



**INSTITUTO LATINO-AMERICANO DE  
TECNOLOGIA INFRAESTRUTURA E  
TERRITÓRIO  
(ILATIT)  
ENGENHARIA QUÍMICA**

**APRENDIZADO DE MÁQUINA NA PREDIÇÃO DO PERFIL DE TEMPERATURA  
DE TROCADORES DE CALOR**

**VITTORIA SANTOS DE VICENTE**

Foz do Iguaçu  
2024

**APRENDIZADO DE MÁQUINAS NA PREDIÇÃO DO PERFIL DE  
TEMPERATURA DE TROCADORES DE CALOR**

**VITTORIA SANTOS DE VICENTE**

Trabalho de Conclusão de Curso apresentado ao Instituto Latino-Americano de Tecnologia, Infraestrutura e Território da Universidade Federal da Integração Latino-Americana, como requisito parcial à obtenção do título de Bacharel em Engenharia Química.

Orientador:

Foz do Iguaçu  
2024

VITTORIA SANTOS DE VICENTE

**APRENDIZADO DE MÁQUINA NA PREDIÇÃO DO PERFIL DE TEMPERATURA DE  
TROCADORES DE CALOR**

Trabalho de Conclusão de Curso apresentado ao Instituto Latino-Americano de Tecnologia, Infraestrutura e Território da Universidade Federal da Integração Latino-Americana, como requisito parcial à obtenção do título de Bacharel em Engenharia Química.

**BANCA EXAMINADORA**

---

Orientador: Prof. Dra. Marlei Roling Scariot  
UNILA

---

Prof. Dr. Luís Antonio Lourenço  
(UFSC)

---

Prof. Dra. Andreia Cristina Furtado  
UNILA

Foz do Iguaçu, \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

## **AGRADECIMENTOS**

Inicialmente, gostaria de agradecer a minha professora e orientadora Marlei R. Scariot por toda parceria e ensinamentos durante o processo de confecção do trabalho e ao longo da graduação. Obrigada por acreditar no meu potencial e pelas nossas reuniões às quartas de manhã.

Quero agradecer os meus amigos da faculdade que nos momentos mais difíceis não deixaram eu desistir dessa jornada e me carregaram nas costas quando eu não tinha forças para continuar. Barbara e Marta sem vocês eu não teria chegado até o fim.

E finalmente, aos meus familiares, pais, irmã e meu namorado que acompanharam e me incetivaram do início ao fim da graduação.

## RESUMO

O aprendizado de máquina pode ser aplicado em estudos de otimização de vários processos industriais, como de controle de processos, predição de demanda, modelagem e simulação de processos. Essa aplicação também se estende para estudos em equipamentos, como os trocadores de calor, presentes na maioria das indústrias. Com a crescente preocupação com a redução de gastos, bem como com o respeito ao meio ambiente, os estudos relacionados a aplicação da IA na indústria se tornam cada vez mais necessários, pois evitam o desperdício de insumos e trazem maior rentabilidade aos negócios. Diante disso, este trabalho tem como objetivo analisar qual algoritmo de Aprendizado de Máquina melhor identificam o comportamento e consegue prever o processo de troca de calor de forma mais precisa, em um trocador de calor de tubo concêntrico. Para realizar essa análise foram coletados dados de um trocador de calor de tubo concêntrico didático, disponível no laboratório de Engenharia Química da UNILA, e estes dados foram utilizados para alimentar o software Orange Data Mining. Este software possui ferramentas para realizar análises estatísticas e de opções de modelos de aprendizado de máquina, além de ser capaz de criar modelos de predição. A variável escolhida para realizar a predição foi a Temperatura de Saída do fluido quente e, como trata-se de uma variável contínua, foram utilizados os indicadores estatísticos RMSE, MSE, MAE, MAPE e  $R^2$  para avaliar o desempenho dos modelos. Os dados experimentais foram utilizados para treinar modelos de predição derivados de três algoritmos de aprendizado de máquina diferentes: *Linear Regression*, *Decision Tree* e *Randon Forest*. A partir das informações coletas e geradas neste estudo, foi possível avaliar que o modelo de predição que melhor se adaptou ao processo de troca de calor foi o gerado pelo modelo de aprendizado de máquina *Randon Forest*. Esse resultado é consistente com a natureza não linear dos dados do processo de trocadores de calor. O Modelo teve um bom desempenho com os dados de teste, evidenciado por erros baixos e um coeficiente de determinação ( $R^2$ ) próximo de 1. Isso indica que o modelo se ajustou bem aos dados de teste, refletindo de forma precisa o comportamento do sistema.

**Palavras-chaves:** Trocador de calor, Machine Learning, Aprendizado de Máquina, Modelo de Predição

## RESUMEN

El aprendizaje automático puede aplicarse en estudios de optimización de varios procesos industriales, como el control de procesos, la previsión de la demanda, la modelización y simulación de procesos. Esta aplicación también se extiende a estudios en equipos, como los intercambiadores de calor, presentes en la mayoría de las industrias. Con la creciente preocupación por la reducción de gastos, así como el respeto al medio ambiente, los estudios relacionados con la aplicación de la inteligencia artificial en la industria se vuelven cada vez más necesarios, ya que evitan el desperdicio de insumos y aumentan la rentabilidad de los negocios. Por lo tanto, este trabajo tiene como objetivo analizar qué algoritmo de aprendizaje automático se adapta mejor y puede prever el proceso de intercambio de calor de manera más precisa en un intercambiador de calor de tubo concéntrico. Para llevar a cabo este análisis, se recopiló datos de un intercambiador de calor de tubo concéntrico didáctico, disponible en el laboratorio de Ingeniería Química de la UNILA, y estos datos se utilizaron para alimentar el software Orange Data Mining. Este software cuenta con herramientas para realizar análisis estadísticos y opciones de modelos de aprendizaje automático, además de ser capaz de crear modelos de predicción. La variable elegida para realizar la predicción fue la temperatura de salida del fluido caliente y, dado que es una variable continua, se utilizaron los indicadores estadísticos RMSE, MSE, MAE, MAPE y R<sup>2</sup> para evaluar el rendimiento de los modelos. Se utilizaron datos experimentales para entrenar modelos de predicción derivados de tres algoritmos de aprendizaje automático diferentes: Regresión Lineal, Árbol y Bosque Aleatorio. A partir de la información recopilada y generada en este estudio, fue posible evaluar que el modelo de predicción que mejor se adaptó al proceso de intercambio de calor fue el generado por el modelo de aprendizaje automático Bosque Aleatorio. Este resultado es consistente con la naturaleza no lineal de los datos del proceso de intercambiadores de calor. El modelo tuvo un buen desempeño con los datos de prueba, evidenciado por bajos errores y un coeficiente de determinación (R<sup>2</sup>) cercano a 1. Esto indica que el modelo se ajustó bien a los datos de prueba, reflejando de manera precisa el comportamiento del sistema.

**Palabras clave:** Intercambiador de calor, Machine Learning, Aprendizaje automático, Modelo de Predicción.

## ABSTRACT

Machine learning can be applied in optimization studies of various industrial processes, such as process control, demand forecasting, process modeling, and simulation. This application also extends to studies on equipment, such as heat exchangers, found in most industries. With increasing concern for cost reduction, as well as environmental respect, studies related to the application of AI in industry become increasingly necessary as they prevent waste of inputs and bring greater profitability to businesses. Therefore, this work aims to analyze which machine learning algorithm best adapts and can predict the heat exchange process more accurately in a concentric tube heat exchanger. To perform this analysis, data from a didactic concentric tube heat exchanger were collected, available in the Chemical Engineering laboratory of UNILA, and these data were used to feed the Orange Data Mining software. This software has tools for performing statistical analyses and options for machine learning models, as well as being capable of creating prediction models. The variable chosen for prediction was the outlet temperature of the hot fluid, and since it is a continuous variable, statistical indicators such as RMSE, MSE, MAE, MAPE, and R<sup>2</sup> were used to evaluate the performance of the models. Experimental data were used to train prediction models derived from three different machine learning algorithms: Linear Regression, Tree, and Random Forest. Based on the information collected and generated in this study, it was possible to evaluate that the prediction model that best adapted to the heat exchange process was generated by the Random Forest machine learning model. This result is consistent with the nonlinear nature of the heat exchanger process data. The model performed well with test data, evidenced by low errors and a coefficient of determination (R<sup>2</sup>) close to 1. This indicates that the model fitted well to the test data, accurately reflecting the system's behavior.

**Keywords:** Heat exchanger, Machine Learning, Machine Learning, Prediction Model

## LISTA DE FIGURAS

<b>Figura 1.</b> Esquema de funcionamento de um trocador de calor de tubo concêntrico	17
<b>Figura 2.</b> Esquema de um projeto de Machine Learning. <b>Erro! Indicador não definido.</b>	
<b>Figura 3.</b> Funcionamento do modelo de aprendizado supervisionado.	20
<b>Figura 4.</b> Funcionamento do modelo de aprendizado não-supervisionado.	20
<b>Figura 5.</b> Modelo 5-folds Cross Validation	25
<b>Figura 6.</b> Fluxograma da metodologia do experimento	28
<b>Figura 7.</b> Trocador de calor de tubo concêntrico pedagógico	30
<b>Figura 8.</b> Módulo de Operação	30
<b>Figura 9.</b> Configurações de fluxo	32
<b>Figura 10.</b> Localização dos termopares no trocador de calor	33
<b>Figura 11.</b> Modelos de ML disponíveis no Orange Data Mining	34
<b>Figura 12.</b> Tabela de dados antes do tratamento	35
<b>Figura 13.</b> Tabela de dados depois do tratamento	35
<b>Figura 14.</b> Fluxo para análise do banco de dados de treinamento	36
<b>Figura 15.</b> Fluxo para análise do banco de dados teste	37
<b>Figura 16.</b> Local para selecionar a variável target no banco de dados	38
<b>Figura 17.</b> Local para selecionar a variável target no Select Columns	39
<b>Figura 18.</b> COnfiguração do ML Random Forest	40
<b>Figura 19.</b> Fluxo de teste dos modelos de ML	40
<b>Figura 20.</b> Configuração do método de Cross Validation	41
<b>Figura 21.</b> Fluxo para salvar dados gerados na predição	41
<b>Figura 22.</b> Dados gerados na predição	42
<b>Figura 23.</b> Fluxo para visualizar as árvores geradas pelos modelos de predição	42
<b>Figura 24.</b> Fluxo completo	43
<b>Figura 25.</b> Árvores de decisão geradas pelo modelo de predição Random Forest	52
<b>Figura 26.</b> Mapa mental de uma das árvores da floresta gerada pelo modelo de predição Random Forest	52
<b>Figura 27.</b> Perfil de temperatura experimental vs real	53
<b>Figura 28.</b> Recorte temporal do perfil de temperatura - Momento 1	55
<b>Figura 29.</b> Recorte temporal do perfil de temperatura - Momento 2	56

## LISTA DE ABREVIATURAS

TC Trocador de calor

ML Machine learning

AM Aprendizado de máquina

TI Tecnologia da informação

ETA Estação de Tratamento de água

MOO *Multi-Objective Optimization* ou Otimização multicritério

MCDM *Multi-criteria decision making* ou Decisor multicritério

## SUMÁRIO

1. INTRODUÇÃO.....	12
2. OBJETIVOS.....	14
2.1 OBJETIVO GERAL:.....	14
2.2 OBJETIVOS ESPECÍFICOS:.....	14
3. REFERENCIAL TEÓRICO.....	15
3.1 CONTEXTO HISTÓRICO DE MÁQUINAS NOS SISTEMAS PRODUTIVOS.....	15
3.2 TROCADORES DE CALOR DE TUBO CONCÊNTRICO .....	16
3.3 APRENDIZADO DE MÁQUINAS.....	18
3.3.1 MODELOS DE APRENDIZADO DE MÁQUINA.....	19
3.3.2 Orange Data Mining .....	21
3.3.3 Modelo de Regressão Linear.....	21
3.3.4 Modelo Árvore de Decisão .....	22
3.3.5 Modelo Floresta Aleatória.....	22
3.6 O APRENDIZADO DE MÁQUINA EM PROCESSOS DE TRANSFERÊNCIA DE CALOR.....	23
3.7 AVALIAÇÃO DOS MODELOS DE APRENDIZADO E PREDIÇÃO .....	24
3.7.1 <i>K-fold Cross Validation</i> .....	24
3.7.2 Métricas de desempenho.....	25
3.8 AVALIAÇÃO DOS BANCOS DE DADOS .....	27
4. METODOLOGIA .....	28
4.1 PLANEJAMENTO DOS ENSAIOS EXPERIMENTAIS.....	29
4.2 ETAPA EXPERIMENTAL - COLETA DE DADOS.....	29
4.4 MODELO DE APRENDIZADO DE MÁQUINA .....	34
4.4 DADOS PARA TREINAMENTO E TESTE DO MODELO AM.....	35
4.5 PROCESSO DE MONTAGEM DO LAYOUT DO ORANGE.....	36
4.6 AVALIAÇÃO DO MODELO AM .....	43

4.7 QUESTÕES PRÁTICAS DO SOFTWARE.....	45
5. RESULTADOS.....	46
5.1 AVALIAÇÃO DOS BANCOS DE DADOS .....	46
5.1.2 Banco de dados de treinamento.....	46
5.1.3 Banco de dados de teste.....	48
5.2 AVALIAÇÃO DOS MODELOS DE APRENDIZADO DE MÁQUINA .....	49
5.2.1 Treinamento dos modelos de aprendizado com o banco de dados de treinamento .....	49
5.2.2 Testes dos modelos de predição.....	50
5.2.3 Árvores de decisão geradas.....	51
5.2.4 Comparação entre modelos .....	53
6. CONCLUSÃO .....	58
7. REFERÊNCIAS.....	59
ANEXOS.....	62
ANEXO I.....	63
ANEXO II.....	64

# 1. INTRODUÇÃO

Estudos recentes têm usado aprendizado de máquina na engenharia química para diversas aplicações, como no trabalho de WANG et al., (2023) em que foi criado um modelo de predição para prever a demanda de ar de estações de tratamento de gua, com objetivo de melhorar a eficiência da aeração, diminuir o consumo de energia e produção de gases do efeito estufa; também no artigo de DAOUTIDIS et al., (2024) foi feita uma análise sobre como o Machine Learning (ML) vem impactando a indústria de processos, tanto como um facilitador de inovação, eficiência e automação, quanto na descoberta de novos produtos e na gestão de operações; no capítulo 5 do livro escrito por GAO, (2024), o autor aborda temas como aplicação de técnicas de modelagem híbrida, combinando modelos baseados em princípios físicos com abordagens de aprendizado de máquina e como essa integração tem sido bem-sucedida na descoberta de cinéticas bioquímicas, otimização de reações complexas e dimensionamento de reatores industriais; no trabalho escrito por WANG et al., (2022) foi desenvolvido um framework abrangente que combina MOO e MCDM (soluções de otimização multicritério) com técnicas de ML e quando é integrado ao controle preditivo baseado em modelo (MPC) permite que o MPC lide eficientemente com múltiplos objetivos e aproveite modelos de ML no controle de processos; dentre outras áreas.

Quando se trata de tornar os processos mais eficientes utilizando ML é preciso analisar vários fatores, principalmente os equipamentos utilizados. Os trocadores de calor (TC) são equipamentos amplamente utilizados em diversas indústrias, onde são o principal equipamento no processo de transferência de energia térmica de um fluido para outro. Aplicações industriais de trocadores de calor são das mais variadas possíveis. Fala-se então de sua utilização para geração de energia elétrica, indústria petroquímica, indústria siderúrgica, indústria de alimentos, ciclos de refrigeração, aplicações espaciais, entre outras (KAKAÇ; LIU; PRAMUANJAROENKIJ, 2012-a apud FARIA, 2015).

Trocadores de calor são equipamentos utilizados com o intuito de que se aproveite eficientemente determinada quantidade de energia térmica, por exemplo uma energia resultante de um processo exotérmico. Tal quantidade de energia é trocada entre dois ou mais fluidos que circulam no interior do equipamento e a

predição do perfil de temperatura desses equipamentos é crucial para garantir sua eficiência e segurança operacional.

De acordo com HODSON (2016) o objetivo do Aprendizado de Máquina (AM) nos processos industriais é a construção de modelos de predição que permitam o mapeamento de novos cenários e comportamentos para aquele processo específico, para revelar problemas antes de acontecerem, otimizar a vida útil dos componentes e reduzir a necessidade de intervenção humana. Estes modelos de predição aprendem com exemplos de cenários anteriores, por isso precisa-se de uma base de dados de um tamanho relevante para que o algoritmo consiga identificar os padrões dos processos e, dessa forma, conseguir criar hipóteses a partir do que foi aprendido com os dados fornecidos.

MATOS (2021) fala sobre a importância do AM nas indústrias e afirma que todas as indústrias serão afetadas pelo desenvolvimento e aplicação de algoritmos de inteligência artificial. A aplicação destes algoritmos em processos, produtos e serviços tornarão o negócio de qualquer empresa mais eficiente, inovador e rentável.

Visto a relevância do AM nos processos industriais e buscando sempre a conexão entre IA e indústria, este trabalho tem como objetivo avaliar se o algoritmo de aprendizado de máquina será capaz de prever as temperaturas de saída do fluido quente do trocador de calor, diante das variações nas variáveis do processo.

## 2. OBJETIVOS

### 2.1 OBJETIVO GERAL:

O objetivo deste trabalho é identificar qual modelo de predição será capaz de prever as temperaturas de saída do fluido quente do trocador de calor com maior precisão diante de perturbações nas variáveis de entrada.

### 2.2 OBJETIVOS ESPECÍFICOS:

1. Coletar, compreender e analisar dados de um trocador de calor.
2. Compreender o funcionamento do Orange Data Mining.
3. Entender o funcionamento dos modelos de predição
4. Treinar um algoritmo de aprendizado de máquina para criar um modelo de predição
5. Validar estatisticamente a precisão e exatidão dos modelos de predição.

### 3. REFERENCIAL TEÓRICO

Neste capítulo, são apresentados conceitos sobre indústria 4.0, aprendizado de máquina, trocadores de calor, a aplicação destes equipamentos em processos industriais, modelos de aprendizado de máquina, software Orange Data Mining e avaliações estatísticas para modelos de predição.

#### 3.1 CONTEXTO HISTÓRICO DE MÁQUINAS NOS SISTEMAS PRODUTIVOS

Com o passar dos séculos, ocorreram avanços tecnológicos nos meios de produção que se mostraram essenciais para a melhoria na qualidade, rapidez e nos custos de produção. A Revolução Industrial foi de grande relevância para a sociedade atual e para o surgimento da revolução tecnológica vivida até hoje (CAVALCANTE; SILVA, 2011). Durante a primeira revolução industrial, iniciou-se a substituição do trabalho manual pelas máquinas. No século XIX, houve a introdução de máquinas mais modernas, juntamente com a utilização de energia elétrica na produção. Já no século XX, a produção enxuta, a automação e o uso intensivo da tecnologia da informação (TI) marcaram a terceira revolução industrial. No século XXI, com o objetivo de fazer com que os sistemas automatizados que controlam os equipamentos industriais pudessem trocar informações entre máquinas e seres humanos para otimizar o processo de produção, foi introduzido o conceito de indústria 4.0, marcando a quarta revolução industrial (SACOMANO et al., 2018).

O conceito de Indústria 4.0, busca explorar os efeitos de um mundo cada vez mais conectado. Isso se dá pela capacidade cada vez maior de processamento, que permite trabalhar com uma grande quantidade de informações disponíveis (Big Data) e ampliação do uso de técnicas de Inteligência Artificial (IA), especialmente de Machine Learning (ML), para extrair informações do comportamento do processo, fornecendo soluções capazes de tornar os processos cada vez mais eficientes, como é o caso do controle de processos (SILVA, 2020).

O aprendizado de máquina é uma ramificação da Inteligência Artificial que utiliza uma grande quantidade de dados para simular o aprendizado humano. Assim como os humanos aprendem com experiência de vida, a máquina aprende a partir dos dados, conseguindo observar padrões e prever resultados (ISHIDA, 2021). Na indústria, o aprendizado de máquina pode auxiliar na otimização de

processos e redução de custos com embasamento nos dados gerados(CARTWRIGHT, 2020).

A Engenharia Química está passando por uma transformação em direção à digitalização e automação total da indústria e pesquisa. O que leva a uma disponibilidade cada vez maior de dados e à necessidade de uma tomada de decisão otimizada e automatizada com base em dados, permitindo operações de processo mais sustentáveis (SCHWEIDTMANN et al., 2021).

### 3.2 TROCADORES DE CALOR DE TUBO CONCÊNTRICO

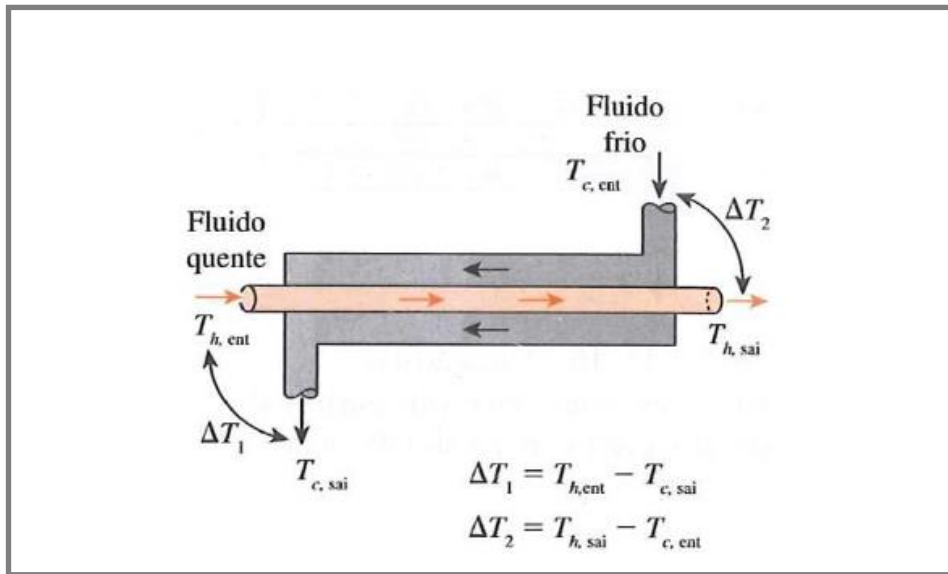
De acordo com INCROPERA; DEWITT, (2014), trocadores de calor são equipamentos que promovem a transferência de calor entre dois fluidos que escoam a diferentes temperaturas, normalmente separados por uma interface sólida. Estes equipamentos podem ser encontrados em diversos segmentos da indústria e também no uso doméstico.

A transferência de calor ocorre entre dois ou mais corpos - líquidos, gases ou sólidos e os meios pelos quais esta transferência pode ocorrer são condução, convecção e/ou radiação térmica (BEJAN, 2004).

Nos trocadores de calor a transferência de calor se dá por condução (na parede que separa os dois fluidos) e por convecção em cada fluido (ÇENGEL; GHAJAR, 2012). Sua principal diferenciação é dada em termos de sua geometria, destacando-se os trocadores dos tipos: tubo duplo (“double pipe”), casco e tubo (“shell and tube”) e de placas (“plate”).

Trocadores de calor de tubos concêntricos (Figura 1) são, geralmente, constituídos por um tubo concêntricamente inserido em outro tubo de diâmetro maior, fazendo com que um dos fluidos passe por dentro do tubo menor, e o outro passe pela cavidade entre os tubos. Assim, pode ocorrer duas configurações de fluxo para esse tipo de trocador, o fluxo contracorrente ocorre quando o fluido que escoar dentro do tubo menor e o fluido que escoar na cavidade entre os tubos fluem em sentidos opostos, e o fluxo paralelo ou concorrente, que ocorre quando os fluidos fluem no mesmo sentido (KAKAÇ; LIU; PRAMUANJAROENKIJ, 2012-b).

**Figura 1.** Esquema de funcionamento de um trocador de calor de tubo concêntrico



Fonte: ÇENGEL; GHAJAR, (2012)

Os trocadores de calor são usados para transferir energia térmica entre um ou mais fluidos que se encontram em um processo de interação térmica. Entre os sistemas mais comuns estão o aquecimento, resfriamento, condensação e evaporação de fontes de fluido de um ou mais componentes. Sendo assim, diversos setores de aplicação podem ser citados, tais como: a indústria do petróleo e gás, fins domésticos e industriais, tais como a refrigeração ar condicionado e geração de energia (BERNUY et al., 2021)

O trocador de calor industrial adequa-se a diversos segmentos, suprimindo as necessidades de indústrias pertencentes a vários setores como: químico, petroquímico, farmacêutico, alimentício, siderúrgico, minerador, energético e de papel e celulose. Além disso, o trocador de calor industrial é também utilizado para gerar energia e vapor em usinas de produção de açúcar e de biocombustível (INCASE, 2022).

### 3.3 APRENDIZADO DE MÁQUINAS

De acordo com Guilhon et al., (2022) aprendizado de Máquina ou Machine Learning é uma técnica que usa algoritmos baseados em matemática e estatística para realizar a tarefa de aprendizado. Surgiu da necessidade de processar e obter informação útil a partir dos dados e, já que é inviável realizar manualmente o processamento e análise da grande quantidade de dados que temos disponíveis atualmente, é necessário automatizar tarefas, simulando o comportamento humano.

O aprendizado é o objetivo principal em machine learning, é a capacidade, dos seres inteligentes (humanos ou não), de se adaptar, modificar e melhorar seu comportamento e suas respostas. Diz-se que se está aprendendo um modelo de conhecimento a partir de um conjunto de dados, quando se procura por padrões nestes dados. Já quando há uma estimativa dos valores desconhecidos para atributos do conjunto de dados é realizada, diz-se que o modelo está sendo aplicado. Uma etapa inicial do processo será necessário coletar e analisar os dados adequados para o problema e prepará-los. Após estas etapas é que podemos construir o modelo de Machine Learning, que deve ser avaliado e criticado. Quando se chegar a um modelo satisfatório para o problema, será necessário apresentar os resultados para o demandante e distribuir o modelo em ambiente produtivo. A Figura 2 resume estas

etapas que devem ser realizadas nos projetos de Machine Learning (GUILHON et al., 2022).

**Figura 2.** Esquema de um projeto de Machine Learning.



Fonte: ESCOVEDO; KOSHIYAMA, (2020)

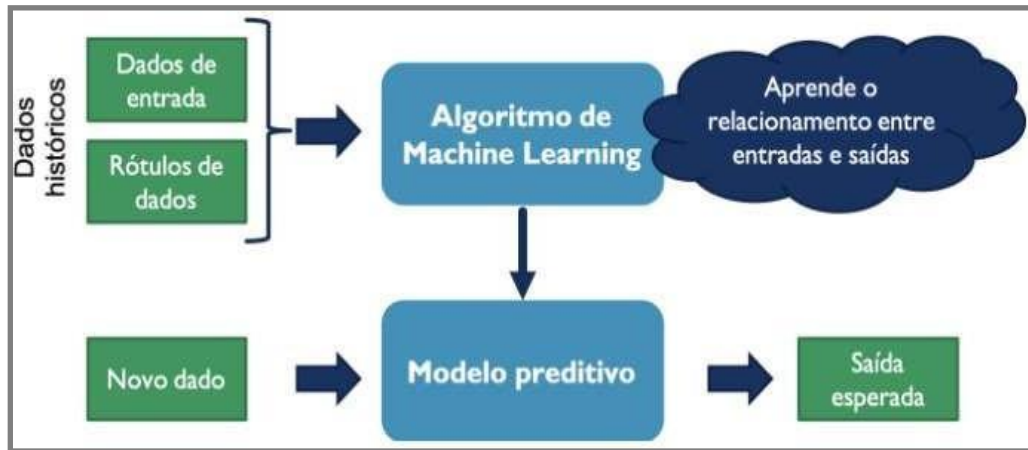
### 3.3.1 MODELOS DE APRENDIZADO DE MÁQUINA

Uma das mais notáveis inovações contemporâneas que impactam significativamente a humanidade é a adoção de algoritmos de aprendizado de máquina, os quais não apenas desempenham um papel crucial em processos industriais, mas também permeiam o dia a dia das pessoas.

Segundo GUILHON et al., (2022) os modelos de aprendizado de máquina são a estrutura lógica dos algoritmos que fazem previsões, classificações e muitas outras tarefas. Estes modelos podem ser do tipo Não Supervisionado ou Supervisionado. No aprendizado supervisionado, o aprendizado se baseia em apresentar para o algoritmo um número suficiente de exemplos de entradas e saídas desejadas, já rotuladas previamente, desta forma, o algoritmo terá como objetivo aprender uma regra geral que mapeie as entradas nas saídas corretamente, o que

consiste no modelo final. A Figura 3 mostra o esquema de funcionamento do modelo de aprendizado supervisionado.

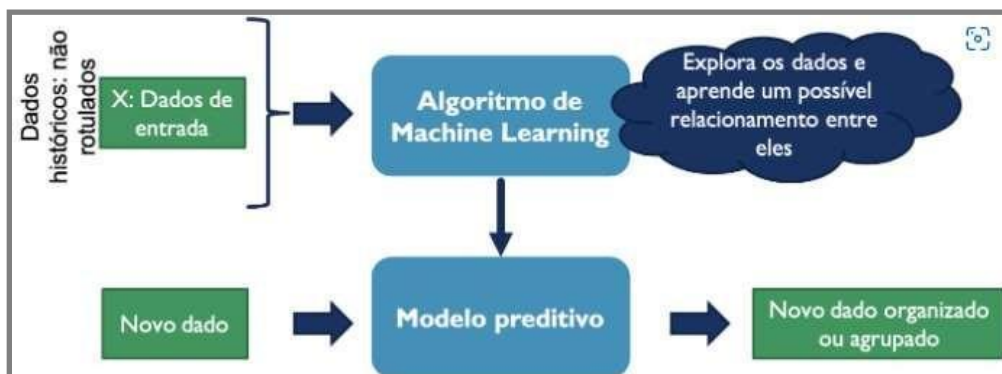
**Figura 3.** Funcionamento do modelo de aprendizado supervisionado.



Fonte: GUILHON et al., (2022)

No aprendizado não-supervisionado, os dados são não-rotulados, pois não se tem as informações das saídas desejadas a serem estimadas. O algoritmo não recebe durante o treinamento os resultados esperados, devendo descobrir por si mesmo, por meio da exploração dos dados, os possíveis relacionamentos entre eles. Assim, o processo de aprendizado busca identificar regularidades entre os dados a fim de organizá-los em função das similaridades que apresentam entre si (GUILHON et al., 2022) A Figura 4 mostra o esquema de funcionamento do modelo de aprendizado não-supervisionado.

**Figura 4.** Funcionamento do modelo de aprendizado não-supervisionado.



Fonte: GUILHON et al., (2022)

### 3.3.2 Orange Data Mining

A ferramenta Orange é uma ferramenta livre (*open source*) de análise de dados direcionada para os mais novos na área e também para especialistas. Trata-se de conjunto de software compreensível e baseado em componentes para *machine learning* e *data mining*, desenvolvido no laboratório de bioinformática na Faculdade de Ciência da Computação e Tecnologias da Universidade de Ljubljana, na Eslovénia, em conjunto com uma comunidade de apoio “*open source*” (GOMES, 2014).

O Orange é uma ferramenta utilizada para visualização e análise de dados de código aberto, na qual é possível extrair dados por programação visual ou scripts Python, além disso, é possível realizar análises estatísticas, box plots ou scatter plots, agrupamento hierárquico, heatmaps e projeções lineares (DEMSAR J et al., 2013). O Orange Canvas oferece uma interface estruturada em funcionalidades agrupadas em nove categorias, entre elas: visualização, avaliação, operações de dados, aprendizado não supervisionado, associação, classificação, regressão, visualização usando “*widgets*” e protótipo de implementações (JOVIC; BRKIĆ; BOGUNOVIĆ, 2015)

O Orange possui diversos modelos de ML. Os modelos específicos utilizados neste estudo são detalhados a seguir:

### 3.3.3 Modelo de Regressão Linear

Segundo James et al., (2013) a regressão linear é uma abordagem conhecida e simples utilizada para se prever uma resposta quantitativa Y a partir de uma variável de entrada X. Ela assume que há uma relação linear entre as duas variáveis, sendo matematicamente expressa como:

$$y = ax + b$$

Como já esperado, na prática, os coeficientes a e b são desconhecidos e para uso da função estatística, é necessário utilizar o conjunto de dados das observações para obter uma estimativa de seus valores. Em outras palavras, quer-se descobrir um a e b tal que a a linha resultante seja a mais próxima possível do conjunto de n observações.

### 3.3.4 Modelo Árvore de Decisão

As árvores de decisão são um popular algoritmo de machine learning que se destaca por sua facilidade visual quando comparado a outros métodos. As etapas desse algoritmo muito lembram um fluxograma e, portanto, possuem certa facilidade de entendimento. Uma árvore de decisão é um algoritmo de aprendizado de máquina supervisionado que é utilizado para classificação e para regressão. Isto é, pode ser usado para prever categorias discretas (sim ou não, por exemplo) e para prever valores numéricos (o valor do lucro em reais). (SACRAMENTO, 2021)

Existem dois tipos de árvores: as de decisão para classificação e as de regressão. Enquanto árvores de regressão são utilizadas quando a variável de saída é numérica, as árvores de classificação são utilizadas quando a variável de saída é categórica. No caso das árvores de regressão, o valor obtido pelos nós terminais dos dados de treinamento é o valor médio das observações, enquanto para as árvores de classificação, esse valor é a moda das observações. Ambas as árvores dividem o chamado espaço preditor em regiões distintas e não sobrepostas. Essas regiões podem ser pensadas como caixinhas com as devidas observações (MARTINS, 2016).

Em ambos os casos, árvores totalmente crescidas são sujeitas ao overfitting, dado que a divisão mais avançada alcançada seria aquela com uma só observação em cada folha. Para lidar com tal situação, pode-se utilizar a poda ou técnicas de ensembles (HAN, 2022).

### 3.3.5 Modelo Floresta Aleatória

Os métodos baseados em árvores para regressão e classificação. Estes envolvem estratificar ou segmentar o espaço preditor em várias regiões simples. Para fazer uma predição para uma determinada observação, normalmente usamos a média ou o modo das observações de treinamento na região a que pertence. Como o conjunto de regras de divisão usado para segmentar o espaço de predição pode ser resumido em uma árvore, esses tipos de abordagens são conhecidos como métodos de árvore de decisão (JAMES et al., 2013).

O algoritmo de floresta aleatória é composto por uma coleção de árvores de decisão, sendo que cada árvore na combinação é composta por uma amostra de dados extraída de um conjunto de treinamento com substituição, chamada de amostra bootstrap. Dessa amostra de treinamento, um terço dela é separado como dados de teste, conhecida como amostra out-of-bag (OOB), a qual iremos abordar mais tarde. Outra instância de aleatoriedade é então injetada por meio do bagging de recursos, incluindo mais diversidade ao conjunto de dados e reduzindo a correlação entre as árvores de decisão. Dependendo do tipo de problema, a determinação da predição irá variar. Para uma tarefa de regressão, as árvores de decisão individuais terão suas médias calculadas, e para uma tarefa de classificação, um voto da maioria, ou seja, a variável categórica mais frequente, produzirá a classe prevista. Por fim, a amostra OOB é então utilizada para validação cruzada, finalizando esse predição. (IBM, 2020).

### 3.6 O APRENDIZADO DE MÁQUINA EM PROCESSOS DE TRANSFERÊNCIA DE CALOR

Alguns estudos merecem uma abordagem mais detalhada, pois contribuem de modo especial no entendimento das questões que serão abordadas ao longo deste trabalho, por exemplo:

KUNELBAYEV; ASSEL; GULDANA, (2021) fizeram uso de algoritmos de aprendizado de máquina para a identificação de anomalias em um sistema de aquecimento solar de água quente com um trocador de calor em um tanque de armazenamento, com o intuito de introduzir uma nova arquitetura para redes, baseada em equações diferenciais ordinárias para predição de acidentes.

O modelo foi criado para um sistema de aquecimento solar, com base em um modelo matemático de um coletor solar plano. Foram instalados sensores de temperatura no tanque do acumulador, na saída do coletor solar plano, sensores de tensão na bomba e de vibração. Com base nos dados dos sensores de pressão e temperatura, foi possível prever problemas futuros na instalação. Porém, houve diferença entre a temperatura e pressão previstas no modelo em comparação com as reais.

Já, CARMO, (2017), propôs uma metodologia para o emprego de redes neurais artificiais no monitoramento da formação de material particulado em

uma caldeira de recuperação. A autora utiliza redes de mapeamento auto-organizável (SOM) para a classificação dos dados em função do número de partículas, utilizando dados históricos de um ano de operação em uma planta industrial. Obteve-se uma rede neuronal artificial de topologia 20x5 em um mapa auto-organizável (SOM) que permite a identificação de regiões de operação normal, regiões de transição e regiões de maior probabilidade de operação anormal.

RODRÍGUEZ et al., (2013), realizou uma modelagem térmica de um trocador industrial casco e tubos, por meio do método  $\dot{q} - NTU$  e a abordagem Bell Delaware com o intuito de estimar o coeficiente de transferência de calor e a queda de pressão no lado do casco. Também, utilizando o software MATLAB R2009, o trocador foi otimizado para maximizar sua eficiência e minimizar o custo total. A eficácia foi maximizada e os custos totais minimizados. Os resultados obtidos por métodos exatos e por métodos heurísticos foram comparados, contabilizando as economias que a otimização sugere. Os resultados mostraram o conflito entre dois objetivos: eficiência e custo total, o que demonstra a necessidade de técnicas de otimização no projeto de trocadores de casco e tubo.

O estudo conduzido por NAIR et al., (2022) “Machine Learning for Prediction of Heat Pipe Effectiveness” utilizou algoritmos de aprendizado de máquina para prever a eficácia de um sistema de tubo de calor em um TC de tubo concêntrico. O estudo realizou experimentos com metanol como fluido de trabalho e nos ensaios os pesquisadores variaram vários fatores e analisaram quanto cada um deles influencia na eficiência da troca térmica.

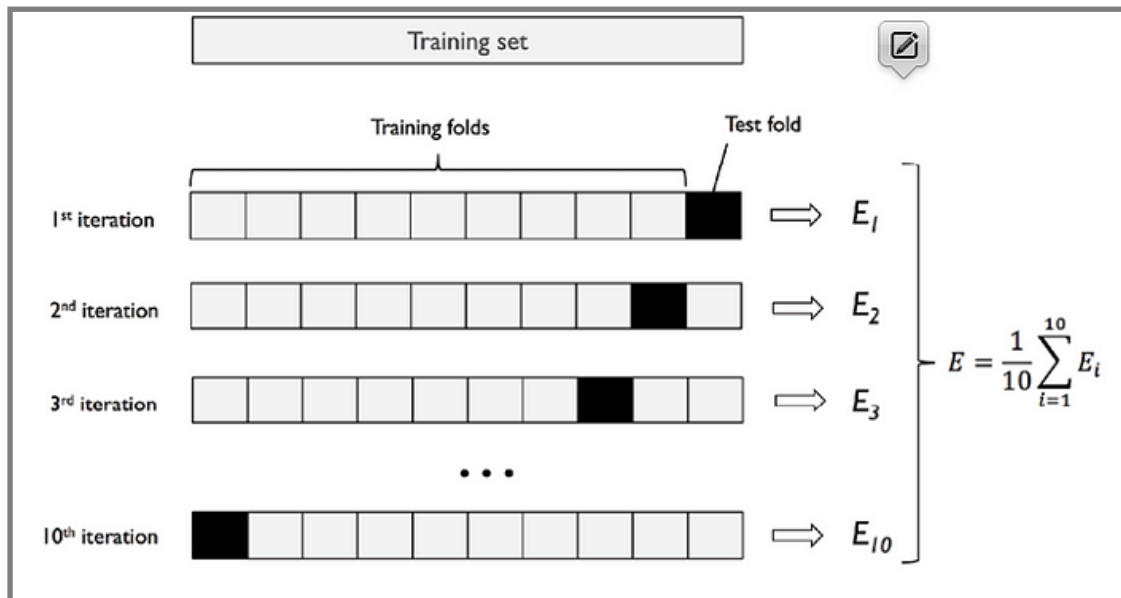
### 3.7 AVALIAÇÃO DOS MODELOS DE APRENDIZADO E PREDIÇÃO

#### 3.7.1 *K-fold Cross Validation*

A validação cruzada é definida por RASCHKA, (2018) como uma das principais e mais comuns técnicas de validação de um modelo de aprendizado. O autor explica que a ideia principal é dividir o banco de dados de treinamento em  $k$  grupos, onde um é separado para testar o modelo de predição criado pelos  $k-1$  grupos que foram usados para treinamento do modelo de aprendizado (learner). A ideia é que cada uma dessas partes seja usada para teste e o resto para treinamento, criando modelos de predição diferentes em cada rodada e analisando se eles se comportam

de forma parecida. Na Figura 5, é possível ver como é a metodologia *5-fold Cross Validation*, mesmo metodo utilizado neste trabalho.

**Figura 5.** Modelo 5-folds Cross Validation



Fonte: (SINGH, 2020)

### 3.7.2 Métricas de desempenho

Para avaliar modelos preditivos é comum utilizar métricas de desempenho que avaliam a distância entre os valores observados e os previstos, logo, é esperado que este valor seja o mínimo possível.

Neste trabalho escolhemos algumas métricas para avaliar os modelos preditivos criados e essas métricas foram descritas JAMES et al., (2013)

- Erro Absoluto Médio (MAE): O Erro Absoluto Médio é a diferença absoluta média entre os valores observados e os valores previstos pelo modelo. Quanto menor o valor do MAE, mais acuradas são as previsões do modelo.

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

- Erro Quadrático Médio (MSE): O Erro Quadrático Médio é a diferença quadrática média entre os valores observados e os valores previstos pelo modelo. Quanto menor o valor do MSE, mais acuradas são as previsões do modelo; erros grandes são “penalizados” pela potenciação. Modelos que minimizam o MSE levam a previsões da mediana.

$$MSE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$$

- Raiz do Erro Quadrático Médio (RMSE): A Raiz do Erro Quadrático Médio é a raiz quadrada da diferença quadrática média entre os valores observados e os valores previstos pelo modelo. Quanto menor o valor do RMSE, mais acuradas são as previsões do modelo; erros grandes são “penalizados” pela potenciação. Modelos que minimizam o RMSE levam a previsões da média.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

- Erro Percentual Absoluto Médio (MAPE): O Erro Percentual Absoluto Médio é a média do percentual absoluto de erros de predição em relação aos valores observados. Um MAPE igual a zero indica que o modelo prevê perfeitamente os dados; quanto maior o valor de MAPE, maior a discrepância entre os valores observados e previstos.

$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{y_j - \hat{y}_j}{\hat{y}_j} \right|$$

- Coeficiente de determinação: Para avaliar o ajuste geral do modelo aos dados é importante utilizar o Coeficiente de determinação. Num contexto de um modelo de regressão linear simples, em que a variável explanatória (ou preditora) é  $x$  e a variável resposta (ou prever) é  $y$ , o coeficiente de

determinação  $R^2$  dá a percentagem de variabilidade dos  $y$ 's (variável a prever) que fica explicada em função da variabilidade dos  $x$ 's (MARTINS, 2018)

$$R^2 = 1 - \frac{\sum_i^N (y_i - \hat{y}_i)^2}{\sum_i^N (y_i - \bar{y}_i)^2}$$

Assim, um valor de  $R^2 \approx 1$  significa que, em princípio, a nuvem de pontos apresentada no diagrama de dispersão está próxima da reta de regressão, considerada para o modelo de regressão. Quando  $R^2 \approx 0$  já não se vislumbra uma estrutura linear (MARTINS, 2018).

### 3.8 AVALIAÇÃO DOS BANCOS DE DADOS

Avaliação do banco de dados é importante para o aprendizado de máquina, pois dessa forma é possível avaliar a qualidade dos dados coletados e o quão abrangente eles são em relação ao processo.

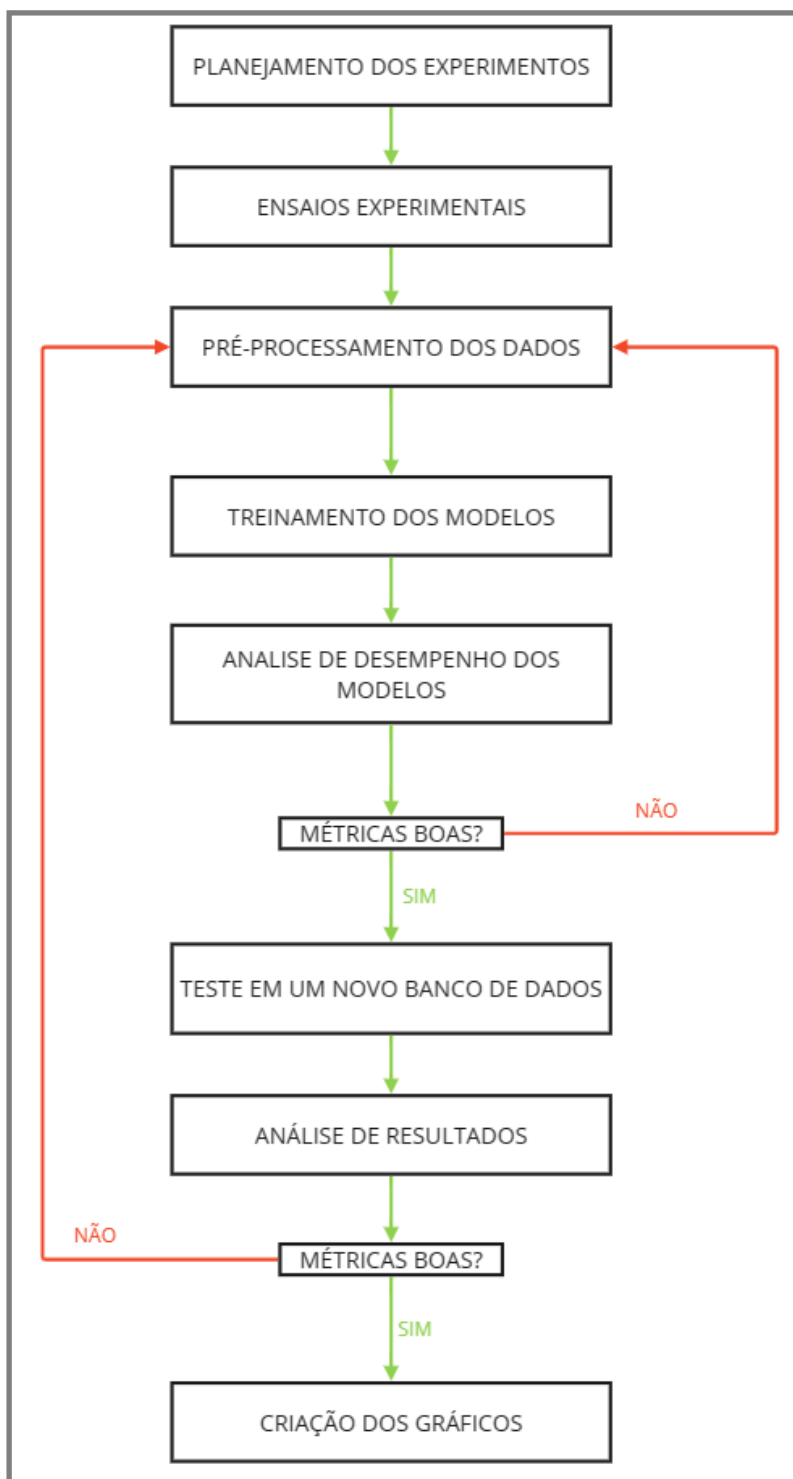
Para avaliação do banco de dados neste trabalho foi utilizada uma exploração estatística básica como: média, mediana, moda, máximo, mínimo, entre outros. Além disso foi utilizado o Coeficiente de Correlação de *Pearson*, o mesmo consegue identificar através de métodos matemáticos, o grau de influência entre duas variáveis, se elas se relacionam de forma linear. Ou seja, esse coeficiente indica o quanto a alteração de uma variável vai influenciar em outra.

O coeficiente varia entre o intervalo  $(-1 \leq R \leq +1)$ . A correlação varia de sentido, de negativo para positivo, e de força, sendo fraca quando se aproxima de zero (à esquerda ou à direita), e moderada ou forte, aproximando-se de -1 ou +1, sendo  $R$  o coeficiente de *Pearson* (BARBETTA, 2003).

## 4. METODOLOGIA

A metodologia do trabalho está representada pelo fluxograma de trabalho, conforme mostra a Figura 6.

Figura 6. Fluxograma da metodologia do experimento



Fonte: Autor (2024)

#### 4.1 PLANEJAMENTO DOS ENSAIOS EXPERIMENTAIS

A partir das possibilidades oferecidas pelo trocador de calor, que é um equipamento didático, analisando os limites operacionais do TC foram planejados diferentes cenários para a coleta de dados. O objetivo era coletar dados com as variáveis manipuladas(MV) em diferentes patamares, no caso as vazões de água fria e quente, em diferentes setpoints (SP) de temperatura de água quente disponíveis no controle do TC, abrangendo ao máximo os limites do TC. Dessa forma, foram montados 8 ensaios em fluxo paralelo com parâmetros diferentes, apresentados na tabela 1, e os dados foram coletados durante 5 minutos para cada ensaio, mantendo-se um intervalo de tempo de 5 segundos.

Tabela 1. Dados de entrada dos dados de partida dos experimentos

<b>Experimento</b>	<b>Setpoint Água quente (°C)</b>	<b>Temperatura Água fria (°C)</b>	<b>Vazão Corrente fria e quente (L/min)</b>
1	40	19	2,5
2	40	19	1
3	45	19	1
4	45	19	2
5	50	19	2
6	50	19	1
7	60	19	1
8	60	19	2

#### 4.2 ETAPA EXPERIMENTAL - COLETA DE DADOS

A seguir será apresentado o aparato experimental bem como o método de coleta dos dados.

O banco de dados utilizado neste trabalho foi obtido a partir de ensaios experimentais, descritos na etapa anterior, em um trocador de calor em escala laboratorial. O equipamento utilizado foi um trocador de calor bitubular de tubo

concêntrico da empresa TecQuipment modelo TD360a, apresentado na Figura 7, disponível no laboratório multidisciplinar de engenharia química da UNILA.

**Figura 7.** Trocador de calor de tubo concêntrico pedagógico



Fonte: TECQUIPMENT

O TC possui um módulo de operação (Figura 8) que interage com uma interface supervisória "VDAS" que pode ser instalada em um notebook, esta interface fez o reconhecimento do TC e permitiu a importação dos dados em um arquivo CSV. O TC está detalhado no esquema abaixo:

**Figura 8.** Módulo de Operação



Fonte: TECQUIPMENT

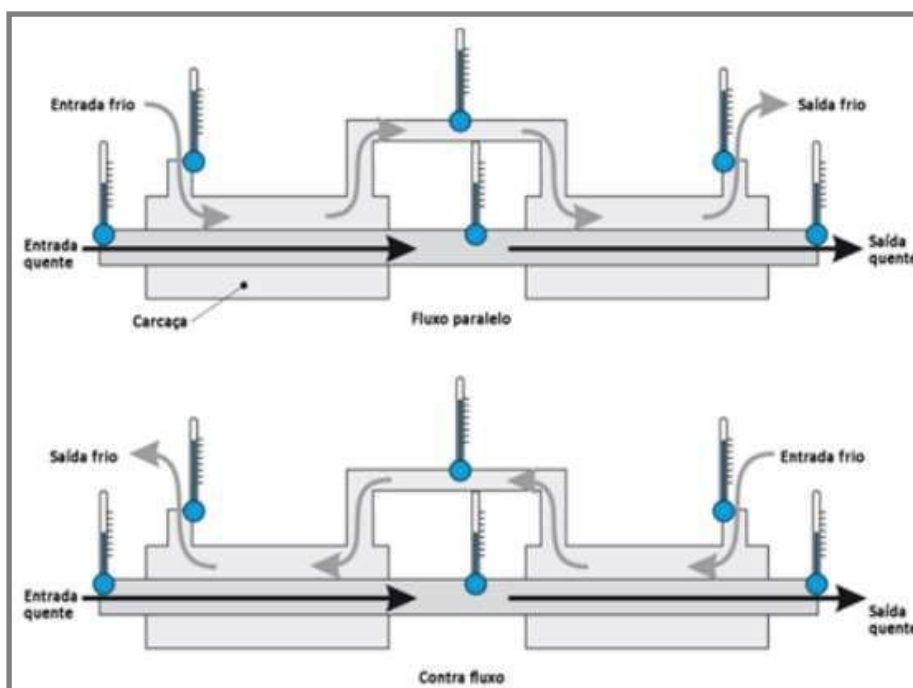
Na tabela 2 foi identificado e detalhado todas as especificações de dimensões do TC (TD360a) disponíveis no Guia do Usuário da Bancada de Trocador de Calor da TecQuipament.

Tabela 2. Especificações do TC tubo concêntrico TD360a

<b>Informações</b>	<b>Valores</b>		<b>Unidade de medida</b>
Largura	500		mm
Altura	260		mm
Profundidade	160		mm
Diâmetro tubo externo	Externo	Interno	mm
	30	20	
Diâmetro tubo interno	Externo	Interno	mm
	1	10	
Área de transferência de calor	0,02		m <sup>2</sup>

Este trocador possui 6 termopares numerados em suas extremidades e ao centro que recolhem os dados de entrada, saída e temperaturas intermediárias, como mostra a Figura 9.

Figura 9. Configurações de fluxo



Fonte: TECQUIPMENT

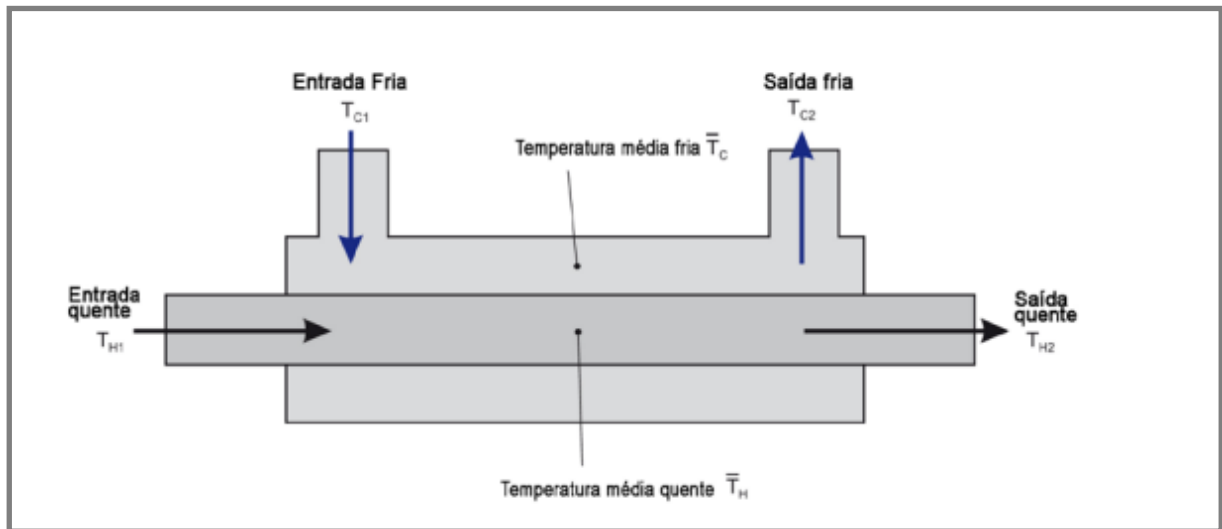
Neste equipamento é possível definir algumas variáveis de entrada no sistema, são estas: Temperatura de entrada quente ( $T_{qe}$ ), Vazão de água quente e vazão de água fria, a última sempre está em temperatura ambiente e a água quente é aquecida em um sistema do próprio módulo de operação, na Tabela 3 foi identificado os limites operacionais dessas variáveis. Estas variáveis de entrada são de extrema importância pois elas que definiram o banco de dados para o modelo realizar a predição, estas estão detalhadas na Tabela 1 no Planejamento Experimental.

Tabela 3. Limites do Trocador de Calor

Informações	Mínimo	Máximo	Unidade de medida
Vazão (fluido frio e quente)	0	3	L.min-1
Temperatura (fluido quente)	Temperatura ambiente	60	°C

Na Figura 10 é possível identificar a localização dentro do tubo onde os termopares fazem a coleta dos dados de cada corrente da temperatura intermediária, de entrada e saída, tanto da corrente fria quanto da quente.

**Figura 10.** Localização dos termopares no trocador de calor



Fonte: (TECQUIPMENT)

E na Tabela 4 é descrito ao que corresponde cada variável no banco de dados coletado do trocador de calor, pois o modulo de operação coloca uma nomeação própria para cada variável.

Tabela 4. Descrições das variáveis coletadas no TC

Variável	Nome no banco de dados
Temperatura Saída Quente	Temp Hot 2
Temperatura Intermediária Quente	Temp Hot 3
Temperatura Entrada Quente	Temp Hot 1
Temperatura Saída Fria	Temp Cold 2
Temperatura Intermediária Fria	Temp Cold 3
Temperatura Entrada Fria	Temp Cold 1
Vazão Fria	Flow Cold
Vazão Quente	Flow Hot
Calor específico Quente	Specific Heat Capacity Hot
Calor específico Fria	Specific Heat Capacity Cold
Diferença Temperatura Quente	D Thot
Diferença Temperatura Fria	D Tcold
Densidade Água quente	Density Hot
Densidade Água fria	Density Cold

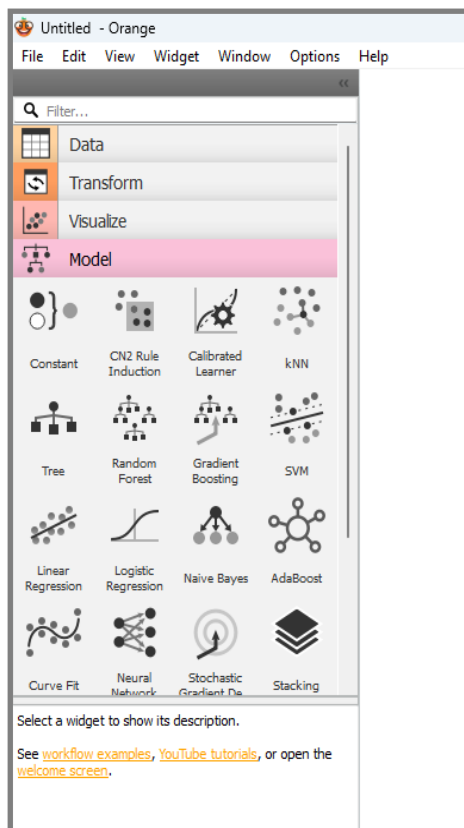
#### 4.4 MODELO DE APRENDIZADO DE MÁQUINA

A eficiência da troca de calor depende de inúmeros fatores em um sistema, como a temperatura de entrada da água quente, a temperatura de entrada da água fria, as vazões da água (quente e fria) entre outros. Essas são variáveis manipuláveis no sistema estudado neste trabalho, porém alguns fatores são inerentes do sistema estudado como área de troca térmica, material de composição do TC. O objetivo da análise foi identificar qual modelo de aprendizado de máquina poderia prever a temperatura de saída do fluido quente do TC, com maior precisão e exatidão.

Para realizar essa predição foram utilizados modelos de ML previamente disponíveis no software *Orange Data Mining*. Os modelos escolhidos para serem analisados foram: *Regressão Linear*, *Tree* e *Random Forest*.

A Figura 11 mostra algumas opções de modelos disponíveis no software, para predição e visualização dos dados, dentre outros.

**Figura 11.** Modelos de ML disponíveis no Orange Data Mining



Fonte: Autor (2024)

#### 4.4 DADOS PARA TREINAMENTO E TESTE DO MODELO AM

Foi necessário realizar um tratamento nas colunas do banco de dados coletado experimentalmente, pois o software reconhece apenas um título por coluna, e, portanto, elas não podem estar mescladas ou com alguma formatação especial, por isso foi necessário realizar uma reformatação do documento CSV gerado pelo módulo de operação do TC que realiza a coleta os dados do processo.

A Figura 12 mostra o banco de dados sem nenhum tratamento, coletado direto do TC, enquanto a Figura 13 mostra os dados após o tratamento.

Figura 12. Tabela de dados antes do tratamento

Time (s)	$\Delta T_{Hot}$ (°C)	$\Delta T_{Cold}$ (°C)	Average $T_{Hot}$ [ $T_3$ ] (°C)	Average $T_{Cold}$ [ $T_3$ ] (°C)	Specific Heat Capacity $_{Hot}$ ( $J.kg^{-1}.K^{-1}$ )	Specific Heat Capacity $_{Cold}$ ( $J.kg^{-1}.K^{-1}$ )	Density $_{Hot}$ ( $kg.m^{-3}$ )	Density $_{Cold}$ ( $kg.m^{-3}$ )	Power Emitted (W)	Power Absorbed Ener (W)
0	1,9	0,3	21,7	18,0	4180,64	4182,91	997,84	998,59	301,19	35,92
5	1,7	0,6	23,4	18,1	4179,88	4182,84	997,44	998,58	268,15	71,01
10	2,1	0,8	24,9	18,2	4179,34	4182,76	997,07	998,56	331,08	94,67
15	3,1	1,0	26,5	18,4	4178,87	4182,62	996,65	998,52	488,47	119,72
20	2,6	1,4	28,6	18,5	4178,41	4182,55	996,06	998,50	414,81	166,63
25	2,7	1,6	30,0	18,7	4178,20	4182,41	995,64	998,46	421,20	170,38
30	3,2	1,9	31,3	18,9	4178,06	4182,27	995,24	998,42	558,86	202,31
35	2,3	2,3	32,3	19,1	4177,99	4182,14	994,93	998,38	390,39	228,88
40	2,3	2,6	33,1	19,3	4177,95	4182,01	994,67	998,35	399,85	309,38
45	2,4	2,9	33,7	19,2	4177,93	4182,07	994,47	998,36	405,51	431,86

Fonte: Autor (2024)

Figura 13. Tabela de dados depois do tratamento

Time (s)	$\Delta T_{Hot}$ (°C)	$\Delta T_{Cold}$ (°C)	Average $T_{Hot}$ [ $T_3$ ] (°C)	Average $T_{Cold}$ [ $T_3$ ] (°C)	Specific Heat Capacity $_{Hot}$ ( $J.kg^{-1}.K^{-1}$ )	Specific Heat Capacity $_{Cold}$ ( $J.kg^{-1}.K^{-1}$ )	Density $_{Hot}$ ( $kg.m^{-3}$ )	Density $_{Cold}$ ( $kg.m^{-3}$ )	Power Emitted (W)	Power Absorbed (W)	Energy Balance Coefficient	Te
0	1,9	0,3	21,7	18,0	4180,64	4182,91	997,84	998,59	301,19	35,92	0,12	
5	1,7	0,6	23,4	18,1	4179,88	4182,84	997,44	998,58	268,15	71,01	0,26	
10	2,1	0,8	24,9	18,2	4179,34	4182,76	997,07	998,56	331,08	94,67	0,29	
15	3,1	1,0	26,5	18,4	4178,87	4182,62	996,65	998,52	488,47	119,72	0,25	
20	2,6	1,4	28,6	18,5	4178,41	4182,55	996,06	998,50	414,81	166,63	0,40	
25	2,7	1,6	30,0	18,7	4178,20	4182,41	995,64	998,46	421,20	170,38	0,40	
30	3,2	1,9	31,3	18,9	4178,06	4182,27	995,24	998,42	558,86	202,31	0,36	
35	2,3	2,3	32,3	19,1	4177,99	4182,14	994,93	998,38	390,39	228,88	0,59	
40	2,3	2,6	33,1	19,3	4177,95	4182,01	994,67	998,35	399,85	309,38	0,77	
45	2,4	2,9	33,7	19,2	4177,93	4182,07	994,47	998,36	405,51	431,86	1,06	
50	2,5	2,6	34,0	19,0	4177,92	4182,21	994,27	998,40	422,75	430,64	1,00	

Fonte: Autor (2024)

Foram utilizados dois bancos de dados, o banco para treinamento, que deve possuir em torno de 5 mil linhas e foi produzido com os dados coletados no planejamento experimento, e o banco de predição, que foi produzido com parâmetros diferentes do banco de dados de treinamento, foi escolhido um banco de dados gerado por um experimento aleatório de uma matéria de Laboratório de Engenharia Química.

## 4.5 PROCESSO DE MONTAGEM DO LAYOUT DO ORANGE

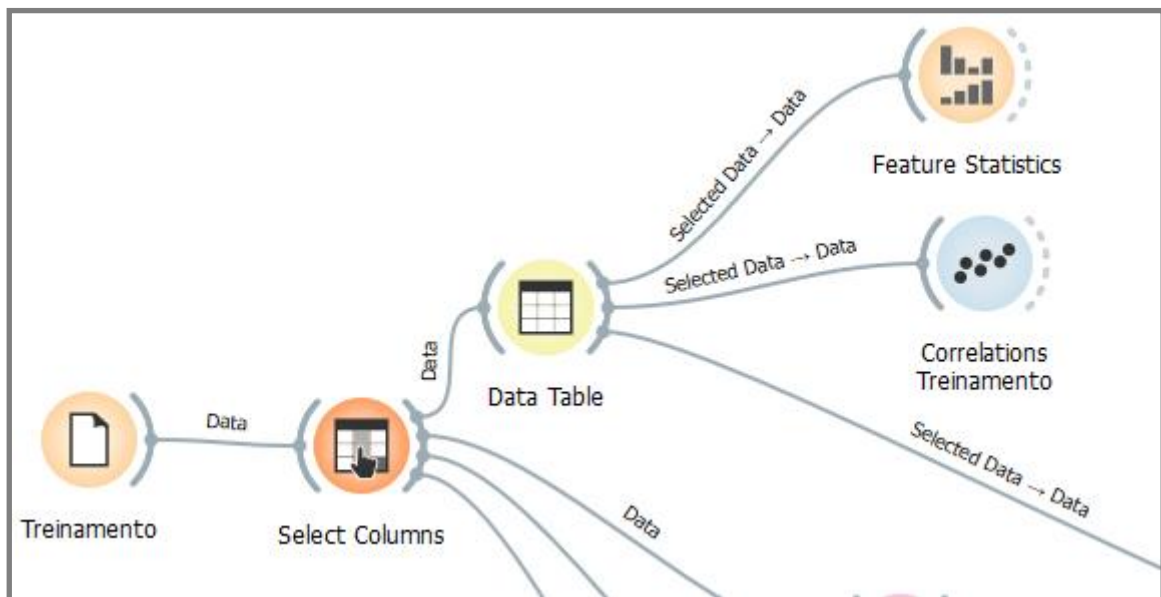
Como condição para que o *software Orange* pudesse fazer predição foi necessário definir três fatores:

1. Um banco de dados de treinamento
2. Um banco de dados de teste
3. Um modelo de aprendizado de máquina

Para demonstrar o uso do *software Orange* foram apresentados alguns prints das telas demonstrando como foi o processo de montagem dos fatores descritos acima.

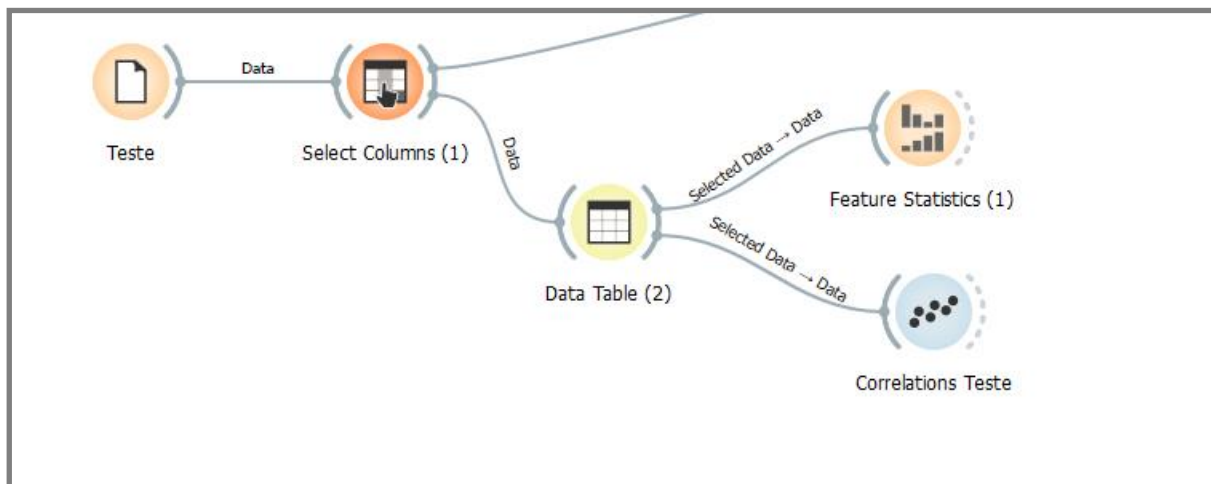
As Figuras 14 e 15 representa a primeira e segunda etapa em relação ao banco de dados de treinamento e teste é feita usando o *widget "File"* que foi nomeado como "Treinamento" para importar o arquivo para o software.

**Figura 14.** Fluxo para análise do banco de dados de treinamento



Fonte: Autor (2024)

Figura 15. Fluxo para análise do banco de dados teste

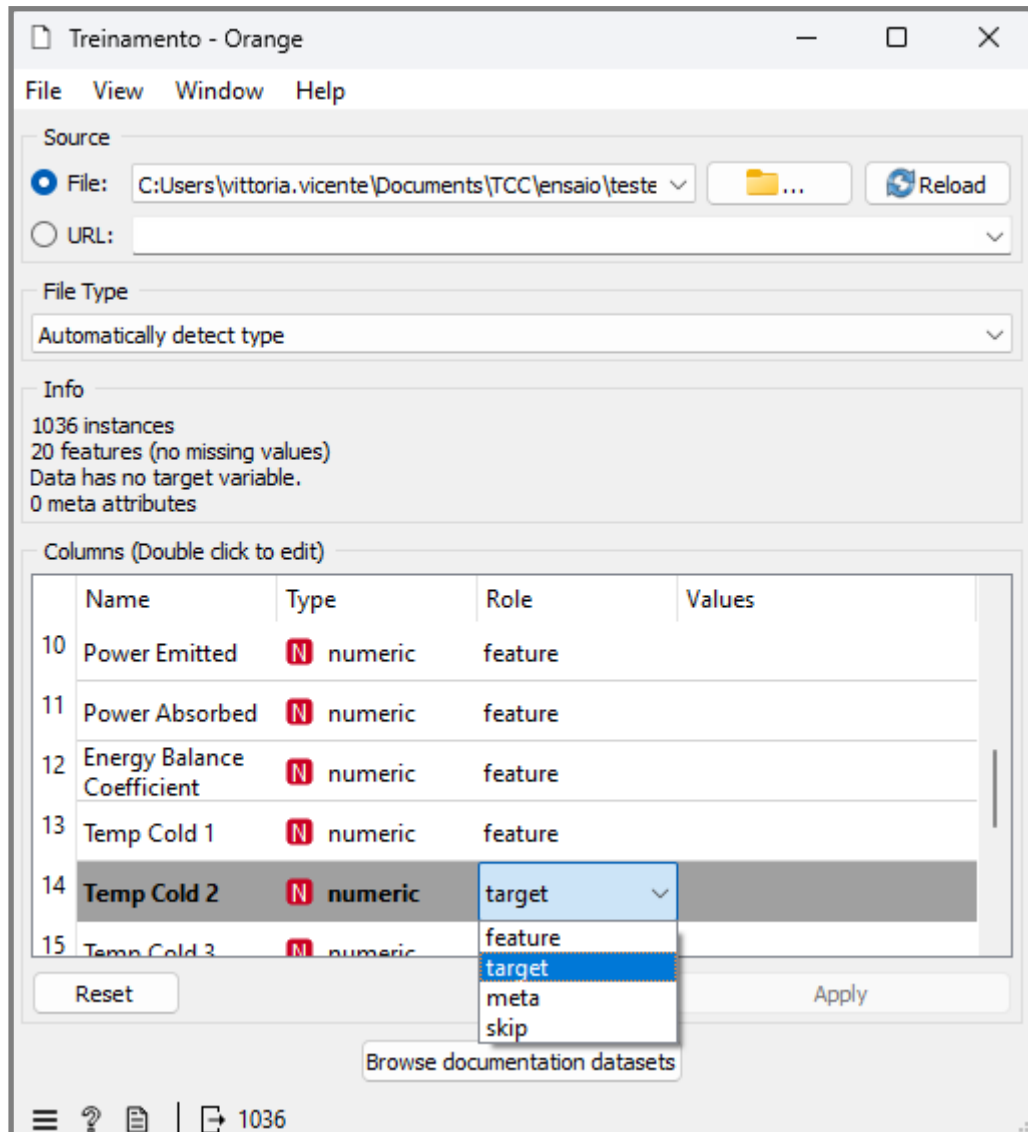


Fonte: Autor (2024)

Depois de alguns testes preliminares notou-se a importância de usar o widget “transform” *Selec Columms* para selecionar a variável *target*, a variável que será prevista pelos modelos de ML, pois foi identificado que o software eventualmente não consegue reconhecer a variável *Target* apenas com a orientação passada no momento da importação do banco de dados. Os widgets *Data Table*, *Feature Statistics* e *Correlations* tem a finalidade de análise do banco de dados, estes que geraram os dados apresentados no ponto 4.3.1 e 4.3.2.

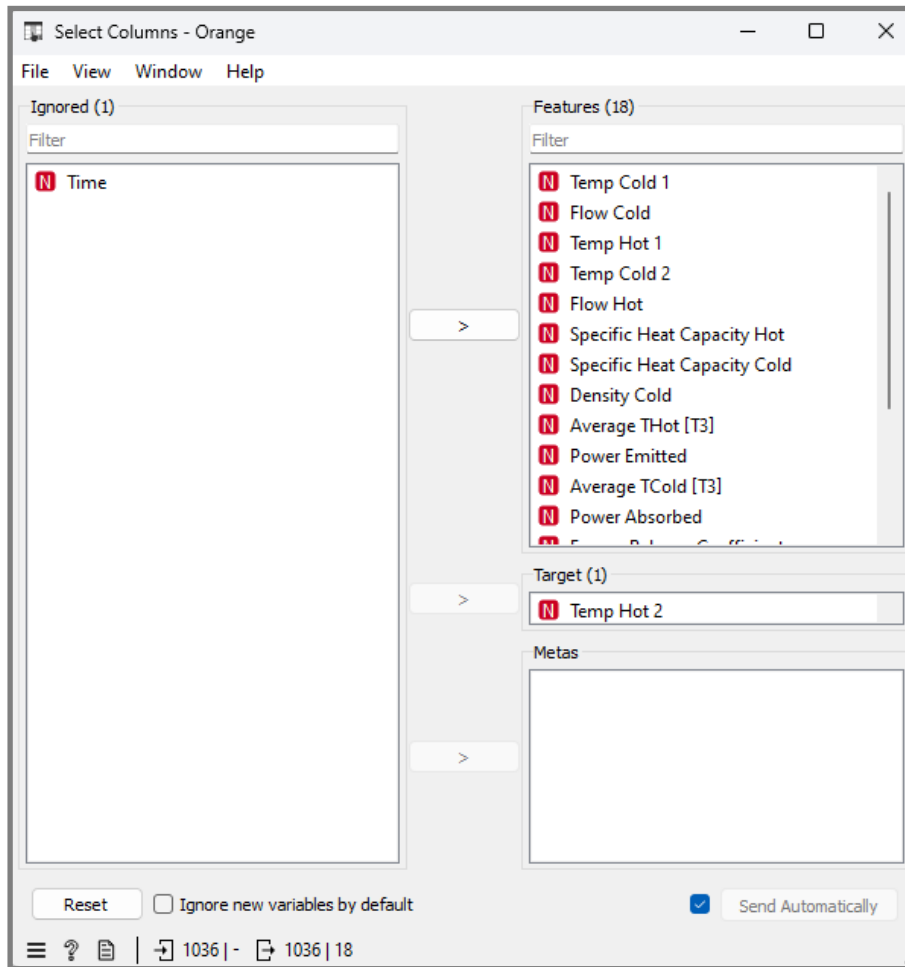
As telas exibidas nas figuras 16 e 17 mostram como foi feita a definição da variável *target* no início do processo, quando é feita a importação do banco de treinamento, e no *widget selec columns*.

Figura 16. Local para selecionar a variável *target* no banco de dados



Fonte: Autor (2024)

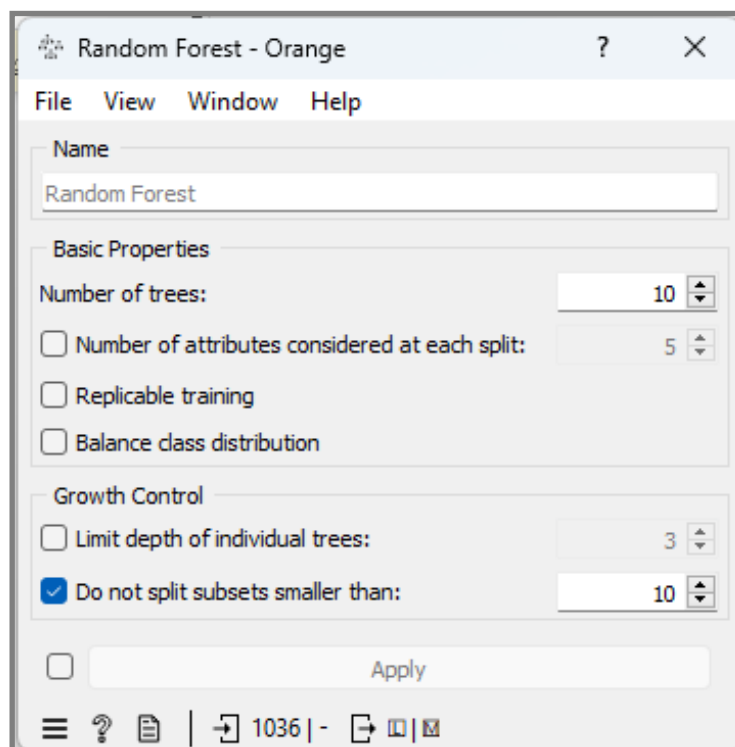
**Figura 17.** Local para selecionar a variável *target* no *Select Columns*



Fonte: Autor (2024)

Em seguida, os modelos de ML foram adicionados ao *layout*, neste caso foram escolhidos os modelos: *Tree*, *Random Forest* e *Linear Regression*. É possível editar parâmetros dentro de cada modelo, como mostra a Figura 18, neste caso foi alterado para 10, tanto para o Random Forest quanto para o Tree, o número de subsets mínimos criados durante a tomada de decisão.

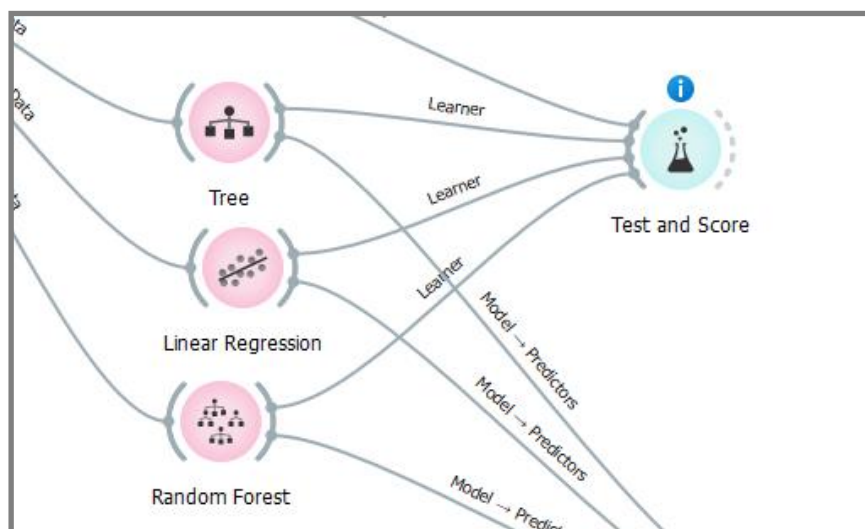
Figura 18. Configuração do ML *Random Forest*



Fonte: Autor (2024)

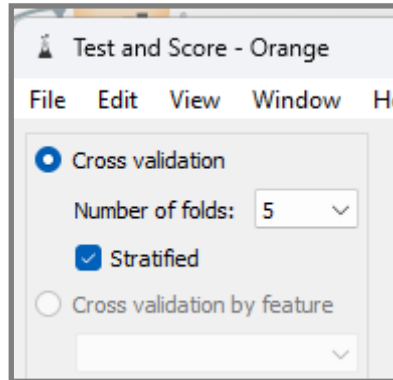
Também foi adicionado o widget *Test and Score*, nele foi possível testar o desempenho dos modelos *Learners* com o banco de dados de treinamento, internamente é possível definir qual tipo de validação fazer, como: *Cross validation* e *Random Simpling*. No caso, neste trabalho foi utilizado o *5 folds Cross Validation*.

Figura 19. Fluxo de teste dos modelos de ML



Fonte: Autor (2024)

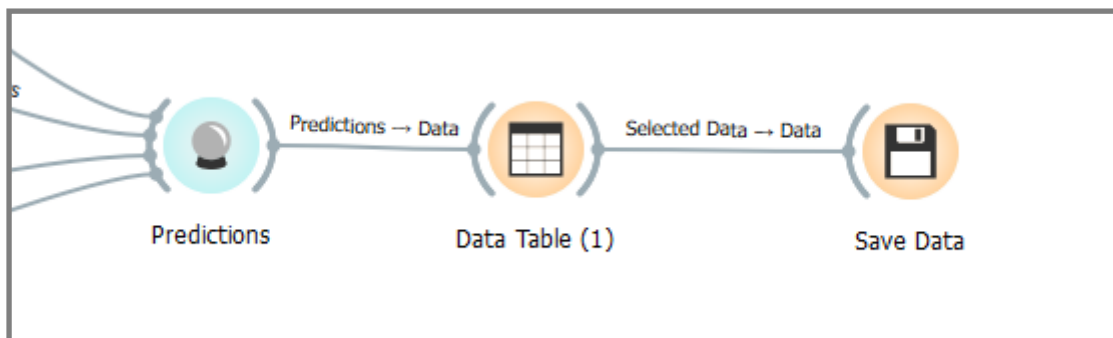
**Figura 20.** Configuração do método de *Cross Validation*



Fonte: Autor (2024)

Para realizar a comparação entre os modelos e a análise de resultados foi utilizado o *Predictions*, ao abri-lo é possível identificar a predição e a sua comparação com o valor real disponibilizado no banco de dados de predição.

**Figura 21.** Fluxo para salvar dados gerados na predição



Fonte: Autor (2024)

A figura 22 mostra em cinza as temperaturas experimentais do fluido quente de saída (Temp Hot 2), enquanto as colunas à esquerda da tabela mostram os resultados das predições dos modelos *Tree*, *Linear Regression* e *Random Forest*, com os respectivos erros associados.

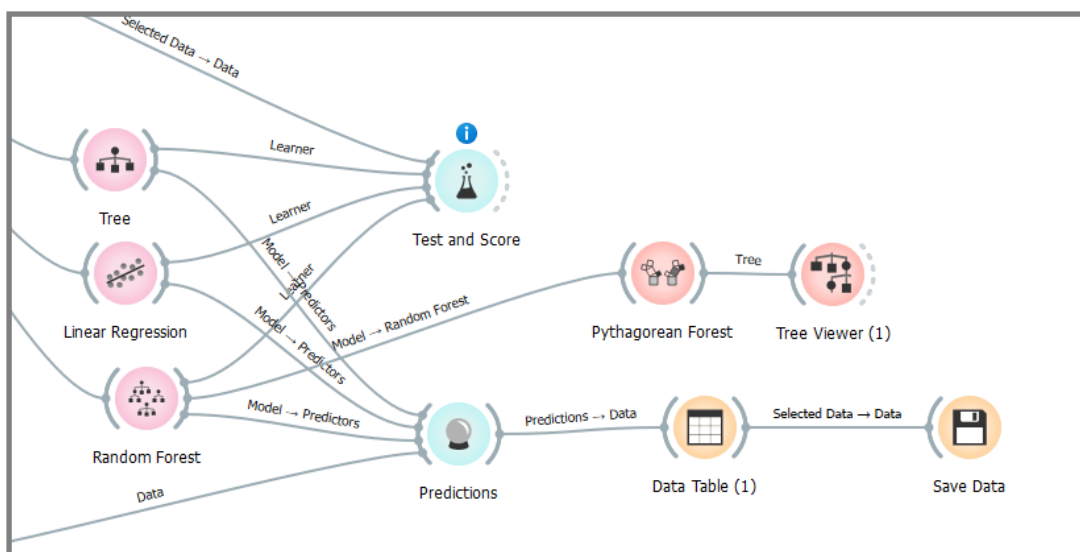
Figura 22. Dados gerados na predição

	Linear Regression	error	Tree	error	Random Forest	error	Temp Hot 2	Temp Cold 1	Flow Cold	Temp Hot 1	Flow Hot
25	42.1	2.1	43.6	3.6	44.1	4.1	40.0	24.4	3.13	47.0	0.41
26	42.1	2.1	43.6	3.6	44.1	4.1	40.0	24.4	3.13	47.0	0.41
27	42.1	2.1	43.6	3.6	44.1	4.1	40.0	24.4	3.13	47.0	0.41
28	42.1	2.1	43.6	3.6	44.1	4.1	40.0	24.4	3.13	47.0	0.41
29	42.1	2.1	43.6	3.6	44.1	4.1	40.0	24.4	3.13	47.0	0.41
30	42.1	2.1	43.6	3.6	44.1	4.1	40.0	24.4	3.13	47.0	0.41
31	42.1	2.1	43.6	3.6	44.1	4.1	40.0	24.4	3.13	47.0	0.41
32	42.1	2.1	43.6	3.6	44.1	4.1	40.0	24.4	3.13	47.0	0.41
33	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.41
34	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.41
35	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.41
36	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.41
37	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.41
38	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.41
39	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.42
40	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.42
41	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.42
42	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.42
43	42.0	2.0	43.6	3.6	44.1	4.1	40.0	24.4	3.13	46.9	0.42

Fonte: Autor (2024)

Na figura 23, é possível identificar alguns widget de visualização, como o Pythagorean Forest e Tree Viewer. Estes widget mostram formas gráficas do modelo de predição, como foi montado o raciocínio para realizar a predição para cada amostra.

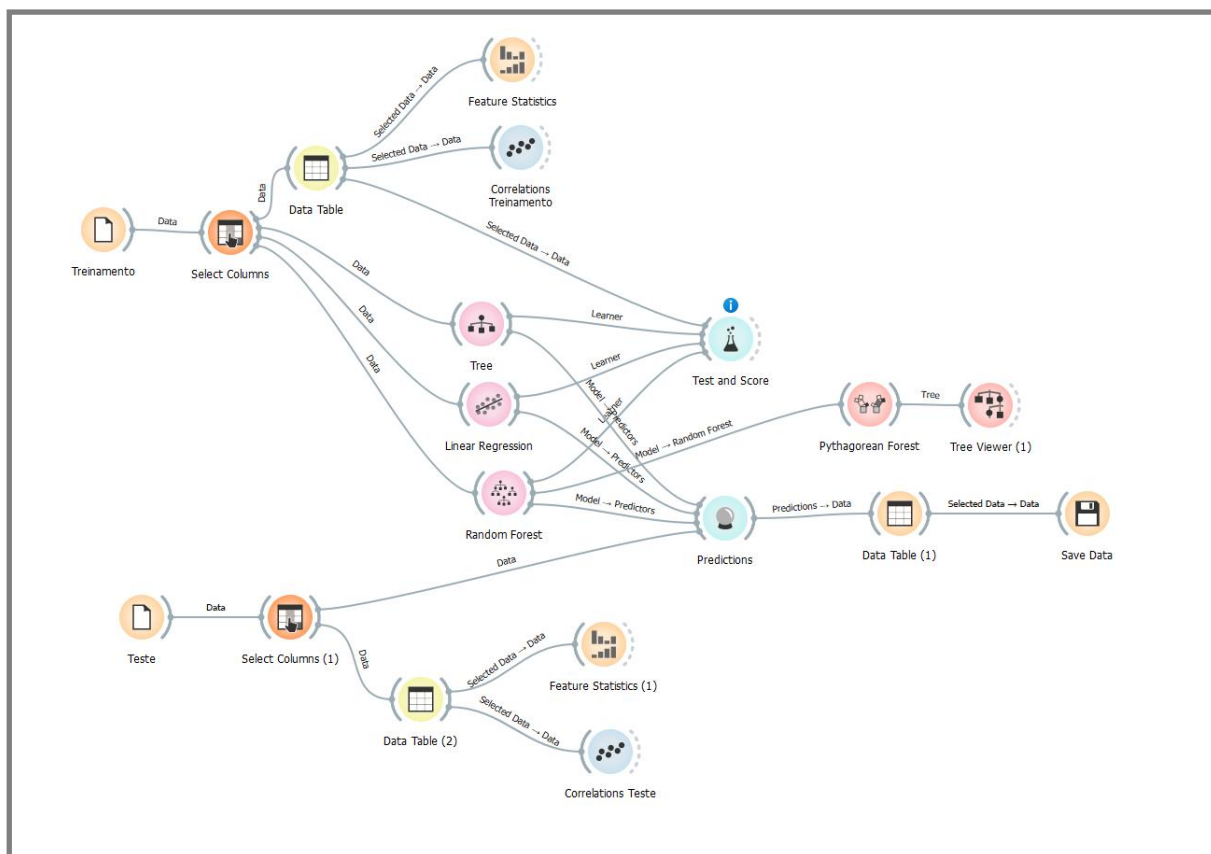
Figura 23. Fluxo para visualizar as árvores geradas pelos modelos de predição



Fonte: Autor (2024)

A figura 24 representa o layout final com todos os *widgets* corretamente conectados.

**Figura 24.** Fluxo completo



Fonte: Autor (2024)

#### 4.6 AVALIAÇÃO DO MODELO AM

Para a análise dos dados coletados foi utilizado o teste de *pearson* para entender a relação de influência entre as variáveis coletadas, o grau de relação linear entre elas, e gráficos de distribuição, para entender o perfil de cada banco de dados. Os resultados foram apresentados no ponto 4.3 Avaliação dos bancos de dados.

Para analisar o desempenho dos modelos de aprendizado de máquina, normalmente utiliza-se a verificação da diferença entre o valor real e previsto pelo modelo, para isso foi utilizado o MAE (erro absoluto médio) e RMSE (erro quadrático médio).

Um erro pode ser definido como a diferença entre o experimental e o real coletado no banco teste. O erro absoluto médio (MAE) é calculado tomando a média do erro. O erro quadrático médio (RMSE) é usado para determinar a proximidade do valor previsto com o valor real. Todas as fórmulas das métricas estão definidas no ponto 3.7.2 Métricas de desempenho.

De acordo com DEMSAR J et.al (2013), pode ser usado algumas análises estatísticas como: Matriz da confusão, acurácia, curva ROC, precisão, entre outros, para sistemas de classificação. E para sistemas de variáveis contínuas, como temperaturas, são utilizados indicadores de precisão diferentes, por exemplo: RMSE, MSE, MAE, R<sup>2</sup>, que são indicadores baseados nos erros e nos desvios dos valores previstos do modelo comparado com o valor real.

Além disso, foram utilizados gráficos de linhas para fazer uma comparação visual dos desempenhos dos modelos de predição e analisar o comportamento de cada um deles.

O processo de troca de calor trata-se de um sistema de variáveis contínuas visto que a temperatura de saída não é uma variável binária, ela tem uma temperatura de objetivo, mas ela sempre está num espectro e o que define um resultado bom ou ruim é o quão perto ela está do objetivo.

## 4.7 QUESTÕES PRÁTICAS DO SOFTWARE

Neste tópico foi pontuado algumas dicas e recursos para facilitar o uso do software Orange Data Mining.

- É importante utilizar o widget Selec Collums para selecionar a variável target, as vezes o modelo não reconhece ela apenas com a seleção no widget “Data” na configuração do banco de dados.
- É possível que se o banco de dados utilizado tiver uma variável para marcar o tempo, esta pode atrapalhar e confundir o modelo, em testes informais é possível verificar um melhor desempenho do modelo quando a variável de tempo não está incluída no banco de dados de treinamento.
- É importante ressaltar que o desempenho dos modelos foi, provavelmente, afetado pela falta de aleatoriedade nos dados de treinamento, além da possibilidade de overfitting. Para futuros trabalhos, recomenda-se a obtenção de um conjunto de dados de treinamento mais diversificado e aleatório, com mais perturbações nas variáveis de entrada, bem como a realização de experimentos adicionais para validar os modelos desenvolvidos.

## 5. RESULTADOS E DISCUSSÕES

Os resultados deste trabalho estão divididos em 4 partes principais. A primeira mostra o desempenho do modelo de aprendizado de máquina com os dados de treinamento, apresentando indicadores estatístico como  $R^2$  e RMSE, na segunda parte a mesma análise é feita, porém agora com o modelo de predição criado pela primeira etapa e com um banco de dados teste.

Na terceira parte são apresentadas as árvores de decisão geradas pelo algoritmo Random Forest, que foi o que teve o melhor desempenho. E na última parte é feita uma comparação visual entre o desempenho dos três modelos com os dados originais.

### 5.1 AVALIAÇÃO DOS BANCOS DE DADOS

Para entender o perfil dos bancos de dados utilizados é importante realizar algumas análises estatísticas como o teste de Pearson, que analisa relação linear entre as variáveis do processo e análises estatísticas para entender em qual espectro dos limites das variáveis do TC cada banco de dados se encontra.

#### 5.1.2 Banco de dados de treinamento

O banco de dados de treinamento segue padrões específicos pré-definidos no planejamento experimental. Analisar esse ponto é importante pois quanto mais próximo o banco de dados de treinamento é do banco teste, melhor serão os resultados do teste final, ou seja, quanto mais situações o banco teste abranger melhor será o modelo de predição gerado. A tabela 5 apresenta os valores dos testes estatísticos relacionados ao banco de dados de treinamento, gerado pelo ensaio feito com o planejamento experimental.

Tabela 5. Análise estatística de cada variável do banco de dados de treinamento

Variável	Média	Moda	Mediana	Dispersão	Min	Máx
Temp Cold 1	21,089	21,000	21,100	0,047	17,500	23,200
Flow Cold	1,051	1,040	1,080	0,395	0,830	2,760
Temp Hot 1	48,512	57,700	48,700	0,131	22,700	59,200
Temp Cold 2	25,463	25,700	26,700	0,065	18,000	20,000
Flow Hot	1,470	0,980	1,040	0,391	0,600	2,520
Specific Heat Capacity Hot	4179,596	4176,260	4179,460	0,001	4177,930	4182,360
Specific Heat Capacity Cold	4179,826	4179,760	4179,730	0,001	4179,210	4182,910
Density Cold	997,370	997,370	997,340	0,001	996,970	998,590
Average THot [T3]	45,870	39,800	46,400	0,121	21,700	55,600
Power Emitted	434,792	201,400	458,045	0,369	35,920	890,940
Energy Balance Coefficient	0,919	0,940	0,930	0,082	0,120	1,550
Temp Hot 3	45,877	39,800	46,400	0,121	21,700	55,800
Temp Cold 3	23,643	23,700	23,800	0,053	18,00	25,300
D TCold	4,394	3,700	4,300	0,291	0,300	6,800
D THot	4,927	7,500	4,600	0,301	1,700	7,600
Density Hot	989,736	992,290	989,620	0,002	985,300	997,840
Temp Hot 2	43,585	50,200	44,400	0,114	20,800	53,000

Analisar a relação das variáveis presentes no banco de dados de treinamento é importante pois essas relações que vão personalizar o modelo de aprendizado de máquina do software, com essas relações os modelos aprendem como o processo de troca de calor funciona e como a variável objetivo (target) se relaciona com cada uma das demais variáveis e cria um modelo de predição.

Analisando a Tabela 6, que apresenta o grau de correlação de pearson entre cada variável do banco de dados e a variável target, é possível notar que as primeiras variáveis, 1 a 5, da tabela têm o grau de correlação mais alto, pois são a mesma corrente em pontos diferentes do percurso e indicadores físico-químicos diretamente ligados com a temperatura da água.

Tabela 6. Grau de relação entre as variáveis auxiliares e a variável target no banco de dados de treinamento

	Variável relacionada	Fator de correlação de Pearson
1	Temp Hot 3	+0,999
2	Density Hot	+0,995
3	Temp Hot 1	+0,994
4	Specific Heat Heat Capacity Hot	+0,940
5	D Tcold	+0,922
6	Temp Cold 2	+0,787
7	Power Absorbed	+0,652
8	Temp Cold 1	+0,610
9	Density Cold	-0,610
10	Specific Heat Heat Capacity Cold	-0,601
11	Power Emitted	+0,581
12	Energy Balance Coefficient	+0,466
13	Temp Cold 3	-0,204
14	Flow Cold	+0,140
15	Flow Hot	-0,073

### 5.1.3 Banco de dados de teste

O banco de dados teste é um banco de dados gerado aleatoriamente e ele não tem nenhuma relação com o banco de dados de treinamento. A tabela 7 apresenta os valores estatísticos das variáveis banco de dados de teste.

Tabela 7. Análise estatística de cada variável do banco de dados teste

Variável	Média	Moda	Mediana	Dispersão	Min	Máx
Temp Cold 1	24,390	24,400	24,400	0,002	24,300	24,500
Flow Cold	2,475	0,630	3,080	0,431	0,000	3,200
Temp Hot 1	52,415	56,500	55,400	0,084	44,400	56,600
Flow Hot	1,469	0,680	0,680	0,806	0,000	3,180
Density Hot	988,237	987,020	987,020	0,0017	986,460	991,030
Temp Hot 2	46,443	49,400	47,900	0,077	39,800	51,300

Na Tabela 8 é possível notar que a variável 5 não tem uma forte relação direta com a variável target, neste caso o peso dela no modelo de predição será mais baixo, pois o software não conseguiu achar uma relação no comportamento da Temperatura de entrada fria e a temperatura de saída quente. Provavelmente, porque a primeira se mantém constante, em temperatura ambiente, enquanto

temperatura quente de saída oscila conforme as variações das MVs no decorrer do experimento.

Tabela 8. Grau de relação entre as variáveis auxiliares e a variável target no banco de dados teste

	Variável relacionada	Fator de correlação de Pearson
1	Density Hot	-0,964
2	Temp Hot 1	+0,889
3	Flow Cold	-0,431
4	Flow Hot	-0,073
5	Temp Cold 1	-0,031

A análise feita no banco de dados de treinamento foi de extrema importância para definir quais variáveis seriam utilizadas no banco de dados teste, pois foi possível identificar as variáveis que possuem a correlação mais forte e pré selecioná-las, como foi o caso da Density Hot. Foi identificada a importância dela nos resultados da Tabela 6 e quando ela foi inserida no banco de dados teste essa correlação forte se repetiu.

## 5.2 AVALIAÇÃO DOS MODELOS DE APRENDIZADO DE MÁQUINA

### 5.2.1 Treinamento dos modelos de aprendizado com o banco de dados de treinamento

Nesta etapa, foram avaliados o desempenho dos modelos de aprendizado escolhidos com o banco de dados teste, usando a *Cross Validation* com 5 *folders*.

Tabela 9. Desempenho dos modelos learners com o banco de dados treinamento

Modelo	MSE	RMSE	MAE	MAPE	R2
Linear Regression	0,000	0,000	0,000	0,000	1,00
Random Forest	0,103	0,320	0,096	0,003	0,996
Tree	0,137	0,370	0,086	0,003	0,994

Analisando a Tabela 9, nota-se que os modelos tiveram um bom desempenho com os dados testes, já que os erros estão baixos e o coeficiente de determinação bem próximo de 1, ou seja, as funções dos modelos se ajustaram praticamente de forma perfeita aos dados.

É importante pontuar que quanto mais detalhado é o banco de dados de treinamento, quanto mais rico forem os dados e variáveis, melhor será o “aprendizado” do modelo sobre aquele banco de dados.

### 5.2.2 Testes dos modelos de predição

Nesta etapa, foi testado o modelo de predição criado pelo treinamento a partir do banco de dados teste. Então, é importante lembrar que o Orange já possui alguns *Learners*, que seriam metodologias de cálculo pré-definidas, estes modelos irão “aprender” como o processo funciona e personalizar a metodologia de cálculo pré-definida. Por exemplo, o *Learner Linear Regression* segue a metodologia de cálculo de uma função linear, essa etapa de personalização serve para definir os coeficientes da função com base nos dados do treinamento; e essa função, agora com coeficientes definidos, será utilizada para calcular os resultados Y usando as entradas do banco de dados teste X.

A tabela 10 apresenta os resultados do desempenho dos modelos de predição, testados com o banco de dados de teste.

Tabela 10. Desempenho dos modelos de predição com o banco de dados teste

Modelo	MSE	RMSE	MAE	MAPE	R2
Random Forest	2,363	1,537	1,249	0,026	0,814
Tree	2,598	1,612	1,283	0,027	0,796
Linear Regression	5,237	2,288	2,079	0,045	0,588

É possível notar que o modelo *Random Forest* teve o maior destaque, o seu erro foi menor que o dos outros modelos, mesmo que a faixa de variação tenha sido pequena para os outros modelo de predição. O  $R^2$  é um fator importante para entender como o modelo explica a variância dos dados, ou seja, o quão sensível o modelo foi para conseguir capturar a dinâmica do comportamento das amostras. Mesmo com os indicadores de erro com resultados bons é importante avaliar como a curva do modelo se ajustou aos dados do processo, pois esta análise também é relevante para avaliar o quão bom será o desempenho do modelo para situações distintas.

Um ponto importante a se destacar é a diferença nítida de desempenho entre os resultados estatísticos dos modelos de aprendizado com os dados de treinamento e na predição, gerado com os dados testes. Esta diferença não necessariamente quer dizer que o modelo não se encaixa nesse processo, por isso é necessário validar como este modelo foi treinado e como ele foi testado, ou seja, realizar uma interpretação dos bancos de dados utilizados.

Analisando os indicadores estatísticos, é possível identificar que não foram gerados *outlier* grosseiros, pois de acordo com WESNER (2016) é possível

verificar essa existência (de *outlier*) quando temos discrepâncias entre as métricas RMSE e MAE, neste caso quando existem muitos *outliers* o RMSE pode ser até três vezes maior que o MAE, o que não foi o caso dos resultados mostrados na Tabela 10.

Como hipótese para a falta de ajuste, pode-se dizer que esse comportamento pode ter ocorrido por causa de um evento de *overfitting*; uma das hipóteses para este tipo de evento é quando na fase de treinamento o modelo entende tão bem a relação entre as variáveis que acaba decorando o comportamento e tenta replicar na predição, que é um cenário nunca visto onde esse comportamento decorado não se aplica, gerando um resultado bem diferente da curva original.

Entende-se que o modelo de aprendizado leu todos os detalhes do banco de dados de treinamento, que era muito rico com 15 variáveis, grande parte diretamente relacionadas com a variável target, por isso o modelo conseguiu aprender os mínimos desvios do processo, mas quando se depara com situações diferentes com dados com menos informações, como é o caso do banco de dados teste, em que foram inseridas apenas 5 variáveis de entrada (no banco teste), ele não conseguiu demonstrar um desempenho tão bom, pois aparentemente precisaria de mais variáveis, que representassem diferentes condições do trocador de calor, para fazer uma predição mais precisa.

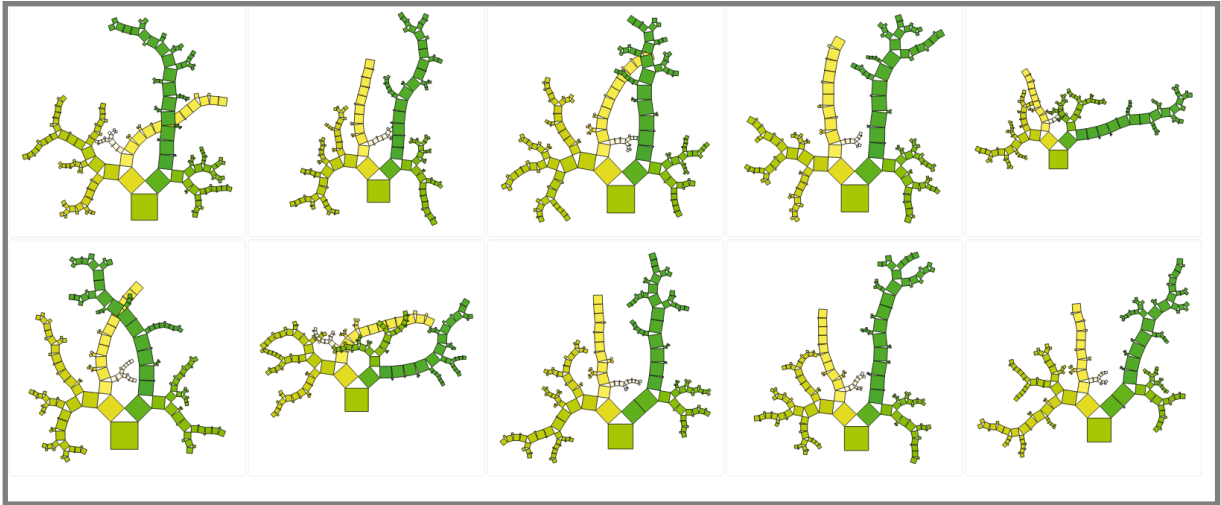
O *overfitting* pode ser corrigido realizando alguns tratamentos no banco de dados testes, identificando e tirando *outliers*, se existir, e entre outros dados que possam confundir o modelo. Além disso, é necessário aumentar o banco de treinamento para abranger outros eventos do processo, além de aumentar a aleatoriedade do banco de dados.

### 5.2.3 Árvores de decisão geradas

Na, figura 25 é possível ver as 10 árvores de decisão geradas pelo algoritmo *Random Florest*. Por meio delas podemos analisar as faixas de valores que o “raciocínio” do modelo seguiu em seu processo decisório.

Além disso, é notável uma tendência de expansão dos ramos em direção à região verde. Esses gráficos são gerados levando em consideração a intensidade da cor, a qual varia de acordo com a média dos valores da temperatura. Com a presença de uma grande quantidade de ramificações nas árvores, podemos inferir uma boa adaptação do modelo, o que reforça a característica não linear dos dados do processo em questão.

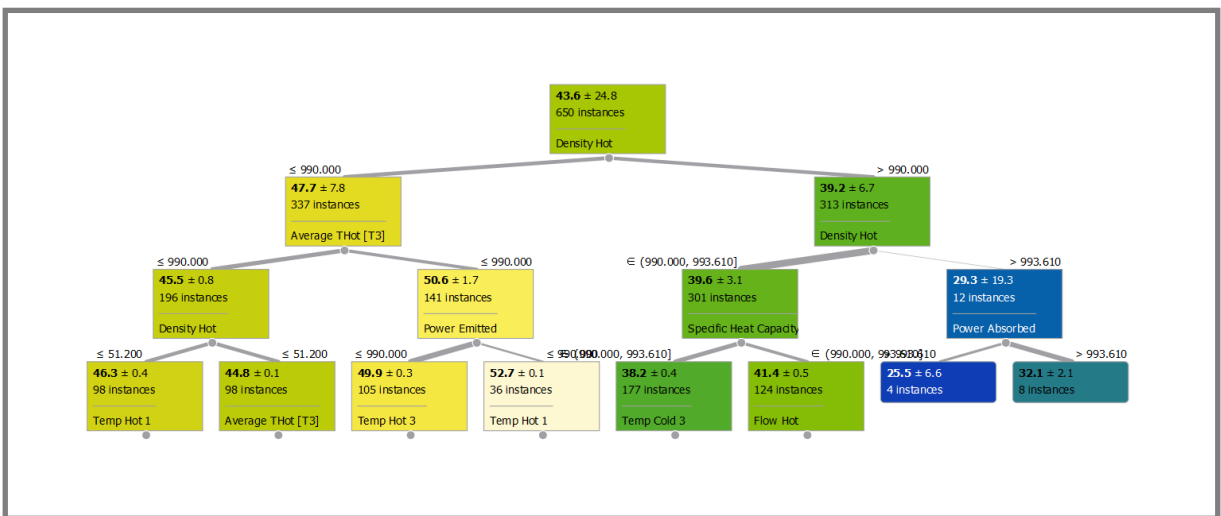
**Figura 25.** Árvores de decisão geradas pelo modelo de predição *Random Forest*



Fonte: Autor (2024)

A figura 26 apresenta a seleção de uma das árvores, e a partir dela é possível verificar como o modelo de predição fez as escolhas e interpretou os dados de entrada para prever o valor final da temperatura de saída do fluido quente (*Temp Hot 2*). Nota-se que a variável mais importante é a *Density Hot*, que aparece sempre como a variável para definir as faixas de densidade para as divisões dos nós.

**Figura 26.** Mapa mental de uma das árvores da floresta gerada pelo modelo de predição *Random Forest*



Fonte: Autor (2024)

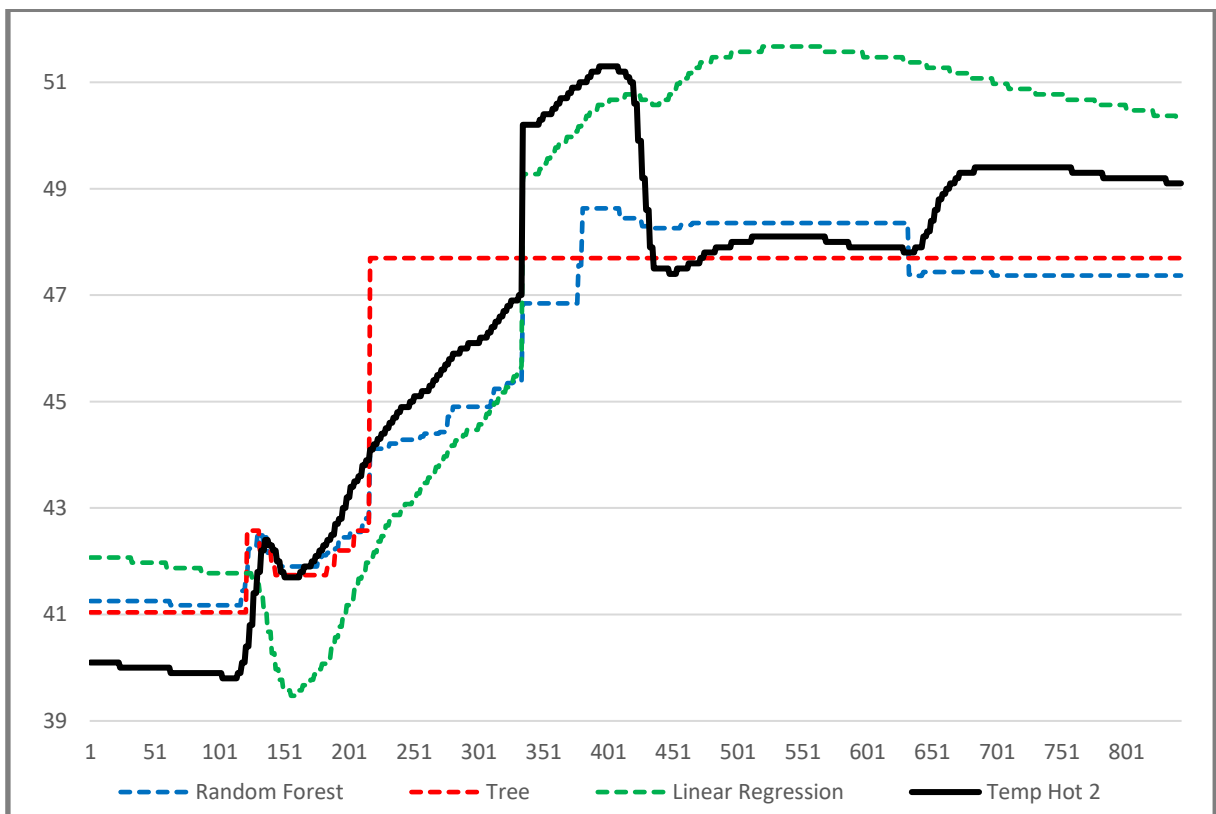
As figuras 25 e 26 estão disponibilizadas em formato maior nos anexos.

## 5.2.4 Comparação entre modelos

Nesta etapa foi feita uma comparação de desempenho entre os três modelos de aprendizado de máquina escolhidos: *Random Forest*, *Linear Regression* e *Tree*.

A partir da figura 27 é possível analisar o comportamento dinâmico dos resultados de cada um dos modelos de predição, criados com base nos modelos de aprendizado, comparando com os dados experimentais de temperatura de saída do fluido quente.

Figura 27. Perfil de temperatura experimental vs real



Fonte: Autor (2024)

Nota-se que o modelo que mais se aproximou do comportamento original foi o *Random Forest*. O modelo de predição *Tree* teve bons resultados estatísticos, porém ao analisar o gráfico é possível identificar que o perfil de temperatura simulado não mostrou valores tão próximos ao experimental.

Para testar se a falta de ajuste do modelo pudesse ser melhorada a partir de um banco de dados de treinamento mais rico e aleatório, seria necessário a obtenção de um novo conjunto de dados experimentais, no entanto, devido a

problemas técnicos com o equipamento utilizado, não foi possível a repetição da etapa experimental de obtenção de um novo banco dados experimentais, para confirmar esta hipótese.

Contudo, os modelos *Random Forest* e *Tree* foram capazes de aprender e a fazer a predição das temperaturas de saída do fluido quente, o que era esperado, tendo em vista que os trocadores de calor apresentarem um comportamento geral de sistema não linear, conforme também pode ser confirmado por PĂTRĂȘCIOIU; MARINOIU, (2010) que afirma que o melhor modelo matemático para representar o comportamento dos trocadores de calor são sistemas de equações não lineares. Por este motivo os modelos *Random Forest* e *Tree* conseguiram se adaptar melhor aos dados já que são baseados em modelos matemáticos não lineares, o oposto do modelo *Linear Regression*, que teve o pior desempenho.

O artigo de NAIR et al., (2022) também confirma o resultado encontrado neste trabalho, em que mostra que o o *learner* que melhor se adapta ao processo de trocador de calor é o *Random Forest*. Os resultados deste artigo, usando o modelo *Random Forest*, mostram que o MAE e RMSE foram 1,1755 e 1,5422, respectivamente; estes indicadores estatísticos foram gerados na fase de treinamento dos modelos, ou seja, eles foram gerados para avaliar qual modelo seria mais adequado para trocadores de calor e não efetivamente fazendo previsões.

Quando comparamos os resultados de Nair et al. (MAE = 1,1755 e RMSE = 1,5422) com os valores obtidos neste trabalho (MAE = 1,249 e RMSE = 1,537) pode-se afirmar que o learner *Random Forest* teve um desempenho bem próximo em ambos trabalhos, reforçando e confirmando a afirmação de (PĂTRĂȘCIOIU; MARINOIU, 2010).

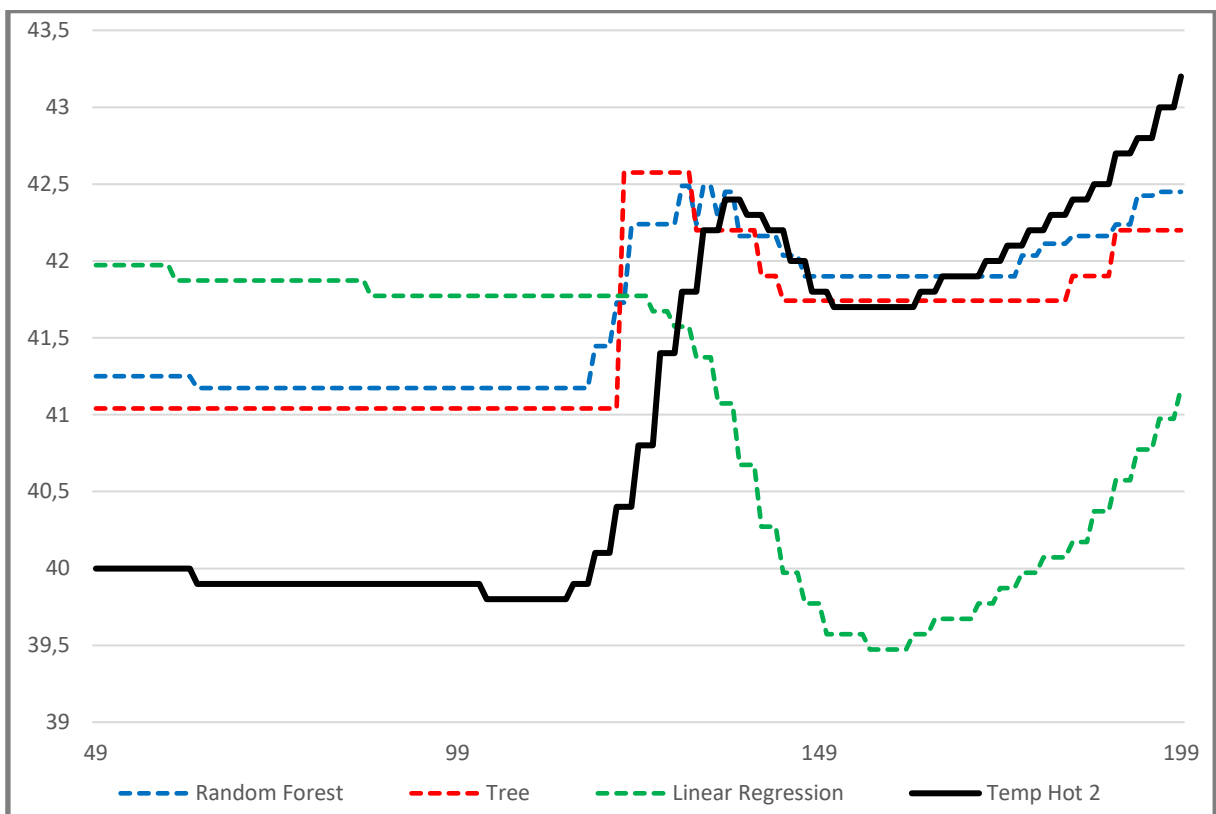
A diferença de desempenho entre o treinamento e o teste dos modelos de predição também pode ter acontecido devido à opção pelos parâmetros padrão do Orange, os quais podem ter normalizado o ajuste dos modelos dentro de faixas que impediram o aprendizado. A falta de aleatoriedade dos dados de treinamento é outro fator que pode ter contribuído para a diminuição da precisão destes modelos, além da possibilidade de ter ocorrido *overfitting*.

A etapa de Planejamento de ensaios experimentais (Item 4.1) pode não ter contribuído para o aprendizado dos modelos, visto que os dados de treinamento foram delineados para abranger todos os limites de temperatura do trocador de calor, porém de forma padronizada e respeitando um tempo de duração e

uma mudança de *setpoint* bem específica, provavelmente gerando um banco de dados de treinamento em que faltou aleatoriedade.

A partir da análise das figuras 28 e 29 é possível visualizar o comportamento dos modelos de predição criados pelos modelos de aprendizado de máquina com o banco de dados de treinamento e testados com o banco de dados teste.

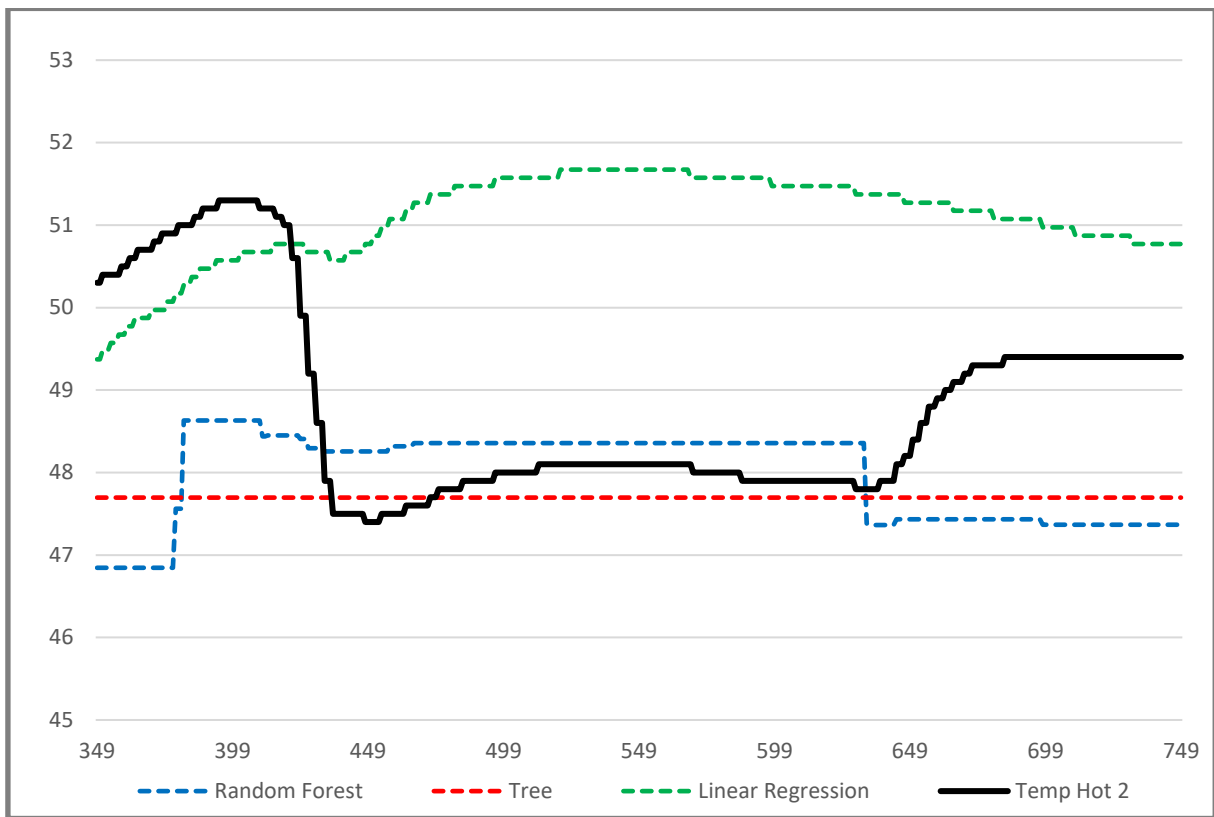
**Figura 28.** Recorte temporal do perfil de temperatura - Momento 1



Fonte: Autor (2024)

A figura 28, mostra que o modelo de *Linear Regression* teve um comportamento totalmente oposto ao experimental, quando Temp Hot 2 está subindo o modelo entende que a temperatura está caindo e este tipo de evento é refletido pelo aumento do erro das análises estatísticas deste modelo.

Figura 29. Recorte temporal do perfil de temperatura - Momento 2



Fonte: Autor (2024)

A partir da análise da figura 29, é possível identificar a relevância de uma análise gráfica, pois apesar do modelo *Tree* ter tido um bom desempenho nos indicadores estatísticos que se deve, provavelmente, ao comportamento do período de tempo entre 499 ao 649 onde as curvas, experimental e simuladas tiveram uma diferença mínima, contudo, analisando um recorte temporal maior, como entre o período de tempo 349 ao 749, é possível identificar que o modelo não respondeu aos estímulos de mudança de temperatura e não aprendeu o suficiente para ter um nível de sensibilidade capaz de simular resultados eficientes, sobretudo quando o sistema é submetido às perturbações nas variáveis de entrada. Esta resposta do modelo é fundamental, já que, em sistemas industriais reais, até mais relevante do que a manutenção das condições do sistema dentro de um patamar estreito está a previsibilidade das respostas às perturbações das variáveis de entrada, as quais precisam ser constantemente serem controladas.

Dessa forma, analisando os dados apresentados e os gráficos de comportamento pode-se concluir que o modelo de predição criado pelo modelo de

aprendizado *Random Forest* foi o que mostrou melhor desempenho estatístico, dentre os modelos selecionados, e que foi o modelo mais capaz de refletir às perturbações das variáveis de entrada do sistema, conforme mostrado na figura 28 (momento 1). Porém, ainda com ressalvas devido à falta de sensibilidade para prever todos os comportamentos das oscilações de temperatura esperados.

## 6. CONCLUSÃO

A adoção da técnica de Machine Learning possibilitou uma análise mais aprofundada e precisa do comportamento do sistema estudado.

Os resultados obtidos demonstraram que os modelos de aprendizado de máquina tiveram um bom desempenho ao prever a temperatura de saída do fluido quente. Em particular, o modelo Random Forest destacou-se por apresentar os menores erros e uma adaptação mais eficaz aos dados experimentais. Por outro lado, o modelo de Regressão Linear mostrou-se menos adequado para lidar com a complexidade do sistema estudado.

Este estudo permitiu a comparação dos resultados obtidos com os de outros trabalhos encontrados na literatura, evidenciando a consistência dos modelos que apresentaram melhores desempenhos. Além disso, foi possível confirmar a não linearidade do comportamento do sistema estudado, o que reforça a necessidade de abordagens mais sofisticadas, como os modelos de Machine Learning, para sua análise e predição com precisão.

A comparação entre os modelos de aprendizado de máquina também revelou a importância de uma análise gráfica detalhada, destacando nuances de desempenho que não são capturadas apenas pelos indicadores estatísticos. Observou-se que, apesar de alguns modelos apresentarem bons resultados estatísticos, a análise visual revelou diferenças significativas em seu comportamento, especialmente em resposta a mudanças nas variáveis de entrada.

## 7. REFERÊNCIAS

BARBETTA, P. A. **Estatística aplicada às Ciências Sociais**. 5. ed. Florianópolis : Editora da UFSC, 2003.

BEJAN, A. **A Transferência de Calor**. São Paulo: Edgard Blucher, 2004.

BERNUY, L. et al. Avaliação do aumento da transferência de calor e de propriedades térmicas em trocadores de calor utilizando nanofluidos. **Workshop de Engenharia mecânica e de Produção**, nov. 2021.

CARMO, E. C. W. DO. **Técnicas de Machine Learning aplicadas ao monitoramento de partículas em caldeira de recuperação Kraft** . Projeto de Final de Curso—Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2017.

CARTWRIGHT, H. M. Machine Learning in Chemistry. **Royal Society of Chemistry**, 2020.

CAVALCANTE, Z. V.; SILVA, M. L. S. **A importância da Revolução Industrial no mundo da Tecnologia**. Matingá: [s.n.].

ÇENGEL, Y. A.; GHAJAR, A. J. **Transferência de calor e massa: uma abordagem prática**. 4. ed. [s.l.] AMGH, 2012.

DAOUTIDIS, P. et al. Machine learning in process systems engineering: Challenges and opportunities. **Computers & Chemical Engineering**, v. 181, p. 108523, fev. 2024.

DEMSAR J et al. Orange: Data Mining Toolbox in Python. **Journal of Machine Learning Research**, 2013.

ESCOVEDO, T.; KOSHIYAMA, A. **Introdução a Data Science — Algoritmos de Machine Learning e métodos de análise**.

FARIA, A. M. **UTILIZAÇÃO DE TROCADORES DE CALOR DE PLACAS EM ESPIRAL EM PROCESSOS DA INDÚSTRIA SIDERÚRGICA**. Guaratinguetá: [s.n.].

GAO, H. Machine learning in reaction engineering. Em: **Artificial Intelligence in Manufacturing**. [s.l.] Elsevier, 2024. p. 139–166.

GOMES, T. **Ferramentas Open Source de Data Mining**. Tese (Doutorado) - Curso de Informática e Sistemas—Coimbra: Instituto Politécnico de Coimbra, 2014.

GUILHON, A. et al. Jornada Python: uma jornada imersiva na aplicabilidade de uma das mais poderosas linguagens de programação do mundo. **Brasport**, 2022.

HAN, S. **Estudo das Aplicações de Data Science e Machine Learning na Engenharia Química**. São Carlos: [s.n.].

HODSON, J. **How to Make Your Company Machine Learning Ready**. Disponível em: <<https://hbr.org/2016/11/how-to-make-your-company-machine-learning-ready>>. Acesso em: 13 mar. 2024.

IBM. What is random forest? **IBM Knowledge base**, 2020.

INCASE. Indústria Mecânica de Equipamentos. 2022.

INCROPERA, F.; DEWITT, D. P. **Fundamentos de Transferência de Calor e de Massa**. 7. ed. Rio de Janeiro : LTC, 2014.

ISHIDA, D. M. T. **Perspectivas do aprendizado de máquina no ensino da Engenharia Química**. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) —São Carlos: Universidade Federal de São Carlos , 2021.

JAMES, G. et al. **An Introduction to Statistical Learning**. New York, NY: Springer New York, 2013. v. 103

JOVIC, A.; BRKIĆ, K.; BOGUNOVIĆ, N. An overview of free software tools for general data mining. 2015.

KAKAÇ, S.; LIU, H.; PRAMUANJAROENKIJ, A. **Heat Exchangers: Selection, Rating, and Thermal Design**. [s.l.] Boca Raton: Crc Press, 2012a.

KAKAÇ, S.; LIU, H.; PRAMUANJAROENKIJ, A. **Heat Exchangers**. 3. ed. [s.l.] CRC Press, 2012b.

KUNELBAYEV, M.; ASSEL, A.; GULDANA, T. Using Machine Learning Algorithms to Detect Anomalies in the Solar Heating System. **International Journal of Mechanics**, v. 15, p. 270–275, 31 dez. 2021.

MARTINS, D. Um tutorial completo sobre modelagem baseada em árvores de decisão. **Vooo-Insights**, 2016.

MARTINS, E. G. M. Coeficiente de determinação. **Revista de Ciência Elementar**, v. 6, n. 1, 30 mar. 2018.

MATOS, G. R. DE. **Machine learning aplicado à gestão de activos físicos industriais**. Dissertação de mestrado—Lisboa: Instituto Superior de Engenharia de Lisboa, 6 maio 2021.

NAIR, A. et al. Machine Learning for Prediction of Heat Pipe Effectiveness. **Energies**, v. 15, n. 9, p. 3276, 29 abr. 2022.

PĂTRĂȘCIOIU, C.; MARINOIU, C. **The applications of the non-linear equations systems algorithms for the heat transfer processes**. mar. 2010.

RASCHKA, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. 13 nov. 2018.

RODRÍGUEZ, M. B. R. et al. Diseño de intercambiadores de calor de tubo y coraza mediante el método de eficiencia – número de unidades de transferencia y optimización de los costos usando matlab. **Revista Centro Azúcar**, 2013.

SACOMANO, J. B. et al. **Industria 4.0**. [s.l.] Editora Bluncher, 2018.

SACRAMENTO, D. Árvore de decisão: entenda este algoritmo de machine learning. **TeraBlog**, 2021.

SCHWEIDTMANN, A. M. et al. Machine Learning in Chemical Engineering: A Perspective. **Chemie Ingenieur Technik**, v. 93, n. 12, p. 2029–2039, 22 dez. 2021.

SILVA, G. D. **Machine learning no controle preditivo de processos**. Maceió: Universidade Federal de Alagoas, 2020.

SINGH, A. **Cross-Validation Techniques**.

TECQUIPMENT. Bancada de Trocadores de calor: Guia do Usuário. **NovaDidacta**, [s.d.].

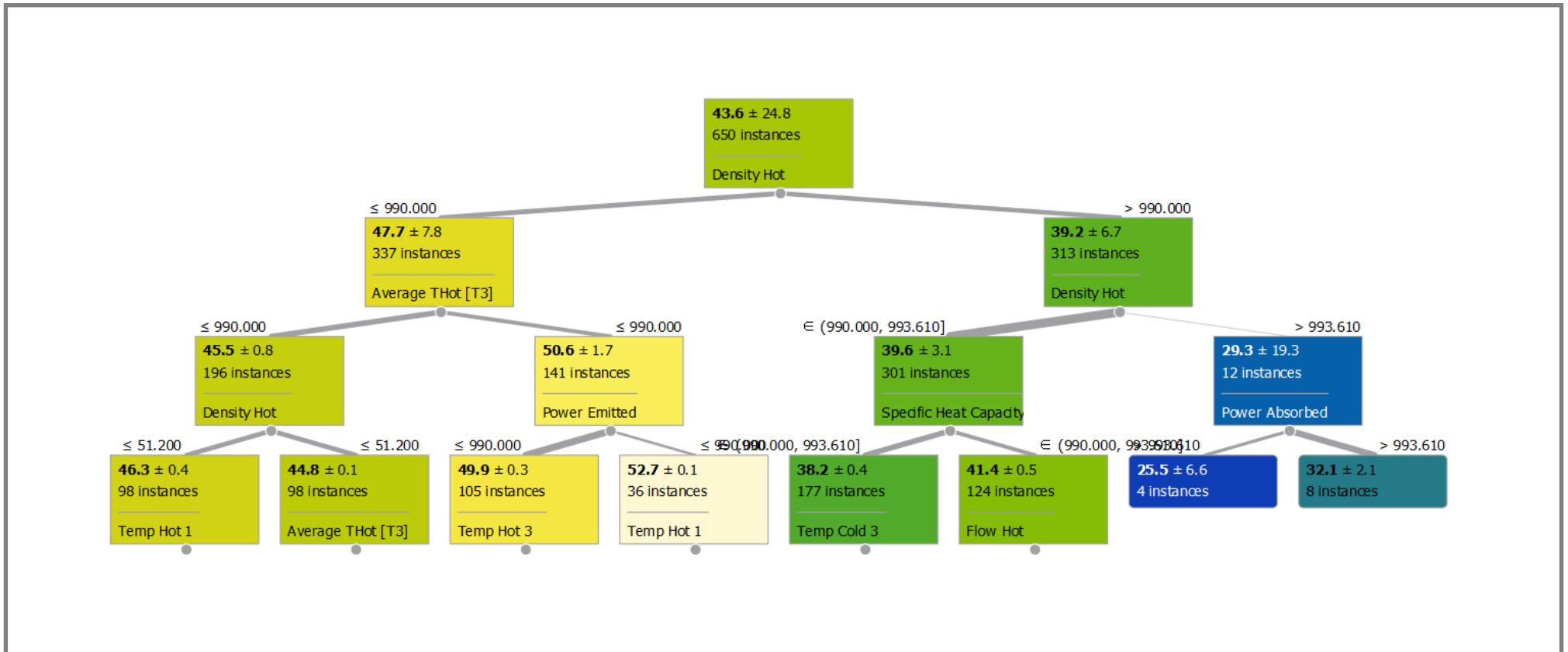
WANG, Y.-Q. et al. Machine learning framework for intelligent aeration control in wastewater treatment plants: Automatic feature engineering based on variation sliding layer. **Water Research**, v. 246, p. 120676, nov. 2023.

WANG, Z. et al. Machine learning aided multi-objective optimization and multi-criteria decision making: Framework and two applications in chemical engineering. **Computers & Chemical Engineering**, v. 165, p. 107945, set. 2022.

WESNER, J. MAE and RMSE — Which Metric is Better? **Medium**, 23 mar. 2016.

# ANEXOS

# ANEXO I



## ANEXO II

