

Avaliação de Modelos Generativos de Dados e Redes Neurais Artificiais para Classificação de *Fake News* na Web

Daniela Deboni Silva de Mello¹, Gabriela Barbosa Oliveira¹,
João Gabriel Rocha Silva¹

¹Instituto Federal de Educação, Ciência e Tecnologia de Brasília (IFB)
Área Especial 2, s/n - São Bartolomeu (São Sebastião), Brasília - DF

`gabriela59016@estudante.ifb.edu.br`, `joao.gabriel@ifb.edu.br`

Abstract. *This work investigates the use of synthetic data generated by generative modeling techniques combined with artificial neural networks to improve fake news detection. Considering the limitations of real datasets, the proposal aims to expand and diversify the training base. Although the accuracies presented are moderate, this is common in fake news detection problems due to the complexity of the topic. The results indicate that the use of synthetic data contributes to enhancing the classifier's performance, making the approach promising for the automatic identification of false news and for strengthening informational reliability.*

Resumo. *Este trabalho investiga o uso de dados sintéticos gerados por técnicas de modelagem generativa combinadas com redes neurais artificiais para melhorar a detecção de fake news. Considerando as limitações dos conjuntos de dados reais, a proposta visa ampliar e diversificar a base de treinamento. Embora as acurácias apresentadas sejam moderadas, isso é comum em problemas de detecção de fake news, dada a complexidade do tema. Os resultados indicam que a utilização de dados sintéticos contribui para o aprimoramento do desempenho do classificador, tornando a abordagem promissora para a identificação automática de notícias falsas e para o fortalecimento da confiabilidade informacional.*

1. Introdução

As notícias falsas (*fake news*) tornaram-se um dos principais desafios da era digital, sobretudo com a expansão das redes sociais e a popularização dos *smartphones* [Kito 2019]. O fácil acesso à Internet e o intenso compartilhamento de informações em plataformas como Facebook, WhatsApp e X (antigo Twitter) modificaram significativamente a forma como notícias são produzidas, distribuídas e consumidas [Prado 2022]. Nesse contexto, cresce o interesse por soluções tecnológicas baseadas em aprendizado supervisionado para identificar, classificar e combater conteúdos falsos, contribuindo para a construção de um ambiente informacional mais confiável [Nogueira and outros 2024].

Um dos principais desafios no desenvolvimento de sistemas eficazes para identificar *fake news* está na limitação dos dados utilizados no treinamento dos modelos de inteligência artificial. A falta de informações adequadas — tanto em volume quanto em qualidade — é apontada como uma das causas mais frequentes do insucesso de projetos

na área, prejudicando a capacidade dos algoritmos de gerar resultados consistentes e realistas [Nogueira and outros 2024]. No caso específico da desinformação, essa limitação é ainda mais evidente: grande parte dos conjuntos de dados disponíveis não contempla a diversidade de formatos, temas e contextos em que as notícias falsas circulam. Além disso, muitos desses conjuntos são compostos por dados antigos ou restritos a determinadas regiões, dificultando a generalização dos modelos para outros cenários e realidades [Horne and Adali 2017].

Uma alternativa que vem ganhando destaque para superar essas barreiras é o uso de dados sintéticos como complemento aos dados reais [De Andrade Júnior and outros 2024]. Esses dados são gerados artificialmente para reproduzir padrões e estruturas presentes em conteúdos autênticos, ampliando a base de treinamento sem depender exclusivamente de fontes reais [Patel 2024]. Ao integrar dados sintéticos aos conjuntos tradicionais, é possível reduzir vieses, aumentar a diversidade dos exemplos e melhorar o desempenho dos classificadores, tornando-os mais aptos a lidar com os desafios impostos pela complexidade e pelo dinamismo da desinformação online [De Andrade Júnior and outros 2024].

Assim, o objetivo deste trabalho é avaliar o uso de dados sintéticos em Redes Neurais Artificiais (RNAs) na detecção de notícias falsas. Para isso, empregam-se técnicas de modelagem generativa a partir de um conjunto de dados reais coletados da internet [Chaves et al. 2023], contendo características próprias de notícias reais e falsas. A partir desses dados, são criadas amostras artificiais que preservam propriedades estatísticas do conjunto original. O banco de dados resultante, composto por exemplos reais e sintéticos, é utilizado para treinar classificadores de *machine learning*. O desempenho obtido é analisado por meio de métricas de classificação, de modo a verificar em que medida a complementação com dados artificiais pode aumentar a precisão dos modelos no reconhecimento de *fake news*.

2. Aspectos Teóricos

2.1. Modelos Generativos de Dados

A geração de dados sintéticos é o processo de criação artificial de novos conjuntos de dados que preservam as propriedades estatísticas dos dados reais, com o objetivo de ampliar e diversificar amostras disponíveis para aplicações em aprendizado de máquina [Lu et al. 2023]. Essa abordagem tem se mostrado eficiente quando os conjuntos reais são escassos, desequilibrados ou apresentam limitações de privacidade e representatividade [Watanuki et al. 2024].

Entre as metodologias mais aplicadas na geração de dados sintéticos, destacam-se as Redes Geradoras Adversariais (*Generative Adversarial Networks – GANs*), introduzidas por Goodfellow et al. (2014), que empregam duas redes neurais treinadas de forma competitiva para capturar com precisão a distribuição de dados reais [Goodfellow et al. 2014]; os *Autoencoders* Variacionais (*Variational Autoencoders – VAEs*), descritos por Kingma e Welling (2014), que modelam um espaço probabilístico para gerar amostras novas e coerentes com o conjunto original [Kingma and Welling 2014]; os Modelos de Difusão com Redução de Ruído (*Denoising Diffusion Probabilistic Models – DDPMs*), popularizados por Ho et al. (2020), que sintetizam dados por meio de processos estocásticos reversos aplicados

a amostras inicialmente ruidosas [Ho et al. 2020]; e métodos estatísticos de sobre-amostragem, como o Método de Sobreamostragem de Minoria Sintética (*Synthetic Minority Over-sampling Technique – SMOTE*), proposto por Chawla et al. (2002), que gera novas instâncias da classe minoritária a partir de interpolação entre vizinhos mais próximos [Chawla et al. 2002]. Tais abordagens são treinadas com bases de dados reais e têm como finalidade gerar amostras que reflitam com fidelidade as distribuições estatísticas, correlações e estruturas presentes no conjunto original, possibilitando a ampliação e diversificação eficaz das bases de dados para tarefas supervisionadas de aprendizado de máquina.

2.2. Redes Adversárias Generativas

As GANs, propostas por Goodfellow et al. (2014), baseiam-se em uma arquitetura de aprendizado competitivo composta por dois módulos principais: o gerador (*generator*) e o discriminador (*discriminator*). O gerador é responsável por criar amostras de dados sintéticos, enquanto o discriminador tem a função de diferenciar dados reais daqueles produzidos artificialmente. Durante o treinamento, ambos os modelos são ajustados de forma iterativa e simultânea: o gerador busca aprimorar continuamente a qualidade das amostras para “enganar” o discriminador, ao passo que este se torna cada vez mais preciso na identificação entre conteúdos autênticos e artificiais [Leão and Figueiredo 2022].

Essa configuração adversarial leva ambos os modelos a evoluírem em conjunto, elevando progressivamente o realismo dos dados sintéticos produzidos. O processo de otimização é formalizado como um jogo de soma zero, no qual o gerador busca maximizar a taxa de erro do discriminador, enquanto este minimiza seu erro de classificação [Goodfellow et al. 2016]. Essa estratégia resulta, idealmente, em um ponto de equilíbrio onde os dados gerados se tornam indistinguíveis dos dados reais, possibilitando aplicações como a síntese de imagens realistas, aprimoramento de resolução visual e simulação de dados para treinamento supervisionado [Farnia and Ozdaglar 2020].

2.3. Sobreamostragem Minoritária Sintética

A técnica de Sobreamostragem Minoritária Sintética (SMOTE) visa resolver o problema de *datasets* desbalanceados por meio da criação de novas amostras da classe minoritária. O procedimento clássico envolve a seleção, para cada instância minoritária, de k vizinhos mais próximos no espaço de características; novos pontos são gerados por interpolação linear entre a instância original e seus vizinhos. Essa abordagem previne o sobreajuste comum na replicação simples de dados e promove regiões de decisão mais genéricas e inclusivas [Chawla et al. 2002].

2.4. Autoencoders Variacionais

Autoencoders Variacionais constituem uma classe de modelos generativos baseados em redes neurais profundas, projetados para aprender representações de dados por meio de uma abordagem probabilística. Diferentemente dos *autoencoders* tradicionais, que apenas comprimem e descomprimem dados, os VAEs incorporam inferência variacional para modelar a distribuição subjacente dos dados. Isso permite que o modelo gere novas amostras realistas ao explorar o espaço implícito contínuo aprendido. O processo envolve dois componentes principais: o codificador (*encoder*), que aproxima a distribuição posterior dos dados, e o decodificador (*decoder*), que reconstrói os dados

originais a partir dessa representação subjacente. A função de perda dos VAEs é composta por dois termos: o erro de reconstrução e a divergência de Kullback-Leibler (KL) entre a distribuição aproximada e a distribuição latente a priori, tipicamente uma normal padrão. Essa formulação garante que o espaço seja regularizado e contínuo, favorecendo a geração de dados sintéticos úteis para tarefas como síntese de imagens, aprendizado semi-supervisionado e compressão de dados [Kingma and Welling 2014]. Recentemente, aplicações de VAEs têm sido exploradas em diferentes domínios, como medicina, processamento de linguagem natural e sistemas de recomendação, destacando sua flexibilidade e poder de generalização [Mozó et al. 2022].

2.5. Modelos Probabilísticos de Difusão com Redução de Ruído

Os Modelos Probabilísticos de Difusão com Redução de Ruído representam uma classe emergente de modelos generativos baseados em processos de difusão estocástica reversível. Inspirados em sistemas físicos, esses modelos geram dados amostrando de uma distribuição simples (como uma gaussiana isotrópica) e aprendendo a revertê-la progressivamente através de uma cadeia de passos de ruídos [Weng 2021]. Cada etapa busca eliminar o ruído adicionado em fases anteriores, aproximando a amostra de uma distribuição de dados reais.

A eficácia dos DDPMs tem sido demonstrada na geração de imagens de alta fidelidade e outras tarefas generativas complexas. O modelo foi inicialmente formalizado por [Ho et al. 2020], que demonstraram que tais arquiteturas podem alcançar resultados comparáveis e em alguns casos superiores a outras abordagens como GANs e VAEs.

3. Materiais e Métodos

3.1. Ambiente e Ferramentas

O desenvolvimento dos códigos foi realizado na plataforma *Google Colab*, utilizando a linguagem de programação *Python*. As bibliotecas importadas para a implementação foram: *tensorflow*, *pandas*, *numpy*, *csv*, *random* e *statistics*.

3.2. Geração de Dados Sintéticos

Para a geração dos dados sintéticos as bases geradas foram derivadas a partir de uma base real composta por 100 notícias disponibilizada por [Chaves et al. 2023]. Essa base original contém informações quantitativas e discretas referentes a notícias, descritas pelas seguintes características (*features*): (1) existência de autor, (2) título da notícia em caixa alta, (3) nota do PageRank do site, (4) posição do site no ranking do Google que veicula a notícia, (5) quantidade de notícias similares, (6) média do PageRank das notícias similares, (7) média das posições dos sites que vinculam as notícias similares.

Para ampliar e diversificar a base de treinamento, foram geradas diferentes conjuntos de dados sintéticos utilizando quatro técnicas distintas de síntese de dados: *Variational Autoencoder* (VAE), *Generative Adversarial Networks* (GAN), *Synthetic Minority Over-sampling Technique* (SMOTE) e *Denoising Diffusion Probabilistic Models* (DDPM).

Cada técnica foi aplicada individualmente para produzir uma base sintética contendo 100 amostras, todas com 50 notícias consideradas falsas e outras 50 consideradas

verdadeiras, preservando as propriedades estatísticas e características relevantes dos dados originais. Dessa forma, foram criadas quatro bases distintas, cada uma representando o padrão gerado por uma das técnicas supracitadas.

Além disso, foi construído um conjunto híbrido que integra os dados gerados por todas as técnicas, contendo 25 unidades aleatórias de cada técnica, totalizando também 100 amostras sintéticas. Essa abordagem visa explorar a complementaridade entre os diferentes métodos de geração, buscando enriquecer a diversidade dos exemplos para o treinamento da rede neural [Smith et al. 2023].

O objetivo central deste trabalho é comparar e avaliar o desempenho dessas técnicas de geração sintética na tarefa de classificação de notícias falsas, investigando como cada método impacta a qualidade do treinamento e a eficácia dos modelos preditivos desenvolvidos. Esses conjuntos sintéticos foram utilizados para treinar os modelos de classificação, permitindo analisar o efeito da complementação com dados artificiais na robustez e precisão dos algoritmos.

Cada uma das bases sintéticas geradas mantém a mesma estrutura de atributos da base real, garantindo a preservação das propriedades estatísticas essenciais para a tarefa de classificação. Ademais, a base original utilizada como referência é composta por 100 notícias, divididas igualmente entre 50 falsas e 50 verdadeiras, mesma proporção mantida nas bases sintéticas produzidas.

3.3. Treinamento e Classificação com RNA

Para a tarefa de classificação de notícias, foi utilizada uma Rede Neural do tipo *Perceptron Multicamadas* (MLP), adequada para problemas supervisionados com dados tabulares não sequenciais [Przybyła-Kasperek and Marfo 2024]. A RNA foi configurada com uma camada oculta contendo 32 neurônios, treinamento realizado ao longo de 150 épocas conforme recomendações para conjuntos tabulares [Goodfellow et al. 2016], e funções de ativação tanto na camada oculta quanto na camada de saída variáveis entre *tanh*, *sigmoid* e *relu* combinadas para a par.

O modelo foi compilado utilizando o otimizador *Adam* e a função de perda utilizada quando *sigmoid* é a função de ativação na camada de saída foi *binary_crossentropy*, quando *tanh* e *relu* a função de perda foi Mean Squared Error (MSE). As principais métricas de avaliação utilizadas foram acurácia, precisão, *recall* positivo e *recall* negativo [Chollet 2018].

3.4. Execuções e Variedade de Parâmetros

Para avaliar o desempenho da RNA, foram criados 45 cenários experimentais, combinando 5 diferentes bases de dados sintéticas geradas (VAE, GAN, SMOTE, DDPM e Híbrida) e 9 cenários de combinação de funções de ativação na camada oculta e na camada de saída (*relu*, *sigmoid* e *tanh* em todas as combinações possíveis de pares).

Os dados foram organizados em listas separadas conforme as classes de notícias falsas e verdadeiras. Essas listas foram embaralhadas aleatoriamente para evitar vieses no treinamento da RNA. Em seguida, os dados foram divididos em conjuntos de treino e teste, utilizando uma proporção de 70% para treino e 30% para teste, buscando equilibrar a necessidade de dados suficientes para o aprendizado do modelo e para sua avaliação

[Gholamy et al. 2018]. Para cada configuração a RNA foi executada 15 vezes devido a natureza estocástica do algoritmo.

3.5. Métricas de Avaliação

Visando alcançar o objetivo deste trabalho, que é a avaliação de modelos generativos de dados e redes neurais para o problema de classificação de *fake news*, foram utilizadas, para cada 15 execuções de cada cenário, as seguintes métricas: acurácia média, acurácia máxima e acurácia mínima, além do desvio-padrão das execuções. Além disso, foi calculado para cada cenário a precisão e o *recall* positivo e negativo. A Tabela 1 apresenta as fórmulas utilizadas, onde VP e VN representam, respectivamente, os verdadeiros positivos e verdadeiros negativos; FP e FN correspondem aos falsos positivos e falsos negativos; e TOTAL é a soma de todos esses valores.

Precisão Positiva	<i>Recall</i> Negativo	<i>Recall</i> Positivo	Acurácia
$VP/(VP + FP)$	$VN/(VN + FP)$	$VP/(VP + FN)$	$(VN + VP)/TOTAL$

Table 1. Fórmulas das métricas de avaliação.

O *recall* representa a capacidade do algoritmo de identificar corretamente casos verdadeiros. Para o *recall* positivo, são considerados os verdadeiros positivos e falsos negativos, ou seja, notícias verdadeiras. Para o *recall* negativo, são levados em conta os verdadeiros negativos e falsos positivos, referentes às notícias falsas [Powers 2011].

A precisão indica a capacidade do algoritmo em classificar corretamente os casos identificados. A precisão positiva avalia a proporção de notícias classificadas como verdadeiras que estão corretas, enquanto a precisão negativa avalia a proporção de notícias classificadas como falsas que estão corretas [Manning et al. 2008].

3.6. Critério de Decisão para Classificação

A contagem dos verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos foi realizada com base em uma regra de decisão definida a partir da probabilidade estimada pela RNA. Caso essa probabilidade fosse maior ou igual a 0,6, assim como utilizado por [Almeida et al. 2021], classifica-se a notícia como verdadeira, o que resultava em: incremento dos verdadeiros positivos, se a notícia fosse realmente verdadeira; incremento dos falsos positivos, se a notícia fosse falsa. Caso a probabilidade fosse menor que 0,6, classificava-se a notícia como falsa, o que resultava em: incremento dos verdadeiros negativos, se a notícia fosse realmente falsa; incremento dos falsos negativos, se a notícia fosse verdadeira.

4. Resultados e Discussão

A Tabela 2 apresenta as maiores acurácias médias obtidas para diferentes bases de dados sintéticas e combinações de funções de ativação na RNA, juntamente com as métricas de precisão positiva, *recall* positivo (Recall P), *recall* negativo (Recall N), função de ativação na camada oculta (Atv. 1) e função de ativação na camada de saída (Atv. 2).

Observa-se que a maior acurácia média foi de 0,7846, obtida quando a RNA foi treinada com dados sintéticos gerados pelo SMOTE, utilizando as funções de ativação ReLU e sigmoid. Nesse cenário, também se verificou a maior precisão (0,8653) e um bom

Ger. Sint.	Acurácia	Precisão	Recall P	Recall N	Atv. 1	Atv. 2
SMOTE	0,7846	0,8653	0,6933	0,8761	relu	sigmoid
SMOTE	0,7730	0,7595	0,8184	0,7292	tanh	sigmoid
VAE	0,7167	0,6847	0,8000	0,6318	tanh	sigmoid
VAE	0,7026	0,6943	0,7336	0,6710	relu	sigmoid
Híbrido	0,6714	0,6968	0,4584	0,8844	tanh	relu

Table 2. Cinco maiores acurácias médias para diferentes geradores sintéticos e combinações de ativação

equilíbrio entre os valores de *recall* positivo (0,6933) e negativo (0,8761). Além disso, destaca-se que o uso do SMOTE com a combinação tanh e sigmoid também apresentou desempenho consistente, alcançando a maior taxa de *recall* positivo (0,8184), indicando maior capacidade da RNA em identificar corretamente as classes positivas.

As bases geradas pelo VAE também apresentaram resultados competitivos. A combinação de funções de ativação *tanh/sigmoid* obteve um *recall* positivo de 0,8000, ainda que com acurácia média de 0,7167, inferior às configurações com o SMOTE. Já com a combinação *relu/sigmoid*, o VAE alcançou desempenho equilibrado, com valores próximos entre *recall* positivo (0,7336) e negativo (0,6710). Vale destacar, ao analisar as execuções de maiores acurácias, a presença recorrente da ativação sigmoid na camada de saída, geralmente acompanhada por funções tanh ou relu na camada oculta, o que sugere que essas combinações foram mais eficazes no treinamento da RNA.

Embora as acurácias observadas na Tabela 2 sejam moderadas, variando entre aproximadamente 67% e 78%, esses resultados refletem a complexidade inerente ao problema de detecção de *fake news*. A tarefa de classificar notícias falsas envolve desafios linguísticos e contextuais que dificultam a obtenção de taxas de acurácia muito elevadas, sobretudo quando há restrições quanto à diversidade e ao volume dos dados disponíveis [Shu et al. 2017].

A predominância da função de ativação *sigmoid* na camada de saída, presente em quatro das cinco melhores configurações, pode ser explicada por sua adequação ao problema de classificação binária, já que suas saídas no intervalo [0,1] podem ser interpretadas como probabilidades, facilitando a tomada de decisão [Chollet 2018]. Nas camadas ocultas, observa-se o uso recorrente tanto da *tanh* quanto da *relu*. A função *tanh* contribui por produzir saídas centradas em zero, favorecendo a convergência e a modelagem de relações não lineares complexas [Goodfellow et al. 2016, LeCun et al. 2015], enquanto a função *relu* oferece simplicidade computacional e ajuda a mitigar o problema do gradiente desaparecendo, promovendo um treinamento mais estável [Nair and Hinton 2010]. Dessa forma, a combinação dessas funções de ativação mostrou-se eficaz para que a rede neural aprendesse representações discriminativas relevantes, justificando os melhores desempenhos observados nos experimentos.

Os valores de *Recall P* variaram entre 0,4584 e 0,8184, sendo este último obtido com a configuração SMOTE associada às ativações *tanh/sigmoid*. Isso indica que, em alguns cenários, a RNA alcançou boa sensibilidade para identificar corretamente as notícias falsas. Por outro lado, os valores de *Recall N* oscilaram entre 0,629 e 0,8844, com destaque para o modelo híbrido, que apresentou a maior taxa de acerto na detecção de

notícias verdadeiras. Essa assimetria nos resultados evidencia um *trade-off* típico da tarefa de detecção de *fake news*: melhorar a identificação de notícias falsas pode levar a uma maior taxa de falsos positivos, enquanto priorizar a identificação de notícias verdadeiras pode reduzir a sensibilidade na detecção de conteúdos falsos. Dependendo do contexto, pode ser mais relevante priorizar o alto *recall* para as classes de interesse, mesmo que isso implique em classificações incorretas em menor proporção.

Visando analisar a estabilidade das variáveis e métricas avaliadas, a Tabela 3 apresenta os menores valores de desvio-padrão obtidos após 15 execuções da rede neural para cada um dos 45 cenários. Além dos valores das métricas, são destacados os respectivos geradores sintéticos e funções de ativação associados a esses resultados. Cada linha indica a configuração em que o desvio-padrão da métrica considerada foi o menor entre todas as combinações testadas. Assim, por exemplo, a linha referente à “Acurácia” corresponde ao cenário em que o desvio-padrão da acurácia apresentou menor variabilidade, apontando para uma execução mais consistente do modelo nessa métrica.

Métrica	Desvio Padrão	Gerador sintético	Atv. 1	Atv. 2
Acurácia	0.0086	Híbrido	tanh	sigmoid
Precisão	0.0079	Híbrido	tanh	sigmoid
<i>Recall</i> P	0.0172	GAN	relu	sigmoid
<i>Recall</i> N	0.0176	DDPM	tanh	sigmoid

Table 3. Geradores sintéticos, ativações 1 e 2 das métricas com os menores desvios-padrão

Observa-se que, para as métricas de acurácia e precisão, os menores desvios-padrão foram obtidos utilizando a base sintética gerada pelo método Híbrido, combinada com as funções de ativação *tanh* na camada oculta e *sigmoid* na camada de saída. Esse resultado indica maior estabilidade e previsibilidade nessas métricas, sugerindo que o modelo, sob essas condições, apresenta desempenho consistente.

Por outro lado, para o *recall* positivo, o menor desvio-padrão foi registrado com dados gerados pelo GAN, em conjunto com a combinação de ativações *relu/sigmoid*, enquanto o *recall* negativo apresentou menor variabilidade ao utilizar o DDPM com a configuração *tanh/sigmoid*. Essa diferença evidencia que a estabilidade varia de acordo com a métrica considerada, refletindo a sensibilidade diferenciada do modelo em reconhecer corretamente notícias falsas (*recall* positivo) ou verdadeiras (*recall* negativo).

5. Considerações Finais

Este estudo demonstrou a eficácia das técnicas de geração de dados sintéticos para o treinamento de classificadores na detecção de *fake news*, especialmente em cenários com dados reais limitados ou desbalanceados. A utilização de múltiplos métodos, VAE, GAN, DDPM e SMOTE, permitiu capturar diferentes aspectos das distribuições dos dados, enriquecendo a variabilidade e contribuindo para uma melhor generalização do modelo.

A análise dos resultados evidenciou que, embora abordagens isoladas como SMOTE e VAE possam apresentar desempenho competitivo, a combinação das diversas técnicas se mostrou superior em termos de desempenho médio e robustez. Além disso, a escolha das funções de ativação revelou-se fundamental para o ajuste fino do classificador,

indicando que a interação entre a arquitetura da rede e o tipo de dado sintético impacta diretamente a eficácia do modelo.

A estabilidade estatística observada em determinadas configurações, refletida pelos menores desvios-padrão das métricas, reforça a importância de considerar não apenas a acurácia máxima, mas também a consistência do desempenho em aplicações práticas. Esses resultados apontam para a necessidade de experimentação empírica e adaptação dos hiperparâmetros conforme o contexto de aplicação.

Como direções futuras, sugere-se aprofundar a investigação sobre a combinação dinâmica de métodos de geração, bem como o desenvolvimento de estratégias automáticas para seleção e integração de geradores durante o treinamento. Adicionalmente, recomenda-se explorar a explicabilidade dos modelos baseados em dados sintéticos para aumentar a transparência e a confiança em sistemas de combate à desinformação.

References

- Almeida, A. L. D., Carrara, G. C., Prates, I. B., Nascimento, L. C., Souza, P. H., Almeida, T. R., Cani, R. C., and Silva, J. G. R. (2021). Modelo matemático apoiado por um algoritmo genético para classificação de fake news na web. In *Encontro Nacional de Computação dos Institutos Federais (ENCompIF)*, pages 17–20. SBC.
- Chaves, V. M., Silva, M. L. P., Brum, F. O., and Silva, J. G. R. (2023). Estudo comparativo entre um algoritmo de evolução diferencial e um algoritmo genético para classificação de fake news na web. In *Encontro Nacional de Computação dos Institutos Federais (ENCompIF)*, pages 17–20. SBC.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- Chollet, F. (2018). *Deep Learning with Python*. Manning Publications.
- De Andrade Júnior, W. T. and outros (2024). Avaliando o desempenho de modelos generativos de dados para classificação de notícias falsas. In *Encontro Nacional de Computação dos Institutos Federais (ENCompIF)*, pages 42–49. SBC.
- Farnia, F. and Ozdaglar, A. (2020). Do gans always have nash equilibria? pages 3029–3039.
- Gholamy, A., Kreinovich, V., and Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. Technical Report UTEP-CS-18-09, University of Texas at El Paso.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, pages 2672–2680.
- Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *arXiv preprint arXiv:2006.11239*.

- Horne, B. and Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 759–766.
- Kingma, D. P. and Welling, M. (2014). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Kito, M. H. I. (2019). Uma abordagem de ciência de dados para identificar fake news no âmbito político.
- Leão, E. T. and Figueiredo, R. C. d. (2022). Um estudo comparativo sobre redes adversárias generativas.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Fu, T., and Wei, W. (2023). Machine learning for synthetic data generation: A review. *arXiv preprint arXiv:2302.04062*.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). Introduction to information retrieval. *Cambridge University Press*.
- Mozó, P. O. et al. (2022). A practical introduction to generative adversarial networks and variational autoencoders with python examples. In *International Conference on Artificial Intelligence and Soft Computing*, pages 244–253. Springer.
- Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.
- Nogueira, A. G. D. and outros (2024). Geração de dados sintéticos tabulares para detecção de malware android: um estudo de caso. In *Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg)*, pages 808–814. SBC.
- Patel, P. (2024). Synthetic data. *Business Information Review*, 41(2):48–52.
- Powers, D. M. W. (2011). *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation*, volume 2.
- Prado, M. (2022). *Fake news e inteligência artificial: o poder dos algoritmos na guerra da desinformação*. Edições 70.
- Przybyła-Kasperek, M. and Marfo, S. (2024). A multi-layer perceptron neural network for varied conditional attributes in tabular dispersed data. *PLOS ONE*, 19(4):e0281234.
- Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36.
- Smith, J., Silva, M., and Oliveira, C. (2023). Hybrid synthetic data generation for enhanced machine learning. *Journal of Artificial Intelligence Research*, 78:123–145.
- Watanuki, S., Edo, K., and Miura, T. (2024). Applying deep generative neural networks to data augmentation for consumer survey data with a small sample size. *Applied Sciences*, 14(19):9030.
- Weng, L. (2021). What are diffusion models? Accessed: 2025-07-20.