

PREVISÃO DE CONSUMO DE GÁS NATURAL NO LTQ DA CSN UTILIZANDO MACHINE LEARNING*

*Tatiane Pretto¹
Lucas Gabriel Seibert²
Paulo José dos Reis Pereira³
Lucas Nobumichi Yshii⁴
Mateus Guimarães⁵
Laura Visintainer Lerman⁶
Gabriel Augusto de Avila Santiago⁷
Vinicius Cardilo Campos Alves⁸
Vitor Hugo Maia Netto⁹
Mattheus Vinicius Cordeiro Vellasco¹⁰*

Resumo

Esse trabalho apresenta o desenvolvimento de um modelo de machine learning (ML) para previsão de consumo diário de gás natural (GN) na laminação a quente de aços planos (LTQ) da Companhia Siderúrgica Nacional (CSN). O modelo, baseado no algoritmo CatBoost, incorpora informações de produção planejada, condições operacionais e variáveis temporais. O objetivo é aumentar a assertividade das previsões diárias, reduzindo as penalidades contratuais associadas a desvios superiores a 5% entre o consumo previsto, informado à distribuidora no início do dia, e o consumo total ao final do dia. Para o desenvolvimento do modelo, foram utilizados dados históricos de 2023 a 2025, além de um período de operação assistida com duração de 46 dias. O desempenho do modelo foi avaliado pelas métricas R^2 , WMAPE e MAE, além de indicadores de negócio, como a proporção de dias dentro da faixa de tolerância contratual e a redução das penalidades financeiras. Os resultados foram comparados ao modelo previamente utilizado pela CSN, onde foi observada uma redução de 43% no valor total das penalidades e diminuição de 48% na média diárias dessas penalidades. Os resultados demonstram que a modelagem por ML é capaz de capturar a complexidade do processo de laminação a quente, proporcionando ganhos financeiros e operacionais sob o processo.

Palavras-chave: Gás natural; Machine learning; Laminação a quente; Indústria siderúrgica.

NATURAL GAS CONSUMPTION PREDICTION IN THE LTQ#2 AT CSN USING MACHINE LEARNING

Abstract

This work presents the development of a machine learning model to predict daily natural gas consumption in the hot strip mill of flat steel production at Companhia Siderúrgica Nacional (CSN). The model, based on the CatBoost algorithm, incorporates planned production data, operational conditions, and temporal variables. The goal is to improve the predictions accuracy, reducing contractual penalties associated with deviations over 5% between predicted consumption, reported to the gas distributor at the beginning of the day, and the total consumption at the end of the day. Historical data from 2023 to 2025 were used for model development, along with an assisted operation period of 46 days. Model performance was evaluated using R^2 , WMAPE and MAE metrics, as well as business indicators, such as the days within the contractual tolerance range and the reduction in financial penalties. The results were compared to the previous model used by CSN, showing a 43% reduction in total penalty costs and a 48% decrease in the average daily penalties. The results

demonstrate that ML-based modeling captures the complexity of the hot rolling process, leading to relevant operational and financial gains..

Keywords: Natural gas; Machine learning; Steel industry; Hot strip mill.

- 1 Mestra em Ciência dos Materiais, Cientista de Dados, EvcomX – CSN, Porto Alegre, Rio Grande do Sul, Brasil.
- 2 Mestre em Matemática, Head de Ciência de Dados, EvcomX – CSN, Porto Alegre, Rio Grande do Sul, Brasil.
- 3 MBA em Gestão de Projetos, Product Owner, EvcomX – CSN, Porto Alegre, Rio Grande do Sul, Brasil.
- 4 Mestre em Engenharia Aeronáutica, Cientista de Dados, EvcomX – CSN, Porto Alegre, Rio Grande do Sul, Brasil.
- 5 Físico, Cientista de Dados, EvcomX – CSN, Porto Alegre, Rio Grande do Sul, Brasil.
- 6 Doutora em Engenharia de Produção, Pesquisadora e Cientista de Dados, Dados, EvcomX, Porto Alegre, Rio Grande do Sul e Brasil.
- 7 Mestre em Engenharia Metalúrgica, Engenheiro de Processos Sênior – CSN, Volta Redonda, Rio de Janeiro, Brasil.
- 8 Mestre em Engenharia Metalúrgica, Engenheiro de Processos Sênior – CSN, Volta Redonda, Rio de Janeiro, Brasil.
- 9 Engenheiro de Produção, Engenheiro de Processos Sênior – CSN, Volta Redonda, Rio de Janeiro, Brasil.
- 10 Químico, Especialista em Inovação - CSN, Volta Redonda, Rio de Janeiro, Brasil.

1 INTRODUÇÃO

A indústria siderúrgica é responsável por um elevado consumo energético, sendo os processos térmicos responsáveis por uma parcela significativa dessa demanda.[1,2] Nesse contexto, o gás natural (GN) é uma das principais fontes energéticas em operações de aquecimento, especialmente no processo de reaquecimento de placas na laminação de tiras a quente (LTQ).[2,3] Na Companhia de Siderurgia Nacional (CSN), a laminação de planos é a maior responsável pelo consumo de GN dentro do fluxo produtivo da Usina Presidente Vargas (UPV). No LTQ, o consumo de GN é influenciado por diversos fatores operacionais, como as características do material processado, o desempenho individual dos fornos, as condições de operação (ativo, abafado ou inativo) e as paradas do laminador. Paradas curtas tendem a manter níveis de consumo próximos ao regime normal de operação, enquanto paradas prolongadas reduzem o consumo, mas apresentam uma curva de reaquecimento. Por outro lado, o retorno dos fornos após períodos de inatividade ou manutenção preventiva resulta em aumento do consumo específico de GN devido ao processo de reaquecimento. Ainda, fornos em estado abafado operam a temperaturas inferiores, mantendo consumo reduzido.

Do ponto de vista contratual, a previsão do consumo diário de GN é um fator crítico para a laminação, visto que a imprecisão na previsão acarreta o pagamento de multas e impacta no custo de transformação da bobina laminada. A distribuidora de GN admite um erro absoluto de até 5% em relação ao valor previsto, com tolerância máxima de 15% no transporte desse volume. Previsões fora desses limites implicam em penalidades financeiras. A previsão inicial deve ser enviada às 8h, com possibilidade de revisão até às 14h. Entretanto, a previsão diária de consumo de GN pode ser um desafio, uma vez que depende de informações de produção planejada, sujeitas a alterações ao longo do dia, além de condições operacionais dinâmicas dos fornos e eventuais imprevistos operacionais. Modelos tradicionais, baseados em relações lineares, apresentam limitações em capturar essas variações do processo. Nesse cenário, o desenvolvimento de modelos de previsão capazes de capturar a dinâmica operacional é relevante. Assim, técnicas de machine learning (ML) se mostram adequadas para esse tipo de problema, uma vez que permitem incorporar múltiplas variáveis, incluindo características de produção, condições operacionais e aspectos temporais.[3, 4, 5] Dentre as possíveis abordagens, modelos baseados em árvores de decisão, como o algoritmo CatBoost, apresentam algumas vantagens frente a outros algoritmos, como a capacidade de lidar com variáveis categóricas e modelar relações não-lineares por meio de conjuntos de árvores de decisão, onde a árvore subsequente aprimora o resultado da árvore anterior.[6]

O objetivo desse trabalho é desenvolver um modelo de ML para prever o consumo diário de gás natural na laminação a quente da CSN, com base na produção planejada e nas condições operacionais, visando à redução de possíveis penalidades contratuais em comparação ao modelo previamente utilizado. Para avaliação do desempenho, são comparadas as diferenças entre as previsões e os valores reais de consumo, tanto para o modelo proposto quanto para o modelo anterior utilizado pela CSN, permitindo quantificar os ganhos obtidos em termos operacionais e financeiros.

2 DESENVOLVIMENTO

2.1 Materiais e Métodos

Foram utilizados dados históricos de consumo de GN, obtidos a partir do medidor oficial da planta, com granularidade horária, garantindo consistência com o valor acumulado diário reportado à distribuidora. Ainda, foram incorporados dados de produção planejada, incluindo a quantidade e característica do material a ser processada, tipo de aço e dimensões das placas e bobinas. Para o desenvolvimento do modelo foi considerado o contexto operacional no momento da previsão, realizada inicialmente às 8h. Assim, foram utilizadas como variáveis de entrada tanto informações planejadas quanto dados realizados até o horário da previsão, incluindo consumo acumulado de GN às 8h e produção realizada às 8h. A partir dessas informações, foram construídas variáveis derivadas que representam o estado operacional dos fornos, incluindo indicadores de eficiência energética e produtiva, como consumo por tonelada e consumo por hora. Também foram incorporadas variáveis temporais, com o objetivo de capturar padrões recorrentes no processo, como a ocorrência de manutenções preventivas em determinados períodos da semana. O conjunto dessas variáveis contribuem para a modelagem da variabilidade do consumo de GN ao longo do tempo.

Para desenvolvimento do modelo, os dados foram divididos em conjuntos de treino e teste, sendo o período de janeiro de 2023 a agosto de 2025 para dados de treino do modelo e os meses de setembro e outubro de 2025 para os dados de teste, sendo a variável resposta o consumo diário de GN.[7] Ainda, o modelo teve como base o CatBoost, um algoritmo de ML baseado em ensembles de árvores de decisão, escolhido por sua capacidade em lidar com variáveis categóricas e modelar relações não lineares.[6] Os hiperparâmetros do modelo foram ajustados por meio de validação em conjunto de treino, visando minimizar o erro de previsão. A seleção das variáveis foi realizada com base na importância atribuída pelo modelo e na análise de correlação, visando reduzir redundâncias e melhorar a generalização. Para avaliação do desempenho, foram utilizadas as métricas estatísticas coeficiente de determinação (R^2), raiz do erro quadrático médio (RMSE) e erro percentual absoluto médio ponderado (WMAPE), calculadas pelas Equações 1, 2 e 3, respectivamente.[8] Além disso, foram considerados indicadores de negócio, como a porcentagem de dias dentro da faixa de tolerância contratual (sem incidência de multas) e a economia (em m³) obtida em relação ao modelo previamente utilizado.

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

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

$$WMAPE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i|} \times 100 \quad (3)$$

Onde y_i representa os valores observados, $\hat{y}_i$ os valores preditos pelos modelos, $\bar{y}$ a média dos valores observados e n o número de observações.

2.2 Resultados

2.2.1 Modelo

O desenvolvimento do modelo teve início com a análise exploratória dos dados, buscando compreender a relação entre o consumo de GN e as variáveis operacionais

da laminação. Foram avaliados, por exemplo, o consumo relativo entre fornos e relação com reformas dos fornos e o impacto disso na redução do consumo, tempo entre enformamentos e influência da laminação de materiais a quente na redução do consumo. Essas análises foram validadas junto aos especialistas do processo, contribuindo tanto para a construção do modelo quanto para o entendimento operacional do sistema, o que possibilita melhorias para a área de laminação. A partir dessas análises, foram construídas, ao longo de todo o projeto, 115 variáveis explicativas, das quais 50 foram selecionadas no modelo final. Essas variáveis representam aspectos de temporalidade, produção e eficiência energética e produtiva. O modelo principal foi desenvolvido para o horário das 8h, momento em que deve ser informada a previsão inicial de consumo diário de GN. Ainda, foram desenvolvidos modelos complementares para outros horários do dia, entre 8 e 14h, sendo o último horário o limite da revisão da previsão de GN no dia. Além disso, um modelo específico para às 13h (utilizando informações atualizadas do dia) e um modelo simplificado para previsão do dia seguinte, no qual há uma menor quantidade de variáveis disponíveis juntamente com uma menor assertividade destas para o dia posterior. Também foi desenvolvido um modelo secundário para cenários atípicos, como limitações de produção na linha de laminação e retorno de fornos após manutenção preventiva, que apresentam comportamento distinto do regime normal de operação. O modelo secundário é acionado em cenários identificados como atípicos, com base em padrões operacionais como elevada fração de paradas ou ritmo de produção. Em todos os modelos, as variáveis mais importantes incluem o tempo de parada dos fornos (considerado de forma agregada entre todos os fornos), o consumo acumulado de GN até o momento da previsão e a quantidade de placas programadas para produção, em toneladas. As métricas de desempenho para os subconjuntos de treino e teste são apresentadas na Tabela 1.

Tabela 1: Métricas de desempenho dos modelos das 8 e 13 horas, porcentagem de acerto e de economia em reais

Subconjunto	R ²	WMAPE	RMSE	% acerto	% R\$	Modelo
Treino	0,994	0,023	13,456	89	85	8 horas
Teste	0,982	0,062	21,175	55	50	
Treino	0,990	0,028	16,737	78	84	13 horas
Teste	0,987	0,053	22,757	61	61	

No período de teste, o modelo das 8h apresentou 55% dos dias dentro da tolerância de 5%, ou seja, sem incidência de multa. Além disso, 75% dos dias apresentaram erro inferior a 10% e 88% inferior a 20%. Na Figura 1, são apresentados os erros do modelo no período de teste (pós-modelo), em comparação com os erros do modelo previamente utilizado pela CSN (pré-modelo), calculados sobre o conjunto de dados disponível para seu desenvolvimento. Observa-se uma concentração maior dos erros do pós-modelo em faixas próximas de zero, indicando previsões mais assertivas. Para o modelo das 13h, foi observada melhora no desempenho, refletindo a incorporação de informações operacionais atualizadas ao longo do dia. Em comparação com o modelo anteriormente utilizado pela CSN, foi observada uma redução de, aproximadamente, 41% no erro em volume (m³). Considerando o impacto financeiro associado às penalidades contratuais, essa melhoria corresponde a uma redução de aproximadamente 51% nos custos relacionados às multas.

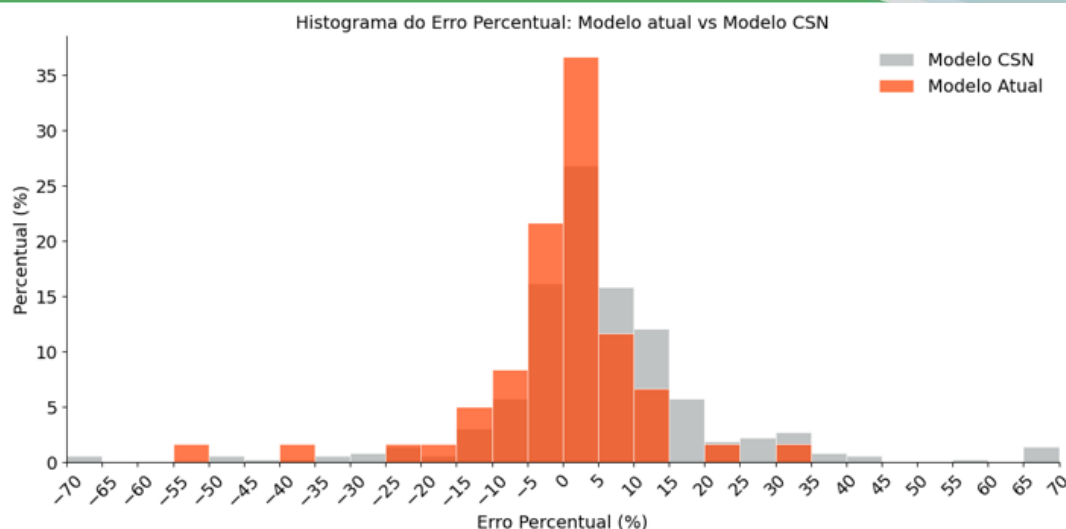


Figura 1: Histograma do erro percentual das previsões de consumo de gás natural, comparando o modelo previamente utilizado pela CSN (pré-modelo) e o modelo atual (pós-modelo).

2.2.2 Operação Assistida

A operação assistida teve duração de aproximadamente dois meses, totalizando 46 dias de previsões realizadas de forma síncrona entre a equipe da EvcomX e a equipe operacional do LTQ da CSN. Nesse período, foram realizadas previsões nos horários de 8h e 13h30, utilizando os respectivos modelos desenvolvidos. Essa etapa teve como objetivo validar a aplicação em ambiente real, bem como promover aprimoramento dos modelos por meio de interações diárias com a equipe operacional. Durante a operação assistida, as previsões do modelo anterior utilizado pelo LTQ foram mantidas, a fim de comparações diretas de desempenho.

Os resultados obtidos durante a operação assistida são apresentados na Tabela 2, bem como na Figura 2. A economia (%) foi calculada com base na redução do erro absoluto em volume (m^3) em relação aos referenciais considerados. No primeiro mês de operação assistida, o processo de laminação apresentou comportamento relativamente estável, sem impactos significativos de paradas não programadas. Nesse cenário, o modelo apresentou taxa de acerto de 68% dos dias dentro da tolerância de 5%, ou seja, sem incidência de multa. Além disso, foi observada uma economia de 70% em relação ao período histórico de referência (jul/24 a jun/25) e de 62% em comparação com as previsões do modelo previamente existente.

Tabela 2: Resultados da operação assistida, considerando porcentagem de acerto e economia de consumo de gás natural em relação ao período histórico de referência (jun/2024 a jul/2025) e ao modelo previamente utilizado pela CSN

Operação Assistida	% acerto	% m^3 vs histórico	% m^3 vs modelo anterior
Fevereiro	68	70	62
Março	20	53	40
Fevereiro a março	43	58	46

Por outro lado, o mês de março foi caracterizado por elevada variabilidade operacional, com ocorrência de paradas não programadas, aumentando o desafio em manter a eficiência e assertividade na previsão de GN. Nesse contexto, pode ser destacado o regime operacional conhecido como “ritmo de uma máquina”, onde mais de 50% dos fornos permanecem ativos, porém, com restrições na linha de laminação,

resultando em redução significativa do ritmo de produção. Esses eventos são considerados cenários atípicos, que, em geral, ocorrem em períodos curtos. Entretanto, durante o período analisado, se estendeu por um período maior que comum. Nesse cenário adverso, o desempenho do modelo foi impactado (Tabela 2), apresentando taxa de acerto de 20% dos dias dentro da tolerância de 5%. Ainda assim, foram observadas economias de aproximadamente 53% em relação ao período de treino, e de 40% em comparação com o modelo previamente utilizado pelo LTQ.

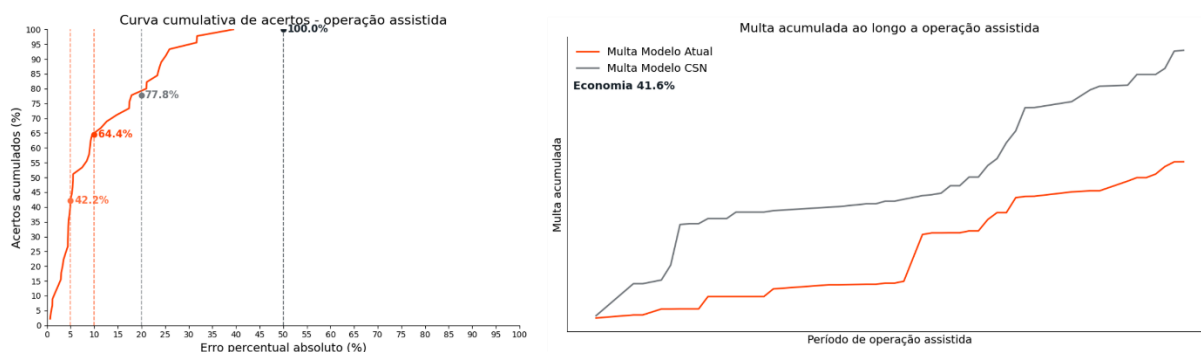


Figura 1: Desempenho do modelo durante operação assistida. À esquerda, curva cumulativa de acertos em função do erro percentual absoluto. À direita, evolução da multa acumulada ao longo do período, comparando o modelo atual com o modelo de referência (CSN).

Apesar do impacto na assertividade no modelo, esse período foi fundamental para a evolução do modelo, possibilitando o desenvolvimento de um modelo secundário voltado a cenários operacionais atípicos. Com a incorporação desse modelo complementar, a taxa de dias dentro da tolerância de 5% aumentou em 8 pontos percentuais, atingindo 28%, enquanto a economia alcançou 62% em relação ao período de treino e 53% em relação ao modelo anterior. Dessa forma, mesmo em condições operacionais diversas, a operação assistida demonstrou a robustez da abordagem proposta e contribuiu diretamente para o aprimoramento do sistema de previsão, especialmente no tratamento de cenários de alta variabilidade.

2.3 Discussão

O modelo anteriormente utilizado pela CSN era baseado em regressão linear múltipla, considerando principalmente a programação de produção diária e os tempos de parada. Nesse contexto, o modelo apresentava 49% dos dias dentro da faixa de tolerância de 5%, ou seja, sem incidência de multas. Dentre as principais limitações observadas, ganham destaque a incapacidade de realizar previsões em cenários com manutenção preventiva, fornos abafados ou inativos, bem como a restrição operacional de prever o consumo de GN apenas quando 3 ou 4 fornos estivessem ativos. O modelo desenvolvido neste trabalho supera essas limitações ao possibilitar previsões em uma maior variedade de cenários operacionais, incluindo condições com menor número de fornos ativos e períodos de manutenção preventiva. Essa maior flexibilidade amplia a aplicabilidade da solução no contexto industrial.

Durante a operação assistida, na qual as previsões foram realizadas em conjunto com a equipe do LTQ – CSN ao longo de 46 dias, foi observada uma redução significativa nas penalidades associadas às previsões de consumo. A média de multa diária foi reduzida em 48%, e o desvio padrão teve redução de 20%, considerando o modelo das 8h, que corresponde à previsão inicial obrigatória e, portanto, a mais

crítica para o processo. Cabe destacar que, para fins de avaliação, foram consideradas também atualizações na programação realizadas ao longo do dia, aplicadas na revisão das previsões, no modelo das 13h. Alterações como mudanças no número de paradas programadas ou na quantidade de placas a serem produzidas foram incorporadas ao modelo das 8h por meio de dados atualizados de planejamento, sem utilização de valores reais. Essa abordagem permite avaliar a capacidade do modelo em refletir corretamente a programação operacional esperada, mesmo diante de ajustes ocorridos após a previsão inicial.

Nesse contexto, o desenvolvimento de múltiplos modelos, específicos para diferentes horários do dia e situações operacionais, foi essencial para capturar a dinâmica operacional do processo. O modelo das 8h é utilizado como referência principal para comparação com o modelo anterior, por se tratar da primeira previsão enviada à distribuidora. Já os modelos posteriores, como o das 13h, permitem incorporar informações atualizadas e aumentar a acurácia da previsão ao longo do dia. Ainda, a utilização de um modelo específico para o período entre 8h e 14h amplia a flexibilidade da solução, possibilitando sua aplicação em horários intermediários, caso necessário. Mesmo em cenários sujeitos a mudanças ao longo do dia, esse conjunto de modelos contribui para uma previsão mais robusta e aderente à realidade operacional.

3 CONCLUSÃO

Com o desenvolvimento do modelo de ML baseado em Catboost, foi possível atingir uma economia de 43% em reais, considerando o período da operação assistida. Ainda, durante esse período, cerca de 42% dos dias ficaram dentro da tolerância de penalidades, ou seja, sem multa alguma. No conjunto de teste, foi observado um R^2 próximo de 0,99 e uma redução de 55% das multas, em termos monetários em comparação aos resultados do modelo prévio utilizado pela CSN.

A abordagem adotada incluiu o desenvolvimento de diversas variáveis, as quais representam o estado operacional do forno, o período da semana em que acontece a previsão, e a produção esperada para o dia. Ainda, além dos modelos principais das 8h e 13h, foi desenvolvido um modelo para o intervalo entre esses horários, e um modelo para cenários atípicos.

A integração entre a modelagem e o conhecimento sobre o processo de laminação a quente discutido ao longo do desenvolvimento do modelo foi essencial para a solução e a economia obtida durante esse período, permitindo melhorias na previsão e maior compreensão operacional.

REFERÊNCIAS

- 1 Carvalho PSL, Mesquita PPD, Araújo EDG. Sustentabilidade da siderurgia brasileira: eficiência energética, emissões e competitividade. Siderurgia, BNDES Setorial, 2015, (41), 181-236.
- 2 International Energy Agency, Iron and Steel Technology Roadmap Towards more sustainable steelmaking, 2020, Disponível em: [Iron and Steel Technology Roadmap – Analysis - IEA](#).
- 3 Zhang X, Yang L, Ma, D. The Importance of Energy Consumption and the Need for Efficiency in the Steel Industry Using Machine Learning. J. Electr. Eng. Technol. 2025, (20), 5097–5112.

- 4 Deba C, Zhangb F, Yanga J, Leea SE, Shaha KW, A review on time series forecasting techniques for building energy. *Renewable and Sustainable Energy Reviews*, 2017, (74), 902-924.
- 5 Eddaoudia Z, Aaraba Z, Boudmena K, Elghazia A, Rahmanian MD. A Brief Review of Energy Consumption Forecasting Using Machine Learning Models. *Procedia Computer Science*, 2024, (236), 33–40.
- 6 Huyen C. *Designing Machine Learning Systems*. Sebastopol: O'Reilly Media; 2022. ISBN 9781098107932. Disponível em: <https://books.google.com.br/books?id=EzhwEAAAQBAJ>.
- 7 Dorogush AV, Gulin A, Gusev G, Kazeev N, Ostroumova Prokhorenkova L, Vorobev A. Fighting biases with dynamic boosting. *CoRR*. 2017; abs/1706.09516. Disponível em: <http://arxiv.org/abs/1706.09516..>
- 8 James G, Witten D, Hastie T, Tibshirani R, Taylor J. *An Introduction to Statistical Learning*, 2023, Springer Cham, ISBN 978-3-031-38747-0. Disponível em: <https://www.statlearning.com/>