

Detecção de *Fake News* em Língua Portuguesa Combinando Redes Neurais Convolucionais e Algoritmos de Aprendizagem de Máquina

Felipe Sousa, Alice Barbosa, Carina Oliveira, Reinaldo Braga

¹Laboratório de Redes de Computadores e Sistemas (LAR)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Abstract. *The popularization of social media has been a facilitator of access to information. However, the growing volume of news sharing has increased concern around “fake news”, due to its potential for manipulation of public opinion. Therefore, this article presents a proposal for the analysis of news in Portuguese and the detection of fake news, using Machine Learning and Convolutional Neural Networks. For this purpose, the Fake.Br database was used, which presents 7,200 news articles in Portuguese. The study carried out focused on analyzing both the texts and their respective metadata. Thus, after an analysis of the selected algorithms, an accuracy of 97% was obtained.*

Resumo. *A popularização das mídias sociais vem sendo um facilitador ao acesso à informação. Porém, o crescente volume de compartilhamento de notícias tem aumentado a preocupação em torno das “fake news”, devido seu potencial de manipulação sobre a opinião pública. Sendo assim, este artigo apresenta uma proposta para a análise de notícias em Português e a detecção de fake news, utilizando Aprendizagem de Máquina e Redes Neurais Convolucionais. Para este fim, foi utilizada a base de dados Fake.Br, que apresenta 7.200 artigos de notícias em português. O estudo realizado concentrou-se em analisar tanto os textos, como também os seus respectivos metadados. Desta forma, após uma análise sobre os algoritmos selecionados, obteve-se uma acurácia de 97%.*

1. Introdução

Nos últimos anos, tem ocorrido um aumento significativo do número de usuários da Internet. Um recente levantamento realizado pela *International Telecommunication Union* (ITU) estima que aproximadamente 4,9 bilhões de pessoas, ou 63% da população mundial, usaram a Internet em 2021, representando um aumento de 17% em relação ao ano de 2019 [ITU 2021]. No Brasil, o número de usuários também apresentou crescimento. A última pesquisa publicada pelo Instituto Brasileiro de Geografia e Estatísticas (IBGE) revela um crescimento de 74,7% em 2018 para 78,3% em 2019, considerando pessoas com 10 anos de idade ou mais [IBGE 2020].

À vista disso, tanto o consumo de notícias quanto a expressão de opiniões através da Internet têm se tornado uma prática cada vez mais comum. A popularização de redes sociais como Facebook, Twitter, Instagram e WhatsApp tem sido outro fator relevante na forma que as pessoas buscam, consomem e divulgam notícias. A facilidade de acesso à informação (com suas diversas formas e estilos textuais) e a celeridade com que a mesma é compartilhada, faz com que os usuários encontrem os meios em que se identifiquem mais e estejam muito mais engajados e informados [Pereira and Marques-Neto 2021].

Entretanto, ao mesmo tempo que facilita o acesso à informação, a popularidade das mídias sociais também tem agravado o problema da disseminação de notícias falsas (*fake news*), ampliando o alcance e a velocidade com que informações falsas se espalham e dificultando a sua identificação e contenção [Sharma et al. 2019]. O termo “*fake news*” tem sido muito utilizado para definir conteúdos que promovam desinformação, com o intuito de influenciar a opinião pública, principalmente, através de mídias sociais [Raymer and Harriss 2017]. Os autores de [Allcott and Gentzkow 2017] definem *fake news* como artigos de notícias cujos conteúdos são intencional e verificadamente falsos, sendo capazes de enganar os leitores, muitas vezes para a obtenção de vantagens financeiras, ou por motivações ideológicas, como o favorecimento de certos candidatos à cargos políticos em campanhas eleitorais [Allcott and Gentzkow 2017].

Devido ao risco que a propagação de *fake news* representa à democracia, muitos esforços têm sido empenhados no sentido de desenvolver estratégias para identificação de notícias falsas e, portanto, minimizar os seus danos à sociedade [Sharma et al. 2019]. Neste contexto, abordagens baseadas em Processamento de Linguagem Natural (PLN) para automatização da tarefa de identificação de notícias falsas podem ser de grande utilidade. O PLN trata computacionalmente os diversos aspectos da comunicação humana, como sons, palavras, sentenças e discursos, considerando formatos e referências, estruturas e significados e contextos e usos [Gonzalez and Lima 2003]. Desta forma, a tarefa de detecção de notícias falsas é uma importante aplicação para o PLN, pois promove uma solução melhor que as tradicionais, que consistem na análise humana das informações e checagem de fatos, o que é caro e demanda muito tempo [Oshikawa et al. 2018].

Diante do exposto, o presente artigo apresenta uma proposta para identificação automática de *fake news*. Mais especificamente, faz-se uso de algoritmos de Aprendizagem de Máquina e de Redes Neurais Convolucionais para realizar a análise de padrões em textos de notícias em Língua Portuguesa a fim de possibilitar a identificação de conteúdos falaciosos de forma acurada. Para este fim, foi utilizada a base de dados *Fake.Br Corpus*, que apresenta 7.200 artigos de notícias escritas em Português. O estudo realizado centralizou-se em analisar tanto os textos quanto os seus respectivos metadados. Todavia, considerando a natureza dos dados, diferentes abordagens foram adotadas. No caso, os metadados, por se apresentarem em grande parte como dados estatísticos, foram treinados com algoritmos de classificação, enquanto as notícias, por se tratarem de dados em formato textual, foram treinados fazendo uso de Redes Neurais Convolucionais, já que com o mesmo é possível realizar uma análise semântica. Desta forma, após uma análise sobre os algoritmos selecionados, obteve-se uma acurácia de 97%.

O presente artigo traz como contribuição a realização de uma análise semântica e também estatística de notícias e metadados disponíveis no *dataset*. É importante destacar que as abordagens existentes sobre soluções inteligentes para detecção de *fake news* estão focadas, em sua maioria, na Língua Inglesa. Portanto, como um diferencial, a pesquisa centralizou-se na Língua Portuguesa.

O restante do artigo está organizado da seguinte maneira: a Seção 2 apresenta os trabalhos relacionados, a Seção 3 detalha a proposta dando destaque para a metodologia aplicada para a identificação de *fake news*, a Seção 4 apresenta os resultados obtidos a partir da aplicação da proposta e, finalmente, a Seção 5 apresenta as considerações finais acerca do trabalho apresentado e os direcionamentos para trabalhos futuros.

2. Trabalhos Relacionados

Esta seção apresenta os trabalhos relacionados à categorização e detecção de *fake news*. Vale destacar que os trabalhos apresentados na literatura estão voltados, em sua maioria, à Língua Inglesa, sendo ainda um desafio o desenvolvimento de estudos focados na Língua Portuguesa.

Em [Wang 2017], são aplicados tanto algoritmos de classificação como de redes neurais para a detecção de *fake news*, como *Majority*, *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Bidirectional Long Short-Term Memory* (Bi-LSTM) e *Convolutional Neural Network* (CNN). Para tal, é utilizado um *dataset* intitulado LIAR, que apresenta mais de 12 mil relatos curtos. Também é utilizado um conjunto de metadados contendo informações acerca dos indivíduos responsáveis por cada relato. Os textos dos relatos são então processados por uma camada de *word embedding*, com o objetivo de obter uma representação vetorial para cada termo. O algoritmo CNN alcançou uma acurácia de 27% no conjunto de testes, o que foi superior a todos os outros algoritmos estudados, apresentando ainda uma eficácia na aplicação dos metadados. Todavia, a pesquisa descrita trabalhou com textos curtos e voltados para a Língua Inglesa.

No trabalho [Rosa et al. 2019], é proposta uma solução para a detecção das *fake news*, comparando 2 técnicas de Aprendizagem de Máquina (AM), no caso, CNN e as Redes Neurais Artificiais de Múltiplas Camadas (RNA). A base de dados utilizada para treinamento dos modelos é a *Fake.BR Corpus*. O estudo exhibe uma análise quantitativa da acurácia, precisão, *recall*, *F1-Score*, bem como da matriz de confusão, na qual é possível observar que a CNN é um pouco melhor que as RNA, com a acurácia de 96% e 95%, respectivamente. Em contrapartida, o trabalho faz análise apenas de textos, ou seja, o mesmo não trabalha os metadados disponibilizados pelo *dataset Fake.Br*.

No trabalho [Barbosa 2020], é apresentado o SIRENE, que consiste em uma solução inteligente para detecção de notícias falsas em Língua Portuguesa. O objetivo do trabalho é a identificação de padrões em notícias na internet, com informações falsas e verdadeiras. Para tal, são aplicados diversos algoritmos de Aprendizagem de Máquina supervisionados. Para o desenvolvimento da proposta, é utilizado um *dataset* intitulado *Sirene-news*, formada por 4.742 notícias extraídas dos portais G1 e Sensacionalista. Também são utilizadas 7.200 notícias da base de dados do *Fake.br*. Estes dados são então submetidos a uma etapa de transformação, com a finalidade de convertê-los a um formato numérico, interpretável pelo classificador. Após esta etapa, é realizado um comparativo dos algoritmos de LR, *Stochastic Gradient Descent* (SGD), *Multilayer Perceptron* (MLP) e SVM. O algoritmo SVM apresenta excelentes resultados, com taxa de acerto superior a 95%, sendo então escolhido para a implementação da proposta. Porém nesta pesquisa, assim como no trabalho anterior, os metadados disponibilizados pelo *dataset Fake.Br* não são analisados.

Em [Kaliyar et al. 2020], o objetivo é apresentar um modelo baseado em uma Rede Neural Convolutiva Profunda (FNDNet) para detecção de notícias falsas. O mesmo busca aprender automaticamente os recursos discriminatórios para classificação de notícias falsas por meio de várias camadas ocultas construídas na rede neural profunda. Os experimentos são realizados utilizando um *dataset* que contém dois arquivos, um contendo dados de treino e outro com dados de teste. Vale frisar que o conjunto de dados está relacionado às notícias falsas espalhadas durante as eleições presidenciais de

2016 nos Estados Unidos. Como resultados, o algoritmo de classificação *Multinational Naive Bayes* (MNB) apresenta uma precisão de 89,97%. Em se tratando da aprendizagem profunda, a precisão é de 94,31%, enquanto o modelo proposto FNDNet apresenta uma precisão de 98,36%. Entretanto, a base de dados utilizada é de Língua Inglesa.

A Tabela 1 apresenta um comparativo dos trabalhos relacionados, dando destaque ao presente estudo. Como será apresentado na próxima seção, um dos grandes diferenciais do presente trabalho é a aprendizagem voltada ao reconhecimento de notícias falsas em português, tanto pelo conteúdo da notícia quanto pelos metadados, fazendo uso de redes neurais e algoritmos de classificação.

Tabela 1. Comparação dos trabalhos relacionados.

Trabalhos	Análise de metadados	Análise de textos	Textos em Português	Uso de Redes Neurais
[Wang 2017]	X	X		X
[Rosa et al. 2019]		X	X	X
[Barbosa 2020]		X	X	
[Kaliyar et al. 2020]		X		X
Presente estudo	X	X	X	X

3. Proposta

Esta seção apresenta a metodologia aplicada para o desenvolvimento da proposta. O processo é ilustrado na Figura 1 e detalhado ao longo da seção. É importante frisar que a presente pesquisa trabalhou com dois níveis de dados, sendo estes: textos e metadados. Desta forma, conforme apresentado na imagem, após a seleção da base de dados, cada nível foi tratado seguindo as seguintes etapas: Pré-Processamento, Transformação, Treinamento e Testes.

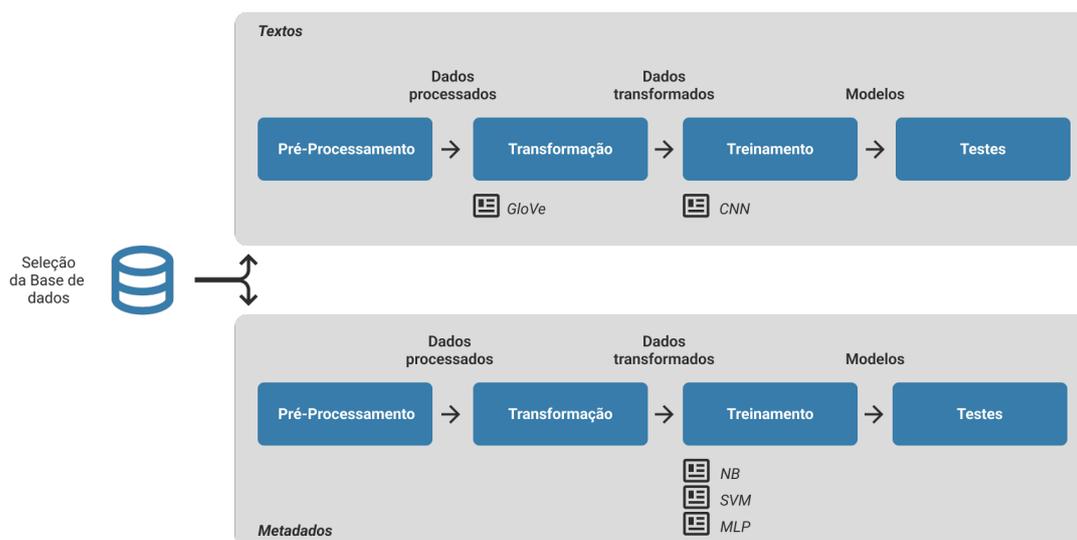


Figura 1. Sequência metodológica aplicada na pesquisa.

3.1. Seleção da Base de Dados

Para compor o *dataset* a ser utilizado na presente pesquisa, foi escolhida a base de dados construída por pesquisadores da Universidade de São Paulo (USP), denominada *Fake.Br Corpus*¹ [Monteiro et al. 2018].

Para a criação deste *dataset*, foram coletados e rotulados 7.200 artigos de notícias escritas em Língua Portuguesa. Destes, exatamente 3.600 notícias são consideradas verdadeiras e 3.600 são consideradas *fake news*. Estes dados foram coletados dentro de um período de dois anos, entre janeiro de 2016 e janeiro de 2018. No geral, as notícias coletadas estão divididas em 6 grandes categorias, conforme apresentado na Tabela 2, na qual é mostrada a distribuição dos artigos do *Fake.Br* por assuntos, sendo estes: política, TV e celebridades, sociedade e notícias diárias, ciência e tecnologia, economia e religião. Dentre estes, o assunto política é o mais frequente.

Tabela 2. Documentos do *Fake.Br Corpus* por categoria [Monteiro et al. 2018].

Categoria	N° de Amostras	%
Política	4.180	58,0%
Tv e celebridades	1.544	21,4%
Sociedade e notícias diárias	1.276	17,7%
Ciência e tecnologia	112	1,5%
Economia	44	0,7%
Religião	44	0,7%

Além dos textos dos artigos, o *Fake.Br* também disponibiliza todos os *links* e metadados, contendo informações sobre os artigos, tais como autor, data de publicação e número de visualizações. Não existe nenhuma outra base de dados semelhante a esta disponível para a Língua Portuguesa, sendo esta a mais utilizada em pesquisas sobre *fake news* para este idioma. Para o desenvolvimento desta proposta, tanto os textos dos artigos quanto os seus metadados foram utilizados.

3.2. Pré-processamento

Esta etapa se faz necessária para padronizar os textos, remover *stopwords* e eliminar quaisquer caracteres que se encontrem em formatos não reconhecíveis pelos algoritmos de AM. A etapa de pré-processamento dos textos ocorreu conforme descrito a seguir:

- **Remoção de ruídos e caracteres estranhos:** a fim de facilitar o trabalho dos algoritmos de aprendizagem, foram removidos caracteres especiais ou que estivessem em um formato não reconhecível pelos algoritmos de aprendizagem, assim como aqueles que não contribuíssem para o seu treinamento. Neste sentido, numerais, *urls*, caracteres de pontuação e outros caracteres especiais foram eliminados. Esta tarefa foi realizada utilizando expressões regulares e *scripts Python*.
- **Remoção de *stopwords*:** são chamadas *stopwords*, palavras consideradas muito comuns e, portanto, irrelevantes para a representação de um texto. São normalmente pertencentes às seguintes classes gramaticais: pronomes, preposições, artigos, advérbios e conjunções. Tendo em vista que grande parte dos termos de

¹<https://github.com/roneysco/Fake.br-Corpus>

um texto são classificados como *stopwords*, a sua eliminação diminui bastante o número de termos a serem analisados, o que reduz o tempo de treinamento dos algoritmos. Esta tarefa foi realizada utilizando a biblioteca *Natural Language Toolkit* (NLTK) e *scripts Python*.

3.3. Transformação

Para que sejam processados pelos algoritmos de AM, é necessário que antes os textos dos artigos sejam convertidos do formato textual para uma representação numérica. Nesta etapa, os textos foram submetidos a um processo de vetorização, que consiste em converter cada termo contido no texto em um valor numérico. Desta forma, cada texto é representado por um vetor contendo todos os valores numéricos que representam cada um dos termos contidos nele. Para isto, foram utilizadas as bibliotecas *Keras* e *Tensorflow*.

Com os textos vetorizados, foi utilizada uma técnica de PLN denominada *word embedding*, que consiste em converter os textos em vetores multidimensionais, de forma a manter os seus valores tanto sintáticos quanto morfológicos. No *word embedding*, palavras que possuem sentidos semelhantes são representadas de formas similares. Para a construção de um *word embedding*, é possível realizar o treinamento de um modelo a partir de um conjunto de dados. Outra possibilidade é a utilização de um *word embedding* pré-treinado. Neste trabalho, foi utilizado um modelo pré-treinado pelo algoritmo *GloVe*². Desta forma, os textos são transformados em matrizes, onde cada linha da matriz corresponde à representação vetorial dos termos contidos no texto original, conforme definido no *word embedding*.

Assim como os textos das notícias, os metadados também precisam ser transformados, pois muitas informações contidas neles (categoria e autor, por exemplo) encontram-se em formato textual. Para executar esta tarefa, foram utilizadas as bibliotecas *Scikit-learn*, *Pandas* e *NumPy*.

3.4. Treinamento

Após a etapa de transformação, iniciou-se a etapa de treinamento dos algoritmos. Como mencionado na seção 3.1, tanto os textos dos artigos quanto os seus metadados foram utilizados para o desenvolvimento da proposta. Desta forma, foi possível realizar uma análise tanto semântica como estatística, uma vez que os metadados apresentam informações estatísticas a respeito das notícias, como, por exemplo, número de substantivos, número de pronomes e tamanho médio de sentenças.

Para o processamento dos metadados dos artigos, foram analisados 3 algoritmos: Probabilístico - *Naive Bayes* (NB), baseado em otimização - *Multilayer Perceptron* (MLP) e *Support Vector Machine* (SVM). Após o treinamento, os algoritmos foram analisados para definir qual deles apresentaria melhores resultados e, portanto, seria utilizado pela aplicação final da proposta. Estes algoritmos foram selecionados por serem largamente utilizados na literatura em abordagens relacionadas à detecção de *fake news*.

Para o processamento dos textos, foi implementada uma Rede Neural Convolutiva (do inglês *Convolutional Neural Networks* - CNN), que é um algoritmo de *Deep Learning*, inspirado em modelos biológicos [Matsugu et al. 2003]. Apesar de largamente

²<http://nilc.icmc.usp.br/nilc/index.php/repositorio-de-word-embeddings-do-nilc>

utilizado em tarefas de Visão Computacional, a aplicação de CNN em tarefas de Processamento de Linguagem Natural também tem alcançado resultados satisfatórios, como, por exemplo, no *parsing* semântico e no processamento de texto [Sharma et al. 2019]. Nesta pesquisa, o CNN foi construído com a utilização de classes das bibliotecas *Keras* e *Tensorflow*, conforme descrito a seguir:

- **Camada de *Embedding***: nesta camada, é configurada a aplicação do *word embedding*. Para isto, utilizou-se o modelo pré-treinado, com 100 dimensões.
- **Camada Convolutiva**: nesta camada, são aplicados filtros sobre os dados de entrada, a fim de extrair as características dos textos. O filtro, ou *kernel*, é uma matriz com quantidade menor de linhas em relação à matriz de entrada, mas com a mesma quantidade de colunas, para que seja capaz de processar todos os valores das linhas da matriz de uma vez. Desta forma, o filtro funciona como uma janela deslizante, deslocando-se de cima para baixo pela matriz. O filtro contém inicialmente valores aleatórios, que são atualizados a cada iteração, a fim de extrair de forma cada vez melhor as características presentes no texto. O vetor resultante do processo de aplicação de um filtro é denominado vetor de características. Este processo é representado na Figura 2.

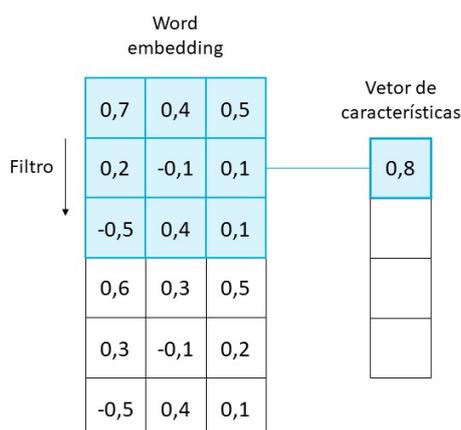


Figura 2. Processo de Convolução.

- **Camada de *Max-pooling***: o processo de *max-pooling*, conforme representado na Figura 3, consiste em selecionar o maior valor de cada vetor de características. Desta forma, é possível identificar as características mais relevantes de um texto.
- **Camada de *Flatten***: o processo de *max-pooling* apresenta como saída vários valores, correspondentes ao número de filtros utilizados. Por este motivo, é necessário agrupá-los novamente na camada de *Flatten*, onde os valores são concatenados, formando um novo vetor.
- **Camada ReLU**: o vetor resultante da camada anterior é processado por uma camada totalmente conectada, com função de ativação ReLU.
- **Camada Sigmoidal**: ao final da rede neural, por tratar-se de um problema de classificação binária, utiliza-se a função de ativação sigmoidal.

Uma vez concluído o treinamento do CNN e selecionado o melhor algoritmo para análise dos metadados, foi feita uma combinação sobre os resultados obtidos por ambos. Para isto, foram atribuídos pesos sobre os seus valores de saída, sendo realizada uma ponderação sobre os números calculados.

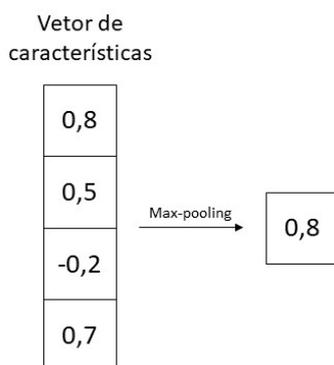


Figura 3. Max-pooling.

4. Resultados e Discussões

Nesta seção são apresentados os resultados obtidos de acordo com a metodologia criada para a detecção de *fake news* em textos de Língua Portuguesa.

Como supramencionado, a presente pesquisa trabalha com dois níveis de dados, sendo estes textos e metadados. Desta forma, foram selecionados diferentes algoritmos. Para o processamento dos textos, foi aplicado o CNN, e para o processamento dos metadados, foram estudados três algoritmos, sendo estes: MLP, SVM e NB, com o objetivo de eleger o algoritmo que mais se adequasse aos dados.

4.1. Análise dos Textos

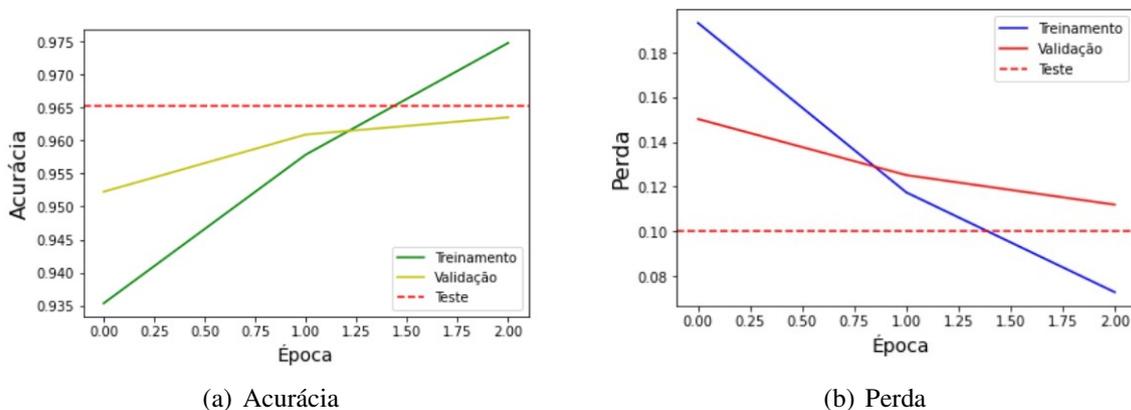


Figura 4. Resultados da CNN.

Inicialmente, para a aplicação dos testes, a base de dados foi dividida em 80% para treinamento e 20% para testes e validação. Em se tratando dos resultados obtidos a partir do treinamento/teste do processamento de textos com o CNN, foi possível constatar que o processo de aprendizagem é rápido, haja vista que o algoritmo é capaz de atingir resultados satisfatórios em apenas 3 épocas. Conforme pode ser observado na Figura 4(a), o CNN alcançou 97,3% de acurácia sobre a base de treinamento, 96,3% sobre os dados de validação e 96,5% nos testes. Observando a Figura 4(b), percebe-se que a perda sobre a base de treinamento foi de 7,4%, enquanto sobre os dados de validação foi de 11,2% e de 10% nos testes.

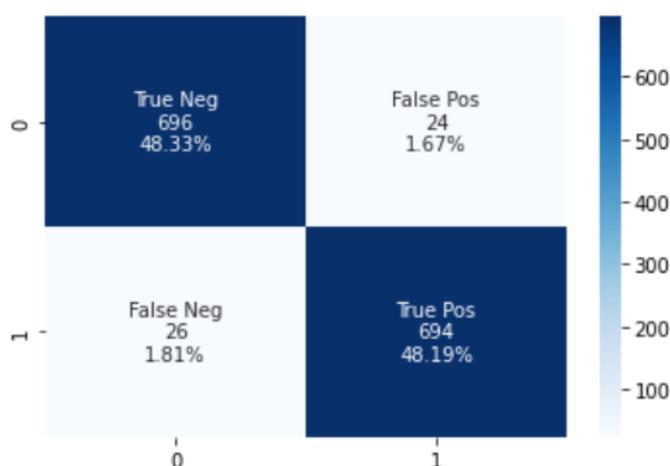


Figura 5. Matriz de confusão de teste do algoritmo CNN.

Quanto à matriz de confusão, Figura 5, nota-se que algoritmo apresenta mais erros em falsos negativos, ou seja, notícias que são verdadeiras e que, no entanto, são classificadas como falsas. Em um estudo aprofundado, buscando detectar a fonte do erro de classificação, o tamanho dos textos foi reconhecido como um grande influenciador, haja vista que o algoritmo apresenta maior dificuldade em aprender sobre textos menores, ou seja, que possuam poucos termos em sua construção. Isto se deve ao fato de que o *Fake.br* apresenta como *fake news* artigos que, em sua maioria, possuem estrutura pequena. Em contrapartida, artigos com estruturas maiores tendem a caracterizar-se como notícias que são verdadeiras.

4.2. Análise dos Metadados

No que se refere aos resultados dos testes no processamento dos metadados, o SVM apresentou melhor desempenho com uma acurácia de 97,5% e precisão de 98,3%. A performance deste e dos demais algoritmos, assim como as métricas de avaliação, podem ser conferidas na Tabela 3.

Tabela 3. Métricas de avaliação.

Algoritmos	Acurácia	Recall	Precisão	F1-Score
SVM	97.5%	96.6%	98.3%	97.4%
MLP	95.8%	94.3%	97.2%	95.7%
NB	96.4%	97.2%	95.7%	96.4%

Em se tratando das matrizes de confusão, apresentadas nas Figuras 6, 7 e 8, é perceptível que, assim como o CNN, os erros também se concentram principalmente no falso negativo, sendo que o NB evidenciou possuir, dentre os três algoritmos, o menor erro de classificação neste quesito, enquanto o MLP apresentou maiores erros de classificação. Vale destacar que, de forma geral, assim como na tabela de métricas de avaliação, o SVM destacou-se com o melhor desempenho e, da mesma forma, o MLP seguiu com uma baixa performance em comparação aos demais algoritmos testados.

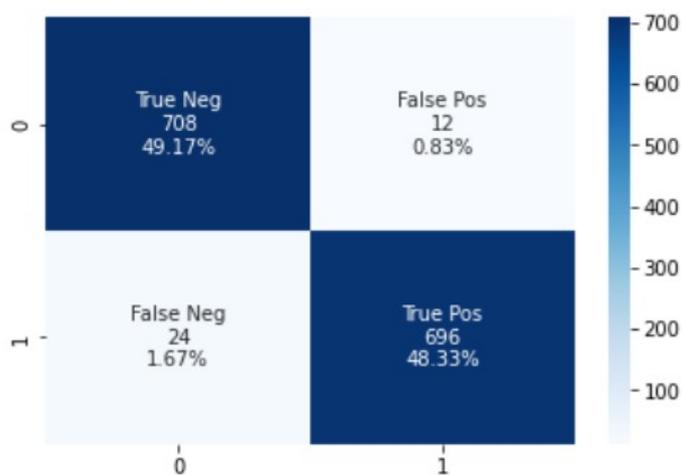


Figura 6. Matriz de confusão SVM.



Figura 7. Matriz de confusão MLP.



Figura 8. Matriz de confusão NB.

4.3. Inferência

Como mencionado na Seção 3.4, após concluído o treinamento do CNN e selecionado o melhor algoritmo para análise dos metadados, no caso o SVM, foi feita uma combinação

sobre os resultados obtidos por ambos. Considerando que ambos os modelos apresentam como saída valores compreendidos entre 0 e 1, esta inferência consiste em atribuir pesos sobre estes valores, o que permite atribuir maior relevância a um modelo em relação ao outro, e então realizar uma ponderação sobre os números calculados. Em relação à atribuição de pesos, fornecer valores iguais, definindo, desta forma, a mesma relevância para ambos os modelos, evidenciou ser a melhor opção, embora os números obtidos tenham sido muito próximos entre si, como pode ser observado na Tabela 4.

Tabela 4. Pesos atribuídos ao modelo.

Pesos (x, y)	Acurácia	Recall	Precisão	F1-Score
0.6 e 0.4	97.3%	97.2%	97.5%	97.3
0.5 e 0.5	97.7%	97.7%	97.7%	97.7%
0.4 e 0.6	97.7%	97.7%	97.6%	97.7%

Com os resultados obtidos, foi possível observar uma melhora considerável no desempenho ao utilizar os dois algoritmos combinados, em comparação às suas execuções individuais, constatando-se, de maneira geral, um aumento na acurácia e na precisão. A matriz de confusão referente à ponderação entre CNN e SVM, apresentada na Figura 9, mostra uma grande redução sobre a porcentagem de falsos negativos, o que significa que a aplicação erra menos ao classificar notícias verdadeiras como falsas.

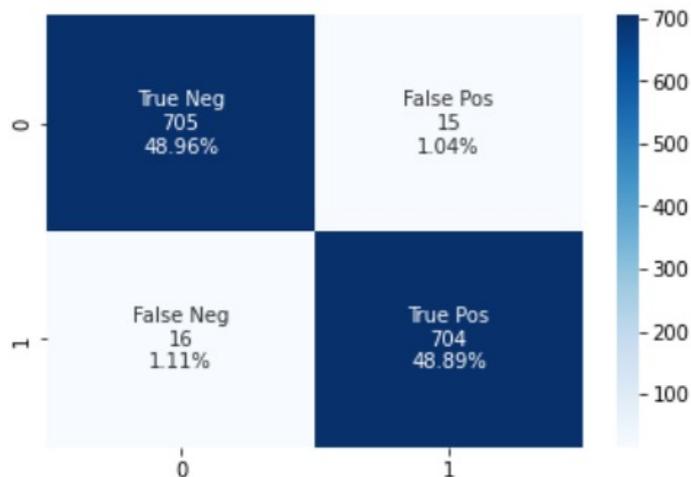


Figura 9. Matriz de confusão de teste do algoritmo CNN + SVM.

5. Considerações Finais

Este trabalho propõe um mecanismo para a análise de notícias em Língua Portuguesa e a detecção de *fake news*, utilizando Aprendizagem de Máquina e o conceito de Redes Neurais Convolucionais. Haja vista que, este mecanismo mostra-se uma importante ferramenta para a identificação de notícias falsas, sua rápida contenção e minimização de seus efeitos danosos à sociedade.

Tendo isso em vista, uma série de estudos foi realizada, como discorrido ao longo deste artigo, sendo possível, desta forma, identificar características de notícias falsas e verdadeiras, assim como a carência de estudos na literatura que abordem a detecção de *fake news* na Língua Portuguesa.

Analisando os algoritmos em sua totalidade, obteve-se uma acurácia acima dos 95% nos algoritmos testados. A solução apresentada por este artigo utiliza dois modelos gerados, respectivamente, pelos algoritmos CNN e SVM. O uso conjunto de ambos obteve resultados satisfatórios, demonstrando ser capaz de analisar padrões em textos e metadados de notícias e classificá-las, de forma acurada, como verdadeiras ou falsas.

Como trabalhos futuros, pretende-se aprimorar o classificador a fim de otimizar o seu desempenho, realizando novos testes com outros algoritmos de Aprendizagem de Máquina e redes neurais, como o *Long Short-Term Memory* (LSTM). Além disto, propõe-se o desenvolvimento de uma aplicação *mobile* que possa proporcionar uma análise quanto à veracidade de textos.

Referências

- Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31:211–236.
- Barbosa, V. N. (2020). Sirene: Solução inteligente para detecção de notícias falsas. Monografia (Bacharelado em Ciência da Computação), Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE.
- Gonzalez, M. and Lima, V. L. S. d. (2003). Recuperação de informação e processamento da linguagem natural. In *XXIII Congresso da Sociedade Brasileira de Computação*.
- IBGE (2020). *Acesso à Internet e à televisão e posse de telefone móvel celular para uso pessoal 2018*. IBGE.
- ITU (2021). Individuals using the internet, 2005-2021. <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>. Acesso em: 2022-02-23.
- Kaliyar, R. K., Goswami, A., Narang, P., and Sinha, S. (2020). Fndnet – a deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61:32–44.
- Matsugu, M., Mori, K., Mitari, Y., and Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16:555–559.
- Monteiro, R., Santos, R., Pardo, T., de Almeida, T., Ruiz, E., and Vale, O. (2018). Contributions to the study of fake news in portuguese: New corpus and automatic detection results. Villavicencio A. et al. (eds) *Computational Processing of the Portuguese Language. PROPOR 2018. Lecture Notes in Computer Science*, vol 11122. Springer, Cham.
- Oshikawa, R., Qian, J., and Wang, W. (2018). A survey on natural language processing for fake news detection. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*.
- Pereira, C. and Marques-Neto, H. (2021). Caracterização da reação de agências de fact-checking às publicações sobre a pandemia da covid-19 em redes sociais. In *Anais do XXXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 113–126, Porto Alegre, RS, Brasil. SBC.
- Raymer, K. and Harriss, L. (2017). Online information and fake news. <https://post.parliament.uk/research-briefings/post-pn-0559/>. Acesso em: 2020-10-10.

- Rosa, M. d. A., Amorim, H. C., Gomes, R. M., and Santos, B. A. (2019). Detecção de fake news: uma abordagem utilizando redes neurais convolucionais.
- Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., and Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology*.
- Wang, W. Y. (2017). "liar, liar pants on fire": A new benchmark dataset for fake news detection. *CoRR*, abs/1705.00648.