



## Aplicação de técnicas de aprendizado de máquina para previsão da velocidade do vento

### Application of machine learning techniques for wind speed forecasting

Leonardo Nathan Chioquetta<sup>1</sup>, Matheus Henrique Dal Molin Ribeiro<sup>2</sup>

#### RESUMO

A energia eólica é um setor de importância entre as fontes de energia, seu crescimento tem sido significativo e sua previsão se torna cada vez mais crucial para operações seguras e estáveis no sistema energético. Entretanto, devido ao atrelamento da energia eólica com a velocidade do vento, encontrar um modelo de previsão que se adapte ao longo do tempo é um desafio. Assim, o presente estudo destaca a aplicação de diferentes modelos de aprendizado de máquina para prever a velocidade dos ventos e avaliar a adequabilidade das previsões desenvolvidas para diferentes estações do ano. Os dados utilizados neste trabalho consistem em quatro séries temporais de valores de velocidade do vento, coletados durante o ano de 2018. Foram selecionadas séries para os meses de janeiro, abril, julho e setembro, que refletem o comportamento das estações climáticas. No presente estudo diversos modelos de aprendizado de máquina foram comparados quando a sua previsão da velocidade dos ventos 1 hora a frente. Ao considerar o menor erro médio absoluto percentual, o modelo RQ-LASSO apresentou a melhor performance em 50% dos casos e ao avaliar o menor valor para a raiz quadrada do erro médio, o modelo LASSO obteve melhor desempenho em 75% dos casos.

**PALAVRAS-CHAVE:** Aprendizado de máquina; previsão de séries temporais; velocidade do vento.

#### ABSTRACT

Wind energy is an important sector between energy sources, its growth has been significant, and its forecast becomes increasingly crucial for safe and stable operations in the energy system. However, due to the link between wind energy and wind speed, finding a forecast model that adapts over time is a challenge. Thus, the present study highlights the application of different machine learning models to predict wind speed and evaluate the suitability of forecasts developed for different seasons. The data used in this work consists of four time series of wind speed values, collected during 2018. Series were selected for the months of January, April, July, September, which reflect the behavior of the climatic seasons. In the present study, several machine learning models were compared when predicting wind speeds 1 hour ahead. When considering the lowest mean absolute percentage error, the RQ-LASSO model presented the best performance in 50% of cases and when evaluating the lowest value for the square root of the mean error, the LASSO model achieved better performance in 75% of cases.

**KEYWORDS:** Machine learning; Time series forecasting; Wind speed.

#### Introdução

De acordo com a Agência Internacional de Energia (IEA) a capacidade de energias renováveis crescerá 2400 Giga Watts até 2027, que equivale à capacidade instalada da China atualmente. Devido à importância do setor eólico nas energias renováveis, a previsão concreta da geração de energia eólica se torna cada vez mais necessária para operações mais seguras e estáveis no sistema de energia (ZHUOY LIU, 2021). No entanto, de acordo com Paula *et al.* (2020) as características de aleatoriedade e de intermitência do vento, tornam a previsão da velocidade do vento uma tarefa desafiadora.

<sup>1</sup> Bolsista da UTFPR (Fundação Araucária). Universidade Tecnológica Federal do Paraná, Pato Branco, Paraná, Brasil. E-mail: Chioquetta@aunos.utfpr.edu.br. ID Lattes: 3063232319869096.

<sup>2</sup> Docente no Curso/Departamento/Programa. Universidade Tecnológica Federal do Paraná, Pato Branco, Paraná, Brasil. E-mail: mribeiro@utfpr.edu.br. ID Lattes: 2531836774466938.



Existem diversos modelos estatísticos utilizados para realizar previsões de séries temporais e entre eles, destacam-se os modelos lineares, tais como AR (Autorregressivo), regressão linear e ARIMA (Média Móvel Integrado Autorregressivo), como cita Altan *et al.* (2021). Entretanto, de acordo com os autores, quando características não lineares são predominantes, a previsão resultante desses modelos pode não ser satisfatória. Esse comportamento é especialmente comum em séries temporais de velocidade do vento, o que pode ser contornado por meio do uso de modelos não lineares, bem como com aqueles relacionados a inteligência artificial, capazes de obter uma melhor performance.

O presente estudo tem como objetivo avaliar de forma comparativa uma variedade de métodos de aprendizado de máquina (ML) tais como regressão LASSO e sua versão quantílica (RQ-LASSO), máquina de vetores suporte (SVR), florestas aleatórias (RF), e *eXtreme gradient boosting* (XGBoost) para previsão da velocidade do vento (uma hora a frente) durante diferentes épocas do ano, com a finalidade de determinar um possível modelo que produza previsões similares aos dados observados.

Dentre as contribuições deste estudo, destacam-se a aplicação de diferentes modelos de ML para previsão da velocidade dos ventos, bem como avaliação da adequabilidade das previsões desenvolvidas para diferentes meses ou estações do ano. Este estudo permitirá obter previsões mais precisas da velocidade do vento e contribuirá para o planejamento de atividades que dependam dessa informação.

## Metodologia

O software mais utilizado no presente estudo foi o RStudio 4.2.2 (R Core Team, 2023), que a homepage POSIT define como sendo um ambiente de desenvolvimento integrado para linguagem R de programação para cálculos estatísticos. Os conceitos iniciais sobre previsões estão compilados no *e-book Forecasting: Principles and Practice* (HYNDMAN, 2021) e os conceitos iniciais sobre series temporais foram coletados na videoaula Séries temporais no R e no livro digital Análise de séries temporais em R: curso introdutório (FERREIRA, 2017).

Os modelos de previsão de dados foram treinados com os métodos supracitados, onde 70% dos dados foram utilizados como conjunto de treinamento e os outros 30% como conjunto de teste, que ao comparados com os próprios valores reais, geraram as métricas de erro médio absoluto percentual (MAPE) e raiz quadrada do erro médio (RMSE), as quais demonstram a acurácia de cada modelo de previsão, avaliando qual apresenta um melhor desempenho. Foram utilizados cinco valores passados para cada modelo, como informações de entrada dos modelos, para prever a velocidade dos ventos uma hora a frente.

No presente estudo, cinco modelos dentre vários foram selecionados para realização da análise comparativa, o SVR, LASSO, RQ LASSO, RF e XGBoost.

O modelo SVR, conforme descrito por Ribeiro (2021), consiste em encontrar pontos próximos a um hiperplano (chamados de vetores de suporte) que maximizem a margem entre as classes de dois pontos, definidos como o maior e o menor valor da variável alvo, através da determinação da diferença entre o valor alvo e um limite.

O método de regressão LASSO incorpora uma penalidade na função de custo do modelo, sendo proporcional à soma dos valores absolutos dos coeficientes de regressão, permitindo que o LASSO controle simultaneamente a variância dos coeficientes ajustados e execute automaticamente a seleção de variáveis relevantes para o modelo, já o método



RQ-LASSO incorpora uma penalidade L1 na função de custo do modelo, reduzindo os coeficientes menos relevantes a zero e promovendo a seleção de variáveis mais importantes. (LI, Y., & ZHU, J. 2008).

O modelo RF é um algoritmo de ML que utiliza o método de comitês de árvores de decisão. Proposto por Breiman (2001), o modelo altera a forma como as árvores de classificação ou regressão são construídas, de forma em que cada nó é dividido de melhor modo entre um subconjunto de preditores escolhidos aleatoriamente naquele nó.

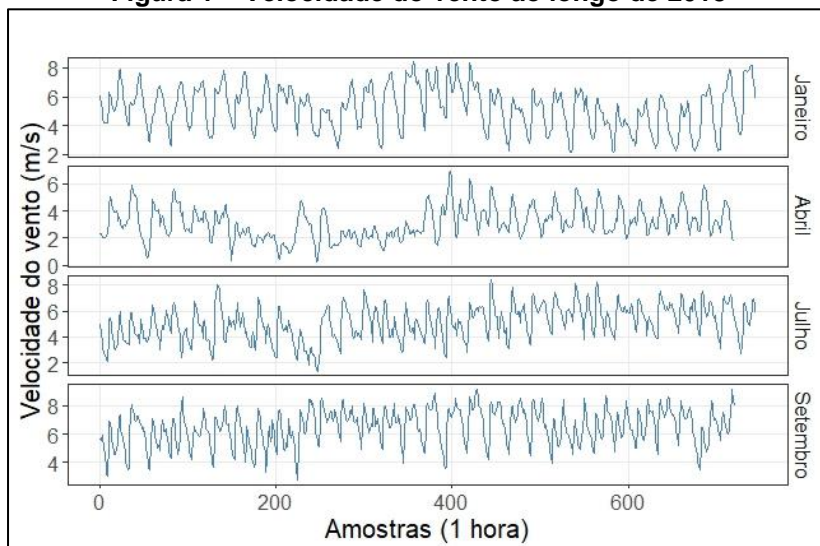
O XGBoost, é um algoritmo de árvores escaláveis projetado para lidar com grandes conjuntos de dados e alta eficiência computacional. Ele é baseado no algoritmo *Gradient Boosting*, porém, utiliza uma abordagem regularizada visando evitar o *overfitting* e aprimorar a precisão preditiva, de acordo com Chen, T., & Guestrin, C. (2016).

Descrição dos dados:

Os dados utilizados no presente trabalho são da série temporal de valores horários de velocidade do vento do ano de 2018 para a cidade de Limoeiro do Norte – CE, originada a partir do MERRA2 (“*Modern Era Retrospective-analysis for Research and Applications*”), que é uma reanálise de dados atmosféricos obtidos por satélites da NASA.

Para realizar o presente estudo os dados adotados serão referentes a um mês de cada estação do ano, sendo eles janeiro para o verão, abril para o outono, julho para o inverno e setembro para a primavera, ilustrado na Figura 1. As medidas estatísticas Máximo, Mínimo, Média e Desvio Padrão de cada mês estão dispostos na Tabela 1:

**Figura 1 – Velocidade do vento ao longo de 2018**



Fonte: Autoria própria.

**Tabela 1 – Medidas estatísticas para as amostras selecionadas**

| Conjunto de dados    | Amostras (n) | Mínimo (m/s) | Média (m/s) | Máximo (m/s) | Desvio Padrão (m/s) |
|----------------------|--------------|--------------|-------------|--------------|---------------------|
| Verão (janeiro)      | 744          | 2,1270       | 5,2240      | 8,3990       | 1,4880              |
| Outono (abril)       | 720          | 1,4620       | 5,1720      | 8,4800       | 1,1890              |
| Inverno (julho)      | 744          | 0,2365       | 3,1297      | 6,9535       | 1,3204              |
| Primavera (setembro) | 720          | 2,1890       | 4,6480      | 8,1480       | 1,2349              |

Fonte: Arquivo próprio.



## Medidas de desempenho

Foram adotadas medidas de desempenho, com a intenção de quantificar a performance dos modelos, sendo elas o MAPE, e o RMSE, onde comparam os valores previstos pelo modelo no conjunto de teste com os valores reais da base de dados, o que é chamado de validação de dados.

$n$  representa o número de elementos do conjunto de dados,  $y_i$  os valores reais e  $\hat{y}_i$  os valores previstos, portanto, as equações que descrevem o MAPE e o RMSE são:

$$MAPE = 100\% \times \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

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

Além das métricas primárias, também foi utilizado o índice de melhora IP, que representa de maneira quantitativa o quanto dois modelos diferem em sua performance, descrito por:

$$Ip = \frac{\text{Métrica do melhor método} - \text{Métrica do método comparado}}{\text{Métrica do método comparado}}$$

Por fim, o tempo de execução do script para cada modelo também foi levado em conta, analisando todos esses fatores, é definido qual modelo de previsão é mais adequado para o horizonte de previsão em questão (1-hora a frente), então, foram gerados gráficos que ilustram o distanciamento dos valores observados e previstos, e tabelas comparativas entre os dados quantitativos.

## Resultados e discussões

As métricas MAPE, RMSE, tempo de execução são levadas em conta para a escolha de um melhor modelo de previsão, onde cada modelo gerou uma medida de erro quando comparado com o valor real observado, dispostos na Tabela 2:

**Tabela 2 – Métricas de erro para cada modelo de previsão para o conjunto de teste.**

| Conjunto de Dados    | Métrica    | RQ-LASSO | SVR   | RF     | XGBoost  | LASSO |
|----------------------|------------|----------|-------|--------|----------|-------|
| Verão (janeiro)      | MAPE (%)   | 5,147    | 5,101 | 12,703 | 13,255   | 6,983 |
|                      | RMSE (m/s) | 0,371    | 0,372 | 0,608  | 0,654    | 0,383 |
|                      | Tempo (s)  | 9,070    | 5,700 | 84,680 | 1265,070 | 0,520 |
| Outono (abril)       | MAPE (%)   | 5,060    | 5,158 | 6,698  | 7,164    | 5,698 |
|                      | RMSE (m/s) | 0,384    | 0,382 | 0,412  | 0,440    | 0,377 |
|                      | Tempo (s)  | 8,780    | 6,540 | 87,130 | 1293,410 | 0,400 |
| Inverno (julho)      | MAPE (%)   | 6,320    | 6,317 | 7,944  | 8,815    | 6,145 |
|                      | RMSE (m/s) | 0,358    | 0,355 | 0,360  | 0,390    | 0,343 |
|                      | Tempo (s)  | 8,520    | 5,110 | 80,760 | 1167,830 | 0,410 |
| Primavera (setembro) | MAPE (%)   | 6,708    | 6,720 | 7,941  | 8,574    | 7,063 |
|                      | RMSE (m/s) | 0,547    | 0,546 | 0,579  | 0,601    | 0,529 |
|                      | Tempo (s)  | 8,370    | 5,150 | 81,390 | 1186,370 | 0,450 |

Fonte: Autoria própria

Ao analisar a métrica de erro MAPE, o método RQ-LASSO teve índices de melhora de 1 a 3% nos meses em que não foi categorizado como modelo de menor métrica, já o





índices de melhora pequenos quando comparados com os modelos XGBoost e RF, o que significa que são mais adequados para construção de um modelo de previsão.

## Agradecimentos

Os autores agradecem a UTFPR – Pato Branco e a Fundação Araucária pelo suporte nesta pesquisa na forma de concessão de bolsa de Iniciação Científica.

## Conflito de interesse

Não há conflito de interesse.

## Referências

- International Energy Agency.** Disponível em: <https://www.iea.org/reports/renewables-2022/executivesummary>. Acesso em abril de 2023.
- LIU, Zhuoyi; HARA, Ryoichi; KITA, Hiroyuki. **Hybrid forecasting system based on data area division and deep learning neural network for short-term wind speed forecasting.** Energy Conversion and Management, v. 238, p. 114136, 2021.
- PAULA, Matheus *et al.* **Predicting long-term wind speed in wind farms of northeast brazil: A comparative analysis through machine learning models.** IEEE Latin America Transactions, v. 18, n. 11, p. 2011-2018, 2020.
- ALTAN, Aytaç; KARASU, Seçkin; ZIO, Enrico. **A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer.** Applied Soft Computing, v. 100, p. 106996, 2021.
- POSIT. **RStudio Desktop.** Disponível em: <https://posit.co/download/rstudio-desktop/>. Acesso em: 24 jul. 2023.
- HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: Principles and Practice.** 2021. Disponível em: <https://otexts.com/fpp3/>. Acesso em: 17 setembro. 2022.
- SILVA, Fernando da. **Séries temporais no R.** 2022. Disponível em: [www.youtube.com/watch?v=8gO48HnCE7o&ab](http://www.youtube.com/watch?v=8gO48HnCE7o&ab). Acesso em: 08 out. 2022.
- FERREIRA, Pedro Guilherme Costa *et al.* **Análise de séries temporais em R: curso introdutório.** 2017. Disponível em: <https://research.ebsco.com/c/x7wivp/>. Acesso em: 08 out. 2022.
- RIBEIRO, Matheus Henrique dal Molin. **Time series forecasting based on ensemble learning methods applied to agribusiness, epidemiology, energy demand, and renewable energy.** 2021. 264 f. Tese (Doutorado) - Curso de Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba, 2021.
- LI, Youjuan; ZHU, Ji. **L 1-norm quantile regression.** Journal of Computational and Graphical Statistics, v. 17, n. 1, p. 163-185, 2008
- BREIMAN, Leo. **Random forests.** Machine learning, v. 45, p. 5-32, 2001.
- CHEN, Tianqi; GUESTRIN, Carlos. **Xgboost: A scalable tree boosting system.** In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016. p. 785-794.