



**UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
CURSO DE ENGENHARIA QUÍMICA**

**RAYSSA CRISTINA DO NASCIMENTO SILVA**

**MODELAGEM COMPUTACIONAL EM PYTHON PARA MINIMIZAÇÃO DA PERDA  
DE FUSÃO NO PROCESSO INDUSTRIAL DE TRANSFORMAÇÃO DO ALUMÍNIO  
UTILIZANDO TÉCNICAS DE MACHINE LEARNING**

**RECIFE**

**2025**

**UNIVERSIDADE FEDERAL DE PERNAMBUCO**  
**CENTRO DE TECNOLOGIA E GEOCIÊNCIAS**  
**DEPARTAMENTO DE ENGENHARIA QUÍMICA**  
**CURSO DE ENGENHARIA QUÍMICA**

**RAYSSA CRISTINA DO NASCIMENTO SILVA**

**MODELAGEM COMPUTACIONAL EM PYTHON PARA MINIMIZAÇÃO DA PERDA  
DE FUSÃO NO PROCESSO INDUSTRIAL DE TRANSFORMAÇÃO DO ALUMÍNIO  
UTILIZANDO TÉCNICAS DE MACHINE LEARNING**

TCC apresentado ao Curso de Engenharia Química da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito para a obtenção do título de bacharel em Engenharia Química.

**Orientador(a):** Fábio Machado Cavalcanti

**RECIFE**

**2025**

Ficha de identificação da obra elaborada pelo autor,  
através do programa de geração automática do SIB/UFPE

Silva, Rayssa Cristina do Nascimento.

Modelagem computacional em python para minimização da perda de fusão no processo industrial de transformação do alumínio utilizando técnicas de machine learning / Rayssa Cristina do Nascimento Silva. - Recife, 2025.

70 p. : il., tab.

Orientador(a): Fábio Machado Cavalcanti

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Engenharia Química - Bacharelado, 2025.

Inclui referências.

1. Python. 2. Machine Learning. 3. Alumínio. 4. Minimização. 5. Escória. I. Cavalcanti, Fábio Machado . (Orientação). II. Título.

660 CDD (22.ed.)

RAYSSA CRISTINA DO NASCIMENTO SILVA

**MODELAGEM COMPUTACIONAL EM PYTHON PARA MINIMIZAÇÃO DA PERDA DE FUSÃO NO PROCESSO INDUSTRIAL DE TRANSFORMAÇÃO DO ALUMÍNIO UTILIZANDO TÉCNICAS DE MACHINE LEARNING**

TCC apresentado ao Curso de Engenharia Química da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito para a obtenção do título de bacharel em Engenharia Química.

Aprovado em: 05/08/2025.

**BANCA EXAMINADORA**

Documento assinado digitalmente  
 **FABIO MACHADO CAVALCANTI**  
Data: 13/08/2025 10:13:55-0300  
Verifique em <https://validar.iti.gov.br>

---

Profº. Dr. Fábio Machado Cavalcanti (Orientador)  
Universidade Federal de Pernambuco

Documento assinado digitalmente  
 **FELIPE PEDRO DA COSTA GOMES**  
Data: 15/08/2025 07:47:40-0300  
Verifique em <https://validar.iti.gov.br>

---

Profº. Dr. Felipe Pedro da Costa Gomes (Examinador Interno)  
Universidade Federal de Pernambuco

Documento assinado digitalmente  
 **HEZROM SAULO DO NASCIMENTO JUNIOR**  
Data: 13/08/2025 13:37:59-0300  
Verifique em <https://validar.iti.gov.br>

---

Profº. MSc. Hezrom Saulo do Nascimento Júnior (Examinador Externo)  
Universidade Federal de Pernambuco

A meu amado avô Antônio (in memoriam) que fez de meu sonho o seu sonho.

## AGRADECIMENTOS

Porque em todo o trajeto que se caminha com alguém cada passo é celebrado e nenhum obstáculo é insuperável. Agradeço aos meus pais Elaine e Ronaldo por todo o esforço, dedicação e apoio, a quem tornou toda essa experiência possível, a quem nunca me deixou faltar nada, a quem dedico todo e qualquer fruto dessa trajetória acadêmica. Agradeço ao meu amado avô Antônio (in memoriam) que fez de meu sonho o seu sonho, quem tanto me apoiou e me incentivou.

A minha avó Maria das Dores por todo o carinho e dedicação. Ao meu irmão Ronald por todo o apoio e ajuda. Aos meus avós Zezinho e Geiza, a minha tia Liparli, meu tio Marcos, meu tio Elisson e a toda a minha família que tanto acreditou em mim, agradeço por toda resiliência e motivação. Agradeço a minha namorada Francieli que tanto veio para somar, agradeço por todo apoio incondicional, por todo o incentivo e por sempre acreditar que eu poderia mais do que imaginava.

Agradeço também aos amigos que tiveram comigo enquanto tudo ainda era sonho e que tanto acreditaram que a consagração dessa trajetória poderia acontecer. Agradeço em especial a Layse e Ricardo.

Aos amigos que chegaram e fizeram dessa trajetória um caminho mais fácil de se percorrer. Agradeço a Maria Eduarda, Taciana, Moises, Laura, Iasmin, Alice e Thiago que tanto somaram nessa extensa equação que é a graduação, obrigada por todo o apoio e ajuda, sem vocês seria tudo mais difícil.

Agradeço ao meu orientador por todo apoio, dedicação e paciência para que esse projeto fosse bem executado em cada mínimo detalhe. Agradeço a San Diego que embarcou na ideia desse projeto e me deu todo auxílio necessário para que este fosse possível.

Agradeço a todo mundo que fez parte indiretamente e diretamente desse rico percurso que tanto me construiu. Agradeço a Paulo meu tutor de iniciação científica, a quem me ensinou a caminhar na vida acadêmica. Agradeço a Joana, minha segunda tutora, por todo auxílio e ajuda.

Agradeço, por fim, a espiritualidade que tanto me guia e está sempre comigo, pois sou uma, mas não ando só. Gratidão aos meus orixás, pois sem eles absolutamente nada disso seria possível e nada disso faria sentido.

## RESUMO

A perda de fusão, dada pela oxidação do alumínio e sua subsequente formação de escória durante o processo de fusão, representa um desafio relevante na indústria de alumínio, sobretudo em razão dos gastos econômicos e energéticos das perdas e do rendimento metálico. Com base nisso, o presente estudo teve como objetivo desenvolver um programa computacional em Python, utilizando técnicas de *Machine Learning*, para analisar e prever a perda de fusão com base em dados operacionais industriais. A metodologia consistiu na coleta e tratamento de dados de diferentes carregamentos de fusão, aplicação de modelos estatísticos e preditivos (Regressão Linear, Random Forest, Gradient Boosting e Support Vector Regression), testados através do parâmetro  $R^2$ , para avaliar o impacto de variáveis como tempo e temperatura de carregamento e retirada da escória na perda de fusão e simulação de cenários de melhores condições de processo. O modelo Random Forest, selecionado por seu alto desempenho preditivo ( $R^2 > 0,9$ ), demonstrou excelente capacidade de identificar interações complexas entre variáveis e apontar condições operacionais ideais. Os testes indicaram que os ajustes nas variáveis quando dado de maneira individual pode reduzir a perda de fusão em até 12%, enquanto a implementação combinada das recomendações geradas pelo sistema pode alcançar reduções superiores a 20%. Por fim, a integração entre ciência de dados com a engenharia química, por meio da modelagem computacional, mostrou-se como um elemento significativo para a otimização no processo de fusão, oferecendo uma ferramenta que pode ser bastante eficaz para redução de perdas metálicas.

**Palavras-chave:** Python; *Machine Learning*; Alumínio; Minimização; Escória.

## ABSTRACT

Fusion loss, caused by aluminum oxidation and the subsequent formation of dross during the melting process, represents a significant challenge in the aluminum industry, mainly due to the economic and energy-related costs of these losses and their impact on metal yield. Based on this, the present study aimed to develop a computational program in Python, using Machine Learning techniques, to analyze and predict fusion loss based on industrial operational data. The methodology involved the collection and preprocessing of data from different melting batches, the application of statistical and predictive models (Linear Regression, Random Forest, Gradient Boosting, and Support Vector Regression), tested using the  $R^2$  metric, to evaluate the impact of variables such as loading time and temperature, as well as dross removal, on fusion loss, and the simulation of scenarios with improved process conditions. The Random Forest model, selected for its high predictive performance ( $R^2 > 0.9$ ), demonstrated excellent capability in identifying complex interactions between variables and in pointing out optimal operational conditions. Tests indicated that adjusting individual variables could reduce fusion loss by up to 12%, while the combined implementation of the system's recommendations could lead to reductions greater than 20%. Finally, the integration of data science with chemical engineering through computational modeling proved to be a significant element in optimizing the melting process, offering a highly effective tool for reducing metal losses.

**Keywords:** Python; Machine Learning; Aluminum; Minimization; Dross.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> –Gráfico de Importância de Variável do Forno 1. ....	43
<b>Figura 2</b> –Gráfico de Importância de Variável do Forno 2. ....	44
<b>Figura 3</b> –Gráfico de Importância de Variável do Forno 3. ....	45
<b>Figura 4</b> –Gráfico de Dependência Parcial da Temperatura de pré-escumagem do Forno 1. ....	46
<b>Figura 5</b> –Gráfico de Dependência Parcial da Temperatura de pré-escumagem do Forno 2. ....	47
<b>Figura 6</b> –Gráfico de Dependência Parcial da Temperatura de pré-escumagem do Forno 3. ....	48
<b>Figura 7</b> –Gráfico de Dependência Parcial da Temperatura de carregamento do Forno 1. ....	50
<b>Figura 8</b> –Gráfico de Dependência Parcial da Temperatura de carregamento do Forno 2. ....	51
<b>Figura 9</b> –Gráfico de Dependência Parcial da Temperatura de carregamento do Forno 3. ....	52
<b>Figura 10</b> –Gráfico de Dependência Parcial do Tempo para pré-escumagem (h) Forno 1. ....	53
<b>Figura 11</b> –Gráfico de Dependência Parcial do Tempo para pré-escumagem (h) Forno 2. ....	54
<b>Figura 12</b> –Gráfico de Dependência Parcial do Tempo para pré-escumagem (h) Forno 3. ....	55
<b>Figura 13</b> –Gráfico de Dependência Parcial do Tempo de carregamento (min) Forno 1. ....	56
<b>Figura 14</b> –Gráfico de Dependência Parcial do Tempo de carregamento (min) Forno 2. ....	57
<b>Figura 15</b> –Gráfico de Dependência Parcial do Tempo de carregamento (min) Forno 3. ....	58
<b>Figura 16</b> –Melhores e Piores Cenários para a Perda de Fusão Forno 1. ....	59
<b>Figura 17</b> –Melhores e Piores Cenários para a Perda de Fusão Forno 2. ....	60
<b>Figura 18</b> –Melhores e Piores Cenários para a Perda de Fusão Forno 3. ....	61
<b>Figura 19</b> – Redução da Perda de Fusão (%) com alteração da Variável Isolada e com Todas as Variáveis. ....	63

## LISTA DE ABREVIações

PF	Perda de Fusão
RBF	Função de Base Radical
SVR	Regressão de Vetores de Suporte
MSE	Erro Quadrático Médio
VIF	Fator de Inflação da Variância

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>12</b>
<b>2 REVISÃO DE LITERATURA</b> .....	<b>16</b>
<b>2.1 O alumínio</b> .....	<b>16</b>
<b>2.1.1 A indústria primária</b> .....	<b>17</b>
<b>2.1.2 A indústria secundária</b> .....	<b>18</b>
<b>2.2 Transformação do Alumínio</b> .....	<b>20</b>
<b>2.3 A escória de Alumínio</b> .....	<b>21</b>
<b>2.4 Machine Learning</b> .....	<b>23</b>
<b>2.4.1 Modelos Estatísticos e de Machine Learning</b> .....	<b>24</b>
<b>2.4.1.1 Regressão Linear</b> .....	<b>24</b>
<b>2.4.1.2 Random Forest</b> .....	<b>25</b>
<b>2.4.1.3 Gradient Boosting</b> .....	<b>26</b>
<b>2.4.1.4 Support Vector Regression(SVR)</b> .....	<b>27</b>
<b>2.5 Linguagem de Programação Python</b> .....	<b>29</b>
<b>2.6 Bibliotecas Python Aplicadas à Modelagem Preditiva</b> .....	<b>30</b>
<b>2.6.1 Pandas</b> .....	<b>30</b>
<b>2.6.2 NumPy</b> .....	<b>30</b>
<b>2.6.3 Seaborn</b> .....	<b>31</b>
<b>2.6.4 Scikit-learn</b> .....	<b>32</b>
<b>3 METODOLOGIA</b> .....	<b>34</b>
<b>4 RESULTADOS</b> .....	<b>42</b>
<b>5 CONCLUSÃO</b> .....	<b>64</b>
<b>REFERÊNCIAS</b> .....	<b>66</b>

## 1 INTRODUÇÃO

Nas últimas décadas, o alumínio consolidou-se globalmente como um dos metais de maior importância e crescimento industrial, impulsionado principalmente por propriedades como baixa densidade, elevada resistência à corrosão, alta condutividade térmica e elétrica, além de sua ampla capacidade de ser reciclado sem perda significativa de sua qualidade (ABAL, 2023; Calegari et al., 2023). Atualmente, o alumínio é o metal não ferroso mais consumido do mundo e a demanda de seus produtos provém especialmente dos setores automotivo, aeronáutico, construção civil e embalagens (Statista, 2023). Esses setores exigem que os processos de produção e reaproveitamento do alumínio sejam continuamente aperfeiçoados, buscando maior eficiência, redução de custos e, sobretudo, sustentabilidade ambiental.

A cadeia produtiva do alumínio é dividida em dois segmentos: o primário e o secundário (Subramanian; Sahu, 2021). O segmento primário corresponde à extração da bauxita e à produção do alumínio metálico através de processos intensivos que demandam alto consumo de energia, como o beneficiamento pelo método Bayer e a subsequente eletrólise ígnea, realizada pelo processo Hall-Héroult (Callister Jr.; Rethwisch, 2020). Em contraste, a produção secundária, também denominada reciclagem de alumínio, baseia-se na fusão de sucatas e lingotes para a reintegração do metal na cadeia produtiva (ASM International, 1993; Schlesinger et al., 2011). Esta via destaca-se pela significativa eficiência energética, uma vez que necessita de apenas cerca de 5% da energia necessária para a produção do alumínio primário, resultando em benefícios ambientais expressivos, como a mitigação das emissões de gases de efeito estufa e a redução da extração de recursos naturais (Calegari et al., 2023; Haraldsson; Johansson, 2018).

Um dos principais desafios encontrado na indústria de alumínio secundário é a perda de fusão, representada pela formação de escória — um subproduto inevitável decorrente da oxidação do alumínio (Freti; J. D., Bornand; K. Buxmann, 1982). A geração de escória nos processos de fusão do alumínio varia conforme a matéria-prima utilizada, os parâmetros operacionais e o controle térmico aplicado durante a produção. Essa escória retém contaminantes, sais e compostos sólidos, o que torna difícil sua recuperação e beneficiamento. A sua geração excessiva

compromete o aproveitamento metálico, tornando-se um indicador essencial da eficiência do processo de fusão (Freti; Bornand; Buxmann, 1982).

Além de refletir diretamente no rendimento metálico do alumínio, o excesso de escória pode aumentar os custos operacionais, devido à necessidade de reposição de material e ao maior consumo energético (Davis, 1993; Totten, 2006). Logo, o controle rigoroso desses fatores é fundamental para otimizar a recuperação do metal e aumentar eficiência produtiva permitindo que a produção secundária de alumínio permaneça no eixo de eficiência energética (ABAL, 2023; Davis, 1993).

Nos últimos anos, o uso de técnicas de Machine Learning tem se destacado como uma abordagem eficaz para a modelagem e otimização de processos complexos (Jordan; Mitchell, 2015; Zhang et al., 2020). De acordo com Domingos (2012), essa área fundamenta-se em três pilares essenciais: representação, avaliação e otimização. A representação define a estrutura do modelo utilizado para identificar padrões nos dados; a avaliação estabelece critérios que diferenciam modelos eficazes dos ineficazes; e a otimização busca os melhores parâmetros para maximizar o desempenho. Essa estrutura permite a construção de modelos robustos, capazes de se adaptar a novos dados — característica fundamental em contextos marcados por alta variabilidade e incerteza.

No setor industrial, essas técnicas vêm sendo amplamente aplicadas para reduzir perdas, prever falhas e aumentar a eficiência dos processos produtivos. Segundo a KPMG (2023), a automação baseada em inteligência artificial e aprendizado de máquina possibilita tomadas de decisão mais rápidas e assertivas, com base em dados em tempo real. Essa capacidade tem impulsionado transformações significativas no escopo da Indústria 4.0.

Modelos de *Machine Learning* como *Random Forest* (RF), *Gradient Boosting* (GB) e *Support Vector Regression* (SVR) são amplamente empregados em tarefas de modelagem e previsão (Hastie; Tibshirani; Friedman, 2009; Chen; Guestrin, 2016). Os métodos de *Random Forest* e o *Gradient Boosting*, ambos baseados em árvores de decisão, oferecem maior capacidade preditiva: o primeiro reduz o overfitting por meio da combinação de múltiplas árvores (Breiman, 2001), enquanto o segundo aprimora o desempenho ajustando erros sequencialmente (Friedman, 2001). O SVR, por sua vez, busca uma função de regressão com erro controlado, sendo eficaz para dados complexos e de alta dimensionalidade (Vapnik, 1995).

A literatura evidencia que o desempenho dos modelos de *Machine Learning* é altamente sensível às características do conjunto de dados e ao tipo de problema analisado. Modelos que se destacam em certos contextos podem apresentar resultados inferiores em outros. Por exemplo, Cenita et al. (2023) mostraram que o *Gradient Boosting* teve o melhor desempenho na previsão de custos, enquanto Bargam et al. (2023) observaram superioridade do SVR em cenários com dados escassos. Em um estudo aplicado, Liaw e Wiener (2002) demonstraram a eficácia do *Random Forest* para problemas de classificação com múltiplas variáveis, destacando sua robustez frente a dados complexos. Esses resultados reforçam que não existe um modelo universalmente superior; assim, uma abordagem com múltiplos modelos torna-se imprescindível para explorar as possibilidades preditivas dos dados e assegurar decisões analíticas mais robustas e fundamentadas no processo em estudo (Sagi; Rokach, 2018; Zhou, 2021).

Nesse contexto, o uso de linguagens de programação como Python se destaca, pela vasta disponibilidade de bibliotecas especializadas que facilitam a implementação e experimentação dos variados modelos preditivos de forma acessível e escalável (Pedregosa et al., 2011). A combinação entre modelos inteligentes e programação eficiente promove a otimização contínua e autônoma dos processos, contribuindo para uma maior eficiência do processo, minimização de perda e o aumento da previsibilidade operacional (Lu; Morris; Frechel, 2016). Assim, o uso de Machine Learning, apoiado por ferramentas computacionais robustas, consolida-se como uma estratégia em potencial para enfrentar os desafios da produção industrial.

Diante disso, o presente trabalho visou desenvolver um programa em Python fundamentado em técnicas de *Machine Learning* para avaliar e compreender os fatores que influenciam a perda de fusão no processo industrial de transformação do alumínio. O programa se baseia na aplicação de diversos modelos preditivos, no qual foi selecionado o que apresentou o melhor desempenho. Com o modelo escolhido, foram realizadas análises estatísticas multivariadas e modelagem preditiva para identificar as variáveis relacionadas ao processo produtivo que mais influenciam a perda de fusão. Esse procedimento visou prever e minimizar a perda de fusão na transformação do alumínio, permitindo promover a otimização do processo.

## 1.1 Objetivos Gerais

O presente trabalho teve como objetivo desenvolver um programa em Python fundamentado em técnicas de *Machine Learning* para avaliar e compreender os fatores que influenciam a perda de fusão no processo industrial de transformação do alumínio.

## 1.2 Objetivos Específicos

- Avaliar diversos modelos preditivos, de *Machine Learning* e Estatística, para selecionar o que melhor se enquadra nos dados de campo.
- Realizar, com o modelo escolhido, análises estatísticas multivariadas e modelagem preditiva para identificar as variáveis relacionadas à matéria-prima e ao processo produtivo que mais influenciam a perda de fusão.
- Utilizar *Machine Learning* para avaliar os cenários que minimizam a perda de fusão.
- Estudar em escala real as modificações sugeridas pelo programa.

## 2 REVISÃO DE LITERATURA

### 2.1 O alumínio

O alumínio é um metal fundamental para a indústria contemporânea, reconhecido por suas propriedades físico-químicas únicas que o diferenciam de outros metais. Sua baixa densidade, aproximadamente  $2,7 \text{ g/cm}^3$ , proporciona leveza sem comprometer a resistência mecânica, enquanto a formação espontânea de uma fina e aderente camada de óxido de alumínio ( $\text{Al}_2\text{O}_3$ ) confere alta resistência à corrosão, protegendo o metal contra ataques químicos e intempéries (Davis, 1993). Além disso, o alumínio apresenta boa condutividade térmica (cerca de  $235 \text{ W/m}\cdot\text{K}$ ) e elétrica (equivalente a 60% da do cobre), mas com peso três vezes menor, o que o torna muito valorizado em aplicações elétricas e térmicas (Asm International, 1990).

Sua elevada maleabilidade e ductilidade permitem que o alumínio seja facilmente conformado por processos industriais como laminação, extrusão, forjamento e estampagem, possibilitando a fabricação de produtos com geometrias complexas e espessuras variadas (Callister; Rethwisch, 2021). A excelente refletividade óptica, refletindo cerca de 90% da radiação visível e infravermelha, o torna ideal para luminárias, espelhos e revestimentos térmicos (Krishnamurthy, 2005). Outro aspecto importante é sua não-toxicidade e impermeabilidade a odores e sabores, que são fundamentais para seu uso na indústria alimentícia e farmacêutica (Polmear, 2006; European Aluminium Association, 2018). Essas características tornam o alumínio especialmente adequado para embalagens, pois ele protege os alimentos e bebidas contra contaminação, umidade, luz e oxidação, preservando a qualidade e prolongando a vida útil dos produtos (Bertrand et al., 2022).

O alumínio é amplamente utilizado em setores industriais estratégicos, como transporte, construção civil, embalagens e energia, em função da crescente demanda por materiais leves, recicláveis e com elevado desempenho mecânico e térmico (Guzmán; Muñoz; Carrasco, 2020). No setor automotivo, seu uso tem aumentado para reduzir o peso dos veículos, melhorando a eficiência energética e diminuindo as emissões de gases de efeito estufa (Zhang et al., 2021). Também é fundamental na indústria aeronáutica, por sua alta relação resistência-peso (Bertrand et al., 2022).

Além de suas propriedades físico-químicas, o alumínio destaca-se por sua sustentabilidade, pois pode ser reciclado infinitamente sem perda significativa de qualidade (European Aluminium Association, 2023). Aproximadamente 75% do alumínio produzido na história ainda está em uso, o que reforça sua durabilidade e relevância para a economia circular (Bertrand et al., 2022). Dessa forma, o alumínio é considerado um “metal verde” essencial para a transição energética e redução da pegada de carbono, com aplicações estratégicas em energias renováveis, veículos elétricos e construção sustentável, alinhando-se às metas ambientais globais, como o Acordo de Paris (Zhang et al., 2021).

A produção do alumínio industrial pode ser feita através de duas maneiras: o segmento primário, que obtém o alumínio direto da bauxita e o refina; e o segmento secundário, que recicla e funde sucatas e lingotes (Subramanian; Sahu, 2021).

### **2.1.1 A indústria primária**

A produção de alumínio primário é um processo energeticamente intenso, no qual se configura pela transformação da bauxita — o principal minério de alumínio — no metal puro, através de um complexo ciclo tecnológico (Subramanian; Sahu, 2021; International Aluminium Institute, 2023). Esse processo envolve essencialmente duas grandes etapas: a obtenção da alumina (óxido de alumínio,  $\text{Al}_2\text{O}_3$ ) por meio do Processo Bayer, e a subsequente eletrólise da alumina fundida utilizando o Processo Hall-Héroult, ambos amplamente padronizados na indústria global (Kaplan, 2019; Davis, 1993).

O ciclo produtivo do alumínio inicia-se com a mineração da bauxita, um minério abundante na crosta terrestre, cuja viabilidade econômica depende do teor de alumina extraível, mineralogia e condições de extração (Subramanian; Sahu, 2021). A bauxita é composta principalmente por gibbsita ( $\text{Al}(\text{OH})_3$ ) e contém impurezas como óxidos de ferro e minerais de titânio, que precisam ser removidos para obter alumina de alta pureza, essencial para o processo seguinte (Martins et al., 2020; International Aluminium Institute, 2023).

Na sequência, no Processo Bayer, desenvolvido em 1888, a bauxita é digerida em solução de hidróxido de sódio a temperaturas entre  $140^\circ\text{C}$  e  $240^\circ\text{C}$ , dissolvendo seletivamente a alumina como aluminato de sódio ( $\text{NaAlO}_2$ ). As impurezas insolúveis, conhecidas como lama vermelha, são separadas, e a alumina

precipita-se como hidrato de alumínio ( $\text{Al}(\text{OH})_3$ ), que é calcinado acima de  $1000^\circ\text{C}$  para formar alumina anidra, pronta para a etapa seguinte (Papanikolaou et al., 2021; Davis, 1993).

A alumina calcinada é então submetida à eletrólise ígnea no Processo Hall-Hérault, patenteado em 1886. Neste processo, a alumina é dissolvida em criolita fundida ( $\text{Na}_3\text{AlF}_6$ ) para reduzir seu ponto de fusão e melhorar a condutividade elétrica. Sob alta corrente contínua, os íons de alumínio são reduzidos no cátodo para formar alumínio metálico, enquanto o oxigênio reage com os ânodos de carbono, liberando dióxido de carbono ( $\text{CO}_2$ ) como subproduto (Roedel et al., 2020; Kaplan, 2019).

O resultado desse processo é a obtenção de alumínio metálico com pureza superior a 99,5%, que se acumula no fundo das cubas eletrolíticas, sendo periodicamente removido por sifonagem (Davis, 1993). Entretanto, o processo Hall-Hérault consome cerca de 14 a 16 MWh por tonelada de alumínio, configurando-se como o principal fator de impacto ambiental e custo na produção primária, o que torna a matriz energética um elemento crucial para a sustentabilidade do setor (International Energy Agency, 2023; International Aluminium Institute, 2023).

Além do elevado consumo energético, a produção primária gera resíduos como a lama vermelha e emite gases de efeito estufa, especialmente  $\text{CO}_2$  e perfluorocarbonos (PFCs), estes últimos com potencial de aquecimento global até 10 mil vezes superior ao  $\text{CO}_2$ , liberados durante os chamados anode effects na eletrólise (United Nations Environment Programme, 2021).

### **2.1.2 A indústria secundária**

O alumínio secundário corresponde ao metal produzido a partir de processos de reciclagem, especialmente da sucata e de lingotes descartados ou reaproveitados de processos industriais (European Aluminium Association, 2020). Diferentemente do alumínio primário, que depende da extração da bauxita e processos intensivos em energia, a produção secundária se fundamenta na transformação de resíduos metálicos em novos produtos, promovendo a redução significativa do impacto ambiental associado à cadeia produtiva do alumínio (European Aluminium Association, 2020).

A reciclagem do alumínio envolve diversas etapas, desde a coleta e classificação da sucata até sua fusão, refino e posterior conformação. Esse ciclo permite que o alumínio seja infinitamente reciclado sem perda significativa de suas propriedades físico-químicas, como resistência mecânica, leveza e resistência à corrosão (Subramanian, 2017). No contexto industrial, duas principais fontes alimentam a reciclagem: a sucata pré-consumo, composta por resíduos industriais e aparas de processos produtivos; e a sucata pós-consumo, oriunda do descarte de produtos no final de sua vida útil, como latas, painéis, cabos e componentes automotivos (Cimpan et al., 2015).

Um dos principais atrativos do alumínio secundário é sua eficiência energética em comparação com a produção primária. De acordo com Haraldsson & Johansson (2018), a reciclagem de alumínio consome apenas cerca de 5% da energia necessária para a produção do alumínio primário a partir da bauxita. Esse diferencial coloca o alumínio reciclado como um dos materiais mais sustentáveis da indústria metalúrgica global, contribuindo diretamente para a mitigação das emissões de gases de efeito estufa e a redução da pegada de carbono da indústria (Kleijn et al., 2016). A redução no consumo energético também se traduz em ganhos econômicos expressivos, tornando a reciclagem não apenas uma prática ambientalmente responsável, mas também economicamente vantajosa (Kleijn et al., 2016).

Nesse contexto, a reciclagem do alumínio é frequentemente apontada como um pilar fundamental da economia circular, um modelo de produção e consumo baseado na maximização da reutilização de recursos e na minimização da geração de resíduos (Geissdoerfer et al., 2017). A valorização da sucata metálica como matéria-prima impulsiona não apenas a competitividade da indústria, mas também reforça compromissos globais com os Objetivos de Desenvolvimento Sustentável (ODS), especialmente no que tange à produção e consumo responsáveis (ONU, 2022).

Apesar dos reconhecidos benefícios ambientais e econômicos da reciclagem do alumínio, o processo ainda enfrenta desafios técnicos relevantes. Um dos principais é a perda de fusão, fenômeno que ocorre durante a etapa de fundição e se refere à fração do alumínio metálico que é oxidada (Zhao et al., 2019). Segundo Ravi et al. (2017), essa perda pode variar entre 1% e 10% do volume total processado.

## 2.2 Transformação do Alumínio

A transformação do alumínio compreende um conjunto de operações metalúrgicas essenciais para converter o metal, seja de origem primária ou secundária, em produtos semiacabados ou acabados (Mondolfo, 1976; Hatch, 1984). O processo inicia-se com a fusão, geralmente realizada em fornos industriais específicos, como fornos de cadinho, fornos rotativos ou reverberatórios, os quais permitem o aquecimento controlado do alumínio até sua temperatura de fusão, situada em torno de 660 °C (Mondolfo, 1976; Hatch, 1984).

A passagem do estado sólido para o líquido possibilita a posterior moldagem do metal, bem como a adição de elementos químicos que resultam na formação de ligas metálicas (Davis, 1993). A escolha dos elementos de liga visa atender às propriedades requeridas para cada aplicação, como o aumento da resistência mecânica, a melhoria da resistência à corrosão ou a adequação à soldabilidade (Davis, 1993; Subramanian, 2017).

Após a fusão e a elaboração das ligas, o alumínio fundido é direcionado aos processos de fundição, que variam conforme a complexidade geométrica e as exigências mecânicas das peças a serem produzidas (Campbell, 2015). Entre os métodos mais utilizados destacam-se a fundição em moldes permanentes, em areia, sob pressão e por centrifugação (Campbell, 2015). A fundição em moldes permanentes, por exemplo, é amplamente empregada na indústria automotiva, especialmente na fabricação de blocos de motores e cabeçotes, devido à sua capacidade de produzir peças com elevada precisão dimensional e boas propriedades mecânicas (García et al., 2021).

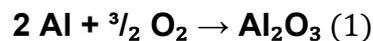
Durante o processo de fusão, especificamente quando se utilizam sucatas, ocorrem reações inevitáveis com o oxigênio presente no ambiente, promovendo a formação de óxidos e escória na superfície do banho metálico (Zhao et al., 2019). Além disso, contaminantes como revestimentos orgânicos, óleos e partículas não metálicas podem comprometer significativamente a qualidade do alumínio fundido e gerar perdas expressivas, conhecidas como perdas de fusão (Bertram et al., 2020).

Na sequência da fundição, o alumínio pode ser submetido a diversos processos de conformação mecânica, como laminação, extrusão ou trefilação, que permitem a obtenção de perfis, chapas, fios ou tubos. A laminação, que pode ser realizada a quente ou a frio, consiste na redução da espessura do alumínio mediante

sua passagem entre cilindros giratórios (Totten & MacKenzie, 2003). A extrusão, por sua vez, envolve a compressão do metal através de um molde, possibilitando a produção de perfis com seções transversais específicas, amplamente utilizados na construção civil e na fabricação de esquadrias e componentes estruturais (Kaufman & Rooy, 2004).

### 2.3 A escória de Alumínio

A escória de alumínio é um subproduto inevitável do processo de fusão desse metal, formada principalmente pela reação do alumínio fundido com o oxigênio do ar, além da incorporação de contaminantes e sais fundentes utilizados no processo metalúrgico (Freti, Bornand & Buxmann, 1982; Santos, 2013). A reação química predominante responsável pela formação do óxido é descrita como na equação 1, esta reação é resultado da elevada afinidade do alumínio pelo oxigênio em temperaturas típicas de fusão (~660 °C) (Santos, 2013).



De acordo com Santos (2013), a escória contém, além dos óxidos de alumínio ( $\text{Al}_2\text{O}_3$ ), diversos outros compostos, como carbeto de alumínio ( $\text{Al}_4\text{C}_3$ ), nitreto de alumínio ( $\text{AlN}$ ), óxidos de ferro ( $\text{Fe}_2\text{O}_3$ ), óxidos de silício ( $\text{SiO}_2$ ) e sais, principalmente cloretos e fluoretos, que aparecem quando fluxantes são utilizados para minimizar a oxidação do metal líquido. A composição média da escória apresenta entre 25% a 30% de óxido de alumínio e de 65% a 75% de alumínio metálico residual, além de pequenas proporções de compostos secundários (ABAL, 2007; Santos, 2013).

No ambiente industrial, a formação de escória é o principal indicador técnico conhecido como Perda de Fusão (PF), que reflete a eficiência do processo de fusão do alumínio. A PF corresponde à fração da massa processada que se transforma em escória, representando de forma objetiva o impacto das reações de oxidação e das impurezas presentes na carga metálica (Santos, 2013). De modo prático e amplamente adotado no setor, a perda de fusão é percentual medido através da massa residual gerada no processo, a escória, pela massa total de alumínio alimentada no forno (Santos, 2013).

A elevada quantidade de alumínio metálico retido na escória representa um dos principais desafios para a eficiência do processo e a sustentabilidade econômica da indústria de transformação do alumínio. Estima-se que a formação de escória possa acarretar perdas metálicas entre 10% a 25% do material processado, especialmente quando há utilização de sucatas com altos níveis de contaminação (Zhao et al., 2019; Santos, 2013).

Além disso, a escória apresenta grande heterogeneidade granulométrica, contendo desde partículas grossas com elevado teor de alumínio metálico — que podem ser recuperadas por processos pirometalúrgicos — até frações finas, com partículas inferiores a 0,15 mm, que apresentam maiores dificuldades de recuperação e são comumente descartadas, ampliando os passivos ambientais (Santos, 2013; Tenório, 2010).

Diversos estudos destacam que o controle operacional é um dos principais fatores determinantes na intensidade da formação de escória e na magnitude das perdas metálicas no processamento do alumínio. Tsakiridis et al. (2013) alertam que a oxidação é intensificada em atmosferas ricas em oxigênio e em temperaturas elevadas, especialmente quando há exposição prolongada do banho metálico.

A temperatura de fusão, por exemplo, exerce influência direta sobre a taxa de oxidação do alumínio líquido. Segundo Zhao et al. (2019), temperaturas superiores a 750 °C promovem uma aceleração exponencial das reações oxidativas, favorecendo a formação de óxidos e o crescimento da camada superficial de escória. Esse comportamento é explicado pela elevada reatividade do alumínio líquido com o oxigênio atmosférico, que aumenta drasticamente com o incremento térmico, conforme descrito por Totten e MacKenzie (2003).

Além da temperatura, o tempo de residência do alumínio no forno está diretamente associado à maior formação de escória. Quanto mais tempo o metal permanece em estado líquido sob atmosfera oxidante, maiores são as perdas metálicas por oxidação superficial. Esse fenômeno foi demonstrado experimentalmente por Santos (2013), que observou um incremento significativo na quantidade de escória formada em ensaios com tempos prolongados de fusão, especialmente na presença de sucatas contaminadas.

Outro fator crítico é o tipo de atmosfera de processamento. De acordo com Li et al. (2021), o uso de atmosferas inertes ou redutoras — como nitrogênio ou argônio — durante a fusão reduz a disponibilidade de oxigênio livre, limitando as

reações de oxidação e, conseqüentemente, diminuindo a formação de escória. Esse mesmo estudo relata reduções superiores a 30% na massa de escória formada quando comparado com processos realizados sob atmosfera aberta ao ar.

A gestão da escória de alumínio é, ainda, uma preocupação ambiental relevante. Classificada como resíduo industrial perigoso, sua disposição inadequada pode causar contaminação do solo e da água, além de riscos à saúde humana pela liberação de substâncias tóxicas (Ambiente Brasil, 2007; Santos, 2013). Por isso, o controle da geração de escória torna-se fundamental não apenas para minimizar perdas no processo, mas também para atenuar os impactos ambientais e garantir a segurança no ambiente industrial (Geissdoerfer et al., 2017).

#### **2.4 Machine Learning**

O Aprendizado de Máquina (*Machine Learning*) é um subcampo da inteligência artificial focado no desenvolvimento de algoritmos capazes de identificar padrões e realizar previsões a partir de dados, sem que sejam explicitamente programados para cada tarefa específica (Mitchell, 1997; Goodfellow, Bengio & Courville, 2016). Essa tecnologia tem papel fundamental na Indústria 4.0, ao viabilizar a automação inteligente de processos, a otimização operacional, o aprimoramento da manutenção preditiva e a tomada de decisão baseada em dados em tempo real (Wuest et al., 2016; Lee et al., 2018; Deloitte, 2023).

O ciclo de implementação de soluções baseadas em aprendizado de máquina envolve algumas essenciais. Inicialmente, ocorre a coleta de dados por meio de sensores industriais, sistemas de controle e bancos de dados corporativos, capturando informações sobre variáveis de processo, parâmetros de qualidade e indicadores de desempenho (Han, Kamber & Pei, 2012; Zhou et al., 2021). Em seguida, o pré-processamento é realizado para garantir a qualidade dos dados, incluindo limpeza, tratamento de valores ausentes, normalização e transformação, tornando-os adequados para o treinamento dos modelos (Kotsiantis, Zaharakis & Pintelas, 2007).

A etapa seguinte é o treinamento, na qual os algoritmos ajustam seus parâmetros com base em dados históricos, buscando extrair padrões subjacentes e relações relevantes (Hastie, Tibshirani & Friedman, 2009; Géron, 2019). Na validação, o modelo é testado com dados independentes para avaliar sua

capacidade de generalização e evitar o sobreajuste (overfitting), garantindo robustez nas previsões (Goodfellow, Bengio & Courville, 2016).

Domingos (2012) descreve que o aprendizado de máquina repousa sobre três pilares fundamentais: representação, avaliação e otimização. A representação diz respeito à forma como o modelo matemático captura os padrões dos dados, que pode ser por meio de regressões, árvores de decisão, máquinas de vetores de suporte ou redes neurais profundas. A avaliação envolve métricas de desempenho, como erro quadrático médio (RMSE) e coeficiente de determinação ( $R^2$ ), que orientam a escolha do modelo mais eficaz. Por fim, a otimização corresponde ao ajuste dos parâmetros do modelo para maximizar sua capacidade preditiva, utilizando métodos como gradiente descendente, busca em grade (grid search) ou algoritmos evolutivos (Domingos, 2012; Goodfellow, Bengio & Courville, 2016).

No ambiente industrial, o aprendizado de máquina é amplamente aplicado em manutenção preditiva — antecipando falhas e reduzindo paradas não planejadas — e na otimização de processos, permitindo o ajuste automático e em tempo real dos parâmetros operacionais para maximizar a eficiência e minimizar desperdícios (Zonta et al., 2020; Lee et al., 2018; Zhang et al., 2022). A integração do machine learning aos sistemas industriais configura-se, assim, como um dos pilares da transformação digital e da competitividade na era da Indústria 4.0, promovendo maior agilidade, precisão e sustentabilidade operacional (Deloitte, 2023; Wuest et al., 2016).

## **2.4.1 Modelos Estatísticos e de *Machine Learning***

### **2.4.1.1 Regressão Linear**

A Regressão Linear é um modelo estatístico que busca descrever a dependência entre uma variável resposta contínua e um conjunto de variáveis preditoras por meio de uma função linear. Segundo Montgomery, Peck e Vining (2012), a regressão linear "modela a relação entre uma variável dependente e uma ou mais variáveis independentes ajustando uma equação linear aos dados observados". A formulação clássica do modelo é dada pela equação 2:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (2)$$

Em que  $\beta_0$  representa o intercepto,  $\beta_1, \beta_2 \dots \beta_n$  são os coeficientes que expressam o efeito marginal de cada variável preditora  $X_i$  sobre a variável dependente  $Y$ , e  $\varepsilon$  é um termo aleatório de erro, assumido como uma variável aleatória com distribuição normal de média zero e variância constante, ou seja,  $\varepsilon \sim N(0, \sigma^2)$  (Montgomery, Peck & Vining, 2012).

O ajuste do modelo se dá pela minimização da soma dos quadrados dos resíduos, conhecida como Resíduos Quadráticos Mínimos (Ordinary Least Squares - OLS). A função objetivo a ser minimizada é:

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

Onde  $y_i$  são os valores observados e  $\hat{y}_i$  são os valores previstos. A solução analítica para o vetor de coeficientes  $\beta$  é obtida através das equações normais:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (4)$$

Onde  $X$  é a matriz de design contendo as observações das variáveis preditoras e  $y$  é o vetor das observações da variável resposta (Seber & Lee, 2012).

O modelo de regressão linear parte de quatro pressupostos fundamentais: (i) linearidade da relação entre as variáveis; (ii) homocedasticidade dos erros; (iii) independência dos resíduos; e (iv) normalidade dos erros (Montgomery et al., 2012). A violação dessas hipóteses compromete a validade estatística do modelo, conforme demonstrado por Draper e Smith (1998).

As principais vantagens da regressão linear residem na sua simplicidade e facilidade de interpretação. O modelo permite identificar de maneira clara como cada variável impacta no resultado, sendo amplamente utilizado em análises exploratórias e em processos industriais bem compreendidos (James et al., 2013). Entretanto, sua principal limitação consiste na incapacidade de capturar relações não-lineares ou interações complexas entre variáveis, reduzindo sua eficácia em contextos em que tais padrões predominam (Hastie, Tibshirani & Friedman, 2009).

#### **2.4.1.2 Random Forest**

O modelo de *Random Forest* é um método de aprendizado de máquina baseado em conjuntos de árvores de decisão, fundamentado na técnica de

*Bootstrap Aggregating (bagging)*, cuja concepção foi apresentada por Breiman (2001). A ideia central do *Random Forest* consiste na geração de múltiplas árvores de decisão a partir de subconjuntos aleatórios com reposição do conjunto original de dados, e na realização de predições por meio da agregação dessas árvores.

Matematicamente, o preditor final  $\hat{f}(x)$  de uma *Random Forest* composta por B árvores é expresso como:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (5)$$

Em que  $T_b$  é a predição da b-ésima árvore sobre a entrada x. Em cada divisão de nó dentro das árvores, uma seleção aleatória de subconjuntos de variáveis é considerada, processo que introduz variabilidade entre as árvores e reduz a correlação entre elas, resultando em menor variância do modelo agregado (Hastie, Tibshirani & Friedman, 2009).

A redução do erro do modelo por meio do *bagging* é sustentada pela decomposição do erro de generalização em termos de viés, variância e ruído irreduzível, como formalizado por Geman, Bienenstock e Doursat (1992). O *Random Forest* atua essencialmente na redução da variância sem aumentar substancialmente o viés, o que confere estabilidade ao modelo (Breiman, 2001).

As principais vantagens do *Random Forest* incluem sua robustez ao *overfitting*, sua capacidade de lidar com dados de alta dimensionalidade e a possibilidade de estimar a importância relativa das variáveis no modelo (Breiman, 2001). Essas características tornam o *Random Forest* uma das técnicas mais eficazes para problemas complexos e com múltiplas variáveis inter-relacionadas (Louppe, 2014).

No setor industrial, o *Random Forest* é utilizado em aplicações como diagnóstico de falhas, previsão de demanda e otimização de processos, especialmente quando os dados apresentam ruído ou estruturas não-lineares complexas (Wuest et al., 2016).

#### **2.4.1.3 Gradient Boosting**

O *Gradient Boosting*, proposto por Friedman (2001), difere conceitualmente do *Random Forest* por construir modelos de forma sequencial, em que cada novo modelo busca corrigir os erros cometidos pelos modelos anteriores. A estrutura aditiva do modelo pode ser representada pela equação 5:

$$F_M(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (6)$$

Onde  $h_m$  são modelos fracos, geralmente árvores de decisão rasas, e  $\gamma_m$  os coeficientes de aprendizado.

O algoritmo do *Gradient Boosting* ajusta cada modelo  $h_m$  ao gradiente negativo da função de perda na iteração  $m-1$ , aproximando-se da direção de máxima descida do erro residual. No contexto de regressão, a função de perda mais comum é o Erro Quadrático Médio (MSE):

$$L = \sum_{i=1}^n (y_i - F(x_i))^2 \quad (7)$$

A atualização do modelo se dá pelo acréscimo de um novo modelo ajustado sobre o resíduo atual:

$$F_m = F_{m-1}(x) + \gamma_m h_m(x) \quad (8)$$

Com a determinação de  $\gamma_m$  realizada por meio da minimização da função de perda em cada iteração (Friedman, 2001).

O *Gradient Boosting* caracteriza-se, portanto, como um método de minimização funcional de gradientes, técnica que converge para funções altamente expressivas, capazes de modelar padrões não-lineares complexos. Contudo, sua complexidade exige o ajuste rigoroso de hiperparâmetros, como o número de iterações, taxa de aprendizado e profundidade das árvores, sob pena de sobreajuste (Natekin & Knoll, 2013).

As principais vantagens do *Gradient Boosting* são seu alto desempenho preditivo e sua capacidade de modelar relações complexas e não-lineares entre variáveis (Chen & Guestrin, 2016). No entanto, essa técnica também apresenta desvantagens, como maior suscetibilidade ao *overfitting* caso os hiperparâmetros não sejam ajustados adequadamente, além de exigir maior capacidade computacional (Natekin & Knoll, 2013).

#### **2.4.1.4 Support Vector Regression(SVR)**

O *Support Vector Regression* (SVR) é uma extensão da *Support Vector Machine* (SVM) para problemas de regressão, formalmente introduzida por Vapnik (1995). O objetivo do SVR é encontrar uma função linear de predição que minimize simultaneamente a complexidade do modelo e o erro, permitindo que as predições se desviem dos valores reais em no máximo uma tolerância pré-definida  $\varepsilon$  (Smola & Schölkopf, 2004).

$$f(x) = w \cdot x + b \quad (9)$$

O problema de otimização primal do SVR pode ser formalizado como:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (10)$$

sujeito às restrições:

$$\begin{aligned} \{y_i - (w * x_i + b) \leq \varepsilon + \xi_i \\ \{(w * x_i + b) - y_i \leq \varepsilon + \xi_i^* \\ \{\xi_i + \xi_i^* \geq 0 \quad (11) \end{aligned}$$

Em que  $C > 0$  é um parâmetro de regularização que controla o trade-off entre a complexidade do modelo e a penalização por desvios superiores a  $\varepsilon$ , enquanto  $\xi_i + \xi_i^*$  são as variáveis de folga que permitem essas violações (Vapnik, 1995; Schölkopf & Smola, 2002).

Um aspecto crucial do SVR é a possibilidade de estender a regressão linear para espaços de maior dimensão, por meio do emprego de funções kernel que realizam uma transformação implícita dos dados. O kernel mais comum é o *Radial Basis Function* (RBF), definido como:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (12)$$

Onde é  $\gamma$  um parâmetro que determina a largura do kernel e controla a complexidade da superfície de decisão (Smola & Schölkopf, 2004). A utilização de kernels permite que o SVR modele relações não-lineares entre as variáveis sem a necessidade de transformação explícita dos dados, tornando-se uma ferramenta poderosa para regressão em espaços de alta dimensionalidade (Smola & Schölkopf, 2004).

A formulação dual do SVR, obtida através do método dos multiplicadores de Lagrange, conduz à seleção de vetores de suporte, ou seja, apenas uma fração das

amostras participa da definição do hiperplano de regressão, o que confere esparsidade ao modelo e melhora sua eficiência computacional (Schölkopf & Smola, 2002).

Na indústria, o SVR é comumente empregado em tarefas de previsão de séries temporais, controle de processos e análise de eficiência energética, especialmente em cenários onde há padrões complexos e variabilidade significativa (Zontak & Cohen, 2013).

## 2.5 Linguagem de Programação Python

O Python consolidou-se, nas últimas décadas, como uma das linguagens mais populares em ciência de dados, aprendizado de máquina e automação industrial, principalmente devido à sua sintaxe simples e legível, que facilita a prototipagem rápida de soluções complexas (Van Rossum & Drake, 2009). De acordo com o Stack Overflow Developer Survey (2023), Python está entre as linguagens mais utilizadas globalmente, especialmente em *data science* e *machine learning*, impulsionada por uma comunidade ativa e pelo desenvolvimento contínuo de bibliotecas especializadas.

As principais vantagens da linguagem incluem: curva de aprendizado acessível, integração com outros sistemas e um vasto ecossistema de bibliotecas, como *NumPy*, *Pandas*, *Scikit-learn* e *Seaborn*, que permitem desde operações matemáticas básicas até técnicas avançadas de modelagem preditiva (VanderPlas, 2016; Raschka, 2015).

No contexto da Indústria 4.0, Python destaca-se como ferramenta essencial para automação de processos, implementação de sistemas inteligentes e integração com tecnologias emergentes, como IoT e manufatura inteligente. Esse protagonismo deve-se, sobretudo, ao seu robusto ecossistema científico, que fornece suporte eficiente para manipulação de grandes volumes de dados, desenvolvimento de algoritmos sofisticados e visualizações avançadas (Oliphant, 2006; Harris et al., 2020).

Segundo VanderPlas (2016), esse ecossistema consolidou Python como uma das linguagens mais utilizadas em ambientes acadêmicos e industriais. As bibliotecas científicas promovem padronização de práticas, eficiência computacional

e reprodutibilidade nos fluxos de trabalho, características essenciais na ciência de dados moderna (McKinney, 2017; Pedregosa et al., 2011). Assim, a chamada "stack científica do Python" tornou-se um padrão de fato para aplicações em aprendizado de máquina, estatística computacional e modelagem preditiva (Raschka & Mirjalili, 2020).

## 2.6 Bibliotecas Python Aplicadas à Modelagem Preditiva

### 2.6.1 Pandas

A biblioteca *Pandas* foi desenvolvida inicialmente por Wes McKinney em 2008, com o objetivo de suprir a necessidade por ferramentas de manipulação de dados estruturados em Python (McKinney, 2010). Sua estrutura de dados mais representativa é o *DataFrame*, um arranjo bidimensional rotulado que permite armazenar, acessar e manipular dados heterogêneos de forma altamente eficiente, reproduzindo funcionalidades anteriormente limitadas a softwares como R e SAS (McKinney, 2017).

Além da manipulação básica de dados, *Pandas* disponibiliza uma série de funções especializadas para tratamento de dados faltantes, conversão de tipos, filtragem, agregação, pivotagem e *merge* de *datasets*. Um recurso amplamente utilizado é o *GroupBy*, que permite realizar operações de agregação e transformação em subconjuntos dos dados, conforme princípios do paradigma "*split-apply-combine*" descrito por Wickham (2011).

Adicionalmente, *Pandas* integra-se nativamente com *NumPy*, o que confere grande velocidade às operações vetorizadas, e com *Matplotlib* e *Seaborn*, o que facilita a geração de gráficos diretamente a partir de estruturas de dados *Pandas* (McKinney, 2017; Hunter, 2007; Waskom, 2021).

Outro aspecto relevante é a capacidade de *Pandas* de lidar com dados temporais através do módulo *datetime*, possibilitando a manipulação eficiente de séries temporais e operações como *resampling*, *rolling windows* e cálculos estatísticos, essenciais para análise preditiva em setores como finanças, energia e processos industriais (VanderPlas, 2016).

### 2.6.2 NumPy

O *NumPy* é a base da computação científica em Python, fornecendo estruturas de dados e algoritmos otimizados para operações numéricas e matriciais (Harris et al., 2020). Introduzido por Travis Oliphant (2006), o *NumPy* surgiu da necessidade de integrar o poder do *Fortran* e do C ao ambiente mais acessível e produtivo de Python, criando uma interface de alto desempenho para manipulação de *arrays* multidimensionais.

O componente central do *NumPy* é o objeto *ndarray*, que permite a execução eficiente de operações matemáticas e algébricas em grandes coleções de dados, explorando vetorização e eliminando a necessidade de laços explícitos em Python, o que resulta em ganhos expressivos de desempenho (Harris et al., 2020).

*NumPy* disponibiliza também uma extensa biblioteca de funções para álgebra linear (como decomposição LU, QR, SVD), operações estatísticas (como média, variância, correlação), geração de números aleatórios e transformadas matemáticas, como a Transformada Rápida de Fourier (FFT), essencial em aplicações de processamento de sinais (Virtanen et al., 2020).

Outro aspecto crucial do *NumPy* é sua interoperabilidade com outras linguagens, como C, C++ e Fortran, via ferramentas como *Cython* e *f2py*, o que permite a incorporação de bibliotecas de baixo nível em projetos de alto desempenho em Python, consolidando-se como uma base indispensável para bibliotecas mais complexas como *SciPy*, *Pandas*, *Scikit-learn* e *TensorFlow* (Oliphant, 2006).

*NumPy* também desempenha papel essencial na definição e manipulação de tensores, uma generalização de matrizes para múltiplas dimensões, que são fundamentais para o desenvolvimento de algoritmos em *deep learning*, como os implementados em *frameworks* como PyTorch e TensorFlow (Goodfellow, Bengio & Courville, 2016).

### 2.6.3 Seaborn

A biblioteca *Seaborn*, criada por Michael Waskom (2014), é uma extensão de alto nível do *Matplotlib*, projetada para simplificar a criação de visualizações estatísticas e promover boas práticas de design gráfico. Um dos principais diferenciais do *Seaborn* é sua capacidade de produzir gráficos estatisticamente informativos com padrões estéticos consistentes, minimizando a necessidade de configuração manual (Waskom, 2021).

*Seaborn* fornece interfaces específicas para análise de distribuição de variáveis, como histogramas kernel *density estimations* (KDE), *boxplots* e *violinplots*, assim como recursos para a exploração de relações bivariadas, por exemplo através do *pairplot*, que exibe todas as combinações possíveis entre variáveis numéricas de um *dataset*, sendo particularmente útil na fase de análise exploratória de dados (Waskom, 2021).

Outra funcionalidade poderosa é a geração de mapas de calor (*heatmaps*), úteis para visualizar matrizes de correlação ou outras representações matriciais, e a criação de gráficos de regressão com intervalos de confiança automáticos, utilizando funções como *lplot* e *regplot* (Wilke, 2019).

Além disso, *Seaborn* oferece suporte nativo a *DataFrames* do *Pandas*, o que permite a criação de gráficos diretamente a partir de *datasets* tabulares, simplificando o fluxo de trabalho na modelagem preditiva (McKinney, 2017). A biblioteca também se integra facilmente a ferramentas interativas, como *Jupyter Notebooks*, viabilizando análises iterativas e relatórios interativos.

#### **2.6.4 Scikit-learn**

A *Scikit-learn* é uma das bibliotecas de *machine learning* mais amplamente utilizadas e respeitadas na comunidade científica e na indústria (Pedregosa et al., 2011; Buitinck et al., 2013). Desenvolvida por Pedregosa et al. (2011), ela oferece implementações eficientes e de fácil utilização para uma ampla variedade de algoritmos supervisionados e não supervisionados, bem como ferramentas para pré-processamento de dados, seleção de atributos e validação de modelos.

Um dos pontos fortes da *Scikit-learn* é sua aderência a um paradigma de API uniforme, no qual todas as classes de estimadores implementam métodos padrão como *fit()*, *predict()* e *score()*, proporcionando uma interface coerente e facilitando a criação de pipelines de modelagem (Buitinck et al., 2013).

Em termos de pré-processamento, *Scikit-learn* disponibiliza ferramentas para normalização, padronização, codificação de variáveis categóricas (como *OneHotEncoder*), bem como técnicas de redução de dimensionalidade (como PCA e t-SNE) e de seleção de atributos baseadas em métodos estatísticos e de aprendizado (Pedregosa et al., 2011).

Na modelagem preditiva, a biblioteca inclui implementações otimizadas de algoritmos clássicos de regressão (como *Linear Regression*, *Ridge*, *Lasso*), classificação (como SVM, k-NN, *Decision Trees*, *Random Forest*, *Gradient Boosting*), e *clustering* (como *K-Means*, *DBSCAN* e *Hierarchical Clustering*).

Para avaliação de modelos, *Scikit-learn* oferece um conjunto abrangente de métricas de desempenho, como *mean squared error* (MSE),  $R^2$  score, precision, recall, f1-score, além de procedimentos de validação cruzada e técnicas para ajuste de hiperparâmetros, como *GridSearchCV* e *RandomizedSearchCV* (Raschka & Mirjalili, 2020).

Além disso, a *Scikit-learn* integra-se diretamente com bibliotecas como *NumPy* e *Pandas*, e é frequentemente utilizada em conjunto com *Matplotlib* ou *Seaborn* para a análise gráfica dos resultados, compondo uma *stack* completa de ferramentas para o desenvolvimento de sistemas de aprendizado de máquina em ambientes acadêmicos e industriais (VanderPlas, 2016).

A clareza, eficiência e abrangência da *Scikit-learn* transformaram-na em uma referência obrigatória na formação de profissionais de ciência de dados e engenharia, como evidenciado por sua adoção em cursos universitários e MOOCs de instituições como Stanford, MIT e Coursera (Zhou, 2021).

### 3 METODOLOGIA

Este trabalho foi desenvolvido em uma planta industrial de transformação de alumínio secundário, onde a fundição da matéria-prima é realizada em fornos revestidos com materiais refratários, próprios para suportar altas temperaturas e condições agressivas do processo. O processo industrial envolve a fusão de sucata e lingotes para a produção de ligas de alumínio.

Adotou-se para esse trabalho uma abordagem quantitativa, fundamentada na modelagem matemática e na análise estatística de dados empíricos coletados em campo, alinhando-se às metodologias consagradas na literatura para análise de processos industriais complexos (Creswell, 2014; Yin, 2018).

#### 3.1 Variáveis de estudo na planta

Os dados utilizados foram obtidos a partir dos registros históricos de operação de três fornos distintos de capacidade de produção (Forno 1, Forno 2 e Forno 3) que produzem a mesma liga de alumínio 3105, uma liga da série 3000, composta principalmente por alumínio com pequenas adições de manganês e magnésio (Leth-Olsen; Nisancioglu, 1997). Esses fornos foram monitorados ao longo de um período de seis meses, o que proporcionou um conjunto robusto e representativo das variações normais do processo. Para cada forno, foram coletadas informações detalhadas sobre os diferentes carregamentos diários, contendo variáveis operacionais essenciais — tais como temperatura e tempo para a pré-escumagem (a etapa de retirada da escória), temperatura e duração do carregamento, quantidades de sucata utilizadas, além do peso total fundido. Essa estruturação dos dados segue recomendações específicas para estudos em processos metalúrgicos secundários, que ressaltam a importância de capturar variabilidades tanto do processo quanto da matéria-prima (Davis, 1993; Bertram et al., 2020).

A temperatura durante o carregamento e a pré-escumagem desempenham um papel crucial na formação da escória, conforme destacado por Davis (1993), desta forma elas foram escolhidas como variáveis explicativas. Além dos fatores mencionados, o tempo de residência do metal líquido dentro do forno e a atmosfera gasosa ambiente impactam diretamente o grau de oxidação, como citado Santos(2013). Assim, os tempos de carregamento e de pré-escumagem foram

também incorporados como variáveis explicativas, visto que influenciam as condições oxidativas.

Neste contexto, a Perda de Fusão (PF) é definida como a variável dependente central do estudo, cuja variação precisa ser explicada e prevista a partir das variáveis operacionais e de matéria-prima mencionadas. O objetivo da modelagem é estabelecer relações quantitativas entre essas variáveis explicativas e a PF, permitindo identificar os fatores que mais impactam as perdas durante o processo de fusão, de modo a subsidiar estratégias para sua minimização.

### **3.2 Coleta e Organização dos Dados**

O banco de dados foi estruturado a partir de duas planilhas eletrônicas compiladas com informações provenientes dos registros industriais. A primeira planilha continha os dados operacionais do processo, incluindo colunas relativas à data, turno, forno (1, 2 e 3), temperaturas de pré-escumagem e carregamento, tempos respectivos, quantidade de escória gerada, total carregado e peso do lingote, obtidos durante o período de 6 meses.

Já a segunda planilha armazenou as informações da matéria-prima, listando a data, o número do carregamento, o tipo de sucata e o peso correspondente. Devido à possibilidade de múltiplos carregamentos por dia, e de cada carregamento ser composto por diferentes tipos de sucata, foi necessário organizar os dados de forma hierárquica para preservar as relações entre essas informações.

### **3.3 Modelagem Computacional**

#### **3.3.1 Ferramentas Computacionais Utilizadas**

Todo o desenvolvimento deste estudo foi realizado em ambiente computacional baseado em nuvem, utilizando o *Google Colab* para a execução dos códigos em Python, garantindo flexibilidade, escalabilidade e reprodutibilidade no processamento dos dados.

Foram utilizadas as seguintes ferramentas e bibliotecas: Python como linguagem principal para análise e modelagem; *Pandas* e *NumPy* para manipulação e tratamento dos dados; *Matplotlib* e *Seaborn* para a geração de visualizações gráficas; *Statsmodels* para a implementação de modelos estatísticos de regressão linear; e *Scikit-learn* para desenvolvimento, treinamento e avaliação de modelos de machine learning, incluindo o algoritmo Random Forest.

Esse conjunto integrado de ferramentas possibilitou a realização de uma análise abrangente e rigorosa, combinando métodos estatísticos tradicionais com técnicas avançadas de aprendizado de máquina, conforme melhores práticas para estudos de otimização de processos industriais (Bishop, 2006; Hastie, Tibshirani & Friedman, 2009).

### 3.3.2 Entrada do Programa

A entrada do programa foi estruturada para permitir o carregamento manual de duas planilhas no ambiente *Google Colab*, utilizando a função *files.upload()*. A primeira planilha contém os dados operacionais do processo de fusão, como temperatura, tempo de carregamento, peso de lingote e perda de fusão (PF). A segunda planilha registra as informações referentes aos tipos de sucata utilizados, suas respectivas massas e a liga correspondente.

Após o upload, os arquivos são lidos diretamente da memória com o auxílio da biblioteca *pandas*, utilizando *io.BytesIO*, o que dispensa a necessidade de salvamento local. Em seguida, os dados são convertidos em *dataframes* e passam por um processo de padronização, com destaque para a conversão de datas para o formato *datetime* e a uniformização do campo “Carregamento” como texto.

Para garantir a correta associação entre as bases de dados, foi criado um identificador único, denominado *ID\_Carregamento*, que vincula as duas planilhas e permite uma análise integrada, mesmo na presença de carregamentos compostos por múltiplos tipos de sucata (Kumar; Raj, 2015). Essa abordagem assegura que os parâmetros operacionais e as características da matéria-prima sejam combinados corretamente em cada ponto de análise, formando a base consolidada para os cálculos de perda de fusão corrigida e para a execução das análises estatísticas.

### 3.3.3 Tratamento e Pré-processamento dos Dados

O tratamento e o pré-processamento dos dados foram conduzidos com o auxílio das bibliotecas *Pandas* e *NumPy*, reconhecidas por sua eficiência e robustez em análise de dados científicos (McKinney, 2017; Oliphant, 2007). As etapas principais incluíram a padronização das unidades de medida e nomenclaturas para uniformização dos dados, fundamental para garantir a consistência ao longo das análises.

Também foram tratadas possíveis falhas nos dados, como valores ausentes, inconsistências e *outliers*. Para valores ausentes, aplicou-se imputação utilizando média ou mediana conforme o contexto. Inconsistências foram verificadas por meio de regras de validação lógica e correção manual quando necessário. Discrepâncias foram identificados com base no critério do intervalo interquartil (IQR) e eliminados ou substituídos para evitar distorções na análise.

Esses procedimentos seguem as melhores práticas recomendadas na literatura para garantir a qualidade e confiabilidade dos dados (Little & Rubin, 2019; Wilcox, 2012). Foi então calculada a Perda de Fusão corrigida (Equação 10), onde é descontado o efeito do lingote, criando-se uma variável resposta mais adequada para a modelagem.

$$PF_{CORRIGIDA} = PF - \left( \frac{\text{Peso do Lingote}}{\text{Peso total}} \times 100 \times 0,075 \right) \quad (13)$$

A correção na variável Perda de Fusão (PF) foi respaldada na literatura técnica que considera que a perda de fusão do lingote é relativamente constante e varia entre 0,5% e 1% do peso adicionado (ASTM B221, 2020). Optou-se, então, por aplicar um desconto fixo de 0,75% do peso do lingote para ajustar a Perda de Fusão bruta.

Adicionalmente, foram criadas variáveis derivadas e de interação, combinando temperatura, tempo e tipo de sucata, o que ampliou a capacidade explicativa dos modelos e permitiu capturar efeitos conjuntos que não seriam evidenciados por variáveis isoladas. As variáveis categóricas, especialmente o tipo de sucata, foram codificadas por meio de técnicas de *One-Hot Encoding*, garantindo a compatibilidade com algoritmos de *machine learning* utilizados posteriormente (Pedregosa et al., 2011).

Por fim, a análise exploratória dos dados foi realizada com a ajuda de bibliotecas de visualização gráfica como *Matplotlib* e *Seaborn*, que possibilitaram a identificação preliminar de padrões, tendências e potenciais relações entre as variáveis. Esta é uma etapa fundamental para orientar a construção de modelos preditivos confiáveis e interpretar os fenômenos observados (Tukey, 1977; Waskom, 2021).

### 3.3.4 Integração de Modelos

A etapa de integração de modelos teve como objetivo principal construir uma estrutura computacional de testes de modelos estatísticos e de machine learning no qual fosse possível identificar qual destes se enquadraria melhor aos dados para que posteriormente fosse possível identificar e quantificar os fatores que mais influenciam a Perda de Fusão (PF), de forma a subsidiar propostas de melhoria e redução dos índices de escória gerada.

Foram adotadas quatro abordagens complementares para análise:

- A regressão linear múltipla foi empregada para examinar, de maneira mais simples, as relações lineares entre as variáveis explicativas e a PF. Nessa etapa, foi realizada análise de multicolinearidade por meio de matrizes de correlação e cálculo do *Variance Inflation Factor* (VIF), para assegurar a robustez e interpretabilidade do modelo, além do teste de normalidade. A biblioteca *Statsmodels* foi utilizada para estimar os parâmetros e conduzir testes de significância estatística, possibilitando a identificação das variáveis com maior impacto no processo (Seabold; Perktold, 2010).
- Para captar relações não lineares e possíveis interações complexas entre as variáveis, aplicou-se o algoritmo *Random Forest*, reconhecido por sua eficácia em problemas de regressão em processos industriais com variáveis heterogêneas e não lineares (Breiman, 2001; Liaw; Wiener, 2002). O modelo foi treinado, validado e testado utilizando a biblioteca *Scikit-learn*, com avaliação do desempenho por meio das métricas de coeficiente de determinação ( $R^2$ ) e erro quadrático médio (MSE, sigla do inglês *Mean Squared Error*), o que possibilitou a comparação objetiva entre os modelos.
- *Gradient Boosting Regressor* (GBR): Implementado com o intuito de explorar ganhos adicionais em precisão preditiva por meio de aprendizado sequencial, onde modelos fracos são combinados para formar um modelo robusto. O GBR é eficaz para evitar *overfitting* e lida bem com variáveis correlacionadas, oferecendo uma abordagem mais sensível para capturar padrões sutis nos dados (Friedman, 2001).

- *Support Vector Regression (SVR):*  
Aplicado como alternativa baseada em métodos de margem máxima, o SVR é eficiente para lidar com conjuntos de dados menores e com ruído, buscando uma função que se desvie o mínimo possível dos valores reais, dentro de uma margem de tolerância. Ele é especialmente útil em problemas com não linearidade moderada e foi ajustado com kernel radial (RBF), também usando Scikit-learn (Drucker et al., 1997).

A comparação dos parâmetros obtidos para cada modelo (como o  $R^2$ , MSE e VIF) permitiu que o programa pudesse identificar qual modelo apresentou melhor adequação aos dados estudados, direcionando este para aplicação das análises sobre as variáveis que afetam a Perda de Fusão.

### 3.3.5 Análises das Variáveis

#### 3.3.5.1 Variável mais Influente para Cada Modelo no Programa Implementado

No programa desenvolvido, a identificação da variável explicativa mais influente ocorre de forma específica para cada abordagem estatística ou de aprendizado de máquina. Após o ajuste dos modelos e a avaliação do coeficiente de determinação ( $R^2$ ) o algoritmo que apresentar melhor desempenho fornece, por meio de métodos estatísticos coerentes, a variável que mais impacta a variável dependente. A seguir, detalha-se como cada modelo, se selecionado, determina essa variável.

##### 1) Regressão Linear

Quando a regressão linear múltipla apresenta  $R^2$  satisfatório, o programa utiliza os coeficientes estimados pelo método dos Mínimos Quadrados Ordinários (OLS) para quantificar o impacto marginal de cada variável explicativa (Montgomery et al., 2012). Assim, a variável com o maior coeficiente absoluto indica maior influência direta sobre a variável dependente, desde que estatisticamente significativa. Essa interpretação é fundamentada no pressuposto de linearidade entre as variáveis, conforme preconizado na literatura de regressão clássica (Draper; Smith, 1998). No programa, esta análise é implementada com a biblioteca *statsmodels*, que gera os parâmetros, intervalos de confiança e valores-p.

##### 2) *Random Forest Regressor*

Se o *Random Forest* for o modelo de melhor desempenho, a variável mais influente é identificada com base em gráficos de Dependência Parcial (Friedman,

2001). Embora o algoritmo calcule internamente a importância pela redução média de impureza (Breiman, 2001), o programa utiliza a Dependência Parcial para explicitar o efeito marginal isolado de cada variável na previsão média. Esta abordagem é reconhecida por melhorar a interpretabilidade de modelos complexos, como recomendam Molnar (2022) e Sculley et al. (2018).

### 3) *Gradient Boosting Regressor*

Para o *Gradient Boosting*, o mesmo princípio se aplica: a variável mais influente é aquela que, segundo a Dependência Parcial, apresenta a maior variação média na previsão do modelo. O método, introduzido por Friedman (2001), permite capturar relações não lineares e interações complexas, justificando a necessidade de técnicas de interpretação como a Dependência Parcial ou valores SHAP (Lundhberg; Lee, 2017).

### 4) *Support Vector Regressor (SVR)*

No caso do SVR, se o kernel for linear, o programa utiliza os coeficientes (coef\_), analogamente à regressão linear (Vapnik, 1995). Quando o kernel é não linear, não há coeficientes diretos, sendo imprescindível recorrer à Dependência Parcial para avaliar qual variável mais altera a previsão média, evidenciando a influência isolada de cada fator (Molnar, 2022).

## 3.3.5.2 Melhores e Piores Cenários De Perda De Fusão

Adicionalmente, foi desenvolvida uma ferramenta baseada em *Random Forest* para simular e identificar os melhores e piores cenários de combinação das variáveis de temperatura e tempo para as ligas produzidas por meio do conceito de árvore de decisão baseada no histórico fornecido. Essa ferramenta permitiu prever, para linha de produção, os conjuntos de condições operacionais que minimizam ou maximizam a Perda de Fusão. Os resultados foram apresentados em gráficos com botões de seleção interativos que facilitam a escolha do tipo de linha de produção escolhida para estudo, mostrando de forma dinâmica as condições mais favoráveis para o processo, orientando desta forma a tomada de decisões práticas para redução de perdas e melhoria contínua do desempenho industrial.

## 3.3.5.2 Saída do Programa

A saída do programa consiste na apresentação de resultados estatísticos e gráficos que auxiliam na compreensão dos fatores que mais influenciam a perda de fusão. Após o tratamento e integração das planilhas, o programa executa análises que resultam em gráficos de dispersão, histogramas, gráficos de interação entre variáveis e gráficos de dependência parcial (PDP), evidenciando relações diretas e combinadas entre variáveis operacionais.

Além dos gráficos exploratórios, o programa também gera saídas numéricas como os coeficientes de regressão, o valor de  $R^2$  e o erro quadrático médio (MSE), VIF e multicolinearidade possibilitando a visualização do desempenho dos modelos utilizados — tanto estatísticos quanto de machine learning. Também são destacados os fatores com maior impacto sobre a perda de fusão, permitindo interpretações técnicas e simulações de cenários operacionais mais eficientes.

### **3.4 Teste de Hipótese**

Para verificar se as recomendações fornecidas pelo programa desenvolvido realmente produzem efeitos práticos significativos no processo de fundição, foi utilizado o conceito de teste de hipótese estatística, que permite avaliar se uma diferença observada entre dois grupos pode ser atribuída ao acaso ou se representa um efeito real (Triola, 2017). Foram adotados parâmetros clássicos de análise, com nível de significância de 5% ( $\alpha = 0,05$ ) e poder estatístico de 80% ( $1 - \beta = 0,80$ ), conforme recomendado na literatura para garantir robustez na tomada de decisão (Campbell; Machin; Walter, 2007).

Com base nesses critérios e assumindo um tamanho de efeito médio (Cohen's  $d = 0,5$ ), foi estimado o número mínimo de testagens necessárias para validação estatística da hipótese, com o auxílio de métodos de cálculo amostral para comparação de médias entre dois grupos independentes (Siegel; Castellan, 2006). O resultado indicou a necessidade de aproximadamente 52 fundições (26 com as recomendações do programa e 26 no método convencional) para detectar diferenças significativas. Considerando a média de 5 fundições por dia, o tempo estimado de coleta seria de cerca de 11 dias de operação normal. Essa abordagem quantitativa fundamenta a avaliação da eficácia prática das recomendações do sistema, com base em evidências estatisticamente confiáveis. Foram avaliados tanto o comportamento da perda de fusão quando se segue a recomendação do programa

para a variável de maneira isolada como também para a combinação de todas as variáveis do processo.

## 4 RESULTADOS

### 4.1 Avaliação do Modelo

Inicialmente obteve-se que o modelo de *Random Forest* se mostrou altamente eficaz para representar a complexidade do processo de fusão de alumínio secundário em escala industrial. A obtenção de coeficientes de determinação ( $R^2$ ) superiores a 0,9 (Tabela 1) para os três fornos analisados evidencia não apenas a robustez do modelo, mas também confirma a natureza intrinsecamente não linear da relação entre variáveis operacionais e a perda de fusão (PF). Este comportamento é coerente com o que é relatado por autores como Zhao et al. (2020) e Wang et al. (2021), que destacam a elevada variabilidade de fatores metalúrgicos.

Tabela 1 –Valores de  $R^2$  do Random Forest para cada forno.

Fornos	Valor de $R^2$
Forno 1	0,9313
Forno 2	0,9217
Forno 3	0,9380

Fonte: A autora (2025).

Para o modelo *Random Forest* nenhuma discrepância de dados foi encontrada pelo programa.

### 4.2 Fatores que Mais Influenciaram a Perda de Fusão

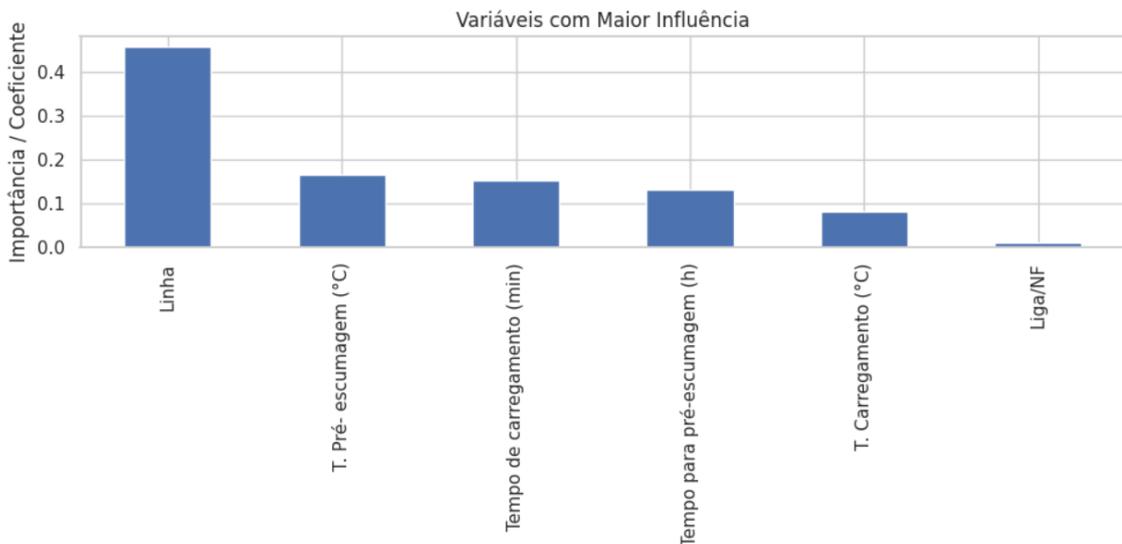
#### 4.2.1 Análise de Importância das Variáveis

A avaliação da importância das variáveis por meio do modelo *Random Forest* revelou que as variáveis operacionais ligadas à linha de produção, às temperaturas e aos tempos de processo foram as mais influentes na determinação da perda de fusão (PF) nos três fornos avaliados. Essa análise permitiu compreender o peso relativo de cada fator no modelo preditivo, funcionando como um indicativo das variáveis que mais explicam a oscilação dos índices de perda de alumínio metálico.

No Forno 1, a variável com maior influência foi a linha de produção, como mostra a Figura 1. Isso sugere que as diferenças operacionais entre linhas – como sequência de carregamento, velocidade de alimentação, configuração do forno, prática dos operadores ou até diferenças nos equipamentos auxiliares – podem ter forte impacto no desempenho da fusão. Tal premissa é coerente com o que afirmam autores como Tsakiridis et al. (2013), que destacam que a eficiência de fusão pode variar significativamente com o layout e a organização operacional de cada linha. Além disso, Fathi et al. (2019) ressaltam que, em plantas com múltiplas linhas, pequenas diferenças em fluxo de ar, tempo de espera da carga, e temperatura residual do refratário podem alterar significativamente o comportamento da oxidação do alumínio.

Ainda no Forno 1, a temperatura de pré-escumagem, o tempo de carregamento e o tempo para pré-escumagem aparecem como variáveis fortemente influentes, o que reforça a ideia de que o controle térmico nas etapas intermediárias do processo é essencial. De acordo com Li et al. (2015), a retirada de escória realizada sob temperaturas inadequadas ou fora do tempo ideal pode provocar reincorporação de óxidos e impedir a remoção eficiente de escória, aumentando as perdas metálicas. A literatura também destaca que o tempo de carregamento prolongado expõe o alumínio ao ar atmosférico por mais tempo em temperaturas crescentes, o que favorece reações oxidativas (Ravi et al., 2017; Zhao et al., 2020).

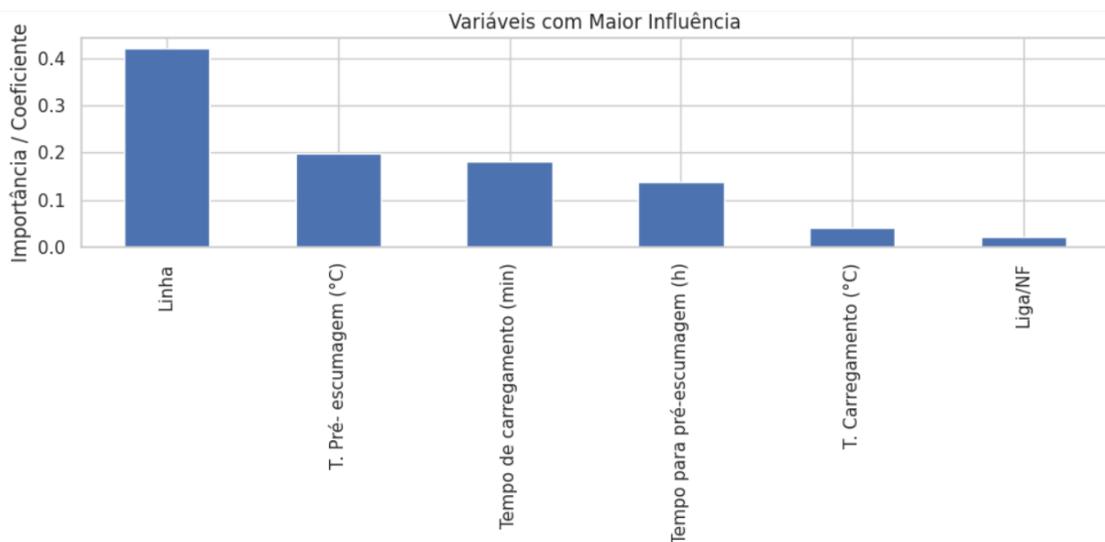
Figura 1 –Gráfico de Importância de Variável do Forno 1.



Fonte: A autora, (2025).

No Forno 2 (Figura 2), a variável mais importante também foi a linha de produção, reforçando a ideia de que há variabilidades operacionais entre as linhas da planta. Isso pode incluir desde diferenças na granulometria da matéria prima até características do refratário, ajustes manuais e perfil de queima. A temperatura de pré-escumagem aparece logo em seguida com destaque, o que sugere que o momento em que a escória é retirada é crítico neste forno. Isso está alinhado com estudos como o de Zhong, Xie e Li (2022), que indicam que a retirada de escória eficiente depende fortemente do ponto térmico em que é realizada, já que temperaturas muito baixas dificultam a remoção da escória líquida, enquanto temperaturas elevadas intensificam a oxidação superficial do banho.

Figura 2–Gráfico de Importância de Variável do Forno 2.



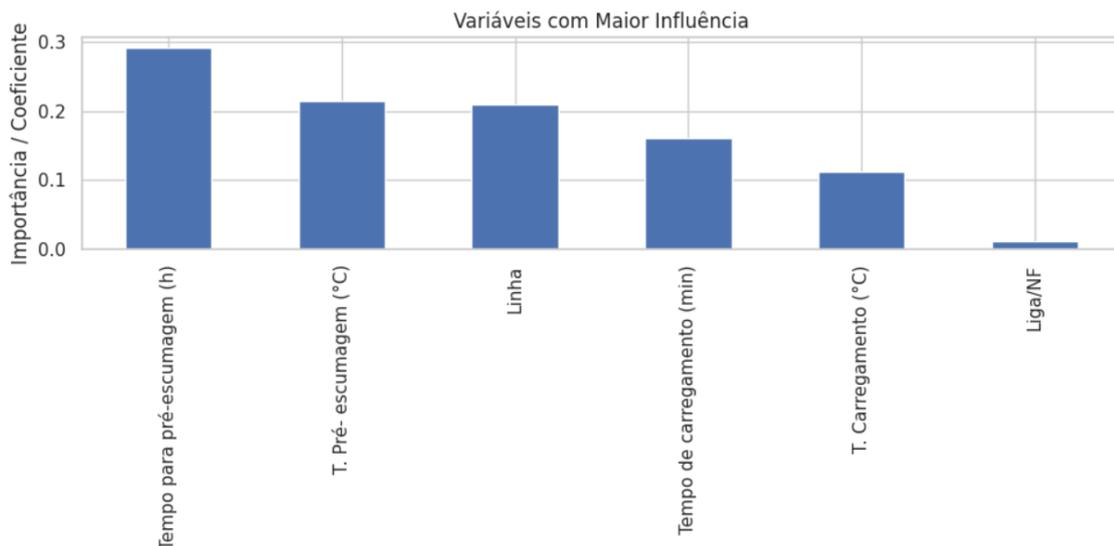
Fonte: A autora, (2025).

Diferentemente do Forno 1, a temperatura de carregamento no Forno 2 teve menor relevância relativa, o que pode ser atribuído ao fato deste forno ter menor dispersão térmica durante essa etapa, possivelmente por possuir melhor isolamento ou por operar com rampas térmicas mais controladas. Fathi et al. (2019) sugerem que fornos com controle mais preciso de queima e menor volume interno são menos afetados por flutuações na temperatura de carregamento, concentrando a sensibilidade em etapas posteriores do processo.

Já no Forno 3 (Figura 3), a variável mais relevante foi o tempo para pré-escumagem, seguido por temperatura de pré-escumagem e tempo de

carregamento. Esse padrão reforça que neste forno o fator crítico não é apenas a temperatura atingida, mas o tempo de exposição do banho metálico antes da remoção da escória, que parece ser decisivo para a perda de massa metálica. Essa observação é coerente com os achados de Zhao et al. (2020), que demonstraram que quanto maior o tempo de retenção em atmosferas ricas em oxigênio antes da retirada da escória, maior a incorporação de óxidos no banho. A presença de óxidos refratários finos, difíceis de remover se a retirada da escória ocorrer fora do tempo ótimo, também pode justificar esse comportamento, como apontado por Li et al. (2015).

Figura 3 –Gráfico de Importância de Variável do Forno 3.



Fonte: A autora, (2025).

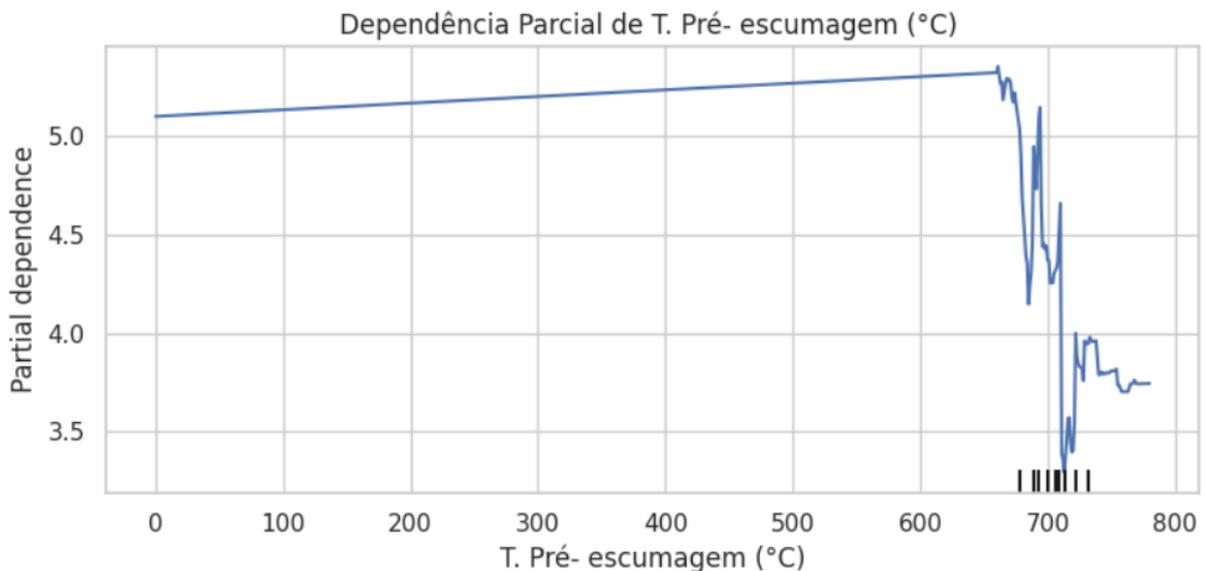
#### 4.2.2 Impacto da Temperatura de Pré-escumagem sobre a Perda de Fusão

Foi observado a influência da temperatura de pré-escumagem sobre a perda de fusão (PF), esta variável que apresentou comportamento distinto entre os três fornos avaliados. A dependência parcial (*Partial Dependence*) é uma métrica interpretativa que mede o efeito médio de uma ou mais variáveis independentes sobre a variável-alvo, mantendo as demais constantes no modelo. Ela permite identificar se o impacto é linear, não linear ou se há pontos de saturação, sendo amplamente empregada em modelos de aprendizado de máquina de difícil interpretação, como *Random Forest* e *Gradient Boosting* (Friedman, 2001).

Embora o efeito oxidativo do alumínio em função da temperatura seja bem documentado na literatura (Zhong, Xie e Li, 2022), os resultados mostram que cada forno possui faixas de sensibilidade térmica próprias, possivelmente devido a diferenças construtivas, operacionais ou no perfil de carga.

No Forno 1 (Figura 4) os dados revelaram que temperaturas de pré-escumagem próximas de 650 °C estão associadas a maior PF, indicando que, abaixo dessa faixa, o sistema não atinge estabilidade térmica suficiente para permitir uma retirada de escória eficiente. Em contrapartida, temperaturas entre 670 °C e 720 °C se mostraram mais favoráveis, apresentando uma tendência clara de redução da PF. Isso sugere que, nesta faixa, há um equilíbrio entre fluidez do banho metálico e controle da oxidação. Temperaturas acima de 730 °C, por sua vez, voltam a gerar aumento da perda, provavelmente por acelerarem reações de oxidação da superfície exposta, conforme já discutido por Ravi et al. (2017).

Figura 4 –Gráfico de Dependência Parcial da Temperatura de pré-escumagem do Forno 1.

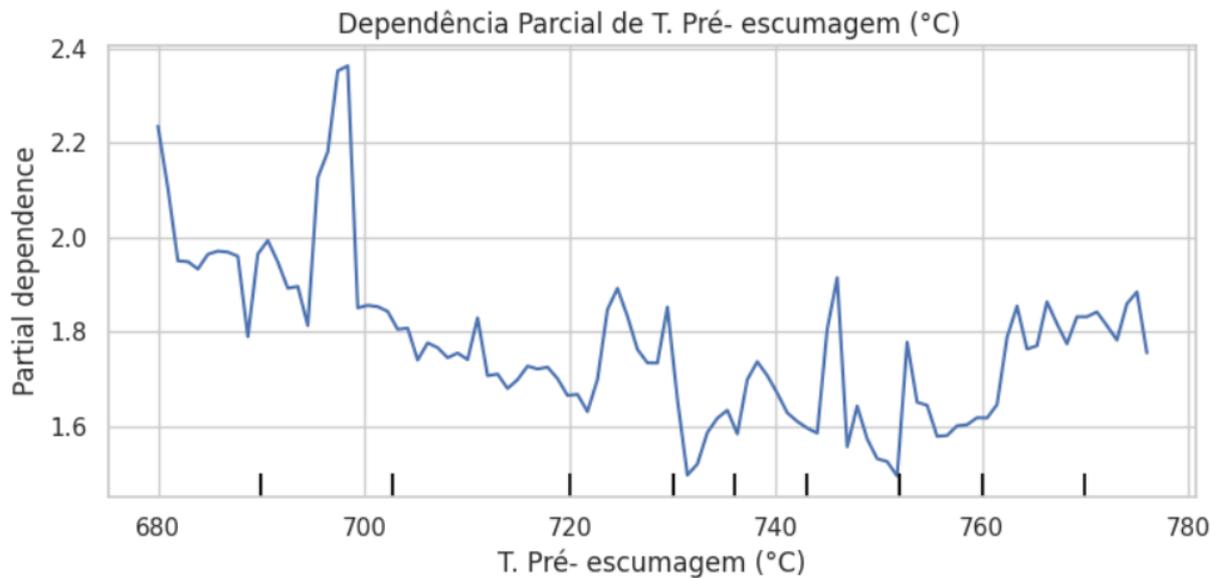


Fonte: A autora, (2025).

Já o Forno 2 (Figura 5) apresentou um comportamento ligeiramente distinto. Nesse caso, o pico de PF foi observado entre 690 °C e 700 °C, faixa intermediária que, em contrapartida com que foi visto, não é crítica para o Forno 1. Essa diferença pode estar relacionada à composição da sucata utilizada ou à geometria do forno, que influencia a distribuição de calor e a formação da escória. As melhores

condições térmicas de pré-escumagem para o Forno 2 foram identificadas entre 730 °C e 740 °C, bem como em torno de 750 °C, faixas que, para o Forno 1, já eram consideradas excessivas. Este dado indica que o Forno 2 pode possuir maior tolerância a temperaturas elevadas, talvez devido a uma taxa de oxigenação do ambiente inferior ou à presença de ligas menos reativas.

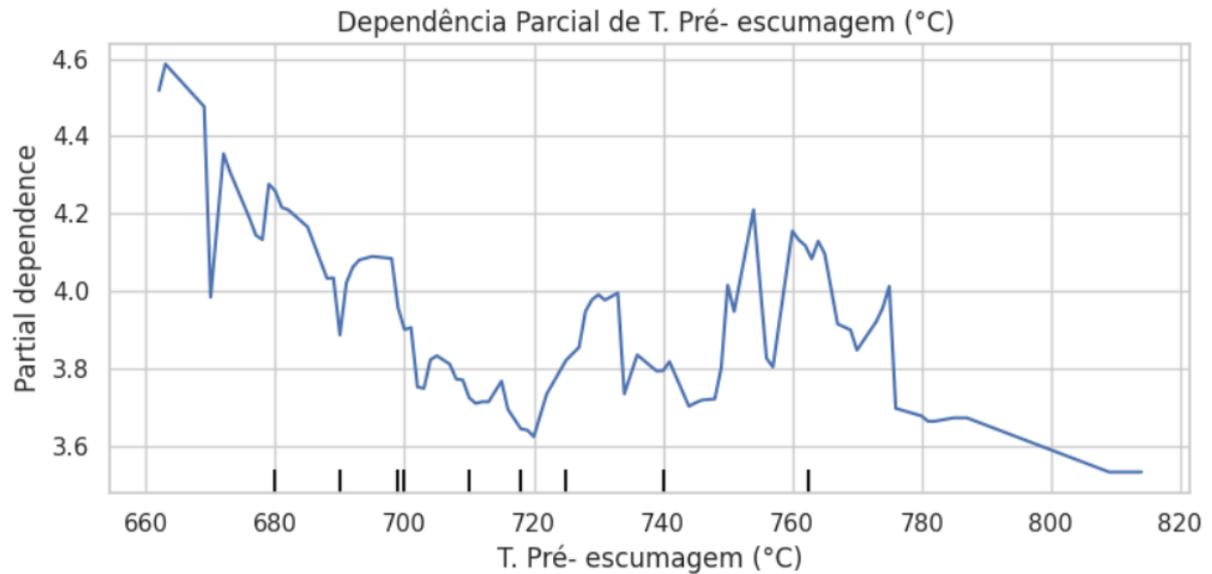
Figura 5 –Gráfico de Dependência Parcial da Temperatura de pré-escumagem do Forno 2.



Fonte: A autora, (2025).

Por sua vez, o Forno 3 (Figura 6) mostrou uma resposta ainda mais sensível à elevação térmica. Temperaturas até 720 °C mantiveram a PF sob controle, mas a partir deste ponto, observou-se um aumento acentuado e contínuo da PF com oscilações, revelando que o forno opera com uma margem térmica mais estreita. Diferentemente dos outros fornos, o Forno 3 parece não apresentar uma zona térmica segura acima de 720 °C, o que sugere um perfil de carga ou controle térmico mais suscetível a variações.

Figura 6 –Gráfico de Dependência Parcial da Temperatura de pré-escumagem do Forno 3.



Fonte: A autora, (2025).

A variação de comportamento térmico entre os fornos pode ser atribuída a fatores como a distribuição de calor, que influencia a homogeneidade térmica e a tolerância a temperaturas mais altas (Ravi et al., 2017). Fornos com melhor convecção evitam zonas de superaquecimento, favorecendo a estabilidade do banho metálico. A composição da sucata também tem papel central, pois ligas com altos teores de manganês, ferro ou silício tendem a oxidar mais intensamente em altas temperaturas, aumentando a perda metálica (Silva; Siqueira; Santos, 2019; Tsakiridis et al., 2013).

Além disso, a eficiência de remoção da escória e a atmosfera interna do forno (oxidante ou redutora) afetam diretamente a taxa de oxidação do alumínio fundido (Zhong; Xie; Li, 2022). Comparativamente, o Forno 1 operou melhor entre 670–720 °C, o Forno 2 se mostrou mais eficiente entre 730–750 °C, e o Forno 3 foi o mais sensível ao calor, com boa estabilidade apenas abaixo de 720 °C, exigindo controle mais rigoroso da temperatura (Zhao et al., 2020).

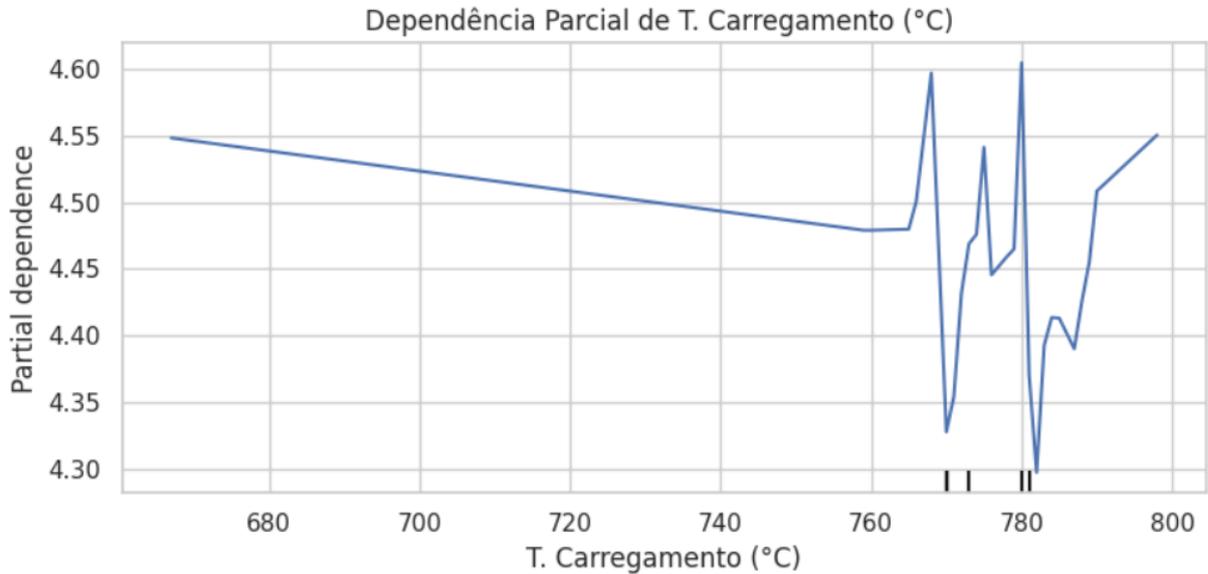
#### 4.2.3 Impacto da Temperatura de Carregamento sobre a Perda de Fusão

A temperatura de carregamento mostrou-se uma variável crítica no comportamento da perda de fusão (PF), com impacto direto sobre a oxidação do alumínio e a eficiência térmica do processo. Através da análise das curvas de dependência parcial foi possível identificar faixas térmicas ideais e críticas distintas para cada forno, refletindo tanto a heterogeneidade operacional quanto o tipo de sucata e liga processadas — elementos que, segundo Tsakiridis et al. (2013), influenciam significativamente a formação de escória e a recuperação metálica durante o processo de fusão.

No Forno 1, temperaturas de carregamento inferiores a 650 °C foram consistentemente associadas a altos níveis de PF, conforme demonstrado nos gráficos de dependência (Figura 7). Em faixas mais baixas de temperatura, o processo de fusão se torna menos eficiente, exigindo maior tempo de permanência no forno, o que aumenta a exposição do metal fundido ao oxigênio atmosférico e intensifica a formação de óxidos de alumínio (Zhong; Xie; Li, 2022).

Por outro lado, temperaturas entre 680 °C e 760 °C configuraram uma zona de estabilidade operacional, em que a PF foi significativamente reduzida, reforçando os achados de Ravi et al. (2017), que apontam essa faixa como ideal para minimizar a formação de escória em fornos rotativos e reverberatórios. A curva de dependência indicou um novo aumento da PF ao ultrapassar os 760 °C, com recuperações pontuais em torno de 775 °C e 781 °C, sugerindo que mesmo em faixas elevadas existem zonas discretas de equilíbrio térmico, possivelmente associadas ao ponto de vaporização de certos elementos da liga (Silva; Siqueira; Santos, 2019).

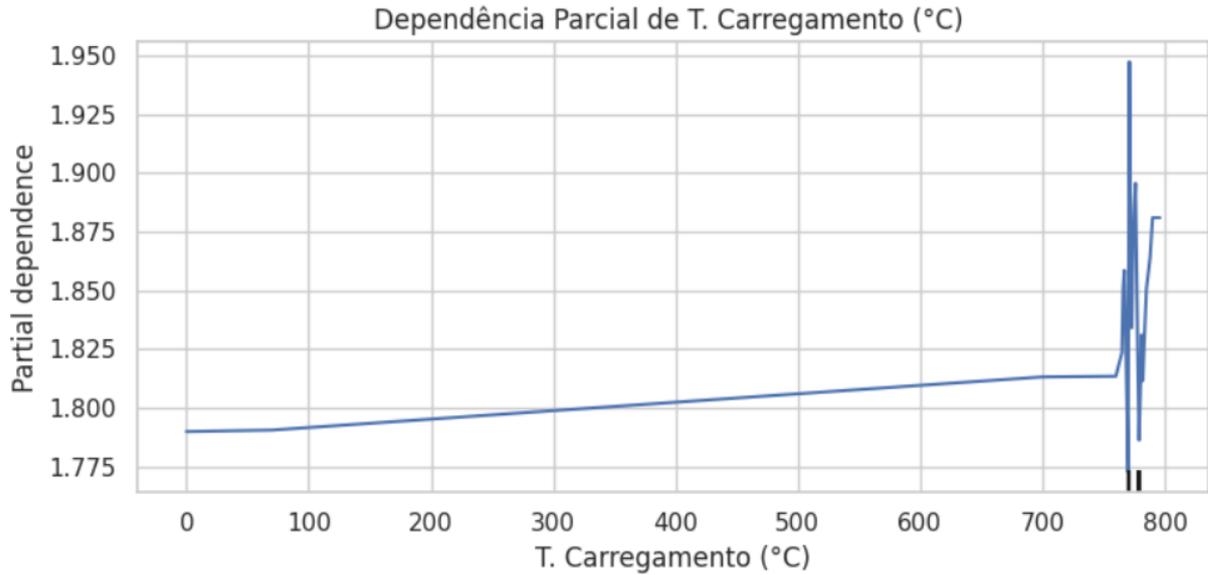
Figura 7 –Gráfico de Dependência Parcial da Temperatura de carregamento do Forno 1.



Fonte: A autora, (2025).

O Forno 2 (Figura 8), por sua vez, apresentou um comportamento de maior sensibilidade ao aumento da temperatura de carregamento. As curvas mostraram que, quanto maior a temperatura, maior a tendência de elevação da PF. Essa resposta sugere que o forno possui menor capacidade de dissipação térmica ou opera com sucatas mais suscetíveis à volatilização em altas temperaturas — um efeito citado por Liu et al. (2021), ao observar a perda de elementos como zinco e magnésio em ligas secundárias. As faixas ideais de operação, de acordo com a dependência parcial, foram identificadas entre 760 °C e 770 °C, acima das quais a perda cresce rapidamente. Isso indica que esse forno possui uma margem térmica estreita, exigindo controle rigoroso da temperatura de entrada e da dinâmica de queima.

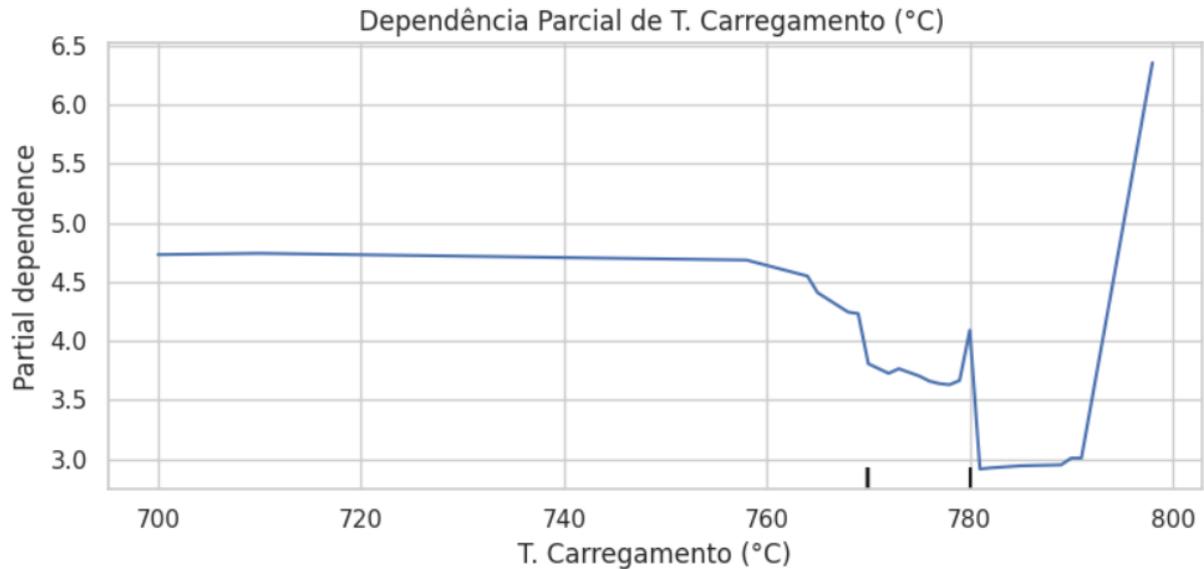
Figura 8 –Gráfico de Dependência Parcial da Temperatura de carregamento do Forno 2.



Fonte: A autora, (2025).

Já o Forno 3 (Figura 9) demonstrou um comportamento intermediário, com ótimo desempenho entre 760 °C e 790 °C, mantendo a PF em níveis baixos e estáveis. No entanto, as curvas apontam que a partir de 790 °C, há uma inflexão acentuada, com rápido crescimento da PF. Esse comportamento confirma o que é descrito por Ravi et al. (2017), que associam temperaturas superiores a 790 °C à intensificação da oxidação superficial e à elevação exponencial na formação de escória. A estabilidade do forno nessa faixa sugere a presença de um sistema mais eficiente de isolamento térmico ou composição de sucata menos reativa, fatores discutidos por Wang et al. (2021) como determinantes para o desempenho térmico em processos de fusão de metais leves.

Figura 9 –Gráfico de Dependência Parcial da Temperatura de carregamento do Forno 3.



Fonte: A autora, (2025).

A comparação entre os três fornos, conforme extraída das curvas de dependência parcial, permite observar que, apesar de todos apresentarem zonas críticas acima de 770 °C, o Forno 1 é mais tolerante à variação térmica, com estabilidade entre 680 °C e 760 °C; o Forno 2 é o mais sensível e exige controle mais rigoroso; e o Forno 3 permite operar com temperaturas mais elevadas, mas apresenta limite técnico claro em 790 °C, a partir do qual os efeitos adversos se intensificam.

Esses dados reforçam que a temperatura de carregamento deve ser ajustada individualmente para cada forno, levando em conta não apenas a temperatura nominal, mas também a taxa de aquecimento, a massa e composição da carga, e a condição atmosférica interna (Tsakiridis et al., 2013; Zhao et al., 2020).

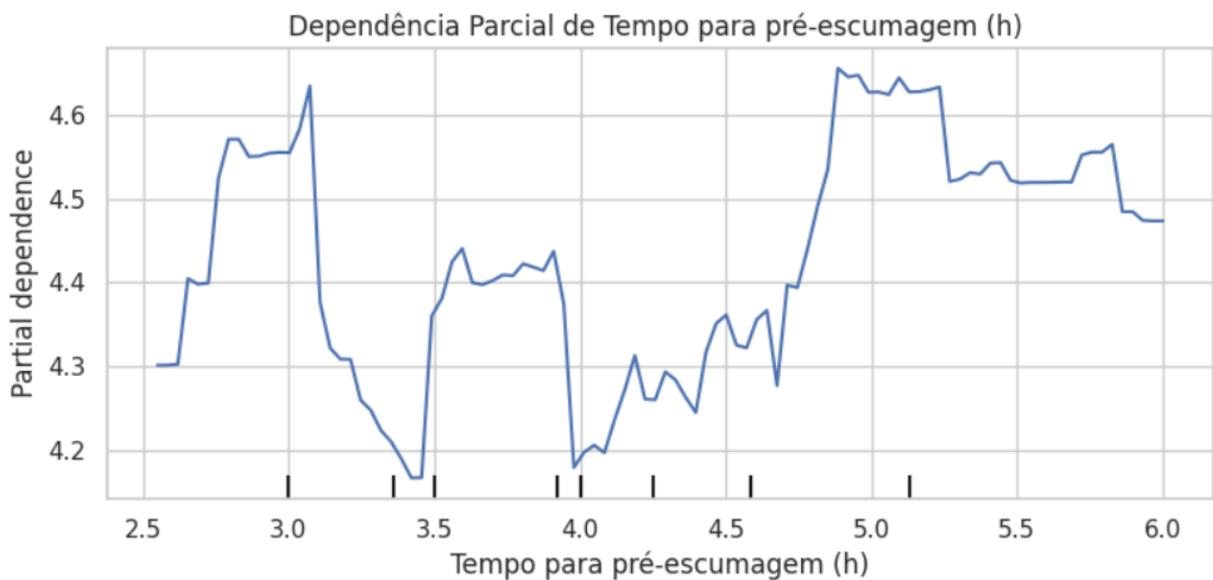
#### 4.2.4 Impacto do Tempo de Pré-escumagem sobre a Perda de Fusão

A análise das curvas de dependência parcial revelou que o tempo para pré-escumagem exerce influência significativa sobre a perda de fusão (PF), demonstrando que não apenas no tempo de retirada de escória, mas sua duração relativa ao ciclo térmico do forno pode alterar de forma substancial os índices de oxidação e reincorporação de impurezas no banho metálico.

No Forno 1, os menores valores de PF foram observados em dois intervalos distintos: entre 2,8 h e 3,4 h, e novamente entre 3,8 h e 4,7 h, como mostra a Figura

10. As curvas indicam que essas faixas representam zonas de equilíbrio em que a escória formada é retirada com eficiência, evitando o acúmulo excessivo e a reincorporação de óxidos. Fora dessas janelas, a PF aumenta consideravelmente, indicando que a retirada da escória muito precoces ou muito tardias afetam negativamente o rendimento metálico. De acordo com Zhong, Xie e Li (2022), escória que permanece tempo excessivo na superfície pode se reverter parcialmente ao banho, promovendo inclusões não metálicas e elevação da oxidação.

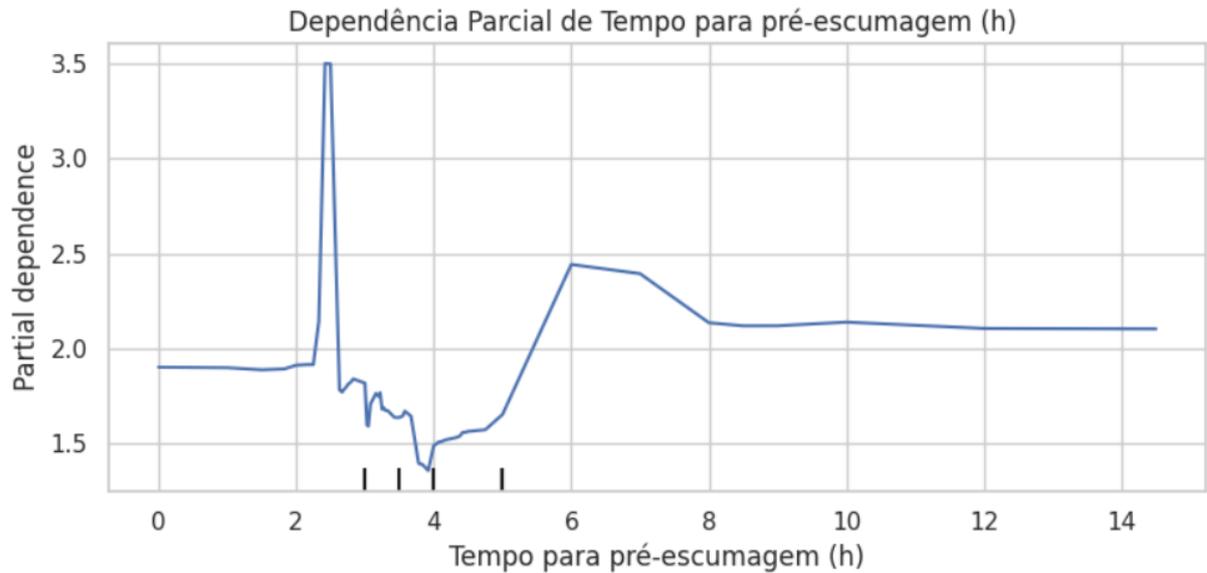
Figura 10 –Gráfico de Dependência Parcial do Tempo para pré-escumagem (h) Forno 1.



Fonte: A autora, (2025).

No Forno 2, o comportamento foi similar em termos de formato, mas com faixas ótimas ligeiramente deslocadas, como mostra a Figura 11. As curvas mostraram que tempos entre 2,8 h e 4,5 h favoreceram menores perdas, enquanto valores inferiores a 2 h ou superiores a 8 h mantiveram a PF em níveis medianos constantes. O pico negativo entre 2 h e 2,8 h indica uma zona crítica de instabilidade, possivelmente ligada a uma retirada de escória incompleta ou à retirada de escória ainda não totalmente formada. Isso reforça o que é discutido por Tsakiridis et al. (2013), que destacam a importância do tempo de maturação da escória para sua separação eficiente do alumínio líquido.

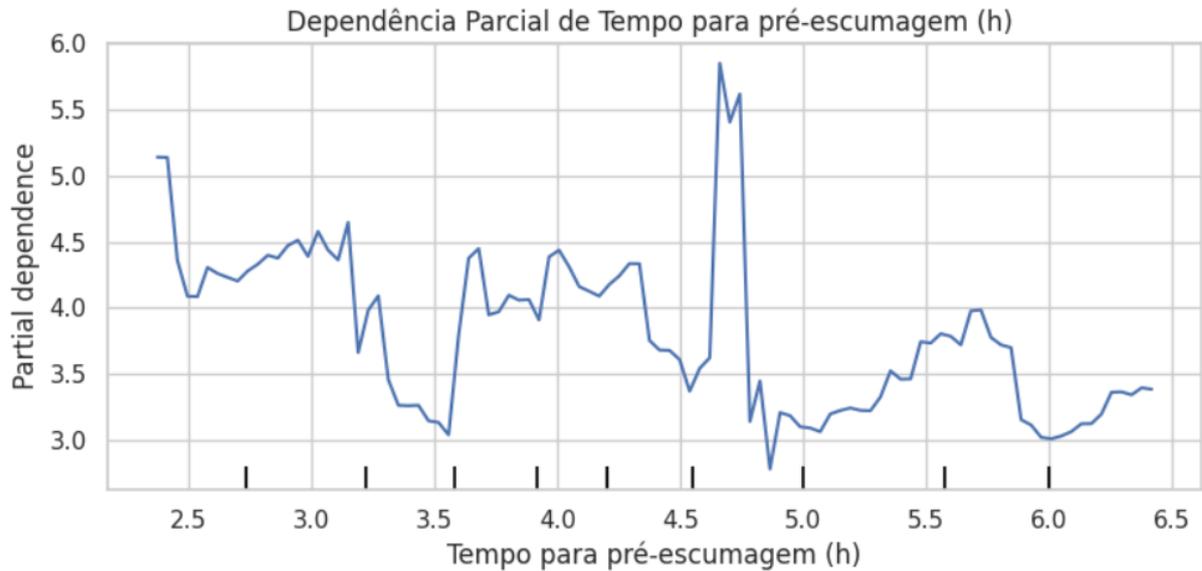
Figura 11 –Gráfico de Dependência Parcial do Tempo para pré-escumagem (h) Forno 2.



Fonte: A autora, (2025).

No Forno 3, as curvas apontaram duas faixas de melhor desempenho: entre 3,0 h e 3,5 h e entre 4,7 h e 5,5 h, como mostra a Figura 12. O padrão bimodal indica que o forno responde bem tanto a retirada de escória mais rápidas, quanto a ciclos mais longos, desde que estejam dentro de uma janela térmica equilibrada. Em faixas intermediárias, como entre 3,6 h e 4,5 h, a PF aumentou, sugerindo possível desestabilização térmica ou interação com outra variável, como temperatura ou composição da carga. Como apontam Ravi et al. (2017), a formação e estabilidade da escória metálica dependem também do tempo de permanência, sendo essencial a sincronização entre esses fatores.

Figura 12 –Gráfico de Dependência Parcial do Tempo para pré-escumagem (h) Forno 3.



Fonte: A autora, (2025).

Comparando os três fornos, observa-se que todos possuem zonas ideais entre aproximadamente 3 h e 4,5 h, mas com respostas diferentes fora dessa faixa. O Forno 1 possui maior tolerância temporal, o Forno 2 apresenta uma zona crítica precoce e o Forno 3 responde com estabilidade em janelas mais amplas, porém intercaladas. Esses resultados demonstram que o tempo de pré-escumagem não pode ser padronizado na planta, devendo ser ajustado com base no perfil térmico e composicional de cada forno.

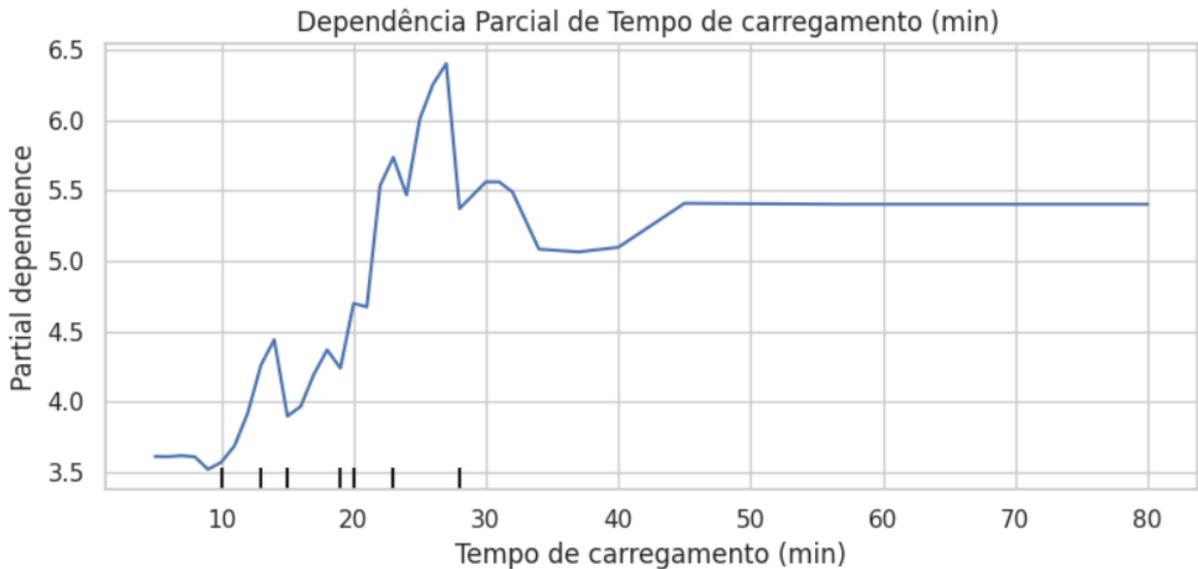
#### 4.2.5 Impacto do Tempo de Carregamento sobre a Perda de Fusão

A variável tempo de carregamento apresentou forte correlação com a PF nas curvas de dependência parcial, confirmando seu papel como um fator operacional crucial. O tempo necessário para inserir a carga no forno influencia diretamente a exposição da sucata ao ambiente oxidante e ao grau de agitação térmica, que são determinantes para a formação de escória e perda metálica.

No Forno 1 (Figura 13) observou-se que tempos de carregamento superiores a 25 minutos estão claramente associados a maiores índices de PF, enquanto tempos inferiores a 20 minutos resultaram em perdas significativamente menores. Esse comportamento é coerente com a literatura, que mostra que o prolongamento da fase de carregamento intensifica a formação de óxidos metálicos por contato prolongado com o ar quente (Zhong; Xie; Li, 2022). Além disso, tempos excessivos

de carregamento indicam lentidão na logística ou falhas no sistema de alimentação, o que pode comprometer a homogeneização térmica do banho (Ravi et al., 2017).

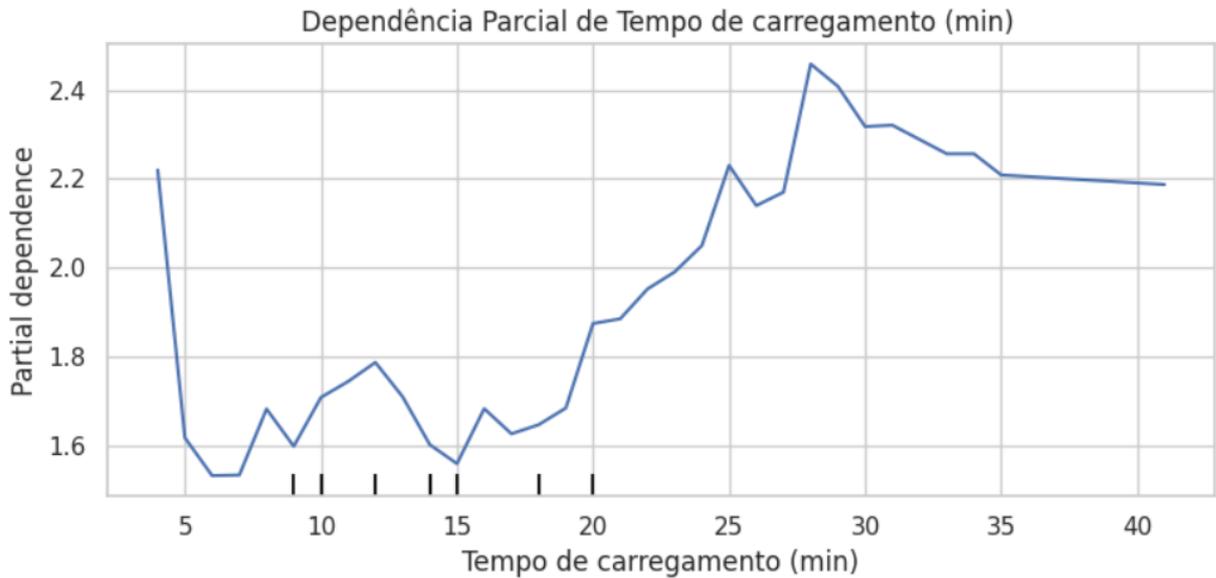
Figura 13 –Gráfico de Dependência Parcial do Tempo de carregamento (min) Forno 1.



Fonte: A autora, (2025).

O Forno 2 (Figura 14) seguiu a mesma tendência. As curvas apontam que quanto mais longo o tempo de carregamento, maior a perda de fusão, mesmo que a temperatura esteja ajustada dentro da faixa ideal. Esse resultado reforça a tese de que a perda não depende apenas do patamar térmico atingido, mas também da trajetória térmica da carga, como discutido por Silva, Siqueira e Santos (2019), que observaram maior escória em ciclos com rampas de aquecimento lentas e prolongadas.

Figura 14 –Gráfico de Dependência Parcial do Tempo de carregamento (min) Forno 2.

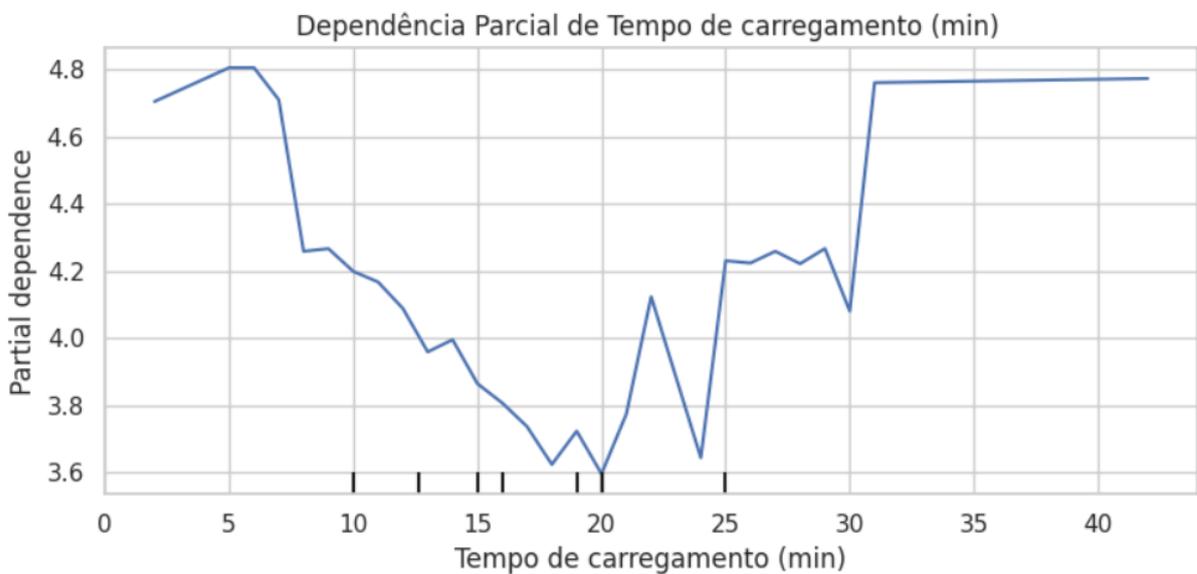


Fonte: A autora, (2025).

Já no Forno 3 (Figura 15), as curvas indicaram um comportamento mais delimitado: a zona ideal se concentrou entre 10 e 20 minutos, sendo que tempos acima de 20 minutos resultaram em crescimento progressivo da PF. Sugere-se que esse forno responde bem a ciclos rápidos de inserção de carga, provavelmente por operar em temperaturas médias mais elevadas, o que exige menor tempo para que o metal atinja o ponto de fusão. No entanto, se o carregamento é estendido, o aumento da perda é imediato, indicando alta reatividade do alumínio em suspensão térmica prolongada (Tsakiridis et al., 2013).

Observa-se, ainda, que tempos extremamente curtos, abaixo de 5 minutos, também se associaram a maiores valores de PF. Esse comportamento pode ser justificado por fatores operacionais, como carregamentos realizados de forma muito acelerada, que tendem a gerar inserção desuniforme da sucata, aumento de turbulência no banho metálico e possível choque térmico entre a carga fria e o banho líquido. Essas condições favorecem a oxidação localizada e a formação de escória, resultando em maior perda metálica, ainda que o tempo total de carregamento seja reduzido.

Figura 15 –Gráfico de Dependência Parcial do Tempo de carregamento (min) Forno 3.



Fonte: A autora, (2025).

Em tese, os dados mostram que carregamentos mais curtos estão associados a menor PF nos três fornos, e que 25 minutos parece ser um limite operacional crítico, acima do qual as perdas se acentuam. Esses achados são coerentes com o princípio de que o alumínio fundido, quando mantido por longos períodos em temperaturas elevadas, sofre oxidação acelerada e aumento de escória (Wang et al., 2021).

#### 4.3 Melhores e Piores Cenários das Variáveis para Perda de Fusão

A avaliação dos melhores e piores cenários operacionais, fornecida pelas simulações do modelo Random Forest, permitiu identificar interações críticas entre variáveis que afetam de maneira conjunta a perda de fusão. Embora os gráficos de

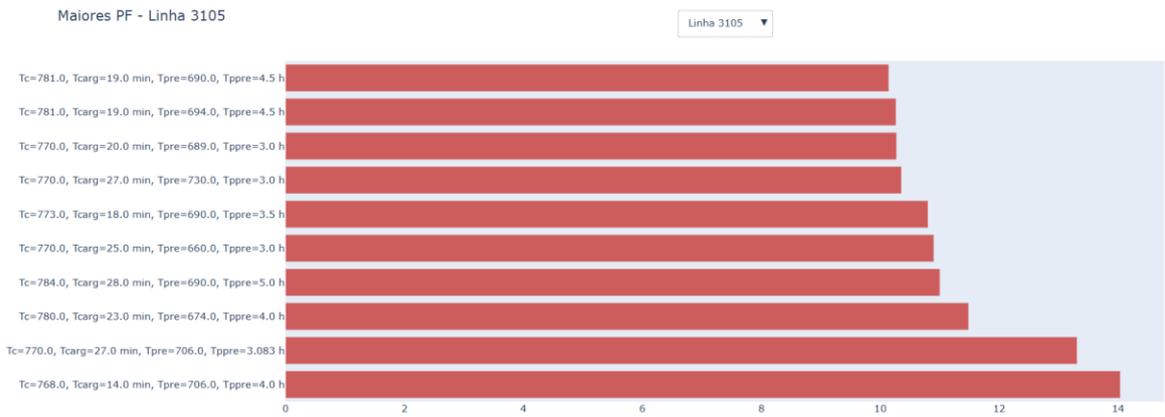
dependência parcial representem o impacto isolado de cada variável, os cenários gerados pela modelagem possibilitaram observar combinações simultâneas de condições térmicas e temporais que resultam, sistematicamente, em maiores ou menores perdas. Essa abordagem é sugerida devido a necessidade de se compreender o processo de fusão de forma integrada, já que a avaliação fragmentada dos parâmetros pode ocultar sinergias ou antagonismos operacionais importantes (Wang et al., 2021).

Para os 3 fornos o programa forneceu dados, tanto de melhores como os piores cenários para perda de fusão com uma alta sensibilidade e flexibilidade. Os valores das variáveis do processo em conjunto demonstraram intensa variação entre si, o que reforça a ideia dos autores Zhao et al. (2020) e Wang et al. (2021), que destacam a elevada variabilidade e a interdependência de variáveis do processo metalúrgicos.

Para os demais gráficos a seguir “Tc” significa temperatura de carregamento, “Tcarg” significa tempo de carregamento, “Tpre” significa temperatura de pré-escumagem e por fim “Tppre” significa temperatura para pré-escumagem.

Figura 16 –Melhores e Piores Cenários para a Perda de Fusão Forno 1.





Fonte: A autora, (2025).

Figura 17 –Melhores e Piores Cenários para a Perda de Fusão Forno 2.



Fonte: A autora, (2025).

Figura 18 –Melhores e Piores Cenários para a Perda de Fusão Forno 3.



Fonte: A autora, (2025).

No Forno 1 (Figura 16), o programa indicou que o cenário que menos gera perda de fusão é combinação entre uma temperatura de carregamento com 788 °C, um tempo de carregamento de 10 minutos, uma temperatura de pré-escumagem em 723°C e um tempo para pré-escumagem de 3,12h. Já o cenário com a maior perda de fusão associada é quando se tem uma temperatura de carregamento de 768 °C, um tempo de carregamento de 14 minutos, uma temperatura de pré-escumagem em 706°C e um tempo para pré-escumagem de 4h.

Já no Forno 2 (Figura 17), o melhor cenário foi uma temperatura de carregamento com 770 °C, um tempo de carregamento de 9 minutos, uma temperatura de pré-escumagem em 761°C e um tempo para pré-escumagem de 1,5h e uma temperatura de carregamento de 770 °C, um tempo de carregamento de

34 minutos, uma temperatura de pré-escumagem em 660°C e um tempo para pré-escumagem de 2,68h.

Por fim no Forno 3 (Figura 18), o melhor cenário foi uma temperatura de carregamento com 770 °C, um tempo de carregamento de 17 minutos, uma temperatura de pré-escumagem em 718°C e um tempo para pré-escumagem de 3,2h e uma temperatura de carregamento de 770 °C, um tempo de carregamento de 19 minutos, uma temperatura de pré-escumagem em 760°C e um tempo para pré-escumagem de 4,67h.

Os dados obtidos sugerem que, embora cada variável apresente individualmente uma faixa propícia à redução da perda de fusão, essas faixas podem não se manter eficazes quando combinadas com outras variáveis do processo, como acontece em escala real na indústria. Isso ocorre porque o comportamento do sistema é influenciado por interações simultâneas e dinâmicas entre temperatura e tempo, o que significa que o efeito de uma variável pode ser potencializado, neutralizado ou até invertido conforme o contexto operacional. Tal concepção reforça a necessidade de abordagens integradas, algo que está respaldado no que Wang et al. (2021) Wang et al. (2021, ressaltam sobre a avaliação isolada dos parâmetros poder ocultar sinergias ou antagonismos que impactam significativamente a performance metalúrgica.

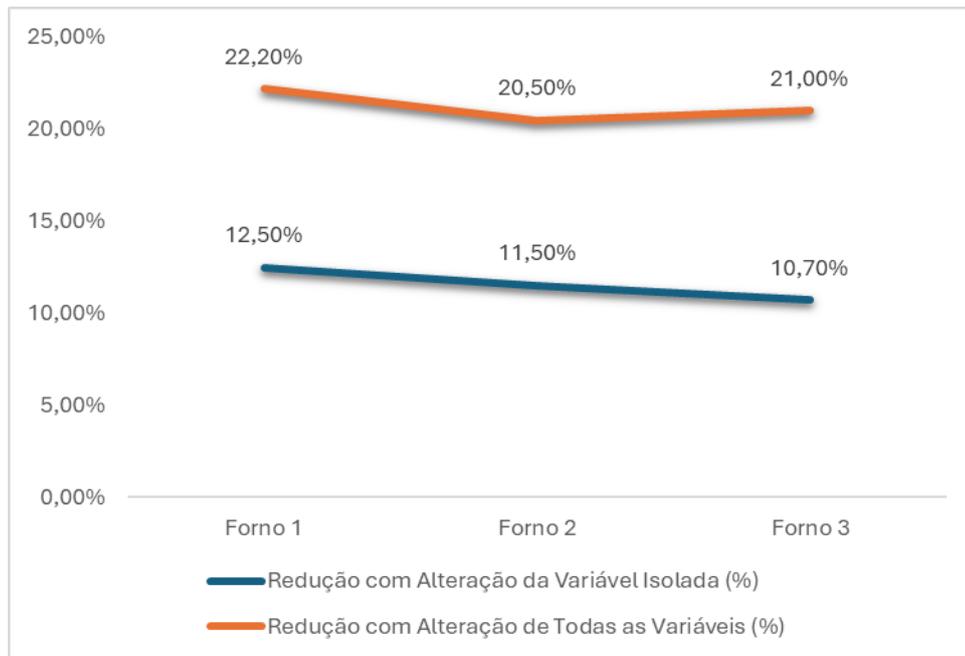
#### **4.4 Validação Operacional dos *Insights* Fornecidos pelo Programa**

Com o objetivo de verificar preliminarmente a aplicabilidade prática e a efetividade do programa computacional desenvolvido, foram conduzidos testes operacionais controlados em ambiente industrial, utilizando os três fornos avaliados neste estudo. A quantidade de testes realizadas foi feita com base no teste de hipótese estatística formulado, que recomendou 52 fusões necessárias para garantir a veracidade das recomendações dadas pelo programa. No caso em estudo para a aplicação da sugestão de melhores e piores cenários realizou-se 52 testes, já para sugestão de variáveis isoladas foram realizadas 22 fundições.

Os ensaios consistiram na comparação entre fundições realizadas sob a alteração de variáveis de maneira isolada e na alteração de todas elas, por meio da combinação dos melhores cenários recomendadas pelo modelo Random Forest.

Apesar das limitações amostrais e do curto período de observação, os testes indicaram uma minimização significativa da perda de fusão (PF) em todos os fornos analisados, como mostra a Figura 19. Quando apenas uma variável crítica — como a temperatura de carregamento ou o tempo de pré-escumagem — foi ajustada conforme a recomendação do modelo, observou-se uma redução média da PF entre 10% e 12%. Entretanto, os maiores ganhos ocorreram quando as fundições seguiram os melhores cenários combinados, considerando simultaneamente as quatro variáveis-chaves do processo (temperatura e tempo de carregamento, temperatura e tempo de pré-escumagem), resultando em reduções superiores a 20% em comparação aos procedimentos convencionais.

Figura 19 – Redução da Perda de Fusão (%) com alteração da Variável Isolada e com Todas as Variáveis.



Fonte: A autora, (2025).

Esses resultados demonstram que o programa não apenas fornece recomendações operacionais coerentes com a literatura técnica (Tsakiridis et al., 2013; Ravi et al., 2017; Wang et al., 2021), mas que também possui efetividade prática na minimização das perdas metálicas, podendo então atuar como uma ferramenta decisória para otimização do processo. A modelagem desenvolvida se mostrou capaz de capturar interações não lineares complexas, muitas vezes imperceptíveis em análises tradicionais, fornecendo orientações específicas por forno que podem levar a ganhos em rendimento e produtividade.

## 5 CONCLUSÃO

O presente trabalho demonstrou que o uso de técnicas de Machine Learning, aplicado por meio de um programa desenvolvido em Python, constitui uma abordagem promissora para a otimização da fusão de alumínio secundário. Diante da relevância industrial do alumínio, da necessidade de reduzir perdas associadas à formação de escória e da crescente demanda por processos sustentáveis e eficientes, os resultados aqui obtidos confirmam que a modelagem computacional pode ser uma aliada estratégica na transformação digital da indústria metalúrgica.

A aplicação do modelo Random Forest, selecionado por seu elevado desempenho preditivo, permitiu compreender em profundidade a complexa interação entre variáveis operacionais — como temperatura e tempo de carregamento, temperatura e tempo de pré-escumagem — e sua influência sobre a perda de fusão (PF). O modelo apresentou coeficientes de determinação ( $R^2$ ) superiores a 0,9, revelando não apenas sua robustez, mas também a natureza não linear e multifatorial do processo de fusão em escala industrial.

A análise das variáveis mais influentes e a simulação de melhores cenários operacionais demonstraram que, embora existam faixas isoladas favoráveis à redução da PF, os maiores ganhos são obtidos por meio da combinação integrada de condições operacionais otimizadas. Os testes realizados indicaram que ajustes pontuais, quando aplicados individualmente, resultaram em reduções médias de até 12% na perda de fusão. No entanto, ao aplicar todas as recomendações combinadas sugeridas pelo programa, foram registradas reduções superiores a 20%, evidenciando o potencial do sistema em promover melhorias significativas e mensuráveis no processo produtivo.

Além disso, os resultados reforçam que a avaliação isolada de variáveis pode ocultar sinergias e antagonismos que só se revelam quando o processo é analisado como um sistema contínuo e interdependente — uma perspectiva que está alinhada com a realidade fabril e com os princípios da Indústria 4.0. A capacidade do programa de oferecer recomendações específicas por forno, considerando suas particularidades construtivas e operacionais, também contribui para a personalização das decisões produtivas.

Assim, conclui-se que a modelagem computacional desenvolvida neste estudo não apenas atendeu aos objetivos propostos, como também se mostrou

capaz de atuar como ferramenta decisória para redução de perdas metálicas, podendo proporcionar uma otimização do processo e aprimoramento da previsibilidade operacional. Sua aplicação prática mostrou que um processo produtivo metalúrgico pode se tornar mais eficiente quando alinhado a uma simbiose entre a engenharia química e à ciência de dados.

### 5.1 Perspectivas para trabalhos futuros

Como perspectivas para trabalhos futuros, recomenda-se ampliar o escopo do modelo desenvolvido por meio da integração de variáveis adicionais, como composição química da sucata, granulometria, umidade e taxa de adição de fluxantes, a fim de refinar as previsões da PF e aumentar a robustez do sistema. A inclusão desses parâmetros pode permitir um diagnóstico mais preciso e sensível das interações entre matéria-prima e condições operacionais.

Sugere-se, ainda, a exploração de modelos híbridos que combinem técnicas de Machine Learning, como *Gradient Boosting* e *Support Vector Machines*, com modelos termodinâmicos e cinéticos, unindo capacidade preditiva e fundamentação físico-química. Outra linha promissora é o desenvolvimento de sistemas de análise em tempo real, alimentados por sensores industriais, viabilizando ajustes imediatos por meio de controle preditivo em malha fechada.

Adicionalmente, a simulação de cenários extremos e a incorporação de análises econômicas podem contribuir para identificar limites de segurança e mensurar o retorno financeiro das otimizações sugeridas. Por fim, o aprimoramento da personalização das recomendações por forno, aliado a técnicas explicativas avançadas como *SHAP values* e *LIME*, pode aumentar a transparência e a aceitação do sistema, consolidando a integração entre engenharia de processo e ciência de dados na indústria metalúrgica.

## REFERÊNCIAS

ASTM INTERNATIONAL. Properties and selection: nonferrous alloys and special purpose materials. 10. ed. **Materials Park (OH): ASM Handbook**, 1993.

ASSOCIAÇÃO BRASILEIRA DO ALUMÍNIO – ABAL. **Manual técnico do alumínio** – Volume 2: Processos metalúrgicos. São Paulo: ABAL, 2007.

ASSOCIAÇÃO BRASILEIRA DO ALUMÍNIO– ABAL. Reciclagem e sustentabilidade do alumínio. São Paulo: **ABAL**, 2023. Disponível em: <<https://abal.org.br/>>. Acesso em: 1 jun. 2025.

BARGAM, D. Predictive modeling in industrial processes: a comparative study. **Journal of Industrial Data Science**, v. 4, n. 2, p. 105–117, 2023.

BERTRAM, M.; MARTCHEK, K. J.; ROMBACH, G. Material flow analysis in the aluminum industry. **Journal of Industrial Ecology**, v. 24, n. 5, p. 1123–1135, 2020.

BERTRAND, V.; MARTIN, G.; MESKERS, C.E.M. **Aluminium in the circular economy**. Resources, Conservation & Recycling, v. 178, p. 106112, 2022.

BREIMAN, L. **Random forests**. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

CALEGARI, M.A. et al. Aluminum recycling and sustainability: a global overview. **Journal of Materials Research and Technology**, v. 13, p. 2934–2945, 2023.

CALLISTER JR., W. D.; RETHWISCH, D. G. Fundamentals of Materials Science and Engineering: an integrated approach. 5. ed. Hoboken (NJ): **Wiley**, 2020.

CAMPBELL, J. Complete casting handbook: metal casting processes, metallurgy, techniques and design. Oxford: **Elsevier**, 2015.

CENITA, T. **Machine learning models for predicting process losses in metallurgy**. Metallurgical Systems Review, v. 6, n. 1, p. 45–62, 2023.

CHEN, T.; GUESTRIN, C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD. **International Conference on Knowledge Discovery and Data**. Mining, 2016. p. 785–794.

CIMPAN, C. et al. Central sorting and mechanical processing of recyclable waste: a review. **Journal of Cleaner Production**, v. 107, p. 333–354, 2015.

DAVIS, J.R. Aluminum and aluminum alloys. Materials Park (OH): **ASM International**, 1993.

DOMINGOS, P. **A few useful things to know about machine learning**. Communications of the ACM, v. 55, n. 10, p. 78–87, 2012.

EUROPEAN ALUMINIUM ASSOCIATION. Aluminium and sustainability. 2023. Disponível em: <https://www.european-aluminium.eu>. Acesso em: 15 jun. 2025.

FATHI, M. et al. Oxidation and dross formation in aluminum melting: influence of furnace configuration and melt treatment. **Metals**, v. 9, n. 6, p. 660, 2019.

FRETI, G.; BORNAND, M.; BUXMANN, R. Losses in aluminum melting: slag and dross formation. **Light Metal Age**, v. 40, n. 4, p. 32–39, 1982.

GARCÍA, M. et al. Analysis of mechanical properties of aluminum alloys under permanent mold casting. **Journal of Materials Research and Technology**, v. 12, p. 318–328, 2021.

GEISSDOERFER, M. et al. The circular economy – a new sustainability paradigm? **Journal of Cleaner Production**, v. 143, p. 757–768, 2017.

GUZMÁN, J. A.; MUÑOZ, F.; CARRASCO, C. Aluminium recycling and sustainability: An overview. **Journal of Materials Research and Technology**, v. 9, n. 3, p. 3404–3415, 2020.

HARALDSSON, L.; JOHANSSON, K. Energy and emissions savings of secondary aluminum production: the Nordic example. **Resources, Conservation & Recycling**, v. 134, p. 84–93, 2018.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The elements of statistical learning: data mining, inference, and prediction. 2. ed. New York: **Springer**, 2009.

INTERNATIONAL ALUMINIUM INSTITUTE. Bauxite, alumina and aluminium production data. Londres: **IAI**, 2023. Disponível em: < <https://international-aluminium.org/> >. Acesso em: 1 jun. 2025.

INTERNATIONAL ENERGY AGENCY – IEA. Tracking industrial emissions: aluminium sector. França: **IEA**, 2023. Disponível em: < <https://www.iea.org> >. Acesso em: 1 jun. 2025.

JORDAN, M.I.; MITCHELL, T.M. Machine learning: trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255–260, 2015.

LETH-OLSEN, Håkon; NISANCIOGLU, Kemal. Filiform corrosion morphologies on painted aluminum alloy 3105 coil material. **Corrosion**, Houston, v. 53, n. 9, p. 705–713, 1997. DOI: 10.5006/1.3280457.

LI, Y. et al. Effect of temperature and holding time on dross generation during aluminum melting. **Journal of Materials Processing Technology**, v. 225, p. 1–8, 2015.

LIAW, A.; WIENER, M. Classification and regression by randomForest. **R News**, v. 2, n. 3, p. 18–22, 2002.

LIU, Y.; ZHANG, L.; CHEN, B. Heat and mass transfer modeling in aluminum recycling furnaces: effects of operational parameters. **Journal of Cleaner Production**, v. 279, p. 123456, 2021.

LU, Y.; MORRIS, K.; FRECHEL, C. Industrial big data analytics in smart manufacturing: current status and future perspectives. **Engineering**, v. 2, n. 2, p. 137–150, 2016.

PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

RAVI, A. et al. Influence of melting practices on aluminum recovery and dross generation. **Materials Science Forum**, v. 879, p. 69–74, 2017.

RAVI, R.; KUMAR, A.; SHARMA, P. Effect of melting parameters on dross formation in aluminum recycling. **Journal of Materials Research**, v. 32, n. 1, p. 55–63, 2017.

SAGI, O.; ROKACH, L. **Ensemble learning**: a survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 8, n. 4, p. e1249, 2018.

SCHLESINGER, M.E. et al. Extractive metallurgy of aluminum. 2. ed. Oxford: **Pergamon**, 2011.

SILVA, J. P.; SIQUEIRA, R. F.; SANTOS, A. C. Influência de ligas e temperaturas na fusão do alumínio reciclado. **Revista Brasileira de Metalurgia**, v. 55, n. 3, p. 112–120, 2019.

STATISTA. Global consumption of aluminum by sector 2023. New York: **STATISTA**, 2023. Disponível em: <<https://www.statista.com/>>. Acesso em: 15 jun. 2025.

SUBRAMANIAN, R.; SAHU, K.K. **Aluminium recycling and environmental benefits**: a global perspective. Resources, Conservation & Recycling, v. 142, p. 58–66, 2021.

TOTTEN, G. E. Handbook of Aluminum: Vol. 1 – Physical Metallurgy and Processes. New York: **CRC Press**, 2006.

TOTTEN, G. E.; MACKENZIE, D. S. Handbook of Aluminum: Volume 1: Physical Metallurgy and Processes. New York: **Marcel Dekker**, 2003.

TSAKIRIDIS, P. E. et al. **Aluminium salt slag characterization and utilization** – A review. Resources, Conservation and Recycling, v. 82, p. 1–10, 2013.

WANG, Q. et al. Data-driven approaches for process optimization in metallurgy: An overview. **Journal of Cleaner Production**, v. 279, 2021.

WANG, Y. et al. Machine learning applications in metal manufacturing: process optimization and fault prediction. **Computational Materials Science**, v. 190, 2021.

ZHANG, X. et al. Aluminum in the era of climate change: progress, challenges and outlook. **Journal of Cleaner Production**, v. 276, p. 123–131, 2020.

ZHAO, X. et al. Application of machine learning in metallurgical process modeling: a review. **Metallurgical and Materials Transactions B**, v. 51, n. 6, p. 3077–3094, 2020.

ZHAO, X. et al. Oxidation behavior of molten aluminum alloys under various conditions. **Journal of Alloys and Compounds**, v. 828, 2020.

ZHONG, H.; XIE, L.; LI, Y. Thermodynamic behavior of oxidation in aluminum recycling. **Journal of Materials Processing Technology**, v. 300, p. 117339, 2022.

ZHONG, Y.; XIE, J.; LI, Y. Slag removal mechanisms in aluminum refining processes. **Metallurgical Research & Technology**, v. 119, n. 1, p. 2022, 117339.

ZHU, H. et al. Thermal behavior and process optimization in aluminum melting furnaces. **Thermal Science**, v. 18, p. 1253–1260, 2014.