O uso de Machine Learning na classificação de textos com ênfase em Fake News.

Marco P. Altieri, Paulo H. Apro, Antonio L. Basile.

Faculdade de Computação e Informática
Universidade Presbiteriana Mackenzie (UPM) – São Paulo, SP – Brasil

marcoaltieri96@hotmail.com, pauloapro@gmail.com,
antonioluiz.basile@mackenzie.br

Abstract. *Currently, the term Fake News is increasingly entering our daily lives, and news travels great distances in a short period of time on the internet and social media. The present work sought to compare the Random forest, Decision Tree and Logistic Regression classification algorithms in the classification of news described as fake news. aiming at exploring these models and their effectiveness.*

Keywords: *Fake News, Random forest, Decision tree, Logistic regression, news classification, Machine Learning.*

Resumo. *Atualmente o termo Fake News vem adentrando cada vez mais em nosso cotidiano, e as notícias percorrem grandes distancias em um período curto de tempo na internet e nas mídias sociais. O presente trabalho buscou comparar os algoritmos de classificação Random forest, Árvore de decisão e Regressão logística na classificação de notícias descritas como fake news. visando a exploração desses modelos e sua efetividade.*

Palavras – chave*: Fake News, Random forest, Árvore de decisão, Regressão logística, classificação de notícias, Aprendizado de Máquina.*

1. Introdução

A deflagração de notícias falsa, popularmente chamadas de fake news, tornaram-se uma grande preocupação devido à sua capacidade de criar impactos devastadores no que se refere às desinformações que são disseminadas no mundo virtual. Nos últimos anos as aplicações de aprendizado de máquina apresentaram sucesso no reconhecimento de padrões, por esse razão elas estão sendo cada vez mais utilizadas para a análise de textos, principalmente em textos de notícias falsas (El Naqa et al., 2015).

Devido à facilidade que as pessoas têm em acessar, compartilhar e comentar, às notícias são mais fáceis e rápidas de serem consumidas. (Shu et al., 2017). Segundo (Bondielli et al., 2019) o compartilhamento de conteúdo falso tem maior adesão dos usuários nas mídias sociais do que páginas que têm conteúdo jornalístico real. No Brasil, 92% dos usuários de internet estão em alguma mídia social (C. Rock, 2018).

Neste contexto, a utilização de ferramentas de machine learning podem contribuir para identificar a disseminação de notícias intencionalmente falsas. Desta forma, o presente trabalho objetivou aplicar e comparar os algoritmos de classificação Random

forest, Árvore de decisão e Regressão logíca utilizadas nos processos de identificação e classificação de Fake News, visando a exploração desses modelos e sua efetividade.

2. Referencial Teórico

2.1. Fake News.

A expressão fake news é relativamente recente e sua caracterização se faz necessária, pois existem diferentes interpretações para fake news. Mesmo com as mídias tradicionais, já existiam pessoas que divulgavam notícias falsas propositalmente. Independentemente do surgimento, fake news apresenta muitas definições que podem ser divididas em dois grupos. (Golbeck et al., 2018).

No primeiro grupo é considerado o aspecto intencional, pois define as publicações intencionais e verificadas como falsas. Assim, não basta a notícia ser falsa para ser caracterizada uma fake news, é preciso ser intencional, podendo também ser divulgada em uma mídia digital (Zhou & Zafarani, 2020).

O segundo grupo, classifica as fake news como notícias falsas independentemente da sua natureza intencional, como, por exemplo, falta de interpretação sobre o conteúdo abordado (Sharma et al., 2019), programadas para determinadas ações, utilizam algoritmos complexos para a melhor tomada de decisão.

2.2. Detecção de notícias falsas usando aprendizado de máquina.

Muitos estudos já estão focados em detectar fake news, segundo Zeba Khanam, pesquisador em Harbin Institute of Technology, Shenzhen, China, em seu artigo " Fake News Detection Using Machine Learning Approaches, 2021" propôs uma análise relacionada à detecção de notícias falsas explorando os modelos tradicionais de aprendizado de máquina para escolher o melhor.

Uma solução poderia ser o desenvolvimento de um sistema para fornecer uma pontuação de índice automatizada confiável, ou classificação de credibilidade de diferentes editores e contexto de notícias (Z Khanam et al., 2021).

2.3. Linguagem natural.

O processamento da linguagem natural é uma área de inteligência artificial que tem por objetivo a interpretação e manipulação das línguas humanas, envolve traduzir linguagem natural em dados numéricos que um computador pode usar para aprender sobre o mundo, é um processo chamado de vetorização. (Lane et al., 2019). O bag of words é uma técnica que transformará um texto original em um conjunto de palavras e a frequência que uma palavra aparece no texto é calculada.

A saída do método é uma matriz em que cada coluna representa uma palavra no vocabulário e a linha corresponde a um texto. E ao final sairá o número de vezes que a palavra aparece. (Ghosh et al., 2019).

De acordo com Freire e Goldschmidt (2019, apud Laura D et al., 2020) a identificação automática de notícias falsas pode ser entendida como um problema de classificação binária, em que uma notícia é dada como uma entrada (ε), a tarefa de

identificar notícias falsas é prever se essa notícia é falsa ou não, ou seja, f: ε de modo que f é a função de previsão.

2.4. Árvore de decisão

Árvore de decisão é um algoritmo de aprendizagem de máquinas que é usado tanto em problemas de regressão quanto de classificação. Neste modelo, por meio de uma regra de decisão, um conjunto complexo de dados é dividido sucessivamente em conjuntos menores, que também serão divididos usando a mesma estratégia. Os resultados dos conjuntos menores podem ser combinados, formando uma árvore. Assim teremos a solução do conjunto complexo de dados. O algoritmo utiliza a representação de uma árvore dividida em raiz, nós internos e folhas (Oliveira, Érick Ritir, 2019).

Cada nó interno da árvore de decisão especifica uma condição ou um "teste" um atributo e a ramificação são feitas com base nas condições de teste e no resultado. O nó folha carrega um rótulo de classe que é obtido após o cálculo de todos os atributos e distância da raiz à folha representa a regra de classificação (Z Khanam et al., 2021).

Segundo Oliveira, Leandro Massetti et al. (2019), em seu exemplo de representação de árvore de decisão para avaliar se o dia está favorável para uma partida de futebol, visto na Figura 1. É escolhido um atributo para ser a raiz da árvore, depois é criado um galho para cada possível valor, o processo se repete recursivamente em cada galho. Em algum instante, se todas as instâncias de um nó tem a mesma classificação, o desenvolvimento da árvore naquele nó é parado.

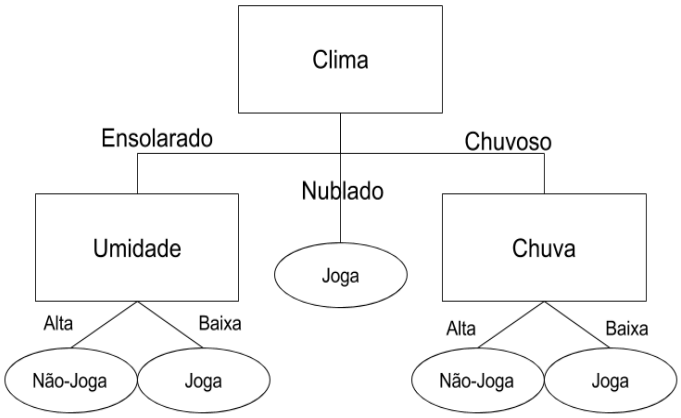

Figura 1. Exemplo de árvore de decisão. Fonte: Oliveira, Leandro M. et al.(2019).

A escolha dos atributos para cada nó é orientada por uma medida que indica o quão bem um atributo divide as classes. Dentre diversas técnicas de divisão de atributos, existe a técnica baseada no Ganho de Informação. (Oliveira, Leandro M.et al.,2019). A ideia dessa técnica segue o conceito de entropia, isto é, descreve o grau de desordem ou aleatoriedade em um sistema. O objetivo em um conjunto de treinamento é encontrar o atributo com pouca aleatoriedade e dividir melhor o problema.

2.5. Random forest.

Random Forests é um tipo específico de ensemble learning no qual treinamos várias árvores de decisão com um único conjunto de dados, ou seja, Random forest é uma

coleção de árvores de decisão. Cada árvore é treinada com um subconjunto do conjunto de dados e as diversas árvores obtidas constituirão um único classificador (Uma Sharma et al., 2021).

De acordo com Oliveira, Érick Ritir (2019), para realizar uma classificação, cada árvore individual apresenta uma previsão de classe e a classe com mais votos torna-se a previsão final do modelo. Na Figura 2, temos um exemplo de Random forest. O algoritmo depende de várias árvores de decisão que são todas treinadas de forma ligeiramente diferente, todas elas são levadas em consideração para a classificação final (R. Silipo, 2019). No exemplo da Figura 2, a ideia é predizer uma cor e cada árvore de decisão efetua a sua classificação.

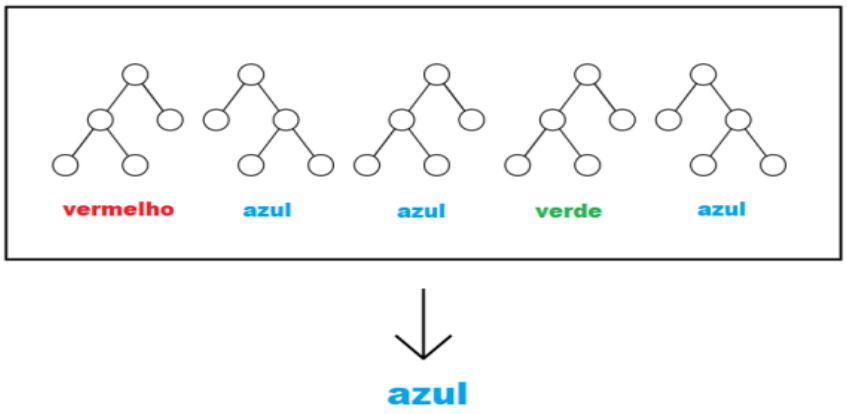

Figura 2. Exemplo de Random forest. Fonte: Oliveira, Érick Ritir (2019).

2.6. Regressão logística.

Também chamado de Regressão Logit, é utilizada para estimar valores discretos (valores binários como 0/1, sim/não, verdadeiro/falso) com base em determinado conjunto de variáveis independentes.

Segundo Aurélien Géron (2019, p.139) estima a probabilidade de uma instância pertencer a uma determinada classe (por exemplo, qual é a probabilidade desse e-mail ser spam?). Se a probabilidade estimada for maior que 50%, então o modelo prevê que a instância pertence a essa classe (chamada de classe positiva, rotulada como "1"), ou então ela prevê que não (isto é, pertence à classe negativa, rotulada "0"). Isso o transforma em um classificador binário.

A seguir temos as equações descritas por Aurélien Géron (2019, p.140). Equação 1. Modelo de probabilidade estimada (forma vetorizada).

Equação 1.

$$p = h_\theta \, \mathrm{x} = \sigma(\theta^T \cdot \mathrm{x})$$

A logística, subscrita $\sigma(\cdot)$, é uma função de formato em S que mostra um número entre 0 e 1. Definida na Equação 2. Função Logística e Figura 4.

Equação 2.

$$\sigma(\mathrm{t}) = \frac{1}{1 + \exp\,[-\mathrm{t}]}$$

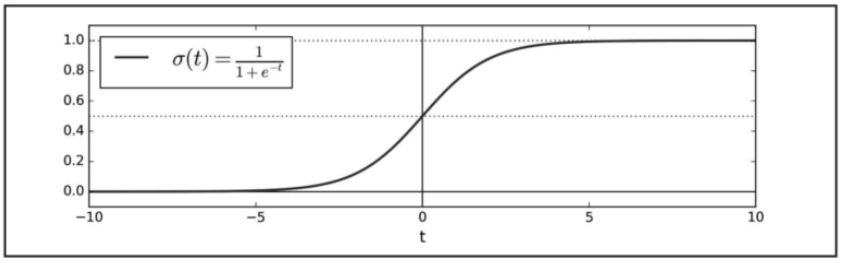

Figura 3. Função Logística. Fonte: Aurélien Géron (2019).

Depois que o modelo estimou a probabilidade p^= hθ(x) que a instância x pertence à classe positiva, ela pode fazer facilmente sua previsão ŷ. Previsão do modelo de regressão logística, Equação 3.

Equação 3.

$$\hat{y} = \begin{cases} 0 \text{ se } p < 0.5, \\ 1 \text{ se } p \geq 0.5. \end{cases}$$

Nota-se que σ(t) < 0,5 quando t < 0, e σ(t) ≥ 0,5 quando t ≥ 0, então um modelo de Regressão Logística prevê 1 se θT · x for positivo, e 0 se for negativo.

3. Metodologia

Para a realização desta trabalho optou-se por dividi-lo em duas etapas. A primeira etapa se inicia com um levantamento bibliográfico sobre o tema da pesquisa. A comparação e análise da fundamentação teórica foi fundamental para a escolha das ferramentas e algoritmos utilizados para a realização deste trabalho. A partir do levantamento de artigos e livros, seguimos para segunda etapa. Identificamos os modelos tradicionais de aprendizado de máquina mais utilizados para classificação de fake news. Os algoritmos de classificação aplicados para comparação e análise foram Árvore de Decisão, Random Forests e Regressão logística. As fases do processo de análise deste trabalho estão em forma de um diagrama na Figura 4.

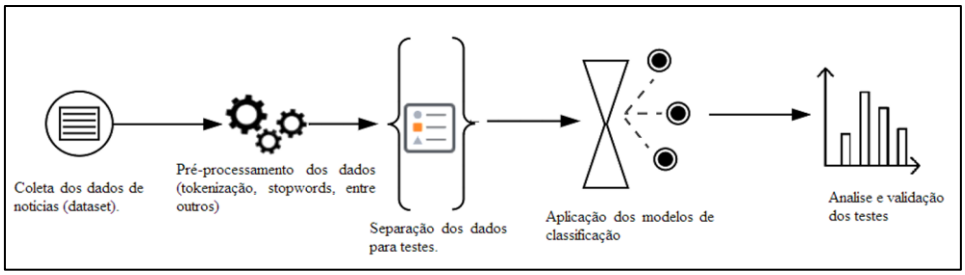

Figura 4. Diagrama com as fases do processo de análise. Fonte: Autoria Própria.

3.1. Desenvolvimento.

Neste estudo, utilizou como linguagem de programação Python e suas bibliotecas Scikit-Learn, que estão ilustradas na Figura 5, e como ambiente de desenvolvimento foi usado o Google Colab, um framework online onde se pode escrever, executar códigos de Deep Learning e Machine Learning. O Python possui um enorme conjunto de bibliotecas e

extensões, que podem ser facilmente usadas em Machine Learning. Segundo Uma Sharma et al., 2021 a biblioteca Scikit-Learn é a melhor fonte para algoritmos de aprendizado de máquina, onde quase todos os tipos de algoritmos de aprendizado de máquina estão prontamente disponíveis para Python, portanto, é possível uma avaliação fácil e rápida de algoritmos de Machine Liearning.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn import feature_extraction, linear_model, model_selection, preprocessing
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
import re
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from nltk.corpus import stopwords
import nltk
nltk.download('stopwords')
```

Figura 5. Bibliotecas Scikit-Learn. Fonte: Autoria Própria.

A base de dados utilizada para treinamento e testes dos algoritmos neste estudo é da Kaggle.com, uma plataforma de comunidade online da Google para cientistas de dados que, disponibiliza conjuntos de dados de notícias falsas e reais para trabalhos e pesquisas, todos os artigos e notícias retirados da base de dados para este trabalho são de origem norte americana de língua inglesa. Segundo Huang, Jeffrey (2020), esses conjuntos de dados têm sido amplamente utilizados em diferentes trabalhos de pesquisa para determinar a veracidade de notícias.

O conjunto de dados estão no formato .CSV denominados Fake.csv e True.csv para teste, sua implementação está na Figura 6.

```python
fake = pd.read_csv("/content/Fake.csv")

true = pd.read_csv("/content/True.csv")
```

Figura 6. Import dos dados. Fonte: Autoria Própria.

Todo o conjunto de dados de notícias contém 23 mil linhas de notícias descritas como falsas e 21 mil linhas de notícias verdadeiras. Os dados referem-se às categorias de notícias variadas de política, governo e mundo, demonstradas em porcentagem na Figura 7.

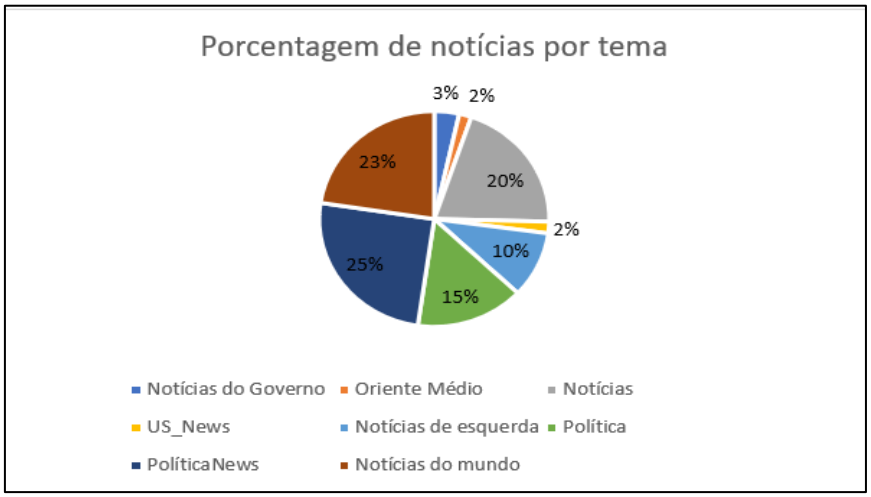

Figura 7. Notícias. Fonte: Autoria Própria.

Após a fase de coleta de dados, é realizado o pré-processamento dos dados utilizando o CountVectorizer, usado para transformar um determinado texto em um vetor com base na frequência (contagem) de cada palavra que ocorre em todo o texto. A limpeza dos dados também inclui Tokenização para separar o texto em unidades, como frases ou palavras. Utilização de Stopwords para remover palavras irrelevantes, Figura 8. O TfidfTransformer é aplicado no corpo do texto, de modo que a contagem relativa de cada palavra nas frases é armazenada na matriz do documento, ou seja, calcula a frequência com que um termo aparece em um documento.

```
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
stop = stopwords.words('english')
```

```
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself
```

Figura 8. Stopwords. Fonte: Autoria Própria.

É adicionado um target, Fake e True, para melhor identificação dos arquivos, Figura 9. A concatenação dos arquivos em um só corpo pode ser identificada na Figura 10. E a padronização do texto para letras minúsculas na Figura 11.

```
fake['target'] = 'Fake'
true['target'] = 'True'
```

Figura 9. Target, Fake e True. Fonte: Autoria Própria.

```
data = pd.concat([fake, true]).reset_index(drop = True)
data.shape
```

	title	text	subject	date	target
0	#NewOrleans: BLACK PATRIOTS Ready To Fight An...	May 7th likely going day clashes protesters co...	politics	May 7, 2017	Fake
1	Islamists lure youngsters in the Philippines w...	MARAWI CITY, Philippines (Reuters) - When saw ...	worldnews	September 21, 2017	True
2	Detained Venezuelan-U.S. Citgo executives to b...	WASHINGTON/CARACAS (Reuters) - Venezuelan Pres...	worldnews	November 22, 2017	True
3	(VIDEO) Female College Students Protesting Bec...	21st Century Wire says US college students con...	Middle-east	November 11, 2016	Fake
4	Melania's Slovenian hometown eyes Trump win as...	SEVNICA, Slovenia (Reuters) - The small Sloven...	politicsNews	November 9, 2016	True

Figura 10. Concatenação dos arquivos. Fonte: Autoria Própria.

```
data['text'] = data['text'].apply(lambda x: x.lower())
data.head()
```

	title	text	subject	target
0	Trump flag-burning tweet leads activists to bu...	new york (reuters) - a small group of hard-lef...	politicsNews	True
1	WHILE OBAMA VACATIONS AND HANDS OUT A BILLION ...	hey barry tell us again about what a great dea...	left-news	Fake
2	Venezuela opposition blames Maduro for detaine...	caracas (reuters) - venezuela s opposition bla...	worldnews	True
3	Bette Midler Sums Up The Republican Party In ...	it really is perfect.and it s from a book that...	News	Fake
4	Seeking to extend martial law in Philippine so...	manila (reuters) - philippine president rodrig...	worldnews	True

Figura 11. Padronização do texto letras minúsculas. Fonte: Autoria Própria.

Ao final do pré-processamento, a quantidade de dados é reduzida para análise de testes, chegando em torno de 12 mil dados analisados. Os imports dos algoritmos de Regressão Logística, Random Forest e Árvore de decisão estão ilustrados na Figura 12.

```
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
```

Figura 12. Imports dos algoritimos . Fonte: Autoria Própria.

Para medir a performance dos três modelos classificadores, utilizou-se da acuracidade para tratar o total de acertos obtidos em relação ao total de experimentos e, avaliando a qualidade dos resultados obtidos, a matriz de confusão apresenta os dados reais e a previsibilidade dos algoritmos. A precisão, Erro, Recall e F1-Scorre, também foram tratados para medir a performance dos modelos. A implementação da matriz de confusão e acuracidade estão na Figura 13.

```
cm = metrics.confusion_matrix(y_test, prediction)
plot_confusion_matrix(cm, classes=['Falso', 'Verdadeiro'])
```

```
print("acuracidade: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
```

Figura 13. Matriz de confusão e acuracidade . Fonte: Autoria Própria.

3.2. Resultados.

Neste estudo, foram analisados os desempenhos de modelos de aprendizagem de máquina, Árvore de decisão, Random Forest e Regressão Logística. Aproximadamente 11.225 dados foram testados para comparação entre os modelos. A Matriz de confusão para todos os algoritmos está representada abaixo na Figura 14, mostrando duas linhas e duas colunas que descrevem a quantidade de falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos.

Nos testes realizados a Árvore de decisão obteve um total de acertos de 11.181, dos quais 5854 identificados como fake news e 5327 identificados como notícias verdadeiras. Random Forest registrou um total de 11.105 acertos, dos quais 5792 identificados como fake news e 5313 identificados como notícias verdadeiras. A Regressão Logística indicou um total de 11.082 acertos, dos quais 5783 identificados como fake news e 5299 identificados como notícias verdadeiras. Quanto ao total de erros registrados nos testes, o modelo de Árvore de decisão obteve os menores erros dentre os

modelos, 44 erros em relação a Random Forest com 120 erros e Regressão Logística com 143 erros.

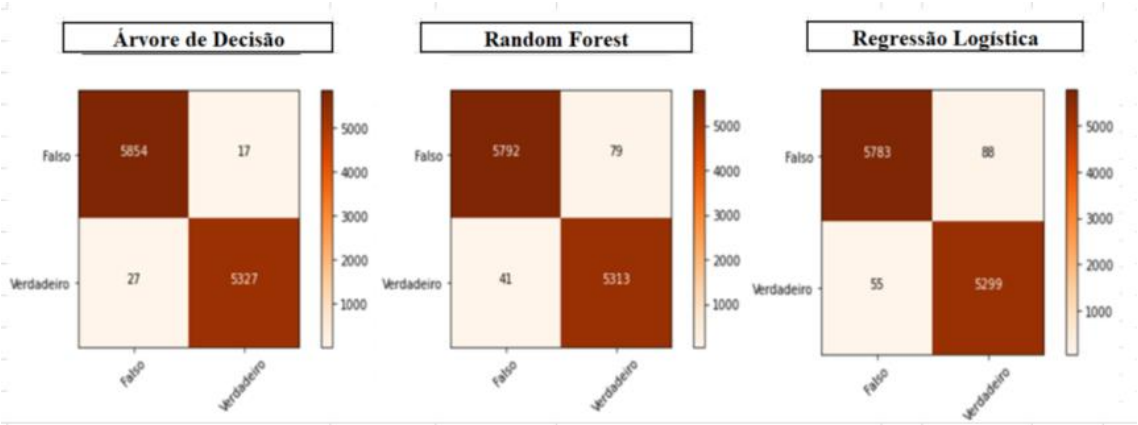

Figura 14. Resultados da Matriz de Confusão. Fonte: Autoria Própria.

Ao medir a performance dos modelos com as técnicas de Acuracidade, Erro, Recall, Precisão e F1-Score, vistos na Tabela 1. Identificou-se uma precisão de 0,99 em todos os modelos analisados quanto a precisão, além do F1-Score que obteve 0,99 para Árvore de decisão e 0,98 para Random Forest e Regressão Logística. Estes valores de precisão poderão ser entendidos como modelos de precisão alta, poucas taxas de falso positivo e falso negativo, vistos na matriz de confusão, Figura 14. Destaca-se a acuracidade do modelo Árvore de decisão com 99,61%, seguido dos modelos de Random Forest 98,93% e Regressão Logística 98,73%.

Tabela 1. Comparação da performance para todos os três modelos.

Modelos	Acuracidade	Total Acertos	Total Erro	Recall	Precisão	F1-Score
Árvore de decisão	99,61%	99%	0,39%	0,99	0,99	0,99
Random Forest	98,93%	98%	1%	0,98	0,99	0,98
Regressão Logística	98,73%	98%	1%	0,98	0,99	0,98

4. Conclusão

O presente trabalho buscou comparar os algoritmos de classificação Random forest, Árvore de decisão e Regressão logística na classificação de notícias descritas como fake news. Visando a exploração desses modelos e sua efetividade.

Ao analisar os resultados obtidos pode-se concluir que, dentre os algoritmos selecionados, a Árvore de decisão foi o que demonstrou o melhor desempenho na classificação de fake news da base supervisionada, registrando uma acuracidade de 99,61% e obtendo um bom desempenho na matriz de confusão com os menores erros. Não muito distantes a acuracidade registradas nos modelos Random Forest e Regressão Logística foram de 98,93% e 98,73% respectivamente.

Com o objetivo de comparar a taxa de acerto entre os modelos de algoritmos citados, os resultados poderão ser entendidos como modelos de precisão alta. Entretanto, na base de dados utilizada da Kaggle verificou-se o fato de que, em relação às notícias reais, as notícias falsas têm muito mais tokens do que as reais, as notícias falsas são uma

mistura de twitters e notícias. O uso gírias também se sobressai na análise, visto que, são dados de usuários em suas redes sociais e são classificadas como notícias falsas. Ao analisar as palavras com maior frequência foi identificado que, as palavras “said" e "reuters" têm grande destaque no conjunto de dados. A simples presença da palavra “reuters” já classificaria como real.

O fato de fake news se misturar com twitters apontam uma tendência para classificar as notícias, porém apenas o fato de menções com @ de usuários do twitter não é suficiente para afirmar que os dados não são adequado para identificar fake new.

Pois com um novo processamento dos dados que remova “@” seria capaz de diminuir uma tendência para notícias falsas, assim como a quantidade de tokens, só trabalharíamos com fake news que fosse menor ou igual a notícia com mais tokens, então esse problema poderia ser administrado hipoteticamente. Seria uma solução para esta base de dados, podendo gerar taxas de acertos mais realistas.

De acordo com outras obras utilizadas como base para o projeto e analisando seus resultados, a variedade de algoritmos de aprendizagem de máquinas demonstram uma grande capacidade de detectar fake news. Acreditamos que a escolha de uma base dados confiável e bem analisada é fundamental.

É possível que no futuro as técnicas de aprendizado de máquinas se tornarão um método padrão na identificação de notícias intencionalmente falsas.

5. Referências

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. Information Sciences, 497, 38-55. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0020025519304372?via%3Dihub. Acesso em: 2 abril 2022.

C. Rock. Social media trends 2018. Technical report, 2018. Disponível em: https://cdn2.hubspot.net/hubfs/355484/Ebooks%20MKTC/Social%20Media.pdf Acesso em: 2 abril 2022.

El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In machine learning in radiation oncology (pp. 3-11). Springer, Cham. Disponível em: https://books.google.com.br/books?id=1N7yCQAAQBAJ&printsec=frontcover Acesso em: 2 abril 2022.

Freire, P. M. S., & Goldschmidt, R. R. (1). Combate automático às Fake News nas mídias sociais Disponível em: http://ebrevistas.eb.mil.br/CT/article/view/8639 Acesso em: 2 abril 2022.

GÉRON, Aurélien. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Sebastopol: O’Reilly, 2017.

Ghosh, S., & Gunning, D. (2019). Natural Language Processing Fundamentals. Disponível em: https://www.perlego.com/book/955577/natural-language-processing-fundamentals-build-intelligent-applications-that-can-interpret-the-human-language-to-deliver-impactful-results-pdf Acesso em: 2 abril 2022.

Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K. H., Bonk, C., Bouzaghrane, M. A., Visnansky, G. (2018). Fake news vs satire: A dataset and analysis. Disponível em: https://www.eecis.udel.edu/~mlm/docs/2018-Golbeck-WebSciFakeNewsVsSatire.pdf Acesso em: 2 abril 2022.

Jeffrey Huang (2020). Detecting Fake News With Machine Learning.Disponível em: https://iopscience.iop.org/article/10.1088/1742-6596/1693/1/012158/meta.Acesso em: 2 Set 2022.

Kaggle.com (2020). Fake and real news dataset .Disponível em: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset?select=True.csv Acesso em: 10 Set 2022.

Lane, Hobson. Howard, Cole. Hapke, Hannes. (2019). Natural Language Processing in Action: Understanding, analyzing, and generatint text with Python. Manning. Disponível em : https://pt.scribd.com/book/511817149/Natural-Language-Processing-in-Action-Understanding-analyzing-and-generating-text-with-Python Acesso em: 4 abril 2022.

Laura D. de Almeida, Victor Fuzaro, Falmer V. Nieto, André L. M. Santana.(2020) Identificação de "Fake News" no contexto político brasileiro: uma abordagem computacional. Disponível em: https://sol.sbc.org.br/index.php/wics/article/view/15966/158 07 Acesso em: 2 abril 2022.

Oliveira, Érick Ritir (2019).Reconhecimento de Bots no Twitter: uma abordagem utilizando aprendizagem de máquina. Disponível em: https://ud10.arapiraca.ufal.br/repositorio/publicacoes/3082 Acesso em: 4 abril 2022.

Oliveira, Leandro Massetti, Costa, Aline Mayara S, Fernandes, Vandecia Rejane M. (2019) Inteligência Artificial Aplicada a Detecção de Fake. Disponível em: https://monografias.ufma.br/jspui/handle/123456789/4251 Acesso em: 4 abril 2022.

R, Silipo (2019) From a Single Decision Tree to a Random Forestem. Disponível em: https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147 Acesso em: 5 dezembro 2022.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Trans. Intell. Syst. Technol. Disponível em: https://arxiv.org/pdf/1901.06437.pdf Acesso em: 2 abril 2022.

Sharma, K., Seo, S., Meng, C., Rambhatla, S., Dua, A., & Liu, Y. (2020). Coronavirus on social media: Analyzing misinformation in Twitter conversations. Disponível em: https://arxiv.org/pdf/2003.12309.pdf Acesso em: 2 abril 2022.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter. Diponível em:https://www.researchgate.net/publication/318981549_Fake_News_Detection_on_Social_Media_A_Data_Mining_Perspective Acesso em: 2 abril 2022.

Uma Sharma, Sidarth Saran, Shankar M. Patil. (2021). Fake News Detection using Machine Learning Algorithms. Disponível em: https://www.ijert.org/research/fake-news-detection-using-machine-learning-algorithms-IJERTCONV9IS03104.pdf Acesso em: 2 nov 2022.

Zhou, X., & Zafarani, R. (2020). Fake news: A survey of research, detection methods, and opportunities. ACM Computing Surveys, 53(5), 1–40. Disponível em: https://arxiv.org/pdf/1812.00315.pdf Acesso em: 2 nov 2022.

Z Khanam, B N Alwasel, H Sirafi and M Rashid (2021). Fake News Detection Using Learning Approaches. Disponível em:https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012040/meta. Acesso em: 2 nov 2022.