

## **Diagnóstico On-Line de Aerogeradores com Uso de Técnicas de Data Science e Inteligência Artificial**

**Autores:** Daniel Carrijo Polonio Araujo, Murilo Marques Pinto, Guilherme Toyoshima, Marcos Alves, Gabriel de Souza Pereira Gomes, Rafael Prux Fehlberg, Rogério Andrade Flauzino, Renan Ferreira Santa Rosa, Iony Patriota Siqueira, Bruno Sardinha, Catia Pedrosa, Daniel Pedrosa Santos, Dara Mendes, Gilberto Amorim Moura, Heber Pedrosa, Márcio Costa

Treetech Tecnologia, Radice Tecnologia, Ibitu Energia, Radice Tecnologia, Radice Tecnologia, Treetech Tecnologia, Radice Tecnologia, Treetech Tecnologia, Radice Tecnologia, Radice Tecnologia, Radice Tecnologia, Radice Tecnologia, Radice Tecnologia, Radice Tecnologia, Radice Tecnologia, Radice Tecnologia, Radice Tecnologia

daniel.carrijo@treetech.com.br

### **RESUMO**

Este trabalho apresenta o desenvolvendo de uma metodologia para monitorar e diagnosticar os principais modos de falha de turbinas eólicas por meio do uso do histórico de manutenção e de dados dos sensores das máquinas, muitas vezes já existentes e conectados no sistema SCADA da usina, associados a técnicas estatísticas avançadas e inteligência artificial. A metodologia foi desenvolvida em uma plataforma de gestão de ativos, um ambiente que permite a integração nativa de algoritmos inteligentes com dados de ordem de manutenção, registro de avarias, dados de cadastros dos equipamentos e dados de sensores em tempo real. A metodologia visa apoiar as equipes de engenharia de manutenção e gestão de ativos no aumento da segurança e confiabilidade dos parques, na otimização dos recursos de manutenção e na redução da perda de receita por indisponibilidade das máquinas. Para avaliar a eficácia do sistema desenvolvido, está sendo realizada uma aplicação piloto da metodologia e do software inteligente em 210 aerogeradores de dois fabricantes diferentes.

### **PALAVRAS-CHAVE**

Aprendizado de Máquina, Turbinas Eólicas, Detecção de Falhas, Diagnóstico de Falhas, Inteligência Artificial

### **ABSTRACT**

This work presents the development of a methodology to monitor and diagnose the primary failure modes of wind turbines through the use of maintenance history and machine sensor data, often already existing and connected to the SCADA system of the plant, associated with advanced statistical techniques and artificial intelligence. The methodology was developed on an asset management platform, an environment that allows native integration of intelligent algorithms with maintenance order data, fault records, equipment registration data, and real-time sensor data. The methodology aims to support maintenance engineering teams and asset management in increasing the safety and reliability of the parks, optimizing maintenance resources, and reducing revenue loss due to machine unavailability. To evaluate the effectiveness of the developed system, a pilot application of the methodology and the intelligent software is being carried out on 210 wind turbines from two different manufacturers.

### **KEYWORDS**

Machine Learning - Wind Turbines - Fault Detection - Fault Diagnosis - Artificial Intelligence

## 1. INTRODUÇÃO

De acordo com o Relatório Internacional de Energias Renováveis da IRENA, em 2020, as turbinas eólicas foram a terceira tecnologia renovável em capacidade instalada e a segunda em geração de energia, sendo responsáveis por quase 20% de toda a produção de energia renovável no mundo [1], conforme [Figura 1](#).

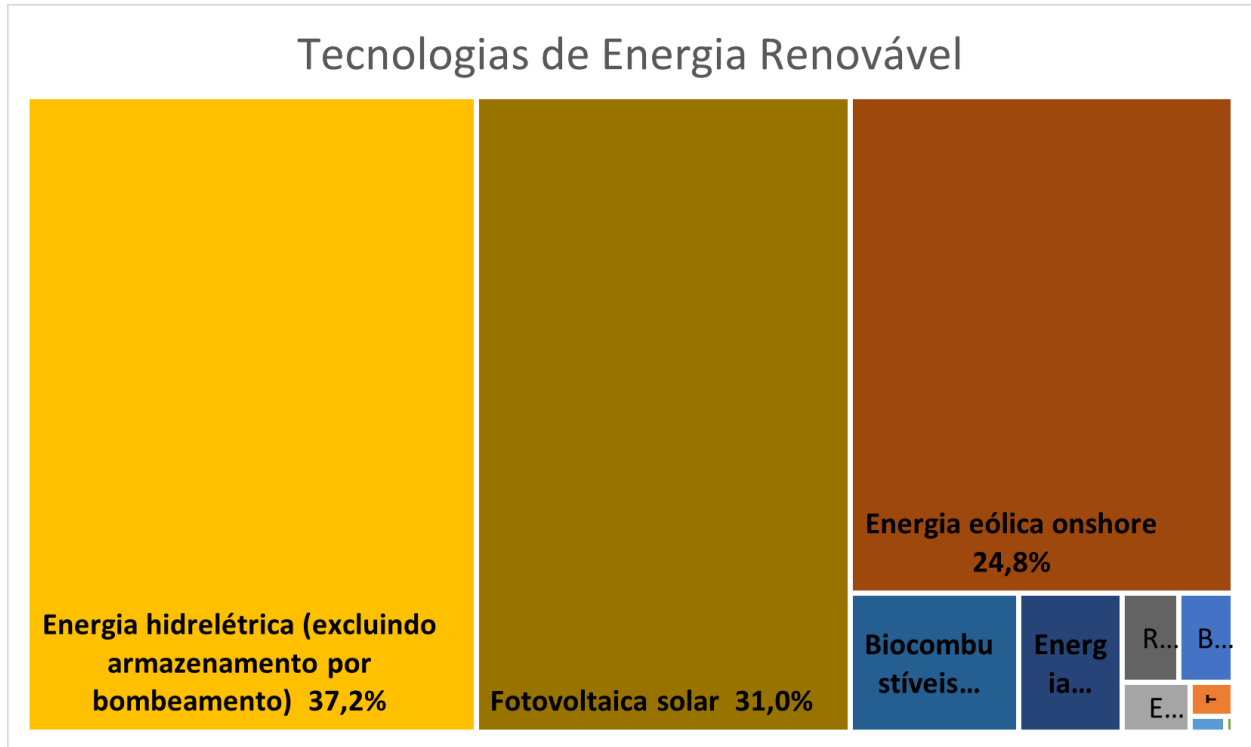


Figura 1: Tecnologias de Energia Renovável, adaptado de [1]

O aumento da geração de energia eólica na matriz energética mundial está relacionado às preocupações ambientais que se concentram na substituição de fontes de energia tradicionais, baseadas em combustíveis fósseis, por fontes de energia renovável e limpa [2].

As turbinas eólicas terrestres e marítimas operam em ambientes extremos, sujeitas a condições climáticas adversas, como luz solar excessiva, areia, chuva, vento e descargas atmosféricas. Essas condições ambientais podem levar a falhas operacionais devido a tensões mecânicas ou elétricas. Em condições críticas, essa situação pode evoluir para interrupção no fornecimento de energia, causando perdas financeiras [3]. A [Figura 2](#) mostra exemplo de aerogeradores terrestres (*onshore*) e marítimos (*offshore*).

Esses problemas aumentam os custos de operação e manutenção (O&M) de parques eólicos, o que pode torná-los menos atraentes financeiramente. Outras partes da usina, como a SE Elevadora, contribuem para esse custo e vem sendo objeto intenso de estudos [4]. Especificamente para as turbinas terrestres, o O&M é responsável por quase 15% do custo de energia, enquanto para as turbinas marítimas, esses valores podem chegar a quase 25% [5]. Além disso, a manutenção dessas turbinas, que muitas vezes estão localizadas em áreas remotas e de difícil acesso, envolve a coordenação complexa de mão de obra e transporte de peças de reposição, com o objetivo de minimizar o tempo de inatividade das turbinas.

Dessa forma, garantir o bom funcionamento desses equipamentos é fundamental para o setor elétrico. Portanto, é cada vez mais necessário o desenvolvimento de novas técnicas e ferramentas para apoiar as equipes de manutenção e gestão de ativos na identificação de defeitos em máquinas rotativas, que ainda se encontram em estágio incipiente, capacitando essas equipes para um melhor suporte à tomada de decisão, de forma a atender aos objetivos organizacionais da concessionária de forma mais eficiente.

A fim de atender a esses desafios e necessidades, este trabalho desenvolveu uma metodologia para monitorar e diagnosticar os principais modos de falha de turbinas eólicas por meio do uso do histórico de manutenção e de dados dos sensores das máquinas, muitas vezes já existentes e conectados no sistema SCADA da usina, associados a técnicas estatísticas avançadas e inteligência artificial. A metodologia foi

desenvolvida em uma plataforma de gestão de ativos, um ambiente que permite a integração nativa de algoritmos inteligentes com dados de ordem de manutenção, registro de avarias, informações de cadastros dos equipamentos e dados de sensores em tempo real.



Figura 2: Aplicações de Aerogeradores em terra e mar

A metodologia visa apoiar as equipes de engenharia de manutenção e gestão de ativos da concessionária de geração no aumento da segurança e confiabilidade dos parques, na otimização dos recursos de manutenção e na redução da perda de receita por indisponibilidade das máquinas.

Para avaliar a eficácia do sistema desenvolvido, está sendo realizada uma aplicação piloto da metodologia e do software inteligente em 210 aerogeradores de dois fabricantes diferentes.

O trabalho foi desenvolvido no âmbito do Projeto de P&D 0119 da ANEEL, intitulado Sistema Inteligente para Gestão Otimizada de Aerogeradores e Hidrogeradores.

## 2. DESCRIÇÃO DOS DADOS

O conjunto de dados usado neste trabalho foi coletado nos anos de 2017 e 2018 em duas usinas de energia eólica (usina A e usina B). As informações utilizadas para compor o conjunto de dados foram dados do sistema SCADA e dados do histórico de manutenção. Em relação às informações coletadas do sistema SCADA, inicialmente foram consideradas 609 características para cada turbina.

As informações presentes no sistema SCADA foram obtidas por vários sensores instalados nas turbinas eólicas. Esses sensores transmitem informações para o sistema SCADA, que registra esses dados usando um intervalo de tempo de 10 minutos. Para cada medição coletada pelos sensores, considerando o intervalo de 10 minutos, quatro parâmetros são calculados: valor máximo e mínimo, valor médio e desvio padrão. Considerando uma abordagem de aprendizado supervisionado, os dados do sistema SCADA foram usados para compor a entrada para o método.

Os dados históricos contendo informações sobre falhas nas turbinas e eventos de indisponibilidade foram usados para construir o conjunto de dados de destino (saída desejada) para a fase de treinamento. Os dados históricos de manutenção indicam a máquina que sofreu o defeito, a data da interrupção, a duração da interrupção e o sistema afetado. Apenas as falhas que causaram períodos de interrupção de fornecimento superiores a um dia foram consideradas, pois interrupções mais curtas foram resolvidas reiniciando a máquina, não exigindo intervenção da equipe de manutenção. A [Figura 3](#) apresenta um histograma com a quantidade de falhas em relação ao local.

Para compor esses dados históricos, foram consideradas as seguintes informações: identificação do parque eólico, índice da turbina eólica, dados de manutenção (tipo de manutenção, ou seja, corretiva não programada, corretiva programada, solicitada pelo cliente, inspeção com restrição, preditiva e preventiva), tipo de falha, tempo de parada, tempo de retorno, duração da falha, atividades de manutenção e sistemas afetados.

A partir dos dados do histórico de manutenção, verificou-se que a maioria das falhas estava relacionada a eventos de manutenção não programados. A porcentagem de cada tipo de manutenção é apresentada na [Tabela 1](#).

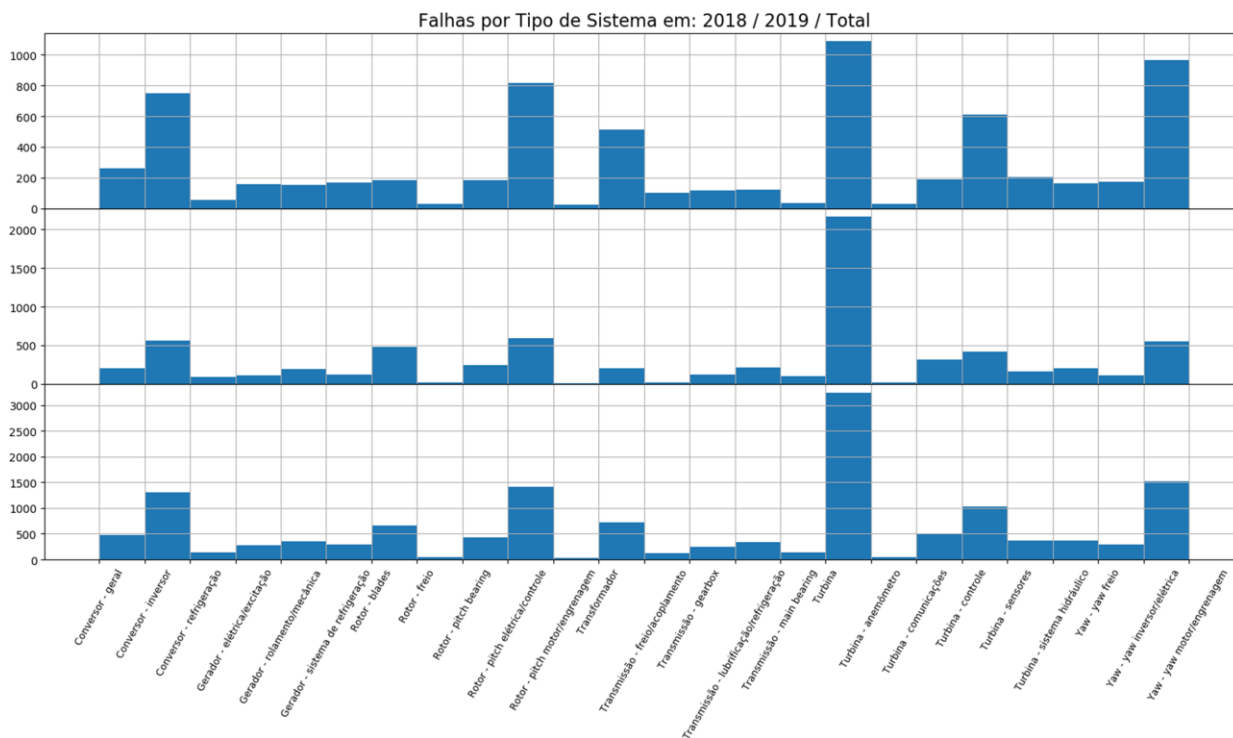


Figura 3: Quantidade de Falhas por Tipo de Sistema

Tabela 1: Tipo de Manutenção

Tipo de Manutenção	Frequência
Corretiva não programada	80,1%
Preventiva	13,7%
Solicitada pelo cliente	3,0%
Corretiva programada	2,87%
Inspeção com restrição	0,235%
Preditiva	0,0281%

O método proposto neste artigo tem como objetivo evitar a ocorrência de eventos de manutenção não programados, que estão relacionados à maioria das falhas e representam os maiores custos de manutenção. Portanto, apenas as amostras relacionadas aos eventos de manutenção não programados foram utilizadas.

Considerando apenas esse tipo de dados, verificou-se que a usina B foi responsável por 79,7% das ocorrências de falhas, enquanto a usina A foi responsável por 20,3%. Essas diferenças podem estar relacionadas a vários fatores, como condições ambientais e diferentes protocolos de manutenção para cada usina. É necessária uma investigação adicional para identificar o motivo dessa discrepância. Essa investigação está fora do escopo deste artigo.

Considerando a porcentagem de falhas por subsistema, a [Tabela 2](#) mostra os resultados para cada usina.

A proporção de falhas por sistema é equivalente para ambas as usinas, por exemplo, o sistema do rotor é responsável pela maioria das falhas, seguido pelo sistema de transmissão e pelo sistema do gerador.

O tempo de inatividade por falha apresenta um comportamento diferente, como mostrado nas duas últimas colunas da [Tabela 2](#). Para a usina A, o sistema do conversor é responsável pelo período mais longo de tempo de inatividade, seguido pelo sistema do rotor. Enquanto na usina B, a classe da turbina é responsável

Tabela 2: Porcentagem de falhas e tempo de inatividade por subsistema

Subsistema	Percentual de falhas		Tempo de inatividade	
	Usina A	Usina B	Usina A	Usina B
Rotor	28,4%	35%	31,7%	20,9%
Transmissão	23,1%	19,2%	5,99%	17,5%
Gerador	17,2%	17,1%	13,3%	14,6%
Conversor	17,1%	12,9%	36,2%	13,2%
Turbina	8,1%	10,4%	10,1%	31,7%
Giro	6,04%	4,19%	2,72%	1,64%
Caixa de engrenagens	-	0,0164%	-	0,257%
Turbina elétrica	-	0,0491%	-	0,0401%
Sistema de controle	-	0,0819%	-	0,03%
Sistema do gerador	-	0,0982%	-	0,0114%
Sistema de azimute	-	0,115%	-	0,0101%
Kabelbruch	-	0,18%	-	0,00563%

A classe definida como falhas "turbina" está relacionada a falhas que não puderam ser identificadas pela equipe de manutenção.

Como pode ser visto na [Tabela 2](#), o rotor foi o sistema mais propenso a falhas e estava entre os sistemas que levaram a um maior tempo de inatividade operacional. Essas características foram consideradas ao selecionar o sistema como foco deste artigo. No entanto, o modelo pode ser generalizado e adaptado a outros sistemas em turbinas eólicas.

## 2.1. CONJUNTO DE DADOS DE SAÍDA DESEJADO

Como mencionado anteriormente, os dados históricos de manutenção foram utilizados para construir o conjunto de dados de saída desejado. Os dados históricos de manutenção apresentam a data em que a falha ocorreu e essa informação foi usada para criar um conjunto de dados no qual a saída desejada era o tempo restante até a falha.

Considerando que os dados do SCADA apresentam uma resolução de 10 minutos, o conjunto de dados de entrada também terá amostras com resolução de 10 minutos. Portanto, o tempo restante até a falha foi obtido para cada uma das amostras.

Por exemplo, se uma falha ocorreu no instante  $t$ , para cada amostra, o tempo restante até o instante  $t$  foi utilizado como saída desejada.

## 3. PIPELINE DE DETECÇÃO DE FALHAS

O pipeline de detecção de falhas é descrito na [Figura 4](#) e consiste no estágio de processamento dos dados, seguido de engenharia de features, seleção de modelo e avaliação dos resultados.

### 3.1. PROCESSAMENTO DOS DADOS

O primeiro passo para implementar um *framework* de aprendizado de máquina é pré-processar o conjunto de dados, removendo erros e analisando se há alguma transformação necessária nos dados. O conjunto de dados foi dividido em três partes: conjunto de treinamento, conjunto de teste e conjunto de validação. Essa divisão é essencial para abordagens de aprendizado de máquina.

O conjunto de treinamento geralmente é o maior e é apresentado ao modelo na fase de aprendizado. Portanto, o modelo aprenderá a generalizar com base nesses dados. O conjunto de validação é usado para verificar o desempenho do modelo em novos dados que ainda não foram apresentados a ele. Se o modelo tiver um desempenho ruim na fase de validação, seus parâmetros devem ser ajustados por meio de um

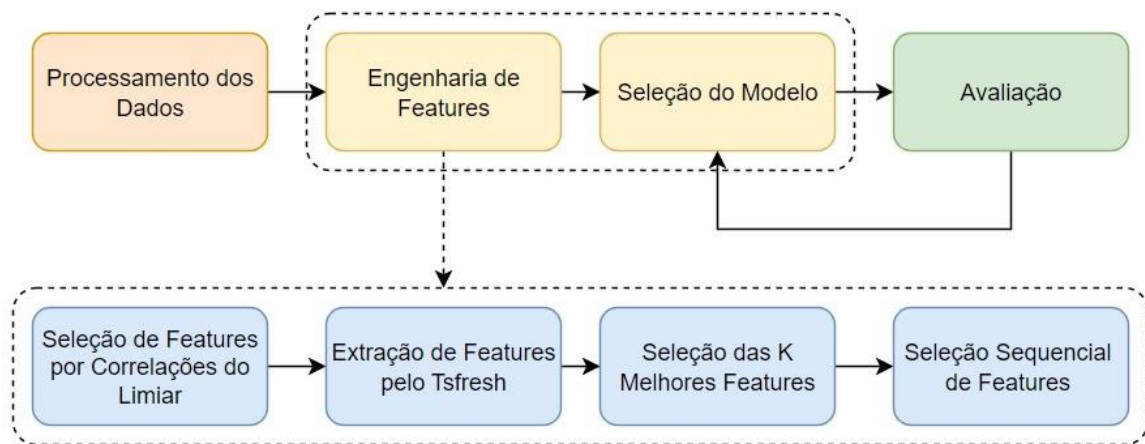


Figura 4: Pipeline de Detecção de Falhas

novo processo de treinamento. O conjunto de teste é usado após o término do processo de treinamento e validação, e os parâmetros do modelo são ajustados. Este conjunto também é novo para o modelo, e o desempenho do modelo nesse conjunto de dados descreverá seu desempenho para dados práticos.

É importante enfatizar que um ano inteiro de dados foi usado para compor o conjunto de treinamento, para que o modelo pudesse aprender o comportamento sazonal das falhas. A razão entre os dados de treino e teste foi de 90%/10%.

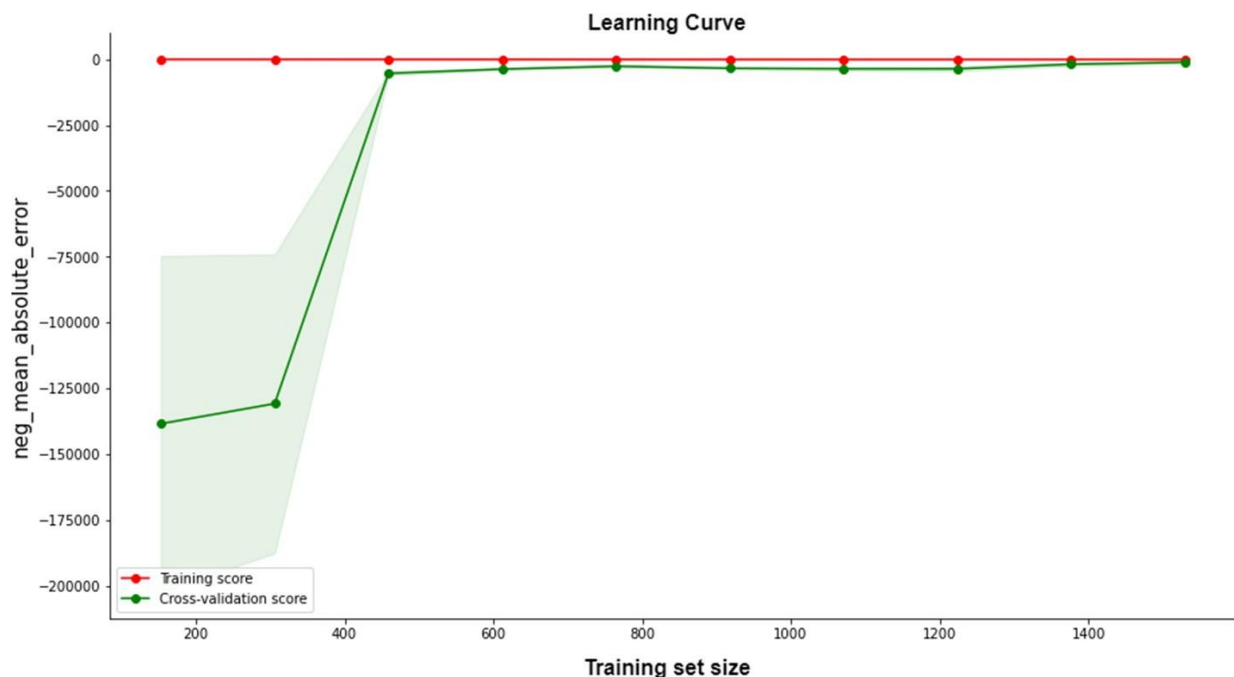


Figura 5: LCA do Modelo de Regressão

### 3.2. ENGENHARIA DE FEATURES

A Engenharia de Features (características relevantes) é realizada, inicialmente, através de uma seleção de Features com base na correlação. Em seguida, é realizada a extração de Features temporais como lag e trend utilizando a biblioteca tsfresh [6]. O próximo passo é realizar uma seleção das melhores Features para diagnóstico do tempo de falha dos ativos. Por fim, é realizado um processo de Seleção Sequencial de

Features.

### 3.3. SELEÇÃO DO MODELO

Embora outros modelos tenham sido testados, este trabalho apresenta os resultados obtidos com uma Regressão Ridge otimizada [7]. A avaliação consistiu na análise da curva de aprendizado (LCA), R-quadrado, erro absoluto médio e gráficos de predição. A [Figura 5](#) mostra a curva de aprendizado do modelo.

## 4. RESULTADOS

Para as métricas de treinamento, o R-quadrado foi de 0,98, o que significa que apenas 2% dos dados não puderam ser explicados pelo modelo.

Na [Figura 6](#), são mostrados o tempo real para uma falha ocorrida (azul) e o tempo previsto nos testes do modelo (vermelho). O comportamento observado nos gráficos nos dá a certeza de que as métricas apresentadas são reais e que os valores previstos seguem o padrão dos reais.

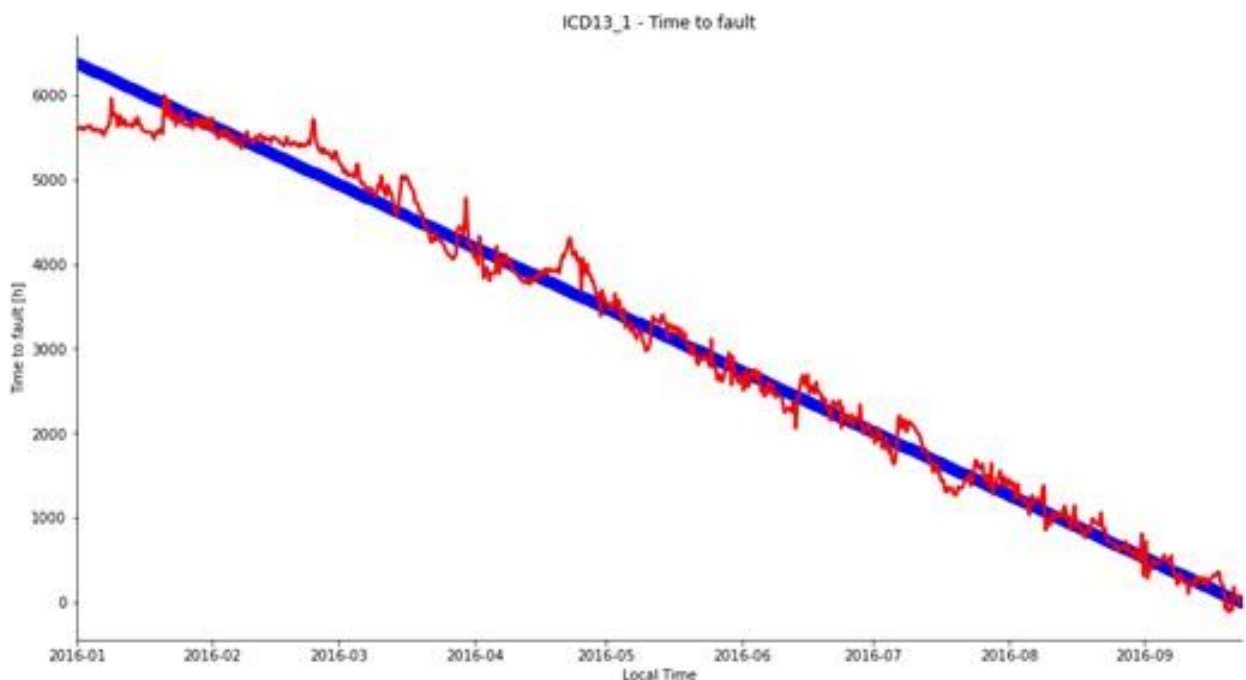


Figura 6: Predição do modelo para os dados de treinamento

Quanto às métricas de teste, foi obtido um valor de R-quadrado acima de 1, que significa um caso anormal, não tem significado lógico e pode ser resultado do pequeno tamanho da amostra. Por outro lado, a [Figura 7](#) mostra que os valores previstos seguem o valor real.

## 5. DASHBOARD

Os resultados dos algoritmos são disponibilizados em um dashboard apresentado na [Figura 8](#), onde é possível acompanhar o estado geral do parque, a evolução das falhas nos diferentes equipamentos e as variáveis que mais impactaram no diagnóstico da máquina. Além disso, o dashboard possui filtros que ajudam a estudar o comportamento do parque em datas, equipamentos, subsistemas e/ou predições específicas.

## 6. CONCLUSÃO

Este trabalho apresentou uma metodologia para diagnóstico da condição em aerogeradores. Os resultados mostraram que a metodologia foi capaz de aprender a identificar o período de falha da turbina. Os resultados dos testes mostraram-se promissores, demonstrando que o modelo entendeu o mecanismo de falha da turbina eólica. Porém, quando se aproximava o momento da ocorrência do defeito, o erro do modelo predito aumentava. Esta é uma indicação de que um modelo de curto prazo poderia ajudar a resolver este

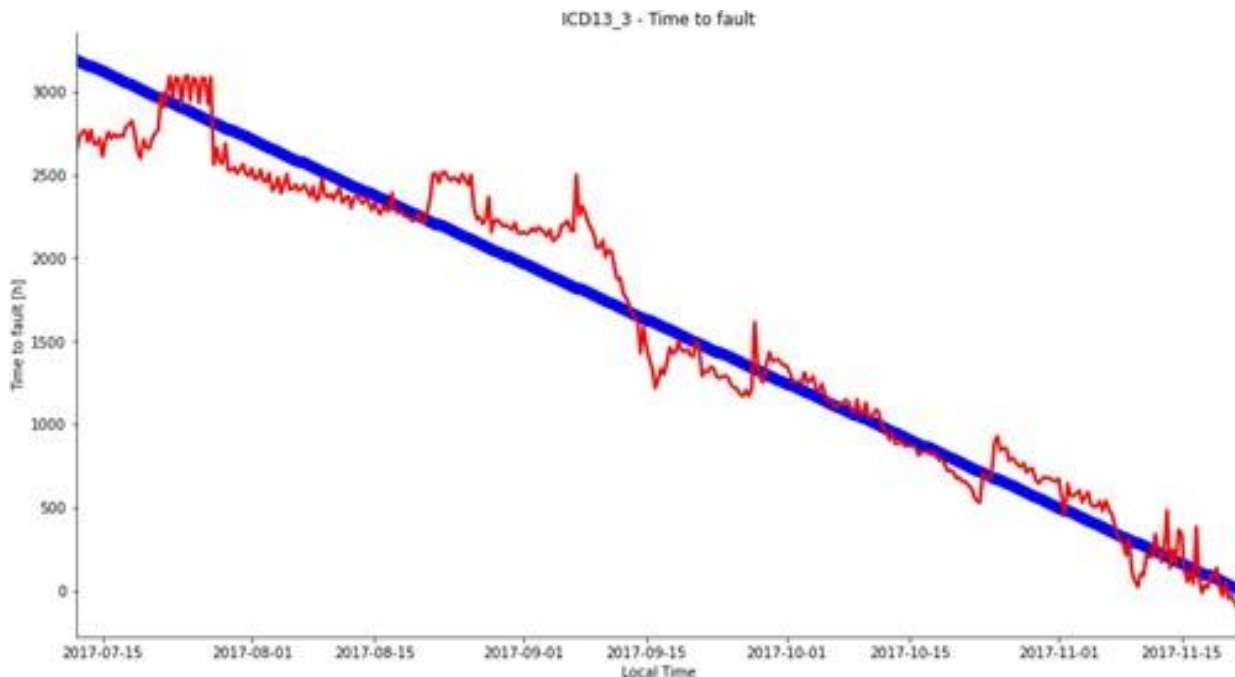


Figura 7: Predição do modelo para os dados de teste

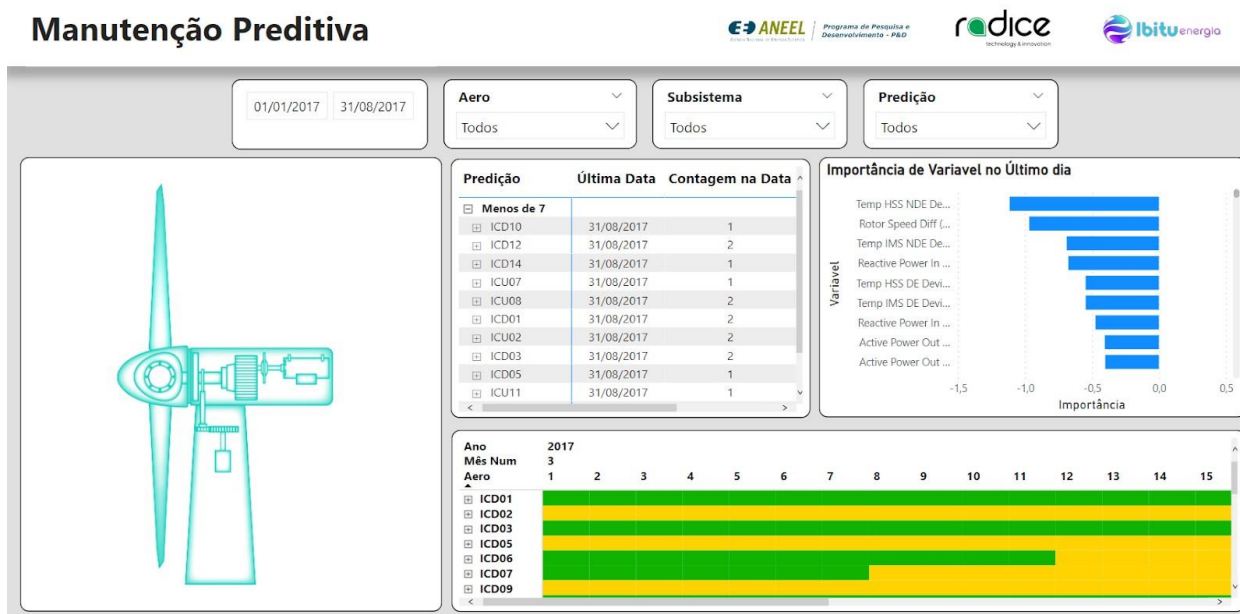


Figura 8: Dashboard de acompanhamento do estado dos Aerogeradores

problema.

O modelo desenvolvido pode ser estendido para qualquer tipo de defeito ou falha. Além disso, este trabalho demonstrou a possibilidade de criar sistemas preditivos de falhas com resultados satisfatórios.

A aplicação dos sistemas desenvolvidos nas rotinas de manutenção e gestão de ativos nos parques de aerogeradores possibilitam a obtenção de diversos benefícios tanto econômicos quanto socioambientais, tais como:

- Redução dos custos indiretos das falhas dos ativos, com a indicação do defeito apresentado pelo equipamento;



- Redução dos custos diretos e indiretos de manutenção, como insumos de manutenção e homem-hora;
- Redução dos riscos de acidentes de trabalho e dos custos diretos e indiretos associados;
- Redução dos riscos de impactos ambientais por falhas de equipamentos e dos custos diretos e indiretos associados;
- Redução da indisponibilidade dos parques e consequentes aumentos na confiabilidade na geração;
- Aumento de confiabilidade do parque gerador brasileiro e do SIN;
- Aumento da vida útil dos ativos.

## 7. AGRADECIMENTOS

O trabalho foi desenvolvido no âmbito do Projeto de P&D 0119 da ANEEL, intitulado Sistema Inteligente de Gestão Otimizada de Aerogeradores e Hidrogeradores.

## REFERÊNCIAS

- [1] IRENA, *Renewable Energy Technologies*, Online; acessado em 6 de julho de 2023. endereço: <https://www.irena.org/Data/View-data-by-topic/Capacity-and-Generation/Technologies>.
- [2] X. Jin, Y. Chen, L. Wang, H. Han e P. Chen, "Failure prediction, monitoring and diagnosis methods for slewing bearings of large-scale wind turbine: A review," *Measurement*, v. 172, p. 108855, 2021, ISSN: 0263-2241. DOI: <https://doi.org/10.1016/j.measurement.2020.108855>. endereço: <https://www.sciencedirect.com/science/article/pii/S0263224120313464>.
- [3] M. Tang, Z. Kuang, Q. Zhao, H. Wu e X. Yang, "Fault detection of wind turbine pitch system based on multiclass optimal margin distribution machine," *Mathematical Problems in Engineering*, v. 2020, pp. 1–10, 2020.
- [4] A. C. Tavares, C. Justi, D. Carrijo e L. P. Fritoli, "Ferramentas Contra Falhas Incipientes de Buchas de Alta Tensão com Cargas Não Lineares," em *XIX ERIAC - Décimo Nono Encontro Regional Ibero-Americano do Cigre*, Cigre, ISBN: 978-65-00-63038-1, 2023.
- [5] L. Williams, C. Phillips, S. Sheng, A. Dobos e X. Wei, "Scalable wind turbine generator bearing fault prediction using machine learning: A case study," em *2020 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, 2020, pp. 1–9.
- [6] M. Christ, N. Braun, J. Neuffer e A. W. Kempa-Liehr, "Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)," *Neurocomputing*, v. 307, pp. 72–77, 2018, ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2018.03.067>. endereço: <https://www.sciencedirect.com/science/article/pii/S0925231218304843>.
- [7] G. C. McDonald, "Ridge regression," *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 1, n. 1, pp. 93–100, 2009.