

## Aplicação de Machine Learning na otimização multiobjetivo de ventiladores centrífugos

Matheus Costa Pereira<sup>1</sup>, Anderson Paulo de Paiva<sup>2</sup>, Matheus Brendon Francisco<sup>3</sup>, Tiago Martins de Azevedo<sup>4</sup>, Ronã Rinston Amaury Mendes<sup>5</sup>

<sup>1</sup>Núcleo de Otimização da Manufatura e Tecnologia da Inovação, Instituto de Engenharia de Produção, Universidade Federal de Itajubá, Brasil. Email: matheusc\_pereira@hotmail.com

<sup>2</sup>Núcleo de Otimização da Manufatura e Tecnologia da Inovação, Instituto de Engenharia de Produção, Universidade Federal de Itajubá, Brasil. Email: andersonppaiva@unifei.edu.br

<sup>3</sup>Núcleo de Otimização da Manufatura e Tecnologia da Inovação, Instituto de Engenharia de Produção, Universidade Federal de Itajubá, Brasil. Email: matheus\_brendon@yahoo.com.br

<sup>4</sup>Núcleo de Otimização da Manufatura e Tecnologia da Inovação, Instituto de Engenharia de Produção, Universidade Federal de Itajubá, Brasil. Email: tiago.deazevedo@yahoo.com.br

<sup>5</sup>Núcleo de Otimização da Manufatura e Tecnologia da Inovação, Instituto de Engenharia de Produção, Universidade Federal de Itajubá, Brasil. Email: rona.rinston@ifsuldeminas.edu.br

### Resumo

Este estudo investiga a combinação de técnicas de Machine Learning, Computational Fluid Dynamics e Multi-Objective Particle Swarm Optimization para otimizar ventiladores centrífugos. Geometrias e simulações são feitas com o software Ansys para alterar variáveis de entrada e obter variáveis de saída. Com base nesses dados, são desenvolvidos modelos de regressão utilizando Random Forest, Decision Tree e Support Vector Machines. A otimização dos hiperparâmetros é conduzida por Grid Search e a avaliação dos modelos é feita com métricas de precisão, acurácia, recall e F1-Score. Após a validação dos modelos, é aplicada uma otimização multiobjetivo para maximizar o Índice de Desempenho e minimizar o Custo do Material. O ventilador estudado é fabricado em aço inoxidável AISI 310. A eficácia da abordagem é confirmada por simulações fluidodinâmicas adicionais com os valores ótimos encontrados, em que o Random Forest apresentou a melhor performance com  $R^2$  de 0,966 para o Índice de Desempenho e 0,946 para o Custo do Material. As simulações fluidodinâmicas apresentaram uma diferença inferior a 10% em relação aos resultados previstos pelo modelo, evidenciando a precisão do método proposto.

**Palavras-chave:** Computational Fluid Dynamics; Machine Learning; Multiobjective Optimization; Meta-Heurística; Particle Swarm; Centrifugal Fan.

### Abstract

This study investigates the combination of Machine Learning, Computational Fluid Dynamics, and Multi-Objective Particle Swarm Optimization techniques to optimize centrifugal fans. Geometries and simulations are performed using Ansys software to alter input variables and obtain output variables. Based on these data, regression models are developed using Random Forest, Decision Tree, and Support Vector Machines. Hyperparameter optimization is conducted via Grid Search, and model evaluation is performed with metrics such as precision, accuracy, recall, and F1-Score. After model validation, a multi-objective optimization is applied to maximize the Performance Index and minimize the Material Cost. The studied fan is manufactured from AISI 310 stainless steel. The effectiveness of the approach is confirmed by additional fluid dynamic simulations with the optimal values found, where Random Forest showed the best performance with an  $R^2$  of 0.966 for the Performance Index and 0.946 for the Material Cost. The fluid dynamic simulations demonstrated a discrepancy of less than 10% compared to the model's predicted results, highlighting the accuracy of the proposed method.

**Keywords:** Computational Fluid Dynamics; Machine Learning; Multiobjective Optimization; Meta-Heuristic; Particle Swarm; Centrifugal Fan.

## 1. Introdução

Ventiladores centrífugos são dispositivos usados para movimentar ar ou outros gases em sistemas de ventilação, aquecimento e resfriamento. Eles funcionam com base no princípio da força centrífuga: o ar é aspirado axialmente para o interior da carcaça do ventilador e é expelido radialmente para fora por meio de pás rotativas. Nesse contexto, tecnologias como o Machine Learning (ML), um subcampo da Inteligência Artificial (IA), têm se mostrado promissoras, permitindo o desenvolvimento de algoritmos capazes de otimizar o desempenho desses ventiladores, aprendendo a partir de dados para realizar ajustes precisos sem a necessidade de programação explícita. A aplicação de ML em ventiladores tem sido amplamente explorada em diversos estudos, o que evidencia seu potencial para aprimorar o desempenho e a eficiência de sistemas de ventilação. Essa integração oferece novas possibilidades para o desenvolvimento de soluções inovadoras e eficientes. Exemplos disso incluem trabalhos sobre a previsão do uso de energia elétrica [1], em veículos aéreos com hélices Ducted [2], em ventiladores inteligentes controlados por voz [3], no diagnóstico de falhas de ventiladores de fluxo axial [4], na detecção de fontes de ruído em ventiladores axiais de baixa rotação [5], na otimização da velocidade de ventiladores em Data Centers [6], no monitoramento das condições de um sistema de ventilação de um radiador automotivo [7], na detecção da vibração das pás do rotor de motores da indústria aeronáutica [8], na verificação da perda combinacional de segmentação de região orientada por dados de um ventilador de camada limite [9] e na análise de falhas de ventiladores centrífugos de arrefecimento [10]. A aplicação de ML em estudos que envolvem variáveis correlacionadas é vista em Souza et al. [11], Almeida et al. [12] e Vasconcelos et al. [13]. O CFD é uma técnica que usa métodos numéricos e algoritmos para resolver e analisar problemas relacionados ao fluxo de fluidos. No estudo dos ventiladores centrífugos, o CFD é utilizado para simular como o ar se movimenta dentro dos ventiladores. A integração entre CFD e ML é abordada por Hammond et al. [14], que revisa como esses dois itens podem ser combinados para melhorar o desempenho e as análises em turbomáquinas.

Com base em numerosas simulações realizadas com o CFD, os resultados obtidos são usados para treinar modelos de regressão de ML. Esses modelos auxiliam na compreensão de padrões nos dados e são úteis para otimizações posteriores. O conjunto é dividido em três etapas: treino, teste e validação. Tal forma permite avaliar o desempenho de cada modelo e identificar qual oferece a melhor performance com o uso de métricas estatísticas.

Após isso, é realizada a otimização multiobjetivo (MO) com o uso do Multi Objective Particle Swarm Optimization (MOPSO), uma técnica de otimização meta-heurística desenvolvida por Kennedy e Eberhart

[15], baseada no comportamento social de animais, que busca encontrar a solução ótima promovendo uma população de soluções candidatas denominadas partículas em seu espaço de busca. Pereira et al. [16] realizam uma revisão de métodos de otimização multiobjetivo focados em problemas da Engenharia Mecânica. Pereira et al. [16–18] utilizam algoritmos de meta-heurística para problemas relacionados à Engenharia Mecânica.

Inicialmente, são realizadas numerosas simulações computacionais por meio do CFD com o uso do Ansys Fluent. Posteriormente, as respostas são armazenadas e os resultados são calculados em relação ao índice de desempenho (*ID*) que avalia a performance do ventilador e ao custo do material (*CM*) que calcula o valor monetário do material para a fabricação. Com essas respostas, são utilizados algoritmos para teste, treino e validação. Após isso, com o MOPSO, encontra-se a fronteira de Pareto das respostas encontradas. O uso da fronteira de Pareto para verificar as soluções ótimas é discutido por Brito et al. [19], Costa et al. [20], Lopes et al. [21], Almeida et al. [22], Bacci et al. [23] e Gaudêncio et al. [24].

O estudo de caso envolve a implementação da abordagem proposta utilizando um ventilador centrífugo, fabricado em AISI 310 e operando em altas temperaturas. A partir dos dados coletados, são calculados as variáveis de saída (ou respostas) *ID* e *CM*, que serão aplicados nos algoritmos de ML. As variáveis de entrada para o modelo incluem o número de pás ( $x_1$ ), o ângulo de entrada das pás ( $x_2$ ), o ângulo de abertura das pás ( $x_3$ ) e o comprimento das pás ( $x_4$ ). Essa abordagem permite uma compreensão mais profunda das variáveis que afetam o desempenho do ventilador e possibilita previsões que podem reduzir a necessidade de simulações futuras, simulações múltiplas e simulações por tentativa e erro. O objetivo é acelerar o processo de planejamento e otimização, além de tornar as melhorias mais precisas e quantificáveis. A utilização de meta-heurísticas para otimizar os parâmetros de entrada visa maximizar o desempenho do ventilador e, ao mesmo tempo, reduzir os custos de produção. Esta metodologia oferece uma inovação significativa em relação às abordagens tradicionais, que frequentemente dependem de simulações extensivas e métodos menos eficientes para ajustes e otimizações.

## 2. Metodologia

Para o uso de ML, é necessário que haja uma quantidade alta de dados para que os resultados sejam mais precisos. Com isso, é imprescindível ter respostas provenientes do CFD para essa realização. Diante disso, com a obtenção das respostas pelo Ansys Fluent, busca-se a construção de modelos de ML para compreender o funcionamento das variáveis de saída dos fluidos e posteriormente otimizar as condições de saída.

Inicialmente, o conjunto de dados é dividido em variáveis de entrada e variáveis de saída. Em seguida, o conjunto é particionado em treino e teste. Neste trabalho, são considerados três modelos para as duas respostas: Random Forest Regressor (RFR), Decision Tree Regressor (DTR) e Support Vector Regressor (SVR). O RFR, DTR e SVR são definidos por:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i \quad (1)$$

Onde:  $\hat{y}$  é a previsão final do RFR,  $N$  é o número total de árvores no ensemble e  $\hat{y}_i$  é a previsão da  $i$ -ésima árvore de decisão.

$$\hat{y} = \frac{1}{|R_j|} \sum_{i \in R_j} y_i \quad (2)$$

Onde:  $\hat{y}$  é a previsão para o ponto de entrada  $\mathbf{x}$ ,  $R_j$  é a região da folha onde  $\mathbf{x}$  “cai”,  $|R_j|$  é o número de pontos de treinamento na região  $R_j$ , e  $y_i$  é o valor de saída dos pontos de treinamento na região  $|R_j|$ .

$$f(\mathbf{x}) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (3)$$

Onde:  $f(\mathbf{x})$  é a previsão para o ponto de entrada  $\mathbf{x}$ ,  $N$  é o número de pontos de treinamento,  $\alpha_i, \alpha_i^*$  são os multiplicadores de Lagrange determinados durante o treinamento,  $\mathbf{x}_i$  são os vetores de suporte,  $K(\mathbf{x}_i, \mathbf{x})$  é a função kernel que mede a similaridade entre os vetores e  $b$  é o termo de intercepto.

Após o treinamento e a avaliação dos modelos, é realizada uma otimização de hiperparâmetros utilizando o método Grid Search. Hiperparâmetros são parâmetros que controlam o processo de treinamento de um modelo de ML. Eles não são ajustados automaticamente durante o treinamento, mas possuem um impacto significativo no desempenho do modelo. O Grid Search busca encontrar uma combinação ideal dos hiperparâmetros para melhorar o desempenho do modelo. Para isso, define-se uma grade de valores possíveis para cada hiperparâmetro e o modelo é treinado e avaliado para cada combinação desses valores. O objetivo desta etapa é identificar a combinação que resulta na melhor performance, com base nas métricas de avaliações escolhidas.

Os modelos são então avaliados utilizando métricas como Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) e Coefficient of Determination ( $R^2$ ), sendo selecionado o modelo com melhor desempenho para a etapa de otimização. As métricas  $R^2$ , MSE, RMSE e MAE são definidas conforme mostrado a seguir:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

Onde:  $y_i$  são os valores observados,  $\hat{y}_i$  são os valores previstos pelo modelo,  $\bar{y}$  é a média dos valores observados e  $n$  é a quantidade de observações.

Em seguida, a importância de cada variável de entrada em relação às variáveis de saída é avaliada. Variáveis de entrada com importância muito baixa para a predição das variáveis de saída são descartadas para simplificar o modelo e melhorar sua eficiência. Em seguida, é aplicada a validação cruzada com K-Fold para avaliar a capacidade de generalização do modelo de ML. Essa técnica envolve o treinamento do modelo várias vezes com diferentes conjuntos de treinamento iterativos e a coleta de métricas de avaliação para cada iteração, o que ajuda a evitar overfitting, especialmente quando há escassez de dados de treinamento.

Após a validação, é realizada a otimização multiobjetivo utilizando o algoritmo de meta-heurística MOPSO. A função objetivo é definida com base no melhor estimador de previsão encontrado para cada resposta, usando o modelo de ML mais eficiente, conforme as métricas de desempenho. Os limites inferiores e superiores das variáveis de entrada são estabelecidos de acordo com as especificações mínimas e máximas do projeto. O peso na combinação dos objetivos foi mantido de forma igualitária, utilizando uma soma ponderada com pesos iguais. Destaca-se que o estudo dos parâmetros como número de partículas, peso ideal e quantidade máxima de iterações não é um objetivo desta análise.

Esse método combina ponderadamente as variáveis de saída para identificar os melhores parâmetros para o ventilador. Os resultados obtidos são então visualizados e analisados, e um novo projeto é criado no CAD. Para validar as melhorias propostas, uma nova simulação é realizada com CFD, confirmando se as otimizações atingiram os resultados esperados.

### 3. Estudo de Caso

A abordagem é aplicada em um estudo de caso que envolve o uso de um ventilador centrífugo industrial operando em condições de altas temperaturas. O ventilador é fabricado com o aço inoxidável AISI 310, um aço austenítico de alta liga conhecido por sua excelente resistência à corrosão e oxidação em altas temperaturas. As características do aço AISI estão detalhadas na Tabela 1.

A convergência das iterações em relação a cada um dos experimentos registrados foi considerada, levando-se em conta a equação da continuidade que descreve a conservação da massa em um volume constante, além dos componentes do vetor de velocidade em um sistema tridimensional, considerando as velocidades nas direções x, y e z, e os parâmetros dos modelos de turbulência cinética ( $k$ ) e dissipativa ( $\Omega$ ).

Na geração das malhas, foi verificada a qualidade ortogonal ( $>0,15$ ), que se refere à qualidade dos

elementos da malha em relação às suas linhas ortogonais, e o skewness ( $<0,95$ ), que trata da distorção geométrica de cada elemento buscando um formato "ideal".

Tabela 1. Características do AISI 310

Característica	Valor	Unidade
Densidade	0,01	$\text{g/mm}^3$
Resistência máxima	0,66	GPa
Módulo de elasticidade	200,00	GPa
Módulo de cisalhamento	78,00	GPa
Coefficiente de Poisson	0,27	-
Condutividade térmica	14,20	$\text{W/m-k}$
Calor específico	50,00	$\text{J/kg}^\circ\text{C}$

Fonte: elaboração própria.

As variáveis de entrada foram ajustadas em cada experimento de simulação foram: a quantidade de pás, variando entre 6 e 18; o ângulo de entrada, entre  $-15,00^\circ$  e  $45,00^\circ$ ; o ângulo de abertura, entre  $-45,00^\circ$  e  $135,00^\circ$ ; e o comprimento das pás, que varia de 19,00 mm a 47,00 mm. Esses ajustes foram realizados dentro dos limites especificados para cada variável. Enquanto outras variáveis, como a espessura e a largura das pás, além do diâmetro externo da voluta, permaneceram inalteradas devido às restrições do projeto. A Figura 1 exemplifica uma das geometrias que inclui as chapas do ventilador centrífugo estudado.

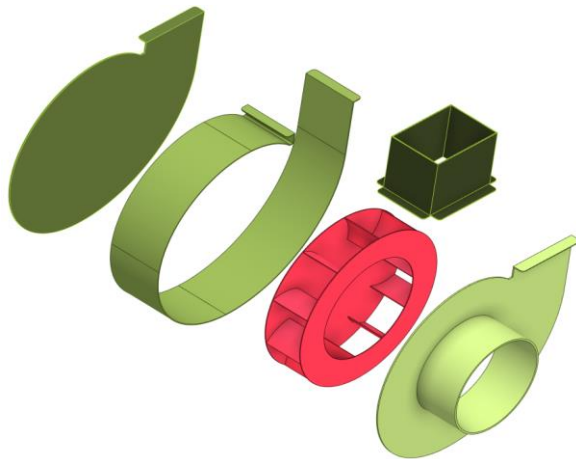


Figura 1. Geometrias e chapas que compõe o componente. Fonte: elaboração própria.

O objetivo é avaliar os resultados obtidos com as variáveis de saída de acordo com a variação das variáveis de entrada. As respostas avaliadas são o índice de desempenho ( $ID$ ), que considera outras respostas diretamente fornecidas pelo software de simulação, como massa, vazão e torque para avaliar a performance do ventilador, e o custo do material ( $CM$ ), que relaciona o valor do material à sua massa para a confecção.

O software Ansys foi utilizado para construir geometrias, realizar experimentos, gerar malhas, obter e armazenar resultados. Além disso, programação em Python foi empregada para implementar algoritmos de ML e MO.

A **Erro! Fonte de referência não encontrada.** apresenta um histograma que ilustra a distribuição dos dados em relação às respostas de  $ID$  e  $CM$ , permitindo verificar o comportamento de ambos os conjuntos de dados.

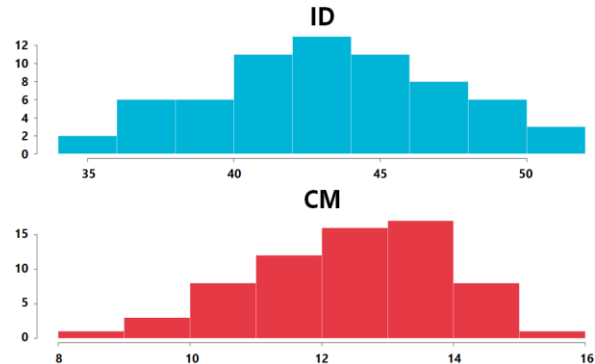


Figura 2. Histograma com os resultados do conjunto de dados em relação a  $ID$  e  $CM$ . Fonte: elaboração própria.

As variáveis de saída estão positivamente correlacionadas, com um coeficiente de correlação de 0,65. No entanto, seus objetivos de otimização são opostos: enquanto o  $ID$  visa a maximização para alcançar o melhor desempenho possível, o  $CM$  busca a minimização para reduzir ao máximo os custos com material.

O  $ID$  é uma medida adimensional que avalia o desempenho geral do ventilador, calculado conforme (8):

$$ID = 1200Q + \frac{11000}{m} + \frac{100}{T} \quad (8)$$

Onde:  $Q$  é a vazão mássica na saída do ventilador,  $m$  é a massa do ventilador,  $T$  é o torque requerido.

O cálculo do  $CM$  refere-se ao custo do material necessário para sua fabricação. O orçamento foi realizado em junho de 2024 e é dado por (9):

$$CM = \frac{31,16m}{1000} \quad (9)$$

Pelas variáveis de saída possuem objetivos de otimização opostos, há a necessidade de um trade-off entre elas, buscando encontrar um equilíbrio, pois são altamente correlacionadas, o que requer uma avaliação cuidadosa das escolhas, dado que uma decisão afeta diretamente a outra.

#### 4. Resultados

Os resultados serão apresentados em subseções, de acordo com cada uma das etapas realizadas durante toda a abordagem do estudo de caso em questão.

#### 4.1. Aplicação dos algoritmos de Machine Learning

Com o conjunto de dados experimentais, implementam-se os algoritmos de Machine Learning que são Random Forest Regressor (RFR), Decision Tree Regressor (DTR) e Support Vector Regressor (SVR) que irão trabalhar com ambas as variáveis de saída (*ID* e *CM*).

O Random Forest é um método de ensemble que usa múltiplas árvores de decisão para melhorar a robustez e a precisão das previsões, o Decision Tree é um algoritmo que divide o conjunto de dados em subconjuntos baseados em condições sobre características dos dados e o Support Vector é uma extensão do Support Vector Machine, mas para problemas de regressão, ele busca encontrar uma função que tenha no máximo um desvio das verdadeiras saídas nos pontos de treinamento.

O conjunto de dados é dividido em 80,00% para treino e 20,00% para teste, nesta etapa o conjunto é dividido de forma aleatória o que garante a replicabilidade dos resultados.

Após o treinamento dos três modelos é realizada a otimização de hiperparâmetros dos estimadores com o Grid Search, nesta etapa busca otimizar os possíveis resultados treinando com inúmeras combinações possíveis e avaliando o desempenho em relação a cada métrica.

Como nenhuma das variáveis de entrada possui a importância inferior a 5,00% para predição das variáveis de saída, todas serão mantidas. Tanto para *ID* quanto para *CM* a variável mais importante é o número de pás, representando um valor superior ao somatório de todas as outras variáveis de entrada.

Por fim, é realizada a avaliação dos modelos que é feita com base em métricas de precisão das previsões, tais como o R-quadrado ( $R^2$ ), o Mean Squared Error (MSE), o Root Mean Squared Error (RMSE) e o Mean Absolute Error (MAE). O  $R^2$  é útil para entender a proporção de variabilidade explicada pelo modelo, o MSE tem o intuito de capturar a magnitude dos erros, o RMSE por sua vez garante uma interpretação direta, enquanto o MAE é menos sensível a outliers e possui fácil interpretabilidade. Os resultados das métricas de avaliação de *ID* são dadas pela Tabela 2, enquanto de *CM* são dadas pela Tabela 3.

Tabela 2. Métricas de regressão para *ID*

	RFR	DTR	SVR
$R^2$	0,966	0,963	0,795
MSE	0,533	0,589	3,243
RMSE	0,730	0,768	1,801
MAE	0,554	0,694	0,978

Fonte: elaboração própria.

Tabela 3. Métricas de regressão para *CM*

	RFR	DTR	SVR
$R^2$	0,946	0,939	0,782
MSE	0,141	0,158	0,563
RMSE	0,375	0,398	0,750
MAE	0,141	0,278	0,346

Fonte: elaboração própria.

A Figura 3 e Figura 4 plotam o comparativo entre os valores reais e previstos de *ID* e *CM*, respectivamente, nelas são apresentados os resultados do Random Forest, o modelo que obteve melhor desempenho entre os três modelos para ambas as variáveis de saída. Essa análise visual permite avaliar a precisão das previsões em relação aos resultados reais. Quanto mais próximos os pontos estão da linha de referência, melhor é a precisão dos resultados.

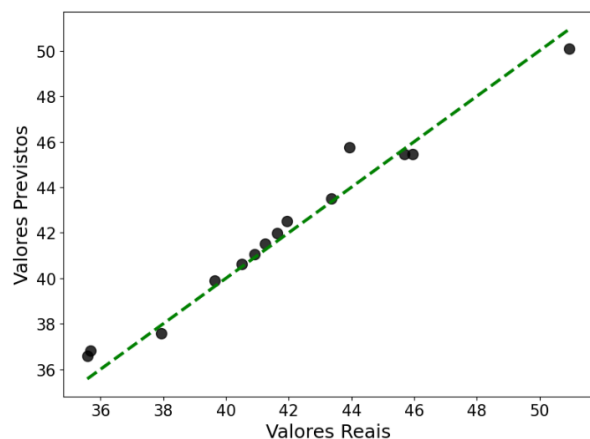


Figura 3. Valores reais e previstos com o Random Forest para *ID*. Fonte: elaboração própria.

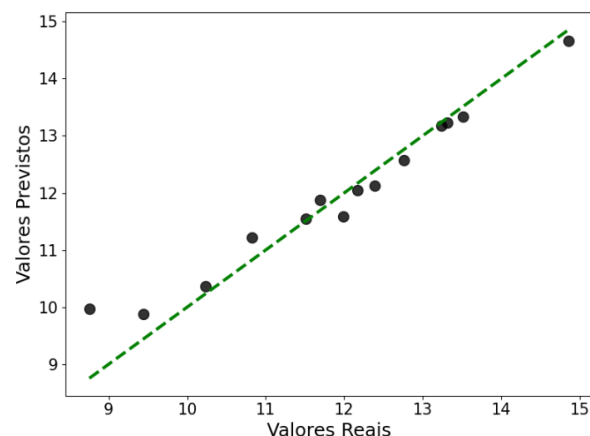


Figura 4. Valores reais e previstos com o Random Forest para *CM*. Fonte: elaboração própria.

O valor do  $R^2$  para ambas as variáveis de saída foi superior a 90,00%, o que indica uma boa capacidade do modelo para realizar previsões precisas.

Na Figura 5 e Figura 6 são exibidas as curvas de aprendizado para  $ID$  e  $CM$ , respectivamente. Essas curvas ilustram a melhoria no desempenho do modelo Random Forest à medida que ele é treinado com um número crescente de dados. A linha vermelha indica o Score de Treinamento, que avalia o desempenho do modelo no conjunto de treinamento para cada valor do hiperparâmetro, refletindo como o ajuste do modelo varia com a quantidade de dados. Já a linha verde representa o Score de Validação Cruzada, que mede o desempenho do modelo em dados não vistos para cada valor do hiperparâmetro, permitindo avaliar a capacidade de generalização do modelo.

Com o aumento no número de exemplos de treinamento, é possível observar um crescimento no valor do Score, indicando uma melhoria no desempenho do modelo.

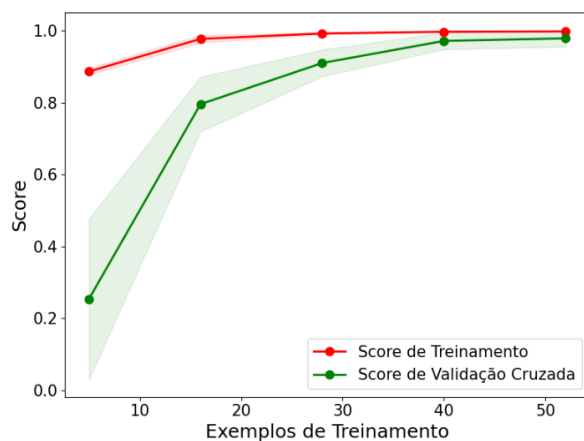


Figura 5. Curva de aprendizado com o Random Forest para  $ID$ . Fonte: elaboração própria.

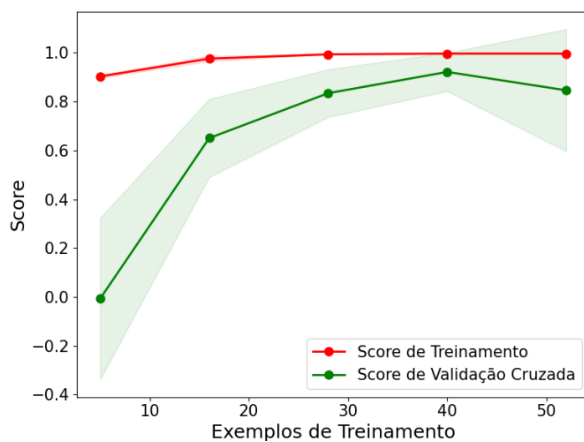


Figura 6. Curva de aprendizado com o Random Forest para  $CM$ . Fonte: elaboração própria.

#### 4.2. Otimização meta-heurística

A otimização meta-heurística com o MOPSO é eficaz para encontrar soluções ótimas em espaços de busca complexos e não-lineares. Inicialmente, definem-se os limites das variáveis de entrada conforme estabelecido no projeto, onde os parâmetros das pás não podem ultrapassar essas definições.

A variável  $ID$  deve ser maximizada, enquanto o  $CM$  deve ser minimizado. Para alcançar esses objetivos, o algoritmo MOPSO ajusta as partículas com base em três componentes: inércia, componente cognitivo e componente social. Os valores para a variável de entrada para  $x_1$  são discretos, enquanto para  $x_2$ , para  $x_3$  e para  $x_4$  são contínuas. No entanto, para melhor visualização, todos os valores são apresentados com até duas casas decimais.

O processo de otimização é realizado com um número máximo de 200 iterações e um tamanho de enxame de 200 partículas. Além disso, as funções-objetivo  $ID$  e  $CM$  têm pesos iguais, cada uma contribuindo com 50% da importância na otimização.

Com essa abordagem, é possível encontrar os valores ótimos para ambas as respostas: um  $ID$  de 509,57 e um  $CM$  de 14,80. Os parâmetros de entrada que levaram a esses resultados são 18 pás, ângulo de entrada de  $-2,60^\circ$ , ângulo de abertura de  $73,90^\circ$  e comprimento das pás de 47,00 mm.

Para visualizar como as funções-objetivo variam em relação às variáveis de entrada, são plotados gráficos de superfície. A Figura 7 ilustra o índice de desempenho em função do número de pás e do ângulo de abertura, enquanto a Figura 8 mostra o custo do material em relação ao número de pás e ao comprimento das pás. Esses gráficos proporcionam uma visualização clara do comportamento das variáveis.

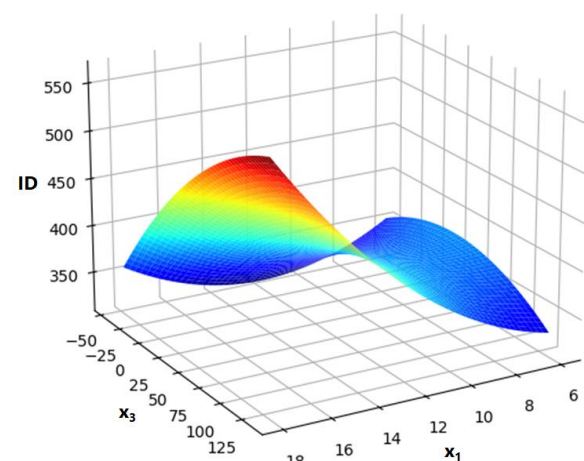


Figura 7. Gráfico de superfície para  $ID$  em relação a  $x_1$  e  $x_3$ . Fonte: elaboração própria.



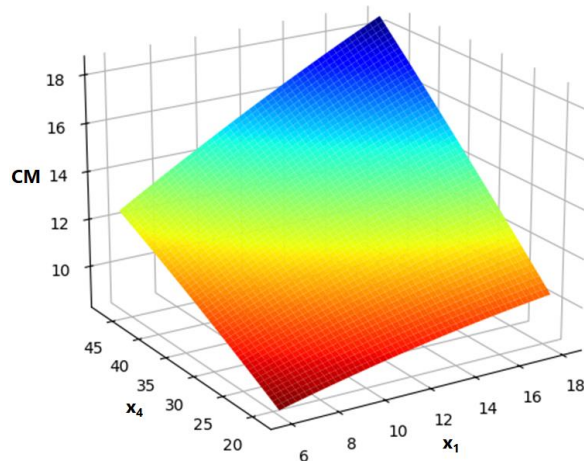


Figura 8. Gráfico de superfície para  $CM$  em relação a  $x_1$  e  $x_4$ . Fonte: elaboração própria.

#### 4.3. Validação dos resultados

Com a obtenção do resultado ótimo proveniente de todo o processo de ML seguido da meta-heurística, realiza-se uma nova configuração geométrica para verificar e validar os resultados obtidos com o MOPSO. Todo o processo de desenho da geometria, geração das malhas, configuração da simulação, simulação e obtenção de resultados é realizado novamente para validar a precisão do modelo obtido. Quanto maior o conjunto de dados, maior será a assertividade do modelo, provavelmente. Vale ressaltar também que se a convergência fosse alterada e o número de iterações fosse maior, o resultado obtido poderia ser diferente.

A Figura 9 exibe o resultado com o gráfico de contorno da magnitude de velocidade, a Figura 10 de pressão total, e Figura 11 de turbulência cinética  $k$ , todos os resultados referentes aos contornos são provenientes do Ansys Fluent. Ao avaliar esses gráficos, quanto mais próximo do vermelho, maior é o valor da escala; por outro lado, quanto mais próximo do azul, menor é esse valor, seguindo a mesma lógica dos gráficos de superfície ilustrados anteriormente.

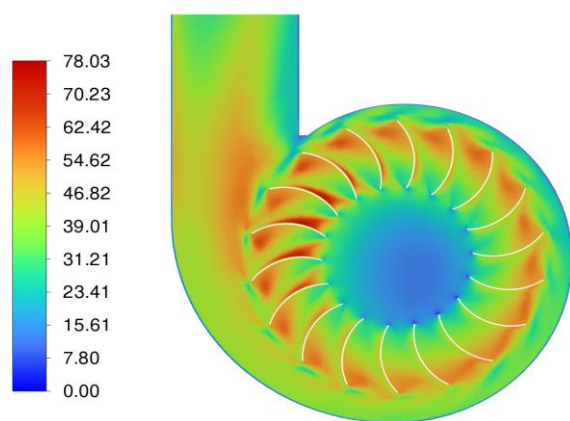


Figura 9. Gráfico de contorno em relação a magnitude de velocidade em m/s. Fonte: elaboração própria.

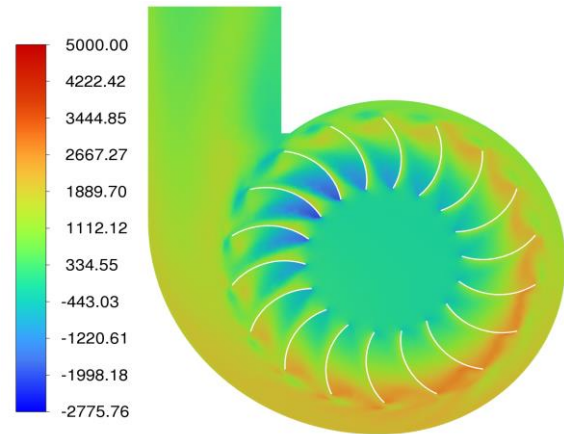


Figura 10. Gráfico de contorno em relação a pressão total em Pa. Fonte: elaboração própria.

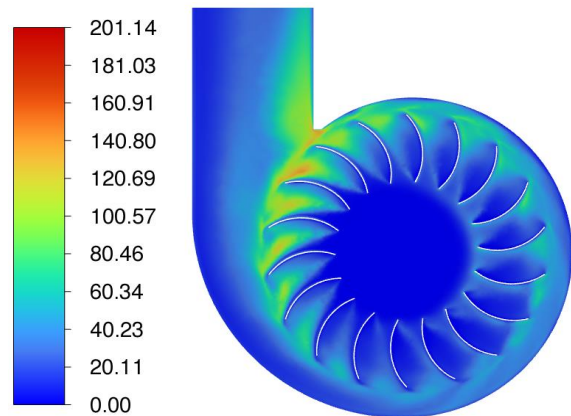


Figura 11. Gráfico de contorno em relação a turbulência cinética  $k$  em  $m^2/s^2$ . Fonte: elaboração própria.

Com o uso de técnicas como o ML, é possível economizar recursos computacionais, tempo de simulação (ou tempo de confecção de protótipo) e melhorar a compreensão do modelo.

Destaca-se que, devido ao conflito entre as variáveis acopladas, não é possível aumentar alguns resultados sem impactar o outro, o que representa um desafio para a MO.

Por fim, a validação dos resultados com o uso do CFD permite confirmar o funcionamento da abordagem. A diferença nos resultados obtidos com a abordagem que combina ML com MOPSO em comparação com os resultados do CFD foi inferior a 10,00. Isso indica que, sem o uso de numerosas simulações por tentativa e erro ou métodos baseados em simulações isoladas, há um custo menor em termos de tempo, custo computacional e custo de mão de obra.

Considerando esses aspectos, o método demonstra eficiência na realização de previsões e otimizações de forma eficaz.

## 5. Conclusões

A integração de Machine Learning (ML) e Computational Fluid Dynamics (CFD) em estudos de ventiladores centrífugos pode representar um avanço na otimização de sistemas complexos. O uso de modelos como Random Forest Regressor (RFR), Decision Tree Regressor (DTR) e Support Vector Regressor (SVR) mostrou a possibilidade de prever o desempenho de ventiladores com base em outras simulações, sem a necessidade de realizar novas, além de possibilitar a otimização das configurações do projeto de forma eficiente.

Os resultados obtidos demonstram que os modelos de ML, treinados com dados de CFD, possuem a capacidade de capturar e compreender as relações entre as variáveis de entrada (como número de pás, ângulo de entrada e comprimento das pás) e as variáveis de saída (como índice de desempenho e custo do material). Após o treinamento, os modelos foram avaliados utilizando métricas como Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) e Coefficient of Determination ( $R^2$ ), com o propósito de verificar quais modelos apresentam melhor desempenho. O RFR destacou-se em relação aos outros, com  $R^2$  de 0,966 e 0,946 em relação às variáveis de saída, sendo, portanto, o modelo escolhido para a etapa de otimização multiobjetivo (MO).

A aplicação do Multi Objective Particle Swarm Optimization (MOPSO) permitiu encontrar soluções ótimas que buscam o equilíbrio entre a maximização do índice de desempenho ( $ID$ ) do ventilador e a minimização do custo do material ( $CM$ ). Esse processo visa reduzir o tempo e os custos computacionais, além de melhorar a exploração do ótimo.

Foi utilizada a validação cruzada (K-Fold) para garantir a robustez dos modelos de ML contra overfitting, assegurando que eles fossem capazes de generalizar bem para novos dados.

Os valores ótimos obtidos com o MOPSO foram 509,57 para o  $ID$  e 14,80 para o  $CM$ . Para alcançar esses valores, a configuração do ventilador, dada pelas variáveis de entrada, deve ser de 18 pás, 2,60° de ângulo de entrada, 73,90° de ângulo de abertura e 47,00 mm de comprimento das pás.

A combinação de CFD, ML e MO proporciona uma abordagem significativa para resolver problemas complexos relacionados a fluidos, com precisão, eficiência e economia de tempo.

Em pesquisas futuras, serão realizadas validações experimentais para garantir a aplicabilidade e confiabilidade das técnicas propostas, além de trabalhar com respostas voltadas para a redução de energia, visando criar projetos sustentáveis.

## 6. Agradecimentos

Agradecimentos são expressos à CAPES, CNPq e FAPEMIG pelo suporte fornecido a este trabalho. Esta pesquisa também foi viabilizada pelo apoio da

NOMATI-UNIFEI, que concedeu acesso aos seus laboratórios, materiais e expertise.

## 7. Referências

- [1] J. Runge, R. Zmeureanu, M. Le Cam, Hybrid short-term forecasting of the electric demand of supply fans using machine learning, *J. Build. Eng.* 29 (2020) 101144. <https://doi.org/10.1016/j.jobte.2019.101144>.
- [2] T. Manzoor, H. Pei, Z. Sun, Z. Cheng, Model Predictive Control Technique for Ducted Fan Aerial Vehicles Using Physics-Informed Machine Learning, *Drones* 7 (2023) 4. <https://doi.org/10.3390/drones7010004>.
- [3] J. Li, Y. Cheng, Design and implementation of voice-controlled intelligent fan system based on machine learning, *Proc. 2020 IEEE Int. Conf. Adv. Electr. Eng. Comput. Appl. AECA 2020* (2020) 548–552. <https://doi.org/10.1109/AECA49918.2020.9213552>.
- [4] P. Tao, J. Liu, T. Liang, Research on Fault Diagnosis Method of Axial Flow Induced Draft Fan of Power Plant Based on Machine Learning, *2019 4th Int. Conf. Syst. Reliab. Safety, ICSRS 2019* (2019) 325–330. <https://doi.org/10.1109/ICSRS48664.2019.8987662>.
- [5] L. Tieghi, S. Becker, A. Corsini, G. Delibra, S. Schoder, F. Czwielong, Machine-Learning Clustering Methods Applied to Detection of Noise Sources in Low-Speed Axial Fan, *J. Eng. Gas Turbines Power* 145 (2023) 031020. <https://doi.org/10.1115/1.4055417>.
- [6] R. Kumar, S.K. Khatri, M.J. Divan, Data Center Air Handling Unit Fan Speed Optimization Using Machine Learning Techniques, *2021 9th Int. Conf. Reliab. Infocom Technol. Optim. (Trends Futur. Dir. ICRITO 2021* (n.d.) Noida. <https://doi.org/10.1109/ICRITO51393.2021.9596311>.
- [7] R. Meena, B.B. Nair, N.R. Sakthivel, Machine learning approach to condition monitoring of an automotive radiator cooling fan system, *Lect. Notes Electr. Eng.* 672 (2020) 1007–1020. [https://doi.org/10.1007/978-981-15-5558-9\\_85](https://doi.org/10.1007/978-981-15-5558-9_85).
- [8] A.N. Viswanatha Rao, T.N. Satish, V.P.S. Naidu, S. Jana, Machine learning augmented multi-sensor data fusion to detect aero engine fan rotor blade flutter, *Int. J. Turbo Jet Engines* 40 (2023) S485–S506. <https://doi.org/10.1515/tjj-2022-0066>.
- [9] T. Pan, K. Shi, H. Lu, Z. Yang, J. Zhang, Q. Li, A region-segmentation combinational loss model based on data-driven machine learning for a boundary layer ingestion fan, *Aerosp. Sci. Technol.* 144 (2024) 108780.



- <https://doi.org/10.1016/j.ast.2023.108780>.
- [10] D.P. Guillen, N. Anderson, C. Krome, R. Boza, L.M. Griffel, J. Zouabe, A.Y. Al Rashdan, A RELAP5-3D/LSTM model for the analysis of drywell cooling fan failure, *Prog. Nucl. Energy* 130 (2020) 103540. <https://doi.org/10.1016/j.pnucene.2020.103540>.
- [11] L.G.P. de Souza, G.A.V.B. Vasconcelos, L.A.R. Costa, M.B. Francisco, A.P. de Paiva, J.R. Ferreira, Roughness prediction using machine learning models in hard turning: an approach to avoid rework and scrap, *Int. J. Adv. Manuf. Technol.* (2024). <https://doi.org/10.1007/s00170-024-13951-8>.
- [12] F.A. de Almeida, E.L. Romão, G.F. Gomes, J.H. de F. Gomes, A.P. de Paiva, J. Miranda Filho, P.P. Balestrassi, Combining machine learning techniques with Kappa–Kendall indexes for robust hard-cluster assessment in substation pattern recognition, *Electr. Power Syst. Res.* 206 (2022) 107778. <https://doi.org/10.1016/j.epsr.2022.107778>.
- [13] G.A.V.B. Vasconcelos, M.B. Francisco, L.R.A. da Costa, R.F. Ribeiro Junior, M. de L.N.M. Melo, Prediction of surface roughness in duplex stainless steel face milling using artificial neural network, *Int. J. Adv. Manuf. Technol.* (2024). <https://doi.org/10.1007/s00170-024-13955-4>.
- [14] J. Hammond, N. Pepper, F. Montomoli, V. Michelassi, Machine Learning Methods in CFD for Turbomachinery: A Review, *Int. J. Turbomachinery, Propuls. Power* 7 (2022) 16. <https://doi.org/10.3390/ijtp7020016>.
- [15] J. Kennedy, R. Eberhart, Particle swarm optimization, *Proc. ICNN'95 - Int. Conf. Neural Networks* 4 (n.d.) 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>.
- [16] J. Luiz Junho Pereira, G. Antônio Oliver, M. Brendon Francisco, S. Simões Cunha, G. Ferreira Gomes, Multi-objective lichtenberg algorithm: A hybrid physics-based meta-heuristic for solving engineering problems, *Expert Syst. Appl.* 187 (2022) 115939. <https://doi.org/10.1016/j.eswa.2021.115939>.
- [17] J.L.J. Pereira, M.B. Francisco, C.A. Diniz, G. Antônio Oliver, S.S. Cunha, G.F. Gomes, Lichtenberg algorithm: A novel hybrid physics-based meta-heuristic for global optimization, *Expert Syst. Appl.* 170 (2021) 114522. <https://doi.org/10.1016/j.eswa.2020.114522>.
- [18] J.L.J. Pereira, M.B. Francisco, F.A. de Almeida, B.J. Ma, S.S. Cunha, G.F. Gomes, Enhanced Lichtenberg algorithm: a discussion on improving meta-heuristics, *Soft Comput.* 27 (2023) 15619–15647. <https://doi.org/10.1007/s00500-023-08782-w>.
- [19] T.G. Brito, A.P. Paiva, J.R. Ferreira, J.H.F. Gomes, P.P. Balestrassi, A normal boundary intersection approach to multiresponse robust optimization of the surface roughness in end milling process with combined arrays, *Precis. Eng.* 38 (2014) 628–638. <https://doi.org/10.1016/j.precisioneng.2014.02.013>.
- [20] D.M.D. Costa, T.I. Paula, P.A.P. Silva, A.P. Paiva, Normal boundary intersection method based on principal components and Taguchi's signal-to-noise ratio applied to the multiobjective optimization of 12L14 free machining steel turning process, *Int. J. Adv. Manuf. Technol.* 87 (2016) 825–834. <https://doi.org/10.1007/s00170-016-8478-7>.
- [21] L.G.D. Lopes, T.G. Brito, A.P. Paiva, R.S. Peruchi, P.P. Balestrassi, Robust parameter optimization based on multivariate normal boundary intersection, *Comput. Ind. Eng.* 93 (2016) 55–66. <https://doi.org/10.1016/j.cie.2015.12.023>.
- [22] F.A. Almeida, A.C.O. Santos, A.P. de Paiva, G.F. Gomes, J.H. de F. Gomes, Multivariate Taguchi loss function optimization based on principal components analysis and normal boundary intersection, *Eng. Comput.* 38 (2022) 1627–1643. <https://doi.org/10.1007/s00366-020-01122-8>.
- [23] L.A. Bacci, L.G. Mello, T. Incerti, A. Paulo de Paiva, P.P. Balestrassi, Optimization of combined time series methods to forecast the demand for coffee in Brazil: A new approach using Normal Boundary Intersection coupled with mixture designs of experiments and rotated factor scores, *Int. J. Prod. Econ.* 212 (2019) 186–211. <https://doi.org/10.1016/j.ijpe.2019.03.001>.
- [24] J.H.D. Gaudêncio, F.A. de Almeida, R.C. Sabioni, J.B. Turrioni, A.P. de Paiva, P.H. da Silva Campos, Fuzzy multivariate mean square error in equispaced pareto frontiers considering manufacturing process optimization problems, *Eng. Comput.* 35 (2019) 1213–1236. <https://doi.org/10.1007/s00366-018-0660-0>.