



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



PREDIÇÃO DE LASCAS DE AQUECIMENTO DURANTE A FABRICAÇÃO DE AÇOS
INOXIDÁVEIS DUPLEX

Fabício Luiz de Alcântara

Gessica Seara Faria

Janeir Ribeiro Dutra

Lucas Soares Cunha

Timóteo, MG
2024

Fabrcio Luiz de Alcântara, Gécica Seara Faria, Janeir Ribeiro Dutra, Lucas
Soares Cunha

PREDIÇÃO DE LASCAS DE AQUECIMENTO DURANTE A FABRICAÇÃO DE AÇOS INOXIDÁVEIS DUPLEX

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

Orientador: Prof. Dr. Thiago Augusto de Oliveira Silva

Timóteo, MG

2024

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

C972p Cunha, Lucas Soares.
Predição de lascas de aquecimento durante a fabricação de aços
inoxidáveis duplex. [manuscrito] / Lucas Soares Cunha. et al. - 2024.
41 f.: il.: color., tab..

Orientador: Prof. Dr. Thiago Augusto de Oliveira Silva.
Produção Científica (Especialização). Universidade Federal de Ouro
Preto. Departamento de Engenharia de Produção.

1. Aço inoxidável - Controle de qualidade. 2. Aço - Defeitos - Previsão.
3. Ciência de dados. 4. Previsão - Processos de fabricação. I. Alcântara,
Fabrício Luiz de. II. Dutra, Janeir Ribeiro. III. Faria, Gessica Seara. IV. Silva,
Thiago Augusto de Oliveira. V. Universidade Federal de Ouro Preto. VI.
Título.

CDU 658.5:519.2

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



FOLHA DE APROVAÇÃO

Fabício Luiz de Alcântara

Géssica Seara Faria

Janeir Ribeiro Dutra

Lucas Soares Cunha

Predição de lascas de aquecimento durante a fabricação de aços inoxidáveis duplex

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 29 de fevereiro de 2024

Membros da banca

Dr. Thiago Augusto de Oliveira Silva - Orientador(a) - Universidade Federal de Ouro Preto

Dr. Alexandre Xavier Martins - Universidade Federal de Ouro Preto

Me. Hélio Braz Loss - Aperam South America

Thiago Augusto de Oliveira Silva, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 06/03/2024



Documento assinado eletronicamente por **Thiago Augusto de Oliveira Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 11/03/2024, às 11:32, 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 **0680758** e o código CRC **8C971A15**.

Dedicamos este trabalho a Deus. Sem ele nada seria possível.

Agradecimentos

Agradecemos primeiro a Deus por ter nos mantido na trilha certa durante este projeto de pesquisa com saúde e força para chegar até o final.

À empresa Aperam South America, em especial aos líderes da Alexandre Farah, Edalmo Magalhães e Tarcísio Reis, por nos proporcionaram a oportunidade de realização desse trabalho.

Agradecemos ao nosso orientador Thiago Augusto pelo incentivo e pela dedicação do seu tempo ao nosso projeto de pesquisa.

Agradecemos também aos colegas de trabalho por todo o apoio durante a realização do projeto, em especial ao nosso colega Mychell Laurindo que sempre se fez disponível ao longo da nossa trajetória.

E também somos gratos às nossas famílias pela compreensão e paciência demonstrada durante o período do projeto.

“A persistência é o caminho do êxito.” Charles Chaplin

Resumo

Os aços inoxidáveis duplex são utilizados em aplicações que necessitam de alta resistência à corrosão associada à elevada resistência mecânica. A fabricação da camada interior de tubos flexíveis, utilizados para extração de petróleo, é uma das principais aplicações deste aço, sendo submetidos à ambientes extremamente agressivos. Portanto problemas de qualidade, tanto internos quanto superficiais, no aço podem reduzir sua vida útil significativamente. Em função disso, esse trabalho propõe um estudo para predição de um dos principais defeitos que acomete os aços duplex, denominado lascas de aquecimento. Para realização do estudo, foram coletadas variáveis de produto e do processo de Aciaria dos aços Duplex. Em seguida, os dados foram tratados e explorados através de quatro algoritmos de aprendizagem de máquina distintos: Regressão logística, *Random Forest*, SVM e *Gradient Boosting*. Os métodos de *Ensemble*, *Stacking Ensemble* e *Voting Classifier*, foram aplicados nos modelos utilizados para melhorar a precisão das previsões. A métrica utilizada para avaliar se a adequação dos modelos se ajusta aos dados foi a acurácia, na qual o melhor desempenho foi obtido a partir do modelo *Gradient Boosting*. A análise SHAP foi aplicada para facilitar a interpretabilidade dos resultados e dessa forma foram elencadas as variáveis com maior influência na ocorrência do defeito lascas de aquecimento. Uma árvore de decisão foi construída para obtenção da receita a ser seguida de forma a atingir menores valores de ocorrência de lascas de aquecimento.

Palavras-chaves: Aços inoxidáveis duplex. Lascas de aquecimento. Gradiente Boosting. Ciência de dados. SHAP.

Abstract

Duplex stainless steels are used in applications that require high resistance to corrosion associated with high mechanical strength. The manufacture of the inner layer of flexible pipes, used for oil extraction, is one of the main applications of this steel, being subjected to extremely aggressive environments. Therefore, quality problems, both internal and superficial, in steel can significantly reduce its service life. As a result, this work proposes a study to predict one of the main defects that affect duplex stainless steels, called heating slivers. To carry out the study, product and steelmaking process variables for Duplex steel were collected. After that, the data was processed and explored through four different machine learning algorithms: Logistic Regression, Random Forest, SVM and Gradient Boosting. The Ensemble, Stacking Ensemble and Voting Classifier methods were applied to the models used to improve the accuracy of the predictions. The metric used to evaluate whether the adequacy of the models fits the data was accuracy, in which the best performance was from the Gradient Boosting model. The SHAP analysis was applied to improve the interpretability of the results and in this way the variables with the greatest influence on the occurrence of the heating slivers defect were listed. A decision tree was built to obtain the recipe to be followed in order to achieve lower occurrence values for heating slivers.

Keywords: Duplex stainless steels. Heating slivers. Gradient Boosting. Data Science. SHAP.

Lista de ilustrações

Figura 1 – Evolução do Custo da Não Qualidade do Aço Duplex.	2
Figura 2 – Representação da configuração de um tubo flexível.	5
Figura 3 – Camadas geológicas e sua extensão: representação de extração de petróleo.	6
Figura 4 – Fluxo de produção dos aços Duplex na Aciaria.	7
Figura 5 – Processo de laminação à quente do Duplex.	7
Figura 6 – Lascas de aquecimento em bobinas laminadas a frio de Duplex.	7
Figura 7 – Trinca associada à lasca de aquecimento, observada na superfície transversal à direção de laminação. Aumento: 500x.	8
Figura 8 – (a) Efeito do forjamento na ductilidade do aço 304 e (b) Influência do B na ductilidade do aço 316.	9
Figura 9 – (a) Efeito de elementos de liga na ductilidade a quente do aço AISI 304L; (b) Efeito do titânio na Faixa de temperatura frágil (Brittle Temperature Range - BRT a $\epsilon = 4\%$) no aço 310S.	10
Figura 10 – Valores de acurácia obtidos de acordo com os modelos utilizados.	21
Figura 11 – Resultado da análise SHAP aplicada ao modelo <i>Gradient Boosting</i>	22
Figura 12 – Reavaliação dos Valores de acurácia obtidos de acordo com os modelos utilizados.	24
Figura 13 – Resultado da análise SHAP final, aplicada ao modelo <i>Gradient Boosting</i>	25
Figura 14 – Árvore de decisão aplicada ao banco de dados otimizado.	26

Lista de tabelas

Tabela 1 – Variáveis referentes ao produto e processo de elaboração, utilizadas nos modelos.	16
Tabela 2 – Variáveis referentes ao processo de lingotamento contínuo, utilizadas nos modelos.	17
Tabela 3 – Resumo das estratégias de modelagem utilizadas.	19
Tabela 4 – Ocorrência de lascas para a condição 1 (Ti e S).	27
Tabela 5 – Ocorrência de lascas para a condição 2 (Ti, S e Cr).	27

Lista de abreviaturas e siglas

AOD-L *argon and oxygen decarburization by lance*

BFs bobinas laminadas a frio

BQs bobinas laminadas a quente

CCC cúbica de corpo centrado

CNQ Custo da Não Qualidade

FP forno panela

FRX Espectrometria de Fluorescência de Raio X

GBM Gradient Boosting Machines

KDD Descoberta de Conhecimento em Base de Dado

K-NN K-vizinhos mais próximos

LC lingotamento contínuo

LTQ laminação de tiras a quente

PLC Controlador Lógico Programável

PTG pré-tratamento de gusa

SHAP SHapley Additive exPlanations

SIP Sistema Integrado de Produção

SVM Máquinas de Vetores de Suporte

VOD *vacuum oxygen decarburization*

ZAC zona afetada pelo calor

Sumário

1	INTRODUÇÃO	1
1.1	Objetivo geral	2
1.1.1	Objetivos específicos	2
2	REVISÃO DA LITERATURA	3
2.1	Aços inoxidáveis Duplex	4
2.1.1	Aplicação e fluxo de produção dos aços Duplex	4
2.2	Defeito lascas de aquecimento	7
2.2.1	Trincas de solidificação	8
2.2.2	Fragilidade a quente	9
2.2.3	Ductilidade a quente	9
2.3	Ciência de Dados e Machine Learning	10
2.3.1	Regressão logística	12
2.3.2	Árvores de decisão	12
2.3.3	Random Forest	12
2.3.4	Máquinas de Vetores de Suporte	13
2.3.5	Gradient Boosting	13
2.3.6	Ensamble	14
3	METODOLOGIA	15
3.1	Aquisição de dados	15
3.2	Estratégia de modelagem	17
3.2.1	Treinamento e teste do modelo	18
3.2.2	Otimização	18
3.2.3	Ensemble	19
3.2.3.1	Stacking Ensemble	19
3.2.3.2	Voting Classifier	19
3.2.4	Configuração da modelagem	20
4	RESULTADOS	21
5	CONSIDERAÇÕES FINAIS	28
	REFERÊNCIAS	29

1 Introdução

Os aços inoxidáveis duplex são definidos como uma família de aços inoxidáveis formados por duas fases dúcteis: ferrita e austenita com uma proporção de cada fase em torno de 50%. Resultado de uma união perfeita entre os aços inoxidáveis austeníticos e os aços inoxidáveis ferríticos, estes aços combinam as melhores características de ambas as ligas, como por exemplo, elevada resistência mecânica, imunidade à corrosão sob tensão e boa soldabilidade. A adição de nitrogênio aos aços inoxidáveis duplex permite aumentar a tenacidade e a resistência à corrosão da [zona afetada pelo calor \(ZAC\)](#) comparativamente à do metal base na condição soldada. Com baixo percentual de níquel em sua composição química, estes aços apresentam seu limite de escoamento muito superior aos dos aços inoxidáveis austeníticos, permitindo a concepção de projetos com espessuras mais finas e com maior resistência à corrosão.

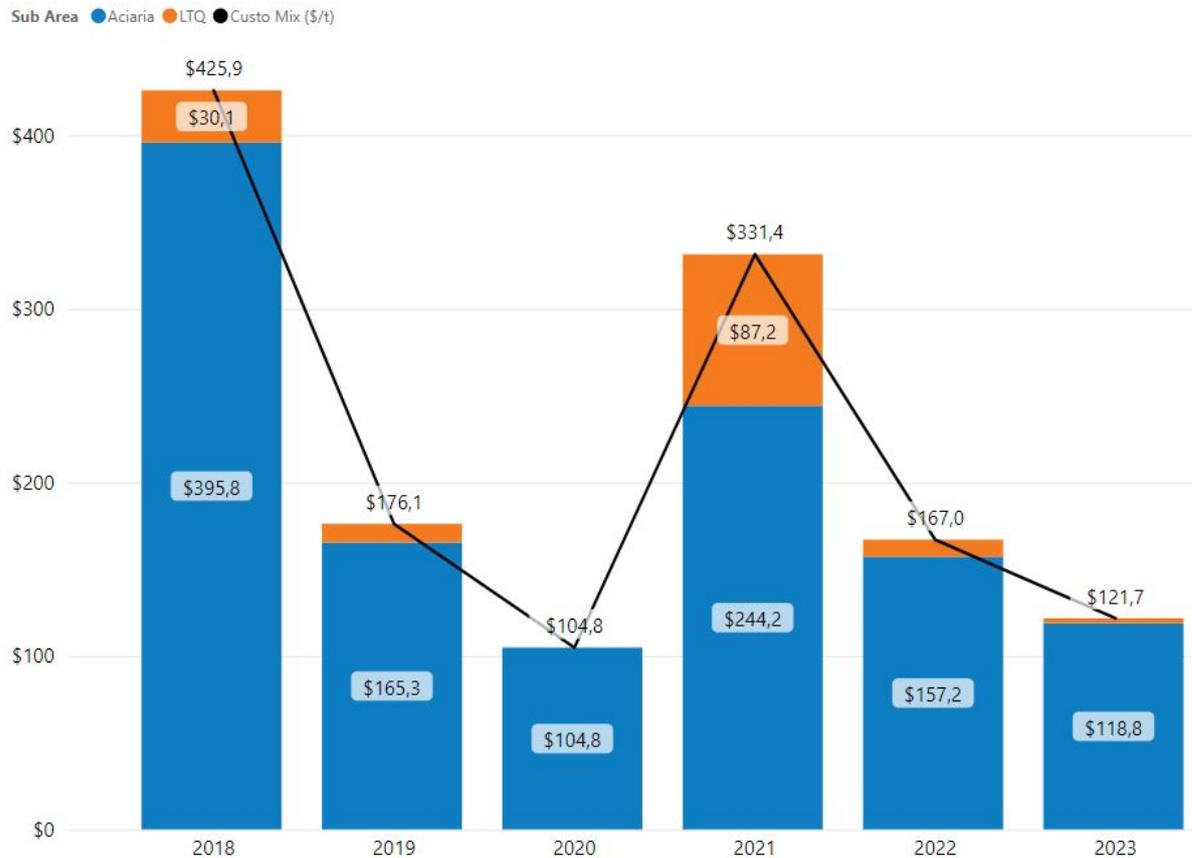
Além da utilização na indústria de petróleo e gás em tubos flexíveis para extração de petróleo, os aços duplex (UNS S32205/UNS S31803/DIN 1.4462) podem também ser aplicados em tanques para armazenagem de produtos químicos e equipamentos para o segmento de papel e celulose. O duplex (UNS S32304/DIN 1.4362), também conhecido como “lean duplex”, possui baixo percentual de níquel em sua composição química, tornando-se uma opção atrativa também para aplicações em implementos rodoviários e tanques para navios.

A ocorrência do defeito lascas de aquecimento é citada por diversos fabricantes de aços inoxidáveis austeníticos planos. Desde a década de 80, várias missões da Aperam (antiga Acesita) ao exterior constataram a presença de lascas de aquecimento nas linhas de inspeção de várias usinas. Durante o seminário do IASMO realizado em 1986 (*International Association of Steckel Mill Operators*) foram apresentadas as dificuldades que algumas empresas tinham para solucionar este defeito. Diversas plantas produtoras de aço inoxidável austenítico inauguradas desde a Europa até o Leste Asiático, como a China por exemplo, enfrentam problemas devido a este defeito ([SANTANA et al., 2006](#)).

Desde o início da produção dos aços duplex na Aperam, grandes avanços foram alcançados no que tange a redução da ocorrência do defeito. Nos últimos 10 anos o sucateamento por lascas de aquecimento reduziu de um patamar inicial de cerca de 15%, para valores próximos a 5% em 2022. Apesar do grande avanço no índice de sucateamento por lascas de aquecimento na Aperam, existem momentos de descontrole deste indicador, sendo as causas desconhecidas.

O custo agregado ao produto, assim como a margem de venda dos aços duplex são extremamente elevados. Conseqüentemente há a necessidade não só de melhor controle deste indicador, como também a redução para patamares ainda mais baixos. O impacto do sucateamento é expressivo e é mensurado através do [Custo da Não Qualidade \(CNQ\)](#), conforme mostrado na [Figura 1](#).

Figura 1 – Evolução do Custo da Não Qualidade do Aço Duplex.



Fonte: Interna Aperam.

1.1 Objetivo geral

O objetivo geral deste trabalho é desenvolver um modelo de aprendizado de máquina que seja capaz de classificar as bobinas quanto a formação do defeito de lasca de aquecimento do aço inoxidável duplex.

1.1.1 Objetivos específicos

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

- Mapeamento das variáveis de processo com potencial de influenciar no defeito;
- criar uma sistemática padrão para levantamento de dados;
- analisar a importância das variáveis na predição do modelo;
- realizar uma análise crítica da importância das variáveis do processo na predição do modelo.

2 Revisão da Literatura

Os aços inoxidáveis possuem uma composição química baseada nos elementos químicos Fe-Cr, sendo que apenas para teores de cromo iguais ou superiores a 10,5% classificados como aços inoxidáveis, conforme norma ASTM (ASTM, 2022).

O Cr contribui para a formação de um filme fino e contínuo de óxido de cromo ao longo da superfície do aço, que o protege contra o ataque corrosivo. Este filme é denominado camada passiva e é impermeável e insolúvel em meios corrosivos usuais. Quando rompido ou danificado, a camada passiva é imediatamente regenerada em atmosfera oxidante, mantendo assim a capacidade de proteção à corrosão (LACOMBE; BAROUX; BERANGER, 1993).

Apesar do Cr ser o elemento principal da composição química dos aços inoxidáveis, diversos outros elementos podem ser adicionados com o objetivo de melhorar suas propriedades, sejam elas metalúrgicas, físicas, químicas ou mecânicas. O Molibdênio, por exemplo, melhora a resistência à corrosão localizada e promove o endurecimento por solução sólida. Já o cobre pode contribuir quanto a resistência à corrosão para determinados reagentes, assim como melhora a estampabilidade do aço. O Nióbio e Titânio são elementos estabilizadores e também evitam a corrosão intergranular. O Silício contribui para a resistência à oxidação em temperaturas elevadas. A adição destes elementos permite a formação de um extenso conjunto de materiais que associado à qualidade superficial, em função de diferentes tipos de acabamentos, explica o crescente uso como material arquitetural, em ambientes internos e externos (SILVA; MEI, 2010).

A microestrutura de um aço inoxidável pode ser influenciada pela sua composição química, sendo possível obter três tipos de microestrutura: Martensítica, ferrítica e austenítica. Cada classe dos aços inoxidáveis possui propriedades mecânicas e físicas únicas e será produzida de acordo com uma norma específica. Existem cinco principais categorias que os aços inoxidáveis podem ser classificados (SILVA; MEI, 2010):

- aços inoxidáveis martensíticos: são ligas compostas essencialmente de Fe-Cr (11 a 18% de Cr), com teor de carbono acima de 0,1% para garantir que a microestrutura seja martensítica. São caracterizados pela estrutura tetragonal e são necessários quando a aplicação requer propriedades de boa resistência à tração, fluência e resistência à fadiga combinadas com resistência à corrosão. Neste grupo podem-se citar os aços AISI 403, 410, 416, 420 e 422;
- aços inoxidáveis ferríticos: são ligas de Fe-Cr que possuem estrutura **cúbica de corpo centrado (CCC)** e que não endurecem por tratamento térmico de têmpera. Geralmente, possuem teor de cromo mais elevado do que os martensíticos, em torno de 10 a 30%, e menor teor de carbono. Neste grupo podem-se citar os aços AISI 405, 409, 430, 446 e 502;

- aços inoxidáveis austeníticos: são ligas a base de Fe-Cr (16 - 30%) - Ni (8 - 35%). Os mais conhecidos são os da série 300, como por exemplo, os aços AISI 301, 304, 308 e 316. Devido ao elevado custo, parte do níquel pode ser substituído por manganês ou nitrogênio, formando outra família de aços inoxidáveis, a série 200. Os principais aços deste grupo são: AISI 201, 202 e 204. Possuem estrutura CCC e são não magnéticos. O teor de carbono é, em geral, inferior a 0,08%;
- aços inoxidáveis endurecíveis por precipitação: são ligas de Fe-Cr (12 - 17%), Ni (4 - 8%), Mo (0 - 2%), contendo adições de elementos que permitem o endurecimento da martensita pela precipitação de compostos intermetálicos à base de alumínio, cobre, nióbio e/ou titânio.
- aços inoxidáveis duplex: possuem microestruturas de ferrita e austenita, obtidas com composições balanceadas de Fe-Cr (18 - 27%) - Ni (4 - 7%) - Mo (1 - 4%) e outros elementos. Os aços que fazem parte desta família são: AISI 329, UNS S32304 e S31803. Esta classe possui menor teor de níquel quando comparado aos aços austeníticos, mas com propriedades mecânicas e de corrosão similares ou superiores aos austeníticos.

2.1 Aços inoxidáveis Duplex

Os aços inoxidáveis duplex são resultados de uma união perfeita entre a estrutura e propriedades dos aços ferríticos e austeníticos, formando uma estrutura bifásica. Formam uma categoria que une eficientemente uma notável resistência à corrosão com uma elevada resistência (VIANA *et al.*, 2015). Suas propriedades físicas posicionam-se entre as dos aços inoxidáveis austeníticos e ferríticos, inclinando-se mais para as características dos aços ferríticos e do aço carbono. A resistência à corrosão por pites e em frestas está diretamente relacionada aos teores de cromo, molibdênio, tungstênio e nitrogênio. Essa resistência pode equivaler à do aço 316 ou superar a dos aços inoxidáveis convencionais, especialmente em aplicações marítimas. Todos os aços inoxidáveis duplex demonstram uma notável resistência à corrosão sob tensão na presença de cloretos, ultrapassando muitas vezes a dos aços inoxidáveis austeníticos da série 300. Além disso, oferecem uma resistência mecânica significativamente superior aos tipos austeníticos, combinada com boa ductilidade e tenacidade (IMOA *et al.*, 2009).

2.1.1 Aplicação e fluxo de produção dos aços Duplex

O início do desenvolvimento na Aperam ocorreu no ano de 2008, com a produção em escala industrial iniciada em 2010 e consolidação do desenvolvimento em 2014. O foco do desenvolvimento foi o atendimento do setor de Oil & Gas, para a fabricação de tubos flexíveis utilizados na extração de petróleo como carro chefe e em seguida desenvolvimento de outras aplicações como a indústria de papel e celulose.

Atualmente a Aperam South America produz o aço lean duplex P398A (UNS S32304), com o foco na indústria de papel e celulose e também os aços Duplex P399A (UNS S32205) e P399B (UNS S1803), que são utilizados para a fabricação da camada interior dos tubos flexíveis, denominada carcaça (Figura 2) e são objetos deste trabalho.

Figura 2 – Representação da configuração de um tubo flexível.

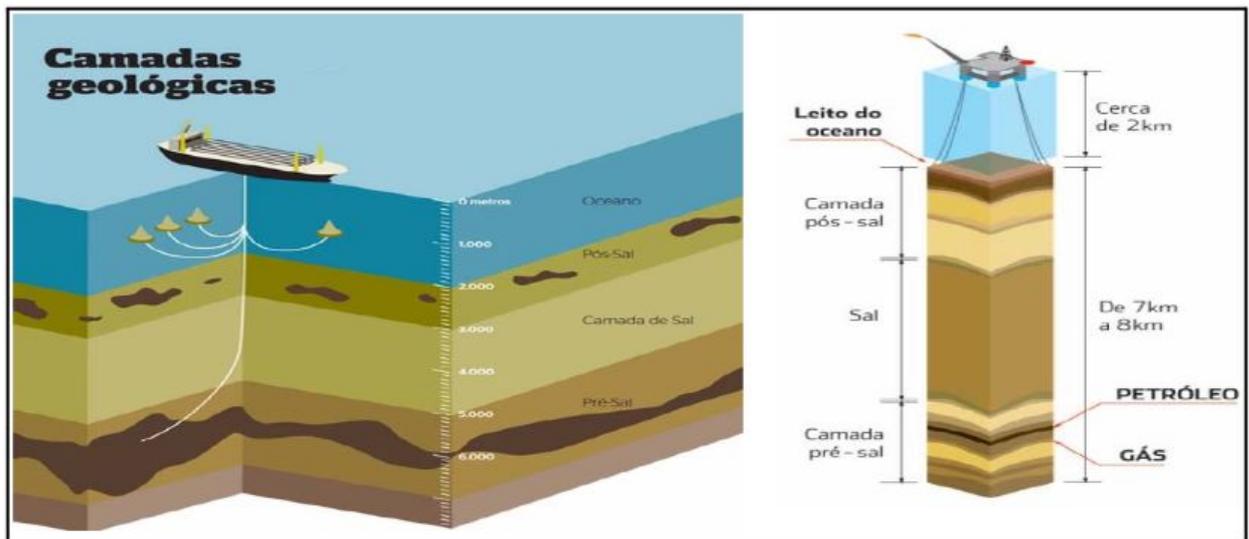


Fonte: [Almiron \(2010\)](#).

Se tratando da aplicação de tubos flexíveis, a Figura 3 apresenta um esquema das camadas geológicas a serem perfuradas para extração do petróleo e a extensão de cada uma delas. Para atingir as camadas pré-sal, que estão entre 5000 e 7000 metros de profundidade, foram desenvolvidos novos projetos de perfuração: mais de 2000 metros de sal foram atravessados. O petróleo da camada pré-sal apresenta alta corrosividade e a profundidade da coluna d'água exerce uma pressão externa podendo chegar a valores superiores a 300 bar.

A primeira camada dos tubos flexíveis (carcaça), tem contato direto com o petróleo e seus contaminantes, sendo submetida a um ambiente de alta pressão e extremamente agressivo. Essa camada tem a função de resistir à corrosão e garantir a resistência do tubo flexível ao colapso ([VIANA et al., 2015](#)).

Figura 3 – Camadas geológicas e sua extensão: representação de extração de petróleo.



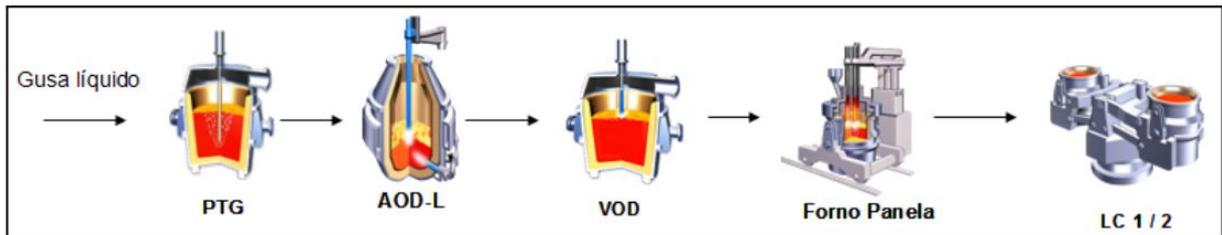
Fonte: [Barbosa et al. \(2015\)](#).

Problemas de qualidade como lascas de aquecimento, esfoliações, arranhões, marcas mecânicas, dentre outros, impedem a utilização destes aços na aplicação final e não são aceitos pelos clientes de tubos flexíveis. Tais problemas podem reduzir a vida útil dos tubos significativamente, bem como deixá-los suscetíveis à corrosão preferencial e, conseqüentemente, ao colapso. O colapso de um tubo flexível utilizado na extração de petróleo pode ter um impacto devastador em relação em níveis de catástrofe ambiental, pois pode resultar em vazamentos descontrolados de petróleo no oceano, causando uma série de problemas.

Dessa forma, a severidade do ambiente de aplicação, associada à garantia mínima de 25 anos que a Aperam fornece para o aço, faz com que as exigências de qualidade superficial dos aços duplex sejam extremamente elevadas. O grau de exigência de qualidade destes aços leva à uma alta complexidade de produção desde a Aciaria até as linhas de acabamento.

A produção na Aciaria é realizada com a utilização de ligas especiais específicas na rota definida como triplex onde se utiliza além do convertedor *argon and oxygen decarburization by lance* (AOD-L), dois equipamentos para o refino secundário *vacuum oxygen decarburization* (VOD) e o forno panela (FP). Antes do AOD-L, o gusa é tratado na estação de *pré-tratamento de gusa* (PTG). Após o refino do aço líquido, o processo de *lingotamento contínuo* (LC) transforma o material em placas com espessura de 200 mm. Para garantia da qualidade superficial das bobinas laminadas a frio, 100% das placas de duplex são esmerilhadas com com tipo de esmerilhamento mais rigoroso existente. A Figura 4 representa o fluxo de produção na Aciaria.

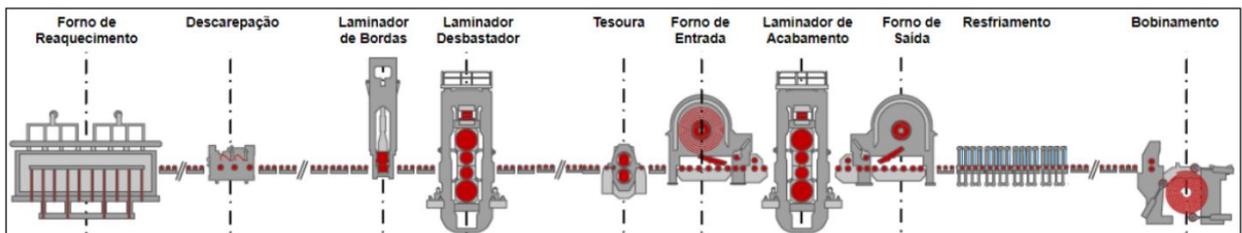
Figura 4 – Fluxo de produção dos aços Duplex na Aciaria.



Fonte: Relatórios Internos Aperam.

A Figura 5 mostra o fluxo de produção do Duplex no processo de **laminação de tiras a quente (LTQ)**. O processo de laminação a quente é realizado iniciando com o aquecimento das placas no forno com atmosfera oxidante, passando em seguida por um processo de laminação de desbaste e finalmente pela laminação de acabamento. Dessa forma a placa com inicialmente 200 mm de espessura, é reduzida em **bobinas laminadas a quente (BQs)**, com espessuras que variam entre 4 a 6 mm.

Figura 5 – Processo de laminação à quente do Duplex.



Fonte: Relatórios Internos Aperam.

2.2 Defeito lascas de aquecimento

O defeito lasca de aquecimento ocorre tipicamente nos aços austeníticos e Duplex e geralmente é identificado, na maioria das vezes, em **bobinas laminadas a frio (BFs)**. Trata-se de um defeito alongado na direção longitudinal na forma aproximada de “V” com dimensões e intensidade variadas (Figura 6). Ocorrem sob a forma de agrupamentos concentrados mais próximos das bordas (de 40 a 200 mm) em ambas as faces e, em casos mais raros, podem se apresentar distribuídos em toda a superfície da bobina (SILVA *et al.*, 2002).

Figura 6 – Lascas de aquecimento em bobinas laminadas a frio de Duplex.

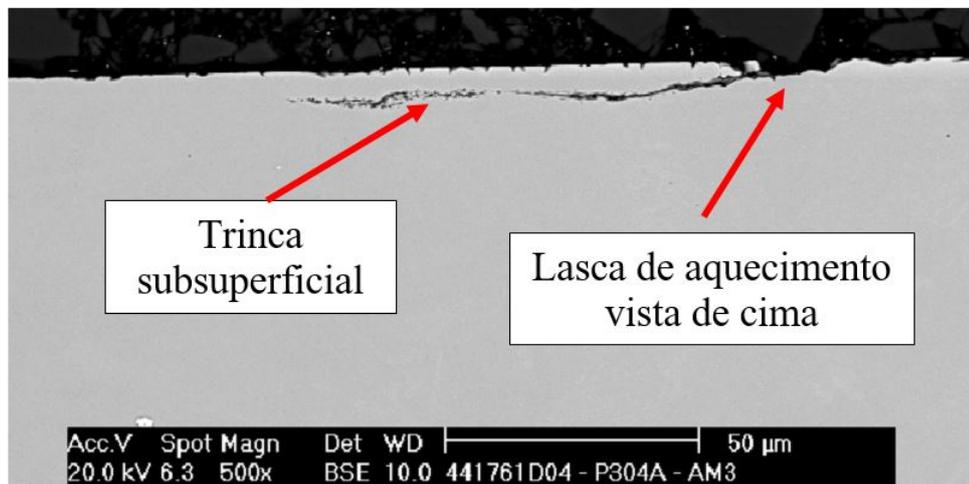


Fonte: Interna Aperam.

O defeito se trata de uma microtrinca interna do material que é revelada no processo de conformação a frio, conforme observado na Figura 7. De forma geral, a origem das lascas são prejudiciais aos aços em função de serem pontos preferenciais de corrosão. Sua origem pode ser classificada em três grandes grupos (SANTANA *et al.*, 2006).

- trincas de solidificação: ocorrem durante a solidificação no lingotamento contínuo;
- fragilidade à quente: elementos contaminantes ou oxidação dos contornos de grão durante o aquecimento de placas;
- baixa ductilidade a quente: relacionado a estrutura do material durante a deformação.

Figura 7 – Trinca associada à lasca de aquecimento, observada na superfície transversal à direção de laminação. Aumento: 500x.



Fonte: Interna Aperam.

2.2.1 Trincas de solidificação

As propriedades mecânicas das placas de aço inoxidável durante e após o lingotamento contínuo são influenciadas pela sequência de solidificação, mais particularmente pela fração de ferrita- δ (primária) presente na estrutura. Esta fração é função da composição química e existem várias equações e diagramas desenvolvidos experimentalmente para seu cálculo. Uma maneira de se estimar a fração de ferrita- δ primária é através da relação Creq/Nieq. A Aperam utiliza a equação desenvolvida pela KSC, que fornece diretamente o valor da ferrita- δ :

$$Ferrita\delta = 3,2 \cdot [Cr + 1,5 \cdot Si + Mo + 0,5 \cdot (Nb + Ti)] - 2,5 \cdot [Ni + 0,5 \cdot Mn + 0,3 \cdot Cu + 30 \cdot (C + N)] - 24,7 \quad (2.1)$$

A alta sensibilidade à trincas nos aços que se solidificam em austenita é devido à baixa solubilidade de elementos de contaminação, tais como o P e S, na austenita, e consequente micro segregação destes elementos nos contornos de grão, tornando o material frágil e susceptível a trincas.

2.2.2 Fragilidade a quente

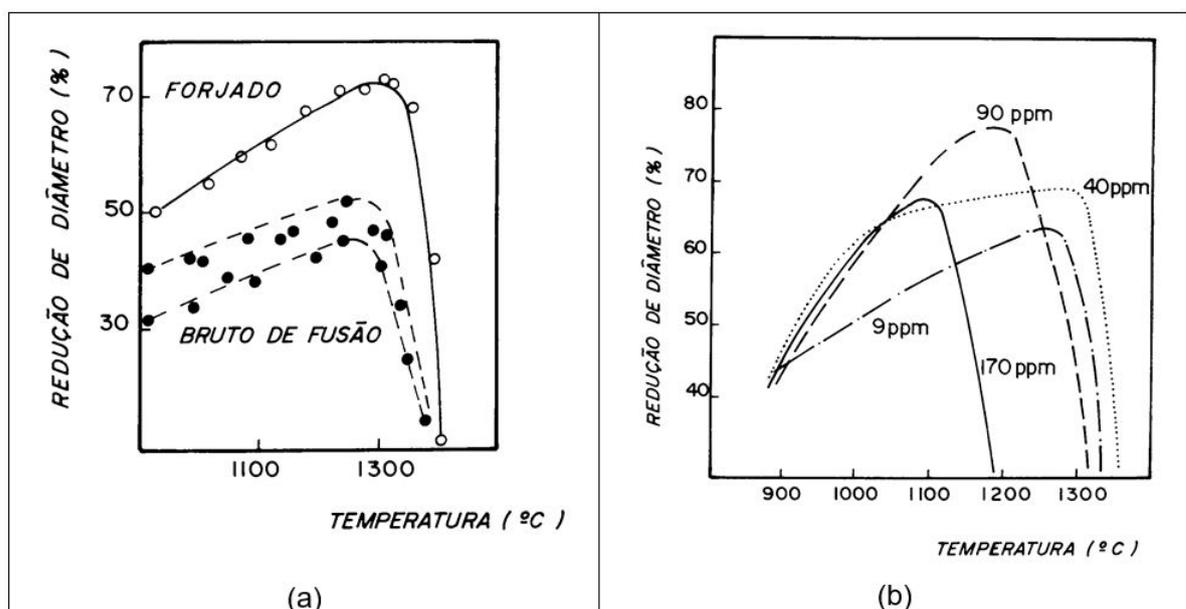
A presença de elementos de contaminação com ponto de fusão inferior as temperaturas normais de laminação a quente provocam o fenômeno conhecido como *hot shortness*, que se caracteriza pela fragilização dos contornos de grão pela presença de material líquido. As contaminações mais comuns são Pb, Sn, Bi, Te, As. O Cu também provoca o fenômeno quando em quantidades superiores a 2%. Outras causas para a fragilidade a quente são: a oxidação dos contornos de grão (queima do material) devido a temperaturas muito altas e tempo excessivo de encharque das placas, e os compostos com baixo ponto de fusão como o NiS, que se forma a partir da presença de S no combustível utilizado no reaquecimento de placas, ou mesmo em função do teor de S residual na composição química do próprio aço.

2.2.3 Ductilidade a quente

Entende-se por ductilidade a capacidade de um material de se deformar plasticamente sem apresentar trincas ou rupturas. A ductilidade a quente de um metal é função das seguintes variáveis: temperatura, velocidade de deformação, estrutura (Figura 8a), capacidade de recristalização e recuperação, sejam estas estáticas ou dinâmicas.

Os aços 316, por exemplo, devido à presença do Mo, apresentam ductilidade a quente inferior aos aços 304, exigindo faixas mais restritas de composição química e controle mais rigoroso na laminação a quente. A adição de B aos aços 316 na faixa de 40 ppm aumenta a ductilidade e é prática comum entre os fabricantes de aço inoxidável (Figura 8b).

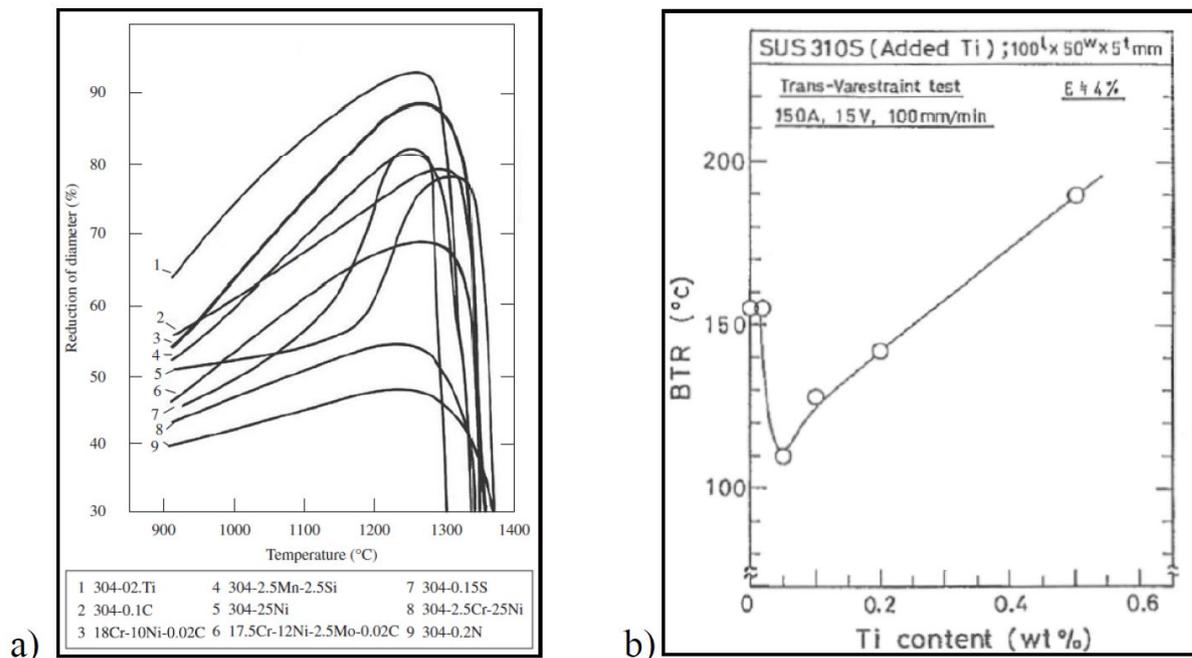
Figura 8 – (a) Efeito do forjamento na ductilidade do aço 304 e (b) Influência do B na ductilidade do aço 316.



Fonte: Santana *et al.* (2006).

Em um estudo realizado em aços inoxidáveis austeníticos por Keown em 1980 e revistado por Plaut et al. (2007), foi identificada uma melhoria significativa da ductilidade a quente do aço 304L após pequenas adições de titânio no 304 (Figura 9a). Outro estudo realizado por Fourmentin (2009), mostra o efeito positivo da adição de titânio na temperatura frágil do aço 310S, em teores variando entre 0,04 a 0,12% (Figura 9b).

Figura 9 – (a) Efeito de elementos de liga na ductilidade a quente do aço AISI 304L; (b) Efeito do titânio na Faixa de temperatura frágil (Brittle Temperature Range - BTR a $\epsilon = 4\%$) no aço 310S.



Fonte: Plaut et al. (2007) e Fourmentin (2009).

2.3 Ciência de Dados e Machine Learning

A Ciência de Dados e o *Machine Learning* têm se tornado pilares fundamentais na era da informação, impulsionando descobertas e aplicações inovadoras em diversas áreas. Quando exploramos esse vasto campo, é impossível não mencionar o trabalho de Christopher M. Bishop, renomado pesquisador em aprendizado de máquina. O autor enfatiza em suas obras a importância da modelagem probabilística. Ele destaca que a incerteza é inerente aos dados e que uma abordagem probabilística fornece uma estrutura sólida para lidar com essa complexidade. Em sua obra "*Pattern Recognition and Machine Learning*", destaca como os métodos probabilísticos, como redes Bayesianas e modelos gráficos, oferecem uma base sólida para representar e inferir sobre incertezas nos dados (BISHOP; NASRABADI, 2006).

A Ciência de Dados, um campo interdisciplinar que utiliza métodos científicos, processos e sistemas para extrair conhecimento dos dados em várias formas, se beneficia enormemente dessas abordagens. A aplicação de modelos probabilísticos na Ciência de Dados permite não apenas prever tendências, mas também avaliar a confiabilidade dessas previsões, o que é crucial em ambientes complexos e dinâmicos.

O processo de aprendizado de máquina busca a forma como os seres humanos assimilam padrões e utilizam conceitos aprendidos para solucionar problemas complexos. Um algoritmo de aprendizado consiste, basicamente, na etapa onde o sistema passando pela fase de aprendizado através de exemplos, é capaz de generalizar para dados que ele não conhece, a informação ou padrão aprendido (NORVIG; RUSSELL, 2013). Com isso, o sistema é capaz de absorver informações e características dos padrões fornecidos a ele e desenha uma representação do problema através de um processo iterativo de ajuste de pesos e conexões entre as unidades de processamento (NORVIG; RUSSELL, 2013).

No domínio específico do *Machine Learning*, Bishop destaca a relevância dos modelos de aprendizado supervisionado e não supervisionado. Os métodos supervisionados, nos quais o modelo é treinado em um conjunto de dados rotulado, são essenciais para tarefas como classificação e regressão. Os métodos não supervisionados também são abordados, como a análise de agrupamento por exemplo. O aprendizado não supervisionado conhece apenas as variáveis do problema e através dos padrões interpretados por ele, ele é capaz de generalizar as informações (BISHOP, 2013).

Um exemplo marcante de aplicação prática é o uso de Redes Neurais, um subconjunto do aprendizado de máquina, em que Bishop também fez contribuições significativas. Redes Neurais profundas, em particular, têm demonstrado um desempenho notável em tarefas complexas, desde reconhecimento de voz até visão computacional. A abordagem não se limita à teoria; ele ressalta a importância da aplicação prática e da validação de modelos. A iteração entre teoria, experimentação e aplicação prática é uma característica marcante de seu enfoque, destacando a necessidade de uma abordagem holística na Ciência de Dados e no *Machine Learning*.

Existem muitos modelos de *machine learning*, cada um projetado para resolver diferentes tipos de problemas. Dentre os modelos supervisionados, por exemplo, podemos citar regressão Linear, regressão logística, *random forest*, Máquinas de Vetores de Suporte (SVM), K-vizinhos mais próximos (K-NN), redes neurais, naive bayes, máquinas de aprendizado extremo (*XGBoost*, *LightGBM*).

Estas são apenas algumas opções e a escolha do modelo depende da natureza do problema, do tipo de dados disponíveis e de outros fatores. A prática comum é experimentar vários modelos para determinar qual funciona melhor para um determinado conjunto de dados e objetivo.

No presente trabalho foram utilizados os modelos regressão logística, *random forest*, SVM, *gradiente boost* e *Ensembles*, que serão detalhados a seguir.

2.3.1 Regressão logística

A Regressão Logística é amplamente estudada e utilizada em diversos domínios, com uma base teórica sólida. Um marco clássico é o trabalho de Cox (1958), que estabeleceu as bases da regressão logística. Desde então, a técnica evoluiu em resposta às necessidades práticas, com avanços notáveis na modelagem de dados longitudinais e na incorporação de penalizações para lidar com *overfitting* (FIRTH, 1993).

Recentemente, tem havido uma crescente aplicação da Regressão Logística em contextos além da simples classificação binária, como a regressão logística multinomial para problemas de classificação multiclasse (AGRESTI; TARANTOLA, 2018). Abordagens bayesianas também têm ganhado popularidade para lidar com incertezas nos parâmetros do modelo (GELMAN; HWANG; VEHTARI, 2014).

Regressão Logística é um modelo de *machine learning* que tem suas raízes na regressão linear, mas é especialmente adequado para tarefas de classificação binária. O cerne da regressão logística reside na função logística, que transforma a combinação linear das características de entrada em uma probabilidade entre 0 e 1.

Este modelo é amplamente utilizado em situações onde a relação entre as variáveis preditoras e a variável resposta é não linear. Sua simplicidade torna-o fácil de interpretar e implementar. No entanto, a regressão logística pode ser sensível a *outliers*, e é mais eficaz em problemas onde as classes são linearmente separáveis.

2.3.2 Árvores de decisão

As Árvores de Decisão é uma técnica amplamente utilizada em aprendizado de máquina, introduzida por Quinlan (1986). Estruturadas hierarquicamente, elas consistem em nós representando decisões e ramos conectando esses nós. Diversos algoritmos, foram desenvolvidos para a construção de árvores, utilizando critérios como entropia, ganho de informação e índices de impureza.

Sua interpretabilidade e capacidade de lidar com dados heterogêneos são destaque. No entanto, desafios como *overfitting* levaram ao desenvolvimento de técnicas como poda de árvores e *Ensemble learning* (ex. *Random Forest*).

2.3.3 Random Forest

A técnica de *Random Forest*, proposta por Breiman (2001), representa uma abordagem eficaz para melhorar a precisão de modelos baseados em árvores de decisão. Desde então, a *Random Forest* tem sido extensivamente estudada em relação à sua robustez, flexibilidade e capacidade de lidar com grandes conjuntos de dados e alta dimensionalidade.

A literatura destaca a capacidade do *Random Forest* em lidar com o *overfitting* e em lidar bem com dados desbalanceados (LIAW; WIENER *et al.*, 2002). Além disso, a aplicação de técnicas de interpretabilidade, como a importância de características, tem sido um foco crescente de pesquisa (STROBL *et al.*, 2008).

É uma abordagem de ensemble baseada em árvores de decisão. Cada árvore é treinada em uma amostra aleatória dos dados, e a combinação das previsões de várias árvores reduz o *overfitting* e melhora a robustez do modelo. Essa técnica é particularmente eficaz em problemas de classificação e regressão, oferecendo alta precisão e capacidade de lidar bem com dados desbalanceados e alta dimensionalidade.

Apesar de sua eficácia, o *Random Forest* é menos interpretável do que árvores de decisão individuais. Além disso, a construção de múltiplas árvores pode tornar o modelo computacionalmente intensivo, especialmente em grandes conjuntos de dados.

2.3.4 Máquinas de Vetores de Suporte

As *SVMs* foram introduzidas por Cortes e Vapnik (1995) e rapidamente ganharam destaque devido à sua eficácia em lidar com problemas de classificação e regressão. O trabalho seminal destacou a importância da maximização da margem de separação entre classes.

Desde então, a *SVM* tem sido estendida para lidar com conjuntos de dados não lineares, através do uso de kernels (CRISTIANINI; SHAWE-TAYLOR, 2000). A literatura atual enfoca a otimização de hiperparâmetros, a interpretabilidade e a aplicação em problemas de aprendizado profundo (COLLOBERT, 2011).

Se trata de um poderoso modelo utilizado tanto em classificação quanto em regressão. A essência das *SVMs* reside na busca por um hiperplano que otimiza a separação entre diferentes classes no espaço de características. A maximização da margem entre as instâncias das classes é fundamental para o desempenho eficaz do modelo.

As *SVMs* são particularmente eficazes em problemas de alta dimensionalidade, sendo utilizadas em diversas aplicações, como reconhecimento de padrões, processamento de linguagem natural e diagnóstico médico. No entanto, a interpretabilidade das *SVMs* pode ser desafiadora, especialmente em espaços de características complexos. A seleção adequada do kernel é crucial para lidar com dados não lineares e explorar a versatilidade do modelo.

2.3.5 Gradient Boosting

Gradient Boosting é uma abordagem de aprendizado de máquina que se destaca por sua capacidade de criar modelos preditivos poderosos através da combinação sequencial de modelos mais simples. Ao longo dos anos, essa técnica evoluiu e deu origem a algoritmos notáveis, como *Gradient Boosting Machines (GBM)*, *XGBoost*, *LightGBM* e *CatBoost*.

O **GBM**, introduzido por **Friedman (2001)**, representa um marco inicial no desenvolvimento da técnica de *Gradient Boosting*. Este método utiliza árvores de decisão como modelos base e visa minimizar a função de perda através da adição iterativa de árvores ponderadas. A flexibilidade do **GBM** o torna adequado para uma ampla gama de problemas, incluindo regressão e classificação.

2.3.6 Ensemble

Os métodos de *ensemble* são técnicas avançadas em aprendizado de máquina que combinam as previsões de múltiplos modelos para melhorar a confiabilidade, robustez e precisão das previsões. Essa abordagem baseia-se na ideia de que a agregação das previsões de um conjunto de modelos pode superar as limitações dos modelos individuais (**OPITZ; MACLIN, 1999**).

Existem várias técnicas de ensemble, cada uma com suas características e aplicações específicas. O *Stacking*, por exemplo, é uma técnica de ensemble que utiliza um modelo de aprendizado de máquina para combinar as previsões de vários outros modelos (**WOLPERT, 1992**). Outra técnica popular é o *Voting Classifier*, que combina as previsões de vários modelos por meio de votação majoritária ou ponderada (**KITTLER et al., 1998**). Estes métodos têm sido amplamente aplicados em diversas áreas do aprendizado de máquina e demonstraram melhorar significativamente o desempenho dos modelos em tarefas complexas de classificação e regressão (**DIETTERICH, 2000**).

3 Metodologia

Para a elaboração deste trabalho utilizamos a abordagem [Descoberta de Conhecimento em Base de Dado \(KDD\)](#). Esta abordagem envolve um processo sistemático e iterativo para descobrir informações úteis e previamente desconhecidas a partir de grandes conjuntos de dados. Ela envolve várias etapas que visam transformar dados brutos em conhecimento compreensível e útil. As etapas da abordagem [KDD](#) são geralmente representadas pelo processo de descoberta de conhecimento em nove fases: definição do problema, seleção de dados, pré-processamento, redução de dimensionalidade e transformação, mineração de dados, interpretação e análise dos resultados, apresentação dos resultados, implementação do conhecimento e consolidação.

3.1 Aquisição de dados

O conjunto de dados brutos inclui 60 variáveis de processo, que são essenciais para analisar e compreender os processos industriais sob investigação. O conjunto de dados incorpora parâmetros de composição química, temperatura e vazão e pressão de água nas zonas do equipamento de lingotamento contínuo, que desempenham papéis cruciais na dinâmica global do processo. Além disso, existem variáveis relacionadas com o produto, tais como espessura, largura, identificação do aço e ocorrência do defeito.

O conjunto de dados utilizado compreende séries temporais e variáveis descritivas, que foram coletadas da elaboração e lingotamento contínuo da Aciaria, de uma usina siderúrgica, compreendendo o período de 21 de janeiro de 2020 a 29 de maio de 2023. Os elementos carbono enxofre e nitrogênio são medidos via combustão com análise de massa. Os demais elementos da composição química são obtidos via [Espectrometria de Fluorescência de Raio X \(FRX\)](#). Ambos os resultados são coletados pelo [Sistema Integrado de Produção \(SIP\)](#).

O lingotamento contínuo possui vários sensores que estão distribuídos por toda a sua estrutura, e todos eles estão conectados a um [Controlador Lógico Programável \(PLC\)](#). O PLC processa os dados e os envia para o Software Historian, que os comprime antes de os armazenar numa base de dados. A frequência com que os dados são armazenados é definida pelo utilizador. Em seguida, os dados são enviados para a nuvem e armazenados num grande repositório de dados.

As variáveis de processo de lingotamento foram coletadas na base do PIMS, as quais foram extraídas considerando o valor médio e o desvio padrão de vazão e pressão, para cada unidade metálica. A compilação das variáveis contidas no conjunto de dados é apresentada nas tabelas [1](#) e [2](#), acompanhada das respectivas descrições e tipos de dados.

Tabela 1 – Variáveis referentes ao produto e processo de elaboração, utilizadas nos modelos.

Índice	Variável (produto e elaboração)	Unidade de Engenharia	Tipo do dado
1	ID do produto (UM)	-	object
2	Sequência da Placa	-	object
3	ID da Corrida	-	object
4	Aço	-	object
5	Desbobinadeira	-	int64
6	Espessura BQ	mm	float64
7	Espessura BF	mm	float64
8	Largura	mm	int64
9	Peso	t	float64
10	Nº processos Forno LTQ	-	int64
11	Pré Laminação (S/N)	-	int64
12	Ocorrência de Lasca	-	int64
13	Intensidade da lasca	-	int64
14	Turno	-	int64
15	Data de Inspeção	-	datetime64[ns]
16	Data de Aciaria	-	datetime64[ns]
17	Data de LTQ	-	datetime64[ns]
18	Teor de C	%	float64
19	Teor de Si	%	float64
20	Teor de S	%	float64
21	Teor de Mo	%	float64
22	Teor de B	%	float64
23	Teor de Co	%	float64
24	Teor de Ti	%	float64
25	Teor de Mn	%	float64
26	Teor de P	%	float64
27	Teor de Cr	%	float64
28	Teor de Ni	%	float64
29	Teor de Al	%	float64
30	Teor de Cu	%	float64
31	Teor de V	%	float64
32	Teor de Nb	%	float64
33	Teor de Sn	%	float64
34	Teor de W	%	float64
35	Teor de N	%	float64
36	Teor de O	%	int64
37	Teor de H	%	float64
38	Teor de Pb	%	float64
39	Cr equivalente (cálculo)	%	float64
40	Ni equivalente (cálculo)	%	float64
41	Ferrita (cálculo)	%	float64

Tabela 2 – Variáveis referentes ao processo de lingotamento contínuo, utilizadas nos modelos.

Índice	Variável (lingotamento contínuo)	Unidade de Engenharia	Tipo do dado
42	Velocidade de Lingotamento	m/min	float64
43	Desv. Pad. Velocidade	-	float64
44	Temperatura do Distribuidor	°C	float64
45	Desv. Pad. Temperatura Distribuidor	-	float64
46	Temperatura Placa	°C	float64
47	Desv. Pad. Temperatura Placa	-	float64
48	Nível do Molde	mm	float64
49	Desv. Pad. Nível do Molde	-	float64
50	Vazão de água Zona 1	l/min	float64
51	Desv. Pad. Vazão de água Zona 1	-	float64
52	Vazão de água Zona 2	l/min	float64
53	Desv. Pad. Vazão de água Zona 2	-	float64
54	Vazão de água Zona 3	l/min	float64
55	Desv. Pad. Vazão de água Zona 3	-	float64
56	Vazão de água Zona 4	l/min	float64
57	Desv. Pad. Vazão de água Zona 4	-	float64
58	Vazão de água Zona 5 Fixa	l/min	float64
59	Desv. Pad. Vazão de água Zona 5 Fixa	-	float64
60	Vazão de água Zona 5 Móvel	l/min	float64
61	Desv. Pad. Vazão de água Zona 5 Móvel	-	float64
62	Vazão de água Zona 6 Móvel	l/min	float64
63	Desv. Pad. Vazão de água Zona 6 Móvel	-	float64
64	Pressão Zona 1	bar	float64
65	Desv. Pad. Pressão Zona 1	-	float64
66	Pressão Zona 2	bar	float64
67	Desv. Pad. Pressão Zona 2	-	float64
68	Pressão Zona 3	bar	float64
69	Desv. Pad. Pressão Zona 3	-	float64
70	Pressão Zona 4	bar	float64
71	Desv. Pad. Pressão Zona 4	-	float64
72	Pressão Zona 5 Móvel	bar	float64
73	Desv. Pad. Pressão Zona 5 Móvel	-	float64
74	Pressão Zona 5 Fixa	bar	float64
75	Desv. Pad. Pressão Zona 5 Fixa	-	float64
76	Pressão Zona 6 Móvel	bar	float64
77	Desv. Pad. Pressão Zona 6 Móvel	-	float64

3.2 Estratégia de modelagem

Neste estudo, foram explorados quatro algoritmos de aprendizagem automática distintos, especificamente Regressão logística, *Random Forest*, *SVM* e *Gradient Boosting*. O treino destes modelos foi realizado utilizando a biblioteca Python de código aberto amplamente utilizada, *scikit-learn* (PEDREGOSA *et al.*, 2011).

3.2.1 Treinamento e teste do modelo

Com o objetivo de obter estimativas confiáveis do modelo e evitar o *overfitting*, duas estratégias foram implementadas para avaliar o modelo. A primeira estratégia consiste em dividir os dados disponíveis em conjuntos de dados de treino e de teste. Cerca de 70% dos dados disponíveis foram utilizados para desenvolver o modelo. Os restantes 30%, foram utilizados para avaliar o desempenho das previsões em dados não conhecidos pelo modelo. Este processo pode fornecer informações sobre a variância do modelo e sua capacidade de generalizar para novos dados.

A segunda estratégia consistiu em utilizar a validação cruzada durante o treino do modelo, a partir da biblioteca scikit-learn utilizando a função `cross_val_score` que divide os dados em k conjuntos, treina o modelo k vezes usando diferentes conjuntos como dados de teste e, em seguida, calcula a média das métricas de desempenho resultantes.

A métrica utilizada para avaliar se a adequação dos modelos se ajusta aos dados foi a acurácia. Trata-se de uma medida amplamente utilizada que fornece um *feedback* importante sobre o desempenho dos modelos utilizados. A acurácia representa a proporção de predições corretas em relação ao total de observações.

3.2.2 Otimização

A fim de melhorar o desempenho do modelo, foi realizada uma otimização dos parâmetros. Foi executado um plano para a otimização de hiperparâmetros para cada modelo (Regressão Logística, Random Forest, SVM, *Gradient Boosting*) usando o método *Grid Search* com validação cruzada, sendo:

- Regressão Logística:
 - Hiperparâmetros a serem otimizados: C (força de regularização), `penalty` (tipo de penalidade).
- *Random Forest*:
 - Hiperparâmetros: `n_estimators` (número de árvores), `max_features` (número de recursos), `max_depth` (profundidade máxima das árvores),
- SVM:
 - Hiperparâmetros: C (força de regularização), `gamma` (coeficiente de kernel para 'rbf', 'poly' e 'sigmoid').
- Gradient Boosting:
 - Hiperparâmetros: `n_estimators` (Número de estimadores), `learning_rate` (Taxa de aprendizado), `max_depth` (Profundidade máxima da árvore).

O uso de GridSearchCV com $cv=5$ indica a utilização da validação cruzada de 5 dobras. O range de exploração para cada um dos algoritmos investigados é apresentado na Tabela 3.

Tabela 3 – Resumo das estratégias de modelagem utilizadas.

Modelo	Hiperparâmetros	Range
Regressão Logística	C	[0.001, 0.01, 0.1, 1, 10, 100]
	penalty	['l1', 'l2']
	n_estimators	[10, 50, 100, 200]
Random Forest	max_features	['auto', 'sqrt', 'log2']
	max_depth	[None, 10, 20, 30, 40, 50]
SVM	C	[0.1, 1, 10, 100]
	gamma	[1, 0.1, 0.01, 0.001]
	n_estimators	[100, 200, 300]
Gradient Boosting	learning_rate	[0.01, 0.1, 0.2]
	max_depth	[3, 4, 5]

3.2.3 Ensemble

Nos modelos de ensemble, combinamos as previsões de vários modelos de aprendizado de máquina para melhorar a confiabilidade e a precisão das previsões. Essa abordagem é baseada no princípio de que um grupo de modelos fracos ou moderados pode se unir para formar um modelo muito forte. Neste trabalho, exploramos dois métodos de ensemble: *Stacking Ensemble* e *Voting Classifier*.

3.2.3.1 Stacking Ensemble

O Stacking é uma técnica de ensemble que combina as previsões de vários modelos base para gerar um conjunto de dados de previsões que é então usado como entrada para um meta-aprendiz. No contexto deste trabalho, implantamos um *Stacking Ensemble* utilizando *Random Forest*, *SVM* e *Gradient Boosting* como modelos base, e a Regressão logística serviu como o meta-aprendiz.

Esta abordagem permite que o meta-aprendiz aprenda como melhor combinar as previsões dos modelos base. Cada um dos modelos base traz suas próprias forças e padrões de captura de dados, e o meta-aprendiz se ajusta para aproveitar ao máximo essas características individuais.

3.2.3.2 Voting Classifier

O *Voting Classifier* é outro método de ensemble onde múltiplos modelos são usados para fazer previsões, e a ‘votação’ final é determinada pela combinação dessas previsões. Neste trabalho, utilizamos um *Voting Classifier* como ‘voting’ suave, que considera a probabilidade prevista para cada modelo. Os modelos incluídos no *Voting Classifier* foram Regressão Logística, *Random Forest*, *SVM* e *Gradient Boosting*.

A principal vantagem do *Voting Classifier* é que ele pode capturar o conhecimento coletivo de múltiplos modelos, cada um contribuindo com suas perspectivas únicas.

3.2.4 Configuração da modelagem

Inicialmente, os dados foram coletados dos sistemas SIP e PIMS, em seguida foram pré processados com a finalidade de limpeza dos dados e remoção de dados faltantes.

Após o pré processamento, foram removidas as colunas não numéricas da base de dados. Este processo resultou na criação de um conjunto de dados preciso, pronto para a modelagem de dados. Em seguida os modelos foram treinados e validados. Cada modelo individual foi submetido a avaliação em um conjunto de teste, facilitando assim uma avaliação do seu desempenho e confiabilidade.

Com o objetivo de melhorar a interpretabilidade do modelo, foi utilizada da biblioteca *shap* do *python* que se baseia no *SHapley Additive exPlanations (SHAP)* para explicar o modelo que apresentou maior acurácia.

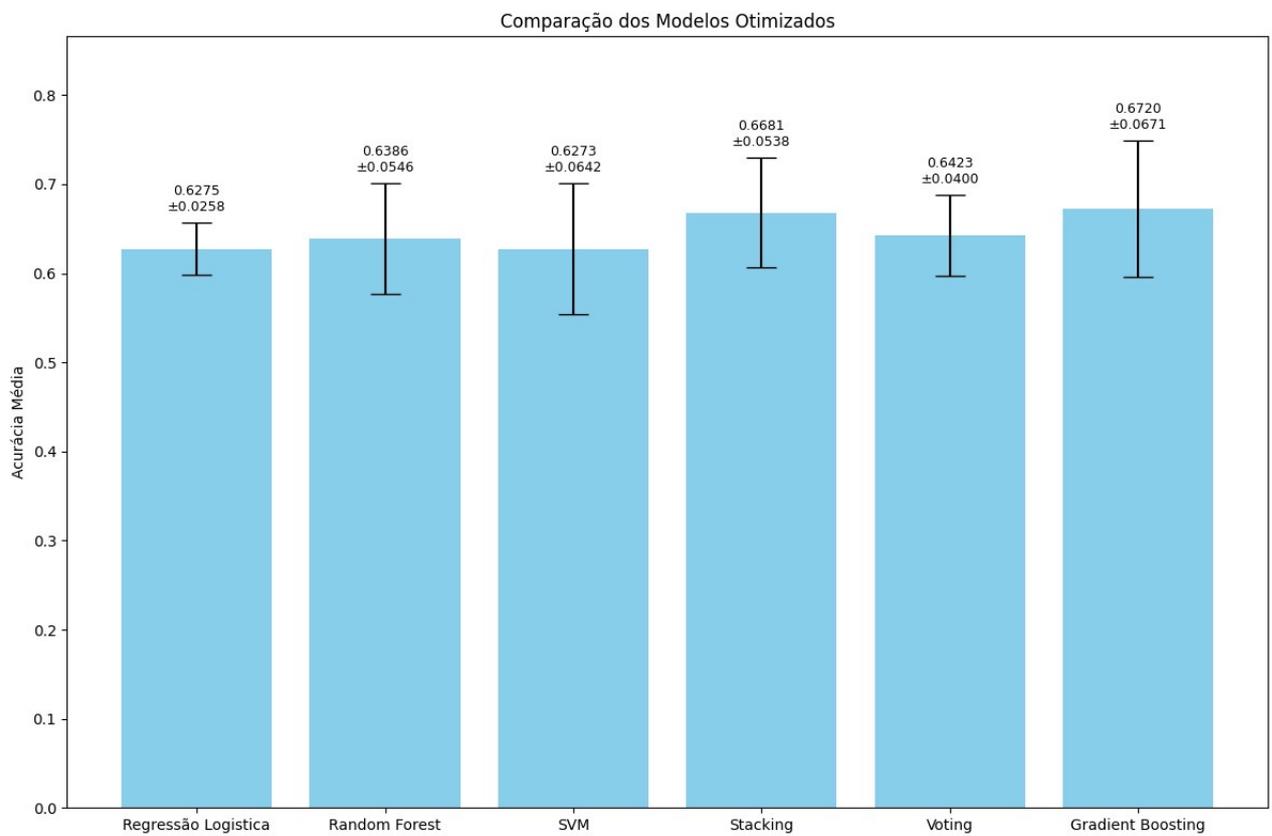
Uma árvore de decisão simples foi construída para definição dos valores dos principais parâmetros a serem trabalhados industrialmente. Dessa forma foi possível obter uma receita otimizada a ser seguida para alcançar baixos patamares de ocorrência dos defeito lascas de aquecimento.

Após a construção da árvore foi realizada uma análise com base do banco de dados utilizados, utilizando os critérios definidos na árvore de decisão

4 Resultados

Os resultados de acurácia obtidos a partir dos modelos utilizados são mostrados na Figura 10. Foi constatado que os modelos testados possuem capacidades estatisticamente iguais de prever a ocorrência de lascas de aquecimento, com resultados de acurácia variando entre 62,75% e 67,2%. Apesar da tendência similar dos modelos, o que mostrou-se o mais acurado, após a otimização, com o resultado de 67,20%, foi o modelo *Gradient Boosting*.

Figura 10 – Valores de acurácia obtidos de acordo com os modelos utilizados.



Fonte: Elaborado pelos autores.

Para melhorar a interpretabilidade do modelo, a estratégia adotada foi a utilização da biblioteca shap do python que se baseia no SHAP para explicar o modelo *Gradient Boosting*, que apresentou maior acurácia. A análise forneceu compreensão mais detalhada da influência das características nas previsões dos modelos, conforme mostrado na Figura 11.

Uma vez que o objetivo do modelo é prever a ocorrência de lascas, a classe positiva é a presença de lascas, e valores SHAP positivos indicam fatores que aumentam a probabilidade de sua ocorrência.

No eixo horizontal é mostrado o impacto da variável na previsão do modelo. Uma vez que o objetivo do modelo é prever a ocorrência de lascas, a classe positiva é a presença de lascas, e valores SHAP positivos indicam fatores que aumentam a probabilidade de sua ocorrência, enquanto que valores negativos indicam um menor probabilidade de ocorrência do defeito. Já no eixo vertical, as variáveis são organizadas por importância, onde a variável no topo é a mais importante, com a importância decrescendo à medida que se move para baixo no gráfico. Os pontos no gráfico são coloridos de acordo com o valor da variável para uma dada observação. Por exemplo, no esquema da Figura 11 foi utilizada a cor azul para valores baixos e vermelho para valores altos das variáveis listadas. Isso ajuda a identificar se valores altos ou baixos de uma variável estão contribuindo para aumentar ou diminuir a previsão do modelo.

Figura 11 – Resultado da análise SHAP aplicada ao modelo *Gradient Boosting*.



Fonte: Elaborado pelos autores.

O foco da discussão será relativo às variáveis que ocupam desde a primeira até a terceira posição da análise **SHAP**, em função da melhor separação e densidade de pontos de cores distintas, sendo elas: teor de titânio, teor de enxofre e espessura.

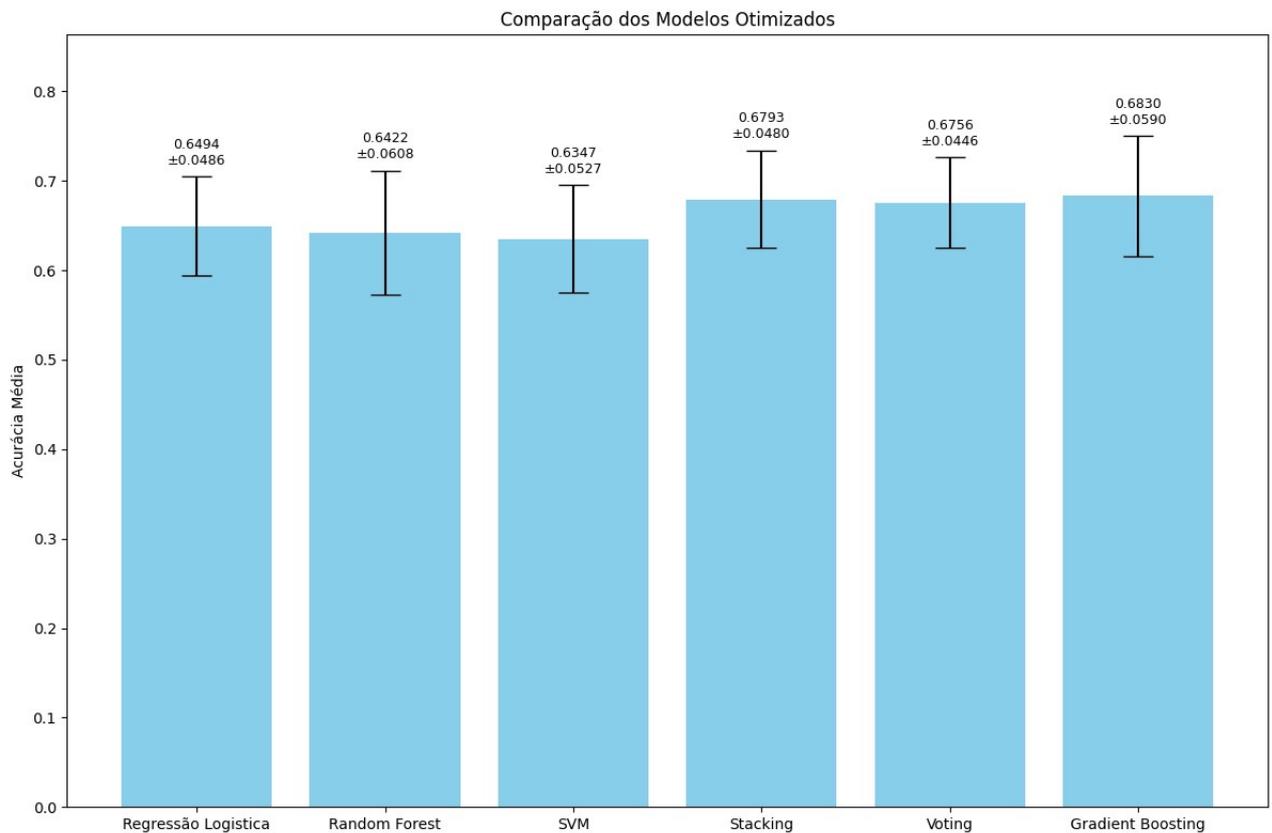
A análise mostrou que a variável com maior efeito na ocorrência do defeito lascas de aquecimento é o teor do elemento titânio. Logo, valores mais elevados de titânio (valores em vermelho) reduzem a probabilidade de ocorrência do defeito lascas de aquecimento. E a redução do teor de titânio, por sua vez, aumenta a probabilidade de ocorrência do defeito. De acordo com a literatura, o titânio tem influência na ductilidade a quente do aço, sendo que o efeito positivo da adição de Ti pode ser atribuído à formação de fosfetos de titânio que apresentam uma temperatura de ponto de fusão com a austenita (γ) próximo a 1280°C comparado a 1085°C para demais fosfetos (M3P). Por outro lado, uma adição excessiva de titânio ($> 0,2\%$) é considerada prejudicial quanto surgimento de trincas a quente devido à formação de compostos eutéticos de baixo ponto de fusão enriquecidos em Ti ((**MATSUDA; KATAYAMA; ARATA, 1983**), (**FOURMENTIN, 2009**)). Em função dos valores de titânio, em geral, serem relativamente baixos, não ultrapassando 0,035%, somente o efeito positivo desse elemento foi percebido.

A segunda variável mais relevante indicada pela análise SHAP foi o teor do elemento enxofre. Ao contrário do observado em relação ao teor de titânio, quanto menor o teor de enxofre menor a probabilidade da ocorrência do defeito lascas de aquecimento. O S é um elemento residual, de difícil remoção e que tem influência na fragilidade a quente do aço. De acordo com **Santana et al. (2006)**, a presença de S no aço permite a formação de compostos com baixo ponto de fusão, como, por exemplo, o sulfeto de níquel (NiS). Durante o reaquecimento de placas esses compostos podem ser eliminados, formando vazios internos na estrutura do material, se revelando após o processamento a frio na forma do defeito lascas de aquecimento.

A variável espessura ocupa a terceira posição da análise **SHAP**, mostrando que quanto menor a espessura do material laminado a frio, maior será a ocorrência do defeito. Essa relação pode ser explicada em função do defeito se tratar de uma micro trinca presente na estrutura interna da placa. Em termos práticos, quanto maior a deformação do material, ou seja, quanto menor a espessura final do produto, maior será a probabilidade de que a micro trinca em questão seja revelada. A espessura, no entanto, se trata de uma característica do produto e pode ser considerada como uma contramedida, não sendo a causa fundamental de geração do defeito. Em função disso o banco de dados foi reavaliado com intuito de desconsiderar variáveis que não sejam relacionadas ao processo produtivo na Aciaria e duas variáveis foram desconsideradas para geração dos modelos: espessura e peso.

Dessa forma, a acurácia dos modelos foi novamente verificada gerando os resultados mostrados na Figura 12. Houve pequena alteração dos valores de acurácia encontrados anteriormente, com uma sutil melhora dos resultados obtidos. Os modelos continuaram apresentando capacidades estatisticamente iguais de prever a ocorrência de lascas de aquecimento, com resultados de acurácia variando entre 63,47% e 68,30%. O modelo *Gradient Boosting* continuou apresentando o melhor resultado quando comparado aos demais, com o valor de acurácia de 68,30%.

Figura 12 – Reavaliação dos Valores de acurácia obtidos de acordo com os modelos utilizados.



Fonte: Elaborado pelos autores.

Em seguida uma nova análise *SHAP* foi gerada, conforme mostrado na Figura 13, na qual foi possível observar que os elementos titânio e enxofre se mantiveram nas mesmas posição da análise anterior. Em terceiro lugar, como o próximo elemento mais relevante, foi listado o molibdênio.

A análise mostrou que, assim como para o titânio, valores mais elevados de molibdênio (valores em vermelho) reduzem a probabilidade de ocorrência do defeito lascas de aquecimento. Apesar do molibdênio reduzir a ductilidade a quente do 304, mostrado no tópico 2.2.3, a contribuição positiva para os aços duplex pode ser relacionada ao fato de ser um elemento ferrizante. Conforme mostrado na Equação 2.1, o molibdênio contribui para o aumento da formação da fase ferrita nos aços da série 3XX. Dessa forma, a solubilidade de elementos contaminantes, tais como S e P é aumentada, reduzindo assim a probabilidade de formação de trincas durante solidificação do aço no lingotamento contínuo.

Figura 13 – Resultado da análise SHAP final, aplicada ao modelo *Gradient Boosting*.

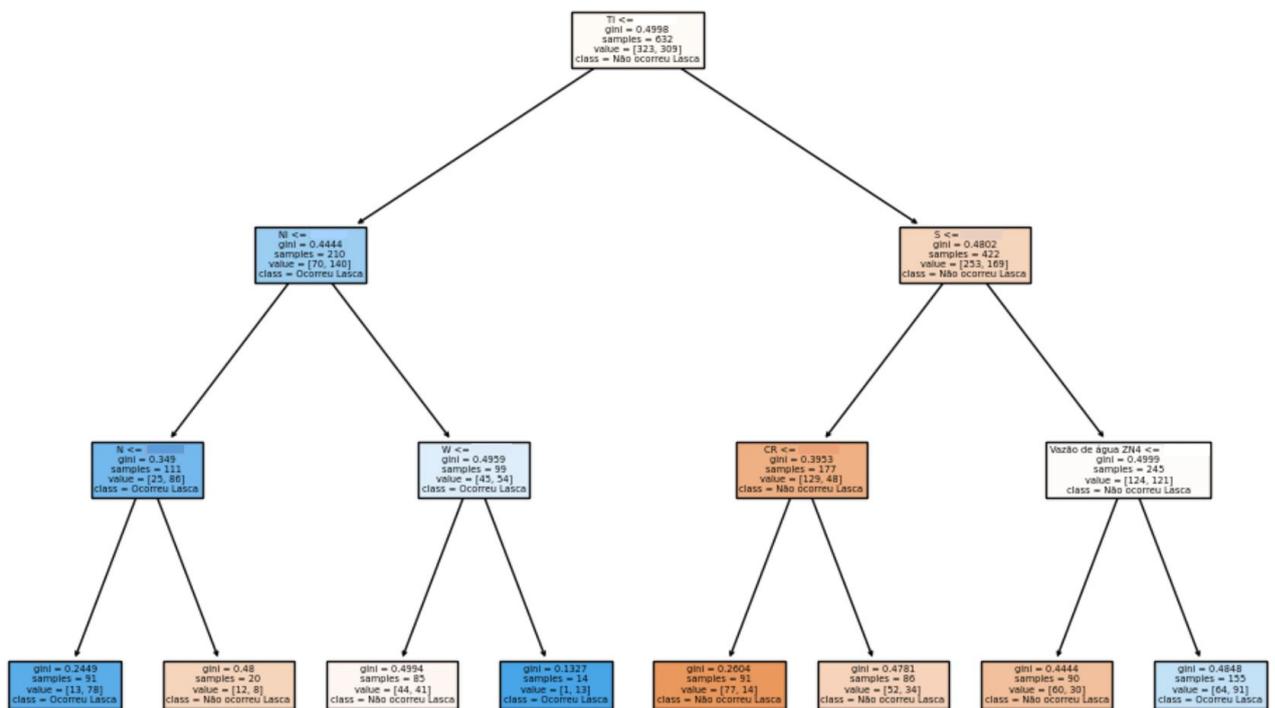


Fonte: Elaborado pelos autores.

Com o intuito de direcionar a otimização de receitas a serem buscadas ou evitadas no processo de aciaria de forma a minimizar a ocorrência de lascas, uma árvore de decisão simples foi gerada a partir dos dados otimizados, mostrado na Figura 14. As cores na árvore de decisão representam a classe predominante em cada nó ou a proporção de classes em cada nó, sendo:

- nós azuis: indicam que a maioria das amostras nesse nó pertence à classe "ocorreu lasca". Quanto mais escura a cor azul, maior a proporção de amostras desta classe no nó;
- nós vermelhos: indicam que a maioria das amostras nesse nó pertence à classe "não ocorreu lasca". Quanto mais escura a cor vermelha, maior a proporção de amostras desta classe no nó;
- nós brancos ou de cores leves: se um nó é branco ou tem uma cor mais clara, isso sugere que as amostras estão mais equilibradas entre as duas classes.

Figura 14 – Árvore de decisão aplicada ao banco de dados otimizado.



Fonte: Elaborado pelos autores.

A acurácia obtida pela árvore de decisão foi de aproximadamente 60,9%. Os valores resultantes de cada nó da árvore de decisão, referentes às variáveis de aciaria, foram omitidos em função de se tratar de informação sigilosa e será abordado no decorrer do trabalho de forma figurada, sendo: teores de Ti, S, Cr e Ni correspondente a "x", "y", "z" e "w", respectivamente.

Considerando-se a classe "não ocorreu lasca", atendendo as condições das variáveis Ti e S propostas é alcançada uma condição em que das 177 amostras, 129 não apresentaram lascas e 48 tiveram o defeito. Isso representa 27,11% de ocorrência. Enquanto que seguindo o caminho oposto, combinando as variáveis Ti e Ni, de um total de 111 amostras, 86 apresentaram o defeito e apenas 25 não tiveram lascas, resultado em 77,48% de ocorrências.

A fim de validar as receitas propostas pela árvore de decisão, o banco de dados foi testado, quanto à ocorrência de lascas, para duas condições:

- condição 1: $Ti > x$ ppm, $S \leq y$ ppm;
- condição 2: $Ti > x$ ppm, $S \leq y$ ppm, $Cr \leq z$ %;

As tabelas 4 e 5 resumizam os resultados da validação do banco de dados, com o número de amostras que apresentam ou não lascas de aquecimento e o percentual de ocorrências para cada caso. Quando os elementos Ti e S atendem, simultaneamente, a condição proposta pela árvore de decisão o percentual de ocorrência de lascas é de 32,0%. Quando a condição não é atendida há um incremento significativo das ocorrências do defeito, chegando a 77,6% (Tabela 4).

Tabela 4 – Ocorrência de lascas para a condição 1 (Ti e S).

Atendimento a condição 1	Com Ocorrência	Sem Ocorrência	% Ocorrência
Atendeu	129	274	32,0%
Não Atendeu	59	17	77,6%

Realizando a mesma análise para a condição 2, na qual o teor de Cr também é considerado, o percentual de ocorrências de lascas de aquecimento, quando a condição proposta é atendida, permanece próximo ao encontrado para a condição 1, com o valor de 32,7%. Quando a condição 2 não foi atendida, as ocorrências do defeito aumentam ainda mais, chegando a 93,8% (Tabela 5).

Tabela 5 – Ocorrência de lascas para a condição 2 (Ti, S e Cr).

Atendimento a Condição 2	Com Ocorrência	Sem Ocorrência	% Ocorrência
Atendeu	85	175	32,7%
Não Atendeu	15	1	93,8%

Em função dos resultados de ocorrência de lascas terem sido similares, quando as condições 1 e 2 foram atendidas, a atuação no processo de Aciaria para controle do defeito pode ser direcionada ao acompanhamento dos teores de Ti e S. Dessa forma a complexidade do processo produtivo é reduzida, e o impacto no defeito será mínimo.

5 Considerações Finais

Os modelos de predição utilizados apresentaram boa aderência à ocorrência do defeito lascas de aquecimento, com os valores de acurácia 63,47% e 68,30%. O modelo *Gradient Boosting* apresentou melhor desempenho, alcançando o valor de acurácia 68,30%, sendo este resultado considerado significativo para aplicação industrial.

A análise *SHAP*, permitiu elencar as variáveis mais significativas para ocorrência do defeito lascas de aquecimento, dentre as quais as mais relevantes foram: teor de titânio, teor de enxofre e teor de molibdênio.

A relação da ocorrência de lascas em função dos teores de titânio, enxofre e molibdênio corroboram com a literatura, sendo correlacionados com os mecanismos que afetam a ductilidade a quente e fragilidade a quente, respectivamente.

A árvore de decisão obtida, trouxe um direcionamento de uma receita a ser utilizada para minimizar a ocorrência do defeito lascas, utilizando-se um teor de $Ti > x$ ppm e $S \leq y$ ppm. Ademais, um caminho a ser evitado também foi evidenciado: $Ti \leq x$ associado ao $Ni \leq w\%$.

A aplicação de ciência de dados foi fundamental para o sucesso do projeto, permitindo uma análise mais detalhada e precisa dos dados obtidos.

A pesquisa em si, não foi ainda esgotada e estudos complementares poderão trazer resultados ainda mais promissores, como por exemplo, a utilização de dados referentes ao processamento a quente dos aços inoxidáveis duplex. Diante deste contexto, seguem algumas sugestões de dados que podem ser estudados em trabalhos futuros:

- Reaquecimento de placas, envolvendo dados de tempo e temperaturas de processo;
- Laminação de desbaste, considerando dados de tempo, temperatura e redução por passe.
- Laminação de acabamento, similarmente ao anterior, com dados de tempo, temperatura e redução por passe e total.
- Aplicação da metodologia à outros tipos de defeitos de grande complexidade na Aperam.

Referências

- AGRESTI, A.; TARANTOLA, C. Simple ways to interpret effects in modeling ordinal categorical data. **Statistica Neerlandica**, Wiley Online Library, v. 72, n. 3, p. 210–223, 2018.
- ALMIRON, C. J. Estudo preliminar comparativo do efeito da espessura no comportamento em fadiga de arames de armadura de tração de risers flexíveis. 2010.
- ASTM, A. A. M. **Standard Specification for Chromium and Chromium-Nickel Stainless Steel Plate, Sheet, and Strip for Pressure Vessels and for General Applications**. [S.l.]: ASTM International West Conshohocken, PA, 2022.
- BARBOSA, R. R. *et al.* **Aços inoxidáveis duplex laminados a frio: uma alternativa para o pré sal**. [S.l.], 2015.
- BISHOP, C. M. Model-based machine learning. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, The Royal Society Publishing, v. 371, n. 1984, p. 20120222, 2013.
- BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [S.l.]: Springer, 2006. v. 4.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.
- COLLOBERT, R. Deep learning for efficient discriminative parsing. In: JMLR WORKSHOP AND CONFERENCE PROCEEDINGS. **Proceedings of the fourteenth international conference on artificial intelligence and statistics**. [S.l.], 2011. p. 224–232.
- CORTES, C.; VAPNIK, V. Support vector machine. **Machine learning**, v. 20, n. 3, p. 273–297, 1995.
- COX, D. R. The regression analysis of binary sequences. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, Oxford University Press, v. 20, n. 2, p. 215–232, 1958.
- CRISTIANINI, N.; SHAWE-TAYLOR, J. **An introduction to support vector machines and other kernel-based learning methods**. [S.l.]: Cambridge university press, 2000.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. **International workshop on multiple classifier systems**. [S.l.], 2000. p. 1–15.
- FIRTH, D. Bias reduction of maximum likelihood estimates. **Biometrika**, Oxford University Press, v. 80, n. 1, p. 27–38, 1993.
- FOURMENTIN, R. **Industrialization of 310S: literature review on the influence of alloying elements on hotcracking susceptibility**. [S.l.], 2009.
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, JSTOR, p. 1189–1232, 2001.
- GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for bayesian models. **Statistics and computing**, Springer, v. 24, n. 6, p. 997–1016, 2014.

- IMOA, I. M. A. *et al.* **Orientações práticas para processamento dos aços inoxidáveis duplex.** [S.l.]: Londres, 2009.
- KITTLER, J. *et al.* On combining classifiers. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 20, n. 3, p. 226–239, 1998.
- LACOMBE, P.; BAROUX, B.; BERANGER. **Stainless steels.** [S.l.]: Les Editions de Physique, Les Ulis, France, 1993.
- LIAW, A.; WIENER, M. *et al.* Classification and regression by randomforest. **R news**, v. 2, n. 3, p. 18–22, 2002.
- MATSUDA, F.; KATAYAMA, S.; ARATA, Y. Solidification crack susceptibility in weld metals of fully austenitic stainless steels (report ix). 1983.
- NORVIG, P.; RUSSELL, S. Inteligência artificial. **Rio de Janeiro: Grupo GEN**, 2013.
- OPITZ, D.; MACLIN, R. Popular ensemble methods: An empirical study. **Journal of artificial intelligence research**, v. 11, p. 169–198, 1999.
- PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011.
- PLAUT, R. L. *et al.* A short review on wrought austenitic stainless steels at high temperatures: processing, microstructure, properties and performance. **Materials Research**, SciELO Brasil, v. 10, p. 453–460, 2007.
- QUINLAN, J. R. Induction of decision trees. **Machine learning**, Springer, v. 1, p. 81–106, 1986.
- SANTANA, C. *et al.* **Estudo e combate ao defeito lasca de aquecimento nos aços inoxidáveis 3XX.** [S.l.], 2006.
- SILVA, A. L.; MEI, P. R. **Aços e ligas especiais.** [S.l.]: Edgard Blucher, 2010.
- SILVA, R. C. R. *et al.* **Catálogo de Defeitos dos Aços Inoxidáveis. Metalurgia de Aços Inoxidáveis.** [S.l.], 2002.
- STROBL, C. *et al.* Conditional variable importance for random forests. **BMC bioinformatics**, Springer, v. 9, p. 1–11, 2008.
- VIANA, A. K. d. N. *et al.* **Redução dos defeitos lasca de aquecimento e esfoliação nos aços inoxidáveis duplex.** [S.l.], 2015.
- WOLPERT, D. H. Stacked generalization. **Neural networks**, Elsevier, v. 5, n. 2, p. 241–259, 1992.