



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



# **MODELO DE PREVISÃO PARA A RESISTÊNCIA MECÂNICA DO COQUE METALÚRGICO**

**Amanda Oliveira Magalhães**  
**Amanda Souza De Avellar**  
**Marcus Antonio Bueno Borges**

João Monlevade, MG  
2023

Amanda Oliveira Magalhães  
Amanda Souza De Avellar  
Marcus Antonio Bueno Borges

# **MODELO DE PREVISÃO PARA A RESISTÊNCIA MECÂNICA DO COQUE METALÚRGICO**

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

Orientador: Prof. Dr. Thiago Augusto de Oliveira Silva

João Monlevade, MG

2023

## SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

B732m Borges, Marcus Antônio Bueno.

Modelo de previsão para a resistência mecânica do coque metalúrgico. [manuscrito] / Marcus Antônio Bueno Borges. Amanda Oliveira Magalhães. Amanda Souza de Avellar. - 2023.

27 f.: il.: color., gráf., tab.. + .

Orientador: Prof. Dr. Thiago Augusto de Oliveira Silva.

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

1. Ciência de dados. 2. Coque. 3. Carvão. 4. Siderurgia. 5. Algoritmos - Random Forest. I. Avellar, Amanda Souza de. II. Magalhães, Amanda Oliveira. III. Silva, Thiago Augusto de Oliveira. IV. Universidade Federal de Ouro Preto. V. Título.

CDU 669.162.16:004.85

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



## FOLHA DE APROVAÇÃO

**Amanda Oliveira Magalhães**  
**Amanda Souza De Avellar**  
**Marcus Antonio Bueno Borges**

### Modelo de previsão para a resistência mecânica do coque metalúrgico

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

Aprovada em 26 de junho de 2023

#### Membros da banca

[Prof. Dr.] - Thiago Augusto de Oliveira Silva  
[Prof. Dr.] - Paganini Barcellos de Oliveira - (Deenp/Icea/Ufop)  
[Eng. Me.] - Bruno Santos Malaquias - (Usiminas)

Prof. Dr. Thiago Augusto de Oliveira Silva, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 10/07/2023



Documento assinado eletronicamente por **Thiago Augusto de Oliveira Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 10/07/2023, às 18:59, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0555958** e o código CRC **E7AD656A**.

# Resumo

Os parâmetros de qualidade das matérias primas metalúrgicas são regularmente controlados afim de que se atinja o controle do processo e de qualidade do produto final. Uma das matérias primas utilizadas como combustível dentro da siderurgia é o coque e sua característica de resistência mecânica (drum index) foi analisada presente estudo. Considerando o alto impacto no custo final do gusa gerado pelo coque metalúrgico, além de suas propriedades estarem ligadas à performance dos altos-fornos, mostra-se importante o aumento da assertividade da previsibilidade de seus parâmetros. Tomando como base pesquisas bibliográficas e conhecimento técnico, o presente trabalho busca a partir de análises estatísticas de dados da qualidade dos carvões individuais e dos parâmetros reais do processo de coqueificação, desenvolver um modelo de previsão para o DI (resistência mecânica). Para modelagem deste trabalho, foi selecionado o algoritmo Random Forest que prediz o DI, combinando o resultado das árvores de decisão através da correlação de variáveis. O estudo contribui para melhor previsibilidade, gestão, desempenho e custo no processo de coqueificação. Os resultados de previsão se mostraram promissores, com um acerto de 96% dos valores preditos em comparação com os resultados reais, além disso a média ponderada do erro apresentou um desvio absoluto de 0,14, demonstrando sua assertividade.

**Palavras-chaves:** Resistência Mecânica. Coque. Random Forest. Carvão. Siderurgia.

# Abstract

The quality parameters of metallurgical raw materials are regularly controlled in order to achieve control of the process and quality of the final product. One of the raw materials used as fuel in the steel industry is coke and its mechanical resistance characteristic (drum index) was analyzed in this study. Considering the high impact on the final cost of pig iron generated by metallurgical coke, in addition to its properties being linked to the performance of blast furnaces, it is important to increase the assertiveness of the predictability of its parameters. Based on bibliographical research and technical knowledge, this work seeks, based on statistical analysis of data on the quality of individual coals and the real parameters of the coking process, to develop a prediction model for DI (drum index). For modeling this work, the Random Forest algorithm was selected, which predicts the DI, combining the result of the decision trees through the correlation of variables. The study contributes to better predictability, management, performance and cost in the coking process. The prediction results were promising, with a 96% accuracy of the predicted values compared to the actual results, in addition the weighted average of the error presented an absolute deviation of 0.14, demonstrating its assertiveness.

**Keywords:** Drum Index. Coke. Random Forest. Coal. Steel industry.

# Lista de ilustrações

Figura 1 – Fluxo da Redução . . . . .	3
Figura 2 – Coqueria . . . . .	5
Figura 3 – Tambor do Ensaio de DI . . . . .	6
Figura 4 – Diagrama de Desenvolvimento do Trabalho . . . . .	9
Figura 5 – Relacionamento da Base de Dados . . . . .	11
Figura 6 – Análise Descritiva das Variáveis Numéricas Seleccionadas . . . . .	17
Figura 7 – Histograma com Todas as Variáveis Numéricas . . . . .	17
Figura 8 – Mapa de Calor com Todas as Variáveis Numéricas . . . . .	18
Figura 9 – Mapa de Calor com as Variáveis Numéricas Seleccionadas . . . . .	20
Figura 10 – Simulação Utilizando o Modelo Salvo . . . . .	26

# Lista de tabelas

Tabela 1 – Comparação DI Obtido x Fórmula . . . . .	10
Tabela 2 – Estrutura da Tabela da Qualidade da Mistura de Carvões . . . . .	12
Tabela 3 – Estrutura da Tabela de Ciclo de Produção . . . . .	13
Tabela 4 – Estrutura das Variáveis Calculadas do Ciclo de Produção . . . . .	13
Tabela 5 – Estrutura da Tabela da Qualidade do Coque . . . . .	14
Tabela 6 – Critérios para Exclusão de Dados da Qualidade da Mistura . . . . .	15
Tabela 7 – Critérios para Exclusão de Dados do Ciclo de Produção . . . . .	15
Tabela 8 – Critérios para Exclusão de Dados da Qualidade do Coque . . . . .	16
Tabela 9 – Variáveis Numéricas da Base de Dados . . . . .	16
Tabela 10 – Variáveis Numéricas Selecionadas . . . . .	19
Tabela 11 – Simulações Realizadas com o RandomizedSearchCV . . . . .	21
Tabela 12 – Precisão do Modelo de Cada Simulação . . . . .	21
Tabela 13 – Parâmetros Selecionados para o Random Forest . . . . .	22
Tabela 14 – Precisão do Modelo na Base de Treinamento e Teste . . . . .	22
Tabela 15 – Resultado das Métricas RMSE e MAPE . . . . .	23
Tabela 16 – Iterações Utilizando o Método K-fold . . . . .	23
Tabela 17 – Precisão do Modelo com Valores Reais . . . . .	25

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>1.1</b>	<b>Objetivo Geral</b>	<b>1</b>
1.1.1	Objetivos Específicos	2
<b>1.2</b>	<b>Justificativa</b>	<b>2</b>
<b>1.3</b>	<b>Organização do Trabalho</b>	<b>2</b>
<b>2</b>	<b>REVISÃO DA LITERATURA</b>	<b>3</b>
<b>2.1</b>	<b>Fluxo da Redução</b>	<b>3</b>
<b>2.2</b>	<b>Processo de Coqueificação</b>	<b>4</b>
<b>2.3</b>	<b>Resistência Mecânica – <math>DI_{(15-150)}</math> (Drum Index)</b>	<b>5</b>
<b>2.4</b>	<b>Definição do Modelo</b>	<b>6</b>
<b>2.5</b>	<b>Trabalhos Relacionados</b>	<b>7</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>8</b>
<b>4</b>	<b>DESENVOLVIMENTO</b>	<b>10</b>
<b>4.1</b>	<b>Cálculo do <math>DI_{(15-150)}</math></b>	<b>10</b>
<b>4.2</b>	<b>Coleta dos Dados</b>	<b>11</b>
4.2.1	Qualidade da Mistura dos Carvões	12
4.2.2	Ciclo de Produção	12
4.2.3	Qualidade do Coque	14
<b>4.3</b>	<b>Tratamento dos Dados</b>	<b>14</b>
4.3.1	Critérios para Exclusão de Dados da Qualidade da Mistura	15
4.3.2	Critérios para Exclusão de Dados do Ciclo de Produção	15
4.3.3	Critérios para Exclusão de Dados da Qualidade do Coque	15
<b>4.4</b>	<b>Seleção das Variáveis</b>	<b>16</b>
<b>4.5</b>	<b>Modelagem</b>	<b>20</b>
4.5.1	Parametrizando o Modelo	20
4.5.2	Divisão da Base de Dados	22
<b>5</b>	<b>RESULTADOS</b>	<b>25</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>26</b>
	<b>REFERÊNCIAS</b>	<b>27</b>

# 1 Introdução

Grande parte dos custos atrelados ao produto gerado em uma indústria siderúrgica integrada são referentes aos combustíveis utilizados no processo de redução do aço. Cerca de 90% destes custos são referentes à produção de coque que utiliza o carvão mineral como matéria prima.

Segundo Giglio (2016) em indústrias siderúrgicas integradas, o processo de coqueificação tem grande impacto na produtividade e na rentabilidade do produto final. O coque, principal combustível dos altos-fornos, é produzido a partir de misturas de carvões metalúrgicos. A mistura de carvões é definida de forma a atender tanto as condições operacionais do processo de coqueificação quanto a qualidade requerida na produção do ferro gusa, visando o menor custo.

*A mistura de carvões para a produção de coque metalúrgico dá-se por diferentes tipos de carvões. Para atender as exigências do processo de coqueificação, os carvões são avaliados segundo os parâmetros de qualidade: análise imediata (cinzas, enxofre, voláteis e umidade), distribuição granulométrica, testes de resistência mecânica a frio (DI), testes de resistência após reação com CO<sub>2</sub> (CSR) e teste de reatividade com CO<sub>2</sub> (CRI). (FERNANDES et al., 2019)*

*A resistência mecânica do coque ou resistência a frio do coque é definida como sendo a sua capacidade em resistir à fragmentação por solicitações mecânicas, impacto e/ou abrasão, em um tambor rotativo. Determina-se a extensão de sua redução granulométrica após ser submetido a um número fixo de revoluções em tambor de características padronizadas. (GIGLIO, 2016)*

A qualidade do coque metalúrgico é definida por fórmulas que consideram as qualidades dos carvões e os parâmetros de processo. Atualmente, é identificado divergências entre as qualidades calculadas através dos parâmetros dos carvões e os valores definidos após a realização dos testes com o coque. A assertividade nos parâmetros de qualidade do coque é importante no aspecto de controle de qualidade, evitando variações na qualidade quando existe a necessidade de fazer alterações na mistura de carvões, auxilia na política de compra de carvões, e proporciona o atingimento das qualidades requeridas pelos altos-fornos.

## 1.1 Objetivo Geral

O objetivo geral do trabalho é desenvolver um modelo de previsão para o DI (resistência mecânica) do coque metalúrgico a partir de análises estatísticas de dados da qualidade dos carvões individuais e dos parâmetros reais do processo de coqueificação.

### 1.1.1 Objetivos Específicos

Como objetivos específicos se tem:

- Levantar dados históricos dos parâmetros de qualidade dos carvões e dados históricos dos parâmetros do processo de coqueificação, como por exemplo as variáveis da mistura e dosagem de carvão, qualidade do carvão e coque, dados de processo e informações térmicas.
- Desenvolver um modelo para prever a resistência mecânica do coque metalúrgico (DI).
- Avaliar a aplicabilidade do modelo de previsão proposto por meio de testes.

## 1.2 Justificativa

O coque metalúrgico possui um alto peso no custo final do gusa e suas propriedades estão diretamente ligadas à performance dos altos-fornos. Com uma melhor assertividade dos parâmetros de qualidade do coque, será possível que a Unidade Técnica da Coqueria tome decisões quanto às mudanças na mistura de carvões, bem como, a Unidade Técnica dos Altos-Fornos preveja o consumo desta matéria-prima. Consequentemente, será possível mensurar o impacto financeiro no processo, podendo assim, auxiliar na seleção e compra de carvões.

## 1.3 Organização do Trabalho

O trabalho está estruturado em 6 Capítulos. Após essa introdução, o Capítulo 2 apresenta conceitos sobre a metodologia de pesquisa adotada neste trabalho. O Capítulo 3 apresenta a base conceitual utilizada para o desenvolvimento do trabalho.

O Capítulo 4 mostra o desenvolvimento do trabalho, onde na Seção 4.1 é apresentada a fórmula atual utilizada para a definição do DI, na Seção 4.2 é apresentado um modelo matemático desenvolvido pela área de Pesquisa e Desenvolvimento da usina siderúrgica onde o trabalho foi desenvolvido para definição do DI do coque, na Seção 4.3 é detalhado como a coleta de dados foi realizada, na Seção 4.4 é descrito sobre o tratamento dos dados e na Seção 4.5 é explicado como foi feita a seleção das variáveis para o desenvolvimento do modelo de previsão para o DI.

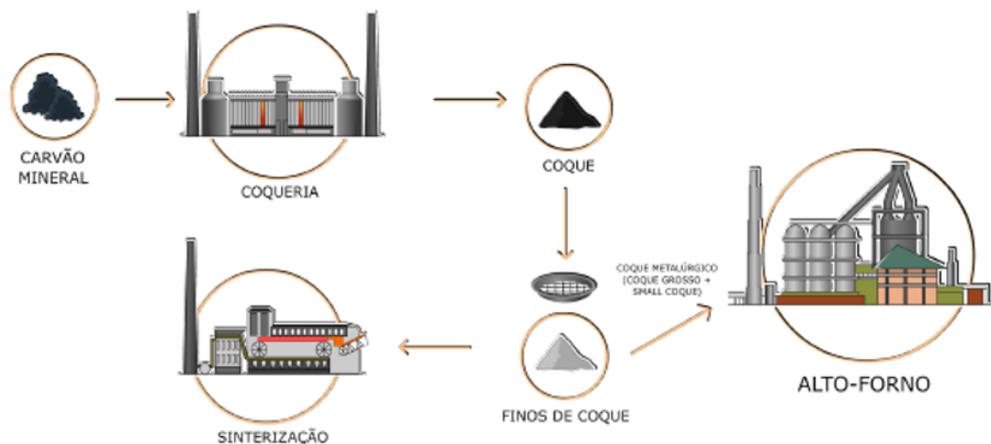
Os resultados são apresentados no Capítulo 5, seguido pelo Capítulo que detalha as considerações finais e por fim, as referências bibliográficas utilizadas no trabalho.

## 2 Revisão da Literatura

### 2.1 Fluxo da Redução

A fim de explicar o fluxo da redução, é necessário compreender o processo de um Alto-Forno, onde de fato o processo de redução ocorre. Para que o ferro gusa seja produzido, o Alto-forno precisa ser preenchido com camadas alternadas de coque e carga contendo minério de ferro. Durante o processo são gerados calor e gás pela gaseificação de coque e carvão por meio de jato quente pré-aquecido. No processo de geração de gás quente, o coque é consumido e o calor do gás é usado para derreter a carga ferrosa, que, na Figura 1, é representada pelo produto final do processo de Sinterização. Ao passo que a carga é consumida cria-se um vazio, e, conforme o processo ocorre, há um movimento da carga gasosa e ferrosa, que será vazada em furos na base do Alto-forno. (GEERDES; CHAIGNEAU; LINGIARDI, 2020)

Figura 1 – Fluxo da Redução



Fonte: Elaboração Própria

O foco do trabalho está no processo de coqueificação, que, conforme destacado anteriormente gera uma das matérias-primas para que o processo de redução ocorra. Contudo, os finos desse processo também são utilizados de combustível para um outro processo, a sinterização.

## 2.2 Processo de Coqueificação

O processo de coqueificação é a transformação do carvão em coque nos fornos da coqueria. De acordo com [Alves \(2017\)](#), o carvão mineral particulado é submetido a temperaturas elevadas na ausência de oxigênio no interior dos fornos da coqueria, e, libera seus voláteis, gerando como produtos gás (gás de coqueria), líquidos (alcatrão, óleos leves) e um produto sólido poroso, resistente e infusível denominado coque. A coqueria pode ser visualizada na [Figura 1](#).

Segundo [Alves \(2017\)](#), o processo de coqueificação pode ser dividido em seis etapas distintas, determinadas em função da temperatura de coqueificação e da fluidez do carvão dentro do forno da coqueria.

a) Desgaseificação: inicia com a formação de gases de baixo peso molecular ( $\text{CH}_4$  e  $\text{H}_2\text{O}$ ), para depois, com o aumento da temperatura inicial, a formação de elementos de peso molecular maior, como o alcatrão, sendo este o principal produto volátil obtido nessa etapa. A desgaseificação se processa até  $500^\circ\text{C}$ .

b) Amolecimento: corresponde aos princípios físicos da fusão de um material termoplástico, o ponto de amolecimento é inversamente proporcional à quantidade de vitrinita no carvão.

c) Zona Plástica: A duração da fase plástica é determinada pela competição das duas reações que acontecem nela, as reações de condensação e as de craqueamento, estas produzem substâncias de alto peso molecular. Os produtos de peso molecular maior permanecem mais tempo no estado líquido em função do calor. Há uma baixa permeabilidade gasosa na zona plástica, isto implicará em um aumento da pressão interior da zona plástica devido à produção de gases nesta fase. Pode-se observar que a pressão interior empurra o coque contra as paredes e comprime o carvão entre elas.

d) Inchamento/Pressão de Coqueificação: Devido ao aumento da quantidade de gás na zona plástica que não consegue escapar rapidamente tem-se o inchamento do carvão. Como informado na zona plástica, em virtude dessa plasticidade térmica dos carvões tem-se um desenvolvimento da pressão contra as paredes dos fornos durante a coqueificação, denominada pressão de coqueificação. Esta pressão de coqueificação é de extrema importância para o processo, porém ela afeta a estrutura dos fornos de bateria, logo está ligada diretamente à vida útil desses. A importância da pressão de coqueificação é que esta causará uma deformação dos grãos sólidos do carvão, que podem se tornar até mesmo fluídos, causando também decomposição da matéria orgânica, implicando diretamente na evolução de produtos voláteis.

e) Ressolidificação: Quando as reações de condensação prevalecem sobre as reações de craqueamento tem-se a decomposição térmica, principal responsável pela massa plástica se converter irreversivelmente ao estado sólido do semicoque ao exceder a temperatura de ressolidificação. Antes da própria ressolidificação a estrutura básica do coque já está formada, isto quer dizer que não existirão mais mudanças físico-químicas significativas na estrutura do mesmo, somente um aumento do grau de ordenação das estruturas moleculares e, conseqüentemente, um aumento da densidade.

f) Contração: O produto final da coqueificação é chamado de semicoque, para completar a produção do coque deve-se aumentar ainda mais a temperatura. A contração do coque é função dos resultados das reações de condensação e liberação de hidrogênio, após o processo de contração pode-se observar o aparecimento de fissuras nas pedras de coque.

Figura 2 – Coqueria



Fonte: Elaboração Própria

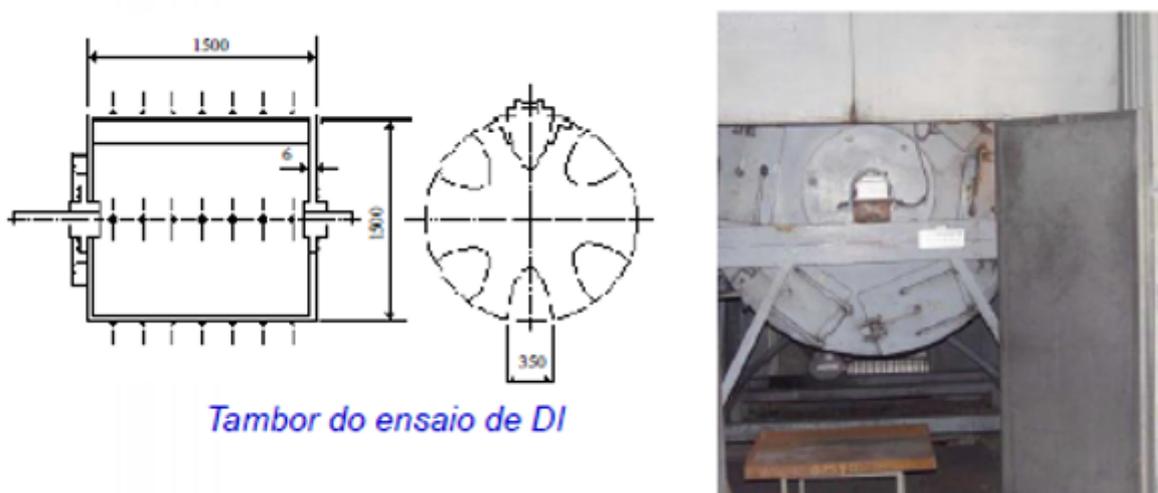
### 2.3 Resistência Mecânica – $DI_{(15-150)}$ (Drum Index)

A Resistência Mecânica é definida como sendo a capacidade em resistir à fragmentação por solicitações mecânicas (impacto e/ou abrasão) em um tambor rotativo. Determina-se a extensão de sua redução granulométrica após ser submetido a um número fixo de revoluções em tambor de características padronizadas. Dois tipos de índices podem ser derivados do teste:

- Um em que se considera a quantidade de finos produzidos (passante e uma determinada malha);
- E outro pela consideração da quantidade de coque graúdo (retido em determinada malha) que permanece presente após o tamboreamento.

Na siderúrgica onde o trabalho foi desenvolvido, o teste é realizado em tambor rotativo a temperatura ambiente, conforme ilustrado na Figura 3.

Figura 3 – Tambor do Ensaio de DI



Fonte: Elaboração Própria

Após as revoluções, a amostra é retirada e peneirada e o somatório da percentagem retida das frações superiores a 15 mm determina o resultado do teste.

## 2.4 Definição do Modelo

Para modelagem deste trabalho, foi selecionado o algoritmo Random Forest, em português "floresta aleatória". Segundo Breiman (2001), são combinações de árvores de decisão ajustadas em subamostras dos dados. Cada árvore depende dos valores de um vetor aleatório amostrado independentemente e com a mesma distribuição para todas as árvores da floresta.

Esse algoritmo possui vários parâmetros, quando um deles não é informado, assume-se o padrão. Neste trabalho, foram selecionados para uso os seguintes parâmetros:

- max\_depth: Profundidade máxima permitida nas árvores de decisão;
- min\_samples\_split: É o número mínimo de amostras necessárias para dividir um nó interno, sendo que o padrão é 2;

- `min_samples_leaf`: É o número mínimo de amostras que deve existir em cada folha, sendo que o padrão é 1;
- `n_estimators`: É o número de estimadores, ou seja, o número de árvores da floresta, sendo que o padrão é 100;
- `random_state`: É usado para fixar a semente para o gerador aleatório, assim, quando houver a divisão dos dados em treinamento e teste, os dados vão estar na mesma base.

## 2.5 Trabalhos Relacionados

Alguns trabalhos na literatura apresentam abordagens similares ao problema descrito nesse trabalho, utilizando aprendizagem de máquina para solucionar problemas na siderurgia.

[Hadavandi e Chelgani \(2019\)](#) estudaram as relações entre os componentes do carvão com a qualidade do coque. O estudo explorou as interdependências entre componentes de carvão medidos pelo método de Mutual Information (MI) e avaliou sua importância na previsão de índices de coqueificação para uma ampla gama de amostras de carvão, para isso, utilizaram o método Adaptive Boosting baseado em regressão de vetores de suporte (SVR), chamado Boosted-SVR.

Os resultados obtidos indicaram que o modelo Boosted-SVR pode ser bastante satisfatório para prever índices de coqueificação. Em geral, os resultados desta investigação demonstraram um potencial de previsão da qualidade do coque com um número limitado de variáveis de entrada e sugeriu que uma combinação de MI com o modelo Boosted-SVR como uma nova ferramenta poderosa que pode ser usada para o cálculo de outros problemas complexos de combustão e processamento com base na medição de propriedades convencionais.

[Cardoso e Felice \(2022\)](#) apresentaram redes neurais para prever a quantidade de impurezas (silício, fósforo e enxofre) na produção de ferro fundido em alto-forno. Nesta pesquisa, o método utilizado para avaliar a qualidade do modelo de rede neural foi o RMSE (root mean square error).

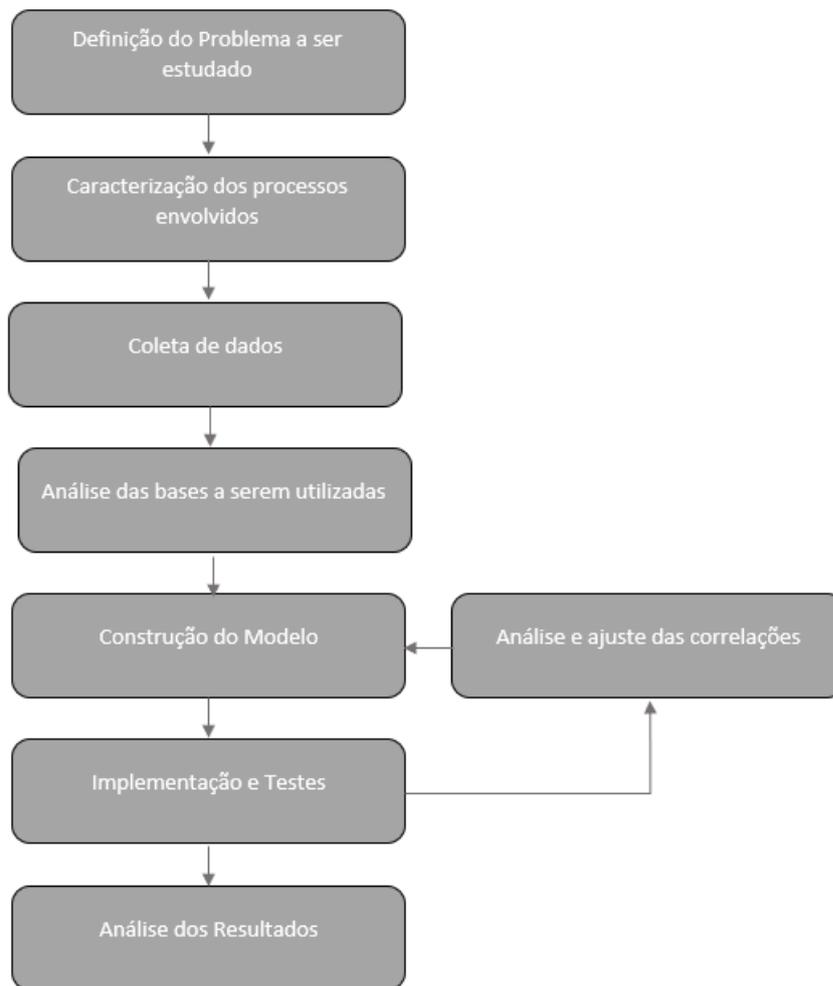
Em outra abordagem, [Li et al. \(2022\)](#) propõem um modelo de previsão do Coke Rate do Alto-Forno baseado em agrupamento fuzzy C-means (FCM) e regressão vetorial de suporte para otimização de pesquisa em grade (SVR) é proposto para obter uma previsão precisa do Coke Rate. Primeiro, os dados do sensor do Alto Forno e os dados de produção da usina siderúrgica foram pré-processados. Em seguida, o algoritmo FCM foi usado para agrupar os dados em diferentes condições de Forno. Finalmente, o modelo SVR otimizado por pesquisa de grade foi usado para prever o Coke Rate sob diferentes condições de Alto-Forno. Os experimentos mostraram que o modelo pode prever o Coke Rate em cada lote quando as condições do Alto-Forno flutuando e tem alta precisão de tempo e estabilidade.

### 3 Metodologia

A metodologia a ser abordada no presente trabalho de pesquisa pode ser classificada como Hipotético-Dedutiva com procedimento estatístico, pois este conhecimento científico segue os passos de definição do problema, que trata da dificuldade de previsão da resistência mecânica do coque, seguido da elaboração de novas teorias que serão testadas através da análise de correlação de variáveis, sendo posteriormente testada estatisticamente para definição do erro. A necessidade da determinação de um modelo de previsão de resistência mecânica do coque consiste em um intuito pré-determinado da pesquisa, desta maneira, sua natureza pode ser classificada como aplicada, experimental e quantitativa, visto que a partir dos objetivos serão levantadas as variáveis envolvidas e sua correlação com os resultados.

A ciência dos dados é uma "ciência que trata de obter conhecimento e informação de forma sistemática, bem como normalizar e organizar esse conhecimento." (AMARAL, 2016) Desta forma, o presente trabalho busca obter conhecimento a partir de dados de processo, além dos dados de qualidade de matéria-prima a fim de gerar um modelo de previsibilidade que auxilie nas tomadas de decisão relacionadas às variáveis do processo de Coqueria, o que corresponde ao cumprimento das etapas de desenvolvimento do trabalho, organizadas conforme o fluxo apresentado na Figura 4.

Figura 4 – Diagrama de Desenvolvimento do Trabalho



Fonte: Elaboração Própria

O processo de coleta de dados será realizado por meio de levantamento dos dados em fontes de automação como PIMS e o SQL Server. Será necessário o pré-processamento dos dados afim de correlacionar um conjunto dedados que possuem informações registradas em etapas diferentes do processo, além da localização de possíveis *outliers* e falhas de gravação de dados.

A confiabilidade do modelo será comprovada com a implementação de testes da correlação, frente os dados reais de resistência mecânica, onde será possível testar o modelo e, por fim, analisar os resultados obtidos.

## 4 Desenvolvimento

### 4.1 Cálculo do $DI_{(15-150)}$

Existe uma fórmula de predição do DI do coque que leva em consideração apenas os parâmetros de qualidade da mistura de carvões (fluidez, matéria volátil e cinza). Essa fórmula é antiga e foi utilizada por um bom tempo, porém, não é aplicada as condições de processo atuais.

$$DI_{(15-150)} = (2,1879 \times \text{Fluidez}) - (0,3317 \times [\text{MatVol} / (100 - \text{Cinza})] \times 100) + 87,7270 \quad (4.1)$$

Em que:

- Fluidez = Fluidez do Carvão
- MatVol = Matéria Volátil
- Cinza = Cinza do Carvão

Com o propósito de comparar o DI obtido com a equação 4.1 mencionada acima, é possível notar na Tabela 1 que há uma divergência significativa entre a coluna DI, que é o resultado real, e a coluna Fórmula.

Tabela 1 – Comparação DI Obtido x Fórmula

DI	MV	CZ	Fluidez	Fórmula
83.4	22.5	5.8	1.2	90.3
83.5	22.8	5.9	1.3	90.5
83.6	22.7	5.9	1.5	91
83.7	22.6	6	1.6	91.2
83.8	22.2	5.9	1.5	90.9
83.9	22.2	6	1.7	91.4
84	22.2	5.8	1.7	91.5
84.1	22.2	5.8	1.9	91.8
84.2	22.1	5.8	1.9	91.9
84.3	22.2	5.9	1.9	91.8
84.4	22	5.8	1.8	91.7
84.5	22.2	5.9	1.8	91.7
84.6	22.1	6	1.8	91.5
84.7	22	6	1.7	91.4
84.8	22.2	5.9	1.7	91.3
84.9	21.8	5.6	1.3	90.4
85	22.2	5.9	2	92

Fonte: Elaboração Própria

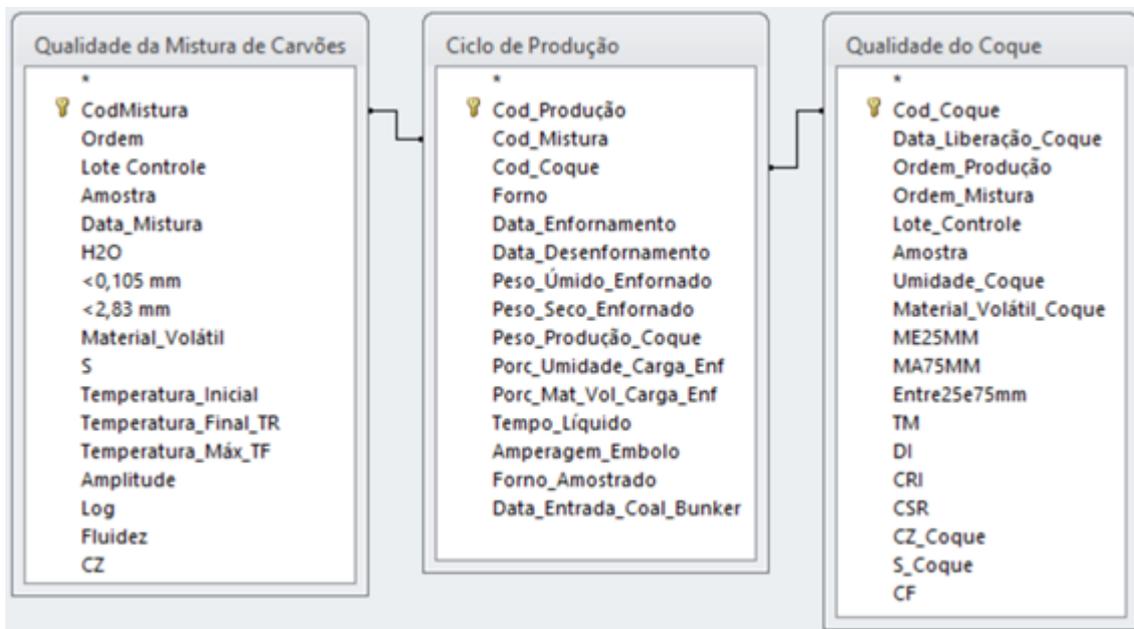
Como referência, o DI deve ser maior ou igual a 84. Ao utilizar a equação 4.1 em dados obtidos do processo, percebe-se uma variação muito grande entre o DI real em comparação com a fórmula antiga. Esse é o motivo dela não estar sendo aplicada atualmente, reforçando a necessidade de um modelo de predição mais assertivo.

## 4.2 Coleta dos Dados

Após realizar o estudo do processo de coqueificação, foi realizado a coleta de dados das etapas necessárias para a produção do coque, dentre elas estão a qualidade da mistura de carvão, ciclo de produção e resultados da qualidade do coque.

Os dados da qualidade da mistura foram extraídos do SAP, já os dados de produção e qualidade do coque foram extraídos do sistema de automação da Coqueria 3 através do SQL Server Management Studio, tendo como parâmetro o período de janeiro de 2020 a setembro de 2021. O relacionamento dos dados desde a qualidade da mistura, produção até os resultados da qualidade do coque resultaram em uma base de dados com 56.994 registros. Os dados são completos, ou seja, representam toda a população do período mencionado e seu relacionamento pode ser visto na Figura 5.

Figura 5 – Relacionamento da Base de Dados



Fonte: Elaboração Própria

### 4.2.1 Qualidade da Mistura dos Carvões

Após realizada a mistura, que é a combinação de tipos de carvões, é feita uma amostragem que ocorre a cada duas horas com o objetivo de verificar a qualidade resultante da dosagem de carvões. Abaixo consta a Tabela 2 que relaciona as variáveis de qualidade presentes na mistura dos carvões.

Tabela 2 – Estrutura da Tabela da Qualidade da Mistura de Carvões

<b>Variável</b>	<b>Descrição</b>	<b>Tipo</b>
Cod_Mistura	Código da mistura (chave primária)	Número (Integer)
Ordem	Sequência de processo	Texto
Lote_Controlo	Lote de controle	Texto
Amostra	Número da amostra	Texto
Data_Mistura	Data e hora da liberação da análise	Data/Hora
H2O	Umidade	Número (Float)
<0,105 mm	Granulometria menor que 0,105 mm	Número (Float)
<2,83 mm	Granulometria menor que 2,83 mm	Número (Float)
Material_Volátil	Material volátil da mistura de carvão	Número (Float)
S	Enxofre	Número (Float)
Temperatura_Inicial_TI	Temperatura inicial do amolecimento do carvão	Número (Float)
Temperatura_Final_TR	Temperatura final do amolecimento do carvão	Número (Float)
Temperatura_Máx_TF	Temperatura máxima do amolecimento do carvão	Número (Float)
Amplitude	Diferença entre TR e TI	Número (Float)
Fluidez	Divisão do dial por minuto (ddpm)	Número (Float)
Log	Logaritmo da fluidez	Número (Float)
CZ	Cinza	Número (Float)

Fonte: SAP

### 4.2.2 Ciclo de Produção

A Coqueria 3 possui 80 fornos, o ciclo de produção de cada forno é iniciado quando este for carregado com a mistura de carvão, após atingir o tempo de coqueificação, transformando o material carregado em coque, este ciclo é encerrado e o coque é desenformado. Há uma relação direta entre a data e hora que a carga entra nos silos de abastecimento da coqueria - Coal Bunker, com a data e hora da mistura de carvões. Abaixo consta a Tabela 3 que relaciona as variáveis resultantes do processo de coqueificação.

Tabela 3 – Estrutura da Tabela de Ciclo de Produção

Variável	Descrição	Tipo
Cod_Produção	Código da produção (chave primária)	Número (Integer)
Cod_Mistura	Código da mistura (chave secundária)	Número (Integer)
Cod_Coque	Código do coque (chave secundária)	Número (Integer)
Forno	Número do forno	Número (Integer)
Data_Enfornamento	Data e hora de carregamento do forno	Data/Hora
Data_Desenfornamento	Data e hora do desenfornamento	Data/Hora
Peso_Úmido_Enfornado	Peso da mistura de carvão sem desconto da umidade	Número (Float)
Peso_Produção_Coque	Produção calculada de coque	Número (Float)
Porc_Umididade_Carga_Enf	Percentual de umidade na mistura	Número (Float)
Porc_Mat_Vol_Carga_Enf	Percentual de material volátil na mistura	Número (Float)
Tempo_Líquido	Tempo gasto em horas para conclusão do processo	Número (Float)
Amperagem_Embolo	Corrente elétrica de esforço gasta para retirar o coque	Número (Float)
Forno_Amostrado	Forno que foi amostrado para qualidade	Número (Integer)
Data_Entrada_Coal_Bunker	Data e hora da entrada da mistura na coqueria	Data/Hora

Fonte: Sistema de Automação da Coqueria 3 - SQL Server Management Studio

Adicionalmente, foram criadas algumas variáveis calculadas para dar apoio na análise dos dados, permitindo identificar possíveis *outliers*, conforme mostrado na Tabela 4.

Tabela 4 – Estrutura das Variáveis Calculadas do Ciclo de Produção

Variável	Descrição	Tipo
Tempo_até_Enfornamento_h	Diferença em horas da data de mistura de carvões pela data de enfornamento	Número (Float)
Tempo_Coqueificação_h	Diferença em horas da data de enfornamento pela data de desenfornamento	Número (Float)
IO	Índice Operacional que é dado pelo cálculo: $24 / \text{Tempo\_Coqueificação\_h} * 100$	Número (Float)
Tempo_até_Qualidade_Coque_h	Diferença em horas da data de desenfornamento pela data da qualidade do coque	Número (Float)

Fonte: Elaboração Própria

### 4.2.3 Qualidade do Coque

Após o processo de produção do coque, é feita uma amostragem que ocorre automaticamente a cada oito horas com o objetivo de medir a qualidade do coque, parâmetros essenciais para garantir um bom desempenho nos Altos-Fornos. Assim, o objetivo deste trabalho é verificar a relação das entradas com a variável DI (resistência mecânica) do coque metalúrgico.

Tabela 5 – Estrutura da Tabela da Qualidade do Coque

Variável	Descrição	Tipo
Cod_Coque	Código do coque (chave secundária)	Número (Integer)
Data_Liberação_Coque	Data e hora da liberação da análise	Data/Hora
Ordem_Produção	Código da ordem de processo de produção	Texto
Ordem_Mistura	Mistura de carvão utilizado	Texto
Lote_Control	Lote de controle do material	Número (Integer)
Amostra	Número da amostra	Número (Integer)
Umidade_Coque	Umidade do coque	Número (Float)
Material_Volátil_Coque	Material volátil do coque	Número (Float)
ME25MM	Percentual de coque com granulometria menor que 25 mm	Número (Float)
MA75MM	Percentual de coque com granulometria maior que 75 mm	Número (Float)
Entre25e75mm	Percentual de coque com granulometria entre 25 e 75 mm	Número (Float)
TM	Tamanho médio do coque	Número (Float)
DI	Resistência mecânica a frio do coque	Número (Float)
CRI	Reatividade do coque ao CO <sub>2</sub>	Número (Float)
CSR	Resistência do coque após a reação	Número (Float)
CZ_Coque	Percentual de Cinza	Número (Float)
S_Coque	Percentual de Enxofre	Número (Float)
CF	Carbono fixo	Número (Float)

Fonte: Sistema de Automação da Coqueria 3 - SQL Server Management Studio

## 4.3 Tratamento dos Dados

Ao fazer análise descritiva das variáveis numéricas, foi verificado pontos discrepantes em algumas variáveis, desta forma, a equipe responsável pela coleta e registro das amostras foi consultada a fim de definir o nível de resultado adequado para cada variável. Após aplicado os critérios mencionados, a base de dados que inicialmente contava com 56.994 registros, passou a contar com 44.324 registros.

### 4.3.1 Critérios para Exclusão de Dados da Qualidade da Mistura

A definição dos critérios apresentada na Tabela 6, foi baseada na análise dos dados do período coletado, e, juntamente com o especialista do setor foi possível estabelecer os níveis aceitáveis para considerar válido os resultados da amostra.

Tabela 6 – Critérios para Exclusão de Dados da Qualidade da Mistura

Variável	Critério	Observação
H2O	> 0	Valores maiores que zero
<0,105 mm	> 0	Valores maiores que zero
<2,83 mm	> 0	Valores maiores que zero
Material_Volátil	> 0	Valores maiores que zero
S	> 0 e < 1	Valores maiores que zero e menores que 1
Temperatura_Inicial	> 0	Valores maiores que zero
Temperatura_Final_TR	> 0	Valores maiores que zero
Temperatura_Máx_TF	> 0	Valores maiores que zero
Amplitude	> 0	Valores maiores que zero
Log	> 0	Valores maiores que zero
Fluidez	> 0	Valores maiores que zero
CZ	> 0 e < 16	Valores maiores que zero e menores que 16

Fonte: Elaboração Própria

### 4.3.2 Critérios para Exclusão de Dados do Ciclo de Produção

Para definir registros que possam ser desconsiderados, através de algumas variáveis criadas, como por exemplo o tempo de coqueificação, todo ciclo de produção maior que 72 horas indica um possível engaiolamento (quando não há saída da carga coqueificada de dentro do forno da coqueria), assim, este dado deve ser descartado. O tempo líquido está alinhado com o tempo de coqueificação, porém, foi definido um intervalo de tempo menor para considerar os dados válidos. O intervalo de amperagem/emboło foi definido pelo especialista de coqueria após análise dos dados, conforme mostrado na Tabela 7.

Tabela 7 – Critérios para Exclusão de Dados do Ciclo de Produção

Variável	Critério	Observação
Tempo_Coqueificação_h	<= 72	Valores menores ou iguais a 72 horas
Tempo_Líquido	Entre 8 e 25	Valores entre 8 e 25 horas
Amperagem_Emboło	Entre 4 e 400	Valores entre 4 e 400

Fonte: Elaboração Própria

### 4.3.3 Critérios para Exclusão de Dados da Qualidade do Coque

O DI é o objeto do estudo, para tanto, o único critério adotado foi que ele não fosse nulo, conforme demonstrado na Tabela 8.

Tabela 8 – Critérios para Exclusão de Dados da Qualidade do Coque

Variável	Critério	Observação
DI	> 0	Valores maiores que zero

Fonte: Elaboração Própria

#### 4.4 Seleção das Variáveis

A base de dados possui variáveis categóricas nominais e variáveis numéricas discretas e contínuas, para dar continuidade ao estudo, somente as variáveis numéricas discretas foram consideradas, totalizando 23 variáveis, como demonstrado na Tabela 9.

Tabela 9 – Variáveis Numéricas da Base de Dados

Tabela	Variável
Qualidade da Mistura de Carvões	<0,105 mm
Qualidade da Mistura de Carvões	<2,83 mm
Qualidade da Mistura de Carvões	H2O
Qualidade da Mistura de Carvões	Material_Volátil
Qualidade da Mistura de Carvões	S
Qualidade da Mistura de Carvões	CZ
Qualidade da Mistura de Carvões	TR
Qualidade da Mistura de Carvões	TI
Qualidade da Mistura de Carvões	Amplitude
Qualidade da Mistura de Carvões	Log
Qualidade da Mistura de Carvões	TF
Qualidade da Mistura de Carvões	Fluidez
Variável Calculada	Tempo_até_Enfornamento_h
Variável Calculada	Tempo_Coqueificação_h
Variável Calculada	IO
Ciclo de Produção	Tempo_até_Qualid_Coque_h
Ciclo de Produção	Peso_Úmido_Enfornado
Ciclo de Produção	Peso_Produção_Coque
Ciclo de Produção	Porc_Umidade_Carga_Enf
Ciclo de Produção	Porc_Mat_Vol_Carga_Enf
Ciclo de Produção	Tempo_Líquido
Ciclo de Produção	Amperagem_Embolo
Qualidade do Coque	DI

Fonte: Elaboração Própria

A qualidade da mistura de carvões é um dado de entrada do modelo por já possuir o resultado da mistura dos carvões selecionados, enquanto a qualidade do coque é a saída. Portanto, pode ser observado que da tabela de qualidade do coque, apenas a variável DI foi adicionada ao modelo devido ela ser o objeto de estudo de trabalho e para que as demais variáveis desta tabela não influenciem nos resultados.

Foi feito a interpretação da análise descritiva (Figura 6) das variáveis numéricas selecionadas e foi feita análise em dimensão univariada a fim de visualizar o comportamento dos dados e suas distribuições, para isso, foi utilizado o histograma conforme apresentado na Figura 7.

	<0,105 mm	<2,83 mm	H2O	Material_Volátil	S	CZ	TR	TI	Amplitude	Log	...
count	44324.00	44324.00	44324.00	44324.00	44324.00	44324.00	44324.00	44324.00	44324.00	44324.00	...
mean	10.29	83.63	8.97	22.21	0.82	5.86	482.23	399.16	83.07	3.14	...
std	1.54	1.37	0.80	1.42	0.05	0.75	4.49	5.77	6.75	0.25	...
min	0.30	73.27	6.03	16.62	0.61	0.85	468.50	372.50	71.00	2.56	...
25%	9.64	83.06	8.47	21.31	0.79	5.37	479.38	395.80	78.00	3.03	...
50%	10.23	83.70	8.87	22.17	0.82	5.82	483.00	399.00	82.00	3.19	...
75%	10.98	84.23	9.30	23.09	0.85	6.30	484.75	403.00	87.25	3.31	...
max	25.57	90.02	15.38	28.49	0.99	11.44	492.00	410.50	112.50	3.74	...

8 rows x 24 columns

Figura 6 – Análise Descritiva das Variáveis Numéricas Selecionadas

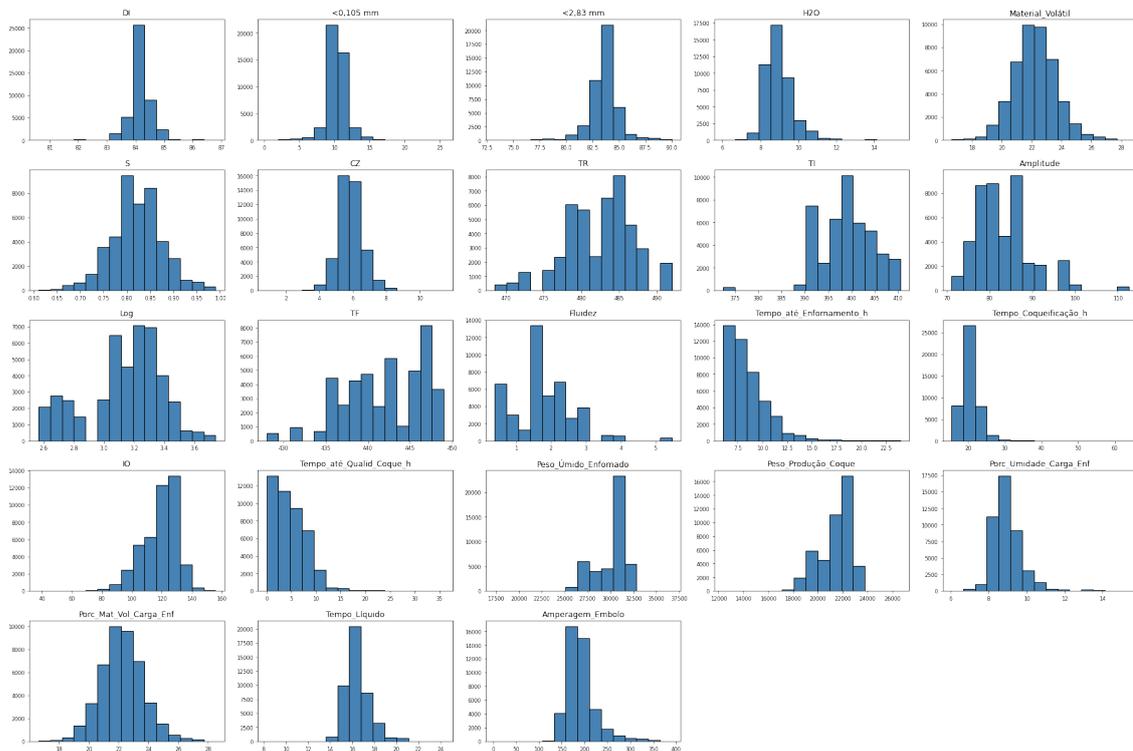


Figura 7 – Histograma com Todas as Variáveis Numéricas

A análise feita pelo histograma é de forma individualizada, para apoiar na definição estatística de variáveis que deveriam ser retiradas do modelo, foi necessário verificar a correlação entre as variáveis, porém, devemos observar que as elas possuem grandezas diferentes, então é importante ressaltar que os dados precisam estar numa escala comum. Existe a técnica de padronização que coloca a média dos dados em 0 e o desvio padrão em 1 e a técnica de normalização que deixa os dados num intervalo entre 0 e 1, essa última, foi a técnica selecionada para o trabalho.

Após realizada a normalização das variáveis, foi criada uma matriz de correlação representada através de um heatmap (mapa de calor), que permite visualização dos dados em duas dimensões, como demonstrado na Figura 8.

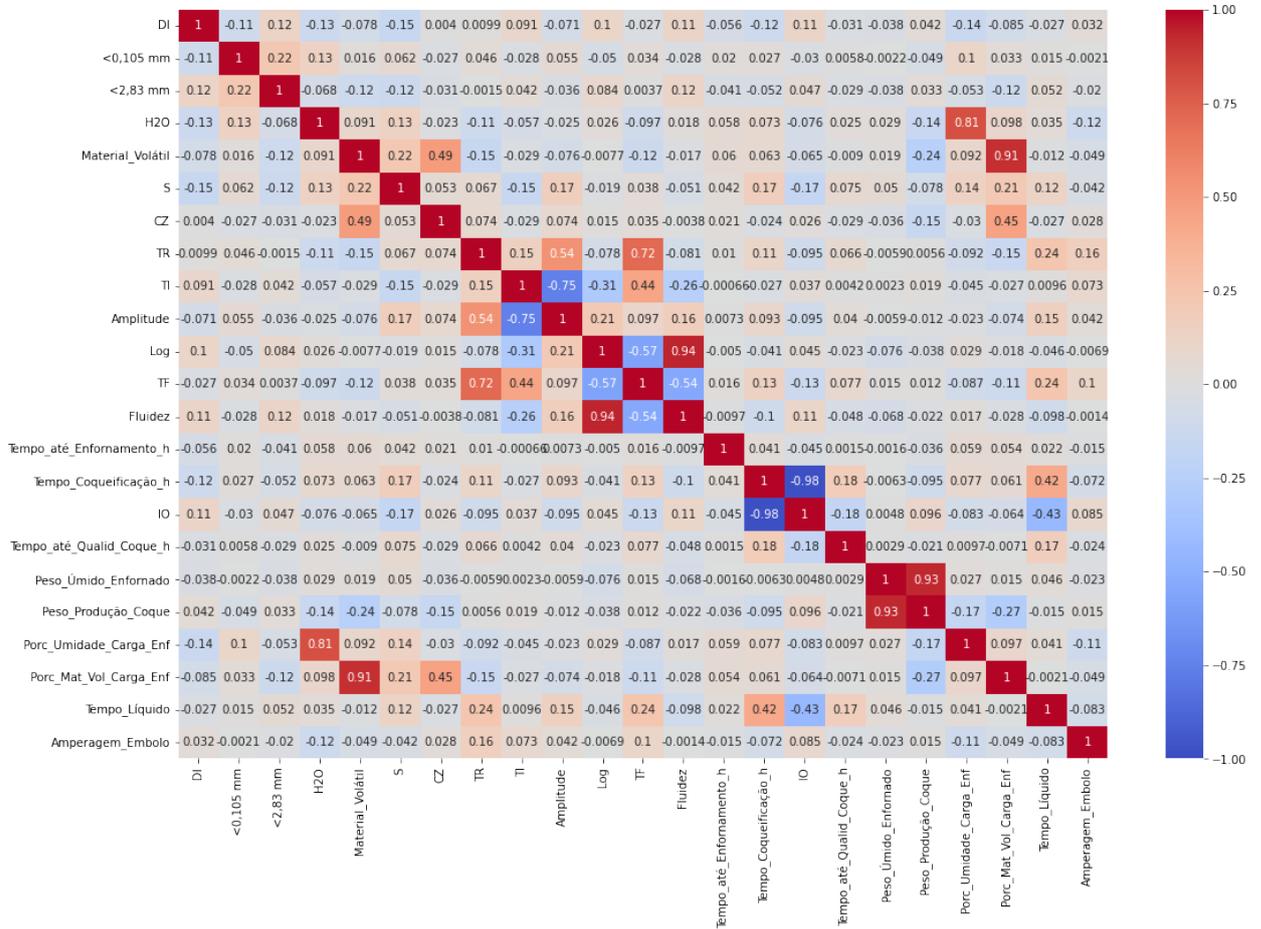


Figura 8 – Mapa de Calor com Todas as Variáveis Numéricas

Na matriz de correlação das variáveis as correlações positivas são representadas pela cor vermelha enquanto a cor azul representa as correlações negativas. Após análise do heatmap foi visto variáveis altamente correlacionadas, ou seja, uma já explica a outra, vimos que o modelo poderia ser simplificado. Foi consultado o especialista da área e foi percebido que as variáveis mais importantes, ou seja, as que vão influenciar no processo são da qualidade da mistura de carvões, sendo assim, elas foram priorizadas, além disso, pensando em implementação do modelo no ambiente produtivo, foi decidido retirar algumas variáveis com grande correlação, assim, foram retiradas 10 variáveis do modelo: Temperatura Inicial (TI), Log, Tempo até o Enfornamento, Tempo de Coqueificação, Tempo até a Qualidade do Coque, Peso Produção Coque e Porc Mat Vol Carga Enf, Porc Mat Vol Carg Enf, Tempo Líquido e Amperagem Embolo, resultando em 13 variáveis numéricas.

A temperatura inicial foi retirada pois o modelo já está contando com outras duas variáveis de semelhantes, a temperatura final e máxima, estas permaneceram, pois o modelo antigo as levava em consideração.

A variável log foi retirada por estar fortemente correlacionada com a variável fluidez, isso, porque ela é uma função logarítmica da fluidez.

A variável Tempo de Coqueificação está negativamente correlacionada com a variável calculada IO (Índice Operacional), isso porque o tempo de coqueificação está presente na fórmula do IO, por isso, alta correlação negativa, assim, foi possível retirar do modelo. Além desta variável de tempo, o Tempo até Enfornamento, Tempo até Qualidade do Coque e Tempo Líquido também foram retiradas, pois seu objetivo era de encontrar possíveis outliers comparando o tempo normal aceitável de cada processo.

Existem duas variáveis de peso no modelo, o Peso Úmido Enfornado e Peso Produção Coque, a diferença entre eles é que o peso úmido é maior, já o peso da produção desconta a umidade e as perdas do processo, por isso, estão fortemente correlacionados, desta forma, foi decidido manter o peso úmido enfornado no modelo.

A variável Porcentagem de Material Volátil na Carga Enfornada possui forte correlação com a variável Material Volátil, da mesma forma a Porcentagem de Umidade na Carga Enfornada possui forte correlação com a variável H2O, como elas são variáveis resultantes do processo, decidiu-se manter as variáveis da qualidade da mistura de carvões.

A variável Amperagem Embolo, está presente no final do processo, está associada a intensidade de uma corrente elétrica em amperes que é feita no processo de desenfornamento, porém, não tem influência significativa no DI.

Tabela 10 – Variáveis Numéricas Seleccionadas

<b>Tabela</b>	<b>Variável</b>
Qualidade da Mistura de Carvões	<0,105 mm
Qualidade da Mistura de Carvões	<2,83 mm
Qualidade da Mistura de Carvões	H2O
Qualidade da Mistura de Carvões	Material_Volátil
Qualidade da Mistura de Carvões	S
Qualidade da Mistura de Carvões	CZ
Qualidade da Mistura de Carvões	TR
Qualidade da Mistura de Carvões	Amplitude
Qualidade da Mistura de Carvões	TF
Qualidade da Mistura de Carvões	Fluidez
Variável Calculada	IO
Ciclo de Produção	Peso_Úmido_Enfornado
Qualidade do Coque	DI

Fonte: Elaboração Própria

Após esse alinhamento, ficou definido seguir com 13 variáveis, conforme Tabela 10 acima. O mapa de calor foi atualizado, conforme apresentado na Figura 9, exibindo com mais clareza as variáveis que influenciam no processo e que serão utilizadas deste ponto em diante no trabalho.

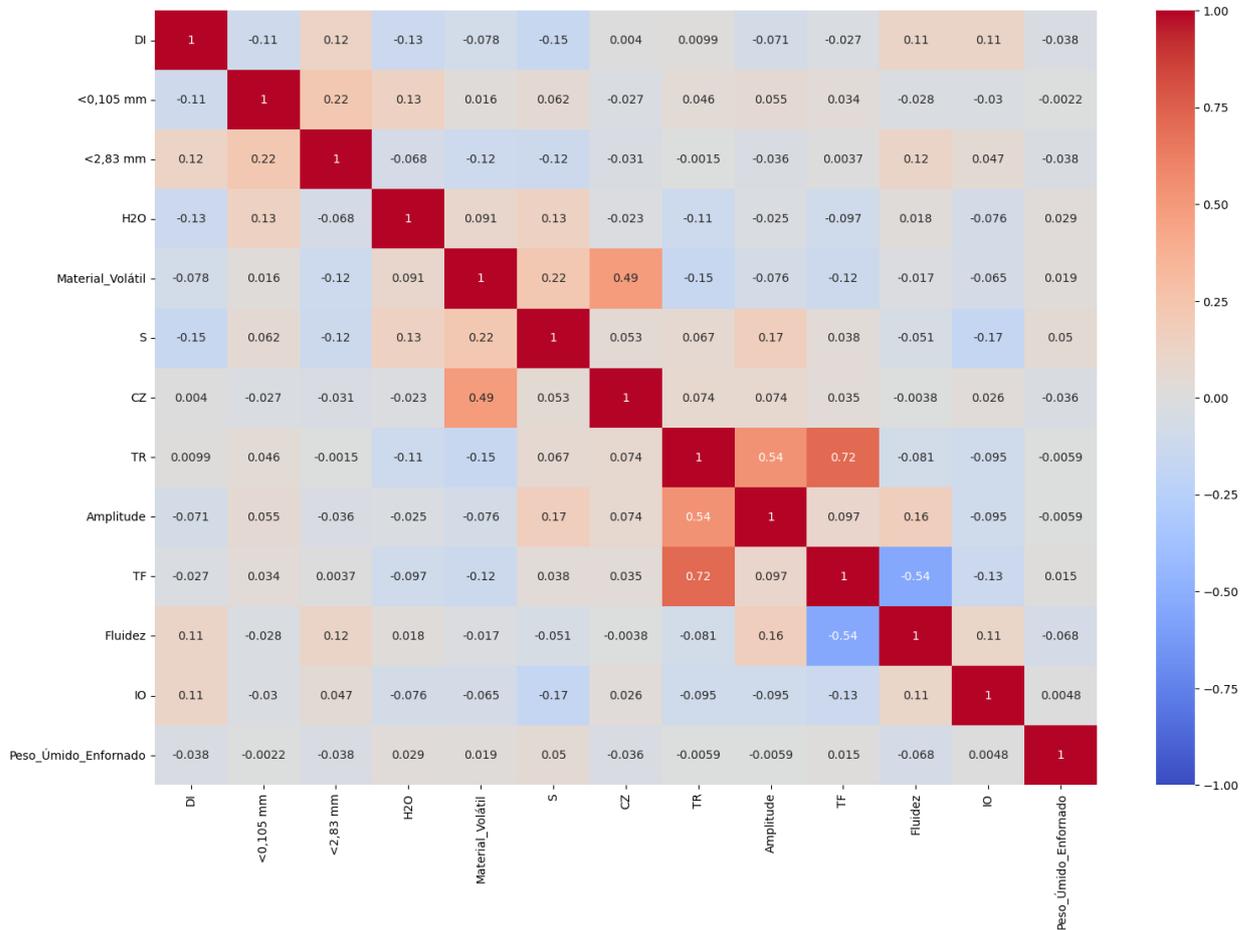


Figura 9 – Mapa de Calor com as Variáveis Numéricas Selecionadas

## 4.5 Modelagem

### 4.5.1 Parametrizando o Modelo

Para encontrar a melhor performance do Random Forest, os parâmetros devem ser ajustados até encontrar um resultado satisfatório. A biblioteca scikit-learn possui dois algoritmos que apoiam essa tarefa, o primeiro é o GridSearchCV que faz todas as combinações possíveis dos parâmetros, o segundo é o RandomizedSearchCV que faz apenas uma quantidade específica de combinações limitada a valores previamente definidos. Os dois algoritmos são excelentes, a diferença é que o GridSearchCV vai fazer todas as combinações, ou seja, terá um alto custo computacional, enquanto, que o RandomizedSearchCV fará uma quantidade de combinações menor por variando os parâmetros com um intervalo menor de possibilidades, assim, o custo computacional é reduzido e o resultado é obtido com maior velocidade.

O RandomizedSearchCV foi o algoritmo selecionado para apoiar na configuração dos melhores parâmetros a serem considerados no RadomForest. Foram realizadas várias simulações para chegar na combinação final variando os parâmetros conforme abaixo:

- max\_depth: Arranjo de 2 até 20 variando de 1 a 1.
- min\_samples\_split: Arranjo de 2 até 11 variando de 1 a 1.
- min\_samples\_leaf: Arranjo de 1 até 7 variando de 1 a 1.
- n\_estimators: Arranjo de 5, 25, 50, 100, 200, 250.
- n\_iter: Iterações de 25, 50 e 100.

A Tabela 11 demonstra os resultados das simulações realizadas:

Tabela 11 – Simulações Realizadas com o RandomizedSearchCV

<b>Arranjos</b>	<b>max_depth</b>	<b>min_samples_split</b>	<b>min_samples_leaf</b>	<b>n_estimators</b>	<b>n_iter</b>	<b>Tempo</b>
1° Teste	9	8	2	50	25	00:04:22
2° Teste	19	2	1	25	25	00:25:12
3° Teste	19	6	2	50	50	00:37:18
4° Teste	19	6	2	50	100	01:11:25
5° Teste	19	3	1	250	25	00:40:15

Fonte: Elaboração Própria

O RadomForest foi parametrizado com os valores encontrados em cada simulação e a precisão do modelo é demonstrada na Tabela 12.

Tabela 12 – Precisão do Modelo de Cada Simulação

<b>Arranjos</b>	<b>R-quadrado</b>
1° Teste	0.586
2° Teste	0.949
3° Teste	0.936
4° Teste	0.936
5° Teste	0.948

Fonte: Elaboração Própria

Utilizar o parâmetro n\_estimators com valores elevados gastam mais tempo computacional, após comparação dos resultados percebeu-se que usar 100, 200 ou 250 não afetaria significativamente a precisão do modelo, portanto, foi definido usar 100 estimadores. Para os parâmetros min\_samples\_split e min\_samples\_leaf foi definido utilizar os valores padrões, 2 e 1 respectivamente. Já para o max\_depth, em vários testes o valor encontrado foi 19 de profundidade máxima. Desta forma, juntando os resultados das simulações com outros testes realizados manualmente, o modelo final ficou configurado com os parâmetros citados na Tabela 13.

Tabela 13 – Parâmetros Selecionados para o Random Forest

<b>max_depth</b>	<b>min_samples_split</b>	<b>min_samples_leaf</b>	<b>n_estimators</b>
20	2	1	100

Fonte: Elaboração Própria

Com os parâmetros definidos, o próximo passo é executar o modelo e verificar o R-quadrado, onde, geralmente, quanto maior, melhor o modelo se ajusta aos seus dados. No presente trabalho o R-quadrado resultante foi de 96,2% indicando que o modelo explica bem a variabilidade dos dados de resposta.

#### 4.5.2 Divisão da Base de Dados

Definido o modelo e seus parâmetros, é preciso verificar se há overfitting nos dados, ou seja, é necessário separar a base de dados em treinamento e teste para confirmar que o desempenho do modelo seja bom tanto nos dados de treino quanto nos dados de teste. Os dados para treinamento com a proporção de 70% da base de dados e 30% para teste, nessa divisão, a base de treinamento ficou com 31026 registros e a base de testes com 13298 registros. Feito a divisão, o modelo foi executado separadamente para a base de treinamento e teste conforme apresentado abaixo na Tabela 14.

Tabela 14 – Precisão do Modelo na Base de Treinamento e Teste

<b>Base</b>	<b>Score</b>
Treinamento	0,965
Teste	0,907

Fonte: Elaboração Própria

Percebe-se que o resultado se manteve na base de treinamento (0.965), já a base de teste houve uma redução (0.907), porém, continua sendo um bom resultado, mas para ter certeza que realmente é um bom resultado, existem algumas métricas de avaliação para medir o desempenho de modelos de regressão, como por exemplo o MSE (média dos erros ao quadrado), RMSE (raiz do erro quadrático médio), MAE (erro médio absoluto) e MAPE (erro médio percentual absoluto).

Para o presente trabalho, foram selecionadas as métricas RMSE e MAPE, onde, enquanto o RMSE avalia o erro nas unidades originais, o MAPE é uma medida relativa do erro, retornando o erro em valores percentuais. Essas duas métricas avaliadas em conjunto ajudam a entender melhor o modelo está se saindo, e como os resultados das duas métricas, tanto na base de treinamento, quanto na base de teste estão próximos de zero, significa que o erro é muito baixo e confirma a precisão do modelo, quando o erro nos dados de treino é baixo e próximo ao erro nos dados de teste indicam que o ajuste está adequado, não ocorre overfitting e que o modelo pode ser aplicado em dados externos (dados atuais).

Tabela 15 – Resultado das Métricas RMSE e MAPE

<b>Base</b>	<b>RMSE</b>	<b>MAPE</b>
Treinamento	0,076	0,055
Teste	0,127	0,084

Fonte: Elaboração Própria

Como observado na Tabela 15, o RMSE que é a raiz quadrada da média dos erros ao quadrado retornou 0,076 para a base de treinamento e 0,127 para a base de teste. Já o MAPE que é o erro médio percentual absoluto retornou 0,055 para a base de treinamento e 0,084 para a base de teste. As duas métricas foram analisadas em conjunto e ambas resultaram em valores próximo de zero, indicando baixo erro no modelo.

Outra técnica utilizada na interpretação dos resultados e a avaliação do modelo para identificar se o mesmo possui boa performance, ou até mesmo se o modelo vai continuar performando bem com outros dados é a técnica de Cross Validation, onde, o presente trabalho vai utilizar essa técnica associada ao método K-fold que tem o objetivo de dividir a base de dados de forma aleatória em K subconjuntos com aproximadamente a mesma quantidade de amostras em cada um deles.

Cada iteração de treino e teste possui, um conjunto formado por K-1 subconjuntos que são utilizados para treinamento e o subconjunto restante será utilizado para teste, para avaliação desse método, o presente trabalho utilizou o score como métrica para avaliar os resultados. O parâmetro n\_splits do método K-fold possui um padrão inicial igual a 5, porém, neste trabalho foi utilizado um valor de 10 iterações apresentadas na Tabela 16.

Tabela 16 – Iterações Utilizando o Método K-fold

<b>Iteração</b>	<b>R-quadrado</b>
1	0,915
2	0,876
3	0,892
4	0,895
5	0,869
6	0,907
7	0,868
8	0,911
9	0,905
10	0,891

Fonte: Elaboração Própria

Essa métrica foi utilizada tendo como saída o R-quadrado que é uma medida estatística para verificar quão próximos os dados estão da linha de regressão ajustada, ou seja, é a porcentagem da variação da variável resposta que é explicada pelo modelo. Neste caso, após 10 iterações com dados de treinos e teste aleatórios, o método k-fold resultou num R-quadrado médio de aproximadamente 90%, confirmando a assertividade do modelo de predição da variável de saída.

## 5 Resultados

Como resultado deste trabalho, o modelo teve um acerto de 96% dos valores preditos em comparação com os resultados reais do DI do processo da planta de coqueria estudada. Além disso, como comprovado nos métodos RMSE e MAPE, o erro é muito baixo, tanto na base de teste quanto na base de treinamento. Isto reforça que o modelo pode ser aplicado em dados reais, já indicando o valor predito do DI com a mistura de carvão que está no processo, podendo assim, antecipar a tomada de decisão.

Com base nisso, é possível buscar uma melhoria significativa no controle das variáveis predictoras, com a finalidade de aumentar a resistência do coque no interior dos Altos Fornos, tendo como consequência a redução de custos através da melhor resistência mecânica do combustível. Para melhor visualizar os resultados, buscou-se a frequência absoluta do DI obtido na base de dados utilizada pelo modelo, em seguida, a média para cada um dos parâmetros de entrada, o DI predito é apresentado na Tabela 17.

Tabela 17 – Precisão do Modelo com Valores Reais

DI	Freq	<0,105	<2,83	H2O	MV	S	CZ	TR	Amp	TF	Fluidez	IO	Peso	DI Predito
83,4	668	10,7	82,5	9,2	22,5	0,9	5,8	482	81,9	443	1,2	112	30170	84,32
83,5	1040	10,7	83,1	9,3	22,8	0,8	5,9	482	83,5	443	1,3	117	30344	84,39
83,6	655	10,5	83	9,4	22,7	0,8	5,9	483	82,9	443	1,5	113	30348	84,31
83,7	938	11,2	83,7	9,4	22,6	0,8	6	482	83,2	442	1,6	116	30172	84,31
83,8	1220	10,2	83,2	9	22,2	0,9	5,9	482	83,7	443	1,5	115	30165	84,41
83,9	1375	10,2	83,5	9,2	22,2	0,8	6	483	84,5	442	1,7	119	30239	84,31
84	8257	10,4	83,6	9	22,2	0,8	5,8	482	83,6	442	1,7	117	30070	84,36
84,1	6197	10,3	83,7	9	22,2	0,8	5,8	482	83,6	442	1,9	120	30131	84,33
84,2	6078	10,2	83,8	8,9	22,1	0,8	5,8	483	84,4	442	1,9	120	30027	84,29
84,3	5078	10,2	83,7	8,8	22,2	0,8	5,9	482	82,7	441	1,9	120	29993	84,36
84,4	3698	10,1	83,7	8,9	22	0,8	5,8	483	82,8	442	1,8	119	29974	84,29
84,5	2673	10,2	83,9	8,8	22,2	0,8	5,9	482	82,1	441	1,8	122	30024	84,37
84,6	1836	9,8	83,8	8,8	22,1	0,8	6	483	81,4	443	1,8	122	29986	84,35
84,7	808	10,1	83,8	8,9	22	0,8	6	483	82,4	443	1,7	123	30122	84,34
84,8	950	10,3	83,8	8,9	22,2	0,8	5,9	481	80,2	442	1,7	120	29918	84,31
84,9	593	10,7	83,8	9,3	21,8	0,8	5,6	482	79,6	443	1,3	121	30028	84,2
85	569	10,3	83,6	8,9	22,2	0,8	5,9	482	79,4	442	2	123	30209	84,41

Fonte: Elaboração Própria

Para uma base de dados com 44324 registros, a tabela acima utilizou 42633 para representar a análise da predição. É percebido que à medida que aumenta a frequência o desvio entre o DI obtido com o DI predito diminui, ao fazer uma média ponderada do erro temos um desvio absoluto de 0,14 confirmando sua assertividade.

## 6 Considerações Finais

O modelo encontrado com esse trabalho foi salvo externamente para poder ser utilizado em simulações, como por exemplo, apresentou-se na Figura 10 uma simulação onde as variáveis necessárias são preenchidas e o modelo é executado, resultando o DI previsto. As 10 primeiras variáveis estão associadas a amostragem da qualidade da mistura dos carvões, já as duas últimas são resultados do processo operacional.

```

modelo_carregado = joblib.load("/content/drive/My Drive/TCC/modelo_random_forest_v1.sav")

[ ] df_simulacao = pd.DataFrame(data = {
    '<0,105 mm': [12.023],          #Entre 10 e 13
    '<2,83 mm': [82.178],         #Entre 80 e 86
    'H2O': [8.182],               #Entre 7 e 11
    'Material_Volátil': [21.89],  #Entre 20 e 26
    'S': [0.73],                  #Entre 0.61 e 1.00
    'CZ': [7.23],                 #Entre 3 e 9
    'TR': [485],                  #Entre 470 e 490
    'Amplitude': [87],            #Entre 70 e 90
    'TF': [439],                  #Entre 430 e 450
    'Fluidez': [1.844],           #Entre 0.4 e 4.0
    'IO': [115.6627],             #Entre 70 e 120
    'Peso_Úmido_Enfornado': [31463] #Entre 25000 e 32000
})
df_simulacao

   <0,105 mm  <2,83 mm  H2O  Material_Volátil   S   CZ   TR  Amplitude  TF  Fluidez   IO  Peso_Úmido_Enfornado
0    12.023    82.178  8.182         21.89  0.73  7.23  485         87  439    1.844  115.6627         31463

[ ] resultado_simulacao = modelo_carregado.predict(df_simulacao)
resultado_simulacao

/usr/local/lib/python3.9/dist-packages/sklearn/base.py:432: UserWarning: X has feature names, but RandomForestRegressor
warnings.warn(
array([83.50243122])

```

Figura 10 – Simulação Utilizando o Modelo Salvo

Para trabalhos futuros, a medição da qualidade do coque engloba outros indicadores além do DI, que foi tratado no presente trabalho. Como sugestão, para futuras investigações poderia ser replicada a estrutura lógica utilizando outros indicadores de qualidade do coque que podem ser explorados para melhorar o desempenho do processo nos Altos Fornos ou até mesmo nas Coquerias.

# Referências

- ALVES, F. I. C. **Produção de coque metalúrgico a partir de carvão densificado por vibrocompactação em forno de soleira aquecida**. Ouro Preto: [s.n.], 2017. Dissertação (Mestrado em Engenharia de Materiais).
- AMARAL, F. **Introdução à ciência de dados: mineração de dados e big data**. [S.l.]: Alta Books Editora, 2016.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.
- CARDOSO, W.; FELICE, R. D. A novel committee machine to predict the quantity of impurities in hot metal produced in blast furnace. **Computers & Chemical Engineering**, Elsevier, v. 163, p. 107814, 2022.
- FERNANDES, D. C. et al. Adição de óleo para aumento da densidade de carga de carvões na coqueria. Universidade Federal de Minas Gerais, 2019.
- GEERDES, M.; CHAIGNEAU, R.; LINGIARDI, O. **Modern blast furnace ironmaking: an introduction (2020)**. [S.l.]: Ios Press, 2020.
- GIGLIO, A. M. **Modelo de Previsão de Qualidade de Coque Metalúrgico para Utilização em Altos-Fornos Utilizando Produção em Escala Piloto**. Tese (Doutorado) — PUC-Rio, 2016.
- HADAVANDI, E.; CHELGANI, S. C. Estimation of coking indexes based on parental coal properties by variable importance measurement and boosted-support vector regression method. **Measurement**, Elsevier, v. 135, p. 306–311, 2019.
- LI, S.; CHANG, J.; CHU, M.; LI, J.; YANG, A. A blast furnace coke ratio prediction model based on fuzzy cluster and grid search optimized support vector regression. **Applied Intelligence**, Springer, v. 52, n. 12, p. 13533–13542, 2022.