



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Especialização em Ciência de Dados



Predição de desgaste refratário para reparo intermediário da linha de escória de panelas de aço em uma aciaria

Alain dos Santos Batista
Lorena Bastos da Silva Schimitz
Mariana Reis Rocha

João Monlevade, MG
2023

Alain dos Santos Batista
Lorena Bastos da Silva Schimitz
Mariana Reis Rocha

**Predição de desgaste refratário para reparo intermediário da
linha de escória de panelas de aço em uma aciaria**

Trabalho de conclusão de curso apresentado ao curso de Ciência de Dados do Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto, como parte dos requisitos necessários para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Dr. George Henrique Godim da Fonseca
Coorientador: Prof. Dr. Carlos Henrique Gomes Ferreira

João Monlevade, MG

2023

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S335p Schimitz, Lorena Bastos da Silva.

Predição de desgaste refratário para reparo intermediário da linha de escória de panelas de aço em uma aciaria. [manuscrito] / Lorena Bastos da Silva Schimitz. Alain dos Santos Batista. Mariana Reis Rocha. - 2023. 57 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. George Henrique Godim da Fonseca.

Coorientador: Prof. Dr. Carlos Henrique Gomes Ferreira.

Produção Científica (Especialização). Universidade Federal de Ouro Preto. Departamento de Engenharia de Produção.

1. Siderurgia - Aciaria. 2. Materiais refratários. 3. Análise de regressão - Regressão linear múltipla. I. Rocha, Mariana Reis. II. Batista, Alain dos Santos. III. Fonseca, George Henrique Godim da. IV. Ferreira, Carlos Henrique Gomes. V. Universidade Federal de Ouro Preto. VI. Título.

CDU 669.18:519.233.5

Bibliotecário(a) Responsável: Sione Galvão Rodrigues - CRB6 / 2526



FOLHA DE APROVAÇÃO

Alain dos Santos Batista
Lorena Bastos da Silva Schimitz
Mariana Reis Rocha

Predição de desgaste refratário para reparo intermediário da linha de escória de painéis de aço em uma aciaria

Monografia apresentada ao Curso de Especialização em Ciência dos Dados da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Especialista em Ciência dos Dados

Aprovada em 30 de março de 2023

Membros da banca

Dr. George Henrique Godim da Fonseca - Orientador Universidade Federal de Ouro Preto
Dr. Carlos Henrique Gomes Ferreira - Coorientador Universidade Federal de Ouro Preto
Dr. Júlio Cesar Soares dos Reis - Universidade Federal de Viçosa
Me. Ronaldo Adriano Alvarenga Borges - Usiminas

George Henrique Godim da Fonseca, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 13/06/2023



Documento assinado eletronicamente por **George Henrique Godim da Fonseca, PROFESSOR DE MAGISTERIO SUPERIOR**, em 13/06/2023, às 15:07, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0540312** e o código CRC **ACE0EFA9**.

Agradecimentos

Agradecemos à Usiminas e à UFOP por incentivar e proporcionar a oportunidade de uma formação que nos agrega profissional e pessoalmente. Aos professores do curso, em especial, ao Prof. Dr. George Fonseca e ao Prof. Dr. Carlos Ferreira, por compartilharem seu tempo e conhecimento. Aos colegas Ronaldo Borges e Marcelo Schimitz por apoiarem na elaboração do trabalho por meio de seus conhecimentos específicos. Por fim, agradecemos aos familiares e amigos pelo apoio e compreensão durante o período do curso.

Resumo

A escassez de recursos naturais, a concorrência de grandes mercados e a demanda crescente por aços com maiores exigências de composição química e limpidez torna necessária a busca por eficiência e economia nos processos siderúrgicos. Neste contexto, a previsibilidade e gestão do consumo de materiais refratários utilizados nas painéis de transporte e tratamento do aço nas aciarias apresenta potenciais ganhos de desempenho, custos e segurança operacional. Tomando como base pesquisas bibliográficas e conhecimento técnico, o presente trabalho busca por meio do conhecimento do comportamento e influência das variáveis reais de processo e operação de uma aciaria, estabelecer um modelo matemático interpretável que realize a predição do desgaste refratário. O modelo proposto dará suporte à tomada de decisão relativa ao afastamento para reparo intermediário da linha de escória, e conseqüentemente aumento do potencial da vida útil das painéis de aço da Aciaria 2 da Usiminas Ipatinga. Por meio de técnicas estatísticas, foram identificadas as variáveis de processo de maior relevância para o desgaste da linha de escória, região mais crítica para as painéis de aço, e estabelecida uma regressão linear capaz de dar suporte às decisões quanto ao momento de afastamento das painéis para manutenção refratária intermediária. O estudo contribui para melhor previsibilidade, gestão, desempenho, custo e segurança operacional no processo de fabricação do aço na Aciaria 2 da Usiminas, Usina de Ipatinga.

Palavras-chaves: Aciaria. Desgaste refratário. Painéis de Aço. Reparo refratário da linha de escória. Regressão Linear Múltipla.

Abstract

The scarcity of natural resources, competition in major markets, and the growing demand for steels with higher chemical composition and clarity requirements make it necessary to search for efficiency and economy in steelmaking. In this context, the predictability and management of the consumption of refractory materials used in ladles for the transport and treatment of steel in steel mills offers potential benefits in terms of performance, cost, and operational safety. Based on literature research and technical knowledge, this work uses operating variables of real processes of a metallurgical plant to build an interpretable statistical model that predicts the wear of refractory materials. The proposed model supports decision making regarding the release for intermediate repair and, consequently, the useful life of the steel ladle of the Usiminas Ipatinga Steel Plant 2. Statistics were used to identify the variables most relevant to slag line wear, the most critical area of the steel ladles, and a linear regression was created to support decisions on when to release the ladles for interim refractory maintenance. The study helps improve the predictability, management, performance, cost, and operational reliability of the steelmaking process at Usiminas's Steel Plant 2, Ipatinga Mill.

Keywords: Steelwork. Refractory wear. Steel ladles. Intermediate refractory maintenance. Multiple Linear Regression

Lista de ilustrações

Figura 1 – Equipamentos e processos da Aciaria 2: Usina de Ipatinga.	5
Figura 2 – Ciclo Operacional das Painelas	7
Figura 3 – Estrutura Painela de Aço	8
Figura 4 – Intercepto em uma regressão linear simples	15
Figura 5 – Tabelas utilizadas como bases de dados	22
Figura 6 – Perfil intermediário de desgaste em painelas	24
Figura 7 – Agrupamento dos dados	25
Figura 8 – Perfil intermediário da painela 17, campanha 185	28
Figura 9 – Variáveis correlacionadas de maior representatividade para a taxa de desgaste do refratário	29
Figura 10 – Gráfico do valor real <i>versus</i> valor predito pelo modelo	34
Figura 11 – Gráfico: Quantil-Quantil Plot	35
Figura 12 – Gráfico de distribuição de pontos dos resíduos	35
Figura 13 – Gráfico de dispersão entre experimento e resíduo	36
Figura 14 – Gráfico do desgaste refratário <i>versus</i> espessura residual	37
Figura 15 – Gráfico de nº de corridas real <i>versus</i> nº de corridas preditas pelo modelo	38
Figura 16 – <i>Boxplot</i> do nº de corridas real <i>versus</i> nº de corridas predito pelo modelo	39
Figura 17 – <i>Boxplot</i> de espessura residual das painelas	40
Figura 18 – Descrição Variáveis: Gerais, Temperatura, Mix, Rota e Tempos	45
Figura 19 – Descrição Variáveis: Condições Painela, Carga e Uso do Plug	46
Figura 20 – Descrição Variáveis: Código Tratamento	47
Figura 21 – Descrição Variáveis: Adições, Taxa e Perfil	48

Lista de tabelas

Tabela 1 – Etapas da Análise Exploratória	13
Tabela 2 – Exemplo de relacionamento entre corrida e panela de aço	23
Tabela 3 – Variáveis utilizadas	23
Tabela 4 – Variáveis transformadas através de <i>One Hot Encoding</i>	26
Tabela 5 – Resultados obtidos pelo modelo de regressão linear múltipla proposto	33
Tabela 6 – Agrupamento das variáveis	49

Sumário

1	INTRODUÇÃO	1
1.1	Objetivos	2
1.1.1	Objetivos específicos	2
1.2	Organização do trabalho	3
2	REVISÃO DA LITERATURA	4
2.1	A indústria siderúrgica e os materiais refratários	4
2.2	Material refratário em painéis de aço	5
2.3	Desgaste refratário em painéis de aço	6
2.4	Papel da ciência de dados na extração de informação útil	10
2.5	Tipos de análise dos dados	11
2.6	Regressão linear múltipla	14
2.6.1	Normalidade, homocedasticidade e independência de resíduos	17
2.6.2	Multicolinearidade	18
2.6.3	Crêrios para seleção de variáveis	19
3	METODOLOGIA	21
3.1	Ambiente computacional e ferramentas	21
3.2	Coleta de dados	21
3.3	Amostragem	24
3.4	Pré-processamento dos dados	25
4	RESULTADOS	29
4.1	Análise de correlação de variáveis	29
4.2	Modelo de regressão linear múltipla	30
4.2.1	Validação do Modelo	34
4.3	Aplicabilidade e discussões	37
5	CONSIDERAÇÕES FINAIS	41
5.1	Limitações e trabalhos futuros	42
	REFERÊNCIAS	43
	APÊNDICE A – DESCRIÇÃO COMPLETA DAS VARIÁVEIS	45
	APÊNDICE B – AGRUPAMENTO DAS VARIÁVEIS	49

1 Introdução

A aciaria é a unidade de uma usina siderúrgica onde o gusa líquido, produto proveniente dos Altos Fornos, e sucata são transformados em diferentes tipos de aço. Na siderúrgica Usiminas, planta de Ipatinga, todo o transporte e tratamento do aço líquido entre o refino primário e o lingotamento, ocorre em reatores conhecidos como painéis de aço, constituídas por uma carcaça metálica revestida internamente por tijolos refratários que sofrem um desgaste natural ao longo da campanha em função das condições do processo, sendo necessário a realização de reparos e substituição de parte dessa estrutura refratária.

De acordo com [Borges \(2016\)](#), a vida útil do material refratário que reveste as painéis, utilizadas nas aciarias, depende de algumas variáveis de processo, tais como: temperatura de vazamento do aço, tempo de painel cheio, da produção do forno painel na produção total da aciaria, adições (cal, desoxidantes, ligas, escorificantes, fluidificantes), tempos de tratamento no refino secundário, entre outras.

Na aciaria da Usiminas de Ipatinga, o método atual para definir o momento de afastamento das painéis para reparo ou demolição do refratário, tem como referência a experiência dos responsáveis pelo processo, a partir de valores fixos do número de corridas transportadas, o que pode levar ao afastamento precoce ou tardio de uma painel, gerando possível ambiente de risco e/ou custo acima do necessário e performance abaixo do potencial.

Além do custo do material refratário em si, as frequentes paradas para manutenção de refratários podem interromper as atividades operacionais, concorrendo indiretamente para elevar os custos de produção do aço ([QUINTELA, 2003](#)).

Do ponto de vista do setor de Custos da Usiminas, os materiais refratários representam em torno de 4,5% do custo do aço. Sendo que na aciaria, corresponde a 33% do custo operacional. Com uma vida útil média de 110 corridas e produção mensal média de 1.200 corridas, as painéis de aço representam em torno de 25% do custo total com refratário da Aciaria 2, que responde por aproximadamente 70% da produção de aço da planta de Ipatinga.

Tendo em vista a relevante participação do material refratário no custo do aço, a racionalidade do seu uso poderá gerar economia financeira, além de segurança operacional, pois um furo de painel pode interromper a operação de uma máquina de lingotamento ou provocar perdas irreparáveis à vida humana. Dessa forma, as principais implicações práticas de explicar os aspectos que mais influenciam no desgaste refratário das painéis de aço da Aciaria 2 da Usiminas e fornecer uma ferramenta preditiva de apoio à tomada de decisão quanto à sua vida útil são:

- **Gestão:** a partir do momento em que se tem uma ferramenta de apoio sobre o nível de desgaste das panelas, os recursos podem ser direcionados de forma a otimizar sua disponibilidade e intervenção, de acordo com as necessidades do processo;
- **Performance / Custos:** com a melhoria da gestão e previsão sobre a vida da panela, há um potencial de aumento da vida média das panelas, com economia de recursos materiais e mão de obra para a execução do serviço de revestimento;
- **Segurança operacional:** a deficiência na previsão e gestão do desgaste refratário da panela de aço pode gerar acidentes no ambiente industrial, caso ocorra um furo na panela que transporta o aço a altas temperaturas, com risco às pessoas e possíveis impactos de perda de produção.

Portanto, foi desenvolvido um modelo de previsão do desgaste do refratário da linha de escória (região mais crítica) por meio do uso de regressão linear, que auxiliará a equipe da Aciaria 2 da Usiminas na avaliação da vida útil de cada panela conforme os valores das variáveis de processo às quais foi submetida. Com base na sugestão do modelo, o operador contará com melhor suporte para decidir pelo afastamento ou não da panela para reparo intermediária do refratário. Dessa forma, o modelo proposto pode contribuir na redução de custos, na melhoria da performance e gestão e no aumento da segurança operacional do processo produtivo.

1.1 Objetivos

O objetivo deste trabalho é desenvolver um modelo de regressão linear capaz de prever o desgaste do refratário da linha de escória das panelas de aço até o reparo intermediário, com base nas variáveis de processo às quais foi submetida, e apoiar a tomada de decisão quanto ao afastamento das panelas.

1.1.1 Objetivos específicos

Para alcançar o objetivo geral proposto, os seguintes objetivos específicos devem ser alcançados:

- Identificar as variáveis do processo às quais a panela é submetida;
- Avaliar a relevância de cada variável no desgaste das panelas;
- Utilizar técnicas estatísticas para escolher as variáveis que mais contribuem para o desgaste;
- Modelar o processo em um modelo de regressão múltipla;

- Validar o modelo estatisticamente e contextualmente, isto é, por meio da comparação dos resultados do modelo com o desgaste real apurado na análise do perfil de demolição da panela.

1.2 Organização do trabalho

O restante do trabalho está organizado da seguinte forma: no Capítulo 2 serão discutidos a relevância do tema e conceitos com base em estudos da literatura; no Capítulo 3 são apresentados os métodos de pesquisa, coleta e análise de dados; o Capítulo 4 apresenta os resultados e análises obtidos na pesquisa, e; por fim, o Capítulo 5 apresenta as reflexões e conclusões finais do trabalho.

2 Revisão da literatura

Neste capítulo será apresentado um debate sobre os temas estudados e necessários para desenvolvimento do trabalho. Para isso, essa revisão será organizada em subseções, conforme segue.

2.1 A indústria siderúrgica e os materiais refratários

A indústria siderúrgica é um setor de notável importância para a economia mundial. Os produtos siderúrgicos são produtos intermediários, utilizados principalmente pelas indústrias de bens de capital, automotiva e de construção civil, e, por esse motivo, segundo CADE (2022), bastante sensível ao desempenho da indústria como um todo.

De acordo com *World Steel Association*¹, em 2021, a indústria siderúrgica produziu cerca de 1,95 bilhões de toneladas de aço bruto, volume 26,8% superior à produção registrada em 2011. Com a produção anual de 36,2 milhões de toneladas, o Brasil ocupa a 9ª posição no *ranking* da produção mundial.

A escassez de recursos naturais, a concorrência de grandes mercados e a demanda crescente por aços com maiores exigências de composição química e limpidez, torna necessário a busca por eficiência e economia nos processos. Neste contexto, encontram-se os materiais refratários, que, de acordo com a área de Custos, representam em torno de 4,5% do custo do aço produzido pela Usiminas, na planta de Ipatinga. Os materiais refratários são estruturas que revestem os diversos reatores das instalações industriais e que dão suporte à fabricação de produtos em condições severas de temperatura, pressão e outros mecanismos de desgaste.

De acordo com a Associação Técnica de Refratário, refratários são materiais não metálicos, desenvolvidos para resistirem a altas temperaturas, abrasão mecânica, corrosão e erosão química, causadas por metal e escória líquida (REFRACTORIES, 1998). De acordo com o *Market Reseach Report* (2022)², a indústria siderúrgica demanda a maior demanda do mercado refratário (59,8%), com aplicação em fornos, reatores e vasos utilizados na transformação do aço.

¹ *World Steel Association* - <<https://worldsteel.org/steel-topics/statistics/world-steel-in-figures-2022/>>

² *Market Reseach Report* - <<https://www.fortunebusinessinsights.com/refractories-market-103287>>

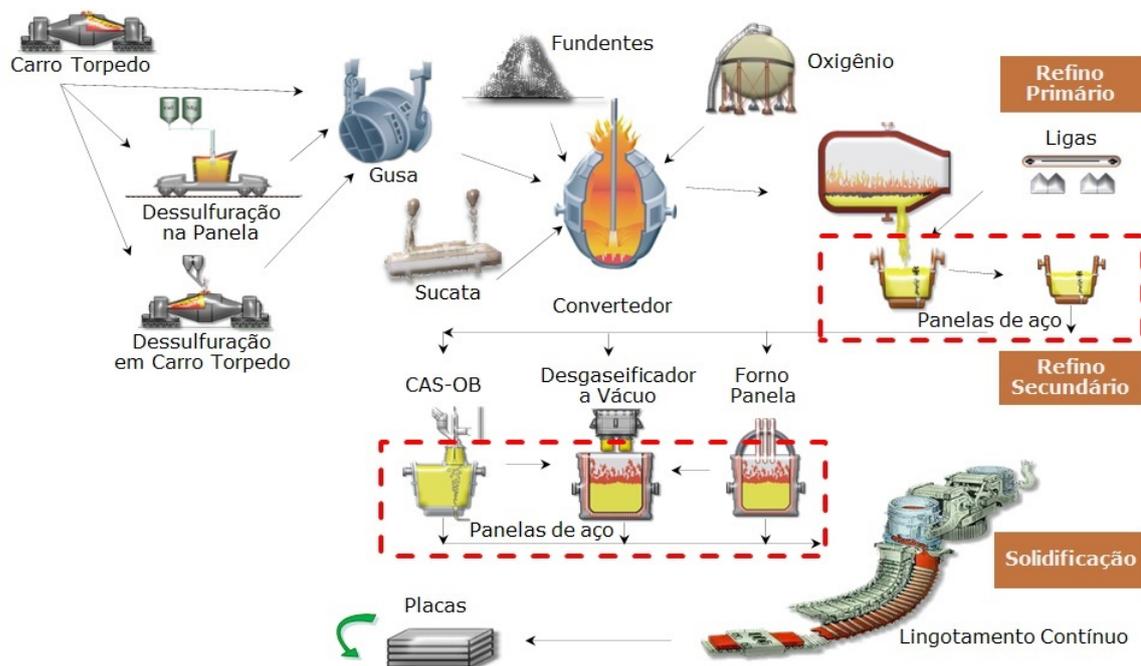
2.2 Material refratário em panelas de aço

A aciaria é a unidade produtiva dentro de uma indústria siderúrgica em que o gusa líquido, proveniente do processo de redução do minério de ferro nos altos fornos e sucata de aço são refinados, de maneira a remover impurezas e realizar adições que confirmam ao aço, produto desse refino, as propriedades físicas e mecânicas requeridas pelos clientes no produto final.

Na aciaria da Usina de Ipatinga, os materiais refratários correspondem a aproximadamente 33% do custo operacional. Dentre diversas aplicações, as chamadas panelas de aço são responsáveis por parcela significativa (25%) do custo da Aciaria 2 com refratários. Estas panelas são constituídas de uma carcaça de ferro fundido, revestidas por tijolos refratários que lhes conferem as propriedades necessárias para as demandas do processo, que consiste em transporte e tratamento do aço líquido nos diversos reatores das etapas de fabricação e refino do aço, até o lingotamento contínuo, em que o aço é solidificado em forma de placas.

A Figura 1 apresenta o modelo esquemático dos equipamentos e processos existentes na Aciaria 2 da Usiminas, área de aplicação do estudo. Em destaque, as panelas de aço e os processos pelos quais elas são submetidas:

Figura 1 – Equipamentos e processos da Aciaria 2: Usina de Ipatinga.



Fonte: Usiminas

As panelas de aço são os recipientes que têm maior tempo de contato com o aço líquido, requerendo do revestimento refratário propriedades químicas, físicas e termomecânicas que garanta o desempenho requerido no processo no que tange à qualidade do aço, segurança operacional, custo e disponibilidade otimizados.

A Aciaria 2 possui disponíveis 18 panelas de aço, sendo em média, 9 em operação, 2 em reserva e 7 em reparo. As panelas possuem, em projeto, capacidade para 187 toneladas de aço e escórias geradas nos processos de refino, sendo o peso médio real do aço transportado de 166,7 toneladas. A variabilidade do volume transportado tem relação com os requisitos operacionais relativos ao tipo de aço e equipamento de refino utilizados.

O ciclo operacional da panela de aço inicia com o recebimento do aço vazado dos convertedores (etapa final do refino primário), também chamado de corrida. Durante o vazamento são adicionadas ligas metálicas, desoxidantes, escorificantes e fluidificantes, de acordo com as exigências do aço final a ser fabricado.

Após o vazamento do aço, as panelas são transportadas aos reatores do refino secundário, onde são feitos os ajustes químicos e de temperatura. O refino direcionado depende do tipo de aço e exigência do produto, sendo os equipamentos disponíveis na Aciaria 2: um Forno Panela, dois CAS-OBs e dois Desgaseificadores à vácuo / RHs.

Posteriormente à etapa de refino secundário, a panela é enviada à torre de uma das duas máquinas de lingotamento, local no qual o aço líquido será vazado das panelas para as estações de solidificação em formato de placas. Durante este processo de esgotamento da panela, o refratário fica exposto à atmosfera.

Em seguida, a panela com residual de aço e com a escória remanescente do processo é transportada para o basculamento da escória e posicionada para limpeza e remoção de cascão (aço e escória) via removedor hidráulico. Este processo, embora desempenhado por máquina, depende da operação humana e pode gerar impactos mecânicos para o refratário, tal como trincas e arrancamento de tijolos.

Após a limpeza, a panela é levada à área de preparação, local em que são verificadas as condições refratárias da mesma, realizados os padrões de troca e eventuais projeções de massa refratária para o retorno aos convertedores, onde receberão a próxima corrida, reiniciando o ciclo operacional, conforme a Figura 2.

Na próxima subseção serão abordadas as variáveis que influenciam no desgaste refratário em panelas de aço.

2.3 Desgaste refratário em panelas de aço

Segundo Leite, Luz e Pandolfelli (2014), durante o ciclo operacional, as panelas de aço são expostas a diferentes solicitações de origem térmica, mecânica e química, as quais exercem forte influência no desempenho final do refratário. Por este motivo, Lenz e Silva (2007), destaca as principais propriedades e características necessárias para a aplicação dos materiais refratários a condições extremas de processo:

Figura 2 – Ciclo Operacional das Painelas



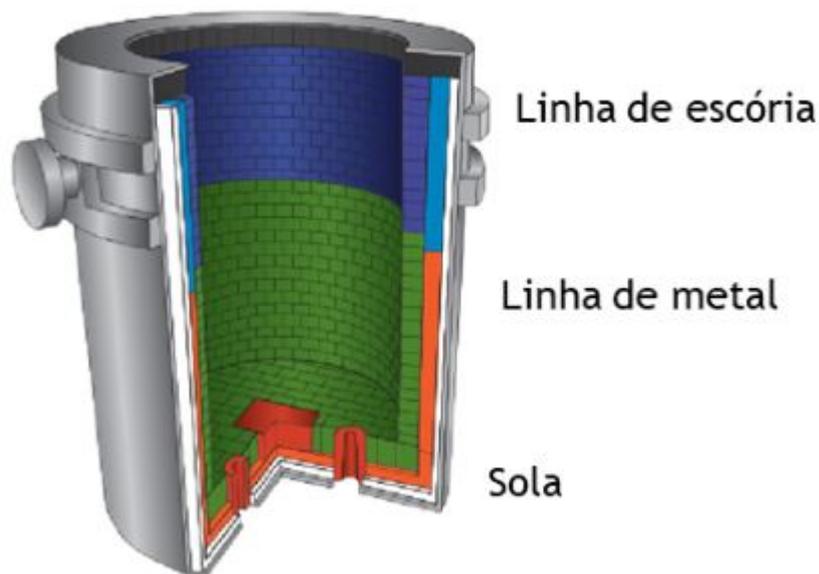
Fonte: Elaboração Própria

- Densidade: interfere diretamente em outras propriedades tais como condutividade térmica e resistência mecânica;
- Porosidade: está relacionada à resistência à corrosão, resistência à penetração de gases, escória e metal líquido, e resistência à erosão;
- Módulo de elasticidade: relação entre tensão e deformação;
- Refratariedade: resistência a altas temperaturas;
- Resistência à abrasão e corrosão: resistência ao atrito com material particulado e contato com fluidos em movimento;
- Resistência mecânica (compressão e flexão);
- Resistência mecânica sob carga em elevada temperatura;
- Resistência a oxidação (refratários com carbono): perda de massa devido reação química com oxigênio;
- Resistência à corrosão por fundentes e escórias: ataque químico durante a etapa de transporte e transformação do aço;
- Resistência ao choque térmico;

- Expansão térmica e dilatação térmica reversível;
- Condutividade térmica: capacidade de retenção de calor durante os processos metalúrgicos.

Comumente, as panelas de aço têm seu projeto de revestimento dividido por zonas, de acordo com as solicitações físicas e químicas de cada região conforme a Figura 3.

Figura 3 – Estrutura Painel de Aço



Fonte: [Saint Gobain](#)

A linha de escória, foco do trabalho, é a região do revestimento que fica em contato direto com a escória, sendo a área de maior solicitação da panela (corrosão, erosão, choque térmico, entre outras), e, portanto, o que em geral, determina a vida da mesma.

A linha de metal é a região da panela que fica em contato com o aço durante o refino secundário. Durante uma pequena parcela de tempo do processo, no lingotamento contínuo, esta região fica em contato com a escória.

A sola refere-se à base da panela, onde localiza-se a zona de impacto, área a qual durante o vazamento recebe o impacto do jato de aço. Nesta região também estão instalados os componentes responsáveis pela injeção de gás inerte no aço (plug poroso) para homogeneização térmica e química do aço, eventuais processos de dessulfuração e globulização de inclusões, e abertura da panela para lingotamento (válvula gaveta).

Ao longo do tempo, os materiais refratários têm sido objeto de pesquisa e desenvolvimento, e por meio de inovações tecnológicas adquiriram maior resistência e qualidade. Uma das ferramentas mais úteis, de acordo com [Lenz e Silva \(2007\)](#), para entendimento dos mecanismos de desgaste e caracterizações sistemáticas dos materiais é o estudo *post-mortem* dos refratários, realizado após o fim de vida útil do refratário, com base na medição do perfil final de desgaste refratário (espessura dos tijolos remanescentes), coleta de amostras e realização de testes (mecânicos, químicos, físicos, microscopia e microanálise, etc.). Entretanto, mesmo com toda a evolução, o desgaste refratário é inerente ao processo metalúrgico.

Por este motivo, na Aciaria 2 da Usiminas, o refratário das painéis de aço passa por um ciclo de vida padrão, segundo o qual, em torno de 70 ciclos operacionais (corridas) a painél é retirada de operação e passa por substituição da linha de escória. Este processo é denominado reparo intermediário. Ao retornar ao processo, a painél realiza em torno de 50 corridas adicionais, e então é afastada para a substituição de todo refratário. Sendo assim, a vida média de uma painél de aço na Aciaria 2 é de 110 corridas.

O número de corridas foi determinado de forma empírica, baseada nos perfis finais de demolição e com base no conhecimento tácito da equipe de especialistas da área. Entretanto, devido à variabilidade das condições do processo, e a inexistência de uma ferramenta de apoio à tomada de decisão, há casos em que: 1) após inspeção durante a etapa de Preparação das painéis, a painél é afastada de maneira antecipada devido à identificação de desgaste superior à média, a fim de garantir a segurança do processo e das pessoas; 2) a painél é afastada de acordo com o padrão estabelecido, mas durante a demolição identifica-se que havia potencial de desempenho superior. Especialmente para o segundo caso, há um subaproveitamento do recurso, e conseqüente não otimização de custos. Por outro lado, no primeiro caso poderia haver uma antecipação para não correr o risco de situações extremas de desgaste ou gerenciar de forma a otimizar a disponibilidade da painél e recursos para intervenção, de acordo com a conveniência do processo. Segundo [Lenz e Silva \(2007\)](#), a performance (vida) refratária, está relacionada basicamente à três grandes fatores:

- Engenharia refratária: relativo ao projeto e aplicação dos materiais, engenharia microestrutural, processamento, reciclagem e disposição;
- Processos e operação: carga recebida, temperaturas e gradientes aos quais é submetido, volume e vazão de gases injetados, adições e escorificantes, tempos de aquecimento, tempos de tratamento, entre outras variáveis operacionais;
- Manutenção: inspeção e reparo refratário, práticas de limpeza de cascões e escória, controle térmico dos refratários em ciclo, posicionamento das lanças de injeção de gás e pós, entre outros.

Tomando-se a premissa de que a engenharia refratária está adequada aos processos requeridos pela Aciaria 2, e dada a indisponibilidade de dados confiáveis relativos à manutenção refratária, o presente trabalho busca compreender a influência das variáveis referentes ao processo e operação e com base nisso, construir um modelo (equação) que estime a taxa de desgaste refratário das painéis de aço.

2.4 Papel da ciência de dados na extração de informação útil

Nos últimos anos, nota-se que as empresas têm gerado um volume significativo e contínuo de dados. [Rautenberg e Carmo \(2019\)](#) indicam que é importante desenvolver ferramentas que extraiam informações para promover o conhecimento no processo de tomada de decisão, e completam que nessa transformação dos dados em informação, e, posteriormente, em conhecimento, é exigida a sinergia de competências interdisciplinares e de especialistas com o objetivo de extrair informação útil por meio de bases de dados complexas. Nesse contexto, surge então uma nova área de atuação chamada ciência dos dados.

Conceitualmente, [Pacheco e Disconzi \(2019\)](#) explicam que a ciência de dados tem como foco estudar o ciclo de vida dos dados e gerar valor comercial por meio de informações provenientes dos dados. Já para [Grady e Chang \(2015\)](#), a ciência dos dados diz respeito à condução da análise de dados de maneira empírica possibilitando o aprendizado direto através dos próprios dados.

Baseados no estudo e tradução de [Bugnion, Manivannan e Nicolas \(2017\)](#), [Rautenberg e Carmo \(2019\)](#), descrevem que o ciclo de vida da ciência dos dados possui sete etapas:

1. Obtenção de dados: avaliação e seleção de dados;
2. Ingestão de dados: transformação dos dados e agrupamento em uma única base;
3. Exploração de dados: estudos iniciais para permitir o relacionamento com as informações requeridas;
4. Definição dos parâmetros: conversão dos dados de entrada para manipulação, conforme algoritmo de manipulação; transformação dos dados de forma a permitir que a saída seja legível; definição dos intervalos; critério de parada e confiabilidade da resposta. Todos os itens estão ligados ao conceito de Aprendizagem de Máquina;
5. Implementação do modelo: empregar, de forma iterativa, os parâmetros mais adequados com o objetivo de ter um modelo estatisticamente representativo;
6. Utilização do modelo: uso do modelo para inferências com os dados em um ambiente real.
7. Tomada de decisão: combinação entre os resultados do modelo e as expertises do especialista do processo, auxiliando na tomada de decisão.

Matos (2015) afirma que a grande quantidade de dados que são gerados e coletados auxiliam as empresas a ganharem vantagem competitiva pela exploração destes recursos e que, quando são desafiadas com um problema de negócio, devem ter a capacidade de avaliar esses dados e entender como eles podem contribuir positivamente para um melhor desempenho. Para isso, Matos (2015) cita que o cientista de dados deve entender os conceitos, definir uma estrutura e organizar o pensamento de maneira analítica com o objetivo de criar oportunidades ou identificar ameaças que irão contribuir na tomada de decisões baseada em dados. Ainda, Rautenberg e Carmo (2019) pontuam que, o inter-relacionamento entre as três principais áreas da ciência dos dados - programação de computadores, estatística e matemática e domínio do conhecimento – são habilidades importantes para os cientistas de dados. Manyika et al. (2011) indicam ainda três oportunidades de atuação para o cientista de dados - análise de dados, tomada de decisão e suporte tecnológico.

Dessa forma, a ciência dos dados e os cientistas de dados, juntamente com os tipos de análise dos dados e a análise exploratória, que serão descritas na próxima seção, possuem um papel significativo na identificação de oportunidades para as empresas no que diz respeito à transformação dos dados em conhecimento. No presente trabalho, foi identificada a oportunidade de melhorar o controle e gestão dos refratários das panelas de aço da aciaria. Por meio de uma equipe multidisciplinar e seguindo as primeiras 5 etapas citadas foi possível extrair e organizar os dados, de forma a gerar conhecimento relativo ao processo por meio da regressão linear, que poderá ser utilizada como ferramenta para tomada de decisão relativa ao uso e afastamento das panelas de aço.

2.5 Tipos de análise dos dados

O grande volume de dados gerados pelas empresas tem motivado o desenvolvimento de metodologias e ferramentas com o objetivo de agregar valor ao negócio e permitir mais uma alternativa para apoiar a tomada de decisão. Ferreira et al. (2021) indicam que para gerar resultados confiáveis e consistentes, é necessário, para a análise dos dados, além da manipulação e cruzamento dos dados, a utilização de conceitos matemáticos, métodos estatísticos e algoritmos que irão auxiliar na descoberta de padrões. Os autores reforçam ainda que um único método de análise de dados e acesso podem não ser suficientes para alcançar os resultados almejados.

De acordo com Neto e Marques (2020), a manipulação de um grande volume de dados, de diversas fontes, pode ser realizada através de metodologias de análise matemática e estatística com o objetivo de agregar valor ao processo de decisão e transformar os dados brutos em informação. Sendo assim, existem quatro tipos de análises: descritiva, diagnóstica, preditiva e prescritiva. De forma resumida, conforme Neto e Marques (2020):

- **Análise descritiva:** tem como objetivo empregar técnicas para sumarizar e/ou sintetizar os dados, permitindo que sua descrição fundamente decisões. Além disso, é demonstrado que, essa análise é considerada por muitos como uma etapa inicial, utilizada como base para as demais análises, e caso mal construída, os erros poderão se propagar ao longo das demais análises.
- **Análise diagnóstica:** o objetivo dessa análise é conduzir um diagnóstico da situação atual não se limitando apenas na descrição detalhada dos dados, mas também em definir uma relação de causa e efeito e entender como os fatos do passado influenciam a situação atual.
- **Análise preditiva:** tem como objetivo prever, conforme determinado grau de certeza, o futuro através da busca de relações de causa e efeito dos acontecimentos atuais, permitindo, por exemplo, que as empresas se antecipem e executem ações para evitar ou potencializar os prováveis acontecimentos.
- **Análise prescritiva:** assim como a preditiva, a prescritiva tem foco no futuro. A diferença consiste em determinar as consequências de uma prospecção ou acontecimento. Essa análise pode indicar, por exemplo, mudança de hábito do cliente ou uma ameaça a determinado tipo de negócio, que não esteja de acordo com a prospecção indicada.

A utilização dessas análises auxilia no processo de transformação dos dados em informação e posteriormente em conhecimento. [Ferreira et al. \(2021\)](#) evidenciam que os dados brutos podem variar significativamente em tipos e formatos e por isso, é importante transformar os dados com o objetivo de realizar a análise de maneira coesa. Essa transformação pode envolver mudanças nos tipos dos dados existentes ou necessidade de criação de novas variáveis.

Os mesmos autores indicam a relevância da análise exploratória para averiguar os dados para aplicação estatística, obtenção de entendimento e estabelecimento de relações entre as variáveis analisadas. A Tabela 1 indica as etapas de uma análise exploratória considerada pelos autores como um roteiro para evitar distorções nas aplicações.

Tabela 1 – Etapas da Análise Exploratória

Etapa	Descrição
Coleta dos dados	Definição de quais dados serão usados e como serão coletados.
Organização dos dados	Permitirá obter as informações a partir dos dados e nessa etapa é importante saber a natureza dos dados para que façam sentido para estudo a ser realizado podendo, por exemplo, unificar ou separar colunas de uma tabela.
Tratamento dos dados	Refere-se à melhoria e adequação do conjunto de dados obtidos e nesse sentido, é importante analisar a base para tratar, por exemplo, os dados faltantes, <i>outliers</i> e normalização dos dados (permite padronizar os dados numéricos em uma mesma escala).
Análise dos dados	Consideram análise univariada – analisar a frequência de possíveis valores; e análise multivariada – estabelecer relacionamento entre duas ou mais variáveis e correlações.
Apresentação dos dados	é considerada como uma das etapas mais importantes e pode ser dividida em apresentação por meio de tabelas (apresentação tabular) ou gráficos (representação gráfica).
Interpretação dos dados	Tem como objetivo, por meio do problema levantado, definir o modelo a ser aplicado.

Fonte: Elaboração própria com base em [Ferreira et al. \(2021\)](#)

Em especial, a análise multivariada, que faz parte da análise dos dados em uma análise exploratória, conforme [Ferreira et al. \(2021\)](#), permite a análise simultânea de mais de duas fontes de informação com o objetivo de alcançar um resultado mais satisfatório. Conforme [Hair et al. \(2009\)](#), existem duas técnicas: técnica de dependência, onde os dados de entrada que alimentam o modelo precisam ser especificados, indicando quais variáveis são independentes e quais são dependentes (que se pretende prever ou explicar) com o objetivo de promover um relacionamento de causa e efeito; e técnica de interdependência, que busca descrever padrões sem determinar suposições sobre as variáveis. A regressão múltipla, que é uma análise preditiva e que foi utilizada no presente trabalho, é um exemplo de técnica dependente e por essa razão, conforme [Ferreira et al. \(2021\)](#) e [Hair et al. \(2009\)](#), permite prever as modificações e/ou relação da variável dependente (critério) com duas ou mais variáveis independentes (preditoras). Na próxima seção será dada ênfase a essa técnica, objeto do estudo.

2.6 Regressão linear múltipla

A regressão linear múltipla permite prever ou prever uma variável dependente com base em duas, ou mais variáveis independentes. [Hair et al. \(2009\)](#) completam essa definição ao afirmar que as variáveis independentes, que possuem valores conhecidos, são utilizadas para prever os valores da variável dependente. Portanto, alguns conceitos, como o coeficiente de correlação, coeficiente de determinação e critérios de validação são importantes para a definição de um modelo de regressão.

O coeficiente de correlação é explicado por [Hair et al. \(2009\)](#) como sendo fundamental na análise de regressão, pois permite associar as variáveis que são ditas correlacionadas se a modificação em uma provoca modificações em outra, ou seja, se uma variável mudar, a outra também mudará. Assim, o coeficiente de correlação (r), também chamado de coeficiente de correlação de Pearson, indica que, quanto mais altas ou fortes as correlações (mais semelhantes), teremos melhores previsões, reforçando que esse coeficiente determina a relação entre as variáveis dependentes e independentes. [Sicsú e Dana \(2012\)](#) mostram que o coeficiente de correlação pode assumir valores entre -1 e 1 , sendo uma correlação positiva quando $r > 0$, correlação negativa quando $r < 0$ e nula quando as variações de uma variável não estão associadas linearmente.

Já o coeficiente de determinação (R^2), conforme [Hair et al. \(2009\)](#), mede a força da relação e pode ser descrita como:

(...) medida da proporção da variância da variável dependente em torno de sua média que é explicada pelas variáveis independentes ou preditoras. O coeficiente pode variar entre 0 e 1. Se o modelo de regressão é propriamente aplicado e estimado, o pesquisador pode assumir que quanto maior o valor de R^2 , maior o poder de explicação da equação de regressão e, portanto, melhor a previsão da variável dependente ([HAIR et al., 2009](#)).

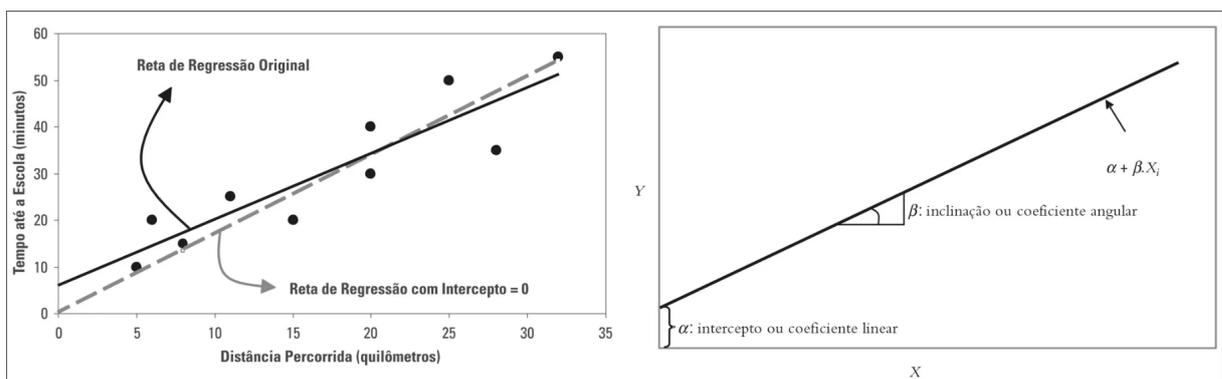
Em relação ao coeficiente ajustado de determinação (R^2 ajustado), os mesmos autores conceituam:

Medida modificada do coeficiente de determinação que considera o número de variáveis independentes incluídas na equação de regressão e o tamanho da amostra. Apesar de a adição de variáveis independentes sempre fazer com que o coeficiente de determinação aumente, o coeficiente ajustado de determinação pode cair se as variáveis independentes acrescentadas tiverem pouco poder de explicação e/ou se os graus de liberdade se tornarem muito pequenos. Essa estatística é muito útil para comparação entre equações com diferentes números de variáveis independentes, diferentes tamanhos de amostras, ou ambos ([HAIR et al., 2009](#)).

É importante ressaltar, conforme explicado por [Hair et al. \(2009\)](#), que o coeficiente de correlação (r) é utilizado para avaliar o relacionamento entre as variáveis independentes e a variável dependente e o coeficiente de determinação (R^2) representa a força dessa relação e uma melhoria na previsão de acordo com a previsão referenciada. [Favero \(2015\)](#) complementa que o R^2 aumenta quando uma nova variável é adicionada ao modelo, o que pode não ser verdade para o R^2 Ajustado que é utilizado quando se tem a intenção de comparar modelos (e nesse caso o que tiver maior R^2 ajustado é o melhor) ou um mesmo modelo com amostragem diferentes.

As previsões podem ser melhoradas por meio da utilização de uma constante de regressão denominada intercepto, conforme mencionado por [Hair et al. \(2009\)](#). Esse intercepto indica o valor da variável dependente quando as variáveis independentes tem um valor nulo. Para facilitar o entendimento, a Figura 4, mostra como funciona esse intercepto em um regressão linear simples.

Figura 4 – Intercepto em uma regressão linear simples



Fonte: Adaptado de [Favero \(2015\)](#)

Já um modelo geral de regressão linear múltipla pode ser definido, conforme Equação (2.1), demonstrada no livro de [Favero \(2015\)](#), onde:

Y representa o fenômeno em estudo (variável dependente quantitativa), α representa o intercepto (constante ou coeficiente linear), b_j , ($j = 1, 2, \dots, k$) são os coeficientes de cada variável (coeficientes angulares), X_j são as variáveis explicativas (métricas ou *dummies*) e u é o termo de erro (diferença entre o valor real de Y e o valor previsto de Y por meio do modelo para cada observação). Os subscritos i representam cada uma das observações da amostra em análise ($i = 1, 2, \dots, n$, em que n é o tamanho da amostra) ([FAVERO, 2015](#)).

$$Y_i = a + b_1X_1 + b_2X_2 + \dots + b_kX_{k_i} + u_i \quad (2.1)$$

É importante destacar, que o termo de erro u é também conhecido como erro (ou resíduo). Favero (2017) justifica seu uso indicando que a relação proposta muito provavelmente não será perfeita. Hair et al. (2009) definem que o erro é a diferença entre os valores reais e previstos da variável dependente. Em outras palavras, é o erro na previsão dos dados da amostra, indicando que dificilmente as previsões serão perfeitas.

Em relação ao intervalo de confiança, Hair et al. (2009), lembram que ao utilizar apenas uma amostra de observações para se prever a equação de regressão, é esperado que os coeficientes de regressão (ou angulares), sofram variações se outra amostra de observações for selecionada. Eles reforçam que, como não vamos considerar repetidas amostras, é necessário um teste empírico para identificar como o coeficiente de regressão se comporta: se for diferente de zero (tem algum valor real) ou se é esperado que ele seja igual a zero quando nova amostra for selecionada. Nesse sentido, é realizado o teste estatístico para definir se o intercepto e os coeficientes de angulares são significativamente diferentes de zero, ou seja, se estes têm um impacto que se pode esperar, com uma probabilidade especificada, que seja diferente de zero em qualquer número de amostras de observações.

Quando uma nova equação de regressão é selecionada por nova seleção de amostras, conforme Hair et al. (2009), assim como acontece com o intercepto e com os coeficientes angulares, é esperado que o valor previsto também sofra variação. Logo, ao invés de confiar apenas em uma estimativa pontual, é estimado o intervalo esperado de valores previstos.

Nesse contexto, Favero (2015) explica que os testes estatísticos apropriados podem determinar a significância estatística geral do modelo e de seus parâmetros estimados, já que o R^2 não é capaz de explicar tal significância. Um exemplo abordado por Favero (2017) é o teste F que avalia significância das variáveis explicativas de forma conjunta, onde é evidenciado hipóteses nula (H_0) e alternativa (H_1), conforme abaixo, onde β é o parâmetro estimado da inclinação do modelo proposto (coeficiente de regressão ou angular):

- $H_0: \beta_1 = 0$ - indicando que as variáveis selecionadas não afetam o comportamento da variável dependente;
- H_1 : existe pelo menos um $\beta_j \neq 0$ - indicando que pelo menos uma variável selecionada é estatisticamente significante para explicar a variabilidade da variável dependente.

Hair et al. (2009) conceituam o teste F como uma contribuição adicional indicando que, quando uma variável independente é adicionada ao modelo, depois que já existem outras variáveis, sua contribuição pode não ser tão relevante, mesmo apresentando correlação alta com a variável dependente já que, essa variável independente adicionada já pode apresentar uma alta correlação com as variáveis já inseridas no modelo. Nas palavras dos autores:

O valor parcial F é calculado para todas as variáveis simplesmente simulando que cada uma, por sua vez, seja a última a entrar na equação. Esse valor fornece a contribuição adicional de cada variável acima de todas as outras na equação. Um valor parcial F pequeno ou insignificante para uma variável que não está presente na equação indica sua contribuição pequena ou insignificante ao modelo como já especificado. (HAIR et al., 2009).

Nesse sentido, Favero (2017) indica que a estatística t é relevante por fornecer e averiguar a significância estatística de cada parâmetro que será considerado no modelo. Ao aplicar essa estatística, é necessário atestar o resultado tabelado conforme a distribuição t de *Student*. Para isso, um dado um nível de significância é definido para verificar se o teste aceita ou rejeita a hipótese nula. Os valores dos níveis de significância do teste t são chamados de valor-p (ou *p-value*). O valor-p, permite identificar, por exemplo, que com 95% de confiança (indicação mais utilizada), a hipótese nula será rejeitada. Dessa forma, espera-se que o valor-p calculado seja inferior ao valor-p tabelado. Em outros termos, a variável inserida é significativa para o modelo e por esse motivo deve permanecer no modelo. Contrário a essa constatação, quando não rejeitamos a hipótese nula (a hipótese nula é aceita), também a um determinado nível de significância ou nível de confiança, temos a indicação de que a correspondente variável independente não se correlaciona com a variável dependente e por esse motivo deve ser excluída do modelo final.

2.6.1 Normalidade, homocedasticidade e independência de resíduos

Outros três conceitos são importantes para a validação de um modelo de regressão, notadamente a normalidade dos resíduos, análise de homocedasticidade e independência de resíduos. A normalidade dos resíduos, que conforme Favero (2015), é requerida para que os testes de hipóteses dos modelos de regressão sejam validados. Caso os resíduos não cumpram os requisitos de normalidade, é indicado que o modelo tenha sido especificado de forma incorreta em relação à funcionalidade e que pode haver variáveis explicativas relevantes não consideradas. De forma análoga, esse erro de especificação (funcional ou falta de variáveis importantes), conforme o mesmo autor, pode gerar erro de heterocedasticidade, que, de acordo com Hair et al. (2009), “se refere à suposição de que as variáveis dependentes exibem níveis iguais de variância ao longo do domínio da(s) variável(is) preditor(a)s”. Logo, é esperado que os erros sejam homocedásticos, indicando que:

Para que essa relação seja completamente capturada, a dispersão (variância) dos valores da variável dependente deve ser relativamente semelhante em cada valor da variável preditora. Se essa dispersão for desigual ao longo de valores da variável independente, a relação é dita heteroscedástica. Apesar de as variáveis dependentes deverem ser métricas, esse conceito de igual extensão de variância entre as variáveis independentes pode ser aplicado quando as variáveis independentes são métricas ou não (HAIR et al., 2009).

Para isso, um gráfico de dispersão entre o resíduo e a resposta predita pode ser elaborado. Este mesmo gráfico também é importante para a análise de independência, que visa identificar tendências indicativas de que um modelo de regressão linear não é adequado para a análise realizada e que, portanto, modelos alternativos devem ser experimentados. Por fim, a questão de independência também deve ser analisada em função do número do experimento. Especificamente, deve-se analisar um gráfico de dispersão entre o número do experimento e o valor do resíduo em que é desejável não haver nenhuma tendência clara, como de crescimento, decréscimo ou agrupamento. Caso contrário, isso pode ser um indicativo de dependência entre as amostras.

2.6.2 Multicolinearidade

Considerando que a regressão múltipla pode melhorar nossa previsão de resultados quando se incrementa mais variáveis independentes ao modelo, surgem os conceitos de colinearidade e multicolinearidade. Quando duas variáveis independentes são linearmente dependentes, elas são chamadas de colineares. A dependência linear entre variáveis é medida por sua correlação. Assim, se a correlação entre duas variáveis independentes difere de zero, elas são colineares e o problema da dependência linear entre muitas variáveis independentes é chamado de problema de multicolinearidade (HAIR et al., 2009).

Hair et al. (2009) explicam que o excesso de multicolinearidade cria uma espécie de variância dividida entre as variáveis, reduzindo dessa forma a capacidade de prever a medida dependente e de apurar a relação de cada variável independente. Ainda, não teremos impacto apenas na previsão do modelo de regressão, mas também sobre os coeficientes de regressão, bem como nos testes de significância.

A avaliação da multicolinearidade, conforme definida por Hair et al. (2009), pode ser feita por meio de uma métrica que determinará o grau da explicação da variável independente de forma individual em relação às demais variáveis independentes. Ainda, em relação aos resíduos, o problema da multicolinearidade, de acordo com Favero (2015), “ocorre quando há correlações muito elevadas entre variáveis explicativas e, em casos extremos, tais correlações podem ser perfeitas, indicando uma relação linear entre as variáveis”.

As medidas mais comuns em relação à avaliação da multicolinearidade são a Tolerância e o Fator de Inflação de Variância, também chamado de VIF, ao qual será dada ênfase. A tolerância refere-se, conforme os mesmos autores, a uma medida direta e irá mensurar a quantidade de variabilidade da variável independente que foi selecionada e não explicada pelas outras variáveis independentes. Nesse processo, a variável independente selecionada se transformará em uma variável dependente prevista pelas demais variáveis independentes. Seu cálculo é definido como $1 - R^2$. No exemplo mostrado pelos mesmos autores, Hair et al. (2009): “(...) se as outras variáveis independentes explicam 25% da variável independente X_1 ($R^2 = 0,25$), então o valor de tolerância de X_1 é 0,75 ($1,00 - 0,25 = 0,75$)”.

Já o VIF, que é uma segunda medida, é definido como o inverso da tolerância, mostrando que, com níveis maiores de multicolinearidade, a tolerância indicará valores baixos e o VIF valores maiores. No exemplo anterior mostrado por [Hair et al. \(2009\)](#), “com uma tolerância de 0,75, o VIF seria 1,33 ($1,00/0,75 = 1,33$)”. Assim, por [Hair et al. \(2009\)](#), indicam que

O VIF tem seu nome devido ao fato de que a raiz quadrada dele (\sqrt{VIF}) é o grau em que o erro padrão aumentou devido à multicolinearidade. (...) Assim, quando o erro padrão aumenta, os intervalos de confiança em torno dos coeficientes estimados ficam maiores, tornando mais difícil a demonstração de que o coeficiente é significativamente diferente de zero. ([HAIR et al., 2009](#)).

Infelizmente, não existe um consenso sobre quais valores de VIFs são toleráveis. [Favero \(2015\)](#) pontua que um VIF maior que 10 indica problemas de multicolinearidade e [Ferreira et al. \(2021\)](#), complementam que:

(...) quanto mais próximo o valor de R^2 ao valor 1, maior o valor de VIF e, conseqüentemente, maior a multicolinearidade com a variável independente específica. Ou seja, se $VIF \leq 1$, então as variáveis não são correlacionadas; $1 < VIF \leq 5$ indica que as variáveis são moderadamente correlacionadas; e $VIF > 5$ indica que as variáveis são altamente correlacionadas. ([FERREIRA et al., 2021](#))

Para o estudo, com base em conhecimento teórico e prático, foi adotado como linha de corte o VIF maior que 5. Assim, assume-se que variáveis com VIF acima de 5 apresentam problemas de multicolinearidade e portanto, não são recomendadas para a regressão gerada.

2.6.3 Critérios para seleção de variáveis

Em relação aos critérios de seleção de variáveis a serem incluídas no modelo, pode-se citar dois métodos de busca sequencial, que irão maximizar a previsão do modelo enquanto utiliza número menor de variáveis. Conforme [Hair et al. \(2009\)](#), temos: (1) estimação *stepwise* e (2) adição *forward* (ou inclusão em avanço) e eliminação *backward* (ou eliminação retroativa).

[Hair et al. \(2009\)](#) explicam que a estimação *stepwise* (procedimento por etapas) se inicia com a seleção do melhor preditor da variável dependente permitindo assim examinar a relevância de cada variável independente adicionada ao modelo. As variáveis só serão adicionadas se os coeficientes de correlação foram significativos para o modelo, ou seja, quando selecionada deverá contribuir de forma incremental sobre as variáveis já presentes no modelo. Dessa forma, a cada adição, é necessário recalcular o coeficiente de determinação.

Já a inclusão em avanço e a eliminação retroativa, é explicada por [Hair et al. \(2009\)](#) como sendo uma técnica de tentativa e erro com o objetivo de descobrir as melhores estimativas da regressão. A adição *forward* diz respeito à inclusão das variáveis independentes em um modelo inicialmente sem variáveis – essa variável é acrescentada com base em sua contribuição para a previsão. A eliminação *backward* refere-se ao método de eliminação de variáveis que não oferece contribuição significativa para a previsão a partir de um modelo inicial que possui todas as variáveis independentes no modelo.

Uma vez que se tem as variáveis mais representativas segundo algum critério (valor-p, correlação ou VIF), dependendo do método escolhido, precisaremos escolher aquele modelo que mais se aproxima da realidade do processo. Portanto, é necessário, conforme [Hair et al. \(2009\)](#) avaliar a precisão dos erros na previsão da variável dependente do modelo selecionado. Apenas somar os erros, conforme evidenciado por [Favero \(2015\)](#), não permitirá obtenção de uma medida satisfatória, uma vez que, pode-se obter condições semelhantes, ou seja, várias retas diferentes, que conforme [Hair et al. \(2009\)](#), irão gerar o mesmo resultado, uma vez que, o somatório desses erros em relação à média, sempre será igual a zero.

De acordo com [Hair et al. \(2009\)](#), para extrapolar essa condição, é preciso elevar ao quadrado cada erro e posteriormente somá-los aos resultados – nesse momento teremos o que chamamos de soma de quadrados dos erros que fornecerá, conforme mesmos autores, uma medida de precisão preditiva em que a meta é obter a menor soma possível dos erros quadrados, ou seja, o melhor modelo será aquele que obedecer essa condição. Já o MSE (*Mean Squared Error*), conforme [Penteado \(2021\)](#) é frequentemente utilizado para verificar a acuracidade dos modelos e acaba gerando peso maior aos erros maiores, uma vez que cada erro é elevado ao quadrado de forma individual e, posteriormente é calculada a média desses erros quadráticos. Outra métrica é o RMSE (*Root Mean Squared Error*), que, de acordo com [Penteado \(2021\)](#) é calculada a partir da raiz quadrada do MSE, em que o erro irá retornar à unidade de medida do modelo, lembrando que, no MSE, a unidade de medida é elevada ao quadrado.

No presente trabalho, tais técnicas foram exploradas no sentido de identificar um conjunto de variáveis que forneçam maximizem o desempenho do modelo e, paralelamente, não sofra com o problema de multicolinearidade, heterocedasticidade e não normalidade dos erros.

3 Metodologia

O presente estudo tem como premissa prever, de maneira quantitativa e por meio de modelagem matemática e dados de processos, a taxa de desgaste refratário da linha de escória das panelas de aço utilizadas no processo da Aciaria 2, da planta de Ipatinga da Usiminas. A pesquisa tem natureza aplicada e em relação aos procedimentos é classificada como experimental e modelagem e simulação. As próximas seções serão utilizadas para descrever o desenvolvimento da metodologia, divididas da seguinte forma: ambiente computacional e ferramentas, coleta de dados, amostragem e pré-processamento dos dados.

3.1 Ambiente computacional e ferramentas

Os dados utilizados no desenvolvimento deste trabalho se dividem em: dados armazenados em planilhas do software Microsoft Excel (Excel) e dados armazenados em banco de dados relacionais, extraídos via software SQL Developer, versão 21.2.1.204. O modelo de regressão proposto foi desenvolvido na linguagem Python na versão 3.9.10, com o auxílio dos pacotes Pandas, Scipy e StatsModel, por meio do software DataSpell, versão 2022.3. A implementação foi feita em um computador com sistema operacional Windows, versão 10 Enterprise, que possui processador Intel i7-1265U Core(TM) de 12ª geração a 1,80 GHz de frequência e 16 GB de memória RAM.

3.2 Coleta de dados

Os dados utilizados na pesquisa são divididos por sua natureza em duas fontes:

- Dados operacionais/ processo, provenientes do banco de dados da Usiminas;
- Dados das medições do perfil de demolição refratária *post mortem*, disponibilizado pelo fornecedor do material via planilhas do Excel.

Os dados relativos às variáveis operacionais e de processo foram coletados por meio do uso da ferramenta SQL Developer, em fontes secundárias, nos sistemas de Nível 3 (Tecnologia da Informação - TI) e contém registros do período de janeiro de 2021 a outubro de 2022 e que também fazem referência às campanhas de panelas de aço revestidas pelo principal fornecedor do material refratário.

As variáveis extraídas estão originalmente distribuídas em 31 tabelas distintas (que geraram 23.354 registros das corridas produzidas) com acessos direto, e para a consolidação em tabelas de saída foi necessária a construção de *procedures*. Toda a construção realizada via SQL Developer foi resumida em duas tabelas, posteriormente exportadas via Excel, conforme indica a Figura 5.

Figura 5 – Tabelas utilizadas como bases de dados

DB_CONTROLE_ACIARIA.TB_TCC_DADOS_CORRIDA	
P * NUM_CORR_ACI	NUMBER
DT_PRODUCAO	DATE
NUM_PAN_ACO	NUMBER
NUM_CAMP_PA	NUMBER
VIDA_PAN_ACO	NUMBER
FAMILIA_ACO	VARCHAR2 (100 BYTE)
ROTA_ACO	VARCHAR2 (1000 BYTE)
COD_TRATAMENTO	VARCHAR2 (100 BYTE)
DT_ABERT_PAN	DATE
DT_ULT_REFINO	DATE
DT_FIM_VAZAM_ACO_LD	DATE
DT_INIC_VAZ	DATE
DT_FIM_LINGNTO	DATE
DT_FIM_LINGNTO_ATUAL	DATE
TEMPO_TOT_PAN_QUENTE	VARCHAR2 (100 BYTE)
TEMPO_PAN_CHEIA	VARCHAR2 (100 BYTE)
TEMPO_PAN_VAZIA	VARCHAR2 (100 BYTE)
TEMPO_CICLO_PAN	VARCHAR2 (100 BYTE)
CARGA_ACO_LING	VARCHAR2 (100 BYTE)
VL_S_VIS	NUMBER
VL_PESO_L8	NUMBER
VL_QTD_BANHO_GUSA	NUMBER
PK_TB_TCC_DADOS_CORRIDA (NUM_CORR_ACI)	
PK_TB_TCC_DADOS_CORRIDA (NUM_CORR_ACI)	

DB_CONTROLE_ACIARIA.TB_TCC_DADOS_REFINOS	
P * NUM_CORR_ACI	NUMBER
P * TRATAMENTO	VARCHAR2 (100 BYTE)
P * SEQ_TRATAMENTO	NUMBER
TEMP_FS	NUMBER
TEMP_VAZ	NUMBER
TEMP_POS_VAZ	NUMBER
TEMP_INICIAL_TRAT	NUMBER
TEMP_FINAL_TRAT	NUMBER
DELTA_AQUEC	NUMBER
TEMP_PANEIA	NUMBER
TEMPO_AQUECIMENTO	NUMBER
TEMPO_TRATAMENTO	NUMBER
TEMPO_VAZ_LD	NUMBER
TEMPO_REFINO	NUMBER
OXIDACAO_FINAL	NUMBER
CARGA_VAZADA_LD	NUMBER
IN_PLUG	VARCHAR2 (100 BYTE)
PLUG_VAZAO	NUMBER
PLUG_TEMPO_INJ	NUMBER
IN_LANCA	VARCHAR2 (100 BYTE)
LANCA_VAZAO	NUMBER
LANCA_TEMPO_INJ	NUMBER
VL_PESO_INJ_PO_OBTD	NUMBER
VL_TEMPO_INJ_PO_OBTD	NUMBER
VL_PESO_INJ_CAO_F2_OBTD	NUMBER
VL_TEMPO_INJ_CAO_F2_OBTD	NUMBER
VL_TEMPO_HOMOGEINIZACAO	NUMBER
ADIC_MNAC	NUMBER
ADIC_MNMC	NUMBER
ADIC_SIMN	NUMBER
ADIC_CASI	NUMBER
ADIC_FLUO	NUMBER
ADIC_CALE	NUMBER
ADIC_ESIN	NUMBER
ADIC_CARB	NUMBER
ADIC_FCAL	NUMBER
ADIC_NEFE	NUMBER
ADIC_FECA	NUMBER
ADIC_BRIQ	NUMBER
ADIC_MNBP	NUMBER
ADIC_MNEL	NUMBER
ADIC_CASV	NUMBER
ADIC_FCAH	NUMBER
ADIC_BRID	NUMBER
PK_TB_TCC_DADOS_REFINOS (NUM_CORR_ACI, TRATAMENTO, SEQ_TRATAMENTO)	
PK_TB_TCC_DADOS_REFINOS (NUM_CORR_ACI, TRATAMENTO, SEQ_TRATAMENTO)	

Fonte: Elaboração própria com o uso do SQL Developer

As variáveis extraídas estão relacionadas ao processo de fabricação do aço. Ao iniciar a produção de uma corrida, é gerado um número sequencial de 6 dígitos, sendo o primeiro sempre relativo ao convertedor onde foi produzida, no caso da Aciaria 2, os convertedores 4 ou 5. Cada corrida possui uma programação relativa ao pedido, que determina a composição química visada, os tipos de tratamento, os equipamentos (rotas) pelos quais deve passar, as adições necessárias, entre outras informações. Após o processamento da corrida, além do planejamento, são registrados os parâmetros reais obtidos, incluindo-se tempos, temperaturas, carga transportada, adições de materiais, utilização de lança e plug poroso, entre outras, que, conforme mencionado anteriormente, podem contribuir para o desgaste refratário das painéis de aço.

Três desses parâmetros reais referem-se à identificação do objeto de estudo: o número da panela, que especifica a carcaça metálica; o número da campanha, que define o ciclo de refratamento em que a panela se encontra; e o número da vida da panela, que indica o sequencial que a corrida representa para a panela de aço e campanha. A Tabela 2 apresenta um exemplo prático do relacionamento entre corrida e panela de aço, onde a corrida de número 546987, processada no convertedor 5 (primeiro dígito), foi a 13ª (vida) vazada na panela (carcaça) número 1, que está em sua 99ª campanha (refratário).

Tabela 2 – Exemplo de relacionamento entre corrida e panela de aço

NUM_CORR_ACI	NUM_PAN_ACO	NUM_CAMP_PA	VIDA_PAN_ACO
546987	1	99	13

Fonte: Elaboração própria

A Tabela 3 apresenta de forma resumida as variáveis utilizadas na base de dados. A relação completa e detalhada está disponível no Apêndice A.

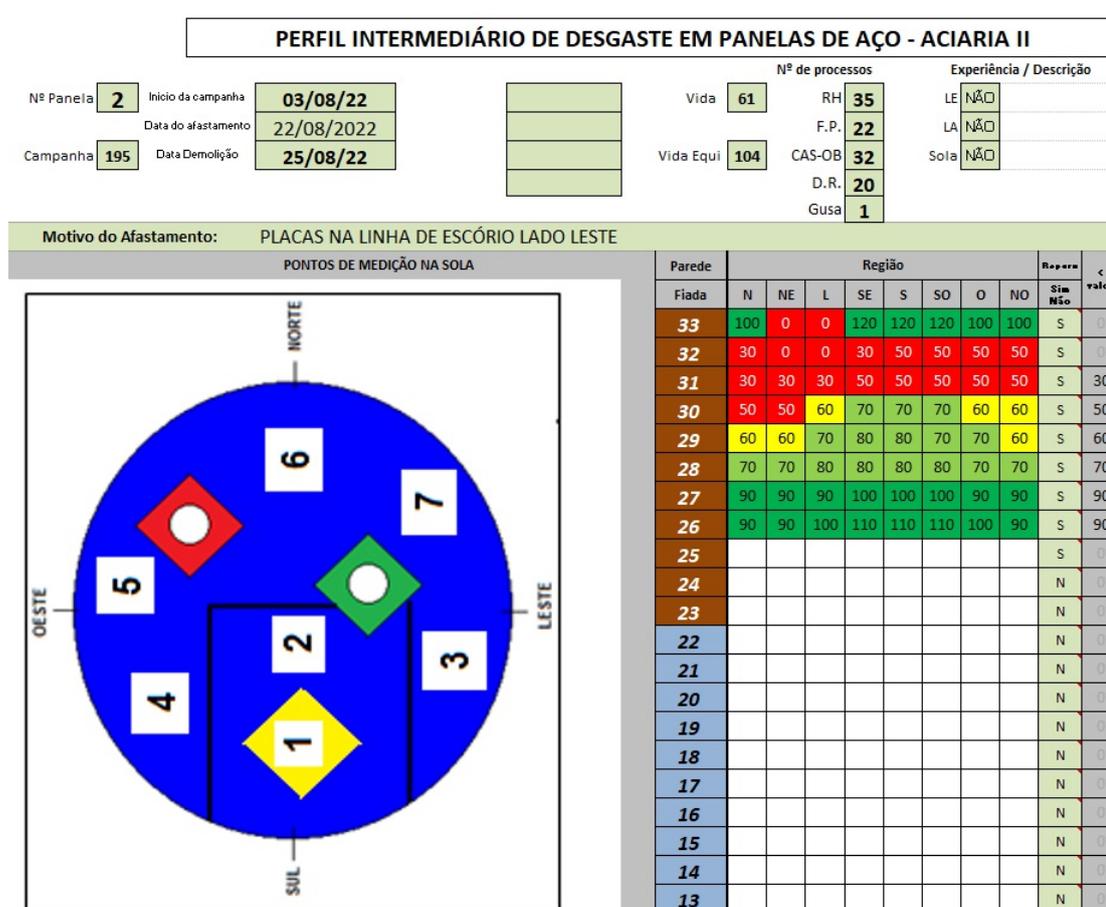
Tabela 3 – Variáveis utilizadas

# Descrição de variáveis	Tipo
1 Temperaturas de processo (fim de sopro, após vazamento, após tratamentos)	Quantitativo
2 Mix de aços fabricados	Qualitativo
3 Tipo de rota / tratamento metalúrgico da corrida	Qualitativo
4 Tempos operacionais (tratamento, vazamento, panela cheia)	Quantitativo
5 Carga das panelas (peso)	Quantitativo
6 Uso de plug	Qualitativo
7 Adições em panela (material e peso)	Quantitativo
8 Perfil de demolição	Quantitativo
9 Taxa de desgaste	Quantitativo

Já os dados relativos às medições das espessuras residuais dos tijolos *post mortem* do refratário, são disponibilizadas em um conjunto de arquivos Excel por panela e campanha. As planilhas são preenchidas manualmente pelo fornecedor, e em cada uma estão apresentadas, em abas distintas, a medição do refratário residual no reparo intermediário e reparo final, com dados relativos à data de demolição e ao número de corridas até o reparo (campo “Vida”).

No reparo intermediário, foco da pesquisa, são apresentadas as medições dos tijolos da linha de escória por fiada (fileira de tijolos) e posição da panela (norte, sul, leste e oeste). Como o objetivo do trabalho é modelar o desgaste máximo da panela, o ponto de menor espessura residual é o foco do desenvolvimento. Como exemplo, a Figura 6, apresenta o perfil da panela 2, campanha 195.

Figura 6 – Perfil intermediário de desgaste em painelas



Fonte: Elaborada pelo fornecedor de refratário

3.3 Amostragem

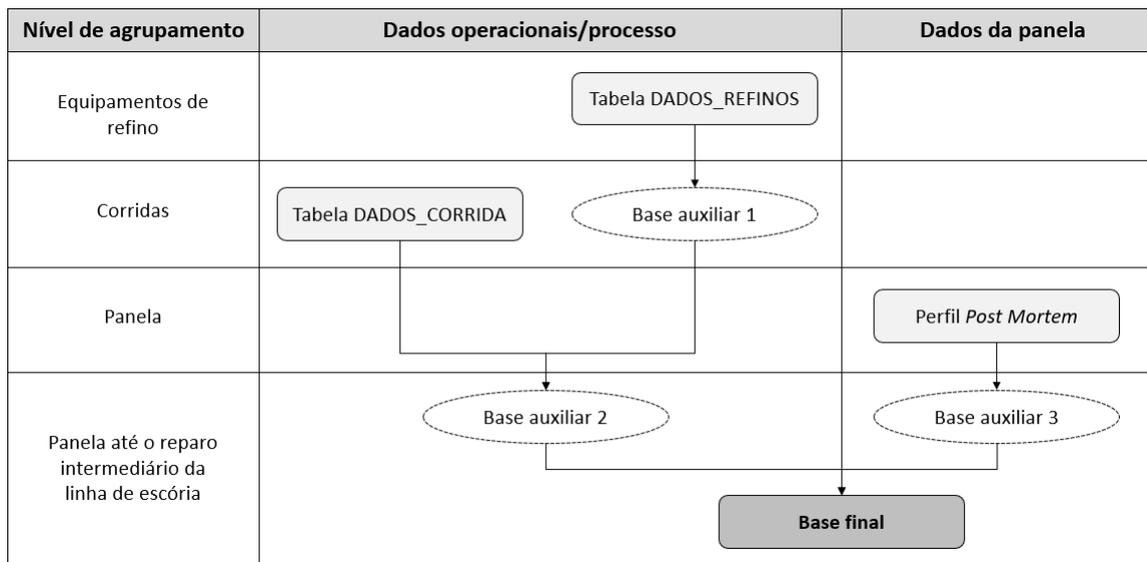
Quanto à amostragem, a escolha da aciaria se deve ao fato de a Aciaria 2 ser responsável por aproximadamente 70% da produção de aço da planta e possuir processos, rotas de produção e adições mais complexas, que permitem a fabricação de aços mais nobres, mix de produtos diversificado, e, portanto, maior complexidade e impacto da solução proposta. Já o período de análise foi determinado com base na vigência do atual projeto de engenharia dos materiais e procedimentos operacionais, que representam as condições atuais do processo. A escolha do fornecedor tem como referência o *market share* e performance superior do material, oferecendo maior potencial de ganhos para a aciaria.

As variáveis operacionais foram selecionadas conforme revisão bibliográfica e alinhamento com a equipe técnica da área estudada. Já o desgaste real do refratário na linha de escória, região crítica que determina o afastamento das painelas, e que validaram o modelo matemático construído, foi definido com base no menor residual medido por meio dos dados primários apurados e registrados em planilha do Excel pelo fornecedor escolhido. No exemplo da Figura 6, o menor residual observado foi 0, nas regiões NE e L, fiadas 33 e 32 do refratário.

3.4 Pré-processamento dos dados

Os dados coletados foram organizados e analisados, questionamentos foram levantados antes da sua exploração, fixando objetivos e levantando hipóteses. Foram então tratados, retirados os *outliers*, codificados, classificados e correlacionados entre si. A Figura 7 demonstra os passos realizados para o agrupamento de três bases de dados originais e obtenção da base de dados final utilizada para a regressão.

Figura 7 – Agrupamento dos dados



Fonte: Elaboração própria

As bases originais DADOS_CORRIDA e DADOS_REFINOS referem-se ao processo, mas possuíam como unidade básica corrida e tratamento por refino, respectivamente. Portanto, inicialmente, foi necessário agrupar os dados de tratamento por refino em corridas, o que foi denominado Base auxiliar 1. Assim foi possível agrupar as bases DADOS_CORRIDA e Base auxiliar 1 em uma única tabela, chamada Base auxiliar 2.

Já a base original Perfil *Post Mortem* possui informações referentes à Panela/Campanha, dividida entre reparo intermediário e reparo final. Como o objetivo do trabalho refere-se ao reparo intermediário, foi selecionada essa subdivisão, no que se chamou Base auxiliar 3. A Base auxiliar 3 contempla informações referentes à menor espessura do tijolo registrada e número de corridas com o qual a panela foi afastada da operação. Com a segunda informação foi possível selecionar da Base auxiliar 2 as corridas processadas até o reparo intermediário, e portanto influenciaram no desgaste, que unida com a Base auxiliar 3, formaram a Base final de treinamento da regressão linear múltipla.

A etapa seguinte consistiu na transformação de dados qualitativos, que classificam as corridas por código de tratamento, família de aço e rota de tratamento, em variáveis quantitativas. Esta transformação gerou 27 variáveis do tipo *One Hot Encoding* com base nos valores unitários de cada coluna, sendo 9 referentes ao código de tratamento, 8 referentes à família de aço e 10 relacionadas aos equipamentos disponíveis na rota de fabricação, conforme Tabela 4.

Tabela 4 – Variáveis transformadas através de *One Hot Encoding*

Variável qualitativa	Variáveis quantitativas
COD_TRATAMENTO	A3, C1, C8, C7, F8, F1, A1, A2, C0
FAMILIA_ACO	ALTOCARBONO, BAIXOENXOFRE, COMUM, DESGASEIFICADA, DRGERAL, PERITETICO, UBCGERAL, UBCUSICORE
ROTA_ACO	LD4, LD5, FP2, OB1, OB2, RH2, RHC, RHD, ML1, ML2

Fonte: Elaboração própria

Novas variáveis foram criadas com o objetivo de demonstrar em colunas distintas as informações de tempo de aquecimento, tempo de refino e variação (delta) de aquecimento por equipamento: RH, CAS-OB e Forno Panela. Posteriormente, os dados das corridas foram sintetizados por panela e campanha para agrupar as informações e formar a base de dados para aplicação da regressão. O agrupamento das variáveis foi realizado por soma, média e percentual, conforme Apêndice B.

Foram agrupadas por soma as variáveis *One Hot Encoding*, por se tratarem de contagens, e as adições por não estarem presentes necessariamente em todas as corridas e o volume total adicionado ter maior relevância que uma média contemplando zeros, que não representaria a ação química dos materiais no refratário.

Foram agrupadas por média todas as variáveis que são comuns a todas as corridas como temperaturas e deltas temperaturas (temperatura final - temperatura inicial) em cada equipamento, pesos de adições de materiais, tempos de homogeneização, pesos de corrida e pesos de injeção de pó por equipamento.

As variáveis referentes a tempos - tratamento, aquecimento, panela cheia, panela vazia - foram agrupadas tanto por média quanto por soma, por serem relevantes, terem sentido lógico e para possibilitar às técnicas estatísticas selecionar da forma que fosse mais representativa para modelo.

Inicialmente foram criadas variáveis que representavam o desvio padrão das consideradas mais representativas para o desgaste, entretanto, chegou-se à conclusão de que o desvio padrão só faria sentido quando associado a uma variável principal, de média ou soma. Tal fato faria com que o número de variáveis da regressão aumentasse significativamente, o que não é considerado interessante para o trabalho, e portanto, estas foram descartadas.

Por fim, outras variáveis foram derivadas, tais como: as variáveis classificatórias (*One Hot Encoding*) foram transformadas em participações percentuais por família de aço, equipamentos e códigos de tratamento (com final `_PART`); variáveis que agrupam as 9 famílias de aço em 4 principais (DR, UBC, FP, COMUM); variáveis que representam a adição de elementos e compostos específicos (MANGANÊS, CÁLCIO e FLUORITA) obtidos por meio da ponderação de pesos de adições pela concentração dos mesmos.

Para a variável resposta (taxa de desgaste médio do refratário) foi necessário realizar um cálculo com as variáveis de desgaste e número de corridas por panela e campanha. O cálculo é dado pela seguinte equação (3.1):

$$Taxa\ Desgaste = \frac{Espessura\ Inicial - Espessura\ Mínima}{Número\ de\ Corridas} \quad (3.1)$$

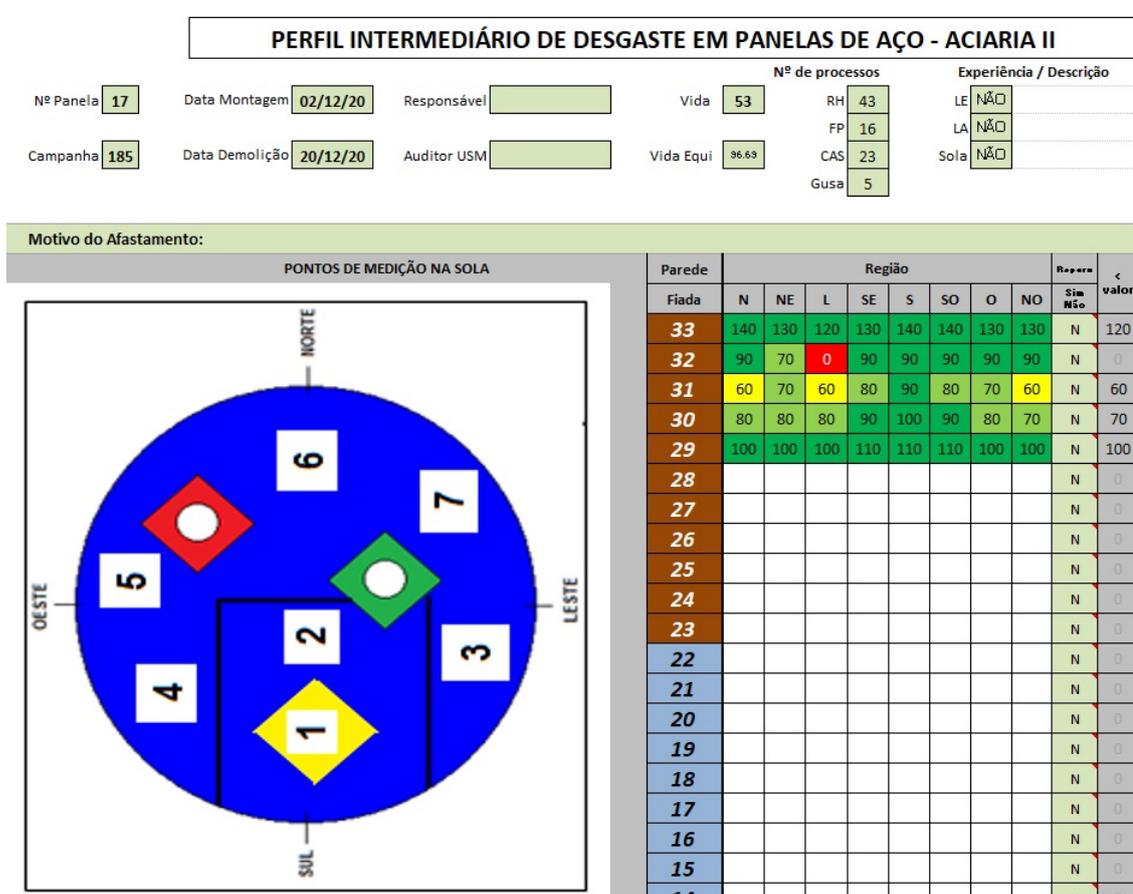
em que:

- Espessura Inicial: A espessura inicial se refere a espessura original do tijolo no início da campanha;
- Espessura Mínima: se refere ao menor valor encontrado na região da linha de escória da panela na medição post mortem do perfil intermediário;
- Número de Corridas: representa quantas corridas foram executadas até a troca intermediária do refratário.

Com relação às observações (panelas e campanhas), foi feita uma análise relativa ao perfil de desgaste, e quatro delas foram excluídas da base devido à suspeita de interferência de variáveis externas e não registradas tal como: lança furada com desvio do fluxo de argônio, plug com excesso de vazão, ou arrancamento de tijolo. Como exemplo, a panela 17, campanha 185 foi considerada *outlier* (Figura 8). Ela apresentou um residual igual a zero na fiada 32 Leste, entretanto, seu entorno não indica a mesma tendência de desgaste acentuado, o que leva à conclusão de que alguma interferência externa atípica à operação e não capturada pelos dados levou a este resultado.

Após realizados os pré-processamentos, tratamentos e retirada de *outliers*, a base contendo 97 variáveis e 131 observações, se mostrou apta para a aplicação das técnicas de regressão linear e seleção de variáveis, que serão apresentadas no próximo capítulo.

Figura 8 – Perfil intermediário da panela 17, campanha 185



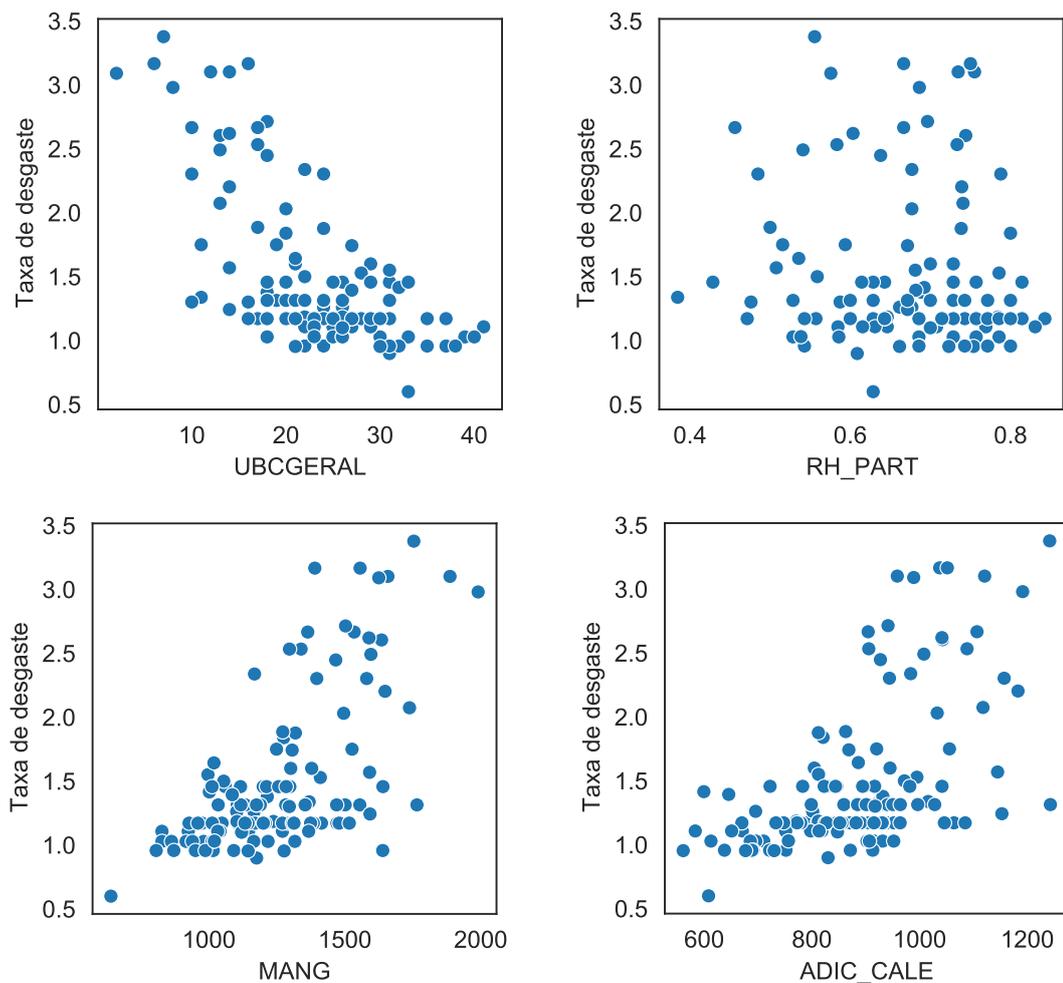
Fonte: Elaborada pelo Fornecedor de Refratário

4 Resultados

4.1 Análise de correlação de variáveis

A base de dados gerada após o pré-processamento continha 97 variáveis que foram inspecionadas por meio de medidas estatísticas e gráficos de dispersão gerados a partir dos pacotes Pandas¹ e Seaborn². A apresentação da exploração de todas as variáveis tornaria este trabalho muito extenso e, a fim de exemplificar a análise realizada, foram selecionadas as 4 variáveis mais correlacionadas com a variável resposta com base no coeficiente de correlação de Pearson, demonstradas na Figura 9.

Figura 9 – Variáveis correlacionadas de maior representatividade para a taxa de desgaste do refratário



Fonte: Elaboração própria com uso do Python

¹ <<https://pandas.pydata.org/>>

² <<https://seaborn.pydata.org/>>

Conforme é possível observar na Figura 9, a variável UBCGERAL, que representa o número de corridas pertencentes à família UBC Geral, possui uma correlação negativa (-0,65) com a variável resposta. Tal influência é esperada, visto que os aços da família UBC Geral são processados apenas no RH, com baixa adição de fluidificantes, e são os materiais que propiciam a formação de uma escória menos agressiva ao refratário. Sendo assim, quanto maior a produção de aços deste tipo, menor o desgaste refratário.

A variável RH_PART representa o percentual de corridas processadas nos equipamentos RH. Observa-se que, assim como UBCGERAL, esta variável possui correlação negativa (-0,788) com a variável resposta taxa de desgaste refratário. Verifica-se através de observações industriais que nas panelas oriundas do processo de RH (exceto as corridas da família Duplo Refino), há formação de uma camada sólida constituída por aço e escória, conhecida como cascão. Esta camada cria uma espécie de revestimento (*coating*) acima dos tijolos, formando uma barreira física entre a escória do processo e o tijolo refratário, que reduz o contato e conseqüentemente o desgaste dos mesmos. Além disso, as corridas processadas neste equipamento, exceto as da Família Duplo Refino, são corridas com baixa adição de fluidificantes, e propiciam a formação de escórias menos agressivas.

A variável MANG refere-se ao peso médio por corrida do elemento Manganês adicionado via ligas metálicas, e o gráfico de dispersão indica, conforme previsto, que quanto maior o volume de manganês adicionado, maior o desgaste refratário (coeficiente de correlação 0,634). Isto ocorre porque o manganês influencia na fluidização da escória, e uma escória mais fluida tem maior potencial de ataque ao refratário.

Da mesma forma, a variável ADIC_CALE, que consiste na média de adição de cal em panela por corrida, indica a necessidade de dessulfuração do aço, por meio de reação química e agitação do aço, além de influenciar na formação de escória, e portanto favorecer o desgaste (coeficiente de correlação +0,570).

Com o objetivo de selecionar variáveis correlacionadas de maior representatividade para a taxa de desgaste do refratário (variável resposta), foi utilizado o coeficiente de correlação de Pearson considerando um p-valor de 0,05 por meio da função `pearsonr` do módulo `scipy`. Após a aplicação da função, 51 variáveis foram selecionadas como entrada para o desenvolvimento do modelo de regressão linear múltipla.

4.2 Modelo de regressão linear múltipla

O uso de todo o conjunto resultante de variáveis com correlação representativa pode resultar em vários problemas. Um dos mais notáveis é o da multicolinearidade que, conforme citado no Capítulo 2, afeta a robustez e validação do modelo. Este problema pode ser tratado por meio de uma técnica de seleção de variáveis atrelada a algum critério em direção a modelos mais parcimoniosos.

Para seleção das variáveis que fariam parte do modelo de regressão, foram avaliadas as seguinte estratégias:

1. Adição *forward*: Inclusão de variável no modelo uma a uma, partindo da variável com maior correlação com o desgaste refratário (coeficiente de Pearson). A cada inclusão de variável foram analisadas as medidas: R^2 , R^2 ajustado, p-valor do teste t, além do VIF para identificar multicolinearidade. Se a variável incluída no modelo não contribuía positivamente para o R^2 , R^2 ajustado, não era considerada significativa de acordo com o teste t ou gerava um problema de multicolinearidade, ela era então descartada;
2. Remoção *backward*: Todas as 51 variáveis foram selecionadas como entrada, e uma a uma foram retiradas aquelas que eram menos significativas para o modelo de acordo com o p-valor do teste t (maior valor). A retirada de variáveis cessou quando todas as variáveis do modelo possuíam um p-valor inferior a 0,05 e o VIF destas era igual ou inferior a 5, conforme descrito no capítulo 2. Por fim, as variáveis retiradas foram inseridas novamente uma a uma. A cada inclusão de variável foram analisadas as medidas de ajuste do modelo: R^2 , R^2 ajustado, p-valor, além do VIF para verificar se elas poderiam permanecer no modelo sem gerar problema de multicolinearidade. Se a variável incluída no modelo não apresentava contribuição positiva para as medidas mencionadas, ela era então descartada.

Dentre as estratégias utilizadas, a que melhor obteve resultado foi o método adição *forward* para seleção das variáveis tendo como base os resultados do R^2 ajustado e menor erro quadrático médio (*mean squared error* - MSE). Após aplicação desta estratégia, foram selecionadas as variáveis a seguir, que, em conjunto, obtiveram o melhor resultado para o modelo: RH, VL_PESO_L8, F1_PART, F8_PART; A3_PART, A3_PART, BAIXOS_PART, BAIXOS_PART, TEMPO_AQUE_FP, e TEMPO_REF_OB_x.

Para validar a coerência das variáveis selecionadas e seus coeficientes, foram realizadas discussões com especialistas do domínio a fim de identificar explicações dentro do processo de produção e mecanismos de desgaste refratário que suportassem a descrição estatística. A variável RH consiste no número de corridas processadas nos RHs, e possui coeficiente negativo. Conforme mencionado na Subseção 4.1, as corridas processadas no RH propiciam a formação de cascão e de escórias menos agressivas e portanto, há uma tendência de que quanto maior o número de corridas processadas neste equipamento, menor o desgaste refratário.

A variável VL_PESO_L8 representa a média do peso de aço transportado e tratado na panela. Espera-se que maiores pesos impliquem em níveis mais elevados de aço dentro da panela. Em situações extremas, esses níveis podem alcançar a região de fechamento entre a última fiada de refratário da linha de escória e a chapa de boca da panela, o que pode provocar infiltrações de escória entre a carcaça e os tijolos. Tais infiltrações provocam dilatação térmica e/ou corrosão e são responsáveis pela degradação das últimas fiadas da linha de escória. As variáveis F1_PART, F8_PART, A3_PART apresentaram coeficientes de correlação positivos, indicando que quanto maiores, maior o desgaste refratário. Elas representam o percentual de corridas processadas com esses códigos de fabricação:

- F1: Dessulfuração no CAS-OB e posterior desgaseificação no RH;
- F8: Dessulfuração no Forno Panela e posterior desgaseificação no RH;
- A3: Descarburização - carbono da sigla ≤ 2 pontos.

Os aços com código de fabricação F1 e F8 são os aços mais complexos produzidos pela Aciaria 2, dado que necessitam duplo processamento (CAS-OB e RH ou Forno Panela e RH) e de forma geral possuem enxofre visado baixo (maiores níveis de dessulfuração em panela), maior volume de adição fluidificantes, ligas de manganês e fontes de cálcio, além de implicar em maior tempo de processo e agressividade para o refratário. Já para a variável A3_PART, inicialmente esperava-se uma correlação negativa com o desgaste refratário devido, principalmente, à característica de formação de cascão e consequente proteção das panelas. Porém, a correlação encontrada foi positiva. Duas explicações possíveis seriam: (1) o aumento da participação de corridas A3 ocorre em substituição à participação de corridas de aços comuns, que possuem desgaste refratário menor; (2) as corridas A3 têm por característica cargas médias de aço maiores, o que pode levar ao mecanismo de infiltração na borda, conforme citado na explicação da variável VL_PESO_L8.

A variável BAIXOS_PART representa o percentual de corridas da família Baixo Enxofre e possui correlação positiva. Estes aços possuem um enxofre visado no aço menor que 0,007%, o que implica em maior tempo de processo, maior adição de fluidificantes e de fontes de cálcio, maior tempo de exposição ao arco elétrico quando produzidas no Forno Panela, maior nível de dessulfuração e maior agitação do banho devido à lança de argônio e plug poroso. Isto explica porque o aumento da variável provoca aumento do desgaste.

Já a variável ADIC_MNEL representa o somatório do peso de adição de Manganês Eletrolítico, que é uma das ligas de manganês utilizadas na Aciaria 2 (37% do peso total de ligas de manganês), e que possui maior concentração do elemento Manganês (98%, enquanto as demais variam entre 68% e 80%). Conforme já discutido no Subtópico 4.1, a adição de manganês e a formação de seus óxidos geram compostos na escória de baixo ponto de fusão, e consequentemente, favorecem o desgaste refratário.

O TEMPO_AQUE_FP representa o tempo médio de aquecimento das corridas no Forno Panela. Quanto maior o tempo de aquecimento neste equipamento, maior a adição de fluidificantes como fluorita e escória sintética, maior a exposição do refratário ao arco elétrico e às escórias com elevada temperatura e fluidez. Por fim, a variável TEMPO_REF_OB_x representa a média dos tempos de refino no CAS-OB e possui coeficiente positivo, indicando que quanto maior o tempo de refino nestes equipamentos, maior o tempo de contato da escória com o refratário e consequentemente, maior probabilidade de reação entre eles.

A Equação (4.1) apresenta o modelo de regressão linear múltipla obtido:

$$\begin{aligned}
 taxa_desg = & -8,4544 + 4,6938 \times F1_PART & (4.1) \\
 & + 4,6690 \times F8_PART \\
 & + 3,2138 \times A3_PART \\
 & + 1,9327 \times BAIXOS_PART \\
 & - 0,0649 \times RH \\
 & + 0,0355 \times TEMPO_AQUE_FP_x \\
 & + 0,0134 \times TEMPO_REF_OB_x \\
 & + 0,0008 \times ADIC_MNEL \\
 & + 0,00005 \times VL_PESO_L8 \\
 & + u_i
 \end{aligned}$$

Conforme sumário de resultado do modelo com métricas da regressão, os resultados alcançados são apresentados na Tabela 5.

Foram realizadas 20 execuções do modelo utilizando observações de treinamento e de testes aleatórias afim de calcular a média e desvio padrão do MSE. Afim de trazer este erro médio para a unidade de medida da variável resposta, foi calculada a raiz quadrada do MSE resultando no valor do RMSE.

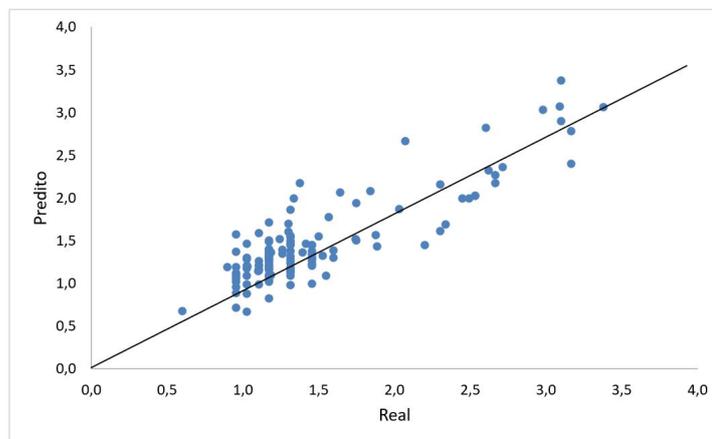
Tabela 5 – Resultados obtidos pelo modelo de regressão linear múltipla proposto

Métrica	Valor
R^2	0,758
R^2 -ajustado	0,740
Teste-F	42,04
Probabilidade (Teste F)	3,70 e-33
MSE	0,1021
Desvio Padrão (MSE)	0,01855
RMSE	0,3195

Com base no gráfico de valor real *versus* valor predito, apresentado na Figura 10, observa-se que o modelo tende a ter maior resíduo (erro) para desgastes reais acima de 2,0 mm/corrída, valores para os quais há menos observações, o que se comprova quando analisados o erro quadrático médio (*mean absolute error* - MSE) para as duas regiões:

- Predição com desgaste real $\leq 2,0$ mm/corrída - MSE: 0,0595
- Predição com desgaste real $> 2,0$ mm/corrída - MSE: 0,20308

Figura 10 – Gráfico do valor real *versus* valor predito pelo modelo



Fonte: Elaboração própria

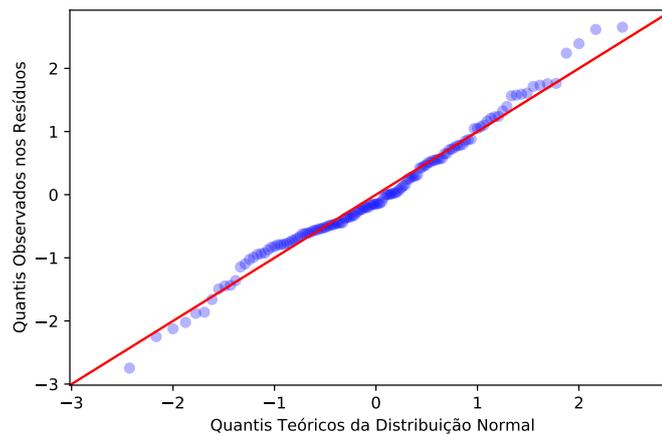
4.2.1 Validação do Modelo

Após estabelecido o modelo de regressão e identificada relação linear entre os valores reais e preditos, conforme mostrado na Figura 10, se faz necessário validar alguns pressupostos, dentre os quais se destacam: normalidade, homocedasticidade e independência dos resíduos. Estes pressupostos são fundamentais para entender a qualidade e validade do modelo.

De acordo com o coeficiente de determinação ajustado (R^2 -ajustado) da Tabela 5 acima, o modelo (4.1) é capaz de explicar 74% da variabilidade na variável resposta com base nas variáveis predictoras. Analisando o resultado do teste F, com base em um nível de significância de 0,05 e p-valor de $3,70 \times 10^{-33}$, a hipótese nula é rejeitada, o que indica que pelo menos uma variável selecionada é estatisticamente significativa para explicar a variabilidade da variável dependente.

Com intuito de verificar se a suposição de normalidade dos resíduos da regressão linear foi atendida, foi utilizado o gráfico Quantil-Quantil (Q-Q), que permite a comparação de distribuições de probabilidade. O interesse aqui é verificar se os erros do modelo construído seguem uma distribuição normal, premissa chave da regressão linear. A Figura 11 apresenta o gráfico Quantil-Quantil em que é esperado que os pontos estejam próximos à reta sem grandes desvios. De fato, é possível observar que a maioria dos pontos está próximos à reta, sugerindo a normalidade dos resíduos.

Figura 11 – Gráfico: Quantil-Quantil Plot

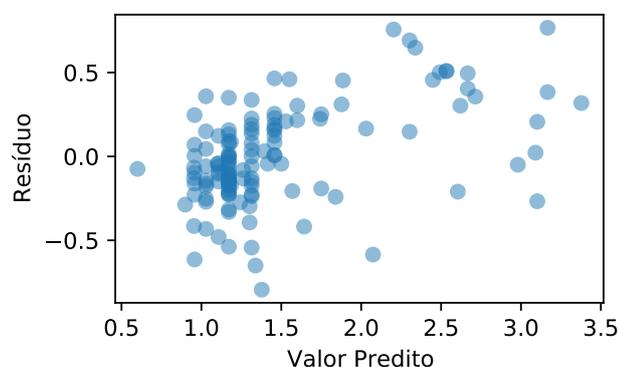


Fonte: Elaboração própria com uso do Python

Com o objetivo de confirmar a percepção da normalidade apresentada no gráfico Quantil-Quantil apresentado na Figura 11, foi aplicado o teste de normalidade D'Agostino-Pearson Omnibus. Para este teste, a hipótese nula significa aceitar que os resíduos seguem uma distribuição normal. Foi considerado um α de 0,05, portanto, 5% de probabilidade de erro quanto a rejeição da hipótese nula. Este teste resultou em um p-valor de 0,2828, valor este acima do α estabelecido de 0,05, portanto, a hipótese nula não pode ser rejeitada indicando que os resíduos seguem uma distribuição normal.

Para validação da premissa de homocedasticidade dos resíduos foi analisado o gráfico de distribuição de pontos. Neste gráfico os dados são exibidos em uma reta numérica, sendo que o eixo X representa o índice do dado exibido e o eixo Y representa o valor do resíduo correspondente ao dado apresentado. Conforme a figura 12, é possível observar que não há um padrão definido na distribuição dos pontos, o que indica uma possível homocedasticidade dos resíduos.

Figura 12 – Gráfico de distribuição de pontos dos resíduos

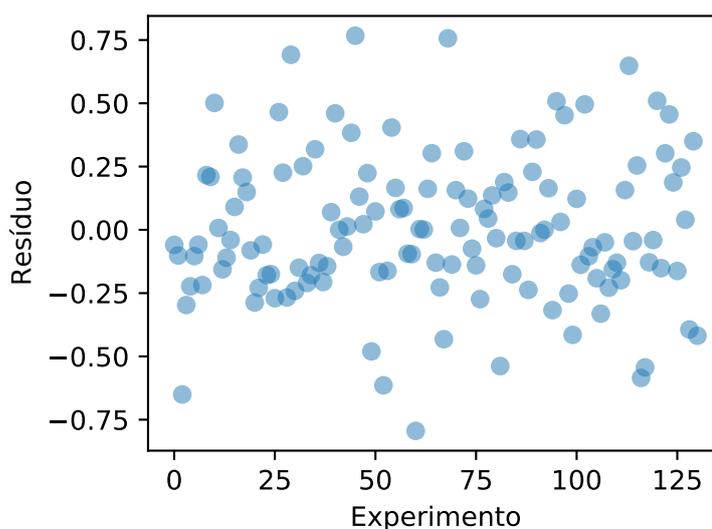


Fonte: Elaboração própria com uso do Python

Para validação da interpretação mencionada quanto à homocedasticidade foi aplicado o teste Goldfeld-Quandt da biblioteca statsmodel³. Para este teste foi aplicado um *alpha* de 0,05, ou seja, de 5% de probabilidade de erro quanto a rejeição da hipótese nula. A hipótese nula para este teste é que os resíduos são homocedásticos e a hipótese alternativa é de que os resíduos não são homocedásticos. Este teste resultou em um p-valor de 0.2654, valor este acima do *alpha* estabelecido de 0,05, portanto, a hipótese nula não pode ser rejeitada e os resíduos são homocedásticos.

Para validação da premissa de independência dos resíduos foi analisado o gráfico de distribuição de pontos, contendo no eixo x os experimentos pela ordem natural (chamado no gráfico de experimento) e no eixo y os valores residuais. Conforme a Figura 13, é possível observar que não há um padrão definido na distribuição dos pontos, o que indica a independência dos resíduos ou em outras palavras, não há uma auto-correlação. Essa premissa também é confirmada pela Figura 10.

Figura 13 – Gráfico de dispersão entre experimento e resíduo



Fonte: Elaboração própria com uso do Python

Com o objetivo de confirmar a independência dos resíduos apresentada no gráfico da Figura 13, também aplicamos o teste Durbin-Watson⁴ da biblioteca statsmodel. O teste retorna um coeficiente *d* em que se *d* for menor que 1,5 ou maior que 2,5, existe um problema severo de autocorrelação positiva ou negativa, respectivamente. Por outro lado, se *d* estiver entre 1,5 e 2,5, a autocorrelação é fraca ou ausente. O resultado do teste retornou um coeficiente $d=1,91$, o que indica pouca ou nenhuma auto-correlação entre os dados residuais.

A próxima seção dedica-se à exploração dos resultados obtidos com base na regressão.

³ <<https://www.statsmodels.org/stable/index.html>>

⁴ <https://www.statsmodels.org/dev/generated/statsmodels.stats.stattools.durbin_watson.html>

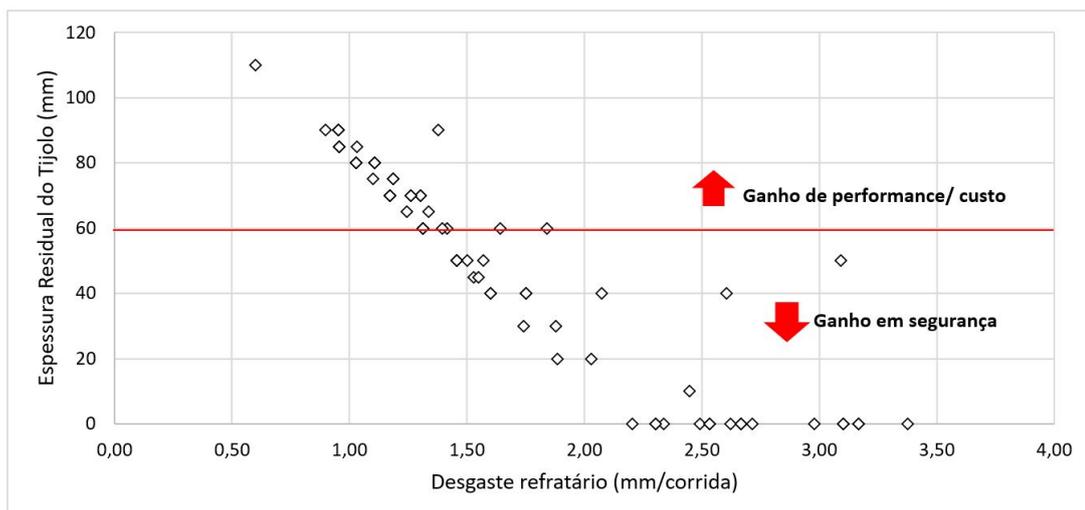
4.3 Aplicabilidade e discussões

A predição do desgaste refratário em milímetros por corrida tem como objetivo, dada uma margem de segurança, apoiar a decisão do número de corridas que uma panela de aço deve processar, de maneira a otimizar o recurso e garantir a segurança do processo e das pessoas. Sendo assim, conforme a espessura inicial dos tijolos refratários, o residual do tijolo até o afastamento da panela que seja considerado seguro dadas as incertezas do processo, e a taxa de desgaste médio por corrida, é possível calcular o número de corridas a serem realizadas com base na Equação (4.2):

$$\text{Número de Corridas} = \frac{\text{Espessura Inicial} - \text{Espessura Residual}}{\text{taxa_desg}} \quad (4.2)$$

A espessura dos tijolos utilizados na linha de escória com base no projeto refratário atual das panelas de aço da Aciaria 2 é de 152mm. Atualmente, utiliza-se como referência, além do número de 70 corridas, a espessura residual dos tijolos de 60mm como parâmetro para afastamento das panelas de aço para reparo intermediário. O valor de referência (60mm) é determinado com uma margem de segurança, dado que a inspeção é feita de maneira visual e a altas temperaturas, o que não garante 100% de assertividade. Por este motivo, algumas panelas atualmente são afastadas com residual mínimo abaixo ou acima deste valor, conforme indica a Figura 14, contendo as observações utilizadas para a construção da regressão:

Figura 14 – Gráfico do desgaste refratário *versus* espessura residual



Fonte: Elaboração própria

A espessura residual acima de 60mm indica que a panela poderia passar por mais ciclos operacionais, gerando economia de material refratário e prolongamento da campanha da panela, antes do afastamento para reparo. Por outro lado, quanto mais próximo de 0mm de espessura residual, maior o risco de furo na carcaça metálica, já que, quando esta medida é 0, todo o tijolo foi desgastado e a carcaça encontra-se sem proteção contra o aço líquido incandescente.

Por meio do gráfico é possível notar que dentro da amostra analisada há painéis que são afastadas com residual superior a 60mm (69 observações), chegando até a 110mm, e também painéis que são afastadas com residual igual a zero (17 observações), o que indica que todo o tijolo foi desgastado no processo. No primeiro caso, há um potencial de ganho de desempenho e custo, visto que estas painéis poderiam processar mais corridas que a quantidade realizada. Por outro lado, as painéis que desgastaram todo o tijolo representam um risco ao processo e às pessoas, visto a possibilidade de furo na carcaça e derramamento de aço líquido incandescente. Percebe-se ainda, que para a faixa de desgaste de 1,5 a 2,0 mm/corrída, a espessura residual fica mais próxima dos limites de 20 a 60mm.

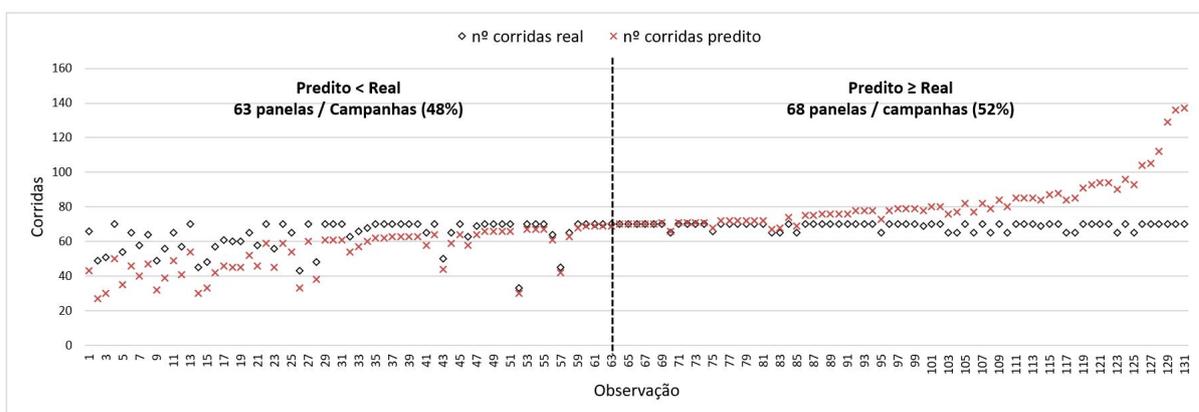
Sendo uma definição, que contempla determinada margem de segurança, o valor da espessura residual pode ser revisado de acordo com a avaliação dos especialistas de processo e a assertividade das ferramentas de decisão. Dessa forma, aplicou-se a regressão para a definição do número de corridas resultante da predição do desgaste refratário via regressão linear múltipla aplicada às observações reais, conforme a Equação 4.2, fazendo-se:

$$\text{Número de Corridas} = \frac{152\text{mm} - 60\text{mm}}{\text{taxa_desg}} \tag{4.3}$$

onde, *taxa_desg* é a taxa de desgaste obtida pela Equação 4.1 que especifica o modelo desenvolvido.

O gráfico da Figura 15 apresenta a diferença entre a política de afastamento praticada e a sugerida pelo modelo desenvolvido. Vale ressaltar que, na amostra analisada, as painéis são afastadas até a corrida de número 70. É possível observar que o modelo de regressão linear desenvolvido indicou o afastamento de 48% das painéis com um número de corridas menor que o praticado na realidade, e 52% das painéis com um número de corridas superior ao praticado na realidade.

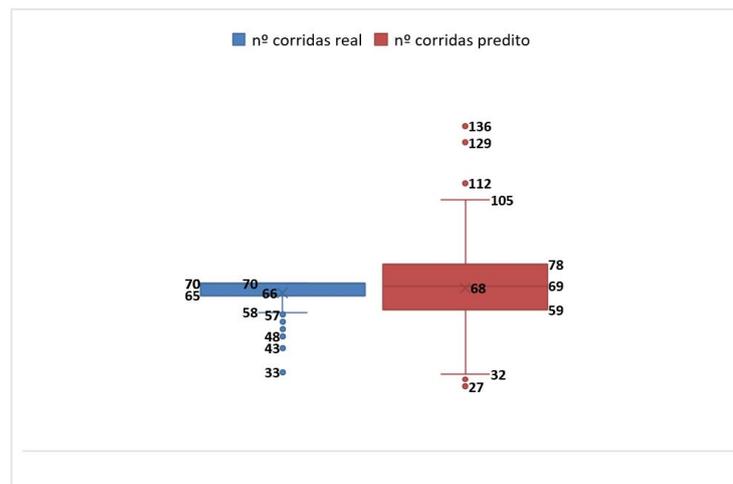
Figura 15 – Gráfico de nº de corridas real versus nº de corridas preditas pelo modelo



Fonte: Elaboração própria

As distribuições do número de corridas realizado e do sugerido pelo modelo são apresentadas no *boxplot* da Figura 16. Nota-se que a média do número de corridas até o reparo intermediário passe de 66 para 68 corridas considerando os valores sugeridos pela regressão. No entanto, percebe-se um aumento na amplitude dos valores sugeridos, atingindo um pico de 136 corridas previstas, ao passo que na política praticada, este valor limita-se a 70 corridas. Este resultado evidencia um potencial de ganho de performance do refratário e ganho em segurança com o uso do modelo em relação aos dados reais.

Figura 16 – *Boxplot* do nº de corridas real *versus* nº de corridas predito pelo modelo



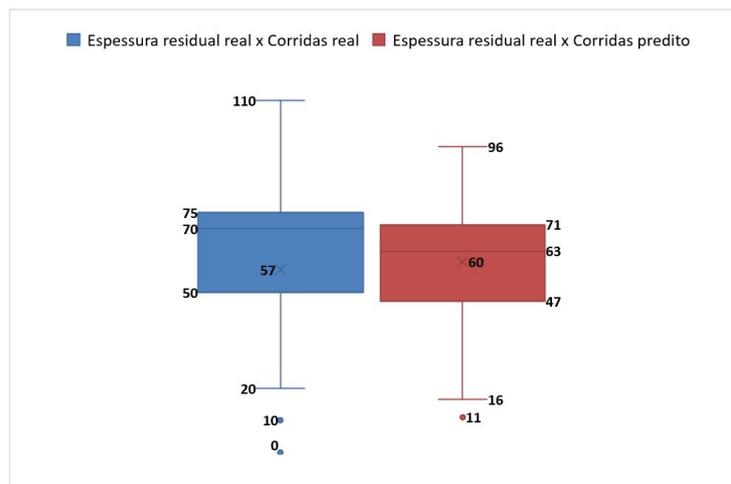
Fonte: Elaboração própria

Por fim, buscou-se quantificar o impacto da política de afastamento das panelas na espessura residual do refratário caso o modelo desenvolvido fosse aplicado. Para isso, foi calculada a espessura resultante utilizando-se a taxa de desgaste real, multiplicada pelo número de corridas previstas, a fim de avaliar as diferenças entre a espessura simulada e a espessura real. Essa avaliação pode ser vista na Figura 17.

Mantendo-se a coerência, o que se conclui é que, a espessura residual calculada com base no número de corridas predito possui média superior à real (3mm a mais) e variância menor, evidenciada pelo achatamento do *boxplot*. Além disso, os pontos de menor espessura, que no real foram de 0 mm para 17 panelas/campanhas, passam a ser 11mm para uma panela, com a aplicação da regressão, indicando maior segurança ao processo.

Sendo assim, considera-se que o batente da espessura residual de 60mm, conforme mencionado, atende com margem de segurança razoável ao processo. Ademais, caso aplicado à amostra, traria um benefício de, em média, 2 corridas a mais por panela, que corresponde a um aumento de 3% na vida das panelas de aço. Dada a representatividade do refratário das panelas de aço para o custo operacional de uma aciaria, este ganho de desempenho representa uma redução considerável no custo do produto final, sem necessidade de investimento, com benefício para segurança e gestão do processo.

Figura 17 – *Boxplot* de espessura residual das painelas



Fonte: Elaboração própria

5 Considerações Finais

As painéis de aço revestidas por material refratário desempenham um papel fundamental no processo de fabricação e tratamento do aço na aciaria. Ao longo da operação, o revestimento dessas painéis vai sofrendo desgaste à medida em que são submetidas às variáveis operacionais e de processo, sendo necessária a realização de reparos intermediários e substituição total do refratário para continuidade da operação. O custo com o material, mão de obra e indisponibilidade das painéis para operação, representam parte significativa do custo operacional da Aciaria 2, o que torna a otimização do seu desempenho objeto relevante de estudo. Além disso, a gestão do desgaste refratário tem impacto direto na produção, qualidade do aço e segurança do processo e das pessoas.

Dividida por áreas de acordo com os mecanismos de desgaste preponderantes, a linha de escória é a região do revestimento mais suscetível aos mecanismos de desgaste presentes no processo, e por esse motivo, em geral, determinam a vida da painela. Conforme mencionado nos capítulos anteriores, atualmente, o reparo intermediário, com a substituição da linha de escória é realizado com base em parâmetros fixos determinados e na expertise de especialistas do processo, que determinam como parâmetro a realização de 70 corridas e/ou espessura residual dos tijolos de 60mm. Entretanto, toda a complexidade e variabilidade do processo, além da dificuldade de inspeção da painela em operação mostram que muitas das vezes estes valores de referência não levam à otimização da performance do recurso e segurança do processo. Por este motivo, o presente trabalho teve como foco avaliar como o uso da ciência dos dados poderia apoiar no processo de decisão do afastamento da painela para troca intermediária do refratário.

A técnica utilizada para a predição do desgaste e conseqüentemente, suporte à tomada de decisão foi a regressão linear múltipla, a fim de que o modelo fosse interpretável, ou seja, que as variáveis de maior relevância sejam identificadas e potencialmente manipuladas, além da fácil aplicação, sem a necessidade de grandes desenvolvimentos de sistemas de informação. Ao definir o modelo de regressão, bem como as variáveis mais importantes para o processo com a utilização de um conjunto de técnicas para análise e validação, concluímos que:

- Para se chegar às variáveis mais relevantes para o modelo, foi necessário um estudo aprofundado utilizando-se do conhecimento da equipe responsável, revisão bibliográfica e técnicas estatísticas;
- O modelo apresentou um R^2 ajustado de 74%, que é considerado alto pelos especialistas da área dada a complexidade dos dados e do processo, e selecionou 9 variáveis de uma base de dados inicial de mais de 23 mil registros sumarizadas após tratamento dos dados em 51 variáveis e 131 observações;

- A variável resposta, contida na base de dados do perfil *post mortem* da troca intermediária, onde é indicado o residual real do tijolo refratário em mm, medido e registrado manualmente foi essencial para o treinamento e validação do modelo gerado;
- A aplicação do parâmetro de 60mm de residual para o cálculo do número de corridas a serem realizadas pela panela, atende, com margem de segurança razoável ao processo, e caso aplicado à amostra, traria um benefício de, em média 2 corridas a mais por panela, o que corresponde a um aumento de 3% na vida das panelas de aço, além do aumento médio do residual, garantindo maior segurança operacional;
- O resultado obtido é significativo frente à representatividade das panelas de aço para o custo operacional da Aciaria 2. Além disso, este ganho de performance representa uma redução no custo do produto final, sem necessidade de investimento, com benefício para segurança e gestão do processo.

O modelo construído tem como alvo fornecer à operação uma ferramenta adicional na definição de afastamento das panelas para reparo intermediário, e em conjunto com as inspeções visuais e o conhecimento tácito dos especialistas, torne o processo mais eficiente e assertivo.

5.1 Limitações e trabalhos futuros

Cabe ressaltar que os resultados do trabalho apresentam limitações sendo necessário, com certa frequência, recalibração e geração de novos dados, frente à mudanças inerentes do processo, como equipes, definição na fabricação dos tipos de aços e utilização de equipamentos, fornecedores do material refratário, engenharia do projeto refratário, etc. Além disso, o modelo é específico para o processo de troca intermediária de refratário das panelas da Aciaria 2 da Usiminas, embora possa contribuir e ajudar outras empresas e estudiosos na escolha das variáveis enquanto lacuna teórica. Por outro lado, o uso da regressão pressupõe relação linear entre as variáveis dependentes e independentes, o que pode ser uma simplificação da realidade. Além disso, a aplicação da regressão está limitada ao intervalo de valores para o qual a mesma foi construída, podendo gerar distorções para valores muito discrepantes.

Em relação à trabalhos futuros, sugere-se aplicar outros métodos de aprendizagem de máquina ao problema, como redes neurais, para possivelmente, obter um resultado diferente e mais assertivo quanto às predições, e que poderá ser utilizado para comparação com o modelo gerado. Além disso, o trabalho pode ser usado como base para definição de um modelo semelhante para a definição do reparo final, no qual é realizada a substituição completa do refratário, quando a panela será 100% demolida, incluindo todas as regiões da panela - sola, linha de metal e linha de escória, além da aplicação para as panelas da Aciaria 1 da Usiminas ou outras empresas.

Referências

- BORGES, R. A. A. **Análise multivariada dos principais fatores que afetam o desgaste (vida) do revestimento refratário das painéis de transporte de aço na Aciaria da Usina de Cubatão-Usiminas**. Tese (Doutorado) — Universidade de São Paulo, 2016.
- BUGNION, P.; MANIVANNAN, A.; NICOLAS, P. R. *Scala: Guide for data science professionals*. Birmingham: Packt Publishing, 2017. Acesso em: 28 Dez. 2022.
- CADE, D. de E. E. D. *Cadernos do cade: Indústria siderúrgica. SEPN 515 Conjunto D, Lote 4, Ed. Carlos Taurisano*, 2022.
- FAVERO, L. P. **Análise de Dados**. Grupo GEN, 2015. ISSN E-book. ISBN 9788595153226. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788595153226/>>. Acesso em: 16 Dez. 2022.
- FAVERO, L. P. **Manual de Análise de Dados - Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®**. Grupo GEN, 2017. ISSN E-book. ISBN 9788595155602. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788595155602/>>. Acesso em: 27 Dez. 2022.
- FERREIRA, R. G. C.; MIRANDA, L. B. A. d.; PINTO, R. A.; AL. et. **Preparação e Análise Exploratória de Dados**. Porto Alegre: Grupo A Educação S.A., 2021. ISBN 9786556902890. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9786556902890/>>. Acesso em: 21 Dez. 2023.
- GRADY, N.; CHANG, W. *Nist big data interoperability framework*. v. 1, 2015. Disponível em: <<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf>>. Acesso em: 28 Dez. 2022.
- HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; AL. et. **Análise multivariada de dados**. Grupo A, 2009. ISSN E-book. ISBN 9788577805341. Disponível em: <<https://integrada.minhabiblioteca.com.br/reader/books/9788577805341>>. Acesso em: 26 Dez. 2022.
- LEITE, F. C.; LUZ, A. P.; PANDOLFELLI, V. C. Características e mecanismos de desgaste dos refratários mgo-c usados na linha de escória de painéis de aço. **Cerâmica**, MCB UP Ltd, v. 60, n. 2, p. 348–365, 2014.
- LENZ; SILVA, G. F. B. Introdução à siderurgia, cap. 9 – refratários para siderurgia. **ABM Associação Brasileira de Metalurgia, Materiais e Mineração**, Marcelo B. Mourão, p. 428, 2007.
- MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; BYERS, A. H. *Big data: The next frontier for innovation, competition, and productivity*. **The McKinsey Global Institute**, 2011. Disponível em: <<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>>. Acesso em: 28 Dez. 2022.
- MATOS, D. *Ciência de dados e soluções*. **Scala: Guide for Data Science Professionals**, 2015. Disponível em: <<https://www.cienciaedados.com/ciencia-de-dados-e-solucoes/>>. Acesso em: 28 Dez. 2022.

- NETO, J. F. C.; MARQUES, E. V. **Tomada de decisões gerenciais com analítica de dados**. Editora Alta Books, 2020. ISSN E-book. ISBN 9788550817101. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788550817101/>>. Acesso em: 16 Dez. 2022.
- PACHECO, B. B. M.; DISCONZI, M. S. Ciência de dados: Enfoque no desafio do processamento. **Research, Society and Development**, v. 8, n. 11, 2019. Disponível em: <<https://www.redalyc.org/journal/5606/560662202012/560662202012.pdf>>. Acesso em: 28 Dez. 2022.
- PENTEADO, K. Métricas de avaliação para séries temporais. **Alura**, Junho 2021. Disponível em: <<https://www.alura.com.br/artigos/metricas-de-avaliacao-para-series-temporais>>. Acesso em: 24 Mar. 2023.
- QUINTELA, M. **Caracterização de cerâmicas refratárias para panela de aço**. [S.l.]: Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Engenharia de Materiais, 2003.
- RAUTENBERG, S.; CARMO, P. R. V. d. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Studies: Research Trends**. **13:1**, p. 56–67, 2019. ISSN 1981-1640. Disponível em: <<https://doi.org/10.36311/1981-1640.2019.v13n1.06.p56>>. Acesso em: 26 Dez. 2022.
- REFRATORIES, T. T. A. of. The technical association of refractories, refractories handbook. Refractories Handbook, Japão, 1998. Acesso em: 15 jan. 2023.
- SICSÚ, A. L.; DANA, S. **ESTATÍSTICA APLICADA: ANÁLISE EXPLORATÓRIA DE DADOS**. Editora Saraiva, 2012. ISSN E-book. ISBN 9788502177574. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788502177574/>>. Acesso em: 26 Dez. 2022.

APÊNDICE A – Descrição completa das Variáveis

Figura 18 – Descrição Variáveis: Gerais, Temperatura, Mix, Rota e Tempos

Descrição de Variáveis
Gerais
NUM_CORR_ACI: Número da Corrida
DT_PRODUCAO: Data de Produção da Corrida
NUM_PAN_ACO: Número da panela de aço
NUM_CAMP_PA: Número da campanha da panela
VIDA_PAN_ACO: Vida da panela de aço (número de corridas da panela)
Temperatura do processo
TEMP_FS: Temperatura fim de sopro (no convertedor)
TEMP_VAZ: Temperatura de vazamento
TEMP_POS_VAZ: Temperatura Pós Vazamento
TEMP_INICIAL_TRAT: Temperatura inicial de Tratamento
TEMP_FINAL_TRAT: Temperatura final de Tratamento
DELTA_AQUEC: Delta Aquecimento (Temperatura Final - Temperatura Inicial)
TEMP_PANELA: Temperatura Panela
Mix Aços Fabricados
FAMILIA_ACO: Alto Carbono, Baixo Enxofre, Comum, Desgaseificada, Duplo Refino Geral, Peritético, Ultra Baixo Carbono Geral
<u>Variações:</u> Alto Carbono, Baixo Enxofre, Comum, Desgaseificada, Duplo Refino Geral, Peritético, Ultra Baixo Carbono Geral
Tipo de Rota / Tratamento
ROTA_ACO: Rota do aço conforme equipamentos em que cada corrida passou
COD_TRATAMENTO: Definido através de combinações de variáveis como Equipamento (tratamento), Composição Química, Adições, etc. - detalhe em Código de tratamento
TRATAMENTO: Equipamento onde ocorreu o trabalhamento da Corrida: DLD- Convertedor, FP - Fomo Panela , RH - Desgaseificador a vácuo, OB - CasOB
SEQ_TRATAMENTO: Quando a corrida passa mais de uma vez no mesmo equipamento
Tempos Operacionais
TEMPO_TOT_PAN_QUENTE: Tempo total de Panela Quente
TEMPO_PAN_CHEIA: Tempo Panela Cheia
TEMPO_PAN_VAZIA: Tempo Panela Vazia
TEMPO_CICLO_PAN: Panela Cheia + Panela Vazia
TEMPO_AQUECIMENTO: Tempo de Aquecimento = Diferença entre Início e Fim do Sopro de Oxigenio (DT_INIC_SOPRO_OXIG_OBTD -DT_FIM_SOPRO_OXIG_OBTD)
TEMPO_TRATAMENTO: Tempo de Tratamento = Tempo de Refino
TEMPO_VAZ_LD: Tempos de Vazamento no Convertedor
TEMPO_REFINO: Tempo de Tratamento = Tempo de Refino

Fonte: Elaboração própria

Figura 19 – Descrição Variáveis: Condições Panela, Carga e Uso do Plug

Descrição de Variáveis
Condições da Panela
CARGA_ACO_LING: Carga (peso) aço lingotado na Máquina
DT_ABERT_PAN: Data da Abertura da Panela
DT_ULT_REFINO: Data do Último Refino
DT_FIM_VAZAM_ACO_LD: Data Fim de Vazamento do Aço no Convertedor
DT_INIC_VAZ: Data Início de Vazamento
DT_FIM_LINGNTO: Data fim de Lingotamento
DT_FIM_LINGNTO_ATUAL: Data fim de Lingotamento Atual
VL_PESO_L8: Calculo: Peso Panela Cheia - Peso Panela Vazia
VL_S_VIS: Valor do Enxofre Visado
Carga das Painelas
CARGA_VAZADA_LD: Carga (peso) vazada no convertedor
VL_QTD_BANHO_GUSA: Quantidade de carregamento de Gusa na Panela
Uso do Plug e Lança de Injeção de Gases
IN_PLUG: Funcionamento do Plug - quando ok CÓD = 7
PLUG_VAZAO: Vazão do Plug
PLUG_TEMPO_INJ: Tempo de Injeção no Plug
VL_PESO_INJ_PO_OBTD: FORNO PANELA: Injeção de pó (mistura dessulfurante) obtida no Forno Panela
VL_TEMPO_INJ_PO_OBTD: FORNO PANELA: TEMPO_INJ_PO_OBTD = Campo DT_FIM_INJ_PO_OBTD – Campo DT_INIC_INJ_PO_OBTD
VL_PESO_INJ_CAO_F2_OBTD: CASOB: Injeção de CaO no CasOB
VL_TEMPO_INJ_CAO_F2_OBTD: CASOB: Campo DT_FIM_INJ_CAO_F2_OBTD – Campo DT_INIC_INJ_CAO_F2_OBTD
VL_TEMPO_HOMOGEINIZACAO: Dados de Injeção de Argônio: TEMPO_HOMOGEINIZAÇÃO = campo DT_FIM_HOMOGENEIZ – campo DT_INIC_HOMOGENEIZ

Fonte: Elaboração própria

Figura 20 – Descrição Variáveis: Código Tratamento

Código Tratamento		
Código Carregado	Objetivo do Tratamento	Código Sequenciado (Rota)
Em Branco	Estação de Argônio - Ajuste Composição Química e Temperatura	Em Branco
	CAS-OB - Ajuste Composição Química e Temperatura	C1
	Forno Panela - Ajuste Composição Química e Temperatura	C7
	Corridas destinadas ao RH para aquecimento de vaso, recuperação de temperatura, teste de vácuo, etc.	A2
A	RH - Desgaseificação - Carbono da sigla > 2 pontos	A1
	RH - Desgaseificação - Carbono da sigla ≤ 2 pontos	A3
C	CAS-OB - Dessulfuração	C1
	CAS-OB - Globalização de inclusões / adição de cálcio	
	Forno Panela - Dessulfuração	C8
	Forno Panela - Globalização de inclusões / adição de cálcio	
	Forno Panela - Aços alto carbono - Carbono sigla > 30 pontos	
	Demais tipos de aços e os recuperados no Forno Panela	
F	Dessulfuração no Forno Panela e posterior Desgaseificação no RH	F8
	Dessulfuração no CAS-OB e posterior Desgaseificação no RH	F1

Fonte: Elaboração própria

Figura 21 – Descrição Variáveis: Adições, Taxa e Perfil

Descrição de Variáveis
Adições em Painela
ADIC_MNAC: Ferro Manganês Alto Carbono
ADIC_MNMC: Ferro Manganês Médio Carbono
ADIC_SIMN: Ferro Silício Manganês
ADIC_CASI: Ferro cálcio silício fio horizontal
ADIC_FLUO: Fluorita na painela
ADIC_CALE: Cal do vazamento do convertedor
ADIC_ESIN: Escória Sintética
ADIC_CARB: Carbureto
ADIC_FCAL: Ferro Cálcio Alumínio
ADIC_NEFE: Nefelina
ADIC_FECA: Ferro Cálcio
ADIC_BRIQ: Briquete dessulfurante
ADIC_MNBP: Ferro Manganês Baixo Fósforo
ADIC_MNEL: Manganês Eletrolítico
ADIC_CASV: Ferro Cálcio Silício fio vertical
ADIC_FCAH: Ferro Cálcio Alumínio fio horizontal
ADIC_BRID: Briquete Desfosforante
Taxa de Desgaste
Variável Resposta: Cálculo: Espessura inicial do refratário - Espessura atual (medida) / N° de corridas da painela (vida útil)
Perfil de Demolição
Variável Validação: Avaliação feita após o fim da vida útil da painela onde ocorre a mensuração do material residual, indicando o desgaste real do material refratário

Fonte: Elaboração própria

APÊNDICE B – Agrupamento das variáveis

Tabela 6 – Agrupamento das variáveis

Agrupamento	Variáveis
Agrupados por soma	TEMPO_AQUE_FP_y, TEMPO_AQUE_OB_y, TEMPO_AQUE_RH_y, TEMPO_PAN_CHEIA, TEMPO_PAN_VAZIA_y, VL_TEMPO_HOMOGEINIZACAO_y, VL_TEMPO_INJ_CAO_F2_OBTD_y, VL_TEMPO_INJ_PO_OBTD_y, TEMPO_REF_FP_y, TEMPO_REF_LD_y, TEMPO_REF_OB_y, TEMPO_REF_RH, DESGASEIFICADA, DRGERAL, F1, F8, FP2, LD4, LD5, ML1, ML2, OB1, OB2, PERITETICO, RH2, RHC, RHD, UBCGE- RAL, UBCUSICORE, ALTOCARBONO, BAIXOENXOFRE, VL_QTD_BANHO_GUSA, RH
Agrupados por média	ADIC_BRID, ADIC_BRIQ, ADIC_CALE, ADIC_CARB, ADIC_CASI, ADIC_CASV, ADIC_ESIN, ADIC_FCAL, ADIC_FECA, ADIC_FLUO, ADIC_MNBP, ADIC_MNEL, ADIC_MNMC, ADIC_SIMN, TEMPO_AQUE_FP_x, TEMPO_AQUE_OB_x, TEMPO_AQUE_RH_x, TEMPO_PAN_VAZIA_x, TEMP_POS_VAZ_x
Agrupados por percentual	IN_PLUG_FP7, IN_PLUG_OB7, A1_PART, A2_PART, A3_PART, C1_PART, C7_PART, C8_PART, COMUM_PART, DESGASEIFICADA_PART, DRGERAL_PART, F1_PART, F8_PART, FP2_PART, UBCGERAL_PART, UBCU- SICORE_PART, PERITETICO_PART, DR, FP, UBC, BAIXO_PART, ALTO_PART, LD_PART, OB_PART, RH_PART, COMUM

Fonte: Elaboração própria