Networks</i>
IA	Inteligência Artificial
LSVC	<i>Linear Support Vector Classifier</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
NILC	Núcleo Interinstitucional de Linguística Computacional
PLN	Processamento de Linguagem Natural
ReLU	<i>Rectified Linear Units</i>
RNC	Redes Neurais Convolucionais
RNN	<i>Recurrent Neural Network</i>
RNN-LSTM	<i>Recurrent Neural Network - Long Short-Term Memory</i>
RSL	Revisão Sistemática da Literatura
SVM	<i>Support Vector Machine</i>
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
UFSCar	Universidade Federal de São Carlos

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Problema de Pesquisa	11
1.2 Justificativa	12
1.3 Objetivo Geral	13
1.4 Objetivos Específicos	13
1.5 Estrutura do Texto	13
2 REFERENCIAL TEÓRICO	15
2.1 Revisão Sistemática da Literatura	15
2.2 Trabalhos Correlatos	17
2.2.1 Soluções e projetos Correlatos	21
2.3 Fake News	23
2.4 Inteligência Artificial	25
2.4.1 Aprendizado de Máquina	26
2.4.2 Redes Neurais Artificiais	27
2.4.3 Aprendizado Profundo	29
2.4.4 Redes Neurais Convolucionais	29
2.4.5 Processamento de linguagem natural.....	31
2.4.6 Hiperparâmetros	32
2.4.7 Avaliação de algoritmos de Inteligência artificial	33
3 METODOLOGIA	36
4 CFN-AI: CLASSIFICADOR DE NOTÍCIAS FALSAS BASEADO EM INTELIGÊNCIA ARTIFICIAL	41
4.1 Modelagem e desenvolvimento do modelo proposto (CFN-AI)	41
5 RESULTADOS E DISCUSSÕES	48
5.1 Comparativo com trabalhos correlatos.....	52
5.1.1 Resumo do Estudo	54
6 CONSIDERAÇÕES FINAIS	55
6.1 Projeto Pampa sem <i>Fake</i>	56
REFERÊNCIAS	59

1 INTRODUÇÃO

O perfil democrático e livre transformou a *Internet* em um local usual para propagação de informações, tornando este um meio comumente utilizado para o compartilhamento de falsas notícias, as ditas *fake news*. No Brasil, a propagação desse tipo de informação já ocorre no meio digital, tendo sido intensificado nos últimos anos principalmente durante o período eleitoral (RUEDIGER; GRASSI; GUEDES, 2018).

Para Poubel (2018), o termo *fake news* reporta à uma tradução literal de notícia falsa, sendo atualmente difundido de maneira intensa na *Internet*. O autor frisa ainda que a projeção da expressão aconteceu nas eleições de Donald Trump para presidente dos Estados Unidos, em 2016 e na saída do Reino Unido da União Europeia (*Brexit*), sendo as notícias falsas, de acordo autor, decisivas para o resultado das respectivas campanhas. Considerando diretamente as redes sociais, Resende et al. (2018) estimam que 48% da população brasileira usa o *WhatsApp* (serviço de troca de mensagens via *smartphone*) para compartilhar e discutir notícias.

O trabalho de Guarise (2019), citando Shariatmadari (2019), mostra que no *Twitter*, rede social de expressão de opinião, as informações falsas têm 70% mais possibilidades de serem compartilhadas e que se propagam mais profundamente pela grande rede. Os autores ainda ressaltam que, apesar dos robôs (do inglês, *bot*) acelerarem a propagação de notícias pela rede, o seu uso tem o mesmo efeito tanto para notícias verdadeiras quanto falsas, colocando o problema da viralidade das *fake news* como um resultado principal do comportamento humano (VOSOUGHI; ROY; ARAL, 2018).

De acordo com o levantamento realizado pelo Grupo de Pesquisa em Políticas Públicas para Acesso à Informação (GPOPPI), em 2017, cerca de 12 milhões de pessoas difundiram notícias falsas sobre política no Brasil. Gomes e Dourado (2019) em seus estudos sobre *fake news*, realizados na campanha eleitoral do Brasil de 2018, analisaram 14 relatos que circularam no mês de outubro de 2018 no Brasil, todos certificados como *fake news* por instituições de verificação de fatos no país. O tema fraude nas urnas foi compartilhado mais de um milhão de vezes (1.006.005) no *Facebook*, *Twitter* e *WhatsApp*, somadas as plataformas. De acordo com Gomes e Dourado (2019), *Facebook* e *WhatsApp* foram as plataformas primárias ou muito relevantes para a disseminação dos relatos.

Por outro lado, partindo de um cenário de pandemia (como o atual de COVID-19), para Neto et al. (2020), refletir sobre as *fake news* na contemporaneidade é pensar nas publicações com base nas evidências científicas, com termos técnicos, próprios

dos centros de pesquisa, que precisam ser decodificadas à população para melhor entendimento. O estudo traz à tona o debate em prol da saúde pública sobre as *fake news* em tempos da pandemia da COVID-19. De acordo com o Governo Brasileiro, através do Ministério da Saúde¹, as informações veiculadas e identificadas desorientam a população ao produzirem efeitos que, ao serem compartilhadas, colocam em risco as condutas diretivas.

Um outro objeto de pesquisa trata da influência do "filtro bolha" na difusão de *fake news* nas mídias sociais (SASTRE; OLIVEIRA; BELDA, 2018). Conforme descrito por Pariser (2011), o "filtro bolha" é um conceito usado para intitular algoritmos que direcionam o acesso ao conteúdo baseado no perfil e hábitos do usuário, dando uma impressão de eficiência na busca. Outro fator é que estes algoritmos restringem a maneira de como a pesquisa é realizada, sendo este método muito utilizado pelo *site* de buscas *Google* e pela rede social *Facebook*. Esses sistemas tendem a dificultar a percepção do usuário, com o intuito de mostrar somente informações de seu interesse e não o contraditório.

Desta forma, é possível afirmar que influência do filtro bolha leva os usuários a bolhas digitais de informações, onde se formam crenças acima do fato verdadeiro, evocando o que se chama de período pós-verdade, onde diminui a realidade atual dos fatos objetivando sustentar ideologias e opiniões próprias (POUBEL, 2018). O Dicionário de *Oxford*, elegeu a expressão "pós-verdade" como o termo no ano de 2016, definindo-o como "relativo ou referente a circunstâncias nas quais os fatos objetivos tem menos influência em moldar a opinião pública do que apelos à emoção e a crenças pessoais".

1.1 Problema de Pesquisa

As mídias digitais são os grandes disseminadores de conteúdos, tornando as redes sociais e os aplicativos para troca de mensagens plataformas estimuladoras dos debates democráticos (SASTRE; OLIVEIRA; BELDA, 2018). São também um local apropriado para a distribuição de notícias falsas, conforme abordam Ruediger, Grassi e Guedes (2018) no estudo do Departamento de Análise de Políticas Pública (DAPP) da Fundação Getúlio Vargas intitulado "O uso de robôs em redes sociais e a política no Brasil".

Ramalho (2018) sugere que a manipulação emocional causada pelas *fake news* pode ser usada para explicar este fenômeno, visto que o conteúdo é escrito com a

¹<https://www.saude.gov.br>

finalidade de provocar fortes emoções ao leitor, como sentimentos de raiva, indignação e frustração. Esses sentimentos dificultam o ato de realizar uma análise racional, ocasionando um senso de urgência no compartilhamento da informação (SIVEK, 2018).

Diante do exposto há de considerar, então, como problema de pesquisa, que as notícias falsas possuem um método de escrita diferenciado de conteúdos jornalísticos focados somente no fato em si (SHARIATMADARI, 2019). Desta forma, será possível assumir que se pode minimizar o impacto das notícias falsas através de uma abordagem computacional baseada em uma análise textual com uso de Inteligência Artificial (IA)?

1.2 Justificativa

Segundo Humby (2006), "Dados são o novo petróleo", e faz referência a importância e valor não só dos dados. Segundo o autor, assim como petróleo, os dados também precisam ser refinados para extração de sua capacidade analítica, neste caso, informação.

Considerando o cenário de pandemia no Brasil, de acordo com Neto et al. (2020), a doença não escolhe pessoa, nem camada social; ela ocorre indiscriminadamente. Tal como o vírus da COVID-19², o compartilhamento de notícias ocorre em paralelo, gerando informações equivocadas em saúde à população, onde metaforicamente, pode-se entender como o vírus que contamina a comunicação e promove atitudes contrárias às orientações das autoridades de saúde, promovendo o que vem se chamando de "infodemia".

Assim, os meios digitais e a massificação de acesso digital e informações em redes sociais, tem se tornado motores propulsores de conteúdos manipulados principalmente por interesses políticos e financeiros produzindo a desinformação.

Uma notícia falsa possui alguns indicadores tais como: ausência de fonte, sensacionalismos, erro de formatação ou ortografia, porém o trabalho de busca de sua veracidade é realizado por pessoas especializadas, demandando tempo para ser finalizado (KRISHNAN; CHEN, 2018).

Por se tratar atualmente de um problema digital conforme destacado no Capítulo 1 e relacionando os principais indicadores de *fake news* com a segurança da informação e a família de normas ISO 27000 que trata do Sistema de Gestão de Segurança da Informação (SGSI) e da segurança dos dados digitais, podemos destacar aqui que de acordo com a norma, os dados são elementos de segurança e os serviços que se deve garantir são a

²<https://pt.wikipedia.org/wiki/COVID-19>

confidencialidade, autenticidade, integridade e não repudição deste elemento.

Desta forma, o uso de inteligência artificial de maneira automatizada para mitigar esta prática se torna importante, visto que mentiras e informações manipuladas sempre existiram, mas nunca foram tão compartilhadas e disseminadas desta forma.

1.3 Objetivo Geral

Analisar algoritmos de inteligência artificial já implementados para classificação de notícias falsas, e desenvolver um modelo que busque classificar uma notícia como verdadeira ou falsa com uso de Inteligência Artificial comparando o modelo desenvolvido com outros trabalhos.

1.4 Objetivos Específicos

Para alcançar o objetivo geral da pesquisa serão seguidos os seguintes objetivos específicos:

- Realizar um estudo preliminar sobre os conceitos relacionados ao tema;
- Analisar e discriminar suas formas de disseminação e ferramentas digitais atuais usadas para classificação;
- Realizar uma revisão sistemática da literatura para levantar pesquisas já realizadas sobre uso de IA na solução do problema;
- Planejar e implementar uma proposta de arquitetura de um modelo de inteligência artificial para solução do problema;
- Comparar modelos de inteligência artificial e seus resultados que permitam classificar um texto como falso ou não;
- Avaliar e comparar os resultados da proposta do modelo de IA com outros trabalhos já desenvolvidos sobre o tema;

1.5 Estrutura do Texto

A partir do exposto até o momento, o restante do trabalho está organizado sequencialmente da seguinte maneira: Capítulo 2 contendo o referencial teórico e seções

que abordam os temas *fake news*, inteligência artificial, tecnologias e trabalhos correlatos. Capítulo 3 com a metodologia; no Capítulo 4 será efetuada uma descrição do projeto, análises e o modelo preditor. Para finalizar, no Capítulo 5 são exibidos os resultados e discussões bem como as conclusões no Capítulo 6.

2 REFERENCIAL TEÓRICO

O referencial teórico deste documento apresenta em sua primeira seção a RSL, seguido de trabalhos correlatos e algumas soluções já desenvolvidas e projetos que estão atuando na formação, educação midiática e conscientização. Para concluir serão apresentados os conceitos relacionados a *fake news* e Inteligência Artificial.

2.1 Revisão Sistemática da Literatura

De acordo com Sampaio e Mancini (2007), as revisões sistemáticas são desenvolvidas de maneira metódica, explícita e passíveis de reprodução. Esse tipo de estudo serve para nortear o desenvolvimento de projetos, indicando novos rumos para futuras investigações e identificando quais métodos de pesquisa foram utilizados em determinada área.

O desenvolvimento de uma RSL inclui caracterizar o tema selecionado, avaliar a qualidade, identificar conceitos, comparar as análises estatísticas apresentadas e concluir sobre o que a literatura informa em relação a determinada intervenção, apontando ainda problemas e/ou questões que necessitam de novas pesquisas. Esses estudos ajudam a sintetizar a evidência disponível na literatura sobre uma intervenção, podendo auxiliar profissionais e pesquisadores no seu cotidiano de trabalho. (SAMPAIO; MANCINI, 2007)

O processo de seleção dos trabalhos correlatos para essa pesquisa busca por documentos científicos, que usem e/ou comparem algoritmos de inteligência artificial para classificação de notícias falsas. Foi tomada então como referência na execução da referida revisão, Neiva (2016), seguindo a metodologia descrita abaixo e numerada de acordo com a literatura citada:

1. Definição das questões de pesquisa, onde a intenção principal é utilização de inteligência artificial para classificação de notícias como falsas ou não.
2. Definição das palavras chave e elencadas de acordo com o Quadro 2.1.
3. Definição de *strings* de busca descritas no Quadro 2.2.
4. A base de busca escolhida foi o *Google Scholar* por ser a plataforma que mais retornou resultados adequados aos objetivos desta pesquisa.
5. Após teste com a primeira *string* de busca na base selecionada, foi verificada a

necessidade de refinamento por retornar poucos resultados(52), sendo reorganizada novamente de acordo com o Quadro 2.3.

6. Foi então executada a busca na base com a *string* refinada a qual retornou 8590 resultados.
7. Armazenamento do resultado da busca.
8. Como critério de inclusão e exclusão, foram filtrados documentos a partir de 2017, por ser o período onde o termo *fake news* começa a ganhar evidência de acordo com *Google Trends*¹, ferramenta que mostra os mais populares termos buscados em um passado recente. Desta forma, ficou então a busca resultando em 8210 resultados.
9. Ainda dentro do processo de inclusão e exclusão, serão selecionados trabalhos em qualquer idioma, onde desses serão selecionados somente os trabalhos com média de citações no período(desde 2017) maior que zero. Para obtenção desta métrica de citações por ano foi necessário utilizar a ferramenta Harzing's Publish or Perish de Harzing et al. (1997), pois o *Google Scholar* não fornece os dados que habilitam esse tipo de classificação. Nesta fase, restaram 2321 trabalhos.
10. Na primeira etapa de seleção, com análise do título das 2321 obras, foram selecionados 95 trabalhos que de acordo com a descrição do título, se enquadram no escopo desta pesquisa, sendo posteriormente tabelados de acordo com a indicação do autor.
11. Como segunda etapa da seleção, após leitura do resumo, introdução e conclusão, foram mantidos os 31 trabalhos, por serem pesquisas que tratam do mesmo tema deste trabalho.
12. E na terceira e última fase de seleção, com leitura completa dos trabalhos e descrição dos principais itens, restaram 17 obras que serão citadas na Seção 2.2.
13. Para encerrar a metodologia de revisão sistemática da literatura, foram extraídas as respostas para a pesquisa que serão descritas e comparadas na Seção 5.1.

Após a definição das palavras chave de pesquisa, foi definida então a primeira versão da *string* de busca conforme Quadro 2.2. Após ser efetuada uma primeira busca na base de dados selecionada com a *string* do Quadro 2.2, a mesma foi refinada por ter ficado muito restrita e retornado apenas 52 resultados. Desta forma então a *string* refinada ficou conforme disposto no Quadro 2.3 a qual retornou 8590 resultados. Com a *string* refinada e o escopo de busca definido, foi dada sequência nos demais itens referentes a RSL. Foi

¹<https://trends.google.com.br/trends/explore?date=allq=fake%20news>

<i>Fake News</i>
Notícias falsas
<i>Machine Learning</i>
Aprendizado de máquina
Inteligência Artificial
<i>Artificial Intelligence</i>
<i>Deep Learning</i>
Aprendizado profundo
<i>Natural Language Processing</i>
Processamento Linguagem Natural
Redes Neurais Convolucionais
<i>Convolutional Neural Networks</i>

Quadro 2.1: Palavras Chave

<i>("fake news" OR "notícias falsas" OR "notícia falsa") AND</i>
<i>("machine learning" OR "aprendizado de máquina") AND</i>
<i>("artificial intelligence" OR "inteligencia artificial") AND</i>
<i>("deep learning" OR "aprendizado profundo") AND</i>
<i>("natural language processing" OR "processamento de linguagem natural") AND</i>
<i>("convolutional neural networks" OR "redes neurais convolucionais")</i>

Quadro 2.2: *String de busca*

realizado o armazenamento dos resultados obtidos, aplicação dos critérios de inclusão e exclusão e extração dos resultados obtidos nas pesquisas selecionadas para comparação com este trabalho no Capítulo 5.

2.2 Trabalhos Correlatos

Nesta seção serão apresentadas pesquisas científicas e trabalhos já desenvolvidos usando IA para classificação de notícias falsas, tendo como critério de seleção dos trabalhos relacionados a partir de revisão sistemática de literatura, descrita na seção anterior. Serão citadas também ferramentas, técnicas e outras tecnologias que podem ajudar na mitigação das *fake news*. Os 17 trabalhos selecionados no item 12 da RSL serão descritos abaixo com o autor e título da pesquisa.

<i>("fake news" OR "notícias falsas" OR "notícia falsa") AND</i>
<i>("artificial intelligence" OR "inteligência artificial") AND</i>
<i>("machine learning" OR "aprendizado de máquina")</i>

Quadro 2.3: *String de busca refinada*

1. **Pérez-Rosas et al. (2017):** *Automatic detection of fake news*
2. **Liu e Wu (2018):** *Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks*
3. **Ma, Gao e Wong (2018):** *Rumor detection on twitter with tree-structured recursive neural networks*
4. **Gravanis et al. (2019):** *Behind the cues: A benchmarking study for fake news detection*
5. **Silva et al. (2020):** *Towards automatically filtering fake news in Portuguese*
6. **Marín e Arroyo (2019):** *Fake news detection*
7. **Umer et al. (2020):** *Fake news stance detection using deep learning architecture (CNN-LSTM)*
8. **Aphiwongsophon e Chongstitvatana (2018):** *Detecting fake news with machine learning method*
9. **Faustini e Covoos (2020):**
textitFake news detection in multiple platforms and languages
10. **Agarwal et al. (2020):** *Fake news detection using a blend of neural networks: an application of deep learning*
11. **Thota et al. (2018):** *Fake news detection: a deep learning approach*
12. **Kong et al. (2020):** *Fake news detection using deep learning*
13. **Asghar et al. (2019):** *Exploring deep neural networks for rumor detection*
14. **Gereme e Zhu (2019):** *Early Detection of Fake News "Before It Flies High"*
15. **Marumo (2018):** *Deep Learning para classificação de fake news por sumarização de texto*
16. **Amine, Drif e Giordano (2019):** *Merging deep learning model for fake news detection*
17. **Sakurai (2019):** *Processamento de linguagem natural-detecção de fake news*

A Tabela 1, apresenta os trabalhos com o número total de citações e a média de citações no período (MCP). A consolidação desses dados foi realizada no item 9 da RSL.

Na Tabela 2 são apresentadas as seguintes características das pesquisas: Autor, algoritmo de IA e acurácia do método.

Analisando as 3 pesquisas mais citadas, o trabalho Pérez-Rosas et al. (2017) se concentra em desenvolver um recurso computacional com dois conjuntos de dados,

Tabela 1 – Média de citações no período do trabalhos correlatos

	Autor	Citações	MCP
1	Pérez-Rosas et al. (2017)	375	93,75
2	Liu e Wu (2018)	192	64
3	Ma, Gao e Wong (2018)	170	56,67
4	Gravanis et al. (2019)	54	27
5	Silva et al. (2020)	20	20
6	Marín e Arroyo (2019)	19	19
7	Umer et al. (2020)	14	14
8	Aphiwongsophon e Chongstitvatana (2018)	36	12
9	Faustini e Covoos (2020)	11	11
10	Agarwal et al. (2020)	11	11
11	Thota et al. (2018)	31	10,33
12	Kong et al. (2020)	7	7
13	Asghar et al. (2019)	13	6,5
14	Gereme e Zhu (2019)	9	4,5
15	Marumo (2018)	8	2,67
16	Amine, Drif e Giordano (2019)	5	2,5
17	Sakurai (2019)	2	1

divididos em 7 assuntos diferentes, usados para identificar propriedades linguísticas que estão presentes nos conteúdos falsos. O estudo compara empiricamente o modelo proposto com a capacidade humana de detecção de uma notícia falsa. O classificador usa *Support Vector Machine* (SVM) com validação cruzada em linguagem R. Os resultados do desempenho são mostrados em separado por assunto, onde o melhor desempenho do classificador foi com notícias relacionadas tecnologia e política, alcançando uma acurácia de 90% e 91% respectivamente. Os piores resultados foram obtidos quando as notícias falsas estão relacionadas a esportes(51%) e negócios (53%). No comparativo com a percepção humana, foram disponibilizadas notícias dos mesmos conjuntos de dados para duas pessoas solicitando as mesma que classificassem como legítimo ou falso de acordo com sua percepção. Na pesquisa, somente quando o tema é celebridades os humanos possuem desempenho melhor que o modelo computacional, nos demais temas, o classificador com uso de IA obteve melhor desempenho. Nas conclusões o autor afirma que o melhor desempenho alcançou precisões comparáveis à capacidade humana de detectar conteúdo falso.

A pesquisa de Liu e Wu (2018) se propõe a identificar notícias falsas de maneira precoce nas redes sociais através da classificação dos caminhos de propagação da notícia. O modelo foi desenvolvido com uso de *Recurrent Neural Network*(RNN) e *Convolutional Neural Network*(CNN), analisando o caminho de propagação de cada notícia como uma

Tabela 2 – Trabalhos Correlatos finais com suas características

	Autor	Algoritmo	Acurácia Geral
1	Pérez-Rosas et al. (2017)	SVM	91%
2	Liu e Wu (2018)	CNN	92%
3	Ma, Gao e Wong (2018)	RNN	72%
4	Gravanis et al. (2019)	SVM	94%
5	Silva et al. (2020)	LR	97%
6	Marín e Arroyo (2019)	SVM	93%
7	Umer et al. (2020)	CNN	97.8%
8	Aphiwongsophon e Chongstitvatana (2018)	CNN	99,9%
9	Faustini e Covoos (2020)	SVM	91%
10	Agarwal et al. (2020)	CNN	99.5%
11	Thota et al. (2018)	CNN	94.31%
12	Kong et al. (2020)	CNN	93.7%
13	Asghar et al. (2019)	CNN	82.12%
14	Gereme e Zhu (2019)	CNN	99%
15	Marumo (2018)	LSTM	78.2%
16	Amine, Drif e Giordano (2019)	CNN	96%
17	Sakurai (2019)	SVM	94.37%

série temporal multivariada em que cada tupla é um vetor numérico que representa as características de um usuário empenhado em espalhar a notícia. Os conjuntos de dados utilizados foram retirados do *Twitter* considerando o histórico de notícias nas redes sociais. Os resultados alcançaram uma acurácia de 92%, onde o autor ressalta que o modelo proposto supera abordagens de detecção de notícias falsas de última geração em termos de eficácia e eficiência, dependendo apenas de características comuns do usuário que são mais disponíveis, confiáveis e robustas do que recursos complexos, como os linguísticos ou estruturais que são amplamente usados em outras abordagens. Frisa ainda que é possível detectar notícias falsas significativamente mais rápido, por exemplo, em cinco minutos após a notícia falsa começar a se espalhar.

Em sua pesquisa, Ma, Gao e Wong (2018) usa dois modelos neurais recursivos baseados em *bottom-up* e *top-down* estruturados em árvore analisando o formato de propagação das postagens no *Twitter*. Os resultados do modelo comparado com outros modelos usando o mesmo conjunto de dados apresentou uma performance melhor alcançando 72% com um tempo de resposta na detecção menor.

Após análise dos dados das pesquisas apresentadas fica indicado, como primeiro foco de uso deste trabalho, para classificação de notícias falsas, o uso de Redes Neurais Convolucionais efetuando uma análise textual das notícias. Por outro lado, é necessário considerar após a revisão apresentada, que a IA pode ser usada de outras maneiras no

combate a disseminação de notícias falsas, seja através da análise de comportamento de usuário para identificação de robôs, mas também com análise do caminho de propagação das notícias falsas.

2.2.1 Soluções e projetos Correlatos

Complementando a seção de soluções e projetos correlatos, serão apresentados alguns sistemas já desenvolvidos para mitigar de alguma forma a disseminação de notícias falsas. Serão apresentados também, alguns projetos listados na Tabela 3, que estão colaborando para o combate a desinformação atualmente através de formação, educação midiática e conscientização, pontos esses considerados também fundamentais para a mitigação do compartilhamento e notícias falsas e desinformação.

Começando pelo *Check*, ferramenta desenvolvida pela Meedan (2020)², que é uma organização sem fins lucrativos, foca na melhora da qualidade e equidade dos dados *online*, voltadas para redações, ONGs e instituições acadêmicas. É um espaço de trabalho *online* que permite aos usuários verificar fotos e texto *online* e criar conjuntos de dados.

O *Hoaxy*³ é uma ferramenta *open source* desenvolvida pela Universidade de Indiana-IUNI e pelo CNetS com objetivo principal visualizar a disseminação de artigos online através do *Twitter* (SHAO et al., 2016).

Cabe citar ainda o NILC-USP – Detecção Automática de Notícias Falsas para o Português⁴, desenvolvido por pesquisadores do Núcleo Interinstitucional de Linguística Computacional (NILC), da USP e da Universidade Federal de São Carlos (UFSCar), através de análise textual, com uso de inteligência artificial, para servir de apoio ao usuário na identificação de notícias falsas com uma precisão de 90% (MONTEIRO et al., 2018a).

Tabela 3 – Projetos e ações de combate a notícias falsas

Nome	Link de acesso
(T3XTO, 2020)	https://anchor.fm/fato-sem-fake
(PAGANOTTI; SAKAMOTO; RATIER, 2019)	https://vazafalsiane.com
(SAYAD, 2019)	https://educamidia.org.br

Já no primeiro item da Tabela 3, o Fato sem *Fake-FsF*, também da Universidade Federal do Pampa, é parte integrante de um projeto de pesquisa e extensão do Grupo

²<https://meedan.com/check>

³<https://hoaxy.iuni.iu.edu>

⁴<https://nilc-fakenews.herokuapp.com/>

t3xto. A proposta visa contribuir socialmente com conhecimentos sobre *fake news* e desinformação, para formar combatentes da "infodemia" que corrompe o bom senso democrático. O projeto é desenvolvido através de produção de *podcasts*.

O projeto Vaza, Falsiane! é um curso *online* de iniciativa de três amigos, jornalistas e professores universitários que ao longo dos últimos dois anos estudam as *fake news*, investigando as melhores formas de produzir conteúdo sobre o assunto para um público amplo. Nesse período, foi incubado pela ONG Repórter Brasil e venceu um edital de financiamento do *Facebook*.

A EducaMídia, programa do Instituto Palavra Aberta com apoio do *Google.org* é uma entidade sem fins lucrativos que advoga a causa da plena liberdade de ideias, de pensamento e de opiniões. Com suas pesquisas, seminários e campanhas, promove a liberdade de expressão, a liberdade de imprensa e a livre circulação de informação como pilares fundamentais para o desenvolvimento de uma sociedade forte e democrática. O projeto EducaMídia foi criado para capacitar professores e organizações de ensino, engajando a sociedade no processo de educação midiática dos jovens, desenvolvendo seus potenciais de comunicação nos diversos meios. Atualmente, atua no desenvolvimento de três competências: interpretação crítica das informações, produção ativa de conteúdos e participação responsável na sociedade. Possui atuação também na formação de professores e educadores, no apoio a formuladores de políticas públicas e na sensibilização para o tema. A plataforma centraliza conteúdos para formação e pesquisa, além de materiais e recursos para a sala de aula alinhados com a Base Nacional Comum Curricular (BNCC).

Outra ação de mitigação das notícias falsas, são as agências checagem de fato, *fact checking*, que dão credibilidade as notícias *online*. De acordo com Fatos (2018), a checagem de fatos é um método jornalístico por meio do qual é possível certificar se a informação apurada foi obtida por meio de fontes confiáveis e, então, avaliar se é verdadeira ou falsa, se é sustentável ou não. Temos no Brasil hoje algumas agências de checagem de fatos como Aos Fatos⁵ e Publica⁶. Alguns portais de notícias também tem oferecido esse tipo de serviço como o G1 que possui o portal Fato ou Fake⁷ que atua de forma parecida as agências de *fact checking*.

Todas as ferramentas, projetos e ações descritas nessa seção, possuem algum método específico de atuação dentro de algum escopo, mas cabe destacar Monteiro et al.

⁵<https://www.aosfatos.org/>

⁶<https://apublica.org/>

⁷<https://g1.globo.com/fato-ou-fake/>

(2018a) que também atua com uso de inteligência artificial para classificação de notícias falsas. Já os projetos descritos na Tabela 3 atuam com educação midiática, digital, geração de conteúdo informativo, técnico e de qualidade para construção de conhecimento sobre o tema.

Após análise dos trabalhos correlatos fica indicado que se tem uma grande gama de projetos, pesquisas e ferramentas buscando soluções para mitigação das notícias falsas. Cada uma focada em um ponto mais específico, usando ou não inteligência artificial. A proposta que será descrita no Capítulo 4, busca apresentar uma solução com uso de inteligência artificial e Redes Neurais Convolucionais que possa ser comparada aos trabalhos correlatos já citados.

2.3 Fake News

Uma das autoras mais citadas quanto a conceituação e especificação do termo *fake news*, Wardle (2017) em seus textos classifica *fake news* como complicado, considerando como um ecossistema de informações e que muito mais do que notícias, o termo falso não descreve a complexidade dos diferentes tipos de desinformação, decompondo em três elementos: As motivações de quem cria este conteúdo; As formas como este conteúdo está sendo disseminado ; E por fim, os diferentes tipos de conteúdo que estão sendo criados e compartilhados, que neste caso são descritos na Tabela 4. Para Quessada e Pisa (2018), a frase atribuída a Goebbels, Ministro da Propaganda de Adolf Hitler, “uma mentira propagada mil vezes torna-se verdade” é um exemplo de como as *fake news* atuam no período pós-verdade e a velocidade que se espalham nas redes sociais para embasar opiniões.

O uso de robôs em redes sociais é um dos escopos de pesquisa que é apontado como influenciador na disseminação de notícias falsas. Ruediger, Grassi e Guedes (2018) alertam para preocupação em entender, filtrar e denunciar a disseminação de informações falsas na *Internet*. Frisam ainda que, a utilização de ferramentas automatizadas (robôs sociais), que usufruem de perfis falsos, não reais, passando-se por seres humanos e participando ativamente de debates políticos de grande repercussão, disseminando informações falsas, promovendo *hashtags* e massificando postagens automatizadas que comprometem o debate espontâneo através de *softwares* que geram este tipo de conteúdo artificialmente.

No estudo sobre “filtros bolha” de Sastre, Oliveira e Belda (2018), foram citadas

Tabela 4 – Classificação das *fake news*

	Tipo	Descrição
1	Sátira ou paródia	Não quer necessariamente causar mal, mas pode enganar o leitor
2	Falsa Conexão	A chamada da notícia não condiz com conteúdo apresentado
3	Conteúdo Enganoso	Uso mentiroso de uma informação para difamar outro conteúdo ou pessoa
4	Falso Contexto	O conteúdo é verdadeiro, mas é compartilhado com contexto falso
5	Conteúdo Impostor	Quando usa o nome de uma pessoa ou marca, mas afirmações irreais
6	Conteúdo Manipulado	O conteúdo verdadeiro é alterado para enganar o público
7	Conteúdo Fabricado	Informações 100% falsas e construídas para causar mal e espalhar boatos

Fonte: Wardle (2017)

as mudanças realizadas pelo *Facebook* com implantação do sistema de *crowdsourcing*, ou colaboração coletiva. O sistema define as prioridades dos *feeds* de notícias nos perfis dos usuários, classificando o que irá aparecer ao usuário por uma maior familiaridade com os conteúdos mais acessados, com propósito de reduzir a difusão de *fake news* através de robôs. Porém, essa configuração gerou uma repercussão negativa com empresas que utilizam estratégias de divulgação por meio de mídias digitais. De acordo com Sérgio Dávila, editor-executivo do jornal a Folha de São Paulo, considerado o maior jornal do Brasil, as redes sociais tendem a criar "bolhas" e "condomínios de convicções" forçando as pessoas a relacionar-se somente com outras que pensam como elas (CAULY, 2018).

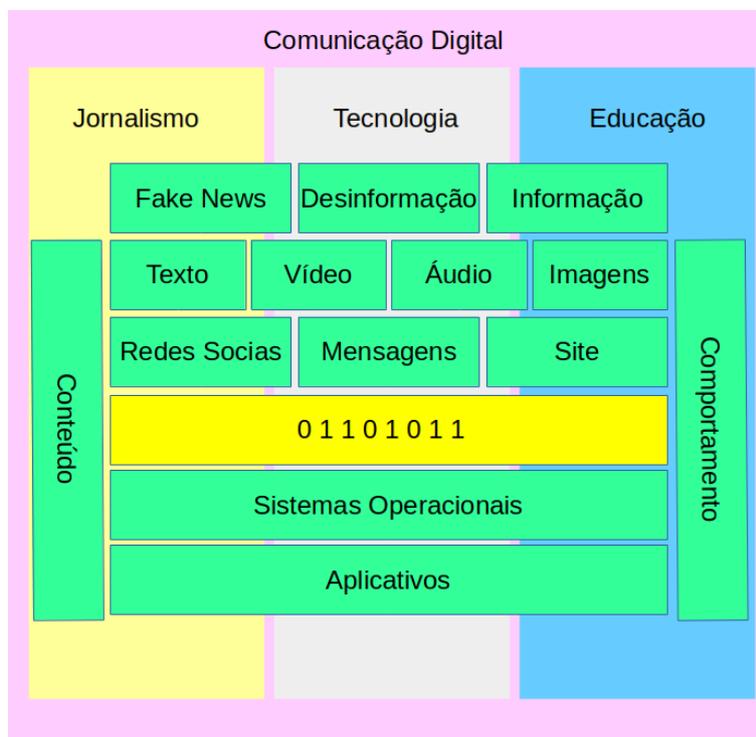
A pesquisa de Ferrara et al. (2016) mostra uma tendência promissora de evolução no combate às *fake news* utilizando o padrão automatizado de *machine learning* (ML) e inteligência humana para diferenciar robôs de pessoas.

Não menos obstante, cabe citar o crescimento das *deepfakes*, uma técnica usada para substituir rostos originais em vídeos e utilizados para disseminar notícias falsas com uso de aplicativos para a troca de rosto, gerando grande quantidade de vídeos sinteticamente manipulados e distribuídos nas redes sociais, representando um grande desafio técnico para detecção e filtragem de tal conteúdo (MORAES, 2019).

Uma das maneiras de visualizar o problema e suas possíveis soluções é a separação dele por "caixas", ilustrado na Figura 1. Nela fica ressaltado o problema, as áreas do conhecimento envolvidas, a forma de atuação, disseminação e como os possíveis métodos

de solução atuariam especificadamente dentro de uma ou mais "caixas".

Figura 1 – Diagrama Conceitual



Fonte: Autor (2021)

O diagrama mostra no fundo um problema de comunicação digital, que no centro na verdade temos os dados em formato de *bits*, ou seja, a sua atuação e modo de operação é digital, tendo como áreas de atuação na sua possível solução, o jornalismo, educação e tecnologia. Podemos compreender também que se trata de um problema segmentado em *Fake News*, desinformação e informação, onde esses dados transitariam em formatos de vídeos, texto, imagem ou áudio, por redes sociais, sites ou aplicativos de mensagens, sendo que estes poderiam ser tratados ou auditados através de uma análise de comportamento ou conteúdo com uso de aplicativos ou diretamente do sistema operacional.

2.4 Inteligência Artificial

O termo Inteligência Artificial foi criado por McCarthy, Minsky e Rochester (1956) nos trabalhos do *Dartmouth Summer Research Project on Artificial Intelligence*, como um novo campo de conhecimento, associando linguagens, inteligência, raciocínio, aprendizagem e resolução de problemas. A comparação de humano e máquina há muito

tempo é objeto de estudo, visto que Turing (1950) já instigava reflexões com a questão: "As máquinas podem pensar?", propondo o jogo da imitação. Em seus estudos, o autor já esperava que máquinas e homens acabariam por competir em campos puramente intelectuais, porém encerrava: "Podemos avistar só um pequeno trecho do caminho à nossa frente, mas ali já vemos muito do que precisa ser feito" (TURING, 1950).

2.4.1 Aprendizado de Máquina

O Aprendizado de Máquina (AM) ou *Machine Learning* (ML) é uma área da inteligência artificial que tem por objetivo a construção de sistemas computacionais capazes de adquirir conhecimento de forma automática, tomando suas decisões baseado em experiências já acumuladas através de problemas anteriores com solução bem sucedida (MONARD; BARANAUSKAS, 2003). De acordo com Murphy (2012), no AM, os algoritmos são capazes de realizar o reconhecimento dos padrões em conjunto de dados pré-estabelecidos, podendo ser, essencialmente, classificados como: supervisionado (preditivo), não-supervisionado (descritivo), semi-supervisionado e ainda por esforço.

Para Monard e Baranauskas (2003), os métodos de aprendizado são dispostos de maneira hierárquica, colocando a indução no topo como uma forma de inferência lógica que leva a conclusões genéricas sobre um conjunto particular de exemplos. Frisa ainda que, na indução, um conceito é aprendido através da inferência indutiva sobre os exemplos apresentados.

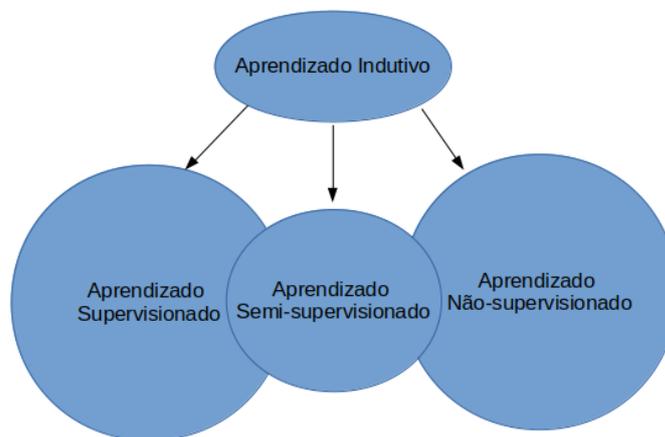
Na sequência, um nível abaixo, o aprendizado indutivo supervisionado e não-supervisionado: No aprendizado supervisionado, é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos para treinamento para os quais o rótulo da classe associada é conhecido (MONARD; BARANAUSKAS, 2003). Para Cheeseman et al. (1990), no aprendizado não-supervisionado, o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou *clusters*.

Ainda neste nível, englobando os dois conceitos anteriores, o aprendizado semi-supervisionado. Para Matsubara, Monard e Batista (2005), no aprendizado supervisionado e necessária uma grande quantidade de dados rotulados para a indução de um bom classificador. Porém, na maioria das vezes, o número de exemplos rotulados é muito pequeno, e às vezes inexistente.

Neste sentido, o aprendizado semi-supervisionado, mescla as características do

aprendizado supervisionado e do aprendizado não supervisionado, sendo uma abordagem útil em alguns casos (GUERREIRO, 2017). A Figura 2 ilustra a relação entre os conceitos dos tipos de aprendizado de máquina.

Figura 2 – Tipos de aprendizado



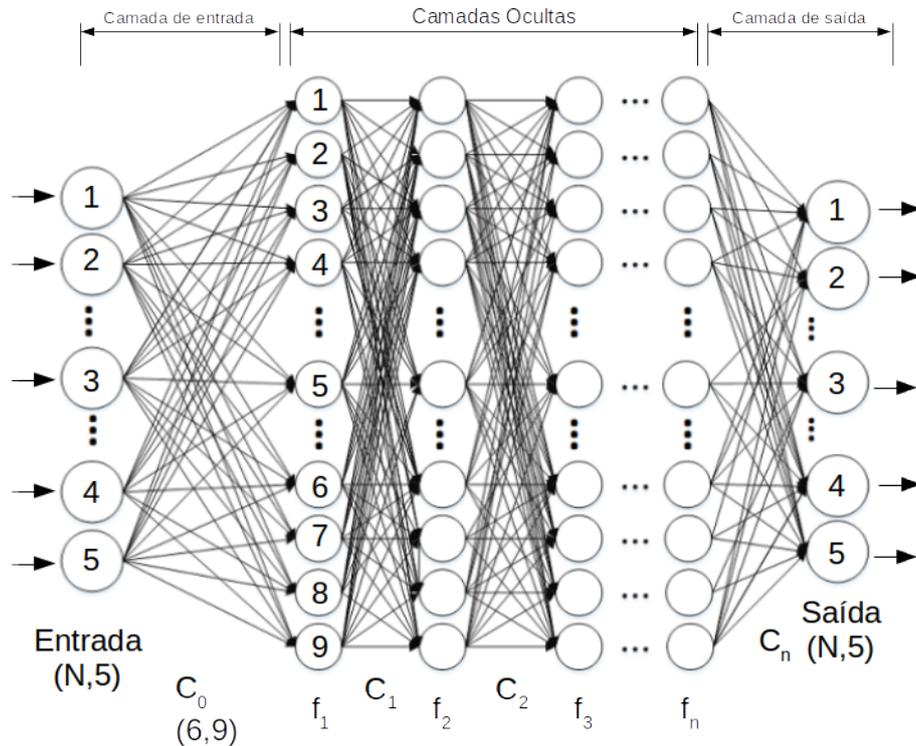
Fonte: Adaptado de Monard e Baranauskas (2003)

2.4.2 Redes Neurais Artificiais

Dentro do aprendizado de máquina temos as redes neurais artificiais. São algoritmos inspirados biologicamente no sistema nervoso de animais, muito utilizados para processar grandes quantidades de dados de treinamento com objetivo de classificar ou identificar novos resultados, semelhantes ao treinamento (BAHDANAU; CHO; BENGIO, 2014).

A arquitetura de uma rede neural artificial apresentada por Bahdanau, Cho e Bengio (2014) descreve camadas de neurônios de processamento e conexões entre eles. A partir da série de camadas conectadas e formadas por neurônios, as conexões transmitem "sinais" de um neurônio para o outro, simulando as sinapses de um cérebro. Cada conexão possui um único de peso que é ajustável durante a fase de treinamento. Cada um dos neurônios possui uma função de ativação, que define a força de propagação do sinal (ANDRADE, 2019). A Figura 3 ilustra o funcionamento da uma rede neural artificial, onde se tem (N) na entradas dos dados. Na sequência, C_0 , C_1 , C_2 e C_n possuem as matrizes de pesos e f_1 , f_2 , f_3 e f_n são as neurônios da camada oculta, finalizando com a camada de saída que fornece o valor previsto a partir da entrada.

Figura 3 – Exemplo de Rede Neural Artificial



Fonte: Adaptado de Agarwal e Mittal (2014)

Durante o processo de treinamento, são atribuídos pesos para cada conexão, que servirão como parte do processo de decisão, simulando as sinapses de um sistema nervoso animal. A função de ativação define a intensidade da ativação, onde a definição da função determina como devem ser as saídas de cada neurônio (ANDRADE, 2019).

Para Sharma e Sharma (2017), após o cálculo dos pesos, é aplicada a função de ativação para obter a saída do neurônio que será fornecida como entrada da próxima camada. O autor cita os seguintes tipos de funções de ativação:

1. *Binary Step Function*
2. *Linear*
3. *Sigmoid*
4. *Tanh*
5. *ReLU*
6. *Leaky ReLU*
7. *Parametrized ReLU*
8. *Exponential Linear Unit*
9. *Swish*

10. SoftMax

2.4.3 Aprendizado Profundo

Como uma evolução das redes neurais, o aprendizado profundo ou *deep learning* (DL), que é considerado por Deng e Yu (2014) uma subárea do *machine learning* e assim como as RNA, utiliza um conjunto de algoritmos buscando aprender diferentes níveis de abstração, utilizando várias camadas de processamento. Com a utilização de várias camadas de processamento, usando por exemplo as funções de ativação citadas, é possível realizar a detecção de padrões dos dados de entrada durante o seu processo de aprendizado e teste (MARUMO, 2018).

Goodfellow et al. (2016) classificam *deep learning* como estruturas computacionais que usam diversas camadas de redes neurais artificiais, que podem ser mescladas, podendo ainda se mesclar diferentes tipos de redes, onde cada camada tem o objetivo de aprender características do conjunto de dados com a finalidade de realizar uma classificação. Segundo Nallapati et al. (2016), a aprendizagem profunda pode ser usada para identificar padrões de texto, analisá-las com uso da técnica de mineração de texto chamada *text summarization* e, posteriormente, classificá-las de acordo com as métricas pré-definidas, não se limitando apenas para esse fim de identificação de padrões textuais, mas também para outros tipos de dados como áudio.

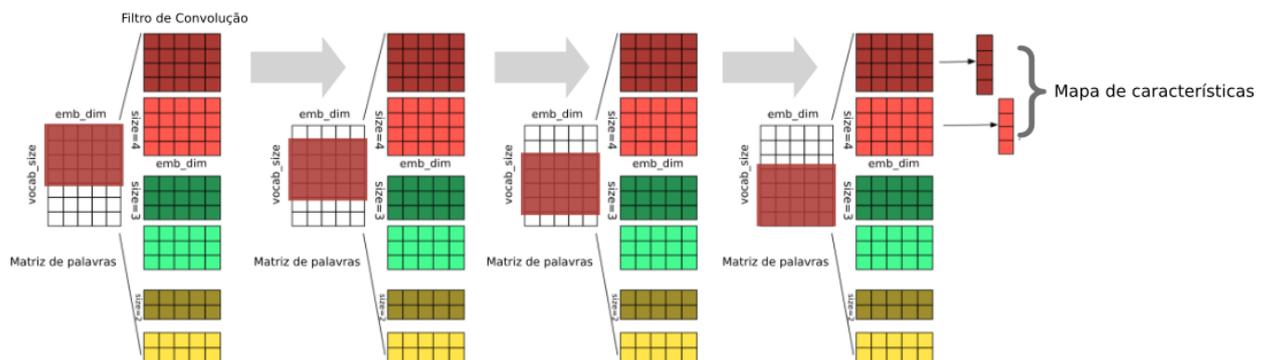
2.4.4 Redes Neurais Convolucionais

Dando continuidade, tem-se as redes as neurais convolucionais ou *Convolutional Neural Network* (CNN), que são redes neurais compostas por uma quantidade maior de multicamadas. As CNN contam inicialmente com uma ou mais camadas de convolução, onde esta camada, que pode ser interpretada como a soma dos produtos de cada elemento da matriz de palavra com seus "vizinhos" locais, pelos elementos da matriz que representa o filtro de convolução (FERREIRA, 2017). Esses filtros que possuem tamanhos que são parametrizados para extração do vetor de características, sendo que este vetor de características, será a saída da camada de convolução.

A Figura 4 ilustra como os dados são selecionados na matriz de palavras para então realizar o processo de soma e produto com o filtro de convolução e posterior

extração do mapa de características. Na entrada é recebida uma matriz de palavras que tem sua altura definida pelo número de palavras únicas, também chamado de *vocab size* e largura pelo tamanho máximo dessas palavras, ou *emb dim*, sendo que, para a matriz não ficar com espaços vazios por conta da diferença de tamanho das palavras, é executado o *post-padding*, que preenche o final de cada linha vazia da matriz com zeros.

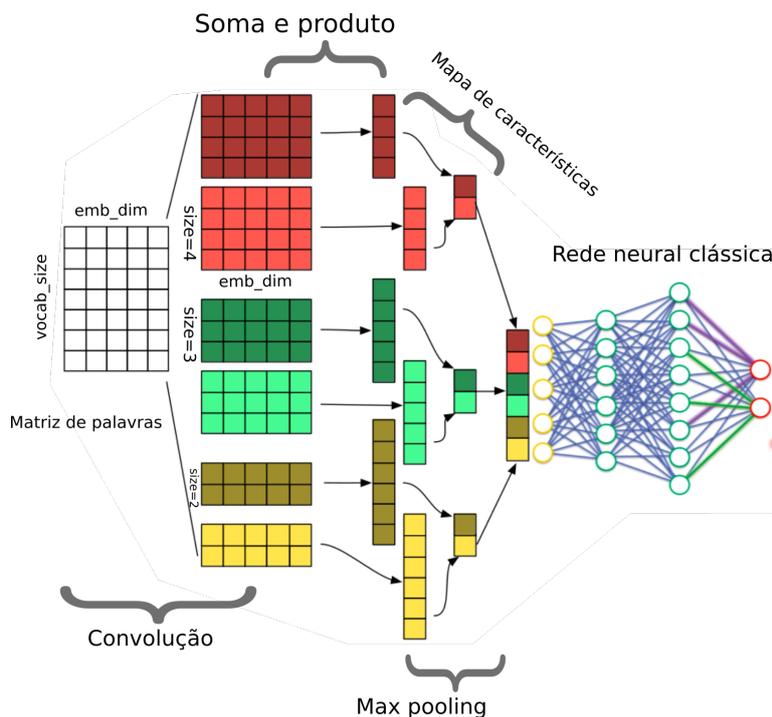
Figura 4 – Processo de seleção dos dados na camada de convolução



Fonte: Adaptado de Zhang e Wallace (2016)

Dando sequência ao processo de redução da dimensionalidade dos dados de entrada com extração das principais características do texto, é realizado o processo de *max pooling*, que retorna o valor máximo de cada vetor de características (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Para finalizar, será realizada a concatenação dos vetores, dando origem ao vetor final do mapa de características de acordo com a Figura 5. Este servirá de entrada para uma rede neural clássica que terá como saída terá duas classes, sendo a classificação da notícia como verdadeira ou falsa.

Figura 5 – Representação do processo inicial da CNN



Fonte: Adaptado de Zhang e Wallace (2016)

2.4.5 Processamento de linguagem natural

Como parte de suma importância neste trabalho, a mineração de texto, que nesta pesquisa será realizada com uso de Processamento de Linguagem Natural (PLN) que, segundo Sakurai (2019), trata da utilização de métodos e recursos computacionais para análise de dados linguísticos. Neste contexto, o PLN auxilia na resolução de ambiguidades estruturando numericamente o texto, separando palavras através do processo de "tokenização", pré-processamento com análises léxicas, sintáticas, semânticas e pragmáticas, para posterior compreensão do algoritmo de IA (DALE; MOISL; SOMERS, 2000).

Uma das formas de estruturação numérica do texto citado acima, neste caso, transformar as notícias em números, é o *Term Frequency - Inverse Document Frequency* (TF-IDF). O valor é definido pelo produto de duas medidas estatísticas, a frequência do termo (TF) ou palavra e o inverso da frequência nos documentos (IDF). O objetivo do TF-IDF é demonstrar a relevância de uma palavra em relação a um conjunto delas no texto (AIZAWA, 2003).

2.4.6 Hiperparâmetros

Todo sistema de Aprendizado de Máquina possui seus hiperparâmetros, que são definidos antes do treinamento da rede visando obter o melhor resultado possível do modelo. Atualmente existem métodos automatizados que realizam esse processo: *Grid search*, *random search* e otimização Bayesiana, definidas, por Feurer e Hutter (2019), da seguinte forma:

O *grid search*, ou pesquisa de grade, é o método mais básico, também conhecido como planejamento fatorial completo. Neste sentido, é especificado um conjunto finito de valores para cada hiperparâmetro e pesquisa de grade avalia o produto cartesiano desses conjuntos. Uma alternativa à pesquisa em grade é a busca aleatória (*random search*) que, como o nome sugere, configurações de amostras são selecionadas aleatoriamente até se encontrar o melhor parâmetro. Já a otimização Bayesiana, de acordo com autor, se trata do estado da arte sobre o tema. É um algoritmo iterativo com dois ingredientes principais: um modelo substituto probabilístico e uma função de aquisição.

De acordo com Zhang e Wallace (2016), na prática, ajustar todos hiperparâmetros simplesmente não é viável, especialmente porque a estimativa de parâmetros é computacionalmente intensiva. No entanto, os métodos de pesquisa sofisticados, acima citados, ainda exigem saber quais hiperparâmetros valem a pena explorar para começar e o intervalo razoável para cada.

Dada a explicação teórica sobre os algoritmos de redes neurais, abaixo os termos que aparecem na execução do código e suas devidas descrições:

1. *kernel size*: Tamanho dos filtros de cada região.
2. *nb filters*: Número de filtros por região.
3. *lossfunction*: Função de perda. Função usada para avaliar a qualidade ou confiança da classificação dos dados, sendo que, neste trabalho será utilizada a função de entropia cruzada ou *cross-entropy*, onde quanto mais próximo de zero o valor de saída, maior a confiança na classificação dos dados (HO; WOOKEY, 2019).
4. *ffn units*: Número de neurônios da rede neural densa.
5. *dropout*: Percentual de neurônios inativos na rede neural densa para evitar *overfitting* ou sobreajuste, que acontece quando o modelo não aprendeu os principais padrões ou tendências dos dados da base, mas sim “decorou” todos os seus casos apresentando um bom desempenho na etapa de treinamento, porém uma

baixa capacidade de generalização para outros conjuntos de dados (CECHINEL; CAMARGO, 2020).

6. *batch size*: Define de quantos em quantos registros serão atualizados os pesos.
7. *epochs*: Define quantas "passagens" para frente e para trás de todos os exemplos de treinamento.

A escolha dos hiperparâmetros a serem estudados, foram definidos também baseados na pesquisa de Zhang e Wallace (2016), pois são diretamente ligados ao tamanho e tipo dos dados de entrada (*dataset*), que neste caso foi usado em língua portuguesa. Já os valores de variação são descritos no Capítulo 3 que apresenta a metodologia.

2.4.7 Avaliação de algoritmos de Inteligência artificial

Nesta seção serão apresentadas métricas de avaliação e interpretação de modelos de classificação. A matriz de confusão é considerada uma representação gráfica das seguintes características:

- Verdadeiro Positivo (VP): Quando o modelo classifica uma notícia como verdadeira e ela é realmente verdadeira.
- Verdadeiro Negativo (VN): Trata-se das notícias classificadas como falsa e de fato são *fake*.
- Falso Negativo (FN): É quando o algoritmo classifica a notícia como *fake news* e ela é uma notícia verdadeira.
- Falso Positivo (FP): São as notícias classificadas como verdadeiras e são na realidade *fake news*.

Com os dados acima é possível realizar a elucidação através do cruzamento dos números de amostras preditas por um classificador ou detector, com a classificação real das amostras. A Tabela 5 ilustra como os dados são apresentados na matriz de confusão.

Tabela 5 – Modelo de Matriz de confusão

	Positivo	Negativo
Positivo	VP	FP
Negativo	FN	VN

A acurácia geral (AG), descrita na Equação 1, de acordo com Ye (2003), citado em Cechinel e Camargo (2020), revela o desempenho do modelo de predição calculado

através de sua exatidão, que é dada pela quantidade de acertos de classificação cometidos dividido pelo número total de casos na amostra utilizada para o teste.

$$AG = \frac{|VP| + |VN|}{|VP| + |VN| + |FP| + |FN|} \quad (1)$$

Ainda com as características apresentadas, é possível medir a precisão para encontrar notícias verdadeiras (PV), falsas (PF) e o *recall* para verdadeiras (RV) e falsas (RF) com o uso das seguintes equações:

$$PV = \frac{|VP|}{|VP| + |FP|} \quad (2)$$

$$PF = \frac{|VN|}{|VN| + |FN|} \quad (3)$$

A precisão analisa as classificações dadas como positivas, ou seja, a proporção de todas as notícias classificadas como FN, de fato eram FN (PF) e também para notícias verdadeiras classificadas como tal (PV). Já o *recall*, ou revocação, analisa a proporção de notícias que de fato são *fake news*, com todas as classificações realizadas e classificadas como *fake news*. Neste caso os Falsos Negativos estão incluídos na formulação, pois a notícia é de fato *fake*, mesmo o detector ou classificador ter predito ao contrário. Abaixo são apresentadas as equações para encontrar a métrica.

$$RV = \frac{|VP|}{|VP| + |FN|} \quad (4)$$

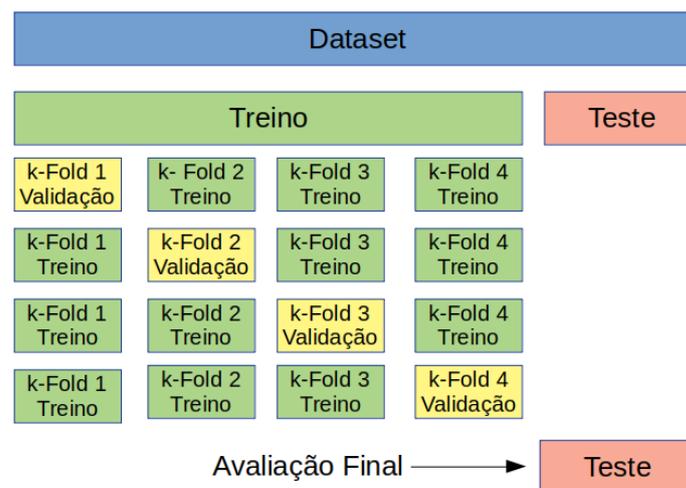
$$RF = \frac{|VN|}{|VN| + |FP|} \quad (5)$$

A função de perda também conhecida como função de custo, tem a finalidade de comparar as pontuações da categoria certa com as outras e com isso determinar o quão assertivo está com o resultado. Mede o nível de imprecisão do modelo, ou seja, se o algoritmo estiver muito ruim, seu valor será alto. Nesta pesquisa será utilizada a função Cross-entropy que se trata da mais utilizada nos modelos de *Machine Learning* e é utilizada com o Softmax que interpreta as pontuações como probabilidades, sempre tentando deixar a probabilidade da categoria correta próximo de um(1).

Outros conceitos importantes relacionados a Inteligência Artificial são o *hold-out* e *cross validation* ou validação cruzada. Essas técnicas dizem respeito ao método de como a rede neural será avaliada. O *hold-out* separa a base de dados em treino, validação

e testes. O conjunto de dados de treinamento é usado para treinar o modelo, que posteriormente é validado com o conjunto da base de dados separado para validação. A parte dos dados separada para teste é usada para avaliar a capacidade de generalização do modelo. Já a validação cruzada separa os dados em um conjunto de treino, dividido em k subconjuntos denominados *k-fold*. O modelo é treinado usando $k-1$ subconjuntos e o conjunto k , não incluso no treinamento, é usado para validação, sendo no final, realizado o teste do modelo conforme ilustrado na Figura 6.

Figura 6 – Validação Cruzada



Fonte: Adaptado de Pedregosa et al. (2011)

Relacionando os conceitos, fica compreendido que dentro do conceito de IA, possuímos o aprendizado de máquina (supervisionado, semi-supervisionado e não supervisionado), as redes neurais artificiais, o aprendizado profundo e as redes neurais convolucionais, que utiliza um conjunto de algoritmos, que buscam aprender diferentes níveis de abstração, utilizando várias camadas de processamento. E ainda, o Processamento de Linguagem Natural (PLN) é usado no pré-processamento dos dados antes de serem inseridos nos algoritmos de *deep learning* para classificação.

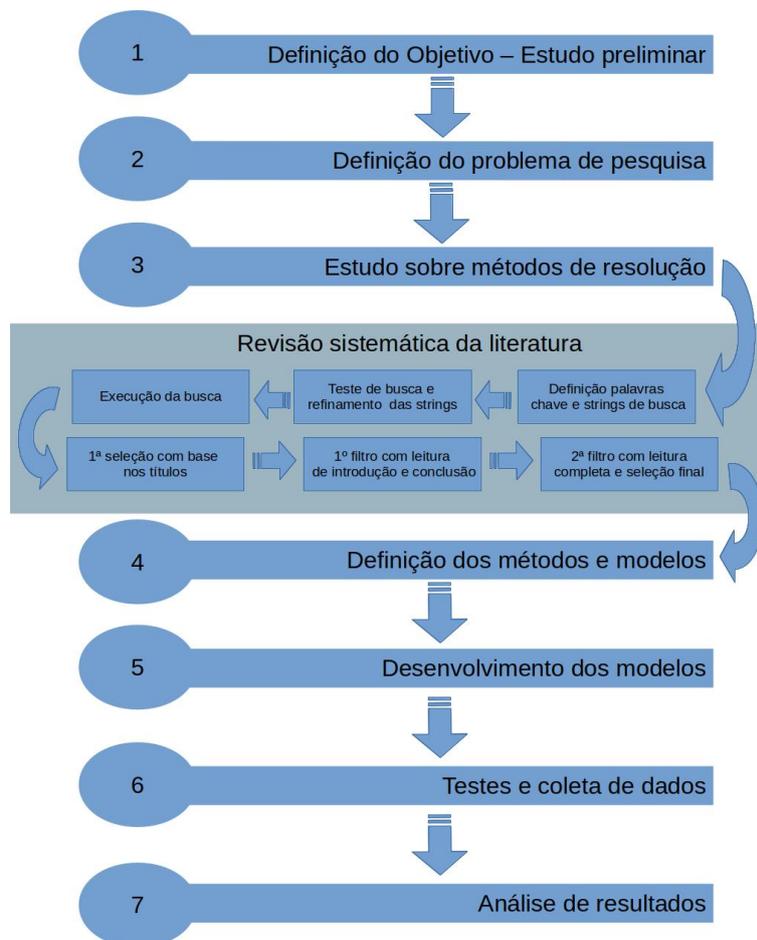
3 METODOLOGIA

Para alcançar os objetivos pretendidos, foi utilizada uma abordagem exploratória, através de procedimentos experimentais, com soluções tecnológicas práticas (algoritmos de inteligência artificial) de natureza aplicada, com uso de uma abordagem qualiquantitativa (GIL et al., 2002; MARCONI; LAKATOS, 2004). As justificativas, abordagens, procedimentos e ferramentas utilizadas, estão descritas abaixo.

Por se tratar de um tema que fica em evidência durante o processo político eleitoral, cabe ressaltar aqui o que diz Ferrari (1974) no que tange às tarefas que a ciência deve cumprir como aumento do conhecimento e descoberta de novos fatos ou fenômenos. Complementando, Demo (2000) refere que no campo científico é sempre mais fácil apontar o que as coisas não são. Onde assim, o conhecimento científico não é ideologia, visto que usa metodologias para salvaguardar a captação da realidade e a ideologia tem por objetivo produzir discurso marcado pela justificação.

A sequência de etapas planejadas para execução deste projeto será descrita na Figura 7 que representa o conjunto de atividades desenvolvidas.

Figura 7 – Atividades de Metodologia



Primeiramente foi realizado um estudo preliminar sobre o tema, objetivando uma aproximação e entendimento do fenômeno, através do procedimento técnico de uma análise bibliográfica. Este processo foi realizado anteriormente a esta pesquisa com a leitura e produção de artigos e pesquisas científicas, com a descrição dos tipos e *modos operandis* das *fake news*, desde como são produzidas e espalhadas, bem como podem ser identificadas.

Para definir o problema de pesquisa, considerando o método de raciocínio indutivo onde generalizações derivam de observações (GIL et al., 2002; MARCONI; LAKATOS, 2004), foi verificada a natureza digital sobre o problema das *fake news*. Visto que a maioria dos autores referenciados classifica que mentiras sempre existiram e que o termo *fake news* foi popularizado pela disseminação de notícias na *Internet*, enquadramos então as *fake news* como um problema que possui uma solução digital, ou seja, por ser um problema digital, sua solução também deve se dar de maneira digital.

Como terceira atividade, foi realizado um estudo, através de uma revisão sistemática da literatura (RSL), sobre os métodos, ferramentas e algoritmos já existentes e

usados no combate as disseminação das notícias. Suas eficiências e características foram analisadas, visando compor o referencial teórico já descrito no capítulo anterior.

Na sequência, atendendo ao quarto item de acordo com a Figura 7, devido à amplitude do tema e as diversas técnicas e métodos existentes e já estudadas como solução, foi realizada a definição do método e modelo de proposta prática que será implementado baseada na arquitetura proposta por Zhang e Wallace (2016), usando Redes Neurais Convolucionais para classificação de textos, onde o *paper* atualmente conta com mais de 10 mil e seiscentas citações e descrito pelo autor como modelo simples e com forte desempenho empírico.

A quinta atividade, trata do desenvolvimento prático do modelo proposto, seguido dos testes e coletas dos dados. Devido ao alto poder computacional necessário para o treinamento dos algoritmos de Inteligência Artificial (*Deep Learning*), foi escolhido o *Google Colab* como ambiente de *notebook Jupyter*, uma vez que disponibiliza o processamento em GPU e é executado na nuvem. A arquitetura do modelo foi desenvolvida em *Python* e bibliotecas do *Numpy*, *Pandas*, *BeautifulSoup*, *Sklearn*, *Tensorflow* e *Keras*. A fim de seguir boas práticas de desenvolvimento e controle de versionamento, foi verificado através da pesquisa de Batista et al. (2017) o uso de Github. Desta forma, o código fonte completo, resultante do estudo, encontra-se disponível em: github.com/pabloalima81/tcc2, através de repositório privado. Caso os membros da banca desejarem acesso, basta solicitar ao autor.

Na sexta atividade metodológica que trata do teste e coleta dos dados, foi utilizado o conjunto de dados disponibilizado por Monteiro et al. (2018b), contendo 7200 notícias, sendo 3600 verdadeiras e 3600 falsas. Esse conjunto de dados é amplamente utilizado em outros projetos de pesquisa sobre o tema em língua portuguesa, obtendo resultados satisfatórios conforme já descrito na Seção 2.2. A amostragem dos dados são de janeiro de 2016 a janeiro de 2018 e estão detalhados na pesquisa (MONTEIRO et al., 2018b). Esta base de dados será usada para treinar o algoritmo e após o treinamento será realizada a coleta dos dados referentes aos resultados alcançados pelo modelo.

Ainda na sexta fase, para adequar o modelo proposto por Zhang e Wallace (2016), será realizada a variação dos seguintes hiperparâmetros: Filtros por região, funções de perda, número de neurônios da rede neural densa, percentual de neurônios inativos na rede neural densa e tamanho do lote que define de quantos em quantos registros serão atualizados os pesos.

A estrutura do modelo conta com três tamanhos de filtro de tamanho: 2, 3 e

4. Zhang e Wallace (2016) realizou testes com mais filtros variando os tamanhos, mas ressaltou que usar mais filtros e também outros tamanhos não gera influencia positiva no desempenho, sendo que somente o número de filtros deve ser ajustado de acordo com os dados de entrada. Esses filtros realizam convoluções na matriz da frase e geram (comprimento variável) mapas de características.

O *max pooling* é realizado em cada mapa, ou seja, o maior número de cada mapa de recursos é registrado, sendo que em suas conclusões Zhang e Wallace (2016) relata não parece necessário gastar recursos avaliando estratégias alternativas para esta etapa. Assim, um vetor de característica é gerado a partir de todos os seis mapas, e estes 6 recursos são concatenados para formar um vetor de recurso para a penúltima camada.

A camada final com a função *sigmoid*, que de acordo com Neal (1992) é a indicada para problemas de classificação binária. Essa camada recebe o vetor de características como entrada e a usa para classificar a frase; aqui se tem classificação binária e, portanto, representam dois estados de saída possíveis para a notícia: verdadeira ou falsa.

Os parâmetros relativos ao tamanho dos filtros utilizados, *max pooling* e uso da função de ativação ReLu são baseados na pesquisa de Zhang e Wallace (2016). No que diz respeito à escolha da função de ativação em CNNs, fica sugerido o uso da função ReLU, tangente hiperbólica (tanh) ou talvez também função de identidade, o que significa não usar nenhuma função de ativação indicando que em alguns conjuntos de dados, uma transformação linear é suficiente para capturar a correlação entre o vetor de palavras e o rótulo de saída (ZHANG; WALLACE, 2016). As três funções citadas foram testadas pelo autor citado e também nesta pesquisa, onde o desempenho mais satisfatório foi apresentado com uso da ReLu.

Em relação ao tamanho dos filtros, fica indicado realizar uma pesquisa mais detalhada para encontrar o "melhor" tamanho para o conjunto de dados em consideração. Já no que diz respeito a quantidade de filtros por região, foi realizada a variação entre 10 e 300 filtros, onde valores menores que 10 apresentaram resultados inferiores e maior que 300 não acrescentou ganho de resultados, somente aumento no tempo de execução.

A variação do número de neurônios na camada densa será baseada na abordagem de Heaton (2008), que segue três abordagens iniciais citadas abaixo e baseadas no tamanho da entrada, que neste caso possui relação com o número de filtros por região multiplicado por 3, que são a quantidade de tamanho de filtros usados.

A primeira abordagem sugere que o número de neurônios escondidos deve estar entre o tamanho da camada de entrada e o da camada de saída. Indicando nesse caso,

usar o número médio entre as duas camadas. Ou seja, quanto temos na entrada 25 filtros por região, o valor número de neurônios seria 38, resultado da multiplicação de 25 por 3, mais 2(camada de saída), dividido por 2.

Na segunda, o número de neurônios escondidos deve ser $2/3$ do tamanho da camada de entrada, mais o tamanho da camada de saída. Assim, a camada escondida no nosso exemplo com 25 filtros por região deve conter $75*(2/3)+2 = 51$ neurônios.

Na terceira e última abordagem, O número de neurônios escondidos deve ser menor que o dobro do tamanho da camada de entrada. Ou seja, no nosso exemplo, a camada escondida deve conter menos que 150 neurônios. As três estimativas acima citadas, foram testadas com as variações de filtros por região entre 10 e 300 conforme citado acima.

A arquitetura da rede com uso da matriz de palavras na entrada e as etapas posteriores de convolução, extração do mapa de características, *max pooling*, concatenação, rede densa e saída foi mantida por já estar intensamente testada, porem neste caso, com uso do *dataset* de notícias falsas em língua portuguesa foi optado por alternar somente alguns parâmetros já descritos acima, mantendo a estrutura base do modelo.

Os resultados serão apresentados na sétima fase, com uma abordagem quali-quantitativa analisando o algoritmo através da eficácia do modelo proposto e quantitativa através da submissão de outras notícias já conhecidas, de outra base de dados, para verificação dos resultados apresentados, classificando a pesquisa do ponto de vista de sua natureza como aplicada (SILVA; MENEZES, 2001). Estas notícias serão selecionadas em um *site* onde as mesmas já foram verificadas manualmente como verdadeiras ou falsas e contendo um tema atual.

A apresentação dos resultados alcançados, comparando a precisão do modelo proposto, com os demais trabalhos correlatos que possuem o mesmo objetivo, métricas e bases de dados usadas para treinamento, será através de matriz de confusão, acurácia geral, precisão para encontrar notícias verdadeiras, precisão de identificação de falsas, *recall* de ambas e análise gráfica da função de perda.

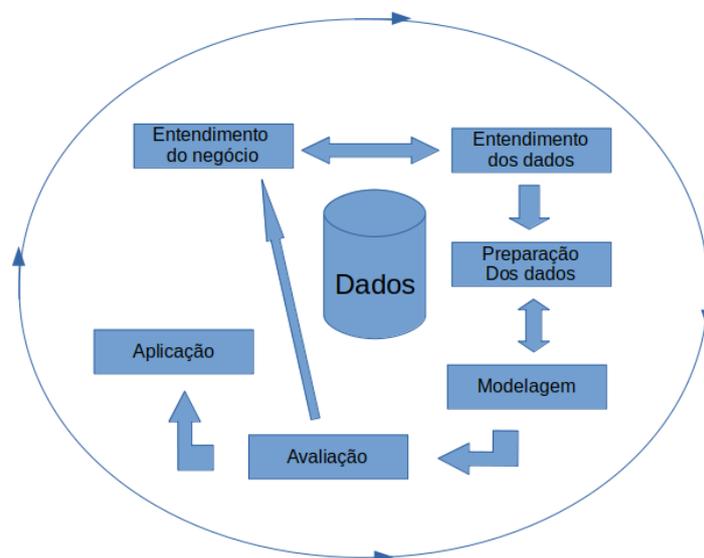
4 CFN-AI: CLASSIFICADOR DE NOTÍCIAS FALSAS BASEADO EM INTELIGÊNCIA ARTIFICIAL

Este projeto propõe um modelo que possa ajudar na mitigação da disseminação de notícias falsas com uso de inteligência artificial propondo uma arquitetura de Rede Neural Convolutacional que possa classificar uma notícia como verdadeira ou falsa.

4.1 Modelagem e desenvolvimento do modelo proposto (CFN-AI)

A partir do estudo sobre os métodos mais utilizados para a construção de conhecimento das empresas, com base em *big data*, identificou-se o modelo CRISP-DM de Moura (2018). Concebido em 1996, apresenta uma visão geral do ciclo de vida de um projeto de *Data Mining*, contendo as fases e as tarefas relacionadas ao projeto conforme a Figura 8 (CHAPMAN et al., 2000).

Figura 8 – Modelo CRISP-DM



Fonte: Adaptado de Chapman et al. (2000)

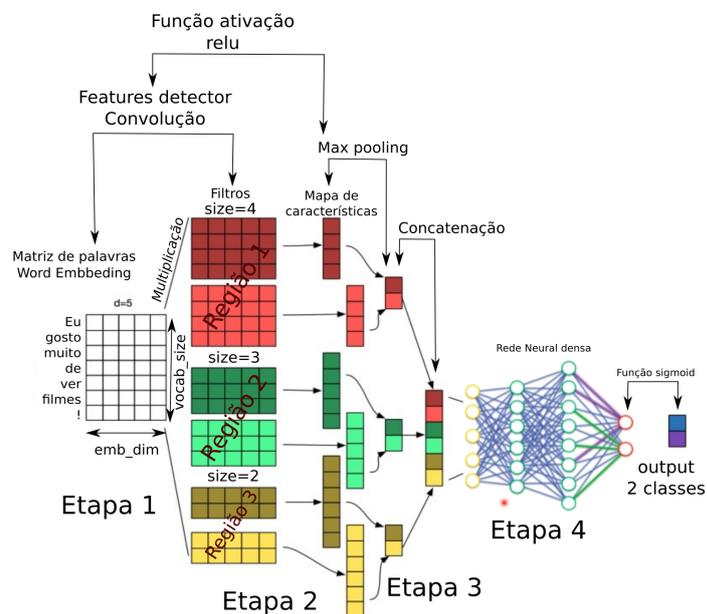
Tomando como base o modelo apresentado é que foi planejado o desenvolvimento da solução proposta e apoiada no auxílio de sistemas qualificados, para que as informações possam ser analisadas e transformadas com mais rapidez, buscando também um processo de tomada de decisão mais preciso e confiável.

Para Zhang e Wallace (2016), o uso da arquitetura proposta, é uma sugestão

e provavelmente fornecerá um ponto de partida razoável, para pesquisadores ou profissionais que procuram aplicar uma CNN simples de uma camada ao mundo real nas tarefas de classificação de frases. Enfatiza ainda que os resultados apresentados por esta CNN simples de uma camada, à posiciona como um novo *base line* padrão semelhante ao *bag-of-words*, *Support Vector Machine* (SVM) e regressão logística.

Sendo assim, a arquitetura proposta implementada está baseada na pesquisa de Zhang e Wallace (2016), ilustrada na Figura 9 que ilustra a arquitetura do modelo implementado já descrito na metodologia.

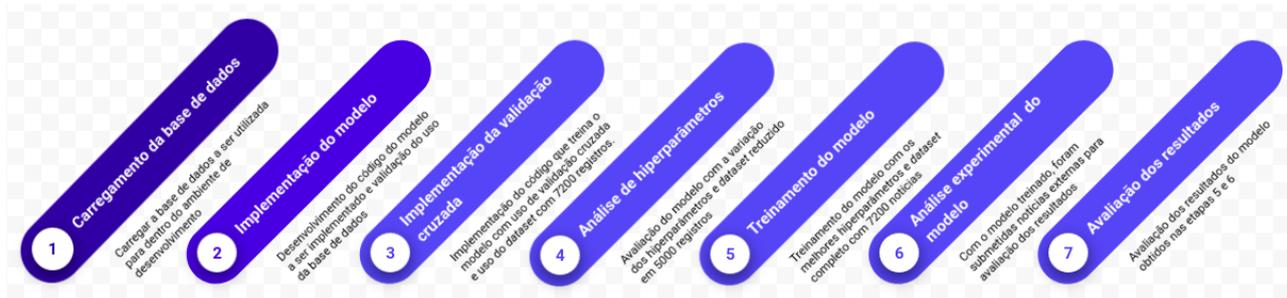
Figura 9 – Proposta de modelo



Fonte: Adaptado de Zhang e Wallace (2016)

O processo de codificação, ocorreu em sua totalidade usando o ambiente em nuvem de desenvolvimento do *Google Colab*. As fases do processo de desenvolvimento são ilustradas na Figura 10. Na fase 1 de codificação foram inseridas as bibliotecas a serem utilizadas e o carregamento da base de dados dentro *Google Colab*, sendo efetuada também avaliação do *dataset* validando os 7200 registro e o balanceamento das classes com 3600 notícias em cada classe.

Figura 10 – Fases de desenvolvimento



Fonte: Autor (2021)

As notícias foram carregadas para a variável `noticias` e as classes de cada uma delas para a variável `classes`, sendo que as notícias verdadeiras são representadas pelo número um (1) e as falsas por zero (0). Dando sequência a primeira fase de desenvolvimento, foi realizado o pré-processamento do texto com a conversão da variável `noticias` em números (“tokenização”) usando *tensorflow datasets* (TF-IDF) e *padding*. A fase 1 está descrita na Figura 11.

Figura 11 – Pré processamento do texto

```
colunas = ['classe', 'noticias']
data = pd.read_csv('/content/drive/My Drive/Colab Notebooks/cfn/Data/pre-processed.csv', header = None,
names = colunas, engine = 'python', encoding = 'utf-8')

noticias = data.iloc[:, 1].values
classes = data.iloc[:, 0].values

classes[classes == 'true'] = 1
classes[classes == 'fake'] = 0

tokenizer = tfds.deprecated.text.SubwordTextEncoder.build_from_corpus(noticias, target_vocab_size=70000)
noticias = [tokenizer.encode(sentence) for sentence in noticias]

max_len = max([len(sentence) for sentence in noticias])

noticias = tf.keras.preprocessing.sequence.pad_sequences(noticias,
value = 0,
padding = 'post',
maxlen=max_len)
```

Fonte: Autor (2021)

O código que implementa a arquitetura proposta na fase 2, está apresentado na Figura 12 e foi desenvolvida usando a biblioteca *Keras*. A estrutura proposta por Zhang e Wallace (2016) foi mantida a mesma, adaptada para receber como entrada o *Dataset* do Monteiro et al. (2018b) armazenado nas variáveis `noticias` e `classes`.

Figura 12 – Codificação do modelo proposto

```

class DCNN(tf.keras.Model):

    #0 Construtor da classe com passagem de parâmetros
    def __init__(self,
                 vocab_size=vocab_size, #Tamanho do vocabulário
                 emb_dim=emb_dim, # Tamanho da matriz de embeddings
                 nb_filters=nb_filters, #Número de filtros por região
                 ffn_units=ffn_units, #Feedforward - Número de neurônios da Rede neural densa
                 dropout_rate=dropout_rate, #Zera percentual de neurônios para evitar overfitting
                 training=True,
                 name="dcnn"):
        super(DCNN, self).__init__(name=name)
        #1 Gera matriz de palavras
        self.embedding = layers.Embedding(vocab_size, emb_dim)
        #2 Gera os filtros de tamanho 2
        self.bigram = layers.Conv1D(filters=nb_filters, kernel_size=kernel_size1, padding='same', activation=activation)
        #2.1 Gera os filtros de tamanho 3
        self.trigram = layers.Conv1D(filters=nb_filters, kernel_size=kernel_size2, padding='same', activation=activation)
        #2.2 Gera os filtros de tamanho 4
        self.fourgram = layers.Conv1D(filters=nb_filters, kernel_size=kernel_size2, padding='same', activation=activation)
        #3 Etapa de maxpooling
        self.pool = layers.GlobalMaxPool1D()

        #4 Rede neural densa com ffn_unit definindo o numero de neuronios e função de ativação relu
        self.dense_1 = layers.Dense(units = ffn_units, activation = activation_dense)
        #4.1 Passagem do parametro de dropout para evitar overfitting
        self.dropout = layers.Dropout(rate = dropout_rate)
        #4.2 Define a camada de saída com 1 bit e função de ativação sigmoid
        self.last_dense = layers.Dense(units = 1, activation = activation_exit)

    #Função que faz a "conexão" dos objetos da rede
    def call(self, inputs, training):
        #Matriz de palavras recebe o texto
        x = self.embedding(inputs)
        #Filtros da região 3 recebem a matriz de palavras
        x_1 = self.bigram(x)
        #Executa o maxpooling no bigram
        x_1 = self.pool(x_1)
        #Filtros da região 2 recebem a matriz de palavras
        x_2 = self.trigram(x)
        #Executa o maxpooling no trigram
        x_2 = self.pool(x_2)
        #Filtros da região 2 recebem a matriz de palavras
        x_3 = self.fourgram(x)
        #Executa o maxpooling no fourgram
        x_3 = self.pool(x_3)
        #Faz a concatenação após o max pooling
        merged = tf.concat([x_1, x_2, x_3], axis = -1) # Return (batch_size, número de camadas (3) * nb_filters)
        #A camada primeira camada densa recebe o vetor concatenado
        merged = self.dense_1(merged)
        #Zera uma porção dos neurônios
        merged = self.dropout(merged, training)
        #Liga a camada de dropout a camada de saída
        output = self.last_dense(merged)

        return output

```

Fonte: Autor (2021)

A fase 3 que descreve o processo de implementação da validação cruzada, que foi configurado para usar 4 k-folds, ou seja, dividindo a base de dados em 4 blocos de notícias. Foi utilizada a técnica de *StratifiedKfold* com *shuffle* que garante que em todos os *folds* a proporção de notícias verdadeiras e falsas é igual e ainda "embaralhando" as amostras. A Figura 13 apresenta o código da fase 3.

Figura 13 – Treinamento do modelo com validação cruzada

```

kfold = StratifiedKFold(n_splits = 4, shuffle = True, random_state = 10)
resultados = []
matrizes = []
history = []
validacao=[]
for indice_treinamento, indice_teste in kfold.split(noticias,
                                                    np.zeros(shape=(noticias.shape[0], 1))):
    print("Kfold: ", indice_treinamento[kfold.n_splits])
    Dcnn = DCNN(vocab_size=vocab_size, emb_dim=emb_dim, nb_filters=nb_filters,
               ffn_units=ffn_units, dropout_rate=dropout_rate)
    Dcnn.compile(loss='binary_crossentropy', optimizer='adam', metrics='accuracy')
    historico = Dcnn.fit(noticias[indice_treinamento], classes[indice_treinamento],
                       batch_size = batch_size_train,
                       epochs = nb_epochs,
                       verbose = 1,
                       validation_split = validation_split)
    precisao = Dcnn.evaluate(noticias[indice_teste], classes[indice_teste], batch_size = batch_size_val)
    previsoes = Dcnn.predict(noticias[indice_teste])
    previsoes = (previsoes > 0.5)
    matrizes.append(confusion_matrix(classes[indice_teste], previsoes))
    resultados.append(precisao)
    history.append(historico)
    precisao.append(validacao)
print(matrizes)
print("Treinamento concluído!")

```

Fonte: Autor (2021)

Com o modelo implementado, base de dados carregada e todo código executando sem erros, foi realizado o processo de avaliação de hiperparâmetros (fase 4). A avaliação foi realizada com a base de dados reduzida em 5000 notícias, com classes balanceadas para diminuir o tempo de execução e estão focados na variação dos hiperparâmetros para avaliar qual o modelo com o melhor desempenho perante o *dataset*. Todos os resultados referentes avaliação dos hiperparâmetros foram armazenados. Os valores de variação são apresentados no próximo capítulo juntamente com os resultados alcançados. Na fase 5, após encontrar o modelo com a maior acurácia, a rede foi treinada com a base de dados completa com 7200 notícias e os respectivos hiperparâmetros que levaram o rede a obter maior acurácia.

Na fase 6, foi realizada análise experimental do modelo, retirando do site da Fiocruz¹, 20 notícias falsas relacionadas a Covid-19 no Brasil e suas versões verdadeiras e submetidas ao modelo treinado para avaliar a eficácia do mesmo com novas entradas. O processo de remoção de caracteres especiais, *stop words* e submissão ao modelo já treinado é descrito na Figura 14.

¹<http://informe.ensp.fiocruz.br/noticias/48548>

Figura 14 – Submissão de novos textos

```
[ ] #Remove caracteres especiais

def clean_text(text):
    text = BeautifulSoup(text, 'xml').get_text()
    text = re.sub(r"[A-Za-z0-9]+", ' ', text)
    text = re.sub(r"https?://[A-Za-z0-9./]+", ' ', text)
    text = re.sub(r"[^a-zA-Z.!?]", ' ', text)
    text = re.sub(r"+", ' ', text)
    return text

#remove stopwords

!python -m spacy download pt
nlp = sp.load('pt')
nltk.download('stopwords')
stopwords = nltk.corpus.stopwords.words('portuguese')

def clean_sw(text):
    text = text.lower()
    document = nlp(text)

    words = []
    for token in document:
        words.append(token.text)

    words = [word for word in words if word not in stopwords and word not in string.punctuation]
    words = ' '.join([str(element) for element in words])

    return text

[ ] text = 'alcool em gel pode ser feito em casa com apenas dois ingredientes'
text = clean_text(clean_sw(text))
print(text)
text = tokenizer.encode(text)

alcool em gel pode ser feito em casa com apenas dois ingredientes

[ ] pred= Dcnn(np.array([text]), training=False).numpy()
pred = (pred > 0.5)
print(pred)

[[False]]
```

Fonte: Autor (2021)

Abaixo são apresentados os resultados da fase experimental (6), onde na Figura 15 é apresentada a matriz de confusão.

Figura 15 – Matriz de Confusão com dados da Fiocruz



Fonte: Autor (2021)

Com os dados da matriz acima foram obtidos os resultados de acurácia geral, precisão para encontrar notícias verdadeiras, precisão de identificação de falsas, e *recall* de ambas, para as notícias relacionadas a pandemia no Brasil.

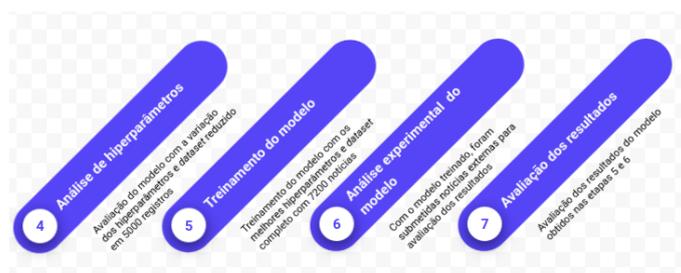
- Acurácia geral: 80%
- Precisão de *News*: 90%
- Precisão de *Fake*: 70%
- Recall de *News*: 75%
- Recall de *Fake*: 87,5%

Os resultados referentes a fase 5, suas avaliações, comparativo dos resultados obtidos nas fases 5 e 6, bem como a comparação com os demais trabalhos correlatos, são apresentas no capítulo seguinte. O controle de versionamento foi feito via repositório GitHub, através de repositório privado. Caso os membros da banca desejarem acesso, basta solicitar ao autor.

5 RESULTADOS E DISCUSSÕES

O presente capítulo é dedicado a apresentação dos resultados e discussão dos resultados a partir dos experimentos realizados. Com base nas decisões metodológicas descritas anteriormente e ilustradas na Figura 16, primeiramente foram realizados os testes com análise e variação dos hiperparâmetros (fase 4), avaliação dos resultados alcançados seguido do treinamento do modelo com melhor acurácia (fase 5). Avaliação dos resultados de treino e teste do modelo preditor e posterior submissão de notícias externas para também avaliar os resultados (fase 6). Para finalizar foi realizada a comparação dos resultados alcançados através do modelo desenvolvido com os outros trabalhos correlatos (fase 7).

Figura 16 – Fases de avaliação dos resultados



Fonte: Autor (2021)

A seguir serão apresentadas as variações dos hiperparâmetros realizadas (fase 4) e resultados do modelo alcançados na fase 5 de desenvolvimento. Os valores de variação dos hiperparâmetros apresentados a seguir, são baseados no trabalho de Zhang e Wallace (2016) e também nas abordagens de Heaton (2008) já descritos na metodologia. Desta forma foi realizada primeira a variação do número de filtros (nb filters) por região e quantidade de neurônios da camada densa (ffn units) de acordo com a Tabela 6. Após a definição do melhor parâmetro para o número de filtros por região (50) e número de neurônios da camada escondida (152), foram realizados testes variando o *dropout* entre 0 e 0,3, ficando o melhor resultado com uso de 0,2. Com relação a função de ativação, foram realizados experimentos com ReLu, tangente hiperbólica (tanh) e função de identidade. O desempenho mais satisfatório foi apresentado com uso da ReLu.

Finalizando a análise dos hiperparâmetros, foram realizados os testes variando o tamanho do lote (*batch size*) com os seguintes valores: 1, 16, 32, 64, 120, 140, 160, 180 e 200, variação esta baseada nas indicações da pesquisa de Keskar et al. (2016), e limitados em 200, pois valores maiores a execução apresenta erro decorrente da falta de memória

Tabela 6 – Variação de parâmetros

nb filters	ffn units	Acurácia
10	16	94,82 %
10	22	95,76 %
10	58	94,48 %
25	38	95,88 %
25	52	95,80 %
25	130	95,17 %
50	76	95,74 %
50	102	96,44 %
50	250	95,96 %
100	151	95,96 %
100	202	96,10 %
100	550	96,02 %
150	226	95,90 %
150	302	96,04 %
150	850	95,92 %
200	301	96,04 %
200	402	95,68 %
200	1150	95,66 %
300	451	95,30 %
300	602	95,34 %
300	1750	95,92 %

no Colab. Os resultados são apresentados na Tabela 7. Analisando a dispersão dos dados

Tabela 7 – Variação do tamanho de lote

batch size	Acurácia
1	94,58%
16	95,80%
32	96,02%
64	95,98%
120	96,16%
140	95,94%
160	95,72%
180	96,06%
200	95,64%

relativos a variação dos hiperparâmetros foi possível observar que a média de acurácia enquanto se varia o número de filtros e neurônios na camada oculta é de 95,88% com uma variância de $19,9e - 6$. Quando se varia o *batch size* a média é de 95,77% com variância de $20e - 6$. A partir da análise de dispersão dos valores de acurácia, é possível afirmar que o modelo proposto, com o uso do *dataset* de Monteiro et al. (2018b), apresenta pouca variação de resultados perante a alteração de hiperparâmetros,

Diante do exposto, o modelo com melhores resultados alcançou um desempenho com acurácia na fase de teste de 95,76% com os hiperparâmetros abaixo. Foi treinando com o *dataset* completo por 100 épocas, com a base dividida com 5760 notícias para treinamento(80%) e 1440 para teste(20%).

- nb filters: 50
- ffn units: 102
- dropout: 0.2
- Função de ativação: reLu
- batch size: 120

Passando para a fase final de avaliação, fase 7, abaixo a Figura 17 ilustra a matriz de confusão do modelo com melhor desempenho. A partir da mesma, fica evidenciado que o modelo conseguiu diferenciar de maneira satisfatória as notícias verdadeiras das falsas, evidenciado pelo número elevado de verdadeiros positivos(705) e verdadeiros negativos(674), e pelo baixo número de falsos positivos(15) e falsos negativos(46). Com

Figura 17 – Matriz de Confusão



Fonte: Autor (2021)

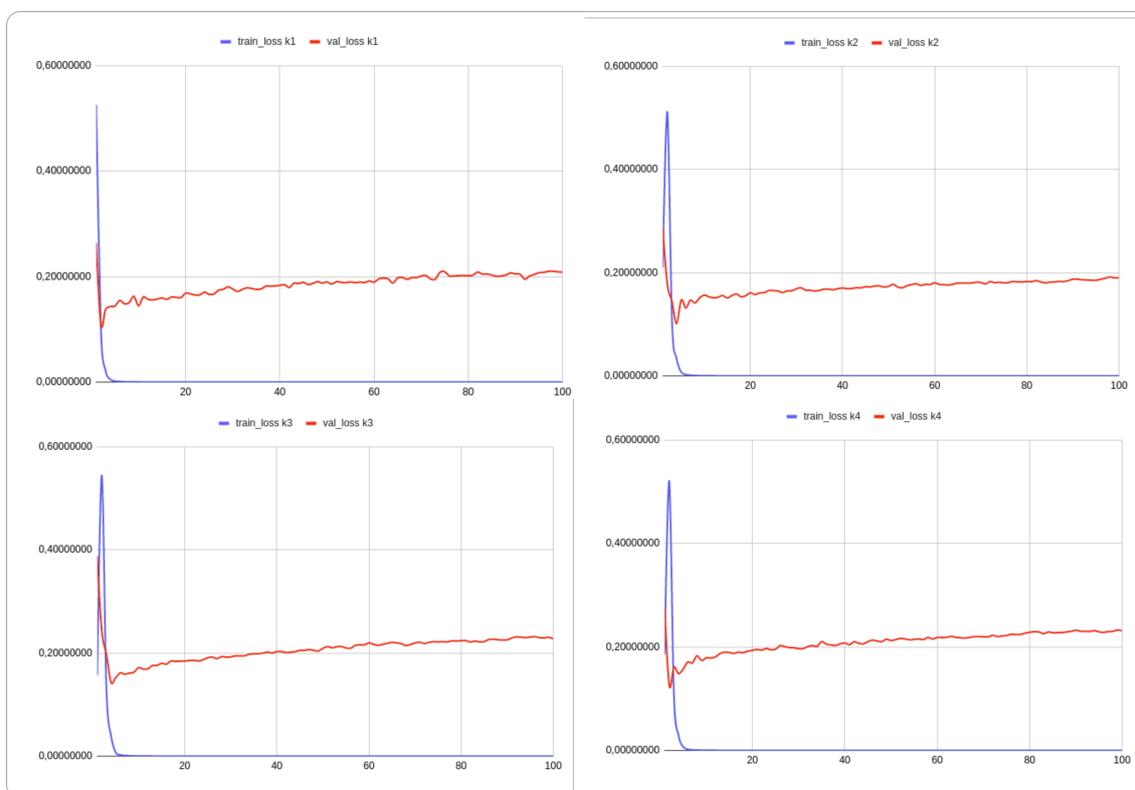
os dados da matriz de confusão foram obtidos os resultados de precisão para encontrar notícias verdadeiras, precisão de identificação de falsas, e *recall* de ambas. A Tabela 8 mostra as métricas, onde o modelo obteve uma melhor precisão na classificação de notícias verdadeiras com 97,98%, sendo que no caso do *recall*, o modelo atingiu resultado de 97,82% para notícias falsas, sendo estes considerados bons indicativos de que o modelo está classificando os textos da maneira esperada. Após cada época durante o treinamento

Tabela 8 – Métricas

Métricas	Verdadeiras (News)	Falsas(Fake)
Precisão	97,98%	93,62%
Recall	93,87%	97,82%

foram medidos os valores de perda e acurácia para os dados de treinamento e de validação. Os resultados obtidos para o treinamento em 100 épocas são mostrados nos gráficos na Figura 18 e Figura 19.

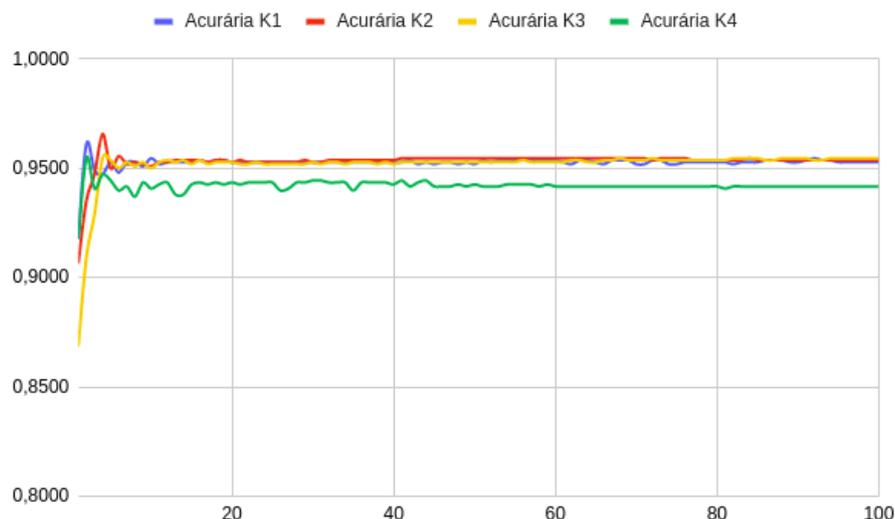
Figura 18 – Função de perda



Fonte: Autor (2021)

Os gráficos da Figura 18 mostram nos 4 quadros os valores da função de perda em cada *fold(k)* da validação cruzada em relação a cada época do treinamento. É possível verificar que o modelo proposto apresentou desempenho alto para os elementos dentro do base de dados, onde o comportamento da função de perda, próxima a zero e a estabilidade da acurácia de acordo com a Figura 19, sinalizando para uma solução viável.

Figura 19 – Acurácia durante o treinamento



Fonte: Autor (2021)

Analisando individualmente o modelo implementado e os resultados alcançados na fase 5 de desenvolvimento, foi possível verificar que a proposta possui nível de assertividade dentro do esperado com uso da base de dados com 7200 notícias. Por outro lado, na análise experimental, fase 6, ao submeter o classificador a outros textos externos após o treinamento da rede, foi possível verificar um resultado mais modesto, ressaltando a dificuldade para encontrar padrões textuais que possam ser confiáveis para classificação de notícias falsas.

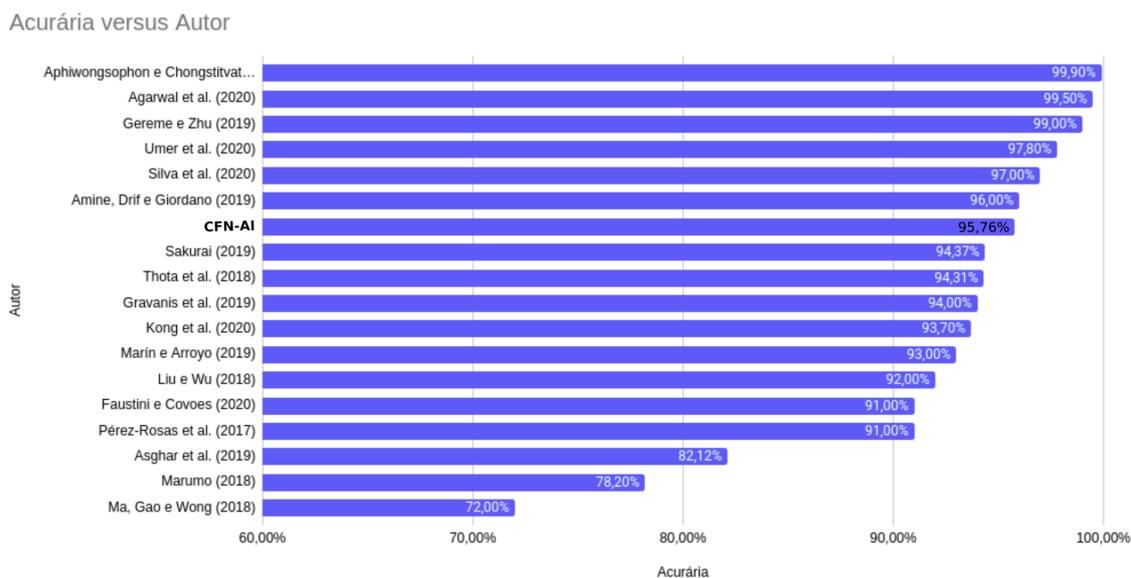
5.1 Comparativo com trabalhos correlatos

Após análise dos dados individuais do modelo, será apresentado nessa seção um comparativo com os demais trabalhos correlatos. A comparação foi realizada usando os resultados obtidos a partir dos dados de treino e teste. O gráfico comparando os valores de acurácia geral *versus* autores é apresentado na Figura 20. Os trabalhos de Silva et al. (2020) com uso de regressão logística e também o de Sakurai (2019) que usa SVM, utilizaram o mesmo *dataset*. O trabalho de Marumo (2018) apesar de tratar notícias falsas em língua portuguesa utilizou outra base de dados.

Na comparação com trabalhos que utilizam o mesmo conjunto de dados usados adotados nesta pesquisa foi possível verificar um desempenho similar aos demais. Os resultados apresentados são mais modestos que o trabalho de Silva et al. (2020), mas com

melhor desempenho que a pesquisa de Sakurai (2019).

Figura 20 – Comparação de acurácia de outros trabalhos



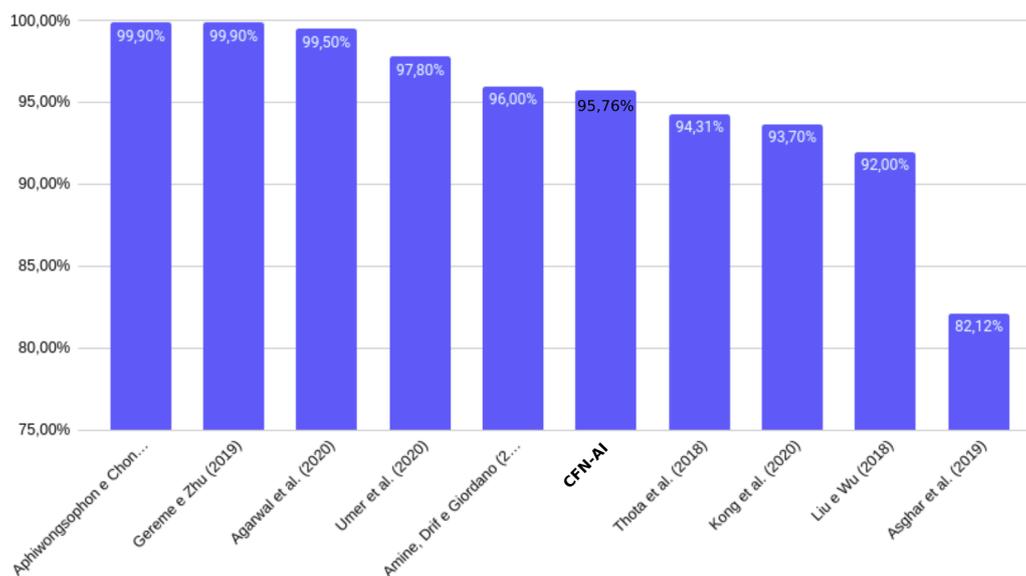
Fonte: Autor (2021)

Assim como o relatado por Zhang e Wallace (2016), o modelo proposto e estudado apresenta resultados semelhantes aos algoritmos de regressão logística e máquina de vetores de suporte sendo uma opção relativamente mais simples e com resultados satisfatórios usando um *dataset* em língua portuguesa.

Os trabalhos com acurácia de 99%, em nenhum deles é citado o uso de validação cruzada ou outro método que possa mitigar a possibilidade de *overfitting* nos resultados. Já os trabalhos de Umer et al. (2020) e Silva et al. (2020) com resultados em 97% apresentam o uso de *cross validation*.

Comparando as pesquisas que também utilizam CNN, na Figura 21 foi possível verificar que o CFN-AI apresentou resultados dentro do esperado e similar as demais pesquisas que utilizam a mesma abordagem.

Figura 21 – Comparação de acurácia de outros trabalhos com uso de CNN



Fonte: Autor (2021)

No comparativo geral com outras pesquisas o modelo se mostrou eficiente, com resultados muito similares e em alguns casos melhor que as demais soluções citadas, sendo uma solução tecnológica com uso de inteligência artificial viável para classificação de notícias falsas com rapidez, buscando também um processo de tomada de decisão mais preciso e confiável.

5.1.1 Resumo do Estudo

Sendo assim, a presente pesquisa, realizou um estudo teórico sobre o modo de operação das *fake news*, formas de mitigação e elaboração do referencial teórico com uso de RSL. Desenvolveu uma proposta de algoritmo de IA, baseado na pesquisa de Zhang e Wallace (2016), com uso de redes neurais convolucionais. O *dataset* utilizado foi de Monteiro et al. (2018b) para classificar notícias falsas ou verdadeiras em língua portuguesa.

Após o treinamento da rede com a base de dados de 7200 notícias, o modelo foi submetido a uma análise experimental com 20 notícias externas. Para finalizar, os resultados foram apresentados, analisados e comparados com outras pesquisas levantadas na RSL.

6 CONSIDERAÇÕES FINAIS

A presente pesquisa apresentou fortes contribuições do ponto de vista teórico sobre o tema, apresentando estudos científicos sobre os métodos utilizados para propagação de notícias falsas e ferramentas atuais usadas para classificação. Realizou uma revisão sistemática da literatura a fim de levantar pesquisas já realizadas que usem inteligência artificial para classificação de notícias falsas.

Entregou um modelo preditor para resolução do problema com uso de Inteligência Artificial, uso de redes neurais convolucionais, através de um modelo básico e de simples implementação, que poderia ser utilizados em outros sistemas para identificação de notícias falsas e mitigação do problema.

Desta forma, foi possível verificar, através dos estudos teóricos, técnicas computacionais, análise e comparação dos resultados, que o modelo proposto através Redes Neurais Convolucionais apresentou bons resultados preliminares, quando comparado às outras pesquisas para classificação de notícias como verdadeiras ou falsas. A pesquisa apresenta limitação em não utilizar pré-processamento na entrada e também testar apenas um *dataset*. Sendo assim, o desafio para trabalhos futuros, seria utilização de outras base de dados maior, mais diversificada, em outras línguas, bem como, o pré-processamento da matriz de entrada com uso de Glove ou Word2vector. Fica compreendido também que há uma margem para melhoria do modelo com a implementação da técnica de parada precoce ou *Early Stopping* visando impedir o *overfitting*.

Sendo assim, diante do exposto, é possível assumir que se pode identificar padrões das notícias falsas através de uma abordagem computacional, baseada em uma análise textual, com uso de Inteligência Artificial (IA). Porém há de se observar algumas inconformidades quando os algoritmos são submetidos a situações reais pós o treino, deixando sugerido que os resultados não são conclusivos.

Cabe citar e ressaltar a importância do estudo conceitual sobre o tema e seu modelo de operação, onde no princípio desta pesquisa, em meados de 2018, o termo *fake news* estava em alta. Em 2020, o conceito de "desinformação", citado por Ramalho (2018), ganhou notoriedade na literatura como um fator preocupante que mantém uma alta polaridade política situações sensíveis a qualquer novo estímulo que dão condições fundamentais para a propagação viral em larga escala de uma narrativa falsa sobre qualquer assunto (GOMES; DOURADO, 2019).

É importante também mencionar que um protótipo do modelo proposto foi apresentado no 5º Workshop Regional de Segurança da Informação e de Sistemas Computacionais (WRSeg)¹ do ano de 2020, na forma de um mini-curso intitulado "Fake News - Conceitos, métodos e aplicações de identificação e mitigação".

A relação entre o uso de IA e a possibilidade de cerceamento da liberdade de expressão tão discutida atualmente no Congresso Nacional através no PL 2630, fica ressaltado que a possibilidade da aplicação de IA pode ter muito a contribuir no processo de classificação de conteúdo que pode ser enquadrado como *fake news*. Isso se deve, principalmente, a possibilidade de alto processamento para classificação de uma grande quantidade de textos, algo humanamente impossível nos dias de hoje, considerando que a propagação de informação do meio digital é muito rápida.

6.1 Projeto Pampa sem *Fake*

O princípio desta pesquisa se deu através do Programa Universidade *Hacker*² em 2018, onde no mesmo ano, após a primeira publicação no 10º Salão Internacional de Ensino, Pesquisa e Extensão (Siepe) foi criado o grupo Pampa Sem *Fake*, Grupo de Pesquisas acadêmicas do Pampa sobre disseminação de notícias falsas na internet que faz parte do Programa Universidade *Hacker* - Unihacker e é compostos por alunos do Curso de Engenharia de Computação da Universidade Federal do Pampa, Campus de Bagé e busca estudar, pesquisar e desenvolver possíveis soluções digitais que possam mitigar a disseminação de notícias falsas na *internet*. O grupo tem como objetivo de estudar, pesquisar e desenvolver possíveis soluções digitais e publicações científicas que possam ajudar a mitigar a disseminação de notícias falsas na *internet*. Desde então o Pampa sem *Fake* participou e publicou em eventos como Siepe, 4º *Workshop* Regional de Segurança da Informação e de Sistemas Computacionais - WRSeg 2019, *TcheLinux* Bagé 2019, 5º Encontro de Ciência e Tecnologia do IFSul - Campus Bagé em 2019, Dia de Segurança em Informática de Bagé - DISI 2019. Em 2020 desenvolveu um módulo focado em *Fake News*, do Curso de Introdução a Segurança da Informação e de Sistemas promovido pelo *Unihacker*.

Promoveu também o evento *online* Maratona de *Tech Talks* sobre *Fake News* com participação de mais de 100 ouvintes em 3 dias de palestras envolvendo diversas área

¹<https://errc.sbc.org.br/2020/programacao.php>

²<https://unihacker.club>

Figura 22 – Logomarca do projeto



Fonte: Projeto Pampa sem *Fake*

do conhecimento, discutindo amplamente o tema com a sociedade de Bagé e região. Recentemente, também através de evento *online*, o projeto Pampa sem *Fake* teve a oportunidade de debater as notícias falsas sob o olhar da educação midiática e como esta pode ajudar no combate a disseminação das notícias falsas com a participação da Educamídia, Programa do Instituto Palavra Aberta com apoio do Google.org criado para capacitar professores e organizações de ensino, e engajar a sociedade no processo de educação midiática dos jovens.

A convite do projeto Fato sem *Fake*, também da Unipampa Campus São Borja, participou de gravação de um episódio de *podcast* sobre as pesquisas realizadas e a importância da computação no processo de mitigação da disseminação das notícias falsas.

Cabe citar também que o grupo já participou de eventos em outras instituições de ensino como Instituto Federal Farroupilha campus São Vicente do Sul (IFFar-SVS) e recentemente conforme já citado na seção anterior, ministrou minicurso no 5º *Workshop* Regional de Segurança da Informação e de Sistemas Computacionais (WRSeg) intitulado "Fake News - Conceitos, métodos e aplicações de identificação e mitigação".

Após as contribuições e ações acima citadas, todas dispostas no *website*³ do grupo, o projeto hoje atua de maneira intercampi com interações e colaborações com alunos e professores dos cursos de Engenharia de Software (Alegrete) e em São Borjá com os cursos de Jornalismo e Comunicação Social - Publicidade e Propaganda, mantendo assim o compromisso da Unipampa descrito no Art. 2 da lei 11.640, de 11 janeiro de 2008 que institui a Fundação Universidade Federal do Pampa, que através do ensino público e de qualidade, desenvolver pesquisa nas diversas áreas do conhecimento e promover a

³pampasemfake.unihacker.club

extensão universitária, caracterizando sua inserção regional, mediante atuação multicampi na mesorregião Metade Sul do Rio Grande do Sul.

REFERÊNCIAS

- AGARWAL, A. et al. Fake news detection using a blend of neural networks: an application of deep learning. **SN Computer Science**, Springer, v. 1, n. 3, p. 1–9, 2020.
- AGARWAL, B.; MITTAL, N. Text classification using machine learning methods-a survey. In: SPRINGER. **Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012**. [S.l.], 2014. p. 701–709.
- AIZAWA, A. An information-theoretic perspective of tf–idf measures. **Information Processing & Management**, Elsevier, v. 39, n. 1, p. 45–65, 2003.
- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. **2017 International Conference on Engineering and Technology (ICET)**. [S.l.], 2017. p. 1–6.
- AMINE, B. M.; DRIF, A.; GIORDANO, S. Merging deep learning model for fake news detection. In: IEEE. **2019 International Conference on Advanced Electrical Engineering (ICAEE)**. [S.l.], 2019. p. 1–4.
- ANDRADE, R. R. Utilização de técnicas de aprendizado de máquina para detecção de fake news. **Ciência da Computação-Tubarão**, 2019.
- APHIWONGSOPHON, S.; CHONGSTITVATANA, P. Detecting fake news with machine learning method. In: IEEE. **2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)**. [S.l.], 2018. p. 528–531.
- ASGHAR, M. Z. et al. Exploring deep neural networks for rumor detection. **Journal of Ambient Intelligence and Humanized Computing**, Springer, p. 1–19, 2019.
- BAHDANAU, D.; CHO, K.; BENGIO, Y. Neural machine translation by jointly learning to align and translate. **arXiv preprint arXiv:1409.0473**, 2014.
- BATISTA, N. A. et al. Gitsed: Um conjunto de dados com informações sociais baseado no github. In: **SBB-Dataset Showcase Workshop**. [S.l.: s.n.], 2017. p. 224–233.
- CAULYT, F. Facebook perdeu importância para a folha. **diz editor. Deutsche Welle Brasil, versão online, Boon (Alemanha)**, v. 9, 2018.
- CECHINEL, C.; CAMARGO, S. da S. Mineração de dados educacionais: avaliação e interpretação de modelos de classificação. **JAQUES, Patrícia Augustin; SIQUEIRA; Sean; BITTENCOURT, Ig; PIMENTEL, Mariano.(Org.) Metodologia de Pesquisa Científica em Informática na Educação: Abordagem Quantitativa. Porto Alegre: SBC**, 2020.
- CHAPMAN, P. et al. Crisp-dm 1.0: Step-by-step data mining guide. **SPSS inc**, v. 9, p. 13, 2000.
- CHEESEMAN, P. et al. Autoclass iii. **Program available from NASA Ames Research Center: Research Institute for Advanced Computer Science**, 1990.

DALE, R.; MOISL, H.; SOMERS, H. **Handbook of natural language processing**. [S.l.]: CRC Press, 2000.

DEMO, P. Definindo conhecimento científico. **Metodologia do conhecimento científico**. São Paulo: Atlas, p. 13–43, 2000.

DENG, L.; YU, D. Deep learning: methods and applications. **Foundations and trends in signal processing**, Now Publishers Inc. Hanover, MA, USA, v. 7, n. 3–4, p. 197–387, 2014.

FATOS, A. O que é checagem de fatos—ou fact-checking. **Acesso em 5 de agosto de 2020**, v. 12, 2018.

FAUSTINI, P. H. A.; COVOES, T. F. Fake news detection in multiple platforms and languages. **Expert Systems with Applications**, Elsevier, v. 158, p. 113503, 2020.

FERRARA, E. et al. The rise of social bots. **Communications of the ACM**, ACM New York, NY, USA, v. 59, n. 7, p. 96–104, 2016.

FERRARI, A. T. **Metodologia da ciência**. [S.l.]: Kennedy Editora, 1974.

FERREIRA, A. d. S. Redes neurais convolucionais profundas na detecção de plantas daninhas em lavoura de soja. 2017.

FEURER, M.; HUTTER, F. Hyperparameter optimization. In: **Automated machine learning**. [S.l.]: Springer, Cham, 2019. p. 3–33.

GEREME, F. B.; ZHU, W. Early detection of fake news "before it flies high". In: **Proceedings of the 2nd International Conference on Big Data Technologies**. [S.l.: s.n.], 2019. p. 142–148.

GIL, A. C. et al. **Como elaborar projetos de pesquisa**. [S.l.]: Atlas São Paulo, 2002. v. 4.

GOMES, W. da S.; DOURADO, T. Fake news, um fenômeno de comunicação política entre jornalismo, política e democracia. **Estudos em Jornalismo e Mídia**, v. 16, n. 2, p. 33–45, 2019.

GOODFELLOW, I. et al. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1.

GRAVANIS, G. et al. Behind the cues: A benchmarking study for fake news detection. **Expert Systems with Applications**, Elsevier, v. 128, p. 201–213, 2019.

GUARISE, L. Detecção de notícias falsas usando técnicas de deep learning. 2019.

GUERREIRO, L. Aprendizado semi-supervisionado utilizando modelos de caminhada de partículas em grafos. Universidade Estadual Paulista (UNESP), 2017.

HARZING, A.-W. et al. **Publish or perish**. [S.l.]: Tarma Software Research Pty Limited, 1997.

HEATON, J. The number of hidden layers. **Heaton Research Inc**, 2008.

HO, Y.; WOOKEY, S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. **IEEE Access**, IEEE, v. 8, p. 4806–4813, 2019.

- HUMBY, C. Data is the new oil. **Proc. ANA Sr. Marketer's Summit. Evanston, IL, USA**, 2006.
- KESKAR, N. S. et al. On large-batch training for deep learning: Generalization gap and sharp minima. **arXiv preprint arXiv:1609.04836**, 2016.
- KONG, S. H. et al. Fake news detection using deep learning. In: IEEE. **2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)**. [S.l.], 2020. p. 102–107.
- KRISHNAN, S.; CHEN, M. Identifying tweets with fake news. In: IEEE. **2018 IEEE International Conference on Information Reuse and Integration (IRI)**. [S.l.], 2018. p. 460–464.
- LIU, Y.; WU, Y.-F. B. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: **Thirty-second AAAI conference on artificial intelligence**. [S.l.: s.n.], 2018.
- MA, J.; GAO, W.; WONG, K.-F. Rumor detection on twitter with tree-structured recursive neural networks. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. [S.l.], 2018.
- MARCONI, M. d. A.; LAKATOS, E. M. **Metodologia científica**. [S.l.]: Atlas São Paulo, 2004. v. 4.
- MARÍN, I. P.; ARROYO, D. Fake news detection. In: SPRINGER. **Computational Intelligence in Security for Information Systems Conference**. [S.l.], 2019. p. 229–238.
- MARUMO, F. S. **Deep Learning para classificação de Fake News por sumarização de texto**. [S.l.]: Londrina, 2018.
- MATSUBARA, E. T.; MONARD, M. C.; BATISTA, G. E. Utilizando algoritmos de aprendizado semi-supervisionados multi-visão como rotuladores de texto. In: **Anais do Workshop em Tecnologia da Informação de da Linguagem Humana (TIL2005)**. [S.l.: s.n.], 2005. p. 2108–2117.
- MCCARTHY, J.; MINSKY, M.; ROCHESTER, N. The dartmouth summer research project on artificial intelligence. **Artificial intelligence: past, present, and future**, 1956.
- MEEDAN. **Check**. 2020. Disponível em: <https://checkmedia.org/>.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, Manole Ltda, v. 1, n. 1, p. 32, 2003.
- MONTEIRO, R. A. et al. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In: VILLAVICENCIO, A. et al. (Ed.). **Computational Processing of the Portuguese Language**. Cham: Springer International Publishing, 2018. p. 324–334. ISBN 978-3-319-99722-3.
- 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**. [S.l.]: Springer International Publishing, 2018. p. 324–334. ISBN 978-3-319-99722-3.

MORAES, C. P. de. “deepfake” como ferramenta manipulação e disseminação de “fakenews” em formato de vídeo nas redes sociais. OSF Preprints, 2019.

MOURA, K. V. d. Data science: um estudo dos métodos no mercado e na academia. 2018.

MURPHY, K. P. **Machine learning: a probabilistic perspective**. [S.l.]: MIT press, 2012.

NALLAPATI, R. et al. Abstractive text summarization using sequence-to-sequence rnns and beyond. **arXiv preprint arXiv:1602.06023**, 2016.

NEAL, R. M. Connectionist learning of belief networks. **Artificial intelligence**, Elsevier, v. 56, n. 1, p. 71–113, 1992.

NEIVA, F. W. **Revisão Sistemática da Literatura em Ciência da Computação Um Guia Prático**. Tese (Doutorado) — Universidade Federal de Juiz de Fora, 2016.

NETO, M. et al. Fake news no cenário da pandemia de covid-19. **Cogitare Enfermagem**, v. 25, 2020.

PAGANOTTI, I.; SAKAMOTO, L. M.; RATIER, R. P. “mais fake e menos news”: resposta educativa às notícias falsas nas eleições de 2018. **Liberdade de Expressão Questões da atualidade**, p. 52, 2019.

PARISER, E. **The filter bubble: What the Internet is hiding from you**. [S.l.]: Penguin UK, 2011.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

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

POUBEL, M. Fake news e pós-verdade. **Infoescola. Sociedade. s/d. Disponível em <<https://www.infoescola.com/sociedade/fake-news>**, v. 15, 2018.

QUESSADA, M.; PISA, L. F. Fake news versus mil: a difícil tarefa de desmentir goebbels. 2018.

RAMALHO, W. O combate às fake news no brasil: Um estudo sobre a checagem de fatos. 2018.

RESENDE, G. et al. A system for monitoring public political groups in whatsapp. In: **Proceedings of the 24th Brazilian Symposium on Multimedia and the Web**. [S.l.: s.n.], 2018. p. 387–390.

RUEDIGER, M. A.; GRASSI, A.; GUEDES, A. L. Robôs, redes sociais e política no brasil: análise de interferências de perfis automatizados de 2014. 2018.

SAKURAI, G. Y. Processamento de linguagem natural-detecção de fake news. 2019.

SAMPAIO, R. F.; MANCINI, M. C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Brazilian Journal of Physical Therapy**, SciELO Brasil, v. 11, p. 83–89, 2007.

SASTRE, A.; OLIVEIRA, C. S. P. de; BELDA, F. R. A influência do “filtro bolha” na difusão de fake news nas mídias sociais: reflexões sobre as mudanças nos algoritmos do facebook. **Revista GEMInIS**, v. 9, n. 1, p. 4–17, 2018.

SAYAD, A. L. V. Educação midiática e pensamento crítico: antídotos contra a “desinformação”. **Liberdade de Expressão Questões da atualidade**, p. 9, 2019.

SHAO, C. et al. Hoaxy: A platform for tracking online misinformation. In: **Proceedings of the 25th International Conference Companion on World Wide Web**. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016. (WWW '16 Companion), p. 745–750. ISBN 9781450341448. Disponível em: <https://doi.org/10.1145/2872518.2890098>.

SHARIATMADARI, D. Could language be the key to detecting fake news. **The Guardian**, 2019.

SHARMA, S.; SHARMA, S. Activation functions in neural networks. **Towards Data Science**, v. 6, n. 12, p. 310–316, 2017.

SILVA, E. L. d.; MENEZES, E. M. Metodologia da pesquisa e elaboração de dissertação. 3. ed. rev. atual, 2001.

SILVA, R. M. et al. Towards automatically filtering fake news in portuguese. **Expert Systems with Applications**, Elsevier, v. 146, p. 113199, 2020.

SIVEK, S. C. Both facts and feelings: Emotion and news literacy. **Journal of Media Literacy Education**, ERIC, v. 10, n. 2, p. 123–138, 2018.

T3XTO, G. **Pampa sem Fake**. 2020. Access date: 04 nov. 2020. Disponível em: <https://anchor.fm/fato-sem-fake>.

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

TURING, A. Mind. **Mind**, v. 59, n. 236, p. 433–460, 1950.

UMER, M. et al. Fake news stance detection using deep learning architecture (cnn-lstm). **IEEE Access**, IEEE, v. 8, p. 156695–156706, 2020.

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

WARDLE, C. Fake news. it’s complicated. **First Draft**, v. 16, 2017.

YE, N. **The handbook of data mining**. [S.l.]: CRC Press, 2003.

ZHANG, Y.; WALLACE, B. **A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification**. 2016.