



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA MECÂNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA**

DAVID MICKELY JARAMILLO LOAYZA

**PREVISÃO DE IRRADIAÇÃO SOLAR USANDO APRENDIZAGEM DE
MÁQUINAS COM MODELOS DE ÁRVORE E A INFLUÊNCIA DO ÍNDICE
OCEÂNICO DO EL NIÑO (ONI)**

**FORTALEZA
2021**

DAVID MICKELY JARAMILLO LOAYZA

PREVISÃO DE IRRADIAÇÃO SOLAR USANDO APRENDIZAGEM DE MÁQUINAS
COM MODELOS DE ÁRVORE E A INFLUÊNCIA DO ÍNDICE OCEÂNICO DO EL
NIÑO (ONI)

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Engenharia Mecânica. Área de concentração: Processos, Equipamentos e Sistemas para Energias Renováveis.

Orientador: Prof. Dr. Paulo Alexandre Costa Rocha.

FORTALEZA
2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- J41p Jaramillo Loayza, David Mickely.
Previsão de irradiação solar usando aprendizagem de máquinas com modelos de árvore e a influência do Índice Oceânico El Niño (ONI) / David Mickely Jaramillo Loayza. – 2021.
37 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Mecânica, Fortaleza, 2021.
Orientação: Prof. Dr. Paulo Alexandre Costa Rocha.
1. XGBoost. 2. Aprendizagem de máquinas. 3. Energia Solar. 4. Oceanic Niño Index. 5. ENOS. I.
Título.

CDD 620.1

DAVID MICKELY JARAMILLO LOAYZA

PREVISÃO DE IRRADIAÇÃO SOLAR USANDO APRENDIZAGEM DE
MÁQUINAS COM MODELOS DE ÁRVORE E A INFLUÊNCIA DO ÍNDICE
OCEÂNICO DO EL NIÑO (ONI)

Dissertação apresentada ao Programa de
Pós-Graduação em Engenharia Mecânica,
do Centro de Tecnologia da Universidade
Federal do Ceará, como requisito parcial
para a obtenção do Título de Mestre em
Engenharia Mecânica. Área de
Concentração: Processos, Equipamentos
e Sistemas para Energias Renováveis.

Aprovada em 23/12/2021

BANCA EXAMINADORA

Prof. Dr. Paulo Alexandre Costa Rocha (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. André Valente Bueno
Universidade Federal do Ceará (UFC)

Prof. Dr. Matheus Pereira Porto
Universidade Federal de Minas Gerais (UFMG)

A Deus.
Aos meus pais, Zoraida e Yofre;
Irmã, Kimberly
e noiva, Daniella.

AGRADECIMENTOS

À CAPES, pelo apoio financeiro com a manutenção da bolsa de auxílio.

Ao Prof. Dr. Paulo Alexandre Costa Rocha, pela excelente orientação.

Aos professores participantes da banca examinadora Prof. Dr. André Valente Bueno e Prof. Dr. Matheus Pereira Porto pelo tempo, pelas valiosas colaborações e sugestões.

Aos colegas da turma de mestrado, pelas reflexões, críticas e sugestões recebidas.

RESUMO

A energia solar representa a coluna vertebral das fontes de energia sustentáveis e pela sua influência sobre as demais, sua participação na mudança da matriz energética é indispensável ainda mais nesse período crítico climático que está vivendo o planeta. No contexto apresentado, a energia solar é uma fonte intermitente e pouco confiável, sendo necessários sistemas complementários como modelos de previsão de irradiação solar que ajudem no controle eficiente da demanda de eletricidade. Os modelos de aprendizagem de máquinas são métodos que podem ser de regressão ou de classificação e estão sendo usados para desenvolver modelos de previsão de irradiação solar global, especificamente em modelos preditivos que podem ser usados para otimizar os sistemas de controle das usinas solares. O objetivo deste trabalho é implementar e comparar o desempenho dos modelos de aprendizagem de máquinas XGBoost e Boosting na previsão de irradiação solar global, usando o Índice Oceânico El Niño (do inglês ONI) como preditor em cinco bancos de dados diferentes: Global com ONI; Global, La Niña, El Niño e Neutro sem ONI, para os horizontes temporais 2 min, 10 min, 30 min, 1h, 1 dia e 7 dias, sendo analisados sob as mesmas condições climáticas na cidade de Fortaleza, no estado do Ceará, 3° 43' 2''S, 38° 32' 35''O. Foram analisados o banco Global que contém todas as observações incluindo o ONI como preditor; e os bancos que não usaram o ONI como preditor: Global, La Niña, El Niño e Neutro, a fim de conhecer o efeito do ONI sobre a previsão de irradiação solar. O banco Global com ONI teve valores menores de RMSE em comparação com o banco Global sem ONI para os horizontes 2 min e 10 min. O modelo XGBoost no banco Global com ONI mostrou um valor do nRMSE 52.08% menor que nos outros bancos no horizonte de 2min. O modelo Boosting apresentou no banco Neutro sem ONI um valor do nRMSE 39.26% menor do que nos outros bancos para o horizonte de 1 h. No desempenho geral dos modelos, o Boosting teve melhores resultados, mas com a análise dos valores da FS percebeu-se que o modelo XGBoost mostrou um desempenho similar com valores de menos de 1% de diferença. Ao observar os valores da FS nos bancos que contêm observações com ocorrência do ONI (Global, La Niña e El Niño) percebe-se a tendência positiva afetando a sensibilidade da previsão de irradiação solar.

Palavras-chave: XGBoost. Aprendizagem de máquinas. Energia Solar. Oceanic Niño Index. ENOS. La Niña. El Niño.

RESUMEN

La energía solar representa la columna vertebral de las fuentes de energías sustentables y por su influencia sobre las demás, su participación en el cambio de matriz energética es indispensable sobre todo en este periodo crítico climático que está viviendo el planeta. En el contexto presentado, la energía solar es una fuente intermitente e poco confiable, siendo necesarios sistemas complementarios como modelos de previsión de irradiación solar que ayuden en el control eficiente de la demanda de electricidad. Los modelos de aprendizaje de máquinas son métodos que pueden ser de regresión o de clasificación y están siendo usados para desarrollar modelos de previsión de irradiación solar global, específicamente en modelos predictivos que poder ser usados para optimizar los sistemas de control de las plantas de energía. El objetivo de este trabajo es implementar y comparar el desempeño de los modelos XGBoost y Boosting en la previsión de irradiación solar, usando el Índice Oceánico El Niño (del inglés ONI) como predictor en cinco bancos de datos diferentes: Global con ONI; Global, La Niña, El Niño y Neutro sin ONI, para los horizontes temporales 2 min, 10 min, 30 min, 1h, 1 día, 7 días, siendo analizados bajo las mismas condiciones climáticas en la ciudad de Fortaleza, en el estado de Ceará, $3^{\circ} 43' 2''\text{S}$, $38^{\circ} 32' 35''\text{O}$. Fueron analizados el banco Global que contiene todas las observaciones incluyendo el ONI como predictor; y los bancos que no usaron el ONI como predictor: Global, La Niña, El Niño y Neutro, con el fin de conocer el efecto del ONI sobre la previsión de irradiación solar. El banco Global con ONI obtuvo valores menores de RMSE en comparación con el banco Global sin ONI para los horizontes 2 min y 10 min. El modelo XGBoost en el banco Global con ONI mostró un valor de nRMSE 52.08% menor que en los otros bancos no horizonte 2 min. El modelo Boosting presentó en el banco Neutro sin ONI un valor de nRMSE 39.26% menor de que en los otros bancos para el horizonte 1 h. En el desempeño general de los modelos, el Boosting tuvo mejores resultados, pero con el análisis de los valores de FS se percibió que el modelo XGBoost mostró un desempeño similar con valores de menos de 1% de diferencia. Al observar los valores de FS en los bancos que contienen observaciones con ocurrencia del ONI (Global, La Niña y El Niño) se percibe la tendencia positiva afectando la sensibilidad de la previsión de irradiación solar.

Palabras clave: XGBoost. Aprendizaje de máquinas. Energía Solar. Oceanic Niño Index. ENOS. La Niña. El Niño.

LISTA DE FIGURAS

Figura 1 – Diagrama de fluxo do processo.	22
--	----

LISTA DE GRÁFICOS

Gráfico 1 – Produção científica anual sobre previsão solar e aprendizagem de máquinas	18
Gráfico 2 – Países com maior número de citações no período 2000 – 2021.	18
Gráfico 3 – Média móvel de três meses do ONI de 1950 até o presente.	19
Gráfico 4 – Variação do nRMSE em função da porcentagem do banco de dados Global com ONI usado na execução dos modelos.	25
Gráfico 5 – Variação do FS em função da porcentagem do banco de dados Global com ONI usado na execução dos modelos.	26
Gráfico 6 – Variação da Habilidade de previsão, conjunto Global com o preditor ONI.....	27
Gráfico 7 – Variação da Habilidade de previsão, conjunto Global sem o preditor ONI	28
Gráfico 8 – Variação do nRMSE, conjunto La Niña sem o preditor ONI.....	30
Gráfico 9 – Variação da Habilidade de previsão, conjunto La Niña sem o preditor ONI	30
Gráfico 10 – Variação do nRMSE, conjunto El Niño sem o preditor ONI	31
Gráfico 11 – Variação da Habilidade de previsão, conjunto El Niño sem o preditor ONI	32
Gráfico 12 – Variação do nRMSE, conjunto Neutro sem o preditor ONI.....	33
Gráfico 13 – Variação da Habilidade de previsão, conjunto Neutro sem o preditor ONI.....	34
Gráfico 14 – nRMSE da previsão realizada com cada banco de dados.....	35

LISTA DE TABELAS

Tabela 1 – Variáveis usadas como preditores em trabalhos de previsão de irradiação solar...	20
Tabela 2 – Resultados dos modelos de previsão, conjunto Global com o preditor ONI.....	27
Tabela 3 – Resultados dos modelos de previsão, conjunto Global sem o preditor ONI	28
Tabela 4 – Resultados dos modelos de previsão, conjunto La Niña sem o preditor ONI	29
Tabela 5 – Resultados dos modelos de previsão, conjunto El Niño sem o preditor ONI.....	31
Tabela 6 – Resultados dos modelos de previsão, conjunto Neutro sem o preditor ONI..	33

LISTA DE ABREVIATURAS E SIGLAS

ONI	<i>Oceanic Niño Index</i> (Índice Oceânico El Niño)
ENOS	El Niño Oscilação Sul
STISF	<i>Short-term Solar Irradiance Forecasting</i> (Previsão de Irradiância Solar a Curto Prazo)
ANN	<i>Artificial Neural Networks</i> (Redes Neurais Artificiais)
SFP	<i>Statistical Feature Parameters</i> (Parâmetro de Recurso Estatístico)
ANN-HDS	<i>ANN model using Historical Data Series</i> (Modelo ANN que usa Series de Dados Históricos)
DOI	Dipolo do Oceano Índico
MAD	<i>Median Absolute Deviation</i> (Desvio Absoluto da Mediana)
SVM	<i>Support Vector Machine</i> (Máquina de Vetor de Suporte)
XGBOOST	<i>Extreme Gradient Boosting</i> (Gradiente de Aumento Extremo)
RFR	<i>Random Forest Regression</i> (Regressão de Florestas Aleatórias)
GBR	<i>Gradient Boosted Regression</i> (Regressão de Gradiente Aumentado)
SVR	<i>Support Vector Regression</i> (Regressão de Vetor de Suporte)
MLP's	<i>Multilayer Perceptrons</i> (Perceptrons Multicamadas)
RMSE	<i>Root Mean Square Error</i> (Raiz do Erro Quadrático Médio)
nRMSE	<i>Normalized Root Mean Square Error</i> (Raiz do Erro Quadrático Médio Normalizado)
MAE	<i>Mean Absolute Error</i> (Erro Médio Absoluto)
nMAE	<i>Normalized Mean Absolute Error</i> (Erro Médio Absoluto Normalizado)
FS	<i>Forecast Skill</i> (Habilidade de Previsão)
Kt	Índice de claridade
SST	<i>Sea Surface Temperature</i> (Temperatura da Superfície do Mar)

SUMÁRIO

1 INTRODUÇÃO.....	14
2 FUNDAMENTAÇÃO TEÓRICA	17
2.1 Bibliometria	17
2.2 Oceanic Niño Index.....	18
2.3 Variáveis de Entrada (Preditores)	19
2.4 Modelo de Persistência.....	19
2.5 Regressão de Gradiente Aumentado (Boosting)	20
2.6 Gradiente de aumento extremo (XGBoost)	21
3 METODOLOGIA	22
3.1 Banco de dados e pré-processamento.....	22
3.2 Validação cruzada	23
3.3 Métricas de erro	23
4 RESULTADOS E DISCUSSÃO.....	25
4.1 Resultados obtidos com o banco de dados completo	26
4.2 Resultados obtidos com o banco La Niña.....	29
4.3 Resultados obtidos com o banco El Niño	30
4.4 Resultados obtidos com o banco Neutro	32
5 CONCLUSÃO.....	36
REFERÊNCIAS.....	37

1 INTRODUÇÃO

A energia solar é uma das fontes renováveis mais abundantes no mundo, o aproveitamento eficiente dela vem sendo um dos maiores desafios na área do desenvolvimento de tecnologias que usam energias renováveis. Nos sistemas de controle das usinas baseadas em energia solar é importante o uso de modelos de previsão de irradiação solar para poder abastecer eficientemente a demanda de eletricidade.

O desenvolvimento de projetos de geração de eletricidade usando fontes renováveis vem aumentando significativamente no Brasil, tendo o estado do Ceará entre os mais destacados e pioneiros na mudança da matriz energética. Na investigação realizada por Freitas *et al.* (2020) foram analisados três cenários (conservador, de transição e 100% renovável) de geração de eletricidade para o estado do Ceará em 2050, tendo como resultado que nos três cenários destacam-se as fontes renováveis eólica e solar com crescimento considerável de participação no mínimo de 50% na matriz energética.

A característica aleatória da radiação solar sob condições climáticas variáveis é um dos principais problemas no desenvolvimento de modelos de previsão de irradiação solar de curto prazo (do inglês STSIF). No estudo realizado por Wang *et al.* (2012), estes desenvolveram um novo modelo de redes neurais artificiais (do inglês ANN) que usa parâmetros de recurso estatístico (do inglês SFP) da irradiância e a temperatura ambiente. Os resultados demonstraram que a precisão do modelo de previsão ANN-SFP melhorou em comparação com o modelo convencional ANN-HDS, para dias ensolarados os valores do MAPE, RMSE e MABE apresentam diferença entre 2 – 16%, enquanto para os dias nublados a diferença foi entre 66-67%.

A fim de estabelecer qual é a influência dos fenômenos climáticos do El Niño Oscilação Sul (ENOS) e do Dipolo do Oceano Índico (DOI) na radiação solar na Austrália, Davy e Troccoli (2012) realizaram uma análise do impacto usando um modelo linear simples de efeitos mistos e uma técnica de bootstrap. Os resultados que eles obtiveram indicaram que o ENOS gerou alterações na energia solar em 10% dos locais que foram analisados.

Atualmente os modelos de previsão não são totalmente confiáveis para escalas de tempo sazonais ou anuais devido a múltiplos fatores que influenciam as mudanças do clima. Mohammadi e Goudarzi (2018) fizeram uma análise da influência do ENOS sobre a irradiação solar total horizontal, velocidade do vento e precipitação em oito localidades com diferentes

condições climáticas na Califórnia. Usando a medida estatística MAD (desvio absoluto da mediana) para determinar a variabilidade dos parâmetros analisados, concluíram que durante os eventos mais fortes do El Niño aconteceram as maiores variações, com diminuição da irradiação solar, tendo um efeito contrário sobre a precipitação para as regiões da Califórnia.

Bakker *et al.* (2019) determinaram que os preditores mais importantes que devem ser usados no desenvolvimento de modelos de previsão probabilística são: irradiação global, irradiação direta, umidade relativa, nebulosidade e precipitação, após terem feito uma comparação de 7 métodos de regressão. Eles compreenderam que o desempenho dos métodos depende principalmente da hora do dia, do tempo de previsão, da localização geográfica, da estação e das condições climáticas. De acordo com os diagramas de confiabilidade, os métodos de árvore foram os mais confiáveis.

Pedro e Coimbra (2015) desenvolveram uma nova metodologia de previsão para modelos intra-horários (de 5 min a 30 min) baseado na otimização do reconhecimento de padrões de telemetria local e imagens do céu, obtendo melhoras consideráveis na previsão, entre 10% a 25%, em comparação com o modelo de Persistência.

Li *et al.* (2016) estudaram o uso dos métodos de aprendizagem de máquinas, Hidden Markov e SVM, na previsão de irradiação solar a curto prazo a fim de demonstrar que o desempenho deles foi ótimo apresentando valores acima de 90% na precisão da previsão e podem também ser usados sob diferentes condições climáticas.

Torres-Barrán *et al.* (2019) realizaram uma comparação entre os métodos *Random Forest Regression* (RFR), *Gradient Boosted Regression* (GBR) e *Extreme Gradient Boosting* (XGBoost) versus os métodos *Support Vector Regression* (SVR) e *Multilayer Perceptrons* (MLP's) aplicados na previsão da energia eólica e irradiação solar. Os resultados que obtiveram para a previsão de irradiação solar foi que os métodos GBR e XGBoost apresentaram o melhor desempenho com MAE de 13.25% na fase de teste, seguidos do SVR (13.46%) e por último o RFR (13.77%), já o MLP foi o método com o pior desempenho tanto para previsão de energia eólica como de energia solar.

O objetivo principal da presente dissertação é avaliar o desempenho dos modelos de árvore XGBoost e Boosting na previsão de irradiação solar sob a influência do ONI para a cidade de Fortaleza. Como objetivos específicos se determinaram:

- Executar e comparar o desempenho dos modelos de árvore com a Persistência na previsão de irradiação solar para os horizontes temporais 2 min, 10 min, 30 min, 1h, 1 dia e 7 dias.
- Analisar o efeito do ONI na acurácia da previsão.
- Avaliar o comportamento da previsão nos diferentes períodos de ocorrência do ENOS.

A presente dissertação está organizada da seguinte forma: A Seção 2 envolve a informação principal sobre os modelos de aprendizagem de máquinas analisados, preditores e demais conceitos abordados no trabalho. A Seção 3 detalha a metodologia usada para o desenvolvimento da previsão de irradiação solar. Já a Seção 4 abrange a análise e discussão sobre os resultados obtidos dos modelos previsão propostos. Finalmente, a Seção 5 apresenta as conclusões e limitações dos modelos estudados nesta dissertação.

2 FUNDAMENTAÇÃO TEÓRICA

Os modelos de aprendizagem de máquinas são conhecidos como métodos de inteligência artificial e são usados de forma extensa na previsão de irradiação solar global, segundo explica Feng *et al.* (2019).

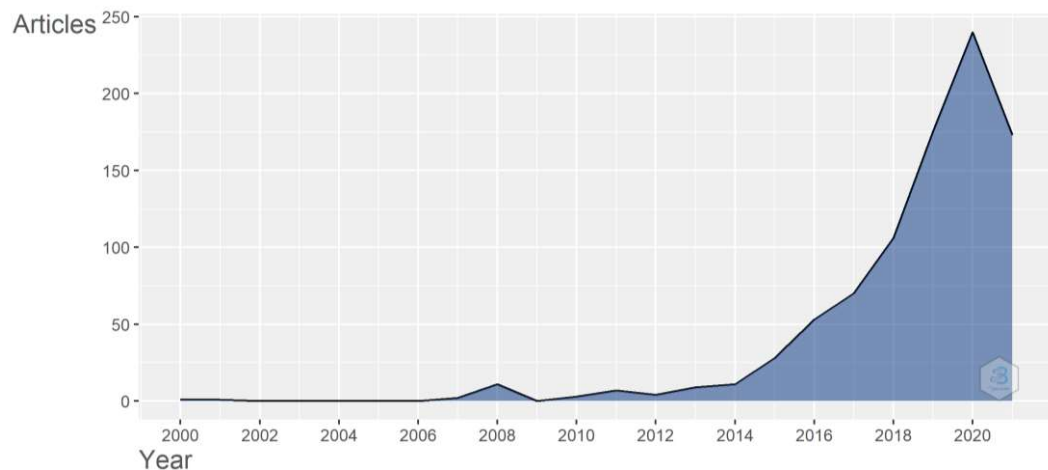
Durante os últimos anos foram desenvolvidos diferentes métodos de previsão de irradiação solar global tais como modelos empíricos, modelos baseados na nebulosidade, na temperatura, entretanto, atualmente se precisa de modelos de previsão de alta qualidade e confiabilidade. Fan *et al.* (2018) realizaram uma análise dos métodos *Support Vector Machine* (SVM) e *Extreme Gradient Boosting* (XGBoost) juntamente com modelos empíricos para determinar qual deles apresentava maior precisão na previsão de irradiação solar global em clima subtropical úmido, tendo como resultado que o SVM e o XGBoost mostraram melhor desempenho que os modelos empíricos. O modelo XGBoost apresentou um aumento médio de 6.3% no RMSE comparado ao 10.5% do SVM. Em referência à velocidade de computação o XGBoost mostrou maior velocidade (3.02 seg e 0.05 seg na fase de treinamento e teste respectivamente) a diferença do SVM (27.48 seg e 4.13 seg).

2.1 Bibliometria

Conforme a pesquisa realizada sobre os tópicos *solar forecasting* (previsão solar) e *machine learning* (aprendizagem de máquinas), no Gráfico 1 pode-se observar o aumento quase exponencial na última década das pesquisas sobre essa área da energia solar, demonstrando a importância de desenvolvimento de estudos sobre previsão de irradiação solar.

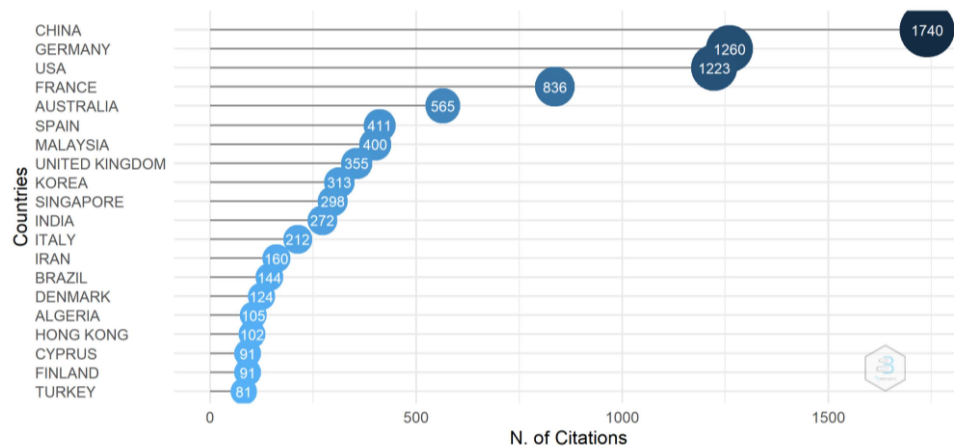
Observa-se no Gráfico 2, com 144 citações, que o Brasil encontra-se entre os 20 países com maior número de citações no período 2000 – 2021. Isso o coloca como um dos pioneiros no estudo da previsão solar usando aprendizagem de máquinas.

Gráfico 1 – Produção científica anual sobre previsão solar e aprendizagem de máquinas



Fonte: elaborado pelo autor com a ferramenta Biliometrix.

Gráfico 2 – Países com maior número de citações no período 2000 – 2021.

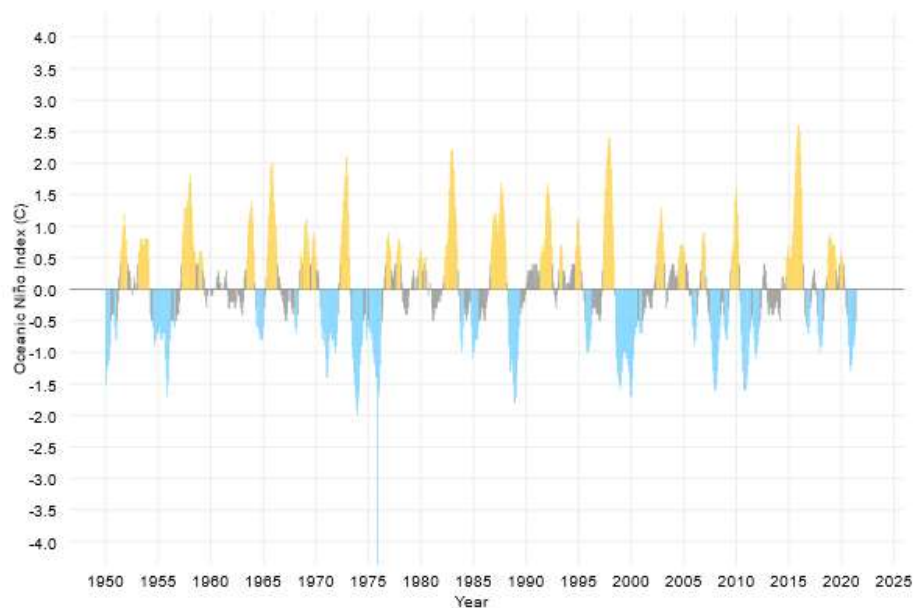


Fonte: elaborado pelo autor com a ferramenta Biliometrix.

2.2 Oceanic Niño Index

Andreoli e Kayano (2005) definem o ONI como a média móvel de 3 meses da variação da *Sea Surface Temperature* (SST) na região 3.4 do El Niño. De acordo com Trenberth (2019), o evento El Niño (fase quente) se apresenta quando existe aquecimento anormal $\geq +0.5$, enquanto o evento La Niña (fase fria) percebe-se ao existir esfriamento anormal ≤ -0.5 . Nos períodos que o ONI é menor que $+0.5$ e maior que -0.5 determina-se ausência dos eventos.

Gráfico 3 – Média móvel de três meses do ONI de 1950 até o presente.



Fonte: Dahlman (2009).

Os valores que o ONI pode assumir se apresentam no Gráfico 3. O evento El Niño (fase quente) se representa pela cor amarela, enquanto La Niña (fase fria) se representa pela cor azul e na cor cinza os períodos de ausência dos eventos.

2.3 Variáveis de Entrada (Preditores)

Além dos dados de irradiação solar, as variáveis comumente usadas como preditores são as medições meteorológicas como umidade relativa, velocidade e direção do vento, temperatura do ar, precipitação, nebulosidade e pressão; conforme explica Antonanzas *et al.* (2016), esse conjunto de preditores é conhecido como dados exógenos. Na Tabela 1 pode-se observar as variáveis usadas na previsão de irradiação solar em alguns trabalhos publicados do período 2016 – 2021.

2.4 Modelo de Persistência

É um modelo de previsão a curto prazo considerado como referência para contrastar o desempenho de outros modelos, geralmente é um modelo impreciso para horizontes acima de 1 hora, conforme explica Diagne *et al.* (2013). A Equação 1 corresponde ao modelo de Persistência do Índice de Claridade, onde K_t no tempo $t + 1$ é a melhor predição do índice de claridade para o mesmo valor no instante t anterior.

$$K_{(t+1)} = K_t \quad (1)$$

Tabela 1 – Variáveis usadas como preditores em trabalhos de previsão de irradiação solar.

Variáveis	Autores/Ano					
	Gutierrez-Corea et al. (2016)	Cervone et al. (2017)	Rocha et al. (2019)	Chandola et al. (2020)	Zambrano e Giraldo (2020)	Lai et al. (2021)
Irradiação Global	✓	✓	✓	✓	✓	✓
Temperatura	✓	✓	–	✓	✓	✓
Temperatura máx.	–	–	✓	–	–	–
Temperatura mín.	–	–	✓	–	–	–
Umidade Relativa	✓	–	✓	✓	✓	✓
Velocidade do vento	✓	–	✓	✓	✓	✓
Direção do vento	✓	–	✓	✓	–	✓
Irradiação extraterrestre	✓	–	✓	–	–	✓
Índice de claridade	✓	–	–	–	–	✓
Ângulo horário solar	✓	–	✓	–	–	–
Ângulo do zênite	✓	✓	–	–	–	✓
Nebulosidade	–	✓	✓	–	✓	–
Ângulo azimute	–	✓	–	–	–	–
Dia do ano	–	–	✓	–	✓	–
Precipitação	–	–	✓	–	✓	–
Orvalho	–	–	✓	✓	–	–
Pressão	–	–	–	✓	✓	✓
Radiação normal direta	–	–	–	✓	–	–
Radiação horizontal difusa	–	–	–	✓	–	–

Fonte: elaborada pelo autor.

2.5 Regressão de Gradiente Aumentado (Boosting)

Segundo Nagy *et al.* (2016), o Boosting é usado em regressões de gradiente aumentado com árvores de decisão. O preditor gerado é uma combinação de aprendizagens de árvores de decisão fracos que são construídos de forma iterativa usando a gradiente negativa de uma função de perda. A Equação 2 representa o modelo boosting, segundo Nielsen (2016).

$$\hat{f}(x) = \hat{f}^{(M)} = \sum_{m=1}^M \hat{\theta}_m \hat{c}_m(x) \quad (2)$$

Onde $\hat{c}_m(x) \in \{-1, 1\}$ são os classificadores fracos e os classificadores forte são $\hat{c}(x) = \text{sign}(\hat{f}(x))$.

2.6 Gradiente de aumento extremo (XGBoost)

O XGBoost é um dos mais conhecidos modelos de gradiente de aumento, funciona adicionando e treinando novas árvores para ajustar os erros residuais de cada iteração, segundo Dong *et al.* (2020). Chen e Guestrin (2016) destacam que o XGBoost é um modelo reconhecido pela sua mais importante característica: poder escalar em todos os cenários enquanto realiza o processo de cálculo até 10 vezes mais rápido do que outros modelos já usados na área. Fan *et al.* (2018) definem a Equação 3 como:

$$f_i^{(t)} = \sum_{k=1}^t f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \quad (3)$$

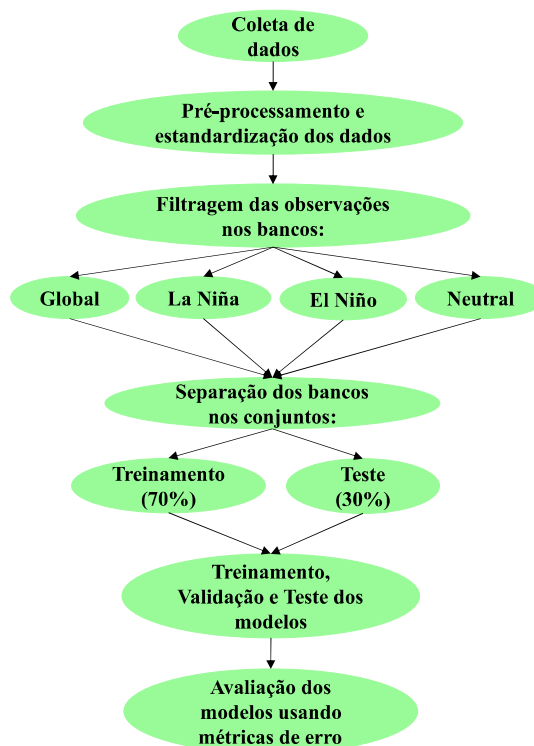
Onde $f_k(x_i)$ é o aprendiz no tempo t ; $f_i^{(t)}$ e $f_i^{(t-1)}$ são as previsões para os tempos t e $t - 1$, finalmente x_i é a variável de entrada.

3 METODOLOGIA

Esta seção apresenta quais foram os preditores usados no treinamento e teste dos modelos, as características do banco de dados, os modelos de aprendizagem de máquinas usados para desenvolver as previsões de irradiação solar.

O desenvolvimento deste trabalho foi realizado no software RStudio usando a linguagem de programação de código aberto R.

Figura 1 – Diagrama de fluxo do processo.



Fonte: elaborada pelo autor.

No esquema mostrado na Figura 1 pode-se observar de forma geral o processo realizado nesta dissertação, que vai ser detalhado nesta seção.

3.1 Banco de dados e pré-processamento

Os dados de irradiação solar foram coletados no Laboratório de Energia Solar e Gas Natural (LESGN) localizado no Campus Pici da Universidade Federal do Ceará na cidade de Fortaleza, Ceará, 3° 43' 2''S, 38° 32' 35''O, enquanto aos dados meteorológicos foram obtidos da Fundação Cearense de Meteorologia e Recursos Hídricos (Funceme).

Após a análise da literatura sobre os preditores usados em trabalhos de previsão de irradiação solar, de acordo Brasil (2020a), foram 25 o número de preditores a serem considerados nesta dissertação, no caso, temporais como: ano, dia juliano e hora da coleção; medições meteorológicas como: velocidade (m/s) e direção do vento (graus), umidade do ar (%), temperatura (°C), precipitação (mm) e ONI; junto à irradiação global horizontal de 16 instantes anteriores.

Realizou-se a padronização da variável dia do ano e o cálculo do índice de claridade (Kt) segundo Duffie e Beckman (2006). O banco de dados compreende as observações desde 2007 até 2019, encontra-se as medições em resolução de 2 min, e é nomeado como Global por conter o total de observações. Derivas-se do banco Global os subconjuntos La Niña, El Niño e Neutro, o critério de subdivisão é baseado no valor do ONI respectivamente para cada evento meteorológico.

3.2 Validação cruzada

Previamente à execução dos modelos XGBoost e Boosting é realizada mais uma divisão aleatória do banco, considerando 70% para treinamento com validação cruzada para 10-fold e 30% para teste. Ressalta-se que esse procedimento não é realizado para o modelo da Persistência.

3.3 Métricas de erro

O desempenho dos modelos é avaliado usando as métricas de erro: raiz do erro quadrático médio (do inglês RMSE), raiz do erro quadrático médio normalizado (do inglês nRMSE), erro médio absoluto (do inglês MAE), erro médio absoluto normalizado (do inglês nMAE), Habilidade de Previsão (do inglês FS), conforme, entre outras fontes, Brasil (2020a).

Para as Equações 4, 5, 6 e 7; $\hat{H}_t$ é o valor previsto pelo modelo, H_t é o valor observado, N é o número de observações que contém o banco de teste e $\bar{H}_t$ é a média dos valores de H_t .

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{H}_t - H_t)^2} \quad (4)$$

$$nRMSE = \frac{RMSE}{\bar{H}_t} \quad (5)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{H}_t - H_t| \quad (6)$$

$$nMAE = \frac{MAE}{\bar{H}_t} \quad (7)$$

Segundo explica Fan *et al.* (2020), quanto menor os valores das métricas mencionadas previamente melhor vai ser o desempenho dos modelos analisados.

Com o propósito de comparar o desempenho dos modelos XGBoost e Boosting ao modelo da Persistência usa-se a Habilidade de Previsão, de acordo Kumari e Toshniwal (2021), Equação 8:

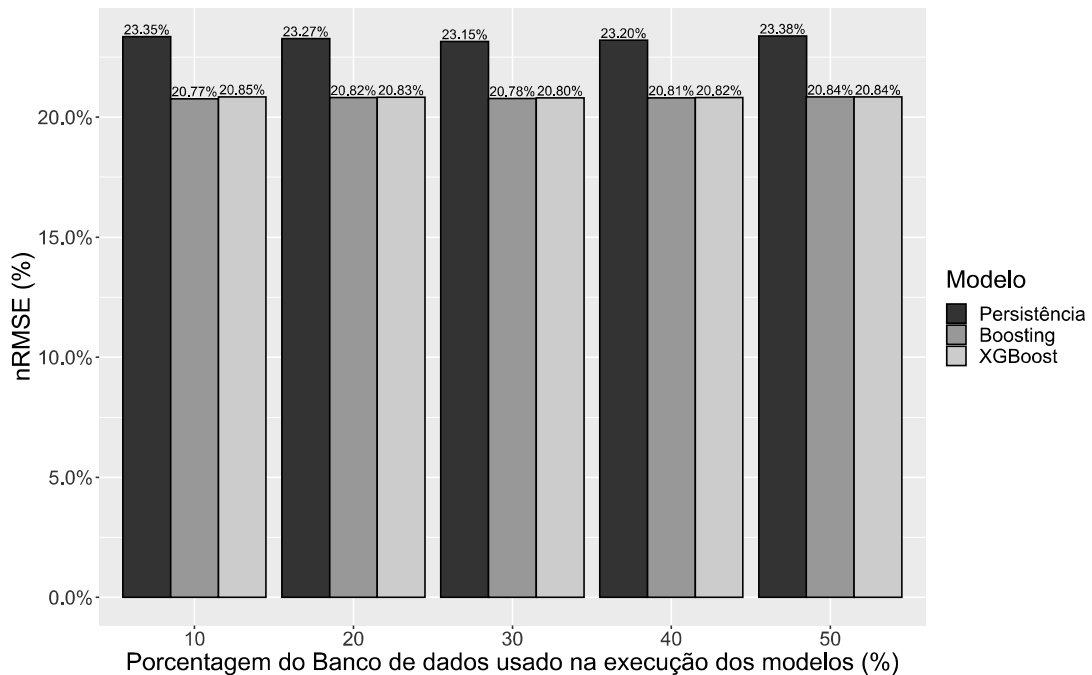
$$FS = 1 - \frac{RMSE_{\text{modelo}}}{RMSE_{\text{persistência}}} \quad (8)$$

4 RESULTADOS E DISCUSSÃO

Previamente à análise dos modelos propostos, foi feito o pré-processamento do banco de dados completo com o objetivo de remover os valores ausentes e padronizar o conjunto de forma geral de acordo com os parâmetros de entrada das funções.

Para facilitar o processamento dos dados pelo software utilizado, especificamente para o caso de resolução temporal de 2 minutos, foi necessário reduzir a quantidade das observações (512.267) usadas no treinamento e teste dos modelos. Realizaram-se execuções dos modelos mudando a porcentagem (10, 20, 30, 40 e 50) do banco de dados Global para determinar uma quantidade adequada de observações a serem usadas sem impor um grande peso computacional, e nos Gráficos 4 e 5 pode-se observar que a variação do nRMSE e FS é menor do que 5%, portanto o uso de 50% das observações é adequado, por não afetar os resultados das métricas de erros e facilitar o desempenho computacional.

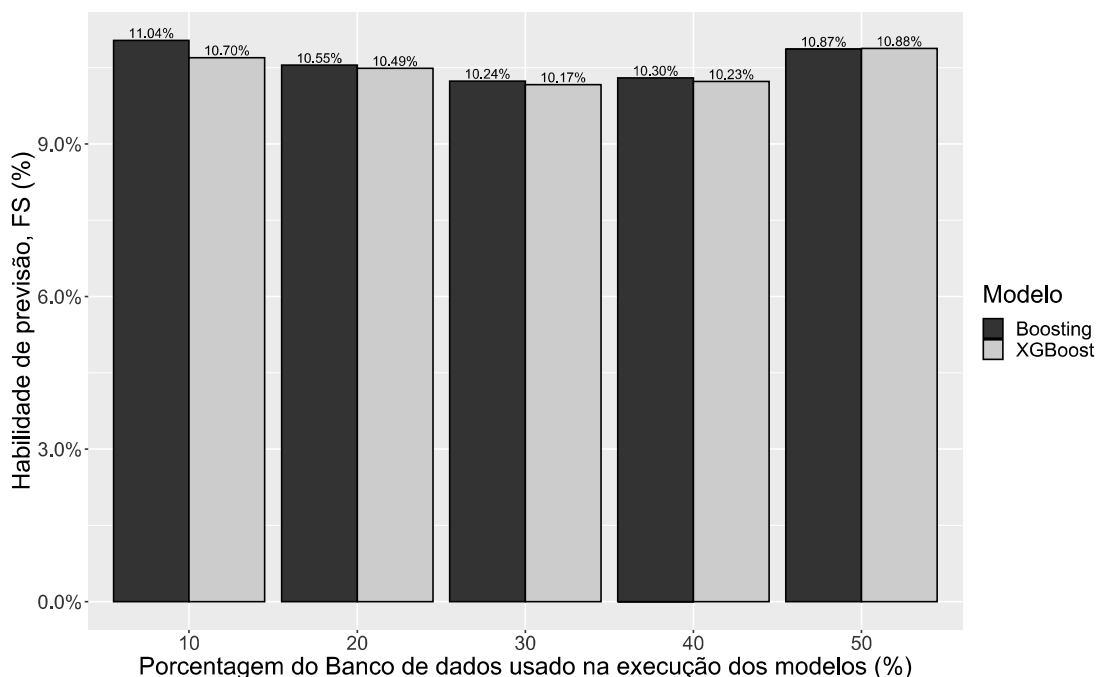
Gráfico 4 – Variação do nRMSE em função da porcentagem do banco de dados Global com ONI usado na execução dos modelos.



Fonte: elaborado pelo autor.

Vale ressaltar novamente que a remoção aleatória dessas observações foi aplicada só para o horizonte de 2 min, uma vez que para os outros horizontes o tamanho do banco de dados não impôs grande peso computacional.

Gráfico 5 – Variação do FS em função da porcentagem do banco de dados Global com ONI usado na execução dos modelos.



Fonte: elaborado pelo autor.

Em adição, foram analisados cinco subconjuntos de observações para cada horizonte temporal, a partir de agora nomeados como o Global (que contém o total das observações) com o ONI como preditor; e sem usar o ONI, tendo os subconjuntos Global, La Niña, El Niño e Neutro; esses três últimos compreendem os períodos onde aconteceram ou não os eventos meteorológicos, respectivamente.

4.1 Resultados obtidos com o banco de dados completo

Conforme se apresenta na Tabela 2 (os valores de RMSE e MAE estão expressos em W/m^2), o modelo XGBoost teve menor valor de RMSE para o horizonte de 2 min, no entanto para os outros horizontes, o modelo Boosting apresenta melhor desempenho na previsão, mas apenas com uma leve diferença sobre o XGBoost. Os menores valores de RMSE foram ressaltados em amarelo.

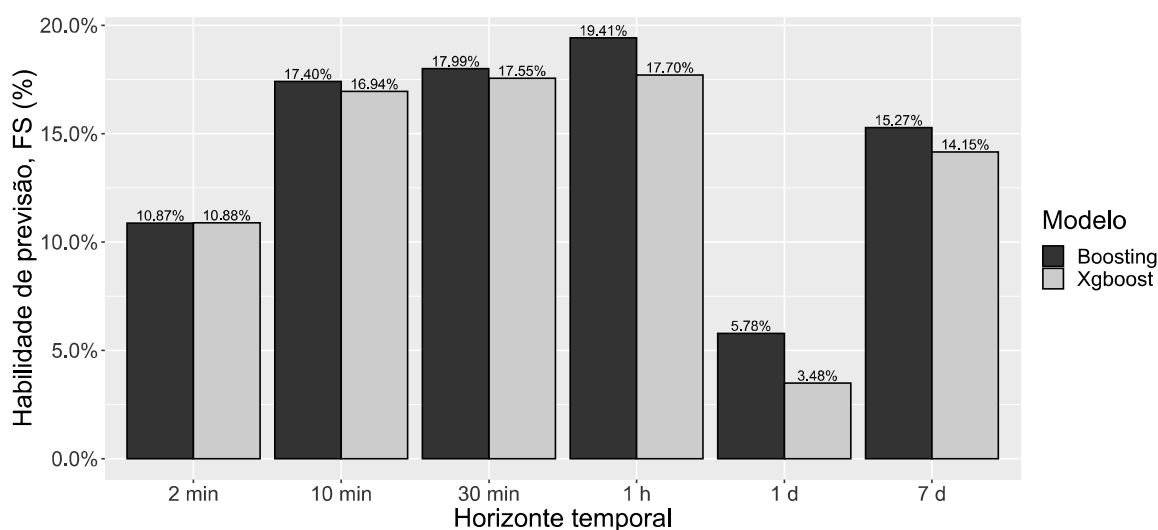
No Gráfico 6 é visível a tendência positiva dos valores de FS tanto para o modelo Boosting quanto para o modelo XGBoost ao longo dos horizontes temporais, com a exceção do horizonte 1 dia em que teve uma diminuição de ~14 pontos percentuais nos dois modelos, após essa queda teve crescimento de ~10 pontos percentuais no horizonte de 7 dias. Para o modelo Boosting os valores de FS variam de 10,87% - 2 min a 15,27% - 7 dias, enquanto para o modelo XGBoost os valores de FS vão de 10,88% - 2 min a 14,00% - 7 dias.

Tabela 2 – Resultados dos modelos de previsão, conjunto Global com o preditor ONI.

Modelo	Métrica de erro	Horizontes temporais					
		2 min	10 min	30 min	1 h	1 dia	7 dias
Persistência	RMSE [W/m ²]	134,2216	185,9741	211,0821	224,4482	353,5532	386,4691
	n.RMSE	23,38%	32,32%	36,62%	39,72%	44,29%	49,63%
	MAE [W/m ²]	68,7020	106,8340	136,0774	163,2498	273,1879	297,1272
	n.MAE	11,97%	18,57%	23,61%	28,89%	34,22%	38,16%
Boosting	RMSE [W/m ²]	119,6337	153,6224	173,0984	180,8886	333,1274	327,4505
	n.RMSE	20,84%	26,70%	30,03%	32,01%	41,73%	42,05%
	MAE [W/m ²]	67,7408	96,1347	115,8064	125,3577	279,5749	265,4295
	n.MAE	11,80%	16,71%	20,09%	22,18%	35,02%	34,09%
	FS	10,87%	17,40%	17,99%	19,41%	5,78%	15,27%
XGBoost	RMSE [W/m ²]	119,6225	154,4736	174,0291	184,7183	341,2422	331,7713
	n.RMSE	11,81%	26,85%	30,19%	32,69%	42,75%	42,61%
	MAE [W/m ²]	67,7756	96,3710	115,4370	127,5722	277,0900	267,9583
	n.MAE	11,81%	16,75%	20,03%	22,58%	34,71%	34,41%
	FS	10,88%	16,94%	17,55%	17,70%	3,48%	14,15%

Fonte: elaborada pelo autor

Gráfico 6 – Variação da Habilidade de previsão, conjunto Global com o preditor ONI.



Fonte: elaborada pelo autor

Nos resultados mostrados na Tabela 3 o modelo com menores valores de RMSE na previsão foi o Boosting para os primeiros cinco horizontes temporais, enquanto o modelo XGBoost teve menor valor de RMSE no horizonte de 7 dias. A ausência do preditor ONI evidencia aumento relativo no valor do RMSE para os primeiros dois horizontes em comparação com os valores obtidos na Tabela 2. Destacam-se em amarelo os valores do RMSE do modelo com melhor desempenho em cada horizonte.

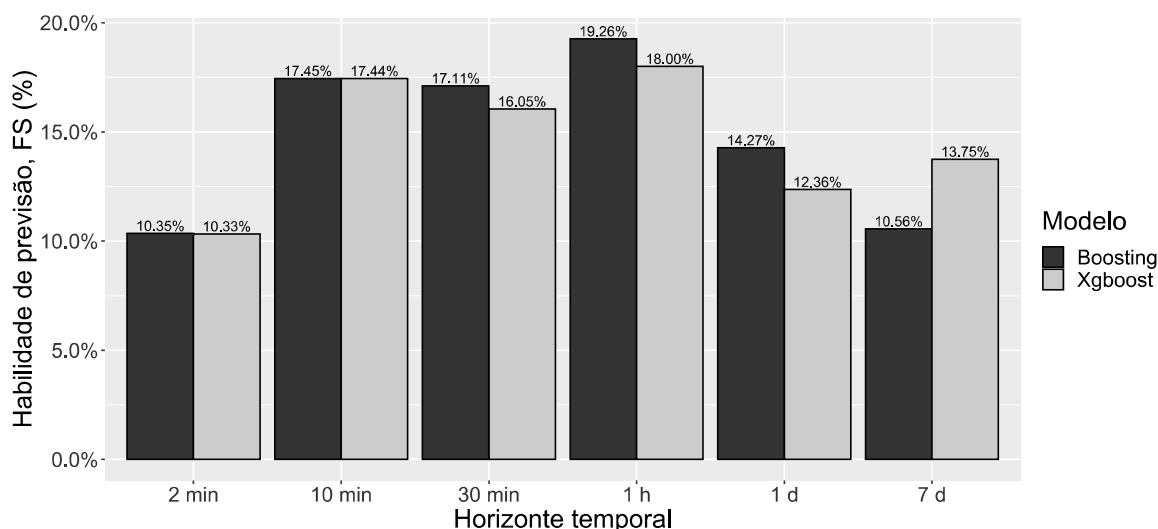
Tabela 3 – Resultados dos modelos de previsão, conjunto Global sem o preditor ONI

Modelo	Métrica de erro	Horizontes temporais					
		2 min	10 min	30 min	1 h	1 dia	7 dias
Persistência	RMSE [W/m^2]	133,9211	187,1050	200,8503	220,4350	386,3919	352,3042
	n.RMSE	23,30%	32,67%	35,18%	39,30%	48,21%	44,86%
	MAE [W/m^2]	69,0152	106,6954	128,5657	161,6070	303,7206	268,1535
	n.MAE	12,01%	18,63%	22,52%	28,81%	37,90%	34,14%
Boosting	RMSE [W/m^2]	120,0591	154,4526	166,4855	177,9720	331,2364	315,0967
	n.RMSE	20,89%	26,97%	29,16%	31,73%	41,33%	40,12%
	MAE [W/m^2]	68,0428	95,9546	108,1667	127,1401	280,2253	256,0396
	n.MAE	11,84%	16,75%	18,95%	22,66%	34,96%	32,60%
	FS	10,35%	17,45%	17,11%	19,26%	14,27%	10,56%
XGBoost	RMSE [W/m^2]	120,0907	154,4739	168,6203	180,7548	338,6287	303,8462
	n.RMSE	20,89%	26,97%	29,54%	32,22%	42,25%	38,69%
	MAE [W/m^2]	68,1032	96,1367	109,8204	127,3951	282,8027	249,1954
	n.MAE	11,85%	16,78%	19,24%	22,71%	35,29%	31,73%
	FS	10,33%	17,44%	16,05%	18,00%	12,36%	13,75%

Fonte: elaborada pelo autor

Os valores de FS mostrados no Gráfico 7 têm um comportamento diferente na ausência do preditor ONI em comparação com o Gráfico 6, apresentando a mesma tendência positiva nos modelos ao longo dos horizontes temporais, tendo valores similares nos primeiros horizontes, crescendo os valores de FS até o horizonte de 1 h onde a diferença é de 1,26 pontos percentuais. Nos horizontes 1 dia e 7 dias se observa que o FS diminuiu em ~4 pontos percentuais em cada horizonte para o modelo Boosting, enquanto o modelo XGBoost diminuiu 5,64 – 1 dia e aumentou em 1,29 – 7 dias.

Gráfico 7 – Variação da Habilidade de previsão, conjunto Global sem o preditor ONI



Fonte: elaborado pelo autor

4.2 Resultados obtidos com o banco La Niña

Os resultados obtidos para o banco que contém as observações para os períodos em que aconteceram os eventos do fenômeno La Niña se apresentam na Tabela 4. Os valores do RMSE são maiores em comparação com o banco Global com e sem ONI. Como explicado por Correia Filho *et al.* (2019), devido ao aumento da precipitação durante o período de La Niña na região Nordeste do Brasil, ocorrem flutuações das condições de claridade do céu, aumentando a variação da radiação solar. Destacam-se em amarelo os valores do RMSE do modelo com melhor desempenho em cada horizonte.

Tabela 4 – Resultados dos modelos de previsão, conjunto La Niña sem o preditor ONI

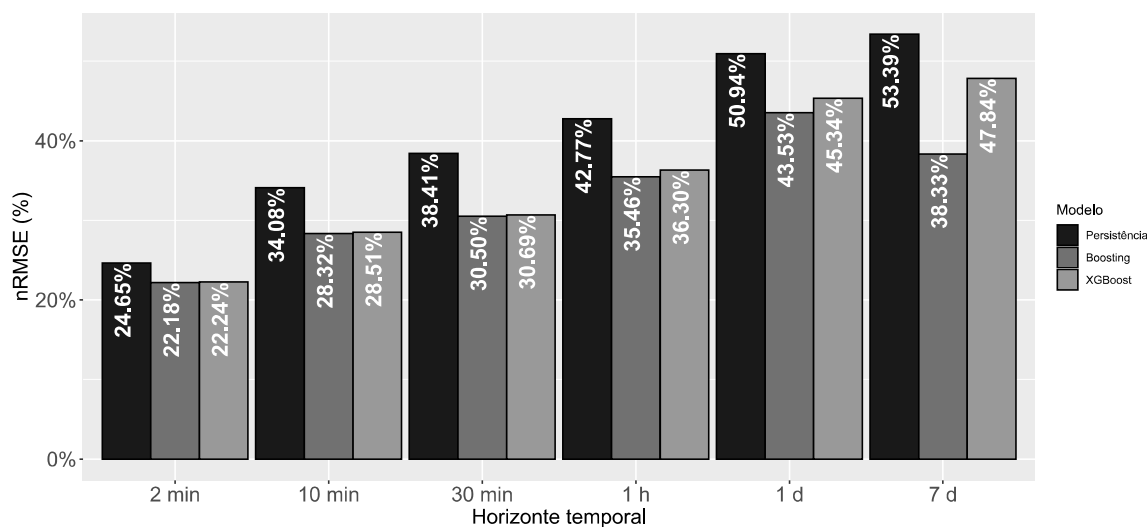
Modelo	Métrica de erro	Horizontes temporais					
		2 min	10 min	30 min	1 h	1 dia	7 dias
Persistência	RMSE [W/m ²]	140,4663	194,0528	222,0314	240,4545	381,8376	397,0776
	n.RMSE	24,65%	34,08%	38,41%	42,77%	50,94%	53,39%
	MAE [W/m ²]	73,7740	113,5658	143,0904	172,4086	295,1925	308,7603
	n.MAE	12,95%	19,95%	24,76%	30,67%	39,38%	41,52%
Boosting	RMSE [W/m ²]	126,4222	161,2441	176,2878	199,3399	326,2993	285,0390
	n.RMSE	22,18%	28,32%	30,50%	35,46%	43,53%	38,33%
	MAE [W/m ²]	73,0978	104,5770	122,1827	144,0712	280,1459	240,2169
	n.MAE	12,83%	18,37%	21,14%	25,63%	37,37%	32,30%
	FS	10,00%	16,91%	20,60%	17,10%	14,55%	28,22%
XGBoost	RMSE [W/m ²]	126,7353	162,3062	177,3827	204,0840	339,8767	355,7857
	n.RMSE	22,24%	28,51%	30,69%	36,30%	45,34%	47,84%
	MAE [W/m ²]	73,9566	104,5844	121,8408	146,0719	289,2923	281,5568
	n.MAE	12,98%	18,37%	21,08%	25,98%	38,59%	37,86%
	FS	9,78%	16,36%	20,11%	15,13%	10,99%	10,40%

Fonte: elaborada pelo autor

A diferença entre os valores do nRMSE dos modelos Boosting e XGBoost é pequena para os primeiros três horizontes, varia de 0,06 - 2 min a 0,19 - 30 min, mas ao longo dos horizontes temporais essa diferença é algo mais notável chegando a 9,51 - 7 dias. No Gráfico 8 se observam em detalhe essas diferenças.

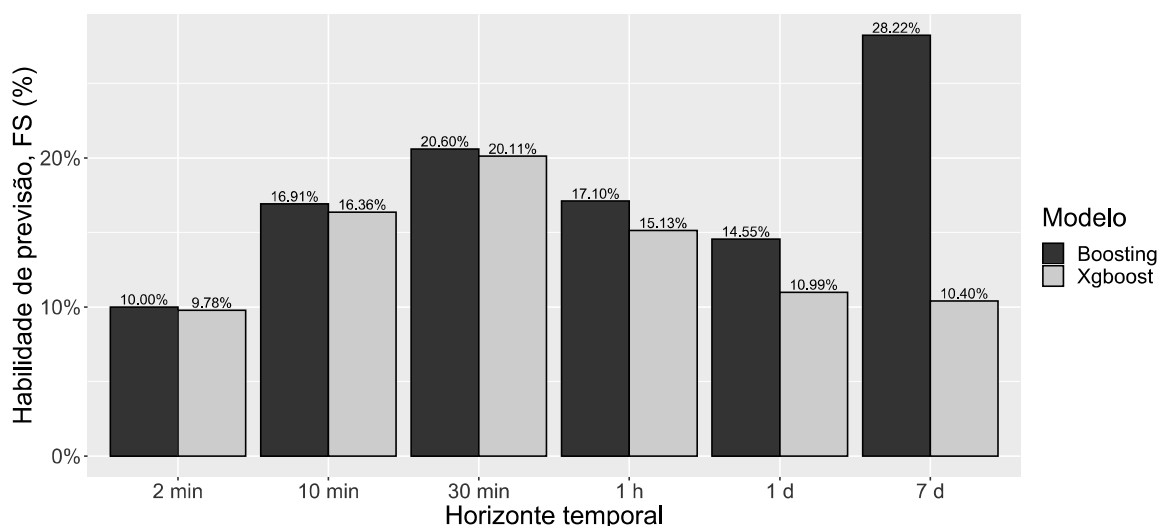
Conforme se observa no Gráfico 9, o valor da FS tanto do modelo Boosting como do modelo XGBoost aumenta nos primeiros três horizontes temporais, diminuindo nos horizontes de 1 h e 1 dia, mas permanecendo positivo. O modelo Boosting apresenta melhor desempenho, com valores de 10,00% - 2 min; 16,91% - 10 min; 20,60% - 30 min; 17,10% - 1 h; 14,55% - 1 dia e o notável aumento até 28,22% - 7 dias.

Gráfico 8 – Variação do nRMSE, conjunto La Niña sem o predictor ONI



Fonte: elaborado pelo autor

Gráfico 9 – Variação da Habilidade de previsão, conjunto La Niña sem o predictor ONI



Fonte: elaborado pelo autor

4.3 Resultados obtidos com o banco El Niño

Os resultados obtidos para o banco com as observações onde aconteceram os eventos do fenômeno El Niño se apresentam na Tabela 5. Os valores do RMSE são menores do que os obtidos com os conjuntos Global e La Niña. Pezzi e Cavalcanti (2001) concluíram que, diferente de La Niña, o efeito do El Niño no padrão das precipitações na região Nordeste do Brasil era negativo, causando a diminuição considerável das chuvas, o que se confere na redução da variação da radiação solar. Destacam-se em amarelo os valores do RMSE do modelo com melhor desempenho em cada horizonte.

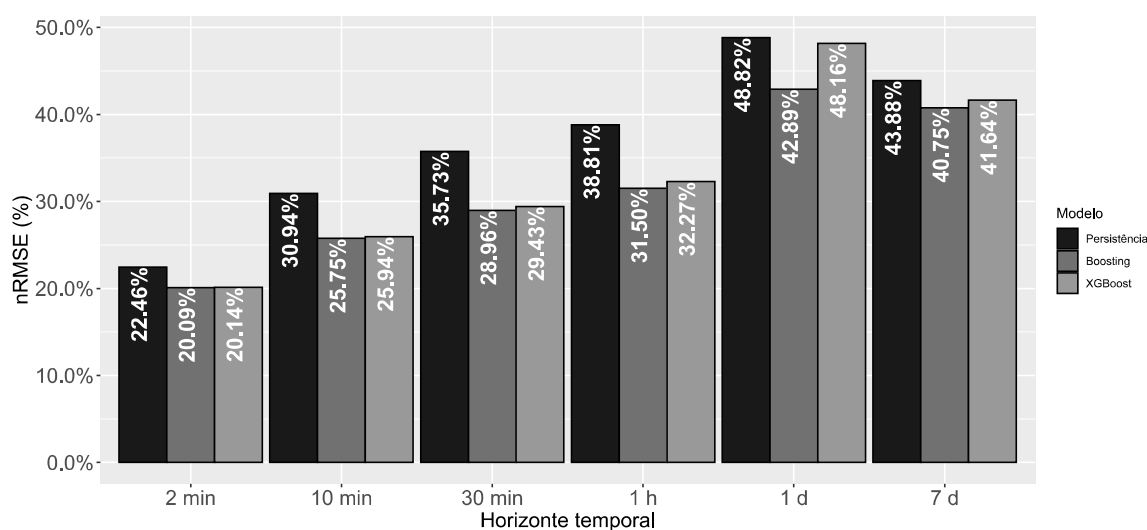
Tabela 5 – Resultados dos modelos de previsão, conjunto El Niño sem o preditor ONI

Modelo	Métrica de erro	Horizontes temporais					
		2 min	10 min	30 min	1 h	1 dia	7 dias
Persistência	RMSE [W/m ²]	129,4075	177,0718	207,6476	222,1358	393,6539	344,7428
	n.RMSE	22,46%	30,94%	35,73%	38,81%	48,82%	43,88%
	MAE [W/m ²]	66,0990	100,9111	135,5069	160,3025	313,2349	265,1908
	n.MAE	11,47%	17,63%	23,32%	28,01%	38,85%	33,75%
Boosting	RMSE [W/m ²]	115,7568	147,3882	168,3222	180,3147	345,8037	320,2183
	n.RMSE	20,09%	25,75%	28,96%	31,50%	42,89%	40,75%
	MAE [W/m ²]	65,0272	90,4298	110,5106	123,8539	287,5488	265,0379
	n.MAE	11,29%	15,80%	19,02%	21,64%	35,66%	33,73%
	FS	10,55%	16,76%	18,94%	18,83%	12,16%	7,11%
XGBoost	RMSE [W/m ²]	116,0511	148,4394	171,0046	184,7281	388,3168	327,1749
	n.RMSE	20,14%	25,94%	29,43%	32,27%	48,16%	41,64%
	MAE [W/m ²]	65,4053	92,0057	110,7271	123,9492	318,9619	258,0529
	n.MAE	11,35%	16,08%	19,05%	21,65%	39,56%	32,84%
	FS	10,32%	16,17%	17,65%	16,84%	1,36%	5,10%

Fonte: elaborada pelo autor

No Gráfico 10 é notável que a tendência do nRMSE dos modelos é positiva até o horizonte de 1 dia, em que atinge a maior diferença de até ~11 pontos percentuais em relação aos valores do horizonte prévio. Os modelos Boosting e XGBoost apresentam desempenhos similares com uma leve variação dos valores, de 0,05 - 2 min, passando por 0,19 - 10 min, 0,47 - 30 min até 0,77 - 1 h e 0,89 – 7 dias, mas no horizonte de 1 dia essa variação atinge 5,27 pontos percentuais.

Gráfico 10 – Variação do nRMSE, conjunto El Niño sem o preditor ONI

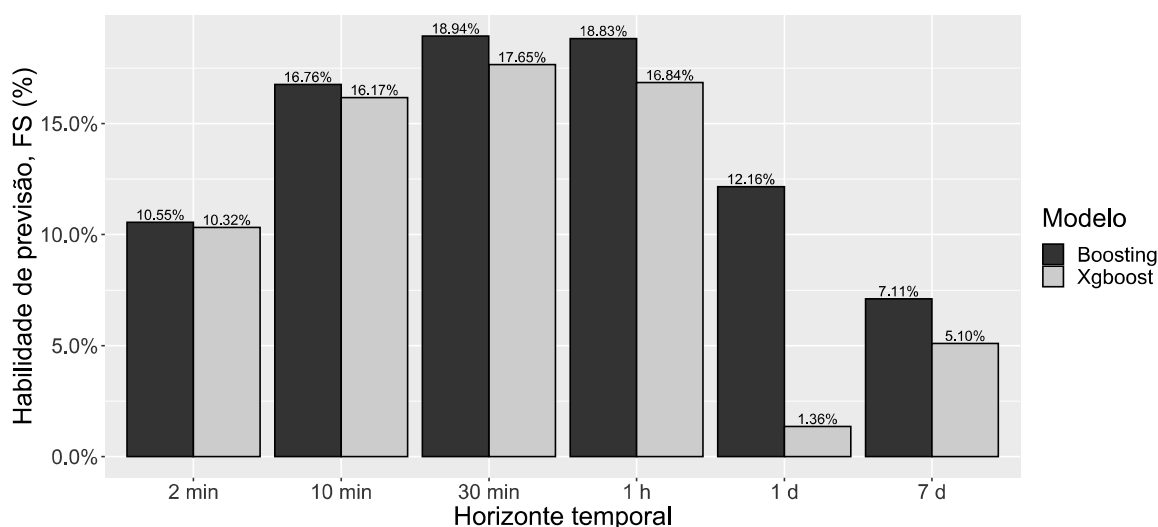


Fonte: elaborado pelo autor

Conforme se observa no Gráfico 11, os valores da FS dos modelos Boosting e XGBoost apresentam até o horizonte de 30 min uma tendência positiva mais uniforme do que

os resultados obtidos com o conjunto La Niña. Para os horizontes 1 h, 1 dia e 7 dias os valores da FS diminuem mas se mantendo positivos, com o destaque da queda considerável para o horizonte 1 dia em que atinge o valor mais baixo da FS (1,36%) do modelo XGBoost.

Gráfico 11 – Variação da Habilidade de previsão, conjunto El Niño sem o preditor ONI



Fonte: elaborado pelo autor

4.4 Resultados obtidos com o banco Neutro

Os resultados obtidos para o conjunto de observações dos períodos em que teve ausência dos eventos La Niña e El Niño se apresentam na Tabela 6. Os valores do RMSE dos três modelos são menores nos horizontes 2 min, 10 min e 30 min, em comparação com os conjuntos Global, La Niña e El Niño. O número de observações resultantes para o horizonte de 7 dias do banco de dados Neutro foi menor do que o número de variáveis usadas para o treinamento e teste dos modelos, razão pela que não foi possível a execução dos modelos para esse horizonte temporal. Destacam-se em amarelo os valores do RMSE do modelo com melhor desempenho em cada horizonte.

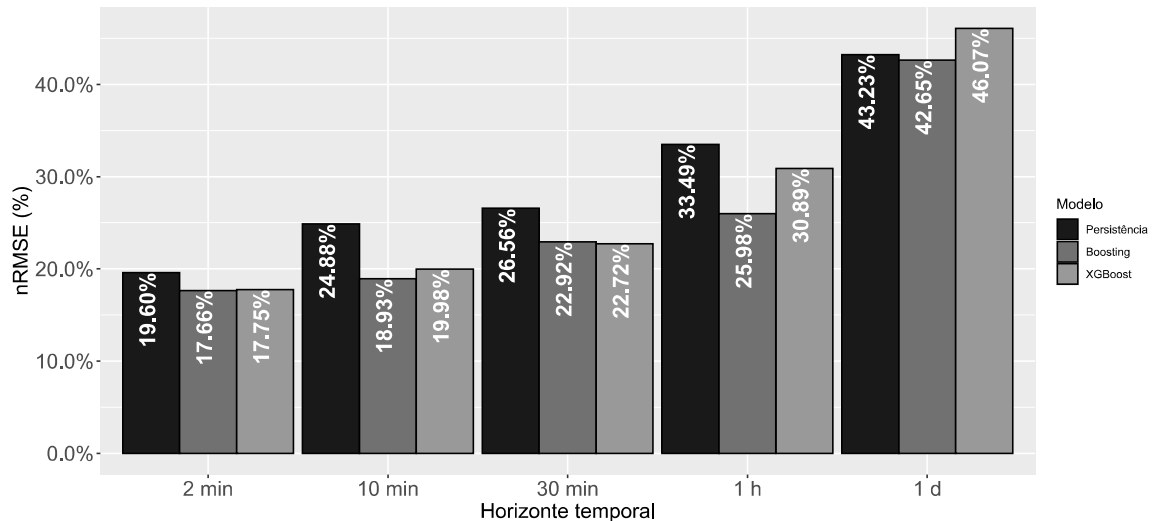
Observa-se no Gráfico 12 que a tendência do nRMSE dos modelos apresenta um comportamento diferente dos conjuntos analisados previamente. Em termos gerais nota-se que existe crescimento dos valores para cada modelo ao longo dos horizontes temporais. Os valores de nRMSE para os modelos Boosting e XGBoost passam desde ter uma diferença quase nula de (0,09 - 2 min, 0,2 - 30 min), até uma diferença considerável (1,05 - 10 min, 4,91 - 1 h, 3,42 - 7 dias).

Tabela 6 – Resultados dos modelos de previsão, conjunto Neutro sem o preditor ONI. A coluna referente a 7 dias não encontra-se preenchida devido ao fato de o número de observações para este horizonte temporal ter sido menor do que o número de variáveis (preditores) usadas para o treinamento e teste dos modelos.

Modelo	Métrica de erro	Horizontes temporais					
		2 min	10 min	30 min	1 h	1 dia	7 dias
Persistência	RMSE [W/m ²]	119,7209	147,9394	154,7095	203,4967	372,3077	**
	n.RMSE	19,60%	24,88%	26,56%	33,49%	43,23%	**
	MAE [W/m ²]	52,6635	77,0976	97,5666	146,9714	280,6358	**
	n.MAE	8,62%	12,97%	16,75%	24,19%	32,59%	**
Boosting	RMSE [W/m ²]	107,8391	112,5787	133,4684	157,8619	367,2751	**
	n.RMSE	17,66%	18,93%	22,92%	25,98%	42,65%	**
	MAE [W/m ²]	53,2743	66,6829	80,9684	103,2870	326,3184	**
	n.MAE	8,72%	11,21%	13,90%	17,00%	37,89%	**
	FS	9,92%	23,90%	13,73%	22,43%	1,35%	**
XGBoost	RMSE [W/m ²]	108,4063	118,8020	132,3360	187,6928	396,9637	**
	n.RMSE	17,75%	19,98%	22,72%	30,89%	46,07%	**
	MAE [W/m ²]	54,6995	68,3933	80,5922	110,4703	313,6374	**
	n.MAE	8,96%	11,50%	13,84%	18,18%	36,42%	**
	FS	9,45%	19,70%	14,46%	7,77%	-6,55%	**

Fonte: elaborada pelo autor

Gráfico 12 – Variação do nRMSE, conjunto Neutro sem o preditor ONI

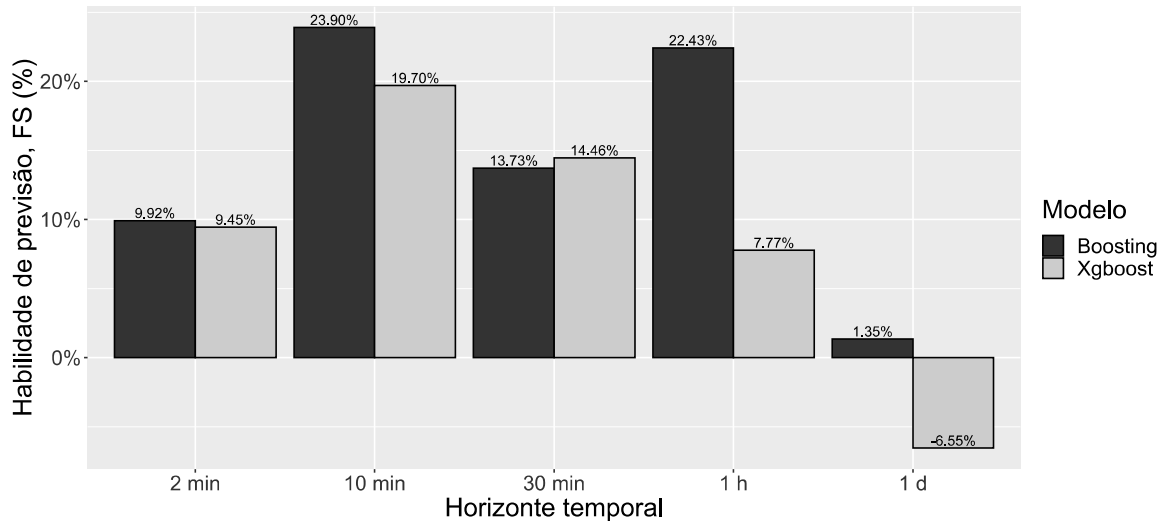


Fonte: elaborado pelo autor

O desempenho dos modelos Boosting e XGBoost com o conjunto Neutro é variável e não apresenta uma tendência uniforme dos valores da FS, conforme se observa no Gráfico 13. Tanto no modelo Boosting como no modelo XGBoost a FS aumenta do horizonte de 2 min até 10 min, mas depois os dois caem subitamente. No caso do XGBoost ele continua caindo até o horizonte de 1 dia atingindo valor negativo de -6,55%, enquanto o Boosting se recupera depois do horizonte de 30 min até 1 h e depois cai até o menor valor obtido de 1,35%. O modelo

Boosting teve melhor desempenho de forma geral, apresentando os valores de 9,92% - 2 min, 23,90% - 10 min, 22,43% - 1 h e 1,35% - 1 dia.

Gráfico 13 – Variação da Habilidade de previsão, conjunto Neutro sem o preditor ONI



Fonte: elaborado pelo autor

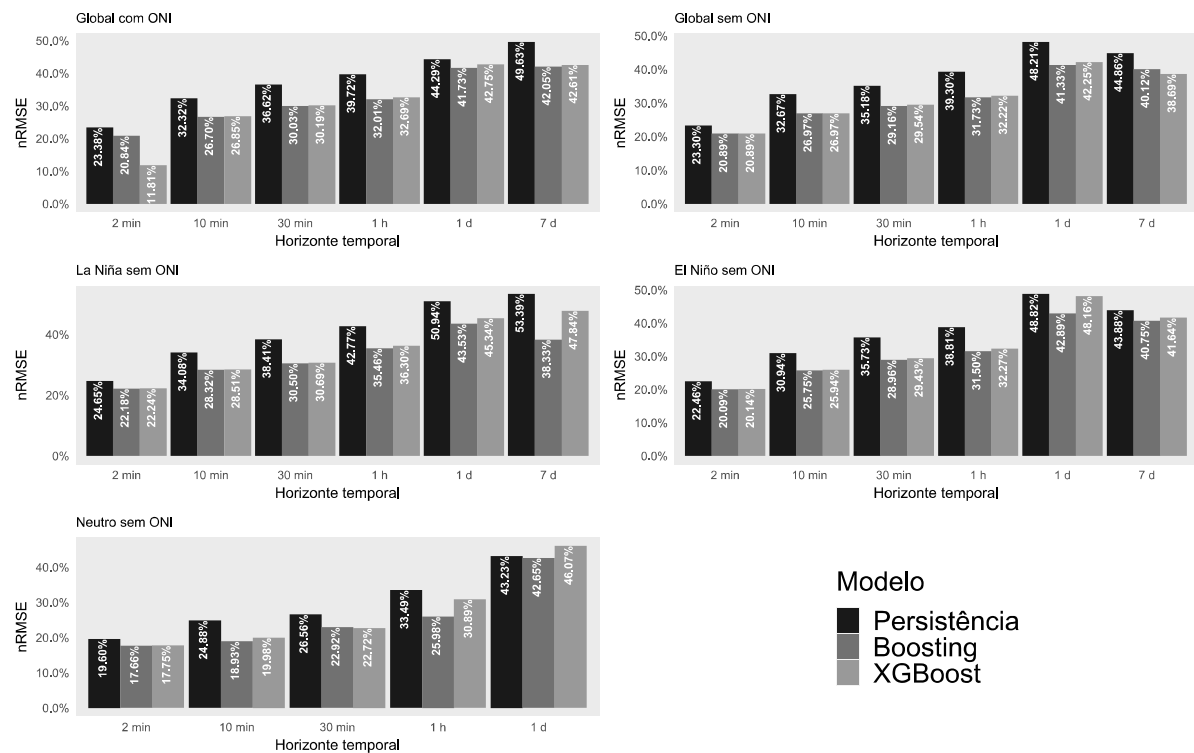
O Gráfico 14 apresenta de forma conjunta a comparação dos valores de nRMSE obtidos com cada banco de dados. Se destaca que o banco de dados Neutro sem ONI apresentou menores valores de nRMSE para os três modelos nos horizontes 10 min, 30 min, 1h, enquanto o modelo XGBoost no banco Global com ONI mostrou um valor 52.08% menor do que nos outros bancos no horizonte de 2min. O modelo Boosting apresentou no banco Neutro sem ONI um valor 39.26% menor do que nos outros bancos para o horizonte de 1 h. Conforme o estudo de Wang *et al.* (2012), se confere o comportamento da previsão, apresentando valores de nRMSE menores para os dias ensolarados como no caso do banco de dados El Niño, enquanto para os dias nublados os valores de nRMSE são maiores como no banco de dados La Niña. Analisando a ocorrência do ENOS, os anos Neutros, anos em que não aconteceu nem La Niña nem El Niño, são poucos ou até raros em termos da frequência, assim, mesmo os resultados de erro sendo baixos não se mostram significativos para se tirar conclusões, isso acontece devido à pouca disponibilidade de dados para o treinamento e teste dos modelos com esse banco de dados.

Os modelos Boosting e XGBoost mostram um desempenho similar nos bancos Global, La Niña e El Niño ao longo dos horizontes temporais, onde os valores de nRMSE apresentam diferença de menos de 1% com a exceção dos casos pontuais especificados no parágrafo anterior. Esse comportamento dos dois modelos está de acordo com os resultados obtidos por Torres-Barrán *et al.* (2019) onde os modelos mencionados obtiveram os melhores

resultados. O desempenho do modelo XGBoost é o esperado segundo Chen e Guestrin (2016), sendo um modelo de aprendizagem de máquinas que pode se desenvolver de maneira igual ou melhor do que os modelos mais comuns usados na previsão de irradiação solar, conseguindo processar bancos de dados de centenas de milhares de observações com menos dificuldade computacional do que os modelos de aprendizagem padrões. Esse desempenho também foi comprovado no estudo desenvolvido por Fan *et al.* (2018).

Existe uma tendência positiva da FS nos resultados obtidos com os bancos Global, La Niña e El Niño, lembrando que esses bancos contêm observações onde teve ocorrência do ENOS. Diante disto, e como sugerido por Brasil *et al.* (2020b), existe influência do ONI na sensibilidade de previsão de irradiação solar.

Gráfico 14 – nRMSE da previsão realizada com cada banco de dados



Fonte: elaborado pelo autor

5 CONCLUSÃO

Na presente dissertação realizou-se a análise da previsão de irradiação solar em Fortaleza para os horizontes 2 min, 10 min, 30 min, 1 h, 1 dia e 7 dias comparando o desempenho dos modelos XGBoost e Boosting. Por razões do desempenho computacional, e mantendo a confiabilidade da análise desenvolvida, usou-se apenas 50% do total das observações para a resolução temporal de 2 minutos. Para as demais resoluções estudadas essa redução não se mostrou necessária.

Observando os resultados obtidos com os modelos de previsão desenvolvidos, pode-se conferir a influência positiva do ENOS tanto de forma direta como preditor ONI, como de forma indireta nos bancos que contêm observações onde aconteceram os fenômenos meteorológicos, aportando mais sensibilidade aos modelos de aprendizagem e melhorando a previsão de irradiação solar, atingindo valores de FS até 28,22% para o horizonte de 7 dias no banco de dados La Niña, em comparação com os 23,90% para o horizonte de 10 min no banco de dados Neutro, em ambos casos com o modelo Boosting.

O Boosting foi o modelo que apresentou melhores resultados de previsão com um valor de FS máximo de 28,22% para o horizonte de 7 dias no banco de dados La Niña, destacando que todos os valores de FS obtidos foram positivos tendo como valor mínimo 1,35% para o horizonte de 1 dia no banco de dados Neutro. Esse método de aprendizagem de máquinas tem a limitante de ser um modelo que ocupa grande quantidade de recursos computacionais, precisando que seja realizado um pré-processamento do banco de dados para diminuir a quantidade de observações o suficiente sem afetar os resultados finais da previsão.

O segundo melhor modelo foi o XGBoost, que mostrou um desempenho similar ao Boosting, com valores de FS similares na maioria dos casos mas também apresentando valores maiores que o Boosting em três casos específicos: 0,01 – 2 min conjunto Global com ONI; 3,19 – 7 dias conjunto Global sem ONI e 0,73 – 30 min conjunto Neutro (diferença em pontos percentuais), atingindo o máximo de 20,11% para o horizonte de 30 min no banco de dados La Niña e o valor mínimo obtido foi -6,55% para o horizonte de 1 dia no banco de dados Neutro. Após ser realizada a análise dos hiperparâmetros adequados para o processamento das observações, o XGBoost apresentou maior velocidade de execução com cada banco, demorando menos da metade do tempo que o Boosting no processamento das observações.

REFERÊNCIAS

- ANDREOLI, Rita V.; KAYANO, Mary T. ENSO-related rainfall anomalies in South America and associated circulation features during warm and cold Pacific decadal oscillation regimes. **International Journal of Climatology**, v. 25, n. 15, p. 2017–2030, 2005. DOI: 10.1002/joc.1222.
- ANTONANZAS, J.; OSORIO, N.; ESCOBAR, R.; URRACA, R.; MARTINEZ-DE-PISON, F. J.; ANTONANZAS-TORRES, F. Review of photovoltaic power forecasting. **Solar Energy**, v. 136, p. 78–111, 2016. DOI: 10.1016/j.solener.2016.06.069.
- BAKKER, Kilian; WHAN, Kirien; KNAP, Wouter; SCHMEITS, Maurice. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. **Solar Energy**, v. 191, n. April, p. 138–150, 2019. DOI: 10.1016/j.solener.2019.08.044.
- BRASIL, Juliana Silva. **Influência do El Niño Oscilação Sul (ENOS) na previsão da irradiação global horizontal de Fortaleza**. 2020a. 49 f. Dissertação (Mestrado em Engenharia Mecânica) - Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2020 Disponível em: <http://www.repositoriobib.ufc.br/000066/000066a4.pdf>. Acesso em: 27 maio 2021.
- BRASIL, Juliana Silva; MARINHO, Felipe Pinto; ROCHA, Paulo Alexandre Costa. Influência do el niño e da la niña na previsão intra-horária da irradiação solar global horizontal. **Brazilian Journal of Development**, v. 6, n. 1, p. 2321–2329, 2020b. DOI: 10.34117/bjdv6n1-169.
- CERVONE, Guido; CLEMENTE-HARDING, Laura; ALESSANDRINI, Stefano; DELLE MONACHE, Luca. Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. **Renewable Energy**, v. 108, p. 274–286, 2017. DOI: 10.1016/j.renene.2017.02.052.
- CHANDOLA, Deeksha; GUPTA, Harsh; TIKKIWAL, Vinay Anand; BOHRA, Manoj Kumar. Multi-step ahead forecasting of global solar radiation for arid zones using deep learning. **Procedia Computer Science**, v. 167, n. Iccids 2019, p. 626–635, 2020. DOI: 10.1016/j.procs.2020.03.329.
- CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). **Association for Computing Machinery**, p. 785–794, 2016.
- CORREIA FILHO, Washington Luiz Félix; DE OLIVEIRA-JÚNIOR, José Francisco; DE BARROS SANTIAGO, Dimas; DE BODAS TERASSI, Paulo Miguel; TEODORO, Paulo Eduardo; DE GOIS, Givanildo; BLANCO, Claudio José Cavalcante; DE ALMEIDA SOUZA, Pedro Henrique; DA SILVA COSTA, Micejane; GOMES, Heliofábio Barros; DOS SANTOS, Paulo José. Rainfall variability in the Brazilian northeast biomes and their interactions with meteorological systems and ENSO via CHELSA product. **Big Earth Data**, v. 3, n. 4, p. 315–337, 2019. DOI: 10.1080/20964471.2019.1692298.
- DAHLMAN, LuAnn; Climate Variability: Oceanic Niño Index; Disponível em: <https://www.climate.gov/news-features/understanding-climate/climate-variability-oceanic-ni%C3%B1o-index>. Acesso em: 18 ago. 2021

DAVY, Robert J.; TROCCOLI, Alberto. Interannual variability of solar energy generation in Australia. **Solar Energy**, v. 86, n. 12, p. 3554–3560, 2012. DOI: 10.1016/j.solener.2011.12.004.

DIAGNE, Maimouna; DAVID, Mathieu; LAURET, Philippe; BOLAND, John; SCHMUTZ, Nicolas. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. **Renewable and Sustainable Energy Reviews**, v. 27, p. 65–76, 2013. DOI 10.1016/j.rser.2013.06.042.

DONG, Wei; HUANG, Yimiao; LEHANE, Barry; MA, Guowei. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. **Automation in Construction**, v. 114, n. March, p. 103155, 2020. DOI: 10.1016/j.autcon.2020.103155.

DUFFIE, John A.; BECKMAN, William A. **Solar engineering of thermal processes**. 3rd ed. New Jersey: J. Wiley, c2006. xix, 908 p. ISBN 9780471698678 (enc.).

FAN, Junliang; WANG, Xiukang; WU, Lifeng; ZHOU, Hanmi; ZHANG, Fucang; YU, Xiang; LU, Xianghui; XIANG, Youzhen. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. **Energy Conversion and Management**, v. 164, n. February, p. 102–111, 2018. DOI: 10.1016/j.enconman.2018.02.087.

FAN, Junliang; WANG, Xiukang; ZHANG, Fucang; MA, Xin; WU, Lifeng. Predicting daily diffuse horizontal solar radiation in various climatic regions of China using support vector machine and tree-based soft computing models with local and extrinsic climatic data. **Journal of Cleaner Production**, v. 248, p. 119264, 2020. DOI: 10.1016/j.jclepro.2019.119264.

FENG, Yu; GONG, Daozhi; ZHANG, Qingwen; JIANG, Shouzheng; ZHAO, Lu; CUI, Ningbo. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. **Energy Conversion and Management**, v. 198, n. May, p. 111780, 2019. DOI: 10.1016/j.enconman.2019.111780.

FREITAS, Breno Bezerra; TELES, Clayton Ferreira; LINS, Davi Ribeiro; LOAYZA, David Mickely Jaramillo; RAMALHO, Dayane Cristina Lima; QUEIROZ, Dayse Maria Benevides de; FREITAS, Deivid Matias de; SILVA, Francisco Eduardo Mendes da; SILVA, Francisco Jeandson Rodrigues da; PALACIO, Gilderlanio Barbosa Alves; RIBEIRO, José Cleison Cassiano; UCHÔA, Kênio Monteles; SANTOS, Leticia De Oliveira; RABELO, Magna Livia Neco; BATISTA, Natasha Esteves; CARVALHO, Paulo Cesar Marques de; LIMA, Raoni Alves de; MENEZES, Victor Augusto Cavalcante Bezerra; GUALBERTO, Victor Soares; NOGUEIRA, Tiago De Oliveira. Cenários para a matriz de geração de eletricidade do Ceará em 2050. **Ciência e Natura**, v. 42, p. e63, 26 maio 2020. DOI: 10.5902/2179460x42805.

GUTIERREZ-COREA, Federico Vladimir; MANSO-CALLEJO, Miguel Angel; MORENO-REGIDOR, Maria Pilar; MANRIQUE-SANCHO, Maria Teresa. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. **Solar Energy**, v. 134, p. 119–131, 2016. DOI: 10.1016/j.solener.2016.04.020.

KUMARI, Pratima; TOSHNIWAL, Durga. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. **Journal of Cleaner Production**, v. 279, p. 123285, 2021. DOI: 10.1016/j.jclepro.2020.123285.

LAI, Chun Sing; ZHONG, Cankun; PAN, Keda; NG, Wing W.Y.; LAI, Loi Lei. A deep learning based hybrid method for hourly solar radiation forecasting. **Expert Systems with Applications**, v. 177, n. February, p. 114941, 2021. DOI: 10.1016/j.eswa.2021.114941.

LI, Jiaming; WARD, John K.; TONG, Jingnan; COLLINS, Lyle; PLATT, Glenn. Machine learning for solar irradiance forecasting of photovoltaic system. **Renewable Energy**, v. 90, p. 542–553, 2016. DOI: 10.1016/j.renene.2015.12.069.

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. DOI: 10.1016/j.renene.2017.12.069.

NAGY, Gábor I.; BARTA, Gergő; KAZI, Sándor; BORBÉLY, Gyula; SIMON, Gábor. GEFCom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. **International Journal of Forecasting**, v. 32, n. 3, p. 1087–1093, 2016. DOI: 10.1016/j.ijforecast.2015.11.013.

NIELSEN, Didrik. *Tree boosting with xgboost-why does xgboost win" every" machine learning competition?*. 2016. 109 f. Dissertação (Mestrado). NTNU.

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. DOI 10.1016/j.renene.2015.02.061.

PEZZI, L. P.; CAVALCANTI, I. F. A. The relative importance of ENSO and tropical Atlantic sea surface temperature anomalies for seasonal precipitation over South America: a numerical study. **Climate Dynamics**, v. 17, n. 2–3, p. 205–212, 2001. DOI: 10.1007/s003820000104.

ROCHA, P. A.Costa; FERNANDES, J. L.; MODOLO, A. B.; LIMA, R. J.Pontes; DA SILVA, M. E.Vieira; BEZERRA, C. A.Dias. Estimation of daily, weekly and monthly global solar radiation using ANNs and a long data set: a case study of Fortaleza, in Brazilian Northeast region. **International Journal of Energy and Environmental Engineering**, v. 10, n. 3, p. 319–334, 2019. DOI: 10.1007/s40095-019-0313-0.

TORRES-BARRÁN, Alberto; ALONSO, Álvaro; DORRONSORO, José R. Regression tree ensembles for wind energy and solar radiation prediction. **Neurocomputing**, v. 326–327, p. 151–160, 2019. DOI: 10.1016/j.neucom.2017.05.104.

TRENBERTH, K. E. (2019). El Niño southern oscillation (ENSO). **Encyclopedia of Ocean Sciences** (pp. 420–432). Elsevier. DOI: 10.1016/B978-0-12-409548-9.04082-3

WANG, Fei; MI, Zengqiang; SU, Shi; ZHAO, Hongshan. Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. **Energies**, v. 5, n. 5, p. 1355–1370, 2012. DOI: 10.3390/en5051355.

ZAMBRANO, Andres Felipe; GIRALDO, Luis Felipe. Solar irradiance forecasting models without on-site training measurements. **Renewable Energy**, v. 152, p. 557–566, 2020. DOI: 10.1016/j.renene.2020.01.092.