

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Escola de Engenharia

Departamento de Metalurgia

AVALIAÇÃO DE MODELOS DE PREVISÃO DA QUALIDADE DE COQUE
BASEADO EM ABORDAGEM DE MINERAÇÃO DE DADOS E APRENDIZADO
DE MÁQUINA

Diestéfano Souza Oliveira

Trabalho de Diplomação

Orientador: Prof. Dr. Bruno Flores

Porto Alegre
2023

CIP - Catalogação na Publicação

Oliveira, Diestéfano Souza
Avaliação de modelos de previsão da qualidade do
coque baseado em abordagem de mineração de dados e
aprendizado de máquina. / Diestéfano Souza Oliveira.
-- 2023.
78 f.
Orientador: Bruno Deves Flores.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Escola de
Engenharia, Curso de Engenharia Metalúrgica, Porto
Alegre, BR-RS, 2023.

1. Siderurgia. 2. Aprendizado de máquina. 3. Coque.
4. CRI. 5. CSR. I. Flores, Bruno Deves, orient. II.
Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os
dados fornecidos pelo(a) autor(a).

Dedico esse trabalho de graduação
aos meus pais e à minha esposa.

AGRADECIMENTOS

Agradeço ao professor Bruno Flores por ter aceitado ser meu orientador e por ter auxiliado de maneira ímpar a estruturar esse trabalho, pelo auxílio no desenvolvimento do senso crítico e pelas correções e ensinamentos técnicos sobre a área de carvão e coque.

À Ternium Brasil por ter permitido a realização desse trabalho, financiando o estudo e fornecendo as bases de dados.

Ao meu amigo Dr. Anderson Azevedo Agra por todos os ensinamentos técnicos sobre a área de siderurgia, bem como a parceria e a amizade desenvolvidas ao longo da minha formação acadêmica.

Ao Eng. Me. Bruno Pinheiro da Silva pela confiança depositada e pelas discussões técnicas que engrandeceram o trabalho.

Ao Eng. Lorenzo Fornasari pelos ensinamentos na área de programação, estatística, aprendizado de máquina, otimização, dentre outros, e pela paciência ao longo do meu desenvolvimento nesse trabalho.

Ao prof. Dr. Antônio Cezar Faria Vilela pela oportunidade de fazer parte do Laboratório de Siderurgia, certamente a decisão que mudou minha vida.

Aos demais colegas e amigos de graduação e pós-graduação do Laboratório de Siderurgia e da Engenharia Metalúrgica que passaram pela minha vida ao longo destes anos.

À minha vó Maria Aparecida Dias de Souza, ao meu vô Osmar Zalla de Souza e à minha tia Josi Dias de Souza por terem contribuído com meu desenvolvimento até aqui.

À minha irmã, futura Eng. Vitória Sthéfani de Souza Oliveira pela amizade e por seguir os ensinamentos do irmão.

Em especial, aos meus pais Getúlio Cardoso de Oliveira e Vivian Dias de Souza Oliveira por todo o amor, o carinho, paciência e suporte necessário para a conclusão dessa graduação.

Por fim, gostaria de agradecer à minha esposa Ana Paula Machado Oliveira por todo o amor, todo o carinho, toda a paciência e toda a parceria em mim depositados desde que entrou na minha vida, assim sendo para todo o sempre.

SUMÁRIO

LISTA DE FIGURAS	VII
LISTA DE TABELAS	IX
LISTA DE ABREVIATURAS E SÍMBOLOS	X
RESUMO	XI
ABSTRACT	XII
1.0 INTRODUÇÃO.....	1
2.0 OBJETIVOS.....	2
3.0 REVISÃO BIBLIOGRÁFICA	3
3.1 Carvão Mineral.....	3
3.2 Coqueificação	4
3.2.1 Coqueificação Industrial	4
3.2.2 Função do Coque no Alto-Forno	5
3.2.3 Testes de Qualidade do Coque.....	6
3.2.4 Qualidade do Coque no contexto da Descarbonização.....	7
3.3 Modelos de previsão de qualidade do coque	7
3.3.1 Modelos de Ammosov, Schapiro e Gray	7
3.3.2 Fator G de Simonis	8
3.3.3 Modelo de previsão de CRI e CSR por comportamento aditivo	9
3.3.4 Modelos baseados no rank.....	10
3.3.5 Modelos baseados nas cinzas e matéria mineral.....	10
3.3.6 Modelos baseados em propriedades termoplásticas	11
3.3.7 Outros modelos	12
3.4 Ferramentas computacionais	12
3.4.1 Python	12
3.4.2 Algoritmos de seleção de parâmetros	13
3.4.2.1 BorutaSHAP.....	13
3.4.2.2 Seleção sequencial de parâmetros (SFS).....	14
3.4.2.3 Seleção exaustiva de parâmetros (EFS)	15
3.4.3 Algoritmos de regressão em aprendizado de máquina	15
3.4.3.1 Florestas aleatórias	15
3.4.3.2 LightGBM.....	17
3.4.3.3 Máquina de vetores de suporte.....	17
3.4.3.4 Extreme Gradient Boosting – XGBoost.....	18

4.0	MATERIAIS E MÉTODOS.....	19
4.1	Tratamento dos dados.....	20
4.2	Criação dos modelos de previsão.....	23
4.2.1	Criação do modelo linear simples.....	24
4.2.2	Criação dos modelos de aprendizado de máquina.....	25
4.3	Validação cruzada.....	26
5.0	RESULTADOS E DISCUSSÃO.....	28
5.1	Modelo linear simples.....	28
5.1.1	Modelagem de CRI.....	30
5.1.2	Modelagem de CSR.....	30
5.2	Modelos de aprendizado de máquina.....	31
5.2.1	Modelo de SVR.....	31
5.2.2	Modelo de florestas aleatórias.....	36
5.2.3	Modelo de LightGBM LT.....	38
5.2.4	Modelo XGBoost.....	40
6.0	CONCLUSÕES.....	46
7.0	SUGESTÕES PARA TRABALHOS FUTUROS.....	48
8.0	REFERÊNCIAS.....	49
	APÊNDICE.....	52
	Parâmetros reológicos.....	63
	Parâmetros petrográficos.....	64
	Parâmetros de química das cinzas.....	66
	Parâmetros de análise imediata.....	67
	Parâmetros de processo.....	67

LISTA DE FIGURAS

Figura 3.1	Esquema de comparação entre fornos de coqueificação	5
Figura 3.2	Esquema de representação das árvores de decisão para regressões	16
Figura 4.1	Resumo da coleta de dados	19
Figura 4.2	Resumo da escolha de modelos e a principal motivação.	23
Figura 4.3	Resumo da sequência de ações para criação do modelo linear	24
Figura 4.4	Resumo da sequência de ações para criação do modelo de aprendizado de máquina	25
Figura 4.5	Esquema da validação cruzada Leave-One-Out.....	26
Figura 5.1	Validação cruzada do modelo linear simples de CRI	30
Figura 5.2	Validação cruzada do modelo linear simples de CSR.....	31
Figura 5.3	Validação cruzada do modelo SVR para CRI.....	32
Figura 5.4	Resumo dos valores SHAP do modelo SVR para CRI	32
Figura 5.5	Validação cruzada do modelo SVR para CSR.....	33
Figura 5.6	Resumo dos valores SHAP do modelo SVR para CSR	33
Figura 5.7	Validação cruzada do modelo de florestas aleatórias para CRI.....	36
Figura 5.8	Resumo dos valores SHAP do modelo de florestas aleatórias para CRI ...	37
Figura 5.9	Validação cruzada do modelo de florestas aleatórias para CSR	37
Figura 5.10	Resumo dos valores SHAP do modelo de florestas aleatórias para CSR ..	38
Figura 5.11	Validação cruzada do modelo LightGBM LT para CRI.....	38
Figura 5.12	Resumo dos valores SHAP do modelo LightGBM LT para CRI	39
Figura 5.13	Validação cruzada do modelo LightGBM LT para CSR.....	39
Figura 5.14	Resumo dos valores SHAP do modelo LightGBM LT para CSR	40
Figura 5.15	Validação cruzada do modelo XGBoost para CRI.....	40
Figura 5.16	Resumo dos valores SHAP do modelo XGBoost para CRI.....	41
Figura 5.17	Validação cruzada do modelo XGBoost para CSR.....	41

Figura 5.18	Resumo dos valores SHAP do modelo XGBoost para CSR.....	42
Figura 5.19	Resumo de resultados para CRI e CSR.....	44

LISTA DE TABELAS

Tabela 4.1	Parâmetros da base de dados da qualidade dos carvões.....	52
Tabela 4.2	Lista de todos os parâmetros selecionados.....	52
Tabela 4.3	Parâmetros selecionados para CRI e CSR no modelo linear simples	60
Tabela 4.4	Parâmetros considerados relevantes pelo algoritmo Boruta	60
Tabela 4.5	Resultado da seleção sequencial de parâmetros	61
Tabela 4.6	Hiperparâmetros otimizados para os modelos de aprendizado de máquina	62
Tabela 5.1	Equações selecionadas para CRI e CSR	27

LISTA DE ABREVIATURAS E SÍMBOLOS

ASTM	<i>American Society for Testing and Materials</i>
CRI	<i>Coke reactivity index</i>
CRISP-DM	<i>Cross-Industry Standard Process for Data Mining</i>
CSR	<i>Coke strength after-reaction</i>
EFS	<i>Exhaustive Feature Selection</i>
FSI	<i>Free-Swelling Index</i>
ISSO	<i>International Organization for Standardization</i>
LT	<i>Linear Trees</i>
M40	<i>Micum Strength</i>
SFS	<i>Sequential Feature Selection</i>
SHAP	<i>Shapley Aditive Explanations</i>
SSR	<i>Sum of Squared Residuals</i>
SST	<i>Sum of Squared Totals</i>
SVR	<i>Support Vector Regressor</i>
XGB	<i>Extreme Gradient Boosting</i>
ϵ	Epsilon

RESUMO

Os coques metalúrgicos precisam possuir alta resistência mecânica e baixa reatividade, devido ao seu papel estrutural indispensável no alto-forno, usado na produção de ferro-gusa. No entanto, os carvões minerais que produzem coques de alta qualidade estão se tornando cada vez mais escassos e caros, ao mesmo tempo em que há pressões para reduzir as emissões de gases de efeito estufa na indústria. O cenário futuro para o uso de coque na metalurgia sugere uma tendência de redução gradual do seu uso no reator. Isso significa que o coque restante precisa ter qualidade superior. Para alcançar esse objetivo, é crucial compreender os parâmetros mais relevantes no processo de coqueificação e otimizar as misturas de carvão. As ferramentas computacionais mais modernas podem ajudar a facilitar esse processo, uma vez que conseguem lidar com um grande volume de dados e esgotar as possibilidades propostas na literatura. Procurou-se, então, avaliar o desempenho de modelos de aprendizado de máquina para previsão de CRI e CSR a partir de dados de carvão e misturas. Com esse objetivo em mente, foi realizado um processo de mineração de dados para limpar e processar dados operacionais de quatro anos. Isso foi seguido por uma etapa de engenharia de parâmetros, visitando diversos métodos de previsão de qualidade dos coques. Então, foi realizada uma seleção de parâmetros relevantes que tentam explicar a variabilidade dos dados. Em seguida, usando técnicas simples de regressão e aprendizado de máquina, foram desenvolvidos cinco modelos de previsão para a reatividade do coque (CRI) e mais cinco para a resistência mecânica após reações (CSR). Além disso, a capacidade dos modelos foi avaliada por meio de validação cruzada, e houve uma tentativa de compreender a importância de cada um dos parâmetros considerados relevantes a partir dos valores SHAP. Os resultados mostraram limitações na base de dados utilizada, sobretudo ao considerar os parâmetros escolhidos para explicar os índices, que destoam da literatura. Ainda assim, os resultados demonstraram que o modelo é suficiente para prever a qualidade do coque com cerca de 60% de confiabilidade, de acordo com a métrica de validação cruzada utilizada. Também foi observado que os valores extremos, onde há poucas amostras de dados e fatores externos podem influenciar a qualidade da mistura, tendem a aumentar os erros observados.

Palavras-chave: coque, aprendizado de máquina, mineração de dados, regressão, máquina preditiva, CSR, CRI, qualidade, carvão mineral.

ABSTRACT

Metallurgical cokes need to possess high mechanical strength and low reactivity, due to their indispensable structural role in the blast furnace used in pig iron production. However, the mineral coals that produce high-quality cokes are becoming increasingly scarce and expensive, at the same time as there are pressures to reduce greenhouse gas emissions in the industry. The future scenario for coke usage in metallurgy suggests a gradual reduction trend in its use in the reactor. This means that the remaining coke needs to have superior quality. To achieve this goal, it is crucial to understand the most relevant parameters in the coking process and optimize coal blends. The most modern computational tools can help facilitate this process, as they can handle a large volume of data and exhaust the possibilities proposed in the literature. Therefore, an evaluation of the performance of machine learning models for predicting CRI and CSR from coal and blend data was sought. With this goal in mind, a data mining process was carried out to clean and process operational data from four years. This was followed by a parameter engineering step, exploring various methods for predicting coke quality. Then, a selection of relevant parameters that attempt to explain the data variability was made. Next, using simple regression and machine learning techniques, five prediction models were developed for coke reactivity (CRI), and another five for mechanical strength after reactions (CSR). Additionally, the models' performance was assessed through cross-validation, and an attempt was made to understand the importance of each of the considered relevant parameters from SHAP values. The results showed limitations in the database used, especially when considering the chosen parameters to explain the indices, which deviate from the literature. Nevertheless, the results demonstrated that the model is sufficient to predict coke quality with approximately 60% reliability, according to the cross-validation metric used. It was also observed that extreme values, where there are few data samples and external factors can influence the blend quality, tend to increase observed errors.

Key words: coke, machine learning, data mining, regression, predictive machines, CSR, CRI, quality, coal.

1.0 INTRODUÇÃO

No processo de fabricação de ferro-gusa via alto-forno, o coque metalúrgico é uma matéria-prima imprescindível, uma vez que seu papel dentro do forno é amplo e, do ponto de vista estrutural, insubstituível. Produto da transformação dos carvões minerais, coques de qualidade estão cada vez mais caros, seja pela escassez dos carvões de qualidade, seja pelo cerco cada vez maior no sentido de diminuição das emissões de CO₂, dentre outros fatores.

A busca por otimizar a qualidade dos coques, que já é uma disputa comercial entre as empresas siderúrgicas, passa a ter uma dimensão ambiental, sobretudo porque muitas das técnicas que visam diminuir as emissões dos gases de efeito estufa se baseiam em substituição do coque convencional dentro do alto-forno e, uma vez que a carga metálica não diminui, a carga de coque remanescente precisará de maior qualidade. Entender as propriedades dos carvões de entrada passa a ter uma validade técnica e financeira ainda mais relevante.

Nesse contexto, as ferramentas de mineração de dados e aprendizado de máquina se mostram bastante promissoras, uma vez que são capazes de lidar com grandes volumes de dados de maneira relativamente rápida, permitindo que se faça um tratamento nos dados brutos e a determinação de máquinas preditivas, baseadas nos dados históricos de entrada dos carvões e os parâmetros desenvolvido ao longo das últimas décadas no sentido de prever a qualidade dos coques formados.

O advento da computação e a facilidade do uso de linguagens computacionais, como o Python, ajuda no desenvolvimento desses modelos de previsão de qualidade dos dados, permitindo abordagens robustas de maneira mais simplificada. Porém, mais importante do que a qualidade dos modelos está a qualidade dos dados, que deve ser garantida a fim de carregar a menor quantidade de erros possíveis para o modelo. Esse trabalho visa o desenvolvimento de modelos de previsão de qualidade, além de uma engenharia e seleção criteriosa de parâmetros para cada um deles, buscando abordar a maioria dos trabalhos relevantes no sentido de prever a qualidade dos coques industriais.

2.0 OBJETIVOS

Objetivo Geral

Esse trabalho tem como objetivo avaliar o desempenho de modelos de aprendizado de máquina na previsão de CRI e CSR de coques com base em dados de carvões e misturas.

Objetivos Específicos

- Coletar e tratar as bases de dados de qualidade dos materiais carbonosos, misturas realizadas e ensaios realizados;
- Realizar uma engenharia de parâmetros abrangendo a maior quantidade possível de parâmetros da literatura;
- Selecionar os melhores parâmetros para cada um dos modelos;
- Desenvolver diferentes modelos de previsão de qualidade para obter uma abrangência maior na previsão;
- Explicar os parâmetros selecionados e a importância deles para cada um dos modelos.

3.0 REVISÃO BIBLIOGRÁFICA

Para fins de entendimento do estudo proposto por essa monografia, será apresentada uma revisão dos principais pontos da literatura sobre os modelos de previsão de qualidade do coque a partir de dados de carvão e possíveis aplicações no contexto da siderurgia.

3.1 Carvão Mineral

O carvão mineral é uma rocha sedimentar de coloração escura ou marrom majoritariamente composta por carbono, contando com relevantes presenças de hidrogênio, enxofre, oxigênio e nitrogênio. É um combustível fóssil formado a partir da deposição e conversão de plantas, árvores e materiais orgânicos ao longo de milhões de anos. É preferencialmente utilizado como fonte de energia a partir da sua queima, mas dependendo das propriedades que possui referentes aos materiais originários e ao tempo de carbonificação, também pode ser matéria-prima para a fabricação de coque metalúrgico, sendo estes chamados de carvões coqueificáveis.

Visando apenas os carvões com capacidade de formar coque, o entendimento das suas características é fundamental para a ampliação da compreensão dos coques formados a partir das misturas ou dos carvões individuais de origem. Para tal, diferentes ensaios foram desenvolvidos e normatizados, garantindo a sua repetibilidade e aplicação, tendo como principais testes o índice de inchamento livre (*Free Swelling Index – FSI*) (ASTM D720), o ensaio de dilatométrica Audibert-Arnu (ASTM D4749) e o teste de plastometria Gieseler (ASTM D2639). Além disso, há ainda ensaios como a análise imediata (ASTM D3172), que mede a matéria volátil, a umidade, o teor de cinzas e o carbono fixo (por diferença); a refletância dos macerais da vitrinita (ASTM D2798) (parâmetro amplamente utilizado para definir o rank dos carvões), bem como a sua composição maceral (ASTM D2799), além de análise química das cinzas (ASTM D3174), discriminando os óxidos presentes nas cinzas do carvão. Esse conjunto de informações permite uma visão ampla das características físicas, químicas e reológicas dos carvões coqueificáveis.

O carvão mineral é a matéria-prima com maior custo associado na rota de produção de ferro-gusa, chegando na casa dos 40% (COELHO, et al, 2004), por isso o

entendimento de seus atributos, seja antes ou depois do processo de coqueificação é fundamental.

3.2 Coqueificação

O coque é o produto da transformação dos carvões minerais coqueificáveis após estes serem submetidos a um aquecimento até temperaturas de cerca de 1100 °C. É um material sólido, poroso, infusível, com alto teor de carbono e com alta resistência mecânica. Durante o processo de coqueificação, diversas reações ocorrem no carvão mineral de origem: amolecimento, craqueamento, desvolatilização primária e secundária, formação de semi-coque e formação de coque. Uma das etapas mais importantes dessa sequência é chamada de estágio plástico, em que ocorrem as principais reações de craqueamento e reorganização das cadeias químicas. (LOISSON, et al. 1989) Esse estágio acontece nas temperaturas entre 350 °C e 500 °C e é responsável por explicar grande parte da qualidade do coque metalúrgico formado, seja em termos de reatividade ao CO₂ (CHAN, et al, 1985) ou resistência mecânica (KIDENA, et al).

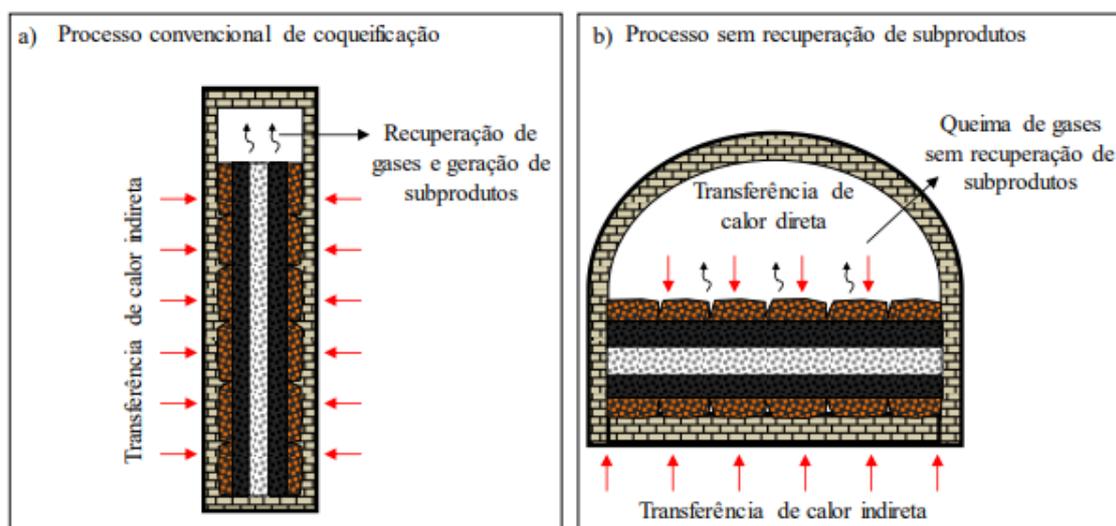
Devido a importância do processo de coqueificação, bem como a análise de sua qualidade, será realizada uma breve revisão sobre alguns aspectos chave do trabalho.

3.2.1 Coqueificação Industrial

No contexto industrial, o coque pode ser produzido de diferentes maneiras, sendo elas divididas em dois grandes grupos: *byproducts*, que tem recuperação de subprodutos e *heat-recovery*, que tem queima dos voláteis para geração de energia.

O primeiro tipo, com recuperação de subprodutos, também é chamado de coqueria convencional. Seus fornos são do tipo vertical, alimentados com a ação da gravidade, e preenchidos completamente no seu interior. Alguns dos subprodutos gerados no processo de coqueificação são o gás de coqueria, o alcatrão, o benzol, o sulfato de amônio e o enxofre (QUITES, 2007). Já o segundo tipo, com recuperação de calor, há a queima do material volátil proveniente dos carvões de entrada, gerando calor que auxilia na própria coqueificação, mas que também pode ser utilizado na geração de energia a partir de caldeiras (BUCZYNSKI, et al, 2018). A figura 3.1 demonstra um comparativo entre os dois tipos de coqueificação industriais:

Figura 3.1 Esquema de comparação entre fornos de coqueificação.



Fonte: FLORES, 2018

3.2.2 Função do Coque no Alto-Forno

O coque desempenha papel fundamental e insubstituível dentro do alto-forno. Primeiramente, ele atua como um agente redutor, participando da reação com o minério de ferro para gerar ferro metálico. Quando submetido a alta temperatura dentro do reator, o carbono contido no coque reage com a atmosfera, que tem presença de CO_2 , e acontece a reação de Boudouard, em que o produto da reação é o CO , o qual então reage com o óxido de ferro para gerar ferro metálico.

Outra função fundamental é fornecer calor suficiente para fundir o minério de ferro e permitir a obtenção do ferro metálico. Quando o coque é queimado no alto-forno, libera calor em grande quantidade, o qual é empregado para aumentar a temperatura do forno e efetuar a fusão do minério de ferro.

O coque também exerce o papel de suporte mecânico para a carga presente no alto-forno, composta pela mistura de minério de ferro, coque e fundentes. Essa função é crucial para garantir que a carga possa suportar o peso dos materiais acima dela, mantendo sua integridade estrutural. Graças a esse papel, os reatores puderam ser cada vez maiores, aumentando então a produtividade da rota. Essa pode ser considerada a mais crítica dentre todas as funções, uma vez que ela faz com que o coque seja insubstituível dentro do reator.

Além dessas funções, o coque é utilizado como carburizador, adicionando carbono ao ferro líquido, sendo a fonte de carbono para o aço. O produto do alto-forno é o ferro-gusa, que contém teores de carbono na casa dos 4% e é a matéria-prima principal do processo de aciaria.

3.2.3 Testes de Qualidade do Coque

Por ser altamente complexo fazer medições e testes de desempenho do coque metalúrgico enquanto está em operação dentro do alto-forno, foram desenvolvidos testes de qualidade visando simular as condições as quais são submetidos enquanto estão em operação dentro dos fornos. Neste tópico, serão abordados os dois principais deles, referentes à reatividade ao CO_2 e a resistência mecânica pós-reação, dois ensaios sequenciais e que são amplamente utilizados nas usinas siderúrgicas: o CRI e o CSR.

O teste CRI é utilizado para determinar a reatividade do coque, um fator importante para o seu desempenho no alto-forno. O teste é realizado em laboratório e envolve os seguintes passos:

Uma amostra de 200 g de coque com partículas na faixa de 19-21 mm é aquecida a 1100°C sob uma pressão de 1 atmosfera de CO_2 por 2 horas. O coque é pré-aquecido e resfriado sob nitrogênio. A perda de massa durante a reação é medida, e a porcentagem de perda de massa é conhecida como reatividade (CRI). O teste é utilizado para determinar a capacidade do coque de resistir à degradação devido à reação com dióxido de carbono no alto-forno.

Já o teste CSR é realizado com a massa residual do ensaio de CRI, submetido a 600 rotações em 30 minutos dentro de um tambor em formato I. Então, toda a massa é passada por uma peneira com dimensão de 9,5mm. O percentual que ficar retido nessa peneira em comparação com a massa inicial é o valor do CSR.

Ambos os ensaios foram desenvolvidos pela Nippon Steel Company e são padronizados pela ASTM International, e o método dos testes é detalhadamente descrito na norma ASTM D5341/D5341M.

3.2.4 Qualidade do Coque no contexto da Descarbonização

O principal desafio da indústria siderúrgica nos próximos anos está na redução das emissões de CO₂, visto que o percentual de emissões globais da produção de aço está entre 7-12% (OWID, 2020). Para tal, existem inúmeras ações que vão na direção de diminuir o *fuel rate* e o *coke rate* do alto-forno, como o uso de coque altamente reativo (NOMURA, et al, 2005), hidrogênio no alto-forno (ZHAO, et al, 2023), biomassa no PCI (WANG, et al, 2015), dentre outras.

Se todas essas medidas forem aplicadas, seja individualmente ou em conjunto, aliado ao fato de que carga metálica do forno não sofrerá diminuição, acarretará a necessidade de aumento da qualidade do coque que irá para o forno, i. e., coques com menores CRI e maiores CSR. Soma-se isso ao fato de que, à medida que as minas de carvão vão sendo mais exploradas, haverá uma maior dificuldade na elaboração de misturas adequadas às novas exigências de reatividade e resistência mecânica.

3.3 Modelos de previsão de qualidade do coque

Prever a qualidade do coque é um tema que faz parte das pesquisas científicas dos carvões há muitos anos. Para fixar uma década, nos anos 1930 o conceito de maceral foi estabelecido por Stopes, sendo o pontapé dos modelos de previsão de qualidade. Inúmeros parâmetros foram desenvolvidos nesse sentido, portanto faremos uma breve revisão sobre eles.

3.3.1 Modelos de Ammosov, Schapiro e Gray

Os parâmetros mais importantes dos modelos de Ammosov (1957), Schapiro (1961) e Gray (1964) são relacionados com a petrografia, mais precisamente com a classificação dos macerais em “reativos” e “inertes”, sendo os primeiros aqueles que participam ativamente do processo de coqueificação, transformando-se, enquanto os segundos permanecem sem quaisquer reações nas suas estruturas. Em resumo, os modelos de Ammosov, Schapiro e Gray sugerem que existe uma relação ótima entre macerais reativos e inertes.

Ammosov e Schapiro consideraram que a concentração relativa de diferentes macerais, bem como a refletância da vitrinita, são melhores indicações do rank do que a

matéria volátil. A partir desse entendimento, separaram em reativos os macerais do grupo da vitrinita, fusinita e 1/3 da semi-fusinita, enquanto os demais foram considerados como inertes. Então, o modelo do coque análogo ao concreto foi desenvolvido por Schapiro, como se a parte reativa fosse o cimento e a parte inerte fosse a areia.

Desses modelos, os principais índices foram o *leaning index* e o *coefficient of coking capacity*, mais tarde redefinidos por Schapiro como *Composition-Balance Index* (CBI) e o *rank index*.

Posteriormente testes realizados baseando-se nesse modelo resultaram em críticas acerca da base do modelo. O comportamento da semi-fusinita depende da origem geográfica dos carvões, uma vez que carvões australianos tem proporção de reativos/inertes nesse grupo maceral maior do que carvões de outras regiões. Além disso, esses modelos foram criticados por levarem em consideração apenas alguns aspectos das petrografias dos carvões, deixando de lado outros que poderiam ser importantes, como a distribuição das partículas dos grupos macerais.

3.3.2 Fator G de Simonis

O modelo do fator G de Simonis (1966) nasceu por conta da crítica anterior de que a matéria volátil não consegue abranger uma ampla gama de carvões coqueificáveis. Nesse modelo, estudado por Simonis, são levados em consideração os parâmetros da dilatométrica.

O fator G é definido pela equação 3.1:

$$G = \frac{T_1 + T_2}{2} \times \frac{C + D}{T_1 D + T_2 C}$$

sendo T_1 – Temperatura de amolecimento; T_2 – Temperatura de ressolidificação; C – Contração máxima; D – Dilatação máxima.

Nos casos em que apenas a contração acontece, o fator G é igual a 0; em casos em que a dilatação é igual à contração, o fator G é igual a 1; se a contração for maior que a dilatação, o fator G é menor do que 1; finalmente, se a dilatação for maior que a contração, o fator G é maior do que 1.

O fator G de Simonis foi desenvolvido visando prever resistência à frio, especialmente o índice M40 (ISO 556), relacionando o G com parâmetros dos carvões de origem.

Assim como os modelos petrográficos mencionados anteriormente, o modelo de Simonis foi criticado por conta da dificuldade de aplicação a carvões fora dos modelos estudados. Segundo NORTH et al (2018), Radek e Gibson criticaram e modificaram o modelo de Simonis, buscando aumentar sua aplicabilidade, mas não encontraram resultados satisfatórios tentando prever os índices de resistência à frio como o M40.

Uma vez que parte do CSR é explicada pela resistência à frio, “G” pode ser um elemento importante para a previsão de qualidade proposta por esse trabalho. O modelo proposto por Simonis encontra certa aceitação dentre os estudos de previsão de qualidade dos coques, mas também teve suas limitações encontradas em estudos posteriores.

3.3.3 Modelo de previsão de CRI e CSR por comportamento aditivo

O modelo mais simples para prever o CSR de uma mistura é a consideração da aditividade do CRI e CSR dos carvões individuais. Diferentes pesquisadores testaram essa aditividade e obtiveram resultados discrepantes entre si. Ao longo do tempo, tornou-se consenso de que existe alguma interação química e física entre os carvões individuais, contestando a aditividade dos índices de qualidade.

Com base na discrepância existente entre os pesquisadores e na compreensão das interações que ocorrem entre os componentes individuais nas misturas, é possível assumir que o comportamento aditivo não é apropriado para todos os tipos de carvão coqueificável. Se fosse viável prever de maneira precisa o CRI e CSR com base nos índices de cada carvão individual, não haveria uma razão significativa para conduzir pesquisas com o intuito de prever as propriedades do coque resultante. Além disso, nos cenários em que os carvões constituintes não conseguem formar coque sem serem misturados com outras variantes, a abordagem aditiva torna-se inviável.

3.3.4 Modelos baseados no rank

Diversos modelos de previsão são baseados no rank dos carvões. Porém, há diferentes maneiras de considerar esse rank, sendo três delas a refletância média das vitrinitas, a matéria volátil e a distribuição da refletância.

O principal método de determinação do rank dos carvões está na refletância média das vitrinitas. Diversos métodos de previsão do CRI e do CSR foram tentados utilizando-se esse parâmetro, com diferentes correlações dependendo do tipo de experimento – laboratorial ou operacional – e da natureza da mistura – individual ou misturado. Os resultados não foram satisfatórios, tendo uma variedade de formatos e tendências nas curvas refletância x CRI e refletância x CSR. Além disso, pesquisadores (NORTH et al, 2018) encontraram relações lineares entre alguns parâmetros que dependem do rank, como o log da máxima fluidez e a refletância das vitrinitas, indicando a possibilidade de utilizar outros parâmetros para a previsão da qualidade.

Outro parâmetro historicamente importante na determinação do rank dos carvões é a matéria volátil. Assim como a refletância da vitrinita, não existem correlações claras, mas sim uma diversidade bastante grande de utilização da matéria volátil, especialmente em base seca isenta de cinzas (bsic), com diferentes resultados entre si.

Por fim, a distribuição das vitrinitas em V-tipos foi analisada na previsão do CSR dos coques. Um dos modelos estudados por Kumar indica que a refletância do grupo dos V-tipos entre V9 e V13 são um método mais eficaz de prever o CSR do que o uso da refletância média, previamente abordado nessa revisão. Ainda assim, nenhum destes modelos considerando os V-tipos foi capaz de encontrar alguma correlação direta e aplicável a todos os tipos de carvões.

3.3.5 Modelos baseados nas cinzas e matéria mineral

Considerando a natureza do ensaio de CRI e a sua relação com a reação de Boudouard, a matéria mineral e as cinzas presentes na mistura ou nos carvões individuais têm um papel importantíssimo, uma vez que tais reações de gaseificação podem ser catalisadas pela presença de alguns elementos. O principal parâmetro derivado dos modelos que levam em consideração a química das cinzas é o índice de basicidade, que é a relação entre os óxidos básicos e os óxidos ácidos.

Entretanto, um parâmetro individual não é capaz de representar a reatividade e os complexos processos que acontecem dentro dos altos fornos. A partir desse entendimento, outros índices relacionados com a química das cinzas foram desenvolvidos, como o MMCI (COIN et al, 1997) e o MCI (YANG et al, 2001), em que há uma hierarquia na importância de cada óxido na propriedade mecânica do coque.

Assim como modelos anteriores, os resultados não eram extrapoláveis a todos os tipos de carvão, seja pela sua natureza, seja pela sua origem geográfica (que tem influência direta nas suas propriedades intrínsecas).

3.3.6 Modelos baseados em propriedades termoplásticas

Assim como no modelo proposto por Simonis, os modelos baseados em propriedades termoplásticas e de inchamento levam em consideração que o estágio plástico dos carvões influencia diretamente na resistência mecânica do coque produzido, uma vez que tem relação direta com a microtextura. Porém, tal qual no modelo de Simonis, uma aplicação direta dos valores dos ensaios de fluidez não encontrou correlação direta com a qualidade dos coques. Então, Jordan (2008) desenvolveu um fator baseado no fator proposto de Simonis, com a intenção de agregar todos os elementos do ensaio de plastometria em um único parâmetro, mostrado na equação 3.2:

$$\text{Fator de Fluidez} = \frac{T_1 + T_3}{2} \times \frac{MF + (T_3 - T_1)}{MF \times T_2}$$

Sendo T_1 – Temperatura de amolecimento; T_2 – Temperatura de máxima fluidez; T_3 – Temperatura de ressolidificação; MF – Máxima fluidez.

Em sua dissertação, Jordan explica que o fator de fluidez desenvolvido tem relação com o M40 e o CSR de maneira distinta para carvões com diferentes ranks, encontrando correlações lineares que dependem da refletância máxima, do intervalo plástico e do fator de fluidez (que, por sua vez, também leva em consideração o intervalo plástico). Entretanto, Jordan não demonstra os resultados encontrados nas previsões nem para M40 nem para CSR. Além disso, é difícil atestar a aditividade desse fator,

considerando que suas relações de previsão mudam com a natureza dos carvões e que não parece ter sido um parâmetro relevante para a comunidade científica da área, não tendo sua aplicação replicada em outros artigos acadêmicos, demonstrando uma certa fragilidade.

3.3.7 Outros modelos

Distintas abordagens de previsão de qualidade do coque foram estudadas e descritas por NORTH et al, 2018 em diferentes cenários. Algumas delas, por exemplo, exploraram a relação entre o CRI e o CSR, notoriamente com uma alta correlação entre si, uma vez que boa parte do CSR é explicada pelo CRI. Entretanto, como não é o foco desse trabalho ter como parâmetro de entrada um ensaio de coque, não serão aprofundados modelos que utilizem essas métricas.

Há, ainda, modelos que utilizam parâmetros de textura, de reatividade, de inertes, de porosidade, de estrutura porosa, dentre outros. Como são abordagens ainda experimentais e com muitas limitações mesmo para carvões individuais e controlados, não serão aprofundados esses modelos.

3.4 Ferramentas computacionais

Considerando o avanço computacional exponencial dos últimos anos, é possível utilizar ferramentas e algoritmos computacionais para selecionar os melhores parâmetros e montar modelos de regressão e classificação. Será feita uma breve revisão acerca das ferramentas e dos algoritmos utilizados para desenvolver os modelos neste trabalho.

3.4.1 Python

Python é uma linguagem de programação de código aberto de alto nível, criada e lançada pela primeira vez em 1991. É uma ferramenta computacional que se destaca pela sua facilidade de uso, sua interface amigável para quem começa na programação e por suas bibliotecas, que podem ser instaladas e carregadas de maneira ampla e fácil. As bibliotecas que tratam de estatística, regressões e aprendizado de máquina são

abundantes, o que facilita sua aplicação em modelos como o proposto por esse trabalho. (PYTHON)

Na área de carvão e coque, a codificação em Python está cada vez mais presente, como na codificação da previsão do índice de inchamento de carvões coqueificáveis (XIANG et al, 2021), avaliação de metodologias alternativas de microespectrografia cabonosa de carvões (RANTITSCH et al, 2014), análise de utilização de carvões para o cumprimento das metas de descarbonização (GOSENS J. et al, 2022), previsão da quantidade de H₂ e CH₄ no gás de coqueria utilizando algoritmos de aprendizado de máquina (LEI Y. et al, 2023), dentre outras aplicações.

Por conta da sua curva de aprendizado relativamente pequena, é possível fazer aplicações na linguagem sem um estudo muito aprofundado de suas nuances, possibilitando aos especialistas nas áreas técnicas utilizarem a linguagem para diminuição de tempo na resolução de problemas relacionados às suas respectivas áreas.

3.4.2 Algoritmos de seleção de parâmetros

Visando a elaboração de modelos de regressão, uma das etapas mais importantes dessa configuração está na seleção dos parâmetros utilizados. Por conta disso, serão abordados alguns algoritmos de seleção de parâmetros que foram utilizados nesse trabalho.

3.4.2.1 BorutaSHAP

O BorutaSHAP é um método de seleção de características em Python que combina o algoritmo de seleção de características Boruta com os valores das explicações de Shapley.

O Boruta é um método utilizado para identificar os parâmetros mais importantes em um conjunto de dados para tarefas de aprendizado de máquina, sejam elas de classificação ou regressão. Seu funcionamento dá-se pela remoção de características com baixos escores de importância, indicando que elas provavelmente são inúteis para a tarefa em questão. Para fazer essa avaliação, o algoritmo cria parâmetros fantasma de maneira aleatória, sem qualquer relação com os dados originais, e realiza os testes de regressão, comparando os parâmetros da base de dados original com estes criados por ele. Então,

aqueles parâmetros que explicarem a variação dos dados de maneira pior do que os criados artificialmente para avaliação são considerados como irrelevantes para a tarefa de regressão. Para a montagem de máquinas preditivas, o Boruta tem um papel fundamental, especialmente em bases de dados que contenham muitos parâmetros avaliados. (Python Package Index, 2023)

O uso do algoritmo de seleção de parâmetros Boruta é algo mais restrito, uma vez que tarefas de aprendizado de máquina são recentes, mas dentro da área metalúrgica já é possível ver sua aplicação no desenvolvimento de novas ligas (HU M et al, 2023), previsão das fases das ligas metálicas (MACHAKA R, 2023), dentre outros.

Por outro lado, os valores de Shapley são uma explicação para a importância de cada característica com base em sua contribuição para a previsão de um modelo. No contexto do BorutaSHAP, os valores de Shapley são usados para medir a importância das características escolhidas pelo algoritmo Boruta. (Shapley Additive Explanations, 2023)

No contexto de carvões e coques, HOMAFAR (2022) utilizou as explicações de SHAP para demonstrar a importância de parâmetros de entrada na modelagem do índice de inchamento livre (FSI) e na máxima fluidez dos carvões, enquanto LIU (2023) utilizou os valores para explicar anomalias nos ensaios de análise imediata. Os valores SHAP servem para diminuir o caráter de “caixa-preta” que modelos de aprendizado de máquina tem, atribuindo valores e análises mais simplificadas para a análise do operador do que valores computacionais complexos.

3.4.2.2 Seleção sequencial de parâmetros (SFS)

A seleção sequencial de características (*sequential feature selection*, SFS) é uma técnica utilizada em aprendizado de máquina para escolher um subconjunto de parâmetros a partir de um conjunto maior. Trata-se de uma abordagem que envolve adicionar ou remover características iterativamente com base em seu impacto no desempenho de um modelo.

A seleção sequencial de características pode ser realizada de várias maneiras, incluindo seleção progressiva, eliminação reversa e busca bidirecional. A seleção progressiva começa com um conjunto vazio de características e adiciona iterativamente a característica mais promissora, a eliminação reversa começa com todas as características e remove iterativamente a característica menos promissora e a busca bidirecional combina

a seleção progressiva e a eliminação reversa, adicionando e removendo características simultaneamente. (MLxtend, 2023)

Embora seja um algoritmo robusto em propostas de aprendizado de máquina, sua aplicação ainda é restrita no meio científico. Algumas aplicações na área médica (ADMASU, 2022) e na área comercial (YULIANTI, 2020) demonstram o potencial do algoritmo.

3.4.2.3 Seleção exaustiva de parâmetros (EFS)

A seleção exaustiva de características (*exhaustive feature selection, EFS*) é uma técnica usada em aprendizado de máquina para escolher o melhor subconjunto de características a partir de um conjunto maior. Ela envolve a busca por todas as combinações possíveis de características e avalia o modelo em cada subconjunto de características. Esse método pode ser computacionalmente custoso, especialmente para conjuntos de dados grandes com muitas características, uma vez que a cada parâmetro que é adicionado, o tempo necessário para realização do processo é dobrado. Por testar todas as combinações possíveis a partir de uma métrica selecionada pelo usuário, o algoritmo garante que o melhor subconjunto de características seja selecionado. (MLxtend, 2023)

3.4.3 Algoritmos de regressão em aprendizado de máquina

Nessa seção, será feita uma breve abordagem a alguns algoritmos de regressão utilizando aprendizado de máquina, bem como o uso dessas técnicas no contexto da siderurgia, suas principais características e sua abordagem em Python.

3.4.3.1 Florestas aleatórias

O algoritmo de florestas aleatórias pertence à família dos *ensembled*, i. e., um modelo que é uma combinação de diferentes modelos individuais para buscar uma melhoria no resultado. O algoritmo base das florestas aleatórias é o modelo de árvore de decisão, em que a classificação ou a previsão baseia-se em nós e ramos, sendo os nós representados pela decisão ou teste e os ramos o caminho para seguir até o próximo nó

ou à previsão final. A figura 3.2 representa uma esquematização visual das árvores de decisão.

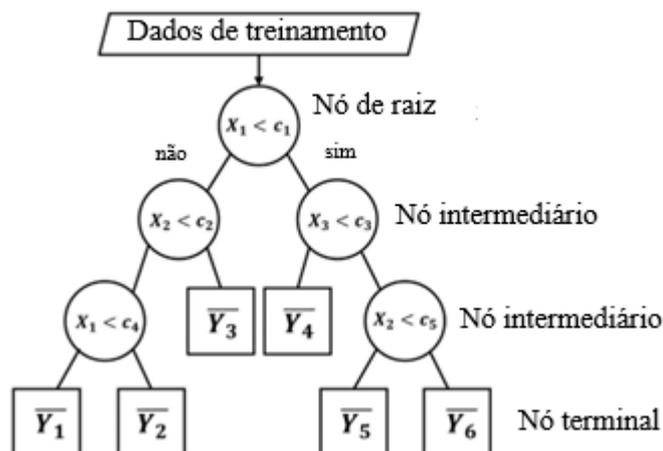


Figura 3.2 Esquema de representação das árvores de decisão para regressões

Fonte: adaptado de SHIN.

Para o algoritmo de florestas aleatórias, são criadas várias árvores de decisões, utilizando diferentes parâmetros nas construções dessas árvores. Por fim, a previsão final é determinada pela média das previsões das árvores individuais, tanto para classificação quanto para regressão.

O algoritmo de florestas aleatórias é um dos modelos mais bem aceitos dentro da comunidade científica, sobretudo na modelagem moderna de propriedades dos carvões e dos coques. CHELGANI et al desenvolveram um bom modelo de previsão do FSI através de parâmetros de entrada dos carvões utilizando, com uma correlação com $R^2 = 0.93$, enquanto HOMAFAR et al utilizaram diferentes modelos, dentre eles as florestas aleatórias, para modelar FSI e máxima fluidez, ainda que esse estudo tenha influência demasiada pela umidade, TIWARY et al desenvolveram um modelo automatizado de classificação dos macerais a partir de árvores aleatórias com uma taxa de assertividade acima de 90%, dentre outros.

3.4.3.2 LightGBM

O LightGBM é um algoritmo de aprendizado de máquina também pertencente à família dos *ensembled*, sendo um modelo de *gradient boosting*, que aprende ao longo do treinamento. Ele também utiliza o algoritmo de árvores de decisão como base, porém de maneira diferente do modelo de florestas aleatórias. Primeiramente, o LightGBM divide os dados na forma de intervalos ou histogramas para construir as árvores e, então, cria árvores de decisão sequenciais, onde cada uma delas se concentra em melhorar o desempenho da anterior, aprendendo com os seus erros. Ao final, é feita uma média ponderada das árvores, dando mais peso para aquelas que tiveram o melhor desempenho, e a previsão é realizada. O LightGBM é um algoritmo paralelizado nativamente, i.e., utiliza todo o poder bruto de processamento disponível do computador do usuário.

TANG et al utilizaram o algoritmo, dentre outros, para fazer uma classificação de partícula no leito de sinterização, enquanto LI et al demonstraram o uso da técnica em comparação com outras para sensores dos altos fornos. Além disso, outros artigos mencionados anteriormente também utilizaram a técnica para desenvolver as regressões e classificações.

3.4.3.3 Máquina de vetores de suporte

A Máquina de Vetores de Suporte (SVM), também conhecida como Support Vector Machine, é um algoritmo de aprendizado que constrói um modelo para atribuir novos exemplos a um grupo específico. As SVMs são empregadas para classificar dois grupos de dados, desenhando linhas (hiperplanos) para separar os grupos conforme os padrões identificados. Em linhas gerais o objetivo do SVM é traçar o melhor plano que separe os dados em dois grupos, para tarefas de classificação.

O Support Vector Regressor (SVR), por sua vez, é uma versão da SVM voltada para tarefas de regressão. O modelo produzido pela classificação de vetores de suporte depende apenas de um subconjunto dos dados de treinamento, uma vez que a função de custo para construção do modelo ignora quaisquer dados de treinamento próximos à previsão do modelo. O objetivo principal do algoritmo SVM é encontrar o hiperplano

ótimo em um espaço N-dimensional que possa separar os pontos de dados em diferentes classes no espaço de características. O hiperplano busca que a margem entre os pontos mais próximos das diferentes classes seja o mais amplo possível. O SVR é usado tanto para tarefas de regressão linear quanto não linear.

XIANG et al utilizaram o modelo de SVM para a previsão do índice de inchamento utilizado na Europa (*Gray-King coke type*) enquanto TUTTLE et al desenvolveram um modelo de classificação da qualidade da combustão dos carvões e uma previsão das emissões dos gases NO_x utilizando o SVM. Embora não seja um modelo tão utilizado quanto os outros, ainda assim apresenta resultados satisfatórios.

3.4.3.4 Extreme Gradient Boosting – XGBoost

Extreme Gradient Boosting ou XGBoost é um algoritmo de aprendizado de máquina da família dos *ensembled*, também com árvores de decisão como modelo-base. Assim como o LightGBM, é um modelo do tipo *gradient boosting*, que tem uma taxa de aprendizado a partir da previsão das árvores individuais anteriores. A diferença entre os dois modelos está na estruturação das árvores, já que o LightGBM utiliza histogramas para separar os dados, enquanto o modelo de XGBoost explora todas as possíveis divisões de dados. Além disso, o modelo LightGBM expande a árvore a partir do nó com menor erro, enquanto o XGBoost preenche todas as árvores até partir para a próxima, o que resulta em um tempo de construção muito mais lento se comparado ao LGBM.

RZYCHON´ et al utilizaram o algoritmo de XGBoost para prever o índice Roga Index (RI), que mede a capacidade de inchamento de maneira análoga e correlacionada ao FSI e ao Grey-King, YUAN et al utilizaram o XGBoost para tentar prever o CRI e o M10, além dos outros artigos citados anteriormente que utilizaram a técnica para comparar com outras. O XGBoost é um dos modelos de aprendizado de máquina mais populares e utilizados na comunidade científica.

4.0 MATERIAIS E MÉTODOS

O processo utilizado para o desenvolvimento do modelo de previsão de qualidade foi realizado em algumas etapas, seguindo como base a metodologia CRISP-DM. Primeiramente, levantamento dos dados necessários de misturas realizadas, qualidade dos materiais carbonosos individuais, tempos de fornos e datas dos desembarques dos navios e ensaios de qualidade do coque produto. A base de dados foi enviada por uma empresa siderúrgica parceira na realização desse trabalho. A disposição da coleta de dados está demonstrada na figura 4.1:

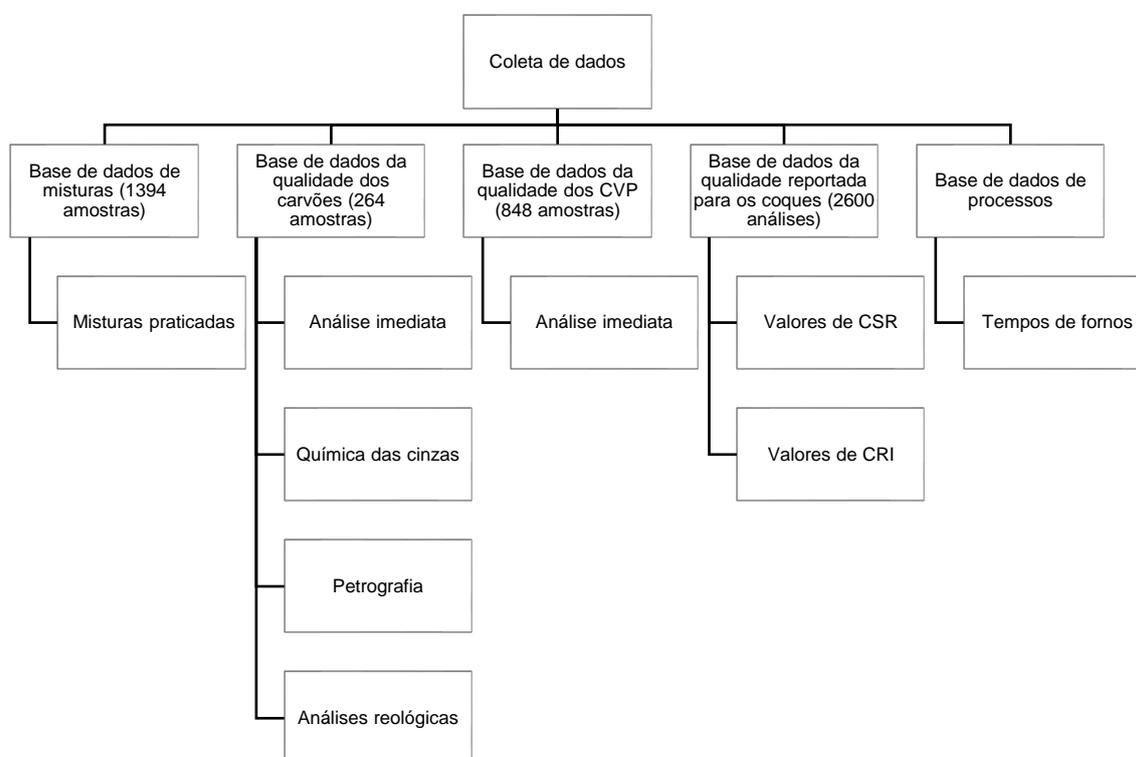


Figura 4.1 Resumo da coleta dos dados.

As bases de dados foram recebidas e tratadas de maneira sequencial, da esquerda para a direita de acordo com a figura 4.1, e serão analisadas ponto a ponto a seguir.

A base de dados das misturas foi retirada de arquivos históricos das trocas de mistura realizadas entre janeiro de 2019 e fevereiro de 2023, documentadas com os

percentuais, o nome da mistura e a data para a troca. Foi considerado o arquivo mais recente como a troca proposta, nos casos em que havia mais de um. Foi desenvolvido um código em Python para extrair as informações dos percentuais das misturas, do nome do navio em que os carvões chegaram e a data da troca da mistura. Além disso, as trocas de misturas propostas nem sempre coincidiam com a data da troca efetiva.

Já a base de qualidade dos carvões vem de dados históricos das análises feitas em laboratório, relacionando os carvões com seus navios de origem. Os ensaios realizados foram reológicos, química das cinzas, petrografias e análise imediata. Grande parte dos carvões entre 2017 e 2022 tinham todas as análises completas, com exceção de alguns períodos em que alguns equipamentos da empresa que forneceu a base de dados não estavam funcionando.

A base de dados do CVP conta apenas com análise imediata e foi realizada em amostras dentro dos silos dos carvões.

A base de dados da qualidade dos coques é baseada na análise do CRI e do CSR dos carvões, que serão as variáveis-alvo, enquanto a base de dados dos tempos de forno foi retirada da operação industrial.

É notório que a quantidade de fontes de erro oriundos das bases de dados pode resultar numa perda de qualidade da construção dos modelos e do resultado, uma vez que muitos dados precisam de tratamento e a confiabilidade não é plena. Entretanto, nas seções a seguir, serão demonstradas as técnicas utilizadas para mitigar esses problemas, ainda que seja relevante mencionar que modelos de aprendizado de máquina não possuem nenhuma informação prévia, sendo altamente dependentes da qualidade dos dados para entregar resultados satisfatórios.

4.1 Tratamento dos dados

Em posse dos dados brutos das bases de dados, foi utilizado o Python para desenvolver a base de dados unificada para a criação dos modelos. O tratamento e a elaboração dessa base foram feitos de maneira sequencial, seguindo a mesma ordem do recebimento das bases de dados brutas e serão demonstrados ponto a ponto a seguir.

Todas as eliminações de valores atípicos que serão demonstradas a seguir foram realizadas seguindo a técnica estatística de intervalos interquartis, considerando 1.5 o

valor positivo e negativo, por ser os valores mais utilizados nas técnicas estatísticas (Towards Science, 2022)

A base de dados das misturas não foi preciso tratar, uma vez que na sua elaboração já foi feito o ajuste disponível.

Iniciando pela base de dados da qualidade dos carvões, primeiro foi realizada uma inspeção visual para identificar se não havia nenhuma anomalia nos dados. Alguns carvões apresentavam erros de digitação, bem como uma ausência de formatação em alguns preenchimentos, mostrando uma falta de padronização na coleta dos dados. Corrigidas as anomalias visuais, foi feita uma análise de dados faltantes. Todos os carvões que tivessem mais do que 30% de dados faltantes foram removidos da base; já aqueles parâmetros que tinham menos do que 70% dos dados preenchidos não foram considerados. Com isso, a base de dados dos carvões ficou com 264 amostras e com os seguintes parâmetros demonstrados na tabela 4.1, no item apêndice.

Então, foi realizada uma inspeção manual na qualidade dos dados, como nos parâmetros que fecham 100% entre si, realizando-se uma transformação para uma mesma base, evitando problemas de formatação. Posteriormente, os dados dos carvões foram completos com a média dos carvões de mesma origem. Apesar de ser uma fonte de erro, foi considerado melhor do que retirar o carvão inteiro, levando em consideração o corte feito anteriormente em carvões com menos dados disponíveis. Por fim, caso não houvesse outros carvões da mesma mina, foi aplicado um algoritmo de aprendizado de máquina chamado *MissForest* para completar os dados faltantes, que completou os dados baseado em carvões de mesmo teor de voláteis e poder coqueificante. Após a base estar completa, foi realizada uma etapa de engenharia de parâmetros, criando-se o maior número de parâmetros possíveis dentre os disponíveis na literatura, independentemente do desempenho reportado pelos artigos acadêmicos. Além disso, uma última etapa de engenharia de parâmetros foi realizada aplicando as operações matemáticas de quadrado, inverso, logaritmo natural e exponencial em cada um dos índices, visando ampliar as possibilidades para o modelo de regressão linear simples. Com essas técnicas, o objetivo foi mitigar ao máximo as limitações da falta de dados na base dos carvões.

Para a base de dados dos CVP, o grande desafio foi a atribuição temporal. Ainda que tenha análises diárias, a base de dados definitiva utiliza apenas trocas de misturas únicas, fazendo com que ainda que tenha mais de 800 amostras, apenas cerca de 200 foram utilizadas. Embora houvesse a possibilidade de se utilizar a base repetindo os

valores das misturas ao longo dos dias, ao final da elaboração da base ainda se diminuiria para as 233 amostras, transformando esse problema em uma das fontes de erro com a maior complexidade associada para se resolver futuramente. Em relação aos dados, foram utilizados ensaios realizados em amostras dos silos utilizados nas misturas.

A base de dados dos CRI e CSR praticados tinham como maior desafio a eliminação de valores atípicos, criação da média dos CRI e CSR praticados durante o tempo em que a mistura esteve sendo realizada na operação e principalmente a atribuição temporal, uma vez que o forno da empresa demora entre 3 e 4 dias para obter o coque a partir do enformamento das misturas. Na etapa de confecção da base unificada será descrito o método utilizado para definir o valor em 4 dias.

Por fim, na base de dados com tempos de fornos foi realizada uma remoção de valores atípicos e uma média dentre todos os fornos e, depois, uma média de tempos de fornos realizados durante o tempo em que a mistura esteve operando no forno.

Em posse de todas as bases, foi criada uma base de dados unificada, onde o objetivo foi associar os carvões praticados com a qualidade reportada na base de dados da qualidade, de maneira aditiva. A atribuição temporal dos carvões só foi feita em carvões cujos navios não possuíam análises e foi testada em um intervalo de 20 a 140 dias, sendo escolhido o valor de 30 dias considerando o melhor R^2 da validação cruzada do modelo linear simples. Ou seja, o modelo considera um intervalo máximo de 30 dias no passado para fazer a atribuição a carvões na base de dados da qualidade e, caso não haja nenhum carvão com aquele nome dentro dos 30 dias, é atribuído o valor da média a ele. Já os carvões cujos navios possuíam análises foram atribuídos pelo nome do navio, sem regra temporal. Depois, outra etapa da engenharia de parâmetros foi aplicada a fim de criar mais parâmetros, relacionados com os dados das misturas e não a aditividade dos parâmetros criados anteriormente. Esses novos parâmetros desenvolvidos a partir da qualidade aditiva recebem o “*blend*” no nome. Por fim, um corte foi realizado nas misturas que ficaram menos do que dois dias em operação na planta industrial. Esse corte também foi otimizado a partir da mesma métrica da atribuição temporal, com o objetivo de evitar análises de CRI e CSR com poucas amostras.

Ao final, após todos os filtros e cortes, a base de dados teve 233 misturas únicas consideradas com características atribuídas e valores médios de tempo de forno, CRI e CSR relacionadas. O total de parâmetros elaborados, seguindo como referência NORTH et al (2018), está demonstrado na tabela 4.2, no item apêndice.

Ainda que tenham sido tomados todos os cuidados e boas práticas estatísticas para mitigar os erros, é notório que há um somatório de problemas ao longo do tratamento das bases e que isso tem potencial de gerar erros. Entretanto, dada a base de dados disponibilizada para a realização desse trabalho, as técnicas aplicadas visam reduzir ao máximo possível a influência dessas incertezas no resultado definitivo.

4.2 Criação dos modelos de previsão

Com a base de dados definitiva, os modelos de previsão foram divididos em lineares e aprendizado de máquina, pois as estratégias para seleção de parâmetros e otimização dos hiperparâmetros foi diferente para cada um dos modelos.

O modelo linear simples foi desenvolvido utilizando a ferramenta *LinearRegression* da biblioteca Scikit-Learn, enquanto os modelos de aprendizado de máquina possuem bibliotecas próprias. Os modelos selecionados foram o *Support Vector Regressor* (SVR), o modelo de florestas aleatórias, o modelo de LightGBM com árvores lineares e o modelo de XGBoost, totalizando 5 para cada um dos parâmetros previstos. Na figura 4.2, um resumo da escolha dos modelos e a principal característica de cada um deles.

Regressão Linear Simples	SVR	Florestas Aleatórias	LightGBM LT	XGBoost
<ul style="list-style-type: none"> • Modelo com alto viés • Fácil explicabilidade 	<ul style="list-style-type: none"> • Lida bem com <i>outliers</i> • Capta relações não-lineares 	<ul style="list-style-type: none"> • Combinação de árvores de decisão • Reduz chance de <i>overfitting</i> 	<ul style="list-style-type: none"> • Combinação de árvores de decisão lineares • Árvores com mais folhas 	<ul style="list-style-type: none"> • Árvores aprendem com o erro anterior • Árvores com maior profundidade

Figura 4.2 Resumo da escolha de modelos e a principal motivação.

4.2.1 Criação do modelo linear simples

A metodologia para a elaboração do modelo de regressão linear foi desenvolvida pensando na velocidade de execução de modelo e a natureza da relação linear. Ao utilizar a base de dados completa inclusive com os parâmetros criados a partir de operações matemáticas, totalizando 3633 parâmetros, a sequência de ações tomadas para elaboração do modelo teve o seguinte passo a passo, demonstrado na figura 4.3.

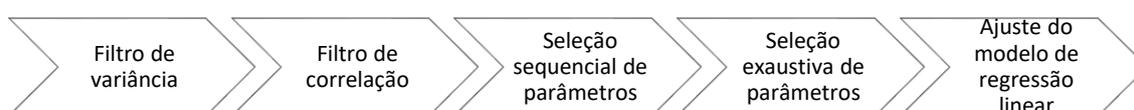


Figura 4.3 Resumo da sequência de ações para criação do modelo linear.

O primeiro filtro realizado nos parâmetros foi o de variância. Qualquer parâmetro cuja variância é menor do que 0.001 foi removido do modelo, uma vez que não tem representatividade estatística para explicar os dados, já que a ideia é que a variação dos parâmetros de entrada explique a variação do CSR e CRI. Com isso, a base passou a ter 560 parâmetros.

Posteriormente, foi realizado um filtro de correlação. Quaisquer parâmetros que possuam correlação linear entre si maior do que 0.70 foram removidos do modelo, uma vez que parâmetros altamente correlacionados tendem a captar redundantes variabilidades nos dados, diminuindo assim a capacidade de previsão do modelo. O total de parâmetros após essa etapa foi de 532.

Então, o algoritmo de seleção sequencial de parâmetros da biblioteca mlxtend foi utilizado com o número máximo de parâmetros definido em 6 e com a seleção regressiva. Isso significa que o algoritmo partiu da base de dados com todos os 532 parâmetros e foi testando a remoção de um a um, a cada rodada diminuindo o número de parâmetros. Então, entregou o melhor conjunto de 6 parâmetros de acordo com a métrica de menor erro médio quadrático.

Por fim, esse melhor conjunto de parâmetros passou pela seleção exaustiva, que testou as 63 combinações possíveis de parâmetros e entregou o resultado também baseado

no menor erro médio quadrático. Os parâmetros selecionados para o CRI e o CSR estão na tabela 4.3, no item apêndice.

4.2.2 Criação dos modelos de aprendizado de máquina

Para os modelos de aprendizado de máquina, a figura 4.4 mostra o resumo das ações tomadas para seleção dos parâmetros, com a exceção do Boruta para SVR, uma vez que o modelo tem um treinamento bastante rápido e viabiliza ir direto para a seleção sequencial de parâmetros. Nos demais itens, será diferenciado o modelo SVR dos demais modelos de aprendizado de máquina.

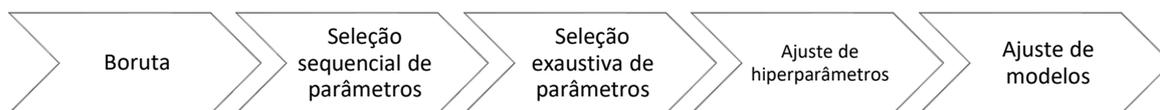


Figura 4.4 Resumo da sequência de ações para criação dos modelos de aprendizado de máquina.

Primeiramente, para todos os modelos, a base de dados foi utilizada sem os parâmetros gerados a partir de parâmetros matemáticos, uma vez que os algoritmos de aprendizado de máquina possuem abordagens mais robustas para lidar com relações não lineares. Então, com exceção do modelo SVR, foi aplicado o algoritmo de seleção de parâmetros Boruta, utilizando como estimador o algoritmo de florestas aleatórias, com o número de florestas definidos em 1000. A tabela 4.4 no item apêndice demonstra os parâmetros considerados relevantes pelo Boruta.

Posteriormente, com os parâmetros selecionados como relevantes pelo algoritmo Boruta, foi realizada uma seleção sequencial de parâmetros, sem limitação do número de parâmetros a ser escolhidos pelo modelo. O estimador selecionado foi o modelo de florestas aleatórias, por conta da sua velocidade de treinamento. Os parâmetros escolhidos para CRI e CSR estão demonstrados na tabela 4.5, no item apêndice.

Ao término da etapa da seleção sequencial, foi feita a seleção exaustiva de parâmetros, testando cada uma das combinações. A melhor combinação para os quatro

cenários foi aquela com todos os parâmetros considerados, portanto não houve alteração nessa etapa. Ao fim, cada um dos modelos de aprendizado de máquina teve seus hiperparâmetros ajustados por uma otimização bayesiana de parâmetros, uma técnica em Python que determina um espaço de busca e testa os resultados em diferentes condições, aprendendo com cada teste anterior e melhorando a previsão o máximo possível. Cada modelo específico tem seu conjunto de hiperparâmetros, que são esmiuçados nas respectivas bibliotecas. O resultado da otimização está demonstrado na tabela 4.6, no item apêndice.

Ao término dos ajustes, foram gerados gráficos das explicações de Shapley, a segunda parte do algoritmo de BorutaSHAP, para explicar a importância de cada um dos parâmetros escolhidos e a sua influência nos resultados das previsões. Os valores SHAP serão demonstrados para cada um dos modelos de aprendizado de máquina na seção de resultados.

Para um melhor entendimento dos parâmetros selecionados, todos eles estão apresentados com suas respectivas equações ou explicações no item apêndice.

4.3 Validação cruzada

Devido à quantidade relativamente limitada de amostras para abordar um problema de regressão, optou-se por não adotar a divisão convencional entre conjunto de treinamento e teste. Em vez disso, uma validação cruzada do tipo Leave-One-Out (LOO) foi empregada.

Na figura 4.5 um esquema representando a validação cruzada LOO.

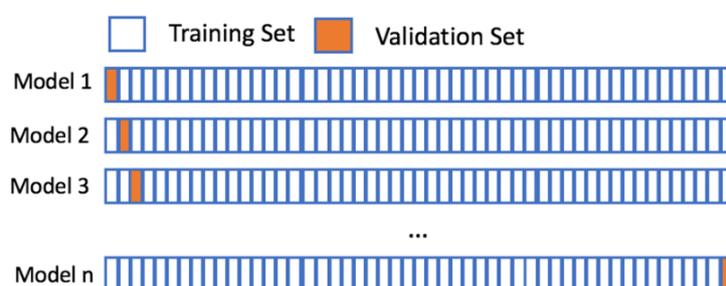


Figura 4.5 Esquema representando validação cruzada Leave One Out.

Fonte: CROSS-VALIDATION, 2023

A validação cruzada é uma ferramenta essencial para avaliar a habilidade preditiva de um modelo de regressão. No caso específico da LOO, a cada iteração, um ponto de dados é excluído do conjunto de treinamento. Posteriormente, os demais pontos são utilizados para ajustar o modelo, procurando prever o resultado com base nos dados do ponto excluído, que por sua vez serve como conjunto de teste. Ao término de cada ciclo, a previsão é comparada com o valor real, avaliando-se o erro resultante. Finalmente, após completar todas as iterações, calcula-se o somatório do quadrado dos erros individuais (SSR), divide-se pelo somatório do quadrado do erro em relação à média dos valores previstos (SST) faz-se 1 menos o resultado dessa razão, obtendo-se assim o coeficiente R^2 da validação cruzada, como demonstrado na equação 4.1.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

5.0 RESULTADOS E DISCUSSÃO

Nessa seção, serão apresentados os resultados encontrados para os modelos de previsão, em todas as métricas mencionadas anteriormente.

5.1 Modelo linear simples

Equação linear simples para CRI	Equação linear simples para CSR
$ \begin{aligned} \mathbf{CRI} = & 21.87506612 + \\ & 0.306587288 * \textit{Fator de fluidez} + \\ & 0.121523104 \\ & * \textit{Tempo bruto médio da mistura} \\ & \quad - \frac{20.00825374}{\textit{Semi fusinita}} \\ & \quad - \frac{83.258577}{\textit{Contração máxima}} \\ & \quad + \frac{0.008142058}{\textit{Teor de umidade do PetCoke}} \\ & \quad - \frac{93.07751605}{\textit{Relação Reativos Inertes modificada}} \end{aligned} $	$ \begin{aligned} \mathbf{CSR} = & 61.15980886 \\ & -0.711053262 * \textit{Fator de fluidez} \\ & \quad -0.156248698 \\ & * \textit{Tempo bruto médio da mistura} \\ & \quad + \frac{22.64451231}{\% \textit{ Semi fusinita}} \\ & \quad + 2.595450077 * \ln(1 + \% \textit{ Exinita}) \\ & \quad - 2423946.789 * (\textit{Contração máxima})^2 \\ & \quad + \frac{125.9399985}{\textit{Relação reativos inertes modificada}} \end{aligned} $

Tabela 5.1 Equações determinadas para CRI e CSR

A seleção de parâmetros do modelo linear simples é uma demonstração da limitação dos modelos lineares em explicar a variação do CRI e CSR a partir dessa base de dados. Ainda que tenham sido selecionados parâmetros com uma certa diversidade, alguns deles não encontram sentido nas teorias estabelecidas para carvão e coque, enquanto outros acabam indo na contramão do esperado para a explicação dos índices de qualidade.

Para CRI, os parâmetros reológicos escolhidos podem ter algum sentido ao considerarmos que elementos da microestrutura dos carvões dependem das transformações no estágio plástico e, portanto, influenciam na reação sólido-gás que

acontece no ensaio. Ainda assim, tais relações não são muito bem estabelecidas na literatura a ponto de se esperar que um modelo matemático escolha parâmetros reológicos para explicar o ensaio. Além disso, como já mencionado, o fator de fluidez proposto por Jordan encontra pouco respaldo na comunidade científica, carece de validações e depende do rank dos carvões. Ao ser considerado na mistura, onde o universo de variação da máxima fluidez e das temperaturas típicas não é muito grande, pode ser uma explicação para a sua seleção, mesmo com a fragilidade do índice.

Em termos de petrografia, a relação reativos inertes modificada aparece invertida e com sinal negativo para CRI, significando que à medida que a quantidade de macerais inertes aumenta, a reatividade diminui. Embora não haja uma relação tão estabelecida na teoria de coque e carvão, é possível argumentar que, por conta de uma maior área superficial dos macerais inertes, a tendência de aumento deles contribui para um aumento de reatividade, o que não é observado no modelo. Já para a semi-fusinita, em que a maioria das vezes tem mais porção inerte do que reativa, a relação vai no sentido correto. À medida que se aumenta o percentual de semi-fusinita nas misturas, maior é a quantidade de inertes e, portanto, pode haver um aumento no CRI.

Os outros dois parâmetros são o tempo bruto médio da mistura e o teor de umidade dos petcoques. O tempo bruto da mistura está com uma relação direta, aumentando o valor numérico à medida que aumenta o tempo de forno, o que não é esperado pela literatura uma vez que tempos de coqueificação maiores podem levar a uma melhoria na estrutura cristalina do carbono presente nos coques. Uma possível explicação pode estar no fato de que tempos de fornos costumam ter uma variação grande quando há problemas operacionais nos fornos, o que pode significar um perfil térmico diferente do esperado, com temperaturas menores de coqueificação. Além disso, a construção da base de dados pode não refletir os tempos de maneira correta, mas caso isso seja realidade, o modelo não deveria captar esse parâmetro como relevante, mostrando uma limitação presente na regressão linear simples ou na metodologia de seleção de parâmetros.

Já o teor de umidade é difícil encontrar algum sentido fenomenológico para explicar o CRI, uma vez que a umidade da base de dados foi realizada em poucos ensaios. Além disso, é possível notar pelo coeficiente que multiplica essa relação que o impacto é absolutamente baixo, mostrando que esse parâmetro tem pouca relevância, atestando mais uma vez as limitações nas escolhas de parâmetros e na base de dados.

Os parâmetros selecionados na correlação que faz a previsão do CSR foram semelhantes ao observado para CRI, exceto o percentual de exinita, que aparece dentro do logaritmo natural e toma o lugar do teor de umidade. O CSR, como dito anteriormente, tem parte da sua explicação na resistência mecânica a frio, que por sua vez tem relação direta com a microestrutura e microtextura resultante da transformação no processo de coqueificação, e parte da sua explicação no CRI, uma vez que a reatividade sólido-gás consome o coque e o fragiliza, reduzindo o valor do CSR. Isso explica parcialmente o porquê de alguns parâmetros estarem nos dois modelos, de maneira invertida como esperado. Porém, ainda que essa presença seja importante, é notória a ausência de parâmetros de química das cinzas, que são utilizados em uma ampla gama de modelos regressivos encontrados na literatura.

5.1.1 Modelagem de CRI

A figura 5.1 mostra a validação cruzada do modelo linear simples de CRI.

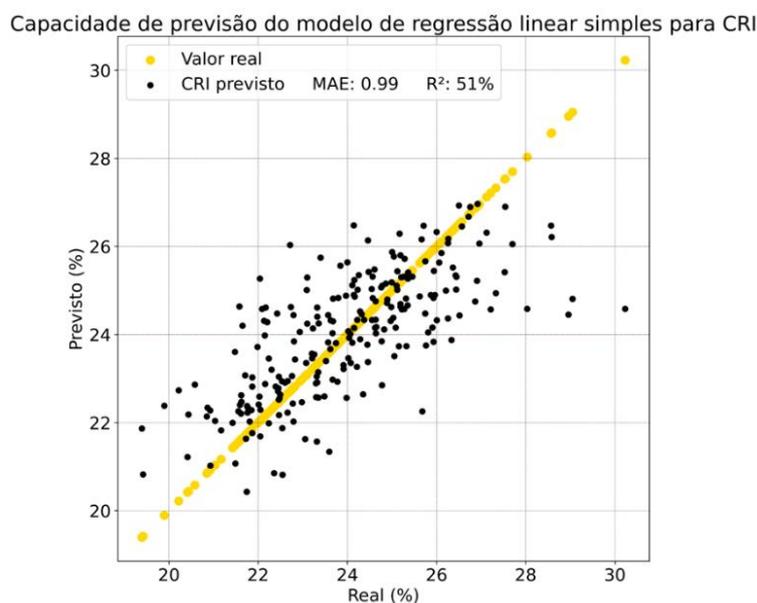


Figura 5.1 Validação cruzada do modelo linear simples de CRI

5.1.2 Modelagem de CSR

Para o CSR, foi avaliada a capacidade preditiva do modelo e os resultados estão na figura 5.2

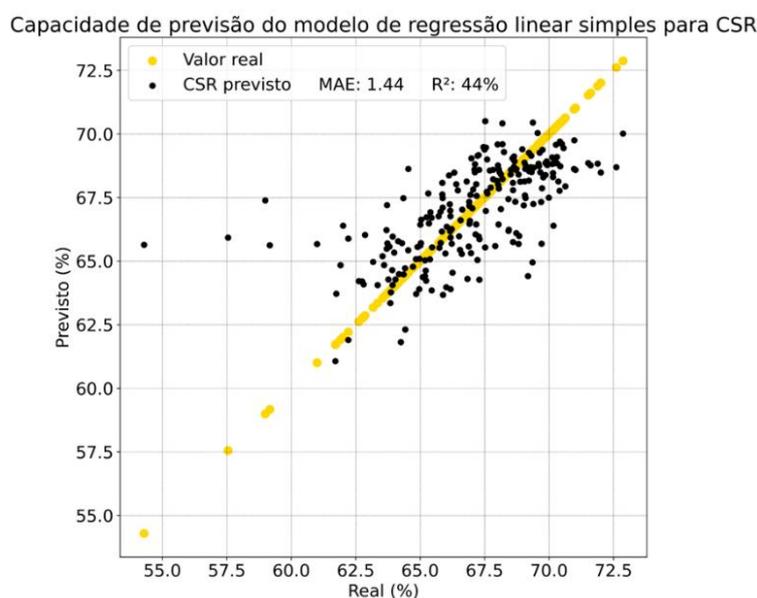


Figura 5.2 Validação cruzada do modelo linear simples de CSR

Em relação aos resultados da capacidade preditiva dos modelos, o R^2 da validação cruzada ficou próximo a 51% para CRI e 44% para CSR, enquanto o erro médio absoluto ficou próximo de 0.99 para CRI e 1.44 no CSR. Para o conjunto de dados demonstrado, dadas todas as observações realizadas anteriormente, os modelos tiveram desempenho regular. São as técnicas de regressão mais simples, com uma dificuldade maior para captar relações não-lineares entre os parâmetros. Um refino maior na seleção de parâmetros, excluindo-se aqueles que são considerados absurdos, pode resultar na elaboração de um modelo útil, sobretudo para os casos em que há uma variação brusca dos dados em relação aos originais, devido ao alto viés do modelo.

5.2 Modelos de aprendizado de máquina

5.2.1 Modelo de SVR

Para o modelo de CRI para SVR, o resultado da validação cruzada está demonstrado na figura 5.3:

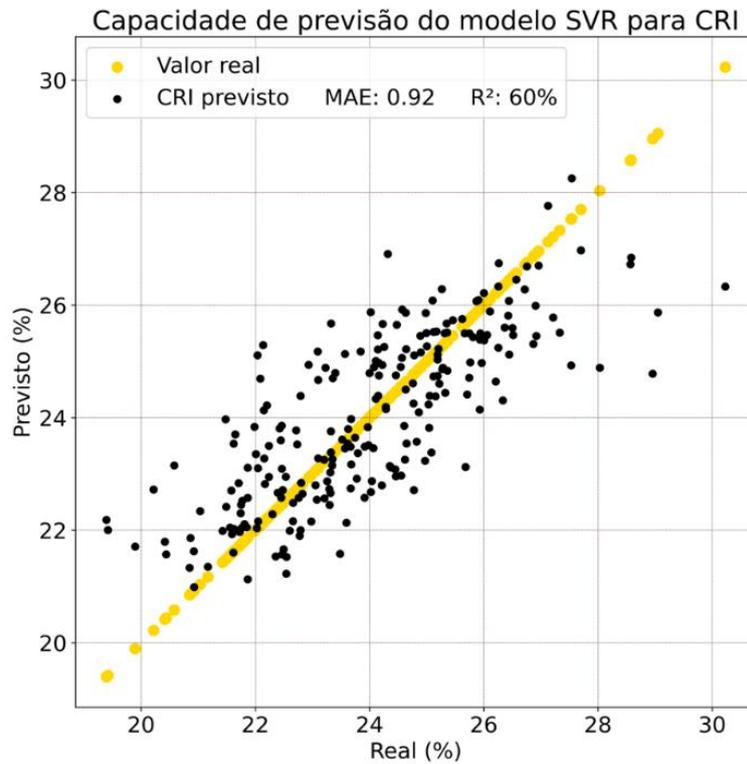


Figura 5.3 Validação cruzada do modelo SVR para CRI

Na figura 5.4, as explicações SHAP demonstram a importância para cada um dos parâmetros.

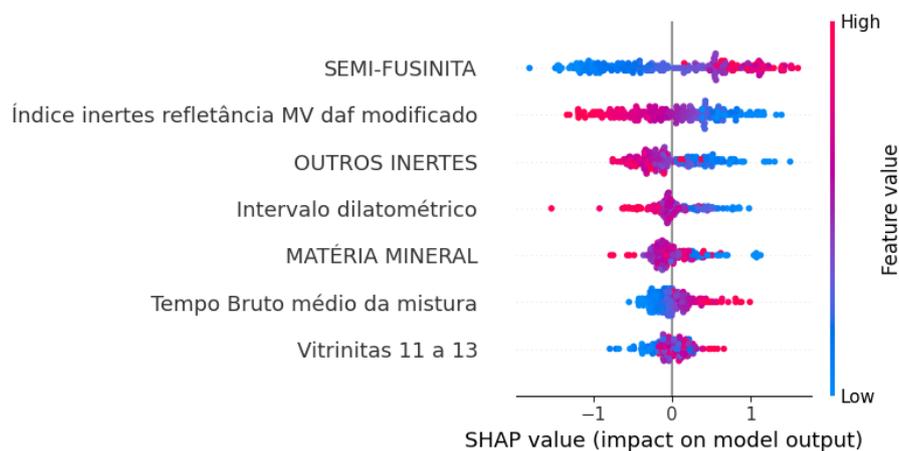


Figura 5.4 Resumo dos valores SHAP para o modelo SVR para CRI

Para o modelo de CSR feito a partir do SVR, o resultado da validação cruzada está demonstrado na figura 5.5:

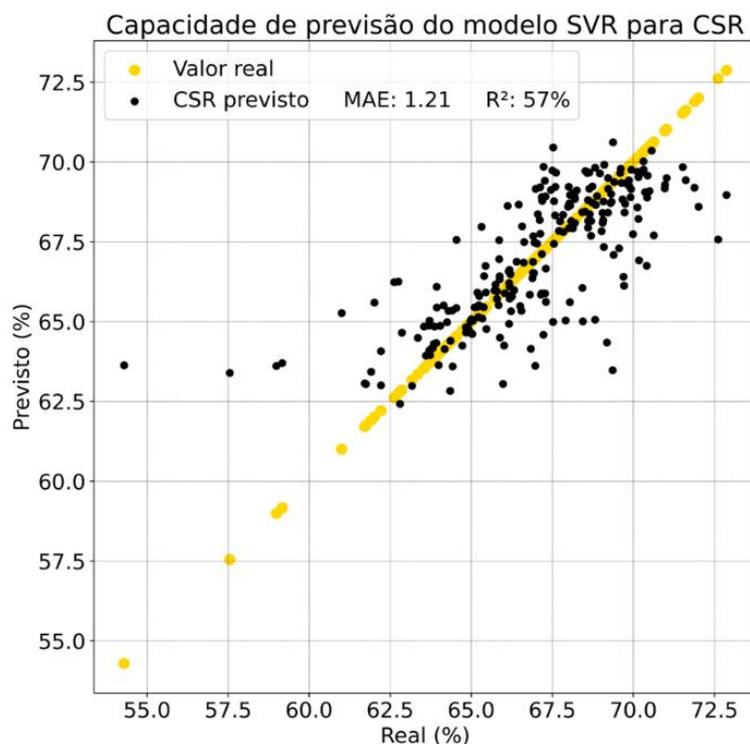


Figura 5.5 Validação cruzada do modelo SVR para CSR

Na figura 5.6, os valores SHAP para os parâmetros escolhidos no modelo SVR.

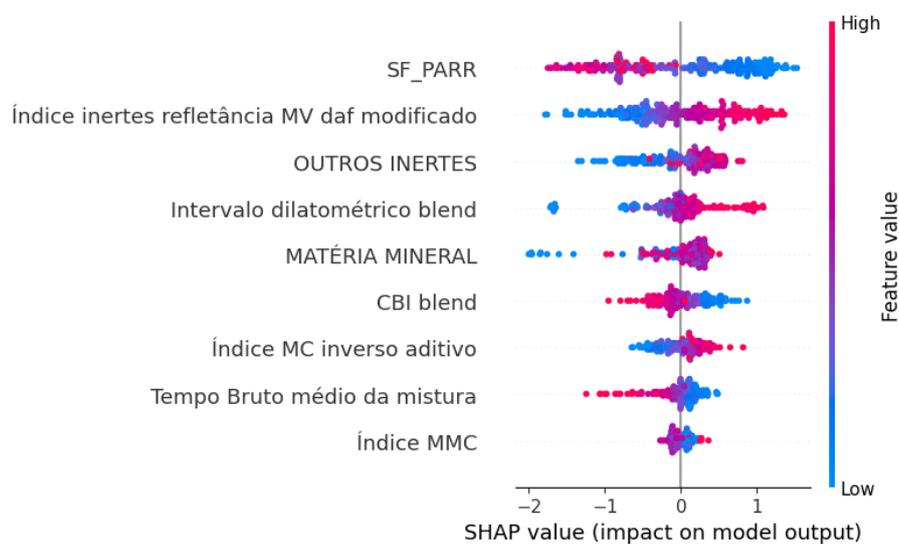


Figura 5.6 Resumo dos valores SHAP para o modelo SVR de CSR

Os parâmetros escolhidos para o modelo SVR foram diferentes, mas tiveram a mesma lógica daqueles escolhidos pelas regressões lineares simples: parâmetros de petrografia, parâmetros reológicos e parâmetros de processos.

Primeiramente, os parâmetros petrográficos de outros inertes, matéria mineral e semi-fusinita são muito co-relacionados entre si: à medida que o valor de um cresce, o outro diminui e vice-versa. Isso pode ser uma limitação do modelo, pois pode dar um peso muito grande a parâmetros que são relacionados. Ainda que tenha sido realizado um filtro de correlação, ele não foi capaz de captar essa relação e remover esses parâmetros. Buscando uma explicação na teoria de carvão e coque, o aumento do teor de semi-fusinita pode tender a aumentar a reatividade, à medida que o maceral da semi-fusinita tem na sua composição grande fração de inertes. Porém, isso contrasta com o índice de outros inertes, que à medida que cresce, tende a diminuir o valor numérico do CRI. Isso indica que não há relações tão diretas e óbvias para esses parâmetros e, do ponto de vista de modelos, não é adequado que esses sejam os parâmetros de maior representatividade na explicação do CRI. Nessa faixa de misturas, com esses dados disponíveis, esses foram os parâmetros que o modelo conseguiu captar, o que indica a incapacidade dos modelos de terem uma relação plena com o que é relevante na área acadêmica para o ensaio de CRI.

Seguindo pelos outros parâmetros, o índice inertes refletância MV (bsic) modificado teve um resultado esperado, uma vez que o aumento da presença de inertes tende a diminuir a reatividade. Já em relação ao parâmetro de vitrinitas 11 a 13, os valores SHAP demonstram uma tendência contrária à encontrada por KUMAR et al, que encontraram que valores de CSR aumentam à medida que as vitrinitas dos V-tipos 9 a 13 aumentam (o que caracterizaria uma diminuição do CRI).

No único parâmetro reológico selecionado para o modelo, o intervalo dilatométrico foi selecionado como relevante. À medida que o valor aumenta, o CRI diminui. Considerando que essa é uma base de dados de mistura de carvões, a variação de intervalos dilatométricos pode ser considerada relativamente pequena. Portanto, o modelo captou esse parâmetro uma vez que houve uma variação relativamente grande que ocasionou em uma diminuição do CRI. Para esse conjunto de dados, pode ser um parâmetro relevante, mas é preciso considerar que as características associadas à fluidez são mais relevantes e bem aceitas dentro da comunidade científica para explicar reatividade e resistência mecânica do que a dilatométrica.

Por fim, o parâmetro de tempo bruto médio da mistura apresentou o mesmo comportamento observado no modelo linear, o que indica que a relação contrária à literatura pode ter explicação devido aos pontos abordados anteriormente. Ainda que seja um conjunto relativamente diverso de parâmetros, existem fragilidades evidentes em

relação à teoria de carvão e coque. Nenhum parâmetro de química das cinzas foi selecionado para explicar reatividade, além de que, como já dito, parâmetros de fluidez costumam ter um desempenho melhor em explicar CRI do que parâmetros de dilatométrica.

Para os parâmetros escolhidos para CSR, podemos dividir em petrográficos, química das cinzas, reológico e de processos.

O parâmetro mais relevante para esse modelo é o SF_{PARR} , índice técnico utilizado pela planta da empresa siderúrgica que forneceu a base de dados e que relaciona o percentual de semi-fusinita com um índice de enxofre e cinzas. Fazendo uma correlação com a semi-fusinita enxergada no modelo de CRI, a relação é contrária, como era esperado. Mais uma vez, há parâmetros correlacionados escolhidos pelo modelo, como outros inertes e matéria mineral. Um parâmetro relacionado com a relação de macerais inertes e reativos é o CBI, onde nessa faixa de misturas estudadas, um aumento no valor indicou uma queda no CSR, enquanto uma diminuição significou um aumento. Isso pode ter relação com a faixa estudada, uma vez que o CBI é um parâmetro que busca encontrar uma relação ótima de inertes e reativos, e misturas que se aproximam de 1 tendem a ser consideradas boas, enquanto os desvios para cima ou para baixo são considerados ruins. Outro parâmetro com relevância dentro da petrografia está o índice inertes refletância MV (bsic), onde há a relação do teor de inertes com a refletância máxima e a matéria volátil em base seca isenta de cinzas. À medida que aumenta a presença dos inertes petrográficos, temos um aumento no CSR, o que é coerente com a literatura.

Novamente, o único parâmetro relacionado à reologia dos carvões é o intervalo dilatométrico. Para explicar CSR, parâmetros relacionados ao ensaio de dilatométrica podem ter relação com a porção relacionada à resistência à frio, microestrutura e microtextura do coque formado, mas ainda assim é muito menos relevante do que parâmetros de fluidez. Fato é que ficaram faltando mais parâmetros reológicos no modelo SVR, dada sua importância para explicar resistência mecânica.

Em relação aos parâmetros de química das cinzas, dois deles foram selecionados, o índice MC inverso aditivo e o índice MMC. Todos eles são variações do índice de basicidade, atribuindo pesos aos diferentes óxidos. É interessante que eles apareçam no modelo de CSR, uma vez que parte da explicação do ensaio vem por conta da reatividade, mas é incoerente que nenhum deles apareça no modelo de CRI, que justamente é o ensaio que é correlacionado à reatividade. Os índices, pelo menos, estão coerentes com o que se

espera, já que o aumento do índice MC inverso ocasiona um aumento do CSR, enquanto o índice MMC não tem uma tendência muito clara, sendo o parâmetro menos relevante dentre todos os selecionados.

À primeira vista, a diversidade dos parâmetros chama a atenção positivamente, mas há uma clara falta de coerência ao se considerar os dois modelos em conjunto. É notória a limitação do modelo, mas também da base de dados, por todos os motivos já discutidos até aqui nesse trabalho.

Em relação à capacidade preditiva, o SVR conseguiu um desempenho satisfatório, tendo um desempenho de 60% no R^2 da validação cruzada para CRI e 57% para CSR. Os erros acabaram sendo mais acentuados nos extremos, especialmente para CSRs abaixo de 62.5% e CRIs acima de 28%.

5.2.2 Modelo de florestas aleatórias

Para o modelo de CRI feito a partir de florestas aleatórias, o resultado da validação cruzada está demonstrado na figura 5.7.

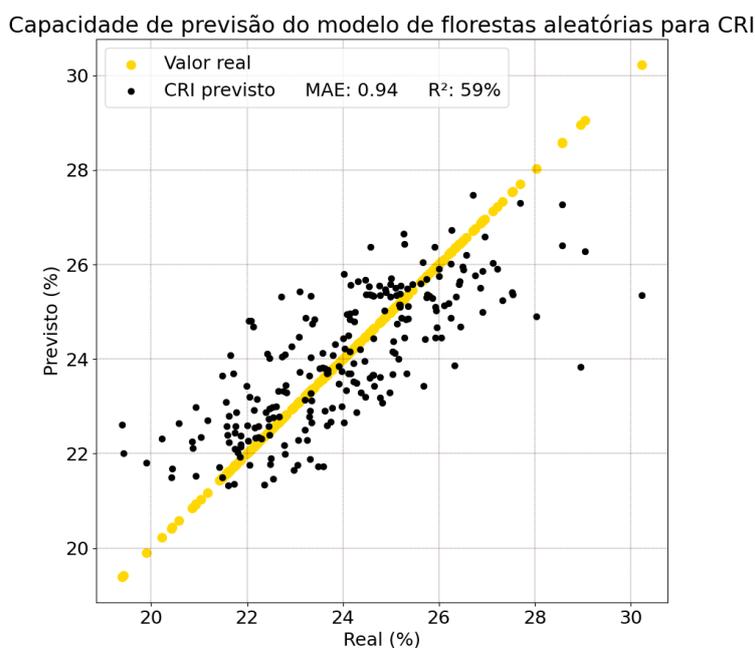


Figura 5.7 Validação cruzada do modelo de florestas aleatórias para CRI

Na figura 5.8, temos os valores SHAP para o modelo de florestas aleatórias para CRI.

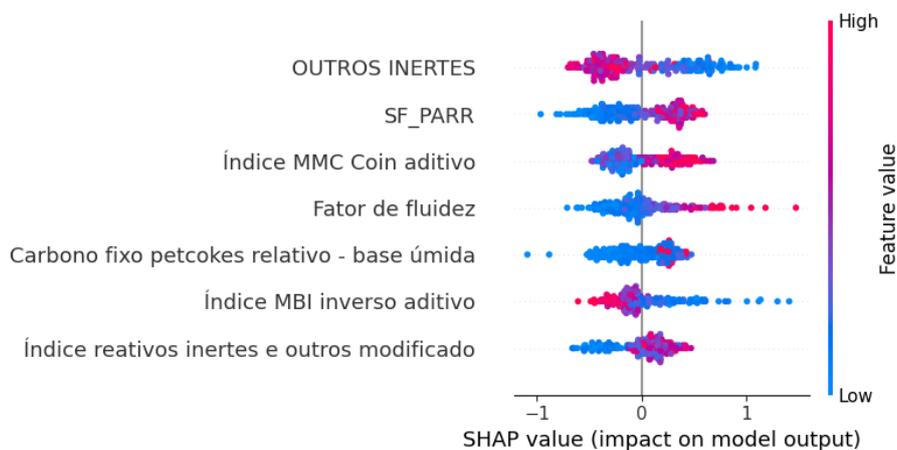


Figura 5.8 Resumo dos valores SHAP para o modelo de florestas aleatórias para CRI

Na figura 5.9, a validação cruzada para o modelo de florestas aleatórias para CSR.

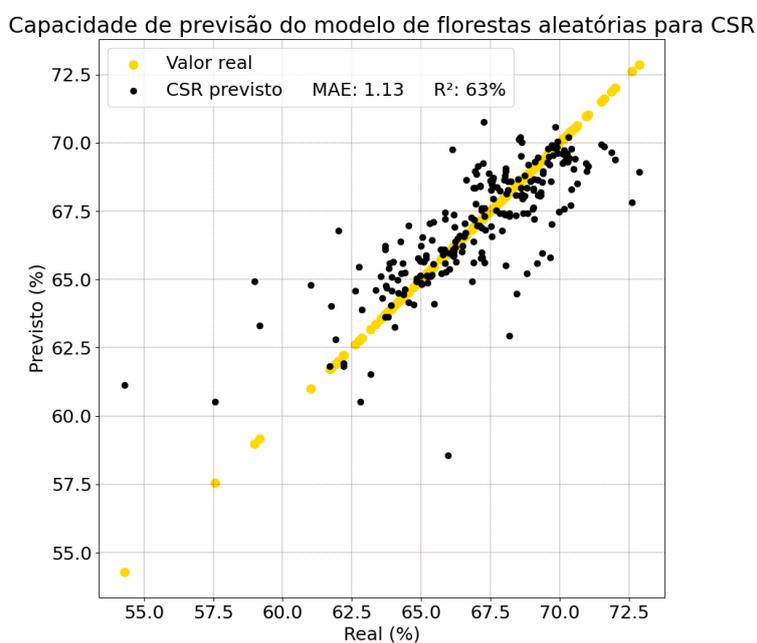


Figura 5.9 Validação cruzada do modelo de florestas aleatórias para CSR.

A figura 5.10 demonstra os valores SHAP para o modelo de florestas aleatórias de CSR.

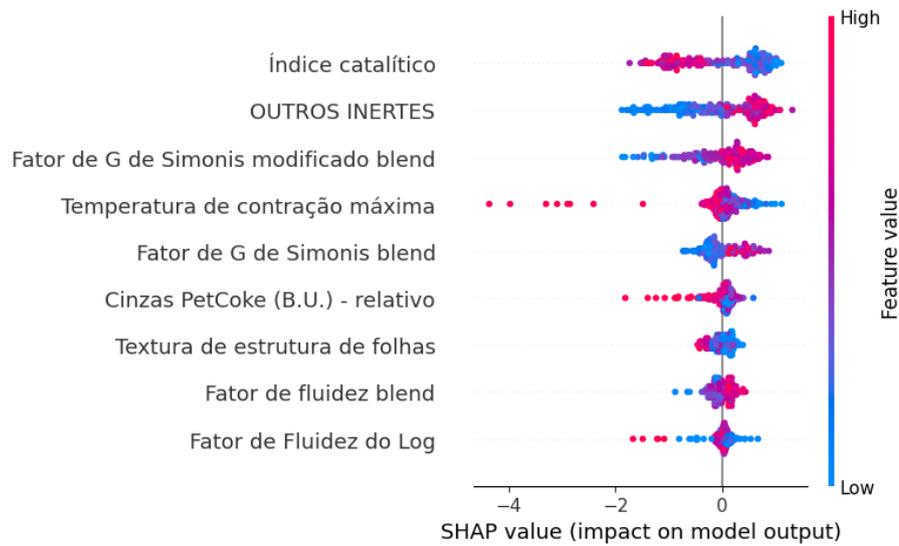


Figura 5.10 Resumo dos valores SHAP para o modelo de florestas aleatórias para CSR

5.2.3 Modelo de LightGBM LT

Para o modelo de LightGBM, foi selecionado o parâmetro “linear-trees”, que faz a previsão a partir de uma floresta de árvores lineares. Na figura 5.11 temos a validação cruzada do modelo.

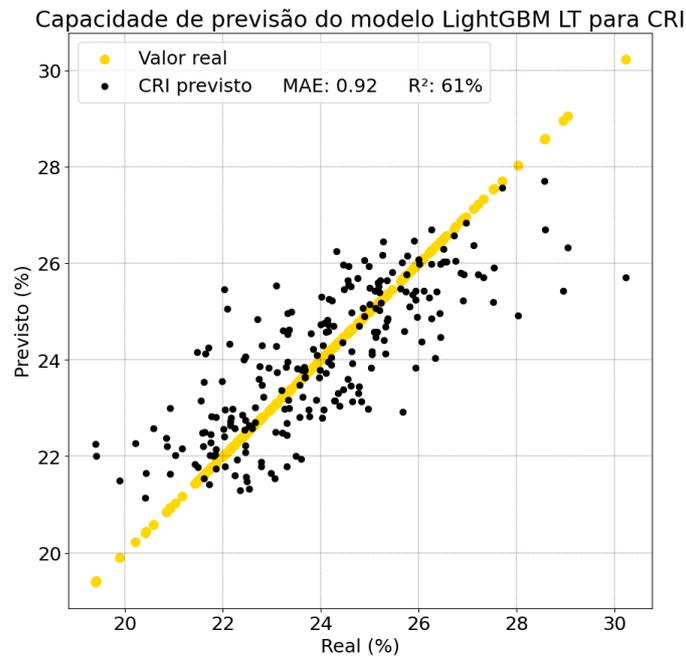


Figura 5.11 Validação cruzada do modelo LightGBM LT para CRI

Na figura 5.12, os valores SHAP para o modelo LigthGBM LT para CRI.

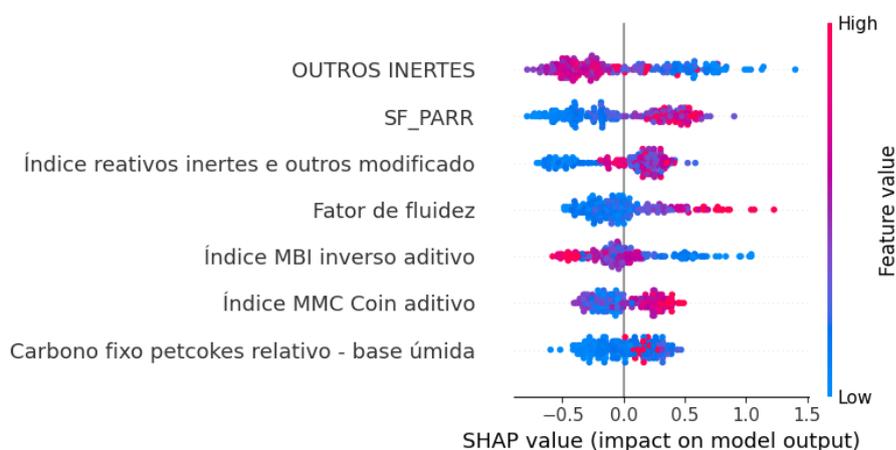


Figura 5.12 Resumo de valores SHAP para o modelo LightGBM LT para CRI

Para o modelo de CSR, o resultado é demonstrado na figura 5.13.

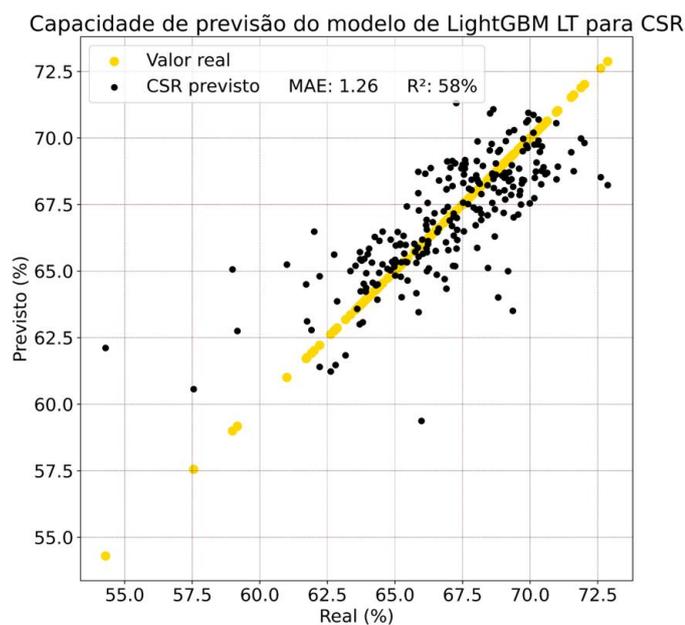


Figura 5.13 Validação cruzada do modelo LightGBM LT para CSR

A figura 5.14 mostra o resumo dos valores SHAP para o modelo LightGBM LT para o CSR.

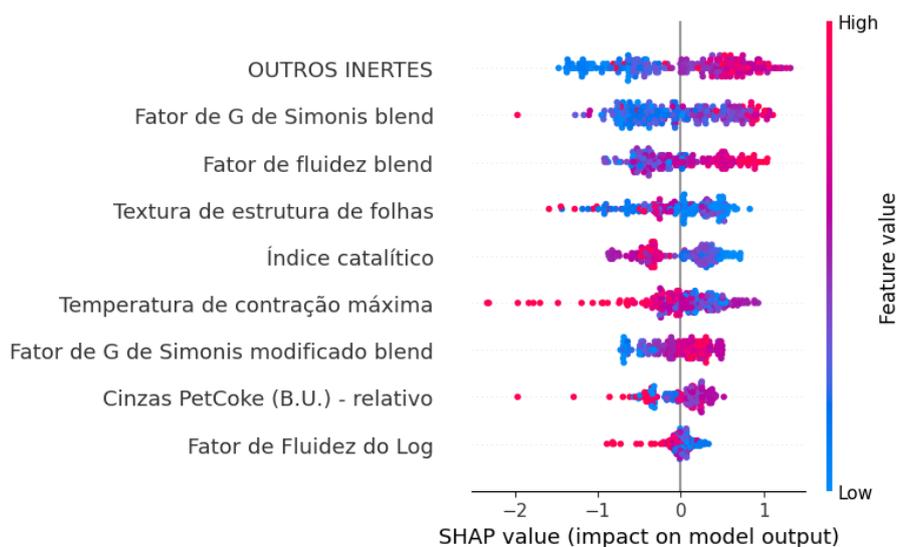


Figura 5.14 Resumo dos valores SHAP para o modelo LightGBM LT para CSR.

5.2.4 Modelo XGBoost

Por fim, o último modelo desenvolvido foi o de *Extreme Gradient Boosting*, ou XGBoost. O resultado da validação cruzada do XGBoost para CRI está demonstrado na figura 5.15.

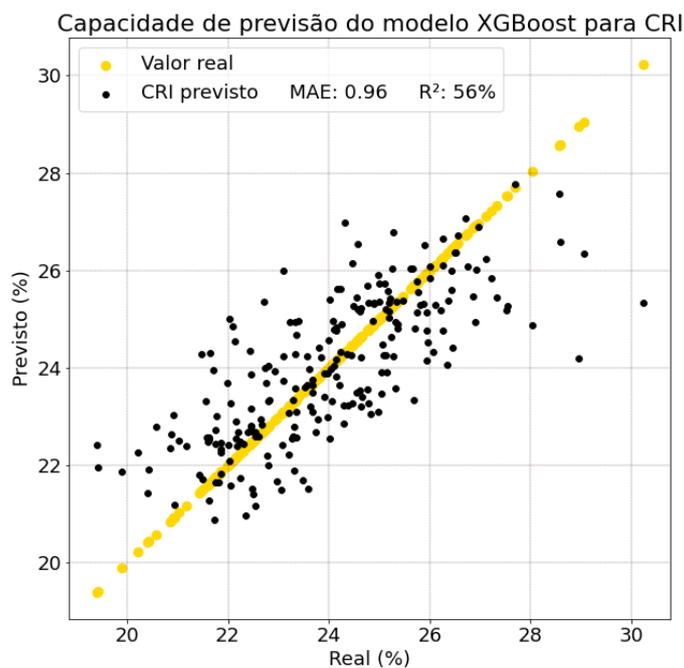


Figura 5.15 Validação cruzada do modelo XGBoost para CRI

Na figura 5.16, o resumo dos valores SHAP para o modelo XGBoost para CRI.

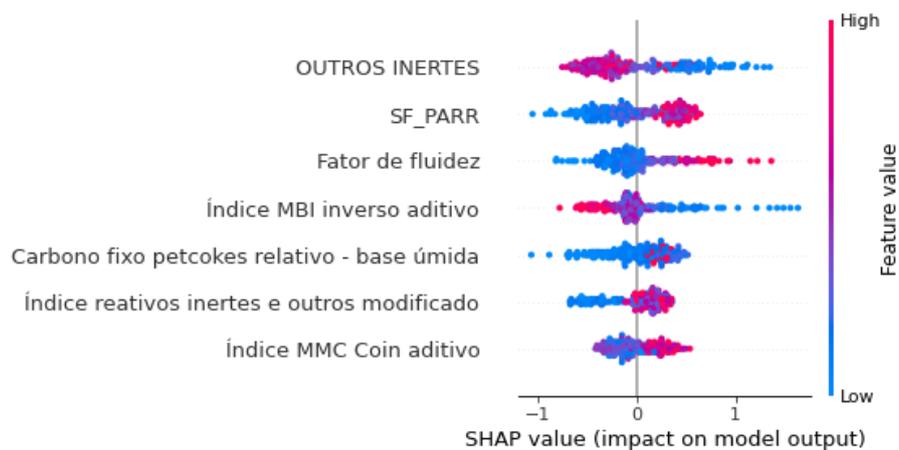


Figura 5.16 Resumo dos valores SHAP para o modelo XGBoost para CRI.

Já para o modelo de CSR, a validação cruzada está demonstrada na figura 5.17.

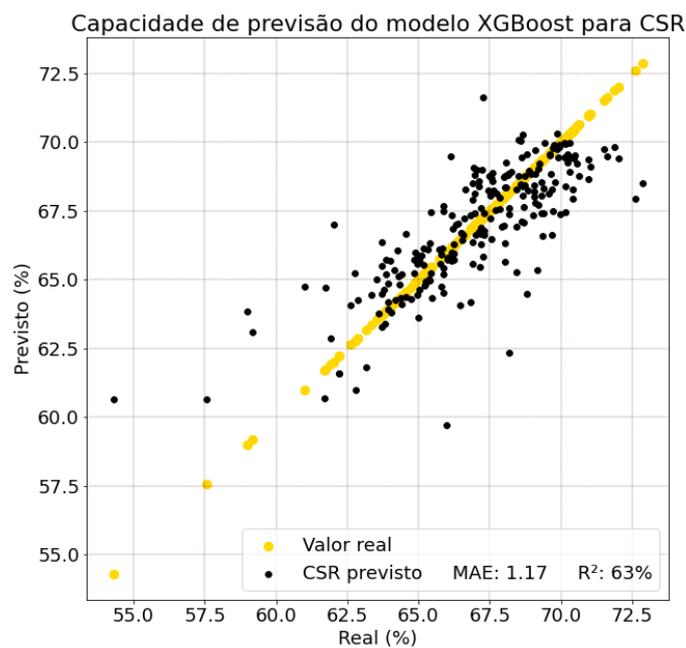


Figura 5.17 Validação cruzada do modelo XGBoost para CSR.

O resumo dos valores SHAP está demonstrado na figura 5.18.

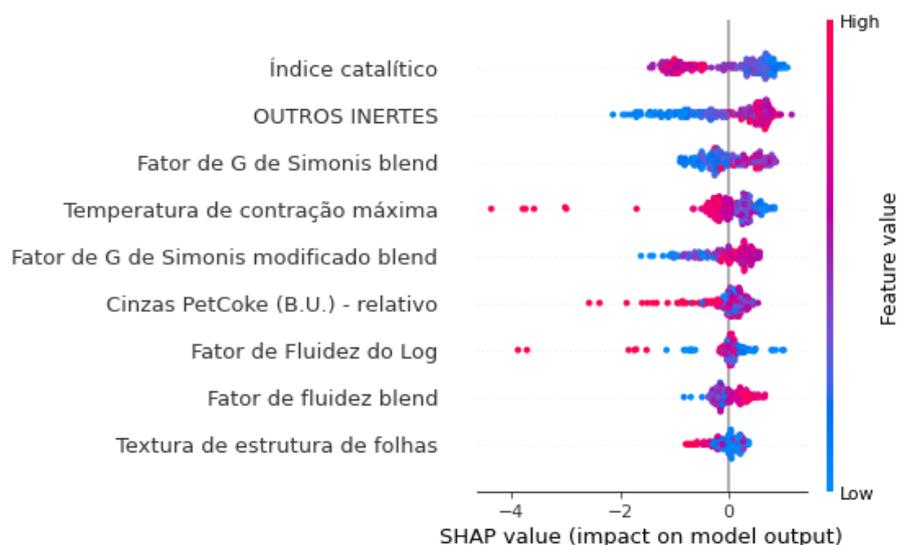


Figura 5.18 Resumo dos valores SHAP do modelo XGBoost para CSR.

Em relação aos parâmetros escolhidos para os demais modelos de aprendizado de máquina, temos mais uma vez a participação de três parâmetros relacionados à petrografia, um parâmetro reológico, dois parâmetros de química das cinzas e um parâmetro de análise imediata dos CVP.

Para a análise desses parâmetros, a mesma explicação dos parágrafos anteriores, com a diferença de que desta vez apenas um parâmetro relacionado diretamente à quantidade de grupos macerais (% de outros inertes). Mais uma vez o parâmetro SF_{PARR} que é relacionado com o maceral da semi-fusinita teve presença importante, e sua relação é inversa à de outros inertes. Uma possível explicação para isso está na origem geográfica dos carvões, onde há diferenças significativas na porção reativos/inertes e na quantidade do maceral semi-fusinita presente. Fato é que todos os modelos consideraram esses parâmetros como relevantes e opostos, o que pode significar mais um ponto da incapacidade do modelo em puxar dados relevantes, dada fragilidade da base de dados.

O fator de fluidez aparece como único parâmetro reológico para CRI. Com a tentativa de Jordan de englobar todos os parâmetros do ensaio de fluidez dentro de um só, o índice foi selecionado e teve um impacto relevante para a explicação da reatividade. Isso só ecoa com a teoria nos termos de que a transformação ocorrida no estágio plástico dos carvões tem influência na estrutura porosa e de área superficial do coque formado, impactando na sua reatividade. O modelo sugere que altos valores para o fator de fluidez aumentam a reatividade, o que é coerente com o esperado por Jordan, uma vez que os

carvões que ele considerou como relevantes tem o seu fator próximo de 1, enquanto um aumento desse fator acarretaria coques problemáticos, mas carece de relevância acadêmica.

Já os parâmetros de química das cinzas que foram escolhidos foram índice MMC Coin e o índice MBI inverso. Ambos os índices são modificações do índice catalítico, atribuindo pesos diferentes aos óxidos presentes nas cinzas dos carvões. O resultado foi coerente com o esperado na literatura, uma vez que o índice MMC teve um impacto numérico negativo à medida que diminui (como o índice de basicidade) e o índice MBI inverso teve o comportamento inverso. Como dito anteriormente, o índice de basicidade ou índice catalítico são relacionados à reatividade da reação de Boudouard, base do ensaio do CRI, portanto era de suma importância que tais parâmetros tenham sido selecionados para explicar esse índice.

Por fim, o parâmetro de carbono fixo dos petcoques acaba sendo pouco relevante do ponto de vista fenomenológico, a não ser que se haja uma relação com a quantidade de cinzas ou matéria volátil (uma vez que o carbono fixo é calculado por diferença). Como há a presença de petcoque em praticamente todas as misturas realizadas, um desvio da qualidade esperada para o petcoque pode ser explicado pelo carbono fixo, mas ainda assim não deveria ser um parâmetro tão relevante a ponto de ser escolhido dentre tantos outros.

Já para os parâmetros de CSR, pela primeira vez os parâmetros escolhidos foram mais reológicos do que de petrografia. Foram 5 parâmetros reológicos, 2 de petrografias, 1 de química das cinzas e 1 de análise imediata.

O mais relevante de todos para o modelo foi o índice catalítico, que embora não tenha sido selecionado no modelo de CRI, teve “representação” semelhante nos outros índices de química das cinzas. Como parte do CSR é explicada pelo CRI, que por sua vez tem dependência forte dos parâmetros dos óxidos das suas cinzas, é um ponto positivo para os modelos de aprendizado de máquina.

Para os parâmetros reológicos, dois fatores G de Simonis foram escolhidos e ambos tiveram o impacto esperado de acordo com o proposto por Simonis, em que valores menores indicam uma queda de CSR e valores próximos de 1 são esperados. O fator do blend performou melhor em alguns modelos, enquanto o fator aditivo performou melhor em outros. Outro parâmetro relacionado à dilatométrica foi a temperatura de contração máxima, que embora tenha uma faixa restrita de variabilidade, resultou na maior

amplitude de perda de CSR à medida que se aumenta o valor. Carvões com maior rank tendem a ter o seu estágio plástico deslocado para temperaturas maiores, e é sabido que carvões de mais alto rank tendem a ter fluidez menor devido aos microdomínios maiores e mais ordenados presentes em carvões com maior grau de carbonificação. Portanto, é um parâmetro que encontra sentido fenomenológico. Por fim, ainda em parâmetros reológicos, dois fatores de fluidez foram selecionados, sendo o fator de fluidez a partir do logaritmo da máxima fluidez o parâmetro que mais vai ao encontro do proposto por Jordan, onde fatores altos e baixos resultam numa diminuição de CSR, enquanto valores intermediários fazem com que o valor de CSR não se altere.

Para os outros parâmetros, outros inertes aparece mais uma vez com alta relevância, a textura de estrutura de folhas, parâmetro relacionado com os V-tipos mais altos, tem uma relevância pequena para o modelo e as cinzas do petcoke seguem a mesma lógica do parâmetro do carbono fixo do petcoke, mas com o valor invertido: à medida que se aumenta as cinzas associadas aos petcokes, menor é o CSR, uma vez que espera-se que petcokes tenham baixíssimas cinzas nas misturas praticadas.

Quanto à capacidade preditiva dos modelos, em geral todos eles desempenharam de maneira satisfatória, considerando todas as limitações já apontadas da base de dados. Para CRI, o melhor modelo foi o LightGBM com árvores lineares, que teve um R^2 de 61% e um erro médio absoluto na casa de 0.92, enquanto para CSR o destaque ficou com os modelos de florestas aleatórias e de XGBoost que conseguiram um R^2 de 63% e um erro médio absoluto de 1.13 e 1.17 respectivamente.

A figura 5.19 demonstra um resumo de todos os modelos em um gráfico.

Modelo	CRI		CSR	
	R^2	MAE	R^2	MAE
Regressão linear	51%	0.99	44%	1.44
SVR	60%	0.92	57%	1.21
Florestas aleatórias	59%	0.94	63%	1.13
LightGBM LT	61%	0.92	58%	1.26
XGBoost	56%	0.96	63%	1.17

Figura 5.19 Resumo de resultados para CRI e CSR.

Considerando a natureza dos ensaios, todas as fontes de erros da base de dados e as limitações dos modelos, os resultados foram satisfatórios e conseguiram explicar a variação dos ensaios, ainda que tenham dificuldade para captar relações esperadas na literatura. Modelos de aprendizado de máquina enxergam apenas relações matemáticas, portanto é preciso elevar a qualidade dos dados, sobretudo para depender menos de técnicas estatísticas para limpeza e tratamento, para que o modelo seja melhorado e consiga captar melhor os parâmetros

6.0 CONCLUSÕES

Esse trabalho teve como objetivo desenvolver diferentes modelos de previsão de qualidade do coque a partir de dados de entrada dos carvões. A partir da otimização das bases de dados, das seleções de parâmetros e do desenvolvimento e avaliação da capacidade preditiva dos modelos, foi possível concluir o seguinte:

- A coleta realizada sobre as bases de dados foi ampla, tendo sido levantadas análises de diferentes áreas operacionais, abrangendo todos os dados de qualidade e relações necessárias durante a sua construção, tratamento estatístico e polimento, ainda que com suas limitações devido a ruídos e suposições feitas durante seu desenvolvimento;
- A elaboração da engenharia de parâmetros, levando em consideração modelos preditivos desenvolvidos e descritos pela literatura, foi realizada, tendo sido gerados mais de 150 parâmetros que visaram explicar as variações de CRI e CSR observadas nessa base de dados.
- A seleção de parâmetros para os modelos lineares simples resultou em parâmetros muito semelhantes para CRI e CSR, mas têm poucas relações com o esperado pela literatura, mostrando possíveis limitações desses modelos em captar relações complexas.
- A seleção de parâmetros dos modelos de SVR tiveram uma performance mais coerente com o esperado pela literatura, sobretudo ao compararmos com o modelo linear simples. Um dos pontos de atenção relacionados a essa seleção está na quantidade de parâmetros que são intercorrelacionados, o que pode gerar um peso associado maior, uma vez que entram no modelo mais de uma vez.
- A seleção de parâmetros dos outros modelos de aprendizado de máquina teve, dentre todos os modelos, a resposta mais comparável com a literatura de carvão e coque. Cabe salientar, porém, que foram os únicos modelos que tiveram três etapas de seleção de parâmetros, incluindo o algoritmo Boruta.
- Todos os modelos tiveram seus parâmetros selecionados por seleção sequencial e seleção exaustiva a partir da métrica de menor erro quadrático médio, tendo sido selecionados os melhores parâmetros para cada uma das regressões realizadas.
- As explicações da influência dos parâmetros nos modelos através dos valores SHAP demonstraram, em linhas gerais, tendências esperadas pela teoria de carvão

e coque. As relações petrográficas foram um pouco mais difíceis de se relacionar com a teoria, uma vez que não há relações tão diretas entre os fenômenos, enquanto os parâmetros relacionados com a fluidez e a química das cinzas tiveram suas tendências coerentes com o esperado pela literatura. Entretanto, em alguns casos a tendência foi inversa à expectativa inicial;

- Em relação à capacidade preditiva, os resultados de R^2 para validação cruzada indicam um desempenho satisfatório para explicar a base de dados, uma vez que há uma tendência de variabilidade próxima aos resultados reais. Entretanto, isso não significa necessariamente que essas relações expliquem outros resultados de carvões e coque em outras coquearias;
- A ordem de classificação do melhor modelo de previsão de qualidade para o pior, para CRI, foi: LightGBM LT, SVR, florestas aleatórias, XGBoost e regressão linear simples, sendo o maior R^2 da validação cruzada de 61% e o pior de 51%. A estrutura de árvores lineares pode ter favorecido o modelo de CRI, enquanto a relação linear simples, como esperado, teve dificuldades por conta da possível complexidade da interação dos parâmetros selecionados.
- A ordem de classificação do melhor modelo de previsão de qualidade para o pior, para CSR, foi: florestas aleatórias, XGBoost, LightGBM LT, SVR e regressão linear simples, sendo o maior R^2 da validação cruzada de 63% e o pior de 44%. As florestas aleatórias, com o algoritmo de melhoria das árvores de previsão baseado na aleatoriedade da construção das demais árvores, explicou melhor a variação do CSR em comparação com os outros.
- Em posse de todos os resultados, foi possível avaliar o desempenho dos modelos de aprendizado de máquina na previsão de CRI e CSR sob a ótica da capacidade de previsão por meio de validação cruzada do tipo *Leave One Out* e sob a ótica das relações estabelecidas na literatura por meio da análise crítica dos parâmetros selecionados pelos modelos.

7.0 SUGESTÕES PARA TRABALHOS FUTUROS

- Replicar a metodologia em outras bases de dados de carvões e ensaios siderúrgicos, também tentando prever outros parâmetros de coque (DI, I40, shatter, etc.)
- Explorar outras metodologias de seleção de parâmetros, em ordens e técnicas diferentes do que as realizadas por esse trabalho;
- Explorar outros modelos de aprendizado de máquina, partindo para outras técnicas computacionais, como redes neurais;
- Expandir a metodologia para outros materiais, sejam eles da siderurgia (como sínter, co-produtos) ou de outras áreas metalúrgicas;
- Replicar a metodologia adicionando-se carvão vegetal ou demais biomassas, considerando o impacto profundo que a adição de materiais dessa natureza tem na qualidade do coque.

8.0 REFERÊNCIAS

- ADMASU T., TAMILARASI D., KUMAR N., **Sequential feature selection for heart disease detection using random forest**, Iraqi Journal of Science 63, p.3947-3953.
- AMMOSOV I., EREMIN I., SUKHENKO S., OSHURKOVA L., **Calculation of coking charges on basis of petrographic characteristics of coals**. Koks Khim. Russia, p.9-12, 1957.
- Boruta**. PyPi Project. <<https://pypi.org/project/Boruta/>> Acesso em 14 Ago. 2023
- BorutaSHAP**. PyPi Project. <<https://pypi.org/project/BorutaShap/>> Acesso em 14 Ago. 2023
- BUCZYNSKI R., WEBER R., KIM R., SCHWÖPPE P., **One-dimensional model of heat-recovery, non-recovery coke ovens**, Fuel 225, p.443-459, 2018.
- CHELGANI S., MATIN S., MAKAREMI S., **Modeling of free swelling index based on variable importance measurements of parent coal properties by random forest method**, Measurement 94, p.416-422, 2016.
- CHAN M., PARKYNS N., THOMAS K., **Thermoplastic behaviour of coal and its relationship to coke structure during high pressure carbonization**, Fuel 70, p.447-453, 1991.
- COIN C., BROOME A., **Coke quality prediction from pilot scale ovens and plant data**, Int Coal Conf, p.325-333, 1997.
- Cross-Validation**. Github <<https://biol607.github.io/lectures/crossvalidation.html#17>> Acesso em 16 Ago. 2023
- FLORES, Bruno. **ESTUDO DO COMPORTAMENTO TERMOPLÁSTICO DO CARVÃO DA CAMADA BARRO BRANCO (SANTA CATARINA - BRASIL) E SUA INFLUÊNCIA NO PROCESSO DE COQUEIFICAÇÃO**. Tese. PPGE3M:UFRGS, 2018.
- GOSENS J., TURNBULL A., JOTZO F., **China's decarbonization and energy security plans will reduce seaborn coal imports: Results from an installation-level model**, Joule 6, p. 782-815, 2022.
- HAYS D., PATRICK J., WALKER A. **Pore structure development during coal carbonization**. Fuel 55, p. 297-302, 1976.
- HU M., TAN Q., KNIBBE R., et al, **Recent applications of machine learning in alloy design: A review**, Materials Science & Engineering R 155, 2023.

HOMAFAR A., NASIRI H., CHELGANI S., **Modeling coking coal indexes by SHAP-XGBoost: Explainable artificial intelligence method**, Fuel Communications 13, 2022.

JORDAN P., **Characterising coals for coke production and assessing coke: predicting coke quality based on coal petrography, rheology and coke petrography**, Dissertação. Faculty of Engineering and the Build Environment: University of Witwatersrand, 2008.

KIDENA K., MURATA S., NOMURA M., **Studies on the chemical structural change during carbonization process**, Energy Fuels 10, p. 672-678, 1996

LEI Y., CHEN Y., CHEN J., LIU X., WU X., CHEN Y., **A novel modeling strategy for the prediction of the concentration of H₂ and CH₄ in raw coke oven gas**, Energy 273, 2023.

LIN M., HONG M., **The effect of coal blend fluidity on the properties of coke**. Fuel 65, p. 307-311, 1986.

LIU X., ALDRICH C., **Explaining anomalies in coal proximity and coal processing data with Shapley and tree-based models**, Fuel 335, 2023.

MACHAKA R., MOTSI G., RAGANYA L., et al, **Machine learning-based prediction of phases in high-entropy alloys: A data article**, Data in Brief 38, 2021.

MLxtend. Rasbt Project. <<https://rasbt.github.io/mlxtend/>> Acesso em 20 Ago. 2023

NORTH L., BLACKMORE K., NESBITT K., MAHONEY M., **Methods of coke quality prediction: A review**, Fuel 219, p. 426-445

NORTH L., BLACKMORE K., NESBITT K., MAHONEY M., **Models of coke quality prediction and the relationships to input variables: A review**, Fuel 219, p. 445-466

NOMURA S., AYUKAWA H., **Improvement in Blast Furnace Reaction Efficiency through the use of Highly Reactive Calcium Rich Coke**, Isij International, 2005.

Our World in Data. Emissions by sector. <<https://ourworldindata.org/emissions-by-sector>> Acesso em 24 Ago. 2023

Python. Python Org. < <https://www.python.org/about/>> Acesso em 23 Ago. 2023

QUITES E. E. C. Coqueificação, **“Introdução à Siderurgia”** (Mourão, M. B, coord.). ABM, São Paulo, volume 1, Cap. 2, p. 50-53, 2007

RANTITSCH G., BHATTACHARYYA A., SCHENK J., et al, **Assessing the quality of metallurgical coke by Raman spectroscopy**, International Journal of Coal Geology 130, p.1-7, 2014.

SCHAPIRO N., GRAY R., EUSNER G., **Recent developments in coal petrography**, Blast Furnace, Coke Oven Raw Mater Committee 20, p. 89-112, 1961.

SCHAPIRO N., GRAY R., **The use of coal petrography in coke making**, Fuel 11, p.234-242, 1964.

Scikit-Learn. Scikit-Learn for Machine Learning. <<https://scikit-learn.org/stable/>> Acesso em 14 Ago. 2023

SHAP. PyPi Project. <<https://shap.readthedocs.io/en/latest/>> Acesso em 14 Ago. 2023

SHIN K., SONG J., BANG W., et al, **Quantitative Precipitation Estimates Using Machine Learning Approaches with Operational Dual-Polarization Radar Data**, Remote Sensing 13, 2021.

SIMONIS W., GNUSCHKE G., **The Effect of Coking Ability on the Strength of Coke**, Deutscher kokereiausschuss Steinkohlenberbauverein, Essen 140, p.1-16, 1966.

TIWARY A., GHOSH S., SINGH R., et al, **Automated coal petrography using random forest**, International Journal of Coal Geology 232, 2020.

Towards Data Science. Why “1.5” in IQR Method of Outlier Detection? <<https://towardsdatascience.com/why-1-5-in-iqr-method-of-outlier-detection-5d07fdc82097>> Acesso em 12 Set. 2023

WANG C., **Biomass as blast furnace injectant – Considering availability, pretreatment and deployment in the Swedish steel industry**, Energy Conversion and Management 102, p.217-226, 2015.

WIRTH R., **CRISP-DM: towards a standard process model for data mining**, Proceedings of the fourth international conference on the practical application of knowledge discovery and data mining, Practical Application Company, p.29-39, 2000.

YANG J., FENG A., DU H., **Relation between mineral catalytic index (MCI) and reactivity of coke**, Iron Steel 36, p. 5-9, 2001.

YULIANTI Y., SAIFUDIN A., **Sequential Feature Selection in Costumer Churn Prediction Based on Naive Bayes**, IOP Conference Series: Materials Science and Engineering 879, 2020.

ZHAO Z., YU X., YUNTAO L., JINMING Z., YANSONG S., **CFD study of hydrogen co-injection through tuyere and shaft of an ironmaking blast furnace**, Fuel 348, 2023.

APÊNDICE

Química das cinzas	Petrografia	Reológicos	Análise Imediata	Outros
CaO MgO Fe ₂ O ₃ Na ₂ O K ₂ O Al ₂ O ₃ SiO ₂ P ₂ O ₅ TiO ₂ Mn ₃ O ₅ SO ₃ SrO BaO ZnO PbO	Vitrinitas Exinitas Fusinitas Semi-fusinitas Matéria mineral Outros inertes V-tipos Ro (max)	Parâmetros de dilatométrica Parâmetros de fluidez Temperaturas típicas	Carbono fixo (bs) Matéria volátil (bs) Cinzas (bs) Umidade	Enxofre Fósforo Relação reativos/inertes

Tabela 4.1 Parâmetros da base de dados da qualidade de carvões

Lista de todos os parâmetros considerados na base de dados
CaO
MgO
Fe ₂ O ₃
Na ₂ O
K ₂ O
Al ₂ O ₃
SiO ₂
P ₂ O ₅
TiO ₂
Mn ₃ O ₅

SO₃

SrO

BaO

ZnO

PbO

Teor de C fixo

Teor de cinzas

Teor de matérias voláteis

Teor de umidade

Ro

OUTROS INERTES

MATÉRIA MINERAL

SEMI-FUSINITA

EXINITA

FUSINITA

VITRINITA

Contração máxima (a)

Dilatação máxima (b)

Max. Fluidez

Range Plástico

Temp Max Fluidez

Temp. Amolecimento (Dilatação)

Temp. Amolecimento (Fluidez)

Temp. Solidificação

Temperatura de dilatação máxima

Temperatura de contração máxima

S

P

V3 final

V4 final

V5 final

V6 final

V7 final

V8 final

V9 final

V10 final

V11 final

V12 final

V13 final

V14 final

V15 final

V16 final

V17 final

V18 final

V19 final

V20 final

Petrografia - razão R/I das semi-fusinitas

Análise imediata - material volátil - base úmida

Análise imediata - cinzas - base úmida

Total de inertes

Fator de fluidez

Fator de fluidez modificado

Logarítmo da máxima fluidez

Fator de G de Simonis

Fator de G de Simonis modificado

Intervalo dilatométrico

SI

CBI

Intervalo plástico

A1_PARR

A2_PARR

MM_PARR

V_PARR

E_PARR

SF_PARR

F_PARR

Reativos_PARR_modificado

Inertes_PARR_modificado

RI_PARR_modificado

MV_daf_carvão

Índice de basicidade AR

Relação básicos ácidos cinzas

Cinzas do coque - aditivo

Cinzas Fe_2O_3+CaO

Ro x IB

Índice MC aditivo

Índice MC inverso aditivo

Índice MC Coin aditivo

Índice MC Coin inverso aditivo

Índice MC sódio aditivo

Índice MC sódio inverso aditivo

Razão Ponderada Vitri aditivo

Razão Vitritinas aditivo

Índice MMC aditivo

Índice MMC Coin aditivo

Índice MMC inverso aditivo

Índice MMC Coin inverso aditivo

Índice de basicidade aditivo

Índice de basicidade inverso aditivo

Índice catalítico aditivo

Índice catalítico inverso aditivo

Índice MBI aditivo

Índice MBI inverso aditivo

Vitritinas 9 a 13

Vitritinas 11 a 13

Textura isotropica

Textura mosaica fina

Textura mosaica média

Textura mosaica grosseira

Textura de estrutura de folhas

Máxima Fluidez do Log

Teor de umidade - média petcokes relativa

Teor de umidade - média petcokes absoluta

Teor de matérias voláteis - média petcokes relativa

Teor de matérias voláteis - média petcokes absoluta

Teor de cinzas - média petcokes relativa

Teor de cinzas - média petcokes absoluta

Análise imediata - material volátil - base úmida - média petcokes relativa

Análise imediata - material volátil - base úmida - média petcokes absoluta

Análise imediata - cinzas - base úmida - média petcokes relativa

Análise imediata - cinzas - base úmida - média petcokes absoluta

<3,35mm

Carbono fixo petcokes relativo

Carbono fixo petcokes absoluto

Carbono fixo petcokes relativo - base úmida

Carbono fixo petcokes absoluto - base úmida

CSR médio da mistura (alvo)

CRI médio da mistura (alvo)

Tempo Bruto médio da mistura

Tempo Líquido médio da mistura

Granulometria média da mistura

Petrografia - distribuição unimodal?

Fator de fluidez blend

Fator de fluidez modificado blend

Logaritmo da máxima fluidez blend

Fator de G de Simonis blend

Fator de G de Simonis modificado blend

Intervalo dilatométrico blend

SI blend

CBI blend

Intervalo plástico blend

Análise imediata - Carbono fixo

Análise imediata - Carbono fixo - base úmida

Relação de Carbono fixo petcoke carvão

Relação de Carbono fixo petcoke carvão - base úmida

Fator de Carbono fixo petcoke carvão

Fator de Carbono fixo petcoke carvão - base úmida

Índice MC

Índice MC inverso

Índice MC Coin

Índice MC Coin inverso

Índice MC sódio

Índice MC sódio inverso

Razão Ponderada Vitri

Razão Vitrinitas

Índice MMC

Índice MMC Coin

Índice MMC inverso

Índice MMC Coin inverso

Índice de basicidade

Índice de basicidade inverso

Índice catalítico

Índice catalítico inverso

Índice MBI

Índice MBI inverso

% petcoke blend

MV_daf_mistura
Índice inertes refletância MV daf modificado
Índice reativos inertes e outros modificado
Relação Reativos Inertes modificada
Fator de correção por plasticidade modificado
Fator de correção por plasticidade modificado 2
Fator de correcao por coque de petróleo
Fator de correção por conteúdo de coque de petróleo
Cinzas do coque - blend
Cinzas médias do coque original
Cinzas médias do coque modificada
Índice de enxofre médio
Conteúdo de álcalis original
Conteúdo de álcalis modificado
Fator de Fluidez do Log
Desvio do CBI

Tabela 4.2 Todos os parâmetros considerados na base de dados

Parâmetros selecionados para CRI no modelo linear	Parâmetros selecionados para CSR no modelo linear
Fator de fluidez	Fator de fluidez
Tempo bruto médio da mistura	Tempo bruto médio da mistura
Semi-fusinita	Semi-fusinita
Contração máxima	Exinita
Teor de umidade do PetCoke	Contração máxima
Relação Reativos Inertes modificada	Relação reativos inertes modificada

Tabela 4.3 Equações para CRI e CSR a partir do modelo linear simples.

Parâmetros relevantes para CRI	Parâmetros relevantes para CSR
Relação reativos inertes modificada, fator de G de Simonis, índice de basicidade, índice MMC Coin aditivo, índice catalítico, carbono fixo petcokes relativo - base úmida, outros inertes, índice de basicidade AR, índice MC Coin, índice MMC inverso aditivo, temperatura de amolecimento, índice MBI inverso aditivo, índice MMC aditivo, fator de fluidez, SF_PARR, índice MC sódio aditivo, índice inertes refletância MV daf modificado, índice de basicidade inverso aditivo, índice reativos inertes e outros modificado, índice MBI aditivo.	Índice MMC aditivo, fator de fluidez (log), Temperatura de contração máxima, Índice MC inverso aditivo, Índice MC sódio, Teor de matérias voláteis - média petcokes relativa, Cinzas - base úmida - média petcokes relativa, fator G de Simonis (blend), índice catalítico, fator de fluidez, Carbono fixo petcokes, índice de basicidade aditivo, Índice inertes refletância MV daf modificado, outros inertes, intervalo dilatométrico, SI, SF_PARR, fator de G de Simonis modificado blend, índice MC sódio inverso aditivo, textura mosaica grosseira, índice MBI aditivo, índice MC Coin,

	<p>índice MMC inverso aditivo, carbono fixo - base úmida, fator de fluidez blend, carbono fixo petcokes relativo - base úmida, textura de estrutura de folhas índice MMC, Temperatura de máxima fluidez.</p>
--	--

Tabela 4.4 Parâmetros considerados relevantes pelo algoritmo Boruta

Parâmetros escolhidos para CRI na seleção sequencial	Parâmetros escolhidos para CSR na seleção sequencial	Parâmetros escolhidos para CRI na seleção sequencial (SVR)	Parâmetros escolhidos para CSR na seleção sequencial
<p>Índice MMC Coin aditivo</p> <p>Carbono fixo petcokes relativo - base úmida</p> <p>Outros inertes</p> <p>Índice MBI inverso aditivo</p> <p>Fator de fluidez</p> <p>SF_PARR</p>	<p>Fator de Fluidez do Log</p> <p>Temperatura de contração máxima</p> <p>Cinzas - base úmida - média petcokes relativa</p> <p>Fator de G de Simonis blend</p> <p>Índice catalítico</p> <p>OUTROS INERTES</p>	<p>Outros inertes</p> <p>Matéria mineral</p> <p>Semi-fusinita</p> <p>Intervalo dilatométrico</p> <p>Vitrinitas 11 a 13</p> <p>Tempo bruto médio da mistura</p> <p>Índice inertes refletância</p> <p>MV daf modificado</p>	<p>Outros inertes</p> <p>Matéria mineral</p> <p>SF_PARR</p> <p>Índice MC inverso aditivo</p> <p>Tempo bruto médio da mistura</p> <p>Intervalo dilatométrico blend</p> <p>CBI blend</p> <p>Índice MMC</p>

Índice reativos inertes e outros modificado	Fator de G de Simonis modificado blend Fator de fluidez blend Textura de estrutura de folhas		Índice inertes refletância MV daf modificado
---	--	--	--

Tabela 4.5 Resultado da seleção sequencial de parâmetros

Modelos	SVR	Florestas aleatórias	LightGBM LT	XGBoost
CRI	C= 6.449 $\epsilon = 0.0$	max_depth=50 max_leaf_nodes=114 max_samples=1.0 n_estimators=302 random_state=42	bagging_freq=1 colsample_bytree=0.386 learning_rate=0.01 max_delta_step=-60.0 max_depth=400 min_child_weight=0 n_estimators=450 num_leaves=466 path_smooth=50.0 reg_alpha=0.08423 reg_lambda=0.18263 seed=1234567	base_score=2.05672 colsample_bylevel=0.8396 colsample_bynode=0.97331 colsample_bytree=0.96610 gamma=0.62223 learning_rate=0.358795 max_bin=256 max_cat_threshold=64 max_cat_to_onehot=4 max_delta_step=12.0896 max_depth=12 min_child_weight=5.25686 n_estimators=20 num_parallel_tree=30
CSR	C= 4.083 $\epsilon = 0.0$	max_depth=87 max_leaf_nodes=216 max_samples=1.0 n_estimators=83 random_state=42	bagging_freq=1 colsample_bytree=0.2482 learning_rate=0.01 max_delta_step=-47.093 max_depth=14	base_score=0.0 colsample_bylevel=1.0 colsample_bynode=0.9 colsample_bytree=1.0 gamma=0.0 learning_rate=0.3 max_bin=256

			min_child_weight=32 n_estimators=1200 num_leaves=466 path_smooth=46.065 reg_alpha=0.8674 reg_lambda=1.4081 seed=1234567	max_cat_threshold=64 max_cat_to_onehot=4 max_delta_step=14.0 max_depth=6 min_child_weight=1.0 n_estimators=31 num_parallel_tree=31
--	--	--	---	--

Tabela 4.6 Hiperparâmetros otimizados para os modelos de aprendizado de máquina

Explicação dos parâmetros considerados por pelo menos algum dos modelos:

Parâmetros reológicos

Fator de fluidez: apresentado na seção 3.3.6.

$$Fator\ de\ Fluidez = \frac{T_1 + T_3}{2} \times \frac{MF + (T_3 - T_1)}{MF \times T_2}$$

Fator de fluidez (log): média ponderada do fator de fluidez calculado com base no log da máxima fluidez para os carvões individuais.

Fator de fluidez (blend): calculado a partir da máxima fluidez aditiva do blend.

Fator G de Simonis: apresentado na seção 3.3.2

$$G = \frac{T_1 + T_2}{2} \times \frac{C + D}{T_1 D + T_2 C}$$

Fator G de Simonis (blend): calculado a partir da contração e dilatação aditiva do blend.

Fator G de Simonis (modificado): calculado a partir das temperaturas do intervalo dilatométrico em vez das temperaturas típicas.

Contração máxima: valor máximo de contração do ensaio de dilatação.

Temperatura de contração máxima: temperatura onde a contração máxima é encontrada.

Intervalo dilatométrico: diferença entre temperatura de máxima dilatação e temperatura de início de amolecimento

Intervalo dilatométrico (blend): calculado em relação às temperaturas aditivas do blend e não dos carvões individuais.

Parâmetros petrográficos

Semi-fusinita: % de semi-fusinita na mistura.

Matéria mineral: % de matéria mineral na mistura.

Outros inertes: % de outros inertes na mistura.

Exinita: % de exinita na mistura.

Relação reativos/inertes modificada (RRIM):

$$RRIM = \frac{100 * \text{Reativos PARR modificado}}{100 - (\text{Reativos PARR modificado} * Ro(\text{Máx}) * \% MV (\text{bsic}))}$$

Índice reativos inertes e outros modificado (IRIOM):

$$IRIOM = \frac{100 * \text{Reativos PARR modificado}}{\text{Índice inertes refletância MV (bsic) modificado}}$$

Índice inertes refletância MV (bsic) modificado (IIRMVM):

$$IIRMVM = \text{Inertes PARR modificado} * Ro (\text{máx}) * \% MV (\text{bsic})$$

Inertes PARR modificado (IPM):

$$IPM = 100 - \text{Reativos PARR modificado}$$

Reativos PARR modificado (RPM):

$$RPM = \left\{ \begin{array}{l} \% \text{ Semi fusinita} > 20: V_{PARR} + S_{PARR} + 0,75 * E_{PARR} \\ \% \text{ Semi fusinita} < 20: V_{PARR} + S_{PARR} + 0,25 * E_{PARR} \end{array} \right\}$$

CVP: 52,5

V_{PARR} :

$$V_{PARR} = \frac{\% \text{ Vitrinitas} * (100 - MM_{PARR})}{100 - \% \text{ Matéria Mineral}}$$

E_{PARR} :

$$E_{PARR} = \frac{\% \text{ Exinitas} * (100 - MM_{PARR})}{100 - \% \text{ Matéria Mineral}}$$

SF_{PARR} :

$$SF_{PARR} = \frac{\% \text{ Semi fusinitas} * (100 - MM_{PARR})}{100 - \% \text{ Matéria Mineral}}$$

MM_{PARR} :

$$MM_{PARR} = \frac{A1_{PARR}}{A2_{PARR} + \left(\frac{A1_{PARR}}{100}\right)}$$

$A1_{PARR}$:

$$A1_{PARR} = 100 * \frac{1,08 * \% \text{ Cinzas} + 0,55 * \% \text{ Enxofre}}{2,8}$$

$A2_{PARR}$:

$$A2_{PARR} = 100 * \frac{1,08 * \% \text{ Cinzas} + 0,55 * \% \text{ Enxofre}}{1,35}$$

Vitrinitas 11 a 13: somatório dos V-tipos 11, 12 e 13

CBI (blend): *Composition-balance index*, relatado na seção 3.3.1

Textura de estrutura de folhas (TEF):

$$TEF = 0,1 * (V16 + V17) + 0,2 * (V17 + V18) + 0,4 * (V18 + V19) + 0,226$$

Parâmetros de química das cinzas

Índice catalítico (IC):

$$IC = 14,04 * \% \text{ Enxofre} + 9,64 * \% \text{ Cinzas} * \frac{(\% \text{CaO} + \% \text{MgO} + \% \text{Fe}_2\text{O}_3 + \% \text{Na}_2\text{O} + \% \text{K}_2\text{O})}{(\% \text{Al}_2\text{O}_3 + \% \text{SiO}_2 + \% \text{TiO}_2)}$$

Índice MC inverso aditivo (MCI aditivo):

$$MCI = \frac{\% \text{ Cinzas}}{\left(100 - \frac{(100 - \% MV)}{(100 - \% CZ)}\right)} * \frac{(\% \text{SiO}_2 + 0,41 * \% \text{Al}_2\text{O}_3 + 2,5 * \% \text{TiO}_2)}{(\% \text{Fe}_2\text{O}_3 + 1,6 * \% \text{CaO} + 0,93 * \% \text{MgO} + 2,2 * \% \text{Na}_2\text{O} + 1,9 * \% \text{K}_2\text{O})}$$

Índice MMC:

MMC

= % Cinzas

$$* \frac{(1,03 * \% \text{CaO} + 0,43 * \% \text{MgO} + 2,85 * \% \text{Na}_2\text{O} + 1,9 * \% \text{K}_2\text{O} + \% \text{Fe}_2\text{O}_3)}{(0,74 * \% \text{Al}_2\text{O}_3 + \% \text{SiO}_2)}$$

Índice MMC Coin:

MMCC

= % Cinzas

$$* \frac{(1,03 * \% \text{CaO} + 2,85 * \% \text{MgO} + 0,43 * \% \text{Na}_2\text{O} + \% \text{K}_2\text{O} + 1,9 * \% \text{Fe}_2\text{O}_3)}{(0,74 * \% \text{Al}_2\text{O}_3 + \% \text{SiO}_2)}$$

Índice MBI inverso:

$$MBI = \frac{100 * \% \text{ Cinzas} * (\% SiO_2 + \% Al_2O_3 + \% TiO_2)}{(100 - \% MV) (\% Fe_2O_3 + \% CaO + \% MgO + \% Na_2O + \% K_2O)}$$

Parâmetros de análise imediata

Teor de umidade do petcoke: % de umidade na análise imediata dos CVP.

Carbono fixo do petcoke: % de carbono fixo na análise imediata dos CVP.

Cinzas petcoke: % de cinzas na análise imediata dos CVP.

Parâmetros de processo

Tempo bruto médio do forno: Tempo médio de enformamento das misturas no intervalo em que elas estiveram no forno.