LVI SBPO SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
DE 04 ATÉ 07 NOVEMBRO 2024
FORTALEZA | CEARÁ

APLICAÇÃO DE REDES ADVERSARIAIS GENERATIVAS NA GERAÇÃO DE DADOS PARA PLATAFORMAS OFFSHORE A SEREM DESCOMISSIONADAS

Carlos Vinicio D. Araújo, Jailon William B. Oliveira da Silva, Pablo Luiz Braga Soares
Laboratório de Pesquisa & Desenvolvimento do NEMO, Universidade Federal do Ceará
Avenida Felipe Santiago – Nº 411, 62.900-420, Russas/CE, Brasil
{carlosvinicio, williambrunocc}@alu.ufc.br, pablo.soarea@ufc.br

Priscila da Cunha Jácome Vidal
Universidade Federal Rural do Semi-Árido
Rua Francisco Mota, 572, 59625-900 Mossoró/RN, Brasil
priscila.jacome@ufersa.edu.br

RESUMO

O descomissionamento de plataformas *offshore* fixas é uma atividade crucial para a indústria de óleo e gás, visando garantir a segurança e minimizar os impactos ambientais. Este trabalho tem como objetivo explorar a aplicação de Redes Adversariais Generativas (GANs) na geração de dados sintéticos para apoiar decisões de descomissionamento. A metodologia envolveu a criação e análise de uma base de dados sintéticos utilizando GANs, treinadas com dados reais de plataformas descomissionadas. Os resultados demonstram que os dados gerados mantêm alta fidelidade em relação aos dados originais, com uma similaridade de $85,77\%$. Conclui-se que o uso de dados sintéticos pode superar a escassez de dados, proporcionando uma base robusta para o aprendizado de máquina e auxiliando na tomada de decisões no descomissionamento de plataformas offshore, promovendo operações mais seguras e sustentáveis.

PALAVRAS CHAVE. Descomissionamento, Aprendizado de Máquina, Dados Sintéticos.

Tópicos: EN&PG – PO na Área de Energia, Petróleo e Gás, IC – Inteligência Computacional

ABSTRACT

The decommissioning of *offshore* fixed platforms is a crucial activity for the oil and gas industry, aiming to ensure safety and minimize environmental impacts. This study aims to explore the application of Generative Adversarial Networks (GANs) in generating synthetic data to support decommissioning decisions. The methodology involved creating and analyzing a synthetic database using GANs, trained with real data from decommissioned platforms. The results demonstrate that the generated data maintains high fidelity to the original data, with a similarity of 85.77%. It is concluded that the use of synthetic data can overcome data scarcity, providing a robust basis for machine learning and aiding decision-making in the decommissioning of offshore platforms, promoting safer and more sustainable operations.

KEYWORDS. Decommissioning, Machine Learning, Synthetic Data.

Paper topics: OR in the Area of Energy, Oil and Gas, CI – Computational Intelligence

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

1. Introdução

O descomissionamento de uma plataforma inicia quando as operações se tornam inviáveis economicamente, devido ao esgotamento dos recursos, aumento dos custos operacionais ou mudanças no mercado [Chandler et al., 2017]. Nesse momento, é crucial escolher um método adequado para a remoção da estrutura, considerando segurança, custos e impactos ambientais.

Durante o planejamento do descomissionamento, é essencial realizar uma avaliação detalhada das condições técnicas e ambientais da plataforma. Esta avaliação inclui a inspeção da integridade da estrutura, análise dos impactos na vida marinha e identificação de potenciais poluentes liberados na atmosfera, entre outros fatores [Krishnamoorthy e Tremblay, 2023]. Além disso, é necessário considerar a reutilização de componentes e o descomissionamento parcial, quando possível, para minimizar desperdícios e maximizar a sustentabilidade.

Diversos modelos foram propostos na literatura para um descomissionamento eficiente e sustentável. Fowler et al. [2014] aplicaram um modelo multicritério, utilizando métodos quantitativos para avaliar e comparar opções de descomissionamento, considerando custos, impactos ambientais, riscos de segurança e benefícios sociais, ajudando na tomada de decisões equilibradas. Kanmkamnerd et al. [2016] utilizaram a Análise Ambiental Líquida (NEBA) para avaliar as consequências ambientais líquidas das opções de descomissionamento, identificando a alternativa com maior benefício ambiental.

Além dos modelos tradicionais, métodos de aprendizado de máquina têm sido aplicados para otimizar o descomissionamento. Martins et al. [2020] reduziram o número de critérios necessários para a seleção de opções, tornando o processo mais eficiente e preciso, utilizando algoritmos de aprendizado de máquina para descomissionamento de dutos, permitindo a análise de grandes volumes de dados e a identificação de padrões e tendências não detectáveis por métodos tradicionais.

Apesar dos avanços, um dos maiores obstáculos no uso de métodos de aprendizado de máquina para decisões de descomissionamento é a falta de dados suficientes [Vuttipittayamongkol et al., 2021]. Os relatórios técnicos disponíveis ao público são limitados, o que dificulta o treinamento adequado de algoritmos de aprendizado de máquina. Para superar essa limitação, este estudo propõe o uso de Redes Adversariais Generativas (GANs) para criar novos registros de plataformas sintéticas, baseados em um conjunto de dados de plataformas reais.

O objetivo desta pesquisa é validar a fidelidade dos dados sintéticos gerados e demonstrar seu potencial uso em aplicações futuras de aprendizado de máquina no processo de descomissionamento de plataformas *offshore* fixas de óleo e gás. Ao comparar os dados artificiais com os originais, busca-se mostrar que os registros sintéticos mantêm alta precisão e confiabilidade.

O restante deste trabalho está estruturado da seguinte forma: na Seção 2, são abordados os conceitos e informações teóricas sobre descomissionamento de plataformas offshore e aprendizado de máquina, incluindo Redes Neurais Artificiais e Redes Adversariais; a Seção 3 apresenta as práticas utilizadas para a criação dos dados sintéticos, desde informações e estrutura da base de dados até a arquitetura das redes geradora e discriminadora da GAN, bem como seu método de treinamento para a geração dos dados artificiais; a Seção 4 apresenta como a base de dados sintética é analisada, quais são os resultados obtidos e o impacto deles na pesquisa; por último, a Seção 5 apresenta uma conclusão sobre o que foi pesquisado e alcançado com este trabalho.

2. Fundamentação Teórica

Este tópico aborda os fundamentos para a compreensão do problema deste trabalho. Na Seção 2.1, são apresentados os conceitos utilizados no descomissionamento de plataformas offshore, bem como os problemas e as soluções que buscam mitigar seus impactos. Já na Seção 2.2, são apresentados os conceitos da área de Aprendizado de Máquina, suas aplicações em diversos campos e as

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

Redes Neurais Artificiais e Adversariais Generativas nas Subseções 2.2.1 e 2.2.2, respectivamente. Esta última rede visa solucionar o problema da falta de dados reais, gerando dados sintéticos que simulam condições reais das plataformas.

2.1. Descomissionamento de plataformas

O descomissionamento de plataformas *offshore* fixas de óleo e gás é um processo complexo que envolve múltiplas etapas e a coordenação de diversas partes interessadas. Este procedimento é crucial para garantir que as operações sejam encerradas de maneira segura e ambientalmente responsável, atendendo às rígidas regulamentações.

A fase de descomissionamento envolve várias atividades, incluindo a paralisação das operações, o tamponamento e abandono dos poços, a remoção das plataformas e estruturas associadas, a limpeza de resíduos com potencial de contaminação e a restauração do ambiente marítimo. Esse processo é essencial para prevenir riscos ambientais e de segurança, além de cumprir as regulamentações governamentais [Kaiser e Liu, 2014].

Bull e Love [2019] discutem que a escolha da estratégia de descomissionamento pode incluir a remoção completa, a disposição no local e a conversão das estruturas em recifes artificiais. Eles destacam que essas estruturas, após anos no mar, desenvolvem uma biomassa marinha rica, composta por infauna, epifauna e peixes, cuja diversidade é influenciada por fatores como a coluna d'água, a complexidade da estrutura da jaqueta, a geografia e a posição da plataforma. Portanto, a decisão sobre o método de descomissionamento deve considerar não apenas os aspectos técnicos e econômicos, mas também os impactos ambientais.

O descomissionamento envolve decisões complexas e variadas. Segundo Tan et al. [2017], a remoção de estruturas pode causar a liberação de poluentes durante a desmontagem e, para mitigar esses impactos, é essencial realizar avaliações ambientais e implementar medidas preventivas. Eles também destacam que a reutilização é a opção mais sustentável, consumindo menos energia do que a reciclagem. Fowler et al. [2014] enfatizam a importância de práticas sustentáveis, como a reutilização de componentes e o desmantelamento parcial, para reduzir desperdícios e maximizar a sustentabilidade. Eles apresentam 13 opções de descomissionamento para plataformas fixas, categorizadas em reutilização, reciclagem e descarte, conforme apresentado na Figura 1.

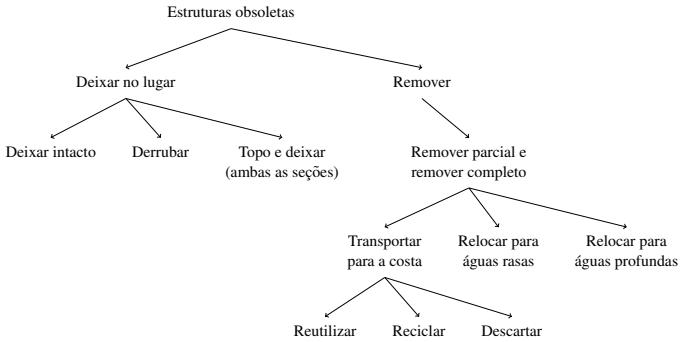


Figura 1: Opções de descomissionamentos de plataformas, adaptado de Fowler et al. [2014]

2.2. Aprendizado de máquina

O Aprendizado de Máquina permite que ferramentas operem de forma mais autônoma, reduzindo a necessidade de intervenção humana, ao ensinar computadores a aprenderem a partir de conjuntos de dados [Géron, 2019]. Analisando as características desses dados, é possível identificar padrões e tendências que revelam informações cruciais sobre o processo que os gerou [Faceli

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

et al., 2011]. Além disso, é possível criar sinteticamente novos dados a partir dos dados originais, ampliando as possibilidades de análise e aplicação.

Diversos modelos de aprendizado de máquina são utilizados em diferentes áreas. Na Medicina, aplicam-se na detecção de doenças cardiovasculares [Oliveira et al., 2023], identificação de marcadores genéticos para a doença de Alzheimer [Alves, 2024], e diagnóstico do Transtorno do Espectro Autista [Artoni, 2020]. Em Problemas de Otimização Combinatória, abrangem o problema do caixeiro viajante [Miki et al., 2018] e o problema da diversidade máxima [Silva, 2023].

Além disso, existem aplicações no descomissionamento, como prever opções para estruturas offshore de óleo e gás [Vuttipittayamongkol et al., 2021] e a redução de dimensionalidade para problemas multicritério [Martins et al., 2020]. Para a geração de dados sintéticos, os modelos de aprendizado de máquina podem criar conjuntos de dados que simulam condições reais, permitindo o treinamento e validação de algoritmos onde os dados reais são escassos ou difíceis de obter.

2.2.1. Rede Neurais Artificiais

A Rede Neural Artificial (RNA) é caracterizada pela sua arquitetura, que determina o tipo de rede e o número de neurônios, e pelo algoritmo de aprendizado, responsável por atualizar os pesos da rede conforme ela aprende com exemplos. O processo de aprendizado envolve a análise das características dos dados e sua generalização por meio do ajuste dos pesos, realizado pelo algoritmo de aprendizagem [Goodfellow et al., 2016]. Além disso, toda RNA possui neurônios artificiais, responsáveis pelo processamento das informações.

Conforme Matsunaga [2012], a unidade de processamento básica em redes neurais é o neurônio artificial, que possui entradas (sinapses) e uma saída determinada pela combinação linear das entradas ponderadas, processadas por uma função de ativação. O modelo de McCulloch e Pitts [1943] serve de base para esses neurônios, e modelos mais avançados incluem um sinal de bias que ajusta a ativação do neurônio. Desta forma, os pesos amplificam os sinais recebidos e a função de ativação define a resposta do neurônio.

2.2.2. Redes Adversariais Generativas

As Redes Adversárias Generativas (GAN) são um tipo de inteligência artificial que emprega duas redes neurais: a rede geradora, que cria novos registros, e a rede discriminadora, que determina se esses registros são genuínos. Essas redes solucionam problemas de modelagem generativa, analisando uma base de dados de treinamento e aprendendo a distribuição de probabilidade que a gerou. Assim, é possível gerar registros adicionais que se assemelham aos exemplos originais através da distribuição de probabilidade estimada [Goodfellow et al., 2020].

Normalmente, uma GAN é treinada utilizando uma base de dados contendo diversas imagens distintas, podendo ser utilizada para gerar imagens [Liu et al., 2017] ou aprimorar a resolução de imagens [Ledig et al., 2017], mas não se limita a essa categoria. Conforme Aggarwal et al. [2021], as aplicações da GAN abrangem diversos campos, incluindo geração de objetos 3D, medicina, pandemias, processamento de imagens, detecção de rosto, transferência de textura e controle de tráfego. Ele acrescenta que, geralmente, a GAN opera seguindo estes passos:

1. Os dados novos são gerados pela rede geradora, que é treinado pela rede discriminador usando dados reais como exemplo;

2. O sistema é treinado de forma que a rede geradora produza dados cada vez mais fiéis aos originais e a rede discriminadora tenha dificuldade em distinguir entre dados reais e dados gerados pelo gerador;

3. Um conjunto de exemplos iniciais é usado para ensinar o discriminador a reconhecer dados reais;

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

4. Para amostras de treinamento, mais exemplos são apresentados ao discriminador até que ele se torne preciso o suficiente;

5. O gerador é ensinado a criar dados que confundem o discriminador quando ele recebe entradas com valores aleatórios, também conhecidos como vetores de ruídos;

6. Por últimos, ambas as redes são ajustadas para melhorar suas habilidades, com a rede gerador criando imagens melhores e o discriminador se tornando mais preciso em distiguir as imagens reais e as geradas.

A Figura 2 mostra a dinâmica de atuação da rede GAN[1]. A rede geradora busca aumentar a probabilidade de erro da discriminadora, visando enganá-la, enquanto esta última procura reduzir essa probabilidade. Ao longo do processo de treinamento, ambas as redes se aprimoram aprendendo com seus erros e competem entre si de forma contínua até alcançarem um estado de equilíbrio.

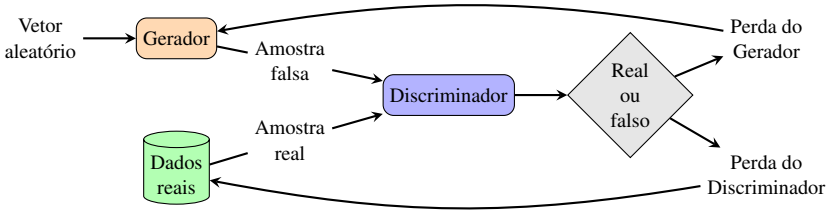


Figura 2: Diagrama de fluxo da GAN

Por último, Bendale [2020] ilustra uma interpretação mais matemática, afirmando que as GANs podem ser compreendidas como dois jogadores que competem entre si em um jogo de minimização-máximização, onde a rede geradora G e a rede discriminadora D são treinadas juntas para otimizar a seguinte função de custo:

$$\min_G \max_D V(D,G) = \mathbb{E}_{jr}[\log D(q)] + \mathbb{E}_{jq}[\log(1 - D(G(p)))] \tag{1}$$

Neste contexto, jr é a distribuição dos dados do conjunto de treinamento (dados reais) e p é o vetor de ruído. A função $G(p)$ representa a saída da rede geradora quando alimentada com um vetor de ruído p. $D(q)$ é a probabilidade de q ser um exemplo real da base de dados. Desta forma, $\mathbb{E}_{jr}[\log D(q)]$ representa o logaritmo da probabilidade de que D prediz q corretamente como um dado real, enquanto $\mathbb{E}_{jq}[\log(1 - D(G(p)))]$ é o logaritmo da probabilidade de D predizer um dado gerado, ou seja, $G(x)$, como um dado real. Portanto, o objetivo de D é maximizar a acurácia da diferenciação dos dados e o objetivo de $G(z)$ é minimizar essa acurácia, gerando dados $G(z)$ tão reais quanto possível.

3. Metodologia

A abordagem central consiste em utilizar redes adversariais generativas (GANs) para criar plataformas sintéticas que se assemelham às reais. A Seção 3.1 apresenta como foi criada a base de dados e como está estruturada e a Seção 3.2 explica como ocorre a geração dos dados sintéticos. Enquanto a Subseção 3.2.1 apresenta as arquiteturas das redes geradora e discriminadora, que são as duas redes neurais artificiais que compõem a GAN. Por fim, a Subseção 3.2.2 aborda o treinamento da rede e a geração dos dados sintéticos.

[1]Fonte: Adapatado de `https://aws.amazon.com/pt/what-is/gan/`

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

3.1. Base de Dados

A base de dados utilizada neste trabalho foi obtida a partir dos relatórios de plataformas descomissionadas fornecidos pelo governo do Reino Unido, disponíveis no site oficial *UK Government: Oil and Gas Decommissioning of Offshore Installations and Pipelines* [UK Government, 2024]. Dos relatórios técnicos disponíveis apenas 7 eram de plataformas do tipo fixa, foco do trabalho, e possuiam informações completas para formar a base dos registros sintéticos, esses relatório comparavam dois tipo de opções de descomissionamento: parcial e completo. A Tabela 1 contém uma breve descrição de cada critério utilizado para avaliação das plataformas.

Tabela 1: Descrição das características

Característica	Descrição
Profundidade da água (m)	Profundidade do mar no local da plataforma.
Peso (t)	Peso total da estrutura a ser removida, medido em toneladas.
Idade da estrutura	Tempo em anos desde a instalação, afetando a condição e a complexidade da remoção.
Tipo de produção	Tipo de recursos anteriormente extraídos, ao qual é atribuído um valor: 1 para óleo e gás, 2 para somente óleo e 3 para somente gás.
Número de pernas	Quantidade de pernas que precisam ser desmontadas ou removidas.
Número de pilares	Quantidade de pilares presentes na estrutura.
Altura da jaqueta (m)	Altura da estrutura de suporte (jaqueta) a ser desmantelada, medida em metros.
Distância da costa (km)	Distância da plataforma até a costa.
Risco para outros usuários	Avaliação do risco que o descomissionamento representa para outros usuários que utilizam o mar, como navegadores e pescadores. O valor foi extraído dos relatórios em formato de porcentagem.
Risco para trabalhadores	Avaliação do risco para os trabalhadores envolvidos diretamente no descomissionamento da plataforma. O valor foi extraído dos relatórios em formato de porcentagem.
Consumo de energia (GJ)	Quantidade de energia necessária para realizar o descomissionamento, medido em gigajoules.
Emissão de poluentes	Quantidade de poluentes emitidos durante o processo de descomissionamento, medido em toneladas.
Impactos ambientais	Avaliação dos impactos ambientais causados pelo descomissionamento. O valor foi extraído dos relatórios em formato de porcentagem.
Viabilidade técnica	Avaliação da viabilidade técnica das operações de descomissionamento. O valor foi extraído dos relatórios em formato de porcentagem.
Impacto no comércio marítimo	Avaliação do impacto do descomissionamento no comércio marítimo, incluindo navegação e pesca. O valor foi extraído dos relatórios em formato de porcentagem.
Impacto na comunidade	Avaliação do impacto do descomissionamento nas comunidades locais, incluindo aspectos econômicos e sociais. O valor foi extraído dos relatórios em formato de porcentagem.
Custo total da remoção	Custo estimado para a remoção completa da plataforma durante o descomissionamento. O valor foi extraído dos relatórios em formato de porcentagem.

Na área de descomissonamento, os aspectos utilizados para avaliação e decisão são comumente referidos como critérios. Neste trabalho, ao aplicar técnicas de aprendizado de máquina, os critérios serão referenciados como características ou atributos. As características apresentadas na Tabela 1 foram extraídas dos relatórios e moldadas em formatos apropriados para a aplicação dos modelos de aprendizado de máquina. As coluna de 2 a 8 da Tabela 2 apresentam os dados de todas as características extraídas das plataformas específicas, a saber: Brae Bravo, Brae Alpha, Ninian Northern, Murchison Field, Miller, North West Hutton e East Brae.

3.2. Geração dos dados sintéticos

Para a criação dos dados sintéticos, foram utilizadas algumas bibliotecas e ferramentas da linguagem de programação *Python3*. A implementação da Rede Adversarial Generativa (GAN) foi realizada utilizando a biblioteca *PyTorch*, que oferece recursos para a construção e treinamento de modelos de aprendizado profundo. Para avaliar a confiabilidade dos dados sintéticos

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

Tabela 2: Dados das plataformas

Característica	Brae Bravo	Brae Alpha	Ninian Northern	Murchison Field	Miller	North West Hutton	East Brae
Profundidade da água (m)	99	112	141	156	103	144	116
Peso (t)	22.000	20.364	15.561	27.647	18.584	17.500	10.054
Idade da estrutura	36	41	46	44	33	41	31
Tipo de produção	1	1	2	2	1	3	1
Número de pernas	8	8	8	8	8	8	4
Número de pilares	8	9	26	32	20	20	5
Altura da jaqueta (m)	114	123	147	166	122	154	136
Distância da costa (km)	269	259	120	240	230	130	282
Risco para outros usuários (parcial)	0	0	$2,3 \times 10^{-5}$	$1,5 \times 10^{-5}$	$7,8 \times 10^{-8}$	$8,80449 \times 10^{-6}$	0
Risco para trabalhadores (parcial)	0,12	0,16	0,010	0,02	0,06	0,13	0,08
Consumo de energia (parcial)	817.000	895.500	530.148	570.818	511.765	568.000	433.179
Emissão de poluentes (parcial)	58.500	64.000	31.064	45.266	41.170	44.000	32.191
Impactos ambientais (parcial)	1	1	1	1	1	0	1
Viabilidade técnica (parcial)	1	1	1	1	0.15	0.7	1
Impacto no comércio marítimo (parcial)	0,78	0,78	0,94	0,66	0	0	0,78
Impacto na comunidade (parcial)	0,5	0,5	1,0	1,0	0	0	0,5
Custo total da remoção (parcial)	0,56	0,56	1	0,57	0,73	0,686	0,41
Risco para outros usuários (completo)	1	1	0	0	0	0	1
Risco para trabalhadores (completo)	0,29	0,32	0,025	0,04	0,09	0,14	0,14
Consumo de energia (completo)	1.110.100	1.180.500	297.654	487.750	733.082	520.000	469.823
Emissão de poluentes (completo)	87.000	92.000	24.277	40.416	59.588	42.000	37.162
Impactos ambientais (completo)	0	0	0.66	1	0	0	0
Viabilidade técnica (completo)	0	0	0,25	0,50	0,39	0,45	0
Impacto no comércio marítimo (completo)	0,79	0,79	1,0	1	0	0	0,79
Impacto na comunidade (completo)	0,5	0,5	1	1	0	0	0,5
Custo total da remoção (completo)	1	1	0.53	1	1	0,602	1

em comparação com os dados reais, fez-se o uso da biblioteca *sdmetrics*, que fornece métricas específicas para essa finalidade. Além disso, as bibliotecas *pandas* e *numpy* foram utilizadas para o tratamento e manipulação dos dados, enquanto as bibliotecas *matplotlib.pyplot* e *seaborn* foram empregadas para a visualização e análise gráfica dos dados. Toda a codificação para a geração dos dados foi conduzida em *notebooks Jupyter* e está disponível no link: `https://encurtador.com.br/Eyqso`.

3.2.1. Arquitetura da Rede Geradora e Discriminadora

A rede geradora recebe em sua camada de entrada um vetor de características com valores aleatórios. Em seguida, essa camada passa essas informações para cada neurônio da próxima camada oculta, que contém 128 neurônios. Os neurônios utilizam a função de ativação ReLU para passar as informações para a camada seguinte, introduzindo não linearidade à rede, o que permite que ela aprenda relações mais complexas entre os dados. Na última camada oculta da rede geradora, as informações são passadas para a camada de saída, que contém um número de neurônios equivalente ao tamanho de cada registro das plataformas reais. A saída desses neurônios corresponde aos valores de cada coluna dos dados sintéticos. Para normalizar os dados, foi utilizada a função de ativação *Tangente Hiperbólica* na camada de saída para transformar os valores de cada coluna do registro entre -1 e 1.

A rede discriminadora, por sua vez, recebe em sua camada de entrada um vetor de características que representa um registro sintético ou real. A camada de entrada passa as informações para cada neurônio da próxima camada, que também contém 128 neurônios. Cada neurônio utiliza a função de ativação ReLU para introduzir não linearidade na rede. A camada de saída possui apenas um único neurônio, que recebe as informações de todos os neurônios da camada anterior. A função de ativação usada na camada de saída foi a *Sigmoid*, que produz uma saída entre 0 e 1, o que representa a probabilidade de o registro ser real.

3.2.2. Treinamento das redes

Para iniciar o treinamento específico de cada rede é necessário definir e ajustar alguns hiperparâmetros: **Épocas**: representa uma passagem completa por todo o conjunto de dados de

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

treinamento; **Taxa de aprendizado**: um hiperparâmetro que determina o tamanho dos passos dados na direção dos gradientes durante a otimização; **Lote**: um subconjunto do conjunto de dados usado para calcular o gradiente e atualizar os pesos; e o otimizador **Adam**: ajusta a taxa de aprendizado para cada parâmetro com base em estimativas de momentos de primeira e segunda ordem (média e variância) dos gradientes.

Com os hiperparâmetros definidos, o treinamento da rede discriminadora e iniciado. Atribui-se zero aos valores dos gradientes do otimizador para evitar a acumulação e em seguida, a rede discriminadora faz previsões com registros reais, calcula o erro e atualiza os pesos com base nos gradientes. Esse processo é repetido para registros falsos (sintéticos). A rede discriminadora faz previsões sobre registros falsos fornecidos pela geradora, calcula a perda associada, e atualiza os pesos com base nessa perda. A função de treinamento retorna a soma das perdas da discriminadora em relação aos registros reais e falsos.

O treinamento da rede geradora inicia de forma semelhante, atribuindo valor zero aos gradientes do otimizador. Dessa forma, a geradora cria novos registros falsos e os passa para a discriminadora, que retorna suas previsões. O objetivo da geradora é enganar a discriminadora, minimizando sua própria perda, calculada com base nessas previsões, e atualizando os pesos de acordo com os gradientes calculados.

O treinamento das redes durou 60.000 épocas, com um lote de tamanho 128, taxa de aprendizado de 0,00001 e o uso do otimizador Adam. Durante cada época, a discriminadora aprimorou sua capacidade de identificar registros falsos, enquanto a geradora se ajustou para criar registros sintéticos mais convincentes. Após o treinamento, a geradora pode produzir novos registros sintéticos semelhantes aos reais.

4. Análise dos Resultados

Para comparar a qualidade dos dados sintéticos gerados a partir da base de dados original, foi utilizada a biblioteca *SDMetrics*, que fornece ferramentas específicas para esse objetivo. O processo consiste em instanciar um objeto da classe *QualityReport*, passando como parâmetros tanto a base de dados original quanto a sintética para seu método *generate*, responsável por fazer a avaliação. Além disso, é necessário fornecer um metadado em formato de dicionário contendo os nomes das colunas e seus tipos, indicando se os valores de cada coluna são ponto flutuante ou inteiros. O resultado das avaliações feitas é mostrado na Tabela 3 em forma de porcentagem.

A avaliação tem como base duas métricas importantes para a comparação. A primeira métrica é denominada *Column Shapes*, que calcula a semelhança entre uma coluna real e uma coluna sintética em termos de formato, também conhecido como distribuição marginal ou histograma 1D da coluna. Para isso, é utilizada a estatística Kolmogorov-Smirnov (KS) na função *KSComplement*, que compara a distribuição de probabilidade da coluna sintética com a original. Em seguida, uma pontuação é atribuída, variando de 0 a 1, onde 1 indica que todos os dados sintéticos são exatamente iguais aos dados reais, enquanto 0 indica o oposto.

A segunda métrica, *Column Pair Trends*, utiliza a função *CorrelationSimilarity* para calcular a correlação entre um par de colunas numéricas e determinar o quão similar é a correlação dos valores da coluna sintética com a coluna original. Essa métrica avalia os coeficientes de correlação de Pearson, que mede se duas colunas são linearmente correlacionadas, e de Spearman, que mede se estão relacionadas monotonicamente. Ambas as métricas são comumente utilizadas para medir a correlação entre colunas.

A função *CorrelationSimilarity* é aplicada a cada par de colunas, comparando, por exemplo, a coluna de peso da jaqueta com todas as demais. Ela associa uma avaliação de 0 a 1, indicando o quão similar é a correlação de cada coluna, sendo 1 indicando correlações idênticas e 0 o oposto. As médias da avaliação de cada coluna são mostradas na Tabela 3 em forma de porcentagem.

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

Tabela 3: Taxa de semelhança entre as colunas dos dados reais e sintéticos

Características	*Column Shapes*	*Column Pair Trends*
Profundidade da água (m)	78,89%	94,08%
Peso (t)	73,61%	95,97%
Idade da estrutura	82,43%	**96, 35%**
Tipo de produção	84,41%	93,37%
Número de pernas	**94, 10%**	**96, 08%**
Número de pilares	73,02%	93,00%
Altura da jaqueta (m)	72,13%	95,11%
Distância da costa (km)	63,58%	92,17%
Risco para outros usuários (parcial)	75,94%	95,34%
Risco para trabalhadores (parcial)	79,43%	91,19%
Consumo de energia (parcial)	76,58%	**96, 23%**
Emissão de poluentes (parcial)	77,33%	95,01%
Impactos ambientais (parcial)	**95, 82%**	91,03%
Viabilidade técnica (parcial)	75,98%	94,05%
Impacto no comércio marítimo (parcial)	65,02%	94,35%
Impacto na comunidade (parcial)	71,61%	90,54%
Custo total da remoção (parcial)	61,36%	91,88%
Risco para outros usuários (completo)	**99, 82%**	93,17%
Risco para trabalhadores (completo)	77,58%	93,52%
Consumo de energia (completo)	73,06%	92,20%
Emissão de poluentes (completo)	74,26%	95,47%
Impactos ambientais (completo)	83,46%	95,35%
Viabilidade técnica (completo)	80,81%	92,37%
Impacto no comércio marítimo (completo)	72,15%	93,91%
Impacto na comunidade (completo)	71,47%	93,77%
Custo total da remoção (completo)	77,36%	89,93%

A média ponderada dessas duas avaliações é utilizada para determinar a fidelidade dos dados sintéticos em relação aos dados originais. No caso de *Column Shapes*, a pontuação geral foi $77,36\%$, enquanto para *Column Pair Trends*, a pontuação geral foi $94,18\%$. Esta última avaliação indica que os dados sintéticos se saíram muito bem em replicar as relações estatísticas entre pares de variáveis, demonstrando a alta qualidade e utilidade dos dados sintéticos para análises que dependem dessas correlações. Adicionalmente, a avaliação dos dados sintéticos atestou uma fidelidade de $85,77\%$ aos dados originais, reforçando sua confiabilidade e utilidade para análises e aplicações.

Analisando especificamente os resultados obtidos na métrica *Column Shapes*, a coluna com o melhor resultado foi "Risco para outros usuários (completo)", com uma taxa de semelhança de $99,82\%$. Esse desempenho evidencia que a distribuição de probabilidade foi replicada com alta precisão pelo modelo de geração de dados sintéticos. Além disso, as colunas "Número de pernas"e "Impactos ambientais"também apresentaram avaliações elevadas, com $94,10\%$ e $95,82\%$, respectivamente. A elevada similaridade dessas três colunas sugere que os dados sintéticos mantiveram quase todas as propriedades estatísticas dos dados reais, o que é crucial para análises de risco e simulações de cenários.

Por outro lado, algumas colunas apresentaram pontuações mais baixas, como "Custo total da remoção (completo)" com $61,36\%$. Essa discrepância pode ser atribuída à complexidade inerente desses dados e as variaçções presentes nos registros reais, o que aumenta a dificuldade para a rede geradora identificar todas as nuances. A arquitetura da GAN utilizada, embora claramente eficaz na maioria das características, pode encontrar dificuldades em modelar dados com distribuições irregulares ou dependências complexas que não são totalmente capturadas pelas camadas atuais da rede.

Na métrica *Column Pair Trends*, com exceção da coluna "Custo total de remoção (com-

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

pleto)", que teve uma taxa de $89,93\%$, todas as colunas foram avaliadas com pontuações acima de 90%. Em destaque, estão as colunas "Idade da estrutura", "Número de pernas"e "Consumo de energia (parcial)", que tiveram as melhores avaliações, com $96,35\%$, $96,08\%$ e $96,23\%$, respectivamente. Isso significa que as relações entre as colunas foram precisamente replicadas, podendo ter diversas aplicações para análises. Por exemplo, as relações entre essas três variáveis podem servir de fundamento para análises de desempenho e previsão de consumo.

4.1. Comportamento da rede GAN no treinamento

Neste modelo de GAN proposto para a geração de dados sintéticos, verificou-se uma relação importante entre o número de épocas e a taxa de aprendizado. A tendência observada no modelo proposto é que a rede discriminadora sempre inicia vencendo a disputa, com uma taxa de acerto maior, enquanto a rede geradora começa perdendo. Entretanto, com o passar das épocas, a rede geradora começa a alcançar a discriminadora, que, por sua vez, começa a perder a disputa, até chegar a um ponto onde a rede geradora se sobressai significativamente à discriminadora. Adicionalmente, a velocidade com que a rede geradora alcança a discriminadora é determinada pela taxa de aprendizado. Portanto, para obter o melhor resultado possível no treinamento, este deve ser interrompido quando as redes alcançarem um equilíbrio em sua disputa. A Tabela 4 ilustra essa relação, mostrando a taxa de fidelidade dos dados conforme passam as épocas.

Tabela 4: Taxa de fidelidade dos dados sintéticos ao decorrer das épocas de treinamento

Características	E_1	E_2	E_3	E_4	E_5	E_6	E_7
Épocas	1.000	5.000	10.000	25.000	50.000	60.000	70.000
Fidelidade	66,51%	79,22%	81,96%	78,68%	62,43%	57,09%	59,35%

Ao todo, foram realizados 8 experimentos, com a taxa de fidelidade dos dados registrada nas colunas E_1, E_2, E_3, E_4, E_5, E_6 e E_7 ao final de cada experimento. Foram consideradas várias configurações de épocas de treinamento, mantendo constantes o tamanho do lote e a taxa de aprendizado, com 32 e $0,0001$ respectivamente. A fidelidade dos dados sintéticos foi medida ao longo dos diferentes números de épocas, conforme mostrado na tabela.

Percebe-se que, inicialmente, a fidelidade dos dados sintéticos aumenta com o número de épocas, alcançando um pico em torno de 10.000 épocas. Entretanto, ao continuar o treinamento além desse ponto, a fidelidade começa a diminuir, demonstrando o desequilíbrio citado anteriormente no treinamento das redes. Desta forma, este experimento reforça a necessidade de interromper o treinamento quando as redes alcançam um equilíbrio, evitando que uma rede se sobressaia excessivamente à outra.

5. Conclusão

O descomissionamento de plataformas *offshore* fixas de óleo e gás é um processo essencial e complexo, que envolve considerações técnicas, econômicas e ambientais. Este estudo explorou a aplicação de Redes Adversariais Generativas (GANs) na geração de dados sintéticos para suprir a falta de dados no planejamento do descomissionamento, demonstrando a viabilidade e a precisão desses dados para futuras aplicações de aprendizado de máquina.

Os resultados obtidos mostram que os dados sintéticos gerados pelas GANs mantêm alta fidelidade em relação aos dados reais, com uma taxa de semelhança de $85,77\%$. Essa alta fidelidade é fundamental para a confiabilidade dos modelos de aprendizado de máquina, garantindo que as decisões tomadas sejam baseadas em informações precisas e representativas das condições reais.

A utilização de dados sintéticos oferece uma solução promissora para superar a escassez de dados no campo do descomissionamento, permitindo uma análise mais robusta e informada.

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

Além disso, as técnicas de aprendizado de máquina podem otimizar o processo de descomissionamento, tornando-o mais eficiente, seguro e sustentável.

Assim, este estudo contribui para o avanço das práticas de descomissionamento, sugerindo que a integração de tecnologias inovadoras, como as GANs, pode melhorar significativamente a gestão e a execução desses projetos complexos. Para trabalhos futuros, espera-se utilizar a base de dados sintética obtida em aplicações de aprendizado de máquina, com o objetivo de fornecer uma decisão mais precisa sobre o tipo de descomissionamento a ser escolhido. Em última análise, espera-se que a adoção dos resultados deste estudo em aplicações futuras promova operações mais responsáveis e eficientes, minimizando os impactos ambientais e maximizando os benefícios sociais e econômicos.

Referências

Aggarwal, A., Mittal, M., e Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1): 100004.

Alves, J. F. (2024). Aplicação de técnicas de aprendizado de máquina na identificação de marcadores genéticos para a doença de alzheimer.

Artoni, A. A. (2020). Aplicação de aprendizado de máquina no auxílio ao diagnóstico do transtorno do espectro autista. Dissertação, Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, BR.

Bendale, S. (2020). A study of 3d-gans and their implementation challenges. *CLIO An Annual Interdisciplinary Journal of History*, p. 194.

Bull, A. S. e Love, M. S. (2019). Worldwide oil and gas platform decommissioning: A review of practices and reefing options. *Ocean & Coastal Management*, 168:274–306.

Chandler, J., White, D., Techera, E. J., Gourvenec, S., e Draper, S. (2017). Engineering and legal considerations for decommissioning of offshore oil and gas infrastructure in australia. *Ocean Engineering*, 131:338–347.

Faceli, K., Lorena, A. C., Gama, J., Almeida, T. A. d., e Carvalho, A. C. P. d. L. F. d. (2011). *Inteligência artificial: uma abordagem de aprendizado de máquina*. Rio de Janeiro: LTC. ISBN 9788521618805.

Fowler, A. M., Macreadie, P. I., Jones, D. O. B., e Booth, D. J. (2014). A multi-criteria decision approach to decommissioning of offshore oil and gas infrastructure. *Ocean & Coastal Management*, 87:20–29.

Goodfellow, I., Bengio, Y., e Courville, A. (2016). *Deep learning*. MIT press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., e Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc., Sebastopol, CA, second edition. ISBN 9781492032649.

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br

Kaiser, M. J. e Liu, M. (2014). Decommissioning cost estimation in the deepwater u.s. gulf of mexico–fixed platforms and compliant towers. *Marine Policy*, 48:185–196.

Kanmkamnerd, J., Phanichtraiphop, P., e Pornsakulsakdi, L. (2016). Neba application for jacket decommissioning techniques. In *International Petroleum Technology Conference*, p. 14–16, Bangkok.

Krishnamoorthy, P. e Tremblay, C. (2023). Offshore underwater fixed structure integrity assessment using lrut, pec & acfm advanced technologies with minimal marine growth removal. In *Abu Dhabi International Petroleum Exhibition and Conference*. OnePetro.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 4681–4690.

Liu, Y., Qin, Z., Luo, Z., e Wang, H. (2017). Auto-painter: Cartoon image generation from sketch by using conditional generative adversarial networks. *arXiv preprint arXiv:1705.01908*.

Martins, I. D., Bahiense, L., Infante, C. E., e Arruda, E. F. (2020). Dimensionality reduction for multi-criteria problems: An application to the decommissioning of oil and gas installations. *Expert Systems with Applications*, 148:113236.

Matsunaga, V. Y. (2012). Curso de redes neurais utilizando o matlab. *Belém do Pará*.

McCulloch, W. S. e Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133.

Miki, S., Yamamoto, D., e Ebara, H. (2018). Applying deep learning and reinforcement learning to traveling salesman problem. In *2018 International Conference on Computing, Electronics & Communications Engineering (ICCECE)*, p. 65–70. IEEE.

Oliveira, M. C., Ferreira, L. V. B., e de Oliveira Barreiros, M. (2023). Classificação de doenças cardiovasculares utilizando aprendizado de máquina. *Multidebates*, 7(1):38–44.

Silva, D. M. A. d. (2023). Aplicação de aprendizagem não supervisionada no problema da diversidade máxima em grafos. Trabalho de Conclusão de Curso (Graduação em Engenharia de Software) - Campus de Russas, Universidade Federal do Ceará, Russas, 2023.

Tan, H., Zhang, Y., Feng, Y., e Wang, J. (2017). Environmental impacts of offshore oil and gas activities: A comprehensive assessment. *Environmental Science & Technology*, 51(12):6957–6965.

UK Government (2024). Oil and gas: Decommissioning of offshore installations and pipelines. URL `https://www.gov.uk/guidance/oil-and-gas-decommissioning-of-offshore-installations-and-pipelines`. Accessed: 2024-05-24.

Vuttipittayamongkol, P., Tung, A., e Elyan, E. (2021). Towards machine learning-driven practices for oil and gas decommissioning – introduction of a new offshore pipeline dataset. In *Proceedings of the 9th International Conference on Computer and Communications Management*, ICCCM '21, p. 111–116, New York, NY, USA. Association for Computing Machinery. ISBN 9781450390071.

DOI: 10.59254/sbpo-2024-193647
https://proceedings.science/p/193647?lang=pt-br