MACHINE LEARNING-BASED TAPPING TEMPERATURE PREDICTION IN AN ELECTRIC ARC FURNACE (EAF)

Mateus Guimarães[1]
Pedro Henrique da Silva Rocha[2]
Desiree Thiely Alves Ramalho Costa[3]
Alfredo Raymundo De Magalhães[4]
Lucas Gabriel Seibert[5]
Lucas Siviero Sibemberg[6]
Cristian Hippler[7]
Henrique Magalhães Grossmann[8]
Pedro Ernesto Gomide Silveira[9]

Abstract

This work presents an industrial application of machine learning techniques for predicting tapping temperature in Electric Arc Furnaces (EAF), carried out at the Piracicaba plant of ArcelorMittal Brasil, aiming to reduce thermal variability and improve energy efficiency in the steelmaking process. The proposed methodology is based on two supervised regression models: a thermal loss model (PDT), responsible for estimating the optimal tapping temperature, and an energy consumption model (CSE), used to define and dynamically adjust the energy setpoint during operation. The models were developed using industrial data from automation systems, MES, and PIMS, structured at both heat-level and measurement-level datasets, with data preprocessing and feature engineering techniques applied to handle missing and noisy data. The results showed an increase of approximately 30% in the percentage of heats within the ±10 °C range and a reduction of about 22% in average energy error per heat, along with reduced thermal variability, indicating improvements in process stability and predictability.

Keywords: Electric Arc Furnace; Machine Learning; Tapping Temperature; Energy Efficiency.

1. *Physicist, Data Scientist, Data & Analytics, EvcomX, Porto Alegre, Rio Grande do Sul, Brazil.*
2. *Metallurgical Engineer, M.Sc., Process Specialist, Steelmaking, ArcelorMittal Longos, Piracicaba, São Paulo, Brazil.*
3. *Statistician, Postgraduate Degree, Data Scientist, Data & Analytics, ArcelorMittal Longos, Belo Horizonte, Minas Gerais, Brazil.*
4. *Systems Analyst, Product Manager, Digital Analytics, ArcelorMittal Longos, Volta Redonda, Rio de Janeiro, Brazil.*
5. *Mathematician, M.Sc., Head of Data Science, Data & Analytics, EvcomX, Porto Alegre, Rio Grande do Sul, Brazil.*
6. *Mathematician, Ph.D. Candidate, CTO, Data & Analytics, EvcomX, Porto Alegre, Rio Grande do Sul, Brazil.*
7. *Systems Analyst, Postgraduate Degree, Business Owner, Data & Analytics, EvcomX, Porto Alegre, Rio Grande do Sul, Brazil.*
8. *Systems Analyst, Data Engineer, Data & Analytics, EvcomX, Porto Alegre, Rio Grande do Sul, Brazil.*
9. *Electrical Engineer, Ph.D. Candidate, Group Product Manager, Digital & Analytics, ArcelorMittal Longos, Belo Horizonte, Minas Gerais, Brazil.*

1 INTRODUCTION

Steel production in Electric Arc Furnaces (EAF) is characterized by high thermodynamic and kinetic complexity, involving multiple interactions among operational variables, metallic charge composition, and energy transfer. In this context, tapping temperature is a critical parameter, directly influencing energy efficiency, metallurgical quality, and process stability, since thermal deviations directly affect scrap melting and slag-metal equilibrium [1]. Traditionally, the control of the electrical energy applied to the EAF is based on operator experience, which introduces subjectivity and results in significant operational variability [1]. This variability is evidenced by the dispersion of tapping temperature, with a relevant share of heats falling outside the ideal operating range [1]. The process also exhibits high sensitivity to operational variables, such as metallic charge and power profile, reinforcing its nonlinear behavior [2]. In parallel, industrial data quality represents a relevant challenge due to the presence of missing values, noise, and inconsistencies, which directly affects the application of machine learning techniques. In this context, preprocessing and feature engineering steps become essential to ensure model robustness. Furthermore, understanding thermodynamic phenomena remains fundamental for controlling the steelmaking process [3]. Given this scenario, this work aims to present a machine learning-based solution for predicting tapping temperature in an EAF, applied to ArcelorMittal's Piracicaba plant, considering the challenges associated with process variability and industrial data quality.

2 DEVELOPMENT

2.1 MATERIALS

The data used in this study were obtained from the industrial automation systems of the Electric Arc Furnace (EAF), as well as from the MES (Manufacturing Execution System) and PIMS (Process Information Management System) systems of the Piracicaba plant. The information was structured into two distinct levels: heat level and measurement level. At the heat level, variables related to the production context of each heat were considered, including historical and planned process information, such as metallic charge profile, steel grade group, and recent performance statistics, including means, standard deviations, and medians. These variables allow the overall behavior of the process to be captured. At the measurement level, dynamic variables obtained throughout the EAF stage were incorporated, such as natural gas, oxygen, and electrical energy consumption, as well as molten bath temperature measurements. This dataset allows the instantaneous state of the process to be represented. The solution was based on the development of two models, whose datasets were specifically structured according to the phenomenon being modeled. For the thermal loss model, predominantly aggregated heat-level variables were used, since this phenomenon is associated with the global conditions of the process. For the energy consumption model, heat-level and measurement-level variables were combined, enabling the operational dynamics throughout the EAF stage to be captured and increasing sensitivity to real-time conditions.

2.2 METHODS

To ensure conformity to the operational practices of the Electric Arc Furnace (EAF), the proposed solution is based on two supervised learning regression models, referred to as PDT (Thermal Loss) and CSE (Energy Consumption). The PDT model aims to estimate the thermal loss between EAF tapping and the ladle furnace (LF) stage, enabling the determination of the optimal tapping temperature. The modeling approach considers operational variables and recent historical values, including alloy additions and statistical indicators of thermal losses. The CSE model is responsible for estimating the energy consumption required to reach the temperature defined by the PDT model, considering the total process energy consumption, including both electrical and chemical energy. Among the information sources used, charge distribution, process time, number of planned stops, and derived metrics stand out, including relationships between yield and energy efficiency.

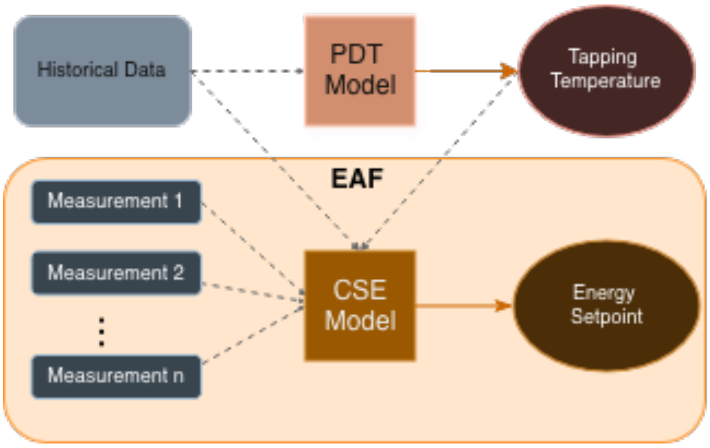

Figure 1. Schematic representation of the two-model solution, PDT and CSE.

First, the PDT model defines the target tapping temperature based on historical process data. Then, the CSE model estimates the energy setpoint required to reach this condition. During EAF operation, successive temperature measurements are incorporated into the CSE model, which iteratively recalculates the energy deficit and dynamically adjusts the setpoint. The models were trained using a split into training and test sets, supporting generalization assessment and reducing overfitting [4]. For the modeling stage, boosting-based algorithms were employed, as they are recognized for their robustness in problems involving high nonlinearity and noisy data [4][5]. The adopted modular approach allows the modeling of thermal loss to be separated from the furnace energy dynamics, resulting in greater robustness and applicability in an industrial environment. Based on the machine learning metrics [4], Table 1 was constructed:

Tabela 1. Model metrics by training and test sets.

Model	Dataset	R^2	RMSE	MAE
PDT	Train	0,98	2,76 (°C)	2,20 (°C)
PDT	Test	0,62	10,82 (°C)	9,10 (°C)
CSE	Train	0,92	0,413 (kwh)	0,23 (kwh)
CSE	Test	0,82	0,636 (kwh)	0,36 (kwh)

The historical dataset used for training comprised 333 heats, carried out between June and September 2025, while the test set consisted of 127 heats, corresponding to October 2025. Both datasets were constructed after applying specific filters to maintain the coherence of the relationships among the variables. The operational evaluation of the solution considered the results observed between October 2025 and March 2026. Additionally, for the specific analysis of adherence to the model recommendations, model-adherent heats were defined as those in which the absolute difference between actual energy consumption and the recommended energy setpoint was less than 500 kWh. To support the interpretation of the PDT model, Figure 2 presents the feature importance plot for a subset of variables, highlighting the relative contribution of the main inputs used by the model.

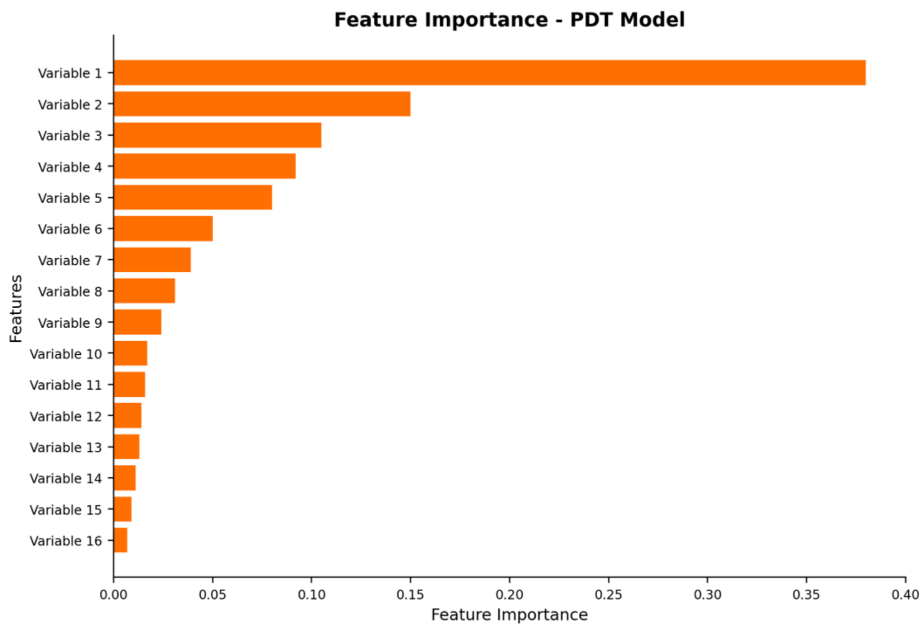

Figure 2. Feature Importance plot for a subset of variables in the PDT model.

2.3 RESULTS

The proposed solution was evaluated by comparing operational indicators before and after the implementation of the PDT and CSE models, considering the overall process performance. An increase of approximately 30% was observed in the percentage of heats within the ±10 °C range, rising from 27% to 35%. Consistently, an increase of approximately 18% was also observed in the more restrictive ±5 °C range, which increased from 17% to 20%. These results indicate greater adherence of the tapping temperature to the process target values. In terms of energy performance, an approximately 22% reduction was observed in the average energy consumption error per heat, indicating reduced energy waste and greater efficiency in the use of electrical energy. The analysis of the absolute thermal error at LF arrival indicated an approximately 10% reduction in the average error, associated with a reduction of about 13% in the standard deviation. These results indicate a simultaneous improvement in thermal accuracy and process stability, which can also be observed in the data distribution presented in Figure 3, where a higher concentration of heats near the target value and a reduction in thermal dispersion can be observed.

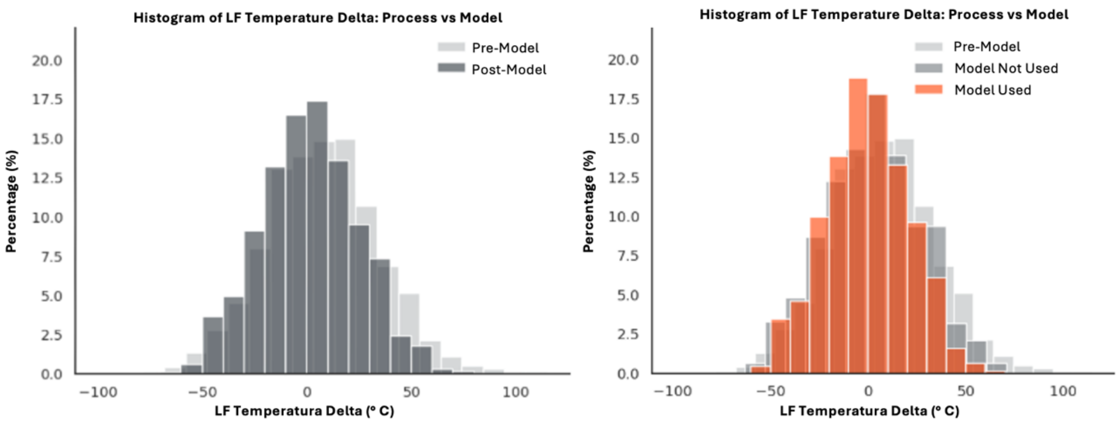

Figure 3. LF arrival temperature delta results comparing the pre- and post-model periods (left) and the comparison between Model Used and Model Not Used heats (right).

When exclusively analyzing the metrics associated with the model recommendations throughout the operational period, an even more expressive performance is observed, as evidenced in Figure 3, which presents the distribution of heats segregated between process operation and model use, indicating a higher concentration around the target value when there is commitment to the recommendations. The percentage of heats within the ±10 °C range relative to the target temperature value increased by approximately 41%, while the ±5 °C range showed an increase of about 24%. In terms of energy savings, an approximately 28% reduction was observed in the average energy consumption error per heat. Additionally, the absolute thermal error showed a reduction of about 22% in the average value, accompanied by an approximately 24% decrease in the standard deviation, indicating high consistency in the estimates generated by the models.

2.4 DISCUSSION

The reduction in thermal dispersion contributes to greater operational stability and predictability, positively impacting process control and product quality. This predictability enables better adjustment of subsequent stages, especially in the ladle furnace, reducing the need for thermal corrections and improving the consistency of refining conditions. From a metallurgical perspective, greater thermal stability may contribute to metallic yield gains by reducing losses associated with overheating, excessive oxidation, and additional operational adjustments. Furthermore, lower variability favors process repeatability, a critical aspect in high-productivity industrial environments. These results are aligned with studies that highlight the importance of thermal control and the understanding of thermodynamic phenomena for optimizing the steelmaking process and improving overall efficiency [2].

3 CONCLUSION

The application of machine learning techniques to tapping temperature prediction in Electric Arc Furnaces, in the context of ArcelorMittal's Piracicaba plant, proved effective in reducing thermal variability and supporting operational decision-making, contributing to improved process energy efficiency. It is worth noting that the development of the solution was conducted in a scenario with relevant limitations in industrial data quality, including the presence of missing values, noise, and inconsistencies across different data sources. Even so, through preprocessing steps and the construction of derived variables based on process knowledge, it was possible to obtain models with good predictive performance. The results showed consistent gains in the main operational metrics, including increased adherence to the target temperature, reduced energy error, and lower thermal dispersion. These advances directly reflect improvements in process control, greater operational stability, and increased predictability of production conditions. The observed performance is directly associated with the integration between model recommendations and operational decision-making, highlighting the importance of adherence to the proposed system for achieving the reported results. In the context of electric steelmaking, where variability is inherent to the process, the reduction in standard deviation represents a significant gain, contributing to greater repeatability, improved performance in subsequent stages, and potential increases in metallic yield. Therefore, it is concluded that the integration of metallurgical knowledge and machine learning techniques constitutes a robust and applicable approach in industrial environments, even under data quality constraints. The observed results indicate that the process is already showing consistent progress toward the optimal conditions estimated by the models, demonstrating that the assisted use of these tools contributes to the continuous improvement of the operation. As future perspectives, improvements in data acquisition systems and model evolution are recommended, including hybrid approaches that integrate physical and statistical foundations.

REFERENCES

1. Choi SW, Lee EB, Kim JH. Development of a machine learning-based tapping temperature prediction model for electric arc furnace. Sustainability. 2023;15:6393.
2. Diniz PHA. Análise do consumo energético de forno elétrico a arco voltaico utilizado em indústria metalúrgica nacional. Universidade Estadual Paulista. 2018.
3. Heck NC, Vilela ACF. Aplicação da termodinâmica computacional na simulação da produção de aço inoxidável em forno elétrico a arco. In: 36º Seminário de Fusão, Refino e Solidificação dos Metais; 2005; Vitória. São Paulo: ABM; 2005.
4. Huyen C. *Designing Machine Learning Systems*. Sebastopol: O'Reilly Media; 2022. ISBN 9781098107932. Disponível em: https://books.google.com.br/books?id=EzhwEAAAQBAJ.
5. Dorogush AV, Gulin A, Gusev G, Kazeev N, Ostroumova Prokhorenkova L, Vorobev A. Fighting biases with dynamic boosting. CoRR. 2017 [acesso em 22 abr. 2026];abs/1706.09516. Disponível em:https://arxiv.org/abs/1706.09516.