



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



Análise do Processo de Recozimento de Aço Inoxidável Utilizando Modelos de Aprendizado de Máquina.

Ana Paula Marques de Oliveira Machado
Angelo Ferreira Costa
Gláucio Barros Barcelos

João Monlevade, MG
2024

Ana Paula Marques de Oliveira Machado
Angelo Ferreira Costa
Gláucio Barros Barcelos

Análise do Processo de Recozimento de Aço Inoxidável Utilizando Modelos de Aprendizado de Máquina.

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: Dr. George Henrique Godim da Fonseca

João Monlevade, MG

2024

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

C837a Costa, Angelo Ferreira.

Análise do processo de recozimento de aço inoxidável utilizando modelos de aprendizado de máquina. [manuscrito] / Angelo Ferreira Costa. Ana Paula Marques de Oliveira Machado. Gláucio Barros Barcelos. - 2024.

53 f.: il.: color., gráf., tab..

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

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

1. Aço - Indústria. 2. Aço inoxidável. 3. Aprendizado do computador. 4. Controle de temperatura. I. Barcelos, Gláucio Barros. II. Machado, Ana Paula Marques de Oliveira. III. Fonseca, George Henrique Godim da. IV. Universidade Federal de Ouro Preto. V. Título.

CDU 658.5:004.85

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



FOLHA DE APROVAÇÃO

Ana Paula Marques de Oliveira Machado
Angelo Ferreira Costa
Gláucio Barros Barcelos

Análise do Processo de Recozimento de Aço Inoxidável Utilizando Modelos de Aprendizado de Máquina

Trabalho de conclusão de curso apresentado ao curso de Especialização em Ciência de Dados da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Especialista em Ciência de Dados

Aprovada em 06 de março do ano de 2024

Membros da banca

Prof. Dr. George Henrique Godim da Fonseca - Orientador - Universidade Federal de Ouro Preto
Dra. Sarah Negreiros de Carvalho Leite - Universidade Federal de Ouro Preto
Dr. Geovane Martins de Castro - Aperam South America

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 01/07/2024



Documento assinado eletronicamente por **George Henrique Godim da Fonseca, PROFESSOR DE MAGISTERIO SUPERIOR**, em 11/07/2024, às 16:56, 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 **0736196** e o código CRC **7CC35512**.

Dedicamos este trabalho a Deus, agradecendo por Sua orientação e inspiração ao longo desta jornada acadêmica. Que este estudo seja uma expressão da nossa gratidão pela Sua presença constante em nossas vidas. Amém!

Agradecimentos

Gostaríamos de expressar nossa sincera gratidão a todas as pessoas que contribuíram para a realização deste trabalho e para nossa jornada acadêmica como um todo.

Gostaríamos de estender nossos mais profundos agradecimentos ao nosso orientador, George Henrique Godim, pela orientação valiosa, apoio constante e insights perspicazes que foram fundamentais para o desenvolvimento deste trabalho. Sua dedicação e paciência foram essenciais para o nosso crescimento acadêmico e profissional.

Agradecemos também às nossas famílias, que estiveram ao nosso lado durante toda essa jornada, oferecendo seu apoio incondicional, compreensão e incentivo. Seu amor e encorajamento foram fundamentais para nos manter motivados e focados em alcançar nossos objetivos.

Aos professores do curso e aos nossos colegas, em especial ao Mychell Laurindo, expressamos nossa gratidão por dedicarem seu tempo e expertise na análise deste trabalho e por fornecerem sugestões valiosas para o seu aprimoramento.

Por fim, gostaríamos de agradecer a todos os participantes da pesquisa, cujas contribuições foram fundamentais para a realização deste estudo.

*Entrega o teu caminho ao Senhor;
confia Nele, e Ele tudo fará.
Salmos 37:5*

Resumo

Os aços inoxidáveis austeníticos desempenham um papel fundamental na indústria devido à sua resistência, sendo amplamente utilizados em diversas aplicações. O controle da temperatura de recozimento do material é de suma importância para o processo. A sua variação pode causar mudanças de propriedades mecânicas e micro-estruturas dos materiais, gerando materiais com não conformidades, o que pode gerar perdas. Este trabalho propõe utilização e comparação de diversos modelos de aprendizado de máquinas, na análise do processo de recozimento dos aços inoxidáveis austeníticos, para classificar e identificar as principais variáveis de processo que impactam no resultado final do material processado. Em primeiro lugar, os dados de várias bobinas foram coletados considerando as variáveis de processo de recozimento e as temperaturas objetivadas alcançadas. Em seguida os dados coletados passaram por um processo de tratamento, limpeza, organização e padronização de todas as informações coletadas. Finalmente, os modelos de aprendizado de máquina são aplicados aos dados para definir as relações entre as variáveis de processo (espessura, largura, temperatura de zonas, entre outras) e a temperatura final do material. Os modelos *Random Forest Classifier* e *XGBoost* demonstraram resultados comparáveis em termos de acurácia, com uma pequena vantagem observada para o *Random Forest Classifier*, registrando uma taxa de acurácia de 0,89%, em comparação com os 0,88% do *XGBoost*. Em relação à AUC, ambos os modelos exibiram desempenhos semelhantes, com valores aproximados de 0,90 em ambas as tabelas. Quando avaliados em termos de sensibilidade, precisão e F1, os resultados indicam que ambos os modelos demonstram eficácia na identificação de classes positivas e negativas, com valores próximos a 0,89 para o *Random Forest Classifier* e 0,88 para o *XGBoost*. Os resultados mostraram que a utilização de aprendizado de máquina revelou *insights* importantes sobre os fatores que influenciam a temperatura final do material.

Palavras-chaves: Aço Inoxidável, Recozimento, temperatura, Controle, aprendizado de máquina.

Abstract

Austenitic stainless steels play a fundamental role in industry due to their resistance, being widely used in various applications. Controlling the annealing temperature of the material is extremely important for the process. Its variation can cause changes in the mechanical properties and microstructures of materials, generating non-conforming materials, which can generate losses. This work proposes the use and comparison of several machine learning models, in the analysis of the annealing process of austenitic stainless steels, to classify and identify the main process variables that impact the final result of the processed material. Firstly, data from multiple coils was collected considering the annealing process variables and the target temperatures achieved. The collected data then went through a process of treatment, cleaning, organization and standardization of all collected information. Finally, machine learning models are applied to the data to define the relationships between process variables (thickness, width, zone temperature, among others) and the final temperature of the material. The *Random Forest Classifier* and *XGBoost* models demonstrated comparable results in terms of accuracy, with a slight advantage observed for the *Random Forest Classifier*, recording an accuracy rate of 89%, compared to 88% of *XGBoost*. Regarding the AUC, both models exhibited similar performances, with approximate values of 0.90 in both tables. When evaluated in terms of sensitivity, precision, and F1, the results indicate that both models demonstrate effectiveness in identifying positive and negative classes, with values close to 0.89 for the *Random Forest Classifier* and 0.88 for the *XGBoost*. The results showed that the use of machine learning revealed important insights into the factors influencing the final temperature of the material.

Keywords: Stainless Steel, Annealing, Temperature, Control, Machine Learning.

Lista de ilustrações

Figura 1 – Layout da linha RB3	7
Figura 2 – Fluxograma dos passos seguidos na metodologia.	16
Figura 3 – Percentual de distribuição para cada classe na base de dados.	21
Figura 4 – <i>Boxplots</i> para cada variável numérica.	23
Figura 5 – Correlação entre as variáveis numéricas	25
Figura 6 – Fluxograma de Aprendizado de Máquina com <i>PyCaret</i>	27
Figura 7 – Matriz de Confusão	30
Figura 8 – Relatório de métricas de avaliação para a classificação	31
Figura 9 – Variáveis mais importantes para modelo	32
Figura 10 – Gráfico SHAP para modelo	33
Figura 11 – Árvore de Decisão	35
Figura 12 – Proporção de amostras quando aplicada nova regra de processo.	37

Lista de tabelas

Tabela 1 – Composições químicas nominais (% em peso) e propriedades mecânicas de algumas classes de aços inoxidáveis austeníticos.	4
Tabela 2 – Dados de Pendências por Mês	11
Tabela 3 – Percentual de dados faltantes para cada variável.	18
Tabela 4 – Parâmetros de processo para o aço P304 com largura $< 1400mm$	19
Tabela 5 – Parâmetros de processo para o aço P304 com largura $\geq 1400mm$	19
Tabela 6 – Parâmetros de tolerância operacional o aço P304 $< 1400mm$	19
Tabela 7 – Parâmetros de tolerância operacional o aço P304 $\geq 1400mm$	20
Tabela 8 – Estatística resumida de variáveis numéricas.	22
Tabela 9 – Resultado da comparação de modelos.	29
Tabela 10 – Resultados do ajuste de hiperparâmetros com validação cruzada de 10 vezes para os modelos selecionados.	29
Tabela 11 – Resultados de predição para modelos ajustados sobre dados de teste.	34
Tabela 12 – Resultados de predição para modelos finalizados sobre dados não vistos.	34
Tabela 13 – Regra extraída da árvore de decisão (rota em azul) para classificação de amostras dentro da norma (SIM).	36
Tabela 14 – Regras de processo onde o resultado foi melhor.	36
Tabela 15 – Regra final de processo.	37

Lista de abreviaturas e siglas

AIAs Aços Inoxidáveis Austeníticos

AM Aprendizado de Máquina

AUC Area Under the Curve

CFC Cúbica de Faces Centradas

CRAVE CRAVE Industry

DT *Decision Tree Classifier*

KNN *K Neighbors Classifier*

ML *Machine Learning*

PCI Poder Calorífico Inferior

PIMS Sistema de gerenciamento de informações de processo

RB3 Recozimento e decapagem 3

RF *Random Forest Classifier*

SHAP *SHapley Additive exPlanations*

XGBOOST *Extreme Gradient Boosting*

Sumário

1	INTRODUÇÃO	1
1.1	Objetivo geral	2
1.2	Objetivos específicos	2
1.3	Contribuições	2
2	FUNDAMENTAÇÃO TEÓRICA	3
2.1	Aços inoxidáveis austeníticos	3
2.2	O estado encruado	5
2.3	Recuperação e recristalização	5
2.4	Descrição do processo produtivo	7
2.5	Variação de temperatura durante o processo	9
2.6	Extração, tratamento e exploração dos dados	12
2.6.1	Extração dos dados	12
2.6.2	Tratamento dos dados	13
2.6.3	Exploração analítica	13
2.7	Treinamento e validação dos modelos	14
2.7.1	Validação Cruzada (<i>Cross-Validation</i>)	15
2.7.2	Parada antecipada (<i>Early stopping</i>)	15
3	METODOLOGIA	16
3.1	Coleta dos dados	16
3.2	Tratamento dos dados	17
3.2.1	Definição das variáveis	17
3.2.2	Valores inválidos	18
3.2.3	Dados faltantes	18
3.2.4	Codificação de variáveis	18
3.3	Análise exploratória dos dados	20
3.3.1	Análise de dados desbalanceados	20
3.3.2	Análise de <i>outliers</i>	21
3.3.3	Análise de correlação	22
3.4	Desenvolvimento dos modelos de aprendizado de máquina	24
4	RESULTADOS	28
4.1	Comparação dos modelos	28
4.2	Criação dos modelos e otimização dos hiper parâmetros	29
4.3	Análise do treinamento	29

4.4	Finalização e predição	34
4.5	Extração de regras	35
5	CONSIDERAÇÕES FINAIS	38
5.1	Trabalhos futuros	39
	REFERÊNCIAS	40

1 Introdução

Os aços inoxidáveis desempenham um papel crucial na indústria devido à sua excepcional resistência à corrosão e oxidação, sendo amplamente utilizados em diversas aplicações. Entre essas ligas, os aços inoxidáveis austeníticos destacam-se pela sua versatilidade e resistência, proporcionando propriedades mecânicas adequadas para trabalho a frio e excelente resistência à corrosão. No entanto, a qualidade e o desempenho das bobinas de aço inoxidável são fortemente influenciados pelo processo de recozimento. Esse tratamento térmico é essencial para aliviar tensões internas, melhorar a ductilidade e a uniformidade da estrutura cristalina, e garantir a eliminação de defeitos de conformação. Portanto, a correta execução do processo de recozimento é fundamental para assegurar a obtenção de propriedades mecânicas desejadas e preservar as características de resistência à corrosão, garantindo a qualidade e durabilidade dos produtos finais fabricados com essas bobinas de aço inoxidável austenítico.

Um modelo capaz de prever a temperatura final das bobinas de aço inoxidável para o processo de recozimento é crucial na indústria, pois a temperatura final das bobinas afeta diretamente sua qualidade e propriedades físicas. A temperatura final das bobinas é realizada através de pirômetros dispostos ao longo do forno de recozimento e são utilizados modelos simplificados. No entanto, esses métodos podem ser imprecisos e sujeitos a variações. Portanto, desenvolver um modelo de aprendizado de máquina preciso e confiável para prever a temperatura final das bobinas pode melhorar significativamente a eficiência e a qualidade dos processos industriais.

Paralelamente, a análise e comparação de modelos em projetos de aprendizado de máquina desempenham um papel crucial na busca por soluções eficazes e precisas. Métricas como acurácia, precisão, *recall* e *F1-score* são essenciais para avaliar o desempenho dos modelos desenvolvidos, proporcionando *insights* valiosos sobre sua capacidade preditiva e comportamento em diferentes cenários. Neste contexto, este estudo se concentra na análise comparativa de quatro modelos - *Random Forest Classifier* (RF), *Extreme Gradient Boosting* (XGBOOST), *Decision Tree Classifier* (DT) e *K Neighbors Classifier* (KNN) utilizando métricas padrão para avaliar seu desempenho.

Além de identificar o modelo com melhor desempenho, este trabalho busca compreender as nuances dos resultados obtidos, especialmente após o ajuste dos hiperparâmetros dos modelos, destacando a importância de uma abordagem criteriosa na escolha e otimização dos algoritmos de aprendizado de máquina. Ao fornecer uma visão abrangente sobre a relação entre as propriedades dos aços inoxidáveis austeníticos e a eficácia dos modelos de aprendizado de máquina, este estudo contribui para o avanço contínuo das aplicações práticas desses materiais e das técnicas de análise de dados.

1.1 Objetivo geral

O presente trabalho tem como objetivo determinar através de variáveis de processo se o material está conforme ou não a uma receita padrão, analisando as variações de temperatura da tira em relação às demais variáveis do forno: temperaturas de zonas, largura, espessura, velocidade, **Poder Calorífico Inferior (PCI)** do gás.

1.2 Objetivos específicos

Para cumprimento do objetivo geral é necessário atender aos seguintes objetivos específicos:

- Identificar os pontos críticos desse processo através da ciência de dados;
- Correlacionar as variáveis de temperatura das zonas com resultado final de temperatura de tira;
- Estabelecer a principal variável que influencia na temperatura de tira.
- Avaliar o desempenho de diferentes modelos de Aprendizagem de Máquina ao problema;
- Interpretar os resultados obtidos e sua aplicabilidade no processo produtivo.

1.3 Contribuições

Este trabalho demonstra a importância da aplicação da ciência de dados na indústria siderúrgica e na otimização do controle de processos. Através do estudo de dados, é possível buscar formas de tornar o processo mais efetivo e com maior repetibilidade. As contribuições deste estudo incluem não apenas o desenvolvimento de um modelo de previsão de temperatura das bobinas de aço inoxidável, mas também a análise detalhada das variáveis envolvidas no processo de recozimento.

Isso proporciona *insights* valiosos para a compreensão e otimização dos processos industriais, contribuindo assim para o avanço contínuo da indústria siderúrgica e das técnicas de análise de dados aplicadas a ela.

2 Fundamentação Teórica

2.1 Aços inoxidáveis austeníticos

Os **Aços Inoxidáveis Austeníticos (AIAs)** foram concebidos para suportar ambientes com níveis médios e severos de corrosão e oxidação. Eles exibem excelentes características de conformabilidade, soldabilidade e tenacidade, embora apresentem um limite de escoamento relativamente baixo. Apesar de possuírem valores modestos de limite de escoamento a temperatura ambiente no estado recozido, os **AIAs** demonstram ductilidade e tenacidade notáveis, o que simplifica o processo de conformação a frio (**PULGARIN, 2006**).

Esses aços austeníticos são monofásicos, possuindo uma estrutura cristalina cúbica de faces centradas que permanece estável em uma ampla faixa de temperaturas. A estabilidade da fase austenítica é alcançada pela incorporação de elementos de liga. Os sistemas de liga mais comuns para os **AIAs** são Fe-Cr-Ni e Fe-Cr-Ni-Mo. Essas ligas contêm aproximadamente 70% de ferro, com teores de cromo variando entre 15% e 26%, níquel entre 5% e 37%, manganês até 15% e nitrogênio. O cromo aprimora a resistência à corrosão, o níquel estabiliza a austenita e, em algumas situações, o nitrogênio substitui o níquel como estabilizador da austenita. Em alguns tipos de aços, pode-se encontrar quantidades de molibdênio entre 2% e 4%, que são adicionadas para melhorar a resistência à corrosão por pites e promover o endurecimento por solução sólida (**PADILHA; RIOS, 2002**) e (**BROOKS; LIPPOLD, 1993**).

As propriedades mecânicas dos **AIAs** são altamente influenciadas pela composição química e pelos tratamentos termomecânicos. Uma vasta gama de combinações de propriedades pode ser alcançada por meio de diferentes composições dentro de um grupo, tornando esses materiais adequados para diversas aplicações (**BROOKS; THOMPSON, 1991**). A **Tabela 1** apresenta as principais características mecânicas de alguns **AIAs** em diferentes estados.

No estado recozido, os **AIAs** exibem limites de escoamento na faixa de 200 a 250 N/mm² e limites de resistência variando de 450 a 750 N/mm², com alongamentos entre 35% e 45%. Em comparação com outros tipos de aços inoxidáveis, os **AIAs** não apresentam propriedades mecânicas particularmente atrativas à temperatura ambiente. No entanto, suas propriedades mecânicas se destacam em altas temperaturas. Por exemplo, a resistência à fluência a 550°C por 100.000 horas pode exceder 100 N/mm², e sua resistência à corrosão é notável (**PADILHA; RIOS, 2002**).

Elementos conhecidos como estabilizadores, como o titânio (AISI 321) e o nióbio (AISI 347), são adicionados para fixar o carbono e dificultar a formação de carbonetos de cromo e a subsequente corrosão intergranular. Uma abordagem adicional para reduzir a susceptibilidade à corrosão intergranular é a diminuição do teor de carbono para menos de 0,03%, resultando nos aços do tipo “L”, como 304L e 316L.

Os **AIAs** exibem elevada resistência à corrosão e à oxidação, bem como tenacidade elevada e excelente ductilidade em uma faixa de temperaturas que vai desde as condições criogênicas até temperaturas elevadas. Essas propriedades excepcionais permitem a aplicação dos **AIAs** em diversos setores, incluindo aplicações nucleares e uma ampla variedade de aplicações não nucleares, tais como indústria química, petróleo, turbinas a gás, entre outros, conforme mostrado na **Tabela 1**.

Tabela 1 – Composições químicas nominais (% em peso) e propriedades mecânicas de algumas classes de aços inoxidáveis austeníticos.

Clase	Composição nominal (%)	Estado	Limite de Escoamento (MPa)	Limite de Resistência (MPa)	Alongamento em 50 mm (%)	Energia Absorvida (J)
AISI 201	16-18Cr 3,5-5,5Ni 0,15C	R	380	760	50	135
	1Si 5,5-7,5Mn 0,06P 0,03S 0,25N	TTM	Até 965	Até 1275	50-8	
AISI 301	16-18Cr 6-8Ni 0,15C	R	275	760	60	135
	1Si 2Mn 0,04P 0,03S	TTM	Até 965	Até 1275	60-8	-
AISI 304	18-20Cr 8-12Ni 0,08C	R	240	585	55	135
	1Si 2Mn 0,04P 0,03S	TTM	Até 1035	Até 1240	55-10	-
AISI 316	16-18Cr 10-14Ni 0,8C	R	240	585	55	120
	1Si 2Mn 0,04P 0,03S 2-3MoN	TTM	Até 860	Até 1035	55-10	-

2.2 O estado encruado

O estado encruado de um aço inoxidável austenítico laminado a quente é uma condição resultante do processo de deformação plástica a alta temperatura, no qual o material é submetido a um trabalho de conformação mecânica intenso, como a laminação. Durante esse processo, ocorrem mudanças significativas na estrutura e nas propriedades do material.

A deformação plástica ocorre quando o material é submetido a tensões que excedem o limite elástico, levando a uma reorganização das posições dos átomos e à formação de discordâncias ou defeitos cristalinos na estrutura cristalina. No caso dos aços inoxidáveis austeníticos, que possuem uma estrutura cristalina **Cúbica de Faces Centradas (CFC)**, a deformação plástica resulta na movimentação de planos de escorregamento ao longo dos planos cristalinos (PADILHA; SICILIANO, 2005).

Durante a laminação a quente, o material é aquecido a temperaturas elevadas, o que reduz a resistência do aço à deformação plástica e permite uma conformação mais eficiente. No entanto, à medida que o material é deformado, ocorre um acúmulo progressivo de discordâncias e outras imperfeições cristalinas, o que leva ao endurecimento do material. Esse endurecimento é conhecido como encruamento. O encruamento resulta em mudanças nas propriedades mecânicas do aço inoxidável austenítico. Por exemplo, o limite de escoamento e a resistência à tração aumentam, tornando o material mais resistente, porém menos dúctil. A tenacidade também pode ser afetada, e a capacidade de conformação a frio pode ser comprometida devido ao aumento da rigidez do material (PULGARÍN, 2006).

Para reverter parcialmente o encruamento e melhorar a conformabilidade, é possível realizar tratamentos térmicos de recozimento. O recozimento envolve o aquecimento do material a temperaturas elevadas, seguido de resfriamento controlado, o que permite que as discordâncias se difundam e diminuam em densidade, restaurando parte das propriedades originais do material (PADILHA; SICILIANO, 2005).

2.3 Recuperação e recristalização

O processo de recuperação e recristalização em um aço austenítico durante o processo de recozimento é uma sequência complexa de eventos microestruturais intrinsecamente associados à deformação plástica prévia. Durante a etapa inicial de recuperação, que ocorre a temperaturas moderadas, as discordâncias cristalinas previamente introduzidas começam a se movimentar e rearranjar em resposta a gradientes de energia elástica e microestrutural. Esse fenômeno provoca uma redução nas tensões internas, conduzindo a uma parcial restauração das características mecânicas e da tenacidade original do material, enquanto a maior parte das deformações plásticas permanece (HUMPHREYS, 2004).

Na etapa subsequente de recristalização, que se desenrola em temperaturas mais elevadas, um processo de nucleação e crescimento de novos grãos cristalinos emerge a partir das áreas mais intensamente deformadas. Esses núcleos cristalinos recém-formados adquirem orientações cristalográficas distintas das regiões circundantes, o que culmina na criação de uma nova estrutura cristalina. Esses grãos recristalizados tendem a possuir dimensões menores e contornos de grão mais bem definidos em comparação com os grãos deformados, o que contribui significativamente para a melhoria das propriedades mecânicas e da microestrutura do material (CHO; YOO; JONAS, 2000).

O processo de recozimento é conduzido mediante o aquecimento controlado do aço austenítico a uma temperatura ótima, capaz de estimular tanto a recuperação quanto a subsequente recristalização. O tempo de residência a essa temperatura crítica é deliberado com base na cinética de formação dos novos grãos recristalizados, que está intrinsecamente ligada aos mecanismos de difusão atômica.

Em conjunto, o resultado desse processo é a restauração das propriedades mecânicas intrínsecas e da estrutura cristalina original do material, conferindo-lhe maior ductilidade e menor endurecimento. Isso, por sua vez, concede ao material uma maior maleabilidade e capacidade de conformação, o que viabiliza a sua subsequente utilização em processos industriais e aplicações diversas (PULGARÍN, 2006).

Além disso, a compreensão detalhada dos mecanismos subjacentes à recuperação e recristalização em aços austeníticos durante o recozimento tem implicações significativas no desenvolvimento de processos de fabricação mais eficientes e na otimização das propriedades dos materiais. A seleção adequada das condições de tratamento térmico, incluindo a temperatura de recozimento e o tempo de permanência, desempenha um papel crucial na obtenção de uma microestrutura desejável e nas propriedades finais do aço austenítico (DUTTA; SELLARS, 1987).

A recuperação e recristalização também estão intimamente ligadas à cinética de difusão, que influencia a taxa de movimento das discordâncias e a formação de novos grãos recristalizados. A mobilidade das discordâncias é afetada pela presença de impurezas e elementos de liga, o que pode resultar em variações nas taxas de recuperação e recristalização em diferentes ligas de aço austenítico.

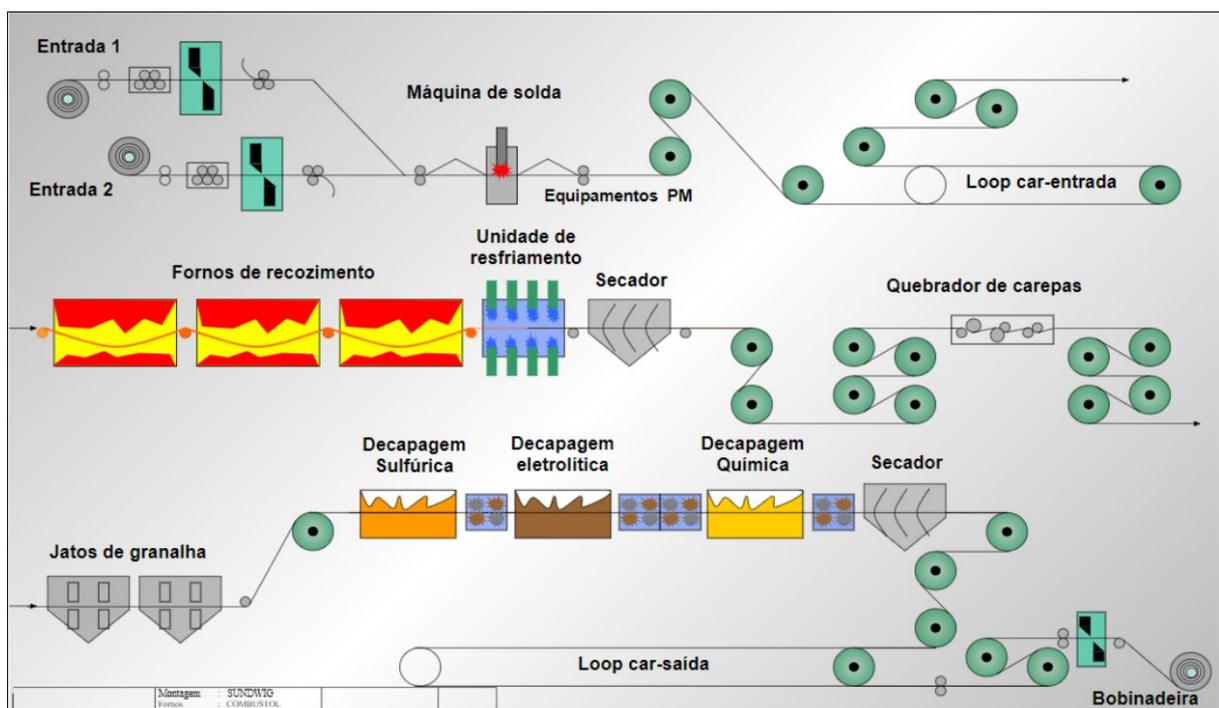
Esses avanços na compreensão dos processos microestruturais durante o recozimento têm contribuído para a melhoria contínua das propriedades e desempenho dos aços austeníticos em uma ampla gama de aplicações industriais e tecnológicas. A capacidade de controlar a recuperação e recristalização por meio de ajustes precisos nas condições de recozimento tem permitido a produção de materiais com características mecânicas otimizadas para requisitos específicos, como ductilidade, resistência e tenacidade.

A compreensão científica dos processos de recuperação e recristalização em aços austeníticos durante o recozimento, respaldada por pesquisas e estudos de renomados especialistas, desempenha um papel vital no aprimoramento das propriedades e na aplicabilidade desses materiais em diversas áreas da engenharia e da indústria. As contribuições da pesquisa nesse campo têm possibilitado avanços significativos na fabricação de produtos de alta qualidade e desempenho, impulsionando a inovação e o progresso tecnológico (PULGARIN, 2006).

2.4 Descrição do processo produtivo

A linha contínua **Recozimento e decapagem 3 (RB3)** é um sistema altamente integrado e eficiente projetado para realizar o processo de recozimento e decapagem de aços inoxidáveis austeníticos, seguindo um fluxo sequencial que envolve múltiplas etapas, conforme mostrado na Figura 1.

Figura 1 – Layout da linha RB3



Fonte: Aperam, 2023

Cada etapa desempenha um papel crucial na melhoria das propriedades mecânicas, da resistência à corrosão e da qualidade superficial do aço inoxidável, preparando-o para uma variedade de aplicações industriais.

- **Desbobinamento:** O processo começa com a desbobinamento da bobina de aço inoxidável laminada a quente. Isso envolve desenrolar a bobina para fornecer uma bobina contínua de material que será processada ao longo da linha.

- **Soldagem:** Se houver a necessidade de emendar duas bobinas, a soldagem é realizada de maneira precisa para garantir a continuidade do material ao longo do processo.
- **Forno de recozimento:** A bobina de aço inoxidável segue para um forno de recozimento, onde é aquecida a temperaturas controladas. Nesse ambiente controlado, as tensões internas acumuladas durante a laminação a quente são aliviadas, ocorrendo a recuperação e recristalização da estrutura cristalina. Isso melhora as propriedades mecânicas, como ductilidade e tenacidade.
- **Jateamento mecânico (jateamento):** Após o recozimento, a superfície do aço é submetida a um processo de jateamento mecânico. Isso envolve o uso de partículas abrasivas para remover quaisquer óxidos, escalas ou contaminantes superficiais, preparando a superfície para a etapa de decapagem subsequente.
- **Decapagem sulfúrica:** A bobina de aço passa por um banho de decapagem sulfúrica, onde é imersa em um ácido sulfúrico diluído. Esse processo remove eficazmente quaisquer óxidos e contaminantes da superfície, preparando-a para as etapas de decapagem subsequentes.
- **Decapagem eletrolítica:** Em seguida, a bobina de aço é submetida a uma decapagem eletrolítica, na qual uma corrente elétrica é aplicada para remover de maneira mais eficiente os óxidos e contaminantes restantes da superfície. Esse processo ajuda a alcançar um alto nível de limpeza e uniformidade na superfície.
- **Decapagem química:** A etapa de decapagem química envolve a imersão da bobina de aço em um banho de ácido, que reage com os óxidos e impurezas, resultando em uma superfície limpa e livre de contaminantes.
- **Lavagem final:** Após as etapas de decapagem, a bobina de aço é submetida a um processo de lavagem final para remover qualquer resíduo ácido e garantir a neutralização adequada.
- **Bobinamento final:** Por fim, a bobina de aço inoxidável é resfriada e bobinada em bobinas finais, prontas para serem utilizadas em diversas aplicações industriais que requerem propriedades aprimoradas, alta resistência à corrosão e qualidade superficial.

A linha contínua RB3 realiza todas essas etapas de maneira integrada e contínua, otimizando a qualidade e as propriedades do aço inoxidável austenítico, tornando-o apto para atender às demandas rigorosas de diversas indústrias, como a indústria química, petroleira, de alimentos, entre outras. Neste trabalho será abordado de forma integrada o processo de recozimento dos aços austeníticos.

2.5 Variação de temperatura durante o processo

Durante o procedimento de recozimento de aços inoxidáveis, a modulação da temperatura nas bobinas de aço desempenha um papel de vital importância na obtenção das características desejadas e no aprimoramento da microestrutura do material. O processo de recozimento é executado mediante a implementação de estágios meticulosamente regulados, englobando a fase de aquecimento, subsequente manutenção em temperaturas específicas e resfriamento cuidadosamente controlado. A seguir, é delineado o perfil típico de variação de temperatura experimentado por bobinas de aço durante o procedimento de recozimento de aços inoxidáveis:

- **Aquecimento:** O processo se inicia com um aquecimento gradual das bobinas de aço inoxidável, conduzindo-as até uma temperatura de recozimento previamente determinada. Este processo de aquecimento é executado de maneira cuidadosamente controlada e gradual, com o intuito de prevenir a ocorrência de tensões térmicas excessivas no material. A faixa de temperatura de aquecimento varia, em geral, de acordo com a natureza específica do aço inoxidável e as propriedades que se almeja alcançar, sendo comumente situada na ordem de 850°C a 1050°C.
- **Manutenção em temperatura:** Após a consecução da temperatura de recozimento, as bobinas de aço permanecem mantidas a esse patamar por um intervalo de tempo pre-determinado. Tal período de permanência à temperatura adequada é fundamental para possibilitar a ocorrência de transformações microestruturais, incluindo a recuperação e a recristalização. A extensão dessa fase temporal pode oscilar, variando desde minutos até algumas horas, em função das características próprias do aço e das propriedades almejadas.
- **Resfriamento controlado:** Subsequentemente à etapa de manutenção em temperatura, procede-se ao resfriamento das bobinas de aço mediante uma abordagem de controle rigoroso. A aplicação de um resfriamento controlado detém importância crucial com o fito de evitar a manifestação de tensões térmicas excessivas e minimizar o potencial de deformações indesejadas no material. Diversos métodos de resfriamento podem ser empregados, tais como resfriamento ao ar, em forno ou através do uso de água, sendo a escolha dependente das características específicas do aço inoxidável e das propriedades visadas.

Evidencia-se que a variação da temperatura e a consecução precisa das etapas de aquecimento, manutenção em temperatura e resfriamento exercem impacto direto nas propriedades finais do aço inoxidável austenítico. O objetivo é a obtenção de uma microestrutura mais refinada e homogênea, traduzindo-se em melhorias na ductilidade, tenacidade e resistência à corrosão do material.

A variação controlada da temperatura no transcurso do procedimento de recozimento reveste-se de um aspecto crítico, demandando, assim, uma meticulosa otimização, com vistas a garantir o alcance das propriedades almejadas. A compreensão das características intrínsecas do aço inoxidável, o domínio dos mecanismos de recuperação e recristalização, bem como a rigidez no monitoramento das temperaturas e dos intervalos de tratamento, são fundamentais para o sucesso do procedimento de recozimento em aços inoxidáveis austeníticos.

A omissão de um controle adequado na variação da temperatura de recozimento em aços inoxidáveis austeníticos pode resultar em implicações substanciais nas características derradeiras do material, colocando em risco a sua qualidade e desempenho. A inadequação na modulação da temperatura de recozimento pode culminar na formação de microestruturas indesejáveis, propriedades mecânicas insatisfatórias e uma susceptibilidade reduzida à corrosão. Algumas das possíveis consequências decorrentes de uma modulação inadequada da temperatura de recozimento abarcam:

- **Microestrutura não otimizada:** A temperatura de recozimento exerce influência direta na formação e no crescimento de grãos recristalizados. A falta de adequação na variação da temperatura pode culminar em uma microestrutura não otimizada, com grãos de dimensões maiores ou recristalização não uniforme, repercutindo negativamente na ductilidade, tenacidade e outras propriedades mecânicas.
- **Inferioridade das propriedades mecânicas:** A deficiência no controle da temperatura de recozimento pode redundar na diminuição das propriedades mecânicas desejáveis, tais como resistência, dureza e capacidade de alongamento. Tal cenário pode impactar a habilidade do aço inoxidável de suportar cargas e deformações sem ocorrência de falhas prematuras.
- **Redução na resistência à corrosão:** A temperatura de recozimento também desempenha um papel na resistência do aço inoxidável à corrosão. A falta de ajuste adequado da temperatura pode culminar em uma superfície mais suscetível à corrosão, comprometendo, por conseguinte, a durabilidade e vida útil do material quando exposto a ambientes corrosivos.
- **Instabilidade dimensional:** A discrepância na modulação da temperatura de recozimento pode originar tensões térmicas excessivas e deformações dimensionais indesejadas nas bobinas de aço. Esse panorama poderá manifestar-se através de distorções, empenamentos ou outras deformações, que impactam adversamente a integridade do produto final.
- **Inconsistência nas propriedades:** A falta de um controle preciso da temperatura de recozimento pode engendrar variações não intencionais nas propriedades do material, refletindo em produtos finais inconsistentes e que não estejam em conformidade com as especificações requeridas.

Com o propósito de evitar tais falhas, torna-se imperativo a consecução de um planejamento criterioso do procedimento de recozimento, levando em consideração as características intrínsecas do aço inoxidável, as propriedades almejadas e as circunstâncias operacionais. A adoção de curvas de recozimento, perfis de temperatura sob controle e uma vigilância rigorosa das temperaturas constituem práticas indispensáveis a fim de assegurar que o processo seja executado de maneira consistente e controlada, culminando na produção de produtos de alta qualidade e desempenho.

A perspicácia para compreender os mecanismos subjacentes à recuperação e recristalização, aliada ao zelo pela modulação precisa da temperatura, são pré-requisitos essenciais para evitar falhas associadas à variação da temperatura de recozimento em aços inoxidáveis austeníticos. O conhecimento profundo das propriedades do aço em questão, a ciência por trás das transformações microestruturais e a aplicação diligente de técnicas de controle térmico são fatores cruciais para o sucesso e a qualidade do processo de recozimento em aços inoxidáveis austeníticos.

Assegurar uma variação adequada da temperatura de recozimento não apenas otimiza as propriedades do aço inoxidável, mas também desempenha um papel vital na produção de materiais que possuam a capacidade de enfrentar os desafios mecânicos, químicos e ambientais impostos por uma diversidade de aplicações industriais e tecnológicas. O conhecimento e a aplicação rigorosa desses princípios científicos e práticos são fundamentais para o sucesso na obtenção de aços inoxidáveis austeníticos com desempenho superior e alta confiabilidade em uma ampla gama de contextos de engenharia e fabricação.

Na [Tabela 2](#) é apresentado os impactos no material devido à variação de temperatura no aço 304 A/I na RB3 em 2023. Pode-se perceber que, na média entre Janeiro e Agosto, 26% do material produzido não atendeu o seu cliente final. Esse material com variação é realocado, reprocessado ou sucitado.

Tabela 2 – Dados de Pendências por Mês

Mês	% Pendências	Peso Pendência (t)
Janeiro/2023	36,1%	1109,38
Fevereiro/2023	24,9%	1159,485
Março/2023	32,7%	1604,875
Abril/2023	19,5%	1164,01
Maiio/2023	27,8%	1548,005
Junho/2023	15,7%	1303,575
Julho/2023	28,0%	459,365
Agosto/2023	19,5%	1276,745

A quantidade de toneladas de aço 304A/I com variação de temperatura durante o período de janeiro a agosto de 2023 totalizou 9560 toneladas, as quais não foram atendidas conforme o fluxo normal de operações. Este valor representa um grande impacto para o negócio pois compromete o atendimento ao cliente, rendimento e o custo do processo.

2.6 Extração, tratamento e exploração dos dados

O conceito de dado se refere a uma representação elementar desprovida de significado intrínseco. Em contrapartida, a informação surge quando os dados são organizados de maneira a transmitir significado, e o conhecimento é adquirido por meio da análise das informações existentes. A presença da informação é crucial para a existência e funcionamento das organizações, permeando todos os processos, dos mais simples aos mais complexos.

No contexto organizacional, há uma demanda constante por dados e informações, que são processados e analisados de forma contínua. Isso destaca a importância não apenas de ter acesso a informações, mas também da capacidade de analisá-las de maneira eficaz para transformá-las em ferramentas que possam melhorar processos e resolver problemas. Portanto, compreender o processo pelo qual os dados se transformam em informações e, posteriormente, em conhecimento, torna-se essencial.

2.6.1 Extração dos dados

A extração de dados implica na identificação, organização e exportação criteriosa de dados para subsequente análise, visando obter informações relativas a assuntos ou problemas específicos. Este procedimento requer análise minuciosa do problema, visando minimizar ambiguidades e erros nos dados. Ademais, é imperativo que os dados coletados sejam representativos e abranjam amplamente o domínio do problema (BOGO *et al.*, 2019). Essa extração pode ser realizada a partir de fontes estruturadas, não estruturadas e/ou semiestruturadas, abrangendo diversos formatos, como arquivos de variados tipos, imagens, e-mails e dados de redes sociais.

As tecnologias modernas oferecem diversas opções para a extração de dados, incluindo o uso de bancos de dados relacionais, frequentemente empregados para armazenar dados heterogêneos. Um banco de dados relacional compreende tabelas interconectadas, cada uma relacionada a várias outras. Esse modelo simplifica a coleta e extração subsequente de dados (BOSCARIOLI *et al.*, 2006).

2.6.2 Tratamento dos dados

No processo de conversão de dados em informação e desta em conhecimento, a etapa crítica de tratamento e análise dos dados surge como destaque. Em muitas bases de dados, é comum a presença de dados faltantes ou inconsistentes, tornando essencial um tratamento prévio. Tal tratamento envolve a eliminação de dados incorretos ou discrepantes, a identificação e abordagem de dados ausentes, bem como a definição de estratégias para lidar com essas lacunas (BROWN; KROS, 2003)).

Existem várias técnicas bem estabelecidas para lidar com dados que podem estar sujeitos a ruído ou incompletude. Uma das abordagens mais comuns envolve a exclusão de observações problemáticas, no entanto, essa estratégia nem sempre é a mais adequada. Isso ocorre porque, ao remover observações problemáticas, corremos o risco de inadvertidamente eliminar informações valiosas.

Em tais situações, uma abordagem recomendada é a imputação de dados. Isso significa que os valores ausentes são substituídos por estimativas ou valores plausíveis. A imputação de dados desempenha um papel fundamental na preservação da integridade do conjunto de dados, permitindo que informações relevantes sejam mantidas e utilizadas de maneira eficaz.

Além disso, é importante mencionar que a escolha da técnica de imputação de dados apropriada pode depender da natureza dos dados, do contexto do problema e das suposições subjacentes. Portanto, é fundamental considerar cuidadosamente qual técnica de imputação é mais adequada para um determinado conjunto de dados e garantir que o processo seja realizado com rigor estatístico e transparência (VERONEZE, 2011).

Adicionalmente, é importante mencionar que a imputação de dados não se limita apenas à substituição de valores ausentes. Em muitos casos, envolve a utilização de métodos estatísticos avançados e técnicas de **Aprendizado de Máquina (AM)** para preencher lacunas nos dados de forma precisa e confiável.

2.6.3 Exploração analítica

Durante a fase de pré-processamento, a transformação dos dados ocorre em quase todos os estágios, e neste cenário, termos cruciais emergem, tais como categorização, padronização e normalização.

A categorização textual, mediante algoritmos, envolve a atribuição de categorias predefinidas a partir de uma base de referência rotulada. Já a normalização e padronização têm como objetivo ajustar os dados a uma escala comum. Variações significativas nas escalas numéricas das variáveis podem acarretar problemas na modelagem dos dados. A normalização e padronização minimizam esses problemas, garantindo a integridade dos dados originais e evitando distorções. Ambas são essenciais para muitos algoritmos de aprendizado de máquina, uma vez que podem prejudicar o desempenho se as variáveis não seguirem distribuição normal (COSTA; MARÇAL; PERUCCI, 2023).

A diferença entre normalização e padronização reside na escala de ajuste. Enquanto a normalização redimensiona os dados usando a fórmula *min-max*, ajustando os valores entre 0 e 1, a padronização é efetuada pela fórmula *z-score*, que centraliza a média em 0 e o desvio padrão em 1 (COSTA; MARÇAL; PERUCCI, 2023).

Não há uma regra rígida quanto à escolha da técnica apropriada, sendo a decisão guiada pelo tipo de dados e pelo problema em questão. Uma abordagem promissora é a combinação de ambas as técnicas, aplicando normalização/padronização e avaliando qual proporciona melhor desempenho ao modelo, ou até mesmo empregando ambas de forma complementar.

Além disso, ao explorar dados, é crucial considerar o tamanho da base de dados. A tecnologia atual possibilita a coleta e armazenamento massivo de dados, resultando em conjuntos de dados cada vez maiores. Porém, conjuntos maiores demandam maior poder de processamento computacional e tempo para análise.

Dessa forma, a seleção cuidadosa de atributos ganha relevância no pré-processamento, objetivando identificar os mais relevantes e otimizar o modelo. A seleção de atributos visa escolher as variáveis mais pertinentes para resolver o problema, sem comprometer o desempenho do modelo (DANTAS, 2017). Essa técnica também ajuda a remover dados ruidosos e redundantes, reduzindo a dimensão da base de dados e, por conseguinte, otimizando o poder computacional necessário (LEE, 2005).

2.7 Treinamento e validação dos modelos

Os estágios iniciais no desenvolvimento de modelos de AM, requerem a coleta de dados relacionados ao problema em questão e sua subsequente separação em conjuntos de treinamento e teste. Os dados de treinamento são empregados para ensinar a rede, enquanto os dados de teste avaliam o desempenho da rede sob condições reais (FURTADO, 2019). Para evitar tendências devido à ordem de apresentação, esses conjuntos são embaralhados.

Durante a etapa de treinamento, métricas de desempenho são aplicadas para comparação dos resultados, permitindo a avaliação eficaz da rede. O uso criterioso dos conjuntos de treinamento e teste e a análise aprofundada dos resultados são cruciais para o desenvolvimento bem-sucedido de modelos de AM.

2.7.1 Validação Cruzada (*Cross-Validation*)

Segundo Stone (1974 apud VIEIRA, 2006, p. 37), a validação cruzada tem como objetivo aumentar a capacidade de generalização em modelos de aprendizagem evitando o sobreajuste (*overfitting*) dos modelos e foi utilizado também para estimar o desempenho dos modelos. O método implementado neste trabalho foi o *k – fold Cross Validation - CV* que consiste em dividir o conjunto de dados original em k subconjuntos. Esses k subconjuntos são utilizados para treinar k redes, onde um subconjunto é usado para verificação de validação, e os outros $k - 1$ subconjuntos são usados para o treinamento. Este processo é repetido durante k vezes (*fold*) e para cada iteração, o conjunto de verificação e validação é selecionado apenas uma vez.

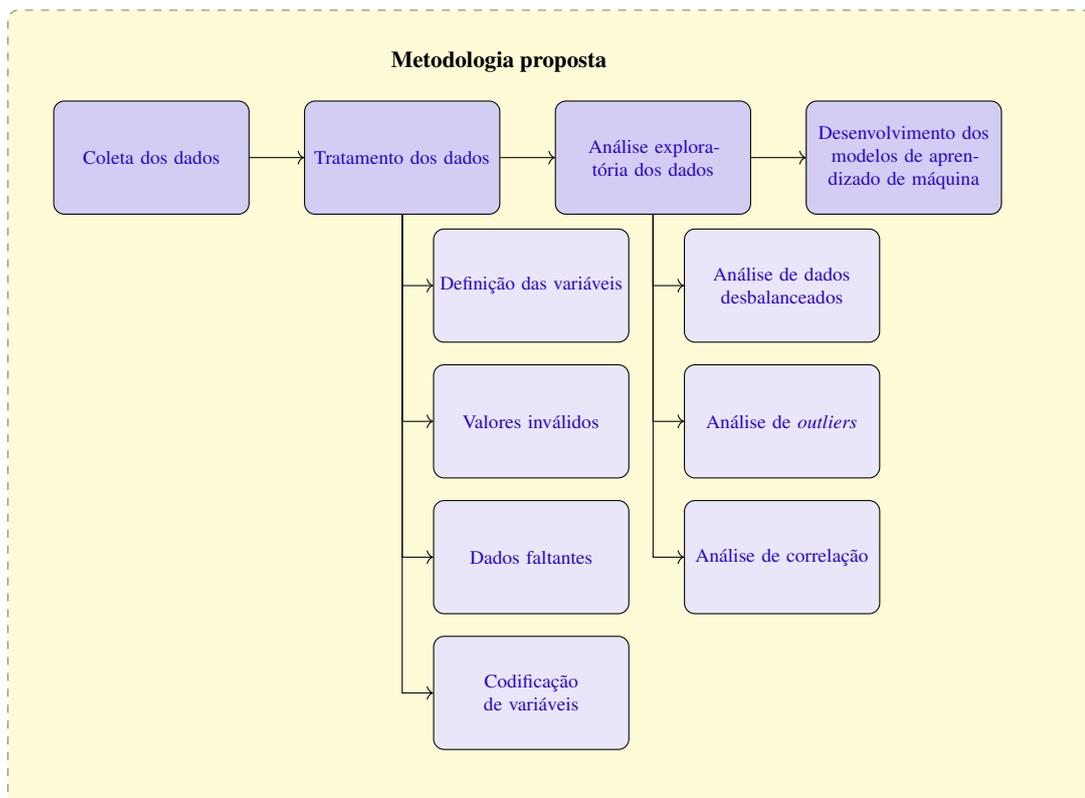
2.7.2 Parada antecipada (*Early stopping*)

Outra estratégia utilizada no treinamento dos modelos para evitar o sobreajuste, foi o de parada antecipada (*Early stopping - ES*), que também objetiva alcançar redes com boa capacidade de generalização (WEIGEND; HUBERMAN; RUMELHART, 1990). Essa é uma técnica baseada na divisão dos padrões de em pelo menos duas partes: a primeira chamada de conjunto de treinamento é utilizada para realizar os ajustes dos parâmetros e a segunda é utilizada para calcular o erro de resposta da rede (validação). Dessa forma, ao final de cada iteração (época) do algoritmo de treinamento, o erro de resposta da rede é calculado e o treinamento é interrompido no instante que antecede o crescimento do erro de resposta do modelo para dados de validação. Assim, os parâmetros de treinamento dos modelos obtidos, na época anterior a interrupção do treinamento, são utilizados como parâmetros finais de treinamento.

3 Metodologia

Este capítulo apresenta a metodologia utilizada no desenvolvimento dos modelos de aprendizado de máquina para previsão da temperatura de saída da linha RB3. A Figura 2 apresenta um fluxograma dos passos seguidos na metodologia.

Figura 2 – Fluxograma dos passos seguidos na metodologia.



Fonte: elaborado pelos autores, 2023.

3.1 Coleta dos dados

A origem dos dados utilizados no presente trabalho são provenientes de fontes internas coletados dentro da organização, gerados por sistemas de automação da planta. Por isso mesmo, o conjunto de dados precisa ter todas as variáveis necessárias para que o problema em questão seja resolvido.

Os dados referentes às variáveis de processo utilizados para construção dos modelos, definidas anteriormente, foram extraídos da plataforma [CRAVE Industry \(CRAVE\)](#) para a linha [RB3](#). Essa plataforma carrega os dados do [Sistema de gerenciamento de informações de processo \(PIMS\)](#) da [Aperam](#). As informações coletadas pela plataforma é realizado considerando um intervalo de tempo de 1 segundo entre cada registro de amostragem e automaticamente já faz um pré-processamento desses dados agrupando-os por bobina. Para cada variável são definidos os valores mínimos, máximos, médias e desvios padrões.

As amostras de dados foram extraídas de aproximadamente 8200 bobinas de um tipo específico de aço (P304A e P304I) e 48 variáveis no período de 01/02/2020 a 20/12/2023. Estes dados passaram por um pré-processamento que será detalhado a seguir.

3.2 Tratamento dos dados

A etapa de tratamento e limpeza dos dados inclui a tarefa de extrair e substituir valores, principalmente os relacionados a dados faltantes no conjunto de informações. Pode-se pensar neste processo como o responsável por organizar os dados, remover o que não é mais necessário, substituir o que está faltando e padronizar o formato em todas as informações coletadas.

3.2.1 Definição das variáveis

Segundo [Ross \(2014\)](#), o pré-processamento eficaz dos dados pode ajudar a melhorar a qualidade dos dados, lidar com valores ausentes ou discrepantes e preparar os dados para análise estatística mais avançada, como a análise de correlação. Assim, para determinação do modelo de previsão de propriedades mecânicas foram definidos todos os fatores que poderiam influenciar no controle de temperatura do material e por consequência nas propriedades mecânicas do material. A escolha destes fatores foram baseadas na experiência dos técnicos e engenheiros responsáveis pelo controle de processo do equipamento e na literatura ([TAKAHASHI; RABELO; TEIXEIRA, 2008](#); [GORNI; SILVA; SILVEIRA, 2009](#)). Desta forma, após algumas reuniões e discussões, as variáveis definidas foram:

- Temperaturas de zonas (1 a 8) dos fornos (Temperatura Mínima, Temperatura Máxima e Desvio padrão)
- Espessura final do material.
- Largura final do material.
- [PCI](#) do gás misto.
- Temperatura final do material.

3.2.2 Valores inválidos

Para garantir que os valores dos dados sejam válidos na base de dados, foram desenvolvidos alguns métodos para verificar a validade do valor com relação ao seu tipo (inteiro ou ponto flutuante) e aplicado a todas as amostras. Caso o valor seja inválido para uma determinada amostra, os métodos desenvolvidos sobrescrevem os valores errados com o valor nulo.

Outra verificação de valores inválidos realizada é referente às bobinas processadas. Bobinas onde a identificação possui letra “T” na sétima posição da identificação ou que contenha “BOB” são materiais colocados na linha de produção para transição de material ou para parada de manutenção e dessa forma foram removidas da base de dados.

3.2.3 Dados faltantes

A qualidade de um conjunto de dados pode estar diretamente relacionada à quantidade de valores ausentes. É importante entender logo no início se esses valores ausentes são significativos comparados ao total de observações. Na [Tabela 3](#) são apresentados o percentual de dados ausentes para as 10 variáveis com o maior percentual de dados ausentes.

Tabela 3 – Percentual de dados faltantes para cada variável.

Variável	Percentual ausente
TEMP_Z1_STD	2,2%
TEMP_P2_STD	2,1%
VELOCIDADE_STD	2,1%
TEMP_Z2_STD	2,1%
PCI_GM_STD	2,1%
TEMP_Z3_STD	1,9%
TEMP_Z5_STD	1,9%
TEMP_Z7_STD	1,9%
TEMP_Z4_STD	1,9%
TEMP_Z8_STD	1,8%

Como o percentual (2%) de dados faltantes é relativamente pequeno em consideração ao número total de amostras, optou-se por remover linhas onde existem dados faltantes, uma vez que não é possível definir uma regra para preenchimento desses dados.

3.2.4 Codificação de variáveis

A maioria dos algoritmos de aprendizado de máquinas, ou do inglês *Machine Learning* (ML) trabalham apenas com dados numéricos, porém os problemas do mundo real muitas vezes trarão variáveis categóricas (ex: sim/não, alto/médio/baixo, ruim/regular/bom entre outros exemplos).

Para que o material tenha as propriedades mecânicas adequadas ele deve ser processado seguindo uma receita padrão, os parâmetros de processo para aço P304 de largura menor que 1400 mm são apresentados na [Tabela 4](#) e para as larguras maiores ou iguais a 1400 mm na [Tabela 5](#). Para as zona 1 e 2 não são definidos valores de temperatura na receita, uma vez que essas zonas operam desligadas e a temperatura medida na zona é consequência do fluxo de gases provenientes das demais zonas proveniente da queima de combustível.

Tabela 4 – Parâmetros de processo para o aço P304 com largura < 1400mm.

ESPESSURA	VELOCIDADE	PCI_GM	TEMP_Z1	TEMP_Z2	TEMP_Z3	TEMP_Z4	TEMP_Z5	TEMP_Z6	TEMP_Z7	TEMP_Z8	TEMP_P2
2,20 a 2,95	41	3200	-	-	1120	1120	1120	1115	1115	1100	1090
2,95 a 3,25	41	3200	-	-	1120	1120	1120	1120	1120	1110	1090
3,25 a 3,75	38	3200	-	-	1130	1130	1130	1130	1130	1120	1090
3,75 a 4,25	34	3200	-	-	1150	1150	1150	1150	1150	1140	1090
4,25 a 4,75	30	3200	-	-	1150	1150	1150	1150	1150	1140	1090
4,75 a 5,25	27	3200	-	-	1150	1150	1150	1150	1150	1140	1090
5,25 a 5,75	25	3200	-	-	1150	1150	1150	1150	1150	1140	1090
5,75 a 6,25	23	3200	-	-	1150	1150	1150	1150	1150	1140	1090
6,25 a 7,25	18	3200	-	-	1140	1140	1140	1140	1140	1130	1090
7,25 a 7,75	17	3200	-	-	1140	1140	1140	1140	1140	1130	1090
7,75 a 8,00	16	3200	-	-	1140	1140	1140	1140	1140	1130	1090

Tabela 5 – Parâmetros de processo para o aço P304 com largura \geq 1400mm.

ESPESSURA	VELOCIDADE	PCI_GM	TEMP_Z1	TEMP_Z2	TEMP_Z3	TEMP_Z4	TEMP_Z5	TEMP_Z6	TEMP_Z7	TEMP_Z8	TEMP_P2
2,20 a 3,25	41	3200	-	-	1140	1140	1140	1140	1140	1130	1090
3,25 a 3,75	41	3200	-	-	1140	1140	1140	1140	1140	1130	1090
3,75 a 4,25	33	3200	-	-	1140	1140	1140	1140	1140	1130	1090
4,25 a 4,75	25	3200	-	-	1140	1140	1140	1140	1140	1130	1090
4,75 a 5,25	24	3200	-	-	1140	1140	1140	1140	1140	1130	1090
5,25 a 5,75	22	3200	-	-	1140	1140	1140	1140	1140	1130	1090
5,75 a 6,25	21	3200	-	-	1140	1140	1140	1140	1140	1130	1090
6,25 a 7,25	18	3200	-	-	1140	1140	1140	1140	1140	1130	1090
7,25 a 7,75	17	3200	-	-	1140	1140	1140	1140	1140	1130	1090
7,75 a 8,00	16	3200	-	-	1140	1140	1140	1140	1140	1130	1090

Para o processo ser considerado conforme a oscilação máxima das variáveis de processo devem ser obedecidas conforme são apresentados na [Tabela 6](#) para para aço P304 de largura menor que 1400 mm e para as larguras maiores que 1400 mm na [Tabela 7](#).

Tabela 6 – Parâmetros de tolerância operacional o aço P304 < 1400mm.

Parâmetros	Valor Objetivado	Tolerância Operacional
Temperatura de Pirômetro 2 (Zona 8)	1090 \pm 15	1090 \pm 25
Temperatura de Zona	Tabela \pm 20	Tabela \pm 30
Velocidade (m/min) \leq 3,24	Tabela -3/+1	Tabela -5/+1
Velocidade (m/min) $>$ 3,24	Tabela \pm 2	Tabela \pm 3
PCI da Mistura	3200 \pm 300	3200 \pm 400

Tabela 7 – Parâmetros de tolerância operacional o aço P304 $\geq 1400mm$.

Parâmetros	Valor Objetivado	Tolerância Operacional
Temperatura de Pirômetro 2 (Zona 8)	1090 \pm 15	1090 \pm 25
Temperatura de Zona	Tabela \pm 20	Tabela \pm 30
Velocidade (m/min) $\leq 3,74$	Tabela -3/+1	Tabela -5/+1
Velocidade (m/min) $> 3,74$	Tabela \pm 2	Tabela \pm 3
PCI da Mistura	3200 \pm 300	3200 \pm 400

Os dados nas tabelas 4 a 7 foram utilizados para criar uma nova coluna que representará quando os valores da variável de processo seguiram ou não a receita padrão (TEMP_P2_IN_NTA). Em seguida todos os registros onde as variáveis de processo operaram dentro da receita padrão e atingiram o objetivo de temperatura foram removidos do conjunto de dados, uma vez que essas amostras estando dentro da faixa da receita padrão não há interferência do operador. No estudo serão consideradas apenas as amostras em que uma ou mais variáveis de processo não se mantiveram as variações dentro do processo conforme para identificar qual a configuração do equipamento mesmo estão fora da faixa de operação da receita padrão ainda assim atinge os objetivos de temperatura estabelecidos.

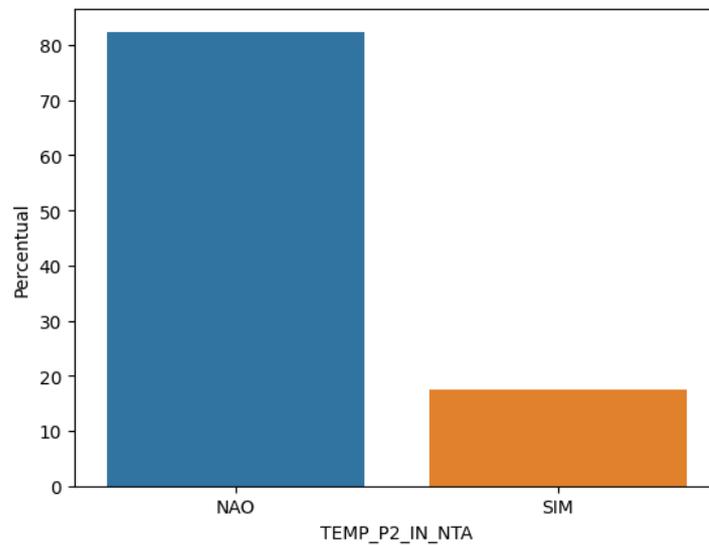
3.3 Análise exploratória dos dados

A análise exploratória de dados verifica conjuntos de dados para resumir suas principais características, geralmente com métodos visual quantitativo. O objetivo é identificar as possíveis fraquezas no processo de aquisição de dados, que suposições eles devem fazer e que modelos podem aplicar para produzir resultados de análise.

3.3.1 Análise de dados desbalanceados

Dados desbalanceados referem-se a uma situação, principalmente no aprendizado de máquina de classificação, em que uma das classes na variável alvo representa uma proporção significativa de observações. Conjuntos de dados desequilibrados são aqueles em que há uma distorção grave na distribuição de classes, como exemplos de 1:100 ou 1:1000 na classe minoritária para a classe majoritária. A Figura 3 representa o percentual de distribuição para cada classe na variável alvo, onde cerca de 82% das instâncias consistem no valor “SIM” e 18% das instâncias consistem no valor “NAO”, evidenciando que há um desbalanceamento entre duas classes.

Figura 3 – Percentual de distribuição para cada classe na base de dados.



Fonte: elaborado pelos autores, 2023.

Portanto, deverá ser verificado se os dados desbalanceados afetarão o desempenho do modelo de aprendizado de máquina, utilizar métricas de avaliação de desempenho mais apropriadas e avaliar a necessidade de utilização de sobreamostragem da classe minoritária ou subamostragem da classe majoritária. Os métodos de sobreamostragem duplicam ou criam novos exemplos sintéticos na classe minoritária, enquanto os métodos de subamostragem excluem ou mesclam exemplos na classe majoritária.

3.3.2 Análise de *outliers*

Em estatística, *outlier*, ou valor atípico, é uma observação que apresenta um grande afastamento das demais da série (que está “fora” dela), ou que é inconsistente. A existência de *outliers* implica, tipicamente, em prejuízos a interpretação dos resultados dos testes estatísticos aplicados às amostras.

Para confirmar, há duas maneiras rápidas que auxiliam a detecção de *outliers*. São elas: resumo estatístico conforme apresentado na Tabela 8 e visualização gráfica através de *boxplots* para cada variável conforme apresentado na Figura 4.

Como pode ser observado na Tabela 8 e na Figura 4, diversas variáveis possuem valores atípicos. Pode-se perceber que as variáveis com maiores variações são os desvios padrões das temperaturas das zonas, PCI e a velocidade do processo. Esses são os parâmetros de controle operacional direto por isso trazem maior variabilidade e podem influenciar na variável resposta. Para remoção dos *outliers*, foi utilizado o método *Z-Score*. Em estatística, o *Z-Score* é usado para identificar valores extremos ou discrepantes em um conjunto de dados. Valores com *Z-Score* acima de um determinado limite (valor de 3 foi aplicado como limite) são considerados *outliers* e foram excluídos da análise estatística.

Tabela 8 – Estatística resumida de variáveis numéricas.

Variável	Média	Padrão	Mínimo	25%	50%	75%	Máximo
ESPESSURA	4,05	1,33	2,60	2,90	4,00	5,00	8,00
LARGURA	1287,75	123,46	1040,00	1240,00	1240,00	1270,00	1556
VELOCIDADE_AVG	31,40	9,35	1,60	23,60	33,00	40,60	43,10
VELOCIDADE_STD	1,83	2,93	0,00	0,10	0,70	2,10	19,20
PCI_GM_AVG	3053,89	206,61	1052,40	3001,53	3117,00	3143,90	3540,60
PCI_GM_STD	107,91	183,33	0,60	26,50	47,80	101,88	3232,60
TEMP_Z1_AVG	926,40	59,49	364,40	893,50	922,20	958,10	1124,60
TEMP_Z1_STD	19,72	23,67	0,00	5,60	11,00	26,10	268,40
TEMP_Z2_AVG	976,80	51,42	372,70	953,80	978,70	1004,40	1106,80
TEMP_Z2_STD	19,75	24,77	0,00	5,50	11,20	25,80	295,30
TEMP_Z3_AVG	1129,54	38,80	438,20	1122,70	1135,10	1145,80	1179,20
TEMP_Z3_STD	13,64	24,37	0,00	3,80	7,30	14,50	339,00
TEMP_Z4_AVG	1131,82	37,86	437,80	1126,10	1136,40	1146,30	1179,80
TEMP_Z4_STD	12,63	24,73	0,00	3,00	6,60	13,40	338,60
TEMP_Z5_AVG	1127,39	42,13	332,30	1121,40	1133,10	1143,30	1180,00
TEMP_Z5_STD	14,66	26,87	0,00	3,80	7,80	15,88	356,70
TEMP_Z6_AVG	1130,23	36,08	381,50	1123,80	1134,40	1144,88	1208,40
TEMP_Z6_STD	12,57	22,15	0,00	3,60	7,40	14,40	304,00
TEMP_Z7_AVG	1128,78	38,55	370,00	1122,50	1133,55	1144,00	1174,70
TEMP_Z7_STD	13,17	24,15	0,00	3,80	7,90	15,00	518,70
TEMP_Z8_AVG	1121,00	40,73	358,10	1114,70	1125,80	1136,30	1185,30
TEMP_Z8_STD	13,19	24,90	0,00	3,40	7,70	15,00	334,80
TEMP_P2_AVG	1086,96	37,12	398,00	1082,10	1090,40	1098,30	1233,70

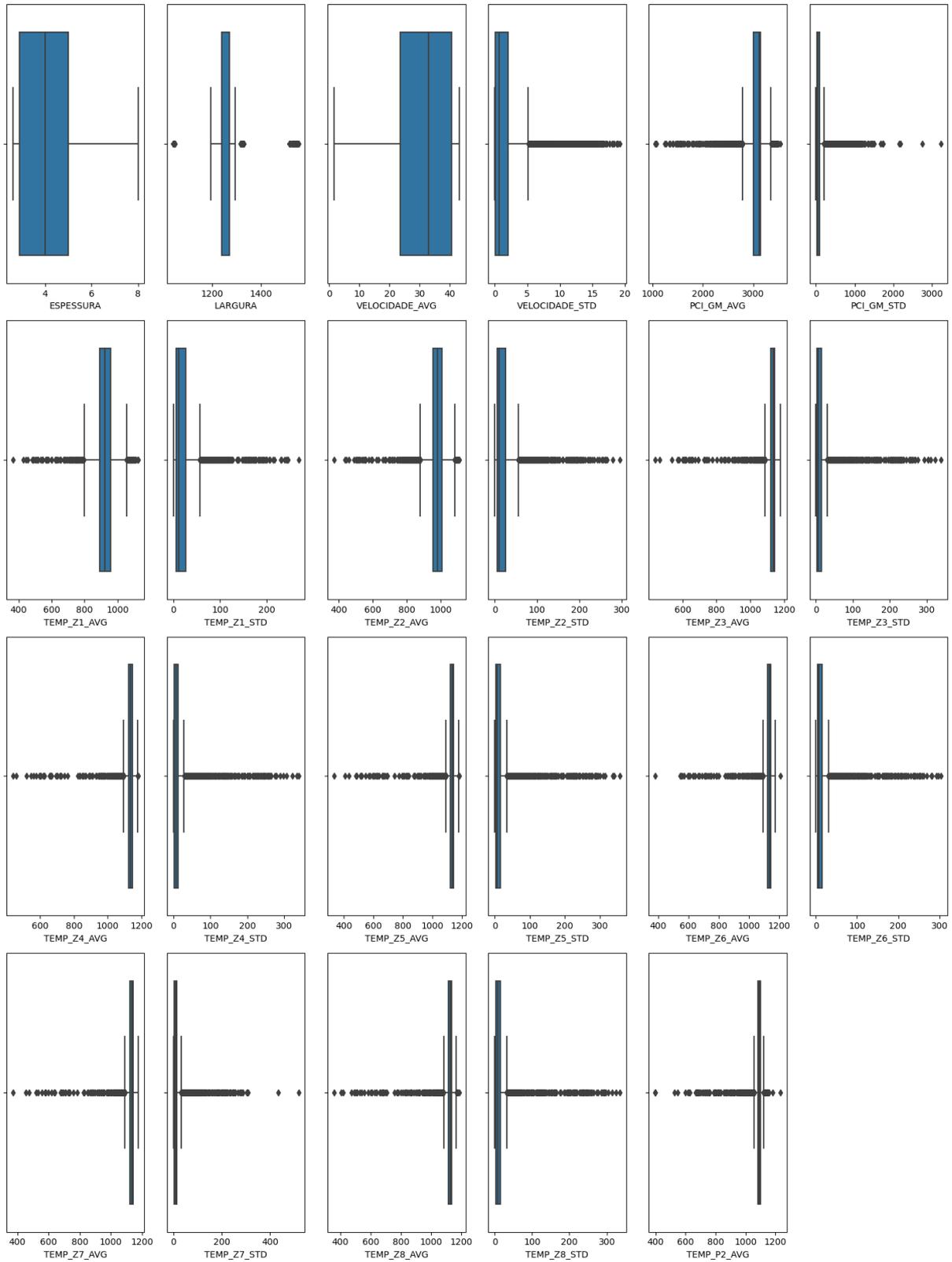
3.3.3 Análise de correlação

O cálculo do coeficiente de correlação linear entre as variáveis é utilizado para definir o grau de correlação e a direção da correlação (negativa ou positiva) entre duas variáveis. O coeficiente de correlação pode variar em termos de valor de -1 a $+1$. Quanto maior for o valor absoluto do coeficiente, mais forte é a relação entre as variáveis. Para a correlação de *Pearson*, um valor de correlação igual a 1 indica uma relação linear perfeita. A correlação perto de 0 indica que não há relação linear entre as variáveis. A correlação igual a -1 indica uma correlação negativa perfeita entre duas variáveis. A correlação de *Pearson* que mede a relação linear entre duas variáveis, pode ser representada da seguinte forma:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Onde:

Figura 4 – *Boxplots* para cada variável numérica.



Fonte: elaborado pelos autores, 2023.

- r_{xy} é a correlação de Pearson entre x e y .

- n é o número de observações.
- x_i e y_i são as observações individuais das variáveis x e y , respectivamente.
- \bar{x} e \bar{y} são as médias de x e y , respectivamente.

Além disso, alguns intervalos comuns usados para interpretar o coeficiente de correlação r ¹ incluem:

- $|r| < 0.3$: Correlação fraca.
- $0.3 \leq |r| < 0.5$: Correlação moderada.
- $0.5 \leq |r| < 0.7$: Correlação substancial.
- $|r| \geq 0.7$: Correlação forte.

Na [Figura 4](#) é apresentado o de todas as variáveis estudadas, esse gráfico mostra a forma, a tendência central e a variabilidade dos dados. Pode-se perceber que as variáveis com maiores oscilações são a espessura e velocidade da linha. O desvio padrão da velocidade está bem deslocado para esquerda, o que indica que a maioria das oscilações são no sentido de redução desse parâmetro.

Foi realizada uma análise de correlação para verificação de multicolinearidade entre as variáveis. A correlação entre as variáveis numéricas do conjunto de dados pode ser visualizada na [Figura 5](#). Observa-se que existe alta correlação entre as variáveis referente a temperatura das zonas dos fornos. Essa correlação faz sentido quando se observa o processo, uma vez que a configuração de temperatura de cada zona do forno é feita de forma a aumentar gradativamente a temperatura do material.

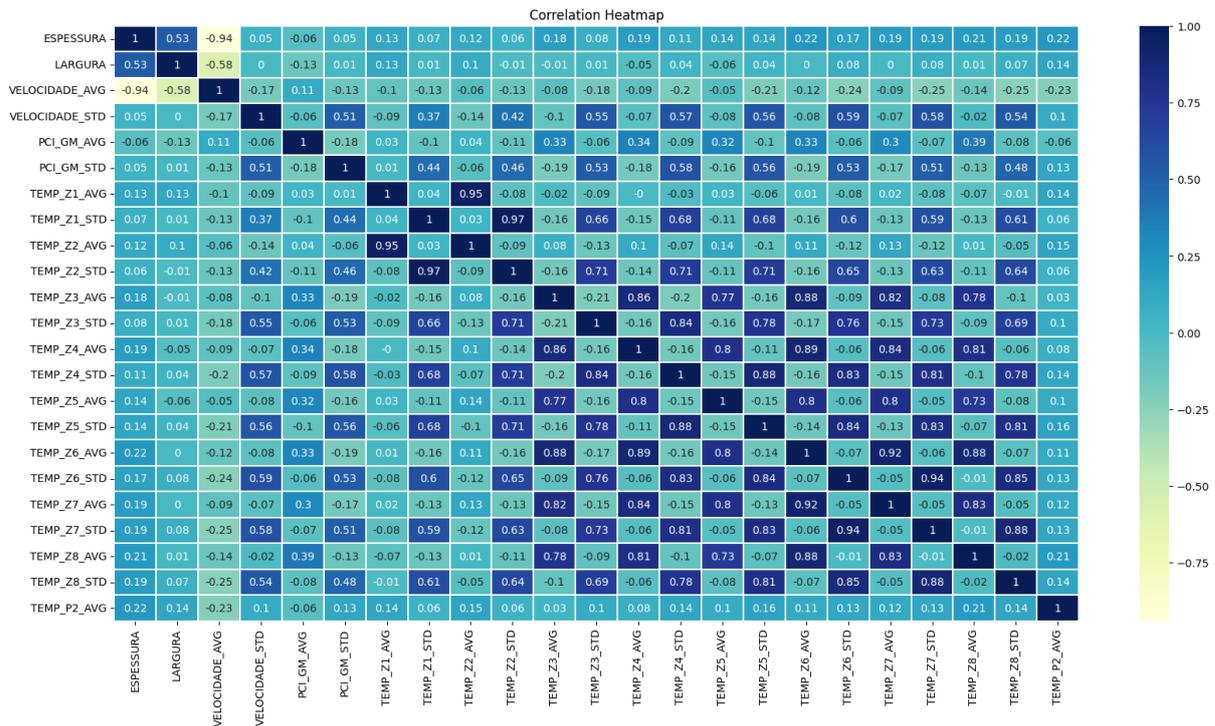
Alta correlação entre variáveis fornecem informações redundantes e a remoção de um dos fatores correlacionados geralmente não resulta em perda de informação, uma vez que a variáveis correlacionada está redundante com outras variáveis. Dessa forma, como existe alta correlação entre as variáveis de valor mínimo, máximo, média e desvio padrão para cada variável, foi definida a remoção das variáveis com valores mínimos e máximo ficando apenas na base de dados as variáveis referente a valores médios e desvio padrão.

3.4 Desenvolvimento dos modelos de aprendizado de máquina

Após as etapas de definição, coleta, tratamento dos dados, partiu-se para o desenvolvimento dos modelos de [ML](#). Os modelos têm o objetivo de classificar se a temperatura final do material está dentro ou fora dos limites.

¹ ([MOORE; MCCABE; CRAIG, 2014](#))

Figura 5 – Correlação entre as variáveis numéricas



Fonte: elaborado pelos autores, 2023.

Para o presente trabalho foram comparados os seguintes modelos:

- **DT** - *Decision Tree*
- **RF** - *Randon Forest*
- **XGBOOST** - *Extreme Gradiente Boosting*
- **KNN** - *K-Nearest Neighbors*

O *Decision Tree* (DT) ou árvore de decisão é um modelo de aprendizado de máquina que toma decisões com base em regras simples aprendidas a partir dos dados. Ela divide o conjunto de dados em subconjuntos menores com base em características específicas, criando uma estrutura semelhante a uma árvore. Cada nó da árvore representa uma característica do conjunto de dados e cada ramo representa uma decisão baseada nessa característica. As árvores de decisão são simples de entender e interpretar, tornando-as indicadas para análise de problemas de classificação e regressão.

O *Random Forest* (RF) é um algoritmo de aprendizado de máquina que utiliza uma coleção de árvores de decisão. Cada árvore de decisão em uma floresta é treinada em uma amostra aleatória do conjunto de dados, e a previsão final é determinada pela média ou pela moda das previsões individuais das árvores. O *Random Forest* é conhecido por sua capacidade de lidar com sobre ajuste e por sua robustez em relação a dados ruidosos, tornando-o uma escolha popular em muitas aplicações de aprendizado de máquina.

O XGBOOST, que significa *Extreme Gradient Boosting*, é uma implementação otimizada de um algoritmo de aprendizado de máquina conhecido como *Gradient Boosting*. Ele constrói uma sequência de modelos de árvores de decisão que tentam corrigir os erros dos modelos anteriores. O XGBoost é amplamente utilizado em competições de ciência de dados devido à sua eficácia e rapidez. Ele é especialmente adequado para conjuntos de dados grandes e complexos.

O *K-Nearest Neighbors* (KNN) é um método de aprendizado supervisionado usado para classificação e regressão. Ele funciona encontrando os *k exemplos* mais próximos no conjunto de dados de treinamento para um ponto de consulta e tomando a média ou o voto da classe desses exemplos para prever a classe do ponto de consulta. O valor de *k*, o número de vizinhos considerados, é um hiper-parâmetro crucial que afeta o desempenho do algoritmo.

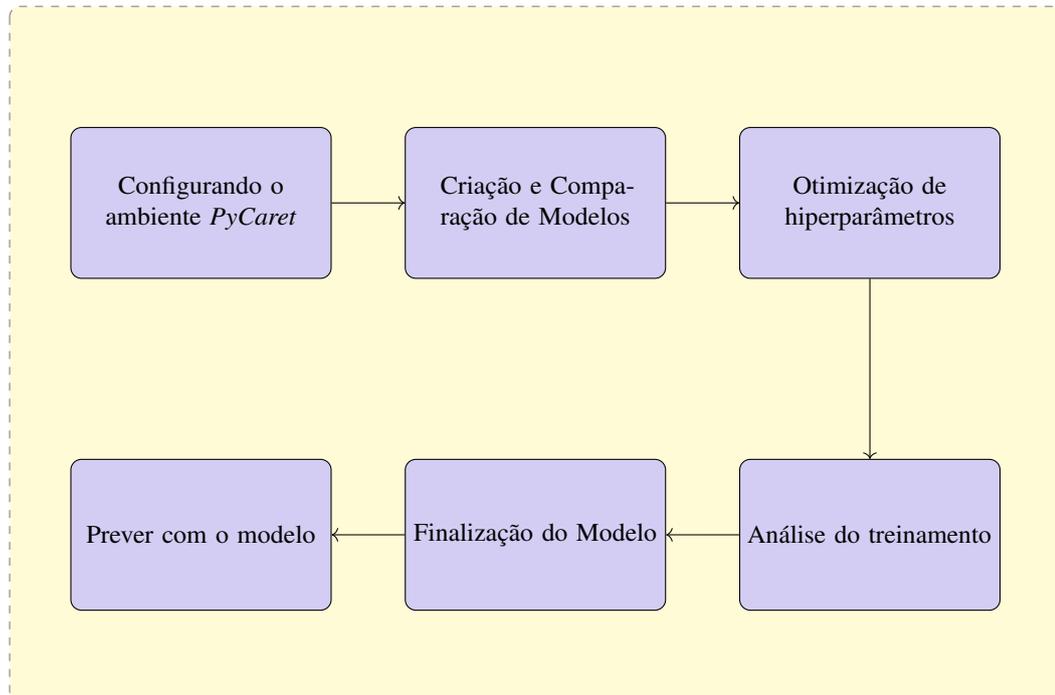
Para criação dos modelos de ML foi utilizada a biblioteca *PyCaret*². O *PyCaret* é uma biblioteca de aprendizado de máquina de código aberto que necessita de pouca programação e automatiza fluxos de trabalho de aprendizado de máquina. É uma ferramenta de aprendizado de máquina e gerenciamento de modelos de ponta a ponta que acelera exponencialmente o ciclo de experimentos e torna o trabalho de execução de experimentos mais produtivo (ALI, 2020).

O fluxo de trabalho de aprendizado de máquina para avaliação dos classificadores envolve as etapas conforme pode ser visualizado na Figura 6.

Inicialmente, uma amostra de 10% dos registros do conjunto de dados original foi retida para uso nas previsões. Para esta amostragem foi realizada uma divisão estratificada, mantendo as mesmas porcentagens de classes. Essa amostra não será utilizada no treinamento e validação dos modelos de classificação e tem como principal objetivo simular um cenário real onde é apresentado aos modelos dados informação que não estão disponíveis no momento em que o experimento de ML foi realizado. Isso se faz necessário, uma vez que o *PyCaret* realiza internamente uma divisão dos dados em conjunto de treinamento validação.

O primeiro passo, no fluxograma, é inicializar e configurar o ambiente *PyCaret* com os dados obtidos. Em seguida, deve-se realizar a criação dos modelos e realizar uma comparação dos modelos para avaliar o desempenho. Nesta etapa os modelos são treinados e os pontua usando uma validação cruzada estratificada para a avaliação das métricas.

² ALI, M. *PyCaret: An open source, low-code machine learning library in Python* [S.l.], 2020. *PyCaret* version 1.0.0. Disponível em: <<<https://www.pycaret.org>>>.

Figura 6 – Fluxograma de Aprendizado de Máquina com *PyCaret*.

Fonte: elaborado pelos autores, 2023.

Após a criação e comparação dos modelos realizar a otimização dos hiperparâmetros de cada modelo avaliado sendo utilizado para esta tarefa a técnica de *GridSearch* que constrói e avalia um modelo para cada combinação de parâmetros e técnica de *Early stopping* evitando assim *overfitting*. Na etapa de análise do treinamento é verificado o desempenho através de diferentes métricas como a área sobre, ou do inglês *glsauc*, matriz de confusão entre outros gráficos.

A finalização do modelo consiste em ajustar o modelo ao conjunto completo de dados, incluindo a amostra de teste. O objetivo desta etapa é treinar o modelo no conjunto completo de dados antes de ser implantado na produção. Em seguida são realizadas as previsões em relação à amostra de dados não foram utilizados durante o treinamento e testes e avaliados as métricas.

4 Resultados

Foram escolhidas as seguintes métricas a fim de medir a performance dos modelos desenvolvidos: acurácia, precisão, *recall* e *f1-score*. Essas métricas auxiliam na avaliação e comparação dos modelos, com o objetivo de escolher o que apresentasse o melhor resultado.

A acurácia indica uma performance geral do modelo, ou seja, de todas as classificações, quantas o modelo classificou corretamente. A precisão mede quão assertiva, dentre todas as classificações de classe Positivo que o modelo fez, quantas estão corretas. O *recall* mede, dentre todas as situações de classe positivo como valor esperado, quantas estão corretas e, por último, o *f1-score* é a média harmônica entre precisão e *recall*.

4.1 Comparação dos modelos

Como pode ser visualizado na [Tabela 9](#), o classificador *Random Forest* **RF** e o *Extreme Gradient Boosting* **XGBOOST** demonstraram a mais alta acurácia, atingindo 86%, indicando que ambos os modelos foram capazes de prever corretamente a classe de 86% das instâncias do conjunto de dados. Enquanto isso, o Classificador de Árvore de Decisão **DT** obteve uma acurácia de 82%, e o Classificador K Vizinhos Mais Próximos **KNN** registrou a menor acurácia, alcançando 81%.

Em relação à Área sob a Curva (AUC), o **RF** apresentou o melhor desempenho com um valor de 0,92, evidenciando sua habilidade em distinguir entre classes positivas e negativas. O **XGBOOST** seguiu de perto com uma AUC de 0,90, enquanto o **DT** e o **KNN** registraram valores mais baixos, com 0,72 e 0,80, respectivamente. Todos os modelos exibiram sensibilidades semelhantes, em torno de 86%, indicando que conseguiram identificar corretamente a maioria das instâncias positivas.

Tanto o **RF** quanto o **XGBOOST** apresentaram os melhores resultados de precisão, atingindo 86%, o que demonstra que 86% das instâncias classificadas como positivas por esses modelos eram realmente positivas. O **DT** e o **KNN** obtiveram precisões de 82% e 80%, respectivamente. Além disso, tanto o **RF** quanto o **XGBOOST** registraram os melhores resultados de F1 (86%), uma medida do equilíbrio entre precisão e sensibilidade. Enquanto o **DT** alcançou um F1 de 81%, o **KNN** apresentou o menor valor de F1, também de 81%.

os modelos **RF** e **XGBOOST** tiveram um bom desempenho com valor de acurácia em torno de 86%, já os modelos **DT** e **KNN** obtiveram um desempenho inferior com valores em torno de 81% de acurácia.

Tabela 9 – Resultado da comparação de modelos.

Modelo	Acurácia	AUC	Sensibilidade	Precisão	F1
<i>Random Forest Classifier</i>	0,86	0,92	0,86	0,86	0,86
<i>Extreme Gradient Boosting</i>	0,86	0,90	0,86	0,86	0,86
<i>Decision Tree Classifier</i>	0,82	0,72	0,82	0,82	0,81
<i>K Neighbors Classifier</i>	0,81	0,80	0,81	0,80	0,81

4.2 Criação dos modelos e otimização dos hiper parâmetros

Após o resultado da comparação dos modelos, o próximo passo foi a criação dos modelos e ajuste dos hiper parâmetros. Os resultados são apresentados na [Tabela 10](#) em comparação aos resultados obtidos com os valores de hiper parâmetros sem os ajustes. Como pode ser observado na [Tabela 10](#), o ajuste dos parâmetros do modelo não resultaram em melhora significativa para alguns modelos e para outros o desempenho foi pior.

Isso significa que, cada combinação de parâmetros utilizada no *GridSearch* para avaliação, tem performance inferior quando o modelo utiliza os hiper parâmetros definidos na criação dos modelos. Quando isso acontece, o *PyCaret* retorna o modelo original sem os ajustes de parâmetros. Os modelos [DT](#) e [KNN](#) foram os únicos modelos que obtiveram uma melhora no desempenho aumentando em torno de 4% o valor de acurácia.

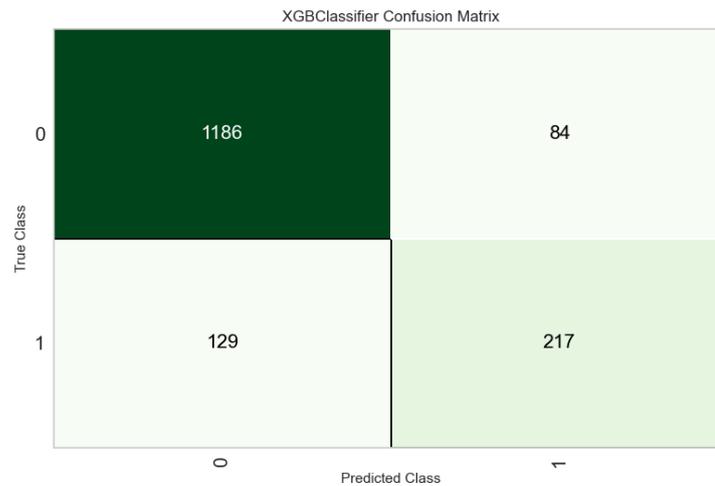
Tabela 10 – Resultados do ajuste de hiperparâmetros com validação cruzada de 10 vezes para os modelos selecionados.

Modelo	Tipo	Acurácia	AUC	Sensibilidade	Precisão	F1
<i>Decision Tree Classifier</i>	Original	0,82	0,72	0,82	0,82	0,81
	Ajustado	0,85	0,87	0,85	0,84	0,84
<i>Random Forest Classifier</i>	Original	0,86	0,91	0,86	0,86	0,86
	Ajustado	0,86	0,91	0,86	0,85	0,85
<i>Extreme Gradient Boosting</i>	Original	0,86	0,90	0,86	0,86	0,86
	Ajustado	0,81	0,88	0,81	0,84	0,82
<i>K Neighbors Classifier</i>	Original	0,81	0,80	0,81	0,80	0,81
	Ajustado	0,84	0,85	0,84	0,83	0,84

4.3 Análise do treinamento

Os modelos [XGBOOST](#) e [RF](#) obtiveram melhores performance, porém, para simplificar as análises, vamos restringir ao classificador [XGBOOST](#). Analisando a [Figura 7](#), pode-se perceber a matriz de confusão, onde 0 é o valor dado para a classificação ‘ruim’ e 1 ‘boa’.

Figura 7 – Matriz de Confusão



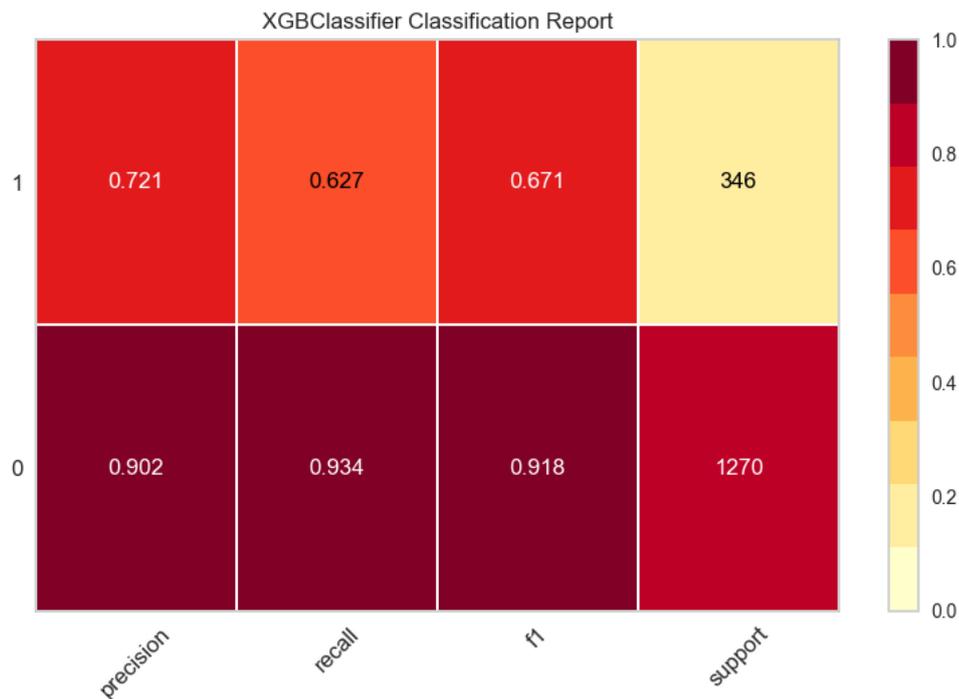
Fonte: elaborado pelos autores, 2023.

Na base de treinamento, dos itens classificados como classe 0, 1186 foram corretamente identificados como classe 0, caracterizando verdadeiros negativos. Entretanto, 129 desses itens foram erroneamente classificados como classe 1, configurando falsos negativos. Além disso, o modelo classificou corretamente 217 itens como classe 1, representando verdadeiros positivos, mas 84 foram incorretamente classificados como classe 0, sendo classificados como falsos positivos. Através das contagens presentes na matriz de confusão é possível calcular métricas de avaliação para a classificação, conforme apresentado na [Figura 8](#). Observa-se que as métricas de avaliação são melhores para a classe 0. Essa métrica é positiva ao trabalho pois o problema analisado é justamente o processo não conforme, como essa população é maior dentro da base de dados é possível identificar os *gaps* e explorar o que tem levado para esse resultado.

Para o recozimento o mais crítico é o que o modelo erre quando o processo esteja não conforme e ele considere que esteja dentro do estabelecido. Pois nesses casos, apesar do processo estar não conforme quando o modelo indicar o contrário o erro poderá não ser identificado e não haverá atuação, levando a consequência direta no produto e falta de qualidade. A precisão é uma métrica que indica, das classificações positivas do modelo quantas foram acertadas ela foi maior dentro do grupo não conforme, o que indica que a precisão do modelo está dentro do desejado para o processo.

O gráfico de importância das variáveis fornece uma pontuação que indica o quão útil ou valioso cada variável foi na construção das árvores de decisão impulsionadas dentro do modelo. Quanto mais um atributo é usado para tomar decisões importantes com árvores de decisão, maior sua importância relativa. Essa importância é calculada para cada atributo no conjunto de dados, permitindo que os atributos sejam classificados e comparados entre si.

Figura 8 – Relatório de métricas de avaliação para a classificação

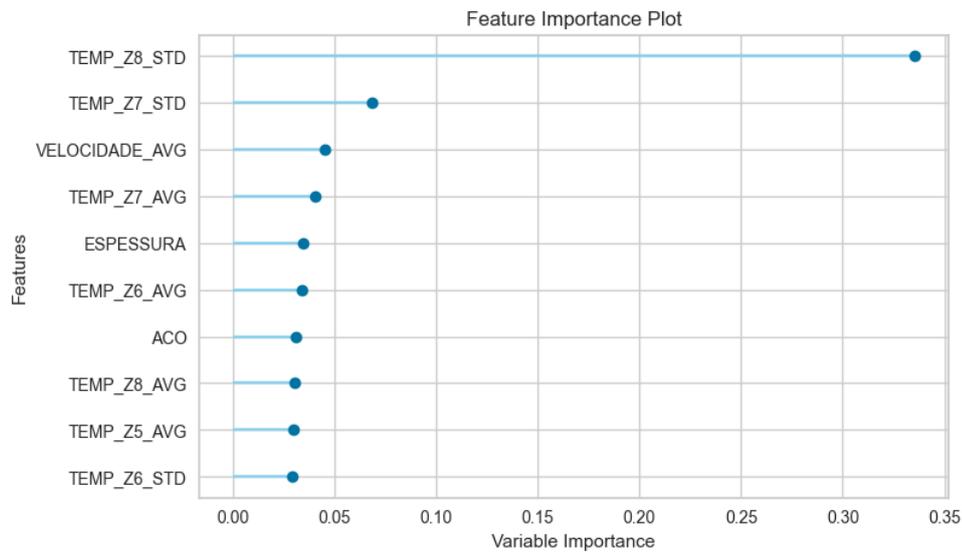


Fonte: elaborado pelos autores, 2023.

A Figura 9 revela que as variáveis com maior influência na temperatura da tira são o desvio padrão das temperaturas das zonas 8 e 7, velocidade média e temperatura média da zona 7. O desvio padrão das temperaturas das zonas 7 e 8 indica a dispersão das temperaturas em torno da média. Um alto desvio padrão pode indicar instabilidade no processo, o que pode afetar a temperatura final da tira. Uma velocidade média mais alta pode contribuir para uma melhor homogeneização do material e, conseqüentemente, para uma temperatura mais uniforme da tira. A temperatura média da zona 7 é um indicador da temperatura do material antes de entrar na zona final de aquecimento. Uma temperatura mais alta na zona 7 pode levar a uma temperatura final da tira mais alta.

Pode-se ressaltar que o gráfico de importância apresenta algumas limitações. Ele não permite determinar se o efeito de cada variável é positivo ou negativo, e a análise foi realizada com base em uma única amostra de dados. Para confirmar os resultados, é necessário realizar análises com mais dados e investigar as dependências entre as variáveis.

Figura 9 – Variáveis mais importantes para modelo



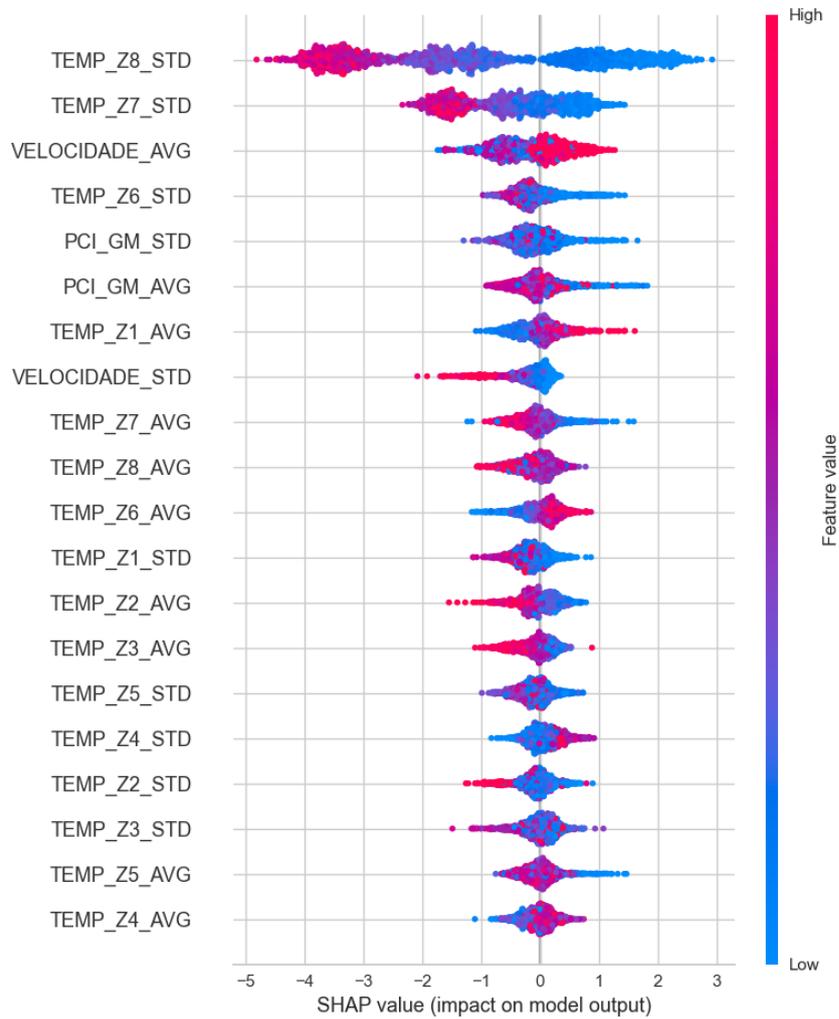
Fonte: elaborado pelos autores, 2023.

Para melhorar a interpretabilidade do modelo, foi utilizado o gráfico de *SHapley Additive exPlanations* (SHAP), que permite uma interpretação melhor dos dados e das variáveis de maior importância para o modelo, quando comparado com o gráfico de importância das variáveis sendo possível não só identificar qual o impacto do valor de uma variável na predição final, mas também realizar uma análise individual de cada resultado. Este gráfico é capaz de explicar esses diversos modelos através da média marginal da contribuição para a predição do modelo de todas as permutações de combinações das variáveis no conjunto de dados. Os resultados são apresentados na [Figura 10](#).

Comparando os resultados do gráfico de importância das variáveis, com os resultados do SHAP, podemos perceber que as variáveis mais importantes não são exatamente a mesma ordem de importância. Pela [Figura 10](#), a ordem de importância são desvio padrão da zonas 8 e 7, velocidade média e desvio padrão da zona 6 (o grau de importância das demais variáveis são apresentados no gráfico de SHAP no sentido de cima para baixo).

Na base de dados deste trabalho, as oscilações de temperatura dentro dos padrões estabelecidos na norma (processo conforme) são apresentadas por 1 e o processo não conforme por 0. Na [Figura 10](#) observa-se que as variáveis de desvio padrão para temperatura das zona 8 e 7 são as de maior importância para o modelo e quanto menor o valor de desvio padrão maior a contribuição para predição da classe 1.

Figura 10 – Gráfico SHAP para modelo



Fonte: elaborado pelos autores, 2023.

A velocidade média alta também tem boa contribuição para atingir a temperatura conforme. Velocidades mais altas, favorecem que a temperatura de tira seja alcançada enquanto valores menores levam a um processo não conforme enquanto o desvio padrão da velocidade é alto tende a contribuir para a classe não conforme. Observa-se também que temperaturas médias das zonas iniciais (2 e 3) mais altas favorecem um processo não conforme. As demais variáveis tem menor força na predição das classes.

4.4 Finalização e predição

Os resultados exibidos na [seção 4.2](#) são baseados no conjunto de treinamento e representam a média de execuções da validação cruzada durante o treinamento. A [Tabela 11](#) apresenta os resultados de predição para modelos ajustados sobre os dados de teste. Entre os modelos avaliados, o *Random Forest Classifier* e o *Extreme Gradient Boosting* destacam-se pela acurácia mais elevada, registrando valores de 0,88 e 0,87, respectivamente. Isso indica que ambos os modelos foram capazes de prever corretamente a classe de aproximadamente 88% e 87% das instâncias do conjunto de teste. Além disso, esses dois modelos também apresentaram valores de AUC semelhantes, indicando uma boa capacidade de distinguir entre classes positivas e negativas, com AUC de 0,90 para ambos.

Em contraste, o *Decision Tree Classifier* e o *K Neighbors Classifier* demonstraram desempenho ligeiramente inferior, com acurácia de 0,84 e 0,82, respectivamente. Embora esses modelos tenham alcançado resultados satisfatórios, suas acurácias são inferiores em comparação com os modelos de *Random Forest* e *XGBoost*. No entanto, é importante ressaltar que todos os modelos exibiram sensibilidade, precisão e pontuação F1 relativamente próximas, indicando uma capacidade semelhante de identificar corretamente instâncias positivas e de alcançar um equilíbrio entre precisão e sensibilidade.

Tabela 11 – Resultados de predição para modelos ajustados sobre dados de teste.

Modelo	Acurácia	AUC	Sensibilidade	Precisão	F1
<i>Random Forest Classifier</i>	0,88	0,90	0,87	0,86	0,86
<i>Extreme Gradient Boosting</i>	0,87	0,90	0,87	0,86	0,86
<i>Decision Tree Classifier</i>	0,84	0,87	0,84	0,84	0,84
<i>K Neighbors Classifier</i>	0,82	0,85	0,82	0,82	0,82

Na sequência foi realizada a finalização dos modelos que tem o objetivo de ajustá-los ao conjunto completo de dados (treino e teste) antes que ele possa liberado para implantação em produção. Após a finalização dos modelos, os dados não vistos reservados no início do experimento foram utilizados para realização da previsão e comparação dos resultados conforme pode ser visualizado na [Tabela 12](#), onde o resultado indica boa capacidade de aplicação do modelo ao problema.

Tabela 12 – Resultados de predição para modelos finalizados sobre dados não vistos.

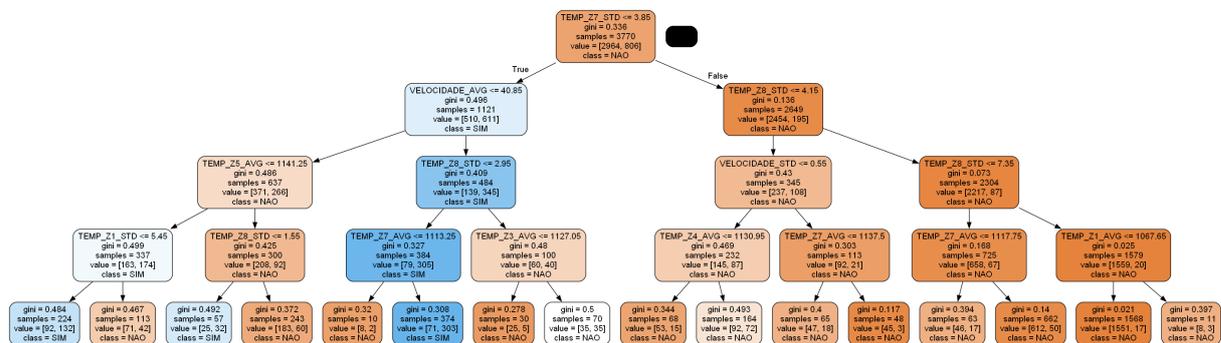
Modelo	Acurácia	AUC	Sensibilidade	Precisão	F1
<i>Random Forest Classifier</i>	0,89	0,92	0,89	0,89	0,89
<i>Extreme Gradient Boosting</i>	0,88	0,92	0,89	0,88	0,88
<i>Decision Tree Classifier</i>	0,86	0,89	0,86	0,87	0,86
<i>K Neighbors Classifier</i>	0,85	0,85	0,85	0,85	0,85

Os modelos Random Forest Classifier e XGBoost demonstraram resultados comparáveis em termos de acurácia em ambas as tabelas analisadas, com uma pequena vantagem observada para o *Random Forest Classifier*, registrando uma taxa de acurácia de 0,89%, em comparação com os 0,88% do *XGBoost*. Em relação à AUC, ambos os modelos exibiram desempenhos semelhantes, com valores aproximados de 0,90 em ambas as tabelas. Quando avaliados em termos de sensibilidade, precisão e F1, os resultados indicam que ambos os modelos demonstram eficácia na identificação de classes positivas e negativas, com valores próximos a 0,89 para o *Random Forest Classifier* e 0,88 para o *XGBoost*.

4.5 Extração de regras

Para aumentar o entendimento do impacto das variáveis no resultado, foi extraído do modelo de DT o gráfico da árvore de decisão. Como é possível verificar na [Figura 11](#), é possível observar claramente as variáveis de maior impacto no modelo, identificando o desvio padrão de temperatura das zonas 7 e 8, bem como a velocidade média, como fatores principais que levam a um aumento no número de amostras classificadas como dentro da norma.

Figura 11 – Árvore de Decisão



Fonte: elaborado pelos autores, 2023.

Ao aplicar as regras derivadas da árvore de decisão, representadas pela rota em azul na [Tabela 13](#), em toda a base de dados, evidencia-se um aumento significativo no número de amostras classificadas como dentro da norma. A progressão da aplicação dessas regras resulta em um total de 491 amostras consideradas dentro da norma, o que corresponde a cerca de 38% do total de amostras avaliadas.

Como na regra da árvore de decisão estão ausentes muitas variáveis do processo, realizou-se uma análise para verificar como as demais variáveis do processo se comportaram quando a regra foi aplicada. O resultado dessa análise é apresentado na [Tabela 14](#).

Tabela 13 – Regra extraída da árvore de decisão (rota em azul) para classificação de amostras dentro da norma (SIM).

Variável	Regra	Percentual Regra (SIM)	Percentual Regra (NAO)
TEMP_Z7_STD	$\leq 3,88$	53,68	46,31
TEMP_Z7_AVG	$> 1113,25$	70,40	29,59
VELOCIDADE_AVG	$> 40,85$	70,40	29,59
TEMP_Z8_STD	$\leq 2,95$	79,58	20,42

Tabela 14 – Regras de processo onde o resultado foi melhor.

Variável	Média	Desvio Padrão	Mínimo	25%	50%	75%	Máximo
ESPESSURA	2,91	0,06	2,60	2,90	2,90	2,90	3,20
LARGURA	1231,07	57,48	1040,00	1240,00	1240,00	1247,00	1279,00
VELOCIDADE_AVG	41,75	0,41	40,90	41,40	41,90	42,00	43,10
VELOCIDADE_STD	0,09	0,19	0,00	0,00	0,00	0,10	1,20
PCI_GM_AVG	3101,26	108,65	2631,40	3028,25	3126,70	3130,90	3466,60
PCI_GM_STD	35,00	25,53	0,80	19,35	26,70	38,40	172,80
TEMP_Z1_AVG	923,03	37,92	845,30	893,60	921,00	940,90	1080,60
TEMP_Z1_STD	5,32	3,29	0,00	3,00	4,50	6,90	19,30
TEMP_Z2_AVG	979,19	29,06	915,50	955,30	979,70	997,15	1065,80
TEMP_Z2_STD	5,29	3,28	0,00	3,00	4,40	6,55	20,60
TEMP_Z3_AVG	1133,21	10,45	1092,40	1128,50	1134,40	1140,40	1158,90
TEMP_Z3_STD	3,21	1,98	0,00	1,90	2,80	4,10	13,60
TEMP_Z4_AVG	1135,41	8,02	1114,20	1130,00	1135,00	1140,00	1169,60
TEMP_Z4_STD	2,18	1,22	0,00	1,30	2,10	2,90	9,20
TEMP_Z5_AVG	1133,53	8,77	1091,00	1128,50	1134,50	1139,70	1156,20
TEMP_Z5_STD	2,59	1,29	0,00	1,70	2,50	3,30	8,00
TEMP_Z6_AVG	1133,28	7,50	1114,10	1129,10	1134,30	1139,00	1153,10
TEMP_Z6_STD	2,08	1,03	0,00	1,40	2,00	2,80	6,40
TEMP_Z7_AVG	1132,42	7,37	1114,00	1127,55	1131,50	1137,00	1154,60
TEMP_Z7_STD	2,03	0,86	0,00	1,40	2,10	2,70	3,80
TEMP_Z8_AVG	1124,58	7,93	1099,40	1119,90	1124,30	1130,25	1145,50
TEMP_Z8_STD	1,59		0,78	0,00	1,00	1,60	2,20 2,90
TEMP_P2_AVG	1090,08	5,07	1077,10	1086,15	1090,00	1093,55	1102,20

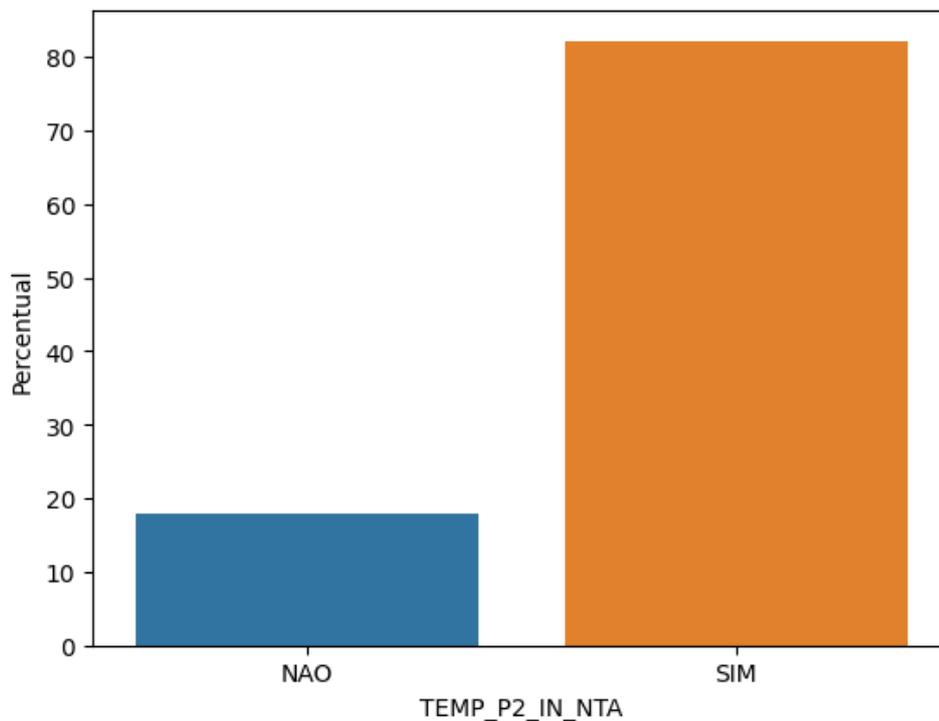
A análise dos desvios padrão revelou que a aplicação da regra final do processo [Tabela 15](#) reduziu o número total de amostras dentro da norma. Porém, essa seleção mais rigorosa resultou em uma proporção maior de amostras que atendem aos critérios de qualidade, conforme pode ser visto na [Figura 12](#). Especificamente, as variáveis de temperatura (exceto “TEMP_Z1_STD”) apresentaram forte impacto na estabilidade do processo. Limites como 6,9 para o desvio padrão de (exceto “TEMP_Z1_STD”) demonstram a importância de controlar a flutuação nessa variável para garantir um processo mais estável.

Além dos desvios padrão, a regra final considera outras variáveis críticas. A velocidade média deve ser superior a 40, possivelmente para garantir um fluxo mínimo que assegure a homogeneidade do processo. O desvio padrão do (exceto “PCI_GM”) precisa ser inferior a 79,60, indicando a necessidade de controlar a variação da rugosidade da superfície do material. A análise também sugere que o modelo funciona melhor quando a espessura está entre 2,64 e 3,2 milímetros e a largura é menor que 1279 milímetros.

Tabela 15 – Regra final de processo.

Variável	Regra
VELOCIDADE_AVG	≥ 40
PCI_GM_STD	$\leq 79,60$
TEMP_Z1_STD	$\leq 6,9$
TEMP_Z2_STD	$\leq 6,55$
TEMP_Z3_STD	$\leq 4,1$
TEMP_Z4_STD	$\leq 2,90$
TEMP_Z5_STD	$\leq 3,3$
TEMP_Z6_STD	$\leq 2,80$
TEMP_Z7_STD	$\leq 2,7$
TEMP_Z8_STD	$\leq 2,2$

Figura 12 – Proporção de amostras quando aplicada nova regra de processo.



Fonte: elaborado pelos autores, 2023.

5 Considerações finais

Os resultados obtidos nesta análise dos modelos desenvolvidos forneceram *insights* valiosos sobre a aplicação dos algoritmos de aprendizado de máquina no contexto abordado. A seguir, são apresentadas considerações finais com base nos resultados apresentados:

Durante a avaliação da performance dos modelos, métricas como acurácia, precisão, *recall* e f1-score foram empregadas. Observou-se que todos os algoritmos alcançaram uma acurácia média em torno de 86%, sugerindo um desempenho com boa capacidade de aplicação do modelo ao problema. No entanto, é essencial ressaltar que a acurácia isoladamente pode não ser suficiente para avaliar completamente a eficácia de um modelo, sendo necessário considerar outras métricas para uma análise mais completa.

A análise comparativa entre os modelos revelou que não houve diferenças significativas em métricas cruciais, tais como [Area Under the Curve \(AUC\)](#), *recall*, precisão e *f1-score*, apesar de apresentarem níveis de acurácia semelhantes. Isso indica que, embora os modelos tenham obtido resultados gerais de previsão similares em termos de acurácia, suas capacidades de identificar corretamente exemplos positivos, evitar falsos positivos e encontrar um equilíbrio entre precisão e *recall* foram consistentes entre eles.

O algoritmo *Random Forest Classifier* demonstrou uma performance ligeiramente superior em comparação com os outros modelos (1%). Os resultados da etapa de ajuste de hiper parâmetros foram variados. Enquanto em alguns casos houve melhora na performance dos modelos, em outros os ajustes não resultaram em benefícios significativos e, em alguns casos, até pioraram o desempenho. Isso ressalta a importância de uma abordagem cuidadosa na seleção e ajuste de hiper parâmetros, levando em consideração as características específicas de cada algoritmo e conjunto de dados.

Embora os modelos tenham apresentado performances semelhantes, uma análise mais detalhada revelou que o algoritmo *XGBoost* é capaz de separar melhor os valores das variáveis de [SHAP](#) de maior importância. Além disso, a análise das variáveis mais importantes destacou que o desvio das temperatura das zonas finais (7 e 8) como o principal fator influenciador na temperatura de bobina.

Os resultados da finalização dos modelos demonstraram uma boa performance na capacidade de classificar corretamente em torno de 89% das amostras corretamente considerando a base de dados com amostras não utilizadas durante o treinamento dos modelos de [ML](#). Isso indica a robustez e a capacidade de generalização dos modelos para o conjunto completo de dados antes da implantação em produção.

Pode-se concluir que o modelo *Random Forest Classifier* obteve um melhor resultado de acuracidade para as análises. Com base nos dados para manter a temperatura de bobina dentro dos parâmetros ideais, foram levantadas as seguintes considerações importantes para o processo:

- Evitar oscilações de temperatura muito altas nas zonas dos fornos;
- Objetivar processo com velocidade mais altas;
- Evitar temperaturas altas na zonas iniciais do forno (2 e 3) pois eles favorecem um processo não conforme.;

Os resultados obtidos fornecem uma base sólida para a utilização dos modelos desenvolvidos na prática. No entanto, é crucial destacar que a análise de modelos de aprendizado de máquina é um processo contínuo e iterativo, sujeito a revisões e refinamentos à medida que novos dados e *insights* se tornam disponíveis. Os parâmetros apresentados auxiliam na tomada de decisão do processo e também apresentam a faixa de trabalho ideal de oscilação que é permitida nas principais variáveis para manter a saída final do modelo (temperatura de bobina) dentro do objetivado. Portanto, recomenda-se a monitorização constante da performance dos modelos e a realização de ajustes conforme necessário para garantir sua relevância e eficácia contínuas.

5.1 Trabalhos futuros

Esses padrões serão integrados ao processo atual e como proposta de trabalho futuro modelo propõe-se:

- Automatizar o forno para realizar esses ajustes de forma automática mantendo o processo conforme para a variável de respostas temperatura de bobina;
- Trabalhar com dados mais detalhados, com informação a cada metro;
- Refazer o experimento gerando modelos para cada faixa de espessura e largura;
- Analisar o impacto das transições de espessura e largura de material na linha.

Referências

- ALI, M. **PyCaret: An open source, low-code machine learning library in Python**. [S.l.], 2020. PyCaret version 1.0.0. Disponível em: <<https://www.pycaret.org>>.
- BOGO, J.; OLIVEIRA, J.; LIMA, E.; MELLO, R.; FURTADO, A. Sport analytics a systematic review. **Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional**, SBC, p. 670–681, 2019.
- BOSCARIOLI, C.; BEZERRA, A.; BENEDICTO, M. d.; DELMIRO, G. Uma reflexão sobre banco de dados orientados a objetos. In: SN. **Congresso de Tecnologias para Gestão de Dados e Metadados do Cone Sul, Paraná, Brasil**. [S.l.], 2006.
- BROOKS, J.; LIPPOLD, J. Selection of wrought austenitic stainless steel. In: **ASM Handbook**. Metals Park, Ohio: ASM International, 1993. v. 1, p. 456–470.
- BROOKS, J.; THOMPSON, A. Microstructural development and solidification cracking susceptibility of austenitic stainless steel. **International Materials Review**, v. 36, n. 1, p. 16–44, 1991.
- BROWN, M. L.; KROS, J. F. Data mining and the impact of missing data. **Industrial Management & Data Systems**, MCB UP Ltd, v. 103, n. 8, p. 611–621, 2003.
- CHO, S.-H.; YOO, Y.-C.; JONAS, J. J. Static and dynamic strain aging in 304 austenitic stainless steel at elevated temperatures. **Journal of Materials Science Letters**, v. 19, n. 22, p. 2019–2022, 2000.
- COSTA, L. H.; MARÇAL, G. H. d. R.; PERUCCI, G. A. Predição do teor de hidrogênio nos refinamentos secundários de uma aciaria usando redes neurais artificiais. 2023.
- DANTAS, C. A. **Seleção de atributos baseado em algoritmos de agrupamento para tarefas de classificação**. Dissertação (Mestrado) — Brasil, 2017.
- DUTTA, B.; SELLARS, C. M. Effect of composition and process variables on nb (c, n) precipitation in niobium microalloyed austenite. **Materials Science and Technology**, Taylor & Francis, v. 3, n. 3, p. 197–206, 1987.
- FURTADO, M. I. V. **Redes Neurais Artificiais: Uma Abordagem Para Sala de Aula**. Ponta Grossa, PR: Atena Editora, 2019.
- GORNI, A. A.; SILVA, M. R. S. da; SILVEIRA, J. H. D. da. Previsão de propriedades mecânicas de bobinas laminadas a quente através de redes neurais. **Tecnologia Metalúrgica e de Materiais**, v. 5, n. 4, p. p. 229–234, abr.-jun 2009.
- HUMPHREYS, J. F. Nucleation in recrystallization. In: TRANS TECH PUBL. **Materials Science Forum**. [S.l.], 2004. v. 467, p. 107–116.
- LEE, H. D. **Seleção de Atributos Baseado em Algoritmos de Agrupamento para Tarefas de Classificação**. Tese (Tese de Doutorado) — Universidade de São Paulo, 2005.

MOORE, D.; MCCABE, G.; CRAIG, B. **Introduction to the Practice of Statistics**. W. H. Freeman, 2014. ISBN 9781464133633. Disponível em: <https://books.google.com.br/books?id=pX1_AwAAQBAJ>.

PADILHA, A.; RIOS, P. Decomposition of austenite in austenitic stainless steels. **ISIJ International**, v. 42, n. 4, p. 325–337, 2002.

PADILHA, A.; SICILIANO, F. J. **Encruamento, recristalização, crescimento de grão e textura**. 3. ed. São Paulo: Associação Brasileira de Metalurgia e Materiais, 2005. 232 p.

PULGARÍN, C. I. H. **Estudo dos fenômenos que ocorrem durante o recozimento dos aços inoxidáveis austeníticos 304L e 316L deformados em várias temperaturas**. 122 p. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, São Paulo, 2006.

ROSS, S. M. **Introduction to Probability and Statistics for Engineers and Scientists**. [S.l.]: Academic Press, 2014.

STONE, M. Cross-validators choice and assessment of statistical predictions. **Journal of the Royal Statistical Society. Series B (Methodological)**, Wiley for the Royal Statistical Society, v. 36, n. 2, p. pp. 111–147, 1974. ISSN 00359246.

TAKAHASHI, H. J.; RABELO, G. B. da C.; TEIXEIRA, R. de A. Aplicação de técnicas de inteligência computacional para predição de propriedades mecânicas de aços de alta resistência microligados. **Tecnologia em Metalurgia e Materiais**, v. 5, n. 2, p. 100–104, out.-dez 2008.

VERONEZE, L. R. **Métodos de imputação de dados aplicados a dados de empresas industriais**. Dissertação (Dissertação de Mestrado) — Universidade Estadual de Campinas, São Paulo, 2011.

VIEIRA, D. A. G. **Rede perceptron com camadas paralelas (PLP - Parallel Layer Perceptron)**. Tese (Doutorado) — Universidade Federal de Minas Gerais, 2006.

WEIGEND, A. S.; HUBERMAN, B. A.; RUMELHART, D. E. Predicting the future: A connectionist approach. **International Journal of Neural Systems**, v. 01, n. 03, p. 193–209, 1990. Disponível em: <[doi:10.1142/S0129065790000102](https://doi.org/10.1142/S0129065790000102)>.