

**UNIVERSIDADE FEDERAL DO PAMPA  
CURSO DE ENGENHARIA QUÍMICA**

**ÁGATHA GLOGER TURMINA**

**APRENDIZADO DE MÁQUINA APLICADO À ANÁLISE DA OPERAÇÃO DE UMA  
PLANTA PILOTO PARA GASEIFICAÇÃO DE CARVÃO MINERAL**

**Bagé  
2021**

**ÁGATHA GLOGER TURMINA**

**APRENDIZADO DE MÁQUINA APLICADO À ANÁLISE DA OPERAÇÃO DE  
UMA PLANTA PILOTO PARA GASEIFICAÇÃO DE CARVÃO MINERAL**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Química da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Bacharel em Engenharia Química.

Orientador: Prof. Dr. Alexandre Denes Arruda

Coorientadores: Dra. Andressa Apio e Prof. Dr. Rodolfo Rodrigues

**Bagé  
2021**

Ficha catalográfica elaborada automaticamente com os dados fornecidos  
pelo(a) autor(a) através do Módulo de Biblioteca do  
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

T941a Turmina, Ágatha Gloger

Aprendizado de máquina aplicado à análise da operação de  
uma planta piloto para gaseificação de carvão mineral / Ágatha  
Gloger Turmina.

81 p.

Trabalho de Conclusão de Curso(Graduação)-- Universidade  
Federal do Pampa, ENGENHARIA QUÍMICA, 2021.

"Orientação: Alexandre Denes Arruda".

1. Gaseificação. 2. Carvão mineral. 3. Aprendizado de  
Máquina. 4. Análise de Correlação. 5. Variáveis do processo.  
I. Título.



SERVIÇO PÚBLICO FEDERAL  
MINISTÉRIO DA EDUCAÇÃO  
Universidade Federal do Pampa

ÁGATHA GLOGER TURMINA

APRENDIZADO DE MÁQUINA APLICADO À ANÁLISE DA OPERAÇÃO DE UMA PLANTA  
PILOTO PARA GASEIFICAÇÃO DE CARVÃO MINERAL

Trabalho de Conclusão de Curso apresentado ao  
Curso de Engenharia Química da Universidade  
Federal do Pampa, como requisito parcial para  
obtenção do Título de Bacharel em Engenharia  
Química.

Trabalho de Conclusão de Curso defendido e aprovado em: 8 de maio de 2021.

Banca examinadora:

---

Prof. Dr. Alexandre Denes Arruda  
Orientador  
Unipampa

---

Prof. Dr. Rodolfo Rodrigues  
Coorientador  
UFSM

---

Profa. Dra. Andressa Apio  
Coorientadora  
Latos

---

Prof. Dr. Paulo Fernando Marques Duarte Filho  
Unipampa

---

Prof. Dr. João Carlos Furtado  
Unisc



Assinado eletronicamente por **ALEXANDRE DENES ARRUDA, PROFESSOR DO MAGISTERIO SUPERIOR**, em 08/05/2021, às 11:12, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **Rodolfo Rodrigues, Usuário Externo**, em 08/05/2021, às 11:14, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **João Carlos Furtado, Usuário Externo**, em 10/05/2021, às 14:22, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **Andressa Apio, Usuário Externo**, em 10/05/2021, às 17:17, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **PAULO FERNANDO MARQUES DUARTE FILHO, PROFESSOR DO MAGISTERIO SUPERIOR**, em 10/05/2021, às 20:35, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



A autenticidade deste documento pode ser conferida no site  
[https://sei.unipampa.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0) informando o código verificador **0518247** e o código CRC **653363AE**.

## **AGRADECIMENTO**

Agradeço a toda minha família pelo amor, acolhimento e união em todos os momentos. Mas principalmente, agradeço aos meus pais Mara Gloger Turmina e Gilson Turmina, por terem me apoiado em todos os meus sonhos, por sempre terem feito tudo que estava ao seu alcance para me proporcionar boas oportunidades e por serem meus exemplos de pessoas íntegras e batalhadoras.

Agradeço a minha irmã Amanda Turmina, por sempre acreditar em mim e me motivar a seguir em frente.

Um agradecimento muito especial ao meu amor, Eduardo Pereira, que esteve comigo em muitos momentos durante a graduação, fez muitos cafés e secou algumas lágrimas, sempre acreditando mais em mim do que eu mesma.

Ao Professor Rodolfo Rodrigues, obrigada por estar presente e contribuir muito desde o início da realização desse trabalho.

A Dra. Andressa Apio, que foi essencial no desenvolvimento computacional desse trabalho. Agradeço por todo o auxílio e por ter me encorajado desde o início a aprender sobre a área de ciência de dados.

Ao Professor Alexandre Arruda, agradeço por ter aceitado assumir a orientação desse trabalho em um período complicado, e por sempre estar disposto a me ajudar em todos os momentos.

Agradeço à Professora Denise Moser, pelas horas em videochamada, revisando e corrigindo a escrita deste trabalho.

Agradeço aos meus amigos, tanto os de longa data quanto os que conheci na UNIPAMPA, pela amizade, e tantos momentos compartilhados que foram essenciais para tornar essa caminhada mais alegre.

*“Sem dados você é apenas uma pessoa  
qualquer com uma opinião”.*

*William Edwards Deming*

## RESUMO

A gaseificação consiste na reação entre o carvão mineral e um agente gaseificante para a geração de um gás energético, chamado de gás de síntese (*syngas*), no qual o reator onde ocorre esse processo é chamado de gaseificador. O processo de gaseificação pode ser estudado com base nas suas condições de operação, sendo as condições de entrada: vazões de ar nas diferentes regiões do reator, vazão de carvão, operação da resistência elétrica, pulsos manuais de pressão e frequência da centrífuga. As condições de saída obtidas no processo são as temperaturas e pressões, obtidas por sensores localizados ao longo do leito. Nesse estudo de caso, são aplicadas técnicas de aprendizado de máquina (*machine learning*), que criam algoritmos capazes de fazer previsões através de modelos baseados em dados, e vêm se destacando para uso em análise de processos industriais. O objetivo deste trabalho é avaliar a aplicação de diferentes algoritmos de Aprendizado de Máquina para a previsão da operação de um gaseificador de leito fluidizado borbulhante instalado no laboratório de Energia e Carboquímica da UNIPAMPA, campus Bagé, utilizando carvão mineral de Candiota-RS. Para construir uma base de dados confiável foram selecionadas 15 corridas experimentais dos registros de experimentos realizados no laboratório, desde o ano de 2016 até o ano de 2020. A coleta dos dados de saída foi obtida por meio de um computador presente na planta piloto, enquanto os dados de entrada foram coletados a partir de um caderno de registros do Laboratório. Para a análise computacional, utilizou-se o ambiente de desenvolvimento *Jupyter Notebook* e a linguagem de programação *Python*. Para implementar o trabalho proposto, foram utilizadas as bibliotecas *Pandas*, para a manipulação e análise de dados; *Numpy*, para a manipulação de estruturas e dados numéricos; *Matplotlib*, para a visualização dos dados; e *Scikit-learn* e *FireTS* para a aplicação de algoritmos de aprendizado de máquina. Utilizando essas ferramentas, foram realizados o pré-processamento dos dados, a análise de correlação entre as variáveis de operação e a aplicação e validação dos seguintes métodos de aprendizado de máquina: árvore de decisão, florestas aleatórias e redes neurais artificiais utilizando rede autorregressiva não-linear com entradas exógenas (NARX). A análise de correlação demonstrou que as variáveis do processo são extremamente correlacionadas entre si, e que a maioria das relações pode ser explicada por motivos físicos e químicos do processo. O Aprendizado de máquina se mostrou parcialmente satisfatório para os

modelos de árvore de decisão e florestas aleatórias, já que foram obtidos valores para  $R^2$  de 0.8263 e 0.8262, respectivamente. No entanto as análises gráficas demonstram que as previsões não condizem com a dinâmica do processo real. Para o modelo NARX, o resultado foi satisfatório, chegando a valores de  $R^2$  de 0.9955 para a previsão das temperaturas e 0.8789 para a previsão das pressões. Pelo fato de o processo de gaseificação ser extremamente dinâmico, as redes dinâmicas do modelo NARX apresentam melhores resultados, por possuírem memória e serem treinadas para aprender a partir de padrões de variação no tempo.

**Palavras-chave:** Gaseificação. Carvão mineral. Aprendizado de Máquina. Análise de Correlação. Variáveis do processo.

## ABSTRACT

The gasification process consists in the reaction between mineral coal and one gasifying agent to generate an energetic gas, called synthesis gas (syngas). The reactor where this process occurs is called gasifier. The gasification process can be studied based in operational conditions. The input conditions are: air flow rate in the different regions of the reactor, coal flow, heating element operation, manual pressure pulses and centrifuge frequency. The output conditions in the process are the temperatures and the pressures obtained through sensors, which are located in the course of the reactor bed. In this case study machine learning techniques are applied to create algorithms capable of making predictions through models based on data. This field has been standing out because of its potential in analysis of industrial processes. The aim of this work is to evaluate the application of different machine learning algorithms to predict the operation of a fluidized bed gasifier with bubbling air, which is installed in the Laboratório de Energia e Carboquímica of UNIPAMPA, Campus Bagé, using mineral coal from Candiota – RS. In order to develop a solid database 15 experimental runings were selected from the experiments records. The data since 2016 until 2020 were collected. The acquisition of the output data was executed through a computer placed in the pilot plant. The input data were obtained from a logbook with laboratory record. To the computational analysis, the development environment chosen was Jupyter Notebook, where Python programming language was employed. To implement the current work Python libraries were applied, Pandas for manipulation and analysis of the data, Numpy for manipulation of structures and numerical data and Matplotlib for data visualization. The libraries Scikit-learn and FireTS were used to employ machine learning algorithms. Using these tools the data was preprocessed, the correlation between the variables of operation was analysed and the following algorithms were applied: decision tree, random forests and non-linear autoregressive neural network with exogenous inputs (NARX). Most of the variables in the process showed high correlations between each other, it can be explained by physical and chemical process characteristics. The machine learning performed partially satisfactory to the decision trees and random forests models, since were obtained  $R^2$  values of 0.8263 and 0.8262 respectively. However the graphical analysis showed that the predictions doesn't match the dynamics of the real process. To the NARX model, the result was satisfactory, getting to  $R^2$  values of 0.9955 to temperature

predictions, and 0.8789 to pressure predictions. The gasification process is extremely dynamic, for this reason the neural networks of the NARX model present better results, since it carries memory and is trained to learn from the patterns of time variation.

**Keywords:** Gasification. Mineral Coal. Machine Learning. Correlation Analysis. Process Variables.

## LISTA DE FIGURAS

Figura 1 – Gaseificação em ciclo combinado.....	19
Figura 2 – Gaseificadores de Leito Fluidizado.....	21
Figura 3 – Tipos de Aprendizado de Máquina.....	26
Figura 4 – Classificação de uma manhã ideal para jogar tênis.....	27
Figura 5 – Vista frontal da planta piloto de gaseificação do LEC.....	32
Figura 6 – Vista lateral da planta piloto de gaseificação do LEC.....	32
Figura 7 – Esboço do gaseificador do LEC.....	33
Figura 8 – Distribuição dos sensores de temperatura e pressão na planta piloto de gaseificação do LEC.....	35
Figura 9 – Dados de Entrada para o experimento de gaseificação realizado na data 01/08/2016.....	40
Figura 10 – Tabela de temperaturas para o experimento realizado na data 01/08/2016.....	40
Figura 11 – Tabela de pressões e temperaturas para o experimento realizado na data 01/08/2016.....	41
Figura 12 – Dinâmica da temperatura (°C) para a corrida N° 1.....	42
Figura 13 – Dinâmica da pressão (mmH <sub>2</sub> O) para a corrida N° 1.....	42
Figura 14 – Dinâmica da temperatura (°C) para a corrida N° 6.....	43
Figura 15 – Dinâmica da pressão (mmH <sub>2</sub> O) para a corrida N° 6.....	43
Figura 16 – Dinâmica da temperatura (°C) para a corrida N° 8.....	44
Figura 17 – Dinâmica da pressão (mmH <sub>2</sub> O) para a corrida N° 8.....	44
Figura 18 – Mapa de calor de correlação entre dados de entrada e saída.....	45
Figura 19 – Dispersão entre dados reais e estimados por árvore de decisão.....	48
Figura 20 – Dispersão entre dados reais e estimados por florestas aleatórias.....	49
Figura 21 – Predição de 100 valores de y por árvore de decisão para T2.....	49
Figura 22 – Predição de 100 valores de y por florestas aleatórias para T2.....	50
Figura 23 – Comportamento previsto por árvore de decisão, para T2 da corrida experimental N°14.....	50
Figura 24 – Comportamento previsto por florestas aleatórias, para T2 da corrida experimental N°14.....	51

Figura 25 – Comportamento previsto por árvore de decisão, para P1 da corrida experimental N°15.....	51
Figura 26 – Comportamento previsto por florestas aleatórias, para P1 da corrida experimental N°15.....	52
Figura 27 – Dispersão entre dados reais e estimados por árvore de decisão.....	53
Figura 28 – Dispersão entre dados reais e estimados por florestas aleatórias.....	53
Figura 29 – Comportamento previsto por árvore de decisão, para T7 da corrida experimental N°15.....	53
Figura 30 – Comportamento previsto por florestas aleatórias, para T7 da corrida experimental N°15.....	56
Figura 31 – Dados reais e estimados por NARX para T5, utilizando a frequência de alimentação e a vazão de ar <i>plenum</i> para treinar o algoritmo.....	57
Figura 32 – Valores de temperatura reais e estimados por NARX para o sensor T5 da corrida experimental N°1.....	59
Figura 33 – Dados reais e estimados por NARX para T5, utilizando a operação da resistência e a vazão de ar <i>plenum</i> para treinar o algoritmo.....	58
Figura 34 – Valores de temperatura reais e estimados por NARX para o sensor T5 da corrida experimental N°14.....	58
Figura 35 – Dados reais e estimados por NARX para T5, utilizando a operação da resistência e a vazão de ar <i>plenum</i> para treinar o algoritmo.....	59
Figura 36 – Dados reais e estimados por NARX para P1, utilizando a frequência de alimentação e a vazão de ar <i>plenum</i> para treinar o algoritmo.....	60
Figura 37 – Dados reais e estimados por NARX para P1, utilizando a operação da resistência e a vazão de ar <i>plenum</i> para treinar o algoritmo, com n=1.....	61
Figura 38 – Dados reais e estimados por NARX para P1, utilizando a operação da resistência e a vazão de ar <i>plenum</i> para treinar o algoritmo, com n=10.....	62
Figura 39 – Dados reais e estimados por NARX para P1, utilizando a operação da resistência e a vazão de ar <i>plenum</i> para treinar o algoritmo, com n=50.....	63

## LISTA DE TABELAS

Tabela 1 – Análise imediata do carvão mineral da jazida de Candiota.....	18
Tabela 2 – Dimensões das diferentes regiões do gaseificador.....	34
Tabela 3 – Distribuição de sensores no gaseificador.....	35

## **LISTA DE ABREVIATURAS E SIGLAS**

AM – Aprendizado de Máquina

CGT Eletrosul – Companhia de Geração e Transmissão de Energia Elétrica do Sul do Brasil

CRM – Companhia Riograndense de Mineração

HCA – Hierarchical clustering

IGCC – *Integrated Gasification Combined Cycle*

LEC – Laboratório de Energia e Carboquímica

LLE – Locally Linear Embedding

ML – Machine Learning

MSE – Mean squared error

NaN – Not-A-Number

NARX – Nonlinear autoregressive exogenous model

PCA – Principal component analysis

$R^2$  – Coeficiente de determinação

RMSE – Root mean square error

RNA – Redes Neurais Artificiais

TCC – Trabalho de Conclusão de Curso

t-SNE – t-Distributed Stochastic Neighbor Embedding

UNIPAMPA – Universidade Federal do Pampa

UTE – Usina Termelétrica

## LISTA DE SÍMBOLOS

$\Delta H$  – Entalpia de reação, kJ/mol

$n$  – Tamanho da amostra

$x$  – Dados de entrada

$y$  – Dados de saída

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>14</b>
<b>2</b>	<b>OBJETIVOS.....</b>	<b>16</b>
<b>2.1</b>	<b>Objetivo Geral.....</b>	<b>16</b>
<b>2.2</b>	<b>Objetivo Específico.....</b>	<b>16</b>
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA.....</b>	<b>17</b>
<b>3.1</b>	<b>Carvão Mineral.....</b>	<b>17</b>
<b>3.2</b>	<b>Gaseificação.....</b>	<b>18</b>
3.2.1	Gaseificadores de Leito Fluidizado.....	20
3.2.2	Processo e Reações Químicas.....	21
<b>3.3</b>	<b>Aprendizado de Máquina.....</b>	<b>23</b>
3.3.1	Base de Dados e Escolha do Algoritmo.....	24
<b>3.4</b>	<b>Classificação de Métodos de Aprendizado.....</b>	<b>25</b>
3.4.1	Aprendizado Supervisionado.....	25
3.4.2	Aprendizado Não-Supervisionado.....	26
<b>3.5</b>	<b>Modelos de Aprendizado de Máquina.....</b>	<b>26</b>
3.5.1	Árvore de Decisão.....	27
3.5.2	Florestas Aleatórias.....	28
3.5.4	Redes Neurais Artificiais.....	28
<b>3.6</b>	<b>Histórico de Trabalhos.....</b>	<b>30</b>
<b>4</b>	<b>METODOLOGIA.....</b>	<b>31</b>
<b>4.1</b>	<b>Detalhes da Planta Piloto de Gaseificação.....</b>	<b>31</b>
4.1.1	Perfil de Pressão e Temperatura.....	34
<b>4.2</b>	<b>Coleta de Dados Experimentais.....</b>	<b>36</b>
<b>4.3</b>	<b>Análise dos Dados, Aplicação do Aprendizado de Máquina e Avaliação do Desempenho.....</b>	<b>37</b>
<b>5</b>	<b>RESULTADOS.....</b>	<b>39</b>
<b>5.1</b>	<b>Análise de Correlação.....</b>	<b>45</b>
<b>5.2</b>	<b>Aprendizado de Máquina.....</b>	<b>47</b>
5.2.1	Árvore de Decisão e Florestas Aleatórias.....	47
5.2.2	Rede Autorregressiva Não-Linear com Entradas Exógenas.....	55
<b>6</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>64</b>

<b>SUGESTÕES PARA TRABALHOS FUTUROS.....</b>	<b>66</b>
<b>REFERÊNCIAS.....</b>	<b>67</b>
<b>APÊNDICES.....</b>	<b>74</b>

## 1 INTRODUÇÃO

O Brasil conta com reservas de carvão mineral que totalizam 32 bilhões de toneladas disponíveis, no qual a jazida de Candiota (RS) é uma das maiores reservas carboníferas, detendo 38% do total em território brasileiro (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA, 2008). Neste contexto, o carvão mineral representa hoje uma das alternativas viáveis para a diversificação da matriz energética do Brasil.

A energia elétrica obtida a partir do carvão é gerada principalmente por meio da combustão direta. Esse método consiste na queima do carvão em ambiente com oxigênio abundante obtendo gases a altas temperaturas para aquecer água e produzir vapor a alta pressão que é então utilizado para mover turbinas e geradores. No entanto, esse processo favorece a formação de compostos poluentes, não sendo considerado um método ambientalmente amigável sem o tratamento adequado de suas emissões.

Uma boa alternativa à combustão é a gaseificação, que é um processo que visa gerar gás de síntese (*syngas*) a partir da reação do carvão com agentes gaseificantes (ar, oxigênio, vapor de água e etc.) a menores temperaturas, emitindo níveis baixos de poluentes e proporcionando maior eficiência energética (COLLOT, 2006). A gaseificação é uma tecnologia mais adequada ao carvão brasileiro, por garantir uma maior eficiência de conversão devido à grande quantidade de material inerte (cinzas e resíduos) na composição do carvão.

O gás de síntese é fundamentalmente uma mistura dos gases: hidrogênio, monóxido de carbono, metano e gás carbônico. Além de gás combustível, este gás pode ser utilizado na produção de combustíveis líquidos por meio da síntese de Fischer-Tropsch e pode ser uma matéria-prima na indústria química para a produção de amônia, metanol e gás natural sintético (HIGMAN; BURGT, 2003).

O desempenho do processo de gaseificação é dependente de uma série de fatores, tais como as características da matéria-prima utilizada, tipo de gaseificador e as variáveis de operação durante as reações químicas. Com base na importância dos parâmetros de entrada para o processo de gaseificação do carvão, faz-se necessário determinar a correlação dessas variáveis através da aplicação de uma ferramenta preditiva adequada.

Como o aprendizado de máquina (*machine learning*) é uma técnica capaz de fazer previsões através de modelos baseados em dados, sua aplicação se adequa

bem ao processo de gaseificação de carvão pela disponibilidade de dados experimentais. Por existirem diversos modelos diferentes de aprendizado de máquina, os estudados neste trabalho foram escolhidos com base nos modelos considerados mais importantes e que costumam gerar melhores previsões, sendo eles: árvore de decisão, florestas aleatórias e redes neurais artificiais.

Portanto, de forma mais específica, o presente trabalho é um estudo sobre o processo de gaseificação de carvão mineral extraído da jazida de Candiota- RS, em gaseificador de leito fluidizado borbulhante presente em uma planta piloto no Laboratório de Energia e Carboquímica (LEC) da Universidade Federal do Pampa (UNIPAMPA), campus Bagé. Para o estudo do processo, foi realizada uma análise de correlação entre as variáveis de operação e aplicação de diferentes métodos de aprendizado de máquina para a previsão de dados de saída (temperatura e pressão), objetivando aumentar o entendimento sobre a operação do processo.

## 2 OBJETIVOS

### 2.1 Objetivo Geral

Avaliar a aplicação de diferentes modelos de Aprendizado de Máquina para a predição de dados de saída de um processo de gaseificação do carvão mineral de Candiota-RS em leito fluidizado borbulhante, utilizando os dados experimentais obtidos de um gaseificador em escala piloto presente no Laboratório de Energia e Carboquímica (LEC) da Universidade Federal de Pampa (UNIPAMPA), situado no campus Bagé.

### 2.2 Objetivos Específicos

- Realizar a coleta de dados de corridas experimentais da planta piloto de gaseificação de carvão mineral no LEC.
- Selecionar as corridas experimentais que serão utilizados para o estudo, a partir da análise dos dados experimentais coletados.
- Realizar o pré-processamento da base de dados selecionados a partir da linguagem *Python*.
- Analisar a correlação entre as variáveis de entrada e saída do processo.
- Aplicar os métodos de Árvore de Decisão, Florestas Aleatórias e Redes Neurais Artificiais, para o treinamento de diferentes modelos, a fim de verificar qual fornece a melhor predição.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Carvão mineral

Carvão mineral é uma rocha combustível, de origem fóssil, que se encontra em jazidas localizadas no subsolo e pode ser extraído por mineração. Tem em sua composição umidade, carbono, oxigênio, hidrogênio, enxofre, nitrogênio e matéria mineral (LUNKES, 2010).

O carvão pode ser subdividido em quatro classes de acordo com suas características, como composição (teor de carbono fixo, teor de matéria volátil, teor de cinzas), poder calorífico e massa específica. São as classes: antracito, betuminoso, sub-betuminoso e linhito (HIGMAN; BURGT, 2008).

As reservas brasileiras são compostas pelo carvão dos tipos linhito e sub-betuminoso. A Associação Brasileira do Carvão Mineral (ABCM) calcula que as reservas conhecidas poderiam gerar hoje 17 mil megawatts. Do volume de reservas, o Rio Grande do Sul responde por 89,25%; Santa Catarina, 10,41%; Paraná, 0,32% e São Paulo, 0,02%. Somente a jazida de Candiota (RS) possui 38% de todo o carvão nacional. O minério é pobre do ponto de vista energético e não admite beneficiamento nem transporte, em função do elevado teor de impurezas (cinzas). Isto faz com que sua utilização seja feita sem beneficiamento e na boca da mina (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA, 2008).

Apesar de grande parte do carvão brasileiro ser classificado como sub-betuminoso, o que indica baixa qualidade, há a viabilidade para implantação de usinas termelétricas (UTES) de carvão mineral no Brasil. A aplicação de novas tecnologias para processo de queima limpa coopera para uma redução expressiva das emissões de SO<sub>x</sub>, NO<sub>x</sub> e material particulado (SILVA, 2016).

Na Tabela 1, pode-se observar os resultados obtidos por diversos autores ao realizarem a análise imediata do carvão mineral de Candiota. Com base nessa tabela, é também possível constatar o alto índice de cinzas no carvão brasileiro e outras características como a umidade, matéria volátil e carbono fixo em sua composição.

Tabela 1 – Análise imediata do carvão mineral da jazida de Candiota

Referência	Umidade (%)	Cinzas (%)	Matéria Volátil (%)	Carbono fixo (%)
KALKREUTH, <i>et al.</i> (2006)	11,0	46,7	18,7	23,6
DEPOI (2008)	6,2	47,9	29,3	16,6
HOELSCHER (2014)	5,65	45,75	17,14	31,46
PIRES E QUEROL (2004)	10	49,7	13,1	27,2
BROWN (2011)	5,3	49	20,4	25,3

Fonte: Brown (2011); Depoi; Pozebon; Kalkreuth (2008); Hoelscher (2014); Pires; Querol (2004); Sá (2018, p. 21)

A utilização dessa matéria-prima é primordial para a região de Candiota-RS, visto que gera diversos empregos na Companhia Riograndense de Mineração (CRM), UTE Candiota III (Eletrobrás-CGT Eletrosul) e UTE Pampa Sul (Engie). Ainda há especulações sobre a expectativa de instalação de novos empreendimentos que utilizarão a tecnologia de gaseificação do carvão mineral da região, como a planta de metanol da Vamtec S.A. e a planta de amônia do grupo TransGas (KLEIN, 2015; NIECKEL; FURTADO, 2014).

### 3.2 Gaseificação

A gaseificação é a conversão de qualquer combustível líquido ou sólido em gás energético por meio da oxidação parcial em temperatura elevada (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA, 2008).

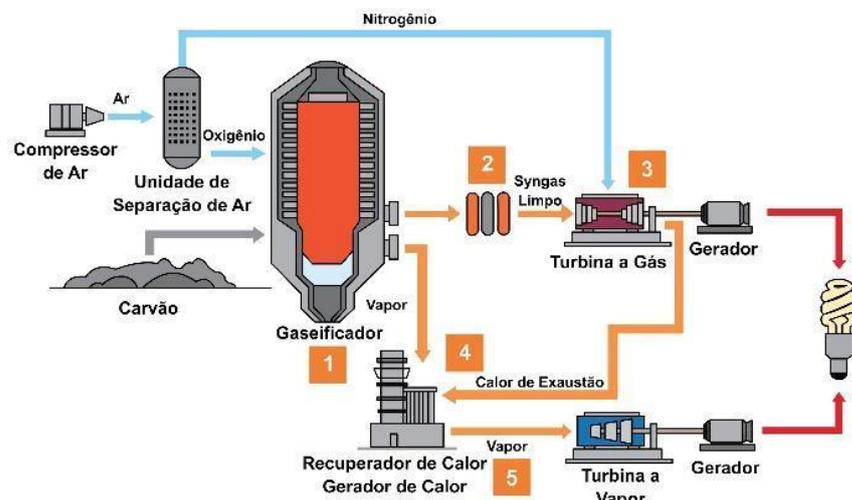
Segundo Higman e Burgt (2003), o gás de síntese (*syngas*) é uma mistura gasosa formada principalmente por H<sub>2</sub>, CO, CH<sub>4</sub> e CO<sub>2</sub>, que pode ser obtido a partir do processo de gaseificação de matérias-primas como carvão mineral, biomassa, coque de petróleo ou até mesmo uma mistura destes. Essa mistura gasosa obtida na gaseificação pode ser utilizada para geração de energia, através de turbinas a gás,

ou para obtenção de produtos químicos como metanol, anidrido acético e hidrocarbonetos de peso molecular variado por meio da síntese de Fischer-Tropsch.

O desempenho do processo de gaseificação é dependente de uma série de fatores, tais como as características da matéria-prima utilizada, tipo de gaseificador empregado e diversas variáveis de processo, como razão de equivalência (razão entre ar e ar estequiométrico), pressão, tempo de residência e temperatura (ROCHA, 2016).

O fluxograma apresentado na Figura 1 se refere ao processo de geração de energia elétrica por gaseificação em ciclo combinado (IGCC, do Inglês “*Integrated Gasification Combined Cycle*”). O processo inicia com a gaseificação do combustível (carvão, biomassa, etc.), obtendo como produto o gás de síntese. Esse gás passa por uma limpeza onde são retidos materiais particulados, derivados de enxofre e impurezas em geral, e o vapor gerado nesse processo é utilizado para movimentar turbinas a vapor e gerar mais energia elétrica. O processo combinado contribui com a melhoria da eficiência de geração de energia utilizando carvão e também reduz as emissões de gases de efeito estufa e material particulado na atmosfera (COLLOT, 2006; SILVA, 2016).

Figura 1 – Gaseificação em ciclo combinado



Fonte: Magnus *et al.* (2017, p. 4 *apud* Duke Energy)

Rodrigues (2015) expõe que é usual classificar os gaseificadores em três classes diferentes, que são determinadas de acordo com a conformação do seu leito reacional, sendo eles: leito fixo, leito fluidizado ou leito arrastado. Essas configurações ainda podem apresentar variações, dependendo do fluxo ou das interações entre as

fases da carga a ser gaseificada e os produtos gasosos produzidos, podendo ser classificados em: leito fixo contracorrente, leito fixo concorrente, leito fluidizado borbulhante, leito fluidizado circulante, leito de arraste com carga superior e leito de arraste com carga lateral.

O gaseificador presente no LEC, que foi utilizado para obtenção dos dados experimentais, é identificado como gaseificador de leito fluidizado borbulhante. Por esse motivo, apenas gaseificadores fluidizados serão discutidos neste trabalho.

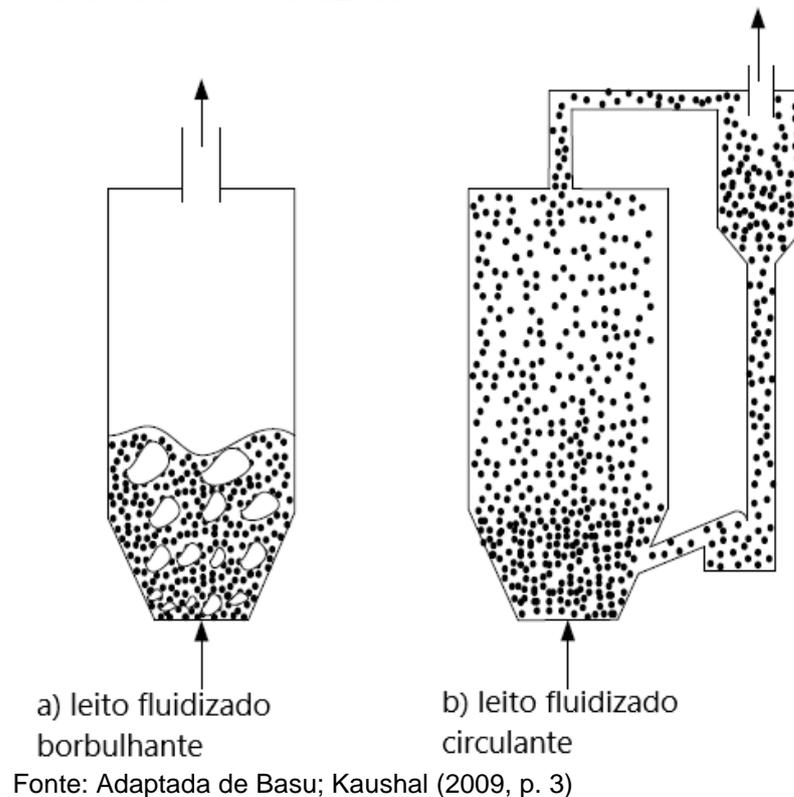
### 3.2.1 Gaseificadores de Leito Fluidizado

Nos gaseificadores de leito fluidizado, sólidos inertes granulados (areia, cinzas, óxidos e etc.) juntamente com o combustível sólido da alimentação são fluidizados por meio de um agente gaseificante, que pode ser ar, oxigênio, vapor de água ou dióxido de carbono. O agente gaseificante, geralmente pré-aquecido por uma fonte de calor externa, é alimentado pela parte inferior do reator, seguindo em sentido ascendente. A temperatura é mantida ao longo do processo por meio das reações químicas exotérmicas. O processo fluidizado promove homogeneidade e um bom contato entre combustível e gás, o que coopera para elevadas taxas de transferência de calor e massa que, além de aumentar a eficiência do processo, auxilia a diminuir as emissões de particulados e poluentes. Esse processo também faz com que a temperatura seja mais baixa e uniforme, fazendo com que as etapas sejam indistinguíveis no interior do gaseificador, diferente do que acontece, por exemplo, em um gaseificador de leito fixo. Os gaseificadores de leito fluidizado são constituídos por duas regiões principais: densa (*plenum*) e livre (*freeboard*). A região densa é formada por um leito de partículas e é onde ocorrem as reações sólido-gás. A região logo acima da região densa, constituída por uma pequena fração de partículas leves em suspensão por gases quentes, é chamada de região livre. Esta é caracterizada por reações gás-gás. (ANDRADE, 2007; BASU; KAUSHAL, 2009; GRÄBNER, 2015; RODRIGUES, 2015).

Os gaseificadores de leito fluidizado se dividem em circulante e borbulhante. Em um leito fluidizado circulante (Figura 2-b), sólidos circulam em torno de um circuito fechado que é caracterizado por uma mistura intensa e um longo tempo de residência dos sólidos particulados. A ausência de quaisquer bolhas evita o problema da ausência de gases em contato com partículas sólidas, que é observado na

configuração de gaseificador de leito fluidizado borbulhante (Figura 2-a) (BASU, 2013).

Figura 2 – Gaseificadores de leito fluidizado



Enquanto nos gaseificadores de leito fluidizado borbulhante as partículas de *char* (carbono sólido não convertido) são coletadas juntamente às cinzas leves, no tipo circulante estas são retornadas ao leito, o que garante maior conversão de carbono devido a um maior tempo de residência, embora sejam necessárias maiores vazões de oxidante. Em ambos tipos, as cinzas pesadas são coletadas no fundo do reator (ANDRADE, 2007; DIAS, 2016; HOFFMANN, 2010; LORA et al., 2012).

### 3.2.2 Processo e Reações Químicas

O processo de gaseificação ocorre em duas etapas: pirólise e gaseificação. A pirólise, também chamada de devolatização, é a primeira etapa do processo, que se dá por meio de reações de decomposição térmica até a formação do *char* (sólido residual constituído por materiais orgânicos e inorgânicos). A transformação das estruturas orgânicas e inorgânicas do carvão acontece a temperaturas superiores a 350°C, quebrando ligações químicas mais fracas e liberando voláteis. Os parâmetros

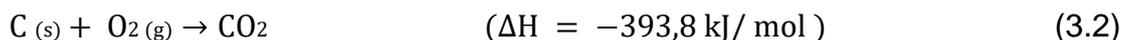
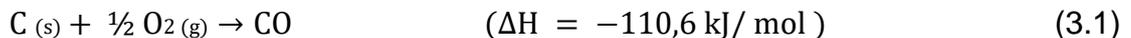
que influenciam a cinética e os mecanismos da pirólise de carvão são: tipo de carvão, tamanho de partícula, taxa de aquecimento, temperatura, tempo de residência, tipo de atmosfera gasosa e pressão. Dependendo da origem da matéria-prima e das condições do processo, os produtos voláteis podem incluir H<sub>2</sub>O, H<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, CO<sub>2</sub>, CO, CH<sub>4</sub>, H<sub>2</sub>S, NH<sub>3</sub>, C<sub>2</sub>H<sub>6</sub> e níveis muito baixos de hidrocarbonetos insaturados, tais como olefinas, acetilenos e compostos aromáticos e alcatrões (COLLAZZO, 2013; JÜNTGEN, 1981; REZAIYAN; CHEREMISINOFF, 2005).

A gaseificação é a segunda etapa do processo, onde acontece a reação entre o *char* resultante da pirólise, os produtos voláteis e o agente gaseificante, formando assim o gás de síntese. As reações químicas decorrentes da gaseificação são complexas e podem variar bastante, dependendo das condições de pressão e temperatura da gaseificação e da matéria-prima utilizada.

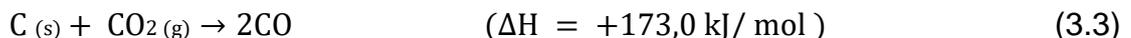
Durante a etapa da gaseificação, ocorrem reações heterogêneas (sólido-gás) e homogêneas (gás-gás), que estão descritas a seguir (PELLEGRINO, 2006).

### Reações Heterogêneas Sólido-Gás

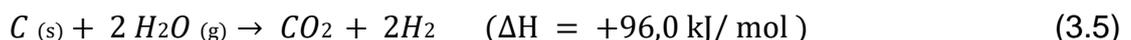
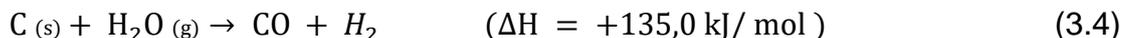
- oxidação do carbono:



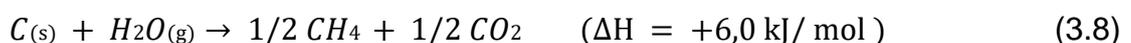
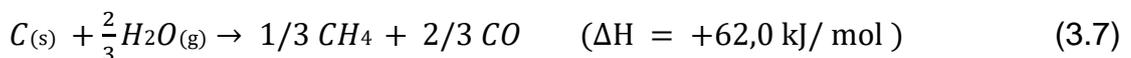
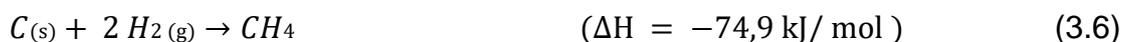
- reação de Boudouard:



- reação carbono-água:

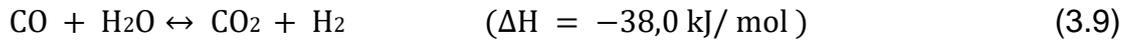


- formação do metano:



### Reações homogêneas Gás-Gás

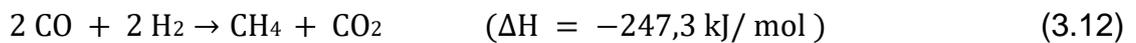
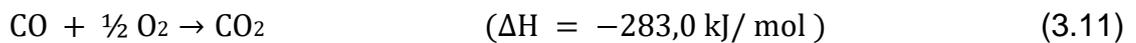
- reação de deslocamento gás-água



- reação de reforma a vapor do metano:



- monóxido de carbono-oxigênio



Ainda segundo Pellegrino (2006), as reações exotérmicas 3.1, 3.2 e 3.6 fornecem o calor necessário para que ocorram as reações endotérmicas 3.3, 3.4, 3.5, 3.7, 3.8 e 3.10. A reação de *Boudouard* (3.3) é a mais importante na faixa de temperatura de 800-900°C e ocorre em todo o leito. As reações de formação de dióxido e monóxido de carbono (equações 3.1 e 3.2) só ocorrem nas regiões mais próximas a entrada do ar no gaseificador, por conta da presença de oxigênio.

### 3.3 Aprendizado de Máquina

*Machine Learning* (ML) ou Aprendizado de Máquina (AM) é um ramo da Inteligência Artificial que fornece um conjunto de ferramentas computacionais para transformar dados em conhecimento prático. Segundo Géron (2017), o AM é simplesmente a ciência de programar computadores de forma que eles aprendam através de dados. De forma mais geral, é possível definir o aprendizado de máquina como o campo de estudo que fornece aos computadores a habilidade de aprender sem serem explicitamente programados. O AM, por ser inerentemente um campo multidisciplinar, pode se basear em resultados advindos de estudos da inteligência artificial, probabilidade, estatística, teoria da complexidade computacional, teoria de controle, teoria da informação, filosofia, psicologia, neurologia e entre outros (MITCHELL, 1997).

O reconhecimento de padrões importantes nos dados permite a resolução de problemas, utilizando a predição. Dessa forma, o termo aprendizado de máquina faz referência ao reconhecimento desses padrões, obtidos pela experiência e transformados em conhecimento por meio de diferentes métodos de aprendizado. De acordo com Géron (2017), para se garantir uma boa predição, é muito importante que os dados a serem estudados tenham uma boa qualidade e sejam adequados para a proposta. Também é determinante a escolha de um algoritmo que se adapte bem aos dados e às tendências dentro deles.

### 3.3.1 Base de Dados e Escolha do Algoritmo

A qualidade dos dados escolhidos para realizar o aprendizado é muito importante para se obter uma boa precisão da previsão. É crucial levar em consideração que a base de dados tenha um tamanho suficiente para um treinamento adequado. Também é importante que a amostragem seja representativa para os casos de teste e que não haja um número expressivo de dados irrelevantes para a amostra, evitando o ruído e tendências equivocadas.

A escolha de um algoritmo coerente para o aprendizado de máquina é muito importante. Algoritmos que não são adequados à base de dados geram previsões não-representativas quando são aplicados para novos dados de teste. Um dos problemas comuns que ocorrem nesses casos são o *overfitting* e *underfitting*.

*Overfitting* acontece quando um modelo descreve muito bem os dados de treino, mas não funciona bem quando aplicado à outra base de dados, ou seja, não faz boas generalizações. De acordo com Ansari (2016), para superar o problema de *overfitting*, somente uma parte dos dados deve ser usada para treinar o modelo. Dependendo da natureza do problema, uma certa porcentagem pode ser utilizada com o propósito de treinar o método e o resto dos dados é usado para validação e teste. A validação é um teste às cegas que é feito enquanto se treina o modelo. Esse processo utiliza o restante dos dados para checar a performance de predição do modelo.

*Underfitting* é um problema que ocorre quando o modelo é muito simples para aprender as estruturas ocultas na base de dados, o que faz com que as previsões não sejam precisas nem nos dados de treino. Uma boa solução para esse problema é escolher algoritmos que utilizem mais parâmetros na predição, para que possam analisar padrões implícitos à base de dados e aprender com eles.

### 3.4 Classificação de Métodos de Aprendizado

Shalev-Shwartz e Ben David (2013) fazem uma extensa referência à classificação dos métodos de aprendizado. Essas metodologias se subdividem em:

- supervisionado e não supervisionado: serão explicados nas seções 3.2.1 e 3.2.2, respectivamente;
- ativo e passivo: o aprendizado ativo interage consulta os dados em todo o tempo de execução do treino, enquanto o aprendizado passivo utiliza apenas os dados fornecidos em momentos específicos e
- online ou batelada: o aprendizado online pode ser realizado concomitantemente à aquisição de informação, podendo ser constantemente incrementado, enquanto o aprendizado em batelada é realizado apenas uma vez utilizando todos os dados disponíveis.

#### 3.4.1 Aprendizado Supervisionado

No aprendizado supervisionado é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido. Em geral, cada exemplo é descrito por um vetor de valores de características, ou atributos, e o rótulo da classe associada. O objetivo do algoritmo de indução é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados, ou seja, exemplos que não tenham o rótulo da classe. Para rótulos de classe discretos, esse problema é conhecido como classificação e para valores contínuos como regressão (MONARD; BARANAUSKAS, 2008).

Géron (2017) determina como os métodos mais importantes de aprendizado supervisionado: *k-Nearest Neighbours*, Regressão Linear, Regressão Logística, Máquina de Vetores de Suporte (SVM), Árvore de decisão e Florestas Aleatórias e Redes Neurais Supervisionadas.

### 3.4.2 Aprendizado Não-Supervisionado

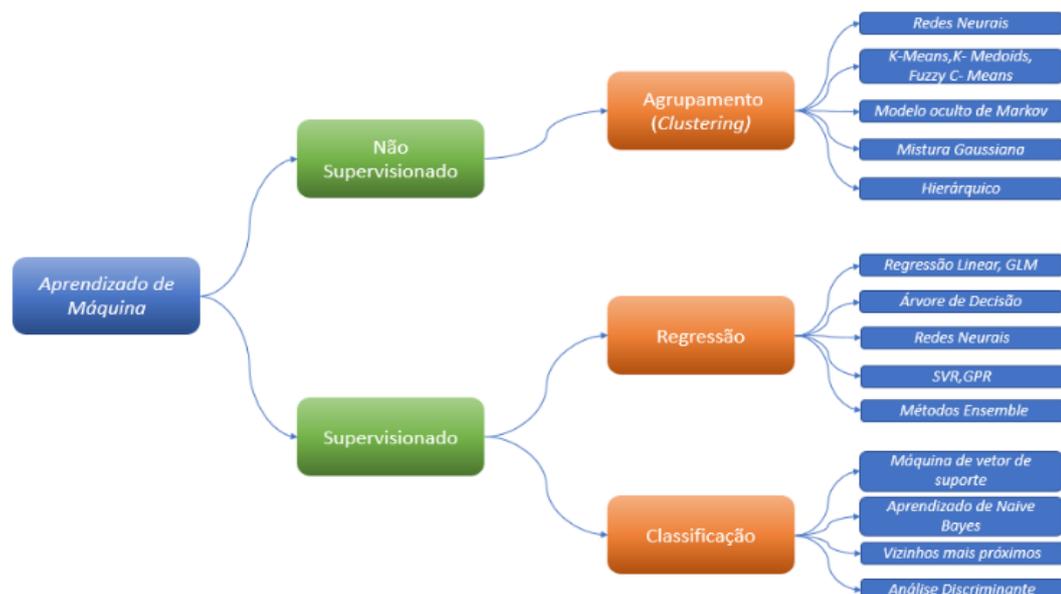
No aprendizado não-supervisionado, o indutor analisa os exemplos fornecidos e tenta determinar se podem ser agrupados de alguma maneira, formando agrupamentos ou *clusters* (HANSON; STUTZ; CHEESEMAN, 1991). Após a determinação dos agrupamentos, normalmente, é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema que está sendo analisado (MONARD; BARANAUSKAS, 2008).

Géron (2017) afirma que os algoritmos mais importantes do aprendizado não-supervisionado são: *Clustering* (*k-Means*, HCA, Maximização da Expectativa), Visualização e redução da dimensionalidade (PCA, *Kernel PCA*, LLE, t-SNE), Aprendizado da regra da associação (*Apriori*, *Eclat*).

### 3.5 Modelos de Aprendizado de Máquina

Na Figura 3, pode-se observar um fluxograma que contém os diferentes tipos de aprendizado de máquina e seus principais métodos. Os métodos de redes neurais e florestas aleatórias podem ser observados no modelo de regressão supervisionado. O método de florestas aleatórias não está no fluxograma, porém esse método é formado por uma combinação de árvores de decisão.

Figura 3 – Tipos de Aprendizado de Máquina



### 3.5.1 Árvore de Decisão

Árvores de Decisão são algoritmos versáteis, que performam tanto problemas de classificação quanto de revisão, e até mesmo problemas múltiplos. São algoritmos poderosos, capazes de ajustar bancos de dados complexos (GÉRON, 2017).

Segundo Hastie, Tibshirani e Friedman (2008), apesar de árvores de decisão serem algoritmos simples e úteis, geralmente não apresentam um bom desempenho quando comparados a métodos mais complexos de aprendizado de máquina. Os métodos de Florestas Aleatórias (*Random Forests*), *Bagging* e *Boosting* combinam um grande número de árvores de decisão e isso resulta em melhoras na precisão da predição, mas podem custar algumas perdas na interpretação.

As folhas da árvore de decisão são os nomes das classes, já os nós representam testes baseados em atributos, com uma ramificação para cada resultado possível. A classificação de um objeto começa na raiz da árvore, e o teste é avaliado e determina o ramo apropriado para a resposta. O processo continua até uma folha ser encontrada. Nesse momento o objeto certamente pertencerá à classe nomeada pela folha (QUINLAN, 1986). Na Figura 4, pode-se observar uma árvore de decisão que classifica uma manhã como ideal ou não para a prática de Tênis.

Figura 4 – Classificação de uma manhã ideal para jogar tênis



Fonte: Adaptada de Mitchell (1997, p. 53)

De acordo com Xu *et al.* (2005), a Árvore de Decisão para regressão é uma variação utilizada para aproximar funções de valor real. A construção desse algoritmo é baseada no recurso de particionamento binário, que é um processo iterativo que

divide os dados. Inicialmente todos os dados de treino são usados para determinar a estrutura da árvore. O algoritmo então quebra a base de dados, utilizando todas as divisões binárias possíveis, a fim de selecionar as duas partes que minimizam a soma do desvio quadrado da média nas partes separadas. O processo de divisão é aplicado a todos os novos galhos até que cada nó atinja o tamanho mínimo especificado pelo usuário e se torne um nó terminal.

### 3.5.2 Florestas Aleatórias

Florestas aleatórias consistem em uma coleção de árvores de decisão, onde cada árvore é construída aplicando um algoritmo nos dados de treino, de acordo com um vetor aleatório que é amostrado de forma ID, ou seja, independente e identicamente distribuído (SHALEV-SHWARTZ; BEN-DAVID, 2014). A predição das florestas aleatórias é obtida a partir do maior número de votos das árvores individuais, realizando uma combinação das suas predições.

Segundo Géron (2017), se uma questão complexa for respondida por milhares de pessoas aleatórias e essas respostas forem agregadas, em muitos casos a resposta agregada será melhor que a resposta de um *expert* no assunto. Essa é a definição do termo Sabedoria de Multidão. Similarmente, se forem agregadas as predições de um grupo de preditores, obter-se-ão melhores predições do que com os melhores preditores individuais. A técnica de agrupar os métodos de Aprendizado é chamada de *Ensemble Learning* ou Aprendizado em Conjunto. Logo o método de florestas aleatórias é um método de Aprendizados em Conjunto que agrega diferentes árvores de decisão.

De acordo com Lantz (2013), os pontos fortes desse método são: boa performance na maior parte dos problemas, bom suporte a bases de dados com ruído ou dados faltantes, seleção dos parâmetros mais importantes e utilização em bases de dados grandes e complexas. Como ponto negativo, é mais difícil de ser interpretado do que Árvore de Decisão e pode requerer trabalho extra para ser ajustado.

### 3.5.4 Redes Neurais Artificiais

Segundo Selau (2000), os modelos neurais procuram aproximar o processamento dos computadores ao cérebro. As Redes Neurais Artificiais (RNA)

possuem um grau de interconexão similar à estrutura do cérebro e em um computador convencional moderno a informação é transferida em tempos específicos, dentro de um relacionamento com um sinal para a sincronização.

Segundo Corrêa (2008), os neurônios que recebem as excitações do exterior são chamados de neurônios de entrada, enquanto os neurônios de saída são os motoneurônios, aqueles usados para alterar o exterior. A utilização de RNAs acontece com intuito de solucionar problemas de previsão, classificação, otimização, controle e outros.

Além disso, redes neurais são sistemas maciços, paralelamente distribuídos e robustos que possuem boa manutenção do desempenho na presença de ruído e capacidade de lidar com dados incompletos. São formados por neurônios (unidades de processamento simples), que armazenam conhecimentos adquiridos através de um processo de aprendizagem. A RNA é inspirada no cérebro humano na forma de adquirir conhecimento, que acontece através de um processo de treinamento. Também lembram cérebros biológicos pela forma de conexão entre os neurônios, conhecidos como pesos sinápticos, que são utilizados para armazenar os conhecimentos adquiridos (HAYKIN, 2001; LIPORACE, 1994).

As técnicas de modelagem a partir de redes neurais podem ser lineares ou não-lineares. Na maioria dos sistemas, os modelos lineares só funcionam bem em uma pequena região de operações ou em processos estacionários, como é o exemplo da rede autorregressiva com entradas exógenas (ARX). Os modelos de redes neurais não-lineares apresentam melhores resultados para processos dinâmicos, já que a saída em um dado instante de tempo depende não apenas de suas entradas atuais, mas do comportamento anterior do sistema.

A rede autorregressiva não-linear com entradas exógenas (NARX, fornece uma representação poderosa para análise, modelagem e previsão de séries temporais devido à sua habilidade de acomodar a natureza dinâmica, complexa e não-linear das aplicações de séries de tempo real (ASYURA; ABDUL; AZIZ, 2011).

Um modelo NARX consiste em entradas exógenas, o que significa que o modelo relaciona o valor atual de uma série temporal com os valores anteriores dessa série para gerar previsões. Ele pode lidar com a maioria dos problemas de simulação e previsão. É uma rede massivamente paralela que pode aprender relações não-lineares muito complexas a partir de um grande conjunto de dados. Por possuírem

memória, as redes dinâmicas podem ser treinadas para aprender a partir de padrões de variação no tempo (DHUSSA *et al.*, 2014).

### 3.6 Histórico de Trabalhos

Brown *et al.* (2006) utilizaram regressão não-linear e redes neurais artificiais para prever as composições da gaseificação de biomassa utilizando as composições de entrada, a razão de equivalência e temperatura. Os autores tentaram diversos tamanhos de camadas ocultas na rede neural *feedforward* e concluíram que a modelagem utilizando Redes Neurais pode desempenhar boas aproximações às modelagens cinética e de equilíbrio, quando aplicada corretamente.

Guo *et al.* (2006) propuseram a modelagem de um estimador inferencial robusto para a predição da composição do *syngas* proveniente da unidade de gaseificação de carvão Texaco em Weinan na China. Foi utilizada análise de componente principal dinâmica aprimorada para o pré-processamento dos dados e aplicada múltiplas redes neurais para o Aprendizado de Máquina. Múltiplas redes neurais combinadas apresentaram resultados mais precisos e robustos do que redes neurais simples. A técnica desenvolvida neste trabalho foi aplicada com sucesso na unidade de gaseificação chinesa.

Elmaz *et al.* (2020) utilizaram Aprendizado de Máquina para aplicar seis técnicas de regressão diferentes, visando prever os parâmetros de saída de um processo de cogaseificação de biomassa e carvão com alto teor de cinzas em gaseificador de leito fluidizado. Foi observado que os métodos que pior performaram foram os de regressão linear e polinomial, já o Máquina de Vetores de Suporte teve sua capacidade de predição diminuída no conjunto reamostrado e a técnica de Árvore de Decisão performou melhor quando a base de dados foi reamostrada, já que os métodos de Árvore de Decisão aprendem melhor utilizando bases de dados maiores.

Debiagi *et al.* (2020) desenvolveram um modelo utilizando Aprendizado de Máquina para reduzir a ordem de um modelo cinética para combustão de carvão mineral a um custo computacional mais baixo do que utilizando a Dinâmica de Fluidos Computacional no contexto de LES (do inglês *large eddy simulation*). A técnica de Aprendizado de Máquina provou ser uma ferramenta promissora para gerar modelos confiáveis e computacionalmente eficientes para manejar a química da combustão do carvão.

## 4 METODOLOGIA

Esta seção apresenta a metodologia experimental e computacional que foi aplicada neste trabalho. Primeiramente é apresentada a planta piloto de gaseificação e o seu procedimento operacional. Posteriormente, são descritas as metodologias utilizadas neste trabalho para análise dos dados, aplicação de aprendizado de máquina e avaliação do seu desempenho.

### 4.1 Detalhes da Planta Piloto de Gaseificação

As Figuras 5 e 6 mostram as vistas frontal e lateral da planta piloto de gaseificação modelo SDC-04A da marca Bioware e capacidade nominal de processamento de 4 kg/h de carvão mineral. Essa planta piloto está localizada no Laboratório de Energia e Carboquímica (LEC) do campus Bagé, da Universidade Federal do Pampa (UNIPAMPA).

Na Figura 5, é possível observar o gaseificador de leito fluidizado borbulhante em (1) e as duas regiões principais: *plenum* (2) e *freeboard* (3), além do sistema de limpeza dos gases em (4) com separação e coleta de coprodutos sólidos e líquidos.

Já na Figura 6, pode-se observar o painel elétrico em (2), que é utilizado para controlar as frequências de giro da rosca de alimentação do carvão mineral e da centrífuga, acionamento dos sopradores de ar (*plenum* e *freeboard*), controle de válvulas rotativas e resistência elétrica externa para partida. Os dados de temperatura e pressão são registrados em um computador, por meio do painel de aquisição de dados (1).

Figura 5 – Vista frontal da planta piloto de gaseificação do LEC



Fonte: Autor (2020)

Figura 6 – Vista lateral da planta piloto de gaseificação do LEC



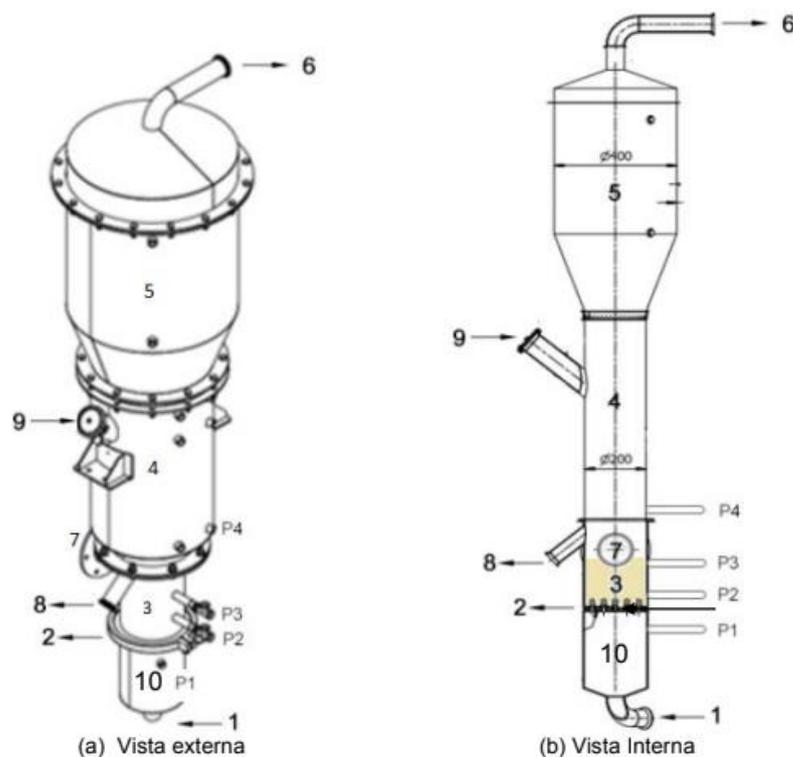
Fonte: Autor (2020)

A Figura 7 mostra um esquema com as vistas interna e externa do gaseificador da planta piloto. Conforme a Figura, o ar entra no gaseificador em (1) e flui até o difusor de ar (2), percolando o leito de partículas (3). Na região *plenum* (4) acontecem as

reações gás-sólido e na região *freeboard* (5), as reações gás-gás. A resistência elétrica externa observada em (3) fornece o calor necessário para a partida da planta piloto. Os gases deixam o gaseificador em (6). Na entrada indicada em (7), o carvão mineral é alimentado, através de uma rosca sem fim, e as cinzas pesadas são retiradas em (8). Em (9) é inserido o agente fluidizante, sendo utilizado neste caso a areia de quartzo (40/50 MESH) e através de (9) é também inserida a carga inicial de carvão mineral para partida da planta piloto. Em (10) é indicada a região abaixo da placa difusora de ar para a região do *plenum*.

A vazão de entrada de ar é controlada por uma válvula globo e ajustada por um rotâmetro. As cinzas pesadas são coletadas no fundo do reator (região *plenum*) e os gases gerados passam por um ciclone do tipo *Lapple*, que coleta as cinzas leves com o auxílio de uma válvula rotativa e as armazena em um tambor de coleta. Os hidrocarbonetos pesados são liquefeitos em um condensador e separados dos gases não condensáveis por meio de uma centrífuga. O *syngas*, é coletado para análise e experimentos ou conduzido a um queimador externo às instalações do LEC.

Figura 7 – Esboço do Gaseificador do LEC



Fonte: Bioware desenvolvimento de tecnologia de energia e meio ambiente LTDA (2015, *apud* Rodrigues 2017, p. 38)

As principais dimensões do gaseificador foram fornecidas pelo seu fabricante, a Bioware Desenvolvimento de Tecnologia de Energia e Meio Ambiente Ltda, aferidas por Sá (2018) e podem ser observadas na Tabela 2.

Tabela 2 – Dimensões das diferentes regiões do gaseificador

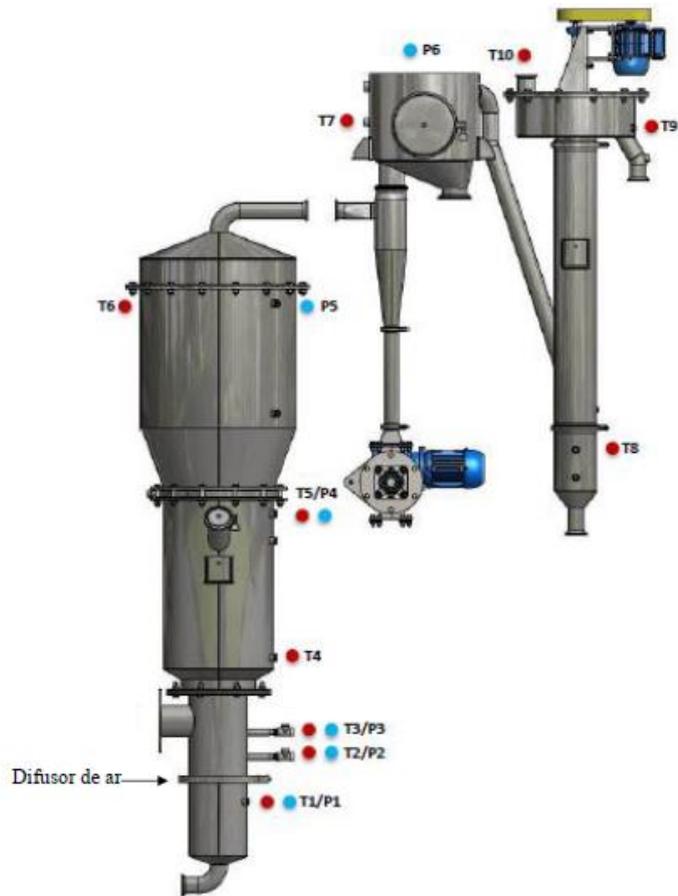
	Dimensões reais (mm)
Diâmetro do <i>plenum</i>	197
Comprimento do <i>plenum</i>	1033
Diâmetro do <i>freeboard</i>	405
Comprimento do <i>freeboard</i>	765

Fonte: Sá (2018, p. 41)

#### 4.1.1 Perfil de Pressão e Temperatura

O registro dos perfis de pressão e temperatura acontece em tempo real. Como pode ser observado na Figura 8, uma série de transdutores de pressão (P1 a P6) e termopares (T1 a T10) que estão distribuídos ao longo da planta piloto de gaseificação. Esses sensores transmitem sinais que são enviados para o sistema de monitoramento do *software FieldChart* (Novus Produtos Eletrônicos, Ltda.), que realiza a leitura e registro dos dados.

Figura 8 – Distribuição dos sensores de temperatura e pressão na planta piloto de gaseificação do LEC



Fonte: Bioware (2015, *apud* Sá 2018, p. 44)

Na Tabela 3, pode-se observar a distribuição dos sensores T2 a T6 e P2 a P5 ao longo da altura do gaseificador, tomando como referência o difusor de ar. Os sensores T1 e P1 estão posicionados abaixo do difusor de ar. Os sensores T7 e P6 estão posicionados no condensador de gases, T9 e T10 estão posicionados na centrífuga e T8, no coletor de condensados e alcatrão.

Tabela 3 – Distribuição de sensores no gaseificador

Sensores	Altura (mm)
T2, P2	73
T3, P3	16
T4	195
T5, P4	248
T6, P5	322

Fonte: Adaptada de Sá *et al.* (2017, p. 45)

## 4.2 Coleta de Dados Experimentais

Os melhores resultados de aprendizado de máquina são obtidos quando se dispõe de uma grande base de dados. Por esse motivo foram utilizados dados de corridas experimentais realizadas anteriormente na planta piloto de gaseificação do LEC, para assim ser obtida uma melhor predição do algoritmo. Os dados de saída de perfil de pressão e temperatura foram coletados no computador do LEC a partir de registros dos anos de 2015 a 2020.

As corridas utilizadas para a análise foram escolhidas com base nos perfis de temperatura, onde foram selecionadas as corridas que chegaram a um estado estacionário acima de 700°C, permitindo assim que todas as reações envolvidas no processo de gaseificação tenham ocorrido. Após isso foram retiradas as corridas que continham inconsistências e dados corrompidos, resultando em um total de 15 conjuntos de dados. A obtenção dos dados de entrada correspondentes foi realizada a partir de um caderno físico onde é feito o registro escrito de cada partida da planta piloto. A partir desse caderno ou diário de operações, os seguintes dados foram obtidos para cada corrida experimental:

- 1) resistência elétrica de partida como ligada ou desligada;
- 2) vazão de ar na região do *plenum* dada em Nm<sup>3</sup>/h;
- 3) vazão de ar na região do *freeboard* dada em Nm<sup>3</sup>/h;
- 4) pulso de ar no leito como realizado ou não;
- 5) alimentação de carvão mineral a partir da rotação da rosca alimentadora dada em Hz;
- 6) alimentação de carvão mineral pela janela de alimentação (posição 9 da Fig. 7) dada em litros e
- 7) rotação da centrífuga dada em Hz.

Todas essas informações estavam escritas manualmente e foram transcritas para planilhas eletrônicas. Assim, dados de entrada e saída do processo foram agrupados em um mesmo formato para facilitar a etapa de pré-processamento.

Para cada corrida experimental foi utilizado para partida um valor padrão de 2 litros de carvão mineral pulverizado e 6 litros de areia de quartzo (SÁ; RAUPP; MUNIZ, 2017). As vazões de alimentação de carvão mineral e ar foram estudadas para que o

processo de gaseificação possa ser estudado em diferentes condições de estado estacionário.

### 4.3 Análise dos Dados, Aplicação do Aprendizado de Máquina e Avaliação do Desempenho

Para realizar a análise computacional, utilizou-se o ambiente de desenvolvimento *Jupyter Notebook* (*Project Jupyter*, <https://jupyter.org>) versão 6.1.4, e a linguagem de programação *Python* (Python Software Foundation, <https://python.org>), versão 3.7.3. A linguagem Python é uma linguagem *opensource*, ou seja, uma linguagem de código aberto, sem custo, e aberta a melhorias constantemente.

Para implementar o trabalho proposto, usaram-se bibliotecas da linguagem *Python*, sendo as principais: *Scikit-learn*, uma biblioteca com algoritmos de aprendizado de máquina; *Pandas*, uma biblioteca para manipulação e análise de dados; *Numpy*, uma biblioteca para manipulação de estruturas e dados numéricos; *Matplotlib* e *Seaborn*, bibliotecas utilizadas para visualização de dados.

A partir da biblioteca *Pandas*, foi feito o carregamento das planilhas, contendo as temperaturas e pressões durante o processo, assim como os dados de entrada para cada um dos 15 conjuntos de dados. Foi programada uma rotina em Python para definir quais conjuntos de dados pertencem ao mesmo experimento, visando definir quais eram os dados de entrada e dados de saída (temperatura e pressão) para cada corrida.

Realizou-se um pré-processamento a fim de garantir a organização e estruturação da base de dados. Para isso retiraram-se os valores incompletos, e foram definidos os tempos de início e fim para cada experimento. Além disso, fez-se uma reamostragem a cada 5 segundos, garantindo que estivessem contidos valores nos mesmos intervalos de tempo para todas as variáveis. Posteriormente realizou-se a concatenação para unificar e criar uma base de dados para a execução do Aprendizado de Máquina.

Após o pré-processamento, analisou-se a correlação entre as variáveis estudadas, a fim de compreender o grau de relação existente entre elas, utilizando a função *heatmap* da biblioteca *Seaborn*. Na próxima etapa, aplicaram-se os algoritmos de Aprendizado de Máquina aos dados obtidos experimentalmente, para a predição

dos dados de saída e geração de gráficos que descrevam os perfis dinâmicos de temperatura e pressão durante a operação. Para isso, separou-se a base de dados em dados de treino e teste, sendo que os dados de treino são utilizados para realizar o aprendizado do algoritmo, e os dados de teste, para evidenciar a qualidade da predição.

Os algoritmos treinados neste trabalho são algoritmos de aprendizado supervisionado. Os métodos de árvore de decisão e florestas aleatórias estão presentes na biblioteca *Scikit-Learn*. Enquanto o algoritmo de redes neurais artificiais utilizando o método NARX está disponível na biblioteca *FireTS*, em <https://pypi.org/project/fireTS>.

A qualidade da predição dos modelos é quantificada pelo coeficiente de determinação ( $R^2$ ), o erro quadrático médio (MSE) e a raiz do erro quadrático médio (RMSE), demonstrados pelas Equações 2, 3 e 4, respectivamente. Esses parâmetros estatísticos são definidos a seguir:

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2} \quad (3)$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

onde,  $\hat{y}_i$  são os valores observados,  $y_i$  são os valores estimados e/ou previstos e  $n$  é o tamanho da amostra de dados.

O coeficiente de determinação ( $R^2$ ) é uma medida de ajuste de um modelo estatístico, sendo definido em um intervalo de 0 a 1. O erro quadrático médio (MSE na sigla em Inglês) é sensível aos grandes erros, por elevar as diferenças individuais ao quadrado. Ele é sempre positivo e quanto mais próximo de zero indica uma melhor predição dos dados observados. A Raiz do Erro Quadrático Médio (RMSE em Inglês), é muito usada para expressar a acurácia dos resultados numéricos e apresenta valores do erro nas mesmas dimensões da variável analisada (HALLAK; PEREIRA FILHO, 2011).

## 5 RESULTADOS

Após a coleta e análise dos dados experimentais, a base de dados foi composta por experimentos que foram numerados de 1 a 15, para facilitar as citações. Os experimentos foram realizados nas seguintes datas: 28/06/2016, 29/07/2016, 01/08/2016, 26/09/2016, 17/11/2016, 30/11/2016, 03/11/2017, 24/03/2017, 12/05/2017, 24/05/2017, 20/09/2017, 10/11/2017, 27/04/2018, 19/11/2020 e 11/12/2020. Foram gerados 45 arquivos de planilhas eletrônicas, sendo que essas contêm temperaturas e pressões obtidas a partir do *software* de aquisição de dados *FieldChart Novus*, disponível no LEC, de modo que as planilhas contendo as alterações nos dados de entrada do processo foram transcritas a partir do diário de operações da planta de gaseificação onde são mantidos os registros de corridas experimentais e operações de rotina desse equipamento pelos usuários do LEC.

As Figuras 9,10 e 11 demonstram o formato dessas planilhas e são relativas aos dados obtidos em um experimento realizado no dia 01/08/2016. Na Figura 9, pode ser observada a tabela correspondente aos Dados de Entrada e às alterações dos parâmetros ao longo do processo, sendo os parâmetros: a operação da resistência elétrica de partida (1 para ligado, 0 para desligado); as vazões de ar nas regiões *plenum* e *freeboard* (Nm<sup>3</sup>/h); os pulsos de ar no leito (1 para pulso, 0 para não pulso); a frequência de alimentação (Hz); a alimentação de carvão a partir da janela de alimentação (L); e a frequência de operação da centrífuga (Hz). Na Figura 10, podem ser observados os valores de temperatura (°C) referentes aos sensores T1 a T8. Já na Figura 11, podem ser observados os valores de pressão (mmH<sub>2</sub>O) referentes aos sensores P1 a P6 e os valores de temperatura (°C) referentes aos sensores T9 e T10.

Figura 9 – Dados de Entrada para o experimento de gaseificação realizado na data 01/08/2016

Data	Hora	Resistência	Vazão Plenum	Vazão Freeboard	Pulso	Alimentação (Hz)	Alimentação janela (L)	Centrífuga
01/08/2016	11:20:00	1	9	3	0	0	0	0
01/08/2016	13:16:00	1	11	3	0	0	0	0
01/08/2016	13:38:00	1	12	3	0	0	0	0
01/08/2016	13:47:00	1	12	3	1	0	0	0
01/08/2016	14:32:00	1	12	3	0	0	2	0
01/08/2016	14:53:00	0	12	3	0	13	0	45
01/08/2016	15:34:00	0	13	3	0	13	0	45
01/08/2016	15:51:00	0	11	3	0	13	0	45
01/08/2016	16:01:00	0	12	3	0	13	0	45
01/08/2016	16:34:00	0	13	3	0	13	0	45
01/08/2016	16:48:00	0	11	3	0	12	0	45
01/08/2016	16:54:00	0	10	3	0	10	0	45
01/08/2016	16:56:00	0	10	3	0	0	0	45
01/08/2016	16:57:00	0	11	3	0	10	0	45
01/08/2016	17:00:00	0	11	3	0	14,3	0	40,67
01/08/2016	17:01:00	0	11	3	0	0	0	40,67
01/08/2016	17:03:00	0	0	0	0	0	0	0

Fonte: Autora (2021)

Figura 10 – Tabela de temperaturas para o experimento realizado na data 01/08/2016

Data	Hora	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8
01/08/2016	11:20:08.5	14	15	15	16	17	18	15	14
01/08/2016	11:20:13.0	14	15	15	16	17	18	15	14
01/08/2016	11:20:18.0	14	15	15	16	17	18	15	14
01/08/2016	11:20:23.1	14	15	15	16	17	18	15	14
01/08/2016	11:20:28.1	14	15	15	16	17	18	15	14
01/08/2016	11:20:33.1	14	15	15	16	17	18	15	14
01/08/2016	11:20:38.2	14	15	15	16	17	18	15	14
01/08/2016	11:20:43.2	14	15	15	16	17	18	15	15
01/08/2016	11:20:48.3	14	15	15	16	17	18	15	15
01/08/2016	11:20:53.3	14	15	15	16	17	18	15	14
01/08/2016	11:20:58.3	14	15	15	16	17	18	15	14
01/08/2016	11:21:03.4	14	15	15	16	17	18	15	14
01/08/2016	11:21:08.4	14	15	15	16	17	18	15	14
01/08/2016	11:21:13.4	14	15	15	16	17	18	15	14
01/08/2016	11:21:18.0	14	15	15	16	17	18	15	14
01/08/2016	11:21:23.0	14	15	15	16	17	18	15	14
01/08/2016	11:21:28.0	14	15	15	16	17	18	15	14
01/08/2016	11:21:33.1	14	15	15	16	17	18	14	14
01/08/2016	11:21:38.1	14	15	15	16	17	18	15	15
01/08/2016	11:21:43.1	14	15	15	16	17	18	15	15
01/08/2016	11:21:48.1	14	15	15	16	17	18	15	15
01/08/2016	11:21:53.2	14	15	15	16	17	18	15	15
01/08/2016	11:21:58.2	14	15	15	16	17	18	15	15
01/08/2016	11:22:03.2	14	15	15	16	17	18	15	15

Fonte: Autora (2021)

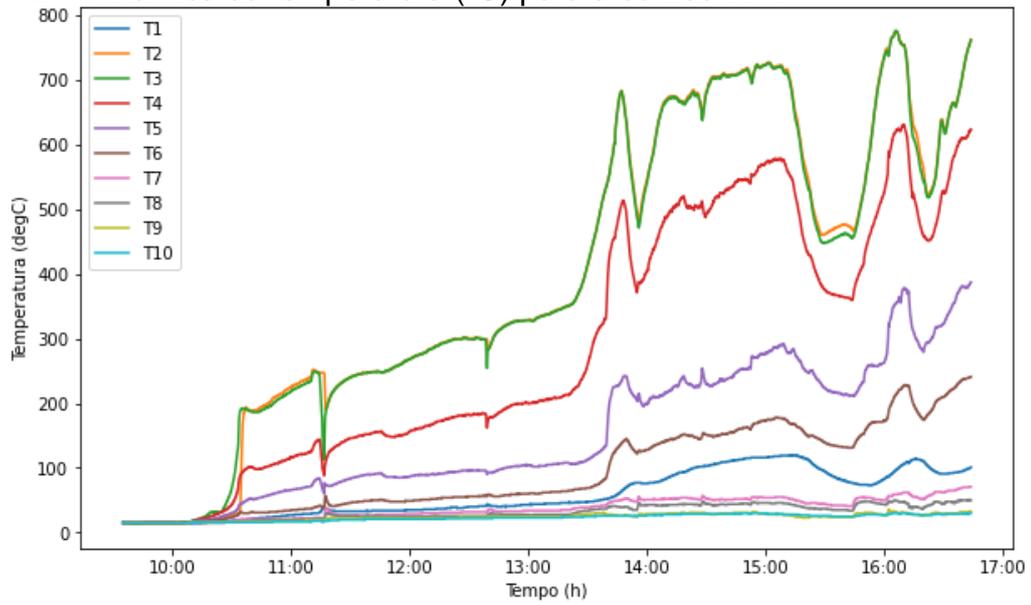
Figura 11 – Tabela de pressões e temperaturas obtidas para o experimento realizado na data 01/08/2016

Data	Hora	P 1	P 2	P 3	P 4	P 5	P 6	T 9	T 10
01/08/2016	11:20:09.5	1	4	1	7	1	6	16	15
01/08/2016	11:20:14.5	1	4	0	-8	2	6	16	15
01/08/2016	11:20:19.6	1	4	0	-8	2	6	16	15
01/08/2016	11:20:24.6	1	5	0	-8	3	6	16	15
01/08/2016	11:20:29.6	165	80	63	-5	0	4	16	15
01/08/2016	11:20:34.7	338	154	96	8	14	8	16	15
01/08/2016	11:20:39.7	335	151	92	9	15	8	16	15
01/08/2016	11:20:44.7	333	150	91	9	15	9	16	15
01/08/2016	11:20:49.8	332	148	90	9	15	8	16	15
01/08/2016	11:20:54.8	331	148	89	9	14	8	16	15
01/08/2016	11:20:59.8	331	148	89	9	15	9	16	15
01/08/2016	11:21:04.9	331	148	89	9	15	9	16	15
01/08/2016	11:21:09.9	330	148	89	9	14	9	16	15
01/08/2016	11:21:14.9	330	157	94	11	16	10	16	15
01/08/2016	11:21:19.5	330	157	94	11	16	10	16	15
01/08/2016	11:21:24.5	331	159	94	11	16	11	16	15
01/08/2016	11:21:29.5	329	158	94	11	16	10	16	16
01/08/2016	11:21:34.6	329	158	94	11	16	11	16	16
01/08/2016	11:21:39.6	329	158	94	11	16	11	16	16
01/08/2016	11:21:44.6	329	158	94	12	17	11	16	16
01/08/2016	11:21:49.6	329	158	94	11	17	11	16	16
01/08/2016	11:21:54.7	329	158	94	11	17	11	16	16
01/08/2016	11:21:59.7	329	158	94	11	17	11	16	16
01/08/2016	11:22:04.7	328	158	94	11	17	11	16	16

Fonte: Autora (2021)

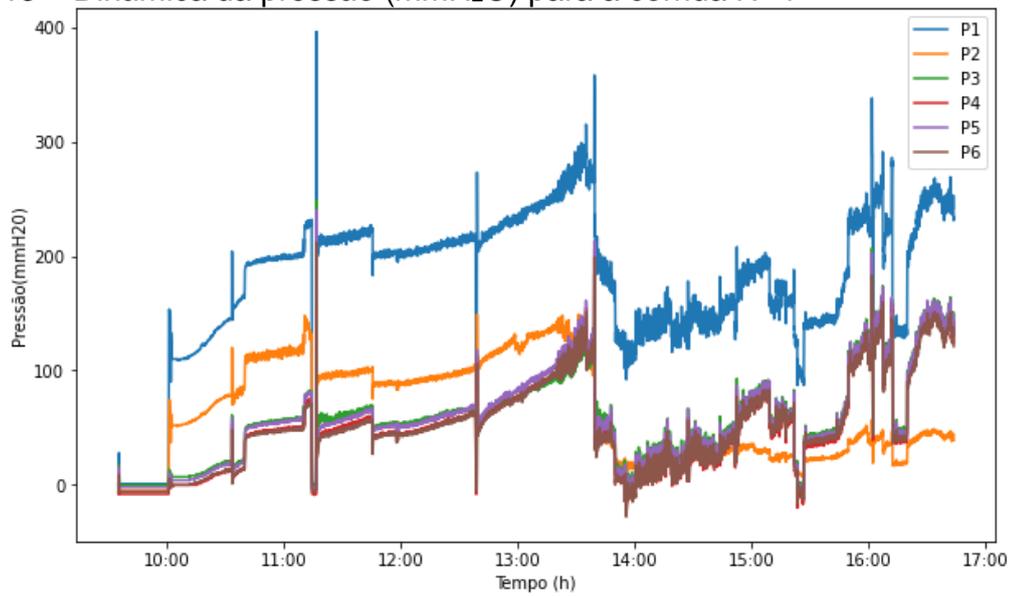
A partir da biblioteca *Matplotlib*, geraram-se os gráficos para os perfis de pressão e temperatura ao longo do tempo. Os gráficos gerados para as corridas experimentais um, seis e oito podem ser observados nas Figuras 12, 13, 14, 15, 16 e 17, onde cada cor de linha representa um sensor ao longo do reator.

Figura 12 – Dinâmica da temperatura (°C) para a corrida N° 1



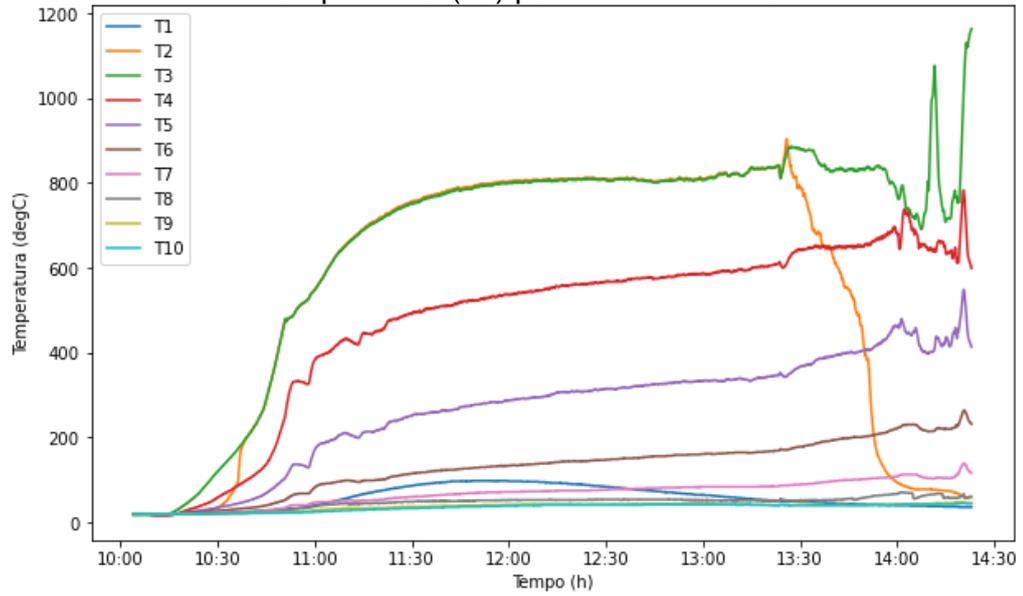
Fonte: Autora (2021)

Figura 13 – Dinâmica da pressão (mmH<sub>2</sub>O) para a corrida N° 1



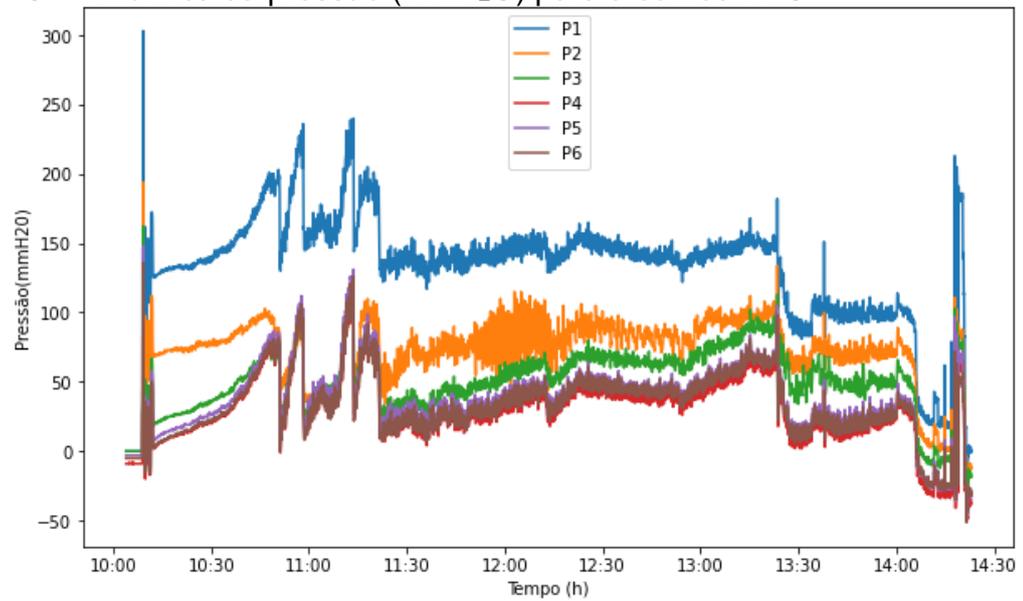
Fonte: Autora (2021)

Figura 14 – Dinâmica da temperatura (°C) para a corrida N° 6



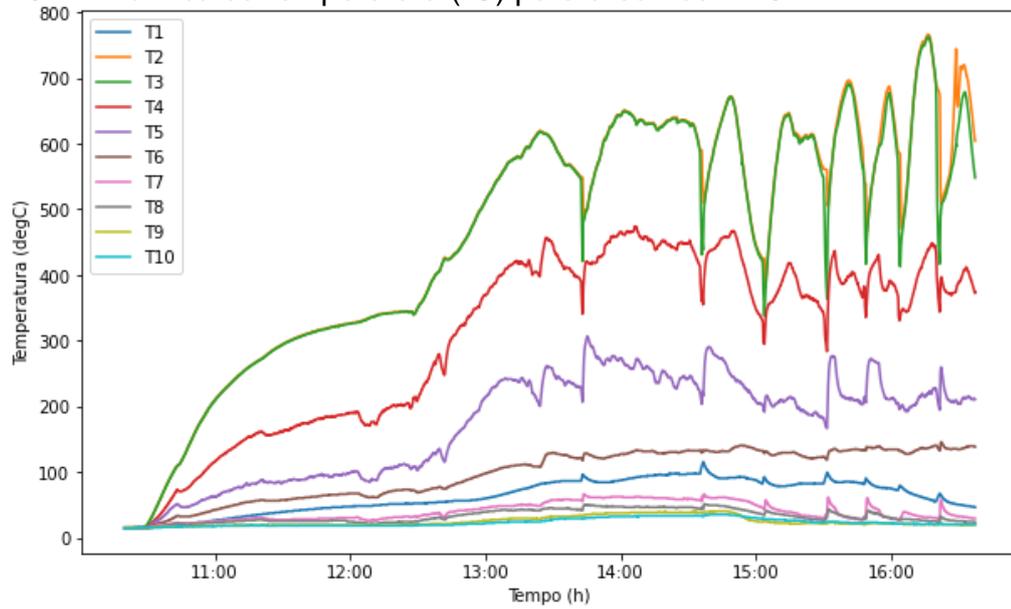
Fonte: Autora (2021)

Figura 15 – Dinâmica da pressão (mmH<sub>2</sub>O) para a corrida N° 6

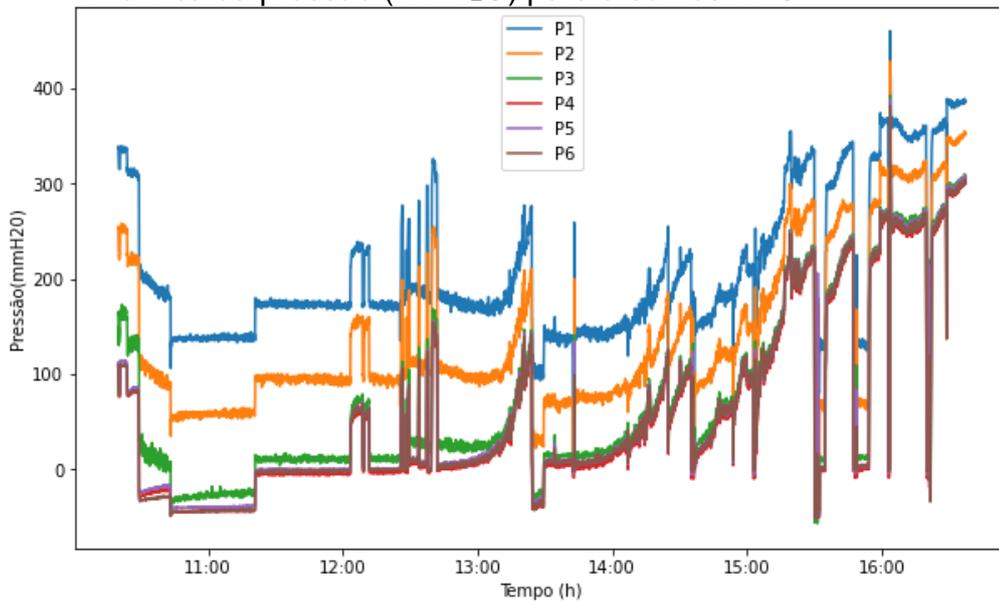


Fonte: Autora (2021)

Figura 16 – Dinâmica da temperatura (°C) para a corrida N° 8



Fonte: Autora (2021)

Figura 17 – Dinâmica da pressão (mmH<sub>2</sub>O) para a corrida N° 8

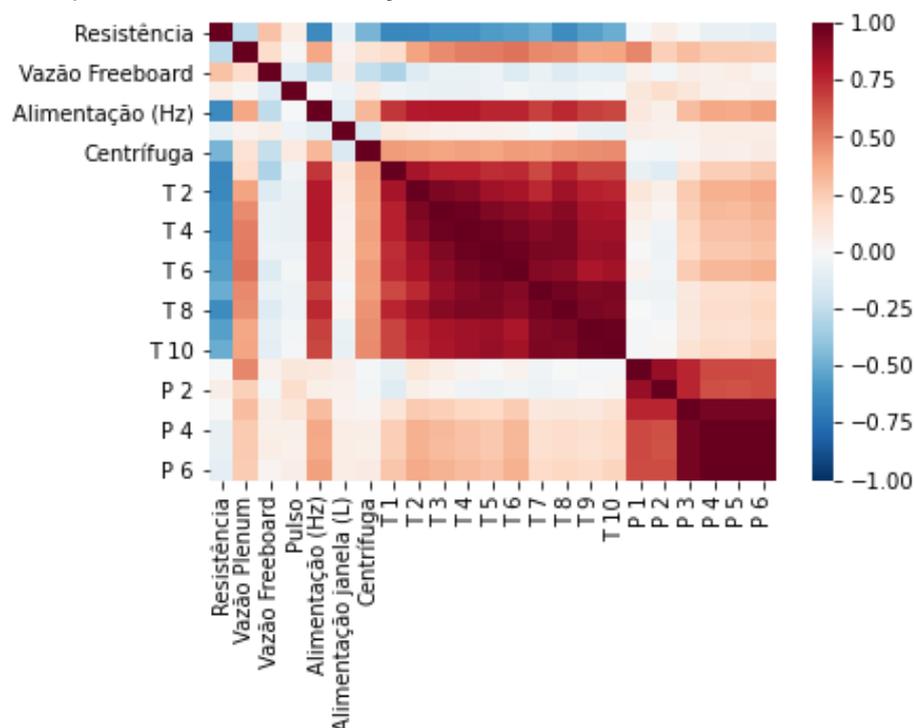
Fonte: Autora (2021)

A Figura 16, referente à corrida N° 8, indica uma operação irregular do gaseificador que pode ser percebida pela oscilação da temperatura, a partir 13h30min, em torno do estado estacionário entre 650 e 750°C. Próximo do final do registro da corrida, pode-se observar o descolamento entre T2 e T3. A partir de Figura 17, pode-se perceber que foram feitos pulsos de ar próximo às 12h30min em uma tentativa de promover uma melhor fluidização do leito e conseqüente elevação da temperatura. Contudo, há momentos onde o fornecimento de ar foi interrompido (13h30min, 15h30min e 15h45min) em uma tentativa de desacelerar o rápido aquecimento do leito.

### 5.1 Análise de Correlação

A partir da biblioteca Seaborn, obteve-se um mapa de calor (*heatmap*). Essa ferramenta auxilia no estudo de correlação das variáveis presentes na base de dados, utilizando cores. Nessa representação, quanto mais escura for a tonalidade, maior é a correlação. Na Figura 18, pode-se observar esse mapa de calor, onde tons de vermelho apresentam correlação direta; já os tons de azul apresentam correlação inversa.

Figura 18 – Mapa de calor de correlação entre dados de entrada e saída



Fonte: Autora (2021)

Ao analisar as cores do mapa de calor, pode-se perceber que a resistência apresenta uma forte correlação inversa em relação às temperaturas durante o processo. Isso pode ser explicado pelo fato de que a resistência é ligada apenas para dar partida no processo. E é desligada quando as reações começam a ocorrer e as temperaturas começam a aumentar.

A vazão *plenum* apresenta correlação direta com as Temperaturas, Pressões, Centrífuga, Alimentações e vazão *freeboard*. Dentre elas, as maiores correlações são com as temperaturas. Isso acontece por conta da vazão *plenum* disponibilizar ar dentro do reator. Como grande parte do ar é composto por oxigênio, a maior disponibilidade desse elemento dentro do reator propicia as reações de oxidação do carbono. Além disso, a vazão *plenum* apresenta correlação direta com as pressões, já que propicia as reações e, conseqüentemente, o aumento da temperatura e expansão dos gases. Além disso, proporciona vazão de substâncias gasosas para dentro do leito, o que contribui para aumentar a pressão.

A vazão *freeboard* apresenta baixa correlação com os outros parâmetros analisados, o que pode ser explicado por conta dessa vazão se encontrar acima do reator e ser constante durante o processo, que é dinâmico.

A frequência de Alimentação apresenta uma forte correlação direta com as temperaturas, indicando que a maior vazão do combustível sólido (carvão mineral) proporciona o aumento das temperaturas já que favorece as reações exotérmicas que ocorrem no processo. Mostra também correlação moderada com as pressões, já que o aumento da temperatura gera expansão dos gases e, por conseqüência, maiores pressões.

A frequência da Centrífuga apresenta correlação direta com as temperaturas, frequência de Alimentação e vazão *plenum*. Isso pode ser justificado por conta da centrífuga ser acionada manualmente quando as temperaturas começam a subir, ou seja, quando começam a ocorrer as reações, que são impulsionadas pelo aumento da frequência de alimentação do carvão e vazão *plenum* de ar.

A variável pulso indica o momento em que é manualmente acionado um pulso de pressão (ar) no reator. Isso acontece quando sensores de temperatura próximos começam a apresentar temperaturas diferentes, indicando que há aglomeração de partículas, o que pode acabar gerando problemas na fluidização do leito e parada no

processo. Por esses motivos, a variável pulso apresenta correlação direta com as pressões e uma leve correlação inversa com as temperaturas.

As temperaturas apresentam alta correlação direta entre si. A correlação é mais alta quanto mais próximos fisicamente são os sensores, sobretudo entre os sensores T2 a T6 que se encontram fixados no gaseificador. Isso pode ser explicado pelo fato de que o leito é fluidizado, e a passagem do ar para fluidização facilita a troca de calor entre as regiões.

As pressões ao longo do leito apresentam no geral forte correlação entre si, porém pode-se observar maior correlação entre os sensores que estão mais próximos fisicamente.

## **5.2 Aprendizado de Máquina**

Para a execução do aprendizado de máquina, é necessário separar os dados de forma aleatória para treino e teste, ou seja, para o aprendizado e para a validação do algoritmo. Para isso, os dados de todas as corridas foram concatenados, o que gerou uma base de dados de 31643 linhas e 23 colunas. Então esses dados foram separados em x e y, onde x são os dados de entrada do processo, e y os dados de saída do processo. A partir disso, os dados foram separados de forma aleatória, sendo que 90% dos dados foram utilizados para treino e 10% para teste. Das 15 corridas, duas foram separadas para testar o comportamento dos modelos obtidos em novos conjuntos de dados. As corridas experimentais N°14 e N°15 não foram utilizadas para o treinamento dos modelos de aprendizado de máquina.

### **5.2.1 Árvore de Decisão e Florestas Aleatórias**

O método de árvore de decisão foi executado para todos os dados disponíveis e foram obtidos os seguintes resultados:

Coeficiente de Determinação ( $R^2$ ) = 0,8263000000248909;

Erro quadrático médio (MSE) = 1326,3748782958091;

Raiz do erro quadrático médio (RMSE) = 36,41942995566802.

O método de florestas aleatórias também foi executado para todos os dados disponíveis. Foi definido que a floresta aleatória contaria com 50 árvores de decisão, e a partir delas foram obtidos os seguintes resultados:

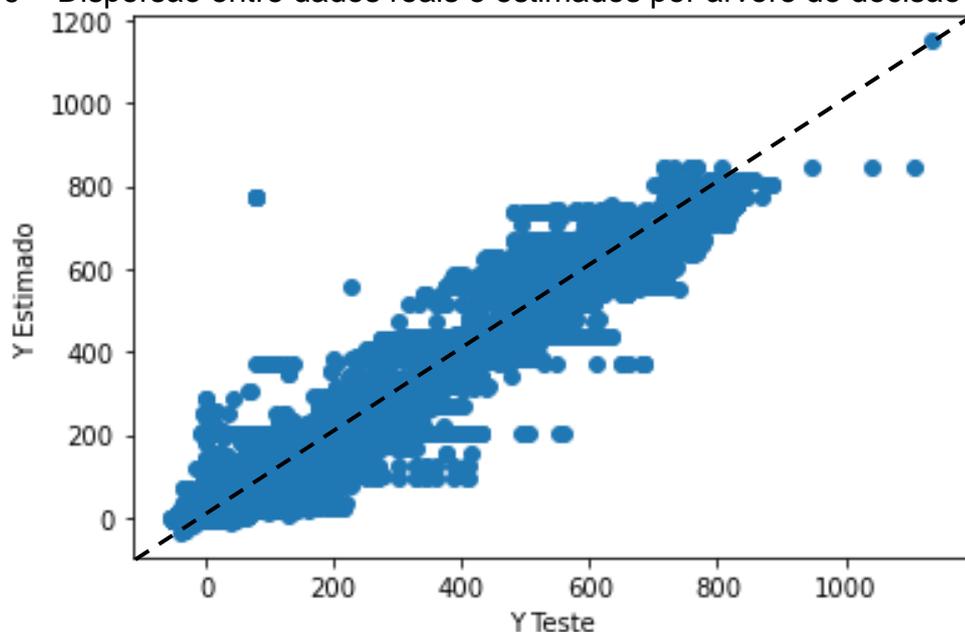
Coeficiente de Determinação ( $R^2$ ) = 0,8262340678430705;

Erro quadrático médio (MSE) = 1327,6498392806247;

Raiz do erro quadrático médio (RMSE) = 36,436929608305704.

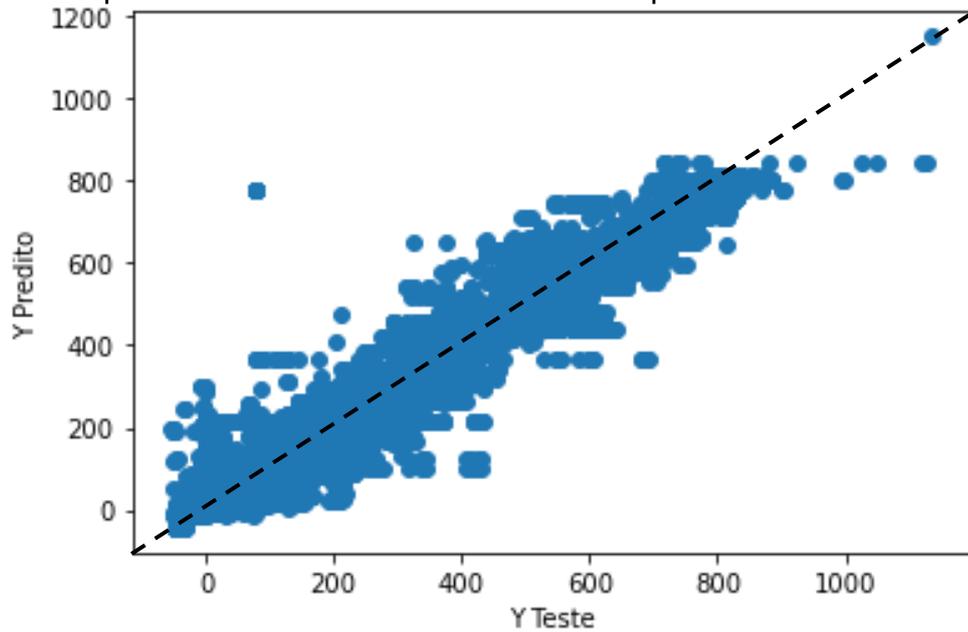
As Figuras 19 e 20 são gráficos de dispersão que relacionam os valores de saída reais e previstos para todas as 15 corridas experimentais. No eixo vertical, estão os valores reais de  $y$  de teste, e no eixo horizontal estão os valores de  $y$  estimado na Figura 19 pelo algoritmo de árvore de decisão, e na Figura 20 pelo algoritmo de florestas aleatórias. Quanto mais próxima à diagonal de  $45^\circ$  for a distribuição entre os dados reais e previstos, melhor é a predição. Ambas as Figuras indicam uma distribuição similar entre os pontos, o que pode ser confirmado pelos valores de  $R^2$ , MSE e RMSE, que apresentaram resultados muito parecidos.

Figura 19 – Dispersão entre dados reais e estimados por árvore de decisão



Fonte: Autora (2021)

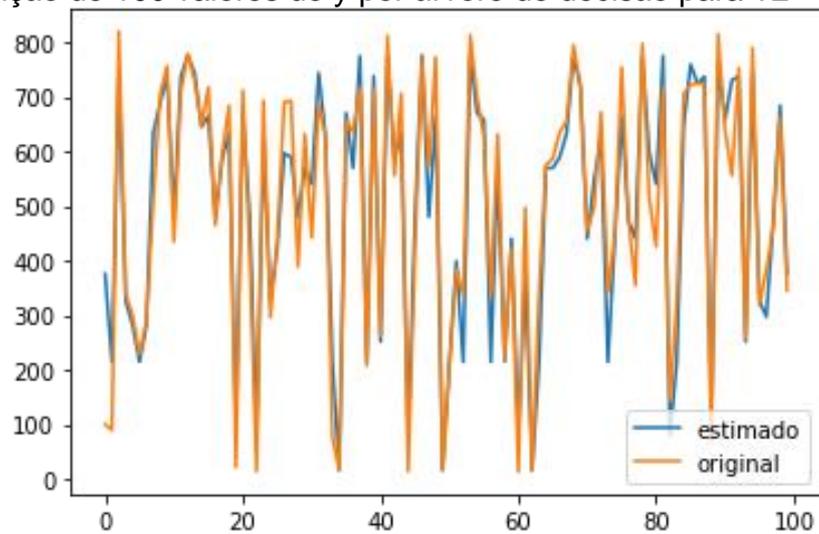
Figura 20 – Dispersão entre dados reais e estimados por florestas aleatórias



Fonte: Autora (2021)

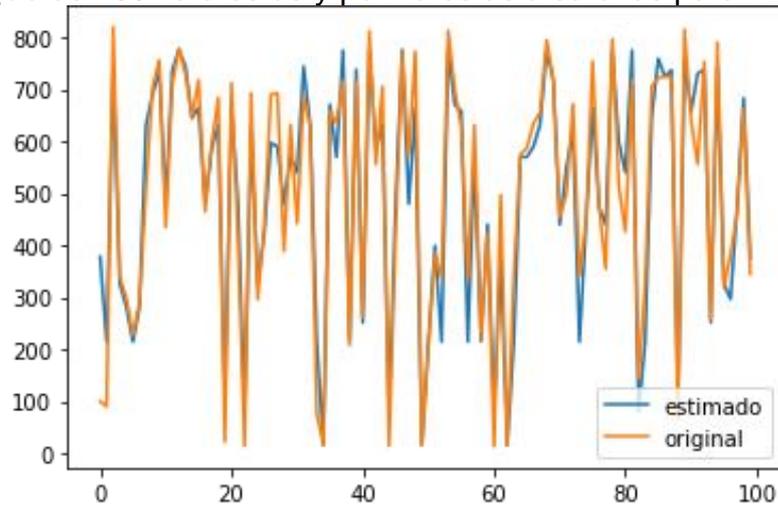
Nas Figuras 21 e 22, as linhas laranjas demonstram 100 valores reais de temperatura selecionados aleatoriamente para a variável temperatura T2, enquanto as linhas azuis representam os valores estimados pelos algoritmos. Na Figura 21, é utilizado o método de árvore de decisão e na Figura 22, o método de florestas aleatórias.

Figura 21 – Predição de 100 valores de y por árvore de decisão para T2



Fonte: Autora (2021)

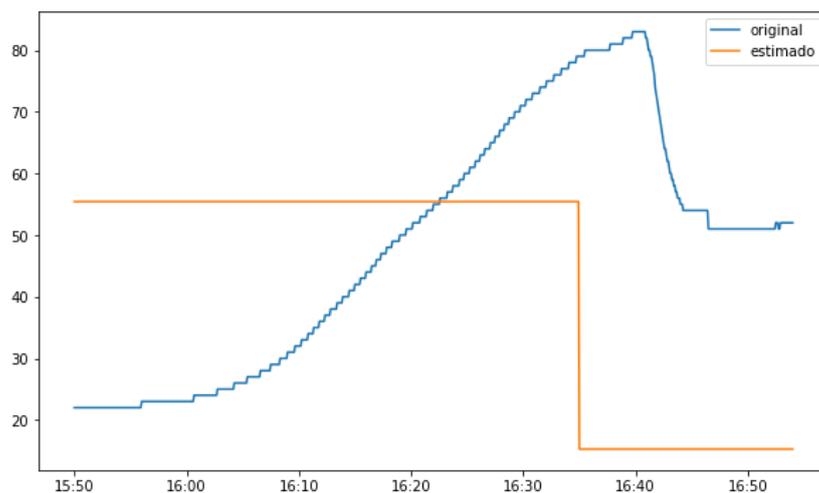
Figura 22 – Predição de 100 valores de y por florestas aleatórias para T2



Fonte: Autora (2021)

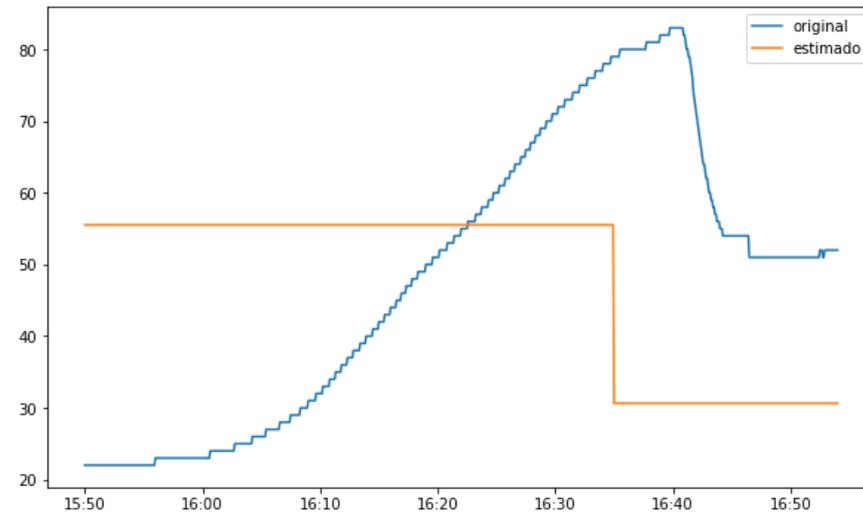
A predição foi testada também para as corridas N<sup>o</sup>14 e N<sup>o</sup>15 que estão fora da base de dados, com objetivo de avaliar o desempenho da predição em corridas experimentais que não foram utilizadas para treino do algoritmo. Nas Figuras 23 e 24, pode-se observar em azul o comportamento real do sensor de temperatura T2, para a corrida experimental N<sup>o</sup>14. Em laranja, são apresentados os comportamentos previstos pelos algoritmos. Na Figura 23, é utilizado o método de árvore de decisão e, na Figura 24, o método de florestas aleatórias.

Figura 23 – Comportamento previsto por árvore de decisão, para T2 da corrida experimental N<sup>o</sup>14



Fonte: Autora (2021)

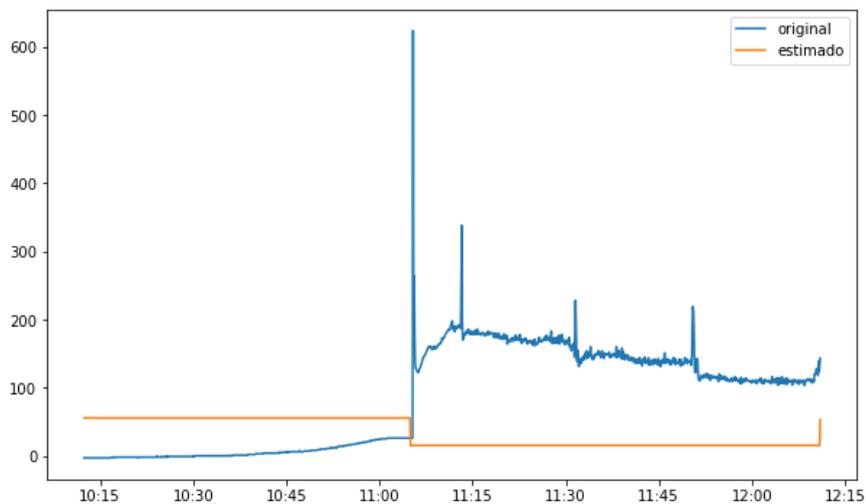
Figura 24 – Comportamento previsto por florestas aleatórias, para T2 da corrida experimental N°14



Fonte: Autora (2021)

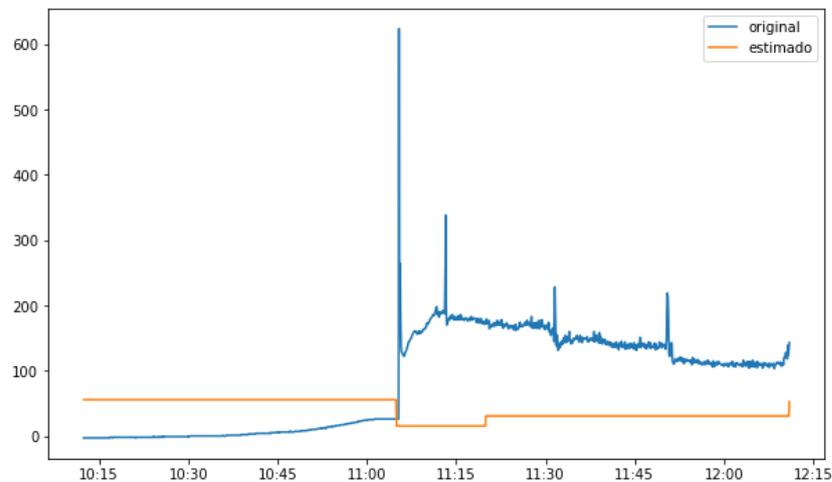
Nas Figuras 25 e 26, pode-se observar em azul o comportamento real de pressão obtido para o sensor P1 da corrida experimental N°15, enquanto em laranja pode-se observar os comportamentos previstos pelos algoritmos. Na Figura 25, é utilizado o método de árvore de decisão e, na Figura 26, o método de florestas aleatórias.

Figura 25 – Comportamento previsto por árvore de decisão, para P1 da corrida experimental N°15



Fonte: Autora (2021)

Figura 26 – Comportamento previsto por florestas aleatórias, para P1 da corrida experimental N°15



Fonte: Autora (2021)

Por terem sido obtidos resultados muito próximos utilizando os dois métodos, testou-se dividir novamente a base de dados, com uma parte bem menor dos dados para realizar o aprendizado de máquina. Para isso reduziu-se a base de dados em torno de 30 vezes, gerando uma nova base de dados com 950 linhas e 23 colunas. Manteve-se a definição de 90% dos dados para treino e 10% dos dados para teste.

A seguir apresentam-se os resultados obtidos para o método de árvore de decisão, utilizando uma base de dados 30 vezes menor:

Coefficiente de Determinação ( $R^2$ ) = 0.8432962049883554;

Erro quadrático médio (MSE) = 1105.9636135650758;

Raiz do erro quadrático médio (RMSE) = 33.25603123592886.

O método de florestas aleatórias também foi executado para a base de dados 30 vezes menor. Definiu-se que a floresta aleatória contaria com 50 árvores de decisão, e a partir delas foram obtidos os seguintes resultados:

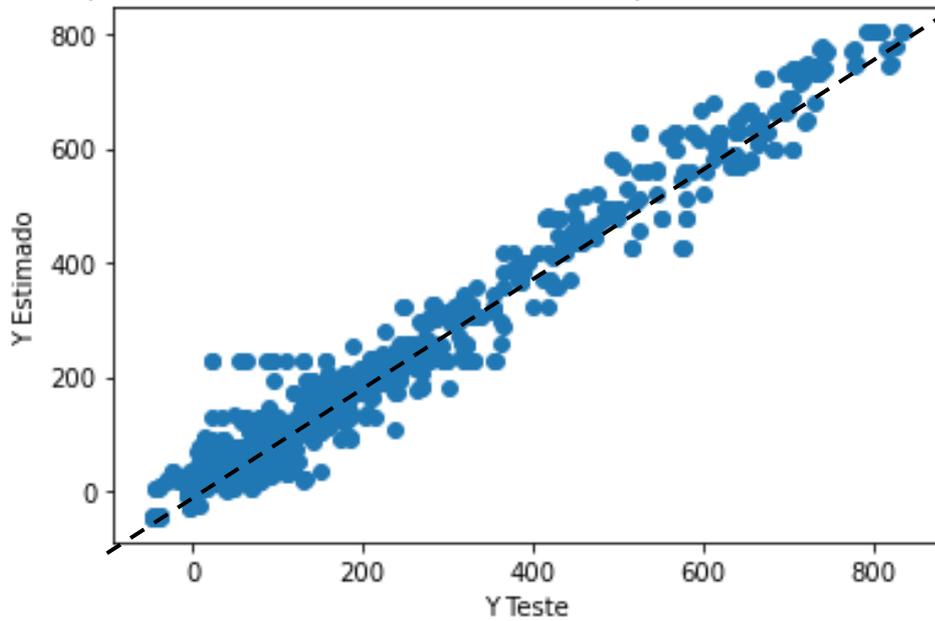
Coefficiente de Determinação ( $R^2$ ) = 0.8525453115237354;

Erro quadrático médio (MSE) = 1074.2275235886818;

Raiz do erro quadrático médio (RMSE) = 32.77541034966125.

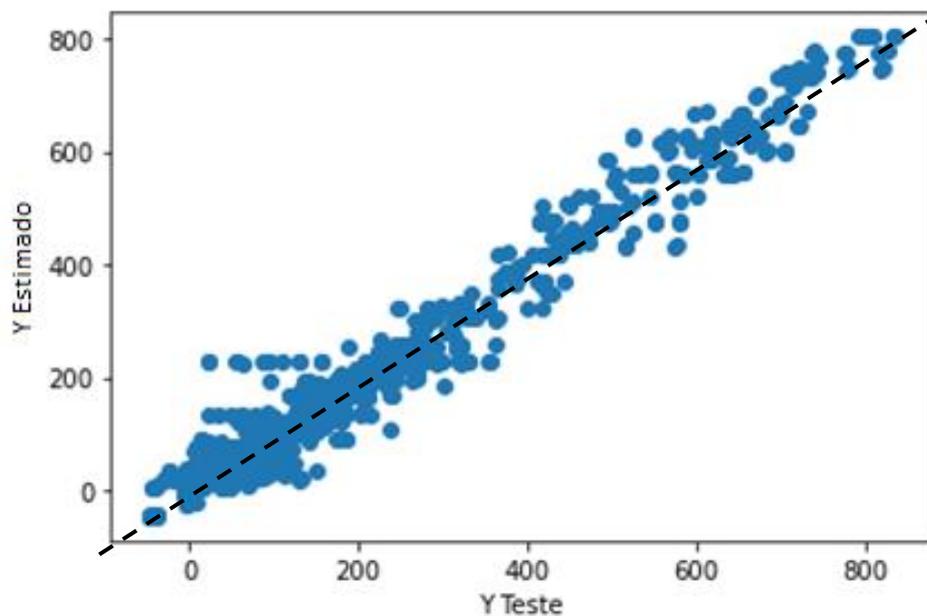
Os gráficos de dispersão para  $y$  real *versus*  $y$  estimado apresentam melhor visualização do comportamento dos algoritmos, visto que contam com um menor número de dados. Esses gráficos podem ser observados na Figura 27 e na Figura 28, para os algoritmos de árvore de decisão e para o algoritmo de florestas aleatórias, respectivamente.

Figura 27 – Dispersão entre dados reais e estimados por árvore de decisão



Fonte: Autora (2021)

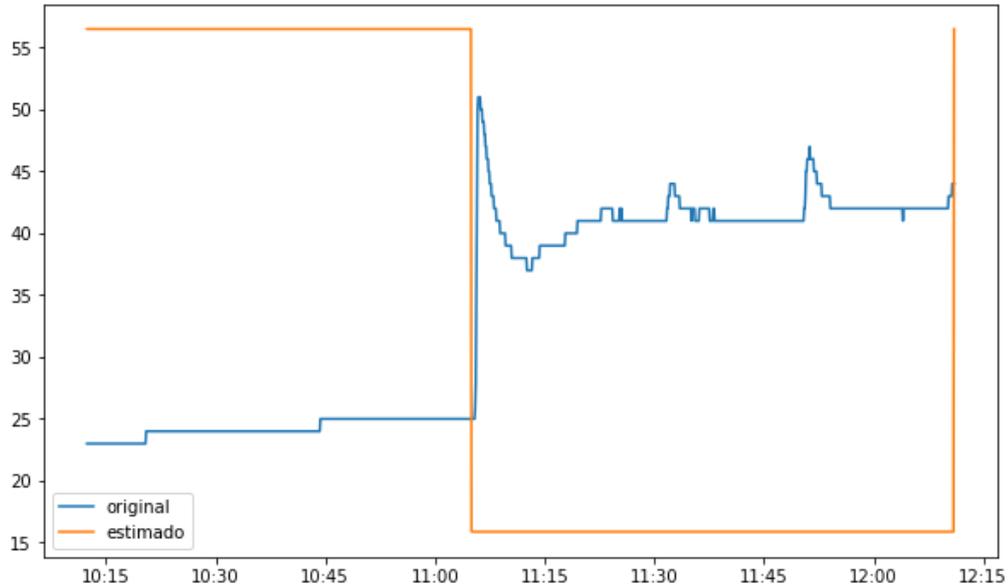
Figura 28 – Dispersão entre dados reais e estimados por florestas aleatórias



Fonte: Autora (2021)

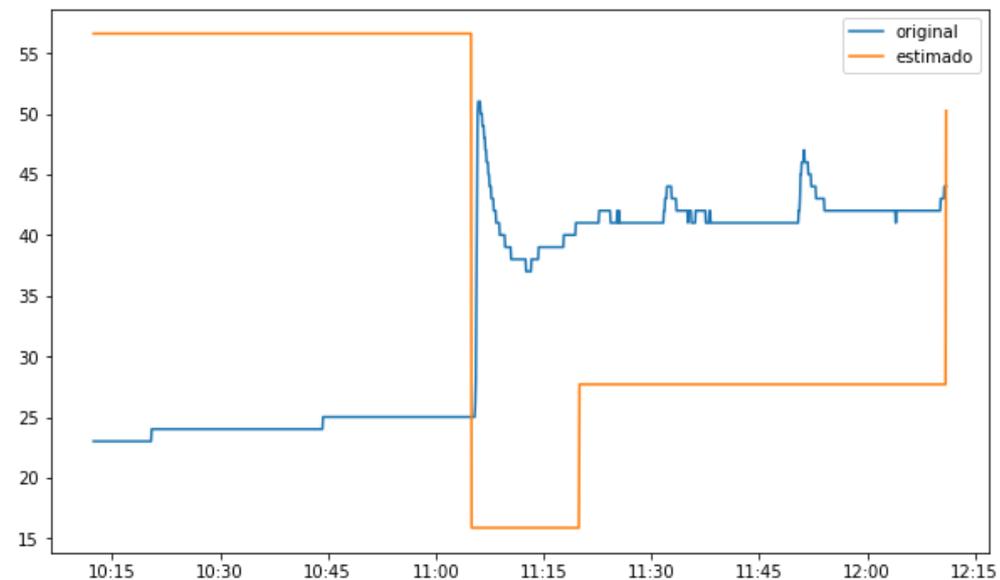
Nas Figuras 29 e 30, pode-se observar em azul o comportamento real do sétimo sensor de temperatura para a corrida experimental 15. Em laranja é apresentado o comportamento previsto para a variável, sendo na Figura 29 pelo algoritmo de árvore de decisão e na Figura 30 pelo algoritmo de florestas aleatórias.

Figura 29 – Comportamento previsto por árvore de decisão, para T7 da corrida experimental N°15



Fonte: Autora (2021)

Figura 30 – Comportamento previsto por florestas aleatórias, para T7 da corrida experimental N°15



Fonte: Autora (2021)

Com base na análise utilizando um número menor de dados, pode-se perceber que o método de florestas aleatórias, contando com 50 árvores de decisão, teve um desempenho ligeiramente melhor do que o método de árvore de decisão unitária. No entanto ainda apresentou erros muito próximos ao rodar o algoritmo com toda a base de dados. Essas informações demonstram que a base de dados é muito robusta, o que faz com que cada árvore de decisão apresente um bom resultado por si só, não apresentando grandes variações quando muitas árvores compõem uma floresta aleatória.

Conforme observado nos resultados dos métodos de árvore de decisão e florestas aleatórias, embora o valor do  $R^2$  esteja relativamente bom, o comportamento dessas variáveis estimadas por essas metodologias, quando plotados, apresentam comportamentos semelhantes a degraus e não condizem com os valores experimentais. Isso acontece uma vez que as metodologias utilizadas não são capazes de identificar e modelar a dinâmica do sistema. Ou seja, existe o fator de acúmulo em relação ao tempo que não é considerado pelo modelo, que existiria em sua modelagem fenomenológica através do balanço de massa e energia (equações diferenciais).

Para contornar essa dificuldade, é necessário utilizar um modelo que considere valores da saída passados como entrada do algoritmo (entradas exógenas), o qual será aplicado no próximo item.

### **5.2.2 Rede Autorregressiva Não-Linear com Entradas Exógenas**

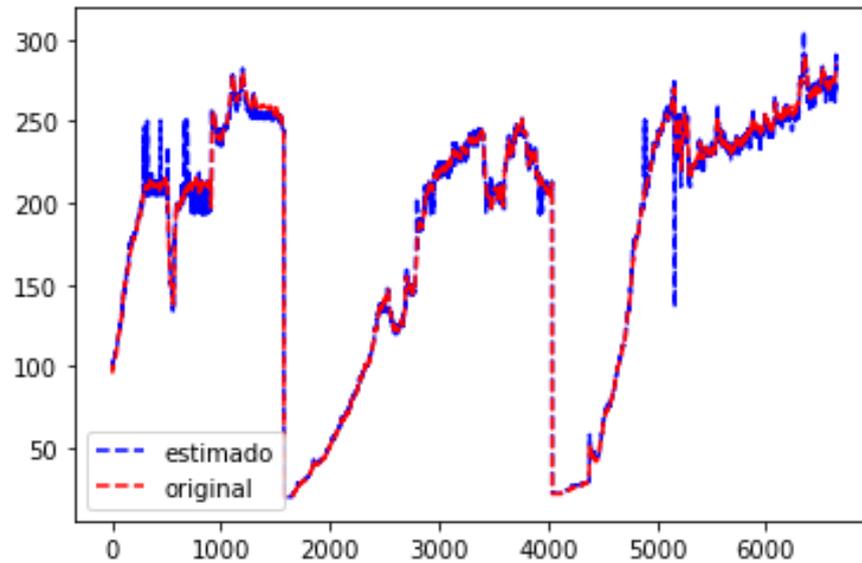
O método de rede autorregressiva não-linear com entradas exógenas não permite que sejam utilizados tantos parâmetros diferentes de entrada e saída. Por esse motivo, devem ser escolhidas 2 variáveis de entrada e 1 variável de saída. O número de neurônios ocultos para todas as análises foi 10. As corridas experimentais N°14 e N°15 foram deixadas fora da base de dados para treino, com objetivo de avaliar as previsões em corridas experimentais desconhecidas pelo algoritmo.

Primeiramente foram estudadas as temperaturas obtidas pelo sensor T5. Para isso, foram escolhidos como dados de entrada a frequência de alimentação e a vazão de ar *plenum*, pois ambas apresentam alta correlação direta com as temperaturas. O número de passos de previsão futura foi definido como 5. Foram obtidos os seguintes resultados:

Coeficiente de Determinação ( $R^2$ ) = 0.9952797526402795;  
Erro quadrático médio (MSE) = 29.567714346255308;  
Raiz do erro quadrático médio (RMSE)= 5.437620283382732.

Na Figura 31, pode-se observar em vermelho os valores de  $y$  utilizados para testar o algoritmo e em azul, os valores de  $y$  estimados a partir do algoritmo.

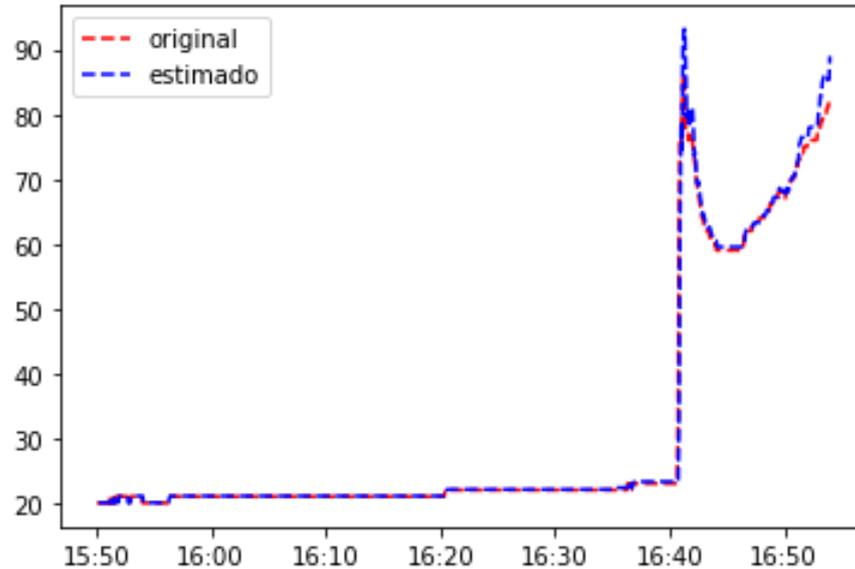
Figura 31 – Dados reais e estimados por NARX para T5, utilizando a frequência de alimentação e a vazão de ar *plenum* para treinar o algoritmo



Fonte: Autora (2021)

Na Figura 32, o algoritmo foi testado para a corrida experimental N°14, que não consta nos dados de treino. As temperaturas reais obtidas pelo sensor T5 podem ser observadas em vermelho, enquanto as estimadas pelo algoritmo estão em azul.

Figura 32 – Valores de temperatura reais e estimados por NARX para o sensor T5 da corrida experimental N°14



Fonte: Autora (2021)

Foram alterados os dados de entrada para vazão de ar *plenum* e resistência, já que ambas apresentam alta correlação inversa com as temperaturas. Os dados de saída continuam sendo as temperaturas obtidas pelo sensor T5. O número de passos à frente também foi definido como 5. Foram obtidos os seguintes resultados:

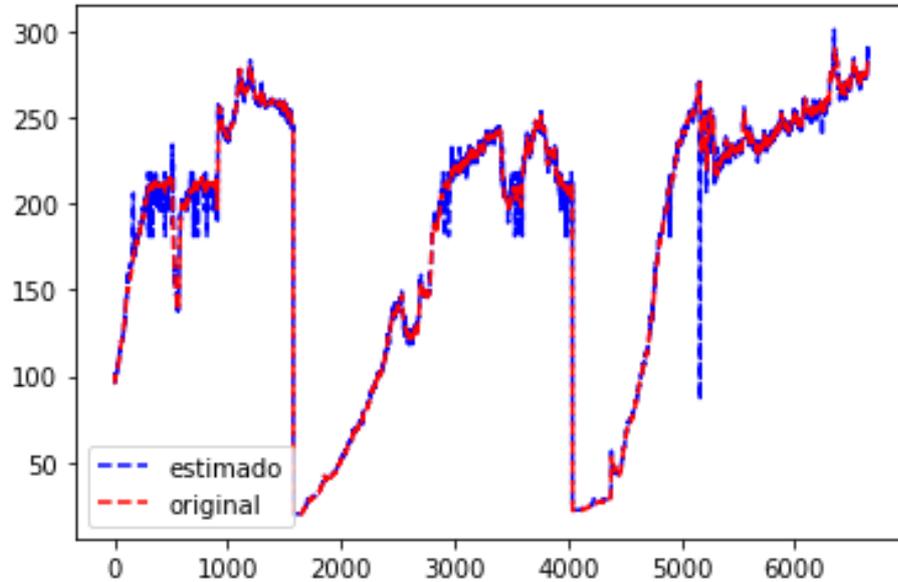
Coeficiente de Determinação ( $R^2$ ) = 0.9955322019225583;

Erro quadrático médio (MSE) = 28.256218503394926;

Raiz do erro quadrático médio (RMSE)= 5.315657861769785.

Na Figura 33, pode-se observar em vermelho os valores de  $y$  utilizados para testar o algoritmo e em azul, os valores de  $y$  estimados a partir do algoritmo.

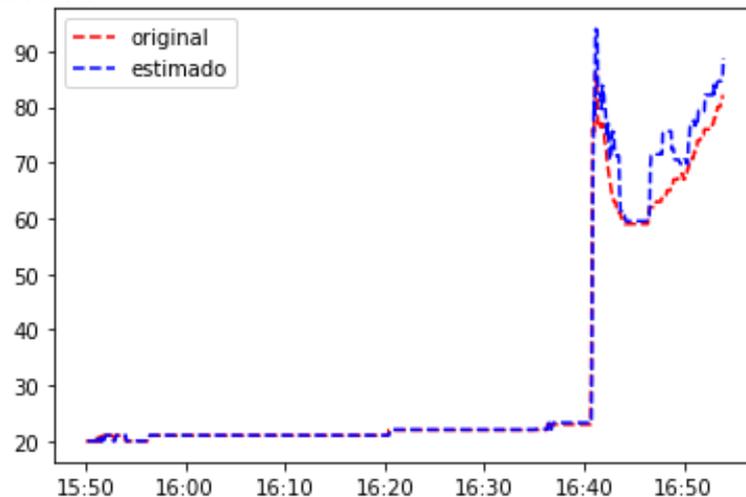
Figura 33 – Dados reais e estimados por NARX para T5, utilizando a operação da resistência e a vazão de ar *plenum* para treinar o algoritmo



Fonte: Autora (2021)

Na Figura 34, o algoritmo foi testado para a corrida experimental N°14 que não consta nos dados de treino. As temperaturas obtidas pelo sensor T5 podem ser observadas em vermelho, enquanto as estimadas pelo algoritmo estão em azul.

Figura 34 – Valores de temperatura reais e estimados por NARX para o sensor T5 da corrida experimental N°14



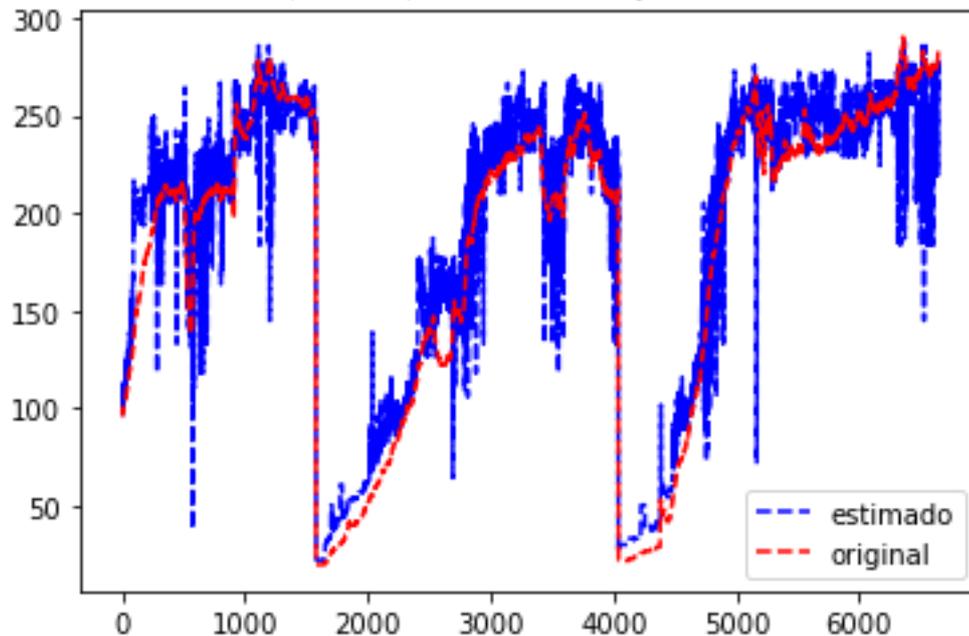
Fonte: Autora (2021)

Com objetivo de testar o desempenho do algoritmo estimando um grande número de passos à frente, foram mantidos os dados de entrada como vazão de ar *plenum* e resistência e a saída como temperaturas para T5. O número de passos de predições futuras foi alterado para 100. Os resultados foram os seguintes:

Coeficiente de Determinação ( $R^2$ ) = 0.912890469992163;  
 Erro quadrático médio (MSE) = 511.9085646140204;  
 Raiz do erro quadrático médio (RMSE) = 22.625396452085.

Na Figura 35, pode-se observar em vermelho os valores de  $y$  utilizados para testar o algoritmo e em azul, os valores de  $y$  estimados a partir do algoritmo.

Figura 35 – Dados reais e estimados por NARX para T5, utilizando a operação da resistência e a vazão de ar *plenum* para treinar o algoritmo



Fonte: Autora (2021)

A variação do Coeficiente de Determinação foi em torno de 8% ao se aumentar em 20 vezes o número de passos à frente executados pelo algoritmo. Isso demonstra que o algoritmo NARX é muito robusto para predição de temperaturas nesse processo e apresenta bons resultados para prever variações futura.

As pressões não apresentam altas correlações com os dados de entrada, mas estão diretamente correlacionadas com as temperaturas. Por esse motivo, o algoritmo foi rodado novamente, utilizando os primeiros dados de entrada (frequência de alimentação e a vazão de ar *plenum*), porém, alterando os dados de saída para as pressões obtidas pelo sensor P1. O número de passos à frente foi 1.

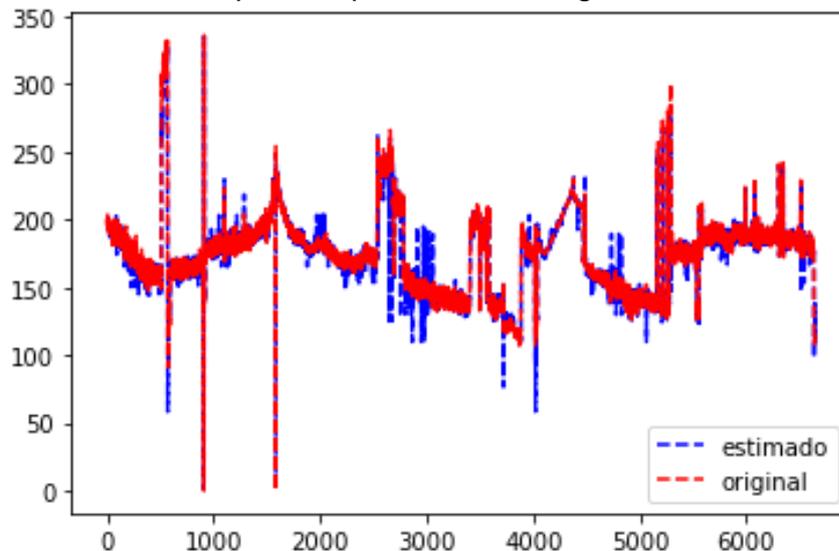
Coeficiente de Determinação ( $R^2$ ) = 0.8714878514383523;

Erro quadrático médio (MSE) = 109.65023378200567;

Raiz do erro quadrático médio (RMSE) = 10.471400755486616.

Na Figura 36, pode-se observar em vermelho os valores de  $y$  utilizados para testar o algoritmo e em azul, os valores de  $y$  estimados a partir do algoritmo.

Figura 36 – Dados reais e estimados por NARX para P1, utilizando a frequência de alimentação e a vazão de ar *plenum* para treinar o algoritmo



Fonte: Autora (2021)

O algoritmo também foi testado, alterando os dados de entrada para resistência e vazão *plenum*, mantendo os dados de saída como as pressões obtidas pelo sensor P1. O algoritmo foi rodado para diferentes números de passos à frente, com o objetivo de observar a variação do desempenho ao aumentar o número de predições futuras. Sendo  $n$  o número de passos à frente, foram obtidos os seguintes resultados:

Na Figura 37, pode-se observar em vermelho os valores de  $y$  originais utilizados para testar o algoritmo e em azul, os valores de  $y$  estimados a partir do algoritmo.

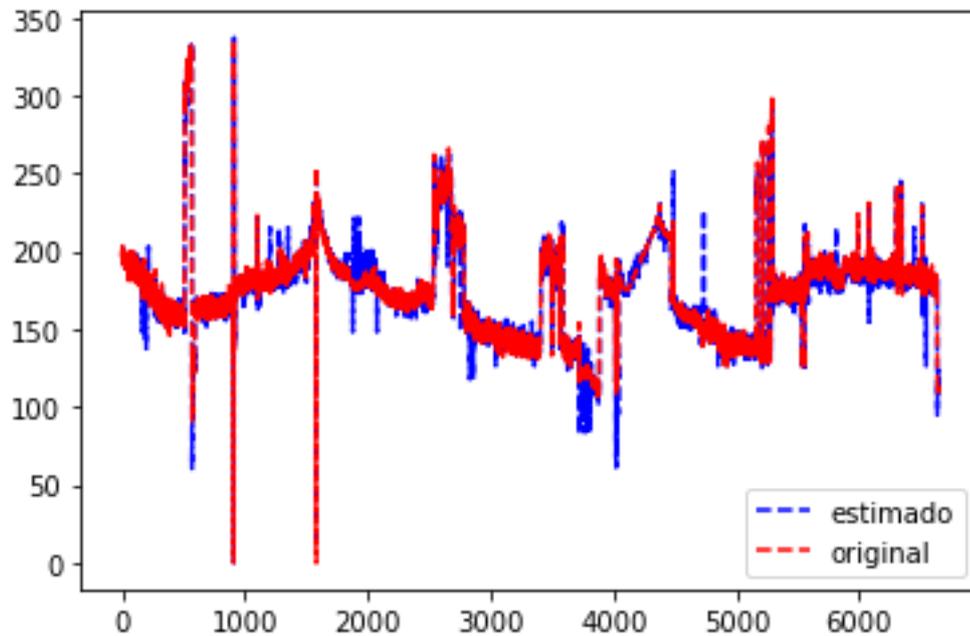
Resultados para  $n=1$ :

Coefficiente de Determinação ( $R^2$ ) = 0.8789933160208453;

Erro quadrático médio (MSE) = 121.88242374446514;

Raiz do erro quadrático médio (RMSE) = 11.040037307204408.

Figura 37 – Dados reais e estimados por NARX para P1, utilizando a operação da resistência e a vazão de ar *plenum* para treinar o algoritmo, com  $n=1$



Fonte: Autora (2021)

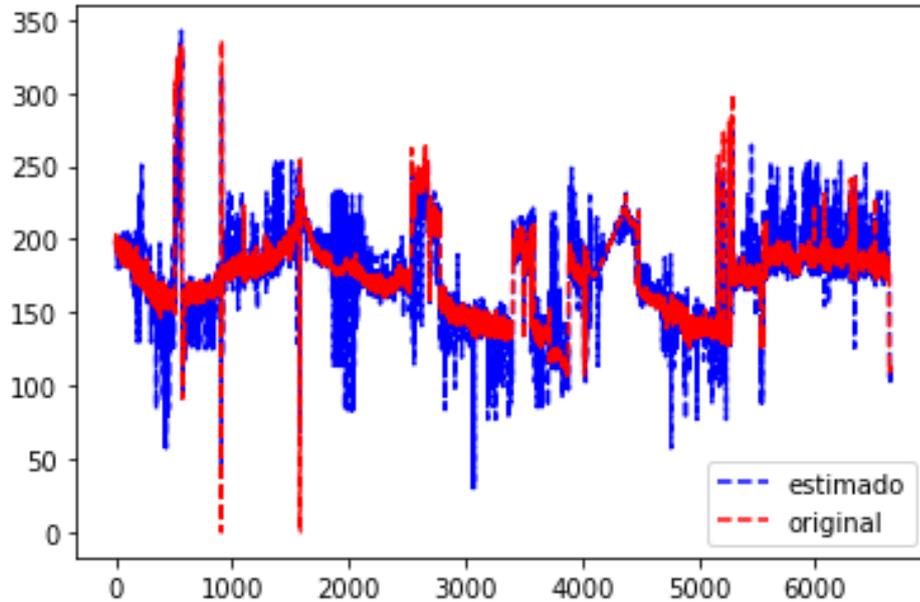
Na Figura 38, pode-se observar em vermelho os valores de  $y$  originais utilizados para testar o algoritmo e em azul, os valores de  $y$  estimados a partir do algoritmo. Resultados para  $n=10$ :

Coeficiente de Determinação ( $R^2$ ) = 0.6923856810175646;

Erro quadrático médio (MSE) = 275.2851360159756;

Raiz do erro quadrático médio (RMSE) = 16.591718898775245.

Figura 38 – Dados reais e estimados por NARX para P1, utilizando a operação da resistência e a vazão de ar *plenum* para treinar o algoritmo, com  $n=10$



Fonte: Autora (2021)

Na Figura 39, pode-se observar em vermelho os valores de  $y$  originais utilizados para testar o algoritmo e em azul, os valores de  $y$  estimados a partir do algoritmo. Resultados para  $n=50$ :

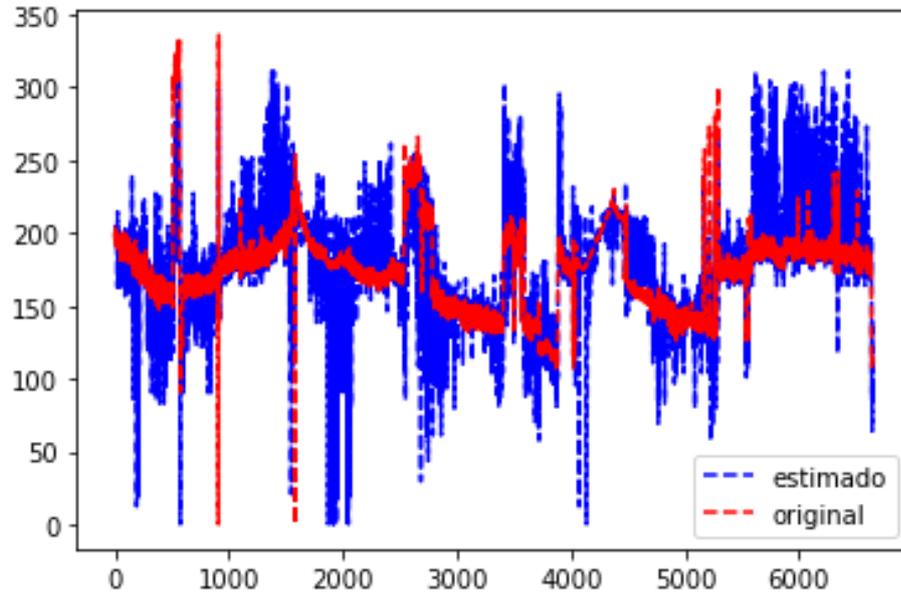
Para  $n=50$ :

Coefficiente de Determinação ( $R^2$ ) = 0.25998077988184254;

Erro quadrático médio (MSE) = 1031.4642279074203;

Raiz do erro quadrático médio (RMSE) = 32.116416797448316.

Figura 39 – Dados reais e estimados por NARX para P1, utilizando a operação da resistência e a vazão de ar *plenum* para treinar o algoritmo, com n=50



Fonte: Autora (2021)

Os resultados obtidos, ao se utilizar o método NARX para prever o comportamento das temperaturas, demonstrou um resultado muito satisfatório, com ajustes muito precisos, que conseguem prever a dinâmica e não-linearidade do processo vários passos à frente. O método apresentou bons resultados para diferentes dados de entrada. Isso pode ser explicado pelo fato de que esse algoritmo relaciona o valor atual de uma série temporal com os valores anteriores dessa série, fazendo com que consiga prever o comportamento das variáveis com o tempo. O método ainda apresentou uma baixa variação entre a predição de 5 e 100 passos, sendo a variação no  $R^2$  em torno de 8%.

Utilizando o mesmo método para prever o comportamento das pressões, os resultados obtidos foram menos satisfatórios, tanto para dados de entrada diretamente quanto inversamente correlacionados. O  $R^2$  foi menor que 0,9 para o primeiro passo à frente e decaiu severamente ao algoritmo ser ajustado para prever mais passos. Ainda sobre as pressões, os resultados, ao avançar alguns passos, mostraram-se melhores ao utilizar a resistência e vazão *plenum* como dados de entrada, que são inversamente correlacionados.

## 6 CONSIDERAÇÕES FINAIS

Os métodos de árvore de decisão e florestas aleatórias tiveram desempenhos muito semelhantes ao serem treinados e testados com toda a base de dados, gerando valores de  $R^2$  de 0.8263 e 0.8262, respectivamente. Os resultados obtidos para uma floresta aleatória, composta de 50 árvores de decisão, só se tornaram melhores que os resultados obtidos para uma única árvore de decisão ao utilizar uma base de dados 30 vezes menor, onde os valores de  $R^2$  foram 0.8525 para florestas aleatórias e 0.8432 para árvore de decisão. Isso pode ser explicado por conta do tamanho da base de dados original, que garante uma grande quantidade de dados de treino e faz com que cada árvore de decisão seja muito robusta por si só.

Fica ainda mais evidente, ao se analisar os gráficos em função do tempo, que os métodos de árvore de decisão e florestas aleatórias não atingem um bom desempenho na predição de processos dinâmicos e não-lineares, apresentando previsões lineares e distantes dos valores reais.

O método de rede autorregressiva não-linear com entradas exógenas apresentou resultados muito positivos ao estimar as temperaturas, com  $R^2$  de 0.9955 para 5 passos à frente e  $R^2$  de 0.9128 para 100 passos à frente. Isso valida a escolha dos dados de entrada e mostra que o algoritmo NARX, treinado para predição de temperaturas nesse processo, é muito robusto e apresenta bons resultados para prever variações futuras.

Já para a predição das pressões, o método NARX demonstrou resultados menos satisfatórios. Isso pode ser explicado devido às variáveis que compõem os dados de entrada do processo não apresentarem altas correlações com os dados de pressão. O método apresentou um desempenho um pouco melhor ao longo dos passos, ao serem utilizados a resistência e a vazão *plenum* como dados de entrada, que são inversamente proporcionais.

O método NARX apresenta como ponto positivo a capacidade de prever bons resultados em processos dinâmicos não-lineares, utilizando seus neurônios estruturados em redes neurais. Esse método consegue guardar informações anteriores sobre o sistema, já que em processos dinâmicos não-lineares, a saída em um dado instante de tempo depende não apenas de suas entradas atuais, mas também do comportamento anterior do sistema.

O ponto negativo desse método é que ele aceita poucas variáveis de entrada e saída, o que torna obrigatória uma extensa análise de correlação para obter boas previsões. Também faz necessário que sejam gerados um algoritmo para cada variável de saída a ser estudada, já que só aceita vetores unitários armazenados na variável  $y$ .

## **SUGESTÕES PARA TRABALHOS FUTUROS**

- Utilizar dados de experimentos de gaseificação para fazer uma Análise de Componente Principal, objetivando a redução de dimensionalidade.
- Utilizar aprendizado de máquina para a prever possíveis falhas de operação no processo do gaseificador do LEC.

## REFERÊNCIAS

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Parte III fontes não-renováveis: derivados de petróleo**. 3. ed. Brasília: ANEEL, 2008. Disponível em: [http://www2.aneel.gov.br/arquivos/pdf/atlas\\_par3\\_cap9.pdf](http://www2.aneel.gov.br/arquivos/pdf/atlas_par3_cap9.pdf). Acesso em: 10 ago. 2020.

ANDRADE, Rubenildo. **Gaseificação de biomassa: uma análise teórica e experimental**. 2007. 150 f. Tese (Doutorado em Engenharia Mecânica) - Universidade Federal de Itajubá, Itajubá, 2007. Disponível em: <https://repositorio.unifei.edu.br/xmlui/handle/123456789/1720>. Acesso em: 28 set. 2020.

ANSARI, Amir. **Developing a smart proxy for fluidized bed using machine learning**. 2016. 90 f. Thesis (Masters of Science in Petroleum and Natural Gas Engineering) - West Virginia University, Morgantown, 2016. Disponível em: <http://search.proquest.com/docview/1865638873/?pq-origsite=primo>. Acesso em: 20 ago. 2020.

ASYURA, Siti; ABDUL, Suhairi; AZIZ, Norashid. Nonlinear autoregressive with exogenous inputs based model predictive control for batch citronellyl laurate esterification reactor. **Advanced Model Predictive Control**, Malaysia, v. 1, n. 1992, p. 267–290, 2011. DOI: 10.5772/16963. Disponível em: <http://www.intechopen.com/books/advanced-model-predictive-control/nonlinear-autoregressive-with-exogenous-inputs-based-model-predictive-control-for-batch-citronellyl->. Acesso em: 12 out. 2020.

BASU, Prabir. Combustion and gasification in fluidized beds. **Journal of Chemical Information and Modeling**, Halifax, v. 53, n. 9, p. 1689–1699, 2013.

BASU, Prabir; KAUSHAL, Priyanka. Modeling of pyrolysis and gasification of biomass in fluidized beds: a review. **Chemical Product and Process Modeling**, Halifax, v. 4, n. 1, p. 45, 2009. DOI: 10.2202/1934-2659.1338.

BROWN, David; FUCHINO, Tetsuo; MARÉCHAL, François. Solid fuel decomposition modelling for the design of biomass gasification systems. **Computer Aided Chemical Engineering**, Tokyo, v. 21, p. 1661–1666, 2006. DOI: 10.1016/S1570-7946(06)80286-5. Disponível em: <https://core.ac.uk/download/pdf/191966438.pdf>. Acesso em: 15 out. 2020.

BROWN, M. T. **Caracterização petrológica e química dos carvões utilizados em usinas termoeletricas brasileiras e as cinzas no processo de combustão**. 2011. 111 f. Dissertação (Mestrado em Geociências) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

COLLAZZO, Gabriela. **Gaseificação de carvões nacionais com vapor d'água visando a produção de gás de síntese**. 2013. 149 f. Tese (Doutorado em Engenharia Química) - Universidade Federal de Santa Catarina, Florianópolis, 2013. Disponível em: <https://repositorio.ufsc.br/handle/123456789/122892>. Acesso em: 25 out. 2020.

COLLOT, Anne Gaëlle. Matching gasification technologies to coal properties. **International Journal of Coal Geology**, London, v. 65, n. 3–4, p. 191–212, 2006. DOI: 10.1016/j.coal.2005.05.003. Disponível em: <https://bitly.com/vKqrh>. Acesso em: 20 set. 2020.

CORRÊA, Fernanda. **Aplicação de redes neurais artificiais no setor de transporte do Brasil**. 2008. 84 f. Dissertação (Mestrado em Ciências Exatas e da Terra) - Universidade Federal de São Carlos, São Carlos, 2008. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/4257?show=full>. Acesso em: 18 out. 2020.

DA SILVA, Dagoberto Cássio. **Análise termoeconômica comparativa de ciclos avançados de usinas termoeletricas a carvão mineral e usinas termonucleares**. 2016. 207 f. Tese (Doutorado em Engenharia Mecânica) - Universidade Federal de Itajubá, Itajubá, 2016. Disponível em: <https://bitly.com/p2iA9>. Acesso em: 20 out. 2020.

DEBIAGI, Paulo Eduardo; NICOLAI, Hendrik; HAN, Wang; JANICKA, Johannes.; HASSE, Christian. Machine learning for predictive coal combustion cfd simulations - from detailed kinetics to HDMR reduced-order models. **Fuel**, Darmstadt, v. 274, n. April, p. 1-12, 2020. DOI: 10.1016/j.fuel.2020.117720. Disponível em: <https://bitly.com/7EhzN>. Acesso em: 15 out. 2020.

DEPOI, Fernanda; POZEBON, Dirce; KALKREUTH, Wolfgang. Chemical characterization of feed coals and combustion-by-products from Brazilian power plants. **International Journal of Coal Geology**, Porto Alegre, v. 76, n. 3, p. 227–236, 2008. DOI: 10.1016/j.coal.2008.07.013. Disponível em: <https://bitly.com/YwDun>. Acesso em: 12 out. 2020.

DHUSSA, Anil; SAMBI, Surinder; KUMAR, Shashi; KUMAR, Sandeep; KUMAR, Surendra. Nonlinear autoregressive exogenous modeling of a large anaerobic digester producing biogas from cattle waste. **Bioresource Technology**, Uttarakhand, v. 170, p. 342–349, 2014. DOI: 10.1016/j.biortech.2014.07.078. Disponível em: <https://bitly.com/5uPQ7>. Acesso em: 10 out. 2020.

DIAS, Yan. **Dessulfurização do gás oriundo da gaseificação do carvão mineral a partir da adsorção por cinzas volantes**. 2016. 59 f. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Química) - Universidade Federal do Pampa, Bagé, 2016.

ELMAZ, Furkan; YÜCEL, Özgün; MUTLU, Ali Yener. Predictive modeling of biomass gasification with machine learning-based regression methods. **Energy**, Izmir, v. 191, p. 116541, 2020. DOI: 10.1016/j.energy.2019.116541. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0360544219322364>. Acesso em: 23 out. 2020.

PEDREGOSA, Fabian, VAROQUAUX, Gaël, GRAMFORT, Alexandre, MICHEL, Vincent, THIRION, Bertrand, GRISEL, Olivier, BLONDEL, Mathieu, PRETTENHOFER, Peter, WEISS, Ron, DUBOURG, Vincent, VANDERPLAS, Jake, PASSOS, Alexandre, COURNAPEAU, David, BRUCHER, Matthieu, PERROT, Matthieu, DUCHESNAY, Édouard. Scikit-learn: Machine Learning in Python, **Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011. Disponível em: <https://jmlr.org/papers/v12/pedregosa11a.html>. Acesso em 10 nov. 2020.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems**. Sebastopol: O'Reilly Media, 2017.

GRÄBNER, Martin. **Industrial coal gasification technologies covering baseline and high-ash coal**. Weinheim: Wiley-VCH, 2015.

GUO, Rong; GUO, Weiwei; HU, Haijun. Texaco coal gasification quality prediction by neural estimator based on Msa and dynamic PCA. *In*: INTERNATIONAL CONFERENCE ON AUTOMATION AND LOGISTICS. 1., Luoyang, 2008. **Proceedings**. Luoyang: IEEE, 2008. p. 2241–2246. DOI: 10.1109/ICAL.2008.4636353. Disponível em: <https://ieeexplore.ieee.org/document/4026446>. Acesso em: 18 nov. 2020.

HALLAK, Ricardo; PEREIRA FILHO, Augusto. Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de são paulo com o modelo arps: sensibilidade a variações com os esquemas de advecção e assimilação de dados. **Revista Brasileira de Meteorologia**, São Paulo, v. 26, n. 4, p. 591–608, 2011. Disponível em: <https://bitly.com/tHPyR>. Acesso em: 1 out. 2020.

HANSON, Robin; STUTZ, John; CHEESEMAN, Peter. Bayesian classification theory. **Technical Report FIA-90-12-7-01**, Moffet Field, p. 1-6, 1991. DOI: 10.1.1.69.8701. Disponível em: <http://hdl.handle.net/2060/19920017651>. Acesso em: 20 nov. 2020.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning**. Stanford, CA: Springer, 2008. v. 2 DOI: 10.1007/978-1-4419-9863-7\_941. Disponível em: <https://web.stanford.edu/~hastie/Papers/ESLII.pdf>. Acesso em: 26 mar. 2021.

HAYKIN, Simon. **Redes neurais princípios e práticas**. 2. ed. Porto Alegre: Bookman, 2001.

HIGMAN, Christopher; BURGT, Marteen van der. **Gasification**. 2. ed. Burlington, USA: Gulf Professional Publishing, 2008.

HIGMAN, Christopher.; BURGT, Marteen van der. **Gasification**. USA: Gulf Professional Publishing, 2003.

HOELSCHER, Fernanda. **Pirólise do carvão mineral: simulação e experimento**. 2014. 89 f., Trabalho de Conclusão de Curso (Bacharelado em Engenharia Química) - Universidade Federal do Pampa, Bagé, 2014.

HOFFMANN, Bettina. Susanne. **O ciclo combinado com gaseificação integrada e captura de CO<sub>2</sub>**: uma solução para mitigar as emissões de CO<sub>2</sub> em termelétricas a carvão em larga escala no curto prazo. 2010. 128 f. Dissertação (Mestrado em Ciências em Planejamento Energético) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

HUNTER, John D. Matplotlib: a 2D graphics environment. **Computing Science & Engineering**, [S. L.], v. 9, p. 90-95, 2007. DOI:10.1109/MCSE.2007.55. Disponível em: <https://matplotlib.org/>. Acesso em: 5 nov. 2020.

JÜNTGEN, Harald. Reactivities of carbon to steam and hydrogen and applications to technical gasification processes. **Carbon**, Baden-Baden, v. 19, p. 167–173, 1981. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/0008622381900397>. Acesso em: 27 out. 2020.

KLEIN, Jefferson. Transgas irá investir US\$ 2,7 bilhões em Candiota. **Jornal do Comércio**, Porto Alegre, 26 mar. 2015. Disponível em: <https://www.jornaldocomercio.com/site/noticia.php?codn=191738>. Acesso em: 30 abr. 2021.

LANTZ, Brett. **Machine learning with R**. Birmingham: Packt Publishing Ltd., 2013.

LIMA JÚNIOR, Gladson Euler. **Dissertação metamodelagem kriging dinâmica aplicada em trocadores de calor**. 2019. 94 f. Dissertação (Mestrado em Engenharia Química) - Universidade Federal de Campina Grande, Campina Grande, 2019.

LIPORACE, Frederico dos Santos. Um sistema neural para monitoração do desflorestamento na região Amazônica utilizando imagens do Landsat. **Technical Report CCR-157, IBM, Rio Scientific Center**, Rio de Janeiro, v. 1, n. 1, p. 38, 1994.

LORA, Electro; ANDRADE, Rubenildo Vieira; MARTINEZ, Juan Daniel; LEITE, Marco Antonio Haikal; ROCHA, Mateus Henrique; SALES, Cristina Vilas Bôas; Mendonza, Mônica Andrea Gualdrón; CORAL, Doris del Socorro Obando. Gaseificação e pirólise para a conversão de biomassa em eletricidade e biocombustíveis. In: LORA, E. E. S.; VENTURINI, O. J. (org.). **Biocombustíveis**. Rio de Janeiro: Interciência, 2012. v. 1, p. 411–498. Disponível em: <https://bityli.com/wuGYt>. Acesso em: 14 out. 2020.

LUNKES, Maiquel Kochhann. **O carvão das camadas superiores e inferiores da jazida de Candiota – RS**: uma contribuição para sua caracterização química, petrológica e tecnológica visando seu potencial na geração de energia. 2010. 83 f. Trabalho de Conclusão de Curso (Bacharelado em Geologia) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/28616/000771591.pdf?sequence=1&isAllowed=y>. Acesso em: 8 nov. 2020.

MAGNUS, Douglas Matos; ABELARDO TAVARES, André; SALVARO, Pedro Vitor. Análise técnico-econômica da modernização do parque termelétrico a carvão mineral com tecnologias de baixas emissões de CO<sub>2</sub>. *In*: CONGRESSO BRASILEIRO DE CARVÃO MINERAL. 5., Criciúma, 2017. **Anais [...]**. Criciúma: CBCM, 2017. DOI: 10.20906/cps/cbcm2017-0041. Disponível em: <https://bit.ly/3eREpX6>. Acesso em: 4 nov. 2020.

MCKINNEY, Wes. **Data Structures for Statistical Computing in Python**, *In*: PYTHON IN SCIENCE CONFERENCE. 9. Texas, 2008. **Proceedings**. Texas: LLC, 2010. p. 51-56. Disponível em: <https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>. Acesso em: 20 out. 2020.

MITCHELL, Tom. **Machine learning**. New York: McGraw-Hill, 1997.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. *In*: REZENDE, Solange Oliveira (org.). **Sistemas inteligentes - fundamentos e aplicações**. Barueri: Manole, 2008. p. 39–56.

NIECKEL, Bianka; FURTADO, André. Potencial de gaseificação do carvão gaúcho é avaliado por investidores internacionais. **Portal do Estado do Rio Grande do Sul**, Porto Alegre, 29 dez. 2014. Disponível em: <https://estado.rs.gov.br/potencial-de-gaseificacao-do-carvao-gaicho-e-avaliado-por-investidores-internacionais>. Acesso em: 30 abr. 2021.

HARRIS, Charles R. MILLMAN, K. Jarrod, VAN DER WALT, Stéfan J., GOMMERS, Ralf, VIRTANEN, Pauli, COURNAPEAU, David, WIESER, Eric, TAYLOR, Julian, BERG, Sebastian, SMITH, Nathaniel J., KERN, Robert, PICUS, Matti, HOYER, Stephan, VAN KERKWIJK, Marten H., BRETT, Matthew, HALDANE, Allan, FÉRNANDEZ, Jaime, WIEBE, Mark, PETERSON, Pearu, GÉRARD-MARCHANT, Pierre, SHEPPARD, Kevin, REDDY, Tyler, WECKESSER Warren, ABBASI, Hameer, GOHLKE, Christoph, OLIPHANT, Travis E. Array programming with NumPy, **Nature**, v. 585, p. 357–362 (2020). Disponível em: <https://www.nature.com/articles/s41586-020-2649-2>. Acesso em 08/04/2021.

PELLEGRINO, Roberto. **Gaseificação de carvão mineral com adição de vapor e remoção de H<sub>2</sub>S, em leito fluidizado**. 2006. 127 f. Dissertação (Mestrado em Engenharia Mecânica) - Universidade Estadual de Campinas, Campinas, 2006. Disponível em: [http://repositorio.unicamp.br/bitstream/REPOSIP/264140/1/Pellegrino\\_Roberto\\_M.pdf](http://repositorio.unicamp.br/bitstream/REPOSIP/264140/1/Pellegrino_Roberto_M.pdf). Acesso em: 18 out. 2020.

PIRES, Marçal; QUEROL, Xavier. Characterization of Candiota (South Brazil) coal and combustion by-product. **International Journal of Coal Geology**, [S. l.], v. 60, n. 1, p. 57–72, 2004. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0166516204000953>. Acesso em: 8 nov. 2020.

PYTHON SOFTWARE FOUNDATION. **Python Language Site**: Documentation, 2020. Disponível em: <https://www.python.org/doc/>. Acesso em: 06 nov. 2020.

QUINLAN, John Ross. Induction of decision trees. **Machine Learning** 1, Sydney, v. 1, n. 1, p. 81–106, 1986. DOI: 10.1007/bf00116251. Disponível em: <https://link.springer.com/article/10.1007/BF00116251#citeas>. Acesso em: 30 out. 2020.

REZAIYAN, John; CHEREMISINOFF, Nicholas. Gasification technologies: a primer for engineers and scientists. Florida: CRC Press, 2005.

ROCHA, Melissa. **Simulação de um reator de combustão e gaseificação através de uma rede de reatores equivalentes baseada em fluidodinâmica computacional**. 2016. 100 f. Dissertação (Mestrado em Engenharia Química) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016. Disponível em: <https://lume.ufrgs.br/handle/10183/149181>. Acesso em: 7 nov. 2020.

RODRIGUES, Filipe César. **Análise experimental e computacional da fluidodinâmica de um gaseificador de leito fluidizado borbulhante**. 2017. 60 f. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Química). Universidade Federal do Pampa, Bagé, 2017.

RODRIGUES, Rodolfo. **Modelagem cinética e de equilíbrio combinadas para simulação de processos de gaseificação**. 2015. 105 f. Tese (Doutorado em Engenharia Química) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015.

SÁ, Fernando; RAUPP, Iris; MUNIZ, Ana Rosa. Influência da granulometria do carvão na gaseificação em leito fluidizado borbulhante. *In: SALÃO INTERNACIONAL E ENSINO, PESQUISA E EXTENSÃO*. 8., Bagé, 2017. **Anais [...]**. Bagé: UNIPAMPA, 2017. p. 1-2, v. 9, n. 5. Disponível em: <https://periodicos.unipampa.edu.br/index.php/SIEPE/article/view/85977#:~:text=O%20carv%C3%A3o%20mineral%20de%20Candiota,50%20como%20agente%20fluidizante%20inerte>. Acesso em: 03 mai. 2021

SÁ, Fernando. **Modelagem e simulação de um gaseificador de leito fluidizado borbulhante através de rede de reatores equivalentes**. 2018. 59 f. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Química) - Universidade Federal do Pampa, Bagé, 2018.

SEALU, Lisiane Priscila. **Redes neurais artificiais no contexto estatístico**. 2000. 108 f. Trabalho de conclusão de Curso (Bacharelado em Estatística) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

SHALEV-SHWARTZ, Shai; BEN-DAVID, Shai. **Understanding machine learning: from theory to algorithms**. New York: Cambridge University Press, 2014. v. 1

## APÊNDICES

### APÊNDICE A - Código desenvolvido em Python

```

1  import pandas as pd
2  import numpy as np
3  import matplotlib.pyplot as plt
4  import pickle
5  import seaborn as sns
6  from sklearn.tree import DecisionTreeRegressor
7  from sklearn.model_selection import train_test_split
8  from datetime import datetime
9  from sklearn import metrics
10 import matplotlib.pyplot as plt
11 from matplotlib.dates import DateFormatter
12 from sklearn.ensemble import RandomForestRegressor
13 from sklearn.preprocessing import scale
14 from fireTS.models import NARX, DirectAutoRegressor
15 from sklearn.ensemble import RandomForestRegressor
16 from xgboost import XGBRegressor
17 from sklearn.metrics import r2_score
18
19 ent_file =
20 ['28-06-2016.xlsx', '29-07-2016.xls', '01-08-2016.xlsx', '26-09-2016.xlsx', '17_11_2016.xls',
21  '30-11-2016.xlsx', '24.03.2017.xlsx', '12-05-2017.xlsx', '24-05-2017.xlsx', '20.09.2017.xlsx',
22  '03_11_2017.xls', '10-11-2017.xlsx', '27.04.2018.xlsx', '19.11.2020.xlsx', '11-12-2020.xlsx']
23 temp_file =
24 ['Temperatura_28_06_2016', 'Temperatura_29_07_2016', 'Temperatura_01_08_2016', 'Temperatura_26_09_2016',
25  'Temperatura_17_11_2016', 'Temperatura_30_11_2016', 'Temperatura_24_03_2017', 'Temperatura_12_05_2017',
26  'Temperatura_24_05_2017', 'Temperatura_20_09_2017', 'Temperatura_03_11_2017', 'Temperatura_10_11_2017',
27  'Temperatura_27_04_2018', 'Temperatura_19_11_2020', 'Temperatura_11_12_2020']
28 pr_file =
29 ['Pressão_28_06_2016', 'Pressão_29_07_2016', 'Pressão_01_08_2016', 'Pressão_26_09_2016',
30  'Pressão_17_11_2016', 'Pressão_30_11_2016', 'Pressão_24_03_2017', 'Pressão_12_05_2017', 'Pressão_24_05_2017',
31  'Pressão_20_09_2017', 'Pressão_03_11_2017', 'Pressão_10_11_2017', 'Pressão_27_04_2018', 'Pressão_19_11_2020',
32  'Pressão_11_12_2020']
33
34 corr = ['01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12', '13', '14', '15']
35
36 df = {}
37 for i in range(0, len(corr)):
38     print('corrida', corr[i])
39     ent1 = pd.read_excel('Desktop/data/'+ent_file[i], header=0)
40     temp1 = pd.read_excel('Desktop/data/'+temp_file[i]+'.xls', header=0)
41     pr1 = pd.read_excel('Desktop/data/'+pr_file[i]+'.xls', header=0)
42
43     var = {'ent1': ent1, 'temp1': temp1, 'pr1': pr1}
44     ini, fim = [], []
45     for v in var:
46         var[v].dropna(axis=1, inplace=True)
47         var[v]['Data'] = var[v]['Data'].astype(str)
48         var[v]['Hora'] = var[v]['Hora'].astype(str)
49
50         var[v]['Timestamp'] = pd.to_datetime(var[v]['Data'] + ' ' + var[v]['Hora'], dayfirst=True) #, format='%Y-%m-%d %H:%M:%S')
51
52     var[v].dropna(inplace=True)
53     var[v].set_index('Timestamp', inplace=True) # seta o timestamp como index

```

```

43
44     keys = var[v].keys()
45     for k in range(0,len(keys)):
46         var[v][keys[k]] = pd.to_numeric(var[v][keys[k]], errors='coerce')
47
48     var[v] = var[v].resample("5S").mean().fillna(method='ffill')
49     ini,fim = np.append(ini,var[v].index[0]), np.append(fim,var[v].index[-1])
50
51     ini,fim = np.max(ini), np.min(fim)
52
53     for v in var:
54         var[v] = var[v][(var[v].index >= ini) & (var[v].index <= fim)]
55
56         df_run = pd.concat([var['ent1'], var['temp1'], var['pr1']], axis=1,
57                             join='inner')
58     df[corr[i]] = df_run
59
60
61     dff =
62     pd.concat([df['01'],df['02'],df['04'],df['05'],df['06'],df['07'],df['08'],df['09']
63               ,df['10'],df['11'],df['13']])
64
65     for i in dff:
66         dff= dff[['Resistência', 'Vazão Plenum', 'Vazão Freeboard', 'Pulso',
67                 'Alimentação (Hz)', 'Alimentação janela (L)',
68                 'Centrífuga', 'T 1','T 2', 'T 3', 'T 4', 'T 5', 'T 6', 'T 7', 'T
69                 8', 'T 9', 'T 10', 'P 1', 'P 2', 'P 3',
70                 'P 4', 'P 5', 'P 6']]
71
72     X1 = np.c_[dff["Resistência"], dff["Vazão Plenum"], dff["Vazão Freeboard"],
73               dff["Pulso"], dff["Alimentação (Hz)"],
74               dff["Alimentação janela (L)"], dff["Centrífuga"]]
75
76     y1 = np.c_[dff["T 1"], dff["T 2"], dff["T 3"], dff["T 4"], dff["T 5"], dff["T 6"],
77               dff["T 8"], dff["P 1"], dff["P 2"],
78               dff["P 3"], dff["P 4"], dff["P 5"], dff["P 6"], dff["T 9"], dff["T 10"]]
79
80     X1_train, X1_test, y1_train, y1_test = train_test_split(X1, y1, test_size=0.10)
81
82     regr_1 = DecisionTreeRegressor()
83     regr_1.fit(X1_train, y1_train)
84     y_pred1 = regr_1.predict(X1_test)
85
86     scl= regr_1.score(X1_test, y1_test)
87
88     metrics.mean_absolute_error(y1_test,y_pred1)
89     metrics.mean_squared_error(y1_test, y_pred1)
90     np.sqrt(metrics.mean_squared_error(y1_test, y_pred1)))
91
92     rfc1 = RandomForestRegressor(n_estimators=50)
93     rfc1.fit(X1_train,y1_train)
94     y_pred3 = rfc1.predict(X1_test)
95
96     sc=rfc1.score(X1_test,y1_test)
97
98     metrics.mean_absolute_error(y1_test,y_pred3)
99     metrics.mean_squared_error(y1_test, y_pred3)
100     np.sqrt(metrics.mean_squared_error(y1_test, y_pred3)))
101
102     rfc1 = RandomForestRegressor(n_estimators=50)
103     rfc1.fit(X1_train,y1_train)
104     y_pred3 = rfc1.predict(X1_test)
105
106     sc=rfc1.score(X1_test,y1_test)
107
108     metrics.mean_absolute_error(y1_test,y_pred3)
109     metrics.mean_squared_error(y1_test, y_pred3)
110     np.sqrt(metrics.mean_squared_error(y1_test, y_pred3)))

```

```

104 X2 = np.c_[dff["Resistência"], dff["Vazão Plenum"], dff["Vazão Freeboard"],
105           dff["Pulso"], dff["Alimentação (Hz)"],
106           dff["Alimentação janela (L)"], dff["Centrífuga"]]
107 y2 = np.c_[dff["T 1"], dff["T 2"], dff["T 3"], dff["T 4"], dff["T 5"], dff["T 6"],
108           dff["T 8"], dff["P 1"], dff["P 2"],
109           dff["P 3"], dff["P 4"], dff["P 5"], dff["P 6"], dff["T 9"], dff["T 10"]]
110 X2_train, X2_test, y2_train, y2_test = train_test_split(X2, y2, train_size=0.03,
111               test_size=0.003)
112 regr_2 = DecisionTreeRegressor()
113 regr_2.fit(X2_train, y2_train)
114
115 y_pred22 = regr_2.predict(X2_test)
116
117 sc22= regr_2.score(X2_test, y2_test)
118
119 metrics.mean_absolute_error(y2_test,y_pred22))
120 metrics.mean_squared_error(y2_test, y_pred22))
121
122 np.sqrt(metrics.mean_squared_error(y2_test, y_pred22)))
123
124 rfc2 = RandomForestRegressor(n_estimators=50)
125 rfc2.fit(X2_train,y2_train)
126 y_pred222 = rfc2.predict(X2_test)
127
128 sc222=rfc2.score(X2_test,y2_test)
129
130 metrics.mean_absolute_error(y2_test,y_pred222))
131 metrics.mean_squared_error(y2_test, y_pred222))
132 np.sqrt(metrics.mean_squared_error(y2_test, y_pred222)))
133
134 dff[['Vazão Plenum', 'Alimentação (Hz)']].shape[0]*0.9
135
136 xtrain2 = dff[['Vazão Plenum', 'Alimentação (Hz)']].iloc[:25000]
137 ytrain2 = dff[['T 5']].iloc[:25000]
138 xtest2 = dff[['Vazão Plenum', 'Alimentação (Hz)']].iloc[25000:]
139 ytest2 = dff[['T 5']].iloc[25000:]
140
141 ns = 5
142 mdl2 = NARX(
143     RandomForestRegressor(n_estimators=10),
144     auto_order=2,
145     exog_order=[2,2])
146
147 mdl2.pred_step = ns
148 mdl2.fit(xtrain2, ytrain2)
149 ypred_narx2 = mdl2.predict(xtest2, ytest2)
150
151 ynarx_test2 = mdl2.predict(xtest2,ytest2)
152 r2_score(ynarx_test2[6:],ytest2[6:])
153
154 metrics.mean_absolute_error(ynarx_test2[6:],ytest2[6:]))
155 metrics.mean_squared_error(ynarx_test2[6:],ytest2[6:]))
156 np.sqrt(metrics.mean_squared_error(ynarx_test2[6:],ytest2[6:]))
157
158 xtrain = dff[['Vazão Plenum', 'Resistência']].iloc[:25000]
159 ytrain = dff[['T 5']].iloc[:25000]
160 xtest = dff[['Vazão Plenum', 'Resistência']].iloc[25000:]
161 ytest = dff[['T 5']].iloc[25000:]
162
163 ns = 100
164 mdl1 = NARX(
165     RandomForestRegressor(n_estimators=10),
166     auto_order=2,
167     exog_order=[2,2])

```

```

167 mdl1.pred_step = ns
168 mdl1.fit(xtrain, ytrain)
169 ypred_narx = mdl1.predict(xtest, ytest)
170
171 from sklearn.metrics import r2_score
172 ynarx_test = mdl1.predict(xtest, ytest)
173 r2_score(ynarx_test[6:], ytest[6:])
174
175 metrics.mean_absolute_error(ynarx_test[6:], ytest[6:])
176 metrics.mean_squared_error(ynarx_test[6:], ytest[6:])
177 np.sqrt(metrics.mean_squared_error(ynarx_test[6:], ytest[6:]))
178
179 xtrain3 = dff[['Vazão Plenum', 'Resistência']].iloc[:25000]
180 ytrain3 = dff[['P 1']].iloc[:25000]
181 xtest3 = dff[['Vazão Plenum', 'Resistência']].iloc[25000:]
182 ytest3 = dff[['P 1']].iloc[25000:]
183
184 ns = 1
185 mdl3 = NARX(
186     RandomForestRegressor(n_estimators=10),
187     auto_order=2,
188     exog_order=[2, 2])
189
189 mdl3.pred_step = ns
190 mdl3.fit(xtrain3, ytrain3)
191 ypred_narx3 = mdl3.predict(xtest3, ytest3)
192
193 ynarx_test3 = mdl3.predict(xtest3, ytest3)
194 r2_score(ynarx_test3[6:], ytest3[6:])
195
196 metrics.mean_absolute_error(ynarx_test3[6:], ytest3[6:])
197 metrics.mean_squared_error(ynarx_test3[6:], ytest3[6:])
198 np.sqrt(metrics.mean_squared_error(ynarx_test3[6:], ytest3[6:]))

```