

# Estudo sobre Modelos de Aprendizado de Máquina para Detecção de Falhas em Turbinas Eólicas

Danielle Pinna,<sup>1</sup> Rodrigo Hamacher,<sup>2</sup> Fernando de Sá<sup>3</sup>

PPCIC/CEFET-RJ, Rio de Janeiro, RJ

Sanderson L. Gonzaga de Oliveira <sup>4</sup>

UNIFESP, São José dos Campos, SP

Raphael Guerra<sup>5</sup>

UFF, Niterói, RJ

Kele Belloze,<sup>6</sup> Diego Brandão<sup>7</sup>

PPCIC/CEFET-RJ, Rio de Janeiro, RJ

**Resumo.** A crescente busca por soluções energéticas renováveis tem trazido destaque para soluções como turbinas eólicas, que são as principais responsáveis pela transformação de energia eólica em elétrica. Assim, o monitoramento, diagnóstico e prognóstico de falhas destas turbinas é fundamental para garantir a produção energética de forma contínua. Estas turbinas são monitoradas por sensores e os dados oriundos deste monitoramento podem ser utilizados para criar modelos capazes de detectar estágios iniciais de degradação dos componentes que formam as turbinas, garantindo assim que falhas sejam identificadas rapidamente, reduzindo custos em manutenção. No presente trabalho é apresentada uma breve revisão sobre o assunto, além da aplicação de duas técnicas de aprendizado de máquina em uma base de dados real.

**Palavras-chave.** Energia Eólica, Aprendizado de Máquina,

## 1 Introdução

Dentre as fontes de energia limpa mais maduras e escaláveis tem-se a energia eólica, que tem apresentado crescente adoção pelos países e custos cada vez menores ao longo dos anos [21]. Em 2020, as novas instalações de energia eólica em todo o mundo ultrapassaram 90 GW, um crescimento de 53% em relação a 2019. Além disso, o Conselho Global de Energia Eólica (GWEC) espera um crescimento médio global de 4% ao ano nos próximos 5 anos[7].

Apesar da ampla adoção de turbinas eólicas para geração de energia, uma grande preocupação é com relação a manutenção de seus equipamentos, visto que a operação e manutenção representam em média 25% do custos totais de uma turbina eólica [18]. Em caso de falhas, a turbina pode ficar inativa por longos períodos, gerando um prejuízo pela interrupção de energia, além dos altos custos de reparos dada a complexidade de seus componentes [3]. Portanto, para diminuir o tempo de ociosidade e reduzir os custos desnecessários com manutenção, torna-se importante a detecção de estágios iniciais de degradação dos componentes antes da ocorrência da falha.

---

<sup>1</sup>danielle.pinna@cefet-rj.br

<sup>2</sup>rodrigo.hamacher@cefet-rj.br

<sup>3</sup>fernando.sa@cefet-rj.br

<sup>4</sup>sgonzaga@unifesp.br

<sup>5</sup>rguerra@ic.uff.br

<sup>6</sup>kele.belloze@cefet-rj.br

<sup>7</sup>diego.brandao@cefet-rj.br

Atualmente, muitas das turbinas eólicas são equipadas com o sistema de monitoramento conhecido como SCADA (*Supervisory Control and Data Acquisition System*), que monitora e armazena dados de todo o funcionamento das turbinas através de sensores instalados em seus componentes [23]. Sistemas baseados no SCADA permitem a implementação de estratégias de manutenção preditiva e gerenciamento das condições das turbinas eólicas. Tais dados podem ser usados posteriormente para o desenvolvimento de modelos de detecção de falhas e diagnóstico em turbinas eólicas. Deste modo, inúmeros trabalhos de detecção de falhas em turbinas eólicas com base em dados do SCADA vem sendo desenvolvidos [11, 19, 24], apresentando distintas abordagens e soluções.

Este artigo apresenta uma revisão bibliográfica de trabalhos relacionados e fundamentos teóricos sobre detecção de anomalias e falhas em turbinas eólicas. Por fim, ele apresenta uma aplicação de dois métodos de aprendizado de máquina em um problema real. O presente trabalho está dividido em mais 4 seções: (i) fundamentação teórica, com as principais técnicas utilizadas, (ii) trabalhos relacionados, onde é apresentada uma discussão sobre os artigos estudados referentes ao tema, (iii) resultados do experimento com dados reais e as (iv) considerações finais com sugestões de trabalhos futuros.

## 2 Fundamentação Teórica

Esta seção apresenta uma breve introdução sobre os conceitos de aprendizado de máquina e as principais técnicas utilizadas desta área no problema de detecção de anomalias em turbinas eólicas.

### 2.1 Aprendizado de Máquina

Aprendizado de máquina pode ser definido como um campo de estudo que capacita os computadores a aprenderem com base em dados [5]. Esta capacidade é de grande utilidade para casos onde se deseja encontrar padrões, relações ou alguma estrutura entre os dados que não sejam explícitos a olhos humanos, ou mesmo que sejam, ao menos otimize esse processo. Ela agrega técnicas das áreas de matemática, estatística e computação para solucionar o problema desejado.

Existem diferentes tipos de aprendizado de máquinas, conforme apresentado na Tabela 1 [1, 5]. Além disso, os tipos de problema podem ser de regressão (quando busca-se prever algum valor numérico) ou de classificação (quando busca-se prever alguma classe).

Tabela 1: Tipos de Aprendizado de Máquina

Tipos de Aprendizado de Máquina	Definição
Supervisionado	Quando os dados de treino contêm dados de entrada associados aos resultados que deseja-se prever ou estudar.
Semi Supervisionado	Quando os dados de treino contêm poucos, ou em quantidade irrisória, dados de entrada associados aos resultados que deseja-se prever ou estudar.
Não Supervisionado	Quando os dados de treino contêm nenhum dado de entrada associados aos resultados que deseja-se prever ou estudar.
Aprendizado por Reforço	Quando os dados de treino contêm dados de entrada associados a algum possível resultado juntamente com uma métrica de o quão bom foi esse resultado.

Diante do exposto, muitos dos trabalhos relacionados a predição de falhas ou detecção de anomalias em turbinas eólicas apresentam como solução técnicas de aprendizado de máquina [8]. Na seção 3 são apresentados com maiores detalhes algumas dessas soluções. No entanto, este trabalho foca em duas técnicas, a Máquina de Vetores de Suporte devido sua versatilidade e desempenho e na técnica de regressão logística pela sua simplicidade.

## 2.2 Máquina de Vetores de Suporte

Este é um dos algoritmos mais eficientes de aprendizado de máquina, no qual o problema pode ser abordado como um problema de classificação linear, não-linear, regressão ou detecção de anomalia.

Primeiramente, o algoritmo busca encontrar o hiperplano de máxima margem que melhor separa os pontos de diferentes classes. Os pontos de treino equidistantes do hiperplano de máxima margem e que se encontram mais próximos dele são chamados de vetores de suporte, que são os principais responsáveis pela determinação deste hiperplano, como mostra a Figura 1. Com intuito de acomodar limites não lineares entre as classes, aumenta-se o espaço dimensional dos dados através do uso de *kernels* deixando o algoritmo mais flexível [9]. Uma variante da Máquina de Vetores de Suporte é o *One-Class SVM* apresentado por [16], com ampla aplicação para casos de detecção de anomalias, sendo o avaliado neste trabalho.

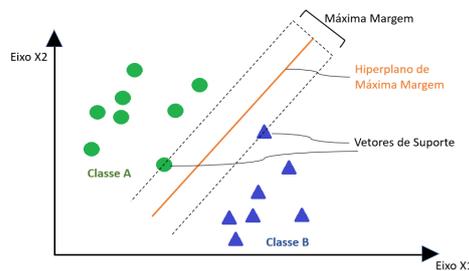


Figura 1: Exemplificação de Máquina de Vetores de Suporte

## 2.3 Regressão Logística

A Regressão Logística é um tipo de modelo linear generalizado (GLM) utilizado para classificação binária, e é também considerado um algoritmo de aprendizagem supervisionado. Tem como objetivo estimar valores discretos (valores binários como 0/1, sim/não, verdadeiro/falso) com base em um determinado conjunto de variáveis explicativas. Normalmente, a regressão logística usa uma função “*Sigmoid*” (função logística), que possui curva em formato “S”, utilizada para a classificação binária que converte valores para o intervalo [0,1], podendo ser interpretados como a probabilidade de determinada instância pertencer ou não a determinada classe.

## 2.4 Matriz de Confusão e Métricas de Desempenho

Para avaliar os resultados dos métodos de classificação existem diferentes métricas. Um método bastante utilizado para essa avaliação é a matriz de confusão e as medidas de desempenho que dela resultam. A matriz de confusão é exibida na Figura 2, que fornece as quantidades previstas e observadas em cada classe da variável resposta.

Figura 2: Matriz de confusão para um problema com duas classes

		Classe prevista	
		0	1
Classe verdadeira	0	VN	FP
	1	FN	VP

Neste caso, os Verdadeiros Positivos (VP) consistem no número de exemplos da classe positiva e que foram corretamente classificados, no problema de interesse representa a detecção correta de falhas. Os Verdadeiros Negativos (VN) consistem no número de exemplos da classe negativa e que foram corretamente classificados, ou seja, os exemplos que não tiveram falhas. Já os Falsos Positivos (FP) são o número de exemplos da classe negativa e que foram incorretamente classificados pelo modelo, compreendendo os falsos alarmes. Por fim, os Falsos Negativos (FN) são o número de exemplos da classe positiva e que foram incorretamente classificados, isto é, as falhas não detectadas.

As medidas de desempenho baseados na matriz de confusão são descritas abaixo:

- **Acurácia** - A acurácia é a razão da soma de todos os casos corretamente classificados pelo número total de objetos no conjunto de teste.

$$acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

- **Precisão** - Corresponde a proporção de resultados positivos classificados corretamente entre todos aqueles preditos como positivos.

$$precisão = \frac{VP}{VP + FP} \quad (2)$$

- **Sensibilidade ou Revocação** - Corresponde à taxa de acerto na classe positiva verdadeira.

$$sensibilidade = \frac{VP}{VP + FN} \quad (3)$$

- **F1-Score** - É a média harmônica entre a precisão e a revocação.

$$F_1-score = 2 * \frac{precisão * revocação}{precisão + revocação} \quad (4)$$

### 3 Trabalhos Relacionados

Os trabalhos relacionados na presente seção foram obtidos através da consulta à base de artigos acadêmicos Scopus utilizando a seguinte expressão ("*wind turbine*" AND "*detection*" AND "*review*"). O resultado retornou 191 referências das quais três foram selecionadas por estarem mais alinhadas com o escopo do artigo e pela ampla gama de artigos e conteúdos que esses três trabalhos apresentaram. As referências encontradas nestes trabalhos também foram utilizadas.

Existem diferentes formas de abordagens para tratar o problema de detecção em turbinas eólicas, [20] apresentam uma revisão de trabalhos cujos resultados tem se mostrado promissores para detecção de anomalia em dados reais de turbinas eólicas armazenados pelo sistema SCADA. Os autores segmentam os métodos entre abordagens de (i) tendência, (ii) agrupamento, (iii) modelagem do comportamento normal, (iv) modelagem de falha, (v) avaliação de alarmes, trazendo as vantagens e desvantagens de cada abordagem. Entretanto, destacam (iii) por ser o foco de pesquisas recentes dada a praticidade e facilidade em modelar o comportamento normal dos componentes das turbinas e detectar anomalia. Como exemplo de modelos desta abordagem na literatura, temos os modelos polinomiais e lineares, redes neurais e sistemas com base na lógica difusa.

Maldonado-Correa *et al.* [12] realizam uma revisão de artigos publicados entre 2017 e 2020 e destacam que redes neurais foi a técnica mais utilizada entre os trabalhos visitados pelos autores aparecendo em 35 artigos, seguido pelo modelo de máquina de vetores de suporte aparecendo em

23 artigos. Entretanto, modelos como árvores de decisão, lógica difusa, *naive bayes*, k-média e redes bayesianas também são apresentados. A justificativa encontrada pelos autores para explicar a popularidade de modelos de redes neurais é que esse tipo de modelo apresenta ótima performance para identificação de padrões não lineares dos dados.

No trabalho realizada por [17], concluiu-se que dois terços dos modelos usavam métodos de classificação e o restante regressão, sendo redes neurais, máquina de vetores de suporte e árvores de decisão os mais utilizados.

Dentre os artigos pesquisados pelo autores, outra abordagem comumente utilizada é o treino de modelos para identificação de falhas em componentes específicos das turbinas eólicas. Enquanto alguns trabalhos analisam falhas nas pás [6, 10, 15], outros focam em analisar falhas na caixa de engrenagem [2, 14, 22].

Atualmente, a falta de uma grande quantidade de dados públicos de turbinas eólicas é um dos limitantes no progresso de maiores trabalhos sobre o tema. A grande parte dos modelos apresentados nos trabalhos utilizam dados simulados ou do sistema SCADA, o que pode ser um risco, pois modelos treinados em dados simulados podem não generalizar bem em dados reais e apresentarem uma performance pior do que esperada em produção. Além disso, pouquíssimas abordagens utilizam dados de imagem ou áudio para detecção de falhas [17].

## 4 Experimento com Dados Reais

Os dados utilizados neste artigo são fornecidos pela empresa de Energias de Portugal (EDP) [4]. Esse é um dos conjuntos de dados gratuitos disponíveis mais completos para análise de recursos eólicos e pesquisa do desempenho de turbinas eólicas [13]. Os registros foram extraídos de um sistema SCADA constituído por 5 turbinas eólicas medidos nos anos de 2016 e 2017. Além dos dados do sistema Scada, a base de dados é constituída por informações relativas a dados meteorológicos (velocidade do vento, temperatura, pressão atmosférica, umidade e precipitação), registro de ocorrência de falhas em cinco componentes da turbina eólica (Transformador, Rolamento do Gerador, Grupo Hidráulico, Gerador e Caixa de Velocidade), histórico dos eventos normais e anormais de cada turbina e a localização de cada turbina (latitude e longitude).

### 4.1 Resultados dos Modelos

A base de dados foi dividida em 80% para treinamento dos modelos e 20% para teste, mantendo a ordem do conjunto de dados. Os experimentos foram realizados através de rotinas computacionais implementadas em Python versão 3, em uma máquina Intel(R) Xeon(R) Gold 5120 CPU 2.20GHz, com 28 núcleos e 192GB de memória RAM. As bibliotecas utilizadas foram *pandas*<sup>1</sup>, *numpy*<sup>2</sup> e *scikit-learn*<sup>3</sup>.

As Tabelas 2 e 3 apresentam as métricas de desempenho dos modelos na base de teste: Acurácia, Precisão, *Recall* e  $F_1$ -Score.

A escolha do melhor algoritmo testado para cada componente da Turbina Eólica se deu pela métrica do  $F_1$ -Score. Pelas tabelas, pode se observar que o algoritmo baseado em Regressão Logística apresentou um desempenho bem inferior ao SVM para todas as componentes avaliadas. Tal resultado era esperado, visto que o SVM utilizado consiste em um classificador de uma classe, conhecido pela sua capacidade em determinar *outliers* de maneira eficiente e o problema abordado consiste em um conjunto de dados de treinamento formado basicamente por uma única classe constituída pelo funcionamento normal das componentes.

<sup>1</sup><https://pandas.pydata.org/>

<sup>2</sup><https://numpy.org/>

<sup>3</sup><https://scikit-learn.org/stable/>

Tabela 2: Métricas da Regressão Logística

Componente	F1-Score	Acurácia	Precisão	Revocação
Caixa de Velocidade	32,14%	69,35%	61,80%	21,72%
Rolamento do Gerador	17,73%	67,90%	62,38%	10,34%
Transformador	23,95%	73,86%	85,56%	34,94%
Gerador	16,88%	74,81%	96,99%	9,24%
Grupo hidráulico	17,49%	62,05%	95,86%	9,62%

Tabela 3: Métricas do SVM

Componente	F1-Score	Acurácia	Precisão	Revocação
Caixa de Velocidade	62,66%	63,79%	99,97%	45,63%
Rolamento do Gerador	44,91%	44,33%	65,72%	34,10%
Transformador	58,70%	53,76%	79,14%	46,66%
Gerador	39,77%	39,80%	71,98%	27,48%
Grupo hidráulico	55,93%	57,75%	71,17%	46,06%

## 5 Considerações Finais

A busca pela redução da emissão de gases de efeito estufa tem trazido destaque para fontes renováveis e de energia limpa como a energia eólica. Entretanto, os altos custos de operação e manutenção de suas turbinas têm sido umas das principais desvantagens em relação a esta fonte energética.

Sistemas de monitoramento e armazenamento de dados de turbinas eólicas, como o SCADA, tem proporcionado o desenvolvimento de técnicas baseadas nestes dados que detectam de falhas nas turbinas, deste modo, reduzindo o tempo e custo de manutenção, prevenindo falhas e aumentando a produtividade. Dentre elas, as mais utilizadas têm sido redes neurais e máquinas de vetores de suporte, devida a maior performance, principalmente das redes neurais em encontrar padrões não lineares nos dados. Neste trabalho foi abordado o uso das técnicas de máquinas de vetores de suporte, como trabalho futuro as redes neurais artificiais serão aplicadas no problema, bem como técnicas de avaliação de hiperparâmetros para modelos de aprendizado de máquina.

Outra vertente a ser avaliada consiste na aplicação de técnicas de aprendizado de máquina com base em imagens ou áudios dada a baixa quantidade de estudos neste sentido. Além disso, grande parte dos trabalhos sobre o tema consideram somente um único parque eólico ou fabricante. A validação de técnicas em dados mais heterogêneos possibilitará métodos com maior poder de generalização para detecção de falhas em turbinas eólicas.

## Referências

- [1] Yaser S Abu-Mostafa, Malik Magdon-Ismail e Hsuan-Tien Lin. **Learning From Data**. Vol. 4. AMLBook, 2012.
- [2] James Carroll et al. “Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques”. Em: **Wind Energy** 22.3 (2019), pp. 360–375.
- [3] Daniel Chan e John Mo. “Life Cycle Reliability and Maintenance Analyses of Wind Turbines”. Em: **Energy Procedia** 110.December 2016 (2017), pp. 328–333.
- [4] **EDP - Open Data**. <https://opendata.edp.com/pages/homepage/>, last accessed on 15/08/21.

- [5] A. Géron. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. O'Reilly Media, 2019.
- [6] Jamie Godwin e Peter Matthews. "Classification and Detection of Wind Turbine Pitch Faults Through SCADA Data Analysis". Em: **International Journal of Prognostics and Health Management** 4 (jan. de 2013).
- [7] GWEC. "Global Wind Report 2021". Em: **Global Wind Energy Council** (2021).
- [8] Georg Helbing e Matthias Ritter. "Deep Learning for fault detection in wind turbines". Em: **Renewable and Sustainable Energy Reviews** 98.September (2018), pp. 189–198.
- [9] Gareth James et al. **An Introduction to Statistical Learning: with Applications in R**. Springer, 2013.
- [10] A Joshuva. e V Sugumaran. "A data driven approach for condition monitoring of wind turbine blade using vibration signals through best-first tree algorithm and functional trees algorithm: A comparative study". Em: **ISA Transactions** 67 (2017), pp. 160–172.
- [11] Xiaoyuan Liu et al. "Wind turbine anomaly detection based on SCADA data mining". Em: **Electronics (Switzerland)** 9.5 (2020).
- [12] Jorge Maldonado-Correa et al. "Using SCADA data for wind turbine condition monitoring: A systematic literature review". Em: **Energies** 13.12 (2020).
- [13] Mateus Mendes et al. "Wind Farm and Resource Datasets: A Comprehensive Survey and Overview". Em: **Energies** 13 (set. de 2020).
- [14] Rafael Orozco, Shawn Sheng e Caleb Phillips. "Diagnostic Models for Wind Turbine Gearbox Components Using SCADA Time Series Data". Em: jun. de 2018, pp. 1–9.
- [15] Taylor Regan, Christopher Beale e Murat Inalpolat. "Wind Turbine Blade Damage Detection Using Supervised Machine Learning Algorithms". Em: **Journal of Vibration and Acoustics** 139.6 (ago. de 2017).
- [16] Bernhard Schölkopf et al. "Estimating the support of a high-dimensional distribution". Em: **Neural Computation** 13.7 (2001), pp. 1443–1471.
- [17] Adrian Stetco et al. "Machine learning methods for wind turbine condition monitoring: A review". Em: **Renewable Energy** 133 (2019), pp. 620–635.
- [18] Bernhard Strack et al. "Ontology for maintenance of onshore wind turbines". Em: **Forschung im Ingenieurwesen/Engineering Research** 85.2 (2021), pp. 265–272.
- [19] Peng Sun et al. "A generalized model for wind turbine anomaly identification based on SCADA data". Em: **Applied Energy** 168 (2016), pp. 550–567.
- [20] Jannis Tautz-Weinert e Simon J. Watson. "Using SCADA data for wind turbine condition monitoring - A review". Em: **IET Renewable Power Generation** 11.4 (2017), pp. 382–394.
- [21] Pierre Tchakoua et al. "Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges". Em: **Energies** 7.4 (2014), pp. 2595–2630.
- [22] Long Wang et al. "Wind Turbine Gearbox Failure Identification With Deep Neural Networks". Em: **IEEE Transactions on Industrial Informatics** 13.3 (2017), pp. 1360–1368.
- [23] Wenna Zhang e Xiandong Ma. "Simultaneous Fault Detection and Sensor Selection for Condition Monitoring of Wind Turbines". Em: **Energies** 9.4 (2016).
- [24] Hongshan Zhao et al. "Anomaly detection and fault analysis of wind turbine components based on deep learning network". Em: **Renewable Energy** 127 (2018), pp. 825–834.