

Uso de Inteligência Artificial na Classificação de Notícias

Tiago de Oliveira Machado¹

1. RESUMO

O fenômeno da desinformação, também conhecido como fake news, vem se estabelecendo na cultura digital e ganhando grande repercussão devido às consequências que traz à sociedade. Neste contexto, a Inteligência Artificial se mostra uma ferramenta capaz de detectar e ajudar a mitigar o compartilhamento dessas informações. Neste projeto, explora-se o uso de modelos de aprendizado profundo de máquina para a construção de um classificador de veracidade de notícias em língua portuguesa. Exploraremos técnicas de processamento e vetorização de texto buscando construir um modelo que seja reproduzível para ser utilizado pela sociedade. Após a construção, o modelo foi treinado e avaliado em uma base de dados já classificada. Ao final, conseguimos uma acurácia maior que 92%, equiparando nosso modelo aos estados de arte existentes.

Palavras-chave: Inteligência Artificial, Aprendizado Profundo, Fake News, Processamento Natural da Linguagem (NLP).

2. INTRODUÇÃO

Há décadas, a evolução tecnológica vem influenciando o comportamento da sociedade, principalmente em relação à comunicação e novas formas de interação social. A democratização da tecnologia, acelerada em grande parte pelas necessidades criadas recentemente com a pandemia da Covid19, acelerou a inclusão da tecnologia no cotidiano de todos. Neste contexto, redes sociais, sites e blogs, que são plataformas onde qualquer cidadão tem autonomia para publicar, tornou comum e rápida disseminação de notícias no meio digital, e com ela, a divulgação de notícias falsas, denominadas fake news.

¹ Aluno de Engenharia Mecatrônica, Centro Universitário ENIAC. e-mail: tiagomachadodev@gmail.com

Esse fenômeno contemporâneo pode parecer inofensivo de acordo com uma perspectiva de impunidade, entretanto vem gerando preocupação por conta de seus impactos sociais, econômicos e políticos. A manipulação emocional promovida pelas fake news visa principalmente provocar fortes reações emocionais no leitor, sentimentos como raiva, indignação, frustração, e dificultam a capacidade do leitor em analisar racionalmente o conteúdo e provoca uma urgência em compartilhar a informação. Por isso, diversos estudos vêm sendo realizados com o intuito de desenvolver estratégias, ferramentas e métodos para o combate e mitigação das fake news.

Neste cenário, a inteligência artificial, que é uma área da ciência que tem como objetivo simular a inteligência humana, por meios tecnológicos, mostra-se uma excelente alternativa para identificação das fake news. O exponencial crescimento de tecnologias associadas ao campo da Inteligência Artificial, tem tornado a aplicação desta ferramenta cada vez mais corriqueira em nosso dia a dia. O uso de ferramentas de IA para classificação em geral é largamente utilizado desde meados da década passada, onde mais precisamente em 1990 o *filtro de spam* melhorou significativamente a vida de muitas pessoas. Tal ferramenta abriu caminhos para que diversos aplicativos de aprendizado de máquina fossem desenvolvidos e que agora, silenciosamente, fornecem centenas de produtos e recursos, desde recomendações a buscas por voz, que utilizamos regularmente (AURÉLIEN GÉRON, 2019).

3. OBJETIVOS

O objetivo deste trabalho é realizar uma prova de conceito sobre o uso da Inteligência Artificial como um aliado para classificação de notícias falsas. Realizaremos a experimentação e construção de um modelo funcional de classificação de notícias através do uso de técnicas de aprendizagem de máquina como ferramenta útil para a sociedade. Nesse contexto, exploraremos o uso de redes neurais na busca daquele que possuir a melhor acurácia, sendo possível representar de forma satisfatória resultados em dados reais e atualizados.

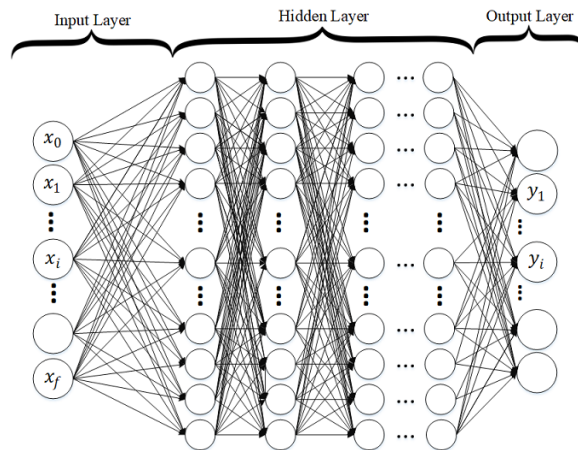
4. METODOLOGIA

A tarefa de entender um texto muitas vezes pode ser bem desafiador, sobretudo por causa da ambiguidade da linguagem natural. Os seres humanos não têm tanta dificuldade em desvendar os duplos-sentidos porque têm algo chamado conhecimento cultural comum e experiência anterior. Para as máquinas, prova-se ser uma tarefa difícil reconhecer a presença de ambiguidades. O desafio torna-se mais evidente quando lembramos que, muitas vezes, é necessário entender situações mais complexas como: reconhecer o contexto, tonalidade da voz falada, extrair informações, interpretar os sentidos, analisar os sentimentos, realizar análise sintática, semântica, lexical e morfológica (PINHEIRO, 2021).

Para modelar a língua e possibilitar que a máquina a entenda, utilizaremos técnicas de Processamento Natural da Linguagem (PNL), o campo da inteligência artificial que estuda a capacidade e as limitações de uma máquina para entender a linguagem dos seres humanos. Essas técnicas geralmente são constituídas a partir de modelos de aprendizado de máquina.

O Aprendizado de máquina é a ciência (e a arte) da programação de computadores para que eles possam aprender com os dados (JOEL GRUS, 2019). De forma geral, o aprendizado de máquina, mais precisamente as redes neurais, são algoritmos computacionais bio inspirados no sistema nervoso de animais . Esse tipo de estrutura é muito usado para processar grandes quantidades de dados a fim de classificar ou identificar dados novos semelhantes aos que foram treinados . As redes neurais são compostas por diversas camadas de neurônios de processamento conectados entre si. A figura 1 ilustra a arquitetura de uma rede neural profunda:

Figura 1: Arquitetura de uma Rede Neural Profunda



Fonte: MITTAL, 2019

Existem diversas arquiteturas de redes neurais. Neste estudo, exploraremos as arquiteturas Perceptron, que é um tipo de arquitetura que apresenta apenas uma camada. Também exploraremos o uso de uma rede neural recorrente (RNN), aliada uma a uma camada de memória (LSTM) para produzir os resultados que esperamos.

Utilizaremos também as funções de ativações para determinar a forma como nosso modelo entende os dados. A cada iteração no conjunto de dados, um peso é atribuído e atualizado, determinando quais neurônios permanecerão ativos na próxima iteração, de forma semelhante à como acontece com as sinapses do sistema nervoso animal. Neste processo, esse limite é estabelecido para caracterizar a aprendizagem daquele neurônio.

Antes de passar os textos para o modelo, serão necessárias etapas de pré processamento que abstraem e estruturam a língua, deixando apenas o que é informação relevante para a máquina. Reduzir o vocabulário, e torná-lo menos esparsos é fundamental uma vez que os conjuntos de documentos de texto são representados por um grande número de caracteres. Dessa forma, o objetivo do pré-processamento é extrair de textos uma representação estruturada e manipulável por algoritmos de classificação que identifique o subconjunto mais significativo para a coleção de texto (Rezende e outros (2011)), ou seja, obter uma representação com qualidade melhor do que a inicial.

As transformações comumente utilizadas consistem em retirar todos os caracteres especiais de uma frase, reduzindo-a a apenas caracteres alfanuméricos.

Depois, é comum a aplicação de um processo chamado tokenização, que consiste em decompor as frases em cada termo que a compõem. Na sequência, extraímos o radical da palavra através do processo de lematização. Ao final, será preciso transformar essa lista de palavras em vetores numéricos, uma vez que o texto é um tipo de dado não estruturado, e o computador ainda não tem capacidade de processar símbolos discretos, operando apenas em linguagem matemática essencialmente.

Para os modelos de aprendizado lineares, trataremos da vetorização do texto com o método “Bag-of-Words”. Este método constrói um vocabulário a partir de um corpus de documentos e conta quantas vezes as palavras aparecem em cada documento. Dito de outra forma, cada palavra do vocabulário torna-se um recurso e um documento é representado por um vetor com o mesmo comprimento do vocabulário (um “saco de palavras”).

Para os modelos de aprendizado profundo, trataremos da vetorização do texto com o método “Word Embedding”, que é o nome coletivo para técnicas de aprendizado de recursos onde as palavras do vocabulário são mapeadas para vetores de números reais. Esses vetores são calculados a partir da distribuição de probabilidade para cada palavra que aparece antes ou depois da outra. Dito de outra forma, palavras do mesmo contexto geralmente aparecem juntas no corpus, então elas também estarão próximas no espaço vetorial.

Dividiremos os dados em três partes: treino, contendo cerca de 70% do total dos dados; teste, contendo cerca de 30% do total dos dados, que ainda serão divididos em duas partes: teste, contendo 70% do total da amostra de teste, e validação, que consiste em 30% da amostra de teste para verificação final do modelo.

Treinaremos um modelo baseado no Teorema de Bayes, que utiliza a probabilidade para fazer previsões com base no conhecimento prévio de condições que podem estar relacionadas. Depois, avaliamos outros modelos e comparamos com este modelo base.

5. DESENVOLVIMENTO

A busca por uma base de dados que aborda o problema das fake news mostrou que, apesar de existir disponível muito conteúdo em inglês, o tema na língua portuguesa ainda foi pouco explorado. Em 2018, um grupo de pesquisadores na área da computação do instituto de Ciências de Matemática e Computação (ICMC) da USP São Carlos e linguistas da Universidade Federal de São Carlos (UFSCar), da Universidade Estadual de São Paulo (UNESP) e da Universidade Estadual de Maringá (UEM), desenvolveram um grandioso trabalho, que além de abordar a classificação de fake news para a língua portuguesa utilizando técnicas de aprendizado de máquina, realizou a construção do corpus Fake.br e disponibilizou o conjunto de notícias previamente classificadas na página web do projeto.

O corpus Fake.br feito pelo projeto Opinando dentro do Núcleo Interinstitucional de Linguística Computacional, ou NILC, reúne 7200 notícias em português, sendo 3600 falsas e 3600 verdadeiras, de diversos gêneros (MONTEIRO et al., 2018). O corpus ainda disponibiliza meta informações sobre cada elemento como autor, link onde foi obtido, gênero e outras métricas semânticas e sintáticas. Esses dados foram então divididos nas seguintes proporções: 70% do total da base de dados para treino, o que totaliza 5040 textos, sendo 2532 notícias verdadeiras e 2508 notícias falsas. Dos 30% dos dados restantes, 70% deste total utilizamos para teste final, o que totaliza 1512 textos, sendo destas 738 notícias verdadeiras e 774 notícias falsas. O restante dos dados utilizamos para validação durante o treinamento do modelo. Esses totalizam 648 textos, sendo 330 notícias verdadeiras e 318 notícias falsas.

Para a implementação do modelo, utilizamos como linguagem principal a linguagem Python, devido ao grande número de ferramentas e bibliotecas para aprendizado de máquinas disponíveis e mantido pela comunidade.

Para o processamento dos textos, utilizamos a biblioteca Spacy, que é um conjunto de ferramentas especializadas em processamento natural da linguagem, e possui um amplo suporte para a língua portuguesa. Nesta etapa as seguintes adequações nos textos são realizadas:

- Remoção de caracteres e pontuações não necessários para a compreensão do texto, ou que não são reconhecidos pelo vocabulário embedding (como ‘?’, ‘[]’, ‘{}’ etc).
- Transformação de todo o texto para letras minúsculas, evitando palavras duplicadas.
- Remoção das palavras que não são necessárias para a compreensão do texto, ou que não são reconhecidas pelo vocabulário embedding (como “o”, “a”, “um”, “uma” etc).
- Extração do radical das palavras

As figuras 2 e 3 abaixo mostram exemplos de um texto processado e um texto não processado.

Figura 2: Exemplo de um Texto Sem Processamento

```
Exemplo de Texto original:
Quantidade de palavras: 1116

'Moro está certo ao não se arrepender de divulgação de áudios de Lula e Dilma. Ele virou um herói. A Revista Veja promoveu o Fórum Amarelas Ao Vivo, que contou com a presença do juiz federal Sérgio Moro. No evento, ele foi indagado sobre a divulgação do áudio entre Lula e Dilma em março do ano passado. Assim ele respondeu: \x93Não me arrependo de forma nenhuma, embora tenha ficado consternado com a celeuma que a divulgação causou.\x94 \x93Na minha opinião eu fiz o que a lei exigia e o que eu achei que era necessário\x94, disse. \x93Não eram exatamente conversas republicanas\x94, disse. \x93Não cabe ao poder Judiciário servir guardião dos segredos sombrios dos nossos governantes\x94, completou. Para quem não se lembra na época Lula e Dilma conversavam ao telefone e sobre a indicação do ex-presidente ao Ministério da Casa Civil. O objetivo estava mais do que óbvio. Dilma o indicou para lhe conceder foro privilegiado, em clara obstrução da Justiça. Foi a partir da revelação destes áudios...'
```

Fonte: O autor, 2022

Figura 3: Exemplo de um Texto Processado

```
Exemplo de Texto após processamento:
Quantidade de palavras: 716

'morar certo arrepender divulgação áudio Lula dilma virar herói revista promover fórum amarelo vivo contar presença juiz federal Sérgio morer evento indagar divulgação áudio Lula dilma março ano passado responder arrepender ter ficar consternado celeuma divulgação causar opinião fiz lei exigir achar necessário dizer ser exatamente conversa republicano dizer caber judiciário servir guardião segredo sombrio governante completar lembrar época Lula dilma conversar telefone indicação expresidente ministério casa civil objetivo óbvio dilma indicar conceder foro privilegiar claro obstrução justiça revelação de este áudio criar clima enfim acaba r governo tirânico dilma rousseff dia Sérgio morer tornar herói nacional'
```

Fonte: O autor, 2022

Para a transformação dos textos em sequências numéricas, processo chamado vetorização, utilizamos as ferramentas da biblioteca Sklearn. Primeiro o vetorizador TfidfVectorizer, em todo nosso dataset de treinamento. Limitamos a quantidade máxima de características a soma total de palavras únicas neste corpus. Depois, realizamos a validação das melhores características com a utilização do método estatístico Chi². Por fim, um total de 385 características foram selecionadas

por terem uma correlação maior que 0.9 em relação a nossa variável alvo. Em termos práticos, isso significa que a palavra tem uma forte correlação com a variável alvo. A figura 4 ilustra as palavras mais comuns nos nossos dataset.

Figura 4: Exemplo de palavras mais comuns

<p>Palavras mais comuns nas notícias verdadeiras:</p> <pre>[('ser', 3196), ('ano', 2867), ('haver', 2862), ('ter', 2776), ('dizer', 2634), ('afirmar', 2536), ('dia', 2448), ('fazer', 2265), ('de', 2245), ('presidente', 2171), ('federal', 2139), ('caso', 2098), ('público', 2063), ('pessoa', 1990), ('ficar', 1948), ('governo', 1894), ('país', 1855), ('chegar', 1785), ('receber', 1754), ('passar', 1750)]</pre>	<p>Palavras mais comuns nas notícias falsas:</p> <pre>[('ser', 1590), ('dizer', 1568), ('presidente', 1060), ('ter', 1038), ('Lula', 910), ('ano', 904), ('país', 899), ('haver', 875), ('Brasil', 849), ('federal', 795), ('dia', 771), ('dilma', 769), ('acordo', 765), ('governo', 761), ('fazer', 720), ('hoje', 698), ('expresidente', 691), ('durante', 682), ('político', 630), ('pessoa', 628)]</pre>
---	--

Fonte: O autor, 2022

Nosso modelo mais simples, a regressão linear, foi implementado com a biblioteca Keras. Ela é composta por uma camada sequencial, que recebe os 774 parâmetros (que são os vetores do nosso vocabulário), e utiliza dois neurônios para realizar a classificação da notícia. Na saída, uma camada densa, de tamanho 2 (quantidade possível de alvos) e a função de ativação softmax. A figura ilustra o sumário da nossa rede:

Figura 5: Modelo Regressão Linear

Linear model on binary vectorized data:		
Model: "sequential_1"		
Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 2)	774
Total params: 774		
Trainable params: 774		
Non-trainable params: 0		
None		

Fonte: O autor, 2022

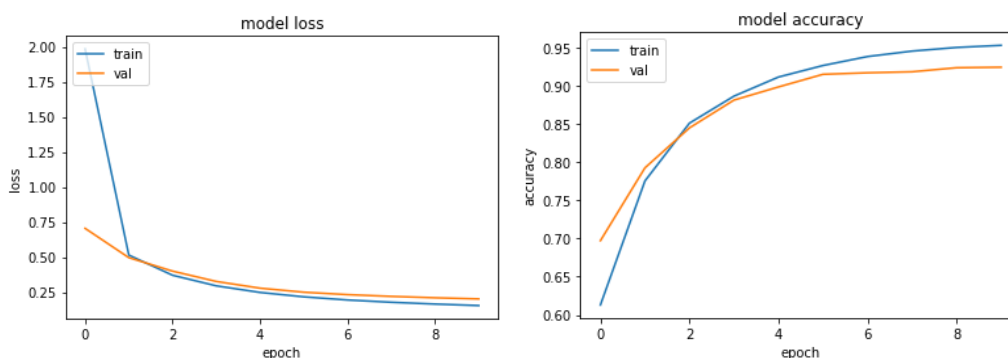
Para o treinamento dos modelos densos, utilizamos uma outra forma de vetorização nos textos: a técnica “Word Embedding”. Para isso, fizemos uso das ferramentas de processamento de texto da biblioteca Gensim. O processamento “Word Embedding” gera um vetor com muitas dimensões, por isso, neste projeto,

limitamo-os a 300. Após a vetorização, extraímos os unigramas e diagramas, e passamos para a rede. A construção da rede neural profunda foi realizada com a biblioteca Keras. Essa possui 4 camadas: a primeira, uma Embedding, para extrair os pesos das sequências de texto. A segunda, uma camada de atenção, que não influencia a predição, mas nos traz informações úteis das palavras que estão sendo consideradas. Na sequência, duas camadas bidirecionais LSTM para ordenar as palavras em sequências em ambas as direções. Por fim, uma camada densa, de tamanho 2 (quantidade possível de alvos) e função de ativação softmax.

6. RESULTADOS

Dentre todos os modelos que testamos, o que obteve o melhor resultado foi a Regressão Linear. Este modelo obteve uma acurácia de 92,33% no conjunto de dados de treinamento, e 92,44% no conjunto de dados de validação. Em cada etapa do treinamento, avaliamos a função perda ('loss') e a acurácia ('accuracy') do modelo. Os resultados são apresentados na figura 6 a seguir.

Figura 6: Acurácia e Perda do Modelo em Relação às Épocas de Treinamento



Fonte: O Autor, 2022

Podemos perceber que o modelo conseguiu obter sua melhor acurácia após 4 épocas de iteração. Notamos também que a curva de aprendizado em treinamento (linha em laranja) está próxima a curva de aprendizado em validação (linha em azul), o que indica que o modelo não está sofrendo de overfitting.

7. CONSIDERAÇÕES FINAIS

Dados os resultados apresentados, concluímos que nosso modelo atingiu com êxito os resultados esperados. A tarefa de classificação de textos não é fácil, mas após a refinação dos dados e ajuste dos parâmetros do modelo, conseguimos no geral um ótimo resultado. Disponibilizaremos este modelo em um portal, para que todos possam ter acesso.

Como contribuição pessoal, este trabalho nos ofereceu a oportunidade de aprender mais sobre as técnicas de aprendizado profundo, especialmente àquelas que envolvem o processamento de textos em língua portuguesa.

Acreditamos ainda que este trabalho será de grande valia para a sociedade como um todo. Não apenas para aqueles que desejam checar a veracidade de uma notícia, como também para aqueles que por ventura possam utilizar este texto como referência para pesquisas futuras.

8. FONTES CONSULTADAS

ENDO PT, Santos GL, de Lima Xavier ME, Nascimento Campos GR, de Lima LC, Silva I, Egli A, Lynn T. **Illusion of Truth: Analysing and Classifying COVID-19 Fake News in Brazilian Portuguese Language. Big Data and Cognitive Computing.** 2022; 6(2):36. Disponível em: <https://doi.org/10.3390/bdcc6020036>. Acesso em: 04 out. 2022.

GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition.** 2019. ISBN: 9781492032649. Disponível em: <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>. Acesso em: 14 out. 2022.

GRUS, Joel. **Data Science do Zero.** 2021. ISBN: 9781492041085. Disponível em: <https://www.oreilly.com/library/view/data-science-from/9781492041122/>. Acesso em: 14 out. 2022.

GUARISE, Lucas. **Detecção de notícias falsas usando técnicas de deep learning.** 2019. Disponível em:

<https://repositorio.usp.br/directbitstream/494173d1-2375-4063-8b9c-8146811ffe6f/lucas%20guarise.pdf>. Acesso em: 14 out. 2022.

HAMID, Abdullah *et al.* **Fake News Detection in Social Media using Graph Neural Networks and NLP Techniques: A COVID-19 Use-case**. 2020. Disponível em: <https://doi.org/10.48550/arXiv.2012.07517>. Acesso em: 04 out. 2022.

MITTAL, S. **Deep Learning Techniques for Text Classification**. 2019. Disponível em: <https://medium.com/datadriveninvestor/deep-learning-techniques-for-text-classification-9392ca9492c7>>. Acesso em: 12/10/2022.

MONTEIRO, R. A.; SANTOS, R. L. S.; PARDO, T. A. S.; ALMEIDA, T. A. de; RUIZ, E. E. S.; VALE, O. A. **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. Acesso em: 12/10/2022.

PIETRO, Mauro di. **Text Classification with NLP: Tf-Idf vs Word2Vec vs BERT**. 2020. Disponível em: <https://towardsdatascience.com/text-classification-with-nlp-tf-idf-vs-word2vec-vs-bert-41ff868d1794>. Acesso em: 14 out. 2022.

PINHEIRO, Nina Maria. **Introdução ao Processamento de Linguagem Natural — Natural Language Processing(NLP)**. 2021. Disponível em: <https://medium.com/data-hackers/introdu%C3%A7%C3%A3o-ao-processamento-de-linguagem-natural-natural-language-processing-nlp-be907cd06c71>. Acesso em: 4 out. 2022.

SANTOS, Cibian Fernanda del Gobbo. **Inteligência artificial na detecção de fake news : uma revisão sistemática**. 2021. Disponível em: <https://hdl.handle.net/1884/72885>. Acesso em: 04 out. 2022.