

## **SOLAR IRRADIATION FORECASTING BY THE APPLICATION OF FIVE MACHINE LEARNING ALGORITHMS**

**Felipe P. Marinho**

**Juliana S. Brasil**

**Juarez P. A. Neto**

**Paulo A. C. Rocha**

**Maria E. V. Silva**

**Ricardo J. P. Lima**

[felipe.pinto.marinho@gmail.com](mailto:felipe.pinto.marinho@gmail.com)

[julianasbra@gmail.com](mailto:julianasbra@gmail.com)

[juarezneto33@hotmail.com](mailto:juarezneto33@hotmail.com)

[paulo.rocha@ufc.br](mailto:paulo.rocha@ufc.br)

[eugenia@ufc.br](mailto:eugenia@ufc.br)

[riponteslima@gmail.com](mailto:riponteslima@gmail.com)

*Mechanical Engineering Department, Technology Center, Federal Univesity of Ceara  
Campus do Pici, 60440-593, Ceará, Brazil*

**Abstract.** In this work, the forecast of global solar irradiation for a one-day ahead forecast horizon was carried out using some machine learning models, namely: Minimal Learning Machine, Support Vector Machine, Random Forests, K- Nearest Neighbors and a regression tree with the application of a Bagging procedure. The Minimal Learning Machine algorithm is a relatively recent method based on the distance calculation between vectors and used for supervised learning purposes in both classification and regression problems. In addition, we used a data set with the presence of attributes (predictors) formed by exogenous variables (insolation, air temperature, precipitation, etc.), endogenous variables (solar irradiation historical data) and temporal variables (year, month and day of measurement) totalizing 44 attributes and 3254 observations. The root mean squared error and forecast skill obtained by applying the Minimal Learning Machine in the validation set were respectively 40.882 W/m<sup>2</sup> and 7.637 %, and the arithmetic mean of the root mean squared error in conjunction with the arithmetic mean of the forecast skill obtained by the use of the other models for the same validation set were 40.752 W/m<sup>2</sup> and 7.93 %. In this way, it can be drawn by the evaluation of the results that the Minimal Learning Machine presents a performance comparable to the classic machine learning methods. Furthermore, it presents the advantage in the training stage of using only a single adjustment parameter.

**Keywords:** Solar irradiation forecast, Machine learning, Global irradiation, Minimal learning machine, Renewable energy

## 1 Introdução

O potencial de inserção das energias renováveis na matriz energética global, em especial da energia solar, torna necessária uma adequada gestão de fornecimento de energia pelo órgão de controle, uma vez que se tratam de fontes de comportamento estocástico e intermitente Notton et al. [1]. Ter uma previsão confiável da radiação futura na região de um parque solar permite ao operador da rede elétrica avaliar em que momentos serão necessárias complementações do abastecimento com outras fontes. O operador da rede precisa de previsões confiáveis em diferentes horizontes temporais, aquelas feitas em intervalos de horas e minutos são adequadas para prever quedas de fornecimento, enquanto as realizadas em intervalos de dias se destacam na programação de manutenções preventivas Qing e Niu [2]. Problemas relacionados à flutuação de tensão, qualidade de energia e estabilidade da rede elétrica podem ser evitados por meio da previsão de irradiação.

Neste sentido, muitas metodologias têm sido utilizadas para que previsões de irradiação solar em diversos horizontes de tempo possam ser realizadas: Uso de modelos numerical weather prediction (NWP) Mejia, Giordano e Wilcox [3]. Nonnenmacher, Kaur e Coimbra [4], análise de séries temporais Trapero, Kourentzes e Martin [5]. Dong et al. [6], aplicações envolvendo processamento de imagens Pedro e Coimbra [7]. Pedro et al. [8]. Pawar et al. [9] utilização de algoritmos de aprendizagem de máquina Koo et al. [10]. Yagli, Yang e Srinivasan [11]. Benali et al. [12], etc.

O presente trabalho, realizou previsões de irradiação solar para um horizonte de um dia a posteriori por meio da aplicação de modelos de aprendizagem de máquina em um conjunto de dados com a presença de atributos formados por variáveis exógenas (insolação, temperatura do ar, precipitação, etc.), variáveis endógenas (dados históricos de irradiação) e variáveis temporais (ano, mês e dia da medição) totalizando 44 atributos 3254 observações, além disso, avaliou-se o desempenho do modelo de aprendizagem supervisionada Minimal Learning Machine (MLM) Junior et al. [13] para a realização de tais previsões, comparando a sua acurácia com a de outros métodos clássicos e bem postos, a saber: K-Nearest Neighbors (KNN), Support Vector Regression (SVR) Carrasco, López e Maldonado [14]. Drucker et al. [15], Random Forests (RF) e uma árvore de regressão com a aplicação de um procedimento de Bagging (BAGGING). Além do modelo de baseline, que foi utilizado como um critério de referência mínimo de desempenho. A comparação da acurácia entre os métodos foi feita pela avaliação das métricas de erro Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Root Mean Squared Error (rRMSE) e o Forecast Skill (s).

Os métodos KNN, SVR e o Random Forests são frequentemente aplicados na literatura para previsões de irradiação Sharma et al. [16]. O MLM, entretanto, é um modelo relativamente recente que ainda carece de trabalhos voltados para a sua aplicação Júnior et al. [13]. Walch et al. [17]. Desta forma, o artigo traz o uso e a avaliação de um algoritmo que ainda não foi utilizado para propósitos de previsão de irradiação.

O restante do artigo é formado pela seção 2, que trata de uma breve explicação sobre a aquisição e a organização do conjunto de dados assim como o procedimento aplicado para a realização do treinamento dos algoritmos também é apresentado. Nesta mesma seção, os modelos serão detalhados e as métricas de erro utilizadas serão definidas. Posteriormente, na seção 3 tem-se a apresentação dos resultados e a seção 4 trata das conclusões obtidas.

## 2 Metodologia

Nesta seção, tem-se uma breve explicação sobre a aquisição e a organização do conjunto de dados bem como os métodos são detalhados, além de se ter a apresentação das métricas de erro aplicadas.

### 2.1 Dados

Os dados foram obtidos pela estação meteorológica da Universidade Federal do Ceará, na cidade de Fortaleza-CE, Brasil, 3°43'6"S e 38°32'36'O, ao nível do mar, entre os anos de 1969 à 1979 e 1995 à 1997. Dentre as variáveis exógenas foi considerado a influência dos fenômenos climáticos do El Niño

e da La Niña. Tal atributo foi adicionado ao conjunto de dados pela atribuição de valores de -4 a 4, no qual os valores extremos representam a intensidade máxima destes, respectivamente, e o zero representa uma condição de neutralidade, como indicado pela tabela 1. As intensidades desses fenômenos foram coletadas pelo Golden Gate Weather Service, serviço estadunidense [18], que avalia seus níveis pelo Índice de Niño Oceânico (ONI), obtido com base na diferença entre a temperatura medida e um valor pré-determinado para a região do pacífico com coordenadas 5°N-5°S, 120°E-170°W.

Tabela 1. Intensidades do El Niño e da La Niña

El Niño			La Niña		
Intensidades	Aumento característico da temperatura	Valor atribuído	Intensidades	Redução característica na temperatura	Valor atribuído
Muito Forte	$T \geq 2^\circ$	4	Muito Forte	$T \leq -2^\circ$	-4
Forte	$1.5^\circ \leq T \leq 1.9^\circ$	3	Forte	$-1.5^\circ \geq T \geq -1.9^\circ$	-3
Moderado	$1^\circ \leq T \leq 1.4^\circ$	2	Moderado	$-1^\circ \geq T \geq -1.4^\circ$	-2
Fraco	$0.5^\circ \leq T \leq 0.9^\circ$	1	Fraco	$-0.5^\circ \geq T \geq -0.9^\circ$	-1
Neutro	$T < 0.5^\circ$	0	Neutro	$T > -0.5^\circ$	0

A irradiação solar global foi captada com o uso de um actinógrafo R. Fuess Berlin-Steglitz, com numeração G6714. Este instrumento apresenta uma precisão na faixa de 15% a 20% com uma sensibilidade que compreende todo o espectro de radiação, com exceção das regiões do espectro que são absorvidas pela cúpula de vidro presente no aparato (ultravioleta e infravermelho distante, bandas que pouco contribuem para a energia total).

Para os preditores pressão, temperatura do ar, temperatura de bulbo úmido, umidade relativa, velocidade do vento e nebulosidade foram consideradas medições às 9h e 15h de cada dia. As irradiações dos 20 dias anteriores e a data da aquisição de cada observação (ano, mês e dia) também foram considerados atributos.

Para o treinamento dos modelos, o conjunto de dados foi dividido aleatoriamente de tal forma que 70% ficou para o treinamento dos algoritmos e o restante constituiu o conjunto de validação. O ajuste dos métodos foi feito por meio da aplicação de validação cruzada 10-Fold ao conjunto de treinamento. Esta etapa de treinamento é representada pela figura 1.

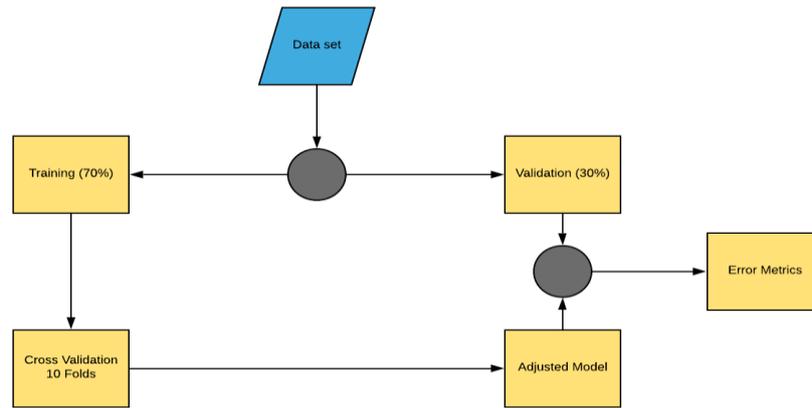


Figura 1. Representação esquemática da etapa de treinamento

## 2.2 Modelo de baseline

O modelo de baseline utilizado atribui para uma nova observação a irradiação média do conjunto de treinamento. Seja  $T = \{G_1, G_2, \dots, G_n\}$  os valores de irradiação do conjunto de treinamento. Desta forma, o valor atribuído para cada nova observação é dado pela eq. (1).

$$G = (\sum_{i=1}^n G_i) / n . \quad (1)$$

## 2.3 K-Nearest Neighbors (KNN)

Este conhecido método é geralmente aplicado em problemas de classificação, porém o mesmo admite uma versão para problemas de regressão, onde o algoritmo identifica os  $K$  vizinhos, que estão no conjunto de treinamento, mais próximos de uma nova observação por meio do uso de uma métrica de dissimilaridade, neste trabalho utilizou-se a distância euclidiana, e atribui a média aritmética das irradiações de tais vizinhos. Neste caso, a irradiação estimada para a nova observação será como na eq. (2).

$$G = (\sum_{i=1}^K G_i) / K . \quad (2)$$

Onde,  $G_i$ , com  $i = \{1, 2, \dots, K\}$  são as irradiações dos  $K$  vizinhos mais próximos. O número de vizinhos  $K$  é, portanto, o único hiperparâmetro deste modelo.

## 2.4 Support Vector Regression (SVR)

Este modelo por mais que seja uma versão do Support Vector Machine (SVM), apresenta a diferença de ser utilizado em problemas de regressão, enquanto o SVM é um modelo comumente aplicado no contexto de classificação. Uma das grandes vantagens em se trabalhar com o SVR se deve a possibilidade de construção de funções não lineares complexas, além de se ter a garantia de otimalidade global via a solução de um programa quadrático convexo (QP), treinamento efetivo e, geralmente, apresentando um desempenho preditivo positivo com um reduzido risco de overfitting graças ao uso do princípio da minimização do risco estrutural. Neste estudo, o algoritmo SVR aplicado foi do tipo  $\varepsilon$ -SVR, onde ocorre a construção de um  $\varepsilon$ -tubo em que os erros são ignorados no seu interior, mas penalizados fora do mesmo Carrasco, López e Maldonado [14]. Drucker et al. [15].

Dado  $x_i \in R^n$  e as saídas correspondentes  $y_i \in R$ , para  $i = \{1, 2, \dots, m\}$  o  $\varepsilon$ -SVR linear busca obter um hiperplano  $f(x) = w^T x + b$  pela construção de um  $\varepsilon$ -tubo como explicado anteriormente, para tanto, faz-se necessário a solução do QP .

$$\begin{aligned} \min_{w,b,\xi,\omega} \quad & (1/2)\|w\|^2 + Ce^T(\xi + \omega) \\ \text{s.t.} \quad & y - (Aw + be) \leq \varepsilon e + \xi, \quad \xi \geq 0 \\ & (Aw + be) - y \leq \varepsilon e + \omega, \quad \omega \geq 0 \end{aligned}$$

Onde,  $A = [x_1 \ x_2 \ \dots \ x_m]^T \in R^{m \times n}$ ,  $y = (y_1, y_2, \dots, y_m) \in R^m$ ,  $e \in R^m$  é um vetor de uns,  $\xi, \omega \in R^m$  são variáveis de folga que penalizam as observações que estão fora do  $\varepsilon$ -tubo e  $C > 0$  é um fator que regulariza o trade-off de maior ou menor flexibilidade do modelo, como pode-se notar tanto a função objetivo como as funções restrições de desigualdade são convexas, portanto, o QP é convexo e pode ser resolvido pela aplicação de algum método eficiente para a solução de problemas convexas, como o método dos pontos interiores.

O  $\varepsilon$ -SVR pode ser estendido para uma configuração não linear pelo uso de kernels, isto pode ser feito pelo uso de um artifício no problema dual correspondente Drucker et al. [15]. A versão de kernels para o  $\varepsilon$ -SVR é dada por

$$\begin{aligned} \min_{\alpha,\tau} \quad & (1/2)(\alpha - \tau)^T K(A, A^T)(\alpha - \tau) - y^T(\alpha - \tau) + \varepsilon e^T(\alpha + \tau) \\ \text{s.t.} \quad & e^T(\alpha - \tau) = 0 \\ & \alpha \geq 0, \quad \tau \leq Ce \end{aligned}$$

Onde,  $K(A, A^T) \in R^{m \times m}$  representa a matriz de funções kernel da forma  $k_{is} = \kappa(x_i, x_s)$ ,  $\kappa : R^n \times R^n \rightarrow R$  e  $\alpha, \tau$  são vetores de variáveis duais (multiplicadores de lagrange do problema primal). O kernel utilizado neste estudo foi o radial gaussiano que apresenta as funções de kernel da forma dada pela eq.(3)

$$\kappa(x_i, x_s) = \exp(-(1/2\sigma^2) \|x_i - x_s\|^2). \quad (3)$$

Desta forma, os parâmetros livres são o  $\varepsilon$ ,  $C$  e  $\sigma$ , vale ressaltar que neste trabalho foi atribuído  $\sigma = 0.016745$  e  $\varepsilon = 0.1$  e o hiperparâmetro  $C$  foi obtido por validação cruzada 10-Fold no conjunto de treinamento.

## 2.5 Árvore de regressão com a aplicação de um procedimento de Bagging

As árvores de decisão envolvem a segmentação do espaço de atributos por meio da aplicação de regras, as quais podem ser resumidas em um formato de árvore, proporcionando o desenvolvimento de métodos simples e de fácil interpretação. Entretanto, estas tipicamente não são competitivas quando se compara sua acurácia de previsão com os de outros modelos clássicos de aprendizagem supervisionada. Desta forma, existem algumas abordagens que envolvem a formação de múltiplas árvores que são combinadas para conduzir a uma única previsão, algo similar ao que ocorre com as máquinas de comitê, tal combinação pode resultar em melhorias significativas na acurácia de previsão, mas com uma perda na capacidade de interpretação James et al. [19]. Uma destas metodologias de combinação de árvores é o Bootstrap Aggregation (Bagging), que é um método ensemble que combina as saídas de múltiplos algoritmos de aprendizagem de máquina em um único valor, esta metodologia é geralmente utilizada com o intuito de redução da variância de modelos de alta variância, como o das árvores de decisão James et al. [19]. Kuhn e Johnson [20]. O primeiro passo neste procedimento é a realização de um bootstrap no conjunto de dados considerado, que consiste na formação de sub-amostras através da realização de sucessivas amostragens com reposição no conjunto de dados original. Posteriormente, para cada sub-amostra desenvolve-se uma árvore de decisão e, por fim, as saídas resultantes de cada árvore serão combinadas por uma média aritmética para a obtenção do resultado final. Todo este procedimento está representado na figura 2

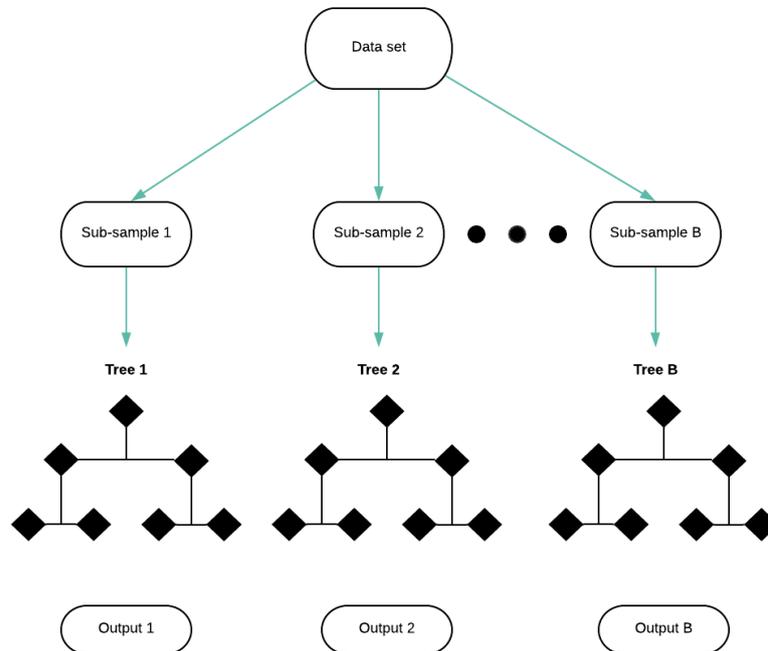


Figura 2. Representação esquemática de um Bagging aplicado em árvore de decisão

Onde o vetor  $O = \{Output\ 1, Output\ 2, \dots, Output\ B\} = \{f_1^*(x), f_2^*(x), \dots, f_B^*(x)\}$ . Assim a saída resultante será dada pela eq. (4).

$$f^*(x) = (\sum_{i=1}^B f_i^*(x))/B. \quad (4)$$

Neste caso, percebe-se que o único hiperparâmetro presente neste método é o número de árvores considerada, onde pode-se assumir o valor arbitrariamente grande para garantir bons resultados, mas tendo sempre cuidado para que o tempo computacional não se torne elevado. Para o presente estudo, considerou-se 200 árvores.

## 2.6 Random Forests

O Random Forests é análogo ao Bagging, mas com o uso de uma pequena modificação visando descorrelacionar as árvores James et al. [19]. Uma vez que, no Bagging para cada sub-amostra ocorre o desenvolvimento de uma árvore por meio da segmentação do mesmo espaço de preditores, este procedimento tende a ocasionar a obtenção de respostas que apresentam um elevado grau de correlação, mas em algumas situações combinações de previsões de vários modelos fornecem melhores resultados quando as saídas dos sub-modelos são descorrelacionadas ou na melhor das hipóteses apresentam um nível fraco de correlação. Basicamente, a proposta do Random Forests é a de que para a formação de cada árvore seja feita a divisão do espaço de  $m$  atributos selecionados aleatoriamente de um total de  $p$  preditores, onde tipicamente seleciona-se  $m = \sqrt{p}$ . Entretanto, por meio da aplicação do processo de aprendizagem é possível obter o valor de  $m$  ótimo.

## 2.7 Minimal Learning Machine

O algoritmo Minimal Learning Machine foi originalmente apresentado nos trabalhos de Junior et al. [13]. A metodologia consiste na construção de um mapeamento linear (regressão linear múltipla) entre as matrizes distâncias de entrada e saída. Posteriormente, o mapeamento linear ajustado (ótimo) é utilizado para estimar as distâncias de  $K$  pontos de referência no espaço das saídas em relação à resposta que deseja-se determinar. Por fim, a saída é estimada via a solução de um problema de multilateração. A formulação geral do MLM é perfeitamente capaz de trabalhar com problemas de regressão não linear, bem como, no contexto em que o espaço das saídas é multidimensional [13]. Seja um conjunto com  $N$  observações com atributos  $X = \{x_i\}$ ,  $i = \{1, 2, \dots, N\}$ ,  $x_i \in R^D$  e o conjunto com as saídas correspondentes  $Y = \{y_i\}$ ,  $i = \{1, 2, \dots, N\}$ ,  $y_i \in R^S$ . Assume-se a existência de um mapeamento contínuo entre o espaço de entrada e o de saída  $f: X \rightarrow W$  dado pela eq. (5)

$$y = f(x) + \varepsilon . \quad (5)$$

Com  $y$  e  $x$  sendo vetores colunas e  $\varepsilon \in R^{N \times S}$  é a matriz de resíduos. O MLM opera em duas etapas

- Ajuste de uma regressão linear múltipla para obter matrizes distância de saída a partir de matrizes distância de entrada
- Estima a resposta desejada por meio da solução de um problema de multilateração

### **Observação 1: Regressão linear múltipla entre matrizes distância**

Esta etapa, inicia-se com a escolha aleatória de algumas observações do conjunto de dados onde as mesmas serão denominadas pontos de referência. Vale ressaltar que considera-se para cada ponto de referência os seus atributos e suas saídas correspondentes de forma separada, formando assim os conjuntos  $\mu = \{m_k\}$ ,  $k = \{1, 2, \dots, K\}$ ,  $\mu \subset X$  que é constituído pelos atributos dos pontos de referência e o conjunto  $T = \{t_k\}$ ,  $k = \{1, 2, \dots, K\}$ ,  $T \subset Y$  das saídas correspondentes. Define-se a matriz  $D_x \in R^{N \times K}$ , tal que sua  $k$ -ésima coluna representa as distâncias  $d(x_i, m_k)$ ,  $i = \{1, 2, \dots, N\}$  entre os atributos  $x_i$  e os preditores do  $k$ -ésimo ponto de referência. De forma análoga, tem-se a matriz  $\Delta_y \in R^{N \times K}$ , tal que sua  $k$ -ésima coluna representa as distâncias  $\delta(y_i, t_k)$ ,  $i = \{1, 2, \dots, N\}$  entre a saída  $y_i$  e a resposta do  $k$ -ésimo ponto de referência. Basicamente, assume-se que há uma relação linear entre as matrizes distância da entrada e matrizes distância de saída, obtendo assim um problema de regressão linear múltipla, como representado pela eq. (6)

$$\Delta_y = D_x B + E . \quad (6)$$

Com  $B \in R^{K \times K}$  sendo a matriz dos coeficientes. O ajuste deste modelo pode ser feito por meio da solução de um problema de mínimos quadrados, ou seja, pela minimização da função, dada pela eq. (7)

$$RSS(B) = tr((\Delta_y - D_x B)^T (\Delta_y - D_x B)) . \quad (7)$$

Para o trabalho em questão, o número de observações é tal que  $N > K$ . Desta forma, a solução por mínimos quadrados fornece a eq. (8)

$$B^* = (D_x^T D_x)^{-1} D_x^T \Delta_y . \quad (8)$$

Para o cálculo da inversa na eq. (8), recomenda-se determinar a pseudo-inversa de Moore-Penrose, Ben-Israel e Greville [21]. Penrose [22]. Com isto, para uma observação de teste com atributos  $x \in R^D$ , calcula-se as distâncias em relação aos vetores  $\{m_k\}$ ,  $k = \{1, 2, \dots, K\}$  obtendo assim o vetor  $d(x, \mu) = [d(x, m_1), d(x, m_2), \dots, d(x, m_K)]$ , onde por meio do modelo de regressão linear múltipla ajustado pode

se estimar as distâncias entre a resposta desejada  $y$  e as saídas conhecidas  $\{t_k\}$ ,  $k = \{1, 2, \dots, K\}$  dos pontos de referência, como ilustrado na eq. (9)

$$\delta^*(y, T) = d(x, \mu)B^* . \quad (9)$$

Obtendo assim o vetor  $\delta^*(y, T) = [\delta^*(y, t_1), \delta^*(y, t_2), \dots, \delta^*(y, t_K)]$

**Observação 2: Estimativa das respostas**

Estimar as repostas conhecendo as saídas  $\{t_k\}$ ,  $k = \{1, 2, \dots, K\}$  dos pontos de referência e as estimativas das distâncias  $\delta^*(y, T) = [\delta^*(y, t_1), \delta^*(y, t_2), \dots, \delta^*(y, t_K)]$  pode ser visto como um problema de multilateração, Niewiadomska-Szynkiewicz e Marks [23]. Existem muitas estratégias que podem ser adotadas para a solução de tal problema, a utilizada busca determinar  $y^*$  por meio de uma interpretação geométrica que gera  $K$  equações não lineares relacionadas a consideração de se ter  $K$  hiperesferas  $S$ -dimensional centradas nos  $t_k$  e passando por  $y^*$  com raios iguais a  $\delta^*(y, t_k)$ , como ilustrado na figura 3.

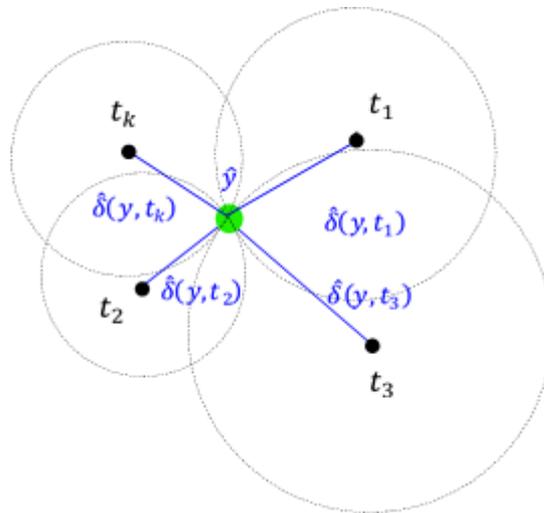


Figura 3. Ilustração da estratégia geométrica utilizada

Desta forma, obtém-se a eq. (10)

$$(y^* - t_k)^T (y^* - t_k) = (\delta^*(y, t_k))^2 . \quad (10)$$

Assim a resposta é estimada pela minimização da função objetivo dada pela eq. (11)

$$J(y) = \sum_{k=1}^K ((y^* - t_k)^T (y^* - t_k) - (\delta^*(y, t_k))^2)^2 . \quad (11)$$

Por ser um problema de mínimos quadrados não linear, recomenda-se o uso do algoritmo de Levenberg-Marquardt, Marquardt [24]. Moré [25]. Para a solução de tal problema de otimização. Assim, percebe-se que o único hiperparâmetro presente no modelo é o número de pontos de referência  $K$ .

**2.8 Métricas de erro**

**Observação 3: Mean Absolute Error (MAE)**

O MAE calcula a média das diferenças absolutas entre o valor previsto  $\hat{y}_i$ , e o valor real,  $y_i$ , isso é, não leva em consideração se o erro é para mais ou para menos e às diferenças absolutas não é atribuído peso, como indica a eq. (12)

$$MAE = (\sum_{i=1}^n |y_i - \hat{y}_i|) / n . \quad (12)$$

**Observação 4: Mean Bias Error (MBE)**

O Mean Bias Error se assemelha ao cálculo do MAE, entretanto se diferencia deste por considerar o sinal do erro, isto é, não calcula o módulo das diferenças. Desse modo, requer prudência em sua análise, uma vez que permite a compensação de erros (erros com sinais distintos). A mesma é dada pela eq. (13)

$$MBE = (\sum_{i=1}^n y_i - \hat{y}_i) / n . \quad (13)$$

**Observação 5: Root Mean Squared Error (RMSE)**

O Root Mean Squared Error calcula a magnitude média do erro pela raiz quadrada da média dos quadrados dos erros, desse modo atribui um peso maior aos erros de maior magnitude, e peso menor aos erros de menor magnitude. É obtido na mesma unidade da variável em análise. Definida pela eq. (14)

$$RMSE = \sqrt{E/n}, \text{ onde } E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 . \quad (14)$$

**Observação 6: Relative Root Mean Squared Error**

O Relative Root Mean Squared Error é a razão entre o RMSE e a média dos dados em análise. Como dado pela eq. (15)

$$rRMSE = RMSE/M, \text{ onde } M = (\sum_{i=1}^n y_i) / n . \quad (15)$$

Tal métrica foi utilizada, porque fornece faixas de classificação de uma previsão: %rRMSE<10% excelente, 10%<%rRMSE<20% bom, 20%<%rRMSE<30% razoável e %rRMSE≥30% ruim como indicado por Li et al. [26].

**Observação 7: Forecast Skill (s)**

Esta métrica avalia o desempenho do modelo aplicado quando comparado com a acurácia do modelo de baseline, sendo definida pela eq. (16)

$$s = 1 - (RMSE/RMSE_B) . \quad (16)$$

Onde,  $RMSE$  representa a métrica raiz do erro quadrático médio do modelo avaliado e o  $RMSE_B$  representa a mesma métrica para o modelo de baseline.

### 3 Resultados e Discussão

Nesta seção serão apresentados inicialmente os resultados relacionados a etapa de treinamento dos modelos. Validação cruzada 10-Fold é utilizada no conjunto de treinamento para se realizar a busca pelos hiperparâmetros ótimos e a métrica  $RMSE$  foi utilizada como critério de seleção, onde aqueles

parâmetros que forneciam menor valor de tal métrica foram escolhidos. Posteriormente, apresenta-se os valores das métricas de erro obtidos pela aplicação dos algoritmos ajustados no conjunto de validação.

### 3.1 Busca por hiperparâmetros

#### *Observação 8: SVR*

Como já havia sido discutido anteriormente, tomou-se  $\sigma = 0,016745$  e  $\varepsilon = 0.1$ . O procedimento de treinamento forneceu um hiperparâmetro de regularização  $C = 0.5$ , os resultados são reportados na tabela 2.

Tabela 2. Busca por hiperparâmetro SVR

C	RMSE ( $W/m^2$ )	MAE ( $W/m^2$ )
0.25	41.031	29.164
<b>0.50</b>	<b>40.090</b>	<b>29.218</b>
1.0	41.158	29.487
2.0	41.573	30.056
4.0	42.331	30.968
8.0	43.717	32.391
16	45.659	34.257
32	47.649	36.004
64	49.600	37.615
128	50.942	38.767

Percebe-se assim que um aumento em  $C$ , ou seja, um aumento no nível de flexibilidade do modelo ocasiona uma piora no desempenho de previsão, como nota-se pelos aumentos nas métricas de erro  $RMSE$  e  $MAE$ . Este fato pode está relacionado à um maior risco de overfitting quando se aumenta a flexibilidade do modelo para o conjunto de dados considerado, esta elevação conduz a menores bias, porém aumenta a variância do método.

#### *Observação 9: Random Forests*

Neste caso, o hiperparâmetro a ser considerado corresponde à dimensão do espaço de preditores que será segmentado,  $m$ , novamente a métrica escolhida para a seleção foi o RMSE. Assim pelo treinamento obteve-se  $m = 15$ . Os resultados são reportados pela tabela 3.

Tabela 3. Busca por hiperparâmetro no RF

m	RMSE ( $W/m^2$ )	MAE ( $W/m^2$ )
2	39.700	30.236
6	39.666	30.192
11	39.694	30.182
15	39.638	30.118
20	39.646	30.102
24	39.699	30.122
29	39.728	30.136
33	39.718	30.135
38	39.767	30.158
43	39.802	30.145

Neste caso, nota-se uma menor sensibilidade do modelo à variação da dimensão do espaço de preditores, como se percebe pela pequena variação no valor das métricas de erro com a alteração do parâmetro  $m$ .

#### **Observação 10: Bagging**

Como já havia sido abordado, o método que utiliza Bagging apresenta como único hiperparâmetro o número de árvores a ser considerado, tal parâmetro não é crítico uma vez que um valor elevado de árvores não irá conduzir a um overfitting James et al. [19]. Entretanto, um número exagerado pode conduzir a um alto tempo de processamento. Neste trabalho, utilizou-se 200 árvores, que foi o valor com a melhor compatibilidade entre desempenho e tempo computacional.

#### **Observação 11: KNN**

Aqui o único hiperparâmetro a ser considerado é o número de vizinhos  $K$ , os resultados para a busca do melhor ajuste é ilustrado na tabela 4.

Tabela 4. Busca pelo hiperparâmetro no KNN

K	RMSE ( $W/m^2$ )	MAE ( $W/m^2$ )
5	42.701	31.484
7	41.903	30.815
9	41.354	30.497
11	41.006	30.216
13	40.776	30.044
15	40.606	29.937
17	40.506	29.853
19	40.430	29.801
21	40.351	29.725
<b>23</b>	<b>40.279</b>	<b>29.671</b>

Para este modelo, percebe-se que o aumento no número de vizinhos melhora a acurácia de previsão do método quando se aplica o mesmo no conjunto de dados considerado, indicando que uma redução na flexibilidade do modelo é benéfica para o problema em questão o que concorda com os resultados obtidos no treinamento do SVR.

**Observação 12: MLM**

Similar ao método KNN, o único hiperparâmetro para este modelo é o número de pontos de referência  $K$ , o qual obteve-se o valor de  $K_{ótimo} = 85$ , os resultados obtidos por validação cruzada 10-fold no conjunto de treinamento são observados na figura 4.

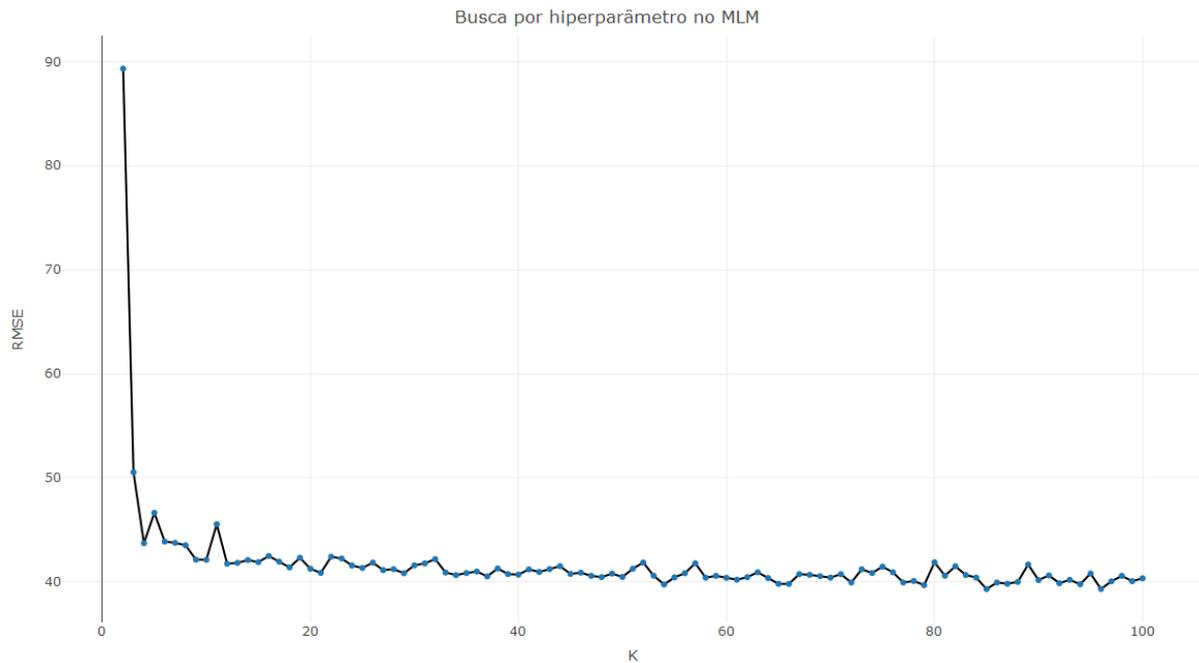


Figura 4. Busca por hiperparâmetro no MLM

### 3.2 Resultados obtidos pela aplicação dos modelos no conjunto de validação

Os valores das métricas de erro obtidas pela aplicação dos algoritmos ajustados no conjunto de treinamento e validação estão reportados na tabela 5.

Tabela 5. Desempenho dos modelos

Aplicação os modelos ajustados no conjunto de treinamento e validação										
Métodos	Conjunto de treinamento					Conjunto de validação				
	MBE (W/m)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	rRMSE (%)	s (%)	MBE (W/m <sup>2</sup> )	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	rRMSE (%)	s (%)
SVR	-7.765	24.769	37.186	18.710	15.987	-8.336	29.433	41.194	20.860	6.930
<b>RF</b>	<b>0.537</b>	<b>12.027</b>	<b>16.270</b>	<b>8.187</b>	<b>63.241</b>	<b>1.077</b>	<b>30.400</b>	<b>40.234</b>	<b>20.374</b>	<b>9.10</b>
Bagging	-0.139	28.376	36.948	18.591	16.523	-0.619	30.614	40.727	20.623	7.987
KNN	-2.776	28.435	38.689	19.466	12.592	-3.679	29.933	40.854	20.687	7.701
MLM	-0.882	29.278	38.993	19.589	12.039	-1.970	30.559	40.882	20.701	7.637

Pela análise dos resultados, percebe-se que estes modelos apresentam um desempenho bem condizente ou até mesmo superior àqueles encontrados nos trabalhos de Mejia, Giordano e Wilcox [3]. Nonnenmacher, Kaur e Coimbra [4] onde nos mesmos também se realizam previsões de irradiação para um horizonte de previsão de 1 dia a posteriori, mas com a utilização de modelos NWP, que se baseiam em princípios derivados da mecânica dos fluidos, termodinâmica e transferência de calor aplicados no contexto de escoamentos de ar atmosférico, tais modelos geralmente apresentam um bom desempenho para a realização de previsões de longo prazo que são aquelas para horizontes de previsão superior à 6 horas, enquanto modelos de aprendizagem de máquinas e também métodos que se baseiam em análise de séries temporais são considerados metodologias dominantes quando se deseja realizar previsões de irradiação de curto prazo, que são aquelas com horizontes temporais na faixa de 5 minutos até 6 horas. De fato o estudo realizado indica que para o conjunto de dados considerado os modelos de aprendizagem

de máquina utilizados apresentam um bom desempenho quando comparado com os obtidos em Mejia, Giordano e Wilcox [3]. Nonnenmacher, Kaur e Coimbra [4] mesmo para um horizonte de 1 dia a posteriori que caracteriza previsão de irradiação de longo prazo, este fato deve-se a utilização de atributos que são estatisticamente significativos para a descrição da irradiação como a insolação e a temperatura ambiente, Medeiros, Silva e Bezerra [27], bem como do fato de se ter condições climáticas estáveis na região nordeste do Brasil, região de aquisição dos dados, o que contribui para uma redução da variabilidade nos valores de irradiação e, portanto, proporciona melhores resultados de previsão. Além disso, percebe-se que para este conjunto de dados o MLM apresenta uma acurácia competitiva com os demais modelos que são considerados clássicos e bem postos, como já havia sido discutido em Júnior et al. [13]. Desta forma, tem-se um novo modelo de fácil interpretação e implementação que apresenta um desempenho competitivo para a previsão de irradiação solar para um dia posteriori.

## 4 Conclusão

No presente trabalho, realizou-se previsões de irradiação solar para um horizonte de um dia a posteriori utilizando cinco modelos de aprendizagem de máquina, a saber: Minimal Learning Machine, K-Nearest Neighbors, Support Vector Regression, Random Forests e uma árvore de regressão com a aplicação de um procedimento de Bagging. A avaliação do desempenho de tais métodos foi feita pela utilização das métricas de erro Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Root Mean Squared Error (rRMSE) e Forecast Skill (s). Ressaltando que o algoritmo Minimal Learning Machine é relativamente recente e ainda não tinha sido aplicado para propósitos de previsão de irradiação solar para um dia a posteriori. Os valores das métricas RMSE e s para o método MLM foram de 40,882 W/m<sup>2</sup> e 7,637%, enquanto os valores das médias aritméticas destas métricas para os demais modelos foram de 40,752 W/m<sup>2</sup> e 7,93%. Desta forma, percebeu-se que este algoritmo apresenta um desempenho competitivo quando se compara com os demais métodos, além disso, os resultados obtidos pela realização das previsões foram superiores ao de alguns trabalhos na literatura indicando o uso de atributos significativos estatisticamente para descrever a irradiação solar, bem como pode refletir o fato de se ter estabilidade climática na região nordeste, local de aquisição das observações.

## Agradecimentos

Esta pesquisa contou com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Conselho Nacional de desenvolvimento Científico e Tecnológico (CNPq), agências governamentais brasileiras.

## Referências

- [1] G. Notton et al. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. *Renewable and Sustainable Energy Reviews*, v. 87, p. 96-105, 2018..
- [2] X. Qing and Y. Niu. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, v. 148, p. 461-468, 2018.
- [3] J. F. Mejia, M. Giordano and E. Wilcox. Conditional summertime day-ahead solar irradiance forecast. *Solar Energy*, v. 163, p. 610-622, 2018.
- [4] L. Nonnenmacher, A. Kaur and C.F.M. Coimbra. Day-ahead resource forecasting for concentrated solar power integration. *Renewable energy*, v. 86, p. 866-876, 2016.
- [5] J. R. Trapero, N. Kourentzes and A. Martin. Short-term solar irradiation forecasting based on dynamic harmonic regression. *Energy*, v. 84, p. 289-295, 2015.

- [6] Z. Dong et al. Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy*, v. 55, p. 1104-1113, 2013.
- [7] H.T.C. Pedro and C.F.M. Coimbra. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. *Renewable energy*, v. 80, p. 770-782, 2015.
- [8] H.T.C. Pedro et al. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. *Renewable Energy*, v. 123, p. 191-203, 2018.
- [9] P. Pawar et al. Detecting clear sky images. *Solar Energy*, v. 183, p. 50-56, 2019.
- [10] C. Koo et al. A novel estimation approach for the solar radiation potential with its complex spatial pattern via machine-learning techniques. *Renewable Energy*, v. 133, p. 575-592, 2019.
- [11] G. M. Yagli; D. Yang and D. Srinivasan. Automatic hourly solar forecasting using machine learning models. *Renewable and Sustainable Energy Reviews*, v. 105, p. 487-498, 2019.
- [12] L. Benali et al. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable energy*, v. 132, p. 871-884, 2019.
- [13] A.H.S. Júnior et al. Minimal learning machine: a novel supervised distance-based approach for regression and classification. *Neurocomputing*, v. 164, p. 34-44, 2015.
- [14] M. Carrasco; J. López and S. Maldonado. Epsilon-nonparallel support vector regression. *Applied Intelligence*, p. 1-14, 2019.
- [15] H. Drucker et al. Support vector regression machines. In: *Advances in neural information processing systems*. p. 155-161, 1997.
- [16] N. Sharma et al. Predicting solar generation from weather forecasts using machine learning. In: *2011 IEEE international conference on smart grid communications (SmartGridComm)*. IEEE, 2011. p. 528-533.
- [17] A. Walch et al. Spatio-temporal modelling and uncertainty estimation of hourly global solar irradiance using Extreme Learning Machines. *Energy Procedia*, v. 158, p. 6378-6383, 2019.
- [18] <<https://ggweather.com/enso/oni.htm>>, Acesso em 17. mai. 2019.
- [19] G. James et al. *An introduction to statistical learning*. New York: springer, 2013.
- [20] M. Kuhn and K. Johnson. *Applied predictive modeling*. New York: Springer, 2013.
- [21] A. Ben-Israel and T.N.E. Greville. *Generalized inverses: theory and applications*. Springer Science & Business Media, 2003.
- [22] R. Penrose. A generalized inverse for matrices. In: *Mathematical proceedings of the Cambridge philosophical society*. Cambridge University Press, p. 406-413, 1955.
- [23] E. Niewiadomska-Szynkiewicz; M. Marks. Optimization schemes for wireless sensor network localization. *International Journal of Applied Mathematics and Computer Science*, v. 19, n. 2, p. 291-302, 2009.
- [24] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, v. 11, n. 2, p. 431-441, 1963.
- [25] J.J. Moré. The Levenberg-Marquardt algorithm: implementation and theory. In: *Numerical analysis*. Springer, Berlin, Heidelberg, p. 105-116, 1978.
- [26] M.F. Li et al. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy conversion and management*, v. 70, p. 139-148, 2013.
- [27] F.J.D. Medeiros; C.M. Silva and B. G. Bezerra. Calibration of Ångström-Prescott Equation to Estimate Daily Solar Radiation on Rio Grande do Norte State, Brazil. *Revista Brasileira de Meteorologia*, v. 32, n. 3, p. 409-416, 2017.