

**MODELAGEM QUÍMICA DO
AÇO INOXIDÁVEL DUPLEX RESISTENTE À
FORMAÇÃO DE LASCAS DE AQUECIMENTO**

VICTOR BERNARDO RODRIGUES JORGE

**MODELAGEM QUÍMICA DO
AÇO INOXIDÁVEL DUPLEX RESISTENTE À
FORMAÇÃO DE LASCAS DE AQUECIMENTO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: ADRIANO A. VELOSO

Belo Horizonte

Julho de 2021

© 2021, Victor Bernardo Rodrigues Jorge.
Todos os direitos reservados.

Rodrigues Jorge, Victor Bernardo

D1234p Modelagem química do aço inoxidável duplex
resistente à formação de lascas de aquecimento /
Victor Bernardo Rodrigues Jorge. — Belo Horizonte,
2021

xviii, 37 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de
Minas Gerais

Orientador: Adriano A. Veloso

1. Computação — Teses. 2. Aprendizado de
máquina. 3. Indústria 4.0 I. Orientador. II. Título.

CDU 519.6*82(043)

[Folha de Aprovação]

Quando a secretaria do Curso fornecer esta folha,
ela deve ser digitalizada e armazenada no disco em formato gráfico.

Se você estiver usando o `pdflatex`,
armazene o arquivo preferencialmente em formato PNG
(o formato JPEG é pior neste caso).

Se você estiver usando o `latex` (não o `pdflatex`),
terá que converter o arquivo gráfico para o formato EPS.

Em seguida, acrescente a opção `approval={nome do arquivo}`
ao comando `\ppgccufmg`.

Se a imagem da folha de aprovação precisar ser ajustada, use:
`approval=[ajuste] [escala] {nome do arquivo}`
onde *ajuste* é uma distância para deslocar a imagem para baixo
e *escala* é um fator de escala para a imagem. Por exemplo:
`approval=[-2cm] [0.9] {nome do arquivo}`
desloca a imagem 2cm para cima e a escala em 90%.

Aos amigos e à família que sempre me apoiaram em toda minha jornada.

“A tecnologia ensinou uma lição à humanidade: nada é impossível”
(Lewis Mumford)

Resumo

A exploração das reservas de petróleo da camada do pré-sal requer o desenvolvimento de novos materiais de aço que sejam resistentes às condições extremamente adversas presentes em águas profundas. Nosso objetivo é projetar aço inoxidável duplex encontrando novos padrões de composição química que atinjam de forma otimizada padrões metalúrgicos desejáveis, permitindo assim a produção de tubos submarinos com alta resistência física e à corrosão. Um desafio particular no desenvolvimento do aço duplex para aplicações em águas profundas é a minimização de defeitos superficiais, como as lascas de aquecimento. Esses defeitos podem aumentar significativamente os custos de produção, porque permanecem indetectados em estágios intermediários da fabricação, sendo observados apenas durante a inspeção final do produto acabado. As lascas podem ser originadas em diferentes estágios do processo de fabricação de aço, mas um importante fator causal das lascas é conhecido por estar relacionado à uma composição química defeituosa. A fim de identificar o mecanismo de formação de lascas em aço inoxidável duplex e encontrar fatores químicos associados à sua formação, propomos uma abordagem de decomposição de atributos que resultou em centenas de milhares de modelos preditivos aprendidos a partir das composições químicas de placas de aço tanto defeituosas quanto não defeituosas, oferecendo diversos modelos e explicações concorrentes para a incidência de lascas no aço duplex. Combinamos modelos com explicações concorrentes para encontrar um conjunto diversificado de padrões de composição química que deveriam levar a chances reduzidas de formação de lascas. Empregamos esses padrões como possíveis ações de melhoria, gerando placas que combinam com esses padrões de composição química.

Palavras-chave: Aço inoxidável, Projeto de materiais, Aprendizado de máquina, Explicabilidade, Indústria 4.0, Análise de causa raiz.

Abstract

Exploring sub-salt oil reservoirs requires the development of novel steel materials that withstand the extremely adverse conditions present in deep water. Our goal is to design duplex stainless steel by finding novel chemical composition patterns that optimally achieve various measurable metallurgical desiderata, thus enabling the production of sub-sea tubes with high strength and corrosion resistance. A particular challenge while designing duplex steel for deep water applications is the minimization of surface defects such as slivers, since these defects may significantly increase production costs as they remain undetected in intermediate processing stages, being observed only during the final inspection. Slivers may be originated in different stages of the steelmaking process, but one important causative factor of slivers is known to be related to defective chemical composition. In order to identify the formation mechanism of slivers in duplex stainless steel, and to find chemical factors associated with sliver formation, we propose a feature decomposition approach that resulted in hundreds of thousands of predictive models learned from the chemical compositions of defective and non-defective duplex plates, thus offering diverse models and competing explanations for the incidence of slivers in duplex steel. We combined models with competing explanations in order to find a diversified set of chemical composition patterns that should lead to reduced chances of sliver formation. We employ these patterns as possible ameliorative actions, generating plates matching these chemical composition patterns.

Keywords: Stainless Steel, Defect Model, Material Design, Machine Learning, Explainability, Root Cause Analysis.

Lista de Figuras

1.1	Cima — Placa de aço inoxidável duplex com uma lasca perto de sua borda. Baixo — Imagem amplificada de uma lasca de aquecimento.	3
2.1	Visão geral do processo de fabricação do aço	11
5.1	T-SNE [van der Maaten, 2009] do espaço amostral de modelos \mathcal{H}' . Cada ponto representa um modelo \mathbf{x}' . Os modelos são posicionados de acordo com as probabilidades de defeito atribuídas a cada placa de aço, de forma que os modelos que atribuem probabilidades semelhantes às mesmas placas sejam colocados mais próximos no espaço. A cor indica o cluster para o qual o modelo foi atribuído.	26
5.2	Análise de silhueta para cada cluster. O valor médio da silhueta sugere coesão e separação.	27
5.3	Distribuição dos valores de AUC em cada cluster. Os modelos performáticos estão distribuídos pelos clusters.	28
5.4	Número de modelos em que cada característica apareceu.	29
5.5	Representação distribuída de cada feature.	29
5.6	Modelos representativos para alguns clusters. Esquerda — Curva ROC mostrando o desempenho do modelo. Direita — O <i>summary plot</i> SHAP correspondente mostrando uma visão geral de quais atributos são mais importantes para o modelo.	30

Sumário

Resumo	xi
Abstract	xiii
Lista de Figuras	xv
1 Introdução	1
1.1 Motivação	2
1.2 Definição do problema	4
1.3 Contribuições	5
1.4 Organização da dissertação	5
2 Processo de fabricação do aço	7
2.1 Fundição inicial	7
2.2 Ajuste de composição química	8
2.3 Lingotamento contínuo	9
2.4 Laminação a quente no moínho Steckel	9
2.5 Inspeção visual	10
3 Referencial teórico	13
3.1 Análise de causa raiz	13
3.2 Modelos de classificação	14
3.3 Interpretabilidade dos modelos	15
3.4 Aprendizado de máquina na siderurgia	17
4 Metodologia	19
4.1 Caracterização do conjunto de dados	19
4.2 Encontrando padrões defeituosos na composição química do aço	20
4.2.1 Estruturas locais nos dados	20

4.2.2	Amostragem do Espaço do Modelos	21
4.2.3	Representando as preferências do modelo	21
4.2.4	Representando as explicações do modelo	22
4.2.5	Prevenindo a formação de lascas evitando padrões defeituosos e impondo padrões não defeituosos na composição química	22
5	Experimentos	25
5.1	Configuração dos experimentos	25
5.2	Resultados	26
5.3	Interpretabilidade dos modelos	29
6	Conclusões e trabalhos futuros	33
	Referências Bibliográficas	35

Capítulo 1

Introdução

O aço inoxidável ou aço inox é uma liga de ferro e cromo que apresenta propriedades físico-químicas superiores aos aços comuns, tendo a alta resistência à oxidação atmosférica como sua principal característica. Os aços inoxidáveis podem ser classificados em diferentes tipos/famílias de acordo com a sua microestrutura cristalina [Deo & Boom, 1993]. As principais famílias de aço inox são:

- Austeníticos: os aços inoxidáveis austeníticos tendem a ter um alto teor de cromo em comparação com outras ligas de aço, conferindo-lhes maior resistência à corrosão. Outra característica comum das ligas de aço inoxidável austenítico é que elas tendem a ser não magnéticas.
- Ferríticos: o aço inoxidável ferrítico é magnético, como sugere o nome. Essas ligas podem ser endurecidas através de encruamento. Eles também tendem a ser mais baratos em relação aos demais devido ao seu baixo teor de níquel.
- Martensíticos: sua resistência à corrosão tende a ser menor do que as ligas ferríticas ou austeníticas, mas eles têm uma alta dureza. As ligas de aço inoxidável martensítico são geralmente ideais para aplicações que requerem uma resistência à tração e ao impacto extraordinariamente alta. Quando as referidas aplicações também requerem resistência à corrosão, essas ligas podem ser usadas com um revestimento de polímero protetor.
- Duplex: possui uma microestrutura austenita e ferrita ao mesmo tempo, geralmente numa proporção igual. São caracterizados por conterem mais cromo e molibdênio e menores teores de níquel do que os aços inoxidáveis austeníticos. Têm aproximadamente o dobro da força de rendimento (resistência à dobra) do aço inoxidável austenítico. Sua microestrutura mista oferece resistência melhorada

à corrosão sob tensão por cloro em comparação com os tipos de aço inoxidável austeníticos.

Em particular, o aço inox da família duplex tem encontrado uso crescente na indústria marítima, principalmente por oferecer um ótimo custo benefício entre resistência física e resistência à corrosão. As primeiras aplicações das ligas de aço inoxidável duplex foram quase exclusivamente os tubos dissipadores de calor, principalmente em serviços de resfriamento que utilizam materiais corrosivos. Atualmente, essas ligas têm uma grande variedade de aplicações potenciais como, por exemplo, a construção de dutos para a exploração de petróleo na camada do pré-sal. O aço inoxidável duplex mais comumente usado hoje em exploração de águas profundas contém aproximadamente 22% de cromo (Cr), contendo também molibdênio (Mo) e nitrogênio (N). Outros elementos presente neste material são: níquel (Ni), cobre (Cu), manganês (Mn) e silício (Si). Estes elementos devem ocorrer em certas proporções especificadas para que o material acabado seja caracterizado como um aço inox do tipo duplex [Gunn, 1997].

1.1 Motivação

A qualidade dessa família de aços é frequentemente ameaçada pela presença de defeitos em sua superfície. Dentre esses defeitos podemos destacar as lascas. A lasca é um defeito superficial que supostamente é originado durante os estágios de solidificação do processo de fundição [Stradomska et al., 2009]. Elas são alongadas na mesma direção da laminação, com comprimento usual de 70mm, e geralmente se concentram mais próximos às bordas das placas de aço [Thomas, 2006], conforme mostrado na Figura 1.1. As lascas causam um aumento nos custos de produção, pois permanecem indetectadas nas etapas intermediárias da fabricação, sendo observadas apenas durante a inspeção final do produto acabado. Uma vez que as lascas são observadas, a placa defeituosa geralmente é descartada. Portanto, a modelagem de novas ligas de aço que são menos suscetíveis à formação de lascas é de suma importância para a eficiência da indústria de aço.

A APERAM South America é a maior fabricante de aços inoxidáveis do Brasil, com uma fatia de mercado de aço estimada em 90% no país. Por ser a única fabricante de aço inoxidável duplex no Brasil, a APERAM South America está expandindo sua presença no segmento de exploração de petróleo do pré-sal com o desenvolvimento de ligas de aço inoxidável ultra-resistentes. Esses novos materiais são impostos a altos padrões qualidade para resistir às condições ambientais extremamente adversas encontradas em águas profundas, e um grande desafio para atingir tais padrões de qualidade

exigidos é prevenção do surgimento de lascas. Por isso, a APERAM South America solicitou uma investigação sobre as possíveis causas do aparecimento de lascas de aquecimento no aço inoxidável duplex. Através da espectrometria de massas foram aferidas as composições químicas de 122 placas de aço inoxidável duplex produzidas durante o ano de 2018 na planta de Timóteo-MG. Das 122 corridas estudadas, 71 resultaram em placas defeituosas enquanto as outras 51 não apresentaram lascas durante a inspeção final. Essas medições correspondendem a um conjunto de dados de escala inédita para o estudo desse defeito.

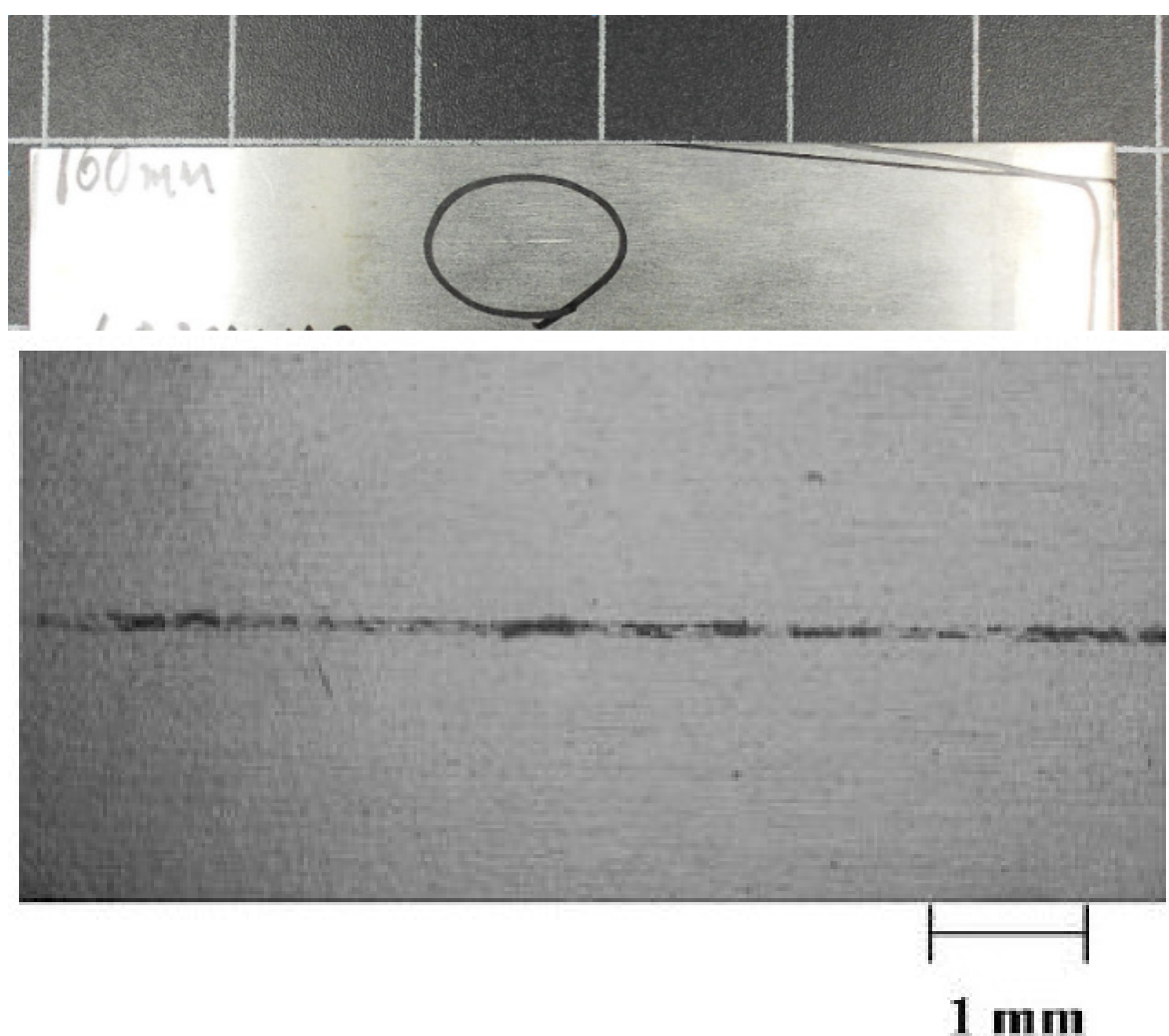


Figura 1.1. Cima — Placa de aço inoxidável duplex com uma lasca perto de sua borda. Baixo — Imagem ampliada de uma lasca de aquecimento.

1.2 Definição do problema

Segundo Barbosa et al. [2007], a determinação dos fatores causais associados à formação das lascas não é trivial, pois elas podem estar associadas a uma combinação de variáveis metalúrgicas e de processo. Variáveis, essas, que são aferidas em diferentes etapas do processo metalúrgico. Portanto, uma determinada composição química pode estar associada ou não à formação de lasca, a depender de variáveis de processo relacionadas a operações posteriores, como reaquecimento e laminação. Nosso principal objetivo neste trabalho é determinar a inter-relação entre a formação da lasca e a composição química do aço inoxidável duplex. Uma vez que os padrões nas composições químicas são encontrados, eles podem ser usados como ações de melhoria a fim de prevenir a formação de lascas.

A intuição que exploraremos neste estudo é: pode existir uma receita química muito específica que é menos suscetível à formação de lascas, desde que as variáveis dos processos posteriores assumam seus valores típicos. No entanto, como estamos nos concentrando apenas nas composições químicas, os dados correspondentes podem apresentar muitas estruturas locais, no sentido de que os fatores químicos que contribuem para a formação da lasca podem variar muito dependendo de variáveis de processo desconhecidas relacionadas às etapas posteriores.

Intuitivamente, se diferentes instâncias - ou seja, placas de aço - estão associados a diferentes estruturas locais nos dados, esperaríamos que cada estrutura fosse melhor descrita por um modelo diferente. Então, podemos obter padrões globais de composição química combinando modelos para todos os tipos de estruturas locais encontrados. Nesse caso, em vez de modelar os dados usando a abordagem única que se ajusta a todos os fatores (ou características) disponíveis em um único modelo, amostramos o espaço de modelos decompondo as características disponíveis. Isso resulta em centenas de milhares de modelos aprendidos a partir das composições químicas. Embora a grande maioria dos modelos seja ineficaz, fomos capazes de produzir vários modelos que geram previsões altamente precisas. Para esses modelos que performam bem, encontramos uma forte ligação entre as características e as previsões do modelo. Mostrando, assim, que algumas características são claramente úteis para detectar diferentes mecanismos de formação de lascas e evidenciando, portanto, que diferentes combinações de características cobrem uma região específica do espaço relativo aos defeitos.

1.3 Contribuições

Existem na literatura diversas diretrizes e guias que relacionam componentes químicos com a ocorrência de lascas no aço inoxidável duplex, como em [Stradomska et al., 2009]. Porém, até onde vai nosso conhecimento, ainda não foi realizada uma análise profunda do problema usando um conjunto de dados exclusivo e de grande escala que considera todo o espectro de elementos químicos presentes no aço. Além disso, a inovação deste trabalho está no processo proposto para se identificar diversas causas raiz que podem ser aplicadas a qualquer tarefa onde pode haver mais de um fator subjacente. As contribuições específicas deste trabalho podem ser resumidas da seguinte forma:

- Dentre nossos principais resultados, destacamos que nossos modelos preditivos são empiricamente precisos para estimar se uma placa de aço qualquer será defeituosa, com um valor de AUC que varia de 0,78 a 0,85. Isso pode ser considerado um resultado impressionante, já que nossos modelos preditivos não levam em consideração as variáveis de processo obtidas nas demais etapas da siderurgia. Ser capaz de antecipar defeitos logo nas primeiras etapas do processo metalúrgico é de grande importância para a redução dos custos operacionais.
- Uma ferramenta implantada e confiável que auxilia a modelagem química de aços inoxidáveis duplex para evitar a formação de lascas. Nossa ferramenta utiliza um grande número de modelos preditivos para encontrar um conjunto diversificado de padrões químicos associados a placas não defeituosas. A siderúrgica produz placas com composição química condizente com os padrões não defeituosos. A formação de lascas também é prevenida ao se evitar padrões defeituosos.

1.4 Organização da dissertação

Esta dissertação está organizada em 5 capítulos além do atual. No Capítulo 2 descrevemos como ocorre a produção de aço inoxidável duplex na planta de Timóteo-MG da APERAM South America. No capítulo 3 é apresentada a sumarização de diversos trabalhos relacionados ao nosso problema. O capítulo 4 descreve, em termos matemáticos, a metodologia adotada em nosso trabalho para se alcançar resultados que apontem caminhos para a melhoria no processo de produção do aço. O capítulo 5 apresenta a configuração e os resultados obtidos em nossos experimentos. O sexto e último capítulo apresenta conclusões e sugestões para futuros trabalhos nesse tema.

Capítulo 2

Processo de fabricação do aço

Neste capítulo são apresentados os conceitos, métodos e procedimentos que permeiam este trabalho. O objetivo é introduzir o conhecimento necessário para compreender o processo em que estamos inseridos e as ferramentas utilizadas. Detalha-se, principalmente, como é realizado todo o processo de fabricação do aço inox do tipo duplex na APERAM South America.

A siderurgia é o processo de produzir aço a partir de minério de ferro e sucata. Esse processo envolve a remoção de impurezas do ferro bruto e a adição de elementos ligamentais como manganês, níquel, cromo e vanádio para produzir diferentes tipos de aço [Deo & Boom, 1993]. Descreveremos brevemente como esse processo é realizado na APERAM South America, na planta de Timóteo-MG. O processo pode ser dividido em cinco etapas fundamentais: fundição inicial, ajuste de composição química, lingotamento contínuo, laminação a quente e inspeção visual.

2.1 Fundição inicial

Nesta primeira etapa, temos materiais ferríticos primordiais vindos de duas fontes principais:

- Ferro-gusa — produto obtido diretamente do minério de ferro transformado em altos-fornos. Possui cerca de 5% de carbono em sua composição e baixa concentração de silício, manganês, fósforo e enxofre.
- Sucata de aço reciclada — materiais obtidos por aquisições externas ou sucata de outros materiais provenientes da linha de produção.

Ao final dessa etapa, o material resultante constitui uma mistura entre o ferro-gusa líquido e a sucata em qualquer proporção.

2.2 Ajuste de composição química

A etapa de ajuste de composição química envolve operações de fusão, purificação e formação de ligas realizadas a cerca de 1600 °C com o ferro derretido. Várias reações químicas são iniciadas para se alcançar as composições químicas de referência do aço inoxidável duplex. Basicamente, o processo de ajuste da composição química é feito através da adição ou remoção certos elementos e/ou manipulação da temperatura, pressão e do ambiente de produção. Este processo é composto de duas etapas detalhadas a seguir:

Etapa AOD (Argon-Oxygen-Decarburization) – O início do processo de fusão consiste em levar a mistura ao equipamento AOD, onde ocorre uma redução significativa da pressão parcial do sistema. Através de um forte fluxo de oxigênio, realiza-se uma redução drástica do teor de carbono da mistura. Toneladas de substâncias que possuem afinidade química com o carbono são complementadas, também com o objetivo de reduzir o teor de carbono.

Além do carbono, componentes como manganês, silício e fósforo também são oxidados, misturados ao óxido de ferro. Esse processo leva à formação de escória, que é um material que flutua sobre o aço fundido devido à sua menor densidade. A escória é então removida da superfície do aço e pode ser utilizada na fabricação de cimento.

Duas amostragens relevantes estão associadas à composição química no AOD. A primeira, e mais imediata, é composta pelos gases emitidos durante o processo. Ela permite que o operador da máquina verifique se todos os elementos químicos medidos estão dentro de uma faixa de referência desejada. Esse cálculo é feito através de fórmulas matemáticas predefinidas. A segunda amostragem é feita por meio de espectrômetros de massa e permite a geração de dados mais detalhados sobre sua composição química.

Há um limite para a redução do teor de carbono do aço durante a fase AOD. Por exemplo, um aumento na quantidade de oxigênio utilizada resultaria na oxidação do cromo, em vez do carbono. Isso implicaria na perda de propriedades importantes do aço. Assim, a fim de se fazer uma redução mais controlada, a mistura de aço é posteriormente processada durante a etapa de VOD, descrita a seguir.

Etapa VOD (Vacuum-Oxygen-Decarburization) – Durante a etapa de VOD, a redução da pressão parcial do sistema é muito mais intensa, atingindo-se o que é considerado um vácuo absoluto. Essas condições permitem uma redução mais controlada

do teor de carbono da mistura. Por exemplo, o vácuo aumenta a reação de descarbonização e preserva o cromo da oxidação excessiva. Essa redução também é feita através de um fluxo de oxigênio e adicionando-se substâncias com afinidade com o carbono. Enquanto no estágio AOD as adições são feitas na ordem de toneladas, no estágio VOD as adições são feitas na ordem de quilos. Para serem realizadas com precisão, as operações VOD dependem da capacidade de se mensurar a composição química do fluxo de gás que sai do forno ao longo do processo. Os sistemas de amostragem baseados no espectrômetro de massa fornecem análises precisas, representativas e em tempo real da composição química do aço que está sendo produzido. A amostragem contínua é necessária porque o estágio VOD é o processo pelo qual a especificação química do aço inoxidável duplex é garantida.

2.3 Lingotamento contínuo

Concluída a etapa de VOD, o aço refinado é reaquecido em um forno panela antes de ser levado ao lingotamento contínuo (LC2). Mais especificamente, o aço refinado é levado para moldes de resfriamento e transformado em placas semiacabadas e solidificadas. A partir deste momento, os lingotes de aço podem ser laminados de acordo com as especificações desejadas por cada cliente. Durante o processo de solidificação, uma tesoura corta os lingotes e já é possível observar o aço em comprimentos adequados para os demais processos.

2.4 Laminação a quente no moínho Steckel

A quarta etapa do processo de fabricação de aço é a laminação. O aço, durante o processo de solidificação, é conformado mecanicamente e transformado em produtos de aço inoxidável duplex utilizados pela indústria de transformação, como chapas e bobinas. Especificamente, o aço inoxidável duplex é conformado em placas e bobinas por laminação a quente usando um laminador Steckel, que é um laminador reversível com uma bobina aquecida em cada extremidade. As duas bobinas são usadas para alimentar o laminador com o material. O material é passado para frente e para trás no moínho até que uma espessura precisa seja alcançada em toda a sua superfície, bem como um nivelamento consistente seja alcançado. Nesse processo, a espessura do aço é drasticamente reduzida enquanto o seu comprimento é expandido.

2.5 Inspeção visual

Após a laminação, as placas de aço inoxidável duplex são movidas para o galpão de inspeção. A identificação e registro de eventuais defeitos que as placas possam apresentar, bem como sua localização e extensão, são feitos por especialistas treinados [Zhao et al., 2017; Neogi et al., 2014]. Em particular, esta etapa é onde a ocorrência do defeito de lasca é reportado. A lasca é caracterizada pela ocorrência de pequenas fissuras na superfície da placa de aço. Geralmente, essas rachaduras se estendem por toda a superfície da placa, de ponta a ponta. Em geral, as fendas são paralelas às laterais da placa e, portanto, no momento em que o inspetor as identifica, é feito um registro da distância entre a lasca e as bordas da placa. Em alguns casos, apenas as bandas laterais de aço, de comprimento suficiente para cobrir a falha, são sucateadas. Outra característica importante do defeito é que ele tende a se manifestar em ambos os lados da placa.

A literatura sugere que algumas das possíveis causas do defeito estão relacionadas à baixa fração de ferrita-delta, ao excesso de elementos de fragilidade como enxofre, chumbo, fósforo, estanho e bismuto ou a um alto nível de oxidação. Porém, mesmo com esse conhecimento, a lasca de aquecimento ocorre entre 30% e 50% das placas fabricadas pela APERAM e resulta no sucateamento de cerca de 10% da produção. De fato, segundo Barbosa e Fagundes Barbosa et al. [2007], determinar a origem desse tipo de defeito não é trivial. Diferentes técnicas de caracterização precisam ser usadas para permitir uma conclusão confiável.

A Figura 2.1 ilustra todo o processo de fabricação do aço.

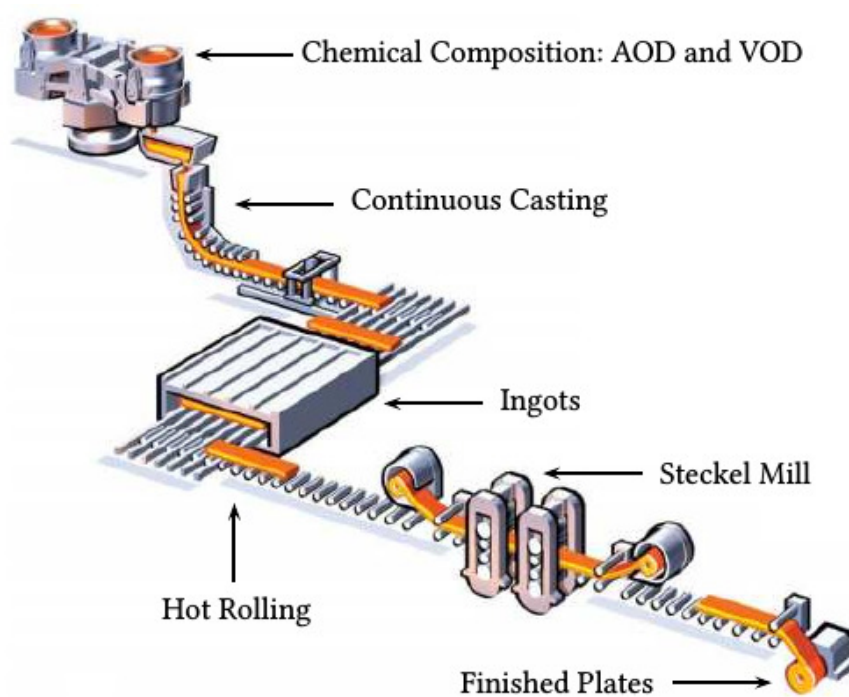


Figura 2.1. Visão geral do processo de fabricação do aço

Capítulo 3

Referencial teórico

Nesta seção discutiremos os principais trabalhos relacionados com nossa proposta de modelagem química de aços duplex resistentes à formação de lascas de aquecimento. A seção 3.1 discorre sobre os métodos tradicionais de análise de causa raiz. A seção 3.2 explica os mecanismos de funcionamento dos modelos de classificação utilizados para gerar as previsões da presença ou não de lasca de aquecimento. A seção 3.3 discute a ferramenta utilizada para gerar interpretações das previsões realizados pelos modelos. A seção 3.4 discorre sobre trabalhos a ausência de trabalhos que tenham empregado aprendizado de máquina nesta tarefa.

3.1 Análise de causa raiz

A análise de causa raiz (RCA) é um método de solução de problemas usado para identificar as causas raiz de falhas ou problemas. O RCA geralmente serve como entrada para um processo de remediação por meio do qual ações corretivas são tomadas para evitar que o problema ocorra novamente. O método está relacionado com inferência e causalidade no sentido de que desejamos compreender um determinado problema e explicar as razões subjacentes que o levaram a ocorrer.

Alguns domínios de aplicação comuns incluem a prevenção e análise de acidentes, bem como aplicações industriais [Rosenfeld, 2014]. Embora haja uma infinidade de diferentes metodologias para alcançar RCA [Solé-Simó et al., 2017], a maioria segue os mesmos cinco princípios básicos:

- Identificação e descrição: definições eficazes dos problemas e descrições dos eventos (como falhas, por exemplo) são úteis e geralmente necessárias para garantir a execução de análises de causa raiz apropriadas.

- Coleta de dados e cronologia: coletar informações sobre o fenômeno avaliado e estabelecer uma sequência de eventos ou cronograma para compreender as relações entre os fatores contributivos (causais), a causa raiz e o problema sob investigação.
- Diferenciação: ao correlacionar esta sequência de eventos com a natureza, a magnitude, a localização e o momento do problema e, possivelmente, também com uma biblioteca de problemas analisados anteriormente, a RCA deve permitir a separação entre a causa raiz, os fatores causais e fatores não causais. Para isso podemos utilizar técnicas de mineração de dados ou comparar a situação sob investigação com situações passadas.
- Grafo causal: extrair das sequências de eventos uma subsequência de eventos-chave que explicam o problema e convertê-la em um grafo causal.
- Medidas preventivas: definição de regras e diretrizes para referência futura.

No entanto, a inferência direta por meio de métodos RCA tradicionais se mostra uma tarefa complicada quando direcionada a encontrar as causas de defeitos no aço inoxidável duplex [Barbosa et al., 2007].

3.2 Modelos de classificação

O aprendizado de máquina e as abordagens baseadas em dados estão se tornando muito importantes em diversas áreas. Dentro das aplicações baseadas em aprendizado de máquina, a tarefa de classificação é uma das mais clássicas. Classificação nada mais é que o processo de prever a classe de determinada instância do problema, baseando-se em observações de instâncias vistas anteriormente. Um sistema antispam, por exemplo, deve prever se um novo email recebido é um spam ou não com base nos email já recebidos que foram reconhecidos ou não como spam. Há dois fatores importantes que impulsionam aplicações desse tipo: o uso de modelos estatísticos efetivos que capturam as complexas dependências entre os dados e os sistemas de aprendizado escalonáveis a ponto de aprenderem bons modelos a partir de grandes conjuntos de dados.

Dentre as diversas possibilidades de modelos de classificação, as técnicas de boosting de árvores tem se destacado por mostrar resultados de estado da arte em vários benchmarks de classificação padrão em comparação com outros métodos. A principal ideia por trás do boosting é usar um conjunto de preditores fracos que podem ser, de alguma forma, combinados para gerar um mais forte. Esta ideia foi proposta pela primeira vez por Michael Kearns como o *Hypothesis Boosting Problem* Kearns [1988]. Ele

afirma que pode haver um algoritmo eficiente que pode converter hipóteses ruins, como preditores fracos que são ligeiramente melhores do que um adivinhador aleatório, em uma única hipótese muito boa. Uma abordagem é filtrar as observações, modificando assim a distribuição dos exemplos de forma a forçar o algoritmo de aprendizado fraco a se concentrar nas partes mais difíceis de aprender da distribuição Schapire [1990].

Portanto, o boosting consiste na utilização de um método de aprendizagem fraco várias vezes para obter uma sucessão de hipóteses. Cada um está focado em aprender a lidar com as observações difíceis restantes com as quais o preditor anterior teve dificuldade. Os preditores não são feitos de forma independente, mas sequencialmente, e cada um aprende com os erros dos preditores anteriores. Isso, por sua vez, faz com que as observações tenham uma probabilidade desigual de aparecer em modelos subsequentes e aquelas com o maior erro aparecem mais. As *Gradient boosting machines* (GBM) são um exemplo de algoritmo de boosting que se originou da observação de Breiman Breiman [1997]: o boosting pode ser interpretado como um algoritmo de otimização sobre uma função de perda adequada.

Sejam y os valores reais da variável de saída, i uma iteração do algoritmo *gradient boosting* e $F_i(x)$ a saída do modelo proposto no tempo i . O algoritmo de *gradient boosting* melhora $F_i(x)$ construindo um novo modelo que adiciona um estimador h para fornecer um modelo melhor, o que leva a $F_{i+1}(x) = F_i(x) + h(x)$. Um h perfeito implicaria em $h(x) = y - F_i(x)$. Portanto, a abordagem de *gradient boosting* tentará ajustar h à perda residual. No entanto, para problemas de classificação e ranqueamento, os resíduos $y - F(x)$ para um determinado modelo são os gradientes negativos em relação a $F(x)$. Com isso, o *gradient boosting* é um algoritmo de descida de gradiente para combinar e treinar preditores fracos. Um preditor comumente usado são as florestas aleatórias.

Neste trabalho, empregamos o XGBoost, que melhora a performance do GBM Chen & Guestrin [2016] original. Por exemplo, ele permite que árvores sejam criadas de maneira gulosa a partir de subamostras do conjunto de dados de treinamento. Isso leva à redução da correlação entre as árvores e evita o excesso de *overfitting*. Esta variação do *boosting* é chamada de *gradient boosting* estocástico.

3.3 Interpretabilidade dos modelos

Existem na literatura alguns métodos para a atribuição de importância a atributos como o LIME [Ribeiro et al., 2016] e o SHAP [Lundberg & Lee, 2017]. Este último tem apresentado os melhores resultados, justamente por ser uma unificação do anterior

com outras ferramentas de mesmo propósito. Sendo assim, utilizaremos o SHAP para realizar a interpretação dos modelos.

O SHAP (SHapley Additive exPlanations) é uma abordagem unificada desenvolvida para explicar a saída de qualquer modelo de aprendizado de máquina. O SHAP combina teoria dos jogos e explicações locais, unindo vários métodos anteriores e representando um robusto método de atribuição de importância a atributos de um modelo com base em expectativas. Basicamente, o SHAP possui uma abordagem baseada em permutações para atribuir a importância de atributos e define um modelo como uma cooperação de características, atribuindo um valor para cada atributo nessa cooperação baseada na sua contribuição para as decisões do modelo. O SHAP possui três propriedades que são essenciais no desenvolvimento deste trabalho:

- **Precisão local:** as explicações explicam o modelo com corretude.
- **Atributos faltantes:** os atributos faltantes não têm impacto atribuído às decisões do modelo.
- **Consistência:** se um modelo mudar para que a contribuição de algum atributo aumente ou permaneça a mesma, independentemente dos outros atributos, o impacto desse atributo não deve diminuir.

O valor de Shapley é um conceito de solução na teoria dos jogos cooperativos Shapley [1953]. Em cada jogo, é fornecida uma distribuição única das recompensas geradas pela cooperação de todos os jogadores, dadas todas as coalizões possíveis de jogadores. Aplicado ao XGBoost, o valor de Shapley dá a contribuição de cada atributo para a explicação das previsões modelo. A coalizão é formada seguindo os nós nas árvores geradas pelo modelo XGBoost.

Basicamente, o valor de Shapley é a contribuição média esperada de um jogador após todas as combinações possíveis terem sido consideradas. Podemos construir uma coalizão com apenas um jogador e então adicionar iterativamente os outros, um de cada vez, até que todo o grupo esteja na coalizão. Adicionar um jogador a uma coalizão existente aumenta seu retorno esperado e, portanto, isso pode ser contabilizado como sua parcela do retorno. No entanto, há muitas maneiras possíveis de formar uma coalizão. A ordem em que cada jogador é adicionado pode influenciar em sua participação no retorno agregado. O valor de Shapley para cada jogador consiste na média dessa contribuição sobre as diferentes permutações possíveis em que a coalizão pode ser formada.

Seja N um conjunto de n jogadores, S uma coalizão de jogadores e v seja uma função característica sobre S . Ou seja, $v(S)$ denota o valor de uma coalizão S e descreve

a soma total esperada de compensações que os membros de S obtêm por cooperação. Adicionar o jogador n_i a uma coalizão S existente aumenta o retorno esperado para a coalizão em $v(S \cup \{n_i\}) - v(S)$. Uma vez que existem $n!$ maneiras possíveis de se organizar os jogadores n e o jogador n_i deve ser precedido por todos os membros de S e seguido pelos jogadores restantes em N , existem $|S|!(n - 1 - |S|)!$ formações em que o jogador n_i junta-se à coalizão existente S . Se somarmos sua contribuição sobre todas as formações em que n_i se junta a S e, em seguida, sobre todas as possíveis coalizões S das quais ele possa se juntar, obteremos sua contribuição total sobre todas as formações possíveis de N .

Por fim, o valor de Shapley $\varphi_{n_i}(v)$ do jogador n_i no jogo cooperativo (v, N) é a média de sua contribuição total em todos os cenários possíveis:

$$\varphi_{n_i}(v) = \sum_{S \subseteq N \setminus \{n_i\}} \frac{1}{n!} |S|!(n - 1 - |S|)! (v(S \cup \{n_i\}) - v(S)) \quad (3.1)$$

A ideia do SHapley Additive Explanations (SHAP) é usar este conceito da teoria dos jogos para interpretar um modelo de aprendizado de máquina [Lundberg & Lee, 2017]. Todos os atributos são jogadores tentando prever uma determinada tarefa que pode ser vista como um jogo cooperativo. A contribuição agregada de cada atributo, sua recompensa, é seu valor real menos a previsão do modelo. O impacto de cada atributo pode, portanto, ser encontrado calculando seu valor de Shapley. No entanto, o cálculo exato dos valores de Shapley é desafiador. O SHAP propõe vários métodos de aproximação. A independência de atributos e a linearidade do modelo são duas suposições opcionais que simplificam o cálculo dos valores esperados.

3.4 Aprendizado de máquina na siderurgia

A natureza dos sistemas de manufatura enfrenta comportamentos cada vez mais complexos, dinâmicos e às vezes até caóticos. Para poder atender com eficiência a demanda por produtos de alta qualidade, é imprescindível utilizar todos os meios disponíveis. Uma área, que teve um desenvolvimento acelerado em termos de resultados promissores, mas também de usabilidade, é o aprendizado de máquina [Wuest et al., 2016]. O aprendizado de máquina e a inteligência artificial conquistaram um papel fundamental no que hoje conhecemos como a quarta revolução industrial. Muito do *hype* em torno da inteligência artificial na manufatura está focado na automação industrial, mas este é apenas um dos aspectos da revolução das fábricas inteligentes. A inteligência artificial também traz para a mesa da indústria e da siderurgia, em especial, a possibilidade de

ganhos de eficiência através de aplicações como a predição de qualidade, a manutenção preditiva e a modelagem generativa dos materiais.

Dentro da seara de aplicações de aprendizado de máquina na siderurgia, Li et al. [2018] apresenta um sistema baseado em Aprendizado de Máquina para se estimar a qualidade do aço no final do processo de fabricação. Eles utilizam um conjunto de dados composto de 51 atributos que variam desde as composições químicas do aço até variáveis de processo, como a temperatura do forno. O problema é apresentado como uma tarefa de regressão em que o objetivo é prever o valor de certas propriedades do produto acabado.

Ainda segundo Li et al. [2018], o problema é melhor tratado agrupando-se vários algoritmos de aprendizagem em forma de *ensemble* aliados à enorme quantidade de dados históricos de produção disponíveis. Os resultados indicam que o sistema de previsão de controle de qualidade do aço com base em um modelo de aprendizagem de máquina pode oferecer resultados promissores, ao mesmo tempo que oferece alta usabilidade para que fabricantes locais resolvam os problemas na produção com auxílio do desenvolvimento de técnicas de aprendizado de máquina. Além disso, a implantação do mundo real deste sistema é demonstrada e discutida.

Seguindo a mesma linha do trabalho anterior, Ruiz et al. [2020] utiliza procedimentos de Aprendizado de Máquina para prever a resistência à tração de barras de aço. O conjunto de dados disponível inclui 5540 observações e 97 atributos monitorados durante as diferentes etapas do processo - forno elétrico a arco, forno panela, lingotamento contínuo e laminação a quente. Diversos algoritmos de regressão foram testados, com o algoritmo *Random Forest*, que utiliza dos mesmos princípios teóricos do XGBoost, obteve a melhor performance com um R^2 de 0,775 no conjunto de dados de teste.

Além disso, técnicas de explicabilidade foram aplicadas para definir quais variáveis possuem maior influência na resistência à tração das barras de aço. Mais precisamente, os algoritmos Feature Importance e Permutation Importance mostraram que as variáveis químicas têm maior influência na resistência do material. A influência quantitativa dessas variáveis foi representada por meio de gráficos de dependência parcial. Por fim, estas informações foram utilizadas para uma tomada de decisão com o objetivo de se otimizar a resistência das barras de aço.

Apesar de existirem algumas diretrizes a respeito dos componentes relacionados à formação de lascas de aquecimento no aço inoxidável duplex, como em Chai & Kangas [2014], não encontramos, até o momento, nenhuma análise de causa raiz profunda e que utilize um conjunto de dados em larga escala considerando toda a gama de elementos químicos presentes no aço inoxidável duplex e nem em outros tipos de ligas de aço.

Capítulo 4

Metodologia

4.1 Caracterização do conjunto de dados

O conjunto de dados fornecido pela APERAM South America é composto pela composição química de placas de aço inoxidável duplex medidas durante o processo de fabricação do aço. Espectrômetros foram utilizados para avaliar as quantidades relativas (%) para cada elemento químico em uma determinada placa. Cada elemento possui um padrão espectral particular e, assim, os picos nos espectros são associados a elementos específicos, com base na comparação com os resultados das amostras de referência. No total, temos 20 elementos químicos que são medidos em diferentes estágios do processo de fabricação de aço.

Enquanto várias medições são realizadas durante o processo de fabricação de aço, consideramos apenas as composições químicas finais registradas para aprender nossos modelos. Os dados foram coletados durante todo o ano de 2018 e, durante esse período, especialistas metalúrgicos da APERAM South America realizaram uma série de estudos, testando composições químicas ligeiramente diferentes, com o objetivo de se evitar a formação de lascas. O conjunto de dados final é composto por 122 placas para as quais as composições químicas correspondentes foram aferidas no início da etapa de lingotamento contínuo. Existem 71 placas defeituosas (que possuem lascas) e as 51 restantes não possuem defeitos (sem lascas).

Como também consideramos as proporções entre esses elementos como atributos para aprender nossos modelos, estendemos o espaço de atributos para 380 atributos.

4.2 Encontrando padrões defeituosos na composição química do aço

Nesta seção, propomos uma nova abordagem para modelar a formação de lascas no aço inoxidável duplex. Essa modelagem é desafiadora porque pode ser definida por várias estruturas locais no espaço dos dados.¹ Como resultado, o problema de aprendizagem correspondente tem uma superfície de erro não-convexa sem um mínimo global óbvio, implicando que podemos gerar múltiplos modelos que são performáticos, cada um deles fornecendo uma explicação diferente para a formação de lascas. Portanto, podem emergir interpretações opostas ou concorrentes para o mesmo fenômeno.

Nossa abordagem, pretende-se encontrar uma explicação para a formação de lascas de aquecimento que seja coerente com muitas explicações concorrentes. Portanto, propomos a decomposição do conjunto original de atributos em vários subconjuntos para que um modelo específico seja construído para cada subconjunto de atributos (elementos químicos, no nosso caso). Por fim, os modelos gerados são agrupados de acordo com suas explicações, promovendo a diversidade em termos de possíveis explicações para a formação de lascas. Espera-se que padrões químicos fortes surjam do conjunto de modelos, e esses padrões quando analisados em conjunto podem fornecer uma visão mais global do mecanismo de formação de lascas de aquecimento.

Podemos resumir nossa abordagem da seguinte maneira:

1. Aprendemos um grande número de modelos com o mesmo conjunto de dados de composição química. Cada modelo utiliza um conjunto diferente de características - ou seja, elementos químicos e suas concentrações relativas.
2. Organizamos os modelos em função das suas previsões e também das explicações fornecidas.
3. Selecionamos padrões defeituosos e não defeituosos presentes em um conjunto de modelos representativos.

4.2.1 Estruturas locais nos dados

Um espaço de dados é definido como um conjunto de n pontos da forma $(\mathbf{x}, y)^n$, de modo que $\mathbf{x} \in \mathcal{R}^d$ é dado como um vetor de característica $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d\}$ e y é a saída verdadeira para \mathbf{x} . Frequentemente, em espaços de dados de alta dimensão, existem

¹Embora a composição química não seja o único fator relacionado à formação de lascas, assumiremos que as características relacionadas às etapas posteriores são constantes até certo ponto.

regiões que mostram correlações complexas entre um conjunto específico de características e o rótulo alvo, e as mesmas correlações não são necessariamente observadas com tanta força em outras regiões do espaço de dados. Assim, a teoria do relacionamento — os dados variam em diferentes regiões do espaço de dados, formando estruturas locais definidas como subespaços abrangidos por um conjunto de pontos de dados e um conjunto de características [Tanay et al., 2005]. As estruturas locais podem se sobrepor e geralmente são o resultado da mistura de diferentes subpopulações ou distribuições no mesmo espaço de dados e, portanto, não é possível separá-las facilmente em vários subespaços. Um tipo específico de estrutura local se assemelha a um esqueleto, no sentido de que há um conjunto de características (também conhecido como características esqueleto) que mostram forte correlação com um outro conjunto específico de características. Portanto, forçar uma característica esqueleto a aparecer no mesmo modelo com características não relacionadas pode resultar em situações confusas.

4.2.2 Amostragem do Espaço do Modelos

Aprender um modelo a partir espaço de dados requer a minimização de uma função objetivo $f(\mathbf{x})$. Em vez de simplesmente misturar várias estruturas diferentes em um único modelo \mathbf{x} e minimizar $f(\mathbf{x})$, amostramos o espaço de modelos minimizando diferentes funções $f(\mathbf{x}')$, de modo que $\mathbf{x}' \subseteq \mathbf{x}$ e $|\mathbf{x}'| \ll |\mathbf{x}|$. As características que compõem cada modelo \mathbf{x}' são selecionadas aleatoriamente, e usamos gradient boosted trees [Chen & Guestrin, 2016] como o algoritmo de aprendizagem² (mas outros algoritmos também podem ser aplicados). Depois de amostrar o espaço de modelos, cada modelo \mathbf{x}' é avaliado em através de uma métrica de erro $\ell(x')$ em um conjunto de validação, de modo que apenas modelos com desempenho mínimo para os quais $\ell(\mathbf{x}') \leq \epsilon$ estão incluídos no espaço de modelos final \mathcal{H}' . Neste ponto, esperamos que \mathcal{H}' contenha modelos performáticos que correspondam a possíveis explicações para o mecanismo de formação de lascas.

4.2.3 Representando as preferências do modelo

Representamos a preferência do modelo como um vetor n -dimensional $\mathcal{P}(\mathbf{x}') = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$, onde \mathbf{p}_i corresponde à probabilidade que o modelo \mathbf{x}' atribuiu ao vetor de dados i . Esperamos que os modelos (performáticos) em \mathcal{H}' sejam representativos das diversas estruturas locais que existem no espaço de dados. Ao filtrar os modelos pe-

²Uma vez que as características de composição química consideradas podem ter uma variedade de interações não lineares complexas, empregamos um algoritmo de classificação com flexibilidade significativa.

formáticos para os quais $\ell(\mathbf{x}') \leq \epsilon$ esperamos que a estrutura local correspondente seja devidamente explicada pelo modelo correspondente \mathbf{x}' .

4.2.4 Representando as explicações do modelo

Representamos como o modelo \mathbf{x}' explica o fenômeno como um vetor d -dimensional $E(\mathbf{x}') = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d\}$ mostrando quais características estão conduzindo a previsão do modelo. Especificamente, \mathbf{e}_i assume um valor que corresponde à influência que a respectiva característica \mathbf{x}_i teve na decisão do modelo. Visto que não assumimos a independência dos atributos enquanto minimizamos $f(\mathbf{x}')$, as características correlacionadas dentro do modelo \mathbf{x}' deveriam compartilhar seu crédito ou importância. Por esse motivo, usamos os valores médios de SHAP para avaliar a importância de cada característica.

4.2.5 Prevenindo a formação de lascas evitando padrões defeituosos e impondo padrões não defeituosos na composição química

Como \mathcal{H}' pode conter modelos com explicações concorrentes sobre suas decisões, queremos construir um modelo sintético de \mathcal{H}' explorando dois conceitos:

- O conceito de diversidade entre modelos individuais. Reconhecemos a diversidade como um elemento central para obter uma compreensão mais geral sobre a formação da lasca. Assumimos que o mecanismo de formação das lascas não está relacionado a um único fator causal e que os fatores causais podem variar dependendo das transformações realizadas em etapas posteriores do processo de fabricação de aço (ou seja, laminação a quente e lingotamento). A fim de promover a diversidade ao encontrar padrões defeituosos, agrupamos modelos em \mathcal{H}' com base na distância entre seus vetores de explicação (ou seja, valores SHAP). Idealmente, isso cria uma série de grupos de modelos que são internamente densos e também separados do resto dos modelos em termos de seus fatores explicativos, ou seja, dentro de cada cluster os fatores explicativos são semelhantes, enquanto os fatores dentro de clusters disjuntos são diferentes.
- O conceito de estabilidade entre a explicação do modelo e as previsões empíricas [Shmueli, 2010]. Definimos uma configuração de clusters como estável se os modelos dentro do mesmo cluster estão associados aos mesmos fatores explicativos e realizam previsões semelhantes. Alcançar a estabilidade dos clusters

é desafiador, pois os modelos que realizam previsões semelhantes podem ser associados a fatores explicativos muito diferentes. A fim de avaliar a estabilidade da previsão-explicação ao encontrar padrões defeituosos, agrupamos o espaço do modelo com base na distância entre o vetor de explicação (ou seja, valores SHAP) associado a cada modelo e, em seguida, projetamos os agrupamentos no espaço de preferências dos modelos. Isso impõe uma configuração estável de clusters contendo modelos que são semelhantes em termos de suas previsões e seus fatores explicativos.

Uma vez que os clusters são encontrados, selecionamos um modelo protótipo dentro de cada cluster, de modo que tenhamos o mesmo número de modelos protótipos e clusters. Em particular, selecionamos o modelo centróide dentro de cada cluster. A intuição é que o mecanismo de formação das lascas pode ter várias explicações, e cada modelo de protótipo é uma possível explicação para o fenômeno. Para selecionar padrões defeituosos e não defeituosos na composição química, separamos caminhos em cada árvore que compõe o ensemble que levam a nós puros. Em seguida, verificamos os padrões incluídos, mantendo os mais específicos. Finalmente, padrões com e sem defeitos são empregados para prevenir a formação de lascas.

Capítulo 5

Experimentos

5.1 Configuração dos experimentos

Durante a amostragem do espaço de modelos, definimos aleatoriamente o número de características que compõem cada modelo, mas garantimos que nenhum modelo utiliza mais que 15 delas. Os atributos que compõem cada modelo também são selecionados aleatoriamente a partir do conjunto de todos os 400 atributos disponíveis. Os modelos são construídos usando implementações do SciKit-Learn e os algoritmos XGBoost [Pedregosa et al., 2011]. Construímos um total de 113.966 modelos de XGBoost¹. Para avaliar o desempenho, usamos a medida AUC (área sob a curva ROC) padrão [Hanley & McNeil, 1982; Fawcett, 2006]. Realizamos uma validação cruzada em cinco conjuntos, ou seja, os dados são organizados em cinco conjuntos e, a cada execução, quatro deles são usados como conjunto de treinamento e o conjunto restante é usado como conjunto de teste. Também utilizamos um conjunto de validação separado usado para selecionar os melhores modelos. Relatamos o valor médio de AUC ao longo das cinco execuções.

Como não há uma comparação de base disponível na literatura, usamos a abordagem com todas os atributos em um único modelo para fornecer uma comparação de base. Essa abordagem utiliza a mesma implementação XGBoost, mas produz apenas um único modelo composto de todos os recursos disponíveis. Os valores médios de AUC obtidos pelo modelo único sem expansão de atributos foi de 0,63. Portanto, consideramos um modelo com performance minimamente aceitável se seu valor médio de AUC for pelo menos 0,63. Este limite de desempenho resultou em um espaço amostral de modelos \mathcal{H}' contendo 40.121 modelos dos 113.966 originais (35,20% dos modelos

¹Executamos o XGBoost 113.966 vezes, e a cada vez um subconjunto diferente de características é considerada.

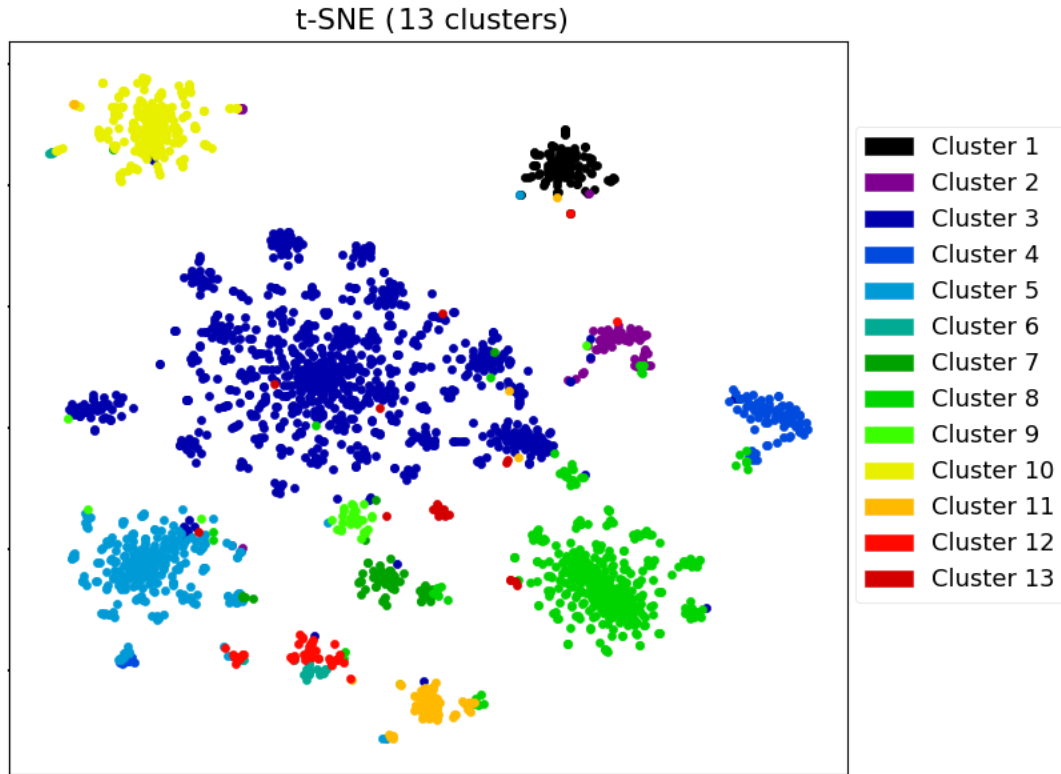


Figura 5.1. T-SNE [van der Maaten, 2009] do espaço amostral de modelos \mathcal{H}' . Cada ponto representa um modelo \mathbf{x}' . Os modelos são posicionados de acordo com as probabilidades de defeito atribuídas a cada placa de aço, de forma que os modelos que atribuem probabilidades semelhantes às mesmas placas sejam colocados mais próximos no espaço. A cor indica o cluster para o qual o modelo foi atribuído.

têm desempenho melhor do que o modelo único). Para uma visualização mais fácil, mostramos apenas os modelos 2% principais em nossos gráficos, o que leva a um limiar de performance de 0,7015.

5.2 Resultados

A Figura 5.1 mostra o espaço amostral dos modelos. Na figura, cada ponto corresponde a um modelo, e as cores indicam o cluster atribuído ao modelo. Usamos o K-Means [Forgy, 1965] para agrupar os modelos em clusters e encontramos um número adequado de clusters maximizando o valor da silhueta. A silhueta é uma medida de quão semelhante um modelo é ao seu próprio cluster (coesão) em comparação com outros clusters (separação). A silhueta varia de -1 a $+1$, onde um valor alto indica que

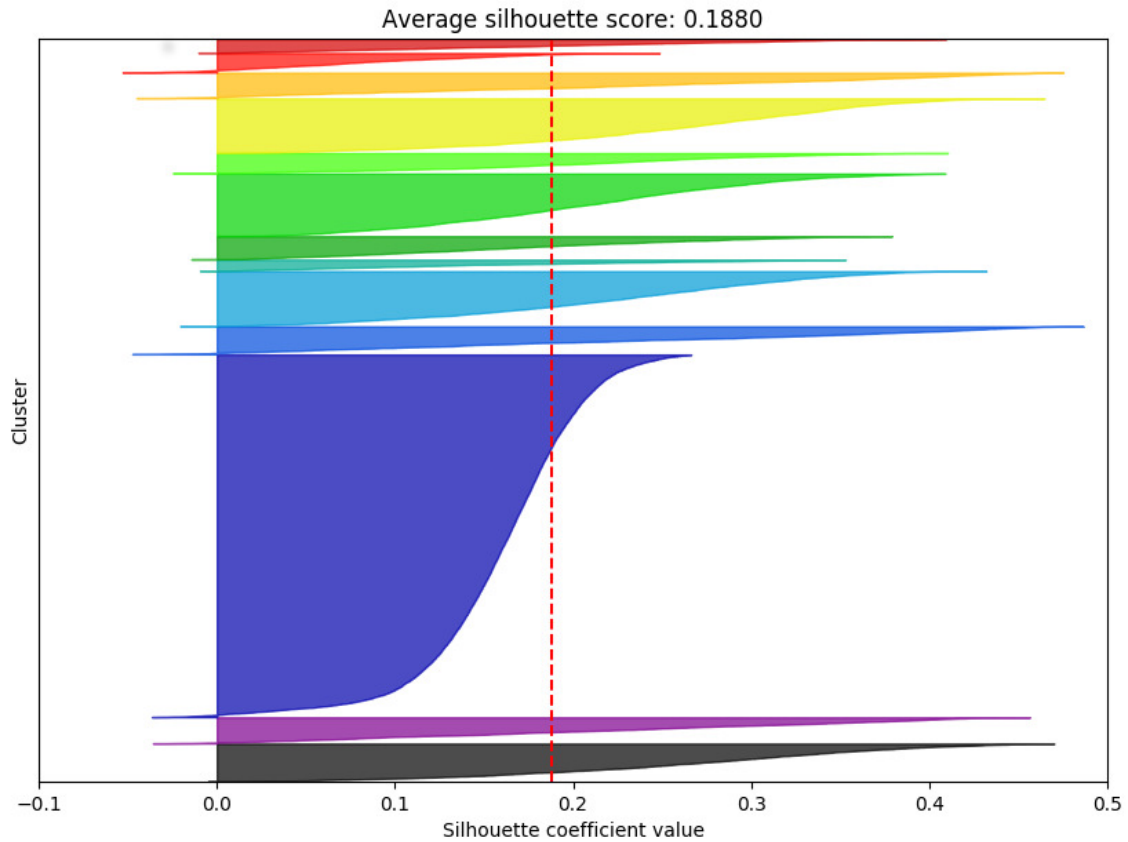


Figura 5.2. Análise de silhueta para cada cluster. O valor médio da silhueta sugere coesão e separação.

o modelo combina bem com seu próprio cluster e não combina com clusters vizinhos. Se a maioria dos modelos possuir um valor alto, a configuração de clusters é considerada apropriada. As melhores pontuações de silhueta são mostradas na Figura 5.2. Após uma inspeção mais detalhada, pudemos determinar que os clusters são divididos principalmente por quais atributos compõem seus modelos. Considere, por exemplo, o segundo cluster. Todos os modelos deste cluster consideram os valores de um mesmo elemento dado na mistura e, portanto, possuem valores de SHAP próximos para este elemento. Isso, por sua vez, os leva a ficarem próximos no espaço de preferências dos modelos, uma vez que os valores de coesão são relativamente altos.

A Figura 5.3 ilustra o valor de AUC de cada modelo. Podemos observar que os modelos performáticos estão bem espalhados pelo espaço de modelos, indicando que existem modelos com preferências diferentes e com fatores explicativos diferentes, mas com desempenhos semelhantes não importando o cluster atribuído a eles. Ou seja, os modelos associados a diferentes fatores explicativos têm números de desempenho

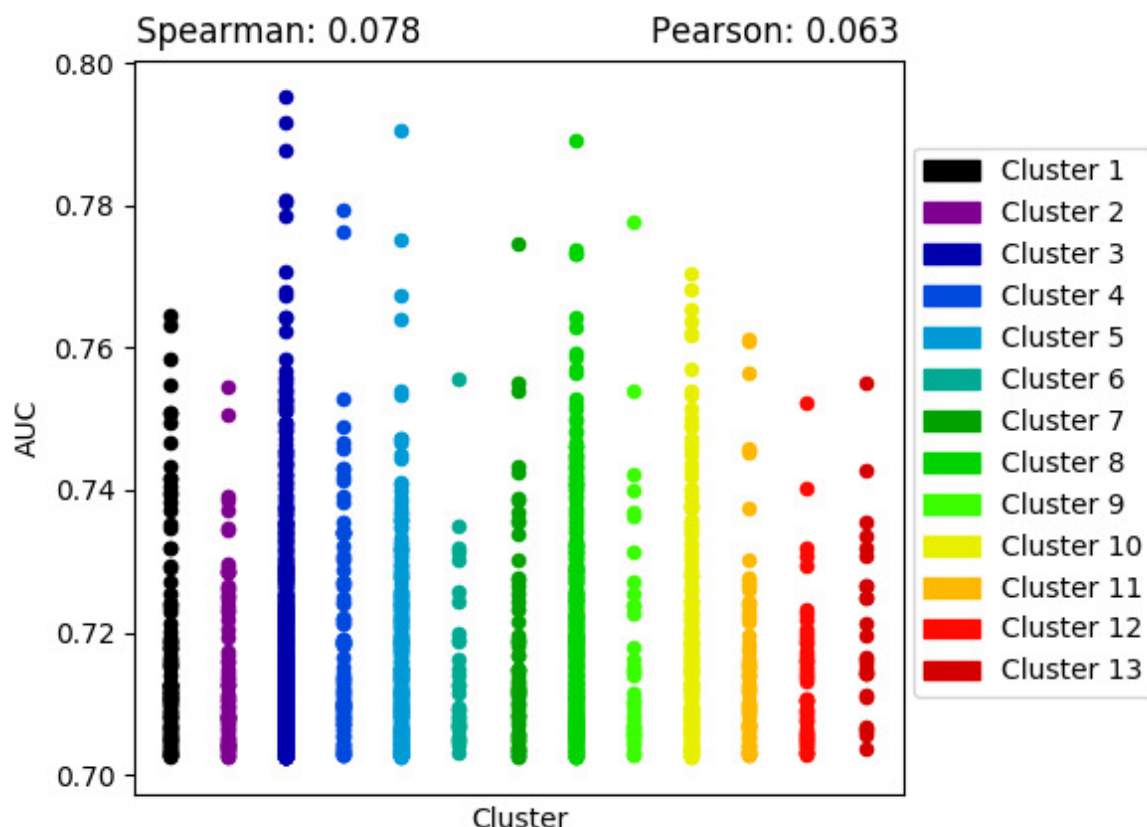


Figura 5.3. Distribuição dos valores de AUC em cada cluster. Os modelos performáticos estão distribuídos pelos clusters.

semelhantes. Não existe um padrão na distribuição de clusters por AUC. De fato, ambas as correlações de Pearson e Spearman entre os clusters e a AUC são inferiores a 0,1, como também mostrado na Figura 5.3.

A Figura 5.4 mostra o número de modelos em que cada característica apareceu. Observa-se que os elementos molibdênio, níquel e nióbio aparecem em muitos modelos, mas geralmente em conjunto com outras características.

O próximo experimento é dedicado a investigar a relação entre as características disponíveis. Especificamente, estamos interessados em investigar quais características aparecem em modelos semelhantes. Isso pode ser útil porque algumas reações químicas podem ser mais difíceis de realizar do que outras e, portanto, características semelhantes podem fornecer alternativas aos elementos originais. Para isso, simplesmente tratamos os modelos como contextos e rodamos o algoritmo SkipGram [Mikolov et al., 2013] para aprender representações distribuídas para as características, de modo que possamos comparar as características em termos de modelos nos quais elas apareceram. Definimos o tamanho do contexto como o número de características dentro do modelo.

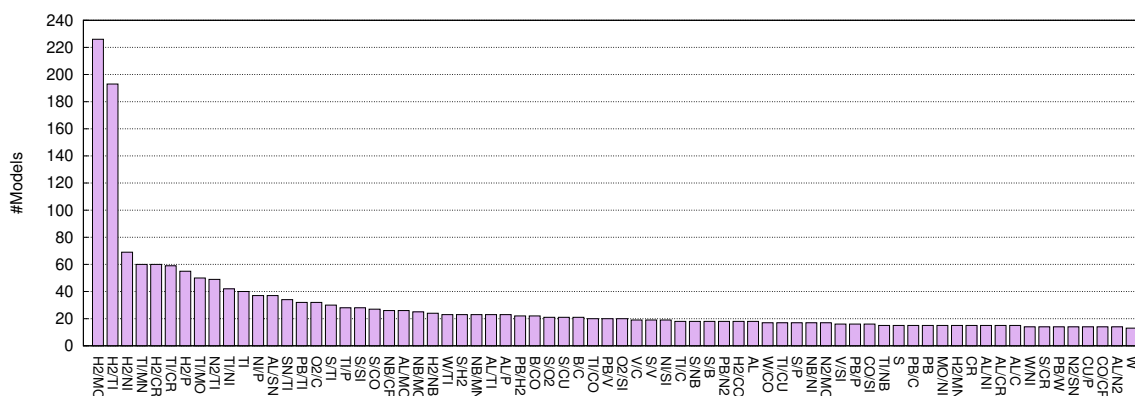


Figura 5.4. Número de modelos em que cada característica apareceu.

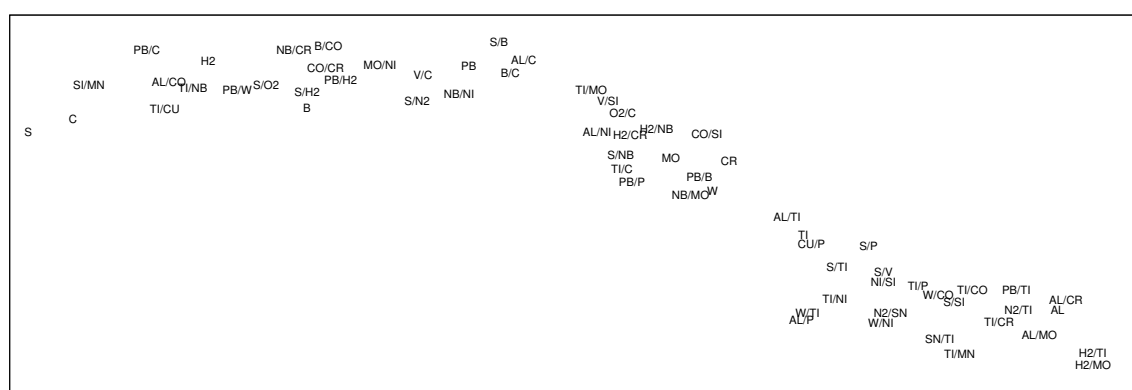


Figura 5.5. Representação distribuída de cada feature.

A Figura 5.5 mostra algumas das características, e a intuição básica aqui é que as características próximas umas das outras apareceram em modelos semelhantes, enquanto as características distantes umas das outras apareceram em modelos diferentes.

5.3 Interpretabilidade dos modelos

A interpretabilidade dos modelos é um dos principais requisitos para que seja possível a modelagem de novos materiais de aço. Um modelo interpretável precisa ser transparente quanto à lógica por trás da sua previsão. A Figura 5.6 (à esquerda) mostra curvas ROC para os modelos centróide dentro de alguns dos clusters. Tomamos esses centróides como modelos representativos para uma investigação mais profunda. A Figura 5.6 (à direita) mostra os gráficos de resumo do SHAP. Esses gráficos mostram os valores de SHAP de cada característica para cada placa de aço no conjunto de dados de composição química. Em cada gráfico, as características são classificadas pela soma

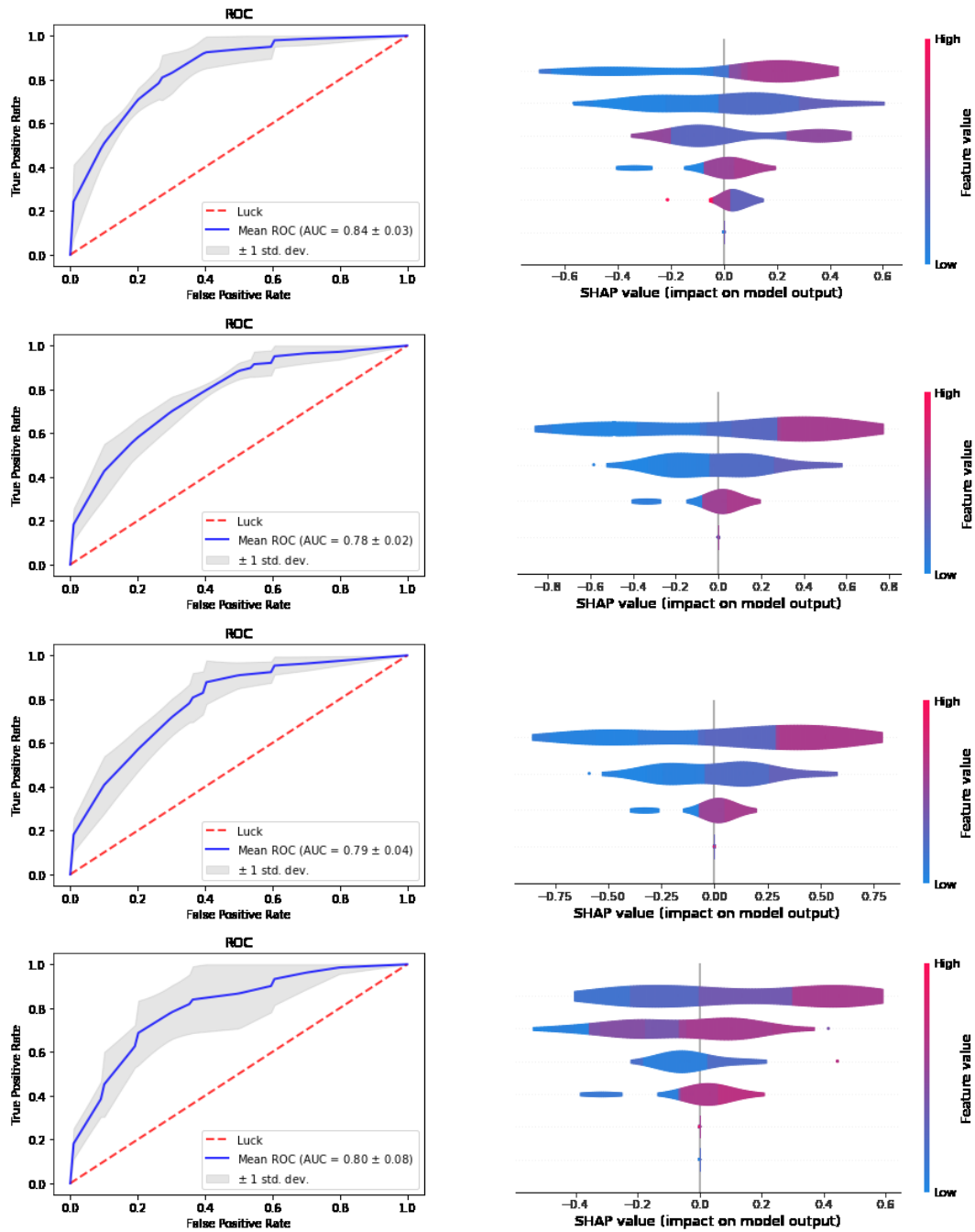


Figura 5.6. Modelos representativos para alguns clusters. Esquerda – Curva ROC mostrando o desempenho do modelo. Direita – O *summary plot* SHAP correspondente mostrando uma visão geral de quais atributos são mais importantes para o modelo.

das magnitudes dos valores SHAP sobre todos as instâncias de dados. A cor representa o valor da característica.

Curiosamente, existem elementos que parecem ser muito relevantes, visto que

ocorrem muito frequentemente em modelos dentro de diferentes clusters. Em geral, no entanto, os gráficos de resumo mostram que as características nos modelos dentro de diferentes clusters são bastante distintas. No todo, o comportamento predominante é que grupos específicos de características estão associadas a espaços de solução distintos e nenhuma relação clara entre conjuntos de características e desempenho pode ser vista.

Depois de selecionar um modelo representativo para cada cluster, pedimos a opinião dos especialistas em metalurgia da APERAM South America. A principal lição é que há casos em que algumas das conclusões encontradas não se enquadram em cenários realistas. Por exemplo, um dos modelos pode alegar que uma grande quantidade de carbono é necessária para evitar a formação de lascas. Porém, para produzir aço inoxidável duplex são necessários baixos percentuais de carbono e tal característica não pode ser aceita pelos especialistas. Felizmente, existem muitos padrões que podem ser acionados, especialmente aqueles relativos aos principais elementos químicos usados durante o processo de produção do aço inoxidável.

Alguns dos padrões descobertos estão sendo usados atualmente para produzir placas não defeituosas. Para aqueles padrões que não se encaixam em cenários realistas, os especialistas estão atualmente avaliando padrões obtidos a partir de características alternativas, conforme mostrado na Figura 5.5.

Capítulo 6

Conclusões e trabalhos futuros

Neste trabalho, estudamos um problema subexplorado de aprendizado de máquina para projetar novos materiais de aço que são menos propensos a um defeito conhecido como lasca de aquecimento. Nosso estudo é realizado em parceria com a APERAM South America, maior fabricante brasileira de aços inoxidáveis. A tarefa foi formulada como um problema de classificação binária para prever quais combinações de componentes químicos podem estar associadas à formação de lascas. Implementamos uma ferramenta que encontra padrões de defeituosos e não defeituosos na composição química do aço inoxidável duplex. A ferramenta implantada forneceu várias regras para orientar a produção de placas de aço sem defeito. Para isso, todos os modelos produzidos pela ferramenta implantada são interpretáveis por humanos e as árvores geradas foram encaminhadas para a linha de produção da APERAM. Placas de aço inoxidável que seguem as diretrizes propostas estão atualmente em construção e os especialistas estão entusiasmados com a análise preliminar.

Referências Bibliográficas

- Barbosa, L. A.; Fagundes, G. A.; Teichmann, L. & Regulý, A. (2007). Evaluation of sliver surface defects in cold-drawn steel bars. *Tecnologia em Metalurgia, Materiais e Mineracao*, 3(4):59.
- Breiman, L. (1997). Arcing the edge. Relatório técnico, Technical Report 486, Statistics Department, University of California at
- Chai, G. & Kangas, P. (2014). Recent developments of advanced austenitic and duplex stainless steels for oil and gas industry. *Energy Materials*, pp. 703--709.
- Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Intl Conf. on Knowledge Discovery and Data Mining*, pp. 785--794.
- Deo, B. & Boom, R. (1993). *Fundamentals of Steelmaking Metallurgy*. Prentice and Hall.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861--874.
- Forgy, E. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21(3):768--769.
- Gunn, R. N. (1997). *Duplex Stainless Steels, Microstructure, Properties and Applications*. Abington Publishing.
- Hanley, J. & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic ROC curve. *Radiology*, 143:29--36.
- Kearns, M. (1988). Thoughts on hypothesis boosting. *Unpublished manuscript*, 45:105.
- Li, F.; Wu, J.; Dong, F.; Lin, J.; Sun, G.; Chen, H. & Shen, J. (2018). Ensemble machine learning systems for the estimation of steel quality control. In Abe, N.; Liu, H.; Pu, C.; Hu, X.; Ahmed, N. K.; Qiao, M.; Song, Y.; Kossmann, D.; Liu, B.;

- Lee, K.; Tang, J.; He, J. & Saltz, J. S., editores, *IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, December 10-13, 2018*, pp. 2245--2252. IEEE.
- Lundberg, S. & Lee, S. (2017). A unified approach to interpreting model predictions. In *Annual Conf. on Neural Information Processing Systems*, pp. 4768--4777.
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pp. 3111--3119.
- Neogi, N.; Mohanta, D. & Dutta, P. (2014). Review of vision-based steel surface inspection systems. *J Image Video Proc*, p. 50.
- Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; VanderPlas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M. & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825--2830.
- Ribeiro, M. T.; Singh, S. & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Intl Conf. on Knowledge Discovery and Data Mining*, pp. 1135--1144.
- Rosenfeld, Y. (2014). Root-cause analysis of construction-cost overruns. *Journal of Construction Engineering and Management*, 140:04013039.
- Ruiz, E.; Ferreño, D.; Cuartas, M.; López, A.; Arroyo, V. & Gutiérrez-Solana, F. (2020). Machine learning algorithms for the prediction of the strength of steel rods: an example of data-driven manufacturing in steelmaking. *International Journal of Computer Integrated Manufacturing*, 33(9):880--894.
- Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2):197--227.
- Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307--317.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3):289--310.
- Solé-Simó, M.; Muntés-Mulero, V.; Rana, A. & Estrada, G. (2017). Survey on models and techniques for root-cause analysis.

- Stradomska, J.; Stradomski, Z. & Soinski, M. (2009). The analysis of solidification process of ferritic-austenitic cast steel. *Archives of Foundry Engineering*, 9(1):83--86.
- Tanay, A.; Sharan, R. & Shamir, R. (2005). Biclustering algorithms: A survey. *Handbook of computational molecular biology*, 9:26--1.
- Thomas, B. (2006). Modeling of continuous-casting defects related to mold fluid flow. *Iron and Steel Technology (AIST Transactions)*, 3(7):128--143.
- van der Maaten, L. (2009). Learning a parametric embedding by preserving local structure. In *Intl Conf. on Artificial Intelligence and Statistics*, pp. 384--391.
- Wuest, T.; Weimer, D.; Irgens, C. & Thoben, K.-D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, 4(1):23--45.
- Zhao, Y. J.; Yan, Y. H. & Song, K. C. (2017). Vision-based automatic detection of steel surface defects in the cold rolling process: considering the influence of industrial liquids and surface textures. *The International Journal of Advanced Manufacturing Technology*, 90(5--8):1665.