

APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA PROJEÇÃO DO PREÇO HORÁRIO DE LIQUIDAÇÃO DAS DIFERENÇAS COMO SUPORTE ÀS ESTRATÉGIAS DE COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA

Cosme Rodolfo Roque dos Santos¹
Luiz Carlos Pereira da Silva¹
Roberto Castro²
Rafael Francisco Marques³

¹Universidade Estadual de Campinas

²Universidade de São Paulo

³Associação Brasileira de Energia Solar Fotovoltaica

DOI: 10.47168/rbe.v28i1.671

RESUMO

Estimar o preço de um ativo financeiro ou de produtos comercializáveis é uma tarefa complexa e que pressupõe a disponibilidade de uma quantidade razoável de amostras de dados. No ambiente de comercialização de energia elétrica, a projeção do preço da energia no mercado de curto prazo em base horária é essencial para a tomada de decisão dentro desse complexo mercado e as particularidades do setor elétrico brasileiro tornam essa tarefa ainda mais complexa do que em outros mercados. Devido ao comportamento estocástico de algumas variáveis, como a vazão afluente às usinas hidrelétricas e a correlação entre as variáveis que afetam a geração de eletricidade, as técnicas estatísticas tradicionais de previsão de séries temporais apresentam uma complexidade adicional, quando se procura prever diferentes horizontes de análise. Para enfrentar essas complexidades dos métodos tradicionais de previsão, neste estudo apresenta-se uma nova abordagem baseada em metodologia de *Machine Learning* para previsão de séries temporais aplicadas no processo de previsão do PLD. A Base de Aprendizado do modelo é obtida a partir de informações públicas dos modelos oficiais do setor: NEWAVE, DECOMP e DESSEM. A aplicação da metodologia para casos práticos, utilizando-se de *back-test* com informações reais do setor elétrico brasileiro, demonstra que a linha de pesquisa é promissora, à medida que a aderência das projeções aos valores realizados é significativa.

Palavras-chave: Inteligência Artificial; Aprendizagem de Máquina; Preço da Liquidação das Diferenças; PLD; Estimativa de Preço; Planejamento Energético; Árvore de Decisão.

ABSTRACT

Estimating the price of a financial asset or tradable product is a complex task that requires the availability of a reasonable amount of data samples. In the electricity trading environment, projecting the price of energy in the short-term market on an hourly basis is essential for decision making within this complex market, and the particularities of the Brazilian electricity sector make this task even more complex than in other markets. Due to the stochastic behavior of some variables, such as the inflow to hydroelectric plants and the correlation between the variables that affect electricity generation, the traditional statistical techniques of time series forecast presents an additional complexity, when trying to predict different analysis horizons. To face these complexities of traditional forecasting methods, this study presents a new approach based on Machine Learning methodology for time series forecasting applied in the PLD forecasting process. The Learning Base of the model is obtained from public information provided by the official models of the sector: NEWAVE, DECOMP, and DESSEM. The application of the methodology to practical cases, using a back-test with real information from the Brazilian electricity sector, demonstrates that the line of research is promising, as the adherence of the projections to the realized values is significant.

Keywords: Artificial Intelligence; Machine Learning; Settlement of Differences Price; PLD; Price Estimation; Energy Planning; Decision Tree.

1. INTRODUÇÃO

A utilização de técnicas computacionais na previsão de Séries Temporais é um dos campos mais ativos em pesquisas acadêmicas. Com o advento de técnicas de aprendizado de máquina, novos algoritmos estão sendo desenvolvidos e muitas vezes disponibilizados através de bibliotecas para uso em larga escala.

Neste estudo são discutidas algumas características da formação do Preço da Liquidação das Diferenças (PLD), e suas aplicações através de uma nova metodologia de previsão de séries temporais, utilizando técnicas de *Machine Learning*, mais especificamente o algoritmo XGBoost.

Estudos mais recentes utilizam abordagens híbridas de métodos tradicionais, como o ARIMA combinado com Redes Neurais (RESTON FILHO, AFFONSO AND DE OLIVEIRA 2012), contudo a utilização do XGBoost é uma abordagem relativamente nova para o setor elétrico. A utilização deste algoritmo na previsão de séries temporais tem sido fonte de estudo em outros setores da economia, e

o XGBoost apresentou resultados relevantes, como em (ZHANG, et al. 2021) e (ALIM, et al. 2020).

Neste estudo, as unidades temporais da série serão passadas para o modelo como variáveis de entrada, permitindo a representação das sazonalidades e ciclos temporais. Também é inserida no modelo uma variável que representa as características do Sistema Interligado Nacional, oriunda dos modelos oficiais NEWAVE, DECOMP e DESSEM. Diferentes horizontes de previsão serão fornecidos, e em especial para o curtíssimo e curto prazo, serão medidos os desempenhos das previsões (DOS SANTOS, CASTRO AND MARQUES 2021).

2. CARACTERÍSTICAS DA FORMAÇÃO DO PLD

O Sistema Interligado Nacional (SIN) é caracterizado pela interligação, através da Rede Básica de Transmissão de Energia Elétrica, de quatro subsistemas do país, definidos como Nordeste, Norte, Sul e Sudeste, este último em conjunto com o Centro-Oeste. O sistema interligado é operado centralizadamente pelo Operador Nacional do Sistema Elétrico (ONS) e é integrado por diferentes empresas de geração e transmissão de energia elétrica, que podem ser públicas ou privadas.

Dados da ANEEL, divulgados em Janeiro de 2022, indicam que 60,12% da demanda de energia elétrica é atendida a partir de Usinas Hidrelétricas (UHE), 8,95% por Usinas Térmicas (UTE) a gás natural, 5,02% por UTEs a combustíveis líquidos originados do petróleo, e 1,97% de UTEs a carvão mineral. A contabilização de todas estas fontes representa cerca de 76% do atendimento da demanda de energia (ANEEL 2022).

As fontes renováveis, como eólicas, pequenas centrais hidrelétricas e solares, possuem um papel importante na composição da matriz de geração de energia elétrica, porém no processo de otimização da operação, a expectativa da geração renovável dessas usinas é abatida diretamente da Demanda Bruta de energia elétrica esperada, não sendo consideradas explicitamente no processo de otimização da operação centralizada.

Devido à natureza estocástica da Demanda Bruta e da geração renovável, esta operação permite apresentar uma parcela única ao modelo de otimização hidrotérmico, sendo esta parcela conhecida como a Demanda Líquida de energia, adotada como determinística (cenário único) no processo de cálculo do PLD.

Esta característica hidrotérmica do sistema tem impactos diretos no preço da energia elétrica gerada, tendo em vista que o custo para se gerar é uma função do despacho ótimo das fontes hidráulicas

e térmicas, com o objetivo de se minimizar o custo ao longo de um horizonte de operação.

Na precificação do PLD, os modelos computacionais utilizados pelo ONS para operar o SIN otimizam a operação e calculam o Custo Marginal de Operação (CMO). Tais modelos levam em consideração as constantes mudanças na condição operativa do sistema, em especial relacionadas às questões meteorológicas - favoráveis ou não à Energia Natural Afluente (ENA) e à Energia Armazenada (EAR) nos reservatórios das hidrelétricas - como também os Custos Unitários de Geração (CVU) de cada usina térmica.

2.1 Variável relacionada ao sistema hidrotérmico

A cadeia de modelos computacionais, mais especificamente o NEWAVE, DECOMP e DESSEM, é utilizada tanto pelo ONS quanto pela CCEE, contudo para objetivos diferentes. O ONS busca a melhor forma de operar o sistema elétrico, no intuito de garantir o suprimento da demanda pelo menor custo possível. Já a CCEE visa determinar o PLD, por submercado e com precificação horária, que será utilizado na contabilização do mercado de curto prazo (i.e. mercado *spot*) (CEPEL).

Para que haja uma conexão entre os resultados gerados pelas duas instituições, o ONS executa os modelos primeiramente, e em seguida a CCEE utiliza tais resultados e trata as restrições elétricas e unidades geradoras em testes, processa novamente os modelos, e por fim publica os preços (CEPEL).

O acoplamento entre os três modelos se dá através da Função de Custo Futuro (FCF). Esta função representa um custo associado a cada uma das diversas possíveis trajetórias das variáveis de estado e turbinagem de água dos reservatórios das usinas hidráulicas, em conjunto com as respectivas complementações térmicas, para atendimento da Demanda Líquida.

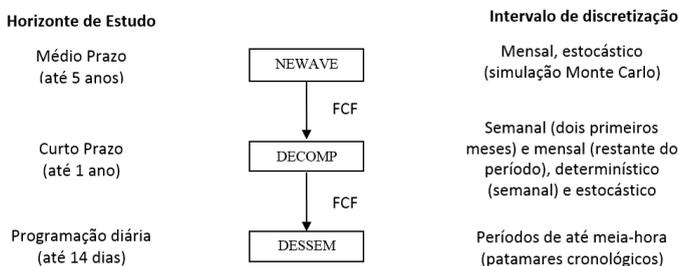


Figura 1 - Estrutura de acoplamento dos modelos

O cenário hidrológico é o principal responsável pela flutuação da FCF, portanto é mandatório incluir uma variável no modelo de *Machine Learning* que possa representar esta oscilação entre os diversos cenários. Neste trabalho será utilizada a variável PLD Semana Patamar como variável independente de entrada, da seguinte forma:

- Curtíssimo e Curto Prazo:
 - PLD Semana Patamar a partir do CMO do deck de saída do DESSEM e do DECOMP, especificamente a parte determinística dos modelos.
- Médio Prazo:
 - PLD Semana Patamar a partir do CMO do deck de saída do DECOMP, especificamente a parte estocástica do modelo.
- Longo Prazo
 - PLD Semana Patamar a partir do CMO do deck de saída do NEWAVE, apresentando uma abordagem estocástica dos cenários hidrológicos em mais longo prazo.

2.2 Variáveis temporais

O modelo de previsão de preço precisa considerar os padrões associados ao perfil temporal dos diversos horizontes considerados. Dessa forma, foram incluídas as seguintes variáveis independentes, conforme Tabela 1:

- Hora, podendo representar as oscilações horárias (i.e. horários de pico e vale dos preços), essenciais para a análise de curtíssimo prazo;
- Dia da Semana e Dia do Mês, podendo representar as oscilações dentro de uma mesma semana e mês, adequadas para análise de curto prazo (i.e. separação de dias úteis e fins de semana, feriados, início do mês, etc.);
- Mês, podendo representar os efeitos de eventos econômicos e meteorológicos anuais (i.e. período seco, período chuvoso, férias escolares, festas de fim de ano, carnaval, etc.).

Tabela 1 - Variáveis do modelo XGBoost do Sudeste, para o dia 17 de Abril de 2018

PLD Horário (R\$/MWh)	PLD Semana Patamar (R\$/MWh)	Ano	Mês	Dia	Dia da Semana	Hora
40,16	118,17	2018	4	17	1	0
40,16	118,17	2018	4	17	1	1
40,16	118,17	2018	4	17	1	2
40,16	118,17	2018	4	17	1	3
40,16	118,17	2018	4	17	1	4
40,16	118,17	2018	4	17	1	5
40,16	118,17	2018	4	17	1	6
40,16	125,33	2018	4	17	1	7
116,86	125,33	2018	4	17	1	8
119	125,33	2018	4	17	1	9
121,45	125,33	2018	4	17	1	10
121,45	125,33	2018	4	17	1	11
119,07	125,33	2018	4	17	1	12
121,41	125,33	2018	4	17	1	13
122,74	125,33	2018	4	17	1	14
123,81	125,33	2018	4	17	1	15
121,35	125,33	2018	4	17	1	16
119,11	125,33	2018	4	17	1	17
120,76	125,33	2018	4	17	1	18
119,16	125,33	2018	4	17	1	19
119,11	125,33	2018	4	17	1	20
119,1	125,33	2018	4	17	1	21
118,31	125,33	2018	4	17	1	22
114,26	125,33	2018	4	17	1	23

2.3 Variável de precificação horária (PLD horário)

Esta é a variável dependente e alvo da previsão. No intuito de se obter a correlação desta variável com todos os possíveis valores atribuídos às demais, ela faz parte da base de dados utilizada no processo de aprendizagem.

O histórico de PLD horário utilizado se inicia no dia 17 de abril de 2018 e se estende até a data de 16 de julho de 2021, com atualização diária, sendo que o valor horário do PLD passou a ser utilizado efetivamente na contabilização da CCEE a partir de 1/1/2021. Entre

abril de 2018 e dezembro de 2020, o parâmetro foi calculado em modo experimental, conhecido como “Operação Sombra”.

3. O MÉTODO DE *MACHINE LEARNING*: ÁRVORES DE DECISÃO

3.1 Conceitos

Árvores de Decisão são estruturas ramificadas com três tipos de nós, utilizadas na classificação dos dados. O nó raiz representa todo o conjunto de dados. Após ele, há os nós internos, que representam as variáveis do conjunto de dados e os critérios de decisão para novas ramificações. Em resumo, cada nó interno irá conter uma comparação de uma determinada variável $x_i \in X$ - também denominada variável independente ou de entrada do modelo - em relação a um valor específico, como por exemplo $x_i \geq 23,7$? (SKIENA 2017)

A partir da resposta a esta comparação apresentada neste nó interior, podendo assumir valores VERDADEIRO ou FALSO, haverá uma nova ramificação à esquerda ou direita. Esta ramificação irá continuar até não haver mais possibilidade de se avançar, seja pelo fato de todas as variáveis terem sido inseridas no modelo, ou por todas as amostras terem sido classificadas corretamente. O último nó da árvore é chamado de folha, e consiste na variável alvo da previsão, comumente denominada variável dependente e com notação y_i .

As estruturas mais comuns e utilizadas são chamadas de CART (*Classification and Regression Trees*) devido sua aplicabilidade em diversas classes de problemas, que envolvem a classificação e regressão.

Como vantagem em se utilizar este tipo de estrutura, podemos listar, conforme apresentado por (SKIENA 2017):

- Não linearidade, devido à possibilidade de poder representar fronteiras complexas de classificação de dados, através de ramificações lógicas;
- Suporte a variáveis categóricas, que resultam em resultados binários, VERDADEIRO ou FALSO, por exemplo;
- Interpretabilidade, devido à facilidade em ter uma estrutura auto-explicativa;
- Robustez, devido à possibilidade de crescimento exponencial de novas variáveis e possíveis testes;
- Aplicação em problemas de regressão, além das tradicionais aplicações de classificação no universo de Ciência de Dados. Esta vantagem é crucial para ser possível obter uma previsão de um valor numérico, alvo deste estudo.

3.2 O Modelo XGBoost

Matematicamente, é possível representar a previsão das variáveis dependentes \hat{y} como função de todas as árvores f_k e variáveis independentes x_i :

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (1)$$

Na Equação 1, $F = \{f(x) = \omega_{q(x)}\}$ é o espaço de árvores CARTs criado a partir do treinamento aditivo detalhado posteriormente no item 3.4, sendo que $q: R^m \rightarrow T$ é a estrutura de cada árvore que mapeia uma entrada ao respectivo indicador do nó folha, e $\omega \in R^T$ representa o peso de cada nó folha.

Uma vez definido o modelo que irá ser utilizado na previsão, a função objetivo que deverá ser minimizada para otimização será:

$$\mathcal{L}(\emptyset) = \sum_i l(y_i - \hat{y}_i) + \sum_k \Omega(f_k) \quad (2)$$

$$\text{Onde } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$$

O componente l da equação acima representa a função de perda e mede a diferença entre a previsão \hat{y}_i e o realizado y_i . O segundo componente Ω controla sua simplicidade, evitando assim o *overfitting*.

O algoritmo XGBoost utiliza a técnica de descida de gradiente para otimizar a função objetivo $\mathcal{L}(y, \hat{y})$ de forma iterativa, calculando $\partial_y \mathcal{L}(y, \hat{y})$ na direção do gradiente que minimiza a função.

Para realizar o cálculo do gradiente, as derivadas de primeira e segunda ordem são utilizadas. Adicionalmente, ao considerar que não há derivadas de segunda ordem para todas as funções objetivo geradas, e que é possível retirar os termos constantes da função objetivo, é possível defini-la conforme a Equação 3 para a iteração t :

$$\mathcal{L}^t = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (3)$$

Matematicamente, é possível representar uma árvore de decisão como uma função f_t , onde x representa cada dado do grupo de treinamento, $q(x)$ representando o nó folha e ω representando o peso ou nota:

$$f_t(x) = \omega_{q(x)} \quad (4)$$

Ao se definir $I_j = \{i \mid q(x_i) = j\}$ como o conjunto de dados associado ao nó folha j , é possível reescrever a Equação 3 da seguinte forma:

$$\mathcal{L}^t = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (5)$$

E ao se considerar uma estrutura fixa $q(x)$, é possível definir o peso ótimo ω_j^* do nó folha j da seguinte forma:

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (6)$$

E por fim, o valor encontrado para o peso na Equação 6 pode ser aplicado na função objetivo. Conforme proposto por (CHEN AND GUESTRIN 2016), o valor da função (3) pode ser utilizado como o índice de impureza, similar ao que o índice Gini representa para outros algoritmos baseados em estruturas de árvores. Contudo, este novo índice possui uma abrangência muito maior, tanto em termos de possibilidade de novas árvores, como também na análise de tendências que indicam se irão otimizar o modelo ou não. A Equação 3 pode ser reescrita conforme (7) ao aplicar ω_j^* para um estrutura fixa $q(x)$.

$$\mathcal{L}^t = -\frac{1}{2} \left(\frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} \right) + \gamma T \tag{7}$$

3.3 O algoritmo ganancioso

O processo de criação da árvore de decisão, baseado na técnica de descida de gradiente, busca resolver dois problemas. O primeiro está relacionado com a busca de uma estrutura otimizada de árvore que facilite o processo de previsão do modelo, e o segundo está relacionado ao processo de associação dos pesos, ou nota, a cada nó folha.

Tendo em vista a dificuldade em enumerar todas as possíveis estruturas de árvore q , o algoritmo procura o melhor ponto de divisão e ramificação da árvore, com o intuito de otimizar a função objetivo. Portanto, para cada nova possível ramificação é calculado o \mathcal{L}_{split} , ou ganho, partindo do nó raiz R e com destino aos nós folha da esquerda E da direita D , conforme (8). Em avanço, o ganho \mathcal{L}_{split} é um importante indicador da importância das variáveis utilizadas, e posteriormente será avaliado o valor obtido.

$$\begin{aligned} \mathcal{L}_{split} &= \mathcal{L}_E + \mathcal{L}_D - \mathcal{L}_R - \gamma \\ &= \frac{1}{2} \left[\frac{(\sum_{i \in I_E} g_i)^2}{\sum_{i \in I_E} h_i + \lambda} + \frac{(\sum_{i \in I_D} g_i)^2}{\sum_{i \in I_D} h_i + \lambda} - \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} \right] - \gamma \end{aligned} \tag{8}$$

Na Equação 8 há dois parâmetros importantes utilizados no ajuste do modelo XGBoost. O primeiro é o γ , com o papel de tornar o algoritmo mais conservador, uma vez que uma nova ramificação só será possível para $\mathcal{L}_{split} > 0$. Portanto, à medida que γ aumenta, maior a pressão para que os termos \mathcal{L}_E e \mathcal{L}_D sejam grandes para se justificar uma nova ramificação. O segundo é o λ , que tem por função criar uma regularização nos pesos, diminuindo assim a sensibilidade que uma única amostra pode apresentar para o modelo.

Uma vez concluída a estrutura da árvore, o próximo passo é associar um peso ou nota para cada nó folha conforme Equação 6.

Posteriormente, após a conclusão do treinamento aditivo apresentado abaixo, o valor final da previsão será definido pela somatória de cada peso, presente em cada árvore criada, multiplicado pela taxa de aprendizagem ν a partir da segunda árvore.

3.4 Treinamento aditivo

O Algoritmo Ganancioso define um método eficiente de criação das árvores $f_t(x)$, possibilitando o avanço em direção à diminuição da função de perda L da Equação 2. O próximo passo é realizar novas iterações deste algoritmo e adotar uma estratégia aditiva para considerar o avanço no aprendizado. Desta forma, é possível escrever o valor de saída do passo t conforme abaixo (XGBoost DEVELOPERS 2021), em linha com o que já foi descrito na Equação 1:

$$\begin{aligned}\hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} &= \hat{y}_i^{(1)} + f_2(x_i) \\ &\dots \\ \hat{y}_i^{(t)} &= \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)\end{aligned}\tag{9}$$

Para controlar a taxa de aprendizado do modelo, é incorporado um parâmetro ν , conhecido com taxa de aprendizado, que multiplica cada nova função f_i . Ele foi proposto por Friedman (FRIEDMAN 1999) e pode adotar valores entre 0 e 1. Valores pequenos tornam o modelo menos suscetível a interferências de novas árvores, facilitando a generalização e consequentemente impedindo o *overfitting*. Contudo, sua diminuição aumenta o risco de não se alcançar o valor ótimo com um número específico de iterações.

Em termos práticos, o treinamento aditivo do XGBoost criou um modelo com 100 árvores CART sequenciais ($n_estimators=100$), de acordo com o conceito de ensemble. Esta estratégia permitiu uma boa generalização do modelo, contudo o processo de criação de todas as árvores demanda um tempo considerável para sua conclusão, alcançando até 7 horas para a conclusão de previsões de longo prazo. Vale destacar que o computador utilizado é uma máquina de uso pessoal e com pouca capacidade de processamento, o que justifica grande parte do tempo utilizado para a definição do modelo de longo prazo.

4. RESULTADOS

4.1 Importância das variáveis

Os resultados abaixo apresentam indicadores importantes sobre a estrutura em árvore obtida a partir do treinamento do método XGBoost. Para facilitar a apresentação dos resultados, foram extraídos indicadores para os modelos de Previsão de Curto e Curtíssimo prazo de dias úteis. Contudo a metodologia seria a mesma para a aplicação em finais de semana e feriados.

4.1.1 Critério de ganho

Em relação aos indicadores, o primeiro é chamado de Ganho, apresentado anteriormente como \mathcal{L}_{split} na Equação 8, e indica o valor médio de Ganho de uma variável utilizada na criação de novas ramificações. Um maior valor de Ganho de uma variável, em relação às demais, representa sua maior importância em gerar previsões. Vale ressaltar que os valores de Ganho estão normalizados de forma que a somatória de todos os Ganhos será igual a 1.

Nas Figuras 3 a 6 temos as variáveis sendo representadas pelas siglas $f0$ (PLD Semana Patamar), $f1$ (Mês), $f2$ (Dia), $f3$ (Dia da Semana) e $f4$ (Hora). Para todos os modelos, foi verificado que o PLD Semana Patamar se sobressai às demais variáveis, com valores de Ganho médio variando de 0,78 para o Nordeste a 0,9 para o Norte.

A segunda variável de destaque neste quesito é o Mês, em especial para o Nordeste, devido a menor participação do PLD Semana Patamar neste submercado.

Com relação às variáveis Dia da Semana e Hora, elas apresentaram os menores ganhos. À medida que o modelo evolui, com a inclusão de novas amostras e outras variáveis independentes do SIN, estas duas variáveis podem ser candidatas a serem removidas, a depender da melhora de dois fatores: precisão e tempo de processamento.

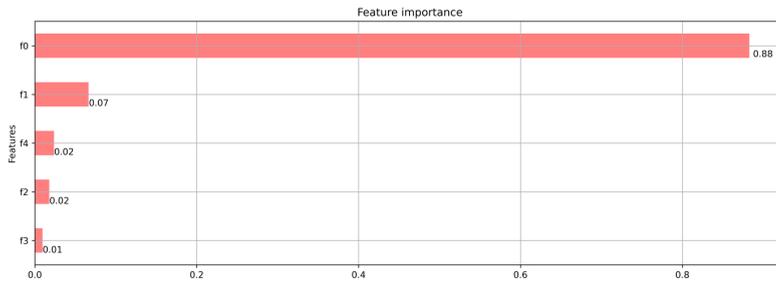


Figura 2 - Importância da variável (critério Ganho) – Sudeste

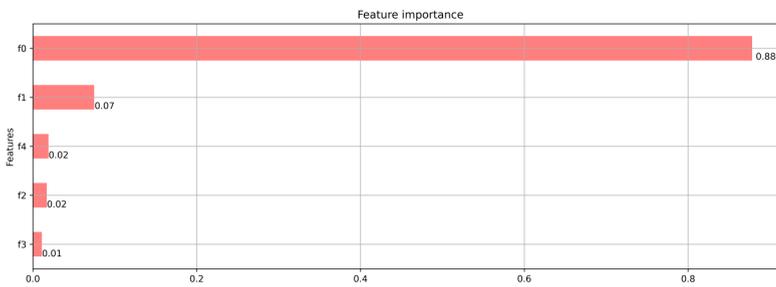


Figura 3 - Importância da variável (critério Ganho) – Sul

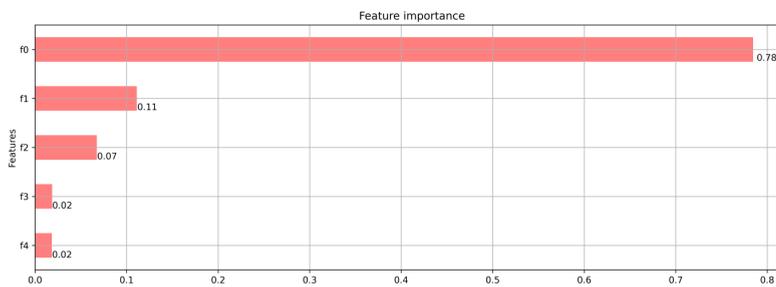


Figura 4 - Importância da variável (critério Ganho) – Nordeste

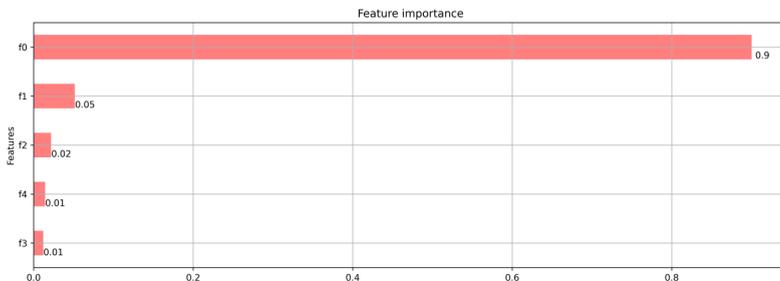


Figura 5 - Importância da variável (critério Ganho) – Norte

4.1.2 Critério de peso

O segundo indicador é chamado de Peso e representa a frequência com que esta variável foi utilizada para se gerar novas ramificações. Um maior valor de Peso de uma variável, em relação às demais, também indica sua maior importância na definição do modelo de previsão.

De forma análoga ao verificado com o Ganho, a variável PLD Semana Patamar foi o destaque, com uma utilização variando entre 607 vezes para o Nordeste a 717 vezes no Sul. E com o pior desempenho neste quesito, a variável Dia da Semana apresentou uma utilização variando entre 86 para o Sudeste e 108 para o Norte.

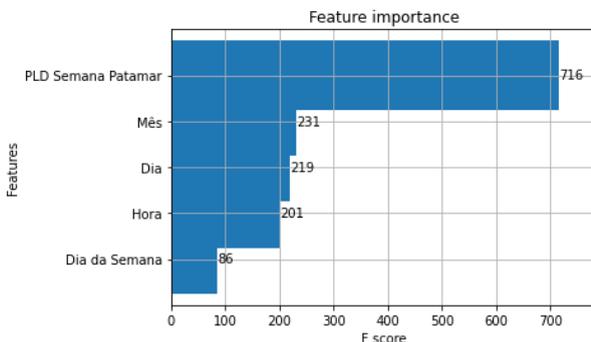


Figura 6 - Importância da variável (critério Peso) – Sudeste

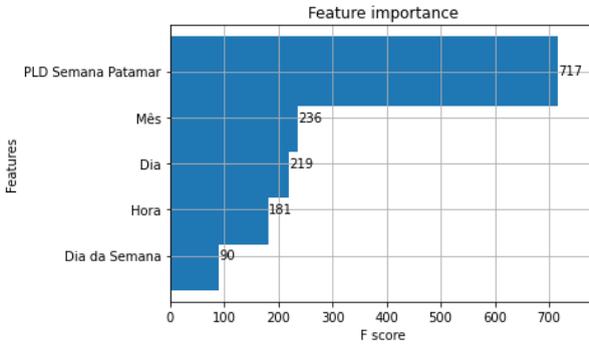


Figura 7 - Importância da variável (critério Peso) – Sul

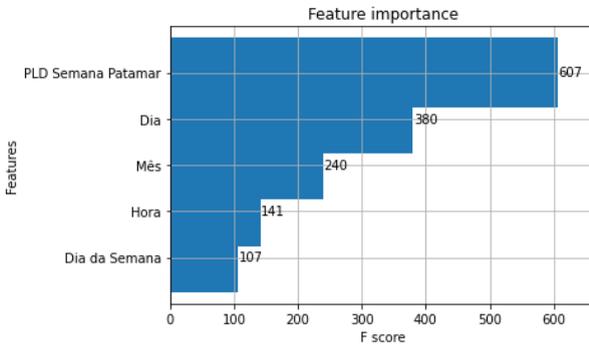


Figura 8 - Importância da variável (critério Peso) – Nordeste

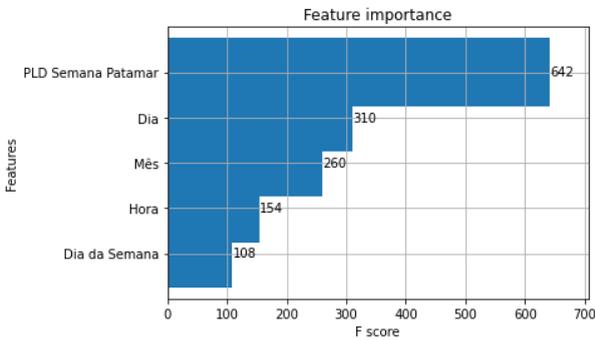


Figura 9 - Importância da variável (critério Peso) – Norte

4.2 Curtíssimo prazo

Para a avaliação do desempenho da previsão de curtíssimo e curto prazo, foram utilizadas duas métricas comumente empregadas em modelos de regressão linear, que são o MAPE e RMSE (Erro Percentual Absoluto Médio e Raiz Quadrada do Erro Médio, respectivamente).

Um terceiro indicador foi especialmente criado, chamado de IAT (Índice de Acerto na Tendência), e representa a aderência entre a variação do PLD real, verificado entre horas subsequentes, com a variação prevista nesses mesmos períodos. Para exemplificar, considere que entre as horas 00:00 e 01:00 de um determinado dia foi previsto um aumento do PLD, e ao se verificar o PLD real para o mesmo período, foi constatado que o PLD também apresentou aumento. Neste exemplo, será atribuída uma nota 1 para este acerto, e caso não fosse similar, seria atribuída a nota 0.

Este indicador visa dotar o decisor da sensibilidade sobre o que esperar do comportamento do preço no horizonte de análise, respondendo com certa acurácia à pergunta: “Nesse horizonte o PLD vai subir ou vai baixar”?

Após a análise de todos os intervalos, todas as notas são somadas e o resultado é dividido pela máxima pontuação possível, que é 24 (número de horas do dia). O Resultado, multiplicado por 100, irá representar uma nota percentual da aderência entre a Previsão e o Realizado, na variação do PLD.

$$nota_i = \begin{cases} 1, & \text{Se } Prev_i > Prev_{i+1} \text{ e } Real_i > Real_{i+1}, \text{ ou } Prev_i < Prev_{i+1} \text{ e } Real_i < Real_{i+1} \\ 0, & \text{caso contrário.} \end{cases} \quad (10)$$

$$IAT = \frac{\sum_{i=1}^{24} nota_i}{24} \times 100$$

Sendo:

Prev_i=Variável do Aumento ou Diminuição do PLD previsto,para o período *i*

Real_i=Variável do Aumento ou Diminuição do PLD real,para o período *i*

nota_i=nota atribuída para o acerto da previsão,para o período *i*

4.2.1 Desempenho da previsão até o primeiro dia à frente

Neste método de análise de desempenho, a previsão de um dia à frente levou em consideração a inclusão do dia anterior na Base de Aprendizado. Esta estratégia permitiu manter o modelo o mais atualizado possível, no que se refere às correlações de variáveis do modelo do SIN, embutidas na variável de PLD Semana Patamar, com o PLD Horário a ser previsto.

O período previsto foi dividido em discretizações horárias das 00h do dia 17 de Julho de 2021 às 23h do dia 23 de Julho de 2021, e as parciais de MAPE, RMSE e IAT podem ser visualizadas na Tabela 2.

Tabela 2 - Resultado das previsões de um dia à frente

Sudeste	MAPE (%)	RMSE (R\$/MWh)	IAT (%)	Sul	MAPE (%)	RMSE (R\$/MWh)	IAT (%)
17-07-2021 sábado	1,67	11,73	65,22	17-07-2021 sábado	3,04	20,79	56,52
18-07-2021 domingo	5,92	40,32	82,61	18-07-2021 domingo	4,49	30,75	73,91
19-07-2021 segunda	1,25	10,6	73,91	19-07-2021 segunda	1,06	8,06	73,91
20-07-2021 terça	1,09	7,08	73,91	20-07-2021 terça	0,92	6	60,87
21-07-2021 quarta	2,38	15,52	82,61	21-07-2021 quarta	2,19	16,52	65,22
22-07-2021 quinta	1,21	8,25	65,22	22-07-2021 quinta	1,26	8,65	56,52
23-07-2021 sexta	0,81	5,15	60,87	23-07-2021 sexta	0,85	5,67	78,26
Médias	2,05	11,3	72,05	Médias	1,97	13,78	66,46
Nordeste	MAPE (%)	RMSE (R\$/MWh)	IAT (%)	Norte	MAPE (%)	RMSE (R\$/MWh)	IAT (%)
17-07-2021 sábado	1,33	10,81	47,83	17-07-2021 sábado	2,18	14,53	60,87
18-07-2021 domingo	5,08	32,12	69,57	18-07-2021 domingo	4,72	30,38	69,57
19-07-2021 segunda	1,30	9,69	69,57	19-07-2021 segunda	1,48	12,65	73,91
20-07-2021 terça	1,49	10,94	73,91	20-07-2021 terça	0,56	4,1	73,91
21-07-2021 quarta	2,48	16,63	52,17	21-07-2021 quarta	1,84	14,97	73,91
22-07-2021 quinta	1,67	11,13	47,83	22-07-2021 quinta	0,81	7,23	65,22
23-07-2021 sexta	1,86	16,31	69,57	23-07-2021 sexta	1,00	6,66	73,91
Médias	2,17	15,38	61,49	Médias	1,80	12,93	70,19

Para exemplificar o perfil de movimento do PLD no período, em comparação com a previsão, seguem abaixo as visualizações por submercado, para o dia 19 de Julho de 2021, uma segunda-feira.

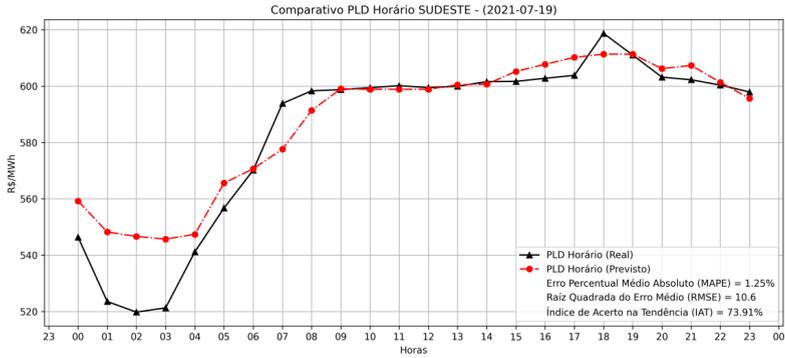


Figura 10 – Exemplo de previsão de curtíssimo prazo Sudeste (19/07/2021)

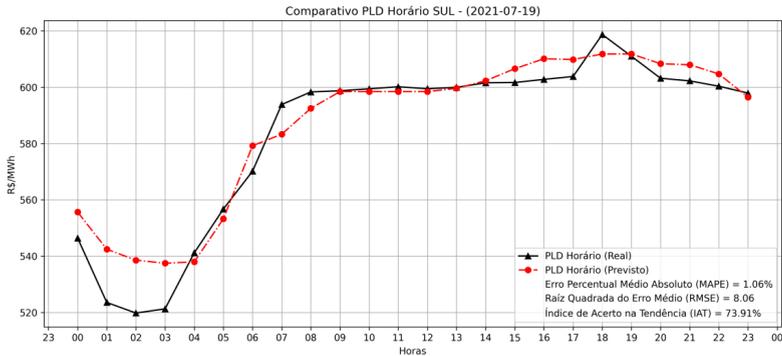


Figura 11 - Exemplo de previsão de curtíssimo prazo Sul (19/07/2021)

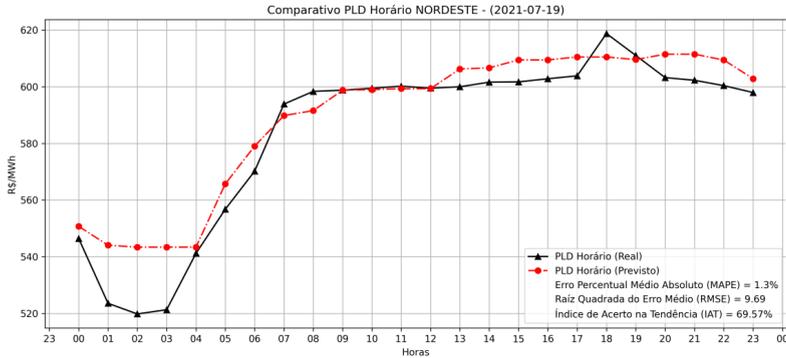


Figura 12 - Exemplo de previsão de curtíssimo prazo Nordeste (19/07/2021)

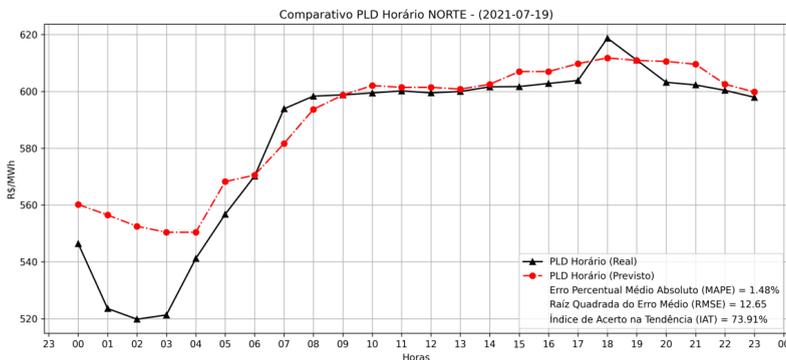


Figura 13 - Exemplo de previsão de curtíssimo prazo Norte (19/07/2021)

4.2.2 Desempenho da previsão até o sétimo dia à frente

Neste método, foi realizada uma previsão do PLD com horizonte de sete dias à frente. Para isso, o modelo foi treinado apenas no início do processo, ou seja, à medida que os dias evoluíam dentro do horizonte não houve novas atualizações do modelo. Ao final do período de sete dias, foi medida sua assertividade com base na visão inicial.

Na Tabela 3 se encontram os resultados da previsão realizada no dia 16 de Julho de 2021, para o período de 17 a 23 de Julho de 2021.

Tabela 3 - Resultado das previsões de uma semana à frente

	MAPE (%)	RMSE (R\$/MWh)	IAT (%)
Sudeste	1,77	15,10	73,81
Sul	2,14	17,22	69,64
Nordeste	2,08	17,32	60,71
Norte	1,84	14,56	72,62

O perfil da previsão em comparação aos valores reais pode ser visualizado nas Figuras abaixo.

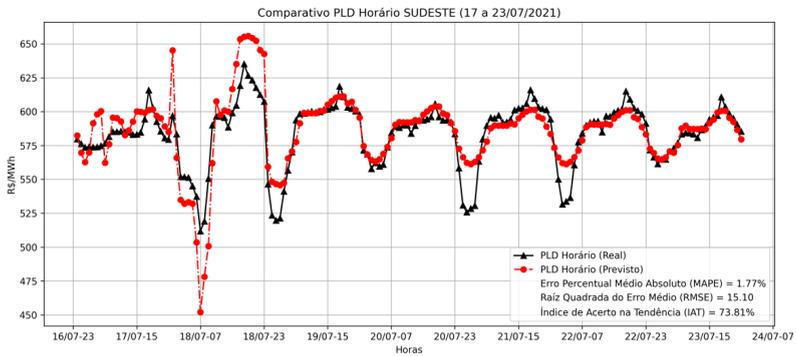


Figura 14 - Previsão sete dias à frente para o Sudeste

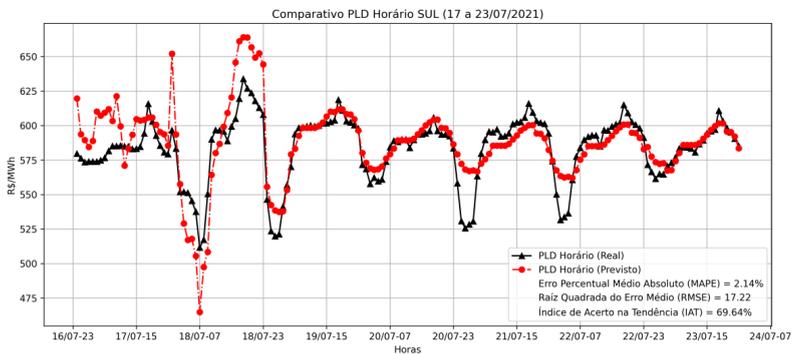


Figura 15 - Previsão sete dias à frente para o Sul

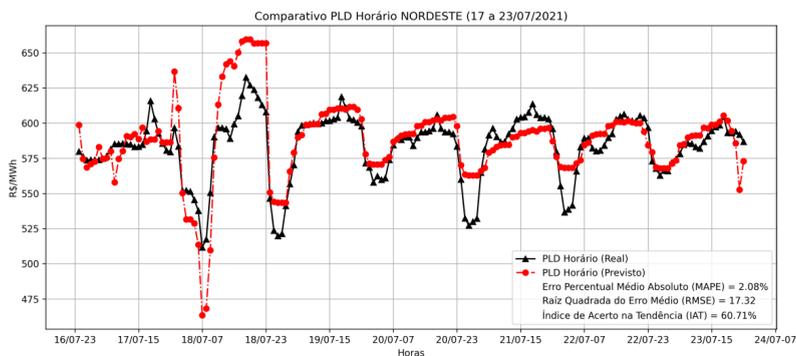


Figura 16 - Previsão sete dias à frente para o Nordeste

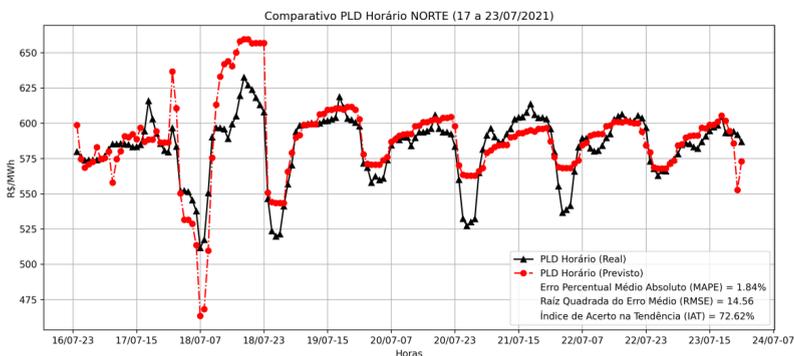


Figura 17 - Previsão sete dias à frente para o Norte

4.2.3 Comentários

As previsões de curtíssimo e curto prazo apresentaram uma boa assertividade, conforme demonstrado pelos índices MAPE, RMSE e IAT.

Para a previsão de um dia à frente, o índice MAPE variou entre 0,56% para o submercado Norte a 5,92% para o submercado Sudeste. O erro médio indicado pelo índice RMSE variou entre 4,10 a 40,32 R\$/MWh. O IAT variou entre 47,83% para o submercado Nordeste e 82,61% para o submercado Sudeste.

Para a previsão de sete dias à frente, os valores de MAPE,

RMSE e IAT se aproximaram do valor médio dos mesmos índices, obtidos pelo cálculo da média das previsões de um dia à frente. O menor MAPE para sete dias à frente, de 1,77%, foi registrado para o Sudeste, e o maior, de 2,14%, foi para o Sul. Em termos comparativos, os MAPEs médios de um dia à frente foram de 2,05% e 1,97%, respectivamente para os submercados Sudeste e Sul. Os outros índices também apresentaram essa similaridade.

Ao se comparar o desempenho das previsões entre os submercados, verifica-se que o Nordeste apresenta o menor desempenho entre todos, apresentando os piores MAPE, RMSE e IAT na previsão de um dia à frente e o piores RMSE e IAT na previsão de sete dias à frente.

Parte do desempenho inferior pode ser explicada pela menor correlação Pearson – entre o PLD Semana Patamar de entrada e o PLD Horário previsto - descrita na Tabela 4, e menores valores dos indicadores de Ganho e Peso do PLD Semana Patamar, descritos nos itens 4.1.1 e 4.1.2, ao se comparar com os demais submercados. Um dos motivos deste comportamento é o maior impacto do perfil estocástico da geração renovável no preço do PLD horário.

Tabela 4 - Correlação entre o PLD Horário e o PLD Semana Patamar

	Correlação Pearson (ρ)
Sudeste	0,929
Sul	0,923
Nordeste	0,851
Norte	0,933

4.3 Curto prazo

4.3.1 Desempenho da previsão até o primeiro mês à frente

De forma análoga à previsão e análise de desempenho de sete dias à frente, neste método a atualização do modelo ocorre no dia da publicação da previsão, e seu desempenho é medido ao final do período do primeiro mês. Para a análise, foi definido um limite de um mês de previsão, contudo esta análise pode ser estendida para até dois meses, conforme arquivos públicos fornecidos pelo DECOMP. A partir do terceiro mês em diante, a previsão adquire caráter informativo de

tendências conjunturais do setor.

Para este horizonte, a data da publicação coincide com a reunião da Programação Mensal da Operação (PMO), realizada na última sexta-feira do mês anterior à vigência desta programação. No caso da previsão abaixo, a data do PMO foi 25 de Junho de 2021, para a programação do mês de Julho de 2021.

Em resumo, as parciais de MAPE, RMSE e IAT obtidas constam na Tabela 5.

Tabela 5 - Resultado das previsões de um mês à frente

	MAPE (%)	RMSE (R\$/MWh)	IAT (%)
Sudeste	14,00	87,27	60,08
Sul	13,00	81,97	46,64
Nordeste	9,06	63,94	46,24
Norte	13,16	82,41	54,03

O perfil da previsão em comparação aos valores reais pode ser visualizado nas figuras abaixo.

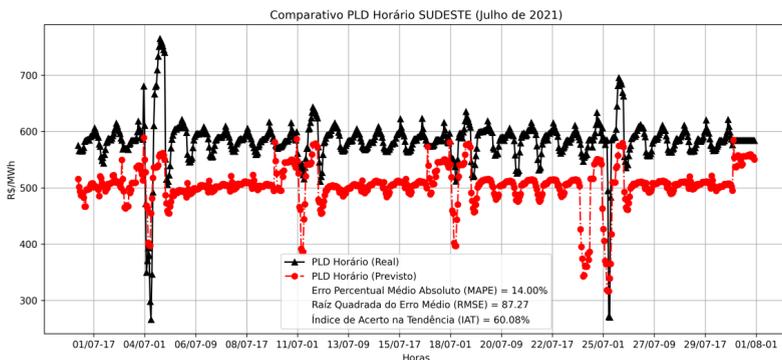


Figura 18 - Previsão um mês à frente para o Sudeste

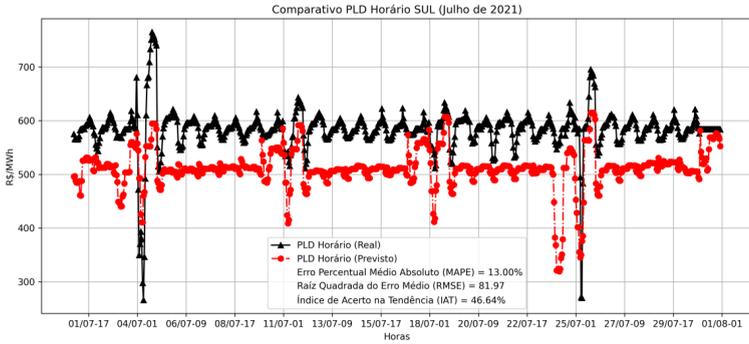


Figura 19 - Previsão um mês à frente para o Sul

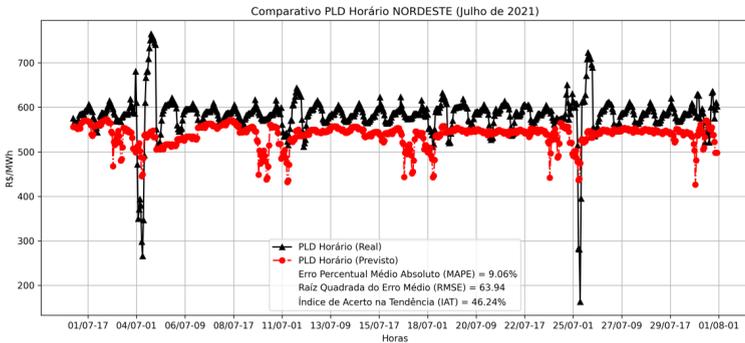


Figura 20 - Previsão um mês à frente para o Nordeste

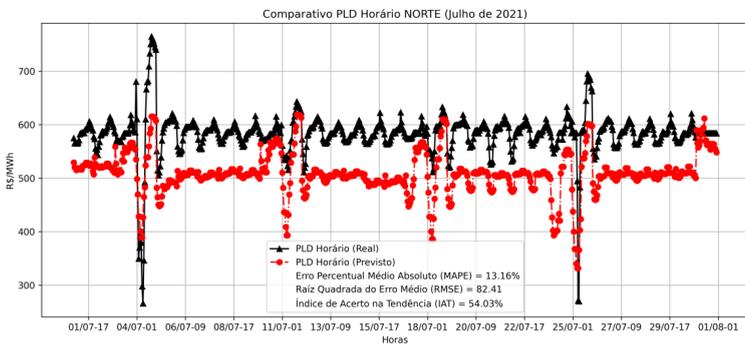


Figura 21 - Previsão um mês à frente para o Norte

4.3.2 Comentários

A previsão de curto prazo apresenta uma volatilidade superior à de curtíssimo prazo, devido às atualizações do PLD Semana Patamar que ocorrem entre a disponibilização do estudo de curto prazo, a partir da rv_0 , até a efetiva divulgação do PLD Semana Patamar que será vigente para a semana em questão.

Para entender um pouco sobre a volatilidade do PLD Semana Patamar, a Tabela 6 disponibiliza os valores destas variáveis, divididas em Patamares Pesado, Médio e Leve, em cada revisão intermediária do DECOMP para o submercado Sudeste. Conforme visto na Tabela 4, o submercado Sudeste é o segundo melhor em termos de Correlação Pearson, portanto um valor informado no início do processo terá um impacto ainda mais relevante.

Tabela 6 - Evolução dos PLDs Semana Patamar (Jul e Ago/21)

Revisões DECOMP	Patamar	Semana 1 (R\$/MWh)	Semana 2 (R\$/MWh)	Semana 3 (R\$/MWh)	Semana 4 (R\$/MWh)	Semana 5 (R\$/MWh)	Semana 6 (R\$/MWh)
rv_0 (26/06/21)	Pesado	824,72	820,92	819,3	823,72	820,74	820,86
	Médio	819,73	817,74	816,9	821,31	819,28	820,86
	Leve	806,22	806,67	807,75	809,47	807,74	803,81
rv_1 (03/07/21)	Pesado	-	823,95	827,28	823,99	821,31	823,93
	Médio	-	822,04	824,1	821,2	821,31	822,02
	Leve	-	807,81	810,19	810,11	808,26	806,82
rv_2 (10/07/21)	Pesado	-	-	902,04	898,89	891,17	885,6
	Médio	-	-	895,29	894,9	891,06	885,58
	Leve	-	-	877,12	877,79	870,98	865,65
rv_4 (17/07/21)	Pesado	-	-	-	933,78	928,62	923,04
	Médio	-	-	-	927,27	927,53	923,04
	Leve	-	-	-	911,76	907,65	902,2
rv_4 (24/07/21)	Pesado	-	-	-	-	1014,81	1009,07
	Médio	-	-	-	-	1008,26	1004,9
	Leve	-	-	-	-	983,31	981,47

Conforme a Tabela 6, os valores de PLD Semana Patamar para a Semana 5 (24 a 30 de Julho de 2021), divulgados na revisão intermediária rv_4 , são 23,65%, 23,07% e 21,74% maiores do que as divulgadas na rv_0 para os Patamares Pesado, Médio e Leve respectivamente.

Este mesmo padrão de volatilidade pode ser visto na Semana 6, que por ser a última representa os PLDs Semana Patamar médios a serem utilizados para todo o segundo mês. Neste caso os índices MAPE, RMSE e IAT, do segundo mês, serão ainda mais impactados, pois todas as semanas do mês irão sofrer com PLDs Semana Patamar aproximadamente 22% menores.

Mesmo com todo este cenário de volatilidade do PLD Semana Patamar utilizado como variável independente do modelo, foi possível obter MAPes entre 9,06% para o Nordeste e 14% para o Sudeste. Para o RMSE, foram obtidos índices que variaram de 63,94 a 87,27 R\$/MWh, respectivamente para o Nordeste e Sudeste. Por fim, para o IAT, foram obtidas parciais entre 46,24% para o Nordeste e 60,08% para o Sudeste.

4.4 Médio prazo

4.4.1 Previsão sem acoplamento de resultados de curto prazo

A previsão de médio prazo se inicia a partir do fim do horizonte definido pela análise de curto prazo, ou seja, se inicia no terceiro mês e se estende até o 14º Mês. Na estratégia de previsão sem acoplamento de resultados de curto prazo, a Base de Aprendizado incluiu apenas os valores reais do PLD Horário, divulgados no momento da realização deste estudo.

Abaixo é possível visualizar o comportamento do PLD Horário ao longo do período, específico para os dias úteis. Para sábados, domingos e feriados, a projeção utiliza a mesma estratégia. A título de comparação, foram incluídos nos gráficos os PLDs Semana Patamar utilizados na entrada.

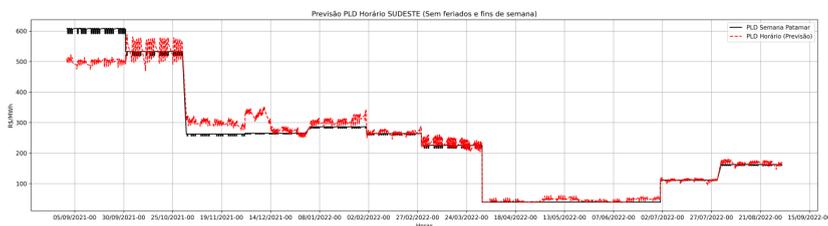


Figura 22 - Previsão doze meses à frente para o Sudeste (dias úteis)

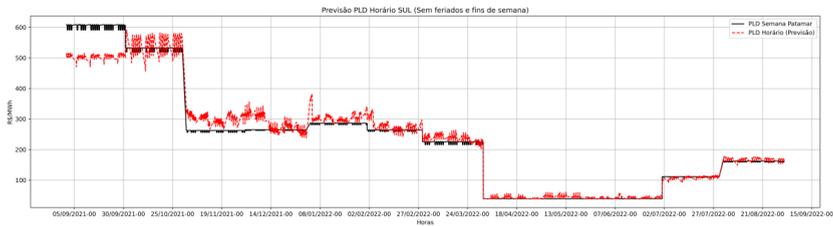


Figura 23 - Previsão doze meses à frente para o Sul (dias úteis)

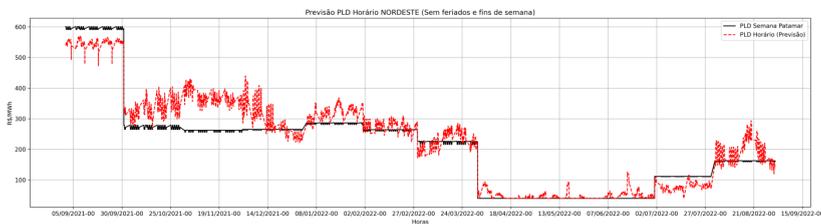


Figura 24 - Previsão doze meses à frente para o Nordeste (dias úteis)

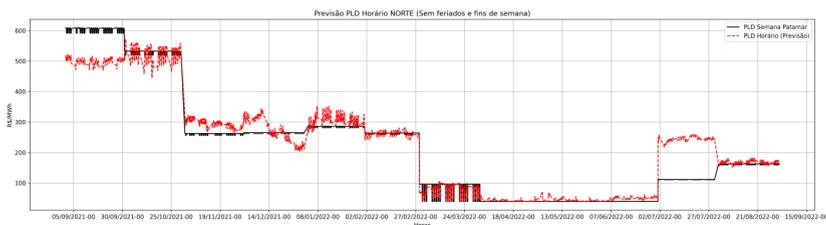


Figura 25 - Previsão doze meses à frente para o Norte (dias úteis)

4.4.2 Efeitos do acoplamento de resultados de curto prazo

Nesta estratégia de previsão com acoplamento de resultados de curto prazo, a Base de Aprendizado incorporou os resultados previstos no curto prazo, conforme apresentado no item 4.3, com isso, os resultados previstos para o primeiro e segundo meses foram incorporados ao modelo.

Na Tabela 7 é possível visualizar o efeito da variação Média do PLD Horário após este acoplamento.

Tabela 7 - PLD Horário com e sem acoplamento de Curto Prazo (Sudeste)

Sudeste	Média PLD Horário com acoplamento (R\$/MWh)	Média PLD Horário sem acoplamento (R\$/MWh)	Variação (%)
Set-21	573,95	500,85	14,60
Out-21	528,11	547,98	-3,63
Nov-21	291,72	301,59	-3,27
Dez-21	284,59	293,05	-2,89
Jan-22	320	306,88	4,28
Fev-22	267,51	269,03	-0,56
Mar-22	238,6	236,2	1,02
Abr-22	42,63	41,93	1,67
Mai-22	45,65	48,12	-5,13
Jun-22	45,84	46,84	-2,13
Jul-22	110,21	111,2	-0,89
Ago-22	165,91	166,24	-0,20
Médias	242,89	239,16	0,24%

Tabela 8 - PLD Horário com e sem acoplamento de Curto Prazo (Sul)

Sul	Média PLD Horário com acoplamento (R\$/MWh)	Média PLD Horário sem acoplamento (R\$/MWh)	Variação (%)
Set-21	582,31	505,36	15,23
Out-21	544,57	548,24	-0,67
Nov-21	296,29	304,19	-2,60
Dez-21	290,86	284,19	2,35
Jan-22	302,81	307,56	-1,54
Fev-22	268,07	281,05	-4,62
Mar-22	233,79	237,96	-1,75
Abr-22	42,09	42,42	-0,78
Mai-22	48,49	45,49	6,59
Jun-22	43,32	43,4	-0,18
Jul-22	108,87	107,94	0,86
Ago-22	172,45	168,51	2,34
Médias	244,49	239,69	1,27

Tabela 9 - PLD Horário com e sem acoplamento de Curto Prazo (Nordeste)

Sul	Média PLD Horário com acoplamento (R\$/MWh)	Média PLD Horário sem acoplamento (R\$/MWh)	Variação (%)
Set-21	579,84	549,43	5,53
Out-21	338,94	331,41	2,27
Nov-21	370,44	367,59	0,78
Dez-21	293,75	292,49	0,43
Jan-22	314,85	322,66	-2,42
Fev-22	277,78	278,01	-0,08
Mar-22	219,96	227,94	-3,50
Abr-22	48,88	49,02	-0,29
Mai-22	48,08	45,03	6,77
Jun-22	59,54	51,02	16,70
Jul-22	85,95	79,98	7,46
Ago-22	206,68	195,56	5,69
Médias	237,06	232,51	3,28

Tabela 10 - PLD Horário com e sem acoplamento de Curto Prazo (Norte)

Sul	Média PLD Horário com acoplamento (R\$/MWh)	Média PLD Horário sem acoplamento (R\$/MWh)	Variação (%)
Set-21	581,55	500,03	16,30
Out-21	517,82	526,42	-1,63
Nov-21	284,84	294,31	-3,22
Dez-21	270,69	270,94	-0,09
Jan-22	317,51	309,2	2,69
Fev-22	270,23	263,22	2,66
Mar-22	72,66	73,79	-1,53
Abr-22	43,98	42,89	2,54
Mai-22	44,37	44,83	-1,03
Jun-22	49,18	49,23	-0,10
Jul-22	234,52	244,57	-4,11
Ago-22	167,53	166,11	0,85
Médias	237,91	232,13	1,11

4.4.3 Comentários

A visão de médio prazo, expressa no PLD Semana Patamar de entrada, apresenta uma projeção esperada pela CCEE, contudo cenários diferentes podem ser criados a partir de novas configurações nos modelos oficiais, gerando assim novos arquivos de entrada.

Analisando a projeção do PLD Horário, a partir desta projeção, é possível verificar um desacoplamento relevante para o mês de setembro, entre o PLD Horário e o PLD Semana Patamar, em todos os submercados. Isto indica que o modelo de XGBoost, através da sua análise de padrão comportamental do preço, aponta para uma necessidade de ajuste do PLD Horário, mesmo que o PLD Semana Patamar não indique esta correção.

Ao se avaliar o desempenho do acoplamento dos resultados de curto prazo na Base de Aprendizagem do modelo, nota-se que houve um aumento médio do PLD Horário para o período.

Em especial para o primeiro mês da série, houve uma correção relevante para cima, variando de 5,53% para o Nordeste a 16,3% para o Norte. Contudo, há meses de diminuição da média, e como resultado final, a média do PLD Horário sofreu um acréscimo, variando de 0,24% para o Sudeste a 3,28% para o Nordeste.

Se considerarmos que a Base de Aprendizagem ainda não é muito grande, esta estratégia de extensão pode ser uma ferramenta útil nas projeções, devendo seus efeitos ser analisados em conjunto com o método sem o acoplamento.

4.5 Longo Prazo

4.5.1 Previsão sem acoplamento de resultados de médio prazo

A previsão de longo prazo se inicia a partir do fim do horizonte definido pela análise de médio prazo, ou seja, no 15º Mês. Contudo, é possível iniciar o horizonte de longo prazo já no terceiro mês, dispensando assim o estudo prospectivo de médio prazo disponibilizado pela CCEE ou outra fonte.

Tendo em vista a incerteza inerente a este horizonte, os arquivos de entrada oficiais do NEWAVE disponibilizam 2000 séries sintéticas, cada uma representando um cenário hidrológico possível, e com isso, as séries foram agrupadas para criar três cenários possíveis de PLD Semana Patamar em cada mês deste horizonte:

- Cenário pessimista: composto por 20% das maiores médias ponderadas de CMOs, calculadas a partir do intervalo entre o Percentil 80 e o maior valor de CMO encontrado,

- Cenário provável: composto por 60% das médias ponderadas de CMOs, calculadas a partir do intervalo entre o Percentil 20 e o Percentil 80,
- Cenário otimista: composto por 20% das menores médias ponderadas de CMOs, calculadas a partir do intervalo entre o Percentil 20 e o menor valor de CMO encontrado.

Como resultado, os gráficos abaixo apresentam três cenários para a projeção de PLD Horário de dias úteis. Os cenários Prováveis foram representados pelas linhas pretas, os cenários Otimistas pelas linhas azuis e por fim, os cenários Pessimistas pelas linhas vermelhas.

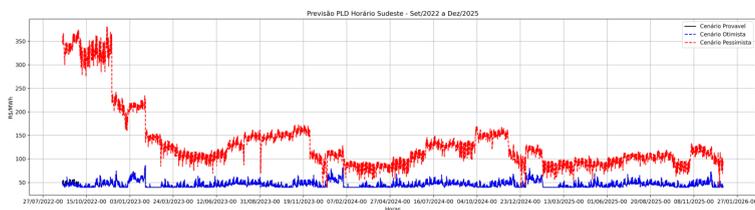


Figura 26 - Previsão Longo Prazo Sudeste sem acoplamento

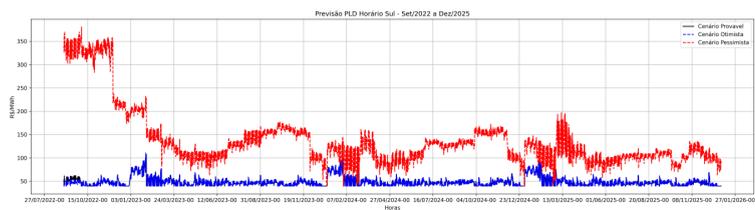


Figura 27 - Previsão Longo Prazo Sul sem acoplamento

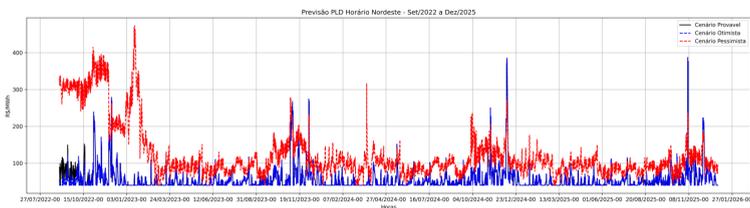


Figura 28 - Previsão Longo Prazo Nordeste sem acoplamento

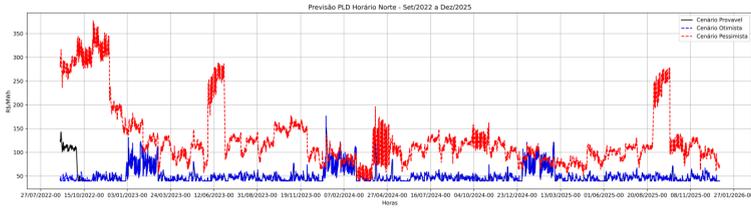


Figura 29 - Previsão Longo Prazo Norte sem acoplamento

Além da visão da projeção com discretização horária para os dias úteis do período, é interessante comparar o PLD Horário previsto médio de cada mês, para cada cenário, frente aos valores Médios, Mínimos e Máximos reais verificados historicamente em cada um deles, que compõe a base de dados que se inicia com o PLD Horário operando como sombra.

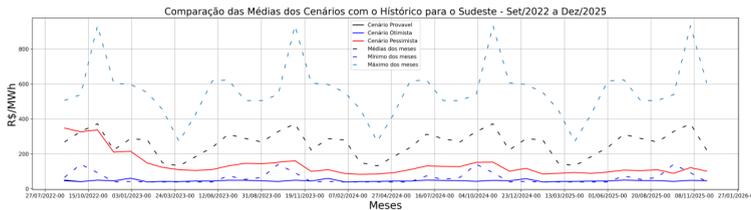


Figura 30 - Comparação Sudeste: Médias da Previsão x Base Histórica mensal

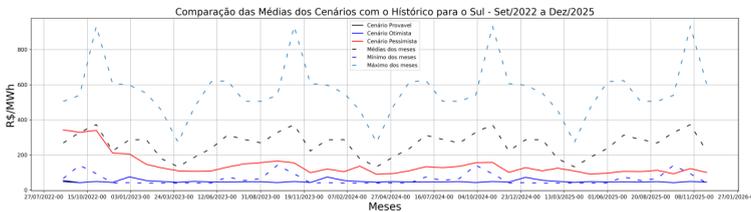


Figura 31 - Comparação Sul: Médias da Previsão x Base Histórica mensal

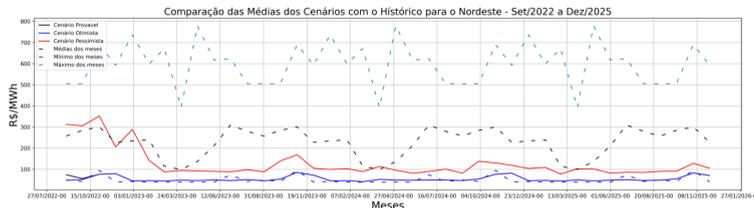


Figura 32 - Comparação Nordeste: Médias da Previsão x Base Histórica mensal

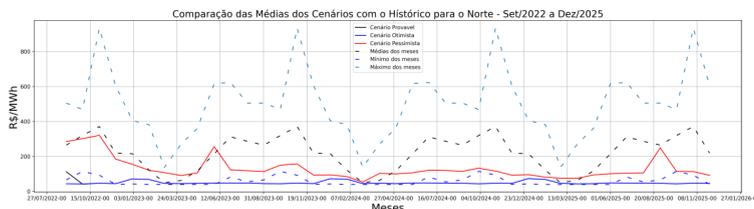


Figura 33 - Comparação Norte: Médias da Previsão x Base Histórica mensal

4.5.2 Efeitos do acoplamento de resultados de médio prazo

De forma análoga ao acoplamento na análise de Médio Prazo, nesta análise foi realizado o acoplamento dos resultados de Médio Prazo na Base de Aprendizado, estendendo a Base de Aprendizado inicial em doze meses.

Com isso, os gráficos abaixo representam a variação percentual dos PLDs Horários médios de cada mês, após o efeito de acoplamento de resultados de médio prazo. Adicionalmente, foram incluídos os três cenários possíveis: otimistas, provável e pessimista.



Figura 34 - Variação do PLD Horário Médio com acoplamento (Sudeste)



Figura 35 - Variação do PLD Horário Médio com acoplamento (Sul)



Figura 36 - Variação do PLD Horário Médio com acoplamento (Nordeste)



Figura 37 - Variação do PLD Horário Médio com acoplamento (Norte)

4.5.3 Comentários

Os resultados de longo prazo confirmam uma característica do NEWAVE que impacta de forma relevante a definição do PLD Semana Patamar, e por consequência o PLD Horário. Em resumo, o NEWAVE apresenta uma tendência otimista para os cenários de preço.

Esta percepção pôde ser confirmada ao se verificar que para todos os submercados, o cenário Provável, que representa 60% das

séries hidrológicas possíveis do NEWAVE, está muito próximo do cenário Otimista, que representa 20% das séries de menor valor. Nos gráficos, as linhas até se confundem por apresentarem resultados muito semelhantes.

Relativo ao acoplamento do estudo prospectivo de médio prazo da CCEE à Base de Aprendizado é possível verificar que apenas para o Sudeste houve um padrão de incremento do PLD Horário médio, corrigindo de forma marginal a tendência otimista. Para os outros submercados há comportamentos diferentes nos três cenários, indicando assim a necessidade de se avaliar esta estratégia em base mensal. De forma consolidada, a Tabela 11 apresenta o efeito do acoplamento, em cada submercado, na variação média para o período deste estudo.

Tabela 11 - Variação percentual após o acoplamento no Longo prazo

	Variação Média (%) com o acoplamento		
	Provável	Pessimista	Otimista
Sudeste	2,01	0,17	2,01
Sul	-4,01	-2,09	4,03
Nordeste	1,57	-1,53	1,05
Norte	-0,62	5,90	-0,67

5. CONCLUSÃO

Devido aos impactos das mudanças climáticas, o mundo tem vivenciado uma transição energética que inclui de forma massiva novas formas de geração de energia através de fontes renováveis. Estas fontes deverão coexistir com outras formas convencionais de geração, garantindo a segurança no atendimento à demanda de energia. E com novas formas de geração de energia, novos padrões de comportamento serão inseridos nos modelos de planejamento e operação de sistemas elétricos.

No caso brasileiro, caracterizado por um sistema hidrotérmico renovável, este estudo oferece uma ferramenta de apoio relevante na previsão de curtíssimo e curto prazo. As métricas indicam que esta estratégia pode ser uma ferramenta de suporte na compra e venda de energia elétrica no curto prazo.

Apesar deste tipo de operação ser pouco comum atualmente, a tendência é que se popularizem à medida que tecnologias de armazenamento de energia, como Sistemas de Armazenamento de Energia por Baterias (BESS em inglês) ou usinas hidráulicas reversíveis,

passem a ser adotadas em larga escala. Este movimento de descentralização do setor elétrico é mundial, e ocorre de forma concomitante a outros movimentos, como a digitalização por exemplo. A expectativa é de que os usuários de tais sistemas descentralizados demandem por soluções de acesso simples e rápido, e que os auxiliem na avaliação de oportunidades de compra e venda de energia.

Para análises de prazos mais longos, esta ferramenta pode se tornar um mecanismo adicional de suporte aos tomadores de decisão, pelo fato de incluir na discussão o perfil histórico do preço, frente ao PLD Semana Patamar apresentado pelos modelos oficiais.

Conforme demonstrado neste estudo, os modelos oficiais possuem um viés otimista no longo prazo, portanto, há muito que se evoluir na estratégia apresentada de se basear em um grupo de séries do PLD Semana Patamar para a criação de cenários.

Novos estudos potenciais podem ser desenvolvidos, no intuito de se antever de forma mais consistente a este viés, ao incluir outras variáveis que possam traduzir potenciais crises hídricas à frente, como o ENA e EAR. Apesar dos modelos oficiais incluírem estes indicadores no PLD Semana Patamar, a inclusão de variáveis específicas para este fim pode ser um campo de estudo promissor.

Do ponto de vista da robustez do modelo, com o aumento de amostras inseridas na Base de Aprendizado, o modelo de XGBoost se tornará gradualmente mais abrangente, servindo como importante ferramenta de apoio na tomada de decisão.

Destaque-se que neste trabalho foram utilizados dados e informações públicas acerca da operação centralizada do SIN e do cálculo do PLD. Uma aplicação da metodologia no suporte à decisão comercial no setor elétrico deve levar em conta informações mais detalhadas e precisas, obtidas a partir da atuação das áreas de inteligência do agente de mercado, o que certamente favorecerá a precisão das projeções realizadas.

REFERÊNCIAS BIBLIOGRÁFICAS

ALIM, M.; YE, GUO-HUA; GUAN. P.; HUANG, DE-SHENG; ZHOU, BAO-SEN; WU, W. "Comparison of ARIMA Model and XGBoost Model for prediction of human brucellosis in mainland China: a time-series study." *BMJ Open*, 2020.

ANEEL. Sistema de Informações de Geração da ANEEL (SIGA). Janeiro 2022. <https://bit.ly/2IGf4Q0> (accessed Janeiro 24, 2022).

CEPEL. "Manual de Treinamento - NEWAVE e DECOMP." Rio de Janeiro.

DOS SANTOS, C. R. R.; Castro, R.; MARQUES, R. F. “Aprendizado de máquina no suporte à decisões no mercado de Energia Elétrica.” xxvi SNPTE. Rio de Janeiro, 2021.

JEROME, H. F. “Stochastic Gradient Boosting.” 1999.

RESTON FILHO, J. C.; AFFONSO, C. M.; OLIVEIRA, R. C. L. “Short-Term Energy Price Prediction Multi-Step-Ahead in the Brazilian Market Using Data Mining.” In *Advances in Data Mining Knowledge Discovery and Applications*, 219-234. IntechOpen, 2012.

SKIENA, S. S. *The Data Science Design Manual*. NY: Springer, 2017.

TIANQI, C.; GUESTRIN, C. “XGBoost: A Scalable Tree Boosting System.” University of Washington, 2016.

XGBoost DEVELOPERS. *Introduction to Boosted Trees*. 2021. <https://xgboost.readthedocs.io/en/latest/tutorials/model.html> (accessed November 23, 2021).

ZHANG, L.; BIAN, W.; QU, W.; TUO, L.; WANG, Y. “Time Series forecast of sales volume based on XGBoost.” 2021. *Journal of Physics: Conference Series*.