

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE QUÍMICA

Danielle Gradin Queiroz



DESENVOLVIMENTO DE SENSORES VIRTUAIS PARA PREDIÇÃO
DA COMPOSIÇÃO DOS EFLUENTES DE UMA UNIDADE DE
TRATAMENTO DE ÁGUAS ÁCIDAS

RIO DE JANEIRO

2023

Danielle Gradin Queiroz

DESENVOLVIMENTO DE SENSORES VIRTUAIS PARA PREDIÇÃO DA
COMPOSIÇÃO DOS EFLUENTES DE UMA UNIDADE DE TRATAMENTO DE ÁGUAS
ÁCIDAS

Trabalho de Conclusão de Curso apresentado à Escola de Química da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do grau de Engenheiro Químico.

Orientador(es): Maurício Bezerra de Souza Jr.

Príamo Albuquerque Melo Jr.

Júlia do Nascimento Pereira Nogueira

Rio de Janeiro

2023

Q3d

Queiroz, Danielle Gradin
Desenvolvimento de sensores virtuais para
predição da composição dos efluentes de uma Unidade
de Tratamento de Águas Ácidas / Danielle Gradin
Queiroz. -- Rio de Janeiro, 2023.
105 f.

Orientador: Maurício Bezerra de Souza Jr.
Coorientador: Príamo Albuquerque Melo Jr.
Trabalho de conclusão de curso (graduação) -
Universidade Federal do Rio de Janeiro, Escola de
Química, Bacharel em Engenharia Química, 2023.

1. Águas ácidas. 2. Sensor virtual. 3.
Aprendizado de Máquina. 4. Banco de dados. 5. Aspen
Plus Dynamics. I. Bezerra de Souza Jr, Maurício,
orient. II. Albuquerque Melo Jr, Príamo, coorient.
III. Título.

Danielle Gradin Queiroz

DESENVOLVIMENTO DE SENSORES VIRTUAIS PARA PREDIÇÃO DA
COMPOSIÇÃO DOS EFLUENTES DE UMA UNIDADE DE TRATAMENTO DE ÁGUAS
ÁCIDAS

Trabalho de Conclusão de Curso apresentado à
Escola de Química da Universidade Federal do
Rio de Janeiro, como parte dos requisitos
necessários à obtenção do grau de Engenheiro
Químico.

Aprovado em 01 de novembro de 2023.

Maurício Bezerra de Souza Jr., D.Sc., Escola de Química

Príamo Albuquerque Melo Jr., D.Sc., PEQ/COPPE/UFRJ

Júlia do Nascimento Pereira Nogueira, M.Sc., Deloitte

Kese Pontes Freitas Alberton, D.Sc., Escola de Química

Francisco Davi Belo Rodrigues, B.Sc., Petróleo Brasileiro S.A.

Rio de Janeiro
2023

Aos meus pais, meu irmão, meu noivo e amigos queridos que, com muito apoio e dedicação, não mediram esforços para que eu chegasse até esta etapa da minha vida.

AGRADECIMENTOS

Gostaria de agradecer primeiramente à minha família que sempre esteve ao meu lado, acreditando nos meus sonhos e dando todo o suporte emocional, financeiro e psicológico que precisei. Aos meus pais, Fátima e Gilberto, vocês são a base da minha vida, serei sempre grata por toda dedicação e amor. Ao meu irmão, Leonardo, obrigada por ter sido um exemplo e uma inspiração para mim desde sempre.

Também gostaria de agradecer ao meu noivo, Carlos, que foi a pessoa mais presente em toda minha trajetória, desde a faculdade até este TCC. Obrigada por sempre me dar segurança para que eu acreditasse em mim e nos meus sonhos, pelo apoio emocional e por sempre me fazer rir e ver o lado bom da vida.

Aos meus grandes amigos que conheci no IFRJ: Beatriz, Gustavo, Micaela, Gabi, Carlos e Rafael, obrigada por terem compartilhado a melhor experiência acadêmica da minha vida.

Aos amigos que fiz durante essa jornada de 6 anos na UFRJ: Leandro, Juliana, Matheus, Yohana, Ariel, Lara e Thiago, vocês foram essenciais para que esse momento se concretizasse.

Quero também expressar minha gratidão aos profissionais que fizeram parte tanto da minha jornada acadêmica quanto profissional. Um agradecimento especial à minha primeira orientadora de Iniciação Científica, Michele, que abriu as portas da pesquisa acadêmica para mim. Também aos meus supervisores de estágio, Larisse e Francesco, que se tornaram grandes amigos e investiram no meu potencial, sempre me desafiando a crescer. Vocês desempenharam um papel fundamental para a engenheira que me tornei hoje.

Por último, mas não menos importante, quero agradecer aos meus orientadores, professor Príamo Melo, professor Maurício Bezerra e Júlia Nogueira, por dedicarem seu tempo, conhecimento e experiência para me guiar nesse processo. Sou muito grata pela oportunidade de aprender com vocês e por todo o suporte oferecido ao longo desta jornada.

“Machines take me by surprise with great frequency”.

- Alan Turing

RESUMO

QUEIROZ, Danielle. **Desenvolvimento de sensores virtuais para predição da composição dos efluentes de uma Unidade de Tratamento de Águas Ácidas**. Rio de Janeiro, 2023. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) - Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2023.

As águas ácidas são um dos principais resíduos aquosos gerados durante o refino do petróleo e necessitam ser processadas em Unidades de Tratamento de Águas Ácidas (UTAA) para remoção de contaminantes como H_2S e NH_3 , a fim de atender às demandas da legislação ambiental brasileira. Dessa forma, é imprescindível o monitoramento em tempo real da composição dos efluentes gerados pela UTAA, como o gás ácido, o gás amoniacal e a água tratada. Portanto, o presente trabalho tem como objetivo criar um novo banco de dados utilizando, como base, a simulação dinâmica de uma UTAA com duas colunas de esgotamento desenvolvida em *Aspen Plus Dynamics*® V10 por Nogueira (2021) e, posteriormente, desenvolver sensores virtuais capazes de prever as frações mássicas de H_2S e NH_3 dos efluentes previamente citados. O novo banco de dados apresenta situações de operação normal da unidade com distúrbios nas variáveis de interesse, e poderá ser utilizado no futuro para outros estudos da área. Para criação dos sensores virtuais, o método Florestas Aleatórias foi escolhido como modelo de Aprendizado de Máquina. Todos os códigos foram desenvolvidos em *Python* no ambiente de execução *Visual Studio Code*. Os resultados encontrados foram validados através da criação de novos modelos via *Gradient Boosting* e Máquinas de Vetores de Suporte, utilizando as funções disponíveis na biblioteca *scikit-learn*. Obteve-se como resultado o desenvolvimento de seis sensores virtuais com R^2 maior que 0,87, MAE menor que 0,12, MSE menor que 0,17 e RMSE menor que 0,41. A análise da importância das variáveis revelou que a temperatura do 2º estágio da Coluna 1 exerce influência significativa no equilíbrio termodinâmico de separação do H_2S e NH_3 das águas ácidas e é essencial para cinco dos seis sensores virtuais. Foi observado também que o aumento desta temperatura melhora a recuperação de H_2S . O estudo desta variável sugere que incrementos adicionais nesta temperatura, acima de 90% de recuperação, resultam em ganhos mínimos na remoção do H_2S e na vaporização acentuada de NH_3 no gás ácido, o que pode levar a problemas operacionais nas Unidades de Recuperação de Enxofre.

Palavras-chave: água ácida; sensor virtual; Aprendizado de Máquina; banco de dados; *Aspen Plus Dynamics*®; Floresta Aleatória; Máquinas de vetores de suporte; *Gradient Boosting*.

ABSTRACT

QUEIROZ, Danielle. **Desenvolvimento de sensores virtuais para predição da composição dos efluentes de uma Unidade de Tratamento de Águas Ácidas**. Rio de Janeiro, 2023. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) - Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2023.

Sour waters are one of the main aqueous byproducts generated during petroleum refining and require processing in Sour Water Treatment Units (SWTUs) to remove contaminants such as H₂S and NH₃ in compliance with Brazilian environmental legislation. Thus, online monitoring of the composition of SWTU effluents including acid gas, ammoniacal gas, and treated water is essential. This study aims to create a new database based on the dynamic simulation of a SWTU with two stripping columns developed in Aspen Plus Dynamics® V10 by Nogueira (2021). Subsequently, it aims to develop virtual sensors capable of real-time prediction of the mass fractions of H₂S and NH₃ in the aforementioned effluents. The new database encompasses normal unit operation scenarios with disturbances in the variables of interest and may be used for future studies in the field. The Random Forest method was selected for virtual sensor creation, implemented in Python within the Visual Studio Code environment. Results were validated by creating additional models using Gradient Boosting and Support Vector Machines functions available in the scikit-learn library. The outcome was the development of six virtual sensors with R² greater than 0,87, MAE less than 0,12, MSE less than 0,17, and RMSE less than 0,41. Variable importance analysis revealed that the temperature of the second stage of Column 1 significantly influences the thermodynamic equilibrium of H₂S and NH₃ separation from sour waters, being critical for five of the six virtual sensors. It was also observed that increasing this temperature improves the recovery of H₂S. However, additional increases in temperature, above 90% recovery, result in minimal gains in H₂S removal and significant vaporization of NH₃ in the acid gas, which can lead to operational issues in Sulfur Recovery Units.

Keywords: sour water; virtual sensor; machine learning; database; Aspen Plus Dynamics®; Random Forest; Support Vector Machines; Gradient Boosting.

LISTA DE ILUSTRAÇÕES

Figura 1.1: Esquema de contextualização da geração e tratamento de águas ácidas.	1
Figura 2.1: Constantes de Henry em função da temperatura para os compostos mais comuns encontrados em sistemas de águas ácidas.	7
Figura 2.2: Esquema de uma Unidade de Tratamento de Água Ácida com uma coluna.	9
Figura 2.3: Esquema de uma Unidade de Tratamento de Água Ácida com duas colunas.	10
Figura 2.4: Diagrama mostrando a relação entre inteligência artificial, <i>machine learning</i> e <i>deep learning</i>	11
Figura 2.5: Esquema geral de um algoritmo de Florestas Aleatórias.	14
Figura 2.6: Esquema geral de um algoritmo do Gradient Boosting.	16
Figura 2.7: Esquema de SVM Linear para um exemplo de classificação binária.	17
Figura 3.1: Fluxograma geral da metodologia.	21
Figura 3.2: Diagrama de blocos da simulação dinâmica de Nogueira (2021).	22
Figura 3.3: Representação da simulação dinâmica de Nogueira (2021).	25
Figura 3.4: Adaptação da simulação dinâmica de Nogueira (2021) com a inclusão dos novos seis controladores: GACID_H2S, GACID_NH3, GAMON_H2S, GAMON_NH3, AGUA_H2S e AGUA_NH3.	28
Figura 3.5: Representação genérica do pop-up de um elemento de divisão.	30
Figura 3.6: Representação do pop-up de erro de integração durante simulação no Aspen Plus Dynamics®.	30
Figura 3.7: Cortes da otimização dos hiperparâmetros do algoritmo Florestas Aleatórias.	34
Figura 3.8: Coordenadas paralelas da otimização dos hiperparâmetros do algoritmo Florestas Aleatórias.	34
Figura 4.1: Variáveis importantes para determinação da fração de H ₂ S no gás ácido.	37
Figura 4.2: Variáveis importantes para determinação da fração de NH ₃ no gás ácido.	38
Figura 4.3: Variáveis importantes para determinação da fração de H ₂ S no gás amoniacal.	38
Figura 4.4: Variáveis importantes para determinação da fração de NH ₃ no gás amoniacal.	39
Figura 4.5: Variáveis importantes para determinação da fração de H ₂ S na água tratada.	39
Figura 4.6: Variáveis importantes para determinação da fração de NH ₃ na água tratada.	40
Figura 4.7: Relação entre recuperação do H ₂ S no gás ácido e temperatura do 2º estágio da Coluna 1.	42
Figura 4.8: Predição do sensor GACID_H2S usando SVM – kernel linear.	47

Figura 4.9 Comparação do Coeficiente de Determinação (R^2) entre os algoritmos de RF, GB e SVM kernel RBF.....	47
Figura 4.10: Comparação da Raiz do Erro Quadrático Médio (RMSE) entre os algoritmos de RF, GB e SVM kernel RBF.	48
Figuras 4.11 A e B: Predição do sensor GACID_H2S usando RF com dados de treinamento (A) e teste (B).....	50
Figuras 4.12 A e B: Resíduos do sensor GACID_H2S usando RF com dados de treinamento (A) e teste (B).....	51
Figura 4.13: Classificação da importância das variáveis para o sensor GACID_H2S.....	52
Figura 4.14: Classificação da importância das variáveis para o sensor GACID_H2S após redução.	53
Figuras 4.15 A e B: Predição do sensor GACID_NH3 usando RF com dados de treinamento (A) e teste (B).....	54
Figuras 4.16 A e B: Resíduos do sensor GACID_NH3 usando RF com dados de treinamento (A) e teste (B).....	56
Figura 4.17: Classificação da importância das variáveis para o sensor GACID_NH3.	56
Figura 4.18: Classificação da importância das variáveis para o sensor GACID_NH3 após redução.	57
Figuras 4.19 A e B: Predição do sensor GAMON_H2S usando RF com dados de treinamento (A) e teste (B).....	59
Figura 4.20: Predição para corrida número 7 (banco de teste) do sensor GAMON_H2S.	60
Figura 4.21: Diferentes faixas de confiabilidade para o sensor GAMON_H2S.	61
Figuras 4.22 A e B: Resíduos do sensor GAMON_H2S usando RF com dados de treinamento (A) e teste (B).....	62
Figura 4.23: Classificação da importância das variáveis para o sensor GAMON_H2S.	63
Figura 4.24: Classificação da importância das variáveis para o sensor GAMON_H2S após redução.	64
Figuras 4.25 A e B: Predição do sensor GAMON_NH3 usando RF com dados de treinamento (A) e teste (B).....	65
Figura 4.26: Predição para corrida número 7 (banco de teste) do sensor GAMON_NH3.....	66
Figura 4.27: Diferentes faixas de confiabilidade para o sensor GAMON_NH3.....	67
Figuras 4.28 A e B: Resíduos do sensor GAMON_NH3 usando RF com dados de treinamento (A) e teste (B).....	69
Figura 4.29: Classificação da importância das variáveis para o sensor GAMON_NH3.....	69

Figura 4.30: Classificação da importância das variáveis para o sensor GAMON_NH3 após redução.	70
Figuras 4.31 A e B: Predição do sensor AGUA_H2S usando RF com dados de treinamento (A) e teste (B).	72
Figuras 4.32 A e B: Resíduos do sensor AGUA_H2S usando RF com dados de treinamento (A) e teste (B).	72
Figura 4.33: Classificação da importância das variáveis para o sensor AGUA_H2S.	73
Figura 4.34: Classificação da importância das variáveis para o sensor AGUA_H2S após redução.	74
Figuras 4.35 A e B: Predição do sensor AGUA_NH3 usando RF com dados de treinamento (A) e teste (B).	76
Figuras 4.36 A e B: Resíduos do sensor AGUA_NH3 usando RF com dados de treinamento (A) e teste (B).	76
Figura 4.37: Classificação da importância das variáveis para o sensor AGUA_NH3.	77
Figura 4.38: Classificação da importância das variáveis para o sensor AGUA_NH3 após redução.	78
Figura A.1: Simulação dinâmica com a inclusão dos novos seis sensores – versão expandida.	90
Figuras C.1 A e B: Predição (A) e resíduo (B) do sensor GACID_H2S usando GB.	92
Figuras C.2 A e B: Predição (A) e resíduo (B) do sensor GACID_NH3 usando GB.	92
Figuras C.3 A e B: Predição (A) e resíduo (B) do sensor GAMON_H2S usando GB.	93
Figuras C.4 A e B: Predição (A) e resíduo (B) do sensor GAMON_NH3 usando GB.	93
Figuras C.5 A e B: Predição (A) e resíduo (B) do sensor AGUA_H2S usando GB.	94
Figuras C.6 A e B: Predição (A) e resíduo (B) do sensor AGUA_NH3 usando GB.	94
Figuras C.7 A e B: Predição (A) e resíduo (B) do sensor GACID_H2S usando SVM kernel RBF.	95
Figuras C.8 A e B: Predição (A) e resíduo (B) do sensor GACID_NH3 usando SVM kernel RBF.	96
Figuras C.9 A e B: Predição (A) e resíduo (B) do sensor GAMON_H2S usando SVM kernel RBF.	96
Figuras C.10 A e B: Predição (A) e resíduo (B) do sensor GAMON_NH3 usando SVM kernel RBF.	97
Figuras C.11 A e B: Predição (A) e resíduo (B) do sensor AGUA_H2S usando SVM kernel RBF.	97

Figuras C.12 A e B: Predição (A) e resíduo (B) do sensor AGUA_NH3 usando SVM kernel RBF.	98
Figuras C.13 A e B: Predição (A) e resíduo (B) do sensor GACID_H2S usando SVM kernel linear.	98
Figuras D.1 A e B: Predição (A) e resíduo (B) do sensor GACID_H2S usando RF após redução das variáveis de entrada.	99
Figuras D.2 A e B: Predição (A) e resíduo (B) do sensor GACID_NH3 usando RF após redução das variáveis de entrada.	99
Figuras D.3 A e B: Predição (A) e resíduo (B) do sensor GAMON_H2S usando RF após redução das variáveis de entrada.	100
Figuras D.4 A e B: Predição (A) e resíduo (B) do sensor GAMON_NH3 usando RF após redução das variáveis de entrada.	100
Figuras D.5 A e B: Predição (A) e resíduo (B) do sensor AGUA_H2S usando RF após redução das variáveis de entrada.	101
Figuras D.6 A e B: Predição (A) e resíduo (B) do sensor AGUA_NH3 usando RF após redução das variáveis de entrada.	101
Figuras E.1: Predições individuais das 12 corridas presentes no banco de teste para o sensor GAMON_H2S usando o melhor modelo do RF.	102
Figuras E.2: Predições individuais das 12 corridas presentes no banco de teste para o sensor GAMON_NH3 usando o melhor modelo do RF.	104

LISTA DE TABELAS

Tabela 3.1: Legenda das correntes da simulação dinâmica de Nogueira (2021).	23
Tabela 3.2: Propriedades físicas das correntes de entrada do sistema.....	24
Tabela 3.3: 14 controladores presentes na simulação dinâmica de Nogueira (2021).....	26
Tabela 3.4: Parâmetros de ajuste dos controladores da simulação dinâmica de Nogueira (2021).	27
Tabela 3.5: Recuperação dos componentes da simulação dinâmica de Nogueira (2021).	27
Tabela 3.6: Limites inferiores e superiores definidos para cada hiperparâmetro.	33
Tabela 4.1: Variações da fração mássica de H ₂ S no gás ácido para cada corrida global.	41
Tabela 4.2: Resultados da análise exploratória na variável da temperatura do 2° estágio da Coluna 1.	42
Tabela 4.3: Comparação dos métodos de integração Implicit Euler (padrão) e Gear (alternativo) utilizando as corridas globais com distúrbios mais intensos.	44
Tabela 4.4: Listagem das corridas presentes no banco de treinamento.....	45
Tabela 4.5: Listagem das corridas presentes no banco de teste.....	45
Tabela 4.6: Hiperparâmetros otimizados para o sensor GACID_H2S usando RF.....	49
Tabela 4.7: Métricas do sensor GACID_H2S após avaliação com dados de treinamento e dados de teste.	50
Tabela 4.8: Métricas do banco de teste do sensor GACID_H2S após redução das variáveis..	53
Tabela 4.9: Hiperparâmetros otimizados para o sensor GACID_NH3 usando RF.	54
Tabela 4.10: Métricas do sensor GACID_NH3 após avaliação com dados de treinamento e dados de teste.	54
Tabela 4.11: Métricas do banco de teste do sensor GACID_NH3 após redução das variáveis.	57
Tabela 4.12: Hiperparâmetros otimizados para o sensor GAMON_H2S usando RF.	58
Tabela 4.13: Métricas do sensor GAMON_H2S após avaliação com dados de treinamento e dados de teste.	59
Tabela 4.14: Métricas calculadas para cada faixa de confiabilidade do sensor GAMON_H2S.	62
Tabela 4.15: Métricas do banco de teste do sensor GAMON_H2S após redução das variáveis.	63
Tabela 4.16: Hiperparâmetros otimizados para o sensor GAMON_NH3 usando RF.....	64

Tabela 4.17: Métricas do sensor GAMON_NH3 após avaliação com dados de treinamento e dados de teste.	65
Tabela 4.18: Métricas calculadas para cada faixa de confiabilidade do sensor GAMON_NH3.	68
Tabela 4.19: Métricas do banco de teste do sensor GAMON_NH3 após redução das variáveis.	70
Tabela 4.20: Hiperparâmetros otimizados para o sensor AGUA_H2S usando RF.....	71
Tabela 4.21: Métricas do sensor AGUA_H2S após avaliação com dados de treinamento e dados de teste.	71
Tabela 4.22: Métricas do banco de teste do sensor AGUA_H2S após redução das variáveis.	74
Tabela 4.23: Hiperparâmetros otimizados para o sensor AGUA_NH3 usando RF.	75
Tabela 4.24: Métricas do sensor AGUA_NH3 após avaliação com dados de treinamento e dados de teste.	75
Tabela 4.25: Métricas do banco de teste do sensor AGUA_NH3 após redução das variáveis.	78
Tabela 4.26: Análise global da importância das variáveis de todos os sensores após redução.	79
Tabela C.1: Análise global de importância das variáveis de todos os sensores virtuais usando Gradient Boosting após redução das variáveis de entrada.	95

LISTA DE ABREVIATURAS E SIGLAS

AI	<i>Artificial Intelligence</i> (Inteligência Artificial)
CART	<i>Classification and Regression Trees</i> (Árvores de Classificação e Regressão)
EDO	Equações Diferenciais Ordinárias
FDD	<i>Fault Detection and Diagnosis</i> (Detecção e Diagnóstico de Falhas)
GB	<i>Gradient Boosting</i>
GPSWAT	<i>Gas Process Association Sour Water Equilibrium</i>
HDT	Unidade de Hidrotratamento
MAE	<i>Mean Absolute Error</i> (Erro Absoluto Médio)
MLP	<i>Multilayer Perceptron</i> (Perceptron multicamadas)
ML	<i>Machine Learning</i> (Aprendizado de Máquina)
MSE	<i>Mean Squared Error</i> (Erro Quadrático Médio)
PID	Proporcional Integral Derivativo
RBF	<i>Radial Basis Function</i> (Função de Base Radial)
RF	<i>Random Forest</i> (Florestas Aleatórias)
RMSE	<i>Root Mean Squared Error</i> (Raiz do Erro Quadrático Médio)
SP	<i>Set-point</i> (Valor alvo)
SVM	<i>Support Vector Machines</i> (Máquinas de Vetores de Suporte)
URA	Unidade de Regeneração de Amina
URE	Unidade de Recuperação de Enxofre
UTAA	Unidade de Tratamento de Águas Ácidas
VI	<i>Variable Importance</i> (Importância das variáveis)

SUMÁRIO

LISTA DE ILUSTRAÇÕES	x
LISTA DE TABELAS	xiv
LISTA DE ABREVIATURAS E SIGLAS	xvi
1. Introdução.....	1
1.1 Motivação	1
1.2 Objetivos.....	3
1.3 Estrutura do trabalho.....	4
2. Revisão de Literatura	5
2.1 Unidades de Tratamento de Águas Ácidas	5
2.2 Inteligência Artificial	11
2.2.1 Florestas Aleatórias (<i>Random Forests</i> , RF)	13
2.2.2 <i>Gradient Boosting</i> (GB)	15
2.2.3 Máquinas de Vetores de Suporte (<i>Support Vector Machines</i> , SVM)	17
2.3 Sensores virtuais	18
3. Metodologia	21
3.1 <i>Hardware e Software</i>	21
3.2 Histórico do Processo Simulado	22
3.3 Modificação da simulação dinâmica base de Nogueira (2021)	27
3.4 Desenvolvimento do banco de dados.....	29
3.5 Avaliação dos métodos de integração do <i>Aspen Plus Dynamics</i> ®.....	30
3.6 Inteligência Artificial	31
3.6.1 Preparação dos dados de treinamento e teste	31
3.6.2 Algoritmo da Inteligência Artificial	32
3.6.3 Hiperparâmetros	33
3.6.4 Métricas	35

4.	Resultados e Discussões.....	37
4.1	Desenvolvimento do banco de dados e estudo fenomenológico do processo	37
4.2	Avaliação dos métodos de integração do <i>Aspen Plus Dynamics</i> ®.....	43
4.3	Preparação dos dados de treinamento e teste.....	44
4.4	Comparação dos algoritmos de Inteligência Artificial	46
4.5	Resultado dos sensores virtuais usando Florestas Aleatórias	49
4.5.1	Sensor GACID_H2S	49
4.5.2	Sensor GACID_NH3.....	53
4.5.3	Sensor GAMON_H2S.....	58
4.5.4	Sensor GAMON_NH3	64
4.5.5.	Sensor AGUA_H2S	70
4.5.6	Sensor AGUA_NH3.....	74
4.6	Análise global da importância das variáveis e estudo fenomenológico do processo	78
5.	Conclusões e Considerações Finais.....	82
	Referências Bibliográficas.....	85
	APÊNDICE A – Adaptação da simulação dinâmica de Nogueira (2021) com a inclusão dos novos seis controladores.....	90
	APÊNDICE B – Banco de dados desenvolvido	91
	APÊNDICE C – Resultados dos sensores virtuais usando <i>Gradient Boosting</i> e SVM – <i>kernel</i> RBF e linear.....	92
C.1	Resultados dos sensores virtuais usando <i>Gradient Boosting</i>	92
C.1.1	Sensor GACID_H2S	92
C.1.2	Sensor GACID_NH3.....	92
C.1.3	Sensor GAMON_H2S	93
C.1.4	Sensor GAMON_NH3	93
C.1.5	Sensor AGUA_H2S.....	94
C.1.6	Sensor AGUA_NH3	94

C.2 Tabela global de importância das variáveis de todos os sensores virtuais usando <i>Gradient Boosting</i> após redução das variáveis de entrada.....	95
C.3 Resultados dos sensores virtuais usando SVM – <i>kernel</i> RBF.....	95
C.3.1 Sensor GACID_H2S	95
C.3.2 Sensor GACID_NH3	96
C.3.3 Sensor GAMON_H2S	96
C.3.4 Sensor GAMON_NH3	97
C.3.5 Sensor AGUA_H2S.....	97
C.3.6 Sensor AGUA_NH3	98
C.4 Resultados dos sensores virtuais usando SVM – <i>kernel</i> linear	98
C.4.1 Sensor GACID_H2S	98
APÊNDICE D – Gráficos de predição e resíduos após redução das variáveis de entrada para os sensores virtuais desenvolvidos com Florestas Aleatórias	99
D.1 Sensor GACID_H2S	99
D.2 Sensor GACID_NH3	99
D.3 Sensor GAMON_H2S	100
D.4 Sensor GAMON_NH3.....	100
D.5 Sensor AGUA_H2S.....	101
D.6 Sensor AGUA_NH3	101
APÊNDICE E – Gráficos de predição para cada uma das 12 corridas do banco de teste para os sensores virtuais GAMON_H2S e GAMON_NH3.....	102
E.1 Sensor GAMON_H2S.....	102
E.2 Sensor GAMON_NH3	104

Capítulo 1

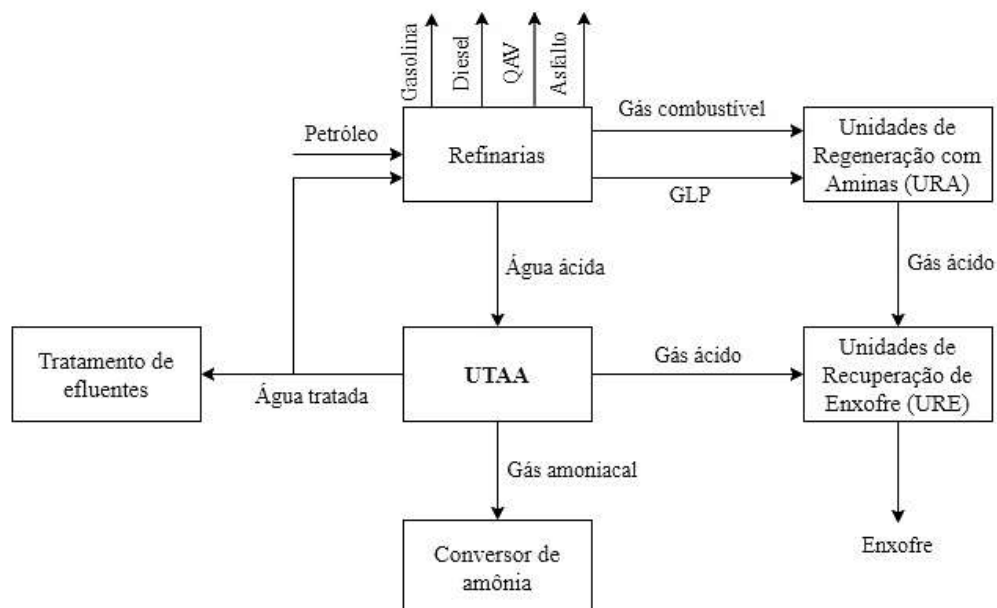
1. Introdução

1.1 Motivação

No processo do refino do petróleo, correntes de vapor/água são utilizadas em diversas unidades como, por exemplo, hidrotratamento, coqueamento retardado e craqueamento catalítico em leito fluidizado. O efluente dessas unidades é chamado de água ácida, uma vez que está contaminado com diversos eletrólitos fracos como fenol, amônia (NH_3), sulfeto de hidrogênio (H_2S) e possíveis traços de dióxido de carbono (CO_2) (HO, CHII-DONG et al., 2021).

Após a etapa do refino, a água ácida é encaminhada para a Unidade de Tratamento de Águas Ácidas (UTAA) com objetivo de reduzir os níveis de contaminantes, concentrando-se principalmente na remoção do NH_3 e do H_2S . Para isso, a água ácida passa por um processo de esgotamento (*stripping*) sendo submetida a um sistema de aquecimento e retificação onde é fornecido o calor necessário para a redução das pressões parciais dos gases e separação dos contaminantes (SOARES et al., 2021). O esquema geral da produção e tratamento das águas ácidas pode ser observado na Figura 1.1.

Figura 1.1: Esquema de contextualização da geração e tratamento de águas ácidas.



Fonte: Adaptado de BARROS (2016).

Esse processo geralmente é realizado com duas colunas quando a concentração de amônia presente na água ácida é alta. Essa configuração é muito comum nas refinarias do Brasil, já que o óleo brasileiro contém uma maior concentração de nitrogênio quando comparado com a média dos óleos mundiais (SOARES et al., 2021).

Nesse modelo, a primeira coluna produz uma corrente de topo rica em H₂S chamada de gás ácido e que é destinada para a Unidade de Recuperação de Enxofre (URE). A segunda coluna é responsável por separar a corrente de fundo da primeira em duas: uma corrente de topo rica em NH₃, chamado de gás amoniacal, que é enviada para o conversor de amônia e uma corrente de fundo denominada de água tratada.

De acordo com a legislação ambiental brasileira, pelo menos 90% da carga de entrada de H₂S presente na água ácida deve ser removida e enviada para a URE. Esse é um ponto crítico do processo, já que se resíduos de H₂S forem enviados para a segunda coluna, eles serão eliminados junto com o gás amoniacal (NH₃), produzindo SO_x que é ambientalmente indesejado, pois contribui para poluição atmosférica e é precursor da chuva ácida (BRASIL, 2007; QUINLAN and HATI, 2010; CONAMA, 2011).

Porém, é importante ressaltar que a alta eficiência na remoção de H₂S pode ter consequências negativas na operação da URE. Isso ocorre porque a alta remoção do H₂S pode levar ao aumento da concentração de NH₃ no gás ácido, causando problemas operacionais na unidade. Esses problemas podem resultar em perda de eficiência devido ao entupimento de linhas e à formação de NO_x, um composto altamente prejudicial ao meio ambiente e à saúde pública, pois é precursor tanto da chuva ácida como do efeito *smog* fotoquímico (EAP, 2021; COLBECK; MACKENZIE, 1994).

Portanto, manter uma alta recuperação de H₂S na primeira torre ao mesmo tempo em que se reduz o teor de NH₃ no gás ácido representa objetivos conflitantes e caracteriza uma operação com uma faixa estreita de tolerância para unidades com essa configuração. Além disso, pequenas perturbações que possam causar variações na concentração das correntes de entrada da UTAA podem resultar em falhas que, em última instância, levam à emissão de compostos prejudiciais ao meio ambiente acima dos limites regulatórios.

Atualmente, um dos grandes desafios operacionais é monitorar de forma *online* a composição dos efluentes dessas unidades. Por isso, o desenvolvimento de sensores virtuais para uma UTAA é uma solução vantajosa para problemas de investimento e precisão com

analisadores em linha ou questões de atraso de medição devido a frequentes análises laboratoriais complexas.

Esses sensores virtuais podem ser empregados para realizar monitoramento em tempo real das variáveis de interesse no processo, possibilitando a identificação de tendências e anomalias e acionando alertas e notificações quando problemas potenciais são detectados. Por exemplo, manutenções preventivas podem ser programadas após uma análise de dados robusta. Além disso, oferecem uma economia substancial em comparação com os sensores tradicionais, uma vez que dispensam instalação e manutenção física e não são afetados pela obsolescência devido à corrosão dos ambientes industriais (BARROS, 2016).

Por outro lado, os sensores virtuais também podem assumir a função dos sensores físicos em situações de falha inesperada, caso estes já existam. Isso garante que os operadores tenham acesso contínuo às informações necessárias para monitorar e controlar os processos, reduzindo o risco de interrupções não planejadas por falta de informações essenciais (HO, CHII-DONG et al., 2021).

Este estudo está inserido na linha de pesquisa conduzida pelo Laboratório de Desenvolvimento de *Softwares* (LADES/COPPE/UFRJ) sobre Unidades de Tratamento de Águas Ácidas, a qual abrange uma série de projetos com temas relacionados. Dentre eles, destacam-se a dissertação de mestrado de Nogueira (2021), o trabalho de conclusão de curso de Pitanga (2022), o artigo científico de Nogueira et al. (2022) publicado na revista acadêmica *Process Safety and Environmental Protection* e os trabalhos de Nogueira et al. (2021) e Filho et al. (2023) apresentados no Congresso Brasileiro de Engenharia Química (COBEQ).

1.2 Objetivos

Este trabalho tem como objetivo principal o desenvolvimento de sensores virtuais para predição da composição dos efluentes de uma Unidade de Tratamento de Águas Ácidas com a configuração de duas torres de esgotamento.

Os objetivos específicos propostos são:

1. Modificação da simulação dinâmica de uma UTAA com duas colunas de esgotamento desenvolvida em *Aspen Plus Dynamics*® V10 por Nogueira (2021) através da inclusão de novos controladores;

2. Desenvolvimento de um novo banco de dados utilizando a simulação dinâmica de Nogueira (2021) previamente modificada;
3. Análise das variáveis estudadas a fim de obter um entendimento fenomenológico abrangente do processo;
4. Desenvolvimento dos algoritmos baseados em Florestas Aleatórias para a criação dos sensores virtuais que determinarão as frações mássicas de H_2S e NH_3 nas correntes de gás ácido, gás amoniacal e água tratada;
5. Elaboração de códigos para aplicar as técnicas de *Gradient Boosting* e Máquinas de Vetores de Suporte com o propósito de validar os resultados obtidos através do método de Florestas Aleatórias.

1.3 Estrutura do trabalho

O presente trabalho foi dividido em cinco capítulos, sendo eles: Introdução, Revisão de Literatura, Metodologia, Resultados e Discussões, e Conclusões e Considerações.

O Capítulo 1 engloba a motivação e os objetivos deste trabalho. No Capítulo 2, são expostos os conceitos teóricos que embasaram o estudo. No Capítulo 3, é detalhada a metodologia para o desenvolvimento do banco de dados e dos sensores virtuais, bem como a aplicação do método de Aprendizado de Máquina. No Capítulo 4, apresentam-se os resultados e sua subsequente discussão. Por fim, o Capítulo 5 trata das conclusões e considerações finais, incluindo sugestões para trabalhos futuros.

Capítulo 2

2. Revisão de Literatura

2.1 Unidades de Tratamento de Águas Ácidas

A água ácida é um efluente líquido originado da indústria petroquímica. Durante o processamento do petróleo, o vapor ou a água entram em contato com correntes de hidrocarbonetos que contêm compostos como o sulfeto de hidrogênio (H_2S) e amônia (NH_3) que são absorvidos por transferência de massa, gerando a água ácida. Além dos compostos mencionados, a água ácida pode conter fenóis, cianeto, gás carbônico (CO_2) e até sais e ácidos (KENSELL; QUINLAN, 1996).

A geração de água ácida em uma refinaria é influenciada por diversos fatores, como a configuração do processo e as características do petróleo processado. Em refinarias de grande porte, é comum observar uma produção típica de água ácida entre 0,2 e 0,5 metros cúbicos por tonelada de óleo bruto processado (COELHO et al., 2006).

A origem do nome água ácida está relacionada à presença dos compostos ácidos como H_2S e CO_2 já mencionados. Porém, como na maior parte dos casos existe a presença significativa de NH_3 , o pH da água ácida pode ser maior que 7 (SHEILAN et al., 2010).

A água ácida pode ser classificada como fenólica e não fenólica. Ambas possuem como principais contaminantes NH_3 e H_2S . A água ácida não fenólica pode conter traços de CO_2 e é normalmente produzida em unidades de hidrotreatamento (HDT), podendo ser reciclada nas HDTs como água de lavagem ou água de *make-up* quando livre destes contaminantes. Já a água ácida fenólica possui como outros contaminantes os fenóis, cianetos e cloretos (WEILAND; HATCHER, 2012).

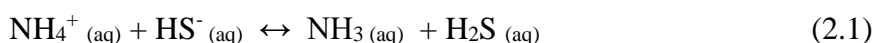
As principais unidades geradoras de água fenólica são: destilação atmosférica e a vácuo, craqueamento catalítico, coqueamento retardado e hidrotreatamento. Além disso, apesar de contribuírem com menos volume, há outras fontes geradoras de águas ácidas nas refinarias, como os vasos de lavagem de carga de unidades de recuperação de enxofre, purga e refluxo de topo de torre regeneradora de amina e torre de resfriamento de unidade de tratamento de gás residual (BARROS, 2016).

A água ácida é purificada em um processo de extração a vapor para remover os contaminantes voláteis (LEE et al., 2002). Esse processo ocorre nas Unidades de Tratamento

de Águas Ácidas (UTAA) onde haverá a recuperação do H₂S e NH₃ e a corrente já tratada, denominada água esgotada, poderá ser reutilizada ou descartada seguindo os padrões ambientais (WEILAND; HATCHER, 2012).

Os teores de H₂S e de NH₃ na água ácida podem variar de traços até valores maiores que 1% p/p. A solubilidade do H₂S em água está diretamente relacionada à presença do NH₃, uma vez que a porcentagem molar de H₂S em geral é inferior à de NH₃, exceto em águas ácidas muito diluídas, onde o teor de H₂S pode ser ligeiramente superior. Normalmente, é observado uma razão molar H₂S/NH₃ entre 0,6 e 1,0 na corrente de água ácida combinada (WEILAND; HATCHER, 2012).

Em meio aquoso, como no caso da água ácida, o sal hidrossulfeto de amônio (NH₄HS) sofre hidrólise dando origem ao H₂S e à NH₃, conforme Equação 2.1 (SHEILAN et al., 2010).



Devido à capacidade de volatilização dos gases H₂S e NH₃ presentes na fase aquosa, estes se deslocam para a fase de vapor, exercendo pressões parciais. As solubilidades dessas substâncias na fase líquida são diretamente proporcionais às suas pressões parciais na fase vapor em equilíbrio, de acordo com a Lei de Henry (SANDLER, 1999).

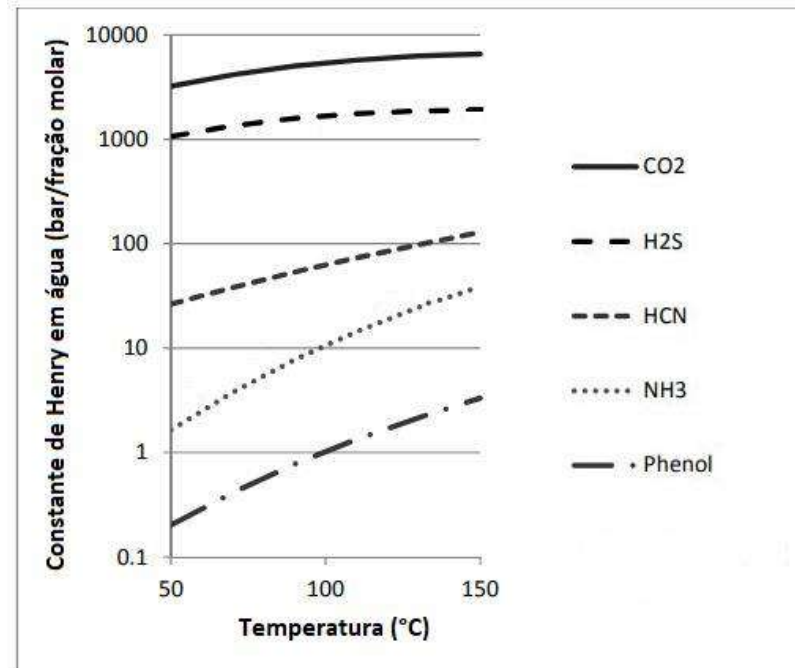
A remoção destes poluentes da fase líquida é realizada através do aumento da temperatura da água ácida favorecendo a vaporização dos contaminantes e da redução das suas respectivas pressões de vapor. Na prática, isso é feito em colunas de esgotamento, seja fornecendo calor (por meio de um refervedor) ou injetando vapor d'água diretamente (WEILAND; HATCHER, 2012).

O aumento da constante de equilíbrio da hidrólise da Equação 2.1 pela elevação da temperatura resulta em uma maior quantidade de NH₃ e H₂S dissolvidos na solução aquosa, permitindo que esses compostos migrem para a fase vapor (SHEILAN et. al., 2010).

Além disso, a elevação da temperatura também aumenta a constante de Henry das espécies de interesse, reduzindo suas solubilidades e aumentando suas volatilidades. A diminuição da pressão de vapor desloca o equilíbrio líquido-vapor na direção da transferência dos solutos da fase líquida para a fase gasosa (ADDINGTON et al., 2013).

Entretanto, esse processo afeta apenas os eletrólitos fracos e voláteis. A Figura 2.1 exhibe a volatilidade, expressa em termos da constante de Henry, dos principais compostos encontrados na água ácida.

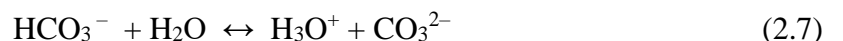
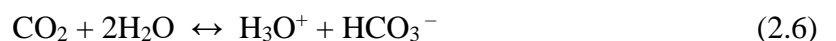
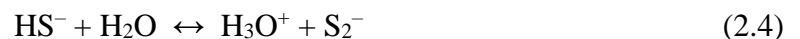
Figura 2.1: Constantes de Henry em função da temperatura para os compostos mais comuns encontrados em sistemas de águas ácidas.

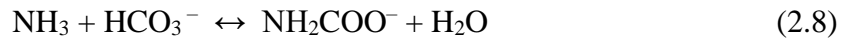


Fonte: Adaptado de ADDINGTON et al. (2013).

Através da Figura 2.1 fica evidente que o H₂S e o CO₂ são mais voláteis do que a NH₃. Essa discrepância sugere que esses compostos podem ser removidos eficientemente em pressões mais elevadas do que as necessárias para retirar a NH₃. Essa diferença de volatilidades abre a possibilidade de projetar sistemas de tratamento de águas ácidas com duas torres de esgotamento: a primeira operando a uma pressão mais alta para remoção seletiva do H₂S, por volta de 10 bar, e a segunda com uma pressão mais baixa, cerca de 2 bar, para eliminar a NH₃ e o restante do H₂S (CHEVRON, 2016).

As principais reações processadas em um sistema com água ácida não-fenólica são descritas pelas reações a seguir: (LEE et al., 2002).



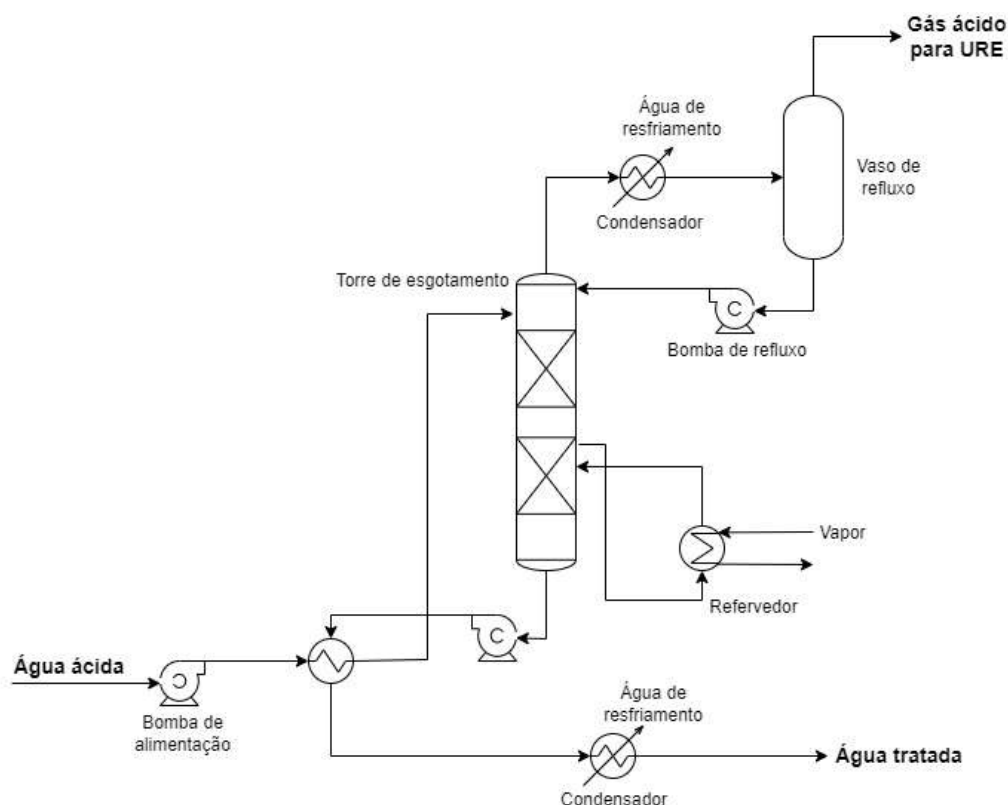


A amônia é o componente reativo do solvente e estando presente na fase gasosa com H₂S, ela é continuamente absorvida enquanto houver protonação decorrente da absorção simultânea do H₂S. Esse comportamento explica o fato de que a solubilidade dos dois componentes juntos é muitas vezes maior do que a solubilidade dos mesmos isoladamente, praticamente não existindo limite para a solubilidade conjunta do H₂S e da NH₃ na água (WEILAND; HATCHER, 2012).

A presença de outros compostos ácidos menos voláteis que o H₂S como o ácido cianídrico (HCN), ácido clorídrico (HCL) e ácidos orgânicos prejudica a remoção da amônia, pois consomem íons OH⁻ da reação de equilíbrio iônico da água, conforme Equação 2.2. Com menor concentração de íons hidroxila, pelo princípio de *Le Chatelier*, a Equação 2.5 é deslocada no sentido da formação de íons amônio (NH₄⁺). Dessa maneira, haverá menor quantidade de NH₃ em fase líquida disponível para a passagem para a fase vapor. Para mitigar este efeito, muitas unidades adotam a injeção de uma base forte, normalmente NaOH, na coluna de esgotamento (MELIN et al, 1975; DARTON e VAN GRINSVEN, 1978).

A configuração do projeto das UTAA's pode ser de uma ou duas torres em série. No caso de apenas uma torre, a corrente de topo da única coluna, neste caso, gera um efluente gasoso rico em H₂S e NH₃ chamado de gás ácido. Este gás pode ser enviado à Unidade de Recuperação de Enxofre (URE) caso essa consiga também processar amônia. Se a amônia não for totalmente convertida, existe a possibilidade da formação de sais de amônio nos pontos frios da URE resultando em grande corrosão dos equipamentos e necessidade de parada da unidade para manutenções devido às obstruções (BARROS, 2016). A Figura 2.2 representa um esquema simplificado de uma UTAA com apenas uma coluna.

Figura 2.2: Esquema de uma Unidade de Tratamento de Água Ácida com uma coluna.

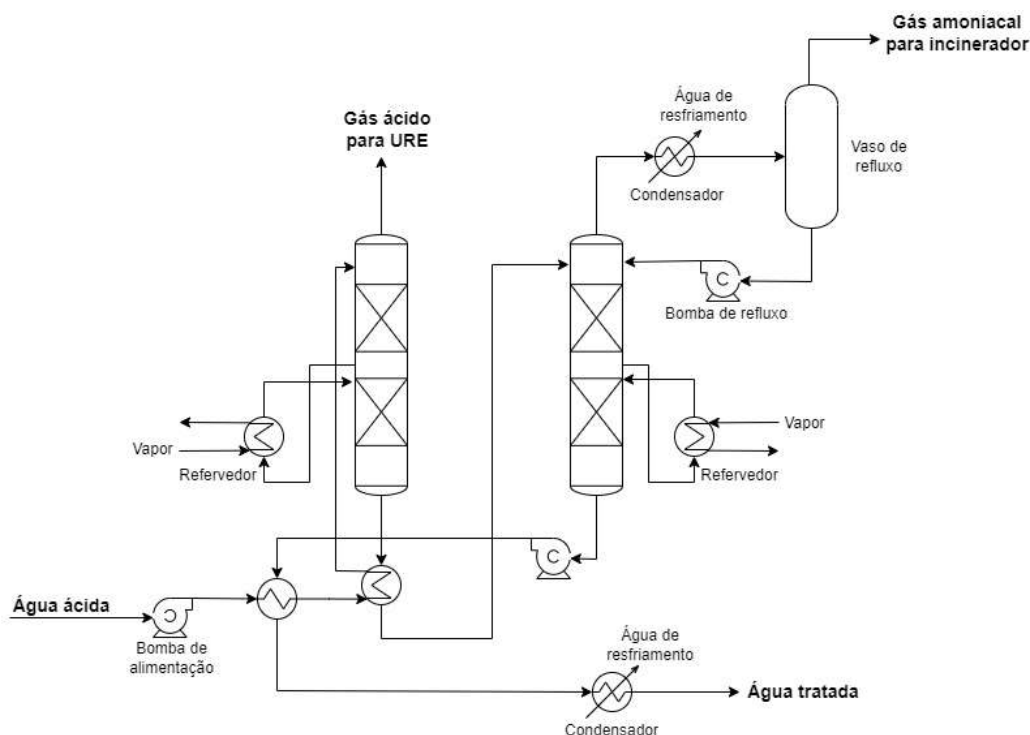


Fonte: Adaptado de QUINLAN e HATI (2010).

Entretanto, uma situação muito comum principalmente nas refinarias brasileiras é o processamento de petróleo com elevado teor de nitrogênio, o que resulta em altos teores de amônia tanto na água ácida como no gás gerado pela UTAA. Esta condição demanda a necessidade de duas torres em série para tratamento da água ácida, visando evitar problemas de operação na URE (CHEVRON, 2016).

No caso da configuração com duas colunas, o esgotamento ocorrerá em duas etapas: na primeira coluna ocorrerá a remoção preferencial do H_2S e na segunda, do NH_3 . Desta forma, a primeira torre gera como corrente de topo um gás ácido rico em H_2S e pobre em NH_3 , que será enviado para a URE, e a segunda torre gera como corrente de topo um efluente gasoso chamado de gás amoniacal rico em NH_3 (QUINLAN e HATI, 2010). A Figura 2.3 representa um esquema simplificado de uma UTAA com duas colunas em série.

Figura 2.3: Esquema de uma Unidade de Tratamento de Água Ácida com duas colunas.



Fonte: Adaptado de QUINLAN e HATI (2010).

O gás ácido produzido nas UTAA's e nas Unidades de Regeneração de Amina (URA) compõe a carga da URE onde ocorre a recuperação do enxofre através do Processo Claus Modificado, convertendo o H_2S em enxofre elementar que possui razoável valor comercial (QUINLAN; HATI, 2010).

Já o gás amoniacal é encaminhado para unidades onde ocorre a purificação de NH_3 ou conversão desta em N_2 . Sendo assim, um parâmetro importante quando se opera com duas torres é a eficiência de remoção de H_2S no topo da primeira coluna, já que se a eficiência for baixa, este componente será removido na segunda coluna, enviando altos teores deste componente para o conversor de amônia, aumentando a emissão de SO_2 . (BRITO et al, 2011; QUINLAN; HATI, 2010).

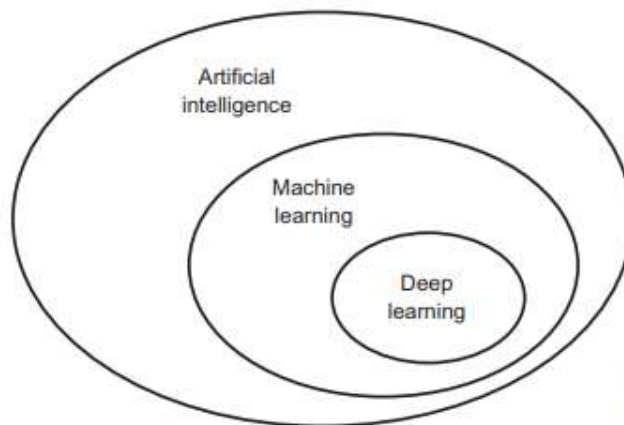
Além disso, para atender os limites de emissão de poluentes atmosféricos descritos na legislação brasileira, as UTAA's que operam com duas torres de esgotamento devem remover pelo menos 90% da carga de entrada do H_2S nas águas ácidas para ser tratado nas URE's (CONAMA, 2011). Pequenos desvios podem causar perturbações no sistema que resultará na liberação de componentes acima do permitido pelo órgão ambiental. Dessa forma, unidades

com duas colunas de esgotamento operam em uma faixa estreita com o objetivo de cumprir metas conflitantes: alta recuperação de H₂S e baixo teor de NH₃ no gás ácido.

2.2 Inteligência Artificial

A inteligência artificial (em inglês, *Artificial Intelligence, AI*) surgiu na década de 1950 quando pioneiros da ciência da computação começaram a questionar se computadores poderiam ser programados para “pensar”. Uma definição concisa para o termo AI seria o esforço para automatizar tarefas intelectualmente realizadas por humanos. Esse campo abrange também o chamado Aprendizado de Máquina (em inglês, *Machine Learning, ML*) e o Aprendizado Profundo (em inglês, *deep learning*) como mostrado na Figura 2.4 a seguir (CHOLLET, 2021).

Figura 2.4: Diagrama mostrando a relação entre inteligência artificial, *machine learning* e *deep learning*.



Fonte: CHOLLET (2021).

Algumas aplicações que podemos citar da AI são: inteligência simbólica, prova de teoremas, robótica, visão computacional, sistemas especialistas, aprendizado de máquina, entre outros (RUSSELL; NORVIG, 2002).

Um dos campos dentro da AI que mais revolucionou foi o Aprendizado de Máquina cujos conceitos estão interligados ao contexto de *Big Data* e *Data Science*. O ML tem sido muito utilizado para resolver problemas tradicionais de estatística multivariada como regressão, redução de dimensionalidade e clusterização (MONTAVON, et al. 2012).

Na área da indústria química, o avanço tecnológico possibilitou a diminuição de gastos com sensores e armazenamento de informações, além de promover uma maior disponibilidade de instrumentos e aprimorar a comunicação entre a equipe e a planta. Com o aumento do histórico de dados *online*, surge uma crescente demanda por análise de informações na área (SOARES, 2017).

ML, em sua definição mais usual, é uma área de pesquisa que possibilita que os computadores aprendam sem serem programados explicitamente (SAMUEL, 1959). Em processos químicos, o Aprendizado de Máquina se classifica como uma abordagem baseada em dados, contrastando com abordagens mais tradicionais que são baseadas em modelos (BAUGHMAN e LIU, 1995).

Dessa maneira, são fornecidos uma grande quantidade de dados para que o algoritmo consiga aprender as relações entre eles e assim executar determinada tarefa. Por isso, é tão importante a qualidade e quantidade dos dados disponíveis, garantindo uma boa variabilidade de exemplos (SOARES, 2017).

Há três tipos de problemas típicos em ML: o aprendizado não supervisionado, aprendizado supervisionado e o aprendizado por reforço. O aprendizado não supervisionado fornece apenas os valores de entrada sem depender de alvos específicos com o propósito de visualização de dados, remoção de ruídos ou para compreender melhor as correlações presentes nos dados disponíveis. Além disso, frequentemente representa um passo necessário para uma compreensão mais profunda de um conjunto de dados antes de tentar resolver problemas de aprendizado supervisionado (ALOM et al., 2019).

O aprendizado supervisionado é o caso mais comum e consiste em fornecer dados de entrada para que alvos (*targets*) sejam alcançados, dado um conjunto de exemplos fornecidos pelo usuário. Por fim, o aprendizado por reforço é uma subárea do aprendizado de máquina que ganhou destaque após a *Google DeepMind* obter sucesso ao aplicá-la no aprendizado de jogos Atari. O algoritmo recebe informações sobre o seu ambiente e aprende a escolher ações que irão maximizar sua recompensa, por exemplo, é possível treinar uma rede neural que analisa a tela de um *videogame* e produz ações de jogo com o objetivo de maximizar sua pontuação (ALOM et al., 2019).

O método de aprendizado supervisionado em ML implica na divisão do conjunto de dados em três grupos distintos: treinamento, validação e teste. O conjunto de treinamento é utilizado para ensinar o modelo, permitindo-lhe aprender padrões e relações nos dados. O

conjunto de validação auxilia na escolha dos melhores hiperparâmetros do modelo, ajudando a evitar o sobreajuste (*overfitting*), enquanto o conjunto de teste é reservado para avaliar o desempenho final do modelo em dados completamente novos e não vistos. Essa divisão cuidadosa e estratégica dos dados é fundamental para garantir que os modelos sejam robustos, generalizem bem e sejam confiáveis em situações do mundo real, fornecendo previsões precisas e úteis (CHOLLET, 2021).

O sobreajuste é o resultado de um modelo que se adapta excessivamente a um conjunto de dados de treinamento específico, perdendo sua capacidade de generalização. Isso pode levar o modelo a apresentar um desempenho excelente nos dados de treinamento, mas um desempenho inferior nos dados de teste ou na aplicação em situações do mundo real. Por outro lado, o subajuste (*underfitting*) também pode ocorrer se o modelo não for treinado o suficiente e for necessário aumentar a quantidade de dados de treinamento para melhorar sua performance (CHOLLET, 2021).

2.2.1 Florestas Aleatórias (*Random Forests*, RF)

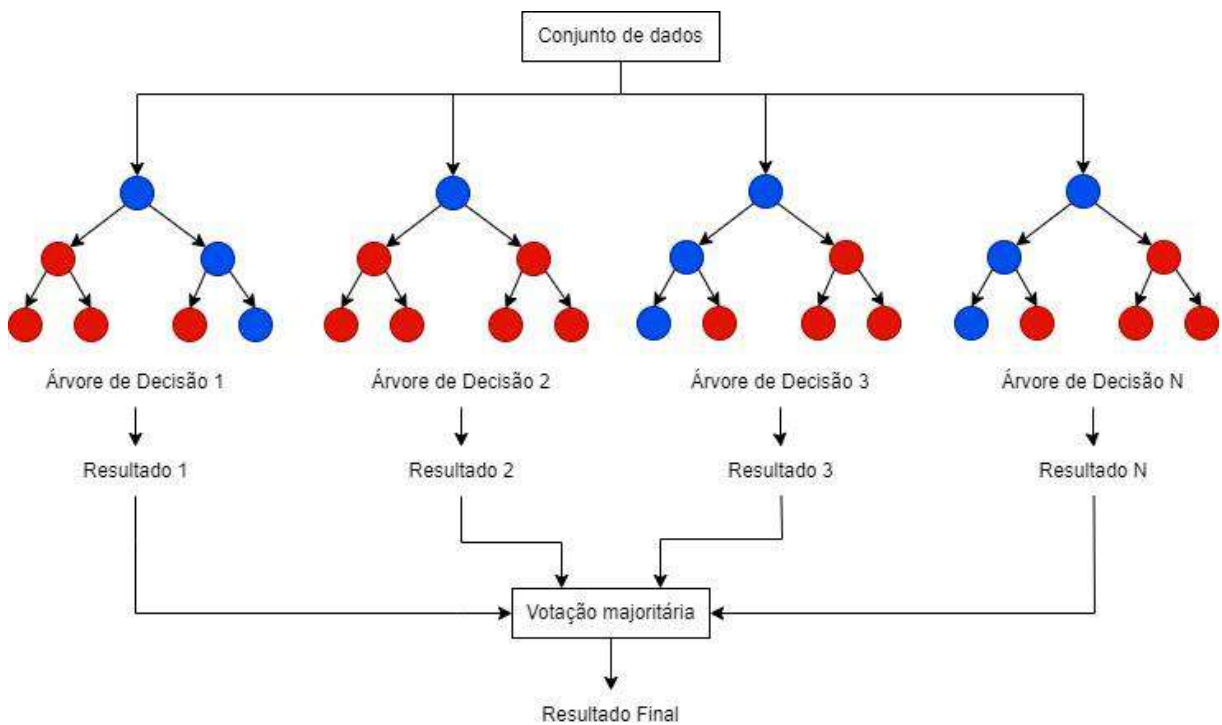
Florestas Aleatórias é uma técnica de Aprendizado de Máquina que foi proposta por Breiman (2001) há 22 anos e é composto por um grupo de estimadores, chamados de árvores de decisão (SPEISER et al., 2019).

As árvores de decisão são estruturas semelhantes a fluxogramas que permitem classificar ou prever valores de saída dados os valores de entrada. Elas são fáceis de visualizar e interpretar. Em particular, o algoritmo Florestas Aleatórias introduziu uma abordagem robusta e prática para o aprendizado de árvores de decisão, que consiste em construir várias árvores de decisão especializadas e combinar suas saídas para produzir um resultado final (CHOLLET, 2021).

As Árvores de Classificação e Regressão, também conhecidas como CART, constituem algoritmos que se baseiam nas características presentes no conjunto de dados para segmentar as amostras. As árvores são construídas de forma iterativa e crescem por meio de divisões binárias sucessivas, criando nós e terminando em folhas quando não há mais divisões. Existe uma hierarquia nessas divisões, que pode ser medida por meio de entropia ou impureza. Essas árvores podem ser profundas e complexas ou rasas e simples, dependendo da necessidade do problema (BREIMAN et al., 1984).

Além disso, a técnica de Florestas Aleatórias utiliza diferentes amostras de inicialização para construir cada árvore, ou seja, diferentes subconjuntos do conjunto de dados. Como exemplo, a Figura 2.5 apresenta um modelo de regressão com RF composto por quatro árvores e um conjunto de dados que inclui múltiplas variáveis dependentes onde cada árvore vai sofrendo divisões binárias até alcançar o nó final. O resultado final é produzido pela média das previsões das árvores individuais.

Figura 2.5: Esquema geral de um algoritmo de Florestas Aleatórias.



Fonte: Elaboração própria.

O método CART é empregado em situações de classificação e também em problemas de regressão, possibilitando previsões tanto quantitativas quanto qualitativas. No caso da classificação, a árvore divide os dados em categorias, isto é, produz a saída que representa a classe escolhida pela maioria das árvores, enquanto na regressão, ela estabelece relações entre variáveis de entrada e saída para prever valores contínuos, sendo que a saída é a média ou a previsão probabilística média das árvores (BREIMAN et al., 1984; CHOLLET, 2021).

No entanto, de acordo com a documentação da biblioteca *scikit-learn*, é implementada uma modificação para que cada árvore produza uma previsão baseada em probabilidade em vez de um único voto. O valor previsto é determinado calculando a média das previsões probabilísticas de todas as árvores. Essa adaptação tem o objetivo de reduzir a variância e

minimizar o risco de sobreajuste do estimador, uma vez que alguns erros podem se anular mutuamente (NOGUEIRA, 2021).

O pacote `randomForest` produz uma informação adicional que pode ser muito útil para avaliar fenomenologicamente um processo: a medida da Importância de Variáveis (*Variable Importance*, VI) preditoras. A importância de uma variável pode ser devido à sua complexa interação com outras variáveis. O algoritmo estima a importância de uma variável observando o quanto o erro de previsão aumenta quando os dados para essa variável são permutados, enquanto todas as outras variáveis são deixadas inalteradas. Os cálculos necessários são realizados árvore por árvore à medida que a floresta aleatória é construída (LIAW e WIENER, 2002).

Essa funcionalidade permite a redução do conjunto de dados de interesse através da remoção de variáveis menos importantes, garantindo a simplificação e otimização do modelo, além de ganho em tempo de processamento computacional (BELGIU e DRAGUT, 2016; SPEISER et al., 2019).

É relevante destacar que, em contextos industriais, a eficiência computacional e o tamanho dos modelos desempenham um papel fundamental. Isso ocorre devido à necessidade de disponibilizar resultados da planta em tempo real, tornando essencial a otimização do tempo de processamento e a criação de modelos menores.

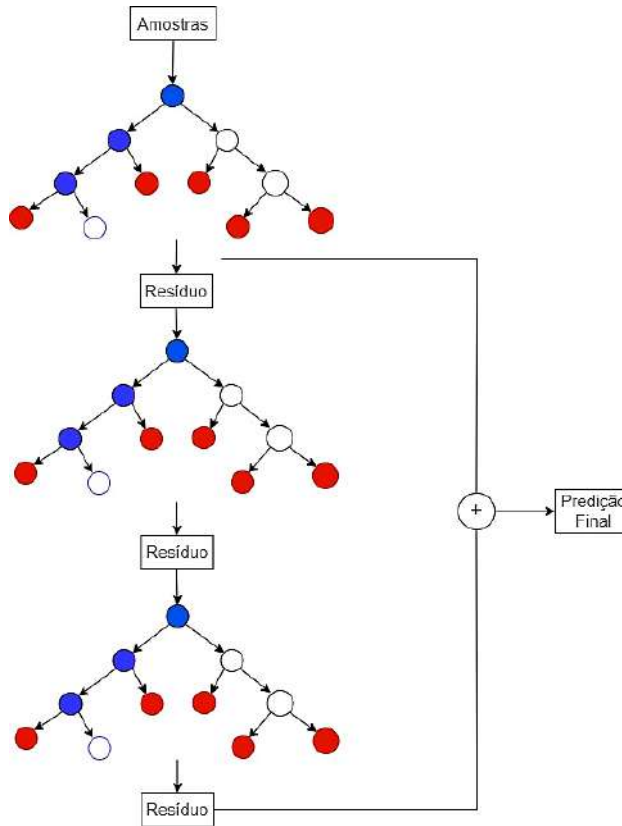
2.2.2 Gradient Boosting (GB)

O *Gradiente Boosting* é uma técnica de Aprendizado de Máquina que se baseia em combinar modelos de previsão fracos, em geral, árvores de decisão, com o objetivo de melhorar o desempenho do modelo de maneira iterativa e aditiva (CHOLLET, 2021).

Essa abordagem utiliza o conceito de “*boosting*” que consiste em treinar repetidamente novos modelos, cada um focado em corrigir as deficiências identificadas nos modelos anteriores. Isso é feito calculando os resíduos, que representam as diferenças entre as previsões do modelo atual e os valores reais. Uma nova árvore é então treinada para minimizar esses resíduos. Esse processo iterativo continua até que uma tolerância de resíduos seja atingida ou outro critério de parada seja satisfeito, e o algoritmo assume a solução encontrada como adequada. Isso ajuda a melhorar o desempenho do modelo, concentrando-se

nas áreas onde ele tem dificuldade em fazer previsões precisas, conforme mostrado na Figura 2.6 (CHOLLET, 2021; ZHOU, 2012).

Figura 2.6: Esquema geral de um algoritmo do *Gradient Boosting*.



Fonte: Elaboração própria.

Quando aplicado às árvores de decisão, o *Gradiente Boosting* frequentemente supera o método de Florestas Aleatórias, oferecendo modelos mais precisos e robustos. Essa técnica se destaca no tratamento de dados não perceptuais e é amplamente reconhecida como uma das melhores abordagens disponíveis (CHOLLET, 2021).

A popularidade do *Gradiente Boosting* não está limitada apenas ao seu desempenho superior, mas também à sua capacidade de lidar com problemas complexos de aprendizado de máquina. É amplamente adotado em aplicações que envolvem desde previsões financeiras até processamento de linguagem natural e visão computacional, destacando-se como uma ferramenta essencial para cientistas de dados e engenheiros de aprendizado de máquina (CHOLLET, 2021).

2.2.3 Máquinas de Vetores de Suporte (*Support Vector Machines, SVM*)

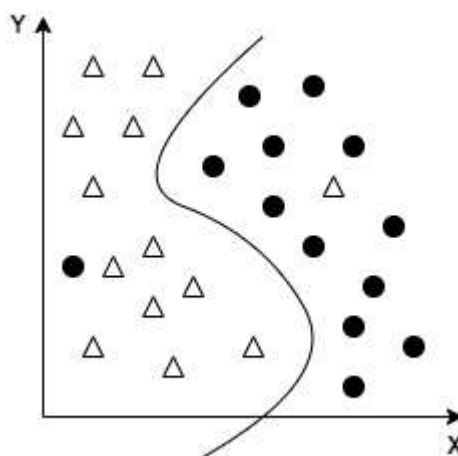
As Máquinas de Vetores de Suporte são algoritmos baseados na teoria do aprendizado estatístico desenvolvido por Vladimir Vapnik e Corinna Cortes em 1995, embora uma formulação mais antiga já havia sido publicada por Vapnik e Alexey Chervonenkis em 1963 (CHOLLET, 2021; TRAFALIS e INCE, 2002).

Esse método de aprendizado supervisionado vem recebendo grande atenção da comunidade de ML por apresentar resultados frequentemente melhores do que os obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (LORENA e CARVALHO, 2007).

Apesar das SVMs serem frequentemente empregadas como ferramentas de classificação, também podem ser facilmente adaptadas aos problemas de regressão (FERRÃO et al, 2007).

Para a resolução dos problemas de classificação, este algoritmo busca encontrar fronteiras de decisão eficazes entre duas categorias diferentes. Essas fronteiras podem ser linhas ou superfícies que separam os dados de treinamento em espaços correspondentes às duas categorias (CHOLLET, 2021). Para classificar novos dados, basta verificar de que lado da fronteira de decisão eles estão, como pode ser observado na Figura 2.7.

Figura 2.7: Esquema de SVM Linear para um exemplo de classificação binária.



Fonte: Adaptado de LORENA e CARVALHO (2007).

De maneira geral, o processo para classificação das SVMs ocorre em duas etapas: primeiro, os dados são mapeados para uma representação de alta dimensão, onde a fronteira

de decisão é expressa como um hiperplano. Em seguida, uma fronteira de decisão eficaz é calculada, maximizando a distância entre o hiperplano e os pontos de dados mais próximos de cada categoria, o que é conhecido como a maximização da margem. Isso permite que a fronteira seja aplicada com sucesso a novas amostras fora do conjunto de treinamento (CHOLLET, 2021).

Já para casos de regressão, o hiperplano é uma superfície (ou linha em problemas unidimensionais) ajustada aos dados de treinamento com o objetivo de minimizar o erro de previsão numérica. A ideia é encontrar um hiperplano que se ajuste o mais próximo possível dos pontos de dados, minimizando o erro de previsão. Isso é geralmente feito minimizando a soma dos quadrados dos erros (método dos mínimos quadrados) ou por meio de outras funções de perda adequadas. De forma geral, o hiperplano representa uma função de regressão que estima os valores alvo com base nas entradas do modelo (TRAFALIS e INCE, 2002).

As SVMs oferecem diferentes tipos de *kernels* que determinam a maneira como elas reconhecem os padrões nos dados de entrada. Entre os mais comuns estão o Linear, que funciona com uma separação linear; o Polinomial, que utiliza uma função polinomial cujo grau pode ser ajustado durante o treinamento; o Gaussiano, que opera em um espaço de características de dimensão infinita e permite que a SVM exiba características semelhantes a uma rede RBF (*Radial Basis Function*) e o Sigmoidal, que possibilita que a SVM se comporte de maneira semelhante a uma rede MLP (*Multilayer Perceptron*). A escolha adequada do *kernel* e de seus parâmetros correspondentes exerce uma influência significativa nos resultados alcançados por uma SVM (FERRÃO et al, 2007).

2.3 Sensores virtuais

Os sensores virtuais (*soft sensors*, em inglês) são modelos matemáticos ou algoritmos que permitem estimar o valor de uma variável de interesse com base em informações disponíveis do processo, sem a necessidade de efetuar medições diretas dessa variável (LOTUFO e GARCIA, 2008).

Essa abordagem é vantajosa principalmente em situações na qual a variável de interesse está relacionada à análise da composição química de uma corrente, podendo gerar um atraso significativo na obtenção de resultados (JOUROWSKI, 2011).

De acordo com Assis e Filho (2000), um sensor virtual é a combinação de um *hardware* que possibilita a medição *online* de certas variáveis do processo e um algoritmo de estimação (*software*) cujo objetivo é fornecer estimativas em tempo real de variáveis não medidas, parâmetros do modelo ou compensar atrasos de medição.

Dito isso, existem três abordagens principais utilizadas para estimar variáveis baseando-se em outras medidas. São elas (FORTUNA et al., 2007; LIN et al., 2007):

1. *Model-driven* (ou caixa branca) na qual um modelo detalhado que descreve o processo em questão é desenvolvido, requerendo um conhecimento sólido dos princípios físicos e matemáticos que governam o sistema.
2. *Data-driven* (ou caixa preta) que se baseia unicamente em dados históricos do processo. Neste caso, os modelos podem ser criados usando técnicas de Aprendizado de Máquina como regressão, redes neurais, árvores de decisão, entre outras.
3. *Gray-box* (ou caixa cinza) que é uma combinação dos dois métodos anteriores. Ele aproveita informações disponíveis sobre o processo para aprimorar e calibrar um modelo *data-driven*, criando uma abordagem híbrida.

Os sensores virtuais têm a capacidade de substituir sensores tradicionais em diversos campos da instrumentação. No entanto, existem áreas específicas em que os sensores tradicionais apresentam limitações significativas, tornando o uso dos sensores virtuais cada vez mais preferível (FORTUNA et al., 2007).

Uma dessas áreas críticas é a dos processos químicos complexos como polimerização, colunas de destilação e bioprocessos. Nesses cenários, os sensores tradicionais enfrentam desafios relacionados a custos, precisão e atrasos operacionais devido a análises laboratoriais mais complexas, o que pode afetar negativamente no controle do processo. Além disso, muitas vezes não há instrumentação disponível para medição de certas variáveis, tornando o sensor virtual a única opção viável (LOTUFO e GARCIA, 2008).

Dessa forma, os sensores virtuais desempenham um papel fundamental na melhoria da confiabilidade e disponibilidade de parâmetros de um sistema. Na ausência dos sensores físicos, eles têm a capacidade de prever e estimar os valores ausentes, garantindo um monitoramento contínuo do processo. Por fim, são uma alternativa econômica em comparação com os sensores físicos, o que permite a expansão do número de pontos de monitoramento disponíveis (SOBREIRA, 2021).

Nas bases de literatura pesquisadas foram encontrados poucos trabalhos que abordam a problemática de utilização de sensores virtuais em Unidades de Tratamento de Águas Ácidas. As principais referências encontradas para este tema específico são as publicações de Barros (2016) e Graziani e Xibilia (2017).

Barros (2016) buscou desenvolver um sensor virtual para predição da eficiência de remoção de H_2S e avaliar a influência de variáveis do processo como vazão e temperatura de alimentação, carga térmica e diferencial de temperatura do refulvedor e diferencial de pressão da coluna. Os dados coletados são referentes ao período de dois anos de operação de uma UTAA constituída por duas torres de esgotamento que processa água ácida não fenólica oriunda de HDT. Para o desenvolvimento do sensor, foram empregadas as abordagens fenomenológica e empírica. A abordagem fenomenológica desenvolvida em *AspenOne*® não apresentou confiabilidade nos resultados obtidos, pois notou-se que os valores estimados pela simulação de processo eram sistematicamente menores que os valores medidos experimentalmente. Este problema na representação fenomenológica pode ser atribuído à incerteza associada na especificação da eficiência global de esgotamento da coluna de águas ácidas em relação à condição de equilíbrio líquido-vapor. Por outro lado, a análise estatística da abordagem empírica, utilizando a técnica de rede neural, se mostrou adequada para a aplicação desejada, pois o desvio relativo médio entre os valores experimentais e calculados foi de apenas 1,9% e a distribuição de resíduos se aproximou de uma distribuição gaussiana.

Graziani e Xibilia (2017) utilizaram uma abordagem de aprendizado semi-supervisionado baseada em redes neurais profundas com objetivo de desenvolver dois sensores virtuais para estimar as concentrações de H_2S e NH_3 na água residual de uma UTAA com apenas uma coluna de esgotamento. O trabalho aborda a dificuldade de rotulagem dos dados, uma vez que, em certos cenários industriais, as variáveis de entrada são medidas constantemente, enquanto as variáveis de saída são quantificadas apenas uma ou duas vezes no dia. Os resultados dos sensores baseados em redes neurais profundas foram comparados com modelos de redes MLP tradicionais. Os sensores desenvolvidos com redes neurais profundas apresentaram melhor desempenho e demonstraram ser uma estratégia eficaz para aplicações industriais onde a escassez de dados rotulados pode afetar significativamente o desempenho dos sensores virtuais orientados por dados.

Capítulo 3

3. Metodologia

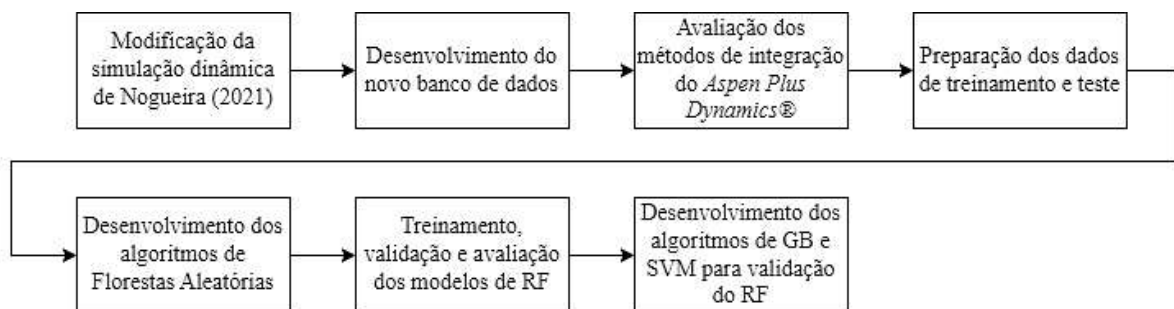
Este estudo teve como ponto de partida a simulação dinâmica de uma Unidade de Tratamento de Água Ácida (UTAA) com duas colunas de esgotamento, originalmente desenvolvida no *software Aspen Plus Dynamics® V10* por Nogueira (2021). A fim de coletar dados das frações mássicas de H_2S e NH_3 nas correntes de gás ácido, gás amoniacal e água tratada, essa simulação foi modificada, incorporando seis novos controladores.

Para criação do novo banco de dados, foram realizadas várias corridas com distúrbios em variáveis de interesse. Em seguida, os métodos de integração disponíveis no *Aspen Plus Dynamics® V10* foram avaliados usando algumas dessas corridas.

Posteriormente, os novos dados gerados foram separados em bancos de treinamento e teste para o desenvolvimento dos algoritmos de Florestas Aleatórias, visando a concepção de sensores virtuais.

Além disso, os resultados obtidos com a técnica de RF foram validados por outros métodos como o *Gradiente Boosting* e Máquinas de Vetores de Suporte. O fluxo geral dessa metodologia pode ser visualizado na Figura 3.1.

Figura 3.1: Fluxograma geral da metodologia.



Fonte: Elaboração própria.

3.1 Hardware e Software

As simulações, cálculos e códigos foram desenvolvidos e executados em um computador com processador Intel(R) Core(TM) i5 - 7200U CPU@ 2.50 GHz 2.70 GHz,

memória instalada (RAM) de 8,00 GB e sistema operacional Windows 10 64 bits. A simulação dinâmica da UTAA foi realizada com o *Aspen Plus Dynamics*® V10.

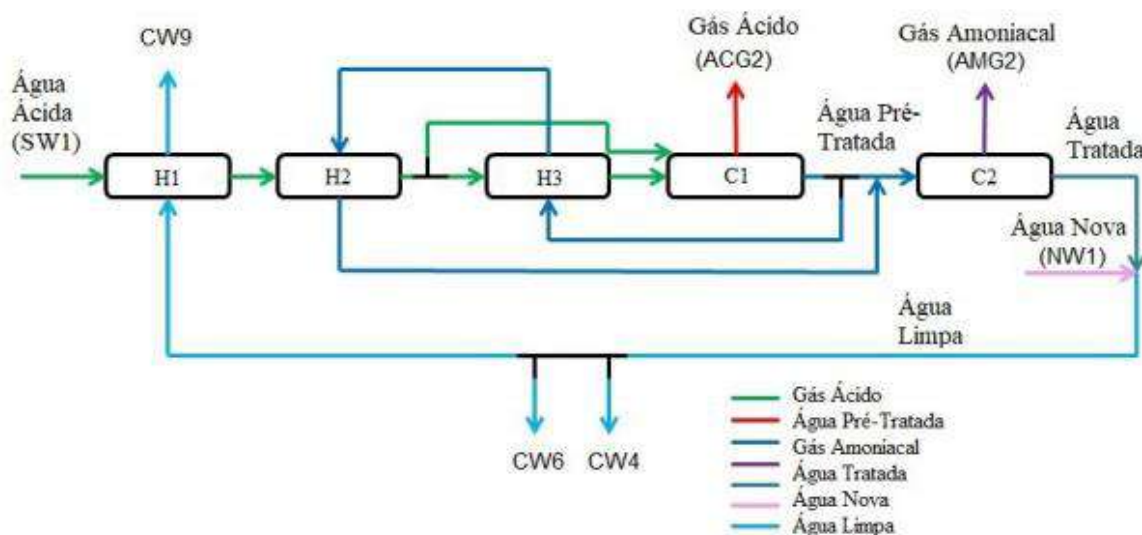
Todo o processamento de dados e algoritmos de Inteligência Artificial foram realizados em *Python* 3.11.4 no ambiente integrado de execução *Visual Studio Code* 1.79.2. De forma geral, as bibliotecas *numpy* (1.25.2), *matplotlib* (2.0.2), *joblib* (1.3.2), *scikit-learn* (0.19.0), *pandas* (2.1.0), *seaborn* (0.12.2) e *optuna* (3.3.0) foram amplamente utilizadas.

3.2 Histórico do Processo Simulado

Nogueira (2021) realizou a simulação dinâmica de uma UTAA em *Aspen Plus Dynamics*® V10 usando o modelo GPSWAT (*Gas Process Association Sour Water Equilibrium*), considerado o mais adequado para aplicações em água ácida.

Nesta simulação, há trinta e quatro correntes materiais e sete diferentes composições, três correntes de calor e vinte e três equipamentos: doze válvulas, três trocadores de calor, três divisores, dois misturadores, uma bomba e duas colunas. O diagrama de blocos a seguir, representado pela Figura 3.2, resume todas as correntes e os principais equipamentos da simulação. Sendo H1, H2 e H3 os trocadores de calor e C1 e C2, as colunas de esgotamento.

Figura 3.2: Diagrama de blocos da simulação dinâmica de Nogueira (2021).



Fonte: NOGUEIRA (2021).

De maneira geral, as sete diferentes composições descritas no processo são: água ácida é a corrente de entrada que será tratada rica em contaminantes, o gás ácido é a corrente de topo da primeira coluna rica em H_2S , o gás amoniacal é a corrente de topo da segunda coluna

rico em NH_3 , a água pré-tratada é a corrente de fundo da primeira coluna, portanto rica em NH_3 , a água tratada é a corrente de fundo da segunda coluna, possuindo pequenas quantidades de contaminantes H_2S e NH_3 solubilizados, a água nova é uma corrente aquosa sem contaminantes e, por fim, a água limpa é a mistura da água nova com a água tratada.

A Tabela 3.1 correlaciona as siglas dessas correntes com sua composição.

Tabela 3.1: Legenda das correntes da simulação dinâmica de Nogueira (2021).

Sigla	Composição da corrente em inglês	Composição da corrente em português
SW	<i>Sour Water</i>	Água Ácida
PW	<i>Pretreated Water</i>	Água Pré-Tratada
TW	<i>Treated Water</i>	Água Tratada
NW	<i>New Water</i>	Água Nova
CW	<i>Clean Water</i>	Água Limpa
ACG	<i>Acid Gas</i>	Gás Ácido
AMG	<i>Amoniacal Gas</i>	Gás Amoniacal

Fonte: NOGUEIRA (2021).

As duas correntes de entrada do processo são SW1 e NW1 enquanto as cinco correntes de saída são ACG2, AMG2, CW4, CW6 e CW9. SW1 é a água ácida a ser tratada, e NW1 é a nova fonte de água sem contaminantes que dilui a corrente de água tratada após passar pelas duas colunas. A Tabela 3.2 mostra as propriedades físicas das correntes de entrada.

Essas correntes tiveram sua vazão mássica descrita como função de SW1, sendo definida como SW1F. Esses valores foram anonimizados devido à confidencialidade de informações que originaram as simulações desenvolvidas por Nogueira (2021).

Por fim, ACG2 e AMG2 são as saídas de gás ácido e gás amoniacal, respectivamente. CW4, CW6 e CW9 são as saídas de água limpa de mesma composição.

Tabela 3.2: Propriedades físicas das correntes de entrada do sistema.

Propriedade	Unidade	SW1	NW1
Temperatura	°C	38,8	122,4
Pressão	bar	10,33	2,29
Vazão mássica	t/h	SW1F	38,5% de SW1F
Fração mássica de H ₂ O	-	0,9928	1,0000
Fração mássica de NH ₃	-	0,0032	0,0000
Fração mássica de H ₂ S	-	0,0040	0,0000

Fonte: NOGUEIRA (2021).

A simulação se inicia com a entrada da corrente SW1 que é pré-aquecida nos trocadores H1 e H2 através de integração energética com as correntes CW7 e PW5, respectivamente, tornando-se SW4. SW4 é dividida em SW5 e SW7. SW5 torna-se SW6 após passar pela válvula V2 e entra no topo da coluna C1. SW7 é pré-aquecida mais uma vez no trocador de calor H3 por integração energética com PW4 e torna-se SW9 depois de passar pela válvula V3, posteriormente entrando na alimentação inferior da coluna C1.

A coluna C1 é aquecida pela carga térmica Q1 do refeedor, separando a corrente de topo ACG1 (gás ácido) da corrente de fundo PW1 (água pré-tratada). ACG1, após a válvula V4, torna-se AGC2 e é retirado do sistema, enquanto PW1 é dividida em duas correntes: PW2 e PW4. PW4 é resfriada no trocador de calor H3 por integração energética com SW7 e torna-se PW5. PW5 é resfriada em H2 por integração energética com SW3 e torna-se PW6. PW2 torna-se PW3 após passar pela válvula V5 e é misturada com PW6, formando PW7.

Nota-se que PW7 tem vazão mássica igual à vazão mássica de PW1, pois as correntes que se dividiram se encontram novamente sem perdas. Após a válvula V6, PW7 torna-se PW8 e entra na alimentação da coluna C2.

A coluna C2 possui um condensador e um refeedor, cujas cargas térmicas são Q2 e Q3, respectivamente. A corrente de topo da coluna 2 é a AMG1 (gás amoniacal) e a corrente de fundo é a TW1 (água tratada). AMG1, após a válvula V7, torna-se AMG2 e é retirado do sistema. TW1, após a válvula V8, torna-se TW2. TW2 é misturado com NW2 tornando-se água limpa (CW1). Após passar pela bomba P1, CW1 passa a se chamar CW2. CW2 é dividida em três correntes: CW3, CW5 e CW7. CW3, após a válvula V10, torna-se CW4 e é retirada do sistema. CW5, após passar pela válvula V11, torna-se CW6 e é retirada do

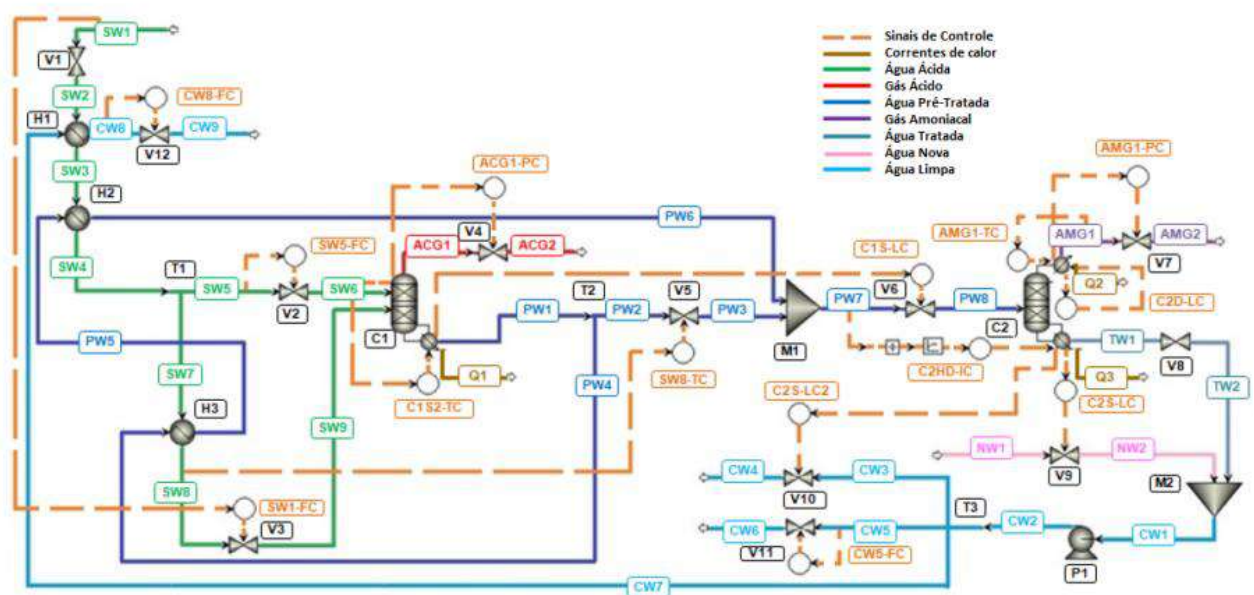
sistema. Enquanto CW7 é resfriada no trocador de calor H1 por integração energética com SW2 e torna-se CW8. CW8, após a válvula V12, torna-se CW9 e é retirada do sistema.

Existem duas colunas do tipo "RadFrac" e três trocadores de calor do tipo "HeatX" na simulação. A coluna C1 é composta por 5 estágios e a coluna C2 por 6 estágios. As entradas ocorrem nos estágios 1 (SW6) e 2 (SW9) em C1, e no estágio 2 (PW8) em C2. Os dois misturadores foram configurados para receber tanto a fase líquida quanto a vapor. As válvulas V4 e V7 foram configuradas apenas para receber vapor, enquanto as outras válvulas aceitam apenas a fase líquida. Todas as válvulas possuem uma perda de carga associada.

A maioria das válvulas é inicialmente ajustada em torno de 50%, com uma faixa de abertura que vai de 0% a 100%. No entanto, existem apenas duas exceções: V10 e V5. A primeira só é aberta quando necessário, funcionando como um dreno para o fundo de C2. Sua abertura é controlada por C2S-LC2, que é o controlador de nível de segurança do fundo de C2. Quanto à V5, seu ajuste inicial fica em cerca de 10%, e ela possui um limite de abertura mínima de 2%, devido às exigências do processo (NOGUEIRA, 2021).

A simulação dinâmica de Nogueira (2021) possui 14 controladores adicionados com a finalidade de garantir a operacionalidade e estabilidade do processo, além de ajudar na convergência da simulação dinâmica. O processo dinâmico com os controladores pode ser visto na Figura 3.3 a seguir.

Figura 3.3: Representação da simulação dinâmica de Nogueira (2021).



Fonte: NOGUEIRA (2021).

Os controladores presentes na simulação são: dois de pressão nos topos das colunas, quatro de nível no fundo e topo das colunas, três de temperatura, incluindo um antes da primeira coluna e um no seu segundo estágio e outro no topo da segunda coluna, quatro controladores de vazão mássica, dois antes da primeira coluna e dois depois da segunda coluna e um controle integrado de vazão mássica e carga térmica na segunda coluna.

A Tabela 3.3 descreve as variáveis controladas e manipuladas para cada um desses controladores.

Tabela 3.3: 14 controladores presentes na simulação dinâmica de Nogueira (2021).

Controlador	Variável controlada	Variável manipulada
SW1-FC	Vazão Mássica de SW1	% Abertura da V3
SW5-FC	Vazão Mássica de SW5	% Abertura da V2
CW5-FC	Vazão Mássica de CW5	% Abertura da V11
CW8-FC	Vazão Mássica de CW8	% Abertura da V12
SW8-TC	Temperatura de SW8	% Abertura da V5
C1S2-TC	Temperatura do 2º Estágio da C1	Carga do Refervedor da C1
AMG1-TC	Temperatura de AMG1	Carga do Condensador da C2
C1S-LC	Nível do Fundo de C1	% Abertura da V6
C2S-LC	Nível do Fundo da C2	% Abertura da V9
C2D-LC	Nível do Vaso de Refluxo da C2	Vazão Mássica do Vaso de Refluxo da C2
C2S-LC2	Redundância do Nível do Fundo da C2	% Abertura da V10
ACG1-PC	Pressão de AGC1	% Abertura da V4
AMG1-PC	Pressão de AMG1	% Abertura da V7
C2HD-IC	Razão entre a Carga do Refervedor de C2 e a Vazão Mássica de PW7	Carga do Refervedor da C2

Fonte: NOGUEIRA (2021).

A sintonização dos controladores foi realizada baseando-se em regras gerais descritas na literatura e através do conselho de especialistas na indústria (NOGUEIRA, 2021). Os parâmetros de ajuste dos controladores podem ser encontrados na Tabela 3.4.

Tabela 3.4: Parâmetros de ajuste dos controladores da simulação dinâmica de Nogueira (2021).

Controlador	Ganho	Tempo integral (min)	Ação
SW1-FC	1,50	0,07	Reversa
SW5-FC	1,50	0,03	Reversa
CW5-FC	1,50	0,07	Reversa
CW8-FC	1,50	0,07	Reversa
SW8-TC	25,00	5,00	Direta
C1S2-TC	1,72	10,20	Reversa
AMG1-TC	1,00	20,00	Reversa
C1S-LC	2,00	22,00	Direta
C2D-LC	2,00	40,00	Direta
C2S-LC	2,00	34,00	Reversa
C2S-LC2	2,00	34,00	Direta
ACG1-PC	30,00	4,00	Direta
AMG1-PC	4,00	20,00	Direta
C2HD-IC	0,30	1,00	Reversa

Fonte: NOGUEIRA (2021).

Como resultado, Nogueira (2021) alcançou recuperação de H₂S igual a 89,08% conforme Tabela 3.5, próximo do valor exigido pela legislação ambiental brasileira que é 90%.

Tabela 3.5: Recuperação dos componentes da simulação dinâmica de Nogueira (2021).

Recuperação	ACG1	AMG1	TW1
H ₂ S	89,08	10,45	0,47
NH ₃	0,01	90,88	9,11
H ₂ O	0,02	0,27	99,71

Fonte: NOGUEIRA (2021).

3.3 Modificação da simulação dinâmica base de Nogueira (2021)

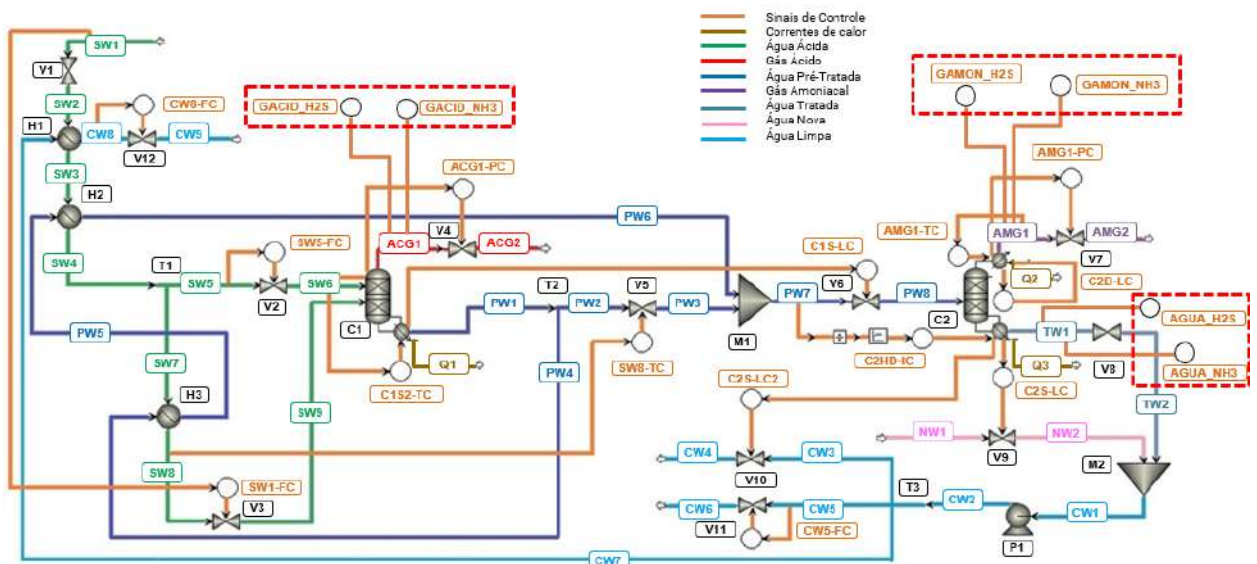
Neste trabalho, o primeiro passo para a criação do novo banco de dados destinado ao desenvolvimento dos sensores virtuais envolveu a inclusão de seis controladores previamente

não utilizados na simulação dinâmica da UTAA com duas colunas de esgotamento de Nogueira (2021).

Desses controladores, dois foram incorporados à corrente de gás ácido (ACG1) no topo da Coluna 1 (GACID_H2S e GACID_NH3), outros dois foram acrescentados à corrente de gás amoniacal (AMG1) no topo da Coluna 2 (GAMON_H2S e GAMON_NH3), e mais dois (AGUA_H2S e AGUA_NH3) foram introduzidos na corrente de água tratada (TW1) que é retirada do fundo da Coluna 2.

Os seis novos controladores podem ser visualizados em destaque na Figura 3.4, que se encontra ampliada no Apêndice A.

Figura 3.4: Adaptação da simulação dinâmica de Nogueira (2021) com a inclusão dos novos seis controladores: GACID_H2S, GACID_NH3, GAMON_H2S, GAMON_NH3, AGUA_H2S e AGUA_NH3.



Fonte: Elaboração própria.

Para viabilizar a coleta dos dados dinâmicos das frações de H₂S e NH₃ presentes nos efluentes descritos acima, os controladores PID (Proporcional, Integral, Derivativo) incluídos foram ajustados com um ganho de 0, um tempo integral de 1 minuto e um tempo derivativo de 0. Isso os configurou para operar unicamente como "indicadores", sem exercer efetivamente funções de controle durante a simulação.

3.4 Desenvolvimento do banco de dados

A simulação dinâmica, com as devidas alterações mencionadas, foi testada em diversos cenários de operação normal com distúrbios em cinco variáveis de interesse, sendo elas: vazão mássica de SW1, vazão mássica de SW5, temperatura da corrente SW1 e frações mássicas de NH₃ e H₂S na corrente SW1. É válido ressaltar que a operação normal se caracteriza quando não ocorrem eventos como a abertura ou o fechamento completo de válvulas, transbordos ou falta de nível nas colunas.

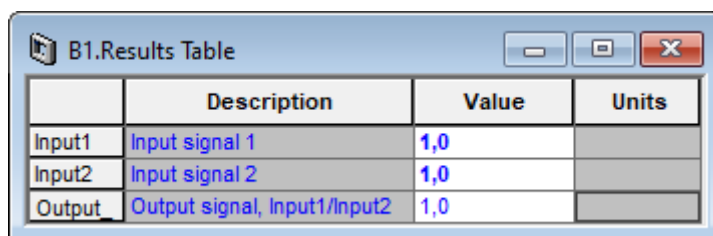
A estratégia inicial para construção do banco de dados foi desenvolver uma corrida global composta por 4 distúrbios aleatórios que variavam entre +1% e -1% do valor do *set point* (SP) de cada uma das cinco variáveis estudadas: vazão mássica de SW1, vazão mássica de SW5, temperatura da corrente SW1 e frações mássicas de NH₃ e H₂S na corrente SW1.

Essas variáveis foram escolhidas, pois são variáveis de entrada do processo que poderiam sofrer alterações em um processo real. SW1 é a água ácida que entra no sistema para tratamento e SW5 representa 10% de SW1, que após passar pelos trocadores de calor, entra no topo da coluna C1.

Paralelamente, foram desenvolvidas corridas parciais que apresentavam apenas os distúrbios de uma variável específica no mesmo tempo em que era dado na corrida global. Isto é, a corrida parcial da vazão mássica de SW1, por exemplo, refere-se à corrida global na qual todos os distúrbios que não estão relacionados à vazão mássica de SW1 foram removidos.

Em seguida, realizou-se uma análise exploratória dentro da simulação dinâmica com o intuito de identificar a variável com o maior potencial de modificar a fração de H₂S no gás ácido. Para isso, foi adicionado um elemento de divisão na simulação, no qual o *Input* 1 era a fração mássica de H₂S na corrente de gás ácido e o *Output* era o *Set Point Remote* do controlador GACID_H2S. Com a simulação rodando, o valor do *Input* 2 foi alterado, a fim de ajustar um novo SP fora da operação normal do controlador para que as respostas das demais variáveis envolvidas no processo pudessem ser avaliadas. A Figura 3.5 é uma representação genérica dos parâmetros que devem ser definidos no elemento de divisão.

Figura 3.5: Representação genérica do *pop-up* de um elemento de divisão.



	Description	Value	Units
Input1	Input signal 1	1,0	
Input2	Input signal 2	1,0	
Output	Output signal, Input1/Input2	1,0	

Fonte: Elaboração própria.

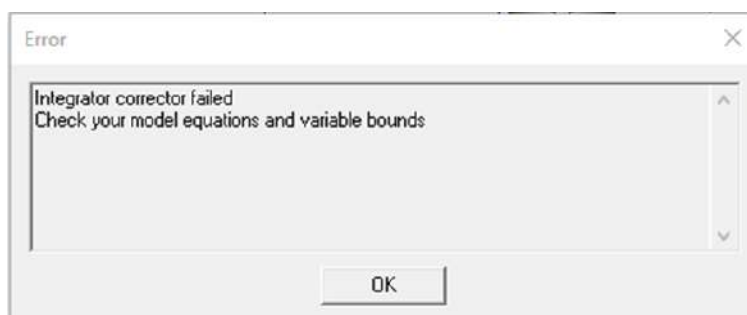
O processamento dos dados incluiu o armazenamento dos dados dinâmicos de todos os 20 controladores após cada simulação (ou “corrida”). A concatenação dessas informações dentro da operação normal resultou no desenvolvimento do banco de dados do trabalho em questão. Vale ressaltar que, neste estudo, não foram considerados ou incluídos atrasos temporais (*delays*) na geração do novo banco de dados.

Posteriormente, o conjunto completo dos dados foi dividido em dados de treinamento e teste para serem utilizados nos modelos de Inteligência Artificial para desenvolvimento dos sensores virtuais. Estes bancos normalizados podem ser encontrados no Apêndice B.

3.5 Avaliação dos métodos de integração do *Aspen Plus Dynamics*®

Durante a busca por uma maior variação na fração de H₂S do gás ácido, diversas corridas foram desenvolvidas. Entretanto, algumas dessas simulações apresentaram erros de integração em certos momentos durante a rodada.

Figura 3.6: Representação do *pop-up* de erro de integração durante simulação no *Aspen Plus Dynamics*®.



Fonte: Elaboração própria.

Portanto, o objetivo foi avaliá-las quanto ao quesito de integração numérica, uma vez que podem ocorrer erros durante o cálculo das soluções numéricas necessárias para resolver

as equações diferenciais que descrevem o comportamento dinâmico de um processo, especialmente em sistemas complexos como o estudado.

A ocorrência de erros de integração em simulações no *Aspen Plus Dynamics*®, assim como em qualquer *software* de simulação dinâmica, pode ser atribuída a diversas causas como: erro na configuração do modelo, complexidade das equações diferenciais que precisam ser resolvidas, condições iniciais que não estão bem definidas, tamanho do passo de integração e até limitações do próprio *software*.

A simulação dinâmica original utilizada por Nogueira (2021) e base para o desenvolvimento do banco de dados do presente trabalho já tinha como modelo de integração padrão o *Implicit Euler*. Com o objetivo de testar se este modelo de integração continuava sendo o melhor para as novas corridas, outros dois diferentes modelos integradores disponíveis (*Runge Kutta 4* e *Gear*) foram testados com as simulações parciais.

O método de integração *Runge Kutta 4* é uma técnica numérica para resolver equações diferenciais ordinárias (EDOs) com alta precisão. Comparado ao método de *Euler*, ele é mais preciso por ser um método de quarta ordem. No entanto, requer mais cálculos por etapa, gerando maior demanda computacional (ROMEO, 2020).

Já o *Gear* pertence à classe de métodos lineares de múltiplos passos. Como método de segunda ordem, ele também é mais preciso do que o *Euler*. Além disso, é particularmente útil para resolver EDOs de alta ordem e é conhecido por sua estabilidade e capacidade de lidar com sistemas de equações diferenciais rígidas, que são aquelas em que as escalas de tempo dos processos envolvidos variam significativamente (CST STUDIO SUITE, 2020).

3.6 Inteligência Artificial

3.6.1 Preparação dos dados de treinamento e teste

O banco de dados recém-criado foi dividido em conjuntos de treinamento e teste. No processo de tratamento dos dados para o algoritmo de AI, os registros relacionados ao controlador C2S-LC2 foram removidos, já que duplicavam as informações do controlador C2S-LC.

Além disso, para simulações que apresentaram erro de integração, foram deletados todos os dados da hora anterior ao erro até o final da simulação. Por exemplo, se o erro de integração ocorreu em 17,6h, os dados foram apagados de 16h até 17,6h.

Para assegurar a representatividade e o treinamento eficiente dos modelos, garantiu-se que cada banco possuía pelo menos um exemplo de cada simulação em operação normal. Além disso, todos os dados relacionados a uma única corrida foram consolidados no mesmo banco, com o objetivo de prevenir qualquer vazamento de dados (*data leakage*). Isso é crucial, pois poderia afetar o desempenho do modelo durante as etapas de treinamento e/ou avaliação, potencialmente levando a resultados enganosos.

Adicionalmente, foi realizado o embaralhamento dos dados, a separação em entradas e saídas, bem como a sua normalização.

3.6.2 Algoritmo da Inteligência Artificial

O método selecionado para o desenvolvimento da AI foi o Florestas Aleatórias (função *RandomForestRegressor*). Foi desenvolvido um algoritmo de RF para cada um dos seis sensores virtuais que receberam o mesmo nome dos novos controladores previamente citados: GACID_H2S, GACID_NH3, GAMON_H2S, GAMON_NH3, AGUA_H2S e AGUA_NH3. Cada algoritmo foi treinado, a validação foi conduzida com 20% dos dados de treinamento, e a avaliação final do melhor modelo para cada sensor foi realizada utilizando os dados do banco de teste.

Para verificação dos sensores produzidos pelo algoritmo Florestas Aleatórias, também foram utilizados os métodos de *Gradient Boosting* (função *GradientBoostingRegressor*) e SVM (função *SVR*, *kernels* linear e RBF) presentes na biblioteca *scikit-learn* (0.19.0). Esses foram os *kernels* escolhidos pois apresentaram os melhores resultados e menores tempos computacionais no estudo de Nogueira (2021).

Para facilitar a avaliação de cada sensor, foram plotados gráficos de predição que comparam o valor verdadeiro presente no banco de teste com o valor predito pelo modelo do sensor e gráficos de resíduo que fornecem a informação da dispersão do erro (diferença entre os valores reais e os valores preditos pelo modelo) ao longo das faixas de valores previstos.

3.6.3 Hiperparâmetros

A biblioteca *Optuna* foi utilizada para otimização dos hiperparâmetros do algoritmo RF, buscando obter o modelo mais eficaz e preciso para cada um dos seis sensores, o que é alcançado pela maximização do valor da função objetivo.

Os hiperparâmetros otimizados foram o número de estimadores (*rf_n_estimators*) que representa o número de árvores na floresta, a profundidade máxima das árvores (*rf_max_depth*) e o número mínimo de amostras necessárias para formar uma folha (nó terminal) da árvore (*rf_min_samples_leaf*).

O aumento do hiperparâmetro *rf_max_depth* pode permitir que as árvores capturem relacionamentos mais complexos nos dados, mas também pode levar a um ajuste excessivo (*overfitting*), onde o modelo se ajusta demais aos dados de treinamento e não generaliza bem para novos dados. Já um valor maior para o hiperparâmetro *rf_n_estimators* geralmente leva a um modelo mais robusto e estável, reduzindo a variância. No entanto, aumentar demais o número de estimadores pode aumentar o tempo de treinamento do modelo. Por fim, um aumento no hiperparâmetro *rf_min_samples_leaf* torna as folhas mais "puras" e pode ajudar a evitar que as árvores tenham muitas divisões pequenas e ruidosas. Um valor maior deste hiperparâmetro pode levar a árvores mais simples e modelos mais robustos contra sobreajuste.

Outra informação importante é que foram definidos limites inferiores e superiores para cada um dos hiperparâmetros do RF como pode ser observado na Tabela 3.6. Isso aumenta a eficiência da otimização, eliminando a necessidade de considerar valores extremos ou inviáveis, auxilia na prevenção de problemas de sobreajuste ou subajuste do modelo, resultando em economia de tempo computacional. Da mesma forma, o número de iterações (ou número de *trials*) para o processo de otimização dos hiperparâmetros foi definido em 40.

Tabela 3.6: Limites inferiores e superiores definidos para cada hiperparâmetro.

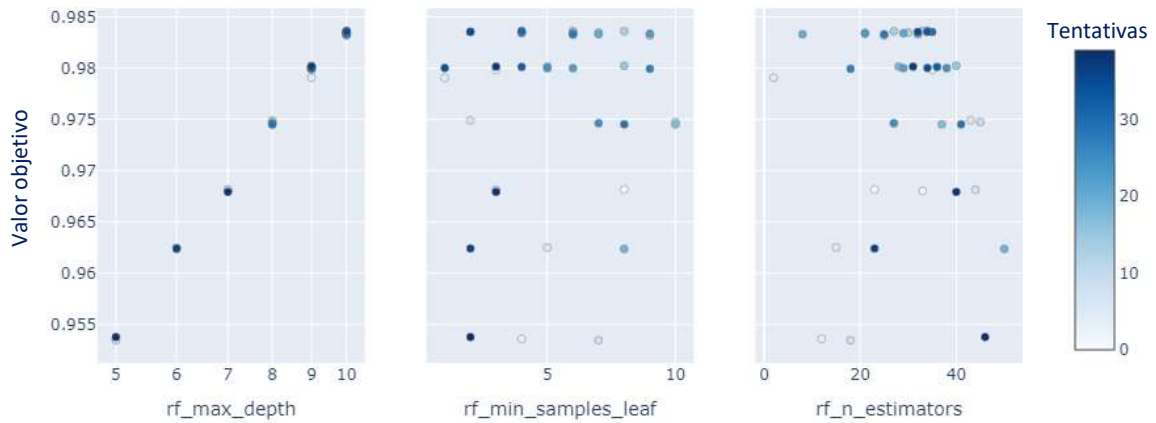
Hiperparâmetros	Limites
<i>rf_n_estimators</i>	1 a 50
<i>rf_max_depth</i>	5 a 10
<i>rf_min_samples_leaf</i>	1 a 10

Fonte: Elaboração própria.

As Figuras 3.7 e 3.8 ilustram um exemplo de resultado da otimização dos hiperparâmetros avaliados no algoritmo do RF. O gráfico de cortes mostra como cada

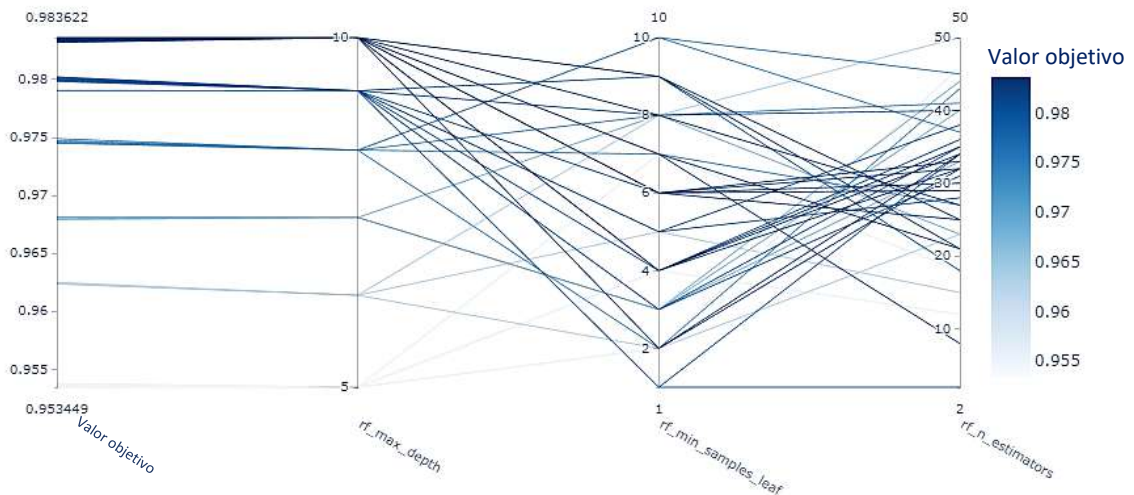
hiperparâmetro afeta individualmente o valor da função objetivo, já o gráfico de coordenadas paralelas representa como a associação dos hiperparâmetros afeta o valor objetivo.

Figura 3.7: Cortes da otimização dos hiperparâmetros do algoritmo Florestas Aleatórias.



Fonte: Elaboração própria.

Figura 3.8: Coordenadas paralelas da otimização dos hiperparâmetros do algoritmo Florestas Aleatórias.



Fonte: Elaboração própria.

Analogamente ao RF, os hiperparâmetros citados na Tabela 3.6 foram otimizados para o algoritmo do GB utilizando o mesmo número de iterações, enquanto que, para o SVM, não houve otimização dos hiperparâmetros (C e γ), já que a otimização destes não apresentou ganhos significativos nos resultados conforme apresentado na dissertação de Mestrado de Nogueira (2021).

3.6.4 Métricas

Uma das métricas utilizadas para avaliar cada modelo do RF e GB foi a Importância de Variáveis, que não está disponível para o método SVM. Após avaliação desta métrica para estes modelos, foi realizada a redução do conjunto de dados de entrada cujas variáveis apresentaram importância menor que 2% para os sensores. Posteriormente, foram gerados gráficos de predição, resíduos, importância das variáveis e tabela de métricas atualizados, utilizando apenas os dados de entrada reduzidos.

Além disso, foram amplamente utilizados indicadores muito comuns em problemas de regressão em ML que estão presentes na biblioteca *scikit-learn*, como: erro absoluto médio (*Mean Absolute Error*, MAE), erro quadrático médio (*Mean Squared Error*, MSE), raiz do erro quadrático médio (*Root Mean Squared Error*, RMSE) e coeficiente de determinação (R^2).

O R^2 representa o percentual da variância dos dados que é explicado pelo modelo de regressão. Quanto mais próximo de 1, maior o ajuste do modelo aos dados. O MAE mede a média das diferenças absolutas entre as previsões de um modelo e os valores reais. Ele fornece uma ideia de quão longe, em média, as previsões estão dos valores verdadeiros e é útil quando não se deseja penalizar erros grandes de forma excessiva (JÚNIOR, 2021).

Já o MSE é semelhante ao MAE, mas em vez de calcular a média das diferenças absolutas, ele calcula a média das diferenças ao quadrado. Isso tem o efeito de penalizar erros maiores de forma mais significativa, sendo comumente usado quando é importante penalizar erros maiores e quando se deseja que o modelo dê mais ênfase à precisão das previsões. O RMSE tem a vantagem de fornecer uma métrica de erro que está na mesma unidade da variável de resposta, tornando-a facilmente interpretável. Quanto menor o valor do RMSE, mais precisas são as previsões do modelo para o conjunto de dados empregado (JÚNIOR, 2021).

A qualidade do banco de treinamento também foi avaliada utilizando os indicadores citados, da mesma forma que foi feito para o banco de teste. Essas avaliações foram essenciais para garantir tanto que o modelo havia conseguido realizar boas generalizações como para verificar a consistência e confiabilidade dos dados de treinamento.

Por fim, após a análise das métricas de cada modelo, foi definida uma faixa de confiabilidade nos gráficos de predição via inspeção visual. Essa faixa abrange o maior número possível de valores previstos que se aproximaram dos valores verdadeiros,

proporcionando assim uma representação visual clara das áreas onde as predições dos modelos foram mais confiáveis.

Capítulo 4

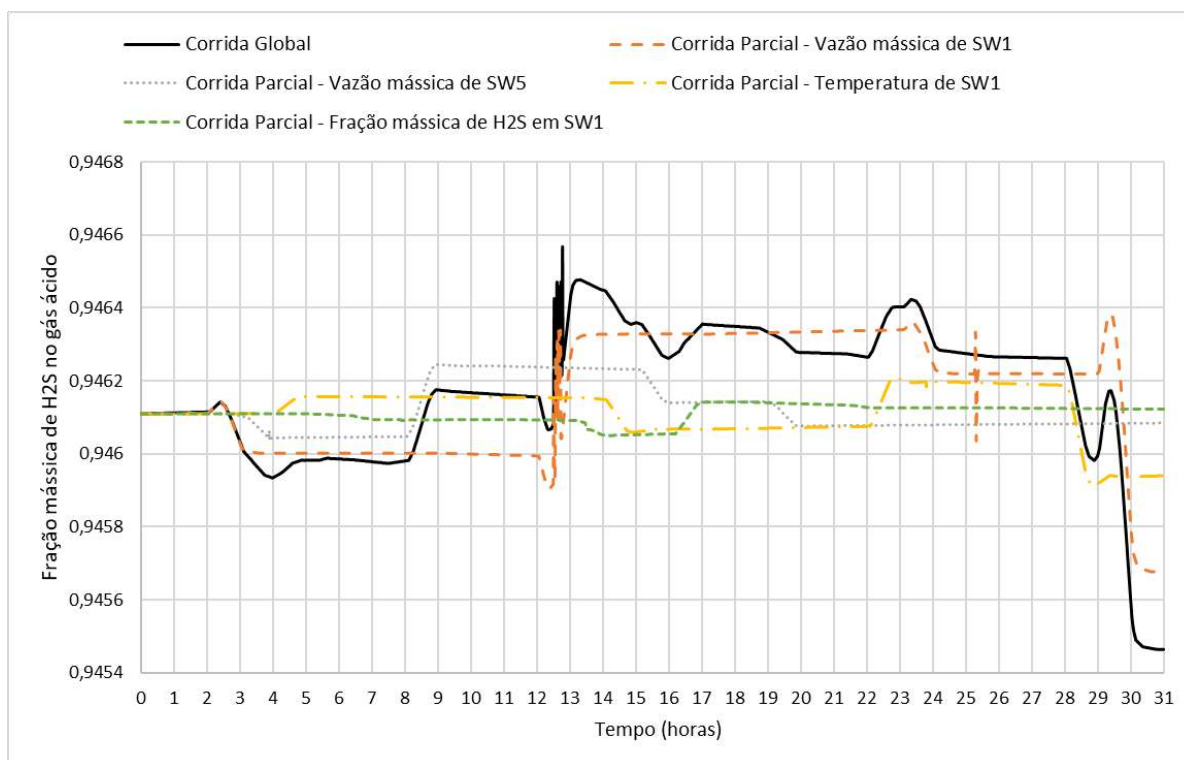
4. Resultados e Discussões

4.1 Desenvolvimento do banco de dados e estudo fenomenológico do processo

Os resultados da corrida global, que engloba distúrbios aleatórios variando entre -1% a +1% do valor do SP das cinco variáveis de entrada analisadas, juntamente com as respectivas corridas parciais, deram origem às Figuras 4.1 a 4.6, que foram apresentadas a seguir.

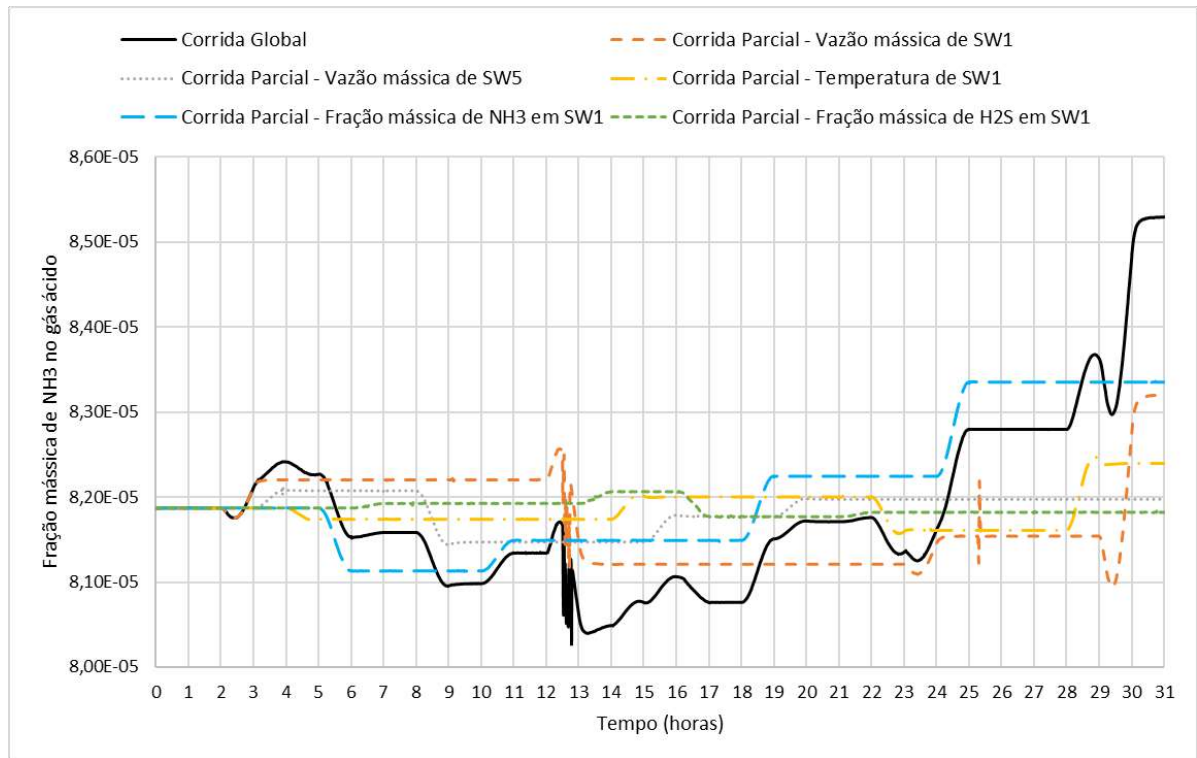
Os gráficos têm como propósito destacar quais variáveis exerceram a maior influência nas frações de H₂S e NH₃ dos efluentes analisados. Isso se tornou evidente quando, em determinado momento, a corrida parcial de dada variável apresentou o mesmo comportamento observado na corrida global.

Figura 4.1: Variáveis importantes para determinação da fração de H₂S no gás ácido.



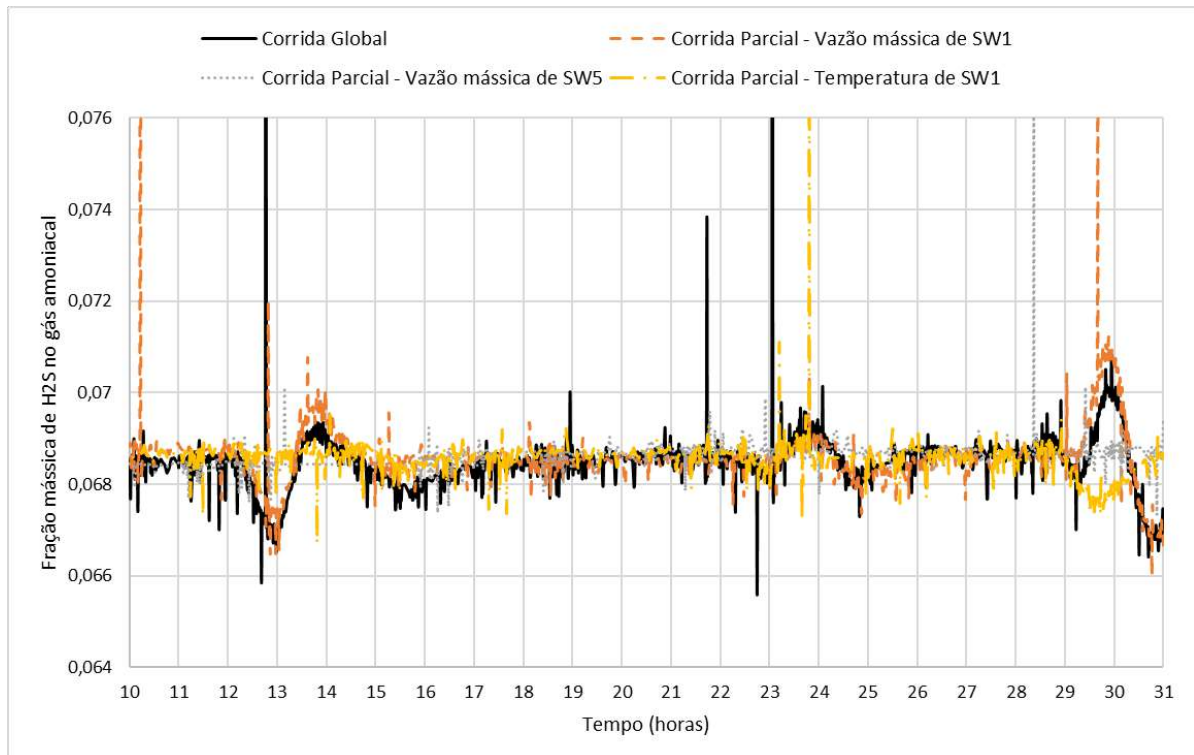
Fonte: Elaboração própria.

Figura 4.2: Variáveis importantes para determinação da fração de NH_3 no gás ácido.



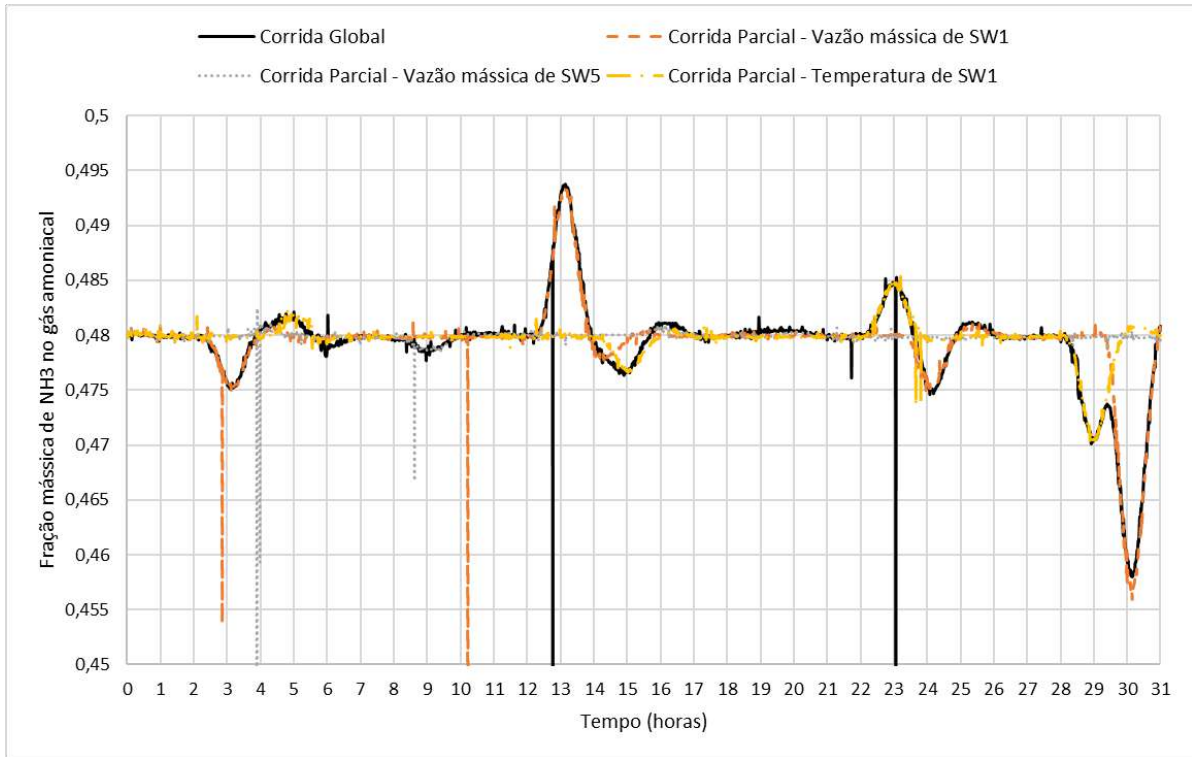
Fonte: Elaboração própria.

Figura 4.3: Variáveis importantes para determinação da fração de H_2S no gás amoniacal.



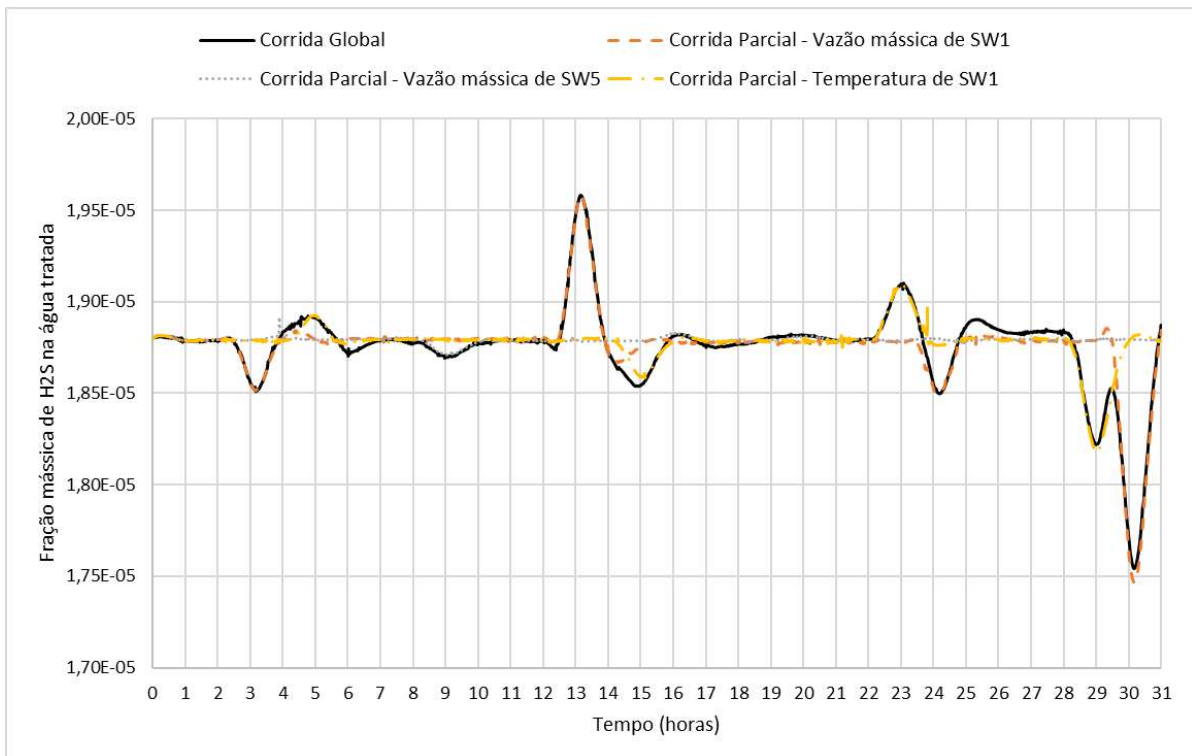
Fonte: Elaboração própria.

Figura 4.4: Variáveis importantes para determinação da fração de NH_3 no gás amoniacal.



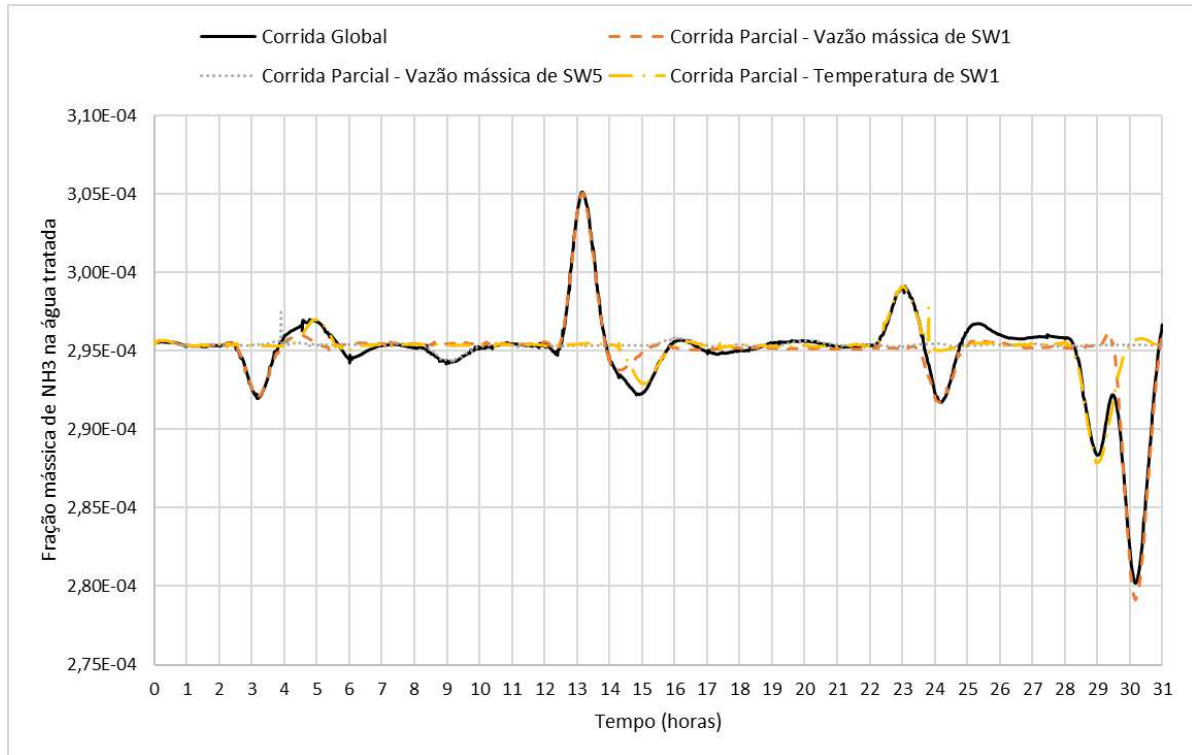
Fonte: Elaboração própria.

Figura 4.5: Variáveis importantes para determinação da fração de H_2S na água tratada.



Fonte: Elaboração própria.

Figura 4.6: Variáveis importantes para determinação da fração de NH_3 na água tratada.



Fonte: Elaboração própria.

Após interpretação das Figuras 4.1 a 4.6, conclui-se que as vazões mássicas de SW1 e SW5 e a temperatura de SW1 são variáveis extremamente importantes para o processo termodinâmico de separação nas duas colunas de esgotamento, uma vez que elas interferem diretamente nas frações de H_2S e NH_3 do gás ácido (efluente da Coluna 1), gás amoniacal e água tratada (efluentes da Coluna 2). É um resultado plausível pois se trata de características ligadas à corrente de entrada do processo.

No entanto, apesar dos desvios aplicados, notou-se que a simulação permaneceu bastante estável, exibindo pequena variação nas frações de H_2S e NH_3 dos efluentes, especialmente na fração de H_2S do gás ácido, que registrou apenas 0,12% de variação. Esse resultado não era o ideal, uma vez que implicaria em uma baixa variabilidade nos dados usados posteriormente no treinamento dos sensores virtuais.

Sendo assim, foram desenvolvidas novas corridas globais com distúrbios substancialmente maiores (entre -25% a +30% do valor do *set point*) em cada uma das cinco variáveis estudadas de forma individual, enquanto se mantinham os distúrbios entre +1% e -1% nas demais variáveis de entrada. O objetivo era obter, pelo menos, uma variação de 1% na fração de H_2S do gás ácido e manter a simulação dentro da operação normal.

As corridas globais com distúrbios mais intensos na vazão mássica de SW1, temperatura de SW1, e frações mássicas de NH₃ e H₂S em SW1 apresentaram novamente pequena variação na fração de H₂S do gás ácido, como exibido na Tabela 4.1. Por outro lado, a corrida global da vazão mássica de SW5 se mostrou mais a promissora entre todas.

Tabela 4.1: Variações da fração mássica de H₂S no gás ácido para cada corrida global.

Corridas globais	Variação da fração mássica de H₂S (%)
Vazão mássica de SW1	0,15
Vazão mássica de SW5	0,66
Temperatura de SW1	0,29
Fração mássica de H ₂ S em SW1	0,24
Fração mássica de NH ₃ em SW1	0,12

Fonte: Elaboração própria.

Portanto, uma nova corrida global foi criada para a vazão mássica de SW5 com distúrbios mais intensos do que os anteriores (entre -30% a +40% do valor do *set point* da variável), o que resultou em 1% de variação na fração de H₂S do gás ácido, mantendo a operação normal.

Com o intuito de ampliar ainda mais a diversidade do banco de dados, o que contribuiria para o desenvolvimento de sensores mais robustos posteriormente, foi realizada uma análise exploratória no controlador GACID_H2S com interesse de buscar a variável com maior capacidade de influenciar a fração de H₂S do gás ácido. A tentativa foi de “enganar” o SP do controlador para entender como as demais variáveis do sistema seriam perturbadas.

Após o teste, identificou-se que a variável mais afetada pela mudança foi a temperatura do segundo estágio da Coluna 1, variável controlada pelo controlador C1S2-TC. Com base nisso, novas corridas foram geradas com foco exclusivo na avaliação desta variável.

Uma delas consistiu em uma rampa crescente da temperatura do 2º estágio da Coluna 1, aplicando apenas distúrbios positivos (entre 2% a 15%) no *set point* do controlador C1S2-TC. Isso resultou em uma variação da fração de H₂S no gás ácido de 2,3%, o que representa mais que o dobro do valor obtido até então.

Com o objetivo de avaliar melhor a influência desta variável no processo de separação dos contaminantes NH₃ e H₂S na Coluna 1, durante a simulação, em alguns pontos

específicos, foi realizada a verificação da vazão mássica total da corrente do gás ácido (ACG1) para que pudesse ser calculada a recuperação do H₂S no gás ácido. As informações obtidas desta simulação podem ser encontradas, em valores desnormalizados, na Tabela 4.2.

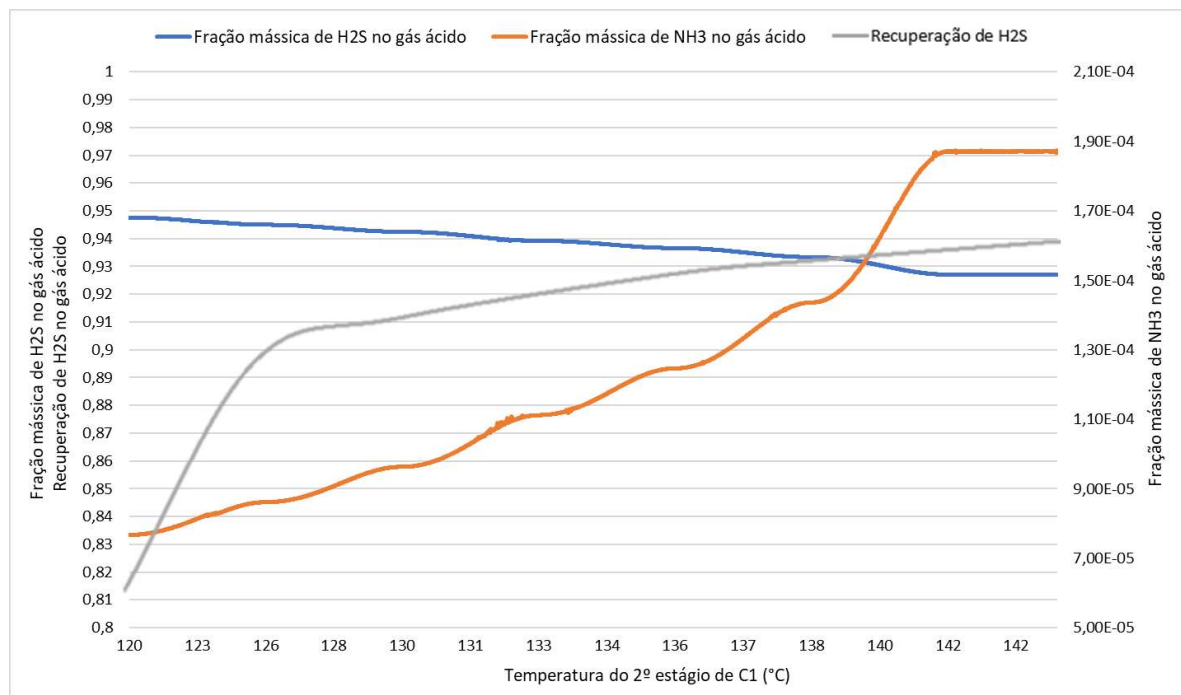
Tabela 4.2: Resultados da análise exploratória na variável da temperatura do 2º estágio da Coluna 1.

Tempo (h)	Fração mássica de H ₂ S no gás ácido	Fração mássica de NH ₃ no gás ácido	Temperatura do 2º estágio de C1 (°C)	Recuperação do H ₂ S no gás ácido (%)	Varição entre os valores de recuperação (%)
2,00	0,9476	7,67.10 ⁻⁵	120,30	81,25	-
7,30	0,9461	8,19.10 ⁻⁵	123,32	89,54	9,25
15,00	0,9445	8,79.10 ⁻⁵	126,65	91,10	1,71
30,00	0,9395	1,10.10 ⁻⁴	132,94	92,68	1,70
45,00	0,9360	1,28.10 ⁻⁴	136,21	93,18	0,54
60,00	0,9277	1,82.10 ⁻⁴	141,54	93,83	0,70
70,00	0,9270	1,87.10 ⁻⁴	141,90	93,85	0,02

Fonte: Elaboração própria.

Através dos dados dinâmicos e das informações da Tabela 4.2 acima, foi possível construir o gráfico presente na Figura 4.7 a seguir.

Figura 4.7: Relação entre recuperação do H₂S no gás ácido e temperatura do 2º estágio da Coluna 1.



Fonte: Elaboração própria.

Os resultados obtidos evidenciaram que a eficiência de remoção do H₂S aumenta com a elevação da temperatura do 2º estágio da Coluna 1, principalmente até se obter cerca de 90% de recuperação, valor que é requerido pela legislação ambiental brasileira.

A partir desse ponto, nota-se que o aumento contínuo da temperatura da coluna produz incrementos pequenos na variação da recuperação de H₂S. Por outro lado, inicia-se a vaporização mais pronunciada do NH₃, resultando em um gás ácido com maior teor de amônia. Conforme discutido anteriormente, elevados teores de NH₃ no gás ácido que será enviado para a URE pode acarretar em depósitos de sais de amônio nos pontos frios da unidade e aumento das emissões de NO_x, o que é ambientalmente indesejado.

Dessa forma, destaca-se a necessidade de um controle efetivo sobre a carga térmica da primeira coluna, a fim de evitar uma sobre especificação da eficiência de recuperação do H₂S e gasto energético desnecessário no refeedor. Essa análise evidencia o potencial ganho de otimização de um processo a se utilizar sensores virtuais para analisar composições dos efluentes de uma UTAA.

4.2 Avaliação dos métodos de integração do *Aspen Plus Dynamics*®

Após a realização dos testes, o método de integração *Runge Kutta 4* se mostrou inviável para as simulações avaliadas, especificamente nas corridas globais com distúrbios mais intensos citadas no tópico 4.1.

Nesses casos, o erro de integração ocorria no começo de cada corrida, sendo impossível iniciá-las mesmo após diversas tentativas. Esse resultado pode ser atribuído à quarta ordem do método, que introduz uma complexidade significativa aos cálculos numéricos, tornando-o inadequado para lidar com as condições desafiadoras das corridas com distúrbios intensos.

Portanto, as comparações foram realizadas utilizando o método padrão *Implicit Euler* e o método alternativo *Gear*. Com a Tabela 4.3 a seguir, foi possível concluir que o método padrão (*Implicit Euler*) já era a melhor opção para o sistema estudado. O método *Gear* apresentou um desempenho pior para três das seis corridas avaliadas: vazão mássica de SW1, vazão mássica de SW5 (2) e fração mássica de NH₃. Para as demais, o desempenho entre os métodos foi o mesmo.

Assim, optou-se por manter o método de integração *Implicit Euler* como o padrão para todas as simulações. Os erros de integração observados podem ser atribuídos à complexidade das muitas equações diferenciais que o *software* precisa resolver, principalmente nas duas colunas de esgotamento da simulação.

Tabela 4.3: Comparação dos métodos de integração *Implicit Euler* (padrão) e *Gear* (alternativo) utilizando as corridas globais com distúrbios mais intensos.

Corridas globais	<i>Implicit Euler</i>	<i>Gear</i>
Vazão mássica de SW1	Sem erro de integração	Erro no início da corrida
Vazão mássica de SW5	Erro no final da corrida	Erro no final da corrida
Vazão mássica de SW5 (2)	Erro no final da corrida	Erro na metade da corrida
Temperatura de SW1	Erro no início da corrida	Erro no início da corrida
Fração mássica de H ₂ S em SW1	Erro na metade da corrida	Erro na metade da corrida
Fração mássica de NH ₃ em SW1	Erro no final da corrida	Erro na metade da corrida

Fonte: Elaboração própria.

4.3 Preparação dos dados de treinamento e teste

Após a divisão do novo banco de dados, o banco de treinamento ficou composto por 45.483 amostras enquanto o banco de teste conteve 31.525 amostras. Isso indica que 59% dos dados totais estão alocados no banco de treinamento e os 41% restantes foram destinados ao banco de teste. A listagem da quantidade e descrição de cada simulação que compôs cada banco pode ser observada nas Tabelas 4.4 e 4.5.

Para esclarecer a divisão das corridas entre bancos de treinamento e teste, conforme explicado no Capítulo 3, cada banco contém um exemplo de cada simulação. Embora compartilhem a mesma descrição, cada simulação é única, apresentando diversas versões em termos de intensidade, amplitude e momento em que os distúrbios foram realizados.

Tabela 4.4: Listagem das corridas presentes no banco de treinamento.

Banco de Treinamento	
Corrida	Descrição
1	Corrida global com distúrbios entre -1% e +1% no SP das 5 variáveis padrão – versão A
2	Corrida global com distúrbios entre -30% a 40% do SP da vazão mássica de SW5 – versão A
3	Corrida global com distúrbios mais numerosos entre -30% a 40% do SP da vazão mássica de SW5 – versão A
4	Corrida global com distúrbios entre -25% a 30% na temperatura de SW1 – versão A
5	Corrida global com distúrbios entre -25% a 30% na fração mássica de H ₂ S de SW1 – versão A
6	Corrida global com distúrbios entre -25% a 30% na fração mássica de NH ₃ de SW1 – versão A
7	Corrida global com distúrbios positivos em rampa crescente na temperatura do 2º estágio de C1 – versão A
8	Corrida global com distúrbios positivos e negativos aleatórios na temperatura do 2º estágio de C1 – versão A
9	Corrida parcial da vazão mássica de SW1 – versão A
10	Corrida parcial da vazão mássica de SW1 – versão B
11	Corrida parcial da vazão mássica de SW5 – versão A
12	Corrida parcial da vazão mássica de SW5 – versão B
13	Corrida parcial da temperatura de SW1 – versão A
14	Corrida parcial da temperatura de SW1 – versão B
15	Corrida parcial da fração mássica de NH ₃ de SW1 – versão A
16	Corrida parcial da fração mássica de NH ₃ de SW1 – versão B
17	Corrida parcial da fração mássica de H ₂ S de SW1 – versão A
18	Corrida parcial da fração mássica de H ₂ S de SW1 – versão B

Fonte: Elaboração própria.

Tabela 4.5: Listagem das corridas presentes no banco de teste.

(continua)

Banco de Teste	
Corrida	Descrição
1	Corrida global com distúrbios entre -1% e +1% no SP das 5 variáveis padrão – versão B
2	Corrida global com distúrbios entre -30% a 40% do SP da vazão mássica de SW5 – versão B
3	Corrida global com distúrbios entre -25% a 30% na temperatura de SW1 – versão B
4	Corrida global com distúrbios entre -25% a 30% na fração mássica de H ₂ S de SW1 – versão B
5	Corrida global com distúrbios entre -25% a 30% na fração mássica de NH ₃ de SW1 – versão B
6	Corrida global com distúrbios positivos em rampa crescente na temperatura do 2º estágio de C1 – versão B

Tabela 4.5: Listagem das corridas presentes no banco de teste.

(conclusão)

Banco de Teste	
Corrida	Descrição
7	Corrida global com distúrbios positivos e negativos aleatórios na temperatura do 2º estágio de C1 – versão B
8	Corrida parcial da vazão mássica de SW1 – versão C
9	Corrida parcial da vazão mássica de SW5 – versão C
10	Corrida parcial da temperatura de SW1 – versão C
11	Corrida parcial da fração mássica de NH ₃ de SW1 – versão C
12	Corrida parcial da fração mássica de H ₂ S de SW1 – versão C

Fonte: Elaboração própria.

4.4 Comparação dos algoritmos de Inteligência Artificial

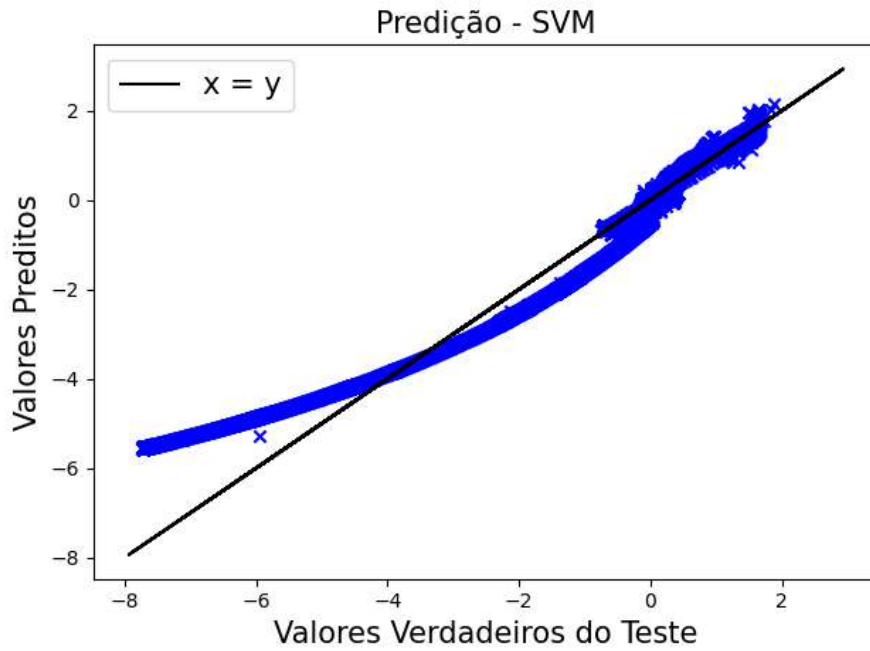
Após aplicar o algoritmo Florestas Aleatórias e criar os seis sensores virtuais, os métodos *Gradient Boosting* e SVM com *kernels* linear e RBF foram avaliados usando os mesmos conjuntos de dados de treinamento e teste, com o objetivo de validar o desempenho do RF.

Na Figura 4.9 ocorreu a comparação do Coeficiente de Determinação (R^2) entre os três métodos avaliados (RF, GB e SVM) para cada um dos seis sensores, enquanto na Figura 4.10 houve uma comparação da Raiz do Erro Quadrático Médio (RMSE).

Uma conclusão inicialmente relevante foi que o método SVM usando o *kernel* linear não se ajustou bem aos dados do novo banco desenvolvido, o que ficou claro pelo gráfico de predição na Figura 4.8. Além disso, o tempo computacional para processamento do algoritmo do sensor GACID_H2S foi significativamente maior, excedendo 20 minutos. O coeficiente de determinação (R^2) encontrado foi de 0,9388, inferior aos resultados alcançados pelos demais métodos conforme Figura 4.9.

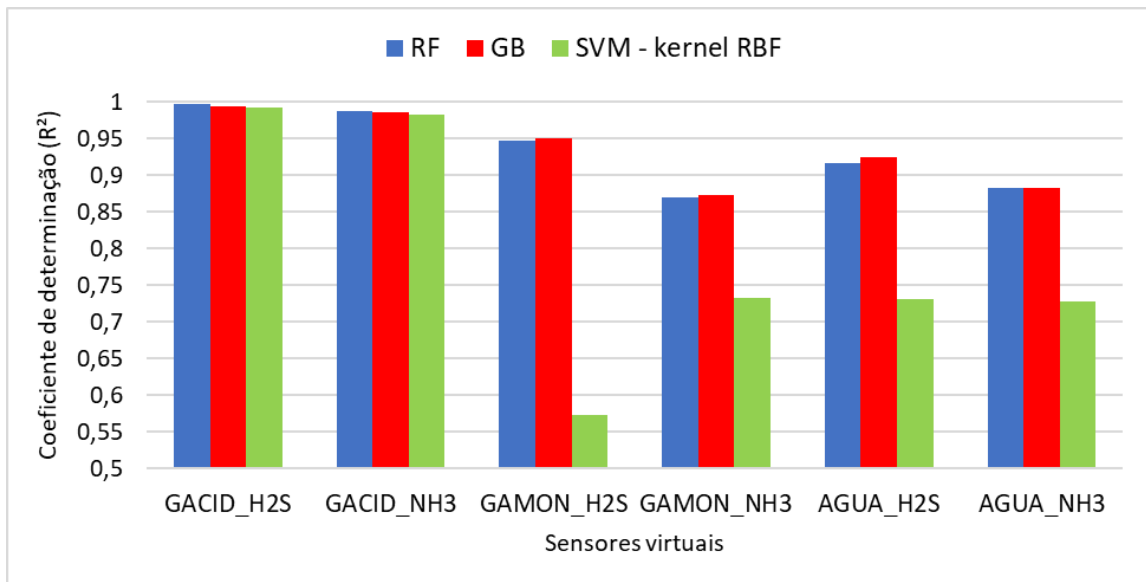
Dessa forma, o *kernel* linear e seus resultados foram descartados para os demais sensores e só foram apresentados os dados relativos ao método SVM – *kernel* RBF.

Figura 4.8: Predição do sensor GACID_H2S usando SVM – *kernel* linear.



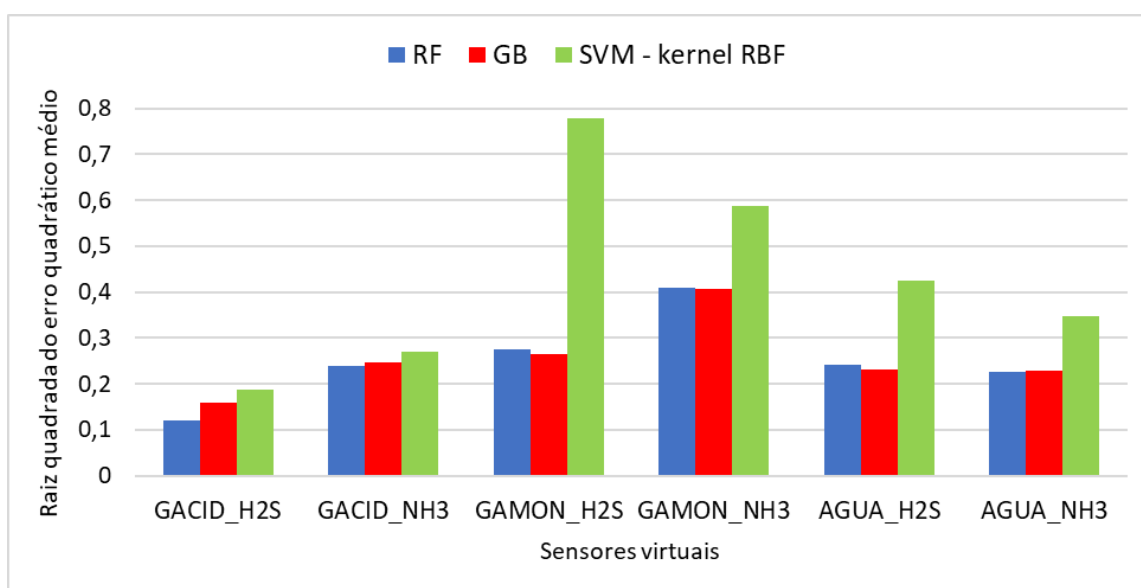
Fonte: Elaboração própria.

Figura 4.9 Comparação do Coeficiente de Determinação (R^2) entre os algoritmos de RF, GB e SVM *kernel* RBF.



Fonte: Elaboração própria.

Figura 4.10: Comparação da Raiz do Erro Quadrático Médio (RMSE) entre os algoritmos de RF, GB e SVM *kernel* RBF.



Fonte: Elaboração própria.

Pelas Figuras 4.9 e 4.10, ficou evidente que as métricas de R^2 e RMSE, para todos os sensores do SVM, apresentaram pior desempenho quando comparadas as dos outros dois métodos. Além disso, é importante ressaltar que esse método não fornece informações sobre a importância das variáveis, um aspecto significativo para a compreensão fenomenológica do processo.

Por outro lado, as métricas do GB são muito similares às encontradas para o RF em todos os sensores. Além das métricas, os gráficos de predição e resíduos e a tabela geral de importância das variáveis são muito semelhantes para ambos os métodos. Esses resultados não foram expostos nesta seção para não prejudicar a fluidez da leitura, porém podem ser encontrados no Apêndice C junto com maior detalhamento do método SVM.

Com base nas observações anteriores, ficou evidente que o GB é um método equivalente ao RF e que o SVM se mostra menos preciso para modelar o banco de dados em questão. Portanto, o modelo RF foi validado pelos demais testes e todas as análises subsequentes dos sensores virtuais foram realizadas com base nesse algoritmo padrão.

4.5 Resultado dos sensores virtuais usando Florestas Aleatórias

Como mencionado, o algoritmo RF foi escolhido para o desenvolvimento da metodologia de AI. Além dos resultados já discutidos na Seção 4.4, esse método demonstrou ser a escolha mais eficaz tanto para o conjunto de dados de Nogueira (2021), alcançando acurácia de 95,2% após a redução das variáveis no caso do estudo de Detecção e Diagnóstico de Falhas (*Fault Detection and Diagnosis, FDD*) da UTAA quanto para o estudo de Pitanga (2022).

Pitanga, partindo do mesmo conjunto de dados de Nogueira (2021), obteve uma acurácia de 94,77% para classificação de cenários de pré-falha, após a separação das amostras de operação normal da região de falha e pré-falha.

4.5.1 Sensor GACID_H2S

Após treinamento e teste do sensor GACID_H2S, foi verificado que o melhor modelo apresentou o maior valor possível para a profundidade das árvores (*rf_max_depth*), enquanto apresentou valores intermediários para o número de árvores (*rf_n_estimators*) e para número mínimo de amostras para separar um nó (*rf_min_samples_leaf*) como indicado na Tabela 4.6.

Tabela 4.6: Hiperparâmetros otimizados para o sensor GACID_H2S usando RF.

Hiperparâmetros	Valor ótimo	Valor Mínimo	Valor Máximo
<i>rf_min_samples_leaf</i>	6	1	50
<i>rf_max_depth</i>	10	5	10
<i>rf_n_estimators</i>	33	1	50

Fonte: Elaboração própria.

Para avaliação do sensor, os gráficos de predição podem ser encontrados nas Figuras 4.11 A e B, os gráficos de resíduos nas Figuras 4.12 A e B, e a Tabela 4.7 apresenta as métricas do modelo.

Nos gráficos de predição, os valores verdadeiros presentes nos bancos de dados de treinamento e teste foram comparados com os valores preditos pelo sensor. Foi possível notar que o sensor apresentou ótima capacidade de previsão dos dados, sendo corroborado pelos altos valores de R^2 iguais a 0,9842 e 0,9966. Além disso, os baixos valores de erro como

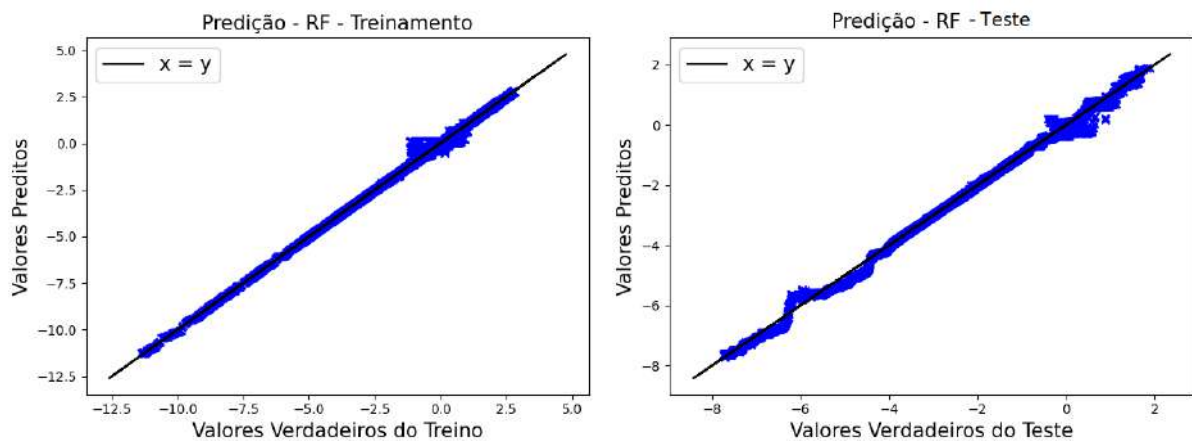
MAE, MSE e RMSE indicaram que o modelo foi capaz de fazer previsões precisas e que as previsões individuais estão, em geral, muito próximas dos valores reais.

Tabela 4.7: Métricas do sensor GACID_H2S após avaliação com dados de treinamento e dados de teste.

Métricas	Treinamento	Teste
R ²	0,9842	0,9966
MAE	0,0526	0,0727
MSE	0,0158	0,0146
RMSE	0,1257	0,1207

Fonte: Elaboração própria.

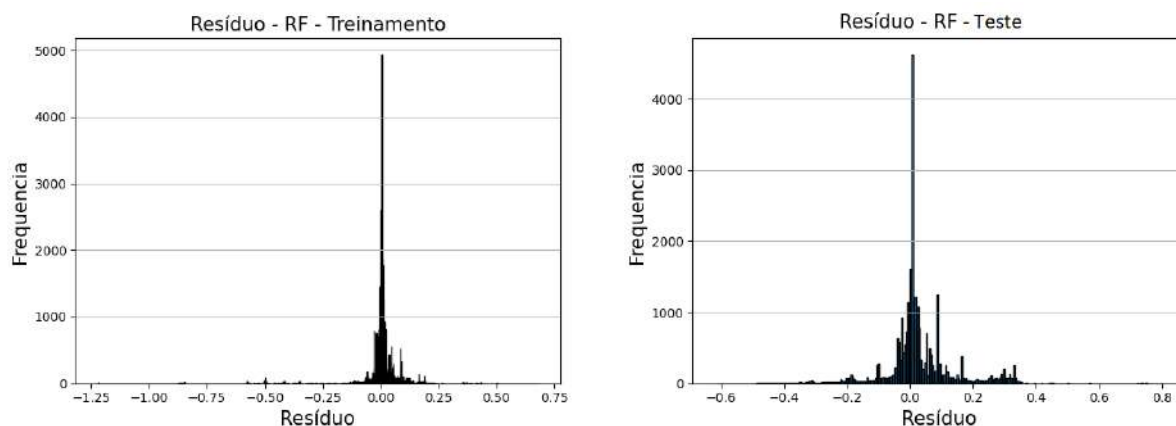
Figuras 4.11 A e B: Predição do sensor GACID_H2S usando RF com dados de treinamento (A) e teste (B).



Fonte: Elaboração própria.

Os gráficos de resíduos foram construídos a partir da diferença entre os valores verdadeiros e os valores previstos pelo modelo. Eles corroboram com as conclusões já comentadas sobre o sensor, uma vez que os resíduos se apresentaram de maneira uniforme em torno do zero, sem apresentar padrões específicos que poderiam indicar problemas no modelo e nem presença de *outliers*.

Figuras 4.12 A e B: Resíduos do sensor GACID_H2S usando RF com dados de treinamento (A) e teste (B).



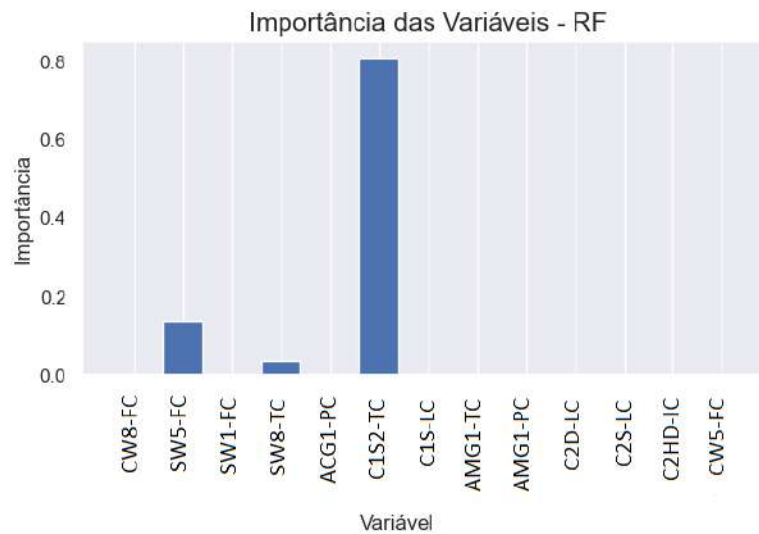
Fonte: Elaboração própria.

Admitindo a ótima capacidade de predição do sensor GACID_H2S, o intervalo de confiabilidade para previsão da fração mássica de H₂S no gás ácido ficou entre 0,9245 e 0,9511, após desnormalização dos dados. Esta faixa foi definida de forma visual, englobando o maior número possível de valores preditos que se aproximaram dos valores reais, como mencionado na metodologia do Capítulo 3. Nesse caso, portanto, a região considerada confiável compreende toda a faixa de predição observada na Figura 4.11B.

O valor da fração mássica de H₂S no gás ácido na operação normal é igual a 0,9461, dessa forma o sensor seria capaz de prever variações entre -2,28% a 0,53% da operação normal.

As variáveis classificadas como mais importantes para o sensor GACID_H2S foram vazão mássica de SW5, temperatura de SW8 e temperatura do 2º estágio da coluna 1 (C1S2) como observado na Figura 4.13. As demais variáveis tiveram uma contribuição menos significativa, portanto não foram incluídas no gráfico.

Figura 4.13: Classificação da importância das variáveis para o sensor GACID_H2S.



Fonte: Elaboração própria.

A redução das variáveis do conjunto de entrada foi realizada excluindo todas as que apresentaram importância menor que 2% para o sensor. Notou-se, na Figura 4.14, que a vazão mássica de SW5, a temperatura de SW8 e a temperatura do 2º estágio de C1 (C1S2) se mantiveram como variáveis mais importantes para o sensor GACID_H2S, sendo a temperatura do 2º estágio de C1 (C1S2) a mais relevante entre as três.

Essa redução do conjunto de dados de entrada, através da remoção de variáveis menos importantes, permite a simplificação e otimização do modelo, além de ganho em tempo de processamento computacional. Como pôde ser observado na Tabela 4.8, houve um ligeiro aumento de 0,13% no R^2 e significativa redução das métricas de erro do banco de teste: 22,69% do MAE, 37,67% do MSE e 20,71% do RMSE, quando comparado aos valores antes da redução das variáveis de entrada.

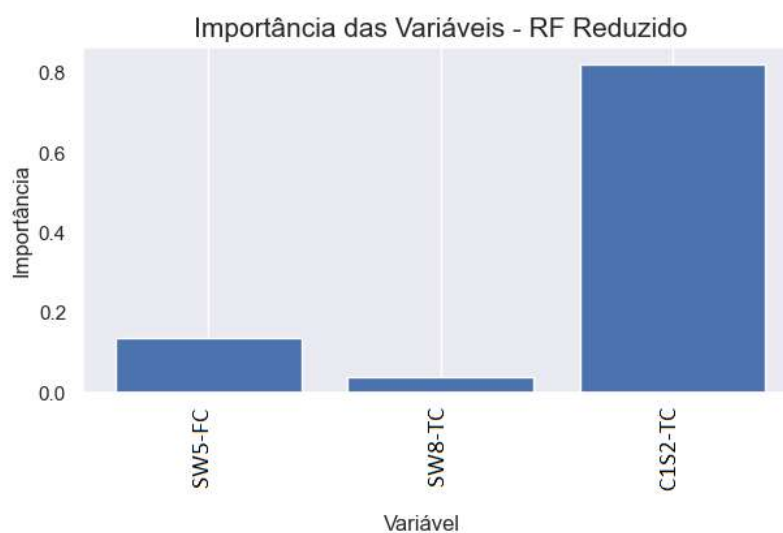
Os gráficos de predição e resíduo após a redução das variáveis de entrada foram suprimidos por serem análogos aos previamente apresentados, entretanto podem ser encontrados no Apêndice D.

Tabela 4.8: Métricas do banco de teste do sensor GACID_H2S após redução das variáveis.

Resultado das métricas de teste após redução das variáveis	
R ²	0,9979
MAE	0,0562
MSE	0,0091
RMSE	0,0957

Fonte: Elaboração própria.

Figura 4.14: Classificação da importância das variáveis para o sensor GACID_H2S após redução.



Fonte: Elaboração própria.

4.5.2 Sensor GACID_NH3

No caso do sensor GACID_NH3, observou-se que o modelo mais eficaz demonstrou ter a profundidade máxima das árvores (*rf_max_depth*). Além disso, ele teve o número de árvores (*rf_n_estimators*) próximo ao limite superior e o número mínimo de amostras para separar um nó (*rf_min_samples_leaf*) próximo ao limite inferior, conforme detalhado na Tabela 4.9.

Tabela 4.9: Hiperparâmetros otimizados para o sensor GACID_NH3 usando RF.

Hiperparâmetros	Valor ótimo	Valor Mínimo	Valor Máximo
<i>rf_min_samples_leaf</i>	3	1	50
<i>rf_max_depth</i>	10	5	10
<i>rf_n_estimators</i>	47	1	50

Fonte: Elaboração própria.

Para avaliação do sensor, os gráficos de predição podem ser encontrados nas Figuras 4.15 A e B, os gráficos de resíduos nas Figuras 4.16 A e B, e a Tabela 4.10 apresenta as métricas do modelo.

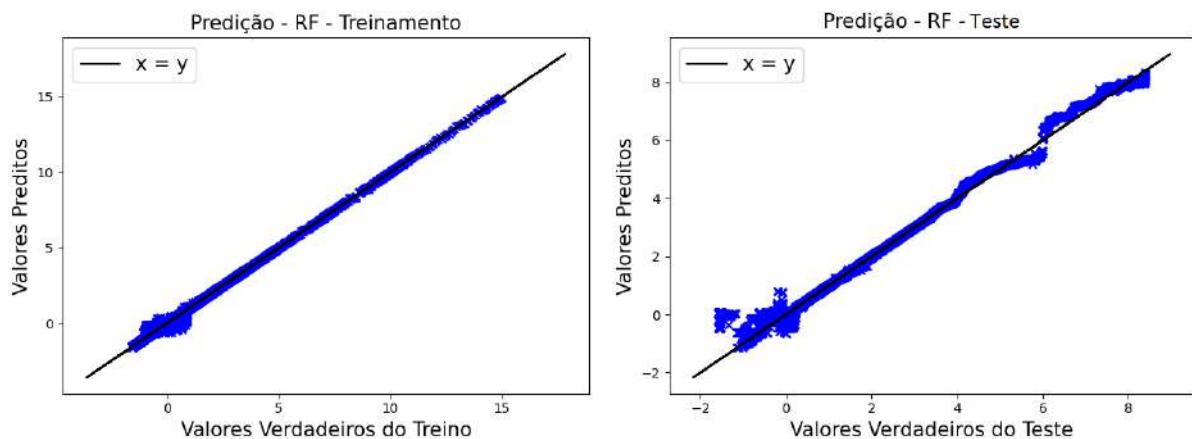
Pelos gráficos de predição, ficou evidente que o sensor GACID_NH3, bem como o GACID_H2S, demonstrou uma excelente capacidade de previsão dos dados, como indicado pelos elevados coeficientes de determinação R^2 de 0,9850 e 0,9867. Adicionalmente, os pequenos valores de MAE, MSE e RMSE sinalizaram que o modelo apresentou boa precisão individual das previsões.

Tabela 4.10: Métricas do sensor GACID_NH3 após avaliação com dados de treinamento e dados de teste.

Métricas	Treinamento	Teste
R^2	0,9850	0,9867
MAE	0,0584	0,1076
MSE	0,0150	0,0572
RMSE	0,1223	0,2392

Fonte: Elaboração própria.

Figuras 4.15 A e B: Predição do sensor GACID_NH3 usando RF com dados de treinamento (A) e teste (B).



Fonte: Elaboração própria.

Entretanto, o menor R^2 e maiores métricas de erro quando comparado ao sensor GACID_H2S, principalmente referente aos dados do banco de teste, mostraram a dificuldade encontrada pelo modelo para prever valores menores da fração mássica de NH_3 no gás ácido entre o intervalo de -1,28 a -0,69 ($6,00 \cdot 10^{-5}$ a $7,34 \cdot 10^{-5}$, em valores desnormalizados), como observado na Figura 4.15B.

Apesar dessa constatação, o sensor GACID_NH3, no geral, apresentou ótima capacidade de predição, de modo que o intervalo de confiabilidade para previsão da fração mássica de NH_3 no gás ácido ficou entre $6,00 \cdot 10^{-5}$ e $2,08 \cdot 10^{-4}$, após desnormalização dos dados. Nesse caso, portanto, a região considerada confiável compreende toda a faixa de predição observada na Figura 4.15B.

O valor da fração mássica de NH_3 no gás ácido na operação normal é igual a $8,19 \cdot 10^{-5}$, portanto o sensor seria capaz de prever variações entre -26,72% a 153,67% da operação normal.

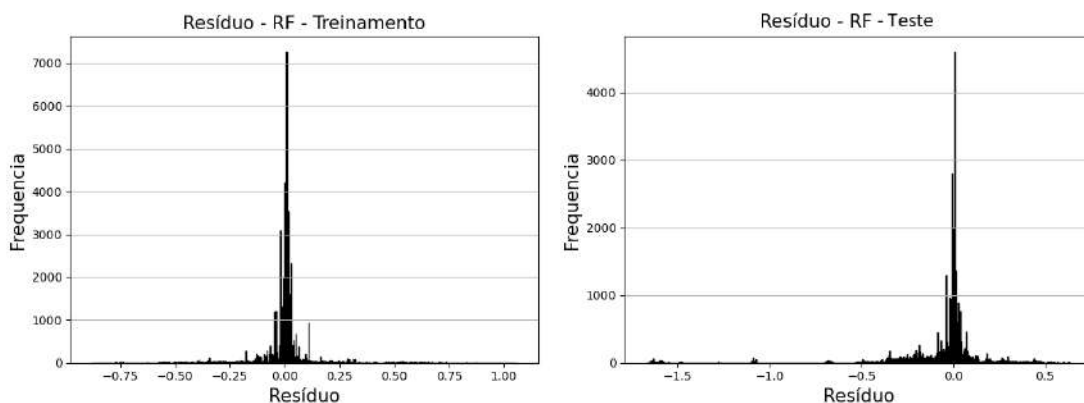
O intervalo confiável de previsão da fração de NH_3 no gás ácido é notavelmente mais amplo quando comparado ao intervalo da fração de H_2S , que varia entre -2,28% e 0,53% da operação normal. Esse comportamento é corroborado pelos resultados observados na Tabela 4.2, onde a fração de H_2S no gás ácido variou cerca de -2,17%, enquanto a fração de NH_3 variou por volta de 143,81% durante o aumento de temperatura do 2º estágio da coluna 1 de 120,3 até 141,9°C.

Isso destaca que, nas situações de operação normal avaliadas, a fração mássica de NH_3 no gás ácido é notavelmente mais sensível a perturbações, especialmente em relação à temperatura do segundo estágio da primeira coluna de esgotamento. Por outro lado, a fração mássica de H_2S no gás ácido é mais estável, apresentando variações menores na presença de distúrbios. Essa análise ressalta a importância relativa dessas duas variáveis no processo de separação dos contaminantes e fornece informações valiosas para o controle e otimização do processo.

Os gráficos de resíduos validam as conclusões previamente discutidas em relação ao desempenho do sensor. Isso porque os resíduos exibem uma distribuição uniforme em torno de zero, não apresentando quaisquer padrões que poderiam sugerir problemas no modelo, como heterocedasticidade que é a presença de variabilidade desigual ou não constante na dispersão (variância) dos erros do modelo de regressão, ou a presença de valores discrepantes

(*outliers*). Essa uniformidade nos resíduos reforça a qualidade do ajuste do modelo e a confiabilidade das previsões, fornecendo evidências adicionais de sua robustez e precisão.

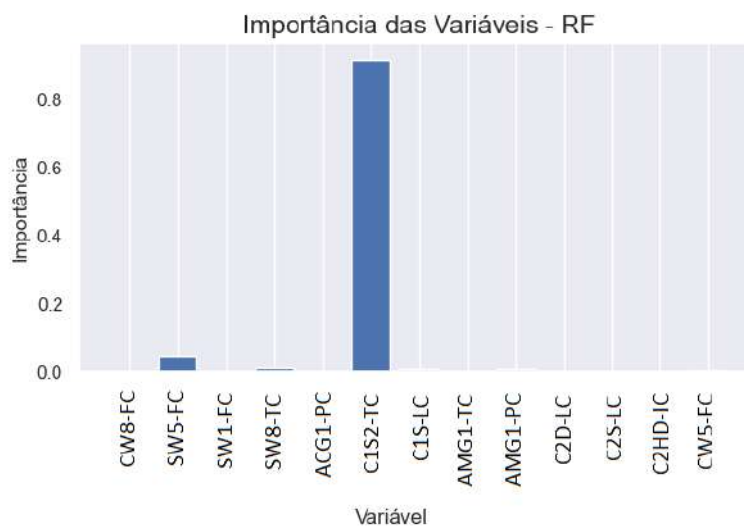
Figuras 4.16 A e B: Resíduos do sensor GACID_NH3 usando RF com dados de treinamento (A) e teste (B).



Fonte: Elaboração própria.

As variáveis classificadas como mais importantes para o sensor GACID_NH3 foram vazão mássica de SW5, temperatura de SW8 e temperatura do 2º estágio da coluna 1 (C1S2) como observado na Figura 4.17.

Figura 4.17: Classificação da importância das variáveis para o sensor GACID_NH3.



Fonte: Elaboração própria.

A redução das variáveis de entrada foi realizada e notou-se que a vazão mássica de SW5 e a temperatura do 2º estágio de C1 (C1S2) se mantiveram como variáveis mais importantes para o sensor GACID_NH3, entretanto a temperatura do 2º estágio de C1 apresentou contribuição muito mais relevante.

Após a exclusão das variáveis menos importantes, a análise da Tabela 4.11 revelou que as métricas do banco de teste não apresentaram mudanças relevantes: o R^2 variou cerca de -0,09%, o valor do MAE reduziu em 13,48% e os valores de MSE e RMSE aumentaram ligeiramente cerca de 6,64% e 3,22%, respectivamente. No entanto, é pertinente prosseguir com a remoção dessas variáveis, uma vez que isso contribuirá para melhorar o tempo de processamento computacional, sem comprometer o desempenho do modelo.

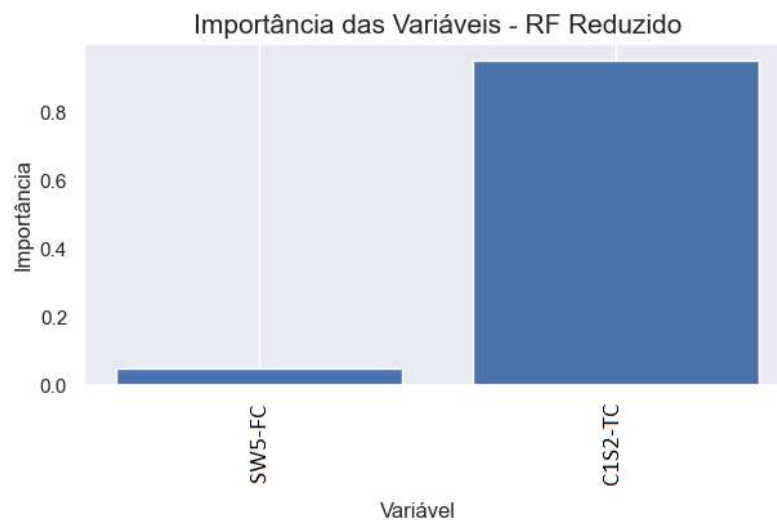
Os gráficos de predição e resíduo após a redução das variáveis de entrada foram suprimidos por serem análogos aos previamente apresentados, entretanto podem ser encontrados no Apêndice D.

Tabela 4.11: Métricas do banco de teste do sensor GACID_NH3 após redução das variáveis.

Resultado das métricas de teste após redução das variáveis	
R^2	0,9858
MAE	0,0931
MSE	0,0610
RMSE	0,2469

Fonte: Elaboração própria.

Figura 4.18: Classificação da importância das variáveis para o sensor GACID_NH3 após redução.



Fonte: Elaboração própria.

4.5.3 Sensor GAMON_H2S

No caso do sensor GAMON_H2S, observou-se que o modelo mais eficaz demonstrou ter a profundidade máxima das árvores (*rf_max_depth*). Além disso, o número de árvores da floresta (*rf_n_estimators*) foi próximo a um valor intermediário e o número mínimo de amostras para formar uma folha (nó final) (*rf_min_samples_leaf*) foi próximo ao limite inferior da faixa, conforme detalhado na Tabela 4.12.

Tabela 4.12: Hiperparâmetros otimizados para o sensor GAMON_H2S usando RF.

Hiperparâmetros	Valor ótimo	Valor Mínimo	Valor Máximo
<i>rf_min_samples_leaf</i>	3	1	50
<i>rf_max_depth</i>	10	5	10
<i>rf_n_estimators</i>	30	1	50

Fonte: Elaboração própria.

Para avaliação do sensor, os gráficos de predição podem ser encontrados nas Figuras 4.19 A e B, os gráficos de resíduos nas Figuras 4.22 A e B, e a Tabela 4.13 apresenta as métricas do modelo.

Inicialmente, ao analisar a tabela de métricas em conjunto com os gráficos de predição do sensor GAMON_H2S, nota-se que, diferentemente do ocorrido para os sensores do gás ácido, os resultados para o banco de treinamento foram bem diferentes dos encontrados para o banco de teste.

Apesar do valor do R^2 entre os dois bancos apresentar uma diferença de apenas 4,89%, os resultados de MAE, MSE e RMSE para o banco de teste apresentaram um aumento significativo de 71,20%, 90,28% e 68,82%, respectivamente, quando comparados ao banco de treinamento. Esse resultado é corroborado pelo gráfico de predição da Figura 4.19B onde se observou uma espécie de histerese dos dados, que não foi tão evidente na Figura 4.19A.

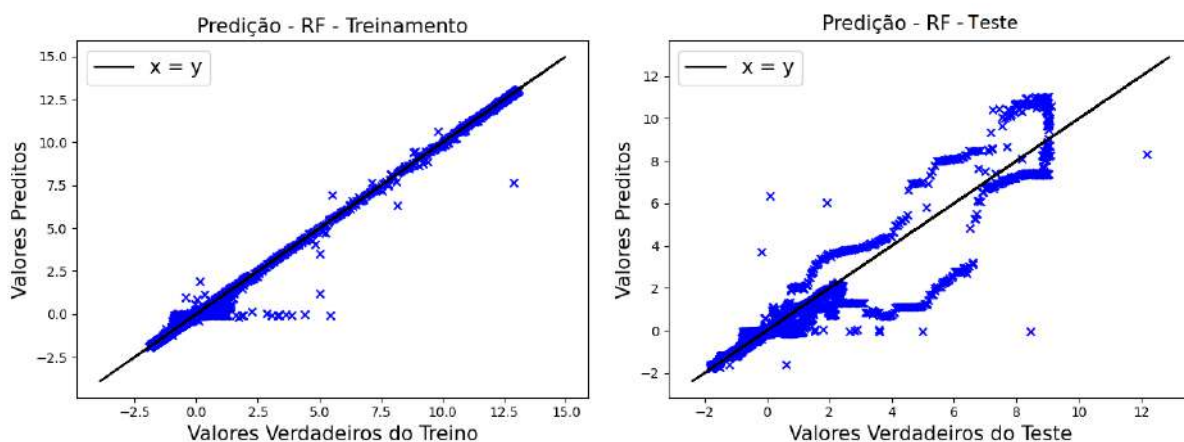
Esta histerese indicou que o modelo predisse diferentes valores de fração de H_2S no gás amoniacal para um mesmo valor verdadeiro presente no banco de teste. Por outro lado, o inverso também foi observado, em que um mesmo valor foi predito para valores verdadeiros distintos e distantes entre si.

Tabela 4.13: Métricas do sensor GAMON_H2S após avaliação com dados de treinamento e dados de teste.

Métricas	Treinamento	Teste
R ²	0,9926	0,9463
MAE	0,0220	0,0764
MSE	0,0074	0,0761
RMSE	0,0860	0,2758

Fonte: Elaboração própria.

Figuras 4.19 A e B: Predição do sensor GAMON_H2S usando RF com dados de treinamento (A) e teste (B).

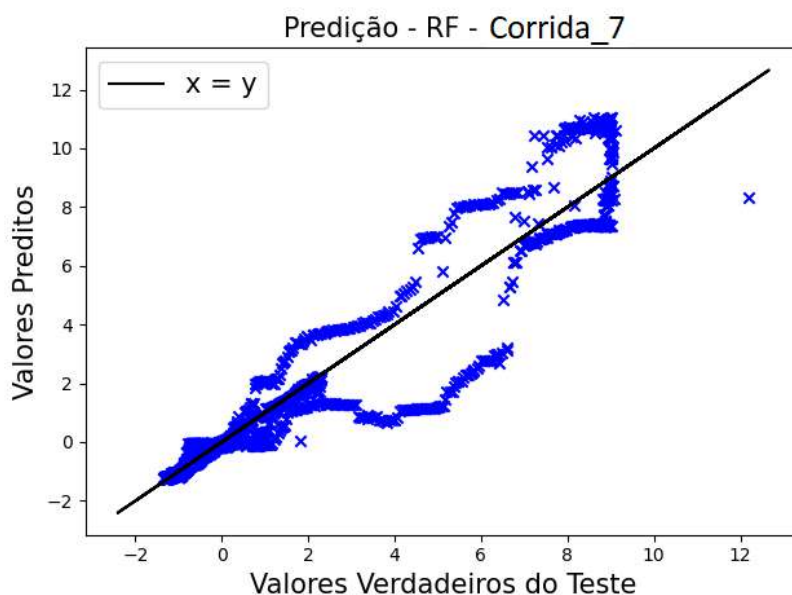


Fonte: Elaboração própria.

Com o objetivo de avaliar melhor este comportamento, cada uma das 12 simulações do banco de teste listadas na Tabela 4.5 foi testada separadamente, utilizando o melhor modelo já com os parâmetros otimizados para o sensor GAMON_H2S, a fim de avaliar a contribuição de cada uma delas para o modelo completo.

A corrida número 7 do banco de teste, que corresponde à corrida global com distúrbios positivos e negativos aleatórios na temperatura do 2º estágio de C1 – versão B, exibiu o comportamento apresentado na Figura 4.20. Os demais gráficos de predição para cada uma das corridas do banco de teste estão disponíveis em maior detalhe no Apêndice E.

Figura 4.20: Predição para corrida número 7 (banco de teste) do sensor GAMON_H2S.



Fonte: Elaboração própria.

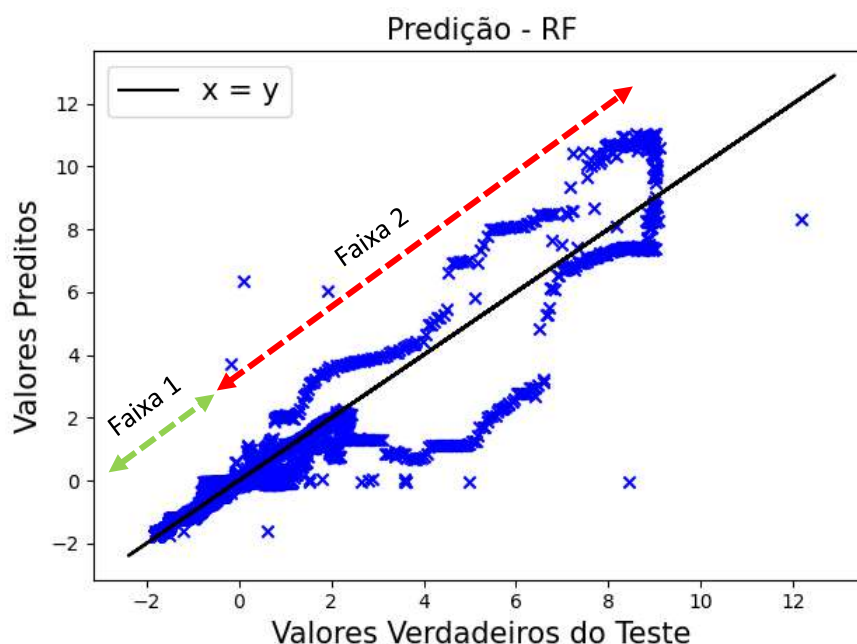
Portanto, ao analisar cada simulação isoladamente, ficou evidente que o conjunto de dados da Corrida 7 contribuiu para a histerese dos resultados, ocasionando previsões menos precisas na faixa de valores entre 0,86 e 12,17, conforme observado na Figura 4.19B.

No entanto, é importante observar que essa simulação em particular abrange apenas 11% do conjunto de dados de teste. Dessa forma, os 89% restante dos dados parecem ter contribuído para que o modelo apresentasse um valor elevado de R^2 igual a 0,9463, enquanto a dispersão causada pelos 11% dos dados da Corrida 7 foi evidenciada somente nas métricas de erro como MAE, MSE e RMSE.

Além disso, vale a pena ressaltar que, no banco de treinamento, a Corrida 8 é a simulação compatível com a Corrida 7 do banco de teste, onde foram realizados distúrbios positivos e negativos aleatórios na temperatura do 2º estágio de C1. Entretanto, no caso do banco de treinamento, o comportamento da histerese não foi tão evidente como observado na Figura 4.19A. Isso sugere que esse resultado pode estar relacionado a uma possível limitação do modelo empírico em relação aos dados de teste examinados, já que não foram identificadas relações substanciais entre as variáveis de entrada e de saída que pudessem explicar essa observação.

Diante disso, o intervalo de confiança para a previsão da fração mássica de H_2S no gás amoniacal foi dividido em duas faixas distintas como representado na Figura 4.21.

Figura 4.21: Diferentes faixas de confiabilidade para o sensor GAMON_H2S.



Fonte: Elaboração própria.

A Faixa 1, que é a faixa de confiança definida visualmente, abrange o intervalo de -1,95 a 0,86, correspondendo a valores da fração mássica de H₂S no gás amoniacal entre 0,0376 e 0,0842, após a desnormalização dos dados. Portanto, o sensor seria capaz de prever mais precisamente variações entre -45,19% a 22,74% do valor da operação normal que é igual a 0,0686.

A Faixa 2, que seria a faixa de menor confiabilidade do sensor, compreende o intervalo de 0,86 a 12,17, representando valores entre 0,0842 e 0,2799, após desnormalização dos dados. Portanto, variações entre 22,74% a 308,02% da operação normal da fração de H₂S no gás amoniacal não são capazes de serem tão bem previstas pelo sensor.

Como comentado anteriormente, o R² não foi uma métrica capaz de explicitar a histerese dos dados de teste tão eficientemente como as métricas de erro MAE, MSE e RMSE. Portanto, apenas as métricas de erro das novas faixas de confiabilidade serão comparadas com as do modelo usando o banco de teste completo.

A Faixa 1 apresentou um acréscimo de 31,67% no valor do MAE, porém uma diminuição de 64,26% no MSE e de 40,21% no RMSE, em comparação com os resultados do modelo de teste completo. Enquanto isso, a Faixa 2 testemunhou um notável aumento de 1385,47%, 2893,43% e 447,24% em relação aos valores de MAE, MSE e RMSE, respectivamente.

Esses resultados mostraram que a Faixa 1, apesar do aumento no MAE, apresentou valores de MSE e RMSE bem menores do que o modelo completo, podendo ser utilizada com mais confiança, enquanto a Faixa 2 claramente é uma região onde os resultados preditos não apresentam tanta confiabilidade.

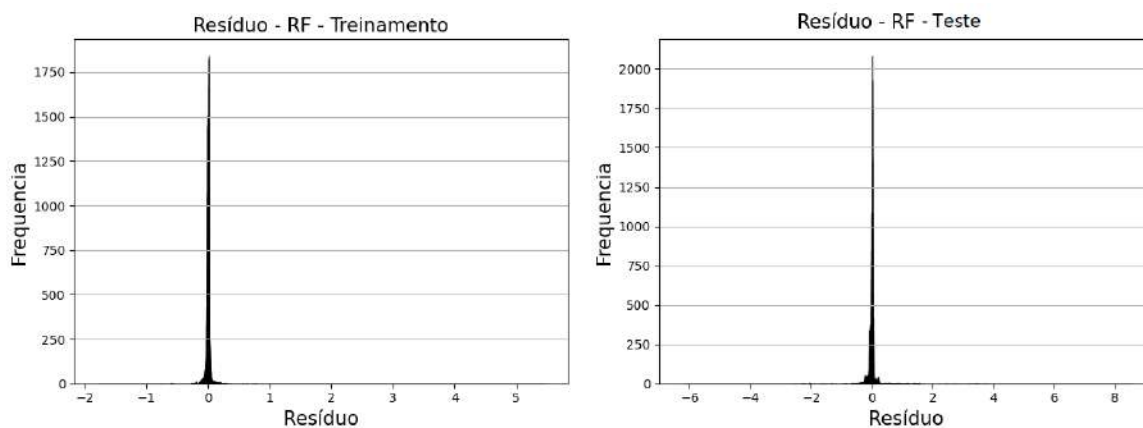
Tabela 4.14: Métricas calculadas para cada faixa de confiabilidade do sensor GAMON_H2S.

Faixa de confiabilidade 1		Faixa de confiabilidade 2	
R ²	0,8709	R ²	0,7687
MAE	0,1006	MAE	1,1349
MSE	0,0272	MSE	2,2780
RMSE	0,1649	RMSE	1,5093

Fonte: Elaboração própria.

Os gráficos de resíduos exibiram uma distribuição uniforme em torno de zero, não apresentando quaisquer padrões que poderiam sugerir problemas no modelo, como heterocedasticidade, ou a presença de valores discrepantes (*outliers*). É importante destacar que, devido ao fato de que a histerese representada no gráfico de previsão abrange apenas 11% do conjunto de dados de teste, esse padrão não se tornou perceptível no gráfico geral de resíduos. Isso ocorreu porque esse subconjunto de amostras é consideravelmente menor em comparação com o conjunto de dados de teste completo.

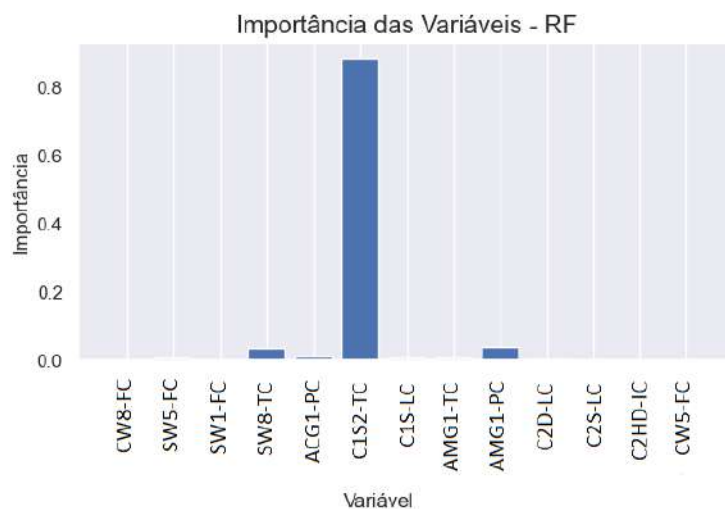
Figuras 4.22 A e B: Resíduos do sensor GAMON_H2S usando RF com dados de treinamento (A) e teste (B).



Fonte: Elaboração própria.

As variáveis classificadas como mais importantes para o sensor GAMON_H2S foram temperatura do 2º estágio da coluna 1 (C1S2), temperatura de SW8, pressão do gás amoniacal (AMG) e do gás ácido (ACG) como observado na Figura 4.23.

Figura 4.23: Classificação da importância das variáveis para o sensor GAMON_H2S.



Fonte: Elaboração própria.

Após redução das variáveis de entrada, a temperatura de SW8, a temperatura do 2º estágio de C1 (C1S2) e pressão do gás amoniacal (AMG) se mantiveram como importantes. Entretanto, a temperatura do 2º estágio de C1 (C1S2) apresentou contribuição muito mais relevante do que as demais.

Após a exclusão das variáveis menos importantes, a análise da Tabela 4.15 revelou que as métricas do banco de teste não apresentaram mudanças relevantes: o R^2 variou cerca de -1,08%, os valores de MAE, MSE e RMSE aumentaram por volta de 13,22%, 18,92% e 9,10%, respectivamente. É pertinente, portanto, prosseguir com a remoção das demais variáveis, uma vez que isso contribuirá para melhorar o tempo de processamento computacional, sem comprometer o desempenho do modelo.

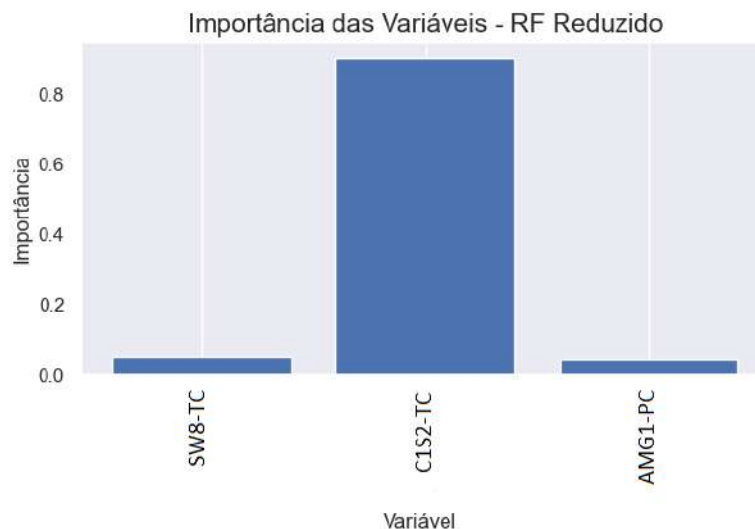
Os gráficos de predição e resíduo após a redução das variáveis de entrada foram suprimidos por serem análogos aos previamente apresentados, entretanto podem ser encontrados no Apêndice D.

Tabela 4.15: Métricas do banco de teste do sensor GAMON_H2S após redução das variáveis.

Resultado das métricas de teste após redução das variáveis	
R^2	0,9361
MAE	0,0865
MSE	0,0905
RMSE	0,3009

Fonte: Elaboração própria.

Figura 4.24: Classificação da importância das variáveis para o sensor GAMON_H2S após redução.



Fonte: Elaboração própria.

4.5.4 Sensor GAMON_NH3

No caso do sensor GAMON_NH3, observou-se que o modelo mais eficaz demonstrou ter a profundidade máxima das árvores (*rf_max_depth*). Além disso, apresentou o valor mínimo para o número de amostras necessárias para formar uma folha (nó final) (*rf_min_samples_leaf*) e um valor intermediário para o número de árvores (*rf_n_estimators*) da floresta, conforme detalhado na Tabela 4.16.

Tabela 4.16: Hiperparâmetros otimizados para o sensor GAMON_NH3 usando RF.

Hiperparâmetros	Valor ótimo	Valor Mínimo	Valor Máximo
<i>rf_min_samples_leaf</i>	1	1	50
<i>rf_max_depth</i>	10	5	10
<i>rf_n_estimators</i>	30	1	50

Fonte: Elaboração própria.

Para avaliação do sensor, os gráficos de predição podem ser encontrados nas Figuras 4.25 A e B, os gráficos de resíduos nas Figuras 4.28 A e B, e a Tabela 4.17 apresenta as métricas encontradas.

Assim como ocorreu para o caso do GAMON_H2S, os gráficos de predição para o sensor GAMON_NH3 apresentaram um comportamento consideravelmente diferente quando

comparado aos sensores do gás ácido: os resultados para o banco de treinamento contrastaram significativamente com aqueles encontrados para o banco de teste.

Além do valor do R^2 entre os dois bancos apresentar uma diferença de 14,30%, os resultados de MAE, MSE e RMSE para o banco de teste apresentaram um aumento significativo de 76,10%, 96,78% e 82,08%, respectivamente, quando comparados ao banco de treinamento. Esse resultado é corroborado pelo gráfico de predição da Figura 4.25B onde se observou uma espécie de histerese dos dados, que não foi tão evidente na Figura 4.25A.

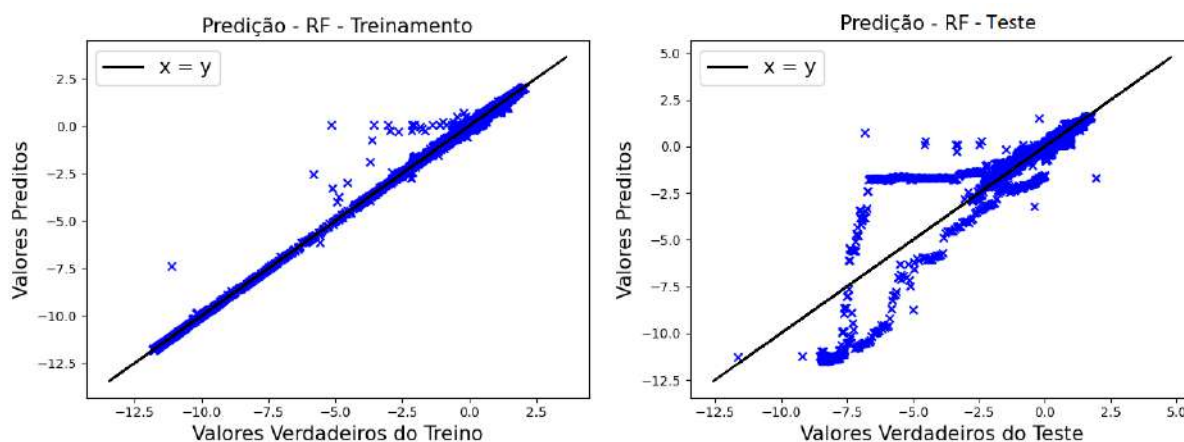
Esta histerese indicou que o modelo predisse diferentes valores de fração de NH_3 no gás amoniacal para um mesmo valor verdadeiro presente no banco de teste. Por outro lado, o inverso também foi observado, em que um mesmo valor foi predito para valores verdadeiros distintos e distantes entre si.

Tabela 4.17: Métricas do sensor GAMON_NH3 após avaliação com dados de treinamento e dados de teste.

Métricas	Treinamento	Teste
R^2	0,9946	0,8702
MAE	0,0283	0,1184
MSE	0,0054	0,1677
RMSE	0,0734	0,4096

Fonte: Elaboração própria.

Figuras 4.25 A e B: Predição do sensor GAMON_NH3 usando RF com dados de treinamento (A) e teste (B).



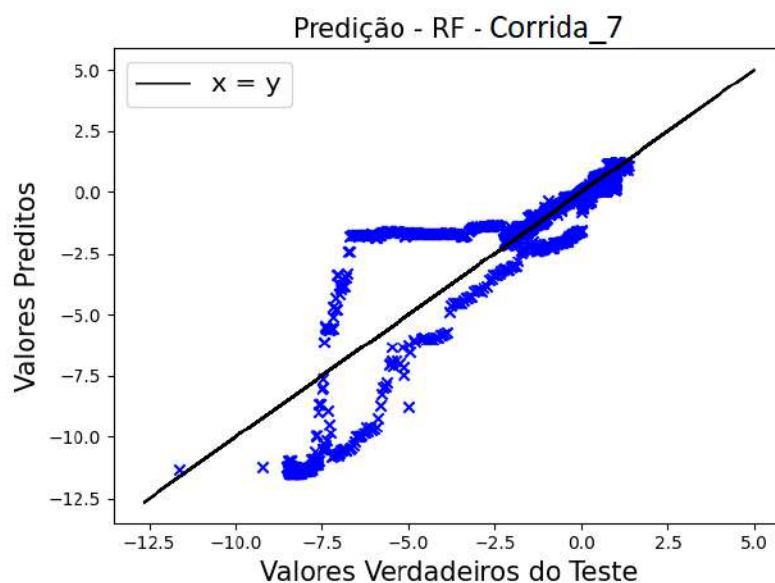
Fonte: Elaboração própria.

Com o intuito de aprofundar a avaliação desse comportamento, da mesma forma que foi feito com o sensor GAMON_H2S, cada uma das 12 simulações do banco de teste foi

examinada individualmente. O objetivo era esclarecer a contribuição de cada simulação para o modelo completo.

A corrida número 7 do banco de teste, que corresponde à corrida global com distúrbios positivos e negativos aleatórios na temperatura do 2º estágio de C1 – versão B, exibiu o comportamento apresentado na Figura 4.26. Os demais gráficos de predição para cada uma das corridas presentes no banco de teste estão disponíveis em maior detalhe no Apêndice E.

Figura 4.26: Predição para corrida número 7 (banco de teste) do sensor GAMON_NH3.



Fonte: Elaboração própria.

Assim como anteriormente, ficou evidente que o conjunto de dados da Corrida 7 contribuiu para a histerese dos resultados, resultando em previsões menos precisas na faixa de valores de -11,56 a -1,89, conforme observado na Figura 4.25B.

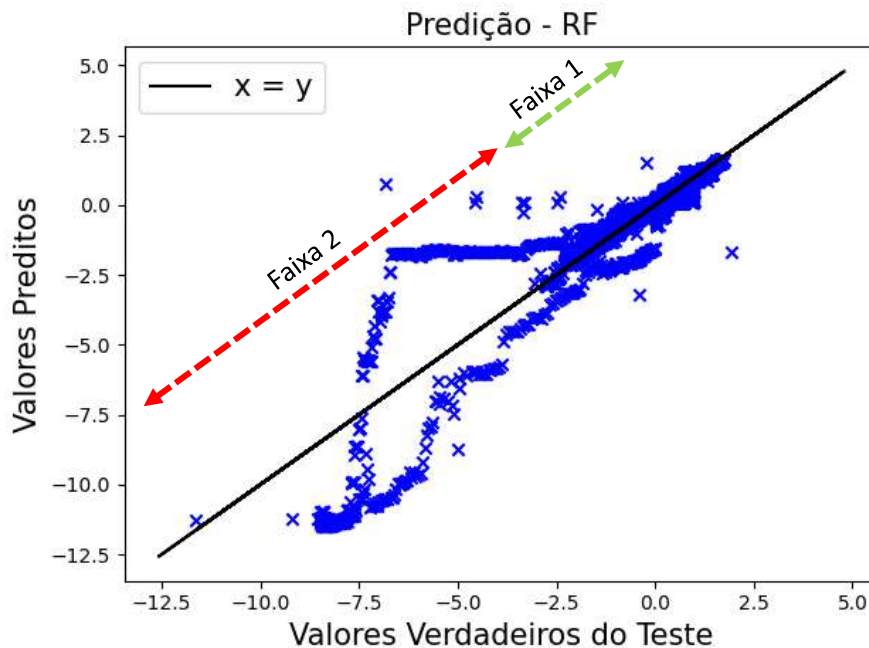
Como já ressaltado, essa simulação em particular abrange apenas 11% do conjunto de teste global, no entanto, estes dados influenciaram diretamente no valor do R^2 assim como nas métricas de erro como MAE, MSE e RMSE obtidas para o sensor GAMON_NH3, resultando no menor R^2 dentre todos os sensores, após avaliação com o banco de teste.

Além disso, vale a pena ressaltar que, no banco de treinamento, a Corrida 8 é a simulação compatível com a Corrida 7 do banco de teste, onde foram realizados distúrbios positivos e negativos aleatórios na temperatura do 2º estágio de C1. Entretanto, no caso do banco de treinamento, o comportamento da histerese não foi tão evidente como observado na Figura 4.25A. Isso sugere que esse resultado pode estar relacionado a uma possível limitação

do modelo empírico em relação aos dados examinados, já que não foram identificadas relações substanciais entre as variáveis de entrada e de saída que pudessem explicar essa observação.

Diante disso, o intervalo de confiança para a previsão da fração mássica de NH_3 no gás amoniacal foi dividido em duas faixas distintas como representado na Figura 4.27.

Figura 4.27: Diferentes faixas de confiabilidade para o sensor GAMON_NH3.



Fonte: Elaboração própria.

A Faixa 1, que é a faixa de confiança definida visualmente, compreende o intervalo -1,89 a 1,93, representando valores entre 0,4507 e 0,5061, após desnormalização dos dados. Portanto, o sensor seria capaz de prever mais precisamente variações entre -6,12% a 5,42% do valor da operação normal da fração de NH_3 no gás amoniacal que é igual a 0,4801.

A Faixa 2, que seria a faixa de menor confiabilidade do sensor, abrange o intervalo de -11,56 a -1,89, correspondendo a valores da fração mássica de NH_3 no gás amoniacal entre 0,3143 e 0,4507, após a desnormalização dos dados. Portanto, variações entre -34,53% a -6,12% da operação normal da fração de NH_3 no gás amoniacal não são capazes de serem tão bem preditas pelo sensor.

Da mesma maneira que ocorreu para o sensor GAMON_H2S, acredita-se que as melhores métricas para avaliar os resultados das faixas de confiabilidade são os valores de MAE, MSE e RMSE.

A Faixa 1 apresentou um acréscimo de 86,15% no valor do MAE, porém uma diminuição de 36,67% no MSE e de 20,44% no RMSE, em comparação com os resultados do modelo de teste completo. Enquanto isso, a Faixa 2 testemunhou um notável aumento de 1899,32%, 4212,76% e 556,57% em relação aos valores de MAE, MSE e RMSE, respectivamente.

Esses resultados mostraram que a Faixa 1, apesar do aumento no MAE, apresentou valores de MSE e RMSE bem menores do que o modelo completo, podendo ser utilizada com mais confiança. Enquanto a Faixa 2 claramente é uma região onde os resultados preditos não apresentam tanta confiabilidade.

Tabela 4.18: Métricas calculadas para cada faixa de confiabilidade do sensor GAMON_NH3.

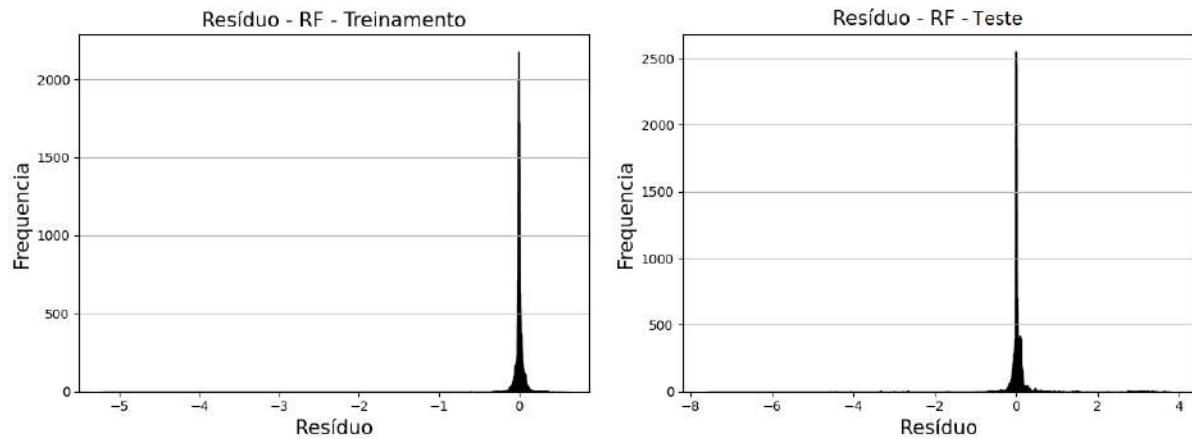
Faixa de confiabilidade 1		Faixa de confiabilidade 2	
R ²	0,7538	R ²	-0,2120
MAE	0,2204	MAE	2,3672
MSE	0,1062	MSE	7,2325
RMSE	0,3259	RMSE	2,6893

Fonte: Elaboração própria.

Os gráficos de resíduos revelaram uma distribuição uniforme, não mostrando qualquer padrão ou a presença de valores discrepantes (*outliers*). Essa uniformidade nos resíduos reforça a qualidade do ajuste do modelo e a confiabilidade das previsões.

É importante destacar que, devido ao fato de que a histerese representada no gráfico de previsão abrange apenas 11% do conjunto de dados de teste, esse padrão não se tornou perceptível no gráfico geral de resíduos. Isso ocorreu porque esse subconjunto de amostras é consideravelmente menor em comparação com o conjunto de dados de teste completo.

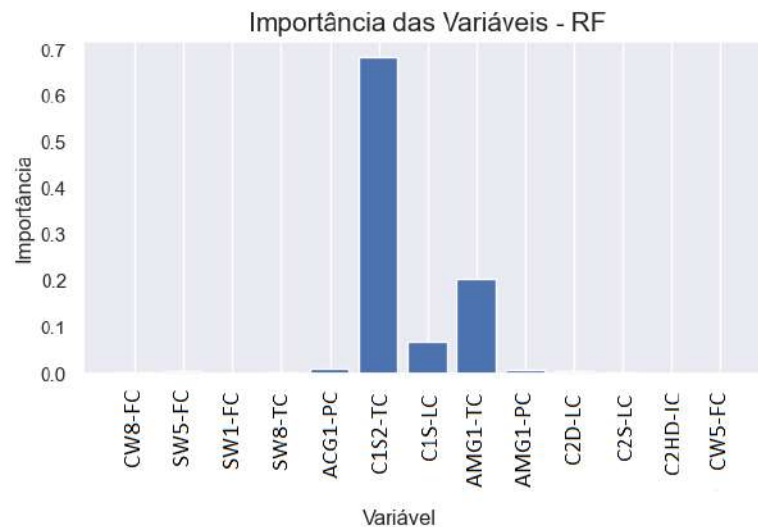
Figuras 4.28 A e B: Resíduos do sensor GAMON_NH3 usando RF com dados de treinamento (A) e teste (B).



Fonte: Elaboração própria.

As variáveis classificadas como mais importantes para o sensor GAMON_NH3 foram pressão do gás ácido (ACG), temperatura do 2º estágio da coluna 1 (C1S2), nível de fundo de C1 (C1S), temperatura e pressão do gás amoniacal (AMG), como observado na Figura 4.29.

Figura 4.29: Classificação da importância das variáveis para o sensor GAMON_NH3.



Fonte: Elaboração própria.

Após redução das variáveis de entrada, notou-se que a temperatura do 2º estágio de C1 (C1S2), o nível de fundo de C1 (C1S) e a temperatura do gás amoniacal (AMG) se mantiveram como variáveis mais importantes. Entretanto a temperatura do 2º estágio de C1 (C1S2) apresentou contribuição muito mais relevante conforme Figura 4.30.

Após a exclusão das variáveis menos importantes, a análise da Tabela 4.19 revelou que as métricas do banco de teste não apresentaram mudanças relevantes: o R^2 variou cerca de

-0,46%, o valor do MAE reduziu cerca de 10,72% e os valores de MSE e RMSE aumentaram por volta de 3,10% e 1,54%, respectivamente. No entanto, é pertinente prosseguir com a remoção das demais variáveis, uma vez que isso contribuirá para melhorar o tempo de processamento computacional, sem comprometer o desempenho do modelo.

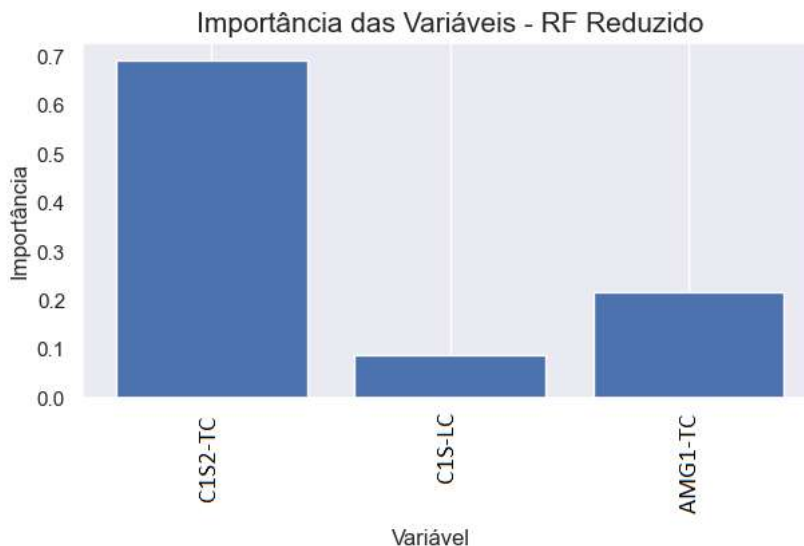
Os gráficos de predição e resíduo após a redução das variáveis de entrada foram suprimidos por serem análogos aos previamente apresentados, entretanto podem ser encontrados no Apêndice D.

Tabela 4.19: Métricas do banco de teste do sensor GAMON_NH3 após redução das variáveis.

Resultado das métricas de teste após redução das variáveis	
R ²	0,8662
MAE	0,1057
MSE	0,1729
RMSE	0,4159

Fonte: Elaboração própria.

Figura 4.30: Classificação da importância das variáveis para o sensor GAMON_NH3 após redução.



Fonte: Elaboração própria.

4.5.5. Sensor AGUA_H2S

Para o sensor AGUA_H2S, o melhor modelo apresentou a maior profundidade para as árvores (*rf_max_depth*) e o valor mínimo para o número de amostras para separar um nó

(*rf_min_samples_leaf*). Por outro lado, apresentou valores intermediários para o número de árvores da floresta (*rf_n_estimators*) como indicado na Tabela 4.20.

Tabela 4.20: Hiperparâmetros otimizados para o sensor AGUA_H2S usando RF.

Hiperparâmetros	Valor ótimo	Valor Mínimo	Valor Máximo
<i>rf_min_samples_leaf</i>	1	1	50
<i>rf_max_depth</i>	10	5	10
<i>rf_n_estimators</i>	34	1	50

Fonte: Elaboração própria.

Para avaliação do sensor, os gráficos de predição podem ser encontrados nas Figuras 4.31 A e B, os gráficos de resíduos nas Figuras 4.32 A e B, e a Tabela 4.21 apresenta as métricas do modelo.

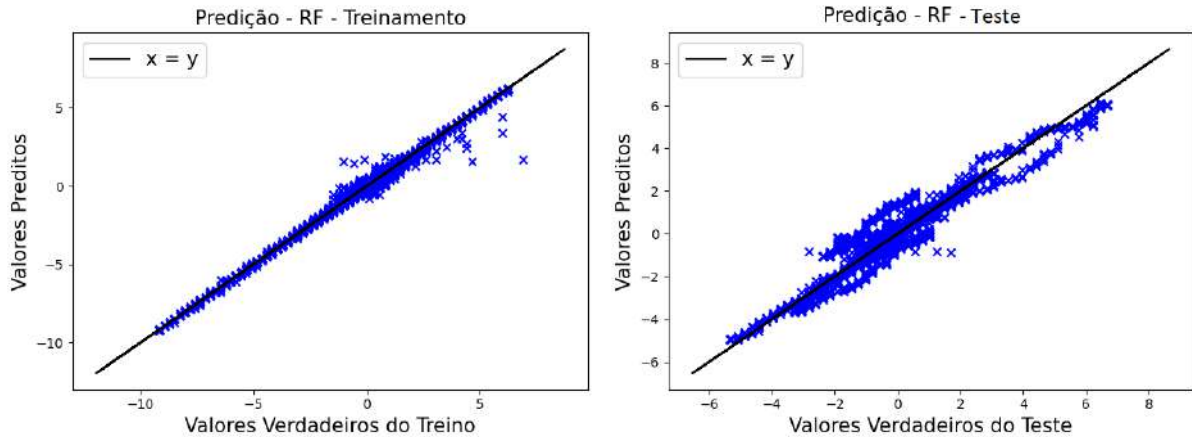
Os gráficos de predição para este sensor relevaram ter uma qualidade de predição razoável quando comparada aos sensores do gás ácido, que foram os melhores modelos desenvolvidos. O valor do R^2 para o banco de treinamento se mostrou ligeiramente superior ao dos sensores de gás ácido, enquanto o do banco de teste foi discretamente inferior. Por outro lado, os resultados de erro foram comparáveis aos do sensor GACID_NH3.

Tabela 4.21: Métricas do sensor AGUA_H2S após avaliação com dados de treinamento e dados de teste.

Métricas	Treinamento	Teste
R^2	0,9913	0,9161
MAE	0,0365	0,1095
MSE	0,0087	0,0590
RMSE	0,0934	0,2429

Fonte: Elaboração própria.

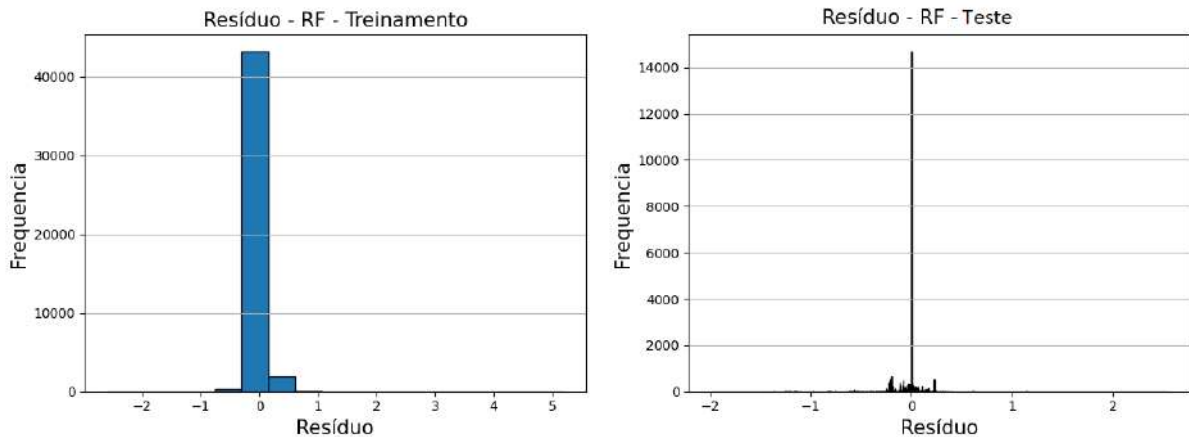
Figuras 4.31 A e B: Predição do sensor AGUA_H2S usando RF com dados de treinamento (A) e teste (B).



Fonte: Elaboração própria.

Os gráficos de resíduos validaram a robustez do sensor AGUA_H2S e sua capacidade de fazer previsões precisas. No conjunto de treinamento, mais de 40.000 amostras (87,9%) das 45.483 foram previstas com um resíduo de 0, o que significa que o valor previsto coincidiu exatamente com o valor real. De forma semelhante, no conjunto de teste, 14.000 amostras (44,4%) das 31.525 foram previstas de forma exata.

Figuras 4.32 A e B: Resíduos do sensor AGUA_H2S usando RF com dados de treinamento (A) e teste (B).



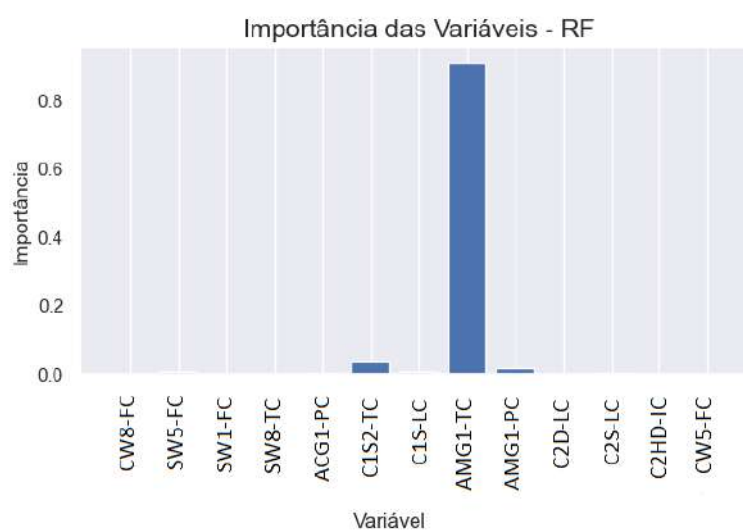
Fonte: Elaboração própria.

Baseado nas métricas e nos gráficos de predição e resíduo, o intervalo de confiabilidade para previsão da fração mássica de H₂S na água tratada ficou entre $2,17 \cdot 10^{-5}$ e $1,64 \cdot 10^{-5}$, após desnormalização dos dados. Nesse caso, portanto, a região considerada confiável compreende toda a faixa de predição observada na Figura 4.31B.

O valor da fração mássica de H₂S na água tratada em operação normal é igual a $1,88 \cdot 10^{-5}$, dessa forma o sensor seria capaz de prever variações entre -12,78% a 15,29% da operação normal.

As variáveis classificadas como mais importantes para o sensor AGUA_H2S foram pressão e temperatura do gás amoniacal (AMG) e temperatura do 2º estágio da coluna 1 (C1S2) como observado na Figura 4.33. As demais variáveis possuem uma contribuição tão pequena que não foram representadas no gráfico.

Figura 4.33: Classificação da importância das variáveis para o sensor AGUA_H2S.



Fonte: Elaboração própria.

A redução das variáveis de entrada foi realizada excluindo todas as que apresentaram importância menor que 2%. Notou-se que a temperatura do gás amoniacal (AMG) e a temperatura do 2º estágio da coluna 1 (C1S2) se mantiveram como variáveis mais importantes, sendo a temperatura do gás amoniacal a variável mais relevante.

A redução do conjunto de dados de interesse teve um efeito positivo nas métricas de desempenho do sensor, refletindo em um leve aumento de 0,7% no coeficiente de determinação R² e uma discreta diminuição dos valores de MAE, MSE e RMSE: 9,41%, 7,29% e 3,71%, respectivamente, como demonstrado na Tabela 4.22. Essa otimização do conjunto de dados contribui para aprimorar a capacidade preditiva do modelo.

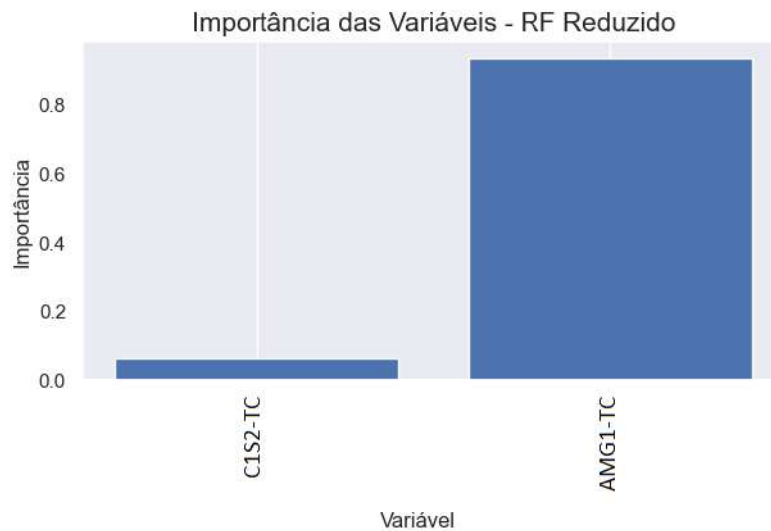
Os gráficos de predição e resíduo após a redução das variáveis de entrada foram suprimidos por serem análogos aos previamente apresentados, entretanto podem ser encontrados no Apêndice D.

Tabela 4.22: Métricas do banco de teste do sensor AGUA_H2S após redução das variáveis.

Resultado das métricas de teste após redução das variáveis	
R ²	0,9222
MAE	0,0992
MSE	0,0547
RMSE	0,2339

Fonte: Elaboração própria.

Figura 4.34: Classificação da importância das variáveis para o sensor AGUA_H2S após redução.



Fonte: Elaboração própria.

4.5.6 Sensor AGUA_NH3

Após treinamento e teste do sensor AGUA_NH3, foi verificado que o melhor modelo apresentou o maior valor possível para a profundidade das árvores (*rf_max_depth*) como todos os anteriores, enquanto apresentou valor próximo do limite máximo para o número de árvores (*rf_n_estimators*) e valor próximo do limite inferior para o número mínimo de amostras para separar um nó (*rf_min_samples_leaf*) como indicado na Tabela 4.23.

Tabela 4.23: Hiperparâmetros otimizados para o sensor AGUA_NH3 usando RF.

Hiperparâmetros	Valor ótimo	Valor Mínimo	Valor Máximo
<i>rf_min_samples_leaf</i>	2	1	50
<i>rf_max_depth</i>	10	5	10
<i>rf_n_estimators</i>	47	1	50

Fonte: Elaboração própria.

Para avaliação deste sensor, os gráficos de predição podem ser encontrados nas Figuras 4.35 A e B, os gráficos de resíduos nas Figuras 4.36 A e B, e a Tabela 4.24 apresenta as métricas dos modelos.

A análise do sensor AGUA_NH3 revelou um coeficiente de determinação (R^2) do banco de teste ligeiramente inferior (3,62%) em comparação com seu homólogo, o sensor AGUA_H2S. No entanto, nota-se que as métricas de erro, como o MAE, MSE e RMSE, apresentaram resultados, na maior parte dos casos, mais favoráveis para o sensor AGUA_NH3 tanto no conjunto de dados de treinamento como no de teste.

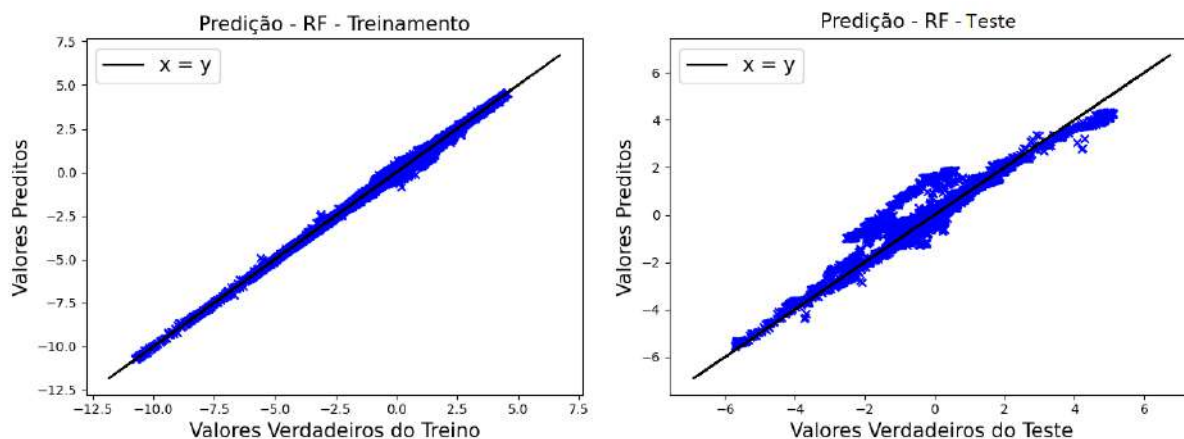
Esses resultados sugerem que, embora o sensor AGUA_NH3 tenha uma capacidade relativamente menor de explicar a variação dos dados de entrada, suas previsões são mais precisas em termos de proximidade dos valores reais.

Tabela 4.24: Métricas do sensor AGUA_NH3 após avaliação com dados de treinamento e dados de teste.

Métricas	Treinamento	Teste
R^2	0,9948	0,8829
MAE	0,0426	0,0937
MSE	0,0052	0,0517
RMSE	0,0723	0,2273

Fonte: Elaboração própria.

Figuras 4.35 A e B: Predição do sensor AGUA_NH3 usando RF com dados de treinamento (A) e teste (B).

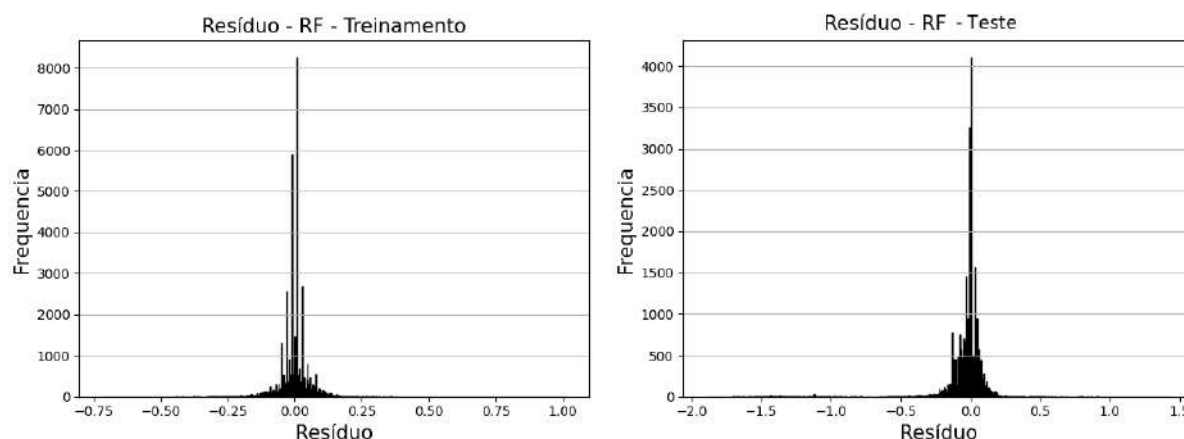


Fonte: Elaboração própria.

Os gráficos de resíduos complementam as conclusões previamente mencionadas sobre o desempenho do sensor, uma vez que os resíduos exibem uma distribuição uniforme em torno do zero e a ausência de padrões discerníveis nos resíduos sugere que o modelo não apresenta vieses sistemáticos que afetariam suas previsões.

Além disso, a falta de valores atípicos nos resíduos indica que não há observações extremamente discrepantes que poderiam distorcer a análise. Isso fortalece a confiança na capacidade do modelo de fazer previsões consistentes e confiáveis para os dados em questão.

Figuras 4.36 A e B: Resíduos do sensor AGUA_NH3 usando RF com dados de treinamento (A) e teste (B).



Fonte: Elaboração própria.

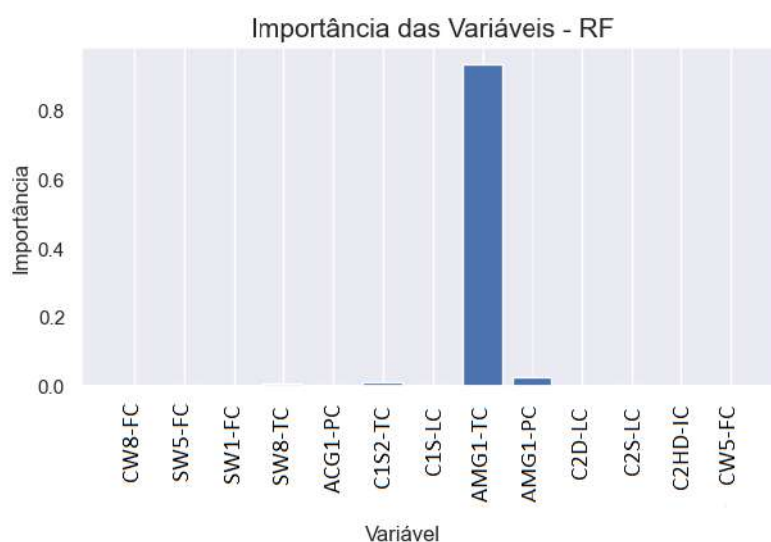
Pelo desempenho geral do sensor, o intervalo de confiabilidade para previsão da fração mássica de NH_3 na água tratada ficou entre $2,66 \cdot 10^{-4}$ a $3,21 \cdot 10^{-4}$, após desnormalização

dos dados. Nesse caso, portanto, a região considerada confiável compreende toda a faixa de predição observada na Figura 4.35B.

O valor da fração mássica de NH_3 na água tratada na operação normal é igual a $2,96 \cdot 10^{-4}$, dessa forma o sensor seria capaz de prever variações entre -10,08% a 8,56% da operação normal.

As variáveis classificadas como mais importantes para o sensor AGUA_NH3 foram temperatura e pressão do gás amoniacal (AMG) e temperatura do 2º estágio da coluna 1 (C1S2) como observado na Figura 4.37. As demais variáveis possuem contribuições muito menores, não sendo representadas no gráfico.

Figura 4.37: Classificação da importância das variáveis para o sensor AGUA_NH3.



Fonte: Elaboração própria.

Após redução das variáveis de entrada, apenas a temperatura e pressão do gás amoniacal (AMG) se mantiveram como variáveis mais importantes para o sensor AGUA_NH3. Sendo a temperatura do gás amoniacal muito mais relevante como evidenciado na Figura 4.38.

No caso deste sensor em específico, a redução do conjunto de dados de interesse teve um efeito negativo nas métricas de desempenho do sensor conforme apresentado na Tabela 4.25. O coeficiente de determinação R^2 apresentou uma redução de 12,80% em relação ao valor anterior à redução das variáveis, e as métricas de erro como MAE, MSE e RMSE aumentaram cerca de 90,72%, 96,32% e 40,21%, respectivamente. Sendo assim, diferente dos

demais sensores, o AGUA_NH3 requer o uso do conjunto de variáveis de entrada original para poder manter a confiabilidade dos resultados.

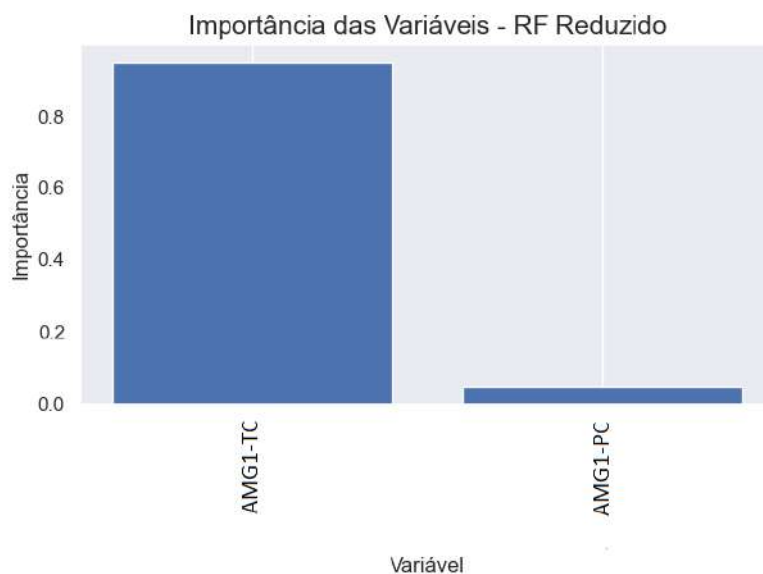
Os gráficos de predição e resíduo após a redução das variáveis de entrada foram suprimidos por serem análogos aos previamente apresentados, entretanto podem ser encontrados no Apêndice D.

Tabela 4.25: Métricas do banco de teste do sensor AGUA_NH3 após redução das variáveis.

Resultado das métricas de teste após redução das variáveis	
R ²	0,7699
MAE	0,1787
MSE	0,1015
RMSE	0,3187

Fonte: Elaboração própria.

Figura 4.38: Classificação da importância das variáveis para o sensor AGUA_NH3 após redução.



Fonte: Elaboração própria.

4.6 Análise global da importância das variáveis e estudo fenomenológico do processo

Após a redução do conjunto de dados de entrada cujas variáveis apresentaram importância abaixo de 2%, avaliou-se as variáveis mais relevantes para cada um dos sensores

e, com base nisso, criou-se a Tabela 4.26. Esse foi o critério utilizado para balizar todas as informações, embora o sensor AGUA_NH3 tenha demonstrado anteriormente que requer todo o conjunto de dados de entrada para um desempenho ótimo.

A tabela criada permitiu a realização de uma análise abrangente da importância das variáveis para os seis sensores. Essa tabela oferece uma visão global da contribuição de cada variável, identificando aquelas que desempenharam um papel mais significativo na simulação. Essa análise é fundamental para embasar decisões e possíveis ajustes posteriores no processo, visando aprimorar sua eficácia.

Tabela 4.26: Análise global da importância das variáveis de todos os sensores após redução.

Sensores	CW8-FC	SW5-FC	SW1-FC	SW8-TC	ACG1-PC	C1S2-TC	C1S-LC	AMG1-TC	AMG1-PC	C2D-LC	C2S-LC	C2HD-IC	CW5-FC
GACID_H2S		X		X		X							
GACID_NH3		X				X							
GAMON_H2S				X		X			X				
GAMON_NH3						X	X	X					
AGUA_H2S						X		X					
AGUA_NH3								X	X				

Fonte: Elaboração própria.

Os resultados encontrados na Tabela 4.26 reforçam que o processo de remoção dos contaminantes H₂S e NH₃ de águas ácidas é totalmente dependente da temperatura. Todas as medições de temperatura na simulação – vazão de entrada na primeira coluna (SW8), 2º estágio da Coluna 1 (C1S2) e gás amoniacal (AMG) – apareceram como variáveis importantes entre os seis sensores virtuais, indicando o papel fundamental desta para o controle do processo termodinâmico de separação do H₂S e NH₃.

Dentre as variáveis mencionadas, a temperatura do 2º estágio na Coluna 1 (C1S2) foi identificada como a mais significativa em cinco dos seis sensores virtuais. Esse resultado enfatiza que esta é uma variável crítica do processo, devendo ser monitorada constantemente, pois impacta tanto na recuperação de H₂S no gás ácido da primeira coluna como consequentemente na recuperação de NH₃ no gás amoniacal da segunda coluna.

Além disso, corrobora com o resultado encontrado no Tópico 4.1, onde os resultados mostraram a necessidade de um controle efetivo sobre a carga térmica da primeira coluna, a fim de evitar uma sobre especificação da eficiência de recuperação do H₂S que além de causar

um gasto energético desnecessário no refeedor, pode levar a problemas operacionais na URE devido ao aumento do teor de NH_3 no gás ácido.

Para ajuste da temperatura do 2º estágio de C1, o controlador C1S2-TC deverá atuar na carga térmica do refeedor de C1, que é a variável manipulada. Consequentemente, é coerente que a vazão mássica de SW5 (corrente que entra no 1º estágio de C1) e a temperatura de SW8 (corrente que entra no 2º estágio de C1) sejam consideradas importantes, uma vez que a quantidade de massa e a temperatura das correntes que são tratadas na coluna afetará diretamente a carga térmica do refeedor de C1, pois este irá trabalhar para manter o equilíbrio termodinâmico dentro da coluna.

Avaliando os efluentes da Coluna 2 (gás amoniacal e água tratada), nota-se que além da grande influência da temperatura representada pela temperatura do gás amoniacal (AMG) e do segundo estágio de C1 (C1S2), a pressão do gás amoniacal (AMG) apareceu como variável importante tanto no sensor GAMON_H2S como no AGUA_NH3. O primeiro mede a fração de H_2S no gás ácido e o segundo, a fração de NH_3 na água tratada.

Uma explicação plausível está relacionada à necessidade de manter a pressão na Coluna 2 inferior à da Coluna 1, com o objetivo de facilitar a remoção seletiva do NH_3 no gás amoniacal. Isso, por sua vez, resulta em níveis mais baixos de NH_3 na água tratada. Qualquer modificação na pressão poderia potencialmente levar a um aumento tanto na fração de H_2S no gás amoniacal quanto na fração de NH_3 na água tratada, o que claramente não é desejável para o processo.

Um ponto importante de ressaltar é o aparecimento do nível de fundo da Coluna 1 (C1S) como variável importante para o sensor GAMON_NH3. O controlador associado a essa variável, C1S-LC, atua diretamente na porcentagem de abertura da válvula V6 que controla a vazão de água pré-tratada que entra na Coluna 2. Conforme mencionado anteriormente, a quantidade de massa presente dentro da coluna exerce uma influência significativa sobre a carga térmica necessária para manter o equilíbrio termodinâmico de separação. Isso é fundamental para atingir a concentração máxima de NH_3 no gás amoniacal, que representa o principal objetivo da operação da Coluna 2.

Por fim, é importante destacar que, dentre as 13 variáveis analisadas, somente 6 foram consideradas mais importantes para os seis sensores virtuais desenvolvidos. Essa observação sugere que a modelagem de futuros trabalhos poderia ser simplificada utilizando apenas as 6

variáveis essenciais como entradas principais, a fim de otimizar recursos e buscar melhorias nos resultados alcançados.

Capítulo 5

5. Conclusões e Considerações Finais

O presente trabalho desenvolveu um novo banco de dados baseado na simulação de uma UTAA com duas colunas de esgotamento desenvolvida em *Aspen Plus Dynamics*® V10 por Nogueira (2021) após inclusão de seis novos controladores nos efluentes analisados: gás ácido, gás amoniacal e água tratada. Este novo banco é formado por 30 simulações independentes que representam mais de 77 mil amostras.

Para garantir uma boa variabilidade dos dados, diversos tipos de distúrbios foram testados em cinco variáveis de entrada: vazão mássica de SW1, vazão mássica de SW5, temperatura de SW1 e frações mássicas de H₂S e NH₃ em SW1, garantindo apenas que a simulação se mantivesse dentro da condição normal de operação.

Concluiu-se que a simulação dinâmica tende a se manter bastante estável com pouca variação, principalmente na fração de H₂S do gás ácido, e que a variável mais influente para este parâmetro é a temperatura do segundo estágio da Coluna 1. Foi também observado que o aumento desta temperatura melhora a recuperação do H₂S no gás ácido.

No entanto, incrementos adicionais na temperatura da Coluna 1, após o sistema atingir 90% de recuperação conforme exigido pela legislação ambiental brasileira, resultam em ganhos mínimos na eficiência de remoção do H₂S e na volatilização acentuada de NH₃ no gás ácido, o que pode levar a problemas operacionais na URE. Isso destaca a importância do controle preciso da carga térmica na primeira coluna para otimizar o processo e a possibilidade de usar sensores virtuais para analisar os efluentes de uma Unidade de Tratamento de Água Ácida.

O estudo dos métodos de integração relevou que o integrador padrão *Implicit Euler* era o mais adequado quando comparado às demais opções disponíveis: *Runge Kutta 4* e *Gear*. Os erros de integração observados puderam ser atribuídos à complexidade das muitas equações diferenciais que o *software* precisa resolver, principalmente nas duas colunas de esgotamento da simulação.

A análise comparativa dos algoritmos RF, GB e SVM revelou que os dois primeiros são equivalentes, exibindo métricas bastante parecidas, ao passo que o método SVM com *kernel* RBF apresentou o pior desempenho. Adicionalmente, o gráfico de predição do sensor

GACID_H2S utilizando o método SVM com *kernel* linear evidenciou a sua limitação em explicar os dados. Portanto, este não foi aprofundado e o RF se manteve como metodologia de AI padrão para desenvolvimento dos sensores virtuais.

Os seis sensores virtuais (GACID_H2S, GACID_NH3, GAMON_H2S, GAMON_NH3, AGUA_H2S e AGUA_NH3) foram desenvolvidos utilizando o algoritmo de RF. Todos os sensores apresentaram coeficiente de determinação R^2 maior que 0,87, MAE menor que 0,12, MSE menor que 0,17 e RMSE menor que 0,41 antes da redução das variáveis de entrada.

Pela análise visual dos gráficos de predição e resíduos e avaliação das métricas encontradas, foi possível concluir que os sensores do gás ácido e da água tratada se mostraram muito eficazes para toda a faixa das frações mássicas de H₂S e NH₃ analisadas. Já os sensores do gás amoniacal tiveram resultados visualmente menos adequados e, portanto, foram considerados dois intervalos de confiabilidade: a faixa 1 onde os resultados preditos são confiáveis e a faixa 2 onde a confiabilidade é menor. Comparando as métricas de erro das duas faixas criadas com as do modelo completo, a faixa 1 apresenta valor de RMSE 40,21% menor, enquanto a faixa 2 apresenta valor de RMSE 447,24% maior.

O estudo da Importância das Variáveis mostrou que os sensores continuaram apresentando boas métricas após a redução das variáveis de entrada com contribuição menor que 2%, sendo útil para melhorar o tempo de processamento computacional, sem comprometer o desempenho dos modelos. A única exceção foi o sensor AGUA_NH3 que revelou precisar do conjunto de variáveis de entrada original para poder manter a confiabilidade dos resultados.

A análise global da Importância das Variáveis destacou a influência significativa da temperatura no processo de remoção dos contaminantes H₂S e NH₃ de águas ácidas. Os resultados apontaram que todas as variáveis de temperatura estudadas foram fundamentais para o controle do processo termodinâmico de separação desses compostos, sendo especialmente crucial a temperatura do 2º estágio da Coluna 1, que se mostrou crítica em cinco dos seis sensores virtuais.

No contexto geral, a análise global identificou que, entre as 13 variáveis examinadas, apenas 6 foram consideradas importantes para os seis sensores virtuais desenvolvidos, sugerindo a possibilidade de reduzir a quantidade de dados que precisam ser coletados em

futuros trabalhos, otimizando recursos e buscando aprimorar os resultados alcançados no processo.

Recomenda-se também para trabalhos futuros que sejam testadas novas simulações capazes de cobrir as faixas de frações mássicas de H_2S e NH_3 no gás amoniacal onde a confiabilidade dos valores preditos não foi tão alta, para que, com um banco de dados mais balanceado, sejam desenvolvidos sensores mais robustos para essas variáveis nestas faixas. Além disso, é recomendado que se avalie a sensibilidade da simulação dinâmica em relação a diferentes intensidades de perturbação, analisando a variância nos dados de entrada e saída, a fim de identificar a faixa de distúrbio mais apropriada para estes dados.

Outras sugestões são: o desenvolvimento de novos sensores utilizando um banco de dados que englobe simulações fora da operação normal, isto é, que apresentem situações de falhas como, por exemplo, válvulas operando totalmente abertas ou fechadas, nível muito alto ou muito baixo nas colunas e cenários de incrustação, testar diferentes proporções da quantidade de dados presentes nos bancos de treinamento e teste e avaliar diferentes hiperparâmetros para os modelos gerados com a técnica de Florestas Aleatórias.

Por fim, é sugerida também a inclusão de controladores “indicadores”, isto é, utilizados apenas para leitura das variáveis, para medição das vazões mássicas instantâneas das correntes SW1, ACG1, AMG1 e TW1 para trabalhos futuros que visem avaliar a eficiência de remoção do H_2S e do NH_3 no sistema estudado.

Referências Bibliográficas

ADDINGTON, L.; FITZ, C.; LUNSFORD, K.; LYDDON, L.; SIWEK, M. *Sour Water: Where it comes from and how to handle it*. 2013. Disponível em: <https://www.bre.com/PDF/Sour-Water-Where-It-Comes-from-and-How-to-Handle-It.pdf>.

Acesso em 20 de fevereiro de 2023.

ALOM, M. Z. et al. A state-of-the-art survey on deep learning theory and architectures. *Electronics, Multidisciplinary Digital Publishing Institute*, v. 8, n. 3, p. 292, 2019.

BARROS, D. J. S. Investigação do efeito de variáveis de processo na eficiência de remoção de H₂S em unidade de tratamento de águas ácidas de duas torres. Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia de Química, 2016.

BAUGHMAN, D. R. e LIU, Y. A. *Neural Networks in Bioprocessing and Chemical Engineering*, Academic Press, 1995.

BELGIU, M.; DRAGUT, L. *Random Forest in remote sensing: A review of applications and future directions*, *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 114, p. 24–31, 2016.

BREIMAN, L. *Random Forests*. *Machine Learning*, v. 45, pp. 5–32, 2001.

BREIMAN, L.; FRIEDMAN, J.; STONE, C. et al. *Classification and Regression Trees*. Taylor e Francis, 1984. ISBN: 9780412048418.

BRASIL. Ministério do Meio Ambiente. Conselho Nacional do Meio Ambiente (CONAMA). Resolução número 382, de 26 de dezembro de 2006. Diário Oficial, nº 01, seção 1, 2 de janeiro de 2007.

BRITO, C. O.; ROSA, C.; MICHELETTO, S. R.; KNUST, C. M.; CARVALHO, H. P. Curso de Tratamento de Águas Ácidas, Universidade Petrobras, 2011.

CHEVRON, *WWT Process*. Disponível em: http://files.engineering.com/download.aspx?folder=19ca821a-5c8e-48f5-a105-63cdb8dccc54efile=chevron_wwt.pdf. Acesso em 20 de fevereiro de 2023.

CHOLLET, F. *Deep learning with Python*, Manning Publications, 2021.

COLBECK, I.; MACKENZIE, A. R. *Air Pollution by Photochemical Oxidants*. Amsterdam [The Netherlands]; New York: Elsevier, 1994.

CONAMA (Conselho Nacional do Meio Ambiente). RESOLUÇÃO N° 436, DE 22 DE DEZEMBRO DE 2011. 2011. Disponível em: (<https://www.udop.com.br/download/legislacao/meio/institucional_site_juridico/Conama_Resolucao%20436.pdf>). Acesso em 20 de fevereiro de 2023.

COELHO, A.; CASTRO, A. V.; DEZOTTI, M.; SANT'ANNA, G.L. Tratamento de água ácida de refinaria de petróleo por processos de oxidação avançada. *Journal of Hazardous Materials*, v. 137, n. 1, p. 178-184, 2006.

CST Studio Suite. *Time Integration Methods*. 2020. Disponível em: (<https://space.mit.edu/RADIO/CST_online/mergedProjects/DES/theory/time_integration.htm>). Acesso em 15 de agosto de 2023.

DARTON. R.C.; VAN GRINSVEN. P.F.A. *Development in Stream Stripping of Sour Water*, The Chemical Engineer, 1978.

EAP (United States Environment Protection Agency). *What is Acid Rain?*. 2021. Disponível em: (<<https://www.epa.gov/acidrain/what-acid-rain>>). Acesso em 20 de fevereiro de 2023.

FERRÃO, M. F.; MELLO, C.; BORIN, A.; MARETTO, D. A.; POPPI, R. J. LS-SVM: uma nova ferramenta quimiométrica para regressão multivariada. comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR. *Química Nova*, v. 30, n. 4, p. 852-859, 2007.

FILHO, M. S. S. et al. *Random forest* para detecção de pré-falha em plantas de tratamento de água ácida. In: ANAIS DO 24° CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA, 2023, Salvador. Disponível em: (<<https://proceedings.science/cobeq-2023/trabalhos/random-forest-para-deteccao-de-pre-falha-em-plantas-de-tratamento-de-agua-acida?lang=pt-br>>). Acesso em: 16 out. 2023.

FORTUNA, L., GRAZIANI, S., RIZZO, A., XIBILIA, M. G., OTHERS. *Soft sensors for monitoring and control of industrial processes*, v. 22. Springer, 2007.

GRAZIANI, S.; XIBILIA, M. G. *A deep learning based soft sensor for a sour water stripping plant*. In: IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2017, Turin, Italy.

HO, CHII-DONG et al. *Evaluation of Process Control Schemes for Sour Water Strippers in Petroleum Refining*. Processes, v. 9, 2021.

JOUCOWSKI, J. Inferência de Propriedades de Correntes para Caracterização da Carga de uma Unidade de Hidrotratamento de Diesel, Dissertação de Mestrado, UFPR, 2011.

JÚNIOR, C. O. Métricas para Regressão: Entendendo as métricas R², MAE, MAPE, MSE e RMSE. Data Hackers. 2021. Disponível em: (<https://medium.com/data-hackers/prevendo-n%C3%BAmoros-entendendo-m%C3%A9tricas-de-regress%C3%A3o-35545e011e70>).

Acesso em 28 de setembro de 2023.

KENSELL, W.; QUINLAN, M. Chapter 11.1: The M.N. Kellogg company refinery sulfur management. In: Handbook of Petroleum Refining Process, 1996.

LEE, D. et al. *Dynamic simulation of the sour water stripping process and modified structure for effective pressure control*. Chemical Engineering Research and Design, Elsevier, v. 80, n. 2, p. 167–177, 2002.

LI, A.; WIENER, M. *Classification and Regression by Randomforest*. R News, v. 2, p. 18-22, 2002.

LIN, B., RECKE, B., KNUDSEN, J. K., JØRGENSEN, S. B. *A systematic approach for soft sensor development*, Computers e chemical engineering, v. 31, n. 5-6, pp. 419– 425, 2007.

LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às *support vector machines*. Revista de Informática Teórica e Aplicada, v. 14, n. 2, p. 43–67, 2007.

LOTUFO, F. A., GARCIA, C. Sensores virtuais ou *soft sensors*: Uma introdução. In: 7th Brazilian Conference on Dynamics, Control and Applications, Presidente Prudente, 2008.

MELIN, G.A.; NIEDZWIECKI, J.L.; GOLDSTEIN, A. M. *Optimum Design of Sour Water Strippers*. Chemical Engineering Process, Vol 71, No 6, 1975.

MONTAVON, G.; ORR, G. B.; MÜLLER, K. R. *Neural Networks: Tricks of the Trade - Second Edition*. Lecture Notes in Computer Science 7700. Springer, 2012. ISBN 978-3-642-35288-1

NOGUEIRA, J. N. P. *Fault Detection and Diagnosis of a Two-Column Sour Water Treatment Unit Based on Artificial Intelligence Algorithms*. Dissertação (Mestrado) - Universidade

Federal do Rio de Janeiro, Escola de Química, Programa de Pós-Graduação em Engenharia de Processos Químicos e Bioquímicos, 2021.

NOGUEIRA, J. N. P.; JUNIOR, M. B. S.; JUNIOR, P. A. M. Detecção e diagnóstico de falhas em unidade de tratamento de águas ácidas com Random Forest. In: ANAIS DO 23º CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA, 2021, Gramado. Disponível em: <<https://proceedings.science/cobeq/cobeq-2021/trabalhos/deteccao-e-diagnostico-de-falhas-em-unidade-de-tratamento-de-aguas-acidas-com-ra?lang=pt-br>> Acesso em: 16 out. 2023.

NOGUEIRA, J. N. P.; MELO, P. A.; JUNIOR, M. B. S. *Faulty scenarios in sour water treatment units: Simulation and AI-based diagnosis*. Process Safety and Environmental Protection, v. 165, p. 716-727, 2022.

PITANGA, G. F. Análise de Cenários de Pré-Falha e Incrustação em Unidades de Tratamento de Águas Ácidas. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) – Escola de Química, Universidade Federal do Rio de Janeiro, 2022.

QUINLAN, M., HATI, A., 2010, *Processing NH₃ Acid Gas in Sulphur Recovery Unit*, Gas, pp. 45–55.

ROMEO, G. *Elements of Numerical Mathematical Economics with Excel*. Static and Dynamic Optimization, 2020.

RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. 2002.

SAMUEL, A. L. *Some studies in machine learning using the game of checkers*. IBM Journal of research and development, IBM, v. 3, n. 3, 1959.

SANDLER. S.I. *Chemical and Engineering Thermodynamics*, 3rd edition, John Wiley, 1999.

SHEILAN. M.H.; SPOONER. B.H.; VAN HOORN. E. *Amine Treating and Sour Water Stripping*, 7th Edition, Sulphur Experts Inc, Canada, 2010.

SOARES, A. de F.; PENTEADO, E. D.; DINIZ, A. A. R.; KOMESU, A. *Influence of operational parameters in sour water stripping process in effluents treatment*. Journal of Water Process Engineering, v. 41, 2021.

SOARES, F. D. R. Técnicas de *machine learning* aplicadas a inferência e detecção e diagnóstico de falhas de processos químicos industriais em contexto *big data*. Dissertação

(Mestrado) - Universidade Federal do Rio de Janeiro, Escola de Química, Programa de Pós-Graduação em Engenharia de Processos Químicos e Bioquímicos, 2017.

SOBREIRA, S. G. A. Aplicação de sensores virtuais baseados em aprendizado de máquina para estimativa da vazão mássica de minério de ferro em correias transportadoras. Dissertação (Mestrado) - Universidade Federal de Ouro Preto, Escola de Minas, Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração, 2021.

SPEISER, J. L., MILLER, M. E., TOOZE, J., et al., 2019, *A comparison of random forest variable selection methods for classification prediction modeling*, Expert Systems with Applications, v. 134, pp. 93–101.

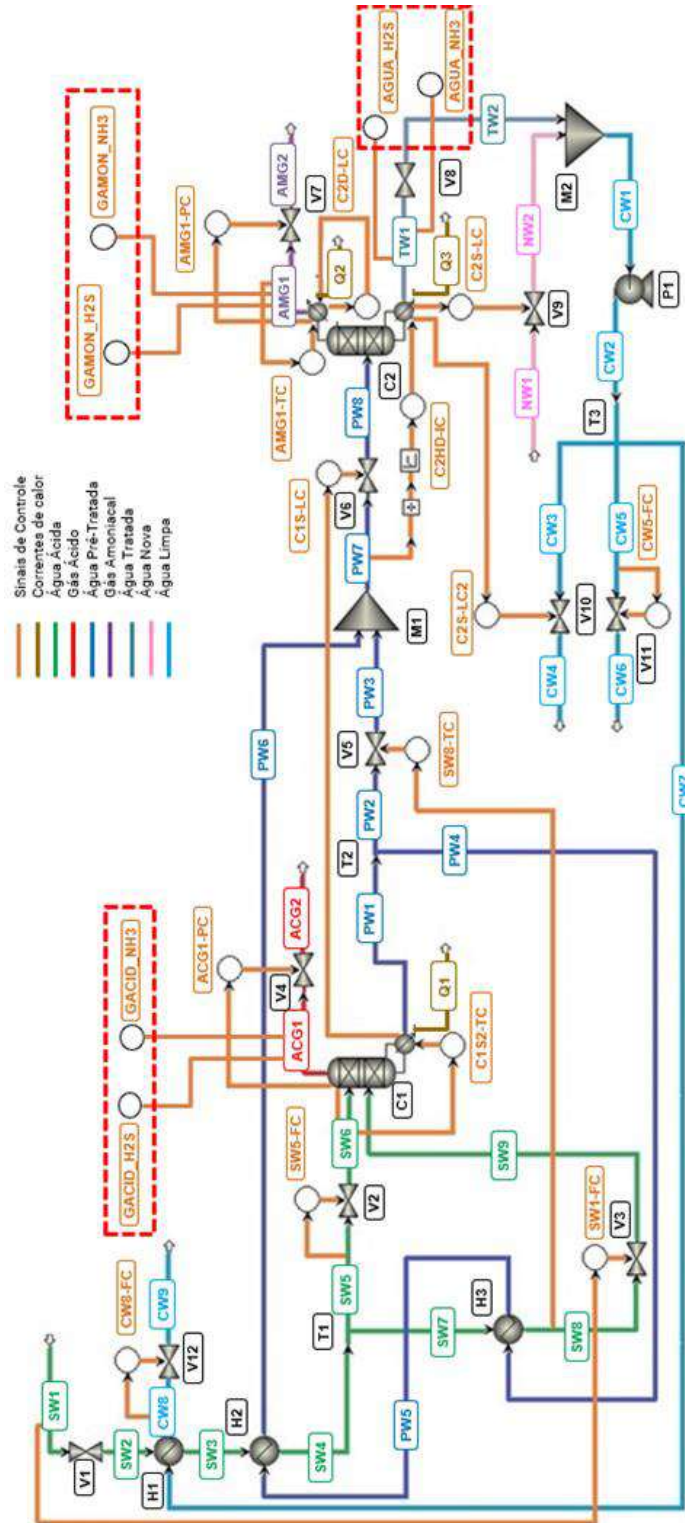
TRAVALIS, T. B.; INCE, H. *Support vector machine for regression and applications to financial forecasting*. In: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, Como, Italy, 2000, p. 348-353, vol. 6.

WEILAND, R. H.; HATCHER, N. A. *Sour water strippers exposed*. In: *Laurence Reid Gas Conditioning Conference*, Norman, Oklahoma. [S.l.: s.n.], 2012.

ZHOU, Z.H. *Ensemble Methods: Foundations and Algorithms*. Abingdon-on-Thames: Taylor e Francis, 2012.

APÊNDICE A – Adaptação da simulação dinâmica de Nogueira (2021) com a inclusão dos novos seis controladores

Figura A.1: Simulação dinâmica com a inclusão dos novos seis sensores – versão expandida.



Fonte: Elaboração própria.

APÊNDICE B – Banco de dados desenvolvido

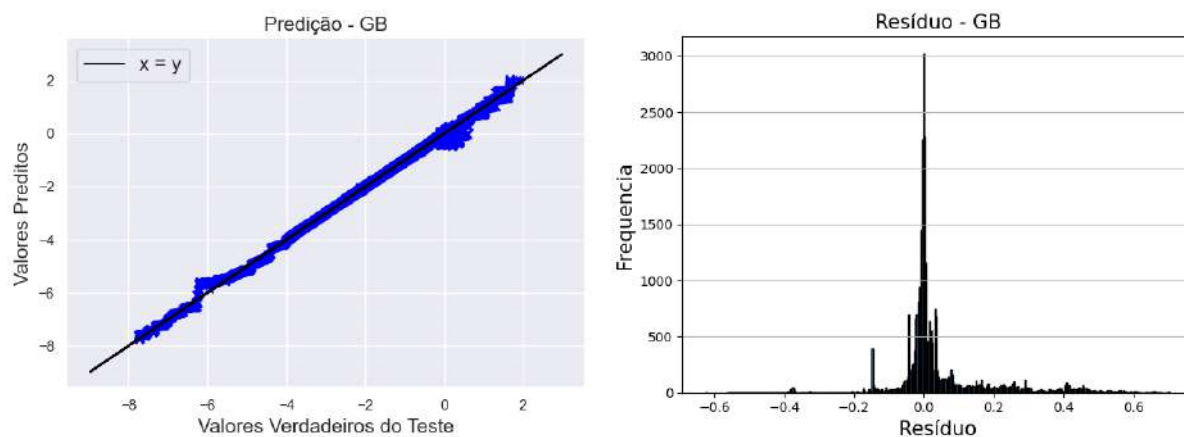
Os bancos de treinamento e teste normalizados desenvolvidos no presente trabalho estão disponíveis em: https://github.com/danigradin/UTAA_Soft_Sensors.

APÊNDICE C – Resultados dos sensores virtuais usando *Gradient Boosting* e SVM – *kernel* RBF e linear

C.1 Resultados dos sensores virtuais usando *Gradient Boosting*

C.1.1 Sensor GACID_H2S

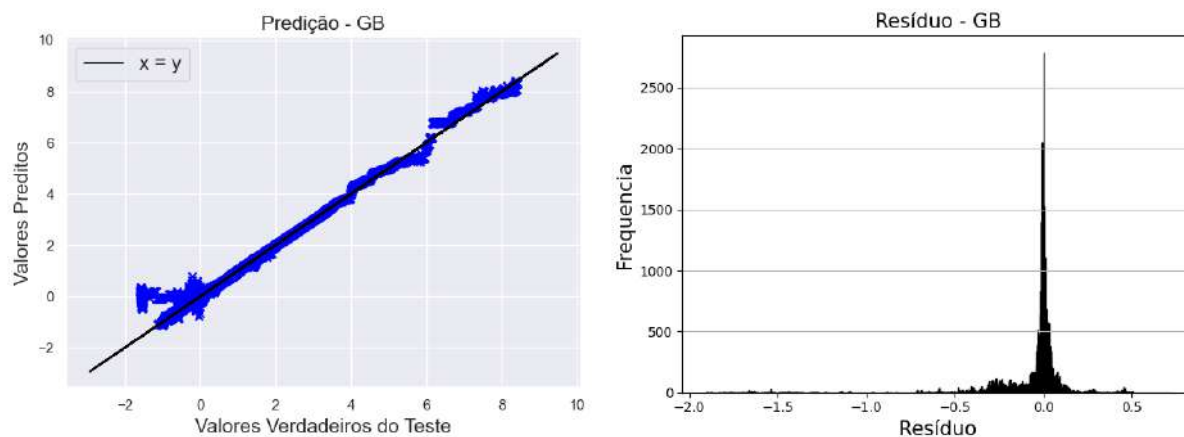
Figuras C.1 A e B: Predição (A) e resíduo (B) do sensor GACID_H2S usando GB.



Fonte: Elaboração própria.

C.1.2 Sensor GACID_NH3

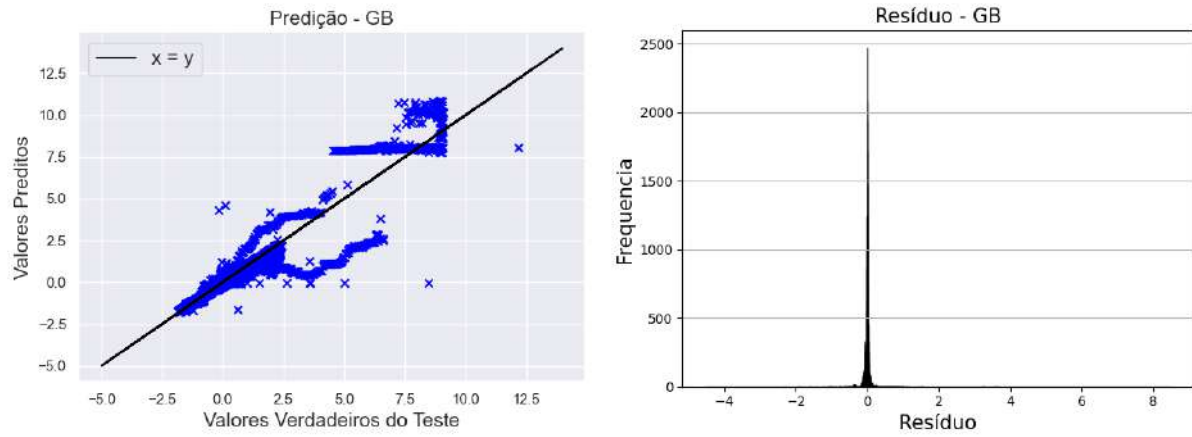
Figuras C.2 A e B: Predição (A) e resíduo (B) do sensor GACID_NH3 usando GB.



Fonte: Elaboração própria.

C.1.3 Sensor GAMON_H2S

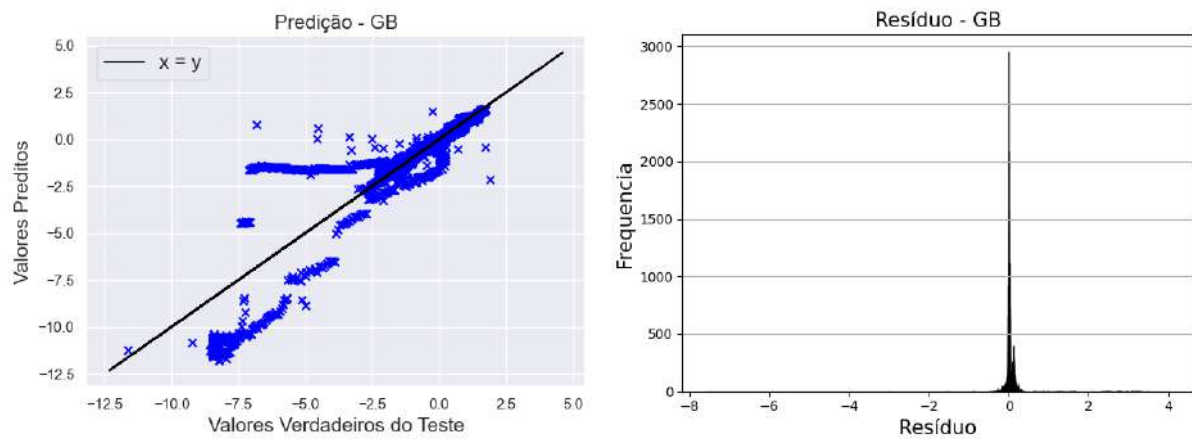
Figuras C.3 A e B: Predição (A) e resíduo (B) do sensor GAMON_H2S usando GB.



Fonte: Elaboração própria.

C.1.4 Sensor GAMON_NH3

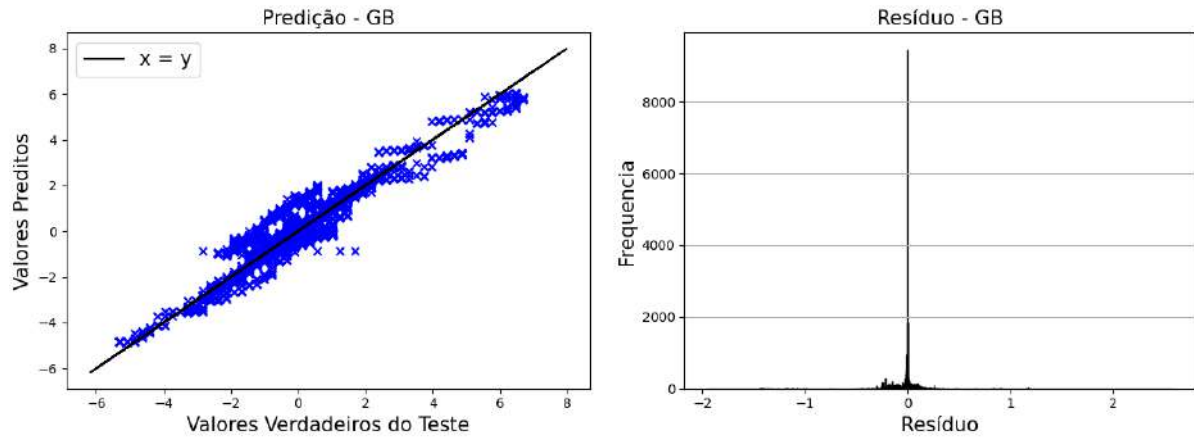
Figuras C.4 A e B: Predição (A) e resíduo (B) do sensor GAMON_NH3 usando GB.



Fonte: Elaboração própria.

C.1.5 Sensor AGUA_H2S

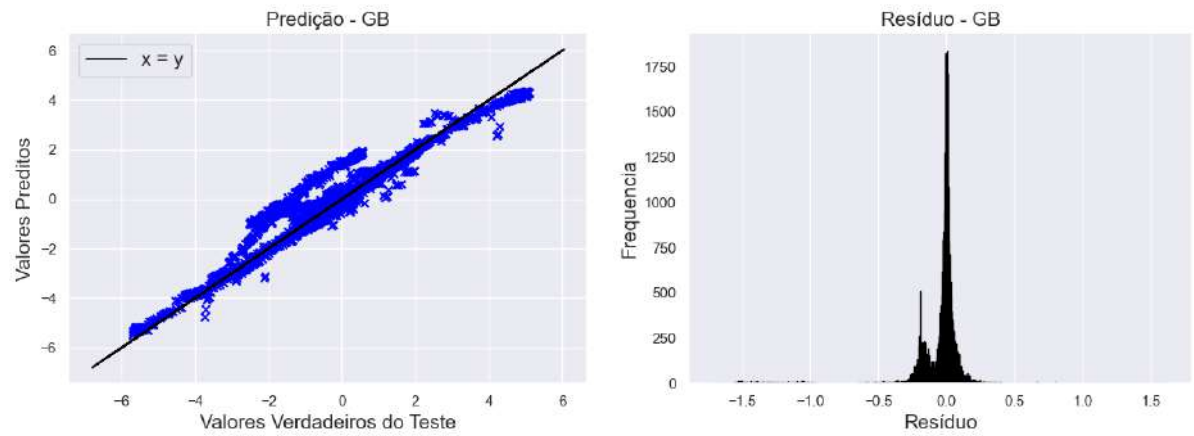
Figuras C.5 A e B: Predição (A) e resíduo (B) do sensor AGUA_H2S usando GB.



Fonte: Elaboração própria.

C.1.6 Sensor AGUA_NH3

Figuras C.6 A e B: Predição (A) e resíduo (B) do sensor AGUA_NH3 usando GB.



Fonte: Elaboração própria.

C.2 Tabela global de importância das variáveis de todos os sensores virtuais usando *Gradient Boosting* após redução das variáveis de entrada

Tabela C.1: Análise global de importância das variáveis de todos os sensores virtuais usando *Gradient Boosting* após redução das variáveis de entrada.

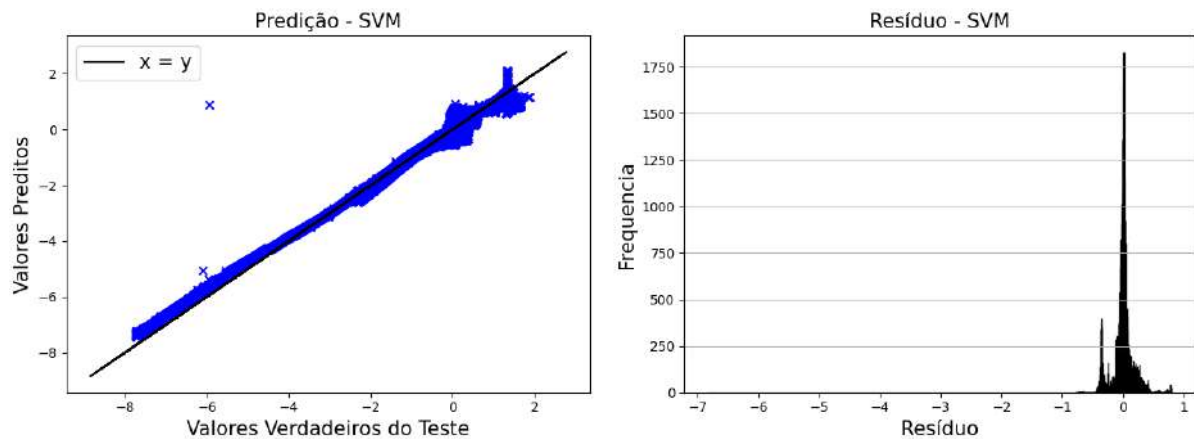
Sensores	CW8-FC	SW5-FC	SW1-FC	SW8-TC	ACG1-PC	C1S2-TC	C1S-LC	AMG1-TC	AMG1-PC	C2D-LC	C2S-LC	C2HD-IC	CW5-FC
GACID_H2S		X		X		X							
GACID_NH3		X				X							
GAMON_H2S					X	X							
GAMON_NH3						X	X	X		X			
AGUA_H2S					X	X		X					
AGUA_NH3					X			X					

Fonte: Elaboração própria.

C.3 Resultados dos sensores virtuais usando SVM – *kernel RBF*

C.3.1 Sensor GACID_H2S

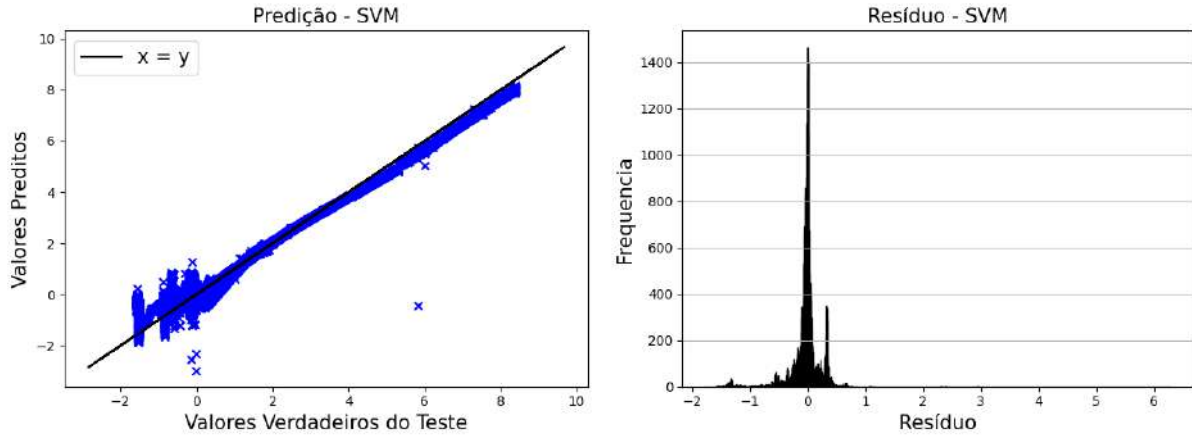
Figuras C.7 A e B: Predição (A) e resíduo (B) do sensor GACID_H2S usando SVM *kernel RBF*.



Fonte: Elaboração própria.

C.3.2 Sensor GACID_NH3

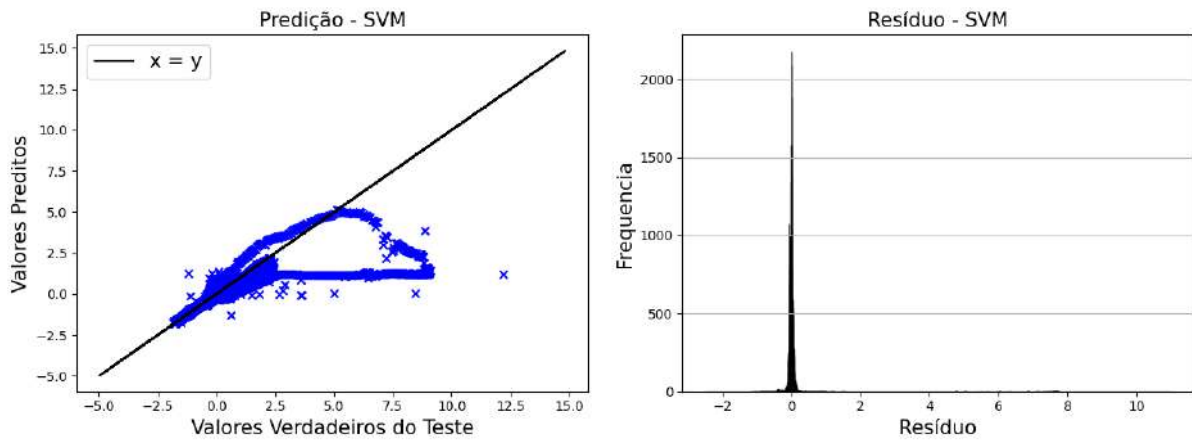
Figuras C.8 A e B: Predição (A) e resíduo (B) do sensor GACID_NH3 usando SVM *kernel* RBF.



Fonte: Elaboração própria.

C.3.3 Sensor GAMON_H2S

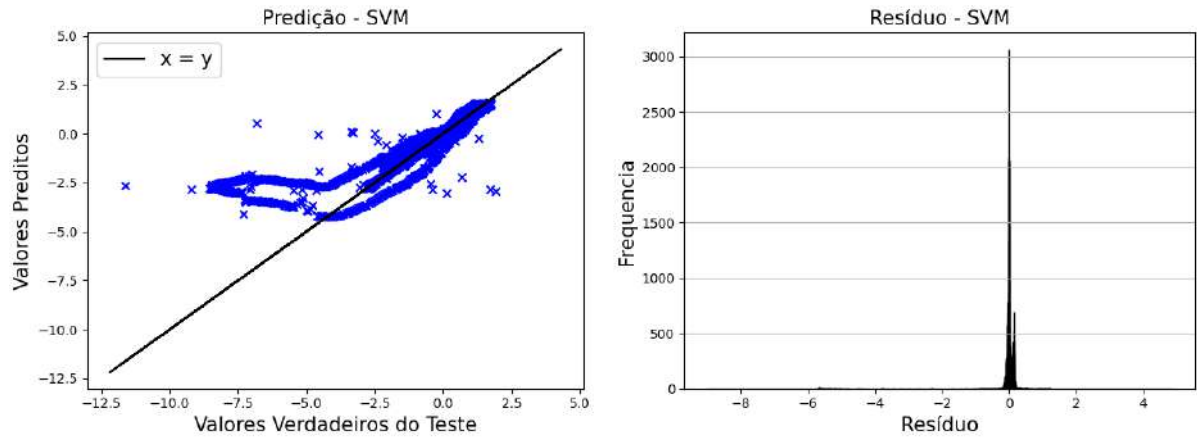
Figuras C.9 A e B: Predição (A) e resíduo (B) do sensor GAMON_H2S usando SVM *kernel* RBF.



Fonte: Elaboração própria.

C.3.4 Sensor GAMON_NH3

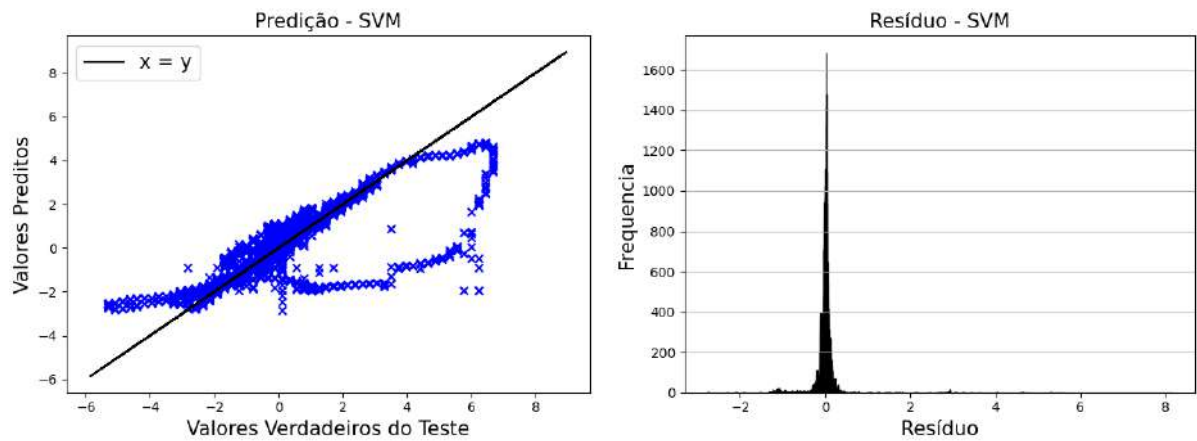
Figuras C.10 A e B: Predição (A) e resíduo (B) do sensor GAMON_NH3 usando SVM *kernel* RBF.



Fonte: Elaboração própria.

C.3.5 Sensor AGUA_H2S

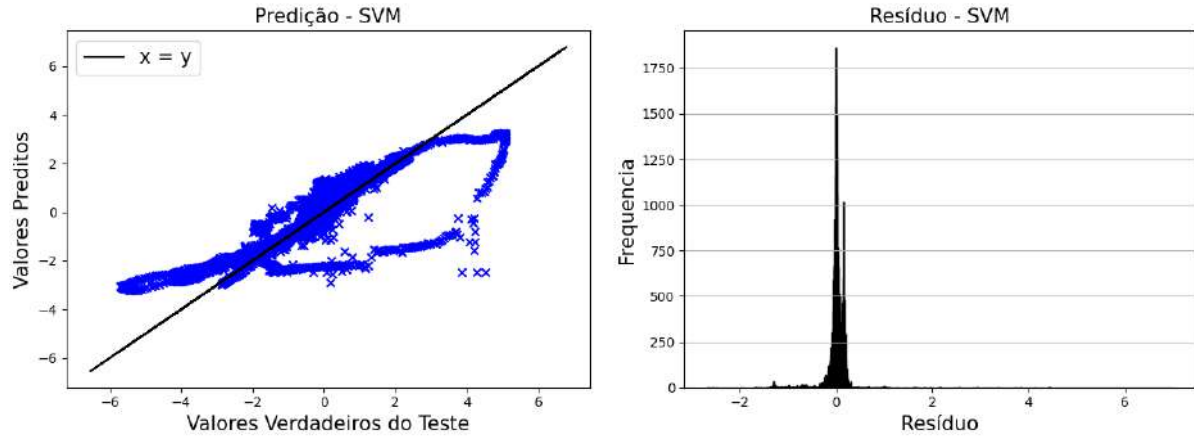
Figuras C.11 A e B: Predição (A) e resíduo (B) do sensor AGUA_H2S usando SVM *kernel* RBF.



Fonte: Elaboração própria.

C.3.6 Sensor AGUA_NH3

Figuras C.12 A e B: Predição (A) e resíduo (B) do sensor AGUA_NH3 usando SVM *kernel* RBF.

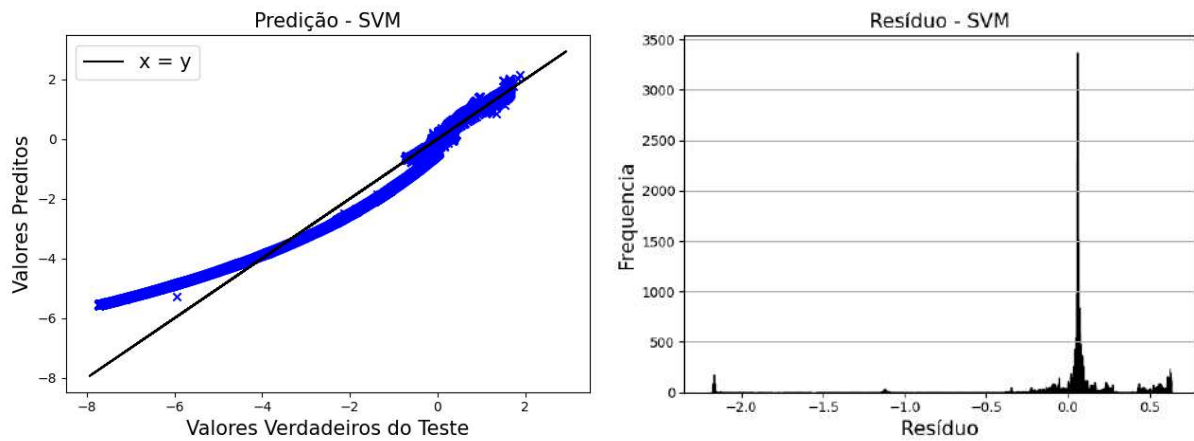


Fonte: Elaboração própria.

C.4 Resultados dos sensores virtuais usando SVM – *kernel* linear

C.4.1 Sensor GACID_H2S

Figuras C.13 A e B: Predição (A) e resíduo (B) do sensor GACID_H2S usando SVM *kernel* linear.

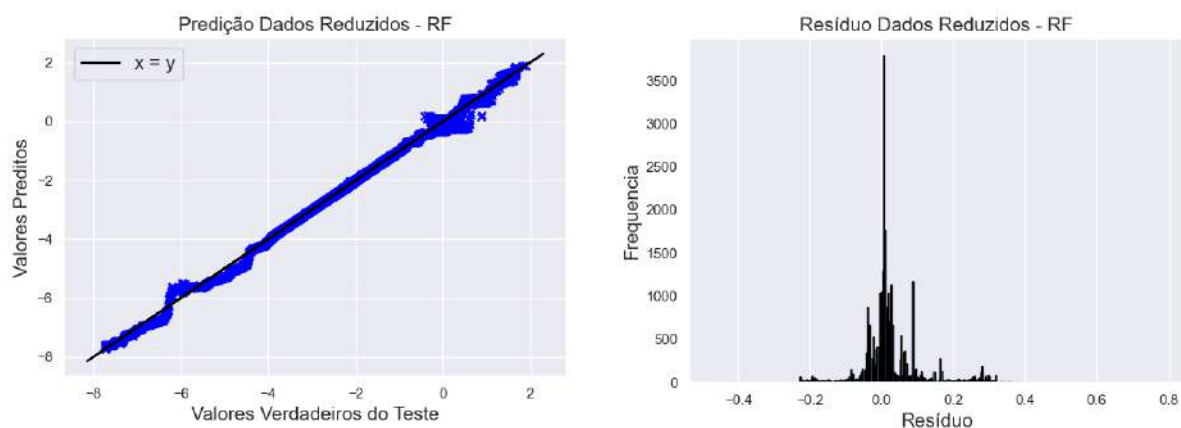


Fonte: Elaboração própria.

APÊNDICE D – Gráficos de predição e resíduos após redução das variáveis de entrada para os sensores virtuais desenvolvidos com Florestas Aleatórias

D.1 Sensor GACID_H2S

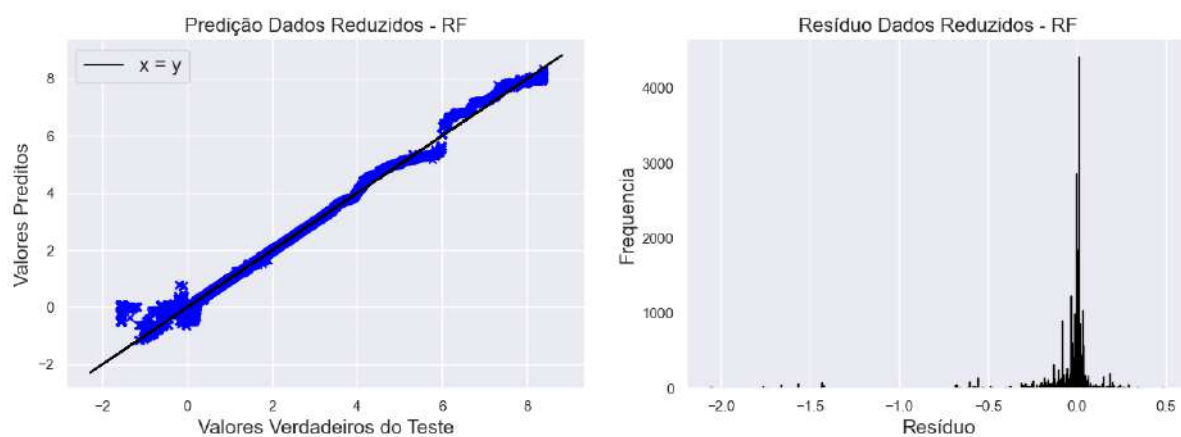
Figuras D.1 A e B: Predição (A) e resíduo (B) do sensor GACID_H2S usando RF após redução das variáveis de entrada.



Fonte: Elaboração própria.

D.2 Sensor GACID_NH3

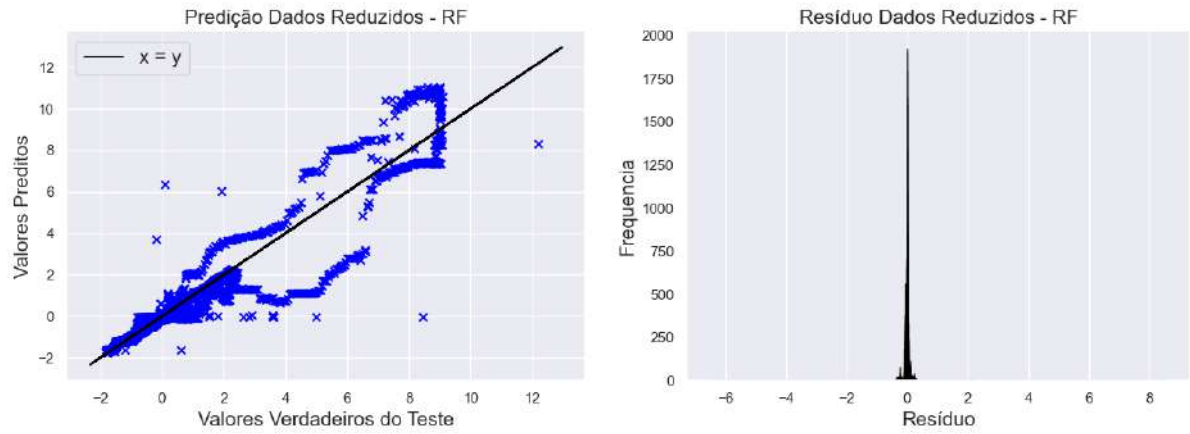
Figuras D.2 A e B: Predição (A) e resíduo (B) do sensor GACID_NH3 usando RF após redução das variáveis de entrada.



Fonte: Elaboração própria.

D.3 Sensor GAMON_H2S

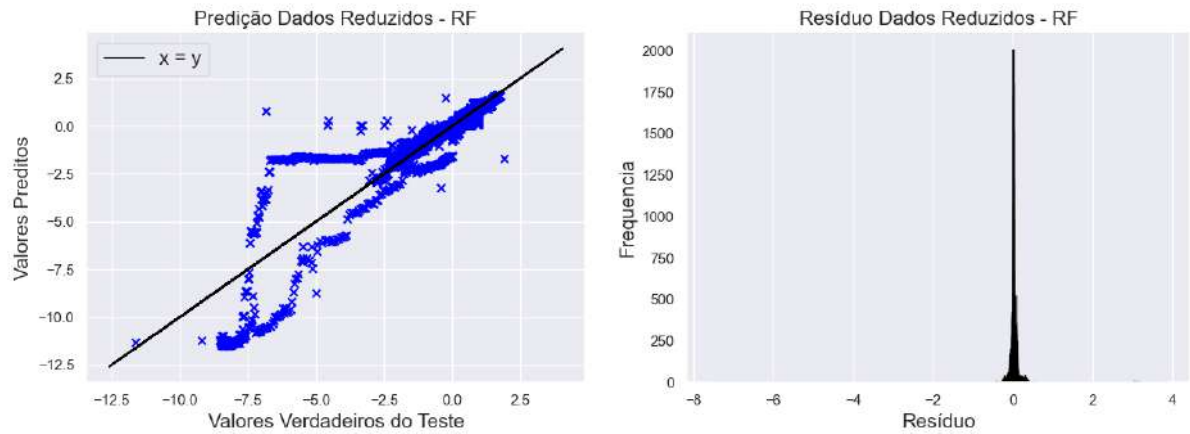
Figuras D.3 A e B: Predição (A) e resíduo (B) do sensor GAMON_H2S usando RF após redução das variáveis de entrada.



Fonte: Elaboração própria.

D.4 Sensor GAMON_NH3

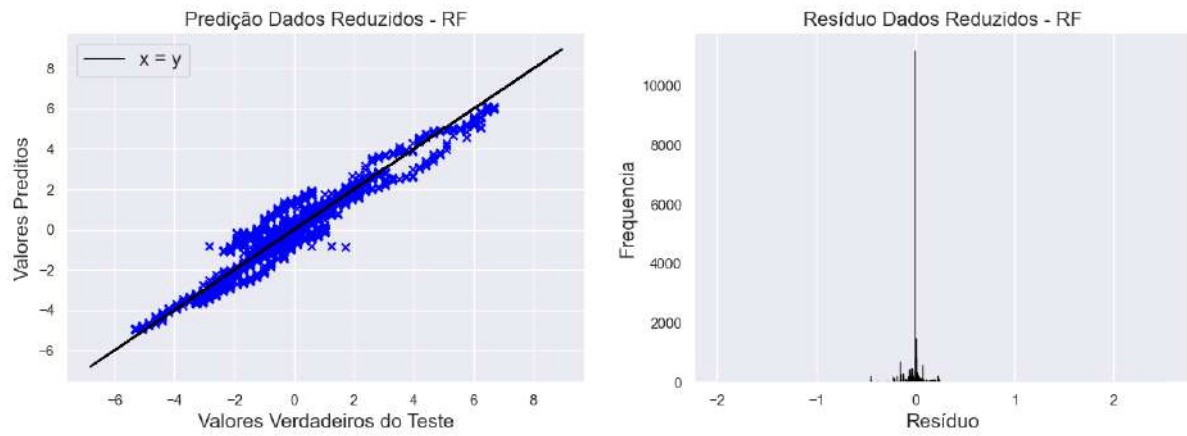
Figuras D.4 A e B: Predição (A) e resíduo (B) do sensor GAMON_NH3 usando RF após redução das variáveis de entrada.



Fonte: Elaboração própria.

D.5 Sensor AGUA_H2S

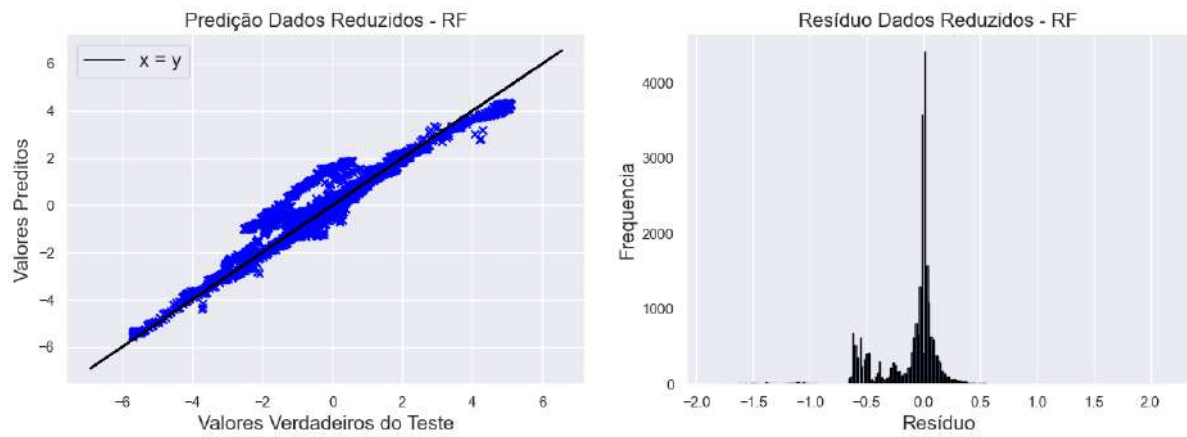
Figuras D.5 A e B: Predição (A) e resíduo (B) do sensor AGUA_H2S usando RF após redução das variáveis de entrada.



Fonte: Elaboração própria.

D.6 Sensor AGUA_NH3

Figuras D.6 A e B: Predição (A) e resíduo (B) do sensor AGUA_NH3 usando RF após redução das variáveis de entrada.

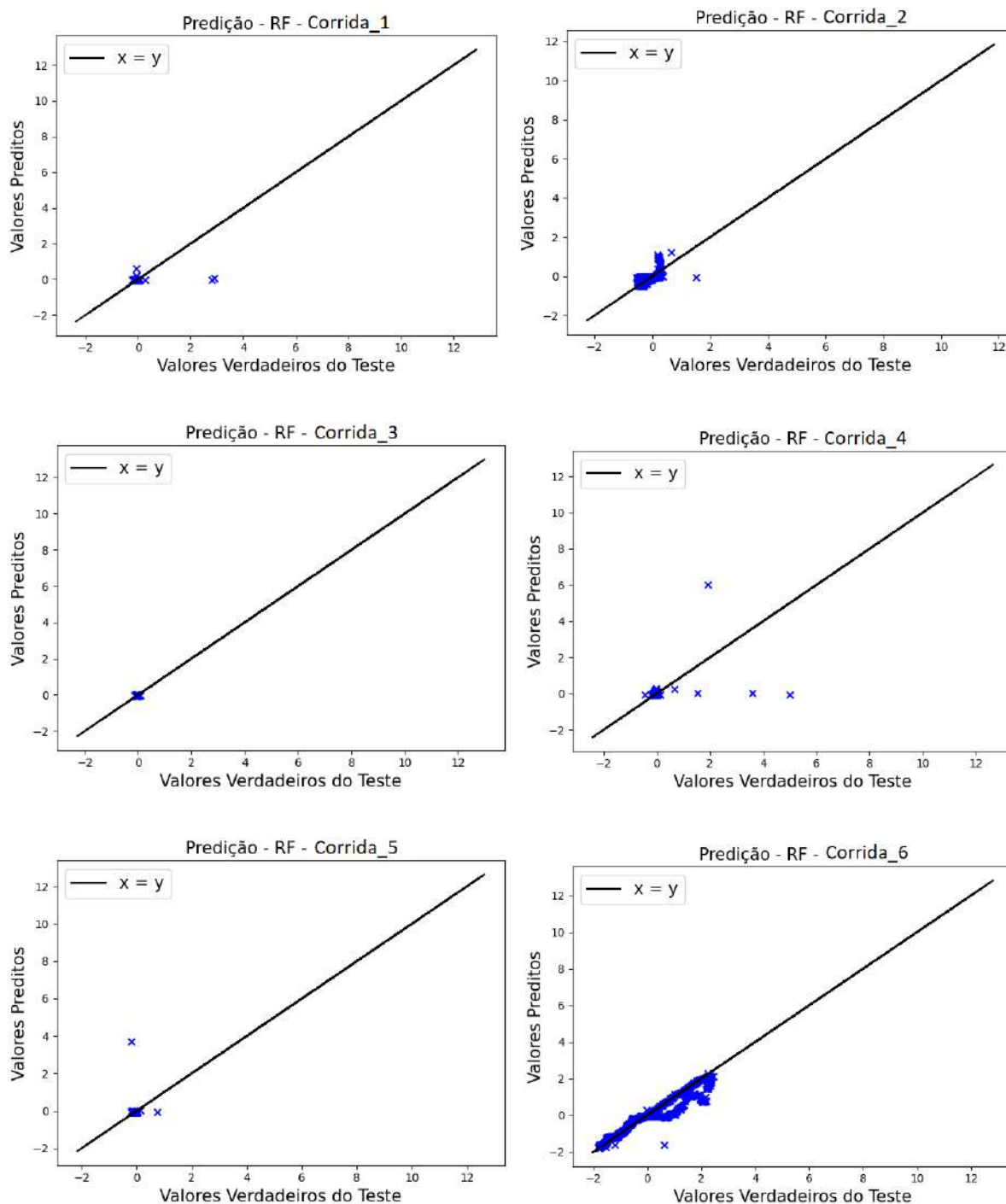


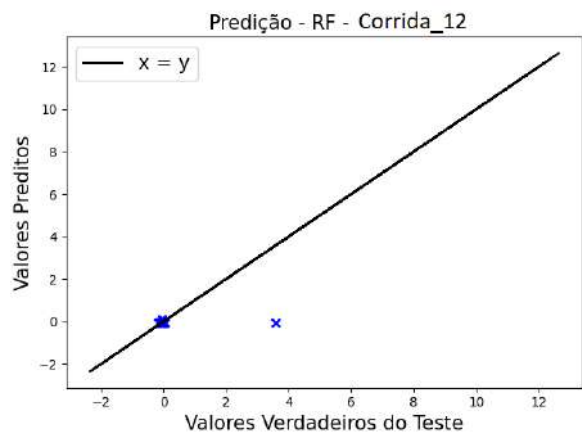
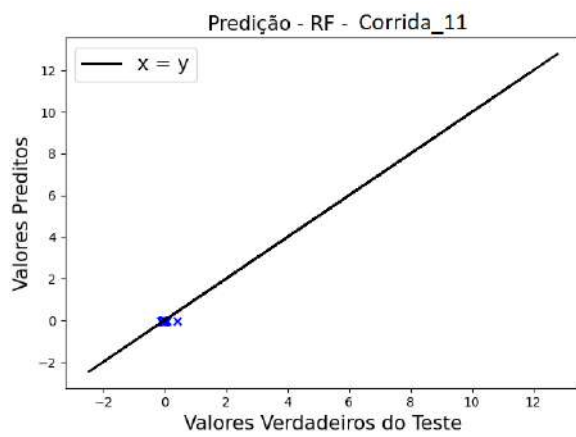
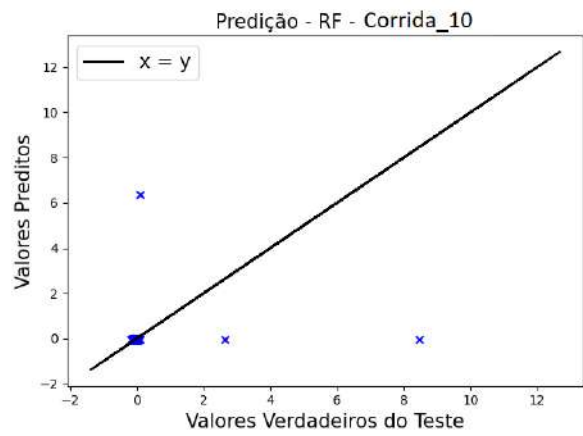
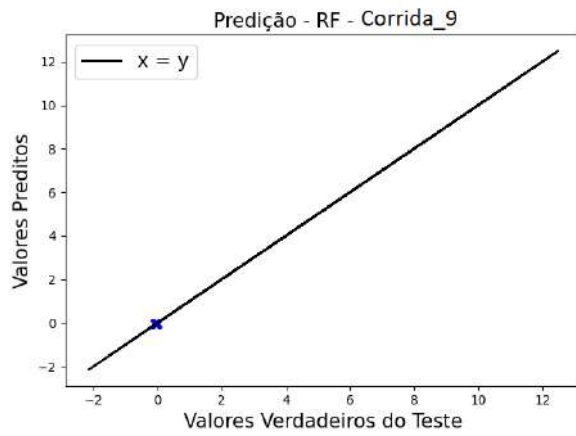
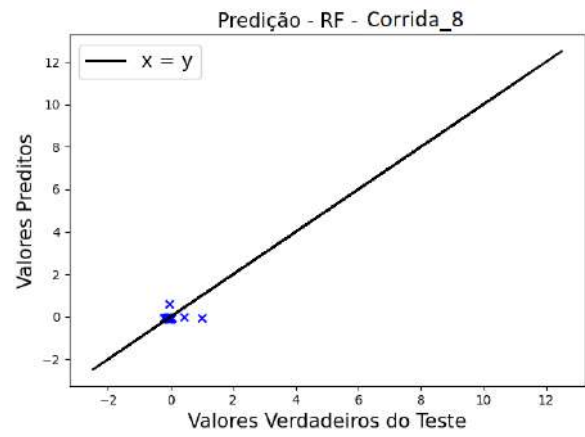
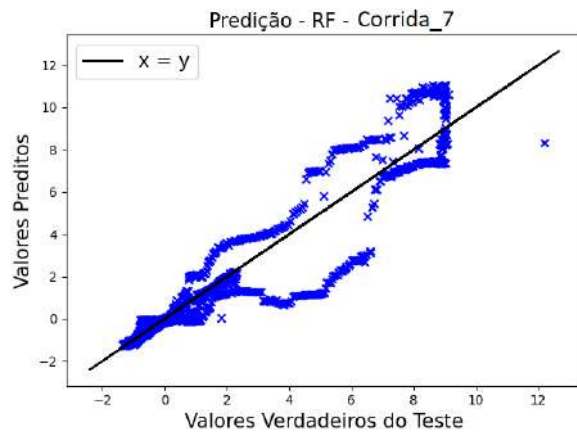
Fonte: Elaboração própria.

APÊNDICE E – Gráficos de predição para cada uma das 12 corridas do banco de teste para os sensores virtuais GAMON_H2S e GAMON_NH3

E.1 Sensor GAMON_H2S

Figuras E.1: Predições individuais das 12 corridas presentes no banco de teste para o sensor GAMON_H2S usando o melhor modelo do RF.

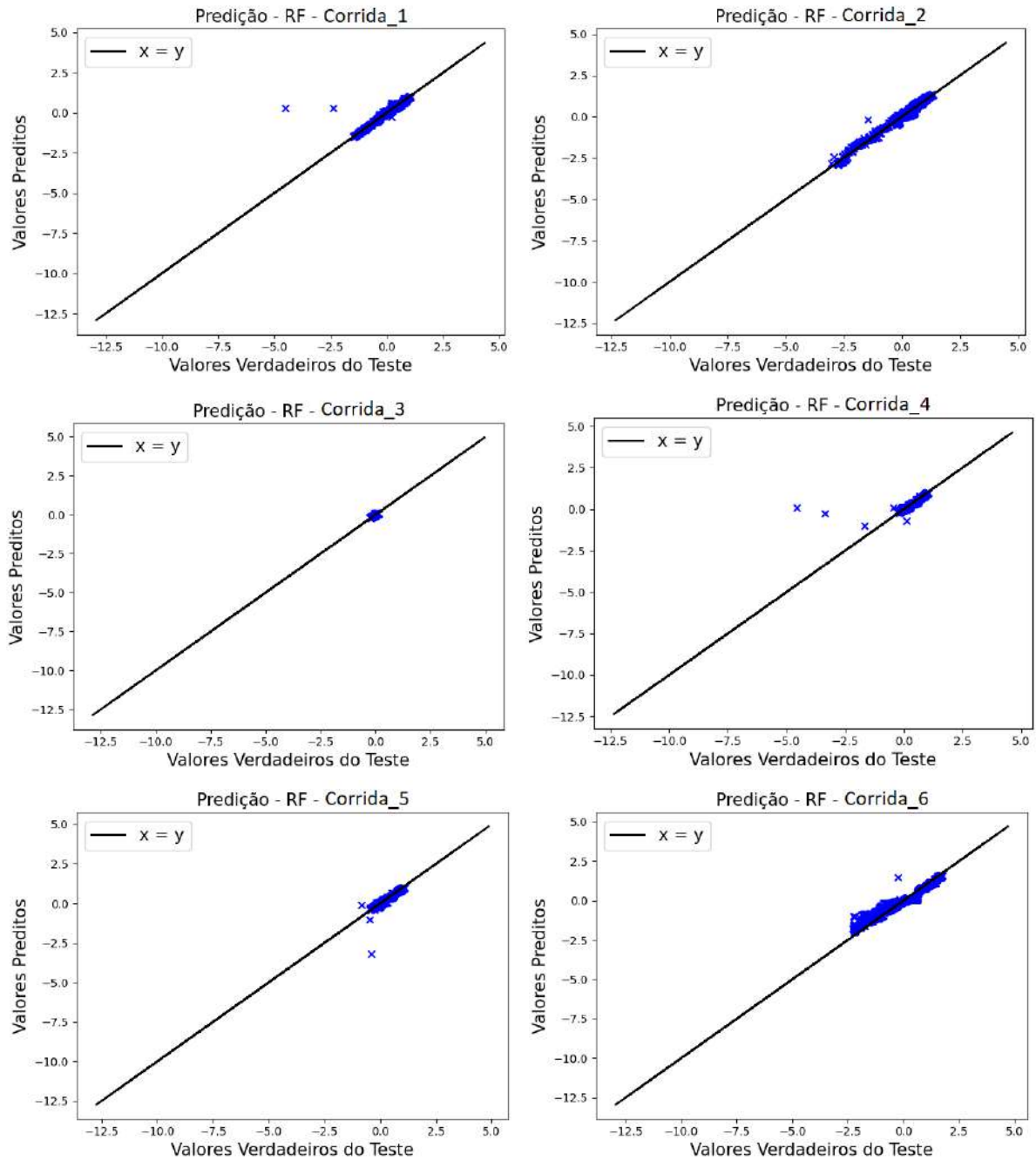


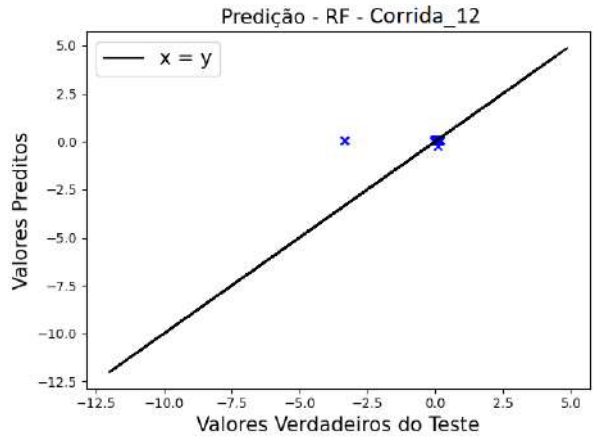
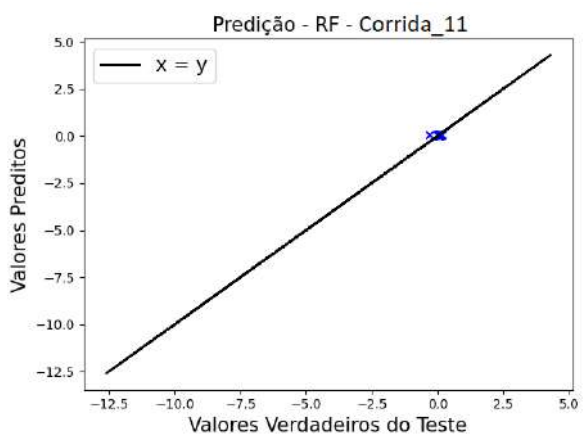
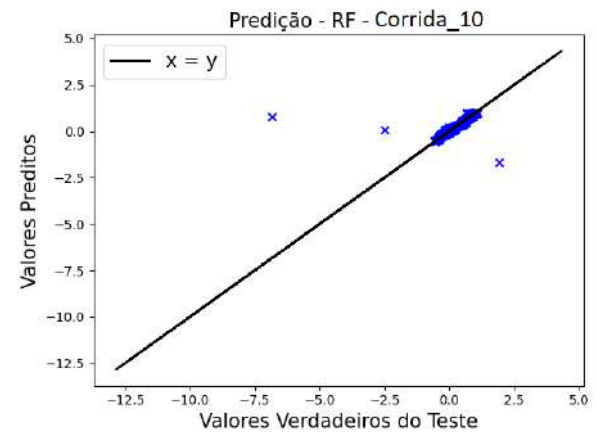
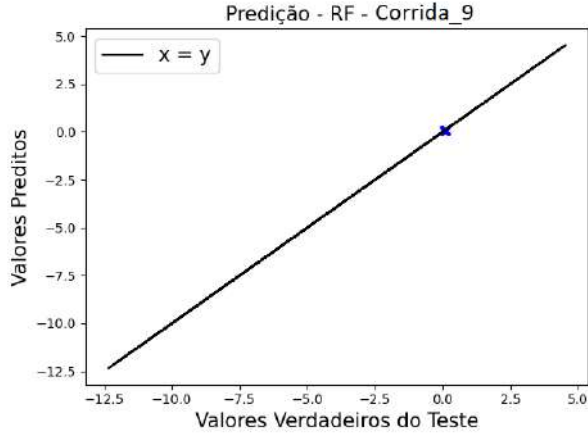
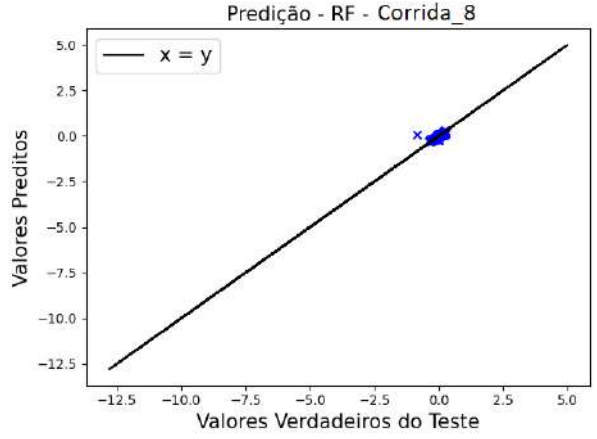
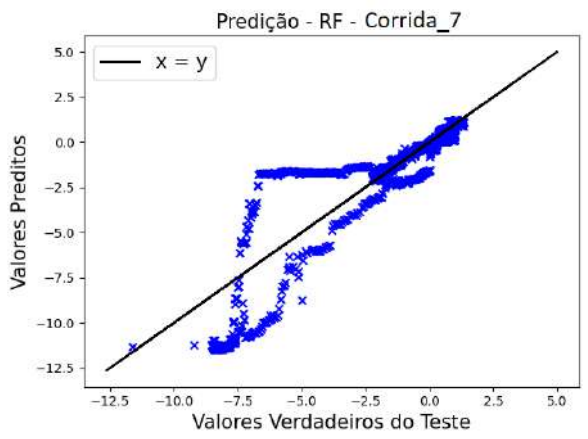


Fonte: Elaboração própria.

E.2 Sensor GAMON_NH3

Figuras E.2: Predições individuais das 12 corridas presentes no banco de teste para o sensor GAMON_NH3 usando o melhor modelo do RF.





Fonte: Elaboração própria.