

UNIVERSIDADE FEDERAL DO PARANÁ

GUILHERME AFONSO GREGORINI TAMANHO

COMPARAÇÃO DE MODELOS E ESTRATÉGIAS DE APRENDIZADO DE  
MÁQUINA PARA PREVISIBILIDADE DE FORMAÇÃO DO ENTUPIMENTO NO  
PROCESSO DE PRODUÇÃO DE GÁS DE SÍNTESE USANDO REFORMA A SECO  
DO BIOGÁS

PALOTINA

2024

Guilherme Afonso Gregorini Tamanho

COMPARAÇÃO DE MODELOS E ESTRATÉGIAS DE APRENDIZADO DE  
MÁQUINA PARA PREVISIBILIDADE DE FORMAÇÃO DO ENTUPIMENTO NO  
PROCESSO DE PRODUÇÃO DE GÁS DE SÍNTESE USANDO REFORMA A SECO

TCC apresentado ao curso de Engenharia de Energia, Setor Palotina da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Engenharia de Energia

Orientadora: Profa. Me. Paola Cavalheiro Ponciano

Coorientador: Prof. Dr. Mauricio Romani

PALOTINA

2024



UNIVERSIDADE FEDERAL DO PARANÁ  
DEPARTAMENTO DE ENGENHARIAS E EXATAS  
Rua Pioneiro, 2153, - - Bairro Jardim Dallas, Palotina/PR, CEP 85950-000  
Telefone: 3360-5000 - <http://www.ufpr.br/>

## ATA DE REUNIÃO

Aos cinco dias do mês de dezembro do ano de 2024, das 10h às 11h , na plataforma virtual Google Meet, link: "<https://meet.google.com/pdn-onap-uad>" compareceram para defesa pública do Trabalho de Conclusão de Curso, requisito obrigatório para a obtenção do título de Engenheiro de Energia o aluno Guilherme Afonso Gregorini Tamanho, tendo como Título do Trabalho de Conclusão de Curso "COMPARAÇÃO DE MODELOS E ESTRATÉGIAS DE APRENDIZADO DE MÁQUINA PARA PREVISIBILIDADE DE FORMAÇÃO DO ENTUPIAMENTO NO PROCESSO DE PRODUÇÃO DE GÁS DE SÍNTESE USANDO REFORMA A SECO". Constituíram a Banca Examinadora os professores: Prof. Me. Paola Cavalheiro Ponciano (Orientadora e Presidente da Banca), Prof. Dr. Fabio Alexandre Spanhol e a Dr. Lígia Gomes. A orientadora e Presidente da Banca concedeu a palavra ao discente, para exposição do seu trabalho. A seguir, foi concedida a palavra em ordem sucessiva aos membros da Banca de Exame, os quais passaram a arguir a discente. Ultimada a defesa, que se desenvolveu nos termos normativos, a Banca de Exame, em sessão secreta, passou aos trabalhos de julgamento, tendo atribuído a discente as seguintes notas: Prof. Me. Paola Cavalheiro Ponciano nota: 90 (noventa), Prof. Dr. Fabio Alexandre Spanhol, nota: 75 (setenta e cinco), e Dr. Lígia Gomes Oliveira, nota: 85(oitenta e cinco). A nota final da discente, após a média aritmética dos três membros da banca de exame, foi 83 (oitenta e três) considerando o discente APROVADO. As considerações e sugestões feitas pela Banca de Exame deverão ser atendidas pela discente sob acompanhamento de seu orientador. Nada mais havendo a tratar foi lavrada a presente ata, que, lida e aprovada, vai por todos assinada eletronicamente.



Documento assinado eletronicamente por **MAURICIO ROMANI, PROFESSOR DO MAGISTERIO SUPERIOR**, em 05/12/2024, às 11:27, conforme art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Fabio Alexandre Spanhol, Usuário Externo**, em 05/12/2024, às 11:29, conforme art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **LÍGIA GOMES OLIVEIRA, Usuário Externo**, em 05/12/2024, às 11:36, conforme art. 1º, III, "b", da Lei 11.419/2006.



A autenticidade do documento pode ser conferida [aqui](#) informando o código verificador **7338515** e o código CRC **B653398D**.

Dedico este trabalho a todas as pessoas com transtornos e déficits cognitivos que, enfrentando barreiras de acessibilidade e preconceitos, continuam lutando pelo conhecimento e pelo aprimoramento intelectual. Vocês são a prova viva de que a determinação supera limites.

## **AGRADECIMENTOS**

Primeiramente, agradeço ao corpo docente do curso de Engenharia de Energias e aos professores da Licenciatura em Computação, cujas disciplinas eletivas contribuíram para minha formação e para a elaboração deste trabalho.

Minha gratidão ao professor Dr. Helton José Alves, com quem realizei duas iniciações científicas, abrindo as portas do LABMATER e da pesquisa científica. Agradeço também à equipe do LABMATER, em especial à mestranda Bruna Machado, pela orientação nos experimentos, e ao mestrando Renan Akira, cuja contribuição foi essencial para este trabalho.

Aos colegas de curso, formados ou que seguiram outros caminhos, sou grato pela parceria em trabalhos acadêmicos e pelo apoio fora das salas de aula.

Minha profunda gratidão à minha neuropsicóloga, Mary Ellen, e à minha psiquiatra, Dra. Nina Ferreira, cujo suporte foi crucial para a superação de uma depressão profunda, impactando positivamente meus resultados, inclusive acadêmicos, e me ajudando a compreender e aceitar minhas condições.

Aos meus pais, devo tudo. Eles, que não tiveram as mesmas oportunidades de estudo, sempre fizeram de tudo para que eu as tivesse. Graças a eles, pude me dedicar aos meus interesses, sem a necessidade de um emprego regular, além de sempre me oferecerem apoio médico com amor e dedicação incondicionais.

Por fim, agradeço especialmente à minha namorada, mestranda Luana Costa, a quem amo muito e foi meu maior suporte. Sou grato por todo o incentivo, apoio emocional nos momentos difíceis e ajuda nas tarefas diárias e pessoais. Ela é a única que conhece todas as minhas responsabilidades e me auxilia fisicamente para que eu consiga atendê-las, mesmo quando a privação de sono e a sobrecarga de atividades se tornam intensas. Além disso, agradeço aos meus quatro gatos, que têm um valor imensurável para mim. Devido às minhas dificuldades de conexão social, a relação com eles se torna essencial, trazendo conforto e uma sensação de companhia incondicional nos momentos mais desafiadores.

"O verdadeiro propósito do aprendizado não é acumular respostas, mas aprender a fazer boas perguntas." – ALBERT EINSTEIN

## RESUMO

O presente estudo investiga a aplicação de aprendizado de máquina (AM) na previsão de entupimentos em reatores de reforma a seco (RS) utilizados na produção de gás de síntese. A pesquisa foi conduzida no Laboratório de Materiais e Energias Renováveis (LABMATER) da UFPR, com o objetivo principal de comparar modelos preditivos, como Random Forest (RF), Support Vector Classifier (SVC) e K-Nearest Neighbors (K-NN). A meta foi identificar o modelo mais eficaz para prever entupimentos, buscando otimizar o controle operacional e melhorar a estabilidade e a eficiência do processo. Foram empregadas bases de dados reais, abrangendo medições realizadas em dez reações do reator piloto. Duas estratégias de particionamento foram adotadas: a primeira organizou todas as reações em uma base unificada, dividida em conjuntos de treino e teste utilizando separação estratificada; a segunda avaliou reações individuais como bases de teste separadas, representando cenários inéditos para os modelos. A normalização linear dos dados foi realizada, e para a validação dos modelos utilizou-se validação cruzada com o otimizador de hiperparâmetros. Os modelos foram avaliados com base em métricas como revocação, precisão e acurácia ponderada, com especial atenção a revocação para o rótulo 1, que identifica a iminência de entupimentos. A priorização do recall visou reduzir a ocorrência de falsos negativos, garantindo maior confiabilidade na aplicação prática. Ferramentas interpretáveis, como Shapley Additive Explanations (SHAP) e Partial Dependence Plot (PDP), foram empregadas para avaliar o impacto das variáveis operacionais sobre as previsões. A análise indicou que variáveis como pressão, umidade e tempo de reação têm forte influência nos resultados, sendo essenciais para a precisão dos modelos. Os resultados mostraram que o uso de ferramentas interpretáveis não apenas aprimora a compreensão do comportamento dos modelos, mas também auxilia na identificação de fatores críticos para o controle do processo. Essas informações permitem ajustes operacionais proativos, reduzindo o risco de interrupções e aumentando a eficiência na produção de gás de síntese. Este estudo fornece uma base robusta para futuras pesquisas, consolidando a relevância do aprendizado de máquina na otimização de processos industriais. Além disso, apresenta estratégias e abordagens para o uso de AM no processo de RS, destacando melhorias realizadas em uma base de dados utilizada para o treinamento de modelos.

**Palavras-chave:** Aprendizado de máquina; Reforma a seco; Gás de síntese; Biogás; Entupimento.

## ABSTRACT

This study investigates the application of machine learning (ML) in predicting fouling in dry reforming (DR) reactors used to produce synthesis gas. The research was conducted at UFPR's Materials and Renewable Energy Laboratory (LABMATER), with the main objective of comparing predictive models such as Random Forest (RF), Support Vector Classifier (SVC) and KNeighbors Classifier (K-NN). The aim was to identify the most effective model for predicting blockages, to optimize operational control and improve the stability and efficiency of the process. Real databases were used, covering measurements taken in ten pilot reactor reactions. Two partitioning strategies were adopted: the first organized all the reactions into a unified base, divided into training and test sets using stratified separation; the second evaluated individual reactions as separate test bases, representing unpublished scenarios for the models. Data normalization was carried out using the MinMaxScaler technique, and model validation included the use of the StratifiedKFold function with cross-validation and hyperparameter optimization by GridSearchCV. The models were evaluated based on metrics such as recall, precision and weighted accuracy, with special attention to the recall of label 1, which identifies imminent blockages. The recall prioritization aimed to reduce the occurrence of false negatives, ensuring greater reliability in practical application. Interpretable tools such as Shapley Additive Explanations (SHAP) and Partial Dependence Plot (PDP) were used to assess the impact of operational variables on predictions. The analysis indicated that variables such as pressure, humidity and reaction time have a strong influence on the results and are essential for the accuracy of the models. The results showed that the use of interpretable tools not only improves the understanding of model behavior, but also helps to identify critical factors for process control. This information allows for proactive operational adjustments, reducing the risk of outages and increasing efficiency in syngas production. This study provides a robust basis for future research, consolidating the relevance of machine learning in the optimization of industrial processes and proposing significant advances for sustainability in the production of renewable hydrogen.

**Keywords:** Machine learning; Dry reforming; Synthesis gas; Biogas; Fouling.

## LISTA DE FIGURAS

FIGURA 1 – DIAGRAMA DE EULER DE TÉCNICAS SELECIONADAS DE CIÊNCIA DE DADOS .....	20
FIGURA 2 – DIAGRAMA DE EULER DE TÉCNICAS SELECIONADAS DE CIÊNCIA DE DADOS .....	25
FIGURA 4 – REPRESENTAÇÃO DO PERCURSO METODOLÓGICO ADOTADO .	30
FIGURA 3 – UNIDADE PILOTO DE REFORMA A SECO .....	32
FIGURA 5 – MATRIZ DE CORRELAÇÃO GERAL .....	37
FIGURA 6 – MATRIZ DE CORRELAÇÃO PARA O RÓTULO ALVO.....	38
FIGURA 7 –GUIA DE ESCOLHA DE ESTIMADORES PARA ABORDAGEM DE PROBLEMAS.....	41
FIGURA 8 – MATRIZES DE CONFUSÃO DO MODELO SVC COM DIVISÃO ESTRATIFICADA.....	45
FIGURA 9 – MATRIZES DE CONFUSÃO DO MODELO RF COM DIVISÃO ESTRATIFICADA.....	46
FIGURA 10 – MATRIZES DE CONFUSÃO DO MODELO K-NN COM DIVISÃO ESTRATIFICADA.....	46
FIGURA 11 – VALORES SHAP DO RANDOM FOREST UTILIZANDO BASE ÚNICA .....	48
FIGURA 12 – DEPENDÊNCIA PARCIAL (PDP) E EXPECTATIVA CONDICIONAL INDIVIDUAL (ICE) DO RANDOM FOREST .....	50
FIGURA 13 – DEPENDÊNCIA PARCIAL BIDIRECIONAL DE SDP002 CORRELACIONADO COM OS ATRIBUTOS SDU002_1 E SDU002_2 .....	52
FIGURA 14 – DEPENDÊNCIA PARCIAL BIDIMENSIONAL DA PRESSÃO CORRELACIONADA COM A UMIDADE E A TEMPERATURA DO GÁS DE SAÍDA.....	53
FIGURA 15 – VALORES SHAP DO RANDOM FOREST, CONJUNTO TESTE BASE 24 .....	55

## LISTA DE TABELAS

TABELA 1 – PRINCIPAIS REAÇÕES ENVOLVENDO A REFORMA A SECO DO BIOGÁS .....	24
TABELA 2 – PARÂMETROS OPERACIONAIS EXTRAÍDOS DA BASE DE DADOS E SEUS VALORES MÁXIMOS E MÍNIMOS .....	35
TABELA 3 – MÉTRICAS PARA OS DADOS DE TESTE GERAIS E DO ROTULO ALVO PARA QUANDO UTILIZADA UMA BASE UNIFICADA .....	47
TABELA 4 – MÉTRICAS PARA OS DADOS DE TESTE GERAIS E DO ROTULO ALVO PARA QUANDO UTILISADA BASES SEPARADAS .....	54

## LISTA DE ABREVIATURAS OU SIGLAS

AM	- Aprendizado de Máquina
IML	- <i>Interpretable Machine Learning</i>
K-NN	- <i>K-Nearest Neighbors</i>
LABMATER	- Laboratório de Materiais e Energias Renováveis
PDP	- <i>Partial Dependence Plot</i>
PDV	- <i>Partial Dependence Value</i>
RF	- <i>Random Forest</i>
RS	- Reforma a Seco
SHAP	- <i>Shapley Additive Explanations</i>
SVC	- <i>Support Vector Classifier</i>
SVM	- <i>Support Vector Machine</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>16</b>
1.1 PROBLEMA .....	17
1.2 OBJETIVOS .....	18
1.2.1 Objetivo geral .....	18
1.2.2 Objetivos específicos .....	18
1.3 JUSTIFICATIVA .....	19
<b>2 REVISÃO DE LITERATURA</b> .....	<b>21</b>
2.1 ENERGIAS RENOVÁVEIS.....	21
2.2 APRENDIZADO DE MÁQUINA.....	24
2.3 TRABALHOS RELACIONADOS .....	27
<b>3 MATERIAL E MÉTODOS</b> .....	<b>30</b>
3.1 CARACTERIZAÇÃO DA PESQUISA .....	30
3.2 PERCURSO METODOLÓGICO .....	30
3.3 REATOR DE REFORMA A SECO .....	32
3.4 BIBLIOTECAS UTILIZADAS .....	33
3.5 CONJUNTO DE DADOS.....	34
3.6 ANÁLISE E TRATAMENTO DOS DADOS.....	34
3.7 PREPARAÇÃO DO CONJUNTO DE DADOS.....	38
3.8 SEPARAÇÃO DOS CONJUNTOS TESTE E TREINO.....	39
3.9 SELEÇÃO E AVALIAÇÃO DE DESEMPENHO DOS MODELOS.....	40
3.10 FERRAMENTAS DE APRENDIZADO DE MÁQUINA INTERPRETÁVEIS .....	43
<b>4 APRESENTAÇÃO DOS RESULTADOS</b> .....	<b>45</b>
4.1 BASE ÚNICA DE DADOS .....	45
4.2 BASES INDIVIDUALIZADAS POR REAÇÕES .....	53
<b>5 CONSIDERAÇÕES FINAIS</b> .....	<b>56</b>
<b>REFERÊNCIAS</b> .....	<b>58</b>

## 1 INTRODUÇÃO

A demanda energética é uma pauta importante no contexto do desenvolvimento econômico e social. Com o aumento da população mundial, o avanço tecnológico e a urbanização, medidas globais vêm sendo tomadas visando a produção de padrões de consumo sustentáveis, a exemplo dos Objetivos para o Desenvolvimento Sustentável - ODS, propostos pela Organização das Nações Unidas - ONU<sup>1</sup> em 2016. O sétimo objetivo descrito no documento prevê que até o ano de 2030 tenha um aumento da participação de energias renováveis na matriz energética global e, conforme previsto no nono objetivo, a industrialização inclusiva e sustentável.

A produção de hidrogênio a partir do biogás é uma solução sustentável que aproveita resíduos orgânicos, como lixo e resíduos agrícolas, para gerar energia limpa. O biogás, composto principalmente de metano ( $\text{CH}_4$ ) e dióxido de carbono ( $\text{CO}_2$ ), é tratado e submetido a processos que resultam em hidrogênio ( $\text{H}_2$ ) e dióxido de carbono. Esse hidrogênio pode ser purificado e utilizado em aplicações industriais, transporte e geração de energia, além de reduzir emissões de gases de efeito estufa, também promove a economia circular. Com o avanço de tecnologias e maior demanda por soluções de baixo carbono, o hidrogênio a partir de biogás tem potencial para ser uma peça central na transição energética global (PUZICKSKI, *et al.*, 2015; SCHAFFNER, 2018).

No método de produção de  $\text{H}_2$  a partir da reforma a seco (RS) do biogás, é aplicada uma técnica para produzir gás de síntese, onde o  $\text{CH}_4$  reage com o  $\text{CO}_2$  em altas temperaturas na presença de catalisadores, formando monóxido de carbono (CO) e  $\text{H}_2$ . Essa abordagem é atrativa por utilizar dois gases de efeito estufa, contribuindo para a redução de emissões, se mostrando promissora em aplicações onde a sustentabilidade e reutilização de  $\text{CO}_2$  são prioridades (SCHAFFNER, 2018; OLIVEIRA, 2021).

Um dos desafios da utilização da reforma a seco é a desativação dos catalisadores e os altos custos energéticos. Durante o processo pode ocorrer entupimento do reator, que se dá devido à deposição de carbono na superfície do

---

<sup>1</sup>- Organização das Nações Unidas – Objetivos do Desenvolvimento Sustentável - <https://brasil.un.org/pt-br/sdgs>

catalisador e excesso de água na reação, reduzindo a eficiência e bloqueando o fluxo de gases.

O presente trabalho apresenta os resultados de uma pesquisa exploratória focada na comparação e teste de modelos de AM para prever o entupimento do leito durante a RS do biogás. Além disso, analisa como diferentes formas de separação e registro dos dados coletados afetam as previsões. A pesquisa foi realizada no Laboratório de Materiais e Energias Renováveis (LABMATER) da Universidade Federal do Paraná, Setor Palotina.

Ainda que o trabalho tenha um caráter de estudos iniciais no contexto do laboratório, observa-se que os resultados obtidos podem servir como referência para pesquisas futuras envolvendo a aplicação de técnicas de AM na otimização dos processos de produção de Hidrogênio Renovável. A interpretação dos modelos de aprendizagem de máquina possibilitou conhecer e compreender os principais fatores que causam o entupimento. Com esses dados, é possível ajustar os parâmetros operacionais, melhorando a estabilidade do processo e evitando interrupções na reação, otimizando o controle e a eficiência na produção de gás de síntese.

Para apresentação dos resultados o trabalho foi estruturado em 4 seções onde inicialmente elenca-se a apresentação do estudo, seguidas da apresentação do problema e dos objetivos, na sequência apresenta-se a revisão de literatura que envolve os conceitos referentes a produção de gás de síntese e a caracterização de como se dá o entupimento de linha, bem como os conceitos sobre aprendizagem de máquina e as possibilidades da sua aplicação na otimização dos processos de produção. Por sua vez, na seção de materiais e métodos é realizada a descrição dos passos adotados tanto do ponto de vista da metodologia científica quanto da perspectiva dos modelos de aprendizagem de máquina comparados no estudo.

## 1.1 PROBLEMA

O entupimento do leito é um dos desafios comuns na RS do biogás, nesse sentido, elencou-se a seguinte questão problema: Quais estratégias e metodologias podem ser adotadas na aplicação de AM, para a otimização de processos de produção de H<sub>2</sub> através da RS?

## 1.2 OBJETIVOS

O presente estudo buscou explorar a aplicação de aprendizado de máquina no contexto do processo de produção de gás de síntese através da reforma a seco. Dada a relevância do controle operacional para evitar o entupimento do reator, esta pesquisa tem como foco principal a comparação de modelos e estratégias, realizando uma análise exploratória para identificar abordagens mais promissoras, utilizando as bases de dados registradas de reações realizadas no último ano no LABMATER da UFPR Setor Palotina.

### 1.2.1 Objetivo geral

Investigar e comparar modelos e estratégias de aprendizado de máquina para avaliar seu desempenho na previsão do entupimento do reator no processo de produção de gás de síntese através da RS, com o intuito de realizar uma análise exploratória utilizando base dados de reações realizadas no LABMATER.

### 1.2.2 Objetivos específicos

- a. Identificar os modelos e parâmetros operacionais utilizados na reforma a seco do processo de geração de gás de síntese.
- b. Analisar o desempenho dos diferentes modelos de aprendizado de máquina: Random Forest, Support Vector Classifier (SVC) e K-Nearest Neighbors, com base em dados produzidos no Labmater.
- c. Explorar e avaliar estratégias distintas de separação de dados, como a separação tradicional em treino e teste e a exclusão de reações específicas para teste.
- d. Identificar o modelo e a estratégia mais adequados, considerando métricas de avaliação e a capacidade de generalização em cenários distintos.
- e. Interpretar os modelos selecionados e comparar os resultados utilizando ferramentas de interpretabilidade, como SHAP (*Shapley Additive Explanations*) e PDP (*Partial Dependence Plot*), para identificar os fatores operacionais mais relevantes para a formação do entupimento.

### 1.3 JUSTIFICATIVA

O processo de produção de gás de síntese desempenha um papel essencial na utilização do biogás como fonte de hidrogênio renovável de baixo carbono. Este processo é realizado a partir da reforma a seco. A RS é uma reação endotérmica, operando em temperaturas entre 600°C e 900°C sob pressão atmosférica, com o auxílio de catalisadores. Durante o processo pode acontecer a formação de coque, um fenômeno que ocorre devido ao surgimento de reações paralelas ou envenenamento do catalisador, resultando no depósito de carbono em sua superfície (ZHANG *et al.*, 2015; OLIVEIRA, 2021). Esse acúmulo de carbono pode levar ao entupimento do reator, exigindo a interrupção do processo reacional, o que compromete sua eficiência e estabilidade.

No contexto do LABMATER, utiliza-se um reator de reforma a seco em escala piloto, representado na (FIGURA 1). Este e os demais módulos desenvolvidos pela equipe de pesquisadores servem de bancada para testar metodologias e ferramentas que possam otimizar e escalar a produção de materiais e energias renováveis. Neste cenário, a comparação de modelos e estratégias de aprendizado de máquina não permite apenas a seleção de um modelo interpretável para prever a formação do entupimento, mas também a avaliação de diferentes estratégias de registro e separação de dados que impactam a eficácia dessas previsões, contribuindo para o trabalho que já vem sendo realizado de aumentar o grau de maturidade tecnológica no laboratório.

FIGURA 1 – DIAGRAMA DE EULER DE TÉCNICAS SELECIONADAS DE CIÊNCIA DE DADOS



FONTE: Adaptado de OLIVEIRA (2021).

Com a previsão precisa do referido fenômeno, é possível ajustar os parâmetros operacionais, aumentando a estabilidade do processo, otimizando a operacionalização do reator e, em alguns casos, prevenindo a interrupção da reação. Além disso, a interpretação dos modelos selecionados contribuirá para a compreensão dos fatores mais relevantes que influenciam o entupimento, promovendo avanços significativos no controle e na eficiência do processo de produção de gás de síntese.

## 2 REVISÃO DE LITERATURA

Este capítulo apresenta uma revisão teórica e conceitual sobre temas essenciais para o entendimento do presente trabalho, incluindo o contexto das energias renováveis, o papel do aprendizado de máquina e sua aplicação em processos industriais, e uma análise de estudos correlatos que utilizam métodos de aprendizado de máquina interpretável para otimização e monitoramento em processos de reforma a seco. Esses tópicos oferecem uma base sólida para a compreensão das técnicas e abordagens aplicadas neste estudo, assim como para situar seu desenvolvimento em meio a pesquisas contemporâneas relevantes.

### 2.1 ENERGIAS RENOVÁVEIS

A iniciativa de descarbonização em todos os setores, estimula investimentos em pesquisa e desenvolvimento de soluções que reduzam emissões sem comprometer a viabilidade econômica. Fontes renováveis, que se regeneram rapidamente e são de manejo simples, surgem como alternativas viáveis para garantir segurança energética e mitigar impactos climáticos (GIELEN *et al.*, 2019).

Neste aspecto, o Brasil se destaca com 49,1% de sua matriz energética proveniente de fontes renováveis, bem acima da média global de 14,7% (EPE, 2024). Isso se deve à abundância de recursos naturais e biodiversidade, que possibilitam a exploração de biomassa, hidrelétricas e energia solar. Essa diversidade diminui a dependência de combustíveis fósseis, aumentando a autonomia energética e a estabilidade econômica (PHILIPPI JR, 2016). Neste cenário, o biogás vem sendo apresentado como uma fonte energética que contribui para o aproveitamento de resíduos e reforça a diversidade da matriz energética brasileira.

O biogás é um biocombustível gasoso, obtido a partir da decomposição de matéria orgânica, como resíduos agropecuários e urbanos, por meio da digestão anaeróbica (ZHANG, Q. *et al.*, 2016). Tem composição variável, principalmente metano e dióxido de carbono, com características semelhantes ao gás natural. Sua capacidade calorífica varia entre 15 e 30 MJ/Nm<sup>3</sup>, dependendo da matéria-prima, e pode conter pequenas quantidades de gases como sulfeto de hidrogênio (H<sub>2</sub>S) e amônia (NH<sub>3</sub>), que afetam a eficiência energética (ROHSTOFFE, FACHAGENTUR NACHWACHSENDE, 2010; ZANETTE, 2009; DO AMARAL *et al.*, 2019).

Apesar do potencial do biogás como fonte de energia, seu uso enfrenta desafios, como a remoção de impurezas e a otimização de processos de conversão para melhorar seu poder calorífico e proteger equipamentos. Nesse contexto, métodos de purificação, como o processo de absorção, destacam-se por sua viabilidade econômica e alta eficiência, sendo capazes, em alguns casos, de promover a remoção completa de compostos como o  $H_2S$  (KWAŚNY *et al.*, 2016; SZARBLEWSKI *et al.*, 2012). Frare (2008) reforça que, mesmo considerando os custos associados à purificação por absorção, o biogás mantém-se competitivo em relação a outros combustíveis, consolidando-se como uma opção sustentável para a matriz energética.

Ademais, além dos avanços na purificação, tecnologias como a reforma a seco ampliam o aproveitamento do biogás, convertendo metano e dióxido de carbono em gás de síntese. Esse processo não apenas aumenta a eficiência do uso do biogás, mas também desempenha um papel essencial na produção de hidrogênio renovável de baixo carbono, sendo assim, uma alternativa vantajosa (SCHAFFNER, 2018).

O gás de síntese é uma mistura gasosa composta principalmente por  $H_2$  e  $CO$ , produzido a partir de matérias-primas ricas em hidrocarbonetos, como carvão, gás natural, biomassa e biogás. Existem dois métodos principais utilizados industrialmente para a obtenção do gás de síntese: a gaseificação e a reforma do metano. A reforma do metano, é classificada em diferentes processos, entre os quais os mais comuns são a reforma a vapor, a oxidação parcial, a reforma autotérmica e a reforma a seco (RAMOS *et al.*, 2011; LORA *et al.*, 2012; ALVES *et al.*, 2013; ZHANG *et al.*, 2016; ABDIN *et al.*, 2020).

No contexto da obtenção de gás de síntese a partir do biogás produzido com resíduos sólidos, pesquisas estão sendo conduzidas especialmente no uso de catalisadores e equipamentos apropriados, devido à alta estabilidade das moléculas (ALVES *et al.*, 2013; AGUIAR, 2017). Seu uso industrial está em expansão, especialmente na geração de energia e na produção de combustíveis sintéticos, como hidrogênio puro e gás natural sintético, tornando-se uma opção promissora para soluções energéticas sustentáveis (ROSSETTI *et al.*, 2016; AL-RAHBI *et al.*, 2017).

No que se refere a produção de combustíveis sintéticos, o hidrogênio (H) é o elemento químico mais simples, leve e altamente inflamável, geralmente encontrado como gás diatômico (GARCIA, 2024). Sua combustão gera apenas água, sem emissões de gases de efeito estufa, o que ajuda a reduzir a poluição. Tem aplicações em transporte e geração de eletricidade, essencial para uma economia de baixo

carbono. O monóxido de carbono resultante também pode ser usado em processos industriais, como a carbonização. (CHEN *et al.*, 2015).

Com a maior densidade energética por unidade de massa entre os combustíveis, atingindo  $141,86 \text{ kJ g}^{-1}$  (FUKUROZAKI, 2011; SANTOS *et al.*, 2016; VARGAS *et al.*, 2006). Sua versatilidade e perfil sustentável o tornam essencial na transição para um sistema energético mais resiliente (SANTOS *et al.*, 2016; ZHANG *et al.*, 2016; STAFFELL *et al.*, 2019).

Entre os métodos de produção de hidrogênio e gás de síntese, a RS tem ganhado destaque ambiental e comercial. Este processo consome  $\text{CO}_2$  e  $\text{CH}_4$ . Contudo, é endotérmico, exigindo altas temperaturas e baixas pressões. A razão  $\text{CO}_2/\text{CH}_4$  varia de 1:1 a 1,5:1, resultando em uma relação estequiométrica de  $\text{H}_2/\text{CO}$  de 1:1 (OYAMA *et al.*, 2012; LUISETTO *et al.*, 2015). Observa-se também, conforme Nakajima *et al.* (2023), que temperaturas mais altas (até 973 K) favorecem a conversão de metano, otimizando a produção de gás de síntese próximo a 1:1 ( $\text{H}_2/\text{CO}$ ).

A RS não envolve a reação de carbono com vapor d'água, comum na reforma a vapor, o que aumenta a tendência à formação de coque, podendo haver o entupimento (SCHAFFNER, 2018).

Um catalisador eficiente para processos industriais deve apresentar alta seletividade, atividade, estabilidade à deposição de carbono e sinterização das partículas metálicas, além de boa capacidade de reciclagem. Para isso, a seleção e obtenção de catalisadores exigem grandes investimentos. As etapas de síntese, caracterização e teste em escala de bancada representam um desafio para os pesquisadores, que buscam materiais de baixo custo, sustentáveis e com boa atividade catalítica (MONTENEGRO, 2006). Além disso, o suporte utilizado na formação do catalisador é crucial para o sucesso nos processos catalíticos.

Para minimizar a formação de coque, e conseqüentemente o entupimento, é crucial ajustar condições operacionais, como otimizar a temperatura para favorecer as reações desejadas, controlar a vazão para impactar a concentração de reagentes e escolher catalisadores resistentes à deposição de coque (SCHAFFNER, 2018). Dentre as principais reações paralelas, que podem comprometer a eficiência da reforma a seco do biogás, três podem causar o entupimento do reator, tais reações estão representadas na (TABELA 1).

TABELA 1 – PRINCIPAIS REAÇÕES ENVOLVENDO A REFORMA A SECO DO BIOGÁS

Reação	Reação	Nome da Reação
1	$\text{CH}_4 + \text{CO}_2 \rightleftharpoons 2\text{CO} + 2\text{H}_2$	Reforma a seco
2	$\text{CO} + \text{H}_2\text{O} \rightleftharpoons \text{CO}_2 + \text{H}_2$	Deslocamento gás-água ( <i>Water-Gas Shift</i> )
3	$\text{CH}_4 \rightleftharpoons \text{C} + 2\text{H}_2$	Decomposição do metano
4	$2\text{CO} \rightleftharpoons \text{C} + \text{CO}_2$	Boudouard
5	$\text{CO} + \text{H}_2 \rightleftharpoons \text{C} + \text{H}_2\text{O}$	Redução do CO
6	$\text{CH}_4 + 2\text{O}_2 \rightleftharpoons \text{CO}_2 + 3\text{H}_2\text{O}$	Combustão
7	$\text{CO}_2 + 4\text{H}_2 \rightleftharpoons \text{CH}_4 + 2\text{H}_2\text{O}$	Metanação

FONTE: Adaptado de ALVES *et al.*, 2013; FIDALGO *et al.*, 2013; VORONTSOV *et al.*, 2014; Oliveira, 2021.

Conforme é possível identificar na tabela, a reação de Boudouard reação 4 envolve a decomposição do CO em carbono sólido e gás carbono, resultando na formação de depósitos que podem bloquear o sistema. A decomposição do metano (REAÇÃO 3) quebra o CH<sub>4</sub> em H<sub>2</sub> e C, gerando resíduos sólidos que obstruem os equipamentos. Além disso, a redução do CO (REAÇÃO 5) também contribui para a formação de carbono sólido, intensificando o risco de entupimento (BEREKETIDOU e GOULA, 2012; LUISETTO *et al.*, 2015).

Essas reações, se não controladas, podem reduzir a produção de hidrogênio e impactar a viabilidade econômica do sistema, tornando essencial o desenvolvimento de estratégias para minimizá-las durante a RS. Tendo isto em vista, é válido estudo de modelos de aprendizado de máquina para o monitoramento operacional do processo de reforma a seco, visando a otimização nas tomadas de decisão do operador/especialista do laboratório durante o processo reacional.

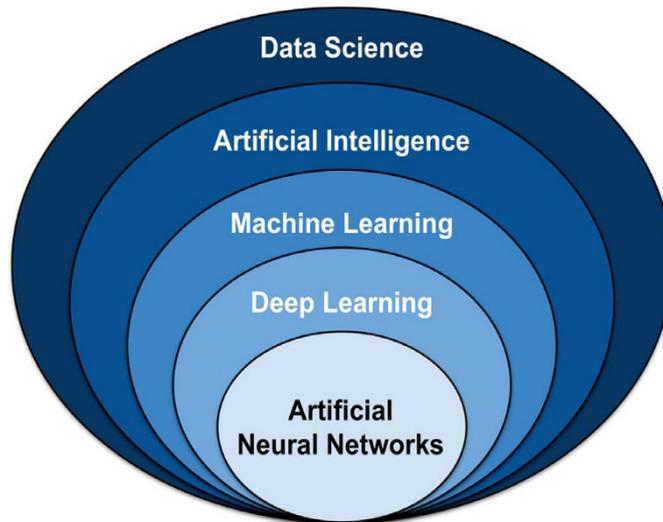
O envenenamento dos catalisadores ocorre quando certos compostos interagem com os sítios ativos do metal, alterando sua superfície e favorecendo a formação de outros compostos (ALMEIDA, 2012). O principal agente causador desse envenenamento é o enxofre, especialmente o H<sub>2</sub>S presente no biogás (OLIVEIRA, 2021; BACH, 2016; PAWAR *et al.*, 2017), o que exige o tratamento do biogás antes da reforma a seco.

## 2.2 APRENDIZADO DE MÁQUINA

O aprendizado de máquina (AM), uma subárea da inteligência artificial, como representado na (FIGURA 2), concentra-se no desenvolvimento de algoritmos e modelos capazes de aprender a partir de dados e aprimorar seu desempenho ao longo

do tempo, sem a necessidade de uma programação explícita para cada tarefa (Hastie *et al.*, 2009). Essa abordagem se diferencia da programação tradicional ao permitir que os sistemas reconheçam padrões em grandes volumes de dados e, a partir disso, façam previsões e decisões de forma autônoma (JAMES *et al.*, 2013).

FIGURA 2 – DIAGRAMA DE EULER DE TÉCNICAS SELECIONADAS DE CIÊNCIA DE DADOS



FONTE: CHOI *et al.* (2020).

Russell e Norvig (2022) ressaltam que, assim como os seres humanos aperfeiçoam suas habilidades por meio de experiências e observações, agentes computacionais utilizam dados para construir modelos que fundamentam a resolução de problemas em diversos contextos. A capacidade de adaptação dos sistemas de aprendizado de máquina é essencial para lidar com a complexidade e a dinamicidade dos dados presentes em áreas como medicina, finanças e engenharia, onde a demanda por soluções inovadoras e eficientes é crescente.

No campo do AM, quatro métodos principais são amplamente empregados para lidar com diferentes tipos de problemas: o aprendizado supervisionado, não supervisionado, semi-supervisionado e o aprendizado por reforço. Cada uma dessas abordagens tem sua utilidade em cenários específicos, dependendo da natureza dos dados e dos objetivos do processo de análise (JAMES *et al.*, 2013; SALIAN, 2018).

A abordagem supervisionada é amplamente utilizada para que o modelo aprenda padrões a partir de exemplos rotulados, onde as respostas corretas já são conhecidas. O modelo infere um algoritmo a partir dos pares de característica-alvo e

é informado se suas previsões estão corretas ao comparar com as saídas rotuladas (SALIAN, 2018).

Diferente do aprendizado supervisionado, o aprendizado não supervisionado identifica padrões em um conjunto de dados sem o uso de um alvo específico, categorizando instâncias de forma autônoma (JAMES *et al.*, 2013; SALIAN, 2018). Nesse caso, os padrões, quando existentes, são descobertos pelo próprio algoritmo. O aprendizado semi-supervisionado, por sua vez, combina elementos dos dois métodos, sendo útil em cenários onde há dados rotulados e não rotulados, permitindo que o algoritmo aprenda a partir de uma base parcial de informações (SALIAN, 2018). Por sua vez, o aprendizado por reforço, que se aproxima da maneira como os humanos aprendem por tentativa e erro, foca em treinar algoritmos para alcançar um resultado desejado sem respostas predefinidas, baseando-se na experiência para aperfeiçoar suas ações (HASTIE *et al.*, 2009; SALIAN, 2018).

Para alcançar uma predição precisa, é fundamental selecionar modelos de AM adequados e combiná-los com um conjunto de dados de treinamento de qualidade. Sawant *et al.* (2023), afirma que na área de ciência dos materiais, a preferência por técnicas de aprendizado supervisionado se dá devido à sua capacidade de aprender padrões diretamente dos dados rotulados, o que melhora significativamente a confiabilidade das previsões.

Com relação ao aprendizado supervisionado, Choi *et al.* (2020) apresentam os passos fundamentais em que o processo inicia com a obtenção de um conjunto de dados, que é então dividido em três partes: treinamento, validação e teste. O conjunto de treinamento é utilizado para ensinar o modelo, enquanto o conjunto de validação ajusta os parâmetros do algoritmo com base no desempenho. Já o conjunto de teste, é empregado para avaliar quão bem o modelo consegue prever resultados com dados não vistos anteriormente. A cada iteração, o desempenho do modelo nos dados de treinamento é comparado com o de validação, o que permite ajustes contínuos. No entanto, como o conjunto de validação pode diferir do conjunto de teste, a capacidade de generalização do modelo pode variar.

Quando aplicado à classificação de documentos, por exemplo, o aprendizado supervisionado tem como objetivo mapear uma entrada a uma anotação de saída, como a identificação de um sentimento positivo ou negativo em um texto. Este método também se estende a tarefas de regressão, nas quais se busca prever um valor contínuo em vez de uma classe, como a probabilidade de um indivíduo gostar de um

livro. A qualidade do aprendizado está intrinsecamente ligada à qualidade dos exemplos de treinamento. Conjuntos de dados que contêm exemplos ambíguos, com características similares, mas pertencentes a classes distintas, podem prejudicar a precisão do sistema (ABDO, 2021).

Sawant *et al.* (2023), enfatiza que entre os modelos mais utilizados na área de ciências dos materiais são Máquinas de Vetor de Suporte (*Support Vector Machine* - SVM), Florestas Aleatórias (*Random Forest* - RF). Estes modelos se destacam pela capacidade de resolver problemas complexos de classificação e regressão.

## 2.3 TRABALHOS RELACIONADOS

Os estudos relacionados, citados abaixo, ilustram o avanço na produção de catalisadores e oferecem contribuições valiosas para o presente trabalho. Já a aplicação de aprendizado de máquina interpretável (*Interpretable Machine Learning* - IML) tem se mostrado uma abordagem promissora no estudo e otimização da RS, especialmente para prever o desempenho de catalisadores e entender as interações complexas entre os parâmetros de processo.

Considerando o exposto, para implementação do estudo teórico, foi realizada a busca por trabalhos relatando experiências de pesquisa semelhantes, no contexto da aplicação de AM em processos reacionais utilizando RS. Foram utilizados os termos: Aprendizado de Máquina, Reforma a seco e Inteligência Artificial na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), que não obteve nenhum retorno.

Sendo assim, realizou-se a busca por trabalhos relacionados utilizando como critério de seleção aqueles que envolviam o uso de aprendizado de máquina para estudos relacionados ao entupimento de leito ou coque na reforma seco do biogás. Para isso foram inseridas as seguintes palavras-chave: "machine learning" AND "dry reforming" AND "clogging". Os 59 resultados não cumpriam com os critérios de busca.

Por fim foram realizadas novas buscas, desta vez com os termos "machine learning" AND "dry reforming", filtrando resultados dos últimos cinco anos, recorte temporal entendido como suficiente pelos pesquisadores já que se trata de um campo de estudo recente. Dos 1.580 retornos, que também não atendiam os critérios, selecionou-se três para mostrar diferentes abordagens de AM no processo de RS.

O artigo "Interpretable Machine Learning Framework for Catalyst Performance Prediction and Validation with Dry Reforming of Methane", apresentado por ROH *et al.*

em 2024, apresenta uma estrutura de aprendizado de IML para prever e validar o desempenho de catalisadores no processo de RS. A pesquisa aborda desafios como a formação de coque e a sinterização dos catalisadores, além dos altos custos e da complexidade do processo. Os autores utilizaram ferramentas interpretáveis, como Explicações Aditivas de Shapley (*Shapley Additive Explanations - SHAP*) e *Partial Dependence Value (PDV)*, aplicadas a modelos como CatBoost, Árvores de Decisão e RF, com o CatBoost alcançando o melhor desempenho. Esse *framework* de IML demonstrou não apenas alta precisão preditiva no desempenho dos catalisadores, mas também permitiu identificar variáveis críticas para esse processo, como a temperatura de reação e a taxa de espaço de gás horário. Isso confirma tanto a precisão das previsões quanto a relevância do modelo na análise e desenvolvimento de catalisadores eficientes (ROH *et al.*, 2024).

No trabalho de ROH *et al.* (2024), foram utilizadas quatro etapas principais do *framework*: preparação dos dados, desenvolvimento do modelo, recomendação de candidatos a catalisadores e análise de interpretabilidade do modelo. Na primeira etapa, de preparação de dados, são realizadas a coleta e o pré-processamento dos dados, incluindo a seleção de características usando o coeficiente de correlação de Pearson e o SHAP. Na etapa de desenvolvimento do modelo, o melhor algoritmo é selecionado entre diferentes alternativas, seguido de otimização de hiperparâmetros para aprimorar o desempenho. A terceira etapa, recomendação de catalisadores candidatos, utiliza SHAP e PDV para gerar um conjunto de dados que minimiza a influência de condições prévias e operacionais dos catalisadores, orientando a escolha dos melhores candidatos. Por fim, na análise de interpretabilidade, o PDV é aplicado para avaliar como as variáveis dos materiais dos catalisadores influenciam os valores de saída do modelo, oferecendo *insights* para o design de catalisadores e garantindo que o modelo respeite o conhecimento estabelecido na área.

Outro estudo importante é o artigo "*Impacts of Catalyst and Process Parameters on Ni-Catalyzed Methane Dry Reforming via Interpretable Machine Learning*", de VELLAYAPPAN *et al.* (2023), que utiliza AM para investigar o impacto de variáveis de processo e características de catalisadores de níquel (Ni) na RS. Utilizando dados de 1.638 pontos obtidos, extraídos de 221 estudos da literatura, os autores aplicaram modelos de árvore de decisão, como CatBoost e RF, para prever a taxa H<sub>2</sub>/CO em função de 11 variáveis. Para este trabalho o modelo de CatBoost também se destacou, tendo alta precisão (R<sup>2</sup> de 0,91). A análise de importância das

variáveis indicou que o tamanho das partículas de Ni e o volume específico de poros são cruciais para o controle da taxa  $H_2/CO$ . A aplicação de gráficos de dependência parcial permitiu visualizar o efeito isolado do tamanho das partículas na RS, sugerindo que modelos de aprendizado de máquina interpretáveis podem fornecer novos *insights* para o ajuste de catalisadores (VELLAYAPPAN *et al.*, 2023).

Por fim, MOSAAD AWAD *et al.*, em 2024, trazem um artigo de revisão "*Recent Developments and Current Trends on Catalytic Dry Reforming of Methane: Hydrogen Production, Thermodynamics Analysis, Techno Feasibility, and Machine Learning*", em que destaca os avanços recentes na RS, abordando a aplicação de aprendizado de máquina para acelerar a seleção e a otimização de catalisadores. A revisão enfoca o potencial do AM para reduzir custos e tempo de desenvolvimento, permitindo identificar catalisadores com alta resistência à deposição de carbono e boa estabilidade. Além disso, a análise termodinâmica sugere que o uso de AM pode aprimorar a seleção de condições reacionais e o desenvolvimento de catalisadores, promovendo a viabilidade econômica e ambiental da RS (MOSAAD AWAD *et al.*, 2024).

Esses estudos reforçam a eficácia do uso de IML para aprimorar a compreensão e a otimização de processos na RS. A utilização de modelos que se baseiam em árvores de decisão, combinados com ferramentas interpretáveis como SHAP e PDP, possibilita uma análise detalhada dos parâmetros de processo e das características dos catalisadores, auxiliando na previsão de desempenho e na escolha de condições ideais de operação.

Dessa forma, esses trabalhos relacionados contribuem para fundamentar a abordagem adotada neste estudo, que também aplica aprendizado de máquina interpretável para detecção e previsão de entupimentos no processo de reforma a seco do biogás.

### 3 MATERIAL E MÉTODOS

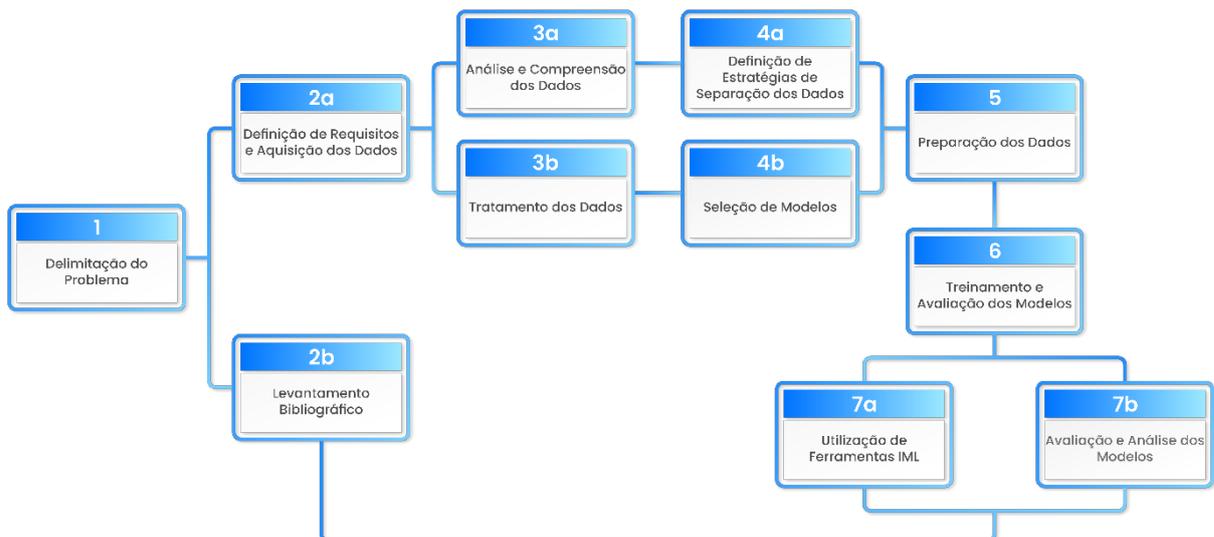
#### 3.1 CARACTERIZAÇÃO DA PESQUISA

Considerando a classificação das áreas de conhecimento pelo CNPq, este estudo trata-se de uma pesquisa do campo das engenharias. No que se refere a finalidade, trata-se de uma pesquisa aplicada, pois abrange estudos elaborados para otimizar processos no âmbito do Laboratório de Materiais e Energias Renováveis - LABMATER da UFPR Setor Palotina. Por sua vez, no que se refere aos propósitos mais gerais a pesquisa é considerada exploratória onde foi realizado o levantamento bibliográfico e análise dos modelos de AM que podem contribuir no processo identificação de entupimento na linha de reforma a seco do laboratório.

#### 3.2 PERCURSO METODOLÓGICO

O percurso metodológico deste estudo, representado na (FIGURA 4), foi estruturado de maneira sequencial, para garantir uma análise robusta e uma modelagem eficiente. A metodologia, apresentada de forma encadeada, reflete as fases essenciais para a delimitação do problema, a construção e a avaliação dos modelos, bem como a análise dos resultados.

FIGURA 3 – REPRESENTAÇÃO DO PERCURSO METODOLÓGICO ADOTADO



FONTE: O autor (2024).

Conforme se observa na (FIGURA 4), após a delimitação do problema e definição dos objetivos, a pesquisa avançou com a definição dos requisitos e da coleta e compreensão dos dados disponíveis. Nos registros consultados foram identificadas variáveis diretamente relacionadas ao comportamento do reator, como tempo de reação, pressão, umidade e temperatura do gás de saída. Paralelamente, um levantamento bibliográfico foi conduzido para embasar as decisões metodológicas, abordando estratégias de aprendizado de máquina e ferramentas de interpretabilidade, devido à complexidade do sistema e a busca de trabalhos correlatos que pudessem complementar o estudo.

Com os dados em mãos, foi analisada a estrutura do conjunto e as características das variáveis. Essa etapa incluiu a identificação de padrões, tendências e valores atípicos, além da avaliação da representatividade dos rótulos.

Na etapa seguinte, foi conduzida uma exploração detalhada do conjunto. Identificaram-se correlações, padrões, tendências e *outliers*, além de se avaliar a representatividade das classes no conjunto. Realizou-se então o tratamento dos dados, onde instâncias irrelevantes foram eliminadas e as variáveis normalizadas, garantindo escalas compatíveis para os algoritmos selecionados. Além disso, aplicou-se uma estratégia de rotulagem com desvio temporal, possibilitando prever o entupimento com antecedência. Essa etapa foi fundamental para preparar os dados para a modelagem subsequente.

Na seleção dos modelos, optou-se por algoritmos robustos para dados não paramétricos e problemas multiclasse, como Random Forest, SVC e K-NN, considerando também metodologias indicadas pela biblioteca Scikit-Learn.

Para a preparação dos dados, a separação foi realizada por meio de duas estratégias. Na primeira estratégia, considerou-se todos os dados das reações foram combinados em uma única base e separados em conjuntos de treino e teste de maneira estratificada. Para a segunda, cada reação foi tratada como uma base independente, com nove reações destinadas ao treinamento e uma utilizada exclusivamente para teste, de forma alternada, entre as que continham o rótulo alvo em suas instâncias.

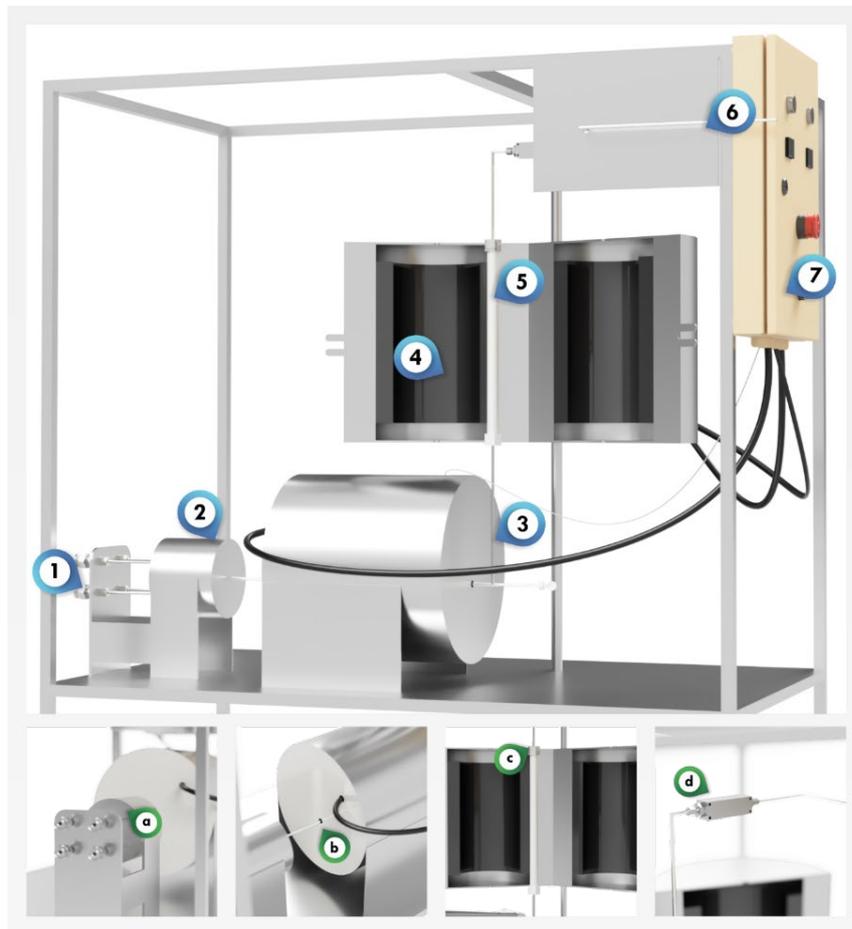
Durante o treinamento, utilizou-se um otimizador de hiperparâmetros aplicado exclusivamente aos dados de treino, validando os ajustes por meio de validação cruzada.

Finalmente, a análise dos resultados foi enriquecida com o uso de ferramentas de interpretabilidade, como SHAP e gráficos do tipo Partial Dependence Plot (PDP), que permitiram avaliar o impacto das variáveis no comportamento dos modelos.

### 3.3 REATOR DE REFORMA A SECO

As reações realizadas e parâmetros abordados no conjunto de dados, são referentes a uma unidade piloto de reforma a seco, representada na (FIGURA 3), pertencente ao LABMATER (UFPR – Setor/Palotina).

FIGURA 4 – UNIDADE PILOTO DE REFORMA A SECO



FONTE: O autor (2024).

LEGENDA: (1) entrada dos gases; (2) misturador de gases; (3) forno de pré-aquecimento; (4) forno principal, (5), reator tubular de aço inox 316; (6) saída do gás; (7) painel de controle e segurança. (a) sensor de pressão; (b) termopar; (c) termopar; (d) sensor de umidade e temperatura.

Da Silva (2024), destaca que a instalação dos sensores acoplados à unidade ocupa posicionamentos que oferecem medições proporcionais às condições

termodinâmicas do sistema. No pré-forno, foram instalados um sensor de pressão e um termopar tipo K, enquanto no forno principal foi instalado outro termopar tipo K. Logo após o forno, há um sensor de umidade e um medidor de vazão, todos conectados a uma placa de prototipagem ESP32 e a um Raspberry Pi para controle e registro dos dados.

A coleta de dados desses sensores foi implementada pelos pesquisadores do laboratório utilizando uma topologia cliente/servidor entre a ESP32 e uma api Node.js, através de protocolo WebSocket, cujos dados ficam armazenados em um banco de dados configurado em máquina virtual específica, ambos hospedados no servidor local. Tais informações podem ser visualizadas pelos especialistas em uma interface Web desenvolvida pela própria equipe.

### 3.4 BIBLIOTECAS UTILIZADAS

Toda a modelagem e os cálculos foram realizados utilizando o *software* de código aberto Python. As bibliotecas empregadas foram as seguintes:

- a) Matplotlib: Biblioteca para a criação de gráficos estáticos, usada para visualizar métricas e dados ao longo das etapas do trabalho;
- b) NumPy: Biblioteca para cálculos numéricos eficientes, como operações vetoriais e manipulação de arrays, essencial para o pré-processamento e ajustes nos dados;
- c) Pandas: Ferramenta voltada para a manipulação e análise de dados tabulares, oferecendo funcionalidades para limpeza, transformação e organização dos dados usados no treinamento do modelo;
- d) Scikit-learn: Biblioteca que fornece ferramentas para a implementação, treinamento e validação de modelos de aprendizado de máquina;
- e) Scipy.stats e Math: Utilizadas para análises estatísticas avançadas e cálculos matemáticos gerais necessários no desenvolvimento do modelo;
- f) Seaborn: Complementa o Matplotlib, fornecendo visualizações estatísticas mais detalhadas para análise dos dados;
- g) SHAP: Aplicada para gerar explicações interpretáveis com base em valores SHAP, permitindo a análise do impacto de cada variável nas previsões dos modelos.

### 3.5 CONJUNTO DE DADOS

Os dados utilizados estão disponíveis através em um repositório no GitHub, publicado por Escribano (2024), e referem-se a registros das reações de reforma a seco para a produção de gás de síntese, que precedem o início deste trabalho. Paralelamente a esta pesquisa, está sendo desenvolvido outro trabalho com o objetivo de explorar de forma mais aprofundada a base de dados, detalhando seu processo de construção, suas características estatísticas e propondo possíveis melhorias. Optou-se pela utilização de uma base de dados preexistente, considerando a longa duração de muitas dessas reações, o que permitiu otimizar o tempo de pesquisa e os recursos envolvidos.

A utilização de dados reais dos fatores operacionais também garante maior representatividade e aplicabilidade dos resultados. O conjunto de dados inclui reações nas quais houve ou não entupimento do reator, proporcionando um cenário robusto para análise.

### 3.6 ANÁLISE E TRATAMENTO DOS DADOS

A amostra de dados utilizada neste estudo consiste em um conjunto de 9 reações, realizadas previamente no LABMATER. Contudo, uma delas foi subdividida em duas partes, pois incluiu uma regeneração pré-programada: uma parte representa o período anterior à regeneração, enquanto a outra se refere ao período posterior, ambas identificadas separadamente. Dessa forma, para este estudo, foram consideradas 10 reações no total.

As reações ocorrem de forma contínua, utilizando como gás de entrada uma mistura sintética de biogás com proporção estequiométrica aproximada de 1:1 entre  $\text{CH}_4$  e  $\text{CO}_2$ . Além disso, as temperaturas do pré-forno e do forno principal foram mantidas próximas a  $650^\circ\text{C}$  e  $800^\circ\text{C}$ , respectivamente, utilizando resistências elétricas.

Além do conjunto de dados, algumas análises exploratórias iniciais, disponibilizadas no GitHub, foram realizadas, aplicando técnicas diversas para examinar as propriedades dos dados e identificar padrões ou tendências relevantes para os modelos. Entre os resultados dessas análises, destaca-se a determinação da

granularidade dos dados e a consistência das coletas, medida pela quantidade média de registros por minuto.

Observou-se uma média de aproximadamente 2,45 registros por minuto, enquanto a mediana foi de 2,88, e o terceiro quartil atingiu cerca de 3 registros por minuto. Contudo, foram identificadas lacunas nos dados, com intervalos médios de aproximadamente 29 minutos sem registros, sendo que, em alguns casos, o intervalo entre um registro e outro ultrapassou 1 minuto, o que pode comprometer a continuidade das análises temporais.

Para avaliar a distribuição dos dados, aplicou-se o teste de Shapiro-Wilk, utilizando a biblioteca SciPy amplamente utilizado para verificar a normalidade de conjuntos amostrais. Os resultados, presentes na plataforma do GitHub, indicam que os dados não seguem uma distribuição normal, caracterizando-os como não paramétricos. Esse diagnóstico reforça a adequação de modelos que não exigem normalidade nos dados, os quais tendem a ter desempenho superior neste cenário.

As variáveis quantitativas de interesse utilizadas para representar os principais parâmetros operacionais do reator encontram-se detalhadas na (TABELA 2). Variáveis sem influência direta no processo químico, como medições de tensão ou corrente do painel e sensores sem registros, foram previamente excluídas. Além disso, para cada reação, todos os registros posteriores ao desligamento do reator foram removidos, garantindo que os modelos trabalhassem apenas com informações pertinentes ao regime de operação do reator.

TABELA 2 – PARÂMETROS OPERACIONAIS EXTRAÍDOS DA BASE DE DADOS E SEUS VALORES MÁXIMOS E MÍNIMOS

<b>Parâmetro</b>	<b>Notação na Base de Dados</b>	<b>Valor Mínimo</b>	<b>Valor Máximo</b>
Pressão no forno principal [bar]	sdp002	0,34	2,58
Temperatura no forno principal [°C]	sdt002	794,00	806,00
Temperatura no pré-forno [°C]	sdt003	650,00	654,00
Umidade relativa após o forno principal [% UR]	sdu002_1	18,00	91,50
Temperatura do gás de saída [°C]	sdu002_2	23,50	43,40
Tempo de reação [h]	tempo_hora	0	28,73

FONTE: O autor (2024).

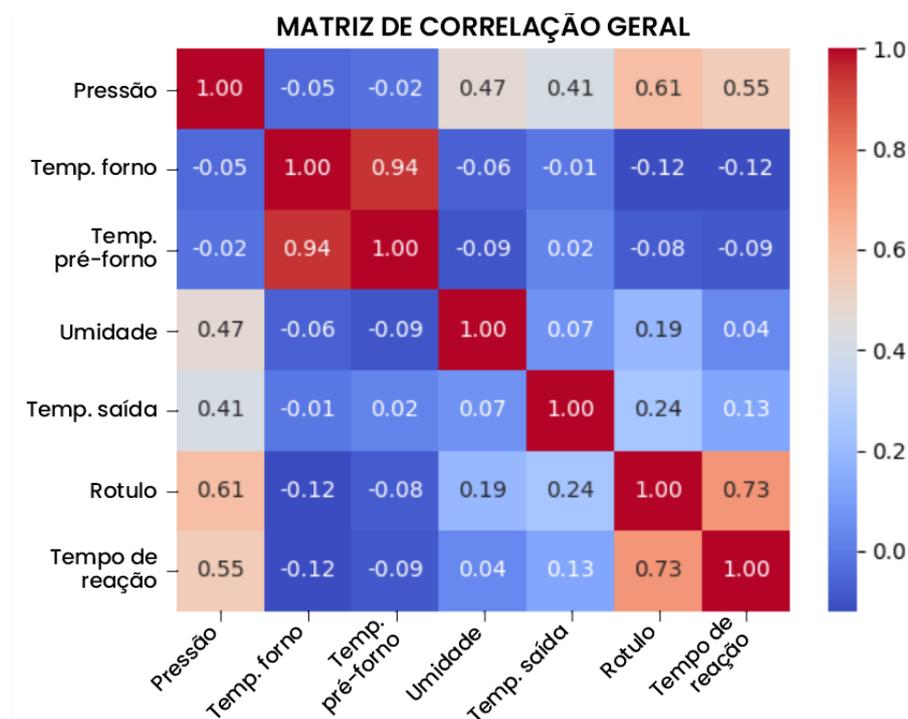
Consoante com o que se mostra na (TABELA 2), apenas três reações possuem registros de vazão do gás na saída do reator, inviabilizando o uso dessa variável nos modelos devido à insuficiência de dados. Além destes atributos conjunto de dados também inclui o identificador único de cada registro, o identificador da reação e o *timestamp* (data e hora) de cada medição. No entanto, essas informações não foram utilizadas diretamente como atributos nos conjuntos de treino e teste.

A base de dados estava previamente rotulada, seguindo um padrão original: rótulo 0 para ativação, rótulo 1 para reação e rótulo 2 para indicar irregularidades (neste caso, entupimentos e regeneração). Entretanto, os dados rotulados como 0, correspondentes à etapa de ativação, foram descartados, considerando que, em implementações iniciais, o modelo será aplicado diretamente à etapa de reação catalítica, sem necessidade de previsões para o período de ativação.

Para ajustar o modelo a prever problemas antecipadamente, a rotulagem foi redefinida. Agora, o rótulo 0 refere-se à reação normal, enquanto o rótulo 2 continua indicando o entupimento. Além disso, todas as linhas de dados registradas até 20 minutos antes do entupimento foram rotuladas com o valor 1, indicando uma janela de previsibilidade do entupimento. Essa modificação, foi feita com o intuito de antecipar a detecção do problema, possibilitando previsões preventivas.

Após o tratamento dos dados, realizou-se uma análise da matriz de correlação geral (FIGURA 5) entre as variáveis, revelando padrões relevantes.

FIGURA 5 – MATRIZ DE CORRELAÇÃO GERAL

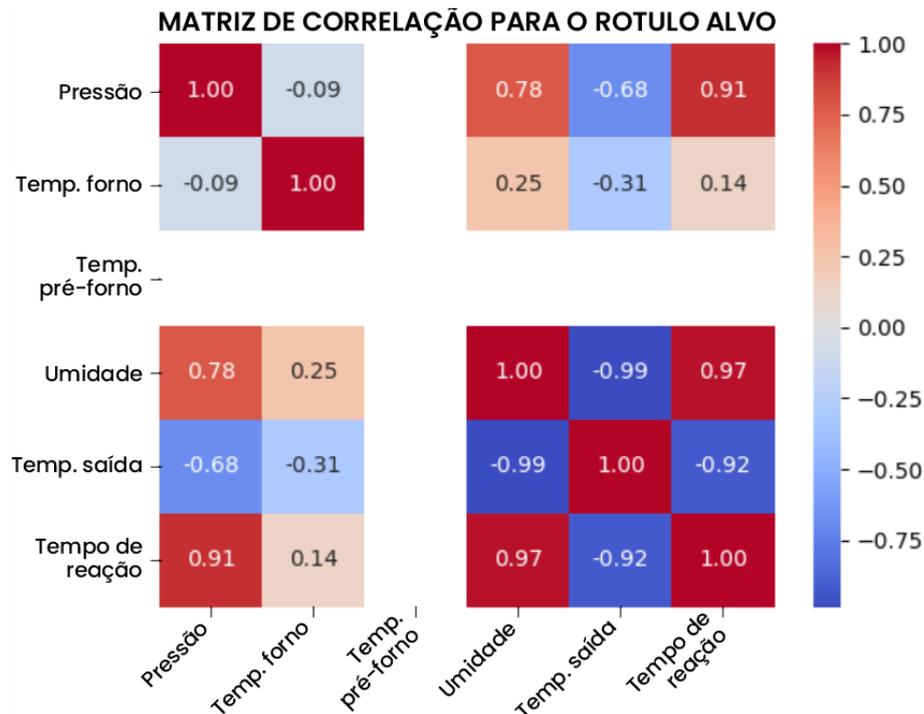


FONTE: O autor (2024).

Observa-se uma forte correlação entre a temperatura do pré-forno e a temperatura do forno principal, refletindo o comportamento sincrônico dessas variáveis durante a operação. Essa correlação é significativa, pois esses pontos de medição estão associados a resistências independentes. Assim, o monitoramento conjunto desses parâmetros pode auxiliar na identificação precoce de falhas operacionais.

A matriz de correlação restrita ao rótulo 1 (FIGURA 6) revelou um aumento nas correlações de algumas variáveis, especialmente entre pressão e umidade. Esse comportamento sugere que essas variáveis apresentam padrões distintos durante a formação do entupimento em comparação ao comportamento geral do processo. Contudo, vale lembrar que essa análise considera todas as reações com previsibilidade de entupimento, e o mesmo comportamento pode não ocorrer de forma uniforme em todas as reações.

FIGURA 6 – MATRIZ DE CORRELAÇÃO PARA O RÓTULO ALVO



FONTE: O autor (2024).

Outro ponto relevante é que a atributo sdt003 (temperatura do pré-forno) apresentou pouca variabilidade, sendo praticamente constante. Esse resultado está de acordo com o controle de temperatura realizado por resistências elétricas, que garantem estabilidade nesse parâmetro.

Das correlações realizadas, seis apresentaram valores significativamente altos. Destaca-se, novamente, a forte correlação entre as temperaturas do forno principal e do pré-forno, confirmando a estabilidade operacional nesses aspectos. Além disso, variáveis como umidade, pressão e tempo de reação demonstraram fortes correlações com o rótulo alvo, sugerindo que o tempo de reação, em particular, é uma variável crucial para a precisão dos modelos preditivos.

### 3.7 PREPARAÇÃO DO CONJUNTO DE DADOS

A normalização dos dados foi realizada através, da biblioteca Scikit-Learn, conforme descrito na documentação oficial e representado na (EQUAÇÃO 1). Essa técnica ajusta os valores das variáveis para um intervalo específico, neste caso, utilizando os parâmetros padrão entre 0 e 1, facilitando a comparação entre diferentes atributos e garantindo que todas sejam tratadas na mesma escala.

(EQUAÇÃO 1):

$$X_{\text{norm}} = X_{\text{min\_norm}} + \left( \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \right) (X_{\text{max\_norm}} - X_{\text{min\_norm}})$$

Onde:

$X_{\text{norm}}$  é o valor normalizado da variável no conjunto de dados antes da normalização;

$X_{\text{max\_norm}}$  é o valor máximo do intervalo desejado após a normalização, neste caso 1;

$X_{\text{min\_norm}}$  é o valor mínimo do intervalo desejado após a normalização, neste caso 0;

$X_{\text{max}}$  é o maior valor no conjunto de dados;

$X_{\text{min}}$  é o menor valor no conjunto de dados;

$X$  é o valor da variável no conjunto de dados antes da normalização.

No processo de normalização, foram excluídas as colunas correspondentes aos identificadores e rótulos, pois estas não fazem parte do conjunto de atributos.

### 3.8 SEPARAÇÃO DOS CONJUNTOS TESTE E TREINO

Para avaliar o desempenho dos modelos de aprendizado de máquina e garantir uma análise completa do conjunto de dados, foram adotadas duas estratégias distintas de separação. Cada abordagem foi escolhida com base em objetivos específicos, buscando explorar tanto a capacidade dos modelos de identificar padrões gerais quanto a sua habilidade de generalizar para cenários inéditos.

A primeira estratégia considerou todas as reações reunidas em uma única base de dados. Essa abordagem buscou maximizar a diversidade do conjunto, permitindo ao modelo aprender padrões globais presentes em todas as reações disponíveis, particionando a base em 70% para treinamento e 30% para teste.

A segunda estratégia, por outro lado, tratou o conjunto de teste como uma base completamente separada. Nesse caso, uma reação inteira foi excluída do treinamento e utilizada como conjunto de teste, enquanto as demais compuseram o conjunto de treinamento. Essa abordagem foi aplicada alternadamente às reações que continham instâncias do rótulo alvo, de forma que cada reação disponível fosse utilizada como base de teste em diferentes execuções. Essa estratégia foi projetada para simular condições reais de aplicação do modelo, em que os dados de uma reação

são completamente inéditos para o treinamento, avaliando, assim, sua capacidade de generalização em cenários desconhecidos.

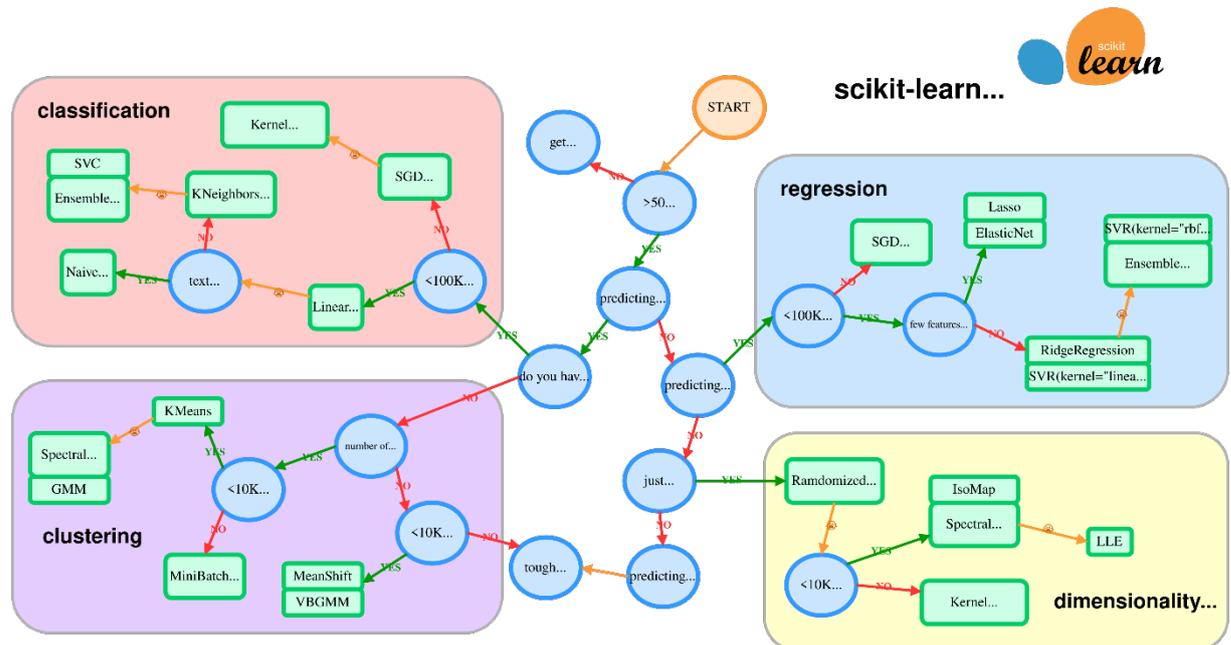
Independentemente da estratégia adotada, o particionamento interno dos conjuntos de treinamento seguiu os mesmos passos. Para ambas as abordagens, a técnica de validação cruzada, utilizando a função `StratifiedKFold` foi utilizada com 3 divisões (*folds*), garantindo a representatividade proporcional das classes em cada subdivisão. Essa metodologia é especialmente relevante para conjuntos de dados desbalanceados, como o utilizado neste estudo, pois assegura que as classes estejam adequadamente representadas durante o treinamento e a validação, quando utilizado o otimizador de hiperparâmetros (PEDREGOSA *et al.*, 2011).

Essas estratégias permitem explorar diferentes aspectos do desempenho dos modelos. Enquanto a primeira estratégia fornece uma visão ampla, aproveitando toda a diversidade do conjunto, a segunda avalia a robustez do modelo em condições mais desafiadoras, como a previsão para reações inéditas. Juntas, essas estratégias garantem uma análise completa e confiável do desempenho dos modelos, tanto em termos de aprendizado quanto de generalização.

### 3.9 SELEÇÃO E AVALIAÇÃO DE DESEMPENHO DOS MODELOS

Para a seleção de modelos apropriados utilizou-se a metodologia recomendada pelo Scikit-learn, na versão 1.5 (2024), representada na (FIGURA 7), considerando o tipo e o volume dos dados. A metodologia sugere que, para problemas com menos de cem mil amostras, onde a relação entre variáveis é não linear e os dados não são textuais, modelos como K-NN, SVC e RF são adequados para obtenção de um bom desempenho. Esta escolha também é ideal do ponto de vista do resultado do teste de Shapiro-Wilk, pois nenhum destes modelos exige normalidade no conjunto de dados.

FIGURA 7 –GUIA DE ESCOLHA DE ESTIMADORES PARA ABORDAGEM DE PROBLEMAS



FONTE: Adaptado de SCIKIT-LEARN, (2024).

Para o ajuste dos hiperparâmetros dos modelos, utilizou-se a função GridSearchCV como otimizador. A avaliação dos modelos priorizou o *recall* como métrica de revocação, com foco específico no rótulo 1. Essa escolha se justifica pelo objetivo principal do modelo: detectar de forma eficaz o rótulo 1, que indica a iminência de entupimento em até 20 minutos. O uso do *recall* é particularmente relevante nesse contexto, pois, para a aplicação prática, é preferível que o modelo emita um alerta indicando a possibilidade de entupimento, mesmo que ele não ocorra (falso positivo), do que deixar de emitir o alerta e um entupimento acontecer sem previsão (falso negativo). Dessa forma, fica a critério do especialista, ao receber a informação e avaliar a situação, definir os próximos passos operacionais.

A métrica de *recall* ponderado, apresentada na (EQUAÇÃO 2), foi utilizada para maximizar a sensibilidade do modelo na detecção de possíveis entupimentos, com foco no rótulo 1. Essa métrica leva em consideração a proporção de instâncias de cada rótulo no conjunto de dados, ponderando o *recall* individual de cada classe pelo número de instâncias da respectiva classe. Esse cálculo é especialmente útil em problemas multiclasse, como neste caso, que possui três rótulos distintos (0, 1 e 2), garantindo que a avaliação do modelo reflita melhor o desbalanceamento entre as classes.

(EQUAÇÃO 2):

$$\text{Recall}_{\text{weighted}} = \sum_{i=0}^n \frac{N_i}{N} \cdot \left( \frac{VP}{VP + VN} \right)_i$$

Onde:

$\text{Recall}_{\text{weighted}}$  é valor da métrica ponderada para a classe alvo, ser de 0 a 1;

$N_i$  é o número de instâncias da classe  $i$ ;

$N$  é o número total de instâncias no conjunto de dados;

$VP$  é o número de predições corretas para o rótulo alvo;

$VN$  é o número de vezes que o modelo não previu o rótulo alvo quando ocorrido.

Embora o *recall* para o rótulo 1 tenha sido utilizado como métrica principal no ajuste dos hiperparâmetros pelo GridSearchCV, a acurácia geral, acurácia para o rotulo 1, e a precisão ponderada, para o rotulo alvo, foram calculadas para fornecer uma avaliação mais ampla do desempenho dos modelos nos dados gerais. As métricas de acurácia e precisão para todos os rótulos são calculadas conforme as (EQUAÇÃO 3) e (EQUAÇÃO 4), respectivamente.

(EQUAÇÃO 3):

$$\text{Acurácia} = \left( \frac{VP + VN}{\text{Total de Casos}} \right)$$

Onde:

Acurácia é valor da métrica podendo ser de 0 a 1;

$VP$  é o número de predições corretas;

$VN$  é o número de predições corretas para as classes não rotuladas como positivas;

Total de Casos é o número de total de amostras avaliadas.

(EQUAÇÃO 4):

$$\text{Precisão}_{\text{weighted}} = \sum_{i=0}^n \frac{N_i}{N} \cdot \left( \frac{VP}{VP + FP} \right)_i$$

Onde:

Precisão<sub>weighted</sub> é valor da métrica ponderada para a classe alvo, ser de 0 a 1;

$N_i$  é o número de instâncias da classe  $i$ ;

$N$  é o número total de instâncias no conjunto de dados;

VP é o número de vezes que o modelo corretamente previu o rótulo alvo;

FP é o número de predições incorretas que classificaram como positivo um rótulo que deveria ser diferente.

Essas métricas foram calculadas utilizando as funções correspondentes da biblioteca Scikit-Learn, que realiza os cálculos de forma automática, considerando as características do conjunto de dados. Isso garante que tanto o *recall* ponderado quanto a precisão ponderada reflitam de maneira precisa o desempenho do modelo em problemas multiclasse.

É possível analisar a matriz de confusão, que fornece uma visão detalhada dos acertos e erros em cada rótulo, permitindo identificar como o modelo está classificando as instâncias de forma geral e para cada classe específica.

### 3.10 FERRAMENTAS DE APRENDIZADO DE MÁQUINA INTERPRETÁVEIS

As ferramentas de IML são técnicas que tornam o funcionamento dos modelos mais compreensível, permitindo que se entenda como cada variável influencia consistentemente nos resultados (KIM; KHANNA; KOYEJO, 2016). MOLNAR em 2020 fala, no livro *Interpretable machine learning: A guide for making black box models explainable*, como essas ferramentas são essenciais em contextos em que a explicabilidade<sup>2</sup> do modelo é fundamental, facilitando a análise de padrões e tendências nos dados.

Neste trabalho, foram utilizadas duas das principais ferramentas interpretáveis: SHAP e PDP, cada uma com abordagens específicas para a análise de variáveis. O SHAP, desenvolvido por SCOTT LUNDBERG e SU-IN LEE em 2017, é uma ferramenta baseada na teoria dos valores de Shapley, originária da teoria dos jogos cooperativos. O SHAP calcula a contribuição individual de cada variável para cada previsão específica, atribuindo a cada uma um valor SHAP que indica sua influência positiva ou negativa. Essa biblioteca permite análises detalhadas em nível

---

<sup>2</sup> Explicabilidade é a capacidade de um modelo de aprendizado de máquina de justificar suas predições de forma clara e compreensível, permitindo identificar como as variáveis influenciam os resultados.

local (para uma previsão específica) e global (para todo o conjunto de dados), fornecendo uma visão clara do comportamento do modelo e da interação entre variáveis (Lundberg e Lee, 2017). A capacidade do SHAP de identificar a influência das variáveis individualmente e suas interações faz dele uma ferramenta robusta e essencial para a interpretabilidade do modelo.

O PDP é uma técnica que mostra a relação média entre uma variável específica e a previsão do modelo, enquanto mantém todas as outras variáveis constantes. Essa ferramenta, disponível na biblioteca scikit-learn, é especialmente útil para entender como mudanças em uma variável de interesse impactam as previsões. O PDP é amplamente utilizado para identificar tendências e comportamentos não lineares no modelo, fornecendo uma análise visual clara de como cada variável influencia o resultado do modelo.

Além disso, neste trabalho, foram utilizados gráficos PDP 3D, que estendem a funcionalidade do PDP tradicional ao analisar a interação entre duas variáveis simultaneamente e suas relações conjuntas com as previsões do modelo. Esses gráficos permitem uma visualização mais rica e detalhada das dependências entre variáveis, capturando interações complexas que não são perceptíveis com gráficos PDP univariados. Essa abordagem é especialmente valiosa em sistemas com múltiplas variáveis correlacionadas, como os analisados neste estudo, fornecendo insights adicionais sobre como as variáveis se influenciam mutuamente e contribuem para o desempenho do modelo.

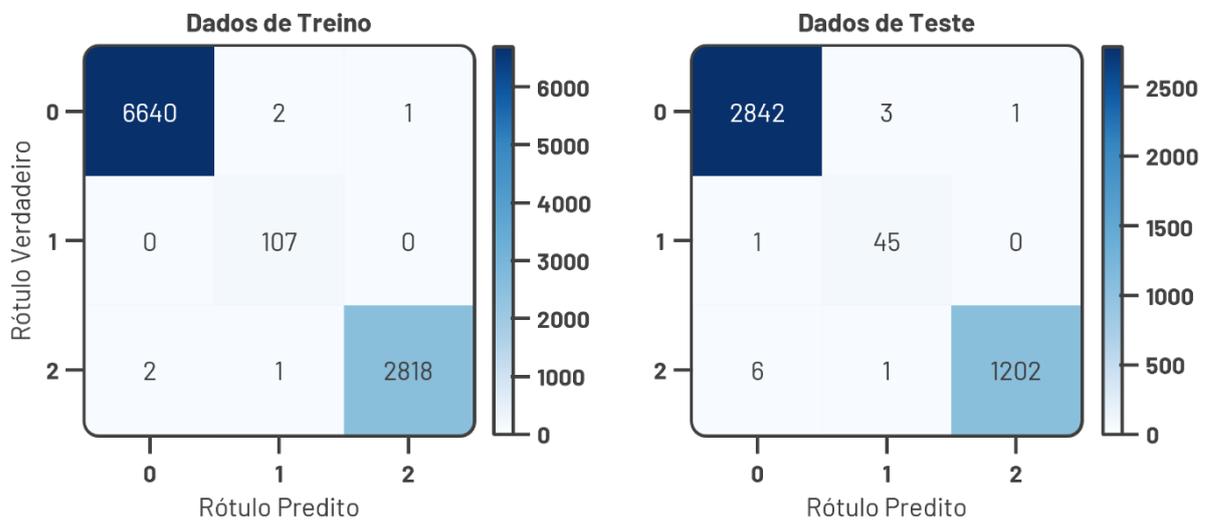
## 4 APRESENTAÇÃO DOS RESULTADOS

Conforme mencionado anteriormente, foram realizados testes utilizando base de dados de reações já realizadas LABMATER. Para três rótulos sendo eles: rótulo 0 - reação, rótulo 1 – faixa para previsibilidade do entupimento e rótulo 2- entupimento, para cada uma das metodologias selecionadas para este estudo: SVC, RF e K-NN.

### 4.1 BASE ÚNICA DE DADOS

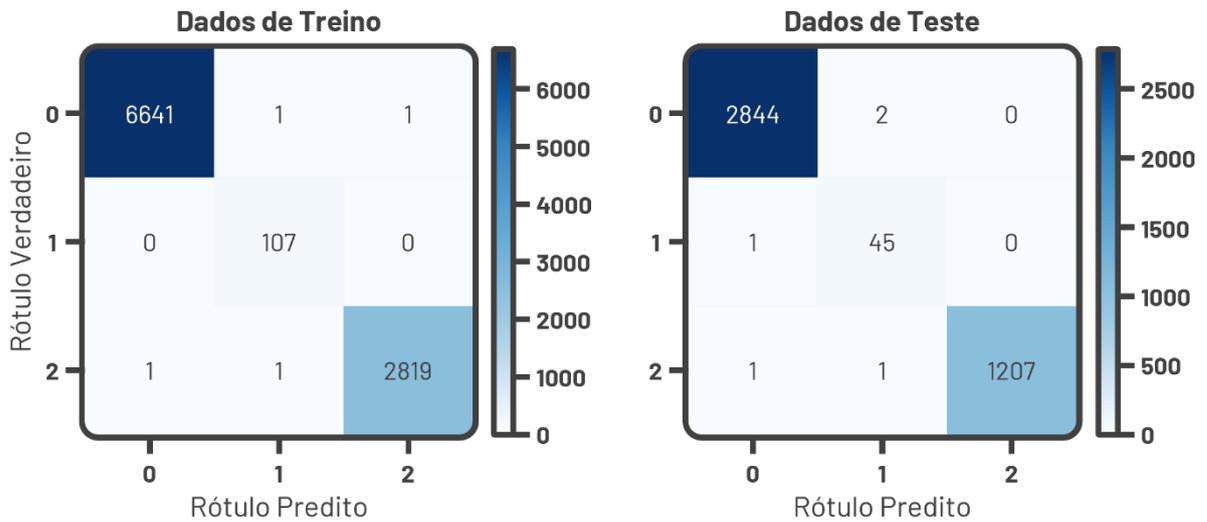
No primeiro momento, os resultados para quando utilizada a estratégia de organizar todas as reações em uma única base e fazer a separação em treino e teste (70/30) mantendo a cardinalidade para os rótulos. Vale lembrar, que para os dados de treino utilizou-se divisão estratificada junto com o otimizador de parâmetros. Conforme é possível visualizar nas (FIGURA 8, FIGURA 9, FIGURA 10) as matrizes de confusão resultantes dos modelos SVC, RF e K-NN, respectivamente.

FIGURA 8 – MATRIZES DE CONFUSÃO DO MODELO SVC COM DIVISÃO ESTRATIFICADA



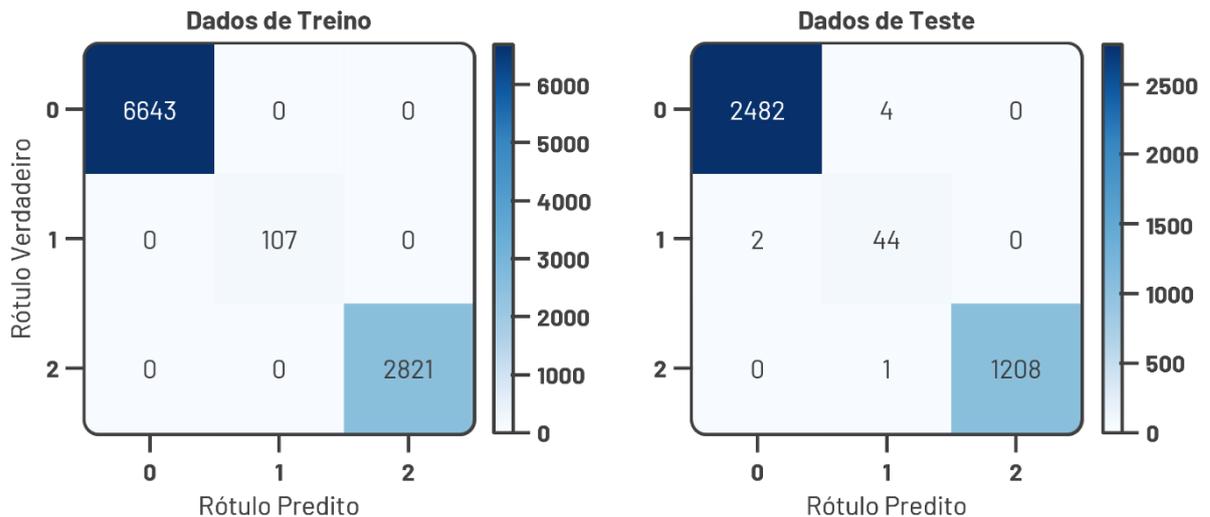
FONTE: O autor (2024).

FIGURA 9 – MATRIZES DE CONFUSÃO DO MODELO RF COM DIVISÃO ESTRATIFICADA



FONTE: O autor (2024).

FIGURA 10 – MATRIZES DE CONFUSÃO DO MODELO K-NN COM DIVISÃO ESTRATIFICADA



FONTE: O autor (2024).

Por meio da análise das matrizes de confusão e das métricas resultantes dos modelos, apresentadas na (TABELA 3), é perceptível que de fato o modelo conseguiu generalizar muito bem para o conjunto de dados proposto, sendo que para os três modelos a assertividade da classificação no treino é replicada para quando os modelos fazem previsões com os dados testes.

Concluiu-se que o modelo mais adequado para uma possível implementação, considerando como prioridade o melhor *recall* para o rótulo 1 e, secundariamente, a precisão, foi o Random Forest. Este modelo obteve precisão de 0,9375, *recall* de

0,97826 e acurácia geral de 0,97826, mesmo que o SVC tenha apresentado um *recall* perfeito de 1,0 para o rótulo 1.

A escolha pelo Random Forest baseia-se no fato de que, embora seu *recall* seja ligeiramente inferior ao do SVC, a diferença é pequena. Por outro lado, a precisão do Random Forest é significativamente maior, superando a do SVC, que foi de apenas 0,7541. Essa superioridade na precisão do Random Forest reforça sua capacidade de reduzir falsos positivos, tornando-o mais equilibrado e confiável para o problema em questão.

TABELA 3 – MÉTRICAS PARA OS DADOS DE TESTE GERAIS E DO ROTULO ALVO PARA QUANDO UTILIZADA UMA BASE UNIFICADA

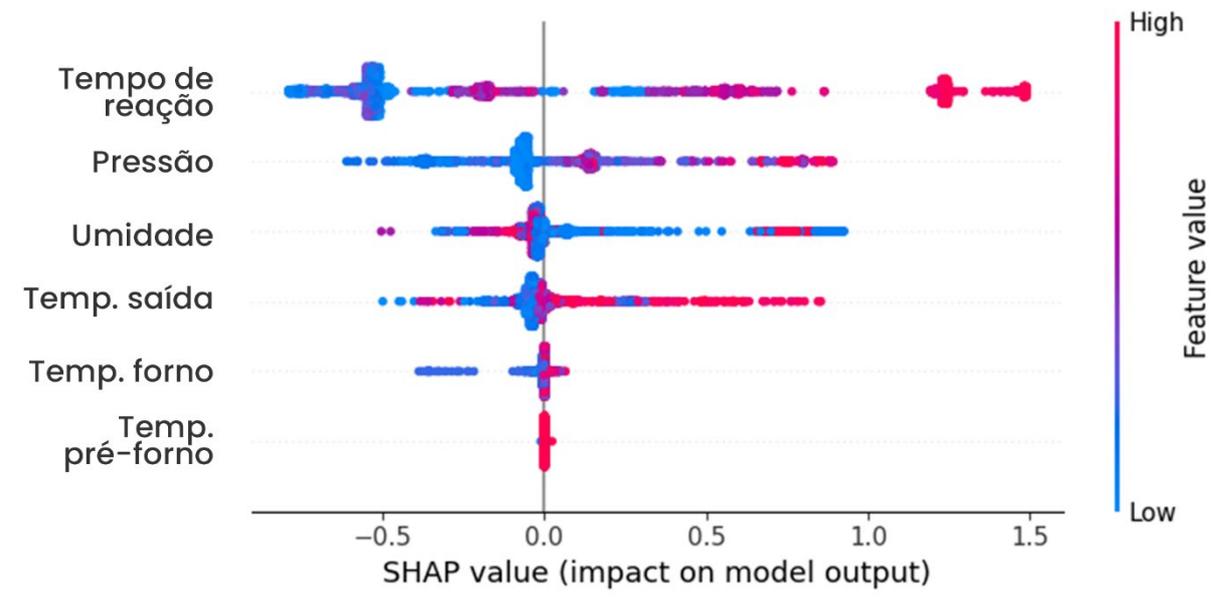
Modelo	Acurácia Geral	Acurácia (Rótulo 1)	Precisão (Rótulo 1)	Revocação (Rótulo 1)
SVC	0,99464	1,0	0,75410	1,0
K-NN	0,99829	0,95652	0,89796	0,95652
RF	0,99878	0,97826	0,93750	0,97826

FONTE: O autor, 2024.

Além disso, o modelo RF também teve resultado notável na predição dos demais rótulos, tendo uma acurácia geral de 0,99878. O K-NN também obteve métricas satisfatórias, sendo que o *recall* para a previsibilidade do entupimento obtido através do K-NN é de 0,95652, enquanto para o SVC é de 1,0 respectivamente.

A (FIGURA 11) apresenta o gráfico de exame de abelhas SHAP (SHAP *bee swarm plots*) gerados pelo RF. No eixo das abscissas, estão representados os diferentes atributos, cuja coloração reflete os valores individuais normalizados de cada uma, utilizando seus valores máximos e mínimos como referência de escala. Já no eixo das ordenadas, encontra-se o impacto do valor individual de cada atributo na predição do modelo.

FIGURA 11 – VALORES SHAP DO RANDOM FOREST UTILIZANDO BASE ÚNICA



FONTE: O autor (2024).

Observa-se, para o atributo que representa o tempo decorrido desde o início da reação (tempo\_hora), apresenta uma alta dispersão nos valores SHAP, variando entre, aproximadamente, -1 e 1,5. Isso indica que o impacto deste atributo na predição é altamente variável, ou seja, o modelo parece reagir de forma distinta a diferentes valores desse atributo. Essa variabilidade pode ser explicada por fatores como o número limitado de reações no conjunto de dados, principalmente, reações que tenham o rótulo alvo, e a diversidade de suas durações.

Além disso, a baixa representatividade de reações contendo todos os rótulos pode ter levado o modelo a associar alguns valores específicos de tempo\_hora a casos particulares, o que pode ter reduzido a capacidade de generalização para um conjunto de dados mais abrangente. Ainda assim, o tempo decorrido de reação, foi identificado como o parâmetro mais relevante para as predições do modelo, evidenciando seu papel crucial na construção das predições.

A pressão, representada pelo atributo sdp002, obteve valores SHAP variando entre -0,5 e 1. Observa-se que, na região negativa do eixo das abscissas, concentram-se dados com valores baixos para essa variável. Esses aglomerados, de coloração homogênea, sugerem que o modelo atribuiu graus de importância semelhantes a valores baixos desse atributo, indicando que tais valores podem ser comuns no conjunto de dados. Em contrapartida, na região positiva, verifica-se uma maior dispersão dos pontos à medida que os valores do atributo aumentam, o que sugere

que valores elevados de pressão possuem maior relevância para a classificação realizada pelo modelo. Ademais, destaca-se que este parâmetro foi identificado como o atributo de segunda maior relevância para as predições do modelo em questão, reforçando seu papel crítico na modelagem.

Os atributos `sdu002_1` e `sdu002_2`, que representam, respectivamente, a umidade e a temperatura do gás de saída do reator, apresentaram padrões de comportamento semelhantes nos valores SHAP. Ambas exibiram dispersão dos impactos variando entre -0,5 e 1,0 no eixo das ordenadas, com um aglomerado de pontos próximo ao valor 0. Esse padrão sugere que, para a maioria dos casos, os valores dessas variáveis possuem impacto moderado ou baixo na predição do modelo, especialmente em regiões próximas a valores médios dos atributos.

No entanto, em relação à temperatura de saída do gás, observa-se que, de modo geral, valores elevados foram associados a um impacto mais significativo nas predições, indicando uma maior relevância desses valores para o modelo. Para a umidade, nota-se que valores medianos possuem pouco impacto na predição. Contudo, os valores extremos, tanto altos quanto baixos, apresentam um impacto significativamente positivo, reforçando sua importância para a modelagem.

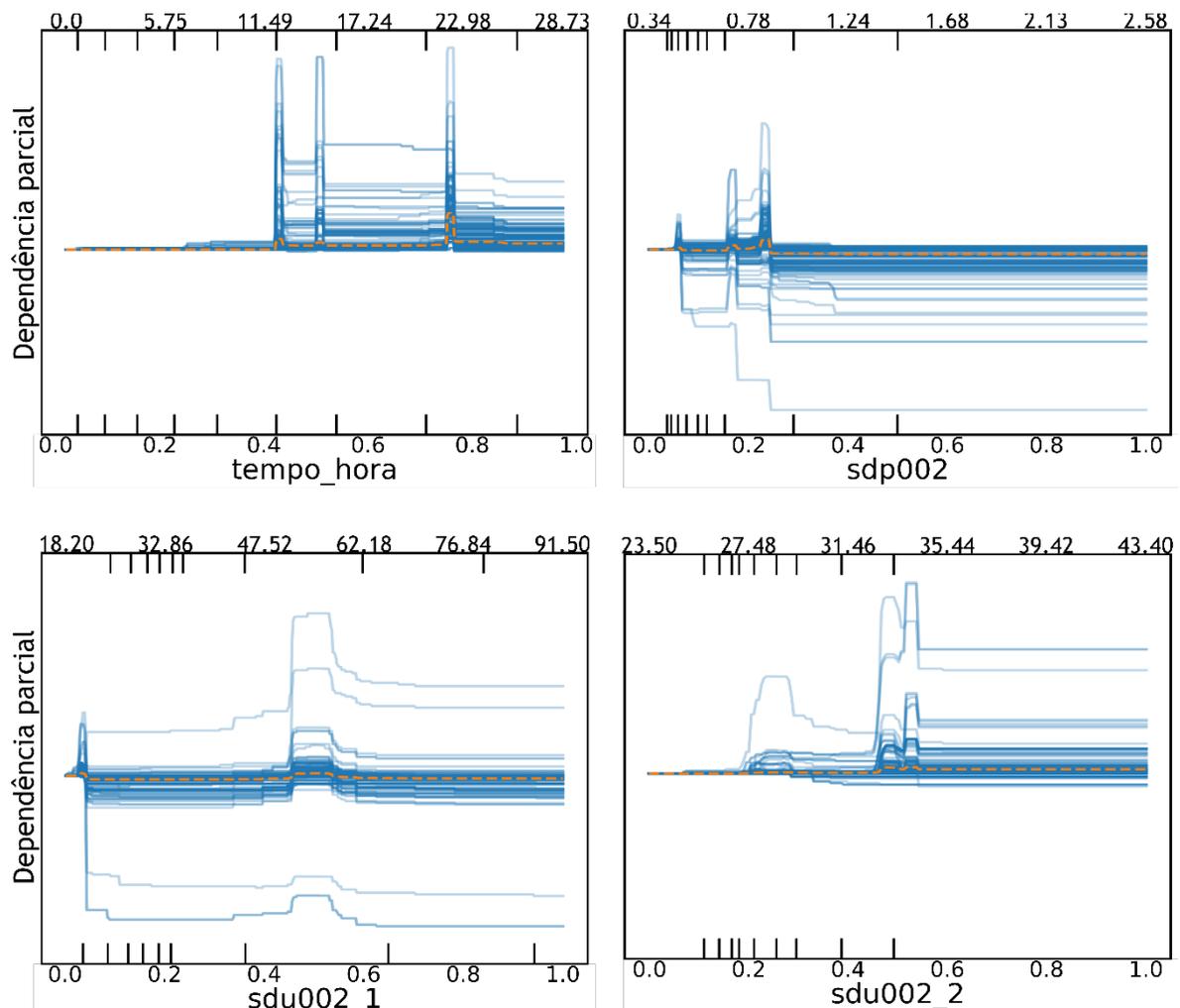
Essa relação está em conformidade com a dinâmica esperada do sistema, conforme indicado pelas reações paralelas apresentadas na (TABELA 1), que mostra o favorecimento do aumento da umidade no sistema através da reação de redução do CO e podendo, conseqüentemente, alterar a temperatura. Esses fenômenos podem causar o entupimento dos poros e o bloqueio dos sítios ativos do catalisador, levando, posteriormente, ao entupimento do reator (FIDALGO *et al.*, 2013). Nesse contexto, o modelo parece identificar essas variáveis como indicadores complementares do fenômeno de entupimento, captando variações significativas associadas a cenários críticos.

Ainda na (FIGURA 11), verifica-se que os atributos de temperatura do pré-forno (`sdt002`) e temperatura do forno (`sdt003`) apresentaram impactos extremamente baixos nas predições, sendo que `sdt003` demonstrou impacto praticamente nulo. Esse comportamento é consistente com as matrizes de correlação geral e do rótulo alvo na (FIGURA 5) e (FIGURA 6), respectivamente, que indicam uma fraca relação dessas variáveis com os rótulos do modelo. Isso reforça a menor contribuição desses atributos para a detecção do entupimento destas reações, uma vez que apresentam

baixa variabilidade, devido ao uso das resistências elétricas no forno, que resultam em um comportamento mais estável dessas variáveis ao longo do processo.

Na (FIGURA 12), estão apresentadas as expectativas condicionais individuais (*ICE - Individual Conditional Expectation*) e a dependência parcial (*PDP - Partial Dependence Plot*) para os quatro atributos que mostraram maior relevância nas predições do modelo, de acordo com os valores SHAP. Em azul, são exibidos os gráficos ICE para cada instância do modelo, enquanto em amarelo está a linha que representa o gráfico PDP, que mostra a correlação de dependência parcial com o rótulo 1, o objetivo de predição.

FIGURA 12 – DEPENDÊNCIA PARCIAL (PDP) E EXPECTATIVA CONDICIONAL INDIVIDUAL (ICE) DO RANDOM FOREST



FONTE: O autor (2024).

Observa-se que o atributo `tempo_hora` apresenta grandes variações no impacto sobre a previsibilidade para o rótulo 1, indicando elevado grau de importância,

especialmente para valores próximos de 11,5, 14,5 e 23,0. Sabendo que essas variações representam diferentes comportamentos do modelo, e considerando que esse atributo apresenta pouco impacto negativo na predição do rótulo 1, podemos inferir que o modelo está sobre ajustando o impacto do atributo para essas instâncias específicas. Destacando mais uma vez, a necessidade de uma base de dados com maior quantidade de reações, que inclua essas mesmas classificações, para melhorar a generalização do modelo.

Com relação ao atributo sdp002, que representa a pressão no reator, percebe-se que, em alguns casos, ela teve um impacto negativo na previsibilidade do entupimento para algumas instancias. Contudo, próximo a valores de 0,8 bar, o atributo mostrou um impacto positivo geral, deslocando a linha média do PDP, o que está alinhado com as análises dos valores SHAP. Isso indica que, quando os valores dessa variável são elevados, ela tem maior influência na predição do entupimento, reforçando o papel deste parâmetro no processo de modelagem.

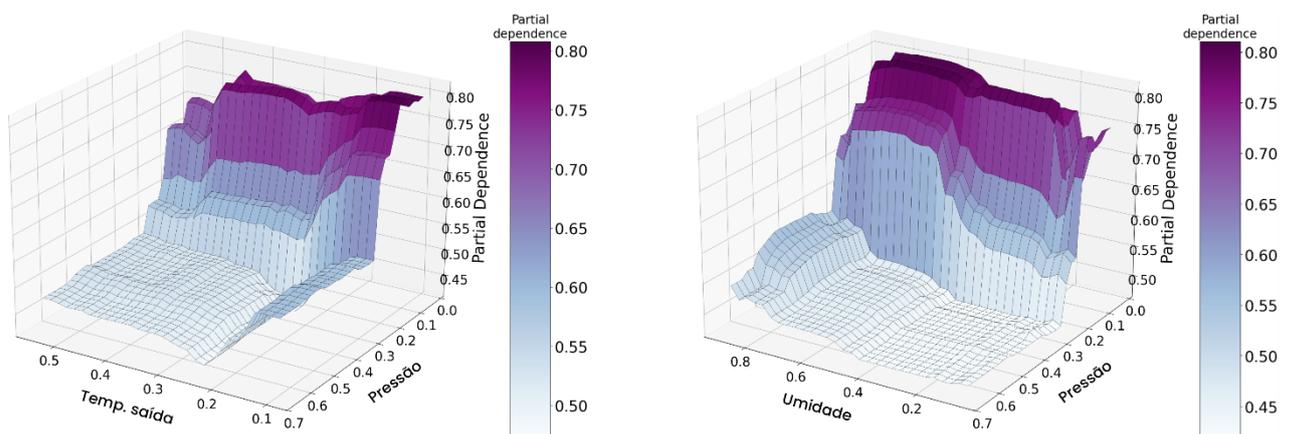
Para umidade do gás de saída (sdu002\_1), observa-se uma alta variabilidade nos dados, com um impacto significativo em várias instâncias, principalmente nos valores próximos a 55,0, onde o atributo demonstrou um bom grau de importância. Vale ressaltar que grande parte dos dados está concentrada entre 24% e 33%, mas também há uma distribuição considerável entre 62% e 67%. Isso indica que o modelo pode estar considerando que a umidade tem uma tendência a se concentrar em torno desses valores para a faixa de previsibilidade do entupimento. A densidade menor de dados nessa faixa pode explicar a maior relevância do atributo em torno de 50%, indicando que o modelo possa estar tentando generalizar para esta faixa considerando que são dados menos representativos.

Por fim, temperatura do gás de saída, representada pelo atributo em sdu002\_2, verifica-se uma variabilidade positiva mais pronunciada entre 27°C e 30°C. No entanto, a quantidade de casos não é suficiente para que essa variação tenha grande influência na linha de PDP. Outra observação se aplica para os valores entre 30°C e 36°C, que agora, apresentam uma variabilidade da linha de dependência parcial. Isso está alinhado com a hipótese de que, com a formação do entupimento, pode ocorrer um aumento da pressão do gás, o que leva a um leve aumento na temperatura, o que explica a correlação, demonstrada na (FIGURA 6), entre os atributos sdu002\_1 e sdp002 (SCHAFFNER, 2018).

A análise de gráficos PDPs bidimensionais é uma ferramenta essencial para compreender como duas variáveis, em conjunto, influenciam as previsões de um modelo (VELLAYAPPAN *et al.*, 2023). Enquanto os PDPs unidimensionais revelam o impacto isolado de uma variável, os gráficos bidimensionais capturam interações entre os atributos, permitindo identificar relações não lineares e efeitos combinados que seriam ignorados em análises individuais. Isso é particularmente relevante em sistemas complexos, como o analisado neste estudo, em que as variáveis estão fortemente interligadas por leis físicas e operacionais como pode-se verificar na lei geral dos gases que diz que para um volume fixo a pressão de um sistema é diretamente proporcional a sua temperatura e a lei de Dalton que mostra que a presença de vapor de água influencia significativamente na pressão parcial gás (JOHN BLAIR RUSSELL; AL, 1994).

Na (FIGURA 13), apresenta-se a dependência parcial bidimensional da pressão correlacionada com a umidade e a temperatura do gás de saída, representada por gráficos tridimensionais. Essa abordagem permite visualizar como as combinações desses atributos influenciam conjuntamente as previsões do modelo.

FIGURA 13 – DEPENDÊNCIA PARCIAL BIDIRECIONAL DE SDP002 CORRELACIONADO COM OS ATRIBUTOS SDU002\_1 E SDU002\_2



FONTE: O autor (2024).

A análise revela que, para valores baixos de pressão ( $sdp002 < 0,78$  bar), o impacto dos valores de umidade e de temperatura de saída aumenta consideravelmente. Para os valores de pressão, o mesmo comportamento é perceptível em situações em que a temperatura do gás é baixa ( $sdu002_2 < 28,5^{\circ}\text{C}$ ). Por outro lado, conforme os valores de umidade aumentam, observa-se um aumento

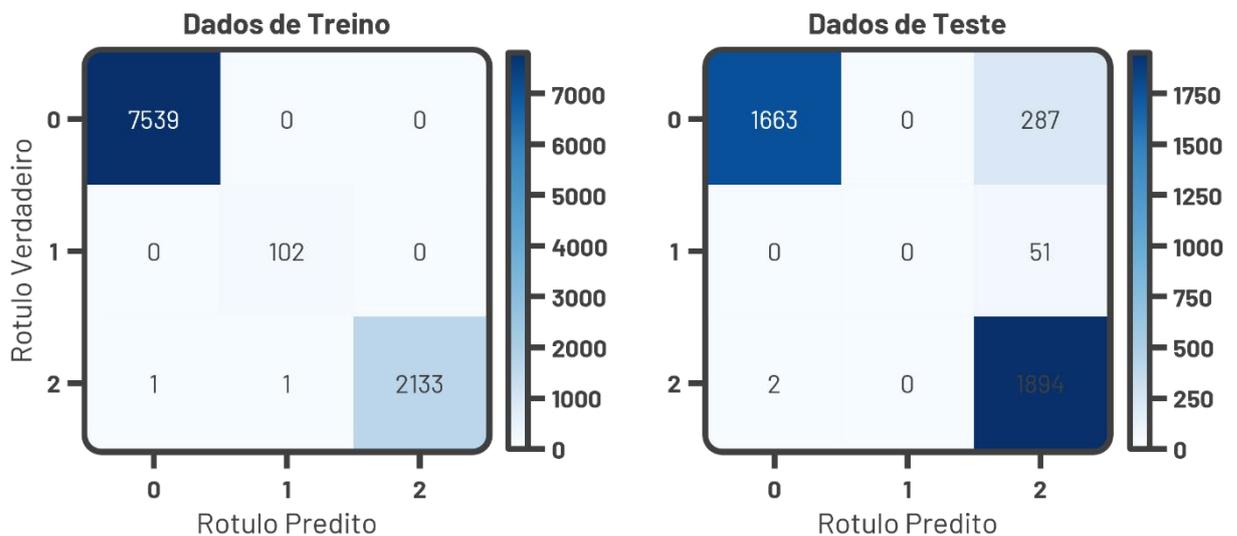
no impacto da pressão, indicando que o modelo considera altos níveis de umidade como um fator relevante para a classificação.

Apesar dessa interação entre pressão, umidade e temperatura, os PDPs também mostram que a variabilidade na dependência parcial da pressão é mais pronunciada em relação às demais variáveis. Isso reforça que sdp002 desempenha um papel mais relevante na modelagem, quando comparada com sdu002\_1 e sdu002\_2 e é, de fato, um dos atributos mais influentes, de acordo com este modelo.

#### 4.2 BASES INDIVIDUALIZADAS POR REAÇÕES

Para a estratégia de exclusão completa dos dados de uma única reação do conjunto de treinamento, utilizando-os como conjunto de teste, observou-se que as predições para o rótulo 1 não apresentaram resultados satisfatórios. Considerando o modelo que apresentou a melhor acurácia geral, disponível na (TABELA 4), no conjunto de teste, as matrizes de confusão do Random Forest, quando a base 24 foi utilizada como conjunto de teste, representadas na (FIGURA 14), evidenciam que o modelo não conseguiu generalizar para os dados dessa reação. Apesar de as predições no conjunto de treinamento terem demonstrado alta assertividade, o modelo falhou em identificar qualquer instância pertencente à classe alvo deste trabalho.

FIGURA 14 – DEPENDÊNCIA PARCIAL BIDIMENSIONAL DA PRESSÃO CORRELACIONADA COM A UMIDADE E A TEMPERATURA DO GÁS DE SAÍDA



FONTE: O autor (2024).

Esse comportamento revela que a base de dados disponível é incapaz de estabelecer correlações robustas que permitam a generalização do modelo para as reações 23, 24 e 262. Apesar de o modelo apresentar certa capacidade preditiva para outras classes, a (TABELA 4) mostra que a acurácia geral de todos os modelos é inferior quando comparada à (TABELA 3).

TABELA 4 – MÉTRICAS PARA OS DADOS DE TESTE GERAIS E DO ROTULO ALVO PARA QUANDO UTILISADA BASES SEPARADAS

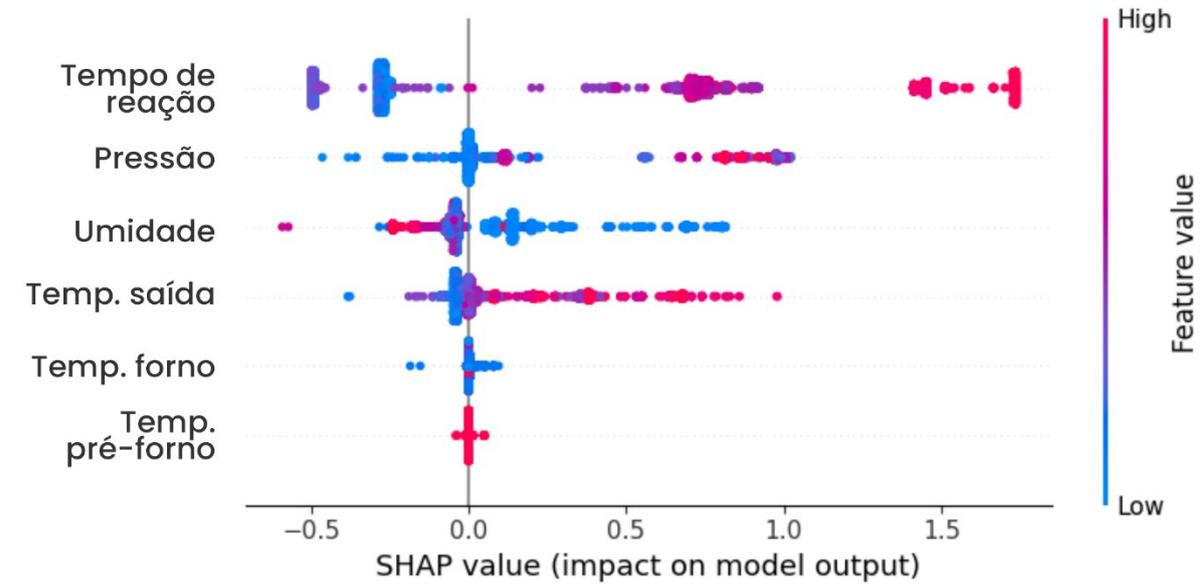
Modelo	Base Teste	Acurácia Geral	Acurácia (Rótulo 1)	Precisão (Rótulo 1)	Revocação (Rótulo 1)
SVC	23	0,83092	0,0	0,0	0,0
	24	0,83092	0,0	0,0	0,0
	262	0,82074	0,0	0,0	0,0
K-NN	23	0,79355	0,0	0,0	0,0
	24	0,79355	0,0	0,0	0,0
	262	0,72829	0,0	0,0	0,0
RF	23	0,83466	0,0	0,0	0,0
	24	0,91275	0,0	0,0	0,0
	262	0,83466	0,0	0,0	0,0

FONTE: O autor, 2024.

Conforme observado na (TABELA 4), o modelo K-NN apresentou os piores resultados, enquanto o Random Forest demonstrou o melhor desempenho em termos de acurácia geral. Esse problema pode ser atribuído ao número limitado de reações que possuem instâncias com o rótulo 1 e à limitação da base de dados com relação aos atributos utilizados.

Adicionalmente, a (FIGURA 15) ilustra os valores SHAP obtidos para o RF com a base 24 como conjunto de teste. A análise revela que a ordem de importância atribuída aos atributos pelo modelo permanece é a mesma dá à estratégia anterior. Contudo, os dados exibem pequenos agrupamentos ao longo do eixo vertical, indicando um sobreajuste ainda mais acentuado.

FIGURA 15 – VALORES SHAP DO RANDOM FOREST, CONJUNTO TESTE BASE 24



FONTE: O autor (2024).

Esse comportamento demonstra que, com o pequeno volume de dados disponíveis e os atributos utilizados, o modelo tende a sobreajustar, comprometendo sua capacidade de generalização. Embora a estratégia atual de exclusão de reações inteiras como base de teste seja relevante para avaliar a robustez do modelo, sua utilização, no contexto do rótulo 1, é inviável com o conjunto de dados estudado. Isso destaca a necessidade de expandir a base de dados, aumentando tanto o número de reações quanto a diversidade dos atributos, para viabilizar a aplicação dessa abordagem de modelagem.

## 5 CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma análise detalhada do desempenho dos modelos de aprendizado de máquina Random Forest, SVC e K-NN aplicados à previsão de entupimento em um reator de reforma a seco, utilizando um conjunto de dados proveniente do LABMATER. Os resultados mostraram que, para a estratégia de utilização de uma única base de dados com divisão estratificada (treinamento/teste), o modelo Random Forest demonstrou maior equilíbrio entre *recall* e precisão, destacando-se como o mais adequado para uma possível implementação inicial. No entanto, quando utilizada a estratégia de exclusão completa dos dados de uma reação específica para utilizá-la exclusivamente no conjunto de teste, o modelo apresentou dificuldades significativas em generalizar para novas reações, falhando completamente na previsão de instâncias do rótulo 1, que representa a faixa de previsibilidade de entupimento.

Esses resultados evidenciam limitações importantes na base de dados utilizada. A ausência de correlações robustas, entre as diferentes reações, aliada ao número reduzido de reações contendo o rótulo 1, comprometeu a capacidade dos modelos de generalizar para situações inéditas. Além disso, os atributos disponíveis, embora relevantes, não capturam informações essenciais que poderiam melhorar a robustez do modelo. Por exemplo, dados como composição do catalisador, granulometria, vazões da reação, composição dos gases na entrada e saída do reator, e taxa de conversão poderiam fornecer correlações mais sólidas entre as reações, permitindo que os modelos identificassem padrões mais consistentes.

Nesse contexto, o enriquecimento da base de dados com os atributos ausentes e com reações que variem os parâmetros operacionais pode, em trabalhos futuros, viabilizar estudos mais aprofundados e uma compreensão ampliada das reações sob diferentes condições operacionais.

Apesar das limitações da base de dados atual, a estratégia de exclusão de reações específicas, apresentadas na (TABELA 4) mesmo sem sucesso neste estudo, deve ser considerada ideal para futuras pesquisas. Essa abordagem oferece a possibilidade de desenvolver modelos mais robustos, capazes de prever cenários desconhecidos, desde que aplicados a bases de dados mais completas e diversificadas. Assim, para trabalhos futuros, é fundamental priorizar a ampliação e o enriquecimento da base de dados, com a inclusão de novas reações e o registro de

variáveis que capturem melhor as condições operacionais e as características físicas e químicas do sistema.

Outra recomendação para estudos futuros é considerar modelos probabilísticos que, além de retornarem à possibilidade de entupimento, retornem as probabilidades associadas a cada classe. Essa abordagem pode oferecer aos operadores informações adicionais sobre o nível de confiança do modelo em suas previsões, permitindo decisões mais informadas e estratégicas.

Por fim, o trabalho trouxe contribuições importantes ao laboratório, que está apenas começando suas investigações na área de aprendizado de máquina. Este estudo serve como um guia inicial, indicando as melhores metodologias e estratégias a serem exploradas, além de destacar os pontos críticos que precisam ser aprimorados. Apesar das limitações da base de dados, muitas das correlações adotadas pelos modelos estão de acordo com as dinâmicas físicas e químicas do sistema, demonstrando que o aprendizado de máquina pode capturar nuances consistentes com os princípios termodinâmicos que regem o comportamento do reator.

Portanto, este trabalho reforça a importância de investir em bases de dados mais robustas e em abordagens metodológicas avançadas, que poderão transformar modelos preditivos em ferramentas úteis e confiáveis para auxiliar na operação de reatores químicos.

## REFERÊNCIAS

- ABDIN, Z. *et al.* Hydrogen as an energy vector. *Renewable and Sustainable Energy Reviews*, 2020.
- ABDO, T. J. Anotaí: uma ferramenta para a anotação de bases textuais utilizando aprendizado de máquina iterativo. p. 51-54. Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, Curitiba (PR), 2021.
- AGUIAR, M. *et al.* Síntese de catalisadores de níquel suportados em si-mcm-41 a partir de diferentes fontes de sílica e sua aplicação no processo de reforma a seco do metano para produção de gás de síntese. *Revista Brasileira de Energias Renováveis*, v. 7, n. 3, 2018.
- ALMEIDA, J. C. Catalisadores Ni/BaO-Al<sub>2</sub>O<sub>3</sub> e Ni/BaO-SiO<sub>2</sub> para Reforma do metano com CO<sub>2</sub>. 116 f. Dissertação (Mestrado em Engenharia Química) Universidade Federal de São Carlos, São Carlos, 2012.
- ALVES, H. J. *et al.* Overview of hydrogen production technologies from biogas and the applications in fuel cells. *International Journal of Hydrogen Energy*. Anais...2013.
- BACH, V. R. Efeito da adição de Mg como promotor em catalisadores Ni/Al<sub>2</sub>O<sub>3</sub> aplicados à reforma a seco do metano. Dissertação de Mestrado. Universidade Federal do Paraná, 2016.
- BERKETIDOU, O. A.; GOULA, M. A. Biogas reforming for syngas production overnickel supported on ceria-alumina catalysts. *CatalToday*, v.195. p. 93-100, 2012.
- BERROCAL, G. J. P. Desenvolvimento de catalisadores baseados em níquel e rutênio para a reforma do metano. Tese de Doutorado, Universidade Federal da Bahia, 2009.
- CHEN, X. *et al.* Biogas dry reforming for syngas production: catalytic performance of nickel supported on waste-derived SiO<sub>2</sub>. *Catalysis Science & Technology*, v. 5, n. 2, p. 860-868, 2015.
- CHOI, R. Y. *et al.* Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science and Technology*, v. 9, n. 2, 2020.
- DA SILVA SZARBLEWSKI, M. *et al.* Métodos para a remoção de sulfeto de hidrogênio de efluentes gasosos aplicáveis a reatores anaeróbios. *Revista Jovens Pesquisadores*, v. 0, n. 1, 2012.
- DA SILVA VERAS, T. *et al.* Hydrogen: Trends, production and characterization of the main process worldwide. *International Journal of Hydrogen Energy*, 2017.
- DO AMARAL, A. C. *et al.* Fundamentos da digestão anaeróbia, purificação do biogás, uso e tratamento do digestato. 2019. Disponível em:

<https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1141297/1/final9895.pdf>. Acesso em: 26 out. 2024.

EMPRESA DE PESQUISA ENERGÉTICA (BRASIL). Balanço Energético Nacional 2023: Ano base 2023 / Empresa de Pesquisa Energética. – Rio de Janeiro: EPE, 2024.

ESCRIBANO, R. A. N. G. *dryReforming\_dataset*. 2024. Disponível em: [https://github.com/RenanAkiraEscribano/dryReforming\\_dataset](https://github.com/RenanAkiraEscribano/dryReforming_dataset). Acesso em: 22 nov. 2024.

FNR. Fachagentur Nachwachsende Rohstoffe e.V. Guia prático do biogás. Ministério da Nutrição, Agricultura e Defesa do Consumidor da Alemanha, v. 5, 2010.

FRARE, L. M *et al*. Processo para remoção de ácido sulfídrico de biogás. Engenharia Sanitária Ambiental. v. 14, n. 2, p.167-172, 2009.

FUKUROZAKI, S. H. Avaliação do ciclo de vida de potenciais rotas de produção de hidrogênio: estudo dos sistemas de gaseificação da biomassa e de energia solar fotovoltaica. Tese de Doutorado. Universidade de São Paulo, 2011.

GARCIA, G. D. S. Hidrogênio verde: seu papel estratégico na transição energética. 2024.

GEELS, F. W. *et al*. Sociotechnical transitions for deep decarbonization. *Science*, 2017.

GIELEN, D. *et al*. The role of renewable energy in the global energy transformation. *Energy Strategy Reviews*, v. 24, 2019.

JOHN BLAIR RUSSELL; AL, E. Química geral. Noções preliminares. São Paulo (Sp): Makron Books, 1994. p.16-64.

JOHN BLAIR RUSSELL; AL, E. Química geral. Gases. São Paulo (Sp): Makron Books, 1994. p.162-225.

JOHN LU, Z. Q. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Journal of the Royal Statistical Society Series A: Statistics in Society*, v. 173, n. 3, 2010.

JAMES, G. M. *et al*. An Introduction to Statistical Learning: With Applications in R. New York: Springer; 2013.

PHILIPPI JR, A. *et al*. Energia e sustentabilidade. Barueri, SP: Editora Manole, 2016. E-book. ISBN 9786555761313.

KIM, Been; KHANNA, Rajiv; KOYEJO, Oluwasanmi O. Examples are not enough, learn to criticize! criticism for interpretability. *Advances in neural information processing systems*, v. 29, 2016.

KWAŚNY, J.; *et al.* Sorbents used for biogas desulfurization in the adsorption process. *Polish Journal of Environmental Studies*, v. 25, n. 1, 2016.

LORA, E. E. S. *et al.* Gaseificação e pirólise para a conversão da biomassa em eletricidade e biocombustíveis. *Biocombustíveis*. Editora Integrada: Rio de, v. 1, 2012.

LUISETTO, I. *et al.* Ni/CeO<sub>2</sub>-Al<sub>2</sub>O<sub>3</sub> catalysts for the dry reforming of methane: The effect of CeAlO<sub>3</sub> content and nickel crystallite size on catalytic activity and coke resistance. *Applied Catalysis A: General*, v. 500, 2015.

MOLNAR, C. *Interpretable machine learning: A guide for making black box models explainable*. Leanpub, 2020.

MONTENEGRO, D. L. Reforma a seco de metano com catalisadores Ni/MCM-41 sintetizados a partir de fontes alternativas de sílica. 2013. Dissertação (Mestrado em Engenharia Química) – Universidade Federal do Rio Grande do Norte, Natal, 2013.

MOSAAD AWAD, M. *et al.* Recent developments and current trends on catalytic dry reforming of Methane: Hydrogen Production, thermodynamics analysis, techno feasibility, and machine learning. *Energy Conversion and Management*, 2024.

NAKAJIMA, E. A. *et al.* Kinetics of dry reforming of methane catalyzed by Ni/Si-MCM-41. *International journal of hydrogen energy*, v. 48, n. 83, p. 32331–32341, 2023.

OLIVEIRA, L. G. Reforma a seco do biogás para obtenção de hidrogênio verde em unidade piloto utilizando catalisadores Ni/Si-MCM-41. 100f. Mestrado - Universidade Federal do Paraná, Setor de Palotina, Pós-Graduação em Bioenergia, Palotina, 2021. 100f.

OYAMA, S. T. *et al.* Dry reforming of methane has no future for hydrogen production: Comparison with steam reforming at high pressure in standard and membrane reactors. *International Journal of Hydrogen Energy*, v. 37, n. 13, 2012.

PAWAR, V. *et al.* Study of the combined deactivation due to sulfur poisoning and carbon deposition during biogas dry reforming on supported Ni catalyst. *Industrial & Engineering Chemistry Research*, v. 56, n. 30, p. 8448-8455, 2017.

PIREZ, C. *et al.* Steam reforming, partial oxidation and oxidative steam reforming for hydrogen production from ethanol over cerium nickel based oxyhydride catalyst. *Applied Catalysis A: General*, v. 518, 2016.

PUZICKI, Diego *et al.* Potencial de produção de hidrogênio para o acionamento de células a combustível a partir da reforma do biogás: estudo de caso. *Revista Gestão & Sustentabilidade Ambiental*, v. 4, p. 394-410, 2015.

RAMOS, A. L. D. *et al.* Atual estágio de desenvolvimento da tecnologia GTL e perspectivas para o Brasil. *Química Nova*, v. 34, n. 10, p. 1704-1716, 2011.

RATHOD, V.; BHALE, P. V. Experimental investigation on biogas reforming for syngas production over an alumina based nickel catalyst. Energy Procedia. Anais...2014.

ROH, J. *et al.* Interpretable machine learning framework for catalyst performance prediction and validation with dry reforming of methane. Applied Catalysis B: Environmental, v. 343, 2024.

ROSSETTI, I. *et al.* Syngas production via steam reforming of bioethanol over Ni-BEA catalysts: A BTL strategy. International Journal of Hydrogen Energy, v. 41, n. 38, 2016.

RUSSELL, Stuart J.; NORVIG, Pedro. Inteligência Artificial: Uma Abordagem Moderna. 4ª edição. Rio de Janeiro: GEN LTC, 2022. E-book. pág.590. ISBN 9788595159495.

SALIAN, I. Difference between supervised, unsupervised, & reinforcement learning. The Official NVIDIA Blog. Published August 2, 2018.

SANTOS, F. M.; SANTOS, F. Combustível "hidrogênio". Millenium - Jornal de Educação, Tecnologias e Saúde. n. (31). p. 252–270, 2016.

SAWANT, V.; DESHMUKH, R.; AWATI, C. Machine learning techniques for prediction of capacitance and remaining useful life of supercapacitors: A comprehensive review. Journal of Energy Chemistry, 2023.

SCHAFFNER, R. A. Catalisadores Ni/Al<sub>2</sub>O<sub>3</sub> para a reforma a seco do metano: efeito da temperatura de calcinação do suporte e uso de promotor alcalino. Dissertação de Mestrado em Engenharia de Energia na Agricultura. Universidade Estadual do Oeste do Paraná, 2018.

SCIKIT-LEARN: Machine Learning in Python. Versão 1.5. Scikit-Learn. 2024. Disponível em: <https://scikit-learn.org/stable/> Acesso em: 26 Out. 2024.

SPATH, P. L.; DAYTON, D. C. Preliminary Screening -- Technical and Economic Assessment of Synthesis Gas to Fuels and Chemicals with Emphasis on the Potential for Biomass-Derived Syngas. National Renewable Energy Laboratory, n. December, 2003.

STAFFELL, I. *et al.* The role of hydrogen and fuel cells in the global energy system. Energy and Environmental Science, 2019.

VARGAS, R. A. *et al.* Hidrogênio: o vetor energético do futuro. São Paulo: Centro de Ciência e Tecnologia de Materiais (CCTM), Instituto de Pesquisas Energéticas e Nucleares (IPEN), 2006.

VELLAYAPPAN, K. *et al.* Impacts of catalyst and process parameters on Ni-catalyzed methane dry reforming via interpretable machine learning. Applied Catalysis B: Environmental, v. 330, 2023.

XUE, Z. *et al.* Autothermal reforming of ethyl acetate for hydrogen production over Ni<sub>3</sub>La<sub>7</sub>O<sub>y</sub>/Al<sub>2</sub>O<sub>3</sub> catalyst. *Energy Conversion and Management*, v. 146, 2017.

ZANETTE, A. L. Potencial do aproveitamento energético do biogás no Brasil. 105 f. Dissertação (Mestre em Planejamento Energético). Universidade Federal do Rio de Janeiro – UFRJ, Rio de Janeiro, 2009.

ZHANG, F. *et al.* The survey of key technologies in hydrogen energy 100 storage. *International Journal of Hydrogen Energy*, v. 41, n. 33, p. 14535-14552, 2016.

ZHANG, Q.; HU, J.; LEE, D. J. Biogas from anaerobic digestion processes: Research updates. *Renewable Energy*, v. 98, 2016.