

Aprendizado de Máquina aplicado na previsão da geração de energia elétrica de uma usina solar fotovoltaica no Ceará

Lucas T. da Silva* Ruth P. S. Leão** André W. B. Silva***
Danielle B. Cavalcante**** Raimundo F. Sampaio†

* Departamento de Engenharia Elétrica, Universidade Federal do Ceará, CE, (e-mail: lucastavaressilva@alu.ufc.br).

** Departamento de Engenharia Elétrica, Universidade Federal do Ceará, CE, (e-mail: rleao@dee.ufc.br)

*** Departamento de Engenharia Elétrica, Universidade Federal do Ceará, CE, (e-mail: andrewagner10@yahoo.com.br)

**** Departamento de Engenharia Elétrica, Universidade Federal do Ceará, CE, (e-mail: daniellebaltazar2@gmail.com)

† Departamento de Engenharia Elétrica, Universidade Federal do Ceará, CE, (e-mail: rfurtado@dee.ufc.br)

Abstract: Given the growing of solar photovoltaic (PV) power in the Brazilian power mix, and considering that PV generation is intermittent, there is great need of PV generation forecasting models for better operational planning of both the PV power plant itself and the power grid. Thus, this paper proposes the application of Machine Learning models to forecast the power generation of a 160 MW PV plant located in the state of Ceará (Brazil). The aim is to forecast the daily power output of the power plant over 365 days based on the historical power data and the weather data of the PV power plant. To accomplish the task, distinct prediction models were tested, such as: sequences recognition, Artificial Neural Networks (ANN), XGBoost and hybrid approaches. The performance of the implemented models was evaluated using error metrics, and the XGBoost model achieved the most accurate results as regard to the prediction error and the execution time followed by the ANN. The Pattern Sequence-based Forecast (PSF), which is more transparent than an ANN or even XGBoost, has proved competitive, having the best performance among the models trained only with historical data of power generation.

Resumo: Tendo em vista a crescente inserção de geração fotovoltaica (FV) na matriz energética brasileira, e considerando que a geração FV é intermitente, surge a necessidade de modelos precisos para previsão da geração FV, de modo a permitir um melhor planejamento operacional da planta FV e do sistema elétrico. Desta forma, neste trabalho é proposta uma aplicação de Aprendizado de Máquina para previsão da geração de uma usina FV de grande porte, com potência instalada de 160 MW e localizada no estado do Ceará. O objetivo é prever a produção FV diária para 365 dias consecutivos, a partir do conhecimento das séries históricas de geração de energia e dos dados meteorológicos da usina. Para realizar essa tarefa, distintos modelos de predição foram testados, tais como: Reconhecimento de Sequências, Redes Neurais Artificiais (RNA), *XGBoost* e Métodos Híbridos. A performance dos modelos implementados foi avaliada com auxílio de métricas de erro, sendo o *XGBoost* o modelo que obteve melhor desempenho e menor tempo de execução, seguido pelo RNA. A previsão baseada em Reconhecimento de Sequências apresentou o melhor desempenho dentre os modelos treinados apenas com dados históricos de geração de energia elétrica.

Keywords: Machine Learning; Artificial Intelligence; Forecasting; PV Power Generation.

Palavras-chaves: Aprendizado de Máquina; Inteligência Artificial; Previsão; Geração FV.

1. INTRODUÇÃO

A geração FV vem ganhando destaque nos últimos anos, tendo a capacidade instalada crescido de aproximadamente 84% a 603% entre 2017 e 2020, tendo, ao final de 2020, a capacidade instalada total mundial alcançado 707,5 GW (BP, 2021). No Brasil, observa-se um crescimento na potência total instalada, centralizada e distribuída, no mesmo período, de cerca de 718%, alcançando em 2020 a marca de 7,9 GW, como ilustra a Figura 1.

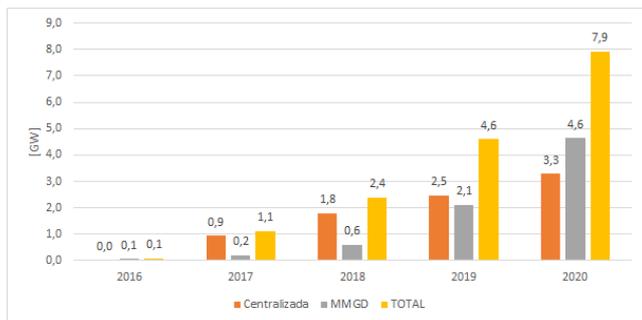


Figura 1. Evolução da capacidade instalada de geração solar fotovoltaica no Brasil.

Levantamentos mais recentes apontam que, em 2022, o Brasil já superou a marca de 13 GW de capacidade instalada de geração FV (Hein, 2022). Esse crescimento se deve em grande parte à ampla integração de Micro e Minigeração Distribuída (MMGD) no país, porém, usinas centralizadas de médio e grande porte também apresentam grande potencial de aplicação (Pereira et al., 2017), tendo o país, ao final de 2021, atingido a marca histórica de 4 GW de capacidade instalada em usinas solares fotovoltaicas de grande porte (Ribeiro, 2021).

Tendo em vista que a geração solar FV é dependente de fatores meteorológicos, como irradiância solar, temperatura ambiente, ocorrências de chuvas, vento, poeira, dentre outros, tem-se que a produção de energia é variável, de modo que a sua integração à rede elétrica constitui-se um desafio. Nesse sentido, surge a necessidade do conhecimento da geração futura das usinas FV de maneira precisa e confiável (Raza et al., 2016). Uma previsão segura da geração para dado intervalo de tempo futuro tende a diminuir incertezas, aumentando a confiabilidade e a viabilidade econômica do sistema elétrico como um todo (Diagne et al., 2013). Diante do exposto, este artigo tem como objetivo implementar, testar e avaliar o desempenho de diferentes modelos computacionais para a previsão 24h à frente da geração de uma usina solar fotovoltaica em um horizonte 365 à frente. Será avaliado o desempenho de 13 métodos distintos, nas categorias de reconhecimento de seqüências, aprendizagem supervisionada e modelos híbridos, para a previsão da geração fotovoltaica. Será também investigada a relevância de dados meteorológicos no desempenho dos métodos de previsão no horizonte de tempo definido.

O restante do artigo está organizado como segue. A seção 2 apresenta conceitos e discute trabalhos relacionados a modelos de previsão. A seção 3 detalha sobre os métodos implementados. Na seção 4 os resultados são apresentados, analisados e discutidos e finalmente a seção 5 traz a conclusão do artigo.

2. REFERENCIAL TEÓRICO

Os modelos de previsão de geração fotovoltaica são em geral classificados em três categorias: modelos físicos, estatísticos e híbridos. Os métodos físicos usam um modelo de simulação teórica para calcular a potência de saída de um sistema fotovoltaico com base em seus principais parâmetros de projeto. Os modelos estatísticos incluem todos os métodos baseados em dados, cobrindo tanto a modelagem estatística clássica, os modelos analíticos, quanto os modelos no campo da Inteligência Computacional (IC). O método híbrido é uma combinação de dois métodos diferentes, um físico e um estatístico, ou dois ou mais métodos estatísticos (Mayer and Gróf, 2021). Os métodos estatísticos são mais comumente usados para previsão de geração fotovoltaica (Antonanzas et al., 2016). Esses métodos orientados a dados são baseados em conjuntos de dados históricos de irradiância e produção de energia, e eles não exigem qualquer informação sobre os parâmetros de projeto do sistema fotovoltaico. Os modelos de previsão no campo da Inteligência Computacional são inúmeros e incluem as RNAs, algoritmos de aprendizado profundo, máquinas de aprendizado e algoritmos de reconhecimento de padrão (Ahmed et al., 2020). Em Sobri et al. (2018), os autores destacam que, dentre as categorias citadas, os principais métodos usados são: Média Móvel Autorregressiva (*Autoregressive Moving Average - ARMA*), Média Móvel Integrada Autorregressiva (*Autoregressive Integrated Moving Average - ARIMA*), regressão linear, Regressão por Vetores Suporte (*Support Vector Regression - SVR*), Árvores de Regressão (*Regression Trees - RT*) e RNAs. Recentemente, a combinação de modelos estatísticos lineares com modelos no campo da Inteligência Computacional também tem sido aplicada na previsão de séries temporais. Em Zhang (2003), uma metodologia híbrida que combina ARIMA com RNA é aplicada para um conjunto de dados real, obtendo-se, ao final, uma acurácia superior à que seria obtida utilizando-se os dois modelos separadamente. Em Júnior et al. (2019), além da combinação com a RNA, a ARIMA é combinada também com um modelo SVR, sendo os resultados em ambos os casos de performance superior quando comparados ao uso dos modelos em separado. Ainda sobre sistemas híbridos, abordagens de ensemble e seleção dinâmica tem sido utilizadas atualmente na combinação de diferentes modelos de previsão, conforme ocorre em de Oliveira et al. (2022), onde um algoritmo de seleção dinâmica escolhe, dentro de um grupo pré-definido, o modelo de aprendizado de máquina mais adequado para a atividade de previsão. Os modelos estatísticos analíticos são aplicações bastante consolidadas, tendo em vista já serem amplamente utilizados para previsões nas mais diversas áreas do conhecimento (Majid and Mir, 2018; Makridakis et al., 2018). Dentre estes, o modelo de persistência, que assume que a geração futura será igual à geração passada, é o mais simples de ser construído, sendo amplamente utilizado como modelo de referência para *benchmark* (Chu et al., 2021). Dada esta característica de simplicidade, a adoção de modelos que não o de persistência só é justificável se houver ganho de precisão. Nesse sentido, as abordagens de Inteligência Computacional têm ganhado notoriedade nos últimos anos, exatamente pelo potencial de produzirem previsões mais precisas e de maneira rápida (Kalogirou and Sencan, 2010). Em Wang et al. (2017) foi

proposta a utilização de modelos de Previsão baseada em Reconhecimento de Sequências (*Pattern Sequence-based Forecast - PSF*) para a previsão de geração de energia de usinas solares, enquanto que Lin et al. (2019) introduziram uma abordagem híbrida pela utilização de PSF combinada a RNAs. Em ambos os casos, os resultados foram promissores e instigaram maior investigação da aplicação destes métodos no tema proposto. Além disso, há trabalhos recentes que visaram outras abordagens de Aprendizagem Profunda (*Deep Learning - DL*) além das redes neurais, sobretudo árvores de decisões e *Extreme Gradient Boosting (XGBoost)*, para previsões de geração eólica (Machado et al., 2021) e de irradiância solar (Kumari and Toshniwal, 2021; Heinen, 2018; Kamarouthu, 2020). Nesse sentido, este trabalho avalia treze diferentes modelos de previsão, que serão melhor detalhados na seção de desenvolvimento.

3. BASE DE DADOS E ALGORITMOS DE PREVISÃO IMPLEMENTADOS

3.1 Base de dados

Os dados utilizados neste trabalho contemplaram o período de dois anos, entre 30 de junho de 2019 e 29 de junho de 2021 (731 dias), e consistiram em: (1) série temporal de geração de energia da usina, com granularidade horária e (2) série temporal de dados meteorológicos, coletados na usina. A escolha dos atributos (*features*), bem como a própria estruturação dos dados em si, foram motivadas visando maior conformidade com os algoritmos PSF implementados em Lin et al. (2019). A Tabela 1 apresenta a estrutura, com especificação das *features* utilizadas, bem como nomenclaturas adotadas para os conjuntos criados.

Tabela 1. Estrutura dos dados para os algoritmos PSF.

Conjunto	Atributos
$PV \in \mathbb{R}^{731 \times 12}$	Dados de geração entre 5:00 e 17:00
$W1 \in \mathbb{R}^{731 \times 10}$	Dados diários: temperaturas ambiente mínima e máxima, precipitação e irradiância média; às 9:00 e 15:00: temperatura ambiente, umidade relativa do ar e velocidade do vento.
$W2 \in \mathbb{R}^{731 \times 4}$	Dados diários: temperaturas ambiente mínima e máxima, precipitação e irradiância média; W2 é um subconjunto de W1.

Para a implementação dos algoritmos de aprendizagem supervisionada, que necessitam de entradas (X) associadas a saídas (y), os dados da Tabela 1 foram adaptados da seguinte forma: para a aprendizagem com utilização apenas dos dados de geração de energia, fez-se a associação de cada dia (X) para seu respectivo dia seguinte (y), em um deslocamento (*shift*) unitário (um dia), conforme Brownlee (2016). Em decorrência dessa adaptação, o número total de dias disponíveis foi reduzido em um, restando um conjunto com 730 dias. Para a aprendizagem com utilização de todos os dados (geração de energia e meteorológicos), a adaptação consiste apenas em tornar a base meteorológica ($W1$) na entrada (X) e os dados de geração na respectiva saída (y); esta associação não ocasiona perda de dados. A Tabela 2 resume a estrutura mencionada.

Toda a base de dados foi pré-processada, consistindo esta etapa em: tratamento de dados faltantes pelo algoritmo de

vizinho mais próximo, normalização dos dados no intervalo $[0,1]$ e identificação de dados anômalos (*outliers*) pelo algoritmo de floresta de isolamento, conforme Liu et al. (2008). Os dados anômalos foram substituídos por dados não anômalos, novamente sendo utilizado o algoritmo de vizinho mais próximo. Foi utilizada a base de dados do primeiro ano como conjunto de treino e a base do segundo ano como conjunto de teste.

3.2 Algoritmos implementados

Conforme dito anteriormente, este trabalho aplicou e avaliou o desempenho de 13 (treze) métodos de previsão distintos, nas categorias de reconhecimento de sequências, aprendizagem supervisionada e híbridos, para a previsão da geração fotovoltaica.

A. Pattern Sequence-based Forecasting (PSF)

O algoritmo *Pattern Sequence-based Forecasting (PSF)* foi definido em Martínez-Álvarez et al. (2011) e consiste em, basicamente, três etapas. Inicialmente, os dados históricos (série temporal) de geração de energia (*PV data*) são clusterizados e, para cada dia, é associado um rótulo (*label*) respectivo ao identificador do *cluster* ao qual aquele dia está associado. Após isso, uma janela de tamanho w , imediatamente anterior ao dia para o qual deseja-se realizar a previsão, é extraída da série temporal de rótulos; então, é feita uma busca ao longo da série temporal de rótulos por outras ocorrências da sequência (padrão) da janela extraída. Finalmente, são identificados os dias imediatamente posteriores às demais sequências localizadas e, com os dados de geração correspondentes (base de dados de potência), é calculada uma média entre os dias e esse valor é atribuído como a previsão de geração (potência de saída da usina) para o dia a ser previsto. A Figura 2 apresenta um esquemático do método PSF. No exemplo, os dados foram clusterizados em 4 *clusters*, de rótulos C1, C2, C3 e C4; a janela tem tamanho w igual a 3 e, com isso, contém o padrão (C3, C1, C2). A busca na série temporal de rótulos localizou a sequência nos dias 1 a 3 e 47 a 49. Assim, o dia a ser previsto ($d + 1$) teve sua geração estimada pela média da geração dos dias 4 e 50, dias imediatamente após o padrão encontrado (C3, C1, C2), conforme representado na curva em azul.

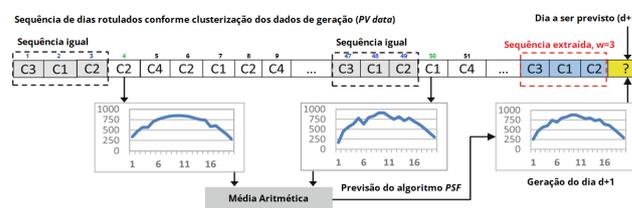


Figura 2. Pattern Sequence-based Forecasting (PSF).

Tabela 2. Estrutura dos dados para aprendizagem supervisionada.

Origem dos dados	Conjunto de <i>features</i>
Dados de geração	$X y \in \mathbb{R}^{730 \times 12}$
Dados de geração +	$X \in \mathbb{R}^{731 \times 10}$
dados meteorológicos W1	$y \in \mathbb{R}^{731 \times 12}$

MLP é amplamente utilizado na literatura para aplicações de previsão de geração de energia elétrica proveniente de fonte solar.

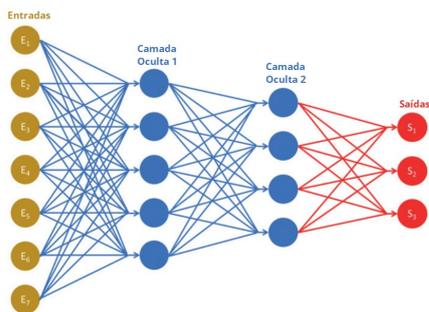


Figura 5. Representação de uma Rede Neural Artificial (RNA).

Foram implementados 2 modelos de RNA MLP, um treinado apenas com dados de potência e outro treinado com dados de potência e meteorológicos (conjunto W1), para avaliar o impacto destes últimos na eficiência dos modelos de previsão. O primeiro modelo foi configurado com função de ativação *tanh*, otimizador *Adam*, taxa de aprendizado constante, regularização L2 igual a 0,05 e duas camadas ocultas de tamanho 10. O segundo, configurado com função de ativação *ReLU*, otimizador *SGD*, taxa de aprendizado adaptativa, regularização L2 igual a 0,05 e três camadas ocultas de tamanhos 5, 10, e 5, respectivamente. Todos estes hiperparâmetros foram selecionados de acordo com a ponderação entre os resultados de três métodos de otimização de hiperparâmetros, a saber, *Grid Search*, *Randomized Search* e *Bayesian Optimization*.

E. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) é um algoritmo baseado em árvores de decisão combinado a *Gradient Boosting* (aumento de gradiente), caracterizado por construir diversas árvores de decisão para extrair os atributos do conjunto de dados e, por meio de aprendizagem supervisionada, implementar o *Gradient Boosting* para favorecimento (impulso/*boosting*) dos parâmetros previsores mais proeminentes, de acordo com a correlação entre entrada e saída (Géron, 2019). O funcionamento geral do algoritmo se dá da seguinte forma: inicialmente, uma árvore de decisão é treinada no conjunto de dados; em seguida, uma segunda árvore é treinada nos resíduos da árvore anterior; o processo se repete até que a condição de parada seja atingida (usualmente um limiar de erro) e os resultados ponderados das árvores determinam a previsão (Antonanzas et al., 2017). A opção pela implementação do algoritmo *XGBoost* neste trabalho se deu por se tratar de um algoritmo em ascensão na literatura de previsão da geração de energia solar fotovoltaica, conforme Sobri et al. (2018), bem como devido o algoritmo apresentar menor custo computacional frente a outros modelos de inteligência artificial, inclusive os demais algoritmos implementados neste trabalho, conforme Machado et al. (2021).

De modo semelhante aos modelos de MLP, foram implementados 2 modelos de *XGBoost*, um treinado apenas com dados de potência e outro tanto com dados de potência

como meteorológicos, em conformidade com o objetivo de avaliar o impacto destes últimos na eficiência dos modelos de previsão. O primeiro modelo foi configurado com peso mínimo, pseudo-regularização e subamostra todas iguais a 1, profundidade máxima por árvore igual a 5, taxa de aprendizado igual a 0,03, regularização L1 igual a 10 e fração de colunas amostradas igual a 1. O segundo, configurado com peso mínimo igual a 5, pseudo-regularização igual a 2, subamostra igual a 0,6, profundidade máxima por árvore igual a 4, taxa de aprendizado igual a 0,03, regularização L1 igual a 0 e fração de colunas amostradas igual a 1. Todos estes hiperparâmetros foram selecionados de acordo com a ponderação entre os resultados de três métodos de otimização de hiperparâmetros, a saber, *Grid Search*, *Randomized Search* e *Bayesian Optimization*.

F. Pattern Sequence Neural Network (PSNN)

Os algoritmos de *Pattern Sequence Neural Network (PSNN)* foram definidos em Lin et al. (2019) e consistem em modificações dos modelos *Pattern Sequence-based Forecasting*. Para os modelos PSNN, as etapas de clusterização e reconhecimento de sequências foram preservadas, porém, no lugar de realizar a previsão do dia a ser previsto ($d + 1$) pela média dos dias identificados pelo algoritmo PSF, utilizou-se estes dias “selecionados” para o treinamento de uma rede neural, sendo a saída da rede a própria previsão desejada. Tendo em vista as razões citadas anteriormente e também visando uma comparação mais coerente entre os diferentes modelos implementados, foi escolhida uma rede neural MLP para cada um dos modelos, doravante denominados PSF-MLP, PSF1-MLP e PSF2-MLP.

O algoritmo PSF-MLP consistiu em executar a clusterização e reconhecimento de sequências do algoritmo PSF e com os dias “selecionados” treinar a rede MLP, utilizando como entrada os dados de geração e meteorológicos dos respectivos dias anteriores e os dados meteorológicos do dia a ser previsto, associados à geração dos dias selecionados (*PV data*); a previsão para o dia $d + 1$ foi então obtida executando o modelo, utilizando os dados de geração e meteorológicos do dia anterior (d) e de previsão meteorológica do dia a ser previsto ($d+1$). A Figura 6 apresenta um esquemático do método PSF-MLP. No exemplo, os dados foram clusterizados em 2 *clusters*, de *labels* C1 e C2; a janela tem tamanho w igual a 3 e, com isso, contém o padrão (C1, C1, C2). A busca na série temporal de rótulos localizou a sequência nos dias 1 a 3 e 66 a 68. Assim, a rede MLP foi treinada com os dados dos dias 3, 4, 68 e 69, conforme supracitado e, então, a previsão da geração (potência de saída da usina) feita pela execução do modelo com os dados dos dias d e $d + 1$.

O algoritmo PSF1-MLP consistiu em executar a clusterização e reconhecimento de sequências do algoritmo PSF1 e com os dias “selecionados” treinar a rede MLP, com os dados estruturados conforme descrito na seção anterior. A Figura 7 apresenta um esquemático do método PSF1-MLP. No exemplo, os dados foram clusterizados em 2 *clusters*, de *labels* C1 e C2; a janela tem tamanho w igual a 3 e, com isso, contém o padrão (C1, C1, C2). A busca na série temporal de rótulos localizou a sequência nos dias 2 a 4 e 66 a 68. Assim, a rede MLP foi treinada com os dados

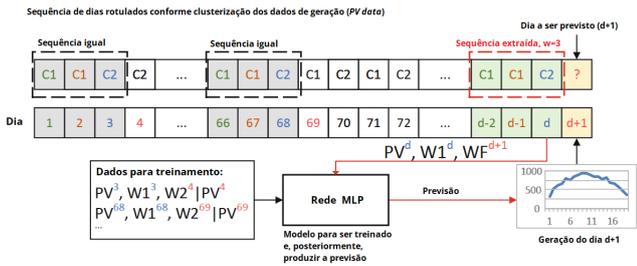


Figura 6. Exemplo de aplicação do algoritmo PSF-MLP.

dos dias 3, 4, 67 e 68, conforme supracitado e, então, a previsão da geração (potência de saída da usina) feita pela execução do modelo com os dados dos dias d e $d + 1$.

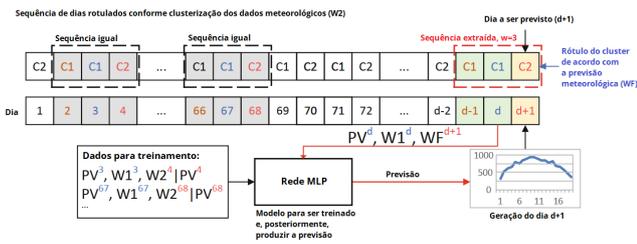


Figura 7. Exemplo de aplicação do algoritmo PSF1-MLP.

O algoritmo PSF2-MLP consistiu em executar a clusterização e reconhecimento de seqüências do algoritmo PSF2 e, com os dias “selecionados” treinar a rede MLP, com os dados estruturados como descrito nas seções anteriores. A Figura 8 apresenta um esquemático do método PSF2-MLP. No exemplo, os dados foram clusterizados, conforme base de dados W1, em 2 clusters, de labels C1 e C2; a janela tem tamanho w igual a 3 e, com isso, contém o padrão (C1, C1, C2). A busca na série temporal localizou a seqüência nos dias 1 a 3, 66 a 68 e 69 a 71, de modo que os dias “pré-selecionados” foram 4, 69 e 72. Conforme a clusterização pela base de dados W2, o dia a ser previsto correspondeu ao cluster K2, enquanto que os dias 4, 69 e 72 da série temporal corresponderam aos clusters K2, K2 e K3, respectivamente. Assim, a rede MLP foi treinada com os dados dos dias 3, 4, 68 e 69, conforme supracitado e, então, a previsão da geração feita pela execução do modelo com os dados dos dias d e $d + 1$.

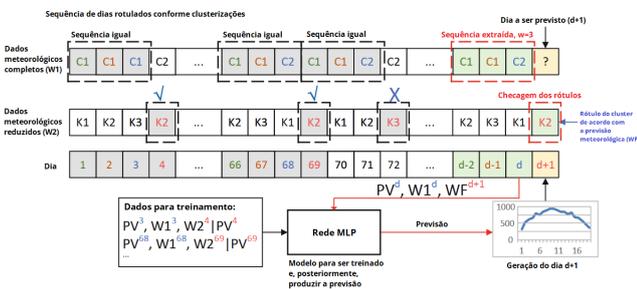


Figura 8. Exemplo de aplicação do algoritmo PSF2-MLP.

Para os algoritmos de *Pattern Sequence Neural Network* (PSNN) as otimizações dos hiperparâmetros dos modelos *PSF*, abordadas anteriormente, foram preservadas, haja vista que a 1ª parte dos algoritmos híbridos consiste na mesma execução do respectivo *PSF* (*PSF*, *PSF1* ou *PSF2*), enquanto que não houve otimização de hiperparâmetros

específica para os modelos de *MLP* utilizados, em razão de que a cada dia de previsão é treinado um novo modelo, específico para a previsão daquele dia, de modo que cada modelo híbrido faz o treinamento de 365 modelos de *RNA/XGBoost*. Caso fossem feitas otimizações para cada modelo e cada algoritmo híbrido, o custo computacional seria impraticável. Assim, preservou-se as configurações selecionadas anteriormente e os modelos diferenciaram-se entre si pelo treinamento.

G. Pattern Sequence Extreme Gradient Boosting

Os algoritmos de *Pattern Sequence Extreme Gradient Boosting* (PSXGB) são propostos pelos autores, explorando o conceito dos algoritmos PSNN (Lin et al., 2019), porém substituindo a aplicação de uma rede neural por um modelo de *XGBoost*, visando usufruir dos benefícios discutidos anteriormente. Em termos práticos, os conceitos discutidos para os algoritmos PSF-MLP, PSF1-MLP e PSF2-MLP se preservam, aplicando-se apenas um modelo diferente no treinamento e previsão da geração. Conseqüentemente, os esquemáticos apresentados nas figuras 6, 7 e 8 são condizentes com os modelos doravante denominados PSF-XGBoost, PSF1-XGBoost e PSF2-XGBoost, a menos, claro, das substituições de rede MLP por modelo *XGBoost*.

Em caráter similar aos algoritmos PSNN, as otimizações dos hiperparâmetros dos modelos *PSF* e *XGBoost* foram preservadas para os algoritmos PSXGB, pela mesma justificativa.

4. RESULTADOS E ANÁLISE DE DESEMPENHO

Para avaliação dos resultados dos algoritmos implementados, utilizou-se como métricas de desempenho a Raiz do Erro Quadrático Médio (*Root Mean Squared Error - RMSE*) e o Erro Absoluto Médio (*Mean Absolute Error - MAE*), tendo em vista serem os indicadores mais utilizados na literatura (Botchkarev, 2019). Seus cálculos foram feitos a partir das funções pré-existentes na biblioteca Python *Scikit-learn*. Como *baseline* de comparação dos métodos foi utilizado um modelo de previsão de persistência no qual a geração do dia $d + 1$ é prevista como igual à do dia anterior (d), em semelhança a Lin et al. (2019). A Tabela 3 apresenta os indicadores supracitados para todos os modelos de previsão implementados, bem como modelo de persistência.

Como principais resultados, verificou-se que:

- (1) O melhor modelo de previsão foi XGBoost2, que foi treinado com todos os dados disponíveis – tanto de geração quanto meteorológicos –, tendo apresentado o melhor desempenho conforme ambos os indicadores, mostrando o grande potencial do algoritmo *XGBoost* para o tema do trabalho.
- (2) O segundo melhor modelo de previsão foi o MLP2, que assim como o XGBoost2 foi treinado com todos os dados disponíveis, o que aponta tanto para o potencial das redes neurais para a aplicação aqui prevista – fato já observado na literatura –, como também para o impacto positivo dos treinamentos utilizando dados meteorológicos, de modo que a correlação entre

Tabela 3. Desempenho dos modelos de previsão implementados.

Modelo	MAE (MW)	RMSE (MW)
Persistência	24,37	34,34
PSF	20,59	26,75
PSF1	22,02	28,20
PSF2	29,13	38,76
MLP1	32,20	38,67
MLP2	21,09	26,40
XGBoost1	31,23	37,73
XGBoost2	17,12	23,50
PSF-MLP	32,33	38,82
PSF-XGBoost	33,72	38,69
PSF1-MLP	22,00	27,68
PSF1-XGBoost	22,53	30,20
PSF2-MLP	32,12	42,21
PSF2-XGBoost	32,07	42,52

condições climáticas e geração de energia em usinas solares se mostrou bastante evidente nos melhores modelos.

- (3) O terceiro melhor modelo foi o PSF, que apesar de não ter reproduzido todo o potencial identificado quando inicialmente proposto, mostrou-se bastante competitivo frente aos demais modelos, sobretudo quando considerados apenas aqueles que não utilizaram dados meteorológicos (entre estes, foi o melhor). Com isso, em situações em que estes dados não estejam disponíveis, o algoritmo PSF apresenta-se como uma boa opção para previsão da geração de energia da planta.
- (4) O melhor modelo híbrido foi o PSF1-MLP, resultado condizente com Lin et al. (2019), o que indica certa vantagem em se utilizar dados meteorológicos na etapa de clusterização. No entanto, a adição de uma segunda etapa de clusterização (PSF2 e modelos híbridos baseados neste) não mostrou ganhos neste trabalho.

Adicionalmente, foram observados também os tempos de execução - *i.e.*, os tempos decorridos desde o início de execução da previsão até o término da mesma - de cada modelo implementado visando quantificar o custo computacional de cada método. A Tabela 4 apresenta os tempos de execução demandados por cada modelo, tanto nas etapas auxiliares, a saber, as etapas de otimização de hiperparâmetros e clusterizações dos dados, como na execução do algoritmo de previsão propriamente dito. É possível observar que, novamente, os modelos de RNA, MLP e *XGBoost* apresentaram os melhores desempenhos e que, apesar de em valores proporcionais existirem diferenças significativas (modelo mais demorado demandou cerca de 13,5 vezes o tempo do modelo mais rápido, por exemplo), em termos de valores absolutos, todos os modelos tiveram tempos de execução, aproximadamente, até 150 segundos (2,5 minutos), de modo que todos os modelos podem ser considerados computacionalmente eficientes para o propósito da aplicação. Todas as execuções foram realizadas no ambiente virtual *Google Colab* (versão gratuita).

5. CONCLUSÕES

Neste trabalho, foi avaliada a aplicação de métodos de Aprendizado de Máquina para a previsão da geração de energia elétrica de uma usina solar fotovoltaica, visando beneficiar a operação e integração à rede de usinas de po-

Tabela 4. Custo computacional (tempos de execução) dos modelos de previsão implementados.

Modelo	Tempo de execução (s)		
	Etapas auxiliares	Execução do modelo	Total
PSF	36,629	9,194	45,823
PSF1	37,675	7,689	45,364
PSF2	38,029	16,662	54,691
MLP1	26,659	62,904	89,563
MLP2	10,599	0,459	11,058
XGBoost1	16,221	52,774	68,995
XGBoost2	44,801	0,298	45,099
PSF-MLP	47,228	35,488	82,716
PSF-XGBoost	81,430	15,906	97,336
PSF1-MLP	48,274	27,037	75,311
PSF1-XGBoost	82,476	56,868	139,344
PSF2-MLP	48,628	43,389	92,017
PSF2-XGBoost	82,830	67,212	150,042

tência de saída variável. Foram implementados e testados diferentes modelos, com diferentes metodologias, notadamente métodos baseados em reconhecimento de sequências, Redes Neurais Artificiais (RNAs), florestas aleatórias (*XGBoost*) e métodos híbridos. Os modelos receberam como entrada os dados históricos de geração da usina e, em alguns modelos, também a série histórica de dados meteorológicos.

Conforme resultados obtidos, verificou-se que o algoritmo *XGBoost* – treinado com dados de potência e meteorológicos – apresentou o melhor desempenho para previsão, seguido pelo modelo MLP, também treinado com dados de geração e meteorológicos. O algoritmo *Pattern Sequence-based Forecast* (PSF), que é mais transparente do que uma RNA ou mesmo o *XGBoost* – o que, em geral, pode ser considerado vantajoso –, mostrou-se competitivo, tendo o melhor desempenho dentre os modelos treinados apenas com dados históricos de geração de energia elétrica. O método *XGBoost*, embora tenha apresentado o melhor desempenho quando comparado aos seus pares, apresentou um erro de cerca de 17 MW, o que representa 10,6% da potência instalada. Isso incentiva a perseguir pela melhora de performance dos métodos ou mesmo investigar outros métodos de Aprendizado de Máquina.

REFERÊNCIAS

- Ahmed, R., Sreeram, V., Mishra, Y., and Arif, M.D. (2020). A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews*, 124, 109792. doi:doi.org/10.1016/j.rser.2020.109792.
- Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martínez-de-Pisón, F.J., and Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. *Solar Energy*, 136, 78–111.
- Antonanzas, J., Urraca, R., Pernía-Espinoza, A., Aldama, A., Fernández-Jiménez, L.A., and Martínez-de-Pisón, F.J. (2017). Single and Blended Models for Day-Ahead Photovoltaic Power Forecasting. In *Hybrid Artificial Intelligent Systems*, Lecture Notes in Computer Science, 427–434. Springer International Publishing, Cham. doi: 10.1007/978-3-319-59650-1_36.
- Botchkarev, A. (2019). Performance metrics (error measures) in machine learning regression, forecasting and

- prognostics: Properties and typology. *Interdisciplinary Journal of Information, Knowledge, and Management*, 14, 45–79. doi:10.28945/4184.
- BP (2021). BP Statistical Review of World Energy 2021. 70.
- Brownlee, J. (2016). Time series forecasting as supervised learning. <https://machinelearningmastery.com/time-series-forecasting-supervised-learning/>. Acesso em: 21 set. 2021.
- Chu, Y., Li, M., Coimbra, C.F., Feng, D., and Wang, H. (2021). Intra-hour irradiance forecasting techniques for solar power integration: a review. *iScience*, 24(10), 103136. doi:10.1016/j.isci.2021.103136.
- de Oliveira, J.F.L., Silva, E.G., and Neto, P.S.G.d.M. (2022). A hybrid system based on dynamic selection for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 33, 3251–3263. doi:10.1109/TNNLS.2021.3051384.
- Diagne, M., David, M., Lauret, P., Boland, J., and Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, 27, 65–76. doi:10.1016/j.rser.2013.06.042.
- Géron, A. (2019). *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn TensorFlow - Conceitos, Ferramentas e Técnicas Para a Construção de Sistemas Inteligentes*. Alta Books, Rio de Janeiro, RJ.
- Hein, H. (2022). Solar atinge novo marco histórico no Brasil: 13 GW de capacidade instalada. <https://cansolar.com.br/solar-atinge-novo-marco-historico-no-brasil-o-de-13-gw-de-capacidade-instalada/>. Acesso em: 6 jan. 2022.
- Heinen, E.D. (2018). *Redes neurais recorrentes e XGBoost aplicados à previsão de radiação solar no horizonte de curto prazo*. Mestrado em ciência da computação, Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de São Carlos, São Carlos, SP.
- Júnior, D.S.d.O., de Oliveira, J., and Neto, P.S.d.M. (2019). An intelligent hybridization of arima with machine learning models for time series forecasting. *Knowledge-Based Systems*, 175, 72–86. doi:10.1016/j.knsys.2019.03.011.
- Kalogirou, S. and Sencan, A. (2010). Artificial intelligence techniques in solar energy applications. In R. Manyala (ed.), *Solar Collectors and Panels*, chapter 15. IntechOpen, Rijeka. doi:10.5772/10343.
- Kamarouthu, P. (2020). *Solar Irradiance Prediction Using Xg-boost With the Numerical Weather Forecast*. Master of science - ms, Utah State University, Logan, Utah.
- Kumari, P. and Toshniwal, D. (2021). Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production*, 318, 128566. doi:10.1016/j.jclepro.2021.128566.
- Lin, Y., Koprinska, I., Rana, M., and Troncoso, A. (2019). Pattern sequence neural network for solar power forecasting. *Communications in Computer and Information Science Neural Information Processing*, 727–737. doi:10.1007/978-3-030-36802-9_77.
- Liu, F.T., Ting, K.M., and Zhou, Z.H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, 413–422. doi:10.1109/ICDM.2008.17.
- Machado, A., Júnior, L., and Nunes, M. (2021). XGBoost na Previsão da Geração de Energia Elétrica em Parques Eólicos. In *Anais da XIV Conferência Brasileira sobre Qualidade da Energia Elétrica*. Galoa. doi:10.17648/cbqee-2021-130586.
- Majid, R. and Mir, S. (2018). Advances in statistical forecasting methods: An overview. *Economic Affairs*, 63, 815–831. doi:10.30954/0424-2513.4.2018.5.
- Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3), e0194889. doi:10.1371/journal.pone.0194889.
- Martínez-Álvarez, F., Troncoso, A., Riquelme, J., and Aguilar-Ruiz, J. (2011). Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, 23(8), 1230–1243.
- Mayer, M.J. and Gróf, G. (2021). Extensive comparison of physical models for photovoltaic power forecasting. *Applied Energy*, 283, 116239.
- Mellit, A. and Pavan, A.M. (2010). A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected pv plant at trieste, italy. *Solar Energy*, 84, 807–821.
- Negnevitsky, M. (2005). *Artificial intelligence: a guide to intelligent systems*. Addison-Wesley, 2 ed. Harlow, England.
- Oliveira, R.F.d. (2018). *Inteligência artificial*. Editora e Distribuidora Educacional S.A, Londrina, Paraná.
- Pereira, E., Martins, F., Costa, R., Gonçalves, A., Lima, F., RÜther, R., Abreu, S., Tiepolo, G., Pereira, S., and Souza, J. (2017). *Atlas Brasileiro de Energia Solar – 2ª Edição*. doi:10.34024/978851700089.
- Raza, M.Q., Nadarajah, M., and Ekanayake, C. (2016). On recent advances in PV output power forecast. *Solar Energy*, 136, 125–144. doi:10.1016/j.solener.2016.06.073.
- Ribeiro, S. (2021). Usinas solares de grande porte atingem 4 GW. <https://revistapotencia.com.br/portal-potencia/energia/usinas-solares-de-grande-porte-atingem-4-gw/>. Acesso em: 21 jan. 2022.
- Silva, I.N., Spatti, D.H., and Flauzino, R.A. (2010). *Redes neurais artificiais para engenharia e ciências aplicadas*. Artliber Editora, São Paulo.
- Sobri, S., Koochi-Kamali, S., and Rahim, N.A. (2018). Solar photovoltaic generation forecasting methods: a review. *Energy Conversion and Management*, 156, 459–497.
- Wang, Z., Koprinska, I., and Rana, M. (2017). Solar power forecasting using pattern sequences. In *Artificial Neural Networks and Machine Learning – ICANN*, Lecture Notes in Computer Science, 486–494. Springer International Publishing. doi:10.1007/978-3-319-68612-7_55.
- Zhang, G.P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50, 159–175. doi:10.1016/S0925-2312(01)00702-0.