

# AVALIANDO O DESEMPENHO DE MODELOS GENERATIVOS DE DADOS PARA CLASSIFICAÇÃO DE NOTÍCIAS FALSAS MEDIANTE MODELO MATEMÁTICO AUXILIADO POR ALGORITMO GENÉTICO

EVALUATING THE PERFORMANCE OF GENERATIVE DATA MODELS FOR FAKE NEWS CLASSIFICATION USING A MATHEMATICAL MODEL AIDED BY A GENETIC ALGORITHM

**William Teles de Andrade Júnior**

wtaj@discente.ifpe.edu.br

**Rodrigo Cesar Lira da Silva (coorientador)**

rodrigo.lira@paulista.ifpe.edu.br

**Antônio Correia de Sá Barreto Neto (orientador)**

antonio.neto@paulista.ifpe.edu.br

---

## RESUMO

Este artigo apresenta uma análise comparativa do desempenho de um sistema de classificação de notícias falsas utilizando um modelo matemático aprimorado por um algoritmo genético. O objetivo deste estudo é investigar o potencial dos modelos generativos de dados sintéticos para esta abordagem de detecção de notícias falsas. A pesquisa compara os resultados obtidos de um conjunto de dados real, contendo informações das notícias, com aqueles obtidos de quatro conjuntos de dados sintéticos gerados usando redes adversárias generativas, autoencoders variacionais, modelo probabilístico de difusão de redução de ruído e técnica de sobre-amostragem minoritária sintética. Os resultados do estudo indicam que o desempenho da classificação obteve uma melhora quando usado os dados artificiais, com uma pontuação de acurácia de, aproximadamente, 87%. Esses resultados sugerem que dados sintéticos, podem servir como ferramentas valiosas para melhorar o desempenho classificação de notícias falsas.

Palavras-chave: Modelos Generativos. Fake News. Algoritmo Genético. Classificação.

## ABSTRACT

This paper presents a comparative analysis of the performance of a fake news classification system using a mathematical model improved by a genetic algorithm. The objective of this study is to investigate the potential of models to generate synthetic

data for this fake news detection approach. The research compares the results obtained from a real dataset, containing news information, with those obtained from four synthetic datasets generated using GAN, VAE, DDPM and SMOTE. The results of the study indicate that classification performance improved when using artificial data, with an accuracy score of approximately 87%. These results suggest that synthetic data can serve as valuable tools for improving fake news classification performance.

Keywords: Generative Models. Fake News. Genetic Algorithm. Classification.

---

## 1 INTRODUÇÃO

As notícias falsas se tornaram um grande problema na era digital. Com o surgimento das mídias sociais e outras plataformas *online*, torna-se mais fácil que informações falsas se espalhem rapidamente, causando danos a indivíduos, comunidades e até nações (Vosoughi; Roy; Aral, 2018). Para combater esse problema, existe um interesse crescente em desenvolver métodos e ferramentas para detectar e classificar notícias falsas, a fim de promover fontes de informação mais precisas e confiáveis (Zhou; Zafarani, 2020).

Os métodos tradicionais de classificação baseiam-se frequentemente em conjuntos de dados reais, que podem ser limitados em tamanho e podem não capturar toda a gama de artigos de notícias falsas. Conforme apontado por Horne et al. ((2017)), a falta de conjuntos de dados confiáveis e abrangentes de notícias falsas é um grande obstáculo no desenvolvimento de algoritmos eficazes para detectar notícias falsas. Além disso, os conjuntos de dados de notícias falsas existentes sofrem geralmente de vieses e limitações, como serem limitados a um período ou região específica. Para resolver este problema, os pesquisadores exploram o uso de conjuntos de dados sintéticos gerados por diferentes modelos como uma ferramenta para melhorar o tamanho do conjunto de dados real e o desempenho da classificação (Salimans et al., 2016). No entanto, mesmo esses conjuntos de dados sintéticos podem não capturar totalmente a complexidade e a variabilidade das notícias falsas do mundo real. Portanto, há uma necessidade premente de conjuntos de dados mais abrangentes e confiáveis de notícias falsas, que possam fornecer uma base para o desenvolvimento de métodos mais precisos e eficazes para detectar notícias falsas.

Conjuntos de dados sintéticos são conjuntos de dados criados artificialmente que imitam as características de dados reais, eles podem ser gerados usando técnicas como Redes Adversárias Generativas (GANs) (Frid-Adar et al., 2018), modelos generativos profundos (Suroso; Cherntanomwong; Sooraksa, 2023) e modelos probabilísticos de difusão de redução de ruído (Carrillo-Perez et al., 2023). Ao gerar dados sintéticos que se assemelham a dados reais, esses métodos podem ajudar a superar algumas das limitações das abordagens tradicionais de Aprendizado de Máquina.

Neste estudo, é explorado o uso de modelagem generativa de dados para criação de um conjunto de dados para detecção de notícias falsas, usando dados sintéticos criados a partir de um conjunto de dados real com informações de características das notícias. O objetivo é avaliar a capacidade dos modelos gerar dados sintéticos para

esta abordagem de detecção de notícias falsas. Para isso, são aplicadas técnicas de modelagem generativa para criar um conjunto de dados sintéticos que imita as propriedades estatísticas do conjunto de dados reais. Em seguida, o modelo de classificação é treinado no conjunto de dados originais e sintéticos e seus desempenhos são comparados com métricas de classificação.

Acredita-se que essa pesquisa tem implicações importantes para combater a disseminação de notícias falsas e desinformação online. Ao usar a modelagem de dados generativa, é esperado superar alguns dos desafios associados às abordagens tradicionais de Aprendizado de Máquina, como a necessidade de abundantes dados rotulados. Isso pode tornar mais fácil para pesquisadores e profissionais desenvolverem métodos mais precisos e eficazes para detectar e classificar notícias falsas, ajudando a promover fontes de informação mais precisas e confiáveis na *Web*.

Este trabalho está dividido da seguinte maneira: na seção 2 é apresentada a fundamentação teórica do trabalho, onde é brevemente discutido a área de detecção de fake news e geração sintética de dados. Na seção 3 são descritos alguns trabalhos que já abordaram temas de classificação e geração de dados sintéticos. Na seção 4 é explicado como foi desenvolvida a metodologia utilizada para atingir os objetivos. Na seção 5 são mostrados os resultados obtidos. Por fim, a seção 6 trata sobre a conclusão e trabalhos futuros.

## **2 FUNDAMENTAÇÃO TEÓRICA**

Nesta seção de fundamentação teórica é apresentada uma visão abrangente da área de pesquisa. Na subseção 2.1 são apresentados conceitos e metodologias sobre classificação e detecção de notícias falsas usadas atualmente. Na subseção 2.2 são apresentados conceitos e técnicas utilizadas para geração de dados sintéticos.

### **2.1 Métodos de classificação e detecção de notícias falsas**

Há diversos métodos de classificação e detecção de notícias falsas. Nesta seção, serão apresentados quatro tipos de métodos já existentes que visam o mesmo resultado, muitos dos trabalhos que abordam a detecção de notícias falsas utilizam-se de técnicas de extração do conteúdo da notícia, conforme mostrado por Parkih et al. ((2018)) e técnicas de mineração de dados de redes sociais como demonstrado por Shu et al. ((2017)).

#### **2.1.1 Métodos baseados em conteúdo**

Métodos baseados em conteúdo dependem da análise do texto do próprio artigo de notícias para determinar se é provável que seja falso ou não. Esses métodos geralmente envolvem técnicas de processamento de linguagem natural (NLP), como análise de sentimento, modelagem de tópicos e análise de estilo de linguagem (Ruchansky; Seo; Liu, 2017).

Embora os métodos baseados em conteúdo tenham se mostrado promissores na detecção de notícias falsas, eles têm suas limitações. Por exemplo, artigos de notícias falsas podem ser projetados para imitar o estilo e a linguagem de artigos de notícias reais, dificultando distingui-los usando apenas esse método. Além disso, os métodos baseados em conteúdo podem ser menos eficazes na detecção de formas visuais de notícias falsas, como imagens ou vídeos adulterados (Sharma, K. et al., 2019).

### **2.1.2 Análise de redes sociais**

Análise de redes sociais é um método que analisa como as informações se espalham por meio de uma rede de conexões sociais. No contexto da detecção de notícias falsas, a análise de redes sociais pode ser usada para identificar padrões na forma como notícias falsas são compartilhadas e propagadas em plataformas de mídia social.

Esse método pode incluir a análise do comportamento do usuário, como a frequência e o tempo das postagens, o número de curtidas e compartilhamentos e os tipos de conteúdo compartilhado (Shu, Kai et al., 2017; Wang; Terano, 2015). No entanto, a análise de redes sociais também tem limitações. Por exemplo, pode ser difícil distinguir entre conexões sociais genuínas e aquelas criadas artificialmente para espalhar notícias falsas. Além disso, a análise de redes sociais pode ser menos eficaz na detecção de notícias falsas que se espalham por meio de plataformas de mensagens privadas ou outros canais não públicos (Kaliyar; Singh, 2019).

### **2.1.3 Métodos baseados em modelagem preditiva**

Uma abordagem mais parecida com a utilizada nesta pesquisa é a modelagem preditiva. Esses modelos aprendem com dados de treinamento rotulados, extraindo características e padrões distintos associados a notícias falsas. Se caracterizam pela utilização de técnicas de Aprendizado de Máquina para treinar o algoritmo e melhorar o modelo (Pérez-Rosas et al., 2017; Rubin; Conroy; Chen, 2015).

É importante que os dados de treinamento sejam de boa qualidade. Se o conjunto de dados de treinamento apresentar rótulos tendenciosos ou imprecisos, o modelo poderá aprender a replicar esses preconceitos e, conseqüentemente, generalizar para dados novos e invisíveis, isso acaba sendo uma limitação para esse método.

### **2.1.4 Métodos híbridos**

Métodos híbridos combinam várias abordagens para melhorar a precisão da detecção de notícias falsas. Por exemplo, uma abordagem pode combinar análise baseada em conteúdo com análise de rede social, aproveitando os pontos fortes de cada método para superar suas respectivas limitações. Outra abordagem pode combinar análise baseada em fontes com técnicas de Aprendizado de Máquina para desenvolver uma compreensão mais abrangente do contexto e da veracidade dos artigos de notícias.

Métodos híbridos oferecem potencial para detecção de notícias falsas mais robusta e confiável, mas também podem exigir algoritmos mais complexos e conjuntos de da-

dos maiores para treinamento (Conroy; Rubin; Chen, 2015). Existem vários artigos de pesquisa que exploram a eficácia de métodos híbridos na detecção de notícias falsas. O estudo de (Nasir; Khan; Varlamis, 2021) propõe um modelo híbrido que combina uma rede neural convolucional (CNN) para análise de conteúdo e uma rede neural recorrente (RNN) para análise de fonte para melhorar a precisão da detecção de notícias falsas.

## 2.2 Geração de dados sintéticos

A geração de dados sintéticos é um processo de criação de novos conjuntos de dados que imitam as propriedades estatísticas dos dados do mundo real. Uma das abordagens mais comuns para geração de dados sintéticos é por meio do uso de técnicas de modelagem generativa, como redes adversárias generativas (GANs) (Goodfellow et al., 2014), autoencoders variacionais (VAEs) (Kingma; Welling, 2022), modelos probabilísticos de difusão de redução de ruído (DDPMs) (Nichol; Dhariwal, 2021) e técnica de sobre-amostragem minoritária sintética (SMOTE) (Mukherjee; Khushi, 2021). Esses diferentes modelos são treinados em dados reais e aprendem a gerar novas amostras de dados estatisticamente semelhantes ao conjunto de dados original. GANs foram usados para gerar imagens realistas (Karras; Laine; Aila, 2019) e VAEs foram usados para gerar fala realista (Sun et al., 2020).

Modelos generativos mostram um potencial significativo no campo de Aprendizado de Máquina para diversas aplicações. Uma das áreas importantes de aplicação dos modelos generativos é a resolução de problemas tabulares (Cullen et al., 2022). Trabalhos recentes desenvolveram inúmeros modelos, incluindo VAEs tabulares (Xu et al., 2019), abordagens baseadas em GANs (Rajabi; Garibay, 2022) e também em DDPMs (Kotelnikov et al., 2022).

Embora a geração de dados sintéticos tenha muitas vantagens, como reduzir a necessidade de abundantes dados rotulados, ela também apresenta algumas limitações. Uma limitação é que os dados sintéticos podem não capturar totalmente a complexidade e a variabilidade dos dados reais, levando a vieses e imprecisões em modelos de Aprendizado de Máquina treinados com eles (Paszke et al., 2019). Outra limitação é que os dados sintéticos podem não ser representativos da população em estudo, levando a uma fraca capacidade de generalização de modelos de Aprendizado de Máquina treinados com esses mesmos dados (Lu; Wang; Wei, 2023). Uma métrica frequentemente usada para medir a representatividade de conjuntos de dados sintéticos é a Divergência de Kullback-Leibler (KL) (Seghouane; Amari, 2007). Ela mede a diferença entre duas distribuições de probabilidade.

### 2.2.1 Redes Adversárias Generativas

Redes Adversárias Generativas (GANs) são um tipo de arquitetura de rede neural que consiste em dois componentes: um gerador e um discriminador (Goodfellow et al., 2014). O gerador é responsável por gerar os dados sintéticos, enquanto o discriminador é treinado para distinguir entre os dados sintéticos e os dados reais. Os dois componentes são treinados simultaneamente competitivamente, onde o gerador

visa produzir dados que enganem o discriminador, e o discriminador visa classificar corretamente os dados (Poudevigne-Durance; Jones; Qin, 2022).

O gerador em um GAN recebe ruído aleatório como entrada e gera dados sintéticos que se assemelham aos dados reais. O gerador é normalmente uma rede neural profunda que aprende a mapear o ruído aleatório para a distribuição de dados reais. Por outro lado, o discriminador também é uma rede neural profunda que aprende a classificar se uma determinada amostra de dados é real ou sintética. O discriminador é treinado usando uma função binária de perda de entropia cruzada, onde tenta minimizar a perda classificando corretamente os dados (MozÓ et al., 2022).

Durante o processo de treinamento, o gerador e o discriminador são atualizados iterativamente. O gerador gera dados sintéticos, sendo então alimentados no discriminador com dados reais. O discriminador dá *feedback* ao gerador classificando os dados como reais ou sintéticos. Esse *feedback* é utilizado pelo gerador para atualizar seus parâmetros e melhorar a qualidade dos dados sintéticos. Este processo continua até que o gerador consiga gerar dados sintéticos indistinguíveis dos dados reais (Poudevigne-Durance; Jones; Qin, 2022).

## 2.2.2 Autoencoders Variacionais

Autoencoders Variacionais (VAEs) são um tipo de modelo generativo que ganharam atenção significativa em vários campos, incluindo finanças, ciência da computação e processamento de imagens (Kenneweg; Stallmann; Hammer, 2022). VAEs são um tipo de autoencoder que aprendem uma representação compactada dos dados de entrada, conhecida como espaço latente, e podem gerar novas amostras de dados que se assemelham aos dados de treinamento (Xiao; Ganguli; Pandey, 2020).

A arquitetura básica de um VAE consiste em uma rede neural codificadora, uma rede neural decodificadora e uma função de perda (Kingma; Welling, 2022). A rede do codificador mapeia os dados de entrada para um espaço latente de dimensão inferior, enquanto a rede do decodificador reconstrói os dados originais do espaço latente. A função de perda é projetada para medir a similaridade entre os dados reconstruídos e os dados originais, incentivando o VAE a aprender uma representação significativa dos dados (Elbattah et al., 2021).

Uma das principais características dos VAEs é a introdução de um componente probabilístico no espaço latente. Em vez de aprender um mapeamento determinístico dos dados de entrada para o espaço latente, os VAEs modelam o espaço latente como uma distribuição de probabilidade. Isso permite a geração de novas amostras de dados por amostragem da distribuição do espaço latente (Singh; Ogunfunmi, 2021).

## 2.2.3 Modelos Probabilísticos de Difusão de Redução de Ruído

Modelos Probabilísticos de Difusão de Redução de Ruído (DDPMs) são um tipo de modelo generativo que aprende a modelar a transição de Markov de uma distribuição simples para a distribuição de dados, eles conseguem gerar diversas amostras por meio de transições estocásticas sequenciais (Ho; Jain; Abbeel, 2020; Choi et al., 2021). Os DDPMs consideram o procedimento de geração como o inverso de um processo

de difusão, adicionando gradualmente ruído às amostras de dados e transformando a distribuição dos dados em uma distribuição gaussiana (Lyu et al., 2022). Isso permite que os DDPMs modelem distribuições de dados complexas (Lam et al., 2022).

Os DDPMs são aplicados a diversas aplicações em diferentes campos. Na ciência da computação, os DDPMs são usados para remoção de ruído e síntese de imagens. No campo da inteligência artificial, DDPMs são utilizados para detecção de anomalias (Wyatt et al., 2022). No campo da medicina, os DDPMs utilizam-se para geração de sinais sintéticos de ECG (Adib et al., 2023). Os DDPMs também foram aplicados em física para modelar fluxo granular e transporte de calor (Zhou; Xu et al., 2022).

Uma vantagem dos DDPMs é que eles podem ser treinados usando um objetivo simples de remoção de ruído, o que os faz ajustar o ruído aos dados. Ao contrário das Redes Adversárias Generativas (GANs), que requerem ajuste cuidadoso de hiperparâmetros, os DDPMs podem usar estruturas de modelo semelhantes e ser treinados com um objetivo simples de remoção de ruído (Liu; Ren et al., 2022). Entretanto, uma limitação dos DDPMs é sua lenta velocidade de inferência, o que os impede de serem usados em aplicações de tempo real (Liu; Su; Yu, 2022).

Para criar dados sintéticos usando DDPMs, os modelos são treinados em dados incorporados e depois amostrados para gerar dados 2D incorporados. Os dados incorporados são então desincorporados e transformados de volta ao formato de dados original usando transformações inversas (Adib et al., 2023). Este processo permite que os DDPMs gerem dados sintéticos que seguem a distribuição dos dados originais.

#### **2.2.4 Técnica de Sobre-amostragem Minoritária Sintética**

A Técnica de Sobre-amostragem Minoritária Sintética (SMOTE) é um algoritmo de sobre-amostragem comumente usada para resolver o problema de conjuntos de dados desequilibrados em Aprendizado de Máquina (Chawla et al., 2002). Conjuntos de dados desequilibrados ocorrem quando o número de instâncias em uma classe é significativamente menor do que o número de instâncias em outra classe, levando a modelos de aprendizagem tendenciosos que favorecem a classe majoritária. O algoritmo SMOTE visa aliviar este problema gerando pontos de dados sintéticos para a classe minoritária, equilibrando assim o conjunto de dados e melhorando o desempenho dos classificadores.

O SMOTE funciona criando pontos de dados artificiais com base nas semelhanças de espaço de recursos entre exemplos minoritários existentes (He; Garcia, 2009). Ele seleciona exemplos que estão próximos no espaço de características, desenha uma linha entre eles e então cria uma nova amostra em um ponto ao longo dessa linha (Chang; Park; Moon, 2021). Esse processo é repetido por um determinado número de vezes ou até que o equilíbrio desejado entre as classes minoritária e majoritária seja alcançado.

Para gerar pontos de dados sintéticos, o SMOTE aplica interpolação linear entre um ponto de classe minoritária e um de seus  $K$  vizinhos mais próximos (Elreedy; Atiya; Kamalov, 2023). O algoritmo calcula a diferença entre os vetores de características do ponto da classe minoritária e seu vizinho mais próximo, multiplica essa diferença por um número aleatório entre 0 e 1 e adiciona o resultado ao vetor de características do

ponto da classe minoritária. Este processo é repetido para cada dimensão de recurso, resultando em um novo ponto de dados sintético.

### 3 TRABALHOS RELACIONADOS

A desinformação é difícil de detectar porque foi propositalmente escrita para deturpar informações. Nesse sentido, existem numerosos estudos de pesquisa sobre a detecção de notícias falsas, utilizando diversos métodos de classificação.

Ferreira et al. ((2020)) em seu trabalho propuseram um modelo matemático capaz de identificar e classificar as notícias como verdadeiras ou não, utilizando-se de metadados obtidos da notícia e do site. O modelo tem como resultado uma pontuação, que indica as chances de uma notícia ser falsa ou verdadeira. Foram realizados testes com 67 notícias, sendo 36 verdadeiras e 31 falsas. Todos os atributos da base foram avaliados com o mesmo coeficiente (i.e. mesmo peso), 0,142 para cada um deles. A partir daí, obtiveram à porcentagem de 81% de acerto para as classificações realizadas automaticamente pela ferramenta.

Posteriormente, Almeida et al. ((2021)) propuseram uma melhoria para o modelo matemático com o auxílio de um algoritmo genético para encontrar o melhor coeficiente (peso) para cada atributo da base. Em seu trabalho utilizaram um novo conjunto de dados, com 100 notícias da *Web*, onde 50 dessas foram utilizadas para o treinamento e as outras 50 para validação dos resultados. Comparando a classificação dos modelos, constatou-se que as modificações apresentadas pela proposta a tornaram mais eficazes que o modelo referência, aumentando a taxa de acertos na classificação para 78% comparados com 70% alcançados pelo modelo original.

Uma das considerações da pesquisa de Almeida et al. ((2021)) para trabalho futuro envolvia uma coleta de um número maior de notícias. O que motivou este trabalho, validar modelos generativos de dados sintéticos para avaliar a capacidade desses modelos para esta abordagem de classificação.

A pesquisa sobre o uso de redes neurais adversárias generativas (GANs) para detectar notícias falsas ganhou atenção nos últimos anos. Ali et al. ((2021)) destacou a ascensão das mídias sociais como fator que contribui para a proliferação de notícias falsas. Enfatizaram a necessidade de avaliar a robustez dos detectores de informações falsas. Ali et al. ((2021)) também aproveitaram redes neurais profundas (DNNs) para classificação de texto para enfrentar esse desafio de classificar falsas notícias. No entanto, o desempenho das DNNs depende fortemente de grandes conjuntos de dados de treinamento. Algoritmos de Aprendizado de Máquina, como Bayes Ingênuo (Naive Bayes), Máquina de Vetores de Suporte (SVM) e Floresta Aleatória (Random Forest), são usados para detecção de notícias falsas, conforme mencionado por Shahzad et al. ((2022)). As técnicas de inteligência artificial, processamento de linguagem natural e Aprendizado de Máquina mostraram-se eficazes na identificação de falsas informações online.

O uso de Redes Neurais Adversárias Generativas (GANs), também é explorado para detectar *deep fakes* e imagens/vídeos falsos, conforme discutido por St et al. ((2022)) e Sharma et al. ((2022)). Os GANs têm se mostrado promissores na distinção entre conteúdo real e falso. Técnicas de geração de dados sintéticos são empregadas



para aumentar dados rotulados limitados para treinar modelos de detecção de notícias inverídicas. Amostras sintéticas que imitam as características de informações fictícias podem melhorar o desempenho e a robustez dos modelos de detecção. Os padrões de propagação e a análise de rede de notícias falsas são estudados para compreender a disseminação da desinformação, conforme destacado por Liu e Wu ((2018)). A análise da rede de propagação pode fornecer sinais sociais refinados para este trabalho de classificação de notícias.

O uso de arquiteturas derivadas de representações codificadoras bidirecionais de transformadores (BERT), é explorado para enfrentar os desafios da desinformação, conforme discutido por Kula et al. ((2021)). O BERT é utilizado em tarefas de detecção de notícias falsas para melhorar o resultado. Pesquisadores também investigaram a detecção precoce de informação falsa nas redes sociais, conforme mencionado por Shu et al. ((2020)). A detecção precoce é crucial em ambientes práticos e estudos têm mostrado resultados promissores nesta área. Em resumo, a investigação sobre a utilização de redes adversárias generativas e de dados sintéticos, juntamente com outras técnicas de aprendizagem automática, contribui para o desenvolvimento de métodos eficazes de detecção de notícias. Essas abordagens aproveitam o Aprendizado Profundo, a Análise de Redes e a Análise de Padrões de Propagação para melhorar a precisão e a detecção precoce de notícias inverídicas.

#### 4 METODOLOGIA

Os dados sintéticos criados nesta pesquisa foram gerados a partir de um conjunto de dados reais composto por 7 critérios e 100 notícias no total, sendo 50 falsas e 50 verdadeiras obtidos pelo trabalho de Ferreira et al. ((2020)) que em seguida foi contribuído também por Almeida et al. ((2021)) quando melhorou o desempenho da classificação com o auxílio de um algoritmo genético. Os critérios deste conjunto de dados verdadeiros escolhidos para serem extraídos e analisados pelo modelo proposto por Ferreira et al. ((2020)) está descrito na tabela 1.

Tabela 1 – Critérios do conjunto de dados reais.

Critérios	Descrição	Tipo da variável
C1	Autor	Lógico
C2	Caixa alta	Lógico
C3	Nota do site	Real
C4	Posição no ranking	Real
C5	Notícias similares	Real
C6	Média das notas de sites com notícias similares	Real
C7	Media do ranking dos sites com notícias similares	Real

Para gerar os dados sintéticos foram utilizados esses modelos generativos: Redes adversárias generativas (GAN); Autoencoder Variacional (VAE); modelos probabilísticos de difusão de redução de ruído (DDPM) e técnica de sobre-amostragem minoritária sintética (SMOTE), cada modelo gerou 100 instâncias a partir do conjunto de dados

original. Essas técnicas têm seus pontos fortes e limitações e, usando-as, é possível obter dados sintéticos diversos.

Nesta pesquisa os experimentos foram feitos utilizando a plataforma do Google Colab com Python 3 e CPU como acelerador de hardware e também em um laptop com CPU AMD Ryzen™ 7 2700U a 2,20 GHz, com GPU AMD Radeon™ 540 com 2 GB de memória e memória RAM de 12 GB no sistema operacional Windows 10. A linguagem de programação utilizada na pesquisa foi *Python*, com auxílio das bibliotecas: *tensorflow*, *keras*, *pytorch*, *imbalanced-learn*, *scikit-learn*, *geneticalgorithm*, *pandas* e *numpy*.

A metodologia utilizada nesta pesquisa pode ser dividida em três etapas, conforme demonstrado na figura 1. Na etapa 1 é realizada a geração dos dados sintéticos, onde o conjunto de dados real é usado em um método de geração de dados e esse método cria dois conjuntos de dados separados, um com notícias da classe falsa e a outra a classe verdadeira, esses conjuntos são agrupados para formar o conjunto de dados sintético final, a subseção 4.1 apresenta mais detalhes sobre essa etapa. Na etapa 2 é iniciado o processo de classificação, onde os dados sintéticos gerados são utilizados no algoritmo genético, para obter os coeficientes de cada critério para ser utilizado no modelo matemático, para o mesmo classificar o conjunto de dados real, este processo é abordado na subseção 4.2. E por fim, na etapa 3 são calculadas as métricas de avaliação da classificação, descritas na subseção 4.3.

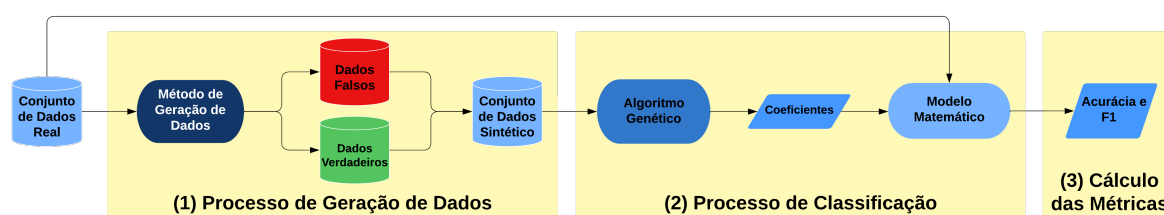


Figura 1 – Fluxograma dos Processos da Metodologia.

#### 4.1 Técnicas de geração de dados sintéticos

Para o modelo GAN, a rede do gerador consiste em várias camadas densamente conectadas com funções de ativação Leaky ReLU, normalização de lote e uma camada final com ativação sigmoide para gerar amostras de dados sintéticos. A entrada para o gerador foi um vetor de ruído aleatório de dimensionalidade 50. A rede discriminadora foi construída com camadas densamente conectadas, utilizando ativações Leaky ReLU, e concluída com um único neurônio com função de ativação sigmoide para classificação binária (falso ou verdadeiro).

O modelo GAN foi treinado de forma adversária com otimizador Adam com taxa de aprendizado de 0,002 e o treinamento foi realizado por um total de 1000 épocas. Em cada iteração de treinamento, um lote de amostras de dados reais e um lote de dados sintéticos de tamanho igual gerado pelo gerador atual foram combinados, o discriminador foi treinado nesse lote combinado com rótulos apropriados. A perda foi calculada

usando entropia cruzada binária, o gerador foi então treinado para minimizar a capacidade do discriminador de distinguir entre dados reais e sintéticos. Esse processo foi repetido duas vezes, uma vez somente com os dados de notícias falsas e outra somente com os dados de notícias verdadeiras, gerando assim dois conjuntos de dados sintéticos, agregados, e posteriormente processados para garantir características dos dados reais, como, por exemplo, colunas binárias. O GAN foi ajustado por meio de experimentação manual com taxas de aprendizado, tamanhos de lote e dimensão de ruído aleatório fornecida como entrada para o gerador para alcançar resultados ideais.

Para o modelo VAE, dado o tamanho limitado do conjunto de dados, a dimensionalidade do espaço latente foi definida como 3, um valor pequeno, para esta implementação para evitar sobre ajuste. O modelo VAE foi treinado por um total de 1000 épocas, usando uma função de perda personalizada que combina perda de reconstrução (entropia cruzada binária) e um termo relacionado à divergência de *Kullback-Leibler* (Seghouane; Amari, 2007), essa função de perda encorajou o espaço latente aprendido a se aproximar de uma distribuição gaussiana unitária. O modelo também foi treinado usando retro propagação e gradiente descendente estocástico, durante o treinamento, o modelo teve como objetivo minimizar a função de perda definida. Hiper-parâmetros como taxa de aprendizagem de 0,001, tamanho de lote de 64 e a perda de reconstrução de 0,5 foram ajustados para otimizar o desempenho. O processo para o modelo VAE também foi repetido duas vezes, da mesma forma que o GAN.

Para o modelo DDPM o codificador consistiu em uma série de camadas totalmente conectadas com ativações de unidades lineares retificadas (ReLU), que transformam os dados de entrada em uma representação de menor dimensão. Já o decodificador foi composto por camadas totalmente conectadas com ativações ReLU, responsáveis por reconstruir os dados do espaço latente. Toda a arquitetura foi projetada para otimizar a capacidade de remoção de ruído e gerar dados sintéticos de alta qualidade. O DDPM durante seu processo de treinamento envolveu a conversão do conjunto de dados reais para uma matriz multidimensional contendo elementos de um único tipo de dados, e também envolveu a minimização da perda do erro quadrático médio (MSE) entre os dados reconstruídos e a entrada original. O otimizador Adam foi utilizado com taxa de aprendizado de 0,001 e o treinamento foi realizado por um total de 1000 épocas. Durante cada época, o modelo foi treinado em mini lotes, de tamanho 8, para melhorar a eficiência computacional. Os gradientes foram zerados antes de cada passagem para trás para evitar acúmulo, e a perda foi calculada e retro propagada para atualizar os pesos do modelo. Igualmente ao GAN e VAE, o processo foi realizado duas vezes com as classes do conjunto de dados separados e posteriormente processados para garantir características definidas.

Para a técnica de geração de dados SMOTE, o processo foi realizado iterativamente desbalanceando as classes do conjunto de dados reais até que o nível desejado de sobre-amostragem fosse alcançado. As amostras sintéticas então foram usadas para criar um novo conjunto de dados com a distribuição das classes equilibradas. A implementação do SMOTE teve seu hiper-parâmetro de quantidade de vizinhos definida como 5 sendo implementada utilizando a biblioteca *imbalanced-learn*, esta biblioteca fornece uma interface eficiente e amigável para aplicação de diversas técnicas de re-amostragem, incluindo SMOTE.

## 4.2 Modelo de classificação dos dados

Para fazer a classificação de uma notícia como verdadeira ou falsa e verificar o desempenho dos dados artificiais criados, foi utilizado o método de classificação proposto por (Ferreira et al., 2020), no qual, a avaliação de uma notícia é realizada por meio da multiplicação do peso de cada critério ( $P_i$ ) pelo valor deste critério estrutural na notícia ( $C_i$ ). Quando o valor da avaliação ( $A_i$ ) é maior que 0,6 (60%) a notícia é classificada pelo modelo como verdadeira e caso contrário, falsa.

Para encontrar o peso de cada critério, para o modelo matemático, foi utilizado o algoritmo genético (GA) proposto por (Almeida et al., 2021). No processo de treinamento para os dados reais foi necessário dividir o conjunto de dados reais pela metade, sendo 50 notícias para treino e 50 notícias de teste, ambos com 25 notícias verdadeiras e 25 notícias falsas. Já o processo de treinamento para os dados sintéticos foi feito com os mesmos e testado no conjunto de dados reais completo, as 100 notícias, um estudo realizado por (Tobin et al., 2017) indicou que realizar o treinamento em dados sintéticos resultaram em melhor desempenho de detecção no conjunto de dados real em comparação ao treinamento com apenas alguns exemplos rotulados do conjunto de dados real.

O algoritmo GA desenvolvido aplicado aos dados sintéticos e o modelo matemático foram executados 30 vezes, pela lei estocástica do algoritmo, visando uma comparação justa entre os conjuntos de dados diferentes. Os parâmetros de operadores genéticos do GA utilizados foram 100 gerações, 100 indivíduos, 5% para a mutação e 50% para o cruzamento.

## 4.3 Métricas de avaliação

Em problemas de classificação, é essencial avaliar o desempenho dos modelos de classificação utilizando métricas adequadas. Duas métricas de avaliação comumente utilizadas para classificação são acurácia e pontuação F1. A acurácia mede a proporção de instâncias classificadas corretamente em relação ao número total de instâncias, fornecendo uma avaliação geral do desempenho do modelo (Gani et al., 2022). Por outro lado, a pontuação F1 considera tanto a precisão quanto a revocação, fornecendo uma medida equilibrada que considera tanto os falsos positivos quanto os falsos negativos (Tan; Lu; Jiang, 2021).

A acurácia definida pela Equação 1 é uma métrica amplamente utilizada em tarefas de classificação. No entanto, a acurácia por si só pode não ser suficiente para avaliar o desempenho de um modelo de classificação, especialmente em cenários onde o conjunto de dados está desequilibrado ou o custo da classificação incorreta varia entre diferentes classes (Fatourehchi et al., 2008). Nesses casos, métricas adicionais como a pontuação F1 são usadas para fornecer uma avaliação mais abrangente.

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Na Equação 1,  $TP$  representa a quantidade de verdadeiros positivos,  $TN$  representa a quantidade de verdadeiros negativos,  $FP$  são os falsos positivos e  $FN$  são os falsos negativos.

A pontuação F1 definida pela Equação 2 é uma média harmônica de precisão e revocação. A pontuação F1 combina essas duas métricas para fornecer um valor único que equilibra precisão e revocação. É particularmente útil quando o conjunto de dados está desequilibrado ou quando falsos positivos e falsos negativos têm consequências diferentes (Tan; Lu; Jiang, 2021).

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (2)$$

## 5 RESULTADOS

Os resultados da avaliação de desempenho do modelo de classificação de notícias falsas no conjunto de dados real utilizando os coeficientes encontrados pelo GA treinado separadamente com os dados reais e com os quatro diferentes conjuntos de dados sintéticos juntamente com o custo de tempo para gerar cada conjunto de dados são apresentados na tabela 2.

Tabela 2 – Resultados da avaliação de desempenho da classificação.

Conjunto de Dados	Acurácia <i>Média</i> ( $\sigma$ )	F1 <i>Média</i> ( $\sigma$ )	Tempo Criação Dados <i>Média</i> ( $\sigma$ )
Reais	0,8020 (0,0059)	0,8347 (0,0042)	Não Aplicável
GAN	0,8736 (0,0048)	0,8878 (0,0038)	150s (4,16s)
VAE	0,8806 (0,0057)	0,8934 (0,0046)	33,8s (6,28s)
DDPM	0,8543 (0,0513)	0,8746 (0,0396)	1002s (128s)
SMOTE	0,8786 (0,0033)	0,8918 (0,0026)	2528e-6s (552e-6s)

Os resultados do modelo de classificação no conjunto de dados real mostraram uma precisão de 80%. A pontuação F1 para o conjunto de dados real também foi calculada e considerada consistente com a medida de precisão. Em contraste, os conjuntos de dados sintéticos produziram resultados de desempenho mais elevados para o modelo de classificação. A precisão média nos quatro conjuntos de dados sintéticos foi de aproximadamente 87%. Isso indica que o modelo alcançou um maior nível de precisão na classificação de artigos de notícias falsas nos conjuntos de dados sintéticos em comparação com o conjunto de dados real. A pontuação F1 para os conjuntos de dados sintéticos também foram calculados e considerados consistentemente superiores aos obtidos para o conjunto de dados real. É importante destacar que o modelo utilizando apenas o conjunto de dados real foi utilizado e avaliado apenas para efeito de informação (*baseline*), pois o seu cenário de treinamento e teste difere dos modelos utilizando os dados sintéticos.

Os resultados deste estudo sugerem que o desempenho do modelo de classificação de notícias falsas foi melhor nos conjuntos de dados sintéticos em comparação

com o conjunto de dados real, isso pode ser devido à base de treinamento dos dados sintéticos possuir mais instâncias. Esta descoberta é consistente com pesquisas anteriores que mostraram a eficácia dos conjuntos de dados sintéticos na melhoria do desempenho dos modelos de detecção de notícias falsas (Shu, Kai et al., 2017; Wang, 2017; Horne; Adali, 2017).

Uma possível explicação para esta observação é que os conjuntos de dados sintéticos foram gerados especificamente para imitar as características dos dados reais. Esses conjuntos de dados podem conter exemplos mais diversos e representativos de notícias falsas e verdadeiras, permitindo que o modelo de classificação desempenhasse de forma mais eficaz e alcançasse maior precisão. Além disso, os conjuntos de dados sintéticos podem ter sido menos tendenciosos ou ruidosos em comparação com o conjunto de dados real, levando a um melhor desempenho.

Apesar dos resultados alcançados, acredita-se que serão necessárias mais pesquisas para investigar as razões por trás dessa diferença de desempenho e para explorar outros fatores potenciais que podem influenciar o desempenho dos modelos de classificação de notícias falsas, como, por exemplo, os hiper-parâmetros dos modelos.

Outro ponto que pode-se notar é em relação ao custo de tempo para cada método, o SMOTE, por ser o mais simples em questão de complexidade, se mostrou ter um tempo muito menor em relação aos outros, enquanto o DDPM, por ser mais complexo e devido a sua lenta velocidade de inferência, foi o que mais demorou para ter seu processo finalizado. É importante mencionar que mesmo o SMOTE sendo o mais rápido, ele possui uma complexidade maior de ser utilizado quando a base de dados não está desbalanceada, tornando-se necessário o desbalanceamento iterativo do conjunto de dados para realizar a sobre-amostragem na quantidade desejada.

O teste Wilcoxon (*Wilcoxon signed-rank test*) (Wilcoxon, 1945) é um teste de hipótese estatística não paramétrico usado para determinar se há uma diferença significativa entre dois grupos de dados relacionados. Usando uma taxa de confiança de 99%, foi aplicado o teste de Wilcoxon para comparar os resultados da métrica F1 entre os algoritmos de geração de dados sintéticos. Conforme mostrado na tabela 3. O símbolo '-' indica não haver diferença estatística entre as soluções, o símbolo '▲' indica que a abordagem obteve resultados melhores que o outro algoritmo e o símbolo '▽' representa que a proposta alcançou resultados piores que o método comparado.

Tabela 3 – Comparação estatística pelo teste Wilcoxon com taxa de confiança de 99%.

Conjunto de Dados	GAN	VAE	DDPM	SMOTE
GAN	-	-	▲	▽
VAE	-	-	▲	▽
DDPM	▽	▽	-	▽
SMOTE	▲	▲	▲	-

Observa-se que os resultados alcançados pelo SMOTE superaram, em geral, o GAN, VAE e DDPM. SMOTE pode não ser o mais eficiente algoritmo para todos os cenários, mas por seus resultados da pontuação F1 durante a classificação terem um desvio padrão menor que os outros, conseguiu ter um resultado estatístico melhor que as

outras abordagens. Isso pode ser devido ao fato que os dados SMOTE devem ter tido uma distribuição mais parecida com as dos dados reais, pela forma que o SMOTE cria os dados sintéticos, indo em busca de vizinhos no espaço e também devido à baixa complexidade e quantidade de dados reais usados. Outra possibilidade se dá pelo fato de que os outros métodos utilizados poderiam ser melhor parametrizados, visto que em comparação ao SMOTE, os outros possuem um conjunto maior de hiper-parâmetros e configurações que não foram investigados profundamente.

## 6 CONCLUSÕES

Este estudo comparou o desempenho de um modelo de classificação de notícias falsas na Web em um conjunto de dados real e também com o auxílio de quatro diferentes conjuntos de dados sintéticos, gerados com técnicas diferentes. As ferramentas utilizadas para geração desses dados foi a linguagem de programação *Python* e as bibliotecas *tensorflow*, *keras*, *pytorch* e *imbalanced-learn*, *scikit-learn*, *geneticalgorithm*, *pandas* e *numpy*.

Os resultados indicaram que a classificação foi superior com os conjuntos de dados sintéticos. Esta descoberta sugere que os conjuntos de dados sintéticos podem ser eficazes na melhoria do desempenho dos modelos de detecção de notícias falsas. No entanto, são necessárias mais pesquisas para compreender os fatores subjacentes que contribuem para esta diferença de desempenho e para explorar outras abordagens potenciais para melhorar a precisão dos modelos de classificação de notícias falsas.

A técnica SMOTE demonstrou uma vantagem em relação aos outros métodos ao ser avaliada pelo teste estatístico Wilcoxon na pontuação F1. Esse resultado pode ser atribuído ao fato de o algoritmo SMOTE ter conseguido criar um conjunto de dados com características mais próximas às do conjunto de dados reais. Dessa forma, ao encontrar os coeficientes com o algoritmo genético e aplicá-los ao modelo matemático, o algoritmo genético variou menos que os outros nos coeficientes, resultando numa pontuação F1 com um desvio padrão menor em comparação com os testes dos outros métodos.

Nos trabalhos futuros, espera-se analisar a parametrização dos modelos já usados como também outros métodos de criação de dados tabulares sintéticos desenvolvidos, bem como classificar os dados com diferentes técnicas e algoritmos de Aprendizagem de Máquinas. Outra análise é a de validar se existe um viés na geração dos dados sintéticos com a mesma base de validação. Dessa forma, aperfeiçoando o conjunto de dados para melhor representar as notícias falsas e estabelecer se algoritmos com o objetivo específico de classificação de dados conseguem melhores resultados.

## REFERÊNCIAS

ADIB, E. et al. Synthetic ecg signal generation using probabilistic diffusion models. **arXiv preprint arXiv:2303.02475**, 2023. DOI: 10.48550/arxiv.2303.02475.

- ALI, Hassan et al. All Your Fake Detector are Belong to Us: Evaluating Adversarial Robustness of Fake-News Detectors Under Black-Box Settings. **IEEE Access**, v. 9, p. 81678–81692, 2021. DOI: 10.1109/ACCESS.2021.3085875.
- ALMEIDA, Ana Luisa et al. Modelo Matemático apoiado por um Algoritmo Genético para classificação de Fake News na Web. In: ANAIS do VIII Encontro Nacional de Computação dos Institutos Federais. Evento Online: SBC, 2021. P. 17–20. DOI: 10.5753/encompif.2021.15945.
- CARRILLO-PEREZ, Francisco et al. RNA-to-image multi-cancer synthesis using cascaded diffusion models. **bioRxiv**, Cold Spring Harbor Laboratory, p. 2023–01, 2023.
- CHANG, Y.; PARK, H.; MOON, I. Predicting the cochlear dead regions using a machine learning-based approach with oversampling techniques. **Medicina**, v. 57, p. 1192, 11 2021. DOI: 10.3390/medicina57111192.
- CHAWLA, Nitesh V et al. SMOTE: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321–357, 2002.
- CHOI, Jooyoung et al. Ilvr: Conditioning method for denoising diffusion probabilistic models. **arXiv preprint arXiv:2108.02938**, 2021.
- CONROY, Nadia K; RUBIN, Victoria L; CHEN, Yimin. Automatic deception detection: Methods for finding fake news. **Proceedings of the association for information science and technology**, Wiley Online Library, v. 52, n. 1, p. 1–4, 2015.
- CULLEN, Drake et al. Evaluation of Synthetic Data Generation Techniques in the Domain of Anonymous Traffic Classification. **IEEE Access**, v. 10, p. 129612–129625, 2022. DOI: 10.1109/ACCESS.2022.3228507.
- ELBATTAH, M. et al. Variational autoencoder for image-based augmentation of eye-tracking data. **Journal of Imaging**, v. 7, p. 83, 5 2021. DOI: 10.3390/jimaging7050083.
- ELREEDY, D.; ATIYA, A. F.; KAMALOV, F. A theoretical distribution analysis of synthetic minority oversampling technique (smote) for imbalanced learning. **Machine Learning**, 2023. DOI: 10.1007/s10994-022-06296-4.
- FATOURECHI, M. et al. Comparison of evaluation metrics in classification applications with imbalanced datasets. **2008 Seventh International Conference on Machine Learning and Applications**, 2008. DOI: 10.1109/icmla.2008.34.
- FERREIRA, Antony L N et al. Um modelo matemático para classificação de fake news na web. In: ANAIS do Simpósio Brasileiro de Pesquisa Operacional. [S.l.: s.n.], 2020.
- FRID-ADAR, M. et al. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. **Neurocomputing**, v. 321, p. 321–331, 2018. DOI: 10.1016/j.neucom.2018.09.013.
- GANI, M. O. et al. Etfpos-idf: a novel term weighting scheme for examination question classification based on bloom’s taxonomy. **IEEE Access**, v. 10, p. 132777–132785, 2022. DOI: 10.1109/access.2022.3230592.
- GOODFELLOW, Ian J. et al. **Generative Adversarial Networks**. [S.l.: s.n.], 2014. arXiv: 1406.2661 [stat.ML].



- HE, H.; GARCIA, E. Learning from imbalanced data. **IEEE Transactions on Knowledge and Data Engineering**, v. 21, p. 1263–1284, 9 2009. DOI: 10.1109/tkde.2008.239.
- HO, Jonathan; JAIN, Ajay; ABBEEL, Pieter. **Denosing Diffusion Probabilistic Models**. [S.l.: s.n.], 2020. arXiv: 2006.11239 [cs.LG].
- HORNE, Benjamin D.; ADALI, Sibel. **This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News**. [S.l.: s.n.], 2017. arXiv: 1703.09398 [cs.SI].
- KALIYAR, Rohit Kumar; SINGH, Navya. Misinformation detection on online social media—a survey. In: IEEE. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). [S.l.: s.n.], 2019. P. 1–6.
- KARRAS, Tero; LAINE, Samuli; AILA, Timo. **A Style-Based Generator Architecture for Generative Adversarial Networks**. [S.l.: s.n.], 2019. arXiv: 1812.04948 [cs.NE].
- KENNEWEG, P.; STALLMANN, D.; HAMMER, B. Novel transfer learning schemes based on siamese networks and synthetic data. **Neural Computing and Applications**, v. 35, p. 8423–8436, 11 2022. DOI: 10.1007/s00521-022-08115-2.
- KINGMA, Diederik P; WELLING, Max. **Auto-Encoding Variational Bayes**. [S.l.: s.n.], 2022. arXiv: 1312.6114 [stat.ML].
- KOTELNIKOV, Akim et al. **TabDDPM: Modelling Tabular Data with Diffusion Models**. [S.l.: s.n.], 2022. arXiv: 2209.15421 [cs.LG].
- KULA, S.; KOZIK, R.; CHORAŚ, M. Implementation of the bert-derived architectures to tackle disinformation challenges. **Neural Computing and Applications**, v. 34, p. 20449–20461, 23 2021. DOI: 10.1007/s00521-021-06276-0.
- LAM, M. F. et al. Bddm: bilateral denosing diffusion models for fast and high-quality speech synthesis. **arXiv preprint arXiv:2203.13508**, 2022. DOI: 10.48550/arxiv.2203.13508.
- LIU, L.; REN, Y. et al. Pseudo numerical methods for diffusion models on manifolds. **arXiv preprint arXiv:2202.09778**, 2022. DOI: 10.48550/arxiv.2202.09778.
- LIU, S.; SU, D.; YU, D. Diffgan-tts: high-fidelity and efficient text-to-speech with denosing diffusion gans. **arXiv preprint arXiv:2201.11972**, 2022. DOI: 10.48550/arxiv.2201.11972.
- LIU, Y.; WU, Y. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. **Proceedings of the Aai Conference on Artificial Intelligence**, v. 32, 1 2018. DOI: 10.1609/aaai.v32i1.11268.
- LU, Yingzhou; WANG, Huazheng; WEI, Wenqi. Machine Learning for Synthetic Data Generation: a Review. **arXiv preprint arXiv:2302.04062**, 2023.
- LYU, Zhaoyang et al. Accelerating diffusion models via early stop of the diffusion process. **arXiv preprint arXiv:2205.12524**, 2022.
- MOZÓ, A. et al. Synthetic flow-based cryptomining attack generation through generative adversarial networks. **Scientific Reports**, v. 12, 1 2022. DOI: 10.1038/s41598-022-06057-2.

- MUKHERJEE, Mimi; KHUSHI, Matloob. SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features. **Applied System Innovation**, MDPI, v. 4, n. 1, p. 18, 2021.
- NASIR, Jamal Abdul; KHAN, Osama Subhani; VARLAMIS, Iraklis. Fake news detection: A hybrid CNN-RNN based deep learning approach. **International Journal of Information Management Data Insights**, Elsevier, v. 1, n. 1, p. 100007, 2021.
- NICHOL, Alexander Quinn; DHARIWAL, Prafulla. Improved denoising diffusion probabilistic models. In: PMLR. INTERNATIONAL Conference on Machine Learning. [S.l.: s.n.], 2021. P. 8162–8171.
- PARIKH, Shivam B; ATREY, Pradeep K. Media-rich fake news detection: A survey. In: IEEE. 2018 IEEE conference on multimedia information processing and retrieval (MIPR). [S.l.: s.n.], 2018. P. 436–441.
- PASZKE, Adam et al. Pytorch: An imperative style, high-performance deep learning library. **Advances in neural information processing systems**, v. 32, 2019.
- PÉREZ-ROSAS, Verónica et al. Automatic detection of fake news. **arXiv preprint arXiv:1708.07104**, 2017.
- POUDEVIGNE-DURANCE, T.; JONES, O. D.; QIN, Y. Mawgan: a generative adversarial network to create synthetic data from datasets with missing data. **Electronics**, v. 11, p. 837, 6 2022. DOI: 10.3390/electronics11060837.
- RAJABI, Amirarsalan; GARIBAY, Ozlem Ozmen. TabFairGAN: Fair Tabular Data Generation with Generative Adversarial Networks. **Machine Learning and Knowledge Extraction**, v. 4, n. 2, p. 488–501, 2022. ISSN 2504-4990. DOI: 10.3390/make4020022.
- RUBIN, Victoria L; CONROY, Niall J; CHEN, Yimin. Towards news verification: Deception detection methods for news discourse. In: HAWAII International Conference on System Sciences. [S.l.: s.n.], 2015. P. 5–8.
- RUZHANSKY, Natali; SEO, Sungyong; LIU, Yan. Csi: A hybrid deep model for fake news detection. In: PROCEEDINGS of the 2017 ACM on Conference on Information and Knowledge Management. [S.l.: s.n.], 2017. P. 797–806.
- SALIMANS, T. et al. Improved techniques for training gans. **arXiv preprint arXiv:1606.03498**, 2016. DOI: 10.48550/arxiv.1606.03498.
- SEGHOUANE, Abd-Krim; AMARI, Shun-Ichi. The AIC criterion and symmetrizing the Kullback–Leibler divergence. **IEEE Transactions on Neural Networks**, IEEE, v. 18, n. 1, p. 97–106, 2007.
- SHAHZAD, K. et al. A scoping review of the relationship of big data analytics with context-based fake news detection on digital media in data age. **Sustainability**, v. 14, p. 14365, 21 2022. DOI: 10.3390/su142114365.
- SHARMA, J. et al. Deepfakes classification of faces using convolutional neural networks. **Traitement Du Signal**, v. 39, p. 1027–1037, 3 2022. DOI: 10.18280/ts.390330.
- SHARMA, Karishma et al. **Combating Fake News: A Survey on Identification and Mitigation Techniques**. [S.l.: s.n.], 2019. arXiv: 1901.06437 [cs.LG].

- SHU, K. et al. Hierarchical propagation networks for fake news detection: investigation and exploitation. **Proceedings of the International Aai Conference on Web and Social Media**, v. 14, p. 626–637, 2020. DOI: 10.1609/icwsm.v14i1.7329.
- SHU, Kai et al. Fake news detection on social media: A data mining perspective. **ACM SIGKDD explorations newsletter**, ACM New York, NY, USA, v. 19, n. 1, p. 22–36, 2017.
- SINGH, A.; OGUNFUNMI, T. An overview of variational autoencoders for source separation, finance, and bio-signal applications. **Entropy**, v. 24, p. 55, 1 2021. DOI: 10.3390/e24010055.
- ST, S. et al. Deep learning model for deep fake face recognition and detection. **Peerj Computer Science**, v. 8, e881, 2022. DOI: 10.7717/peerj-cs.881.
- SUN, Guangzhi et al. Fully-hierarchical fine-grained prosody modeling for interpretable speech synthesis. In: IEEE. ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). [S.l.: s.n.], 2020. P. 6264–6268.
- SUROSO, D.; CHERNTANOMWONG, P.; SOORAKSA, P. Synthesis of a small fingerprint database through a deep generative model for indoor localisation. **Elektronika Ir Elektrotehnika**, v. 29, p. 69–75, 1 2023. DOI: 10.5755/j02.eie.31905.
- TAN, L.; LU, J.; JIANG, H. Tomato leaf diseases classification based on leaf images: a comparison between classical machine learning and deep learning methods. **AgriEngineering**, v. 3, p. 542–558, 3 2021. DOI: 10.3390/agriengineering3030035.
- TOBIN, J. et al. Domain randomization for transferring deep neural networks from simulation to the real world. **2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, 2017. DOI: 10.1109/iro.2017.8202133.
- VOSOUGHI, Soroush; ROY, Deb; ARAL, Sinan. The spread of true and false news online. **Science**, v. 359, n. 6380, p. 1146–1151, 2018. DOI: 10.1126/science.aap9559. eprint: <https://www.science.org/doi/pdf/10.1126/science.aap9559>.
- WANG, Shihan; TERANO, Takao. Detecting rumor patterns in streaming social media. In: IEEE. 2015 IEEE international conference on big data (big data). [S.l.: s.n.], 2015. P. 2709–2715.
- WANG, William Yang. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. [S.l.: s.n.], 2017. arXiv: 1705.00648 [cs.CL].
- WILCOXON, F. Individual comparisons by ranking methods. **Biom. Bull.**, 1, 80–83. [S.l.: s.n.], 1945.
- WYATT, J. et al. Anoddpm: anomaly detection with denoising diffusion probabilistic models using simplex noise. **2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)**, 2022. DOI: 10.1109/cvprw56347.2022.00080.
- XIAO, X.; GANGULI, S.; PANDEY, V. Vae-info-cgan. **Proceedings of the 13th ACM SIGSPATIAL International Workshop on Computational Transportation Science**, 2020. DOI: 10.1145/3423457.3429361.

XU, Lei et al. Modeling tabular data using conditional gan. **Advances in Neural Information Processing Systems**, v. 32, 2019.

ZHOU, S.; XU, J. et al. Numerical simulation on the larger concentration difference characteristics of dense granular jet in a coaxial gas stream. **The Canadian Journal of Chemical Engineering**, v. 101, p. 477–491, 1 2022. DOI: 10.1002/cjce.24361.

ZHOU, Xinyi; ZAFARANI, Reza. A survey of fake news: Fundamental theories, detection methods, and opportunities. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 53, n. 5, p. 1–40, 2020.