

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

INFLUÊNCIA DOS FENÔMENOS CLIMÁTICOS DO EL NIÑO E DA LA NIÑA NA PREVISÃO DA MÉDIA DIÁRIA DE IRRADIAÇÃO GLOBAL NA CIDADE DE FORTALEZA

INFLUENCE OF CLIMATE PHENOMENA OF EL NIÑO AND LA NIÑA ON THE FORECAST OF THE DAILY AVERAGE OF GLOBAL IRRADIATION IN THE CITY OF FORTALEZA

Felipe Pinto Marinho, MSc[1], Juliana Silva Brasil, MSc[2], Paulo Alexandre Costa Rocha, Ph.D[3], Maria Eugênia Vieira da Silva, Ph.D[3], Juarez Pompeu de Amorim Neto, MSc[4]

e2253
https://doi.org/10.47820/acertte.v2i2.53

RESUMO

Neste trabalho, previsões da média diária de irradiação solar global foram obtidas pela aplicação de algoritmos de aprendizagem de máquina em dois conjuntos de dados formados por variáveis exógenas (insolação, temperatura do ar, precipitação), variáveis endógenas (série temporal da média diária de irradiação solar global) e variáveis temporais (ano, mês e dia da medição). A diferença entre os conjuntos de dados está relacionada ao fato de que em um se considera as intensidades dos fenômenos climáticos do El Niño e da La Niña como preditores para os modelos de aprendizagem utilizados, enquanto no outro não se considera. Desta forma, foi possível avaliar se a adição do preditor relacionado ao El Niño/La Niña contribui para uma melhor acurácia de previsão por parte dos modelos aplicados: Máquina de Aprendizagem Mínima, Regressão por Vetor Suporte, Florestas Aleatórias, K-Vizinhos mais próximos e uma árvore de regressão com o uso de *Bootstrap.* As métricas de erro Erro Médio Absoluto, Erro de Viés Médio, Raiz do Erro Quadrático Médio, Raiz do Erro Quadrático Médio Relativo e Habilidade de Previsão foram utilizadas para a análise do desempenho dos algoritmos. As médias aritméticas da Raiz do Erro Quadrático Médio e da Habilidade de Previsão, para o caso em que se considerou o El Niño/La Niña como atributos, foram de 40.78 W/m² e 7,87%, respectivamente. Para o caso em que não se considera tais preditores os valores obtidos foram de 40.86 W/m² e 7.69%, indicando que o uso destes preditores aumenta a acurácia de previsão dos algoritmos em questão.

PALAVRAS-CHAVE: Energia Solar. Aprendizagem de Máquina. Previsão.

ABSTRACT

In this work, predictions of the daily average of global solar irradiation were obtained by the application of machine learning algorithms in two data sets formed by exogenous variables (insolation, air temperature, precipitation), endogenous variables (time series of the daily average of irradiation global solar) and temporal variables (year, month and day of measurement). The difference between the data sets is related to the fact that in one the intensities of El Niño and La Niña weather phenomena are considered as predictors for the learning models used, while in the other it is not considered. In this way, it was possible to evaluate whether the addition of the predictor related to El Niño/La Niña contributes to a better prediction accuracy by the applied models: Minimum Learning Machine, Support Vector Regression, Random Forests, K-Nearest Neighbors and a regression tree using Bootstrap. The error metrics Mean Absolute Error, Mean Bias Error, Root Mean Square Error, Root Mean Square Error and Forecast Ability were used to analyze the performance of the algorithms. The arithmetic mean of the Root Mean Square Error and the Prediction Ability for the case in which El Niño/La Niña were considered as attributes were 40.78 W/m² and 7.87%, respectively. For the case in which such predictors are *not considered, the values obtained were 40.86 W/m² and 7.69%, indicating that the use of these predictors increases the prediction accuracy of the algorithms in question.*

KEYWORDS*: Solar Energy. Machine Learning. Forecast*

[1] Departamento de Engenharia de Teleinformática da Universidade Federal do Ceará
[2] Departamento de Engenharia Mecânica da Universidade de São Paulo
[3] Departamento de Engenharia Mecânica da Universidade Federal do Ceará
[4] Mestre em Engenharia Mecânica pela Universidade Federal do Ceará



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

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 pelos órgãos de controle, uma vez que se trata de fontes com comportamento estocástico e intermitente (BELLMAN, 1961). Ter previsões confiáveis do recurso solar 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 (QUING; NIU, 2018). Problemas relacionados a flutuações de tensão, qualidade de energia e estabilidade da rede elétrica podem ser evitados por meio da previsão de irradiação solar.

Neste sentido, muitas metodologias têm sido utilizadas para que previsões de irradiação solar em diversos horizontes de tempo possam ser efetuadas: Uso de modelos *Numerical Weather Prediction* (MEJIA; GIORDANO; WILCOX, 2018; NONNENMACHER; KAUR; COIMBRA, 2016), análise de séries temporais (TRAPERO; KOURENTZES; MARTIN, 2015; DONG *et al*., 2013), aplicações envolvendo processamento de imagens (PEDRO; COIMBRA, 2015; PEDRO *et al.,* 2018; PAWAR *et al*., 2019), utilização de algoritmos de aprendizagem de máquina (KOO *et al*., 2019; YAGLI; YANG; SRINIVASAN, 2019; BENALI *et al*., 2019), entre outras.

No presente trabalho, realizaram-se previsões da média diária de irradiação solar global para um horizonte de um dia a posteriori por meio da aplicação de modelos de aprendizagem de máquina em conjuntos de dados com a presença de atributos formados por variáveis exógenas (insolação, temperatura do ar, precipitação), variáveis endógenas (série temporal da média diária de irradiação solar global) e variáveis temporais (ano, mês e dia da medição) totalizando 44 atributos e 3254 observações, além disso, avaliou-se a influência da intensidade do *El Niño/La Niña* na previsão da média diária irradiação solar por meio de seu uso como preditor para os modelos utilizados.

Fez-se análise estatística por meio da implementação de 5 algoritmos de aprendizagem de máquina, a saber: Máquina de Aprendizagem Mínima (MLM) (DE SOUZA JÚNIOR, 2015), K-Vizinhos mais Próximos (KNN), Regressão por Vetor Suporte (SVR) (CARRASCO; LÓPEZ; MALDONADO, 2019; DRUCKER *et al.,* 1997), Florestas Aleatórias (RF) e árvores de decisão com a aplicação de um procedimento de *Bootstrap Aggregation (BAGGING)*. Além desses, foi utilizado modelo de *baseline* como critério de referência mínimo de desempenho. A comparação da acurácia de previsão dos métodos foi feita pela avaliação das métricas de erro: Erro Absoluto Médio (MAE), Erro de Viés Médio (MBE), Raiz do Erro Quadrático Médio (RMSE), Raiz do Erro Quadrático Médio Relativo (rRMSE) e Habilidade de Previsão (s).



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Os algoritmos KNN, SVR e RF são frequentemente aplicados para previsão de irradiação solar como encontrado na literatura (SHARMA, 2011). Entretanto, o MLM é um modelo relativamente recente que ainda carece de trabalhos voltados para a sua aplicação (DE SOUZA *JÚNIOR et al*., 2015; WALCH *et al.,* 2019), o que é uma das contribuições do presente trabalho. Assim, o artigo traz o uso de um algoritmo que ainda não foi utilizado para propósitos de previsão de irradiação solar, além da avaliação do impacto que os fenômenos climáticos *El Niño* e *La Niña* proporcionam sobre o desempenho dos modelos.

O restante do artigo é formado pela seção 2, que trata de uma breve explicação sobre os fenômenos climáticos estudados, além de abordar sobre a aquisição e a organização do conjunto de dados, assim como o procedimento aplicado para realizar o treinamento dos algoritmos. Na seção 3, os modelos de aprendizagem são detalhados e as métricas de erro são apresentadas. Posteriormente, na seção 4 os resultados encontrados são ilustrados e discutidos e por fim, a seção 5 trata-se das conclusões obtidas.

2. MÉTODO

O *El Niño* e a *La Niña* caracterizam-se, respectivamente, por variações positivas e negativas da temperatura superficial do oceano pacífico, na região próxima à linha do Equador. Esses fenômenos podem provocar alterações tanto no regime de ventos quanto no de chuvas, uma vez que alteram o clima regionalmente ou até mesmo, globalmente. Os dois eventos são definidos quando há alteração da temperatura por, pelo menos três meses consecutivos e ocorrem, em média, a cada 2 ou 3 anos, durando de 12 a 18 meses. A literatura aponta que, devido às alterações climáticas ocasionadas pelo *El Niño* e pela *La Niña* muitas variáveis meteorológicas sofrem modificações no seu padrão sazonal típico, tais como a irradiação solar global, velocidade do vento e precipitação (MOHAMMADI; GOUDARZI, 2018).

2.1 Dados

Os dados foram obtidos pela estação meteorológica da Universidade Federal do Ceará, na cidade de Fortaleza-CE, Brasil, 3º43''6'S e 38º32''36'O, entre os anos de 1969 e 1979 e 1995 e 1997.

Ao preditor que mensura a intensidade do *El Niño* e da *La Niña* foram atribuídos valores de -4 a 4, no qual os valores extremos representam as intensidades máximas destes, respectivamente, e o zero representa uma condição de neutralidade, como indicado pela Tabela 1. As intensidades desses fenômenos foram obtidas pelo *Golden Gate Weather Service*, serviço estadunidense (LI *et al*., 2013), que avalia a intensidade desses fenômenos pelo Índice de *Niño* Oceânico (ONI), obtido com base na variação entre a temperatura medida e um valor de temperatura pré-determinado para a região do pacífico determinada por 5ºN-5ºS, 120º-170ºO.



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Tabela 1: Código de valores indicando as intensidades do *El Niño* e da *La Niña*

El Niño				La Niña		
Intensidades	Aumento característico da temperatura	Valor atribuído		Intensidades	Aumento característico da temperatura	Valor atribuído
Muito Forte	≥ 2°	4		Muito Forte	≥ -2°	-4
Forte	1,5° ≤ t ≤ 1,9°	3		Forte	-1,5° ≤ t ≤ -1,9°	-3
Moderado	1° ≤ t ≤ 1,4°	2		Moderado	-1° ≤ t ≤ -1,4°	-2
Fraco	0,5° ≤ t ≤ 0,9°	1		Fraco	-0,5° ≤ t ≤ -0,9°	-1
Neutro	< 0,5º	0		Neutro	< -0,5°	0

Fonte: Autoria própria.

A média diária da irradiação solar global foi coletada com o uso de um actinógrafo *R. Fuess Berlin-Steglitz,* com numeração de 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).

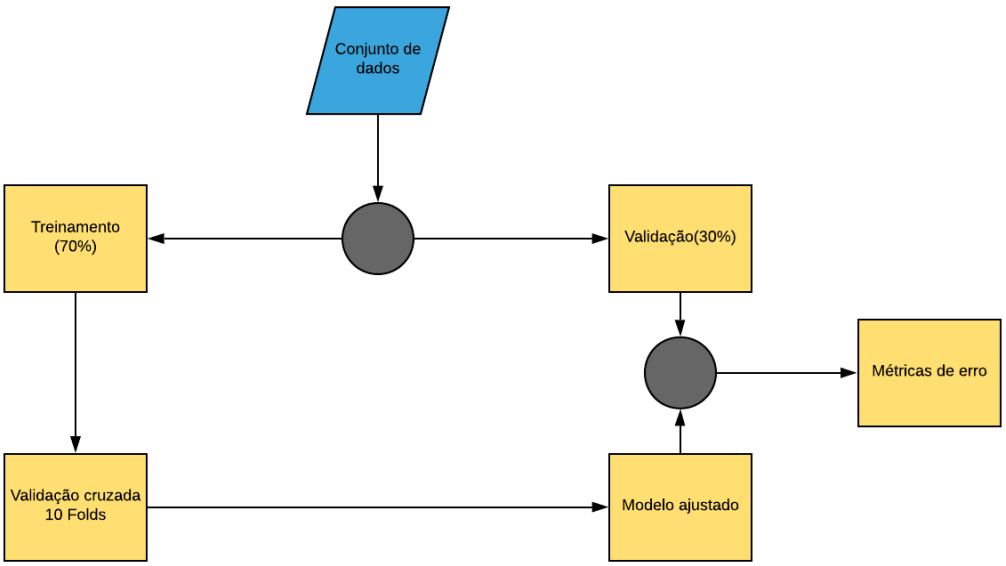
Para os preditores pressão, temperatura do ar, temperatura de bulbo úmido, umidade relativa, velocidade do vento e nebulosidade foram consideradas medições às 9h e 15h de cada dia. Os valores das médias diárias de irradiação solar dos 20 dias anteriores e a data de aquisição de cada observação (ano, mês e dia) também foram considerados preditores. A correlação existente entre os atributos, bem como sua correlação com a resposta pode ser observada pelo mapa de calor no link https://plot.ly/~fpmarinho90/1.embed que representa a matriz de correlação entre todas as variáveis tanto de entrada como de saída presente no conjunto de dados.

Para o treinamento dos modelos, o conjunto de dados foi dividido aleatoriamente de tal forma que 70% ficaram para o treinamento dos algoritmos e o restante constituiu o conjunto de validação. O ajuste dos métodos foi feito pela aplicação de validação cruzada 10-*Folds* (JAMES *et al*., 2013) ao conjunto de treinamento. Esta etapa é representada pela Figura 1.



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Figura 1: Representação da etapa de treinamento



Fonte: Autoria própria.

3. MODELOS UTILIZADOS

3.1 Modelo de *Baseline*

O modelo de *baseline* escolhido atribui para uma nova observação a irradiação média do conjunto de treinamento (HYNDMAN; ATHANASOPOULOS, 2018). Seja $T = \{G_1, G_2, \dots, G_n\}$ os valores de irradiação do conjunto de treinamento. Desta forma, o valor atribuído para uma nova observação é dado pela Equação (1).

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

Funciona como modelo que fornece desempenho mínimo, atuando como cota inferior para a acurácia dos demais modelos utilizados. Espera-se que sua aplicação forneça os maiores valores para as métricas de erro RMSE e MAE.

3.2 K-Vizinhos Próximos

Este conhecido método é geralmente aplicado em problemas de classificação, porém ele admite uma versão no contexto 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 dos



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

valores de irradiação de tais vizinhos. Neste caso, a irradiação estimada para uma nova observação será como na Equação (2).

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

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.

3.3 Regressão por Vetor Suporte

Este modelo, por mais que seja uma versão de uma Máquina de Vetor Suporte (SVM), apresenta a diferença de ser utilizado em problemas de regressão, enquanto o SVM é um método comumente aplicado no contexto de classificação. Uma das grandes vantagens em se trabalhar com o SVR se deve à possibilidade de construção de funções não lineares complexas, além de se ter a garantia de otimalidade global via a solução de um problema quadrático convexo (QP), treinamento efetivo e, geralmente, apresentando um desempenho preditivo positivo com um reduzido risco de *overfitting* graças ao uso do princípio de minimização do risco estrutural. Neste estudo, o algoritmo SVR aplicado foi do tipo ε-SVR, onde ocorre a construção de um ε-tubo em que os erros são ignorados no seu interior, mas penalizados fora do mesmo (CARRASCO; MALDONADO, 2019; DRUCKER *et al.,* 1997).

$$min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + Ce^T(\xi + \xi^*)$$

$$s.t. \, y - (Aw + be) \leq \varepsilon e + \xi$$

$$(Aw + be) - y \leq \varepsilon e + \xi^*$$

Onde $A = [x_1 x_2 \dots x_m]^T \in R^{mxn}$, $y = (y_1, y_2, \dots, y_m) \in R^m$, $e \in R^m$ é um vetor de uns, $\xi, \xi^* \in R^m$ são variáveis de folga que penalizam as observações que estão fora do ε-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 programas convexos, como o método dos pontos interiores (MEHROTRA, 1992).

O ε-SVR pode ser estendido para uma configuração não linear pelo uso de *kernels*. Isto pode ser feito pela aplicação de um artifício no problema dual correspondente (DRUCKER *et al.,* 1997). A versão de *kernels* para o SVR é dada pelo programa convexo.



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

$$min_{\alpha,\alpha^*}\frac{1}{2}(\alpha-\alpha^*)^T K(A,A^T)(\alpha-\alpha^*) - y^T(\alpha-\alpha^*) + \varepsilon e^T(\alpha+\alpha^*)$$

$$s.t.\, e^T(\alpha-\alpha^*) = 0$$

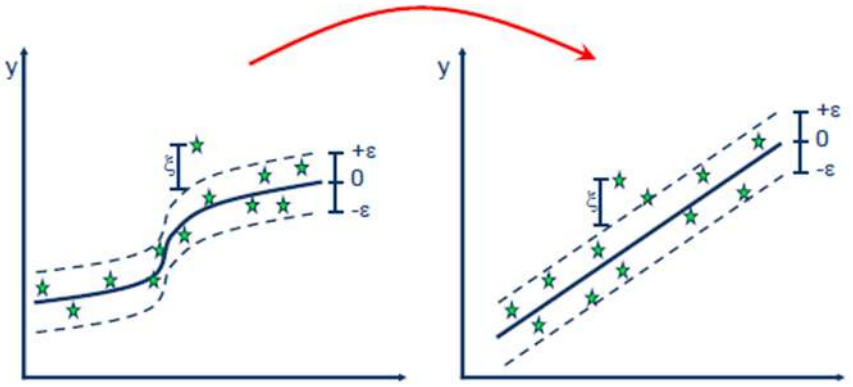
$$0 \leq \alpha, \alpha^* \leq Ce$$

Onde $K(A,A^T) \in R^{mxn}$ representa a matriz das funções *kernel* da forma $k_{is} = \kappa(x_i, x_s)$, $\kappa: R^n \times R^n \to R$ e α, α^* 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 Equação (3).

$$\kappa(x_i, x_s) = \exp\left(\frac{-\|x_i - x_s\|^2}{2\sigma^2}\right) \quad (3)$$

Desta forma, os parâmetros livres para este método são o ϵ, C e σ, vale ressaltar que neste trabalho foi atribuído $\sigma = 0.016745$ e $\epsilon = 0.1$e o hiperparâmetro C foi obtido por validação cruzada 10-*fold* no conjunto de treinamento. Uma representação esquemática do uso de *kernels* para se trabalhar com funções não lineares é ilustrado na Figura 2.

Figura 2: Representação do uso de *kernels* no SVR



Fonte: https://www.saedsayad.com/support_vector_machine_reg.htm.

3.4 Árvores de decisão 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 as 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 nas máquinas de



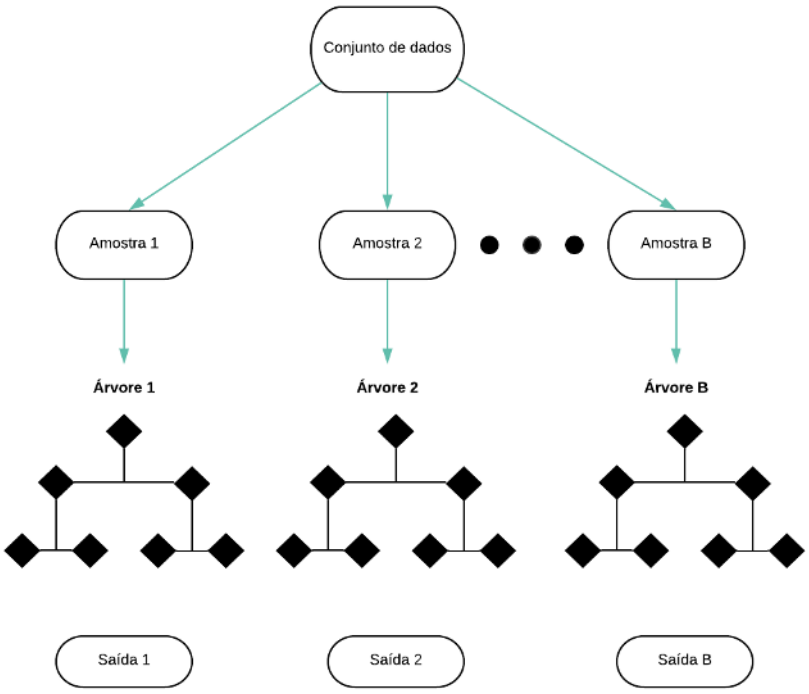
REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

comitê. Tal combinação pode resultar em melhorias significativas na acurácia de previsão, mas com uma perda na capacidade de interpretação (JAMES *et al*., 2013).

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 os das árvores de decisão (JAMES *et al*., 2013; KUHN; JOHNSON, 2013).

O primeiro passo neste procedimento é a realização de um *Bootstrap* no conjunto de dados considerado, que consiste na formação de várias amostras através da realização de múltiplas amostragens com reposição no conjunto de dados original. Posteriormente, para cada 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. Todo este procedimento está representado na Figura 3.

Figura 3: Representação esquemática de um *Bagging* aplicado em árvores de decisão



Fonte: Autoria própria.

Onde tem-se o vetor $O = \{Saída1, Saída2, \dots, SaídaB\} = \{f_1^*(x), f_2^*(x), \dots, f_B^*(x)\}$. Desta forma, a saída resultante do *Bagging* será dada pela Equação (4).



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

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

Neste caso, percebe-se que o único hiperparâmetro presente neste método é o número de árvores considerada, tal parâmetro não é crítico uma vez que um valor elevado de árvores não conduzirá a problemas relacionados à *overfitting* (JAMES *et al.,* 2013). Entretanto, um número exagerado pode proporcionar um alto tempo de processamento. Neste trabalho, para todos os conjuntos de dados foram utilizadas 500 árvores, que foi o valor que proporcionou um melhor balanceamento entre desempenho e tempo computacional.

3.5 Florestas Aleatórias

O modelo *Random Forest* é análogo ao *Bagging*, mas com o uso de uma pequena modificação visando descorrelacionar as árvores (JAMES *et al.,* 2013). 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 submodelos são descorrelacionadas ou na melhor das hipóteses apresentam um nível fraco de correlação. Basicamente, a proposta do *Random Forest* 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.

3.6 Máquina de Aprendizagem Mínima

Este algoritmo foi originalmente apresentado em trabalhos recentes (DE SOUZA JÚNIOR *et al.,* 2015). 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 se deseja 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 (DE SOUZA JÚNIOR *et al*., 2015).

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 dado pela Equação (5).

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



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

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.

1. Ajuste de uma regressão linear múltipla para obter matrizes distância de saída a partir de matrizes distância de entrada;

2. Estima a resposta desejada por meio da solução de um problema de multilateração.

3.6.1 Regressão linear múltipla entre matrizes distância

Esta etapa se inicia com a escolha aleatória de algumas observações do conjunto de dados denominadas pontos de referência. Vale ressaltar que se considera 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 Equação (6).

$$\Delta_y = D_x B + E \tag{6}$$

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 custo, dada pela Equação (7).

$$RSS(B) = tr((\Delta_y - D_x B)^T(\Delta_y - D_x B)) \tag{7}$$

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

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

Para o cálculo da inversa na Equação (10), recomenda-se determinar a pseudoinversa de *Moore-Penrose* (BEN-ISRAEL; GREVILLE, 2003; PENROSE, 1955). 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



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

desejada y e as saídas conhecidas $\{t_k\}, k = \{1,2,\dots,K\}$ dos pontos de referência, como ilustrado na Equação (9).

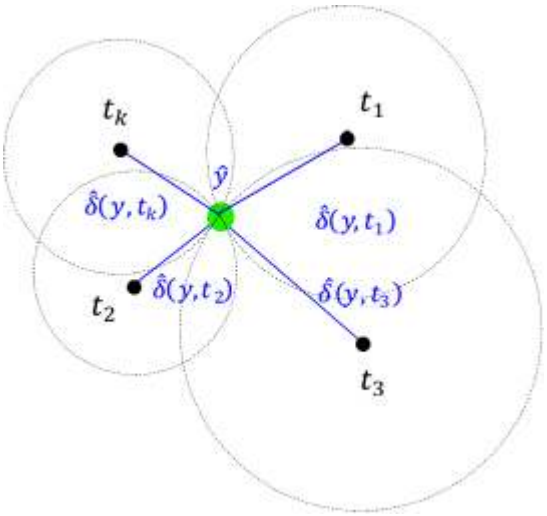
$$\delta^*(y,T) = d(x,\mu)B^* \tag{9}$$

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

3.6.2 Estimativa das respostas

Estimar as respostas 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 [28]. Existem muitas estratégias que podem ser adotadas para a solução de tal problema, a utilizada neste estudo busca determinar $\hat{y}$ por meio de uma interpretação geométrica, que gera K equações não lineares relacionadas à consideração de se ter K hiperesferas S-dimensional centradas nos t_k e passando por $\hat{y}$ com raios iguais a $\delta^*(y,t_k)$, como ilustrado na Figura 4 para o caso bidimensional.

Figura 4: Interpretação geométrica utilizada para a solução do problema de multilateração



Fonte: Autoria própria.

Desta forma, tem-se a Equação (10).

$$(\hat{y} - t_k)^T(\hat{y} - t_k) = (\delta^*(y,t_k))^2 \tag{10}$$

Assim a reposta é estimada pela minimização da função objetivo dada pela Eq. (11).

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

Por ser um problema de mínimos quadrados não linear, recomenda-se o uso do algoritmo de Levenberg-Marquardt (MARQUARDT, 1963; MORÉ, 1978) 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.

3.7 Métricas de erro

As métricas de erro utilizadas para realizar a avaliação do desempenho dos modelos de aprendizagem de máquina utilizados são apresentadas nesta seção.

3.7.1 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 Equação (12).

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

3.7.2 MBE

O MBE se assemelha ao cálculo do MAE, mas se diferencia deste por considerar o sinal do erro, isto é, não calcula o valor absoluto das diferenças. Desse modo, requer prudência em sua análise, uma vez que permite a compensação de erros (erros com sinais distintos). Ela é dada pela Equação (13).

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

3.7.3 RMSE

O RMSE calcula a magnitude da 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, sendo definida pela Equação (14).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \quad (14)$$

3.7.4 rRMSE

O rRMSE é a razão entre o RMSE e a média dos valores da variável em análise, como dado pela Equação (15).

$$rRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}}{\bar{y}} \quad (15)$$



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Onde $\bar{y} = \sum_{i=1}^{n} y_i$. Tal métrica foi utilizada porque fornece faixas de classificação do desempenho das previsões: $rRMSE < 10\%$ excelente, $10\% < rRMSE < 20\%$ bom, $20\% < rMRSE < 30\%$ razoável, $rRMSE > 30\%$ ruim [21].

3.7.5 s

Essa métrica avalia o desempenho do modelo aplicado quando comparado com a acurácia do modelo de *baseline* (WOLD; ESBENSEN; GELADI, 1987), sendo definido pela Equação (16).

$$s = 1 - \frac{RMSE}{RMSE_B} \tag{16}$$

Onde o $RMSE$ representa a o valor da métrica de erro para o modelo em análise e o $RMSE_B$ representa o valor da métrica de erro para o modelo de *baseline*.

4 RESULTADOS E DISCUSSÃO

Nesta seção os resultados obtidos são apresentados para ambos os casos, quando se considera a intensidade do *El Niño/La Niña* como um preditor para os modelos aplicados e quando não se considera tal efeito sobre os métodos. Vale ressaltar que a busca pelos hiperparâmetros na etapa de treinamento foi apresentada apenas para o caso em que se considerou os fenômenos climáticos como atributos, pois um procedimento análogo foi feito para a situação que não se considera.

4.1.1 SVR

Como já discutido, tomou-se $\sigma = 0.016745$ e $\epsilon = 0.1$. Já o hiperparâmetro de regularização C é obtido via validação cruzada 10-*fold* no conjunto de treinamento. A métrica de erro escolhida para ser o critério de seleção do hiperparâmetro ótimo foi o RMSE. Os resultados são reportados pela Tabela 2.



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Tabela 2: Busca pelo hiperparâmetro no SVR

C	$RMSE(W/m^2)$	MAE (W/m^2)
0.25	41.031	29.164
0.50	40.090	29.218
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

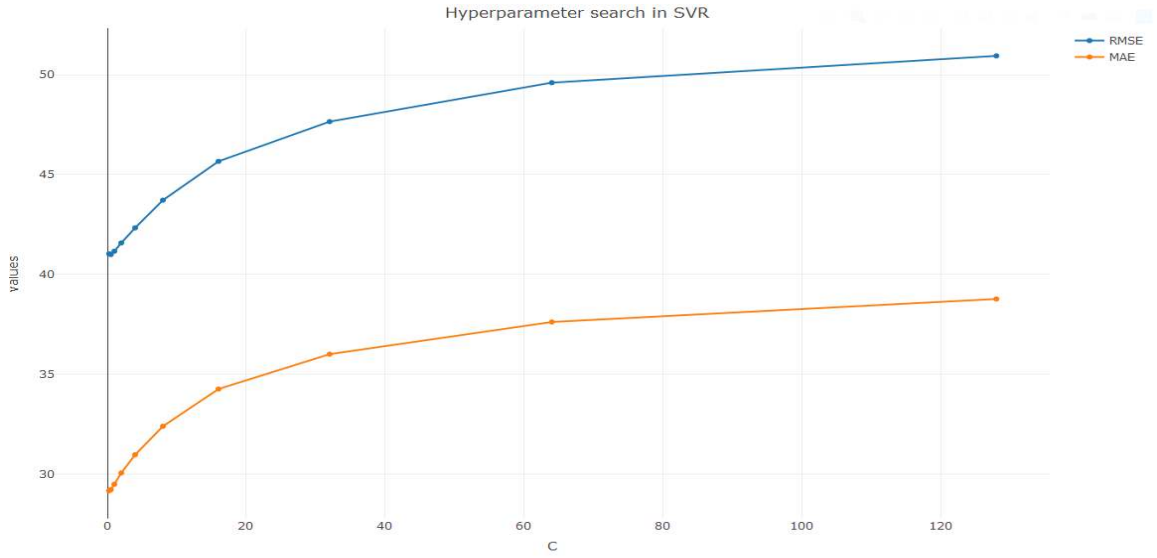
Fonte: Autoria própria.

Estes mesmos resultados podem ser observados de uma forma gráfica pela Figura 5. Nota-se que o valor ótimo para o parâmetro de regularização foi de $C = 0.5$. Observa-se que a partir de $C = 0.5$ o aumento no valor do hiperparâmetro ocasiona um aumento no RMSE indicando que um aumento na flexibilidade do modelo SVR para o conjunto de dados considerado proporciona uma piora no seu desempenho. Os valores ótimos encontrados para o caso em que não se considera os fenômenos climáticos do *El Niño/La Niña* como preditor para os modelos também foi de $C = 0.5$.



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Figura 5: RMSE e MAE versus parâmetro C



Fonte: Autoria própria.

4.2 Florestas Aleatórias

Neste caso, o hiperparâmetro a ser considerado corresponde à dimensão do espaço de preditores que será segmentado, m, e novamente a métrica escolhida para a seleção foi o RMSE. Os resultados são reportados pela Tabela 3 e pela Figura 6.

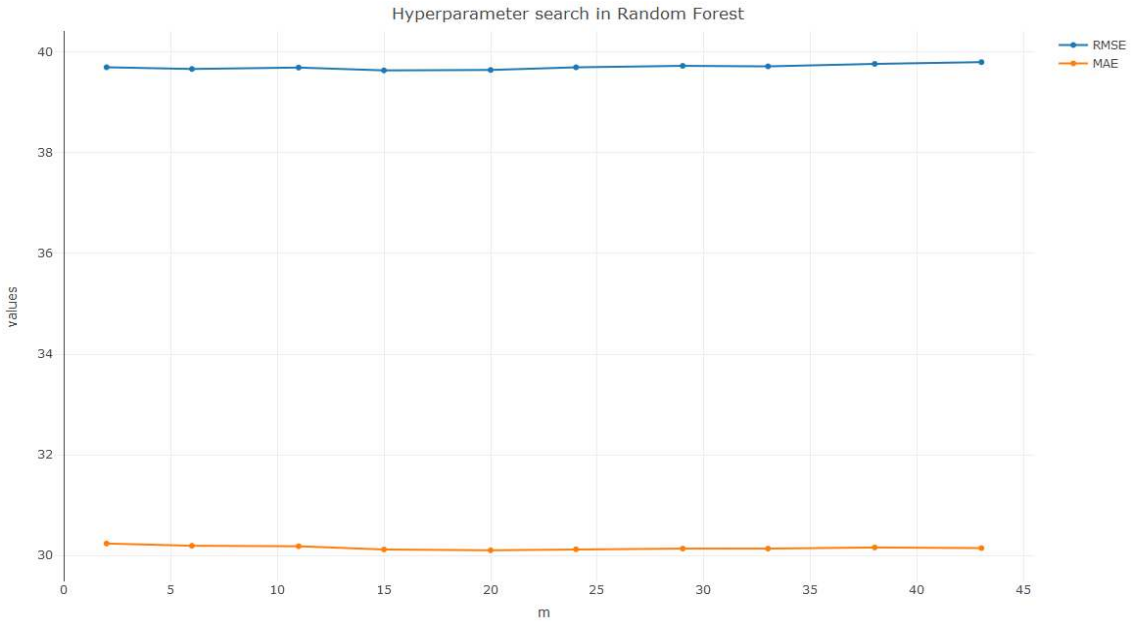
Tabela 3: Busca por hiperparâmetro no *Random Forests*

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

Fonte: Autoria própria

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Figura 6: RMSE e MAE em função do hiperparâmetro m



Fonte: Autoria própria.

Para este modelo, nota-se pela observação dos resultados que não há grandes variações nos valores das métricas de erro RMSE e MAE quando ocorre a variação nos valores do parâmetro livre m, indicando que, para este modelo, não há um preditor que se destaque no que diz respeito à significância estatística na descrição da média diária de irradiação solar global. Além disso, o valor ótimo para o hiperparâmetro foi de $m = 15$, este número também foi encontrado para o caso em que não se considerou o preditor *El Niño/La Niña*.

4.3 KNN

Aqui o único hiperparâmetro a ser considerado é o número de vizinhos K, os resultados para a busca do melhor ajuste são ilustrados na Tabela 4 e na Figura 7.



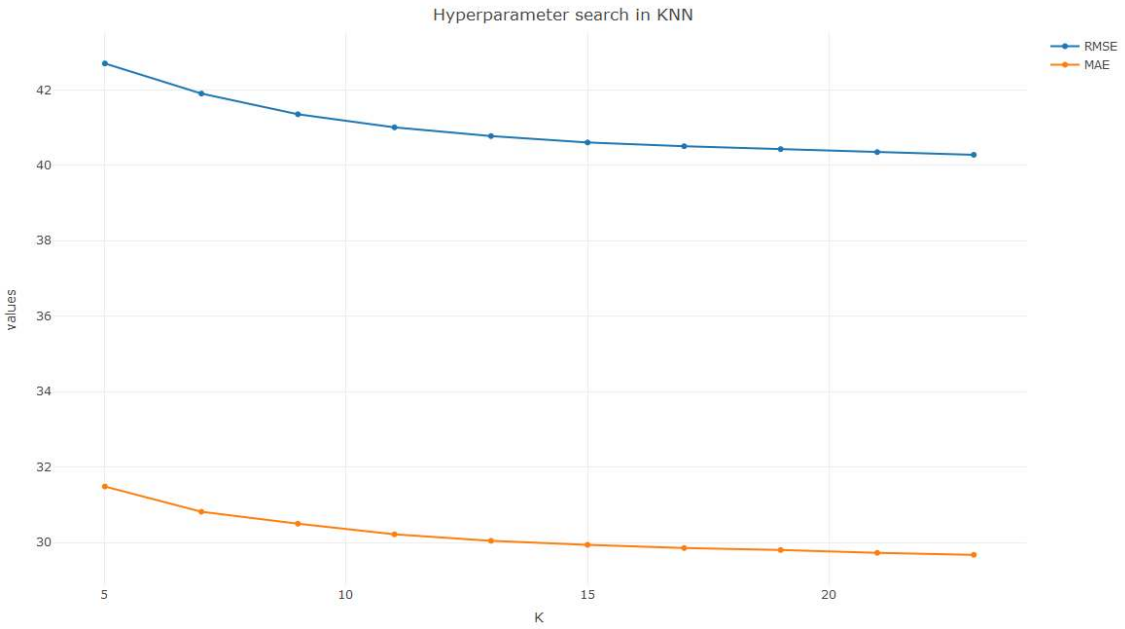
REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

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
23	40.279	29.671

Fonte: Autoria própria.

Figura 7: RMSE e MAE versus o número de vizinhos K



Fonte: Autoria própria.

Para este algoritmo, percebe-se uma redução inicial nos valores das métricas de erro RMSE e MAE com o aumento do valor de K, indicando que no início uma redução no nível de flexibilidade do modelo proporciona melhorias na acurácia de previsão do método quando se aplicado no conjunto de

treinamento do conjunto de dados considerado. O que coincide com o comportamento apresentado pelo método SVR como discutido. O número de vizinhos ótimo foi de $K = 23$ tanto para caso de *El Niño/La Niña* como atributo como para o caso em que não se considera.

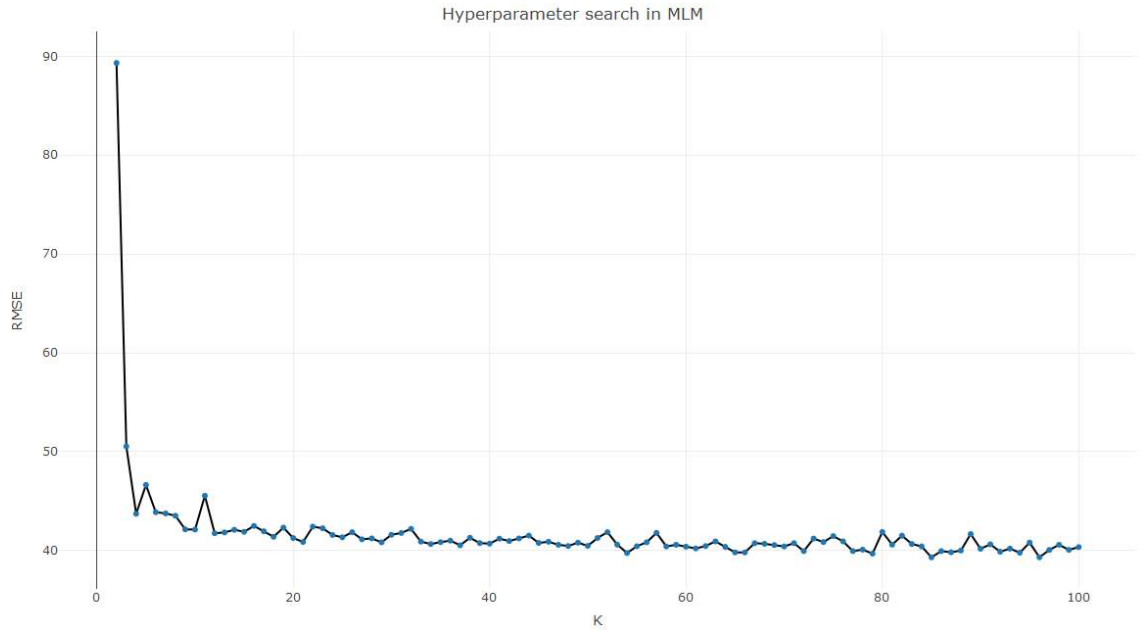
4.4 Árvore de decisão com a aplicação de um procedimento de *Bagging*

Como já havia sido abordado, o método que utiliza *Bagging* apresenta como único hiperparâmetro o número de árvores a ser considerado, e tal parâmetro não é crítico uma vez que um valor elevado de árvores não conduzirá a um *overfitting* (JAMES *et al*., 2013). Entretanto, um número exagerado pode conduzir a um alto tempo de processamento. Neste trabalho utilizou-se 500 árvores, que foi o valor com o melhor custo-benefício entre desempenho e tempo computacional.

4.5 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 ótimo de $K = 84$, os resultados obtidos por validação cruzada 10-*fold* no conjunto de treinamento são observados na Figura 8, que mostra os valores da métrica RMSE em função do número de pontos de referência.

Figura 8: RMSE versus o número de pontos de referência K



Fonte: Autoria própria.

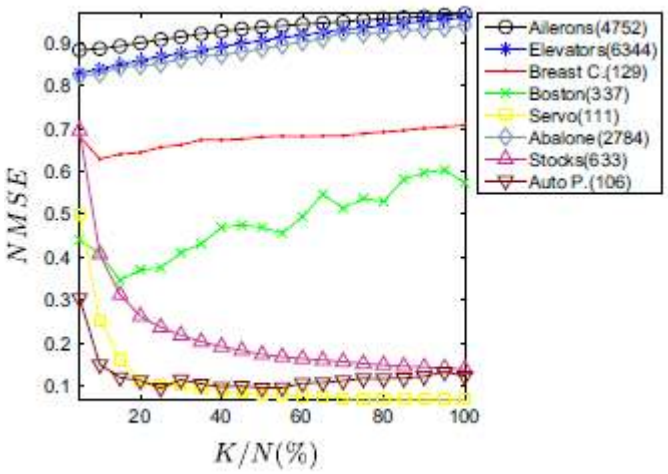
Desta forma, percebe-se que o comportamento geral é de redução dos valores do RMSE com o aumento dos pontos de referência até ocorrer estabilização com pequenas oscilações. Este comportamento não é geral e dependendo do conjunto de dados em que o MLM é aplicado as



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

tendências nos resultados podem ser modificadas. No trabalho de apresentação do método (DE SOUZA JÚNIOR, 2015), os autores aplicaram o MLM em doze conjuntos de dados considerados de *benchmarking* disponibilizados no repositório da Universidade da Califórnia (www.ics.uci.edu/~mlearn/), tanto para propósitos de regressão como de classificação. No contexto de regressão havia oito conjunto de dados, onde para todos a seleção do número ótimo de pontos de referência se deu por validação cruzada 10-*fold* no conjunto de treinamento. Os resultados são reportados na Figura 8 onde o $NMSE$ representa a métrica *Normalized Mean Squared Error* e o número de pontos de referência é normalizado pelo número de observações no conjunto de treinamento.

Figura 9: Busca por hiperparâmetro no MLM para outros conjuntos de dados



Fonte: https://www.sciencedirect.com/science/article/abs/pii/S0925231215003021.

Os valores das métricas de erro obtidas pela aplicação dos modelos ajustados no conjunto de validação estão adicionados na Tabela 5, para quando se considera a intensidade do *El Niño/La Niña* como preditor, e na Tabela 6 para o caso em que não se considera.



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Tabela 5: Desempenho dos modelos quando se considera o preditor *El Niño/ La Niña*

Métodos	*Treino*					*Teste*				
	MBE (W/m²)	MAE (W/m²)	RMSE (W/m²)	rRMSE (%)	s (%)	MBE (W/m²)	MAE (W/m²)	RMSE (W/m²)	rRMSE (%)	s (%)
SVR	-7.765	24.769	37.186	18.710	15.987	-8.336	29.433	41.194	20.860	6.930
RF	0.537	12.027	16.270	8.187	63.241	1.077	30.400	40.234	20.374	9.10
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

Fonte: Autoria própria.

Tabela 6: Desempenho dos modelos quando o preditor *El Niño/ La Niña* é desconsiderado

Métodos	*Treino*					*Teste*				
	MBE (W/m²)	MAE (W/m²)	RMSE (W/m²)	rRMSE (%)	s (%)	MBE (W/m²)	MAE (W/m²)	RMSE (W/m²)	rRMSE (%)	s (%)
SVR	7.767	24.780	37.197	18.716	15.963	-8.337	29.433	41.196	20.860	6.928
RF	0.417	11.974	16.205	8.153	63.390	1.280	30.482	40.332	20.423	8.879
Bagging	-0.102	28.750	37.539	18.888	15.189	-0.707	30.477	40.650	20.584	8.161
KNN	-2.776	28.435	38.689	19.466	12.592	-3.679	29.933	40.854	20.687	7.701
MLM	-1.245	29.161	39.154	19.700	11.540	-2.158	30.809	41.268	20.897	6.764

Fonte: Autoria própria.

Como se pode perceber pela avaliação dos resultados, a presença ou ausência do atributo relacionado a intensidade do *El Niño/La Niña* proporciona uma influência distinta no desempenho dos modelos de aprendizagem. De fato, para os métodos MLM, SVR e RF a remoção de tal preditor contribui para uma piora na acurácia, quando se baseia na métrica RMSE. Vale ressaltar que essa redução é mais significativa no algoritmo MLM e bem mais branda no SVR e RF, isto pode estar relacionado ao fato de que o MLM se baseia em uma regressão linear múltipla de matrizes de distância e a remoção de um atributo ocasionou alterações nos valores de distâncias calculados.

De uma forma geral, os resultados indicam que a presença de tal atributo contribui para uma melhor generalização destes modelos de aprendizagem. Não houve alteração nos resultados obtidos pelo método KNN, para a árvore de decisão com um procedimento de *Bagging* a remoção de tal atributo proporcionou uma melhora em seu desempenho o que indica a presença de maldição da dimensionalidade, problema relacionado ao aumento exponencial no volume associado com a adição de dimensões extras em um espaço matemático (BELLMAN, 1961), quando se utiliza tal modelo, o que



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

pode ser um indicativo da necessidade de se realizar um pré-processamento dos dados por meio da utilização de análise de componentes principais (PCA) (WOLD; ESBENSEN; GELADI, 1987) visando uma redução na dimensionalidade, bem como a realização de uma transformação de coordenadas que elimina a correlação entre os atributos. Apesar de se ter alterações nos valores dos RMSE para cada modelo, nota-se que estas mudanças são pequenas o que é reflexo de se ter uma correlação entre o atributo relacionado à intensidade do *El Niño/La Niña* com a média diária de irradiação solar global do dia a posteriori assumindo um valor reduzido, no caso $corr = 0.0103$.

Por mais que este estudo preliminar indique que exista uma pequena influência dos fenômenos climáticos considerados nos desempenhos dos algoritmos para a previsão de irradiação, é importante destacar que tais fenômenos proporcionam um aumento na variabilidade de muitas variáveis meteorológicas (KOO *et al.,* 2019) e acredita-se que os resultados obtidos estejam relacionados à maneira de como tais atributos foram considerados nos modelos, atribuindo apenas um valor numérico de acordo com a codificação representada na Tabela 1.

Para estudos posteriores pretende-se utilizar as categorias de *El Niño* e/ou *La Niña* para a realização de um *clustering* inicial e em seguida aplicar os algoritmos para cada agrupamento formado, avaliando possíveis alterações nas métricas de erro. 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 (DE SOUZA JÚNIOR *et al*., 2015). Desta forma, tem-se um novo modelo de fácil interpretação e implementação que apresenta um desempenho competitivo para a previsão da média diária irradiação solar para um dia a posteriori.

CONSIDERAÇÕES FINAIS

No presente trabalho realizou-se previsões de médias diárias da irradiação solar gobal para um horizonte de um dia a *posteriori* utilizando modelos de aprendizagem de máquina, a saber: *Minimal Learning Machine, K-Nearest Neighbors, Support Vector Regression, Random Forests* e uma árvore de decisão com um procedimento de *Bagging*. Observou-se a influência que os fenômenos climáticos *El Niño/La Niña* ocasionam sobre o desempenho dos métodos aplicados quando são considerados como preditores. A acurácia de tais algoritmos foi avaliada pelo cálculo das métricas de erro MAE, MBE, RMSE, rRMSE e s.

Quando se considerou o preditor relativo aos fenômenos climáticos, os valores das médias aritméticas para as métricas RMSE e s para os cinco modelos foram de 40.778 W/m² e 7.871%, respectivamente. No caso em que não se considerou o *El Niño/La Niña,* os valores médios do RMSE e s foram de 40.860 W/m² e 7.686%, portanto, na média, a retirada do atributo em estudo ocasionou uma pequena perda de acurácia de previsão. Entretanto, como já foi discutido, esta influência é distinta para cada método. Com relação ao MLM, uma das contribuições desse estudo por ser um algoritmo relativamente novo e pouco testado, o mesmo forneceu resultados que se equiparam àqueles



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

fornecidos pelos demais modelos que são considerados clássicos e bem-postos, assim, verificou-se que este método recente é adequado para a realização de previsões das médias diárias de irradiação solar global para um dia a posteriori na cidade de Fortaleza.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível – Brasil (CAPES) – Código de Financiamento 001 e com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil.

REFERÊNCIAS

BELLMAN, Robert. **Curse of dimensionality** - **Adaptive control processes:** a guided tour. Princeton, NJ: [*S. n.*], 1961.

BENALI, L. *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.

BEN-ISRAEL, Adi; GREVILLE, Thomas NE. **Generalized inverses:** theory and applications. Berlim: Springer Science & Business Media, 2003.

CARRASCO, Miguel; LÓPEZ, Julio; MALDONADO, Sebastián. Epsilon-nonparallel support vector regression. **Applied Intelligence**, v. 49, p. 4223-4236, 2019.

DE SOUZA JÚNIOR, Amauri Holanda et al. Minimal learning machine: a novel supervised distance-based approach for regression and classification. **Neurocomputing**, v. 164, p. 34-44, 2015.

DONG, Zibo et al. Short-term solar irradiance forecasting using exponential smoothing state space model. **Energy**, v. 55, p. 1104-1113, 2013.

DRUCKER, Harris et al. Support vector regression machines. *In:* **Advances in neural information processing systems**. [*S. l.: s. n.*], 1997. p. 155-161.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting:** principles and practice. [*S. l.*]: OTexts, 2018.

JAMES, Gareth et al. **An introduction to statistical learning**. New York: springer, 2013.

KOO, Choongwan 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.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. New York: Springer, 2013.

LI, Mao-Fen 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.

MARQUARDT, Donald W. 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.



REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

MEHROTRA, Sanjay. On the implementation of a primal-dual interior point method. **SIAM Journal on optimization**, v. 2, n. 4, p. 575-601, 1992.

MEJIA, John F.; GIORDANO, Marco; WILCOX, Eric. Conditional summertime day-ahead solar irradiance forecast. **Solar Energy**, v. 163, p. 610-622, 2018.

MOHAMMADI, Kasra; GOUDARZI, Navid. Study of inter-correlations of solar radiation, wind speed and precipitation under the influence of El Niño Southern Oscillation (ENSO) in California. **Renewable energy**, v. 120, p. 190-200, 2018.

MORÉ, Jorge J. The Levenberg-Marquardt algorithm: implementation and theory. *In:* **Numerical analysis**. Berlin, Heidelberg: Springer,1978. p. 105-116.

NIEWIADOMSKA-SZYNKIEWICZ, Ewa; MARKS, Michał. Optimization schemes for wireless sensor network localization. **International Journal of Applied Mathematics and Computer Science**, v. 19, n. 2, p. 291-302, 2009.

NONNENMACHER, Lukas; KAUR, Amanpreet; COIMBRA, Carlos FM. Day-ahead resource forecasting for concentrated solar power integration. **Renewable energy**, v. 86, p. 866-876, 2016.

NOTTON, Gilles 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.

PAWAR, Prathamesh et al. Detecting clear sky images. **Solar Energy**, v. 183, p. 50-56, 2019.

PEDRO, Hugo T. C. *et al.* Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. **Renewable Energy**, v. 123, p. 191-203, 2018.

PEDRO, Hugo T. C.; COIMBRA, Carlos F. M. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. **Renewable energy**, v. 80, p. 770-782, 2015.

PENROSE, Roger. A generalized inverse for matrices. *In:* **Mathematical proceedings of the Cambridge philosophical society**. Cambridge: Cambridge University Press, 1955. p. 406-413.

QING, Xiangyun; NIU, Yugang. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. **Energy**, v. 148, p. 461-468, 2018.

SHARMA, Navin 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.

TRAPERO, Juan R.; KOURENTZES, Nikolaos; MARTIN, Alberto. Short-term solar irradiation forecasting based on dynamic harmonic regression. **Energy**, v. 84, p. 289-295, 2015.

WALCH, Alina 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.

WOLD, Svante; ESBENSEN, Kim; GELADI, Paul. Principal component analysis. **Chemometrics and intelligent laboratory systems**, v. 2, n. 1-3, p. 37-52, 1987.

YAGLI, Gokhan Mert; YANG, Dazhi; SRINIVASAN, Dipti. Automatic hourly solar forecasting using machine learning models. **Renewable and Sustainable Energy Reviews**, v. 105, p. 487-498, 2019.