
# Avaliação de Sistemas de Detecção e Diagnóstico de Falhas Baseados em Aprendizado de Máquina em Cenários de Pré-Falha e Incrustação em Unidades de Tratamento de Águas Ácidas

**Guilherme Faustino Pitanga**

## Projeto Final de Curso

**Orientadores**

Prof. Príamo Albuquerque Melo Junior, D. Sc.

Prof. Maurício Bezerra de Souza Jr., D. Sc.

Júlia do Nascimento Pereira Nogueira, M. Sc.

**Fevereiro de 2022**

# Avaliação de Sistemas de Detecção e Diagnóstico de Falhas Baseados em Aprendizado de Máquina em Cenários de Pré-Falha e Incrustação em Unidades de Tratamento de Águas Ácidas

***Guilherme Faustino Pitanga***

Projeto de Final de Curso submetido ao Corpo Docente da Escola de Química, como parte dos requisitos necessários à obtenção do grau de Bacharel em Engenharia Química.

Aprovado
por:

---

Prof. Bruno Didier Olivier Capron, D. Sc.

---

Maurício Melo Câmara, D. Sc.

Orientado
por:

---

Prof. Príamo Albuquerque Melo Junior, D. Sc.

---

Prof. Maurício Bezerra de Souza Jr., D. Sc.

---

Júlia do Nascimento Pereira Nogueira, M. Sc.

PITANGA, GUILHERME FAUSTINO

Avaliação de Sistemas de Detecção e Diagnóstico de Falhas Baseados em Aprendizado de Máquina em Cenários de Pré-Falha e Incrustação em Unidades de Tratamento de Águas Ácidas.[Rio de Janeiro] 2022

XIII, 95 p. 29,7 cm

Orientador: Príamo Albuquerque Melo Junior

Coorientadores: Maurício Bezerra de Souza Jr.

Júlia do Nascimento Pereira Nogueira

Projeto final de curso - Universidade Federal do Rio de Janeiro, Escola de Química, 2022

1.Água Ácida

2.Pré-Falha

3.Incrustação

4.Inteligência Artificial

5.*Random Forest*

6.Máquina de Vetores de Suporte

I.Albuquerque Melo Junior, Príamo, orient. II.B. de Souza Jr., Maurício, coorient. III.do Nascimento Pereira Nogueira, Júlia, coorient. IV. Título

## AGRADECIMENTOS

Primeiramente, agradeço a Deus, a Ele seja toda a glória eternamente, por me permitir contemplar os mistérios da Sua obra.

Agradeço aos meus pais, Giovana e Leonardo, pelo amor que sempre tiveram por mim desde o princípio. Agradeço por terem fornecido as condições necessárias para que eu pudesse me desenvolver como cidadão e como profissional. Caso os resultados do meu trabalho venham a contribuir de maneira positiva na sociedade, esta contribuição será graças à dedicação dada ao meu ensino. Contudo, e mais importante, sou perpetuamente grato por seus ensinamentos inestimáveis que recebi ora por instrução, ora por exemplo.

Agradeço aos meus avós, Luiz Carlos, Madalena e Regina, pois vossa presença tornou tudo muito mais fácil. Desde a mais tenra infância, saber que podia contar a todo instante com a confiança e o apoio dos meus avós, permitiu-me vislumbrar horizontes mais longos. Agradeço também às minhas tias, Jaqueline e Roberta, por terem me dado amor, carinho e me educado como um filho.

Agradeço também às minhas amizades. Aos meus contemporâneos do Colégio Militar do Rio de Janeiro, eu sou grato por, há uma década, terem me acompanhado em todas as minhas jornadas. Aos meus colegas de curso e companheiros de profissão que estiveram ao meu lado durante esse longo percurso, sou grato por, nos momentos de alegria, terem junto a mim celebrado as vitórias e, nos momentos de dificuldade, terem tornado o fardo mais leve.

Agradeço à Universidade Federal do Rio de Janeiro, aos seus funcionários e, em particular, ao corpo docente da Escola de Química pelo excelente trabalho prestado no ensino de Engenharia Química.

Por fim, sou grato aos meus orientadores. Ao prof. Príamo Melo, por ter confiado em mim desde o começo da minha vida profissional e por ter aberto as portas para o conhecimento acadêmico. Ao prof. Maurício Bezerra, por guiar meus primeiros passos na disciplina e por ter acreditado nas minhas capacidades como engenheiro. À Júlia Nogueira, por me oferecer ensinamentos do início ao fim deste projeto e pelo seu apoio durante toda esta trajetória.

*À minha família,*
*o tesouro mais valioso que possuo na terra.*

*"A medida de um homem é o que ele faz com o poder."*
*- Platão*

**Resumo do Projeto de Final de Curso apresentado à Escola de Química como parte dos requisitos necessários para obtenção do grau de Bacharel em Engenharia Química.**

## Avaliação de Sistemas de Detecção e Diagnóstico de Falhas Baseados em Aprendizado de Máquina em Cenários de Pré-Falha e Incrustação em Unidades de Tratamento de Águas Ácidas

Guilherme Faustino Pitanga
Fevereiro, 2022.
Orientadores: Prof. Príamo Albuquerque Melo Jr., D. Sc.
Prof. Maurício Bezerra de Souza Jr., D. Sc.
Júlia do Nascimento Pereira Nogueira, M. Sc.

Este trabalho tem como objetivo analisar os cenários de pré-falha e incrustação em uma Unidade de Tratamento de Água Ácida aplicado métodos de Inteligência Artificial. Foram utilizados os dados provenientes da simulação dinâmica do processo feito em *Aspen Plus Dynamics*® V10 realizado por Nogueira (2021), contendo mais de sessenta mil amostras das variáveis da unidade em operação normal e seis diferentes condições de falha. Este banco de dados foi previamente tratado para adição de ruídos e tempos de atraso. Para esses dados, os métodos que obtiveram melhor desempenho foram *Random Forest* (RF) e Máquinas de Vetores de Suporte (SVM) (NOGUEIRA, 2021). Os resultados foram apresentados e analisados mediante as métricas estatísticas adequadas para o problema de classificação em Aprendizado de Máquinas. Os resultados apontam que, para o cenário de incrustação, o método SVM Linear apresenta a maior acurácia, 88,45%. Para o cenário de pré-falha, foi feita primeiramente a separação das amostras de operação normal da região de falha e pré-falha. Nesta etapa, o melhor método foi RF, com uma acurácia de 94,77%. Em seguida, foi feita a classificação para dois cenários distintos: uma única região de pré-falha e uma região de pré-falha para cada falha. Os melhores métodos para cada um desses cenários foram: SVM Linear (98,20%) e SVM Gaussiano (95,59%), respectivamente.

**Palavras-chave:** *Água Ácida, Pré-Falha, Incrustação, Inteligência Artificial, Random*

*Forest, Máquinas de Vetores de Suporte*

**Abstract of a Final Course Project presented to Escola de Química as partial fulfillment of the requirements for the degree of Bachelor of Chemical Engineering.**

## Machine Learning-Based Fault Detection and Diagnosis Systems Evaluation in Pre-Fault and Fouling Scenarios in Sour Water Treatment Units

Guilherme Faustino Pitanga
February, 2022.
Supervisors: Prof. Príamo Albuquerque Melo Jr., D. Sc.
Prof. Maurício Bezerra de Souza Jr., D. Sc.
Júlia do Nascimento Pereira Nogueira, M. Sc.

This work aims to study pre-fault and fouling scenarios in a Sour Water Treatment Unit using Artificial Intelligence methods. The benchmark applied in the present work, developed by Nogueira (2021) using *Aspen Plus Dynamics*® V10, consists of more than sixty thousand samples, encompassing both normal operation conditions and six fault conditions. This database was previously treated to add noise and delay times. For these data, the methods that performed better were Random Forest (RF) and Linear and Gaussian Support Vector Machines (SVM) (NOGUEIRA, 2021). The results are presented and analyzed through the suitable statistical metrics for classification problems in Machine Learning and through Python's scikit-learn library. Linear SVM presented the greater accuracy, 88.45%, for the fouling scenarios analysis. Regarding the pre-fault scenarios study, the samples were initially split between normal operation and the fault and pre-fault region. At this stage, the best method was RF, presenting accuracy of 94.77%. Then, the classification was performed in two distinct scenarios: a single pre-fault region and one pre-fault region for each existing fault. The best methods for each scenario were Linear SVM (98.20%) and Gaussian SVM (95.59%), respectively.

**Key-words:** *Sour Water, Pre-Fault, Fouling, Artificial Intelligence, Random Forest, Suport Vector Machines*

# Lista de Símbolos e Abreviações

| | |
|---|---|
| ARU | *Amine Regeneration Unit* (Unidade Regeneradora de Amina) |
| AI | *Artificial Intelligence* (Inteligência Artificial) |
| CNN | *Convolutional Neural Network* (Rede Neuronal Convolucional) |
| DS | *Data Science* (Ciência de Dados) |
| FDD | *Fault Detection and Diagnosis* (Detecção e Diagnóstico de Falhas) |
| FDDC | *Fault Detection, Diagnosis and Correction* (Detecção, Diagnóstico e Correção de Falhas) |
| GA | *Genetic Algorithm* (Algoritmo Genético) |
| KNN | *k-Nearest Neighbour* (k-Vizinhos Próximos) |
| NN | *Neural Network* (Rede Neuronal) |
| ML | *Machine Learning* (Aprendizado de Máquina) |
| MLP | *Multilayer Perceptron* (Perceptron Multicamadas) |
| PCA | *Principal Component Analysis* (Análise de Componentes Principais) |
| RF | *Random Forest* |
| SP | *Set-point* (Valor-alvo) |
| SPC | *Statistical Process Control* (Controle Estatístico de Processo) |
| SOM | *Self-Organizing Map* (Mapa Auto-Organizável) |
| SVM | *Support Vector Machine* (Máquina de Vetores de Suporte) |
| URE | Unidade Recuperadora de Enxofre |
| UTAA | Unidade de Tratamento de Águas Ácidas |
| VI | *Variable Importance* (Importância de Variável) |

# Sumário

**Lista de Figuras** **xiv**

**Lista de Tabelas** **xvii**

**1 Introdução** **1**
1.1 Motivação . . . 1
1.2 Objetivos . . . 3
1.3 Estrutura do trabalho . . . 3

**2 Revisão da Literatura** **4**
2.1 Unidades de Tratamento de Águas Ácidas . . . 4
2.2 Detecção, Diagnóstico e Correção de Falhas . . . 8
2.3 Inteligência Artificial . . . 13
2.3.1 Máquinas de Vetores de Suporte . . . 16
2.3.2 *Random Forest* . . . 17

**3 Metodologia** **18**
3.1 Histórico de Processo Simulado . . . 18
3.1.1 Simulação . . . 18
3.1.2 Simulação das Falhas . . . 21
3.1.3 Banco de Dados . . . 23
3.2 Cenários de Incrustação . . . 26
3.3 Cenários de Pré-Falha . . . 27
3.4 Inteligência Artificial . . . 29
3.4.1 Algoritmos de Inteligência Artificial . . . 30
3.4.2 Hiperparâmetros . . . 31
3.4.3 Métricas . . . 31

**4 Resultados e Discussão** **35**
4.1 Avaliação dos Algoritmos Utilizando os Bancos de Dados Originais . . . 35
4.2 Cenários de Incrustação . . . 41

4.2.1 Cenário de Incrustação com 13 Falhas . . . . . 41
4.2.1.1 Banco de Dados para o Cenário de Incrustação . . . . . 41
4.2.1.2 *Random Forest* . . . . . 42
4.2.1.3 Máquinas de Vetores de Suporte Linear . . . . . 46
4.2.1.4 Máquina de Vetores de Suporte Gaussiano . . . . . 48
4.2.2 Cenário de Incrustação com 12 Falhas . . . . . 50
4.2.2.1 Banco de Dados para o Cenário de Incrustação . . . . . 50
4.2.2.2 *Random Forest* . . . . . 51
4.2.2.3 Máquinas de Vetores de Suporte Linear . . . . . 55
4.2.2.4 Máquinas de Vetores de Suporte Gaussiano . . . . . 57
4.2.3 Análise Geral dos Cenários de Incrustação . . . . . 58
4.3 Cenários de Pré-Falha . . . . . 60
4.3.1 Banco de Dados para os Cenários de Pré-Falha . . . . . 60
4.3.2 *Random Forest* . . . . . 61
4.3.3 Máquinas de Vetores de Suporte Linear . . . . . 71
4.3.4 Máquinas de Vetores de Suporte Gaussiano . . . . . 76
4.3.5 Análise Geral dos Cenários de Pré-Falha . . . . . 82

**5 Conclusões e Considerações Finais** **84**

**Bibliografia** **85**

**A Fluxograma da Simulação Dinâmica** **90**

# Lista de Figuras


2.1 Unidade de Tratamento de Águas Ácidas com uma coluna . . . . . 5
2.2 Unidade de Tratamento de Águas Ácidas com duas colunas . . . . . 7
2.3 Esquema de um sistema de Detecção, Diagnóstico e Correção de Falhas . . 10
2.4 Diagrama de Venn definindo a Ciência de Dados como a interseção de três competências . . . . . 14
2.5 Diagrama representando os principais campos de estudo da Inteligência Artificial e da Ciência de Dados . . . . . 15
2.6 Esquema de SVM Linear para um exemplo de classificação binária . . . . . 16
2.7 Esquema geral de um algoritmo de *Random Forest* . . . . . 17

3.1 Fluxograma simplificado da simulação dinâmica da Unidade de Tratamento de Águas Ácidas com duas torres . . . . . 19
3.2 Fluxograma da simulação dinâmica da Unidade de Tratamento de Águas Ácidas com duas torres . . . . . 20
3.3 Comportamento dinâmico de C2DL com a) Sigma = 0,003 e b) Sigma = 0,001 amplitude de adição de ruído . . . . . 25
3.4 Esquema para obtenção do conjunto de dados de treino para o cenário de pré-falha . . . . . 29
3.5 Comparação dos métodos de AI aplicados para o banco de dados original . 30
3.6 Matriz de Confusão para um Problema de Classificação Binária . . . . . 32

4.1 Matriz de Confusão obtida por RF para o Banco de Dados de Nogueira (2021) . . . . . 36
4.2 Importânica de Variável obtida por RF para o Banco de Dados de Nogueira (2021) . . . . . 36
4.3 Matriz de Confusão obtida por RF com Redução de Variáveis para o Banco de Dados de Nogueira (2021) . . . . . 38
4.4 Importânica de Variável obtida por RF com Redução de Variáveis para o Banco de Dados de Nogueira (2021) . . . . . 38
4.5 Matriz de Confusão obtida por SVM Linear para o Banco de Dados de Nogueira (2021) . . . . . 39

4.6 Matriz de Confusão obtida por SVM Gaussiano para o Banco de Dados de Nogueira (2021) . . . 40
4.7 Matriz de Confusão obtida por RF para o Cenário de Incrustação de 13 Falhas . . . 43
4.8 Importância de Variável obtida por RF para o Cenário de Incrustação de 13 Falhas . . . 44
4.9 Matriz de Confusão obtida por RF com Redução de Variáveis para o Cenário de Incrustação de 13 Falhas . . . 45
4.10 Importância de Variável obtida por RF com Redução de Variáveis para o Cenário de Incrustação de 13 Falhas . . . 46
4.11 Matriz de Confusão obtida por SVM Linear para o Cenário de Incrustação de 13 Falhas . . . 48
4.12 Matriz de Confusão obtida por SVM Gaussiano para o Cenário de Incrustação de 13 Falhas . . . 49
4.13 Matriz de Confusão obtida por RF para o Cenário de Incrustação de 12 falhas . . . 52
4.14 Importância de Variável obtida por RF para o Cenário de Incrustação de 12 falhas . . . 53
4.15 Matriz de Confusão obtida por RF para o Cenário de Incrustação de 12 falhas . . . 54
4.16 Importância de Variável obtida por RF para o Cenário de Incrustação de 12 falhas após a redução das variáveis . . . 55
4.17 Matriz de Confusão obtida por SVM Linear para o Cenário de Incrustação com 12 falhas . . . 56
4.18 Matriz de Confusão obtida por SVM Gaussiano para o Cenário de Incrustação com 12 Falhas . . . 58
4.19 Matriz de Confusão obtida por RF para o Cenário de uma Região de Pré-Falha . . . 62
4.20 Importância de Variável obtida por RF para o Cenário de uma Região de Pré-Falha . . . 62
4.21 Matriz de Confusão obtida por RF para o Cenário de Cinco Regiões de Pré-Falha . . . 64
4.22 Importância de Variável obtida por RF para o Cenário de Cinco Regiões de Pré-Falha . . . 64
4.23 Matriz de Confusão obtida por RF no Cenário de Falhas e Pré-Falhas (Etapa 1) . . . 66
4.24 Importância de Variável obtida por RF no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1) . . . 66

4.25 Matriz de Confusão obtida por RF no Cenário de única Região de Pré-Falha Sem Operação Normal . . . . . 68
4.26 Importância de Variável obtida por RF no Cenário de única Região de Pré-Falha Sem Operação Normal . . . . . 68
4.27 Matriz de Confusão obtida por RF no Cenário de única Região de Pré-Falha Sem Operação Normal . . . . . 70
4.28 Importância de Variável obtida por RF no Cenário de única Região de Pré-Falha Sem Operação Normal . . . . . 70
4.29 Matriz de Confusão obtida por SVM Linear no Cenário de uma única região de pré-falha . . . . . 72
4.30 Matriz de Confusão obtida por SVM Linear no Cenário de seis regiões de pré-falha . . . . . 73
4.31 Matriz de Confusão obtida por SVM Linear no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1) . . . . . 74
4.32 Matriz de Confusão obtida por SVM Linear no Cenário de uma única região de pré-falha sem operação normal . . . . . 75
4.33 Matriz de Confusão obtida por SVM Linear no Cenário de seis regiões de pré-falha sem operação normal . . . . . 76
4.34 Matriz de Confusão obtida por SVM Gaussiano no Cenário de uma única região de pré-falha . . . . . 77
4.35 Matriz de Confusão obtida por SVM Gaussiano no Cenário de seis regiões de pré-falha . . . . . 79
4.36 Matriz de Confusão obtida por SVM Gaussiano no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1) . . . . . 80
4.37 Matriz de Confusão obtida por SVM Gaussiano no Cenário de uma única região de pré-falha sem operação normal . . . . . 81
4.38 Matriz de Confusão obtida por SVM Gaussiano Cenário de seis regiões de pré-falha sem operação normal . . . . . 82

A.1 Versão ampliada do Fluxograma da simulação dinâmica da Unidade de Tratamento de Águas Ácidas com duas torres . . . . . 91

# Lista de Tabelas


2.1 Classificação das técnicas de Detecção e Diagnóstico de Falhas com base no conhecimento . . . . . 12

3.1 Legenda das correntes da simulação em Nogueira (2021) . . . . . 19
3.2 Controladores da simulação dinâmica em Nogueira (2021) . . . . . 21
3.3 Descrição e Limites das Falhas em Nogueira (2021) . . . . . 22
3.4 Variáveis responsáveis pela região de falha . . . . . 23
3.5 Distribuição das classes das amostras entre os conjuntos treino e teste do banco de dados de Nogueira (2021) . . . . . 26
3.6 Descrição das classes do conjunto de treino e teste de (NOGUEIRA, 2021) modificadas para o cenário de incrustação . . . . . 27
3.7 Descrição das classes do conjunto de treino e teste de Nogueira (2021) modificadas para o cenário de pré-falha (1ª Abordagem) . . . . . 28
3.8 Descrição das classes do conjunto de treino e teste de Nogueira (2021) modificadas para o cenário de pré-falha (1ª Abordagem) . . . . . 29

4.1 Relatório de Classificação gerado para RF no Banco de Dados de Nogueira (2021) . . . . . 35
4.2 Relatório de Classificação gerado para RF com Redução de Variáveis no Banco de Dados de Nogueira (2021) . . . . . 37
4.3 Relatório de Classificação gerado para SVM Linear no Banco de Dados de Nogueira (2021) . . . . . 39
4.4 Relatório de Classificação gerado para SVM Gaussiano no Banco de Dados de Nogueira (2021) . . . . . 40
4.5 Distribuição das 13 classes das amostras entre os conjuntos treino e teste do banco de dados de Nogueira (2021) modificado para o cenário de incrustação 41
4.6 Relatório de Classificação gerado para RF no Cenário de Incrustação de 13 Falhas . . . . . 42
4.7 Relatório de Classificação gerado para RF com Redução de Variáveis no Cenário de Incrustação de 13 Falhas . . . . . 45

4.8 Relatório de Classificação gerado para SVM Linear no Cenário de Incrustação de 13 Falhas . . . . . 47
4.9 Relatório de Classificação gerado para SVM Gaussiano no Cenário de Incrustação de 13 Falhas . . . . . 49
4.10 Descrição das classes do conjunto de treino e teste de Nogueira (2021) modificadas para o cenário de incrustação . . . . . 50
4.11 Distribuição das 12 classes das amostras entre os conjuntos treino e teste do banco de dados de Nogueira (2021) modificado para o cenário de incrustação 51
4.12 Relatório de Classificação gerado para RF no Cenário de Incrustação de 12 falhas . . . . . 51
4.13 Relatório de Classificação gerado para RF no Cenário de Incrustação de 12 falhas após a redução de variáveis . . . . . 53
4.14 Relatório de Classificação gerado para SVM Linear no Cenário de Incrustação com 12 falhas . . . . . 56
4.15 Relatório de Classificação gerado para SVM Gaussiano no Cenário de Incrustação . . . . . 57
4.16 Acurácia dos métodos aplicados para 12 e 13 falhas do cenário de incrustação 59
4.17 Relatório de Classificação gerado por RF no Cenário de uma Região de Pré-Falha . . . . . 61
4.18 Relatório de Classificação gerado por RF no Cenário de Cinco Regiões de Pré-Falha . . . . . 63
4.19 Relatório de Classificação gerado por RF são apresentados no Cenário de Falhas e Pré-Falhas (Etapa 1) . . . . . 65
4.20 Relatório de Classificação gerado por RF no Cenário de única Região de Pré-Falha Sem Operação Normal . . . . . 67
4.21 Relatório de Classificação gerado por RF no Cenário de Cinco Regiões de Pré-Falha Sem Operação Normal . . . . . 69
4.22 Relatório de Classificação gerado por SVM Linear no Cenário de uma única região de pré-falha . . . . . 71
4.23 Relatório de Classificação gerado por SVM Linear no Cenário de seis regiões de pré-falha . . . . . 72
4.24 Relatório de Classificação gerado por SVM Linear no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1) . . . . . 73
4.25 Relatório de Classificação gerado por SVM Linear no Cenário de uma única região de pré-falha sem operação normal . . . . . 74
4.26 Relatório de Classificação gerado por SVM Linear no Cenário de seis regiões de pré-falha sem operação normal . . . . . 75
4.27 Relatório de Classificação gerado por SVM Gaussiano no Cenário de uma única região de pré-falha . . . . . 77

4.28 Relatório de Classificação gerado por SVM Gaussiano no Cenário de seis regiões de pré-falha . . . . . 78
4.29 Relatório de Classificação gerado por SVM Gaussiano no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1) . . . . . 79
4.30 Relatório de Classificação gerado por SVM Gaussiano no Cenário de uma única região de pré-falha sem operação normal . . . . . 80
4.31 Relatório de Classificação gerado por SVM Gaussiano no Cenário de seis regiões de pré-falha sem operação normal . . . . . 81
4.32 Acurácia dos métodos aplicados para as etapas 1 e 2 do cenário de pré-falha 83

# Capítulo 1

# Introdução

## 1.1 Motivação

As correntes de água ácida são geradas nas refinarias de petróleo a partir do contato da fase orgânica, contendo compostos de enxofre e nitrogênio, com correntes aquosas que, por transferência de massa, absorvem esses componentes na forma de sulfeto de hidrogênio ($H_2S$) e amônia ($NH_3$). A água ácida é um dos principais rejeitos aquosos gerados em uma refinaria, por isso seu tratamento se faz necessário frente ao impacto ambiental que pode ser causado pelo efluente (KENSELL; QUINLAN, 1996).

Como resultado do aumento do processamento de petróleos pesados e das crescentes restrições ambientais, há um aumento na quantidade de compostos sulfonados e nitrogenados possuindo a exigência de um destino ambientalmente correto. As correntes de água ácida são águas de processo, por conseguinte são águas de alto custo para uma refinaria. Logo, o processo de esgotamento da água ácida tem surgido como um importante processo na indústria de refino de petróleo pela remoção de contaminantes, pois além de mitigar o impacto ambiental da exploração, a água tratada é reutilizada dentro da refinaria (WEILAND; HATCHER, 2012).

As Unidades de Tratamento de Águas Ácidas são responsáveis por retirar os contaminantes das correntes de água ácida, permitindo assim seu reuso ou descarte. As Unidades com a configuração de duas torres ou colunas de esgotamento são as mais adotadas para o tratamento de água ácida com alto teor de $NH_3$. Nesse esquema, a primeira coluna retira o enxofre, gerando uma corrente de topo rica em $H_2S$, denominada gás ácido. A segunda coluna separa a corrente de fundo da primeira torre em uma corrente de topo rica em amônia, denominada gás amoniacal, e uma corrente de fundo constituída de água tratada. O gás ácido é destinado para a Unidade Recuperadora de Enxofre e o gás amoniacal é destinado para a Unidade Regeneradora de Amina (BELATO; LIMA; ODDONE, 2002; QUINLAN; HATI, 2010).

Em caso de baixa eficiência de remoção de $H_2S$ na primeira torre, o enxofre será des-

tinado para a Unidade Regeneradora de Amina, onde o processo de combustão acarretará na emissão deste na forma de $SO_x$, contribuindo para a poluição atmosférica dado que esse composto é precursor da chuva ácida (EAP, 2021). Por outro lado, a elevada eficiência na remoção de $H_2S$ leva também ao aumento da concentração de $NH_3$ no gás ácido, o que resulta em problemas operacionais na Unidade Recuperadora de Enxofre, podendo levar à perda de eficiência relacionado ao entupimento e à formação de $NO_x$. Este é outro composto altamente danoso ao meio ambiente que, além de também ser precursor da chuva ácida, é precursor do efeito *smog* fotoquímico (EAP, 2021; COLBECK; MACKENZIE, 1994). Dessa forma, há um impacto direto da operação da Unidade de Tratamento de Águas Ácidas nas emissões de $SO_x$ da refinaria tanto pelo $H_2S$ não recuperado na primeira torre e destinado para a Unidade Regeneradora de Amina, quanto por falhas operacionais na Unidade Recuperadora de Enxofre (KNUST, 2013)

A separação seletiva de $H_2S$ e $NH_3$ na primeira coluna deve eliminar pelo menos 90% do $H_2S$ presente na água ácida de acordo com as leis brasileiras mantendo o limite operacional que pode ser destinado para a Unidade Recuperadora de Enxofre (CONAMA, 2011). Manter a alta recuperação de $H_2S$ na primeira torre e reduzir o teor de $NH_3$ no gás ácido são metas conflitantes e caracterizam uma estreita faixa de operação para as unidades com essa configuração. Pequenos distúrbios levam a uma variação na concentração das correntes, resultando em falhas que, em última análise, levam a compostos prejudiciais serem emitidos ao meio ambiente acima dos limites regulatórios.

Por vezes, as falhas nessas unidades atingem níveis catastróficos que não apenas comprometem o meio ambiente. Em casos mais graves, as falhas na planta podem comprometer a saúde dos operadores. Há registro de casos no Brasil em que houve intoxicação dos funcionários da planta devido ao vazamento desses gases e a interdição da refinaria a fim de evitar acidentes (G1 Rio, 2013; Agência Brasil, 2018).

A fim de mitigar estes problemas, um sistema de detecção e diagnósticos de falhas é uma alternativa para melhorar a operação da Unidade de Tratamento de Águas Ácidas. A capacidade da identificação das possíveis falhas nessa unidade possui grande potencial de reduzir a quantidade de $SO_x$ e $NO_x$ emitidas por uma refinaria, diminuindo o impacto da atividade petroquímica ao meio ambiente. Além disso, existe a melhoria das condições de segurança de processo e prevenção de perdas econômicas.

Essa crescente tendência pelo aumento da eficiência econômica e ambiental é cada vez mais presentes nos processos industriais. Assim, o estudo de sistemas de Detecção, Diagnóstico e Correção de Falhas tem se tornado cada vez mais importante. Esses sistemas têm como principal objetivo assegurar o sucesso das operações planejadas através da identificação de anormalidades no comportamento do processo e colaborar na assistência da tomada de decisões de correção (SARTORI et al., 2012).

A detecção e o diagnóstico precoce de falhas de processo enquanto a planta está operando dentro da região de controle pode evitar a progressão de condições de falha e reduzir

a perda de produtividade. Em particular, a indústria petroquímica perde anualmente 20 bilhões de dólares devido a falhas de processo e, atualmente, a Detecção, Diagnóstico e Correção de Falhas se tornou o principal problema a ser resolvido nessa indústria (VENKATASUBRAMANIAN et al., 2003). Esses sistemas podem aplicar diversas técnicas, contudo destaca-se o uso de técnicas baseadas em estatística e em Inteligência Artificial. No contexto de processos químicos, a Inteligência Artificial tem sido amplamente utilizada como método para resolver problemas de controle preditivo por vezes complexos e caracterizados pela não-linearidade.

## 1.2 Objetivos

Este trabalho tem como objetivo a análise de cenários para o estudo de caso de uma Unidade de Tratamento de Águas Ácidas com a configuração de duas torres. O método utilizado será o desenvolvimento de sistema de Detecção e Diagnóstico de Falha utilizando ferramentas de Inteligência Artificial. A análise será feita para os cenários os quais apresentaram a mais difícil identificação por esses sistemas no trabalho de Nogueira (2021), são eles: os cenários de pré-falha e de incrustação da unidade em questão.

Seguem abaixo listados os objetivos específicos propostos:

1. Desenvolvimento de códigos para a solução de problemas de classificação utilizando diferentes métodos de Inteligência Artificial, como Florestas Aleatórias (*Random Forests*) e Máquinas de Vetores de Suporte (*Support Vector Machines*);
2. Aplicação das técnicas no banco de dados desenvolvido por Nogueira (2021) para diferenciar falhas com e sem incrustação;
3. Análise das variáveis selecionadas;
4. Aplicação das técnicas no banco de dados desenvolvido por Nogueira (2021) para identificação da região de pré-falha;
5. Análise considerando uma única região de pré-falha; e
6. Análise considerando regiões de pré-falha para um evento específico.

## 1.3 Estrutura do trabalho

O trabalho foi dividido em cinco capítulos, a saber, Introdução, Revisão Bibliográfica, Metodologia, Resultados e Discussões, e Conclusões e Considerações Finais.

Primeiramente são apresentados no Capítulo 2 os conceitos teóricos que foram utilizados no desenvolvimento do trabalho. No Capítulo 3, são apresentados o estudo de caso e a metodologia para a resolução do mesmo, definindo a abordagem utilizada, assim como os métodos de aprendizado de máquinas aplicados. Os resultados e a discussão dos mesmos são apresentados no Capítulo 4. As conclusões e considerações finais são feitas no Capítulo 5.

# Capítulo 2

# Revisão da Literatura

## 2.1 Unidades de Tratamento de Águas Ácidas

A água ácida é um efluente da indústria petroquímica, e esse termo foi atribuído devido a presença de $H_2S$ e $CO_2$ na corrente aquosa. Porém, isto não é um indicativo do pH da solução, pois a presença de grandes quantidades de $NH_3$ eleva o pH da água ácida para valores acima de 7. Durante o refino do petróleo, correntes de água ácida são geradas a partir do contato de correntes aquosas com correntes de hidrocarbonetos contendo enxofre e nitrogênio, que são absorvidos pelas correntes aquosas na forma de $H_2S$ e $NH_3$. Contudo, além desses compostos, a água ácida pode também conter $CO_2$, fenóis e cianetos. Normalmente, a água ácida contém entre 300 e 12.000 ppm de $H_2S$ e entre 100 e 8.000 ppm de $NH_3$. Já a concentração de fenol não costuma ultrapassar 200 ppm. A concentração de cianetos varia consideravelmente e o $CO_2$ tende a ser apenas vestígios da composição (WEILAND; HATCHER, 2012).

Essa variação da composição da água ácida se deve a diferentes composições de petróleo ao redor do mundo. Fatores locais durante a formação do óleo alteram a sua composição, por isso petróleos de diferentes origens, isto é, de diferentes reservatórios, possuem composições diferentes. No contexto de uma refinaria, correntes de água podem entrar em contato com fases orgânicas em diversas etapas do processo, seja pela injeção direta de água líquida ou pela condensação de vapor d'água. Assim, diferentes composições de água ácida são possíveis dependendo da etapa do processo em que a corrente aquosa entrou em contato com o óleo (WEILAND; HATCHER, 2012).

Existem duas principais classificações para a água ácida: a água ácida fenólica e a água ácida não-fenólica. Para ambos os casos $H_2S$ e $NH_3$ são os principais contaminantes. A água ácida não-fenólica é geralmente produzida em unidades de hidrotratamento e pode conter também traços de $CO_2$. A água ácida fenólica possui, além dos contaminantes já mencionados em sua composição, fenóis e cianetos. As principais fontes de água ácida fenólica são a destilação atmosférica, a destilação a vácuo, o craqueamento catalítico e

o coqueamento retardado. Além das etapas já mencionadas, existem outras fontes que contribuem em menor quantidade na geração de água ácida nas refinarias. São elas: vasos de lavagem de carga de Unidades de Recuperação de Enxofre, purga e refluxo do topo de torre regeneradora de amina e torre de resfriamento de unidade de tratamento de gás residual (BARROS, 2016).

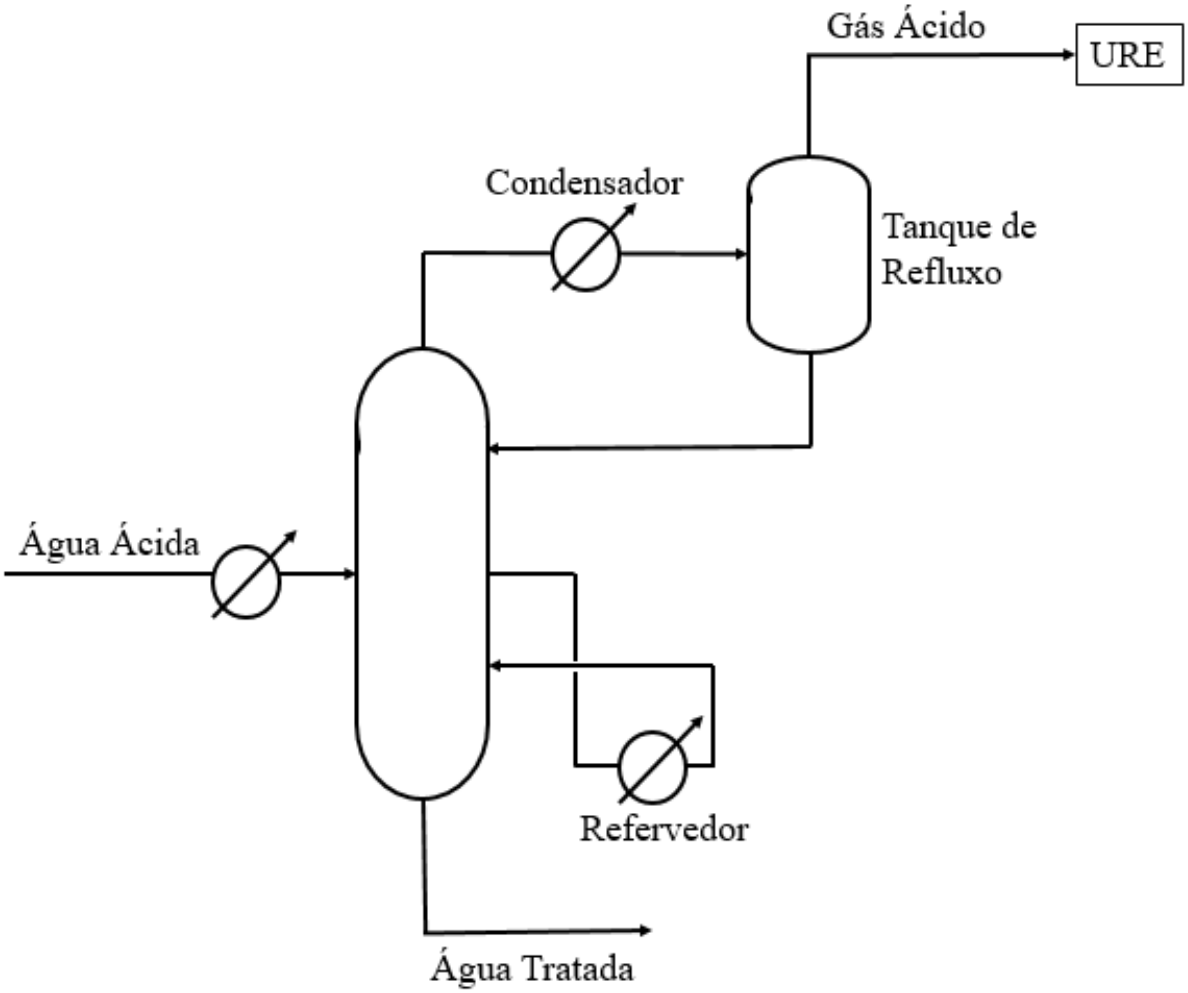


Figura 2.1: Unidade de Tratamento de Águas Ácidas com uma coluna baseado em Quinlan e Hati (2010).

A remoção dos contaminantes da água ácida é feita na Unidade de Tratamento de Águas Ácidas (UTAA). Dependendo da concentração dos contaminantes, pode ser recomendado o projeto de uma ou de duas torres. O projeto de UTAA de uma única coluna é o mais comum. Nesta configuração, a coluna realiza o processo de esgotamento (*stripping*) separando a fase líquida (corrente do fundo da coluna) dos contaminantes (corrente de topo da coluna). Os contaminantes são removidos pela elevação da temperatura e pelo abaixamento da pressão de vapor desses compostos. Isso é realizado através de um refervedor ou da injeção direta de vapor (BARROS, 2016). A corrente de topo da coluna, rica em $H_2S$, é chamada de gás ácido e é destinada para a Unidade Recuperadora de

Enxofre (URE). A Figura 2.1 apresenta um esquema simplificado da UTAA com uma única coluna.

As reações para sistemas com água ácida não-fenólica contendo apenas os contaminantes principais, $H_2S$ e $NH_3$, são descritos pelas Reações 2.1, 2.2, 2.3 e 2.4. Caso $CO_2$ esteja presente, o sistema é descrito pelas Reações 2.5, 2.6 e 2.7 também. A Reação 2.8 descreve a dissociação do cianeto (LEE et al., 2002).

$$2H_2O \longleftrightarrow H_3O^+ + OH^- \tag{2.1}$$

$$H_2S + H_2O \longleftrightarrow H_3O^+ + HS^- \tag{2.2}$$

$$HS^- + H_2O \longleftrightarrow H_3O^+ + {S_2}^- \tag{2.3}$$

$$NH_3 + H_2O \longleftrightarrow {NH_4}^+ + OH^- \tag{2.4}$$

$$CO_2 + 2\,H_2O \longleftrightarrow H_3O^+ + {HCO_3}^- \tag{2.5}$$

$${HCO_3}^- + H_2O \longleftrightarrow H_3O^+ + {CO_3}^{2-} \tag{2.6}$$

$$NH_3 + {HCO_3}^- \longleftrightarrow NH_2COO^- + H_2O \tag{2.7}$$

$$HCN + H_2O \longleftrightarrow H_3O^+ + CN^- \tag{2.8}$$

Uma informação importante é que $NH_3$ e $H_2S$ têm quase solubilidade ilimitada em água quando estão presentes juntos. Esta é uma consequência do fato de que o componente reativo do solvente, $NH_3$, é volátil e continuará absorvendo à medida que, se presente na fase gasosa, estiver sendo protonado como resultado da co-absorção de $H_2S$. Assim, é concebível que uma corrente de água ácida em particular possa ser muito mais concentrada em $NH_3$ e $H_2S$ do que a solubilidade de apenas um dos componentes sozinho (HATCHER; WEILAND, 2012).

O gás ácido é enviado para a URE, responsável pela produção de enxofre elementar que, além de reduzir a contaminação para o meio ambiente por esse processo, pode então ser comercializado. No Brasil, a maior parte desse enxofre elementar (mais de 90%) será utilizada na produção de ácido sulfúrico, cuja principal finalidade é a produção de

fertilizantes agrícolas fosfatados. O processo mais comumente utilizado para conversão do enxofre em enxofre elementar é o processo Claus. Neste processo, cerca de um terço do $H_2S$ é queimado em $SO_2$ usando ar. O $SO_2$ produzido então reage com $H_2S$ não queimado para formar enxofre elementar (QUINLAN; HATI, 2010).

Há ainda uma limitação operacional na URE em relação à concentração de $NH_3$ presente no gás ácido. Se o $NH_3$ não for suficientemente completamente convertido na combustão, os sais de $NH_3$–$H_2S$ podem se formar nos pontos frios, por exemplo, no condensador de enxofre final, e podem obstruir a URE. Além disso, a combustão incompleta de $NH_3$ aumenta a geração de $NO_x$ da refinaria. A alternativa para compensar a perda de eficiência no processo devido a altas concentrações de $NH_3$ é o aumento das dimensões da unidade envolvendo maior investimento financeiro (QUINLAN; HATI, 2010).

Refinarias tratando petróleos com alto teor de nitrogênio apresentam água ácida com alta concentração de $NH_3$. Neste caso, o projeto da UTAA de duas colunas é o mais recomendável. A primeira coluna separa a água ácida em gás ácido, rico em $H_2S$, e em água pré-tratada, rica em $NH_3$. A segunda coluna é alimentada com a corrente de fundo da primeira torre. Ela separa a água pré-tratada em gás amoniacal, rico em $NH_3$ na corrente de topo e em água tratada na corrente de fundo. A Figura 2.2 apresenta um esquema simplificado da UTAA com duas colunas.

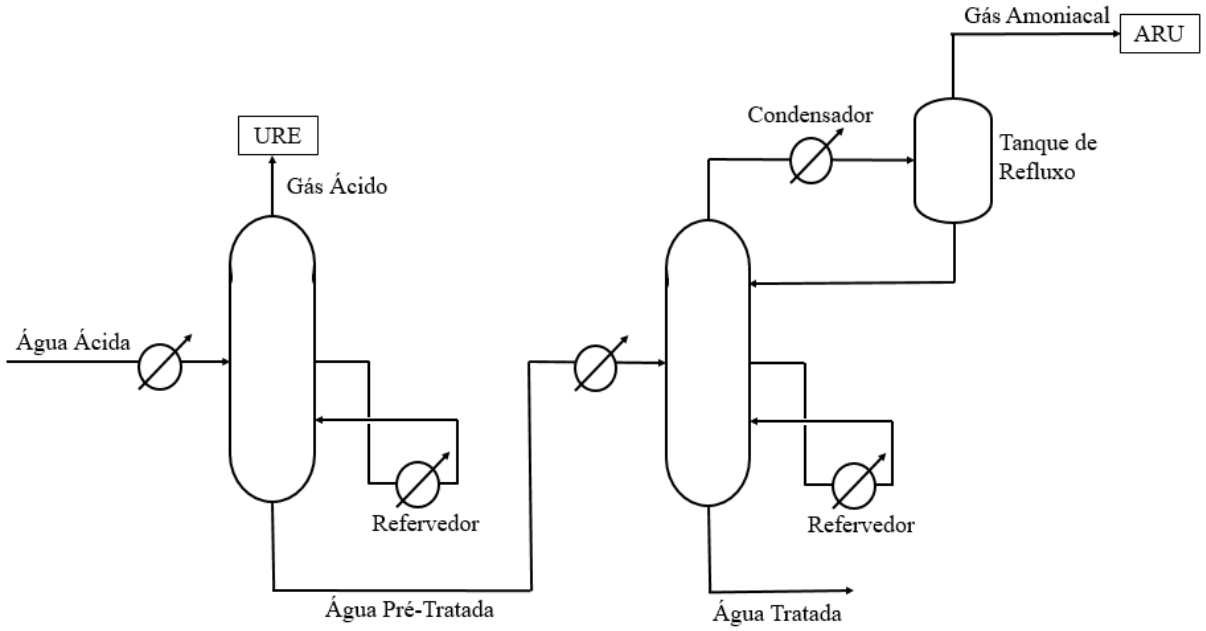


Figura 2.2: Unidade de Tratamento de Águas Ácidas com duas colunas baseado em Quinlan e Hati (2010).

O gás amoniacal é enviado para a Unidade de Regeneração de Amina (*Amine Regeneration Unit*, ARU). Nessa unidade, $NH_3$ é convertida em $N_2$. Contudo, em caso de falha na primeira coluna, $H_2S$ pode ser carregado para a segunda torre, removido na corrente de topo e destinado para ARU, onde será convertido em $SO_x$. Esse efeito também não é

desejado devido ao seu impacto ambiental. Logo, reduzir o conteúdo de $NH_3$ no gás ácido e a manutenção da alta recuperação de $H_2S$ na primeira coluna são objetivos em conflito.

De acordo com a legislação brasileira, a separação seletiva de $H_2S$ e $NH_3$ na primeira coluna deve eliminar pelo menos 90% do $H_2S$ presente em águas ácidas, mantendo os limites operacionais pré-determinados para a Unidade Recuperadora de Enxofre (CONAMA, 2011). Pequenas perturbações podem causar alterações significativas nas concentrações atuais que podem levar a falhas e levar à liberação de compostos prejudiciais ao meio ambiente acima dos limites regulatórios. Por isso, para unidades com configuração de duas colunas em particular, a faixa de operação é consideravelmente limitada.

Na literatura, são encontrados alguns trabalhos que utilizam a simulação de processos aplicada ao estudo de caso da UTAA. Para todos os trabalhos citados a seguir o *software Aspen Plus* foi a ferramenta escolhida para a realizar a simulação.

Lee et al. (2002) examinam o controle do processo da unidade em simulação estática e dinâmica. Knust (2013) utiliza a simulação para a realização da análise de superfícies de resposta visando desenvolver preditores aptos a gerar estimativas confiáveis da recuperação % de $H_2S$ e do teor de $NH_3$ no gás ácido. Silva, Alencar e Danielski (2014) fazem a modelagem do sistema pelo módulo RADFRAC e os modelos termodinâmicos ELECNRTL para fase líquida e Redlich-Kwong para fase vapor e buscam realizar uma análise das principais variáveis de controle do processo. Barros (2016) investiga o efeito das variáveis de processo na eficiência de remoção de $H_2S$, tendo como principal objetivo implementar um sensor virtual para monitoramento e controle da mesma. Morado (2019) utiliza modelos de superfície de resposta baseados em dados de pseudo-experimentos obtidos por simulação em estado estacionário, os modelos obtidos foram utilizados como ferramenta de monitoramento on-line e para implementação de estratégias de controle avançado.

## 2.2 Detecção, Diagnóstico e Correção de Falhas

Nos últimos 50 anos, a área de controle de processos químicos passou por grandes avanços. O controle regulatório, caracterizado por ações de controle simples como o abrir e fechar de válvulas, antes realizado exclusivamente por seres humanos, atualmente é realizado via automação, com o auxílio de computadores. Embora ainda haja muitas atividades manuais na indústria, isto é, atividades realizadas por operadores humanos, muito progresso tem sido feito nas áreas de sistemas de controle distribuído e de modelos de controle preditivo visando o aumento da eficiência na indústria química (VENKATASUBRAMANIAN et al., 2003).

Essa crescente necessidade da indústria em melhorar a qualidade dos produtos, enquanto satisfaz as especificações dos clientes e as rigorosas restrições ambientais e de segurança, torna diversos processos, que antes possuíam operação considerada adequada,

inaceitáveis. Para alcançar esse novo padrão, a indústria de processos químicos possui um determinado número de variáveis de processo consideradas em malhas de controle, como temperatura, pressão, vazão e composição das correntes de processo, assim como o nível de tanques e a carga térmica de colunas de destilação.

As variáveis de processo que se tem a intenção de manter em um determinado valor são chamadas de variáveis controladas. Esse controle é feito através das variáveis de processo que estão à disposição para serem manipuladas livremente, elas são denominadas de variáveis manipuladas. Enquanto aquelas sobre as quais não se tem controle são chamadas de variáveis distúrbio (ou variáveis perturbação) (BABATUNDE; RAY, 1994).

Os controladores de processo típicos, como por exemplo os PID, são projetados para manter a operação próxima dos valores-alvo (*set-points*, SP) através da compensação de distúrbios ou de alterações na magnitude de outras variáveis. Porém, há situações em que o controlador não é capaz de solucionar o problema, quando ocorrem desvios de uma ou mais variáveis de processo que não são permitidas e fogem à região controlável. A esses desvios dá-se o nome de falhas de processo (CHIANG; RUSSELL; BRAATZ, 2001).

Essas falhas podem ser causadas por distúrbios no processo devido a falhas em equipamentos como bombas, compressores, caldeiras, entre outros. Há ainda falhas que não estão relacionadas ao processo em si, mas à instrumentação (sensores, controladores e atuadores) responsável pelo controle do processo (VENKATASUBRAMANIAN et al., 2003).

Em busca de garantir a operação que satisfaça essas especificações de desempenho, a Detecção e Diagnóstico de Falhas (FDD) surge como uma alternativa para identificar quando e qual falha específica ocorreu. De uma maneira mais extensa, a Detecção, Diagnóstico e Correção de Falhas (FDDC) não apenas mantém os operadores e os funcionários de manutenção melhor informados do status da operação, mas, também, tem como objetivo colaborar na tomada de decisão para resolver esse comportamento anormal. Como resultado, espera-se a diminuição dos custos de produção e do número de paradas técnicas por falha de operação e a melhoria da segurança nas operações realizadas na planta (CHIANG; RUSSELL; BRAATZ, 2001).

As três atividades associadas à FDDC formam um sistema de malha fechada (Figura 2.3). A detecção de falhas indica que existe um comportamento inaceitável, sendo realizada por meio do registro de informações, do reconhecimento e da indicação de anormalidades no comportamento do sistema em um determinado instante. O diagnóstico de falhas é a determinação de qual ou quais falhas ocorreram, sendo assim responsável por indicar a causa da falha. O diagnóstico de falhas também pode determinar o tipo, a localização, a magnitude, o instante e o comportamento com o tempo. A correção de falhas consiste na remoção do efeito causador da falha, tomando as medidas apropriadas que vão variar de acordo com a gravidade do evento. Possíveis respostas do sistema incluem parada, mudanças na operação, reconfiguração, manutenção e reparo do sistema.

O objetivo final é a recuperação do processo e o retorno à operação ótima (CHIANG; RUSSELL; BRAATZ, 2001).

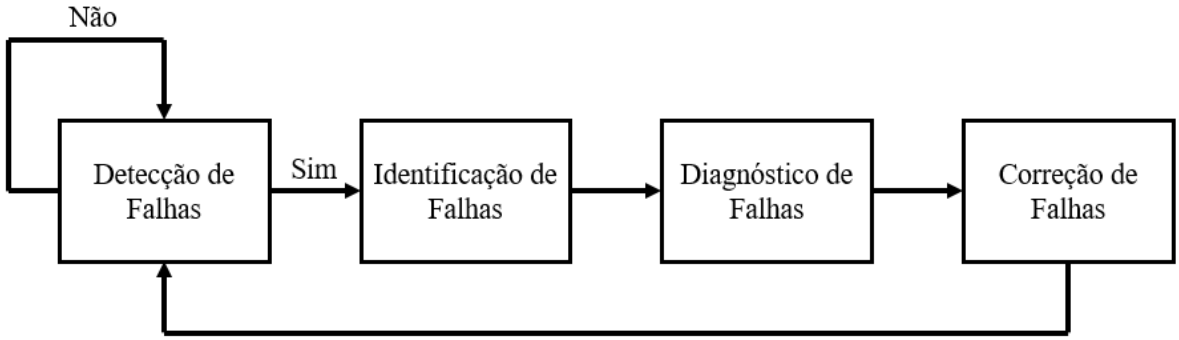


Figura 2.3: Esquema de um sistema de Detecção, Diagnóstico e Correção de Falhas baseado em Chiang, Russell e Braatz (2001).

Todos esses procedimentos são passíveis de implementação em um sistema, contudo não são todos simultaneamente mandatórios. Por exemplo, a falha pode ser diagnosticada sem a correção das variáveis de processo imediatamente afetadas. Além disso, apesar de cada vez mais desejada, a automação não é necessária para a implementação das três atividades (CHIANG; RUSSELL; BRAATZ, 2001).

Para Park, Fan e Hsu (2020), existem três diferentes abordagens para a Detecção e Diagnóstico de Falhas, são elas: a abordagem baseada em histórico de processo, a abordagem analítica e a abordagem baseada no conhecimento. A primeira é baseada em dados adquiridos diretamente da operação do processo. Conforme já mencionado, o aumento da instrumentação e a evolução dos computadores tem sido de principal destaque na indústria. Essa crescente instrumentação de processos industriais tem permitido que cada vez mais dados sejam obtidos de plantas de processos químicos. Aliado a esse fator, a capacidade de armazenamento e processamento dos computadores foi ampliada nesse mesmo período (VENKATASUBRAMANIAN et al., 2003). A complexidade e a quantidade de dados tornam inviável um operador ou engenheiro efetivamente extrair informações que não estejam explícitas sobre o processo. Para resolver isso, são utilizados métodos estatísticos computadorizados para reduzir a dimensão dos dados, assim capturando informações relevantes para o monitoramento de processo. A principal desvantagem dessa abordagem consiste na necessidade de um extenso banco de dados e na qualidade das informações registradas (PARK; FAN; HSU, 2020).

A abordagem analítica é baseada em modelos matemáticos gerados a partir de equações de balanço e conceitos teóricos. Pela necessidade de modelos, ela é particularmente útil para processos químicos que possuem dinâmica bem conhecida. Ela exige a compreensão completa do processo dentro do sistema de FDD, pois essa abordagem usa as relações entre as diversas variáveis de processo para agregar informação sobre os distúrbios para detectar cada falha (PARK; FAN; HSU, 2020). A grande vantagem desses métodos é o

entendimento físico e químico do processo de FDD. Contudo, essa abordagem apresenta algumas impraticabilidades, pois é difícil aplicar esses métodos em sistemas de grande escala, justamente por precisarem de um modelo bem detalhado do processo. Além disso, esses modelos são financeiramente dispendiosos para a indústria considerando todas as relações que precisam ser feitas nesse sistema de múltiplas variáveis (CHIANG; RUSSELL; BRAATZ, 2001).

A abordagem baseada em conhecimento busca aplicar a compreensão do processo, pois em processos com um grau de complexidade muito alto pode ser difícil desenvolver um sistema de FDD apenas baseado nos dados. Dessa maneira, essa abordagem utiliza conhecimentos e regras para aplicar sistemas de FDD para os processos industriais mais complicados. Essas técnicas são baseadas em modelos qualitativos, que podem ser obtidos a partir do conhecimento especializado, da análise causal e do reconhecimento de padrões. Assim como a abordagem analítica, sua aplicação é restrita a sistemas de menores dimensões e complexidade (PARK; FAN; HSU, 2020).

Há ainda muitos desafios para a abordagem baseada em histórico de processo. As características dos dados de processos químicos incluem: alta dimensionalidade, distribuição não gaussiana, não linearidade, comportamentos variados no tempo e autocorrelações entre os dados. Essas características tornam o diagnóstico de falha do processo químico uma tarefa desafiadora (SHU et al., 2016).

Além disso, a falta de dados de histórico de processo com a correta classificação do *status* de operação em determinado instante tem limitado a sua aplicação em FDD. Uma maneira de enfrentar esse obstáculo é utilizar dados de processos gerados por meio de simulações. Entretanto, nota-se que há um desvio intrínseco entre essa abordagem e a abordagem unicamente baseada em histórico de dados (LI et al., 2020).

Há uma necessidade em se trabalhar com um número de dados de falhas e operação normal bem definidos. Além disso, há também a crescente complexidade das plantas industriais e exigências ambientais e de segurança do trabalho já mencionadas. Ambos os fatores têm influenciado no surgimento de técnicas cada vez mais específicas e sofisticadas. Em Sartori et al. (2012), foram avaliados mais de 500 artigos a respeito de métodos para FDD. Na Tabela 2.1 abaixo, estes métodos foram separados entre baseados em modelos e em histórico de processo, sendo em ambos os casos divididos em métodos quantitativos e qualitativos.

Além de NN e PCA, destacam-se as seguintes técnicas: lógica fuzzy (difusa, nebulosa), filtro de Kalman, máquina de vetores de suporte (*Support Vector Machine*, SVM), algoritmo genético (*Genetic Algorithm*, GA) e controle estatístico de processos (*Statistical Process Control*, SPC). Essas técnicas respondem por quase 90% dos artigos avaliados, indicando uma grande relevância de técnicas consideradas de Inteligência Artificial (*Artificial Intelligence*, AI) como ferramentas de FDD (SARTORI et al., 2012).

Nesse contexto, Sartori et al. (2012) observou que em torno de 59% dos trabalhos

Tabela 2.1: Classificação das técnicas de Detecção e Diagnóstico de Falhas com base no conhecimento *a priori* utilizado em (SARTORI et al., 2012).

| **Modelos** | |
|---|---|
| Quantitativo | Qualitativo |
| • Observadores de estados e de saídas<br>• Equações e espaço de paridade<br>• Filtro de Kalman estendido (EKF, *Extended Kalman Filter*)<br>• Identificação e estimação de parâmetros | • Árvores de falha<br>• Simulação qualitativa (QSIM, *Qualitative Simulation*)<br>• Teoria qualitativa de processo (QPT, *Qualitative Process Theory*)<br>• Grafos direcionados com sinais (SDG, *Signed Directed Graph*) |
| **Histórico de Processo** | |
| Quantitativo | Qualitativo |
| • Classificadores estatísticos<br>• Redes neuronais (NN, *Neural Networks*)<br>• Análise de componentes principais (PCA, *Principal Component Analysis*)<br>• Método dos mínimos quadrados parciais (PLS, *Partial Least Squares*) | • Sistemas especialistas (ES, *Expert System*)<br>• Análise qualitativa de tendências (QTA, *Qualitative Trend Analysis*) |

foram aplicados a equipamentos da indústria de processos (reatores, colunas, sensores e atuadores) nos quais as técnicas de AI eram as mais utilizadas.

Ademais, na literatura, alguns trabalhos estudam o FDD em processos químicos aplicando Inteligência Artificial. Aqui serão citados os mais relevantes e recentes nesse tema.

Park, Fan e Hsu (2020) realizaram uma revisão da aplicação de sistemas de FDD em processos industriais, onde comenta muitos exemplos de Inteligência Articial apesar de poucos serem processos químicos. Fan et al. (2020) aplicaram *Random Forest* para determinar as variáveis mais importantes no desempenho do sistema de FDD. Aruntha-vanathan et al. (2020) buscaram identificar falhas desconhecidas através de aprendizado não-supervisionado. Ge, Song e Gao (2013) utilizam CNN para desenvolver um sistema de FDD em uma simulação de produção de ácido fórmico.

Os únicos trabalhos encontrados que aplicam Inteligência Artificial em um sistema de FDD em UTAA foram Nogueira, de SOUZA Jr e Melo (2021) e Nogueira (2021). No primeiro, o compartamento dinãmico da unidade foi simulada no *software Aspen Plus* para 5 condições de falha e foi aplicado *Random Forest* nos dados para a realização do FDD. No segundo, Nogueira (2021) utilizou o mesmo *software* para simular a unidade em 6 condições de falha e aplicou, além de *Random Forest*, outros métodos de inteligência artificial com o objetivo de identificar os melhores algoritmos para a identificação das classes, desenvolvendo o sistema de FDD aplicando esses métodos.

## 2.3 Inteligência Artificial

A 4ª Revolução Industrial, também chamada de Indústria 4.0, também está inserida no contexto de sistemas de FDD, pois está ligada à implementação desses métodos no ambiente industrial. A Indústria 4.0 é caracterizada pelos fatores que vêm tornando os sistemas de FDD mais acessíveis e viáveis. Na indústria química, ela tem permitido a redução no custo de sensores e estocagem de dados, aumentando a instrumentação disponível e a comunicação entre planta e operadores. Esse aumento nos dados de histórico de processo e a disponibilidade on-line está aumentando a demanda pela análise de dados (SOARES, 2017).

Apesar de se sobreporem, AI e Ciência de Dados (*Data Science*, DS) são campos diferentes. A DS é a área interdisciplinar com três principais competências: habilidades computacionais, conhecimento matemático e conhecimento de domínio, conforme ilustrado na Figura 2.4. As habilidades computacionais (*Hacking Skills*) estão relacionadas ao armazenamento e manipulação dos dados e envolve, por vezes, profunda programação, em particular com ênfase nas operações vetorizadas. O conhecimento matemático e estatístico (*Math & Statistics Knowledge*) se refere à implementação de métodos matemáticos e estatísticos para tratar desses dados. Por fim, o conhecimento de domínio (*Substantive Expertise*) destaca a importância do conhecimento especializado da aplicação que permite a plena compreensão dos dados estatísticos e é essencial para a análise dos resultados.

Um subtópico da DS é conhecido pelo termo em inglês *Big Data*, que descreve um conjunto de dados caracterizado pelos 3 V's do *Big Data*: Volume, Velocidade e Variedade. Volume representa a grande quantidade de dados que atualmente tem estado disponíveis, Velocidade representa a alta taxa na qual esses dados são capazes de serem gerados e Variedade são diferentes formas e fontes das quais os dados são coletados (texto, som, vídeo, sensores, etc). Esse tipo de informação torna praticamente inviável antigas metodologias senão aquelas introduzidas pela DS (CHIANG; LU; CASTILLO, 2017).

A AI é um conceito amplo que abrange múltiplos campos de conhecimento. Esses campos, por muitas vezes, se sobrepõem uns aos outros. Segundo Rich (1983), a AI pode ser definida como o estudo de como fazer computadores realizar tarefas nas quais, no momento, pessoas são melhores. Dessa maneira, algoritmos de AI procuram de alguma forma mimetizar habilidades nas quais os seres humanos possuem maior aptidão que as máquinas. O campo mais afluente em AI é o Aprendizado de Máquinas (*Machine Learning*, ML) e esses conceitos estão interligados ao contexto de *Big Data* e DS. Na Figura 2.5, estes são organizados na forma de um diagrama. Algumas aplicações de AI são: inteligência simbólica, prova de teoremas, robótica, visão computacional, sistemas especialistas, aprendizado de máquina, entre outros (RUSSELL; NORVIG, 2002).

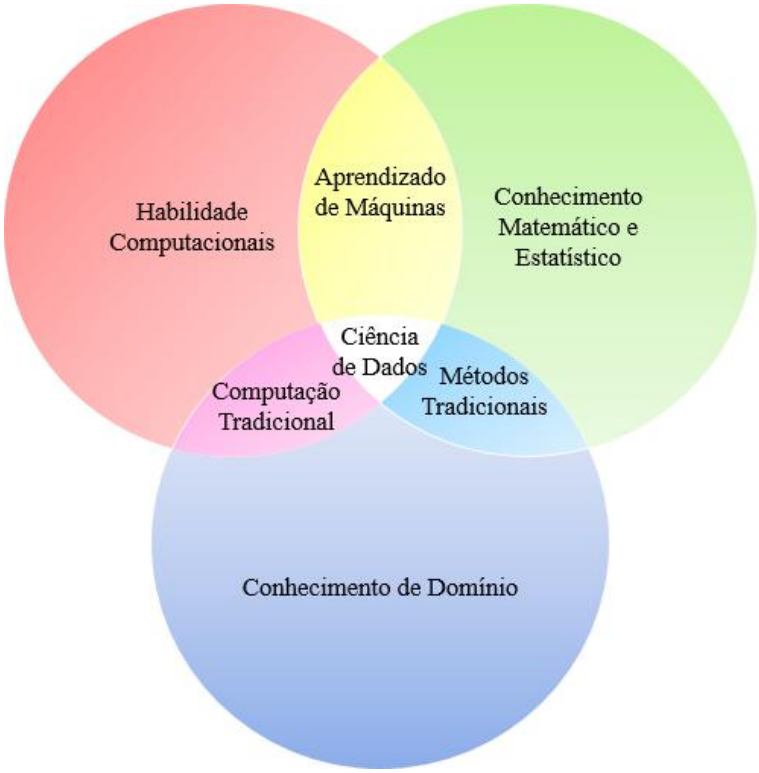


Figura 2.4: Diagrama de Venn definindo a Ciência de Dados como a interseção de três competências baseado em Conway (2015).

A definição mais difundida de ML é: o campo de estudo que fornece aos computadores a habilidade de aprender sem serem explicitamente programados (SAMUEL, 1959). Em processos químicos, técnicas de Aprendizado de Máquinas são classificadas como abordagem baseada em dados, pois essa habilidade de aprender advém de técnicas que fazem uso de bancos de dados para treinar um algoritmo para uma determinada tarefa, sem que haja a programação tradicional para aquela tarefa específica.

O aprendizado em ML pode ser dividido em três tipos: supervisionado, não supervisionado e por reforço. O aprendizado supervisionado consiste na alimentação do algoritmo com os dados e seus alvos (*targets*), isto é, o algoritmo recebe as entradas e as saídas esperadas. Através desta abordagem, o algoritmo deverá aprender as informações que permeiam este banco de dados. O aprendizado não supervisionado alimenta o algoritmo com as informações sem os alvos, é esperado que ele sozinho consiga encontrar os padrões dentro do próprio banco de dados. Já no aprendizado por reforço, o algoritmo possui um determinado objetivo e é premiado quando consegue alcançá-lo ou punido caso não obtenha êxito na tarefa (ALOM et al., 2019).

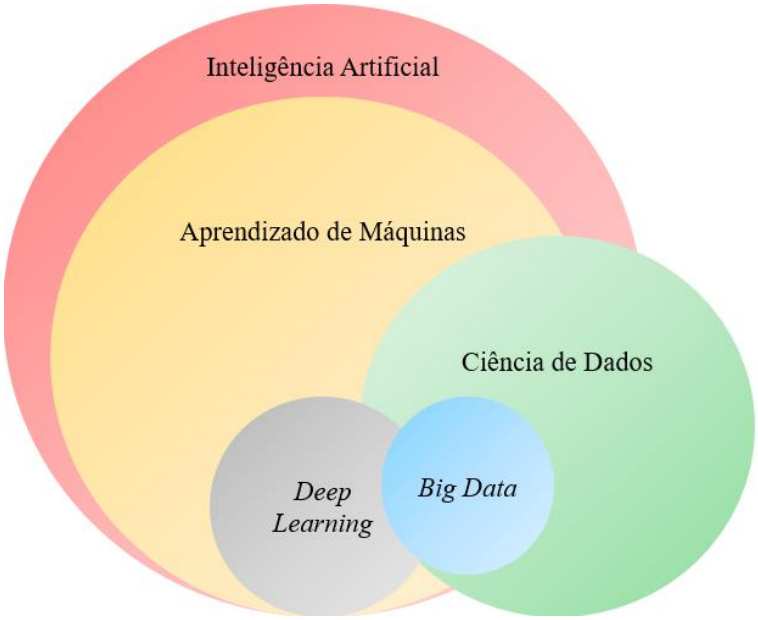


Figura 2.5: Diagrama representando os principais campos de estudo da Inteligência Artificial e da Ciência de Dados baseado em Thakur (2020).

Os problemas de aprendizado supervisionado em ML ainda podem ser separados em duas classificações principais: classificação e regressão. De forma simplificada, para problemas de classificação, o modelo retorna uma classe, também chamada de rótulo. Já para problemas de regressão, o modelo retorna um valor numérico contínuo.

A aplicação de um método de ML estabelece a divisão do conjunto de dados em três grupos: treinamento, validação e teste. Na etapa de treinamento, o conjunto de treinamento é apresentado como entrada a um novo modelo. Seus parâmetros são ajustados para prever os valores corretos através da minimização do erro. A depender do tipo de problema, diferentes métricas são utilizadas com essa finalidade. O conjunto de validação permite ajustar diferentes hiperparâmetros e é aplicado em algumas metodologias antes de realizar o teste para evitar o sobreajuste. Por fim, é realizada a entrada do conjunto de teste para avaliar o modelo (CHOLLET, 2021).

O sobreajuste ocorre quando um modelo estabelece uma análise muito próxima ou idêntica a um determinado conjunto de dados de treinamento, isto é, o modelo se ajusta de forma exacerbada aos dados de treinamento. Quando isso ocorre, o modelo não consegue mais ajustar dados adicionais ou prever observações futuras de forma confiável. Essa perda da capacidade de generalização de um modelo pode acontecer em qualquer problema de ML. O oposto do sobreajuste também pode acontecer, o subajuste ocorre caso o modelo não seja treinado o suficiente e é necessário aumentar a quantidade de dados de treinamento (CHOLLET, 2021).

Como visto anteriormente, há muitos métodos ou algoritmos de Aprendizado de Máquinas. Todos apresentam algumas semelhanças entre si, eles são programas que apresentam uma estrutura básica simples e flexível que se adapta a diferentes problemas

propostos. A metodologia de aplicação para diferentes métodos recebe poucas alterações em relação à divisão do conjunto de dados (SOARES, 2017).

Aqui serão apresentados apenas dois métodos que foram utilizados neste trabalho: *Random Forest* e Máquinas de Vetores de Suporte, que foram os melhores métodos para identificar as classes do banco de dados utilizado por Nogueira (2021).

### 2.3.1 Máquinas de Vetores de Suporte

Máquina de Vetores de Suporte é uma técnica de ML de aprendizado supervisionado. Esse método foi desenvolvido por Cortes e Vapnik (1995) nos laboratórios AT&T Bell. Para resolver problemas de classificação, este algoritmo busca encontrar o melhor hiperplano que separa os dados em diferentes categorias. Os dados são organizados em uma representação dimensional, que é seu espaço ambiente. Um hiperplano é um subespaço cuja dimensão é uma a menos que a do seu espaço ambiente. Por exemplo, para um espaço ambiente de duas dimensões, o hiperplano de separação seria uma reta (Figura 2.6). Para classificar novos dados, o algoritmo apenas verifica a posição da amostra em relação ao hiperplano (CHOLLET, 2021).

Um bom hiperplano de separação é calculado maximizando a distância entre o hiperplano e os dados mais próximos. Isso é chamado de maximização da margem e permite que o modelo seja mais capaz de generalizar dados além do conjunto de treinamento (CHOLLET, 2021).

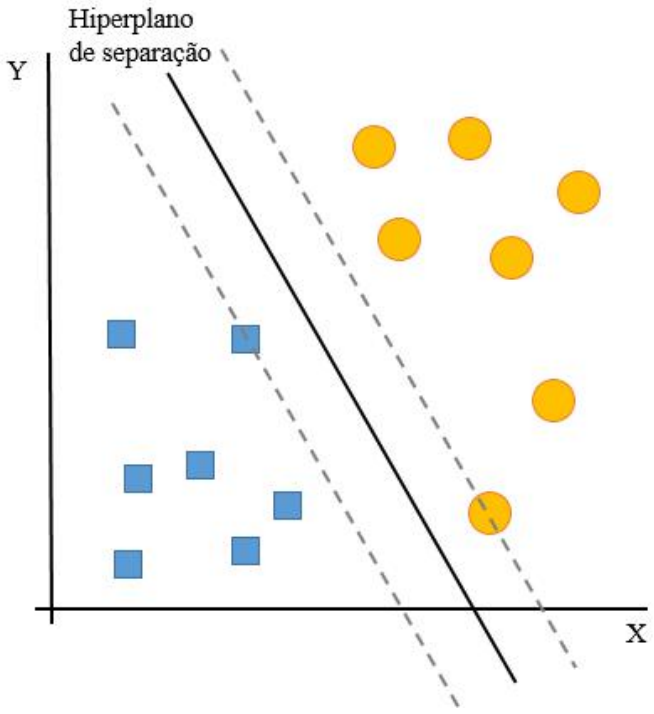


Figura 2.6: Esquema de SVM Linear para um exemplo de classificação binária baseado em Lorena e Carvalho (2007).

### 2.3.2 *Random Forest*

*Random Forest* (RF) é uma técnica de ML de aprendizado supervisionado, proposta pela primeira vez por Breiman (2001). Esse algoritmo constrói múltiplas árvores de decisão (*Decision Trees*) durante o treinamento. Árvores de decisão são estruturas similares a fluxogramas, cada nó possui uma entrada e duas saídas, podendo possuir muitos nós até retornar os valores de saídas finais, nas folhas. Uma árvore de decisão realiza a classificação ou predição do valor de saída a partir das entradas. A RF corrige o sobreajuste, que frequentemente ocorre com as árvores de decisão, justamente por contar com uma grande quantidade de árvores (CHOLLET, 2021). A Figura 2.7 ilustra uma RF e suas árvores de decisão.

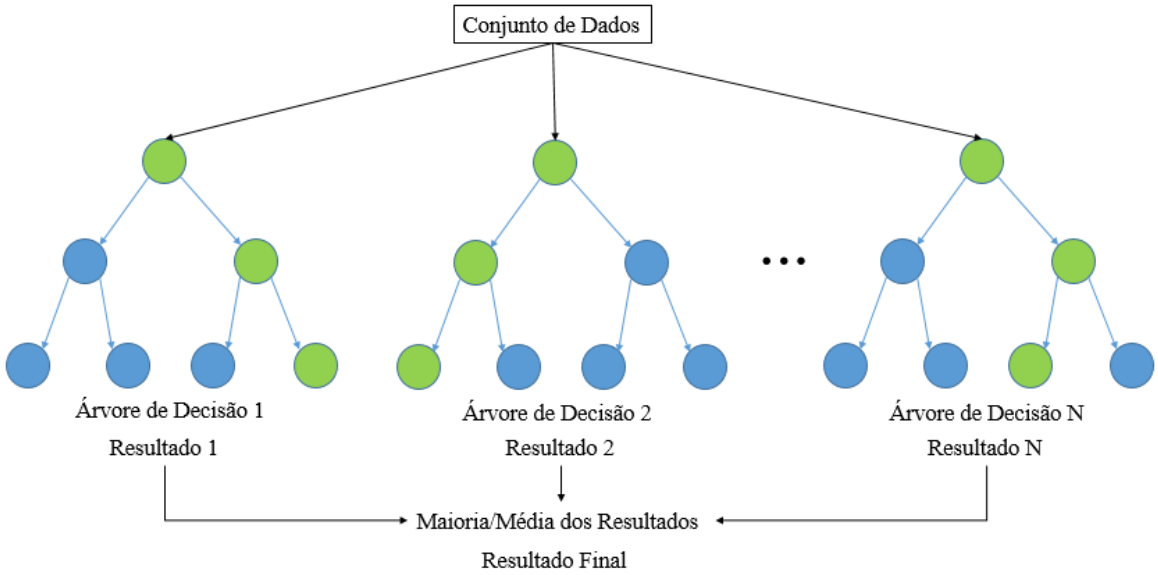


Figura 2.7: Esquema geral de um algoritmo de *Random Forest*.

Para tarefas de classificação, a saída da RF é a classe selecionada pela maioria das árvores. Para tarefas de regressão, a saída é a média ou previsão média das árvores (CHOLLET, 2021). Porém, a biblioteca *scikit-learn* faz uma adaptação a essa técnica, ao invés de cada árvore ter um voto, cada árvore passa a ter uma probabilidade de predição. Essa modificação diminui as chances de sobreajuste (PEDREGOSA et al., 2011).

# Capítulo 3

# Metodologia

O presente trabalho se baseia nos dados de histórico de processo simulados por Nogueira (2021). Nogueira (2021) simulou uma UTAA com duas colunas de esgotamento para tratar uma corrente de água ácida não-fenólica. Devido à sigilosidade dos dados históricos de processos industriais, a alternativa foi utilizar dados provenientes de uma simulação dinâmica. Esse processo foi simulado através do *software de simulação* Aspen Plus®. A simulação de Nogueira (2021) produziu um extenso banco de dados com ampla faixa de operação normal e as principais falhas de processo avaliadas. A primeira parte deste capítulo descreve essa simulação e detalhes do banco de dados.

A partir desse banco de dados, foram feitas alterações para possibilitar a aplicação de métodos de ML para cenários além dos estudados em Nogueira (2021). Esses cenários foram de condição de incrustação e de pré-falha. Em seguida foram aplicadas as metodologias e suas métricas foram analisadas. A segunda parte deste capítulo descreve em detalhes a aplicação das técnicas e suas principais métricas.

## 3.1 Histórico de Processo Simulado

### 3.1.1 Simulação

Nogueira (2021) propôs a simulação em Aspen Plus Dynamics® V10 da UTAA usando o modelo GPSWAT (*Gas Process Association Sour Water Equilibrium*) para a realização da simulação. Esse modelo é apresentado como sendo o mais adequado para aplicações em água ácida. Nessa simulação, há trinta e quatro correntes materiais e sete diferentes composições descritas no processo. A Tabela 3.1 correlaciona as siglas dessas correntes com sua composição.

Além das correntes de água ácida, gás ácido e gás amoniacal, que já foram mencionadas, existem outros 4 tipos de correntes aquosas. A Água Pré-Tratada é a corrente de fundo da primeira coluna, logo é uma corrente rica em $NH_3$. A Água Tratada é a corrente

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

| Sigla | Composição da Corrente em Inglês | Composição da Corrente em Português |
|---|---|---|
| SW | *Sour Water* | Água Ácida |
| PW | *Pretreated Water* | Água Pré-Tratada |
| TW | *Treated Water* | Água Tratada |
| NW | *New Water* | Água Nova |
| CW | *Clean Water* | Água Limpa |
| ACG | *Acid Gas* | Gás Ácido |
| AMG | *Amoniacal Gas* | Gás Amoniacal |

de fundo da segunda coluna. Essa corrente possui pequenas quantidades de contaminantes solubilizadas. A Água Nova é uma corrente aquosa sem contaminantes. A Água Limpa é a mistura das últimas duas correntes. Todas essa correntes e os principais equipamentos estão presentes na versão simplificada da simulação apresentada na Figura 3.1, onde H1, H2 e H3 são os trocadores de calor e C1 e C2 são as colunas de esgotamento.

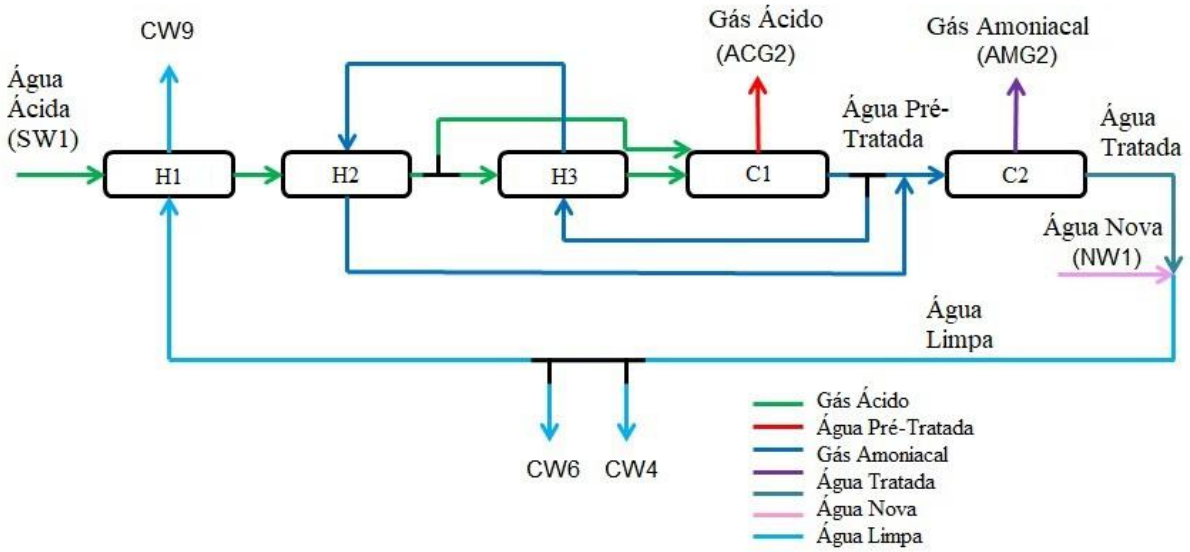


Figura 3.1: Fluxograma simplificado da simulação dinâmica da Unidade de Tratamento de Águas Ácidas com duas torres. Fonte: Nogueira (2021).

A simulação dinâmica é composta por um total de 14 controladores. Os controladores foram adicionados com o objetivo de garantir a estabilidade e a convergência da simulação. A Figura 3.2 apresenta o fluxograma da simulação dinâmica com esses controladores. Essa imagem pode ser encontrada em uma versão ampliada no Apêndice A.

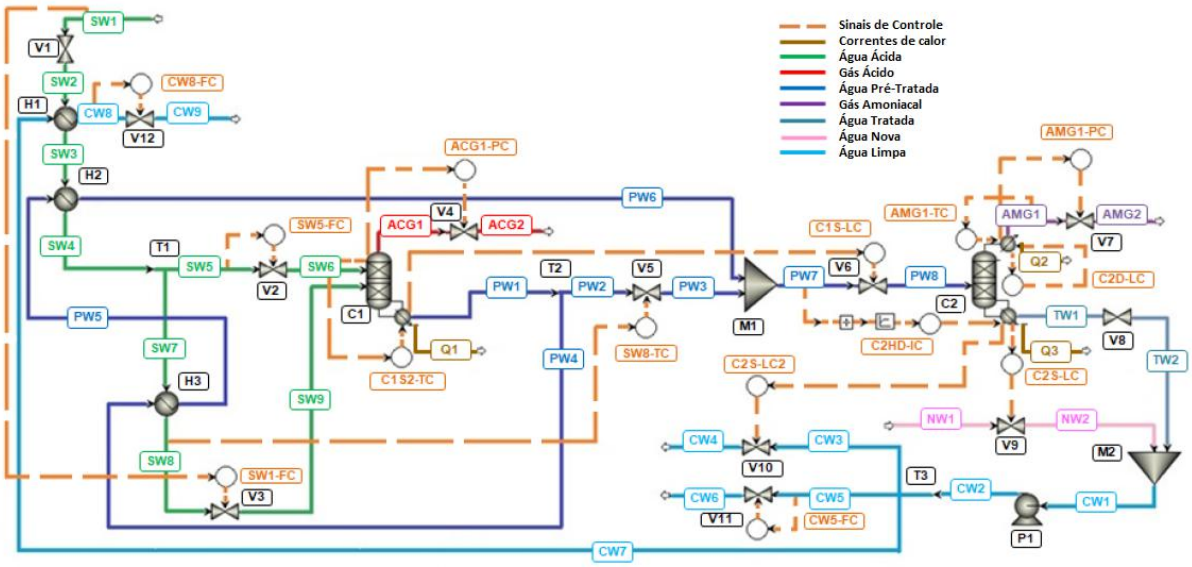


Figura 3.2: Fluxograma da simulação dinâmica da Unidade de Tratamento de Águas Ácidas com duas torres. Fonte: Nogueira (2021) (legenda traduzida).

A simulação possui uma entrada de SW1, que após a válvula V1, torna-se SW2. SW2, por integração energética com CW7, é aquecida no trocador de calor H1 e torna-se SW3. SW3, por integração energética com PW5, é aquecida no trocador de calor H2 e torna-se SW4. A SW4 é dividida em duas correntes, SW5 e SW7. SW5, após a válvula V2, torna-se SW6 e entra na alimentação superior da coluna C1. SW7, por integração energética com PW4, é aquecida no trocador de calor H3 e torna-se SW8. SW8, após a válvula V3, torna-se SW9 e entra na alimentação inferior da coluna C1.

A coluna C1 é aquecida pela carga térmica do refervedor, Q1, e separa a corrente de topo AGC1 da corrente de fundo PW1. AGC1, após a válvula V4, torna-se AGC2 e é retirado do sistema, enquanto PW1 é dividida em duas correntes, PW2 e PW4. PW4, por integração energética com SW7, é resfriada no trocador de calor H3 e torna-se PW5. PW5, por integração energética com SW3, é resfriada em H2 e torna-se PW6. PW2, que após a válvula V5, torna-se PW3 é misturada com a PW6, formando PW7. Nota-se que PW7 tem vazão mássica igual à vazão mássica de PW1, pois as correntes que se dividiram foram misturadas novamente sem haver perdas. Após a válvula V6, PW7 torna-se PW8 e entra na alimentação da coluna C2.

A coluna C2 possui um condensador (carga térmica do condensador é Q2) e um refervedor (carga térmica do refervedor é Q3). A corrente de topo da coluna C2 é AMG1 e a corrente de fundo é TW1. AMG1, após a válvula V7, torna-se AMG2 e é retirado do sistema. TW1, após a válvula V8, torna-se TW2. Nesse ponto entra no sistema NW1, que após a válvula V9, torna-se NW2. TW2 é misturada à NW2, formando a corrente CW1. Após passar pela bomba P1, a corrente 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, por integração energética, é resfriada no trocador de calor H2 e torna-se CW8. CW8, após a válvula V12, torna-se CW9 e é retirada do sistema.

A simulação possui duas colunas do tipo "RadFrac", método rigoroso para destilação do *software*, e três trocadores de calor do tipo "HeatX", que exige duas correntes para realização da transferência de calor entre elas. As colunas C1 e C2 são constituídas de 5 e 6 estágios, respectivamente. 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 aceitar as fases líquida e vapor. As válvulas V4 e V7 foram configuradas para aceitar apenas vapor. As demais válvulas da simulação aceitam apenas a fase líquida. Todas elas possuem uma perda de carga associada.

O controle foi feito baseado na literatura e através do conselho de especialistas na indústria (NOGUEIRA, 2021). Após a sintonização dos parâmetros dos controladores, Nogueira (2021) alcançou a recuperação de $H_2S$ um pouco acima de 89%, próximo ao valor exigido pela legislação. A Tabela 3.2 descreve as variáveis controladas e manipuladas para cada um desses controladores.

Tabela 3.2: Controladores da simulação dinâmica em 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 Reboiler da C1 |
| AMG1-TC | Temperatura de AMG1 | Carga do Condensador da C2 |
| C1S-LC | Nível do *Sump* da C1 | % Abertura da V6 |
| C2D-LC | Nível do Vaso de Refluxo da C2 | Vazão Mássica do Vaso de Refluxo da C2 |
| C2S-LC | Nível do *Sump* da C2 | % Abertura da V9 |
| C2S-LC2 | Redundância do Nível do *Sump* 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 de C2 |

### 3.1.2 Simulação das Falhas

Após o ajuste do controle e a realização de diversas simulações alcançando os valores esperados para o projeto, Nogueira (2021) usou essa simulação para obter dados de falha do sistema. As falhas estão descritas na Tabela 3.3 incluindo os limites de detecção da falha em relação ao SP da variável.

Tabela 3.3: Descrição e Limites das Falhas em Nogueira (2021).

| Classe | Descrição | Limites de SP |
|---|---|---|
| Operação Normal | Pequenos distúrbios | 99% - 101% SP |
| Falha 1 | Aumento significativo na vazão mássica de SW1 | 104,5% SP |
| Falha 2 | Aumento significativo na concentração de $H_2S$ na SW1 | 460% SP |
| Falha 3 | Aumento significativo na concentração de $NH_3$ na SW1 | 262% SP |
| Falha 4 | Falha no sensor de pressão do gás amoniacal | -0,92% SP |
| Falha 5 | Falha no sensor de pressão do gás ácido | -5,61% SP |
| Falha 6 | Superaquecimento da Coluna C1 | 175% SP |

A operação normal é o cenário de operação sem falhas, onde pequenos distúrbios foram estimulados. Existe essa necessidade em particular para tornar a simulação mais próxima da realidade na qual pequenos distúrbios durante a operação do processo ocorrem com regularidade. Os mesmos distúrbios presentes na operação normal ocorrem também durante falhas de processo, conforme é esperado em um processo químico real. Esses pequenos distúrbios podem ocorrer nas seguintes variáveis:

1. Vazão Mássica de SW1;
2. Concentração de $H_2S$ de SW1;
3. Concentração de $NH_3$ de SW1;
4. Temperatura de SW1;
5. Vazão Mássica de SW5
6. Vazão Mássica de CW5; e
7. Temperatura de AMG1.

Os limites de detecção da falha em relação ao SP da variável foram definidos por Nogueira (2021) através de diversas simulações. As classes foram determinadas como os cenários nos quais os controladores não são mais capazes de reverter o desvio do SP das variáveis. Dessa forma, a região de falha é caracterizada pela variável sair da região controlável. É afirmado que, na prática, isso ocorre quando uma válvula abre ou fecha completamente ou quando o vaso de refluxo ou o fundo das colunas se encontram totalmente secos. Isso é importante para o estudo das falhas. Toda simulação que apresenta uma determinada falha tem seus dados considerados livres das demais pertubações que ocasionalmente poderiam permanecer devido à outra falha aplicada na simulação.

As falhas 1, 2 e 3 estão associadas ao efeito de inundação de coluna causado por alterações significativas na corrente de alimentação SW1. As falhas 4 e 5 estão relacionadas com falha nos instrumentos de pressão no topo das duas colunas. A falha 6 está associada ao cenário no qual a coluna "ferveria", isso ocorre quando a carga térmica é muito alta e a água é evaporada na coluna. Esse efeito é expresso pelos valores da temperatura de fundo e de topo da coluna se aproximarem, enquanto a maior parte da água é evaporada.

Isso leva a um aumento significativo na temperatura da corrente de topo e, no caso da coluna C1, à perda da eficiência, pois aumenta a concentração de amônia no ACG1.

Nogueira (2021) realizou as simulações com tarefas programáveis para alterar as variáveis automaticamente. A alteração sempre era feita utilizando a função "SRAMP" do *software* de simulação Aspen Plus®. Essa função promove o incremento ou o decréscimo de uma variável com uma rampa de formato "S", isto é, uma rampa senoidal.

Todas as classes foram simuladas com e sem o efeito de incrustação, inclusive a operação normal. Para gerar esse efeito, o coeficiente global de transferência de calor (U) do trocador H3 foi reduzido de 3% a 3,5% no início das simulações que possuem esse efeito. Nogueira (2021) verificou que a magnitude dessa alteração no valor de U de H3 foi suficiente para causar falhas diante de pequenos distúrbios. A simulação de incrustação foi realizada nesse equipamento, pois, quando aplicado nos demais trocadores, não houve impacto significativo no processo.

Como mencionado, a região de falha é descrita pela afastamento de uma variável da região controlável, porém outras variáveis podem sair da região controlável nesse procedimento. Na Tabela 3.4, Nogueira (2021) relaciona as variáveis que são as primeiras a atingir o limite da região controlável para cada classe, com e sem falha.

Tabela 3.4: Descrição e Limites das Falhas. Fonte: Nogueira (2021).

| Classe | Sem Incrustação em H3 | Com Incrustação em H3 |
|---|---|---|
| Operação Normal | - | SW8T |
| Falha 1 | SW8T | SW8T |
| Falha 2 | ACG1P | ACG1P |
| Falha 3 | C2SL | SW8T |
| Falha 4 | C2SL | C2SL |
| Falha 5 | SW5F | SW5F |
| Falha 6 | SW8T | SW8T |

Nota-se que, para as falhas 2, 4, 5 e 6, a variável na região de falha não se alterou devido a presença da incrustação em H3. Já na falha 3, ambas as variáveis C2SL e SW8T atingem o limite da região controlável, porém a primeira variável a atingir demonstrou depender da presença da incrustação (NOGUEIRA, 2021).

Como já mencionado, a operação normal pode apresentar falhas quando na presença de incrustação em H3, nesse caso a variável que se encontra na região de falha é sempre SWT8. Durante o desenvolvimento do banco de dados esse cenário foi classificado como falha 1, devido a semelhança operacional do comportamento dinâmico observado.

### 3.1.3 Banco de Dados

Após as simulações, Nogueira (2021) salvou os dados em planilhas que passaram por algumas etapas até chegar no banco de dados que foi utilizado neste trabalho. O primeiro

passo foi a adição de ruído nos dados. Wojsznis, Mehta e Thiele (2010) sugerem a adição de ruído aleatório de média zero com amplitude máxima de 0,2% a 0,5% da faixa da magnitude dos dados de saída do processo. Nogueira (2021) aplicou como média o SP das variáveis, pois foi considerado que a magnitude das variáveis variava muito nos cenários de falha. A Equação 3.1 exibe o cálculo feito do vetor de ruído como uma distribuição normal de média zero com uma amplitude máxima *Sigma* vezes o SP de cada variável de processo $i$.

$$Noise_i = N(0, Sigma * SP_i) \tag{3.1}$$

Porém, verificou-se visualmente que mesmo para o valor 0,2%, a amplitude do ruído no SP estava muito alto. Esse valor implicava em uma descaracterização significativa dos dados. Por exemplo, como nota-se para o nível do vaso de refluxo (Figura 3.3), valores acima de 0,1% comprometeram a leitura da dinâmica do processo e acarretam em perda de informações relevantes. Buscando manter as características dinâmicas do processo, (NOGUEIRA, 2021) fixou o valor de Sigma em 0,1% nesse banco de dados. Além disso, a razão sinal/ruído (SNR) foi mantida em pelo menos 10 seguindo a recomendação de Poe e Mokhatab (2016).

Após a inclusão do ruído, cada linha de dado da planilha recebeu a sua classe de acordo com o status da operação naquele instante. A operação normal é a classe 0 e as demais falhas seguem a numeração apresentada na Tabela 3.3. Nogueira (2021) utilizou PCA dinâmico para definir o número de atrasos (*delays*) necessários (KU; STORER; GEORGAKIS, 1995). Um atraso, nesse contexto, significa incluir em cada linha de dado, além do valor da variável naquele instante (t), também seu valor no passado, isto é, a variável com um atraso (t-1). Isso cria uma noção de dinâmica nos conjuntos de dados (RAMLI, 2017). Os bancos de dados de atrasos únicos e duplos foram gerados, bem como um estático, sem atraso. Esses três bancos de dados foram inseridos independentemente de um algoritmo PCA para calcular o número de componentes principais. O valor que mais atendeu essa aplicação foi um valor de atraso, ou seja, cada amostra possui o valor da variável no presente momento e o valor da variável com atraso.

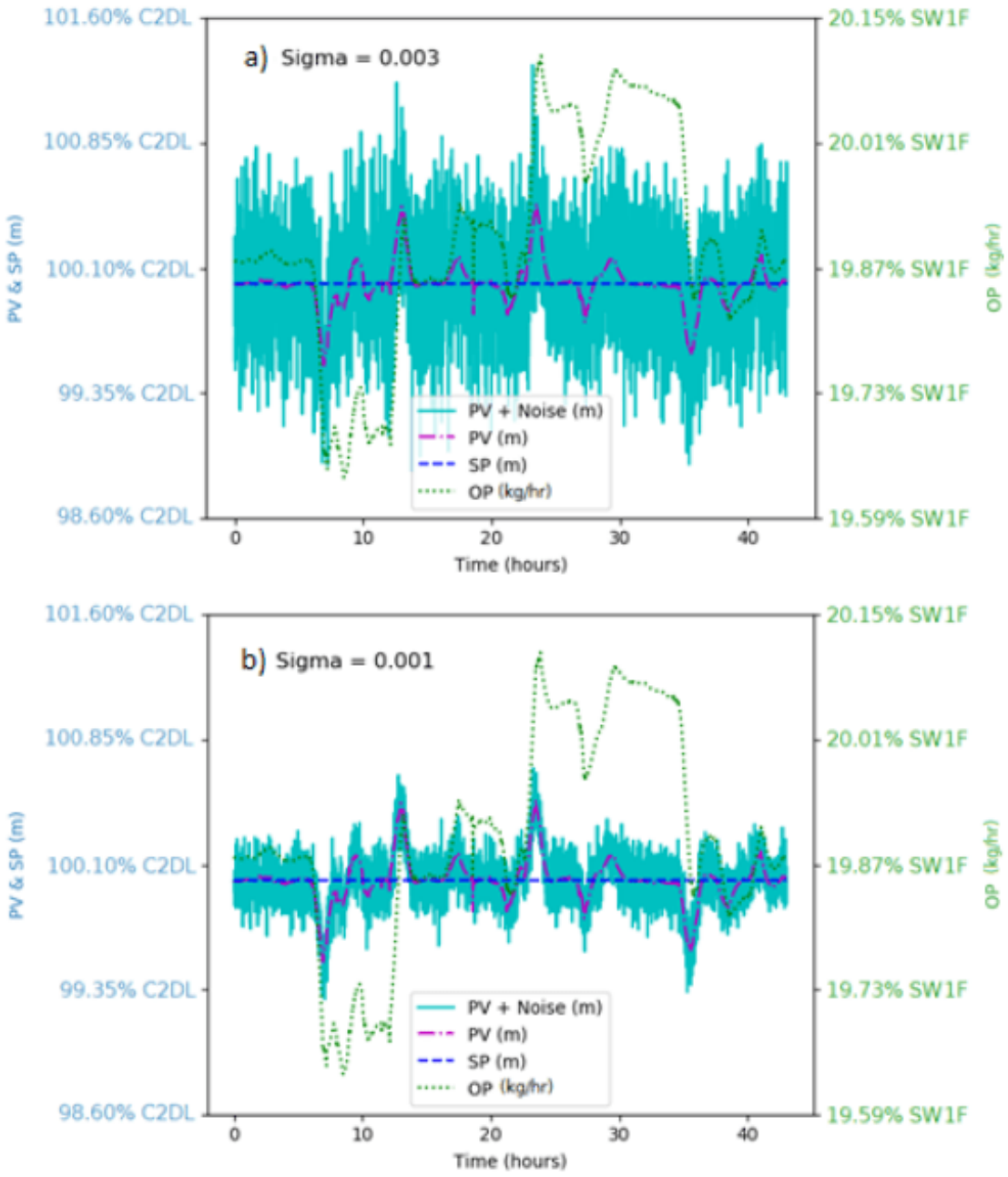


Figura 3.3: Comportamento dinâmico de C2DL com a) Sigma = 0,003 e b) Sigma = 0,001 amplitude de adição de ruído. Fonte: Nogueira (2021).

Após essas etapas, os dados foram separados em dois conjuntos, treino e teste, para aplicação em algoritmos de AI. Em cada rodada, foi simulada no máximo uma falha. Então, os dados de uma mesma simulação estão presentes em apenas um dos conjuntos (treino ou teste), isto é, todos os dados de uma mesma rodada de simulação foram para o mesmo conjunto de dados. Isso foi feito para evitar que houvesse vazamento de informações (*data leakage*), pois isso poderia gerar um viés para os algoritmos utilizados. Além disso, todas as classes estão presentes em ambos conjuntos de dados, com e sem incrustação, garantindo assim que o treinamento seria eficiente para classificar o

conjunto de teste. Nogueira (2021) montou os conjuntos com a maior parte das amostras pertencentes à operação normal, de tal maneira que as operações de falhas fossem distribuídas em menores proporções. Essa é a distribuição esperada de um histórico de processo químico na realidade. O número de amostras para cada classe em cada conjunto de dados é apresentado na Tabela 3.5.

Tabela 3.5: Distribuição das classes das amostras entre os conjuntos treino e teste do banco de dados de Nogueira (2021).

| | **Treino** | | **Teste** | |
|---|---|---|---|---|
| Classe | Número de Amostras | % | Número de Amostras | % |
| Operação Normal | 22245 | 59,29% | 14657 | 65,07% |
| Falha 1 | 3262 | 8,69% | 2742 | 12,17% |
| Falha 2 | 5348 | 14,25% | 2659 | 11,80% |
| Falha 3 | 1392 | 3,71% | 663 | 2,94% |
| Falha 4 | 1875 | 5,00% | 470 | 2,09% |
| Falha 5 | 2209 | 5,89% | 719 | 3,19% |
| Falha 6 | 1188 | 3,17% | 615 | 2,73% |
| **Total** | **37519** | **100%** | **22525** | **100%** |

No total, os dois conjuntos possuem 60044 amostras, das quais 62,49% pertencem ao conjunto de treino e 37,51% ao conjunto de teste.

Dessa maneira, Nogueira (2021) montou o banco de dados com as variáveis controladas apresentadas na Tabela 3.2, exceto a variável C2SL2, que foi retirada por se tratar de uma redundância da variável C2SL. Isso totaliza 27 variáveis listadas nos conjuntos de dados: essas 13 variáveis controladas, os valores dessas respectivas variáveis com atraso e a classe dos dados. Assim, as 27 variáveis que constam no banco das amostras são: ACG1P(t-1), AMG1P(t-1), C1SL(t-1), C2DL(t-1), C2SL(t-1), SW8T(t-1), C1S2T(t-1), AMG1T(t-1), SW1F(t-1), SW5F(t-1), CW5F(t-1), CW8F(t-1), C2HDI(t-1), ACG1P(t), AMG1P(t), C1SL(t), C2DL(t), C2SL(t), SW8T(t), C1S2T(t), AMG1T(t), SW1F(t), SW5F(t), CW5F(t), CW8F(t), C2HDI(t) e *CLASS* (em português, classe).

## 3.2 Cenários de Incrustação

A incrustação é um fenômeno comum em um trocador de calor. Ela ocorre com a deposição de material orgânico ou inorgânico nas paredes do trocador. No caso de um trocador casco-tubo, um dos modelos mais populares, a incrustação pode ocorrer na parte interna ou externa dos tubos. Geralmente, o fluido com maiores características de incrustação é escolhido para passar pelos tubos pois é mais fácil a realização da manutenção (MITROVIC, 2012).

De maneira geral, a incrustação é um problema industrial recorrente que afeta o processo e pode ser origem de falhas. O procedimento padrão é a avaliação das variáveis en-

volvidas nessa etapa para verificar possíveis perdas de eficiência na troca térmica. Logo, é de interesse da indústria que um algoritmo possua a capacidade de detectar essa condição, especialmente em situações de falhas.

O banco de dados Nogueira (2021), conforme já mencionado, classifica o comportamento de operação em 7 classes, a operação normal, livre de falhas, e as 6 falhas simuladas. A incrustação foi aplicada ao trocador H3 para algumas das simulações realizadas, porém esse banco de dados não discrimina quais amostras apresentam o cenário de incrustação. Por isso, foi necessário a modificação dos dados para o conjunto de treino e o conjunto de teste.

Para a análise do cenário de incrustação, o banco de dados deve ser reclassificado. As amostras classificadas como operação normal (classe 0) mantém a sua classificação. As falhas 1, 2, 3, 4, 5 e 6 são separadas de acordo com a presença ou não de incrustação durante a simulação. A falha 1 no banco de dados de Nogueira (2021) agrupa as falhas 1 com e sem incrustação, mas também as operações normais com distúrbios leves com incrustação, como já mencionado. Dessa maneira, devem ser adicionadas novas 7 falhas (Tabela 3.6).

Tabela 3.6: Descrição das classes do conjunto de treino e teste de (NOGUEIRA, 2021) modificadas para o cenário de incrustação.

| Classe | Falha Simulada | Incrustação |
|---|---|---|
| Operação Normal | Distúrbios leves | Não |
| Falha 1 | Falha 1 | Não |
| Falha 2 | Falha 2 | Não |
| Falha 3 | Falha 3 | Não |
| Falha 4 | Falha 4 | Não |
| Falha 5 | Falha 5 | Não |
| Falha 6 | Falha 6 | Não |
| Falha 7 | Distúrbios leves | Sim |
| Falha 8 | Falha 1 | Sim |
| Falha 9 | Falha 2 | Sim |
| Falha 10 | Falha 3 | Sim |
| Falha 11 | Falha 4 | Sim |
| Falha 12 | Falha 5 | Sim |
| Falha 13 | Falha 6 | Sim |

## 3.3 Cenários de Pré-Falha

O exato instante que uma operação livre de falhas se torna uma operação com falha não é sempre evidente. Entretanto, Nogueira (2021) concluiu que essa região está presente no conjunto de dados, pois os resultados das métricas utilizadas indicam que os algoritmos tiveram dificuldade de reconhecer parte das amostras como operação com falha, sendo

todas estas classificadas como operação normal. Essa região de difícil determinação entre a operação normal e as falhas sugere a existência de uma região de pré-falha.

Uma forma para lidar com esse problema é classificar essas amostras em questão com outra classe. Duas abordagens foram propostas para esse cenário. A primeira abordagem é classificar todas as amostras como uma única classe definida como região de pré-falha. Apesar de sua limitação prática numa planta industrial, esta maneira poderia indicar que o sistema fugiu da operação normal e que, caso nada for feito, este entrará em uma região de falha. Com isso, seria adicionada apenas uma nova classe aos conjuntos de dados, totalizando agora 8 classes (numeradas de 0 a 7). Na Tabela 3.7, está a primeira proposta de nova classificação para lidar com esse problema.

Tabela 3.7: Descrição das classes do conjunto de treino e teste de Nogueira (2021) modificadas para o cenário de pré-falha (1ª Abordagem).

| Classe | Falha Simulada |
|---|---|
| Operação Normal | Distúrbios leves |
| Falha 1 | Falha 1 |
| Falha 2 | Falha 2 |
| Falha 3 | Falha 3 |
| Falha 4 | Falha 4 |
| Falha 5 | Falha 5 |
| Falha 6 | Falha 6 |
| Pré-Falha | Válida para todas as simulações |

A segunda abordagem trata-se em classificar cada amostra suspeita de região de pré-falha pela região de pré-falha do evento em questão. Essa alternativa é mais rica em termos de informação, pois, além de indicar que o processo está migrando da região controlável, ela é capaz de identificar qual falha está ocorrendo. Dessa maneira, o banco de dados passaria a ter treze classes, as sete classes originais em adição às 6 regiões de pré-falha. A Tabela 3.8 descreve essa segunda abordagem para lidar com esse cenário.

Para que essa nova classificação seja possível, a alternativa encontrada é a aplicação dos métodos de AI para que essas amostras sejam identificadas. O primeiro passo desse procedimento é a aplicação do método escolhido nos conjuntos de dados originais. Primeiro, o modelo é alimentado com o conjunto de treino, seguido pelo conjunto de teste. Espera-se que o observado seja o mesmo resultado encontrado por Nogueira (2021).

Então, o conjunto de treino é alimentado como entrada novamente. Os dados são reclassificados de modo que as amostras identificadas erroneamente como operação normal, mas que na verdade pertenceriam às classes de falhas, sejam marcadas como uma única classe de pré-falha ou às classes de pré-falha específicas, de acordo com a abordagem escolhida. Assim é construído um novo conjunto de treino, supostamente, capaz de identificar a classe de pré-falha. Faz se um novo treinamento para gerar um novo modelo, agora capaz de classificar a(s) classe(s) de pré-falha (Figura 3.4).

Tabela 3.8: Descrição das classes do conjunto de treino e teste de Nogueira (2021) modificadas para o cenário de pré-falha (2ª Abordagem).

| Classe | Falha Simulada |
|---|---|
| Operação Normal | Distúrbios leves |
| Falha 1 | Falha 1 |
| Falha 2 | Falha 2 |
| Falha 3 | Falha 3 |
| Falha 4 | Falha 4 |
| Falha 5 | Falha 5 |
| Falha 6 | Falha 6 |
| Pré-Falha 1 | Falha 1 |
| Pré-Falha 2 | Falha 2 |
| Pré-Falha 3 | Falha 3 |
| Pré-Falha 4 | Falha 4 |
| Pré-Falha 5 | Falha 5 |
| Pré-Falha 6 | Falha 6 |

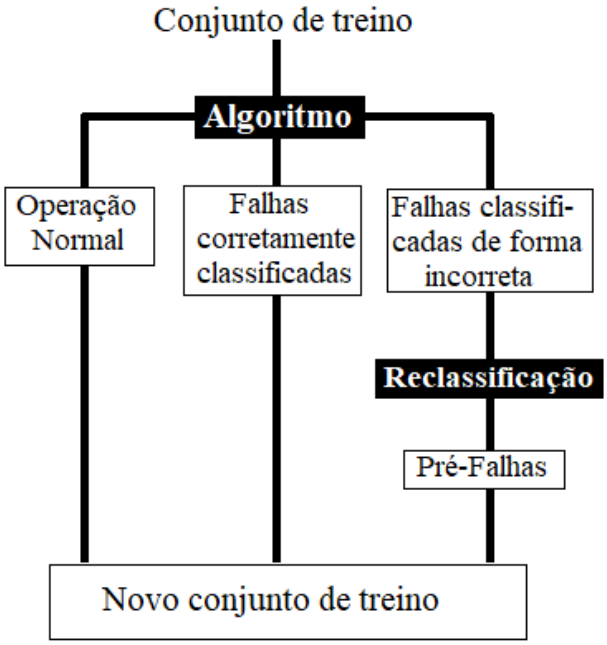


Figura 3.4: Esquema para obtenção do conjunto de dados de treino para o cenário de pré-falha.

## 3.4 Inteligência Artificial

Os códigos foram feitos em um computador com processador Intel(R) Core(™) i7-7500U CPU@ 2.70 GHz, memória instalada (RAM) de 8,00 GB e sistema operacional Windows 10 64 bits. Todo o processamento de dados e algoritmos de Inteligência Artificial

foram realizados em Python 3.5.4 no ambiente de execução de código aberto Jupyter Notebook. De forma geral, as bibliotecas *matplotlib* (2.0.2), *numpy* (1.13.1), *pandas* (0.20.3), *scikit-learn* (0.19.0), *scipy* (0.19.1) *seaborn* (0.9.0) e *tensorflow* (1.3.0) foram amplamente utilizadas.

### 3.4.1 Algoritmos de Inteligência Artificial

Nogueira (2021) aplicou no banco de dados múltiplas metodologias de ML, a saber: RF, o algoritmo de k-Vizinhos mais próximos (*k-Nearest Neighbours*, KNN), SVM (Linear, Polinomial, Gaussiano e Sigmoid), percetron multicamada (*Multi-Layer Perceptron*, MLP), rede neuronal convolucional (*Convolutional Neural Network*, CNN) e mapa auto-organizável (*Self-Organizing Maps*, SOM). Dentre essas, RF e SVM (Linear e Gaussiano) foram os que apresentaram as melhores acurácias, acima de 93% (Figura 3.5). Após a realização da redução de variáveis (processo pelo qual as variáveis que possuem menor relevância para solução do problema são removidas da entrada de dados), RF foi o melhor método analisado para o banco de dados proposto, alcançando 95,2% de acurácia. Logo, esses modelos foram os escolhidos para o desenvolvimento da metodologia de AI no presente trabalho.

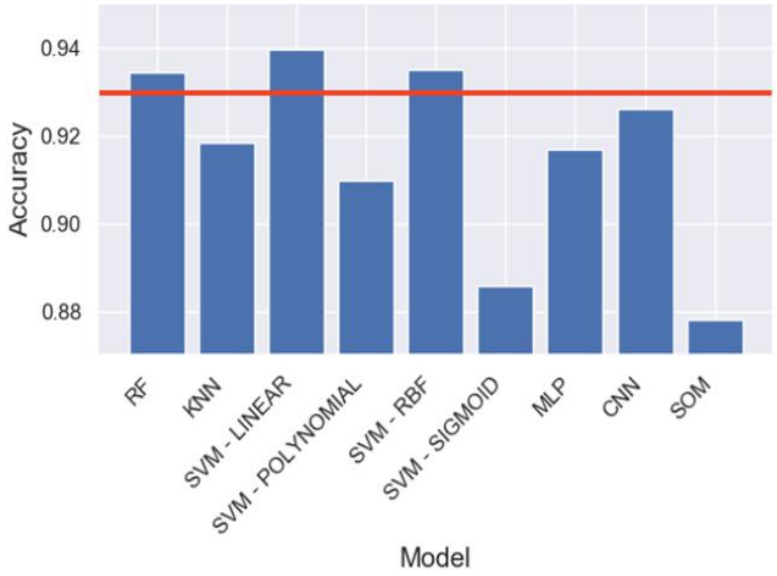


Figura 3.5: Comparação dos métodos de AI aplicados para o banco de dados de teste original. Fonte: Nogueira (2021).

Conforme já mencionado no Capítulo 2, os algoritmos de ML possuem poucas alterações em sua metodologia de utilização. Eles devem ser configurados segundo os melhores hiperparâmetros para o conjunto de dados utilizados. Os hiperparâmetros são parâmetros ajustáveis que permitem controlar o processo de treinamento do modelo. Dessa forma, o desempenho dos modelos depende dos hiperparâmetros. Em seguida é feito o treinamento do modelo, o conjunto de dados de treino é alimentado como entrada

nesse algoritmo. Cada algoritmo possui uma técnica para reduzir o erro da sua previsão. Neste caso, não há conjunto de validação, logo o algoritmo é testado através do conjunto de teste a partir do qual pode ser avaliado usando as métricas apropriadas.

A grande diferença entre os métodos sugeridos é o artifício da Importância de Variável (*Variable Importance*, VI) presente na biblioteca *scikit-learn* para o algoritmo RF. A VI seleciona as variáveis mais significativas para realizar aquela classificação. A partir dessa informação, as variáveis de menor importância podem ser descartadas. Esse procedimento chamado de redução de variáveis simplifica o modelo, tende a melhorar a classificação e reduz o tempo de processamento.

### 3.4.2 Hiperparâmetros

Nogueira (2021) realizou a otimização dos hiperparâmetros através da biblioteca *Optuna*. *Optuna* é uma estrutura de otimização de hiperparâmetros de código aberto para automatizar a pesquisa de hiperparâmetros. Os hiperparâmetros escolhidos para a otimização do algoritmo RF foram o número de estimadores (*rf_n_estimators*) que representa o número de árvores de decisão na RF, a profundidade das árvores (*rf_max_depth*) e o número mínimo de amostras para separar em um nó (*rf_min_samples_leaf*).

Foi verificado que, para valores mais altos de *rf_n_estimators* e *rf_max_depth*, a acurácia aumentou. Já *rf_min_samples_leaf* apresentou melhor acurácia próximo ao limite inferior do hiperparâmetro. O resultado foi que a escolha ótima para os hiperparâmetros é $rf\_n\_estimators = 97$, $rf\_max\_depth = 30$ e $rf\_min\_samples\_leaf = 3$ (NOGUEIRA, 2021).

Nogueira (2021) utilizou o mesmo procedimento para o algoritmo SVM. O hiperparâmetro otimizado para SVM Linear foi apenas *c_lin* e para SVM RBF (Gaussiano) foram *c_rbf* e *gamma_rbf*. O hiperparâmetro $C$ é um parâmetro de regularização. Ele informa o quanto deve-se evitar a classificação incorreta de cada exemplo de treinamento. Para altos valores de C, o algoritmo escolherá um hiperplano de menor margem se esse hiperplano obtiver todos os pontos de treinamento classificados corretamente. Já com valor baixo de C, o algoritmo escolherá um hiperplano de separação de maior margem, mesmo que esse hiperplano classifique incorretamente mais pontos. Já *Gamma* define o impacto de cada amostra nas fronteiras de classificação.

Foi verificado que para os SVM estudados, os melhores valores para *c_lin* e *c_rbf* foram entre 0.1000 e 0.115. Já para *gamma_rbf* a opção ótima é *auto*. Nesta opção, o gamma é calculado pelo inverso do número de variáveis, desconsiderando a variância da matriz de dados.

### 3.4.3 Métricas

Além da VI para RF, a biblioteca *scikit-learn* possui alguns indicadores bastante difundidos para analisar os resultados do problema de classificação utilizando ML.

O primeiro conceito a ser apresentado é o da matriz de confusão. Uma matriz de confusão é um modelo específico de tabela que permite a visualização da performance de um algoritmo. Geralmente, cada linha representa uma classe real e cada coluna representa uma classe predita pelo modelo. Dessa maneira, a classe real de cada amostra pode ser comparada com a classe prevista pelo algoritmo.

Em uma classificação binária, a matriz de confusão permite que uma amostra classificada seja agrupada em uma das quatro definições listadas abaixo e ilustradas na Figura 3.6.

1. Verdadeiro Positivo (VP): classe real positiva, corretamente classificada como positiva.
2. Verdadeiro Negativo (VN): classe real negativa, corretamente classificada como negativa.
3. Falso Positivo (FP): classe real negativa, erroneamente classificada como positiva. Equivalente a um erro tipo I pelo Teste de Hipóteses em Estatística, também considerado alarme falso.
4. Falso Negativo (FN): classe real positiva, erroneamente classificada como negativa. Equivalente a um erro tipo II pelo Teste de Hipóteses em Estatística.

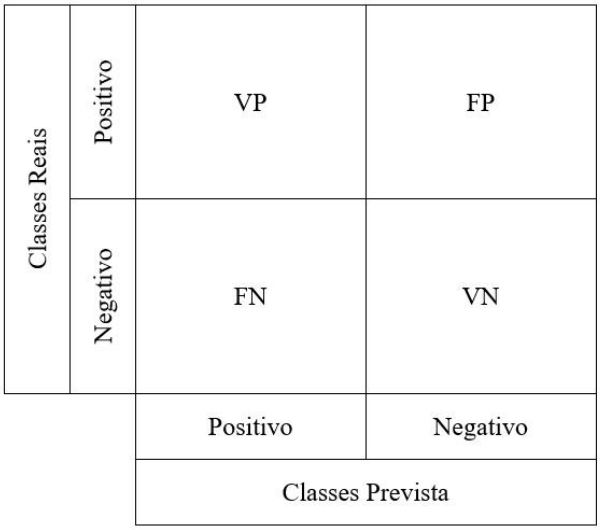


Figura 3.6: Matriz de Confusão para um Problema de Classificação Binária.

Contudo, isso não impede que esse conceito seja utilizado para uma quantidade maior de classes. É importante destacar aqui que existe um juízo de valor para cada uma dessas situações. Para diferentes casos, os resultados FP e FN possuem impactos diferentes. Por exemplo, em testes médicos, um erro do tipo I (FP) faria parecer que um tratamento para uma doença tem o efeito de reduzir a gravidade da doença, quando, na verdade, não tem. Neste caso, um erro tipo II (FN) seria de que o tratamento parece não ter efeito, quando na verdade tem. Este último é considerado menos grave que a situação em que um paciente recebe um tratamento o qual não possui efeito algum. O oposto ocorre para

a aplicação em sistemas de FDD, um alarme falso (FP) pode não ter um forte impacto em uma unidade, porém um falso negativo (FN), isto é, uma falha classificada como operação normal pode ter efeitos prejudiciais na planta industrial.

A seguir são apresentadas as principais métricas aplicando essas classificações. Uma das medidas mais utilizadas para esse tipo de problema é a acurácia. Acurácia é a razão entre a quantidade de amostras corretamente classificadas e a quantidade total de amostras, em outras palavras é o percentual de classificações corretas feitas por um dado modelo (Equação 3.2).

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \tag{3.2}$$

A precisão é a proporção de amostras preditas como positivas que são positivas na realidade. Segue a Equação 3.3:

$$Precisão = \frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{FP}} \tag{3.3}$$

A sensibilidade é a proporção de amostras realmente positivas que foram preditas como positivas (Equação 3.4).

$$Sensibilidade = \frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{FN}} \tag{3.4}$$

Existe ainda um indicador que opera com os valores de precisão e sensibilidade. Na análise estatística da classificação binária, a medida F (*F-score*) é uma medida da acurácia de um teste. A medida F-beta aplica pesos adicionais para cada indicador, valorizando mais precisão ou mais sensibilidade. A medida F-1 é a média harmônica entre precisão e sensibilidade (Equação 3.5).

$$Medida\ F\text{-}1 = \frac{2}{\frac{1}{Precisão} + \frac{1}{Sensibilidade}} \tag{3.5}$$

Bancos de dados equilibrados apresentam quantidades similares de amostras com diferentes rótulos. Já em banco de dados desequilibrados, a maioria das amostras possuem o mesmo rótulo. Nesse contexto, para bancos de dados equilibrados, Chollet (2021) sugere o uso de Característica de Operação do Receptor e Área Sobre a Curva. Enquanto a pre-

cisão e a sensibilidade são indicadas como melhores métricas para a análise de problemas com bancos de dados desequilibrados. O problema proposto nesse trabalho para ambos os cenários apresenta bancos de dados desequilibrados, por isso foram escolhidas as últimas métricas mencionadas.

# Capítulo 4

# Resultados e Discussão

## 4.1 Avaliação dos Algoritmos Utilizando os Bancos de Dados Originais

Primeiramente, os algoritmos foram testados para o banco de dados original. Isso se fez tanto para ratificar a escolha de hiperparâmetros quanto para utilizar seus resultados na comparação da análise. Todas as métricas apresentadas nesse capítulo foram aplicadas no resultado para o conjunto de teste.

Utilizando RF, a acurácia encontrada foi de 94,35%. O relatório de classificação é encontrado na Tabela 4.1. A matriz de confusão é exibida na Figura 4.1. A VI foi obtida e plotada na Figura 4.2.

Tabela 4.1: Relatório de Classificação gerado para RF no Banco de Dados de Nogueira (2021).

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,93 | 0,98 | 0,96 |
| Falha 1 | 0,93 | 0,84 | 0,88 |
| Falha 2 | 0,99 | 0,82 | 0,90 |
| Falha 3 | 0,97 | 0,90 | 0,94 |
| Falha 4 | 0,98 | 1,00 | 0,99 |
| Falha 5 | 1,00 | 0,99 | 0,99 |
| Falha 6 | 1,00 | 0,90 | 0,95 |

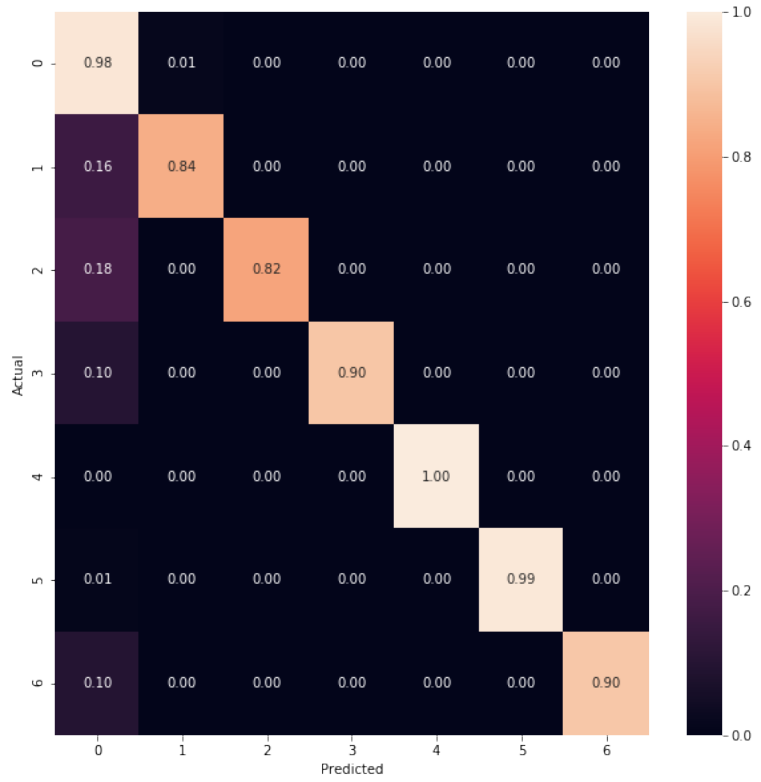


Figura 4.1: Matriz de Confusão obtida por RF para o Banco de Dados de Nogueira (2021).

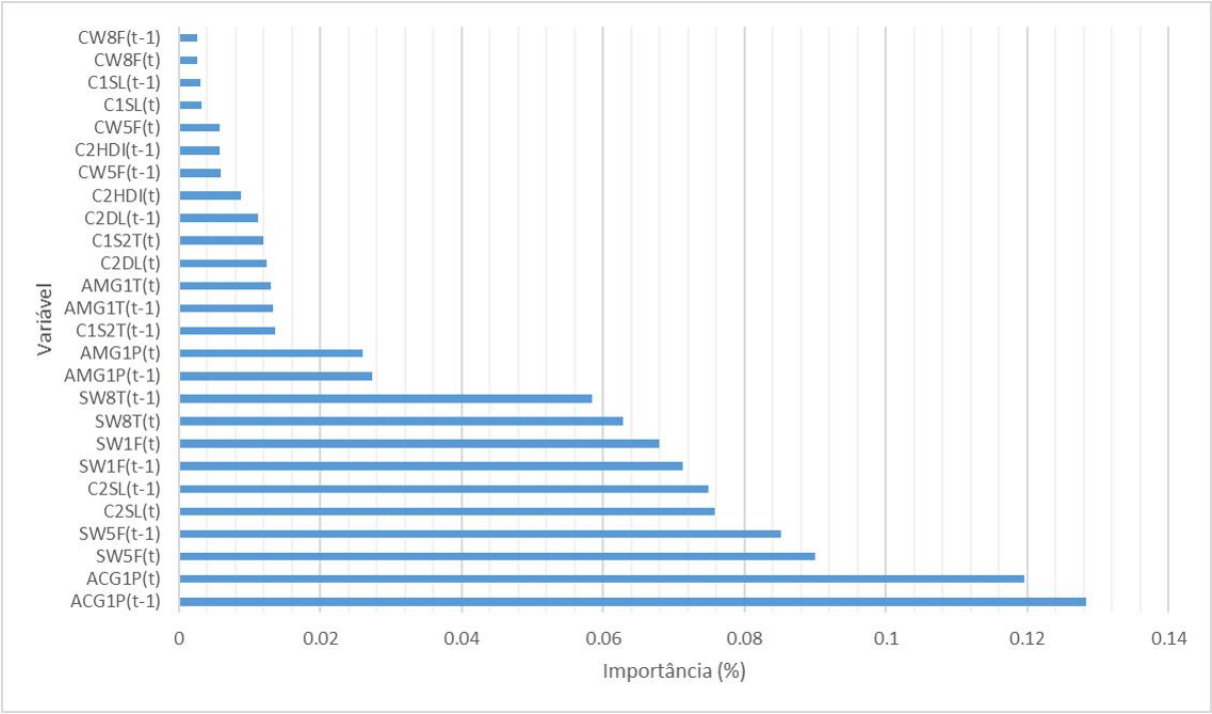


Figura 4.2: Importância de Variável obtida por RF para o Banco de Dados de Nogueira (2021).

Foi realizada em seguida a redução de variáveis. (NOGUEIRA, 2021) estabeleceu

que as variáveis que tiveram importância menor que 2% foram removidas e se repetiu o mesmo processo.

Utilizando RF com redução de variáveis, a acurácia encontrada foi de 95,21%. O relatório de classificação e a matriz de confusão são encontrado na Tabela 4.2 e na Figura 4.3, respectivamente. Além disso, foi feita a importância das variáveis após a redução (Figura 4.4).

Tabela 4.2: Relatório de Classificação gerado para RF com Redução de Variáveis no Banco de Dados de Nogueira (2021).

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,94 | 0,99 | 0,96 |
| Falha 1 | 0,96 | 0,85 | 0,90 |
| Falha 2 | 0,98 | 0,87 | 0,92 |
| Falha 3 | 0,97 | 0,91 | 0,94 |
| Falha 4 | 0,99 | 1,00 | 0,99 |
| Falha 5 | 1,00 | 0,99 | 0,99 |
| Falha 6 | 1,00 | 0,89 | 0,94 |

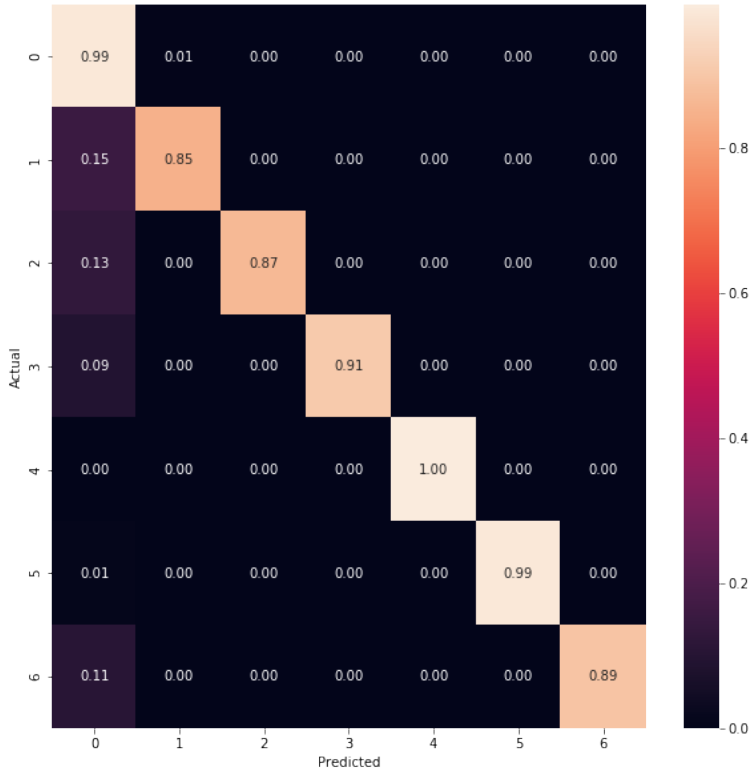


Figura 4.3: Matriz de Confusão obtida por RF com Redução de Variáveis para o Banco de Dados de Nogueira (2021).

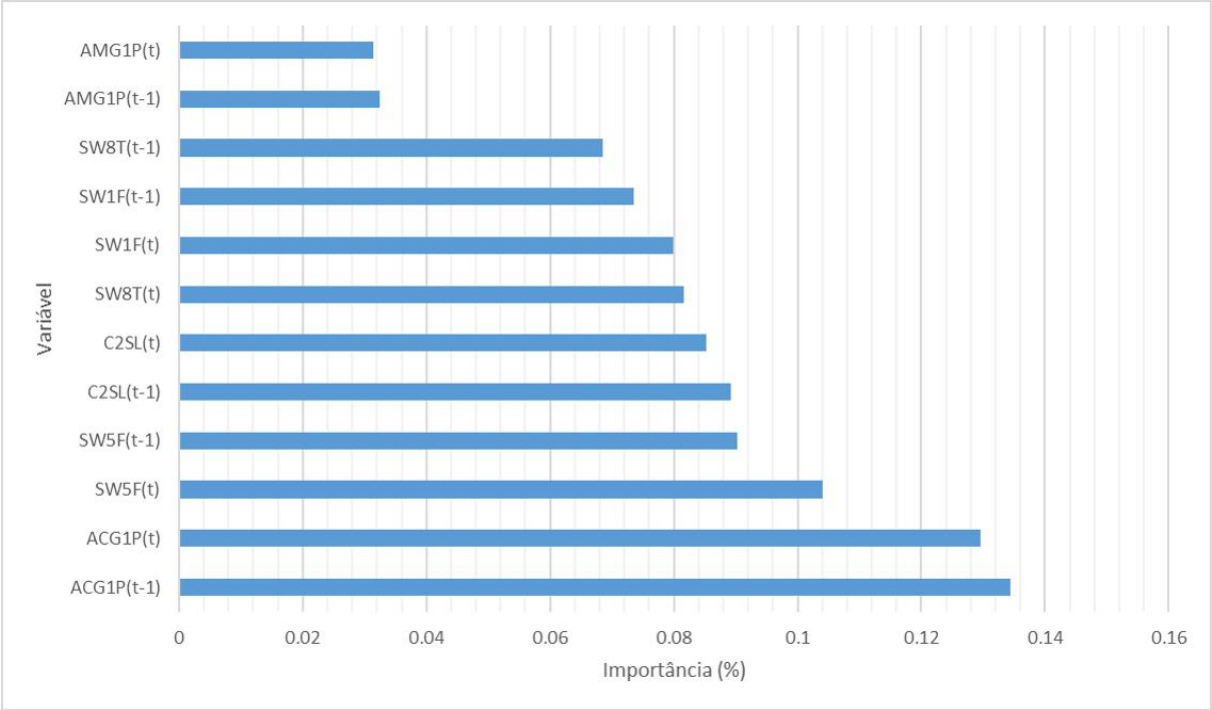


Figura 4.4: Importânica de Variável obtida por RF com Redução de Variáveis para o Banco de Dados de Nogueira (2021).

Para SVM Linear, a acurácia obtida foi 93,97%. O relatório de classificação e a matriz

de confusão para esse modelo são mostrados na Tabela 4.3 e na Figura 4.5, respectivamente.

Tabela 4.3: Relatório de Classificação gerado para SVM Linear no Banco de Dados de Nogueira (2021).

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,92 | 0,99 | 0,96 |
| Falha 1 | 0,97 | 0,78 | 0,87 |
| Falha 2 | 1,00 | 0,81 | 0,90 |
| Falha 3 | 0,96 | 0,89 | 0,92 |
| Falha 4 | 0,97 | 1,00 | 0,98 |
| Falha 5 | 0,93 | 0,97 | 0,95 |
| Falha 6 | 1,00 | 0,91 | 0,95 |

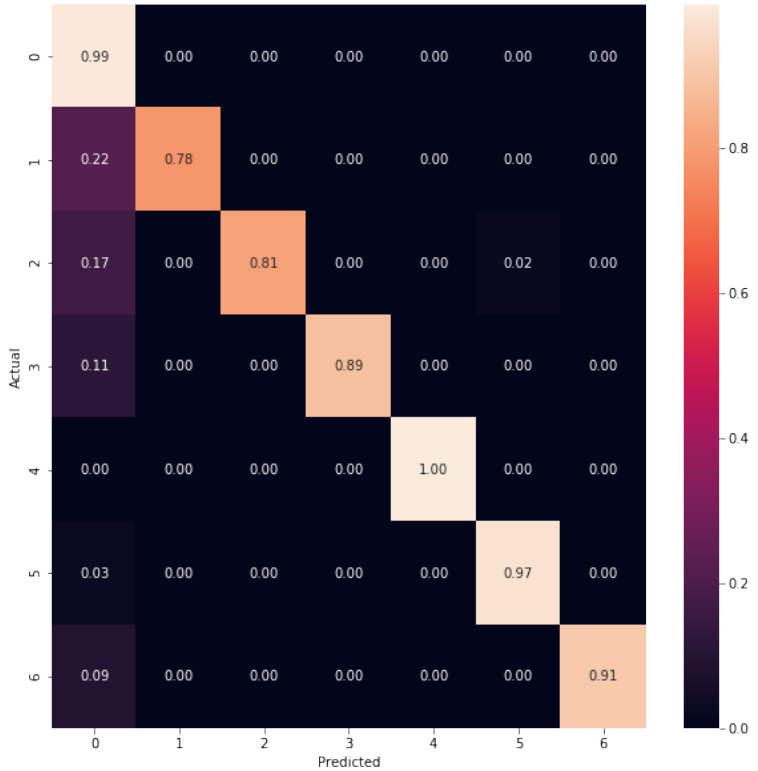


Figura 4.5: Matriz de Confusão obtida por SVM Linear para o Banco de Dados de Nogueira (2021).

Para SVM Gaussiano, a acurácia obtida foi 93,50%. O relatório de classificação e a mariz de confusão para esse modelo são mostrados na Tabela 4.4 e na Figura 4.6, respectivamente.

Tabela 4.4: Relatório de Classificação gerado para SVM Gaussiano no Banco de Dados de Nogueira (2021).

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,91 | 1,00 | 0,95 |
| Falha 1 | 0,99 | 0,76 | 0,86 |
| Falha 2 | 1,00 | 0,80 | 0,89 |
| Falha 3 | 0,94 | 0,87 | 0,91 |
| Falha 4 | 1,00 | 0,93 | 0,97 |
| Falha 5 | 1,00 | 0,94 | 0,97 |
| Falha 6 | 1,00 | 0,90 | 0,95 |

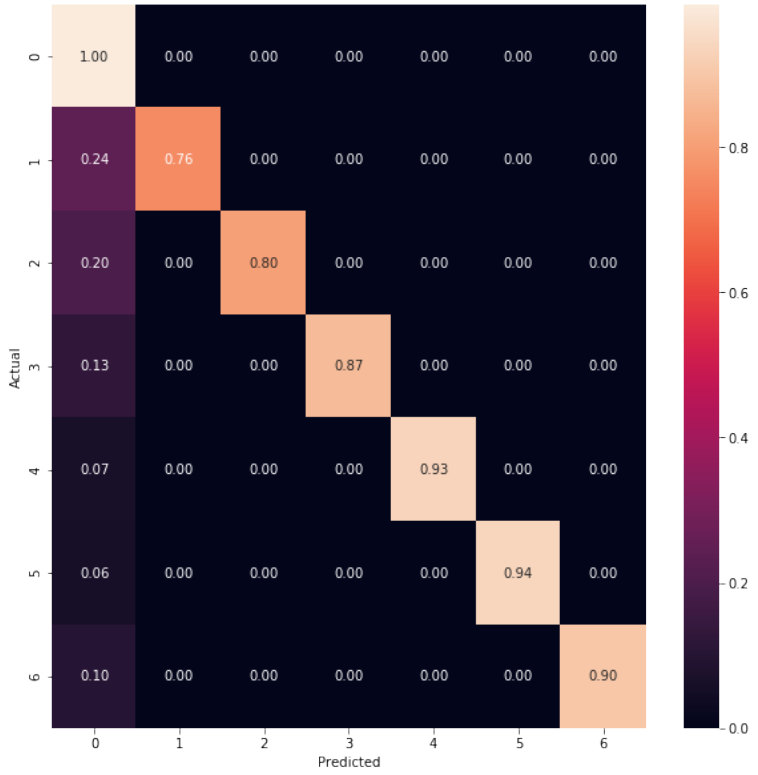


Figura 4.6: Matriz de Confusão obtida por SVM Gaussiano para o Banco de Dados de Nogueira (2021).

Os resultados para todos os métodos foram praticamente idênticos aos de Nogueira (2021), ratificando que os algoritmos desenvolvidos no presente trabalho estão funcionando corretamente e que os hiperparâmetros estão corretamente otimizados para essa aplicação.

## 4.2 Cenários de Incrustação

### 4.2.1 Cenário de Incrustação com 13 Falhas

#### 4.2.1.1 Banco de Dados para o Cenário de Incrustação

A partir de informações das simulações realizadas, foi possível realizar a modificação no banco de dados de Nogueira (2021). A Tabela 4.5 lista o número de amostras que passou a ter cada falha.

Tabela 4.5: Distribuição das 13 classes das amostras entre os conjuntos treino e teste do banco de dados de Nogueira (2021) modificado para o cenário de incrustação.

| | **Treino** | | **Teste** | |
|---|---|---|---|---|
| Classe | Número de Amostras | % | Número de Amostras | % |
| Operação Normal | 22244 | 59,29% | 14657 | 65,07% |
| Falha 1 | 562 | 1,50% | 541 | 2,40% |
| Falha 2 | 2747 | 7,32% | 1420 | 6,30% |
| Falha 3 | 504 | 1,34% | 259 | 1,15% |
| Falha 4 | 1124 | 3,00% | 256 | 1,14% |
| Falha 5 | 594 | 1,58% | 378 | 1,68% |
| Falha 6 | 530 | 1,41% | 197 | 0,87% |
| Falha 7 | 1781 | 4,75% | 589 | 2,61% |
| Falha 8 | 919 | 2,45% | 1612 | 7,16% |
| Falha 9 | 2602 | 6,94% | 1239 | 5,50% |
| Falha 10 | 888 | 2,37% | 404 | 1,79% |
| Falha 11 | 751 | 2,00% | 214 | 0,95% |
| Falha 12 | 1615 | 4,30% | 341 | 1,51% |
| Falha 13 | 658 | 1,75% | 418 | 1,86% |
| **Total** | **37519** | **100%** | **22525** | **100%** |

Comparando os dois conjuntos de dados originais com os modificados para a aplicação no cenário de incrustação em H3, nota-se que a falha que sofre a maior redução na quantidade de amostras é a falha 1. Isso é o esperado, pois, ao contrário das demais falhas, a falha 1 no banco de dados originais correspondia às atuais falha 1, falha 7 e falha 8.

A redução significativa de certos cenários no treinamento é uma informação relevante para o estudo, pois quanto menor a quantidade de amostras mais o algoritmo poderá sofrer um subajuste, isto é, o algoritmo pode não ser capaz de "aprender" essa classe com uma quantidade insuficiente de dados.

#### 4.2.1.2 *Random Forest*

Primeiramente, foi aplicado o algoritmo de RF nesses conjuntos de dados. Após ser realizado o treinamento e o teste, obteve-se a acurácia de, em média, 79,42% para esses dados. Esse resultado foi consideravelmente inferior ao obtido em Nogueira (2021), possivelmente indicando que o modelo não foi capaz de reconhecer fielmente as classes com incrustação. Foi gerado então, o relatório de classificação exposto na Tabela 4.6

Tabela 4.6: Relatório de Classificação gerado para RF no Cenário de Incrustação de 13 Falhas.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,93 | 0,99 | 0,96 |
| Falha 1 | 0,99 | 0,99 | 0,99 |
| Falha 2 | 0,52 | 0,39 | 0,44 |
| Falha 3 | 0,91 | 0,85 | 0,88 |
| Falha 4 | 0,66 | 0,22 | 0,33 |
| Falha 5 | 0,89 | 0,36 | 0,51 |
| Falha 6 | 0,82 | 0,25 | 0,39 |
| Falha 7 | 0,00 | 0,00 | 0,00 |
| Falha 8 | 0,11 | 0,03 | 0,05 |
| Falha 9 | 0,50 | 0,46 | 0,48 |
| Falha 10 | 0,89 | 0,82 | 0,86 |
| Falha 11 | 0,47 | 0,86 | 0,61 |
| Falha 12 | 0,57 | 0,94 | 0,71 |
| Falha 13 | 0,76 | 0,89 | 0,82 |

É possível, através desse relatório, identificar falhas que foram corretamente e as que foram mais dificilmente identificadas. Nota-se que poucas classes foram satisfatoriamente classificadas, apenas a operação normal e as falhas 1, 3 e 10 tiveram precisão e sensibilidade acima de 0.80. Em particular, destaca-se as falhas 7 e 8 que tiveram os piores resultados entre as falhas observadas. Porém, não é possivel analisar apenas através dessa tabela entre quais falhas o algoritmo confundiu uma classe. Para analisar esta condição, será utilizada a matriz de confusão apresentada na Figura 4.7.

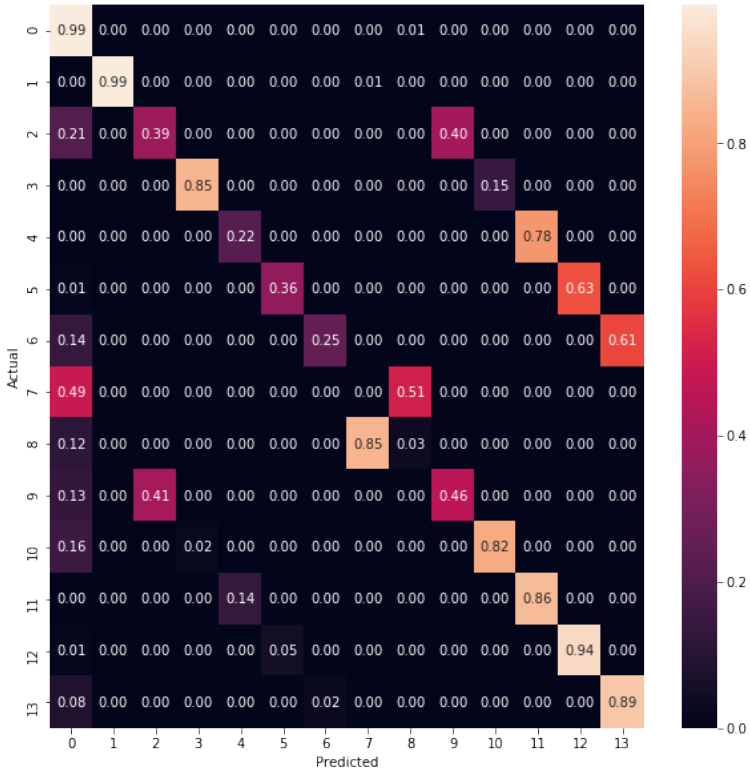


Figura 4.7: Matriz de Confusão obtida por RF para o Cenário de Incrustação de 13 Falhas.

A partir da matriz é possível afirmar que o algoritmo RF conseguiu classificar bem bem algumas falhas. As falhas 10 a 13 foram classificadas corretamente em mais de 80% dos casos. Porém de maneira geral, especialmente das falhas 4, 5 e 6, o algoritmo teve uma baixíssima acurácia. O algoritmo errou a falha 7, caracterizada por distúrbios leves com incrustação, em todas as amostras. Praticamente metade dessas falhas foi classificada como operação normal e a outra parte como falha 8 (falha 1 com incrustação). A princípio isso indica que Nogueira (2021) estava correta em afirmar que o comportamento dinâmico da falha 7 e da falha 8 eram muito similares, dado que o modelo não foi capaz de diferenciar.

Foi então analisada a VI para esse cenário e os seus valores foram plotados na Figura 4.8.

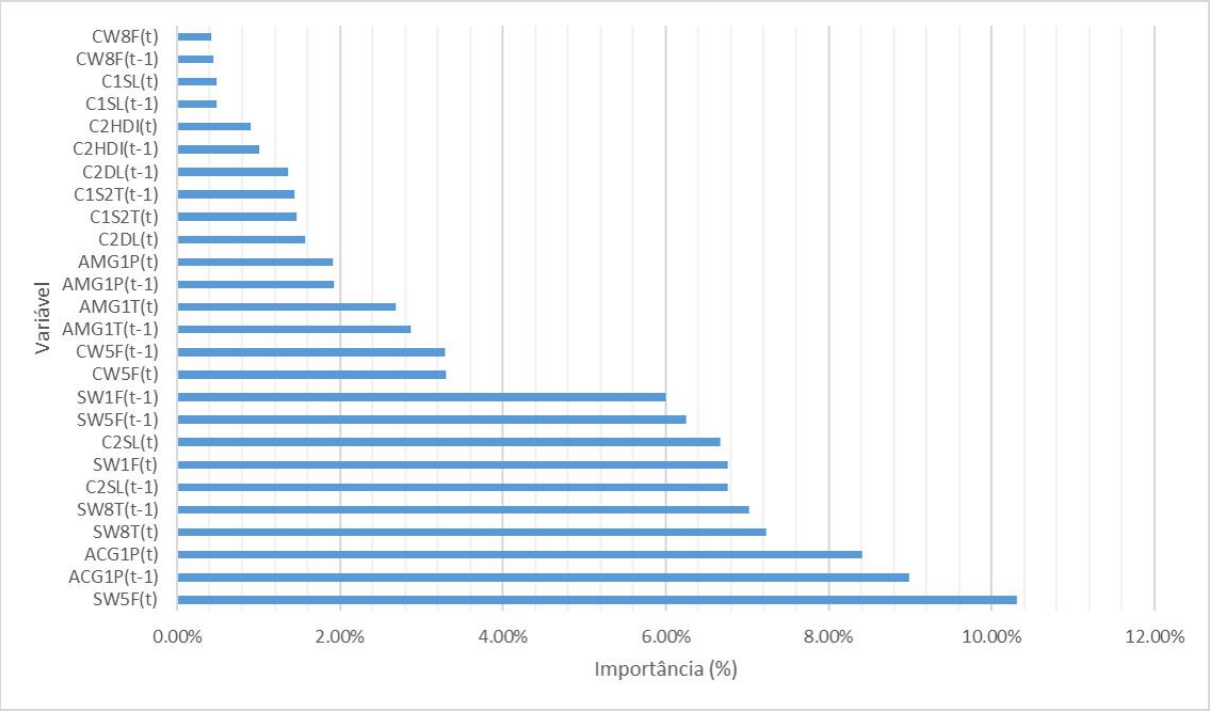


Figura 4.8: Importância de Variável obtida por RF para o Cenário de Incrustação de 13 Falhas.

A VI indica que as variáveis mais importantes para o banco de dados original continuaram sendo as mais importantes para o cenário de incrustação, SW5F(t), ACGP(t-1) e ACGP(t). Porém nota-se que as variáveis SW8T(t) e SW8T(t-1) que antes estavam em nona e décima colocação, agora se mantem na quarta e quinta colocação. Esse aumento na importância relativas às demais variáveis faz sentido, pois elas estão associadas à temperatura da corrente do trocador H3, onde foi aplicado o efeito de incrustação.

A fim de aumentar a acurácia desse algoritmo foi realizada a redução de variáveis. Após alguns testes, foi avaliado que as variáveis com importância menor que 4,00% deveriam ser removidas para buscar uma melhor acurácia do modelo. Após a remoção das variáveis, foi repetido o treinamento do algoritmo. No entanto, não houve mudanças expressivas no valor da acurácia, isto é, seu valor permaneceu próximo a 79%. O relatório de classificação está na Tabela 4.7.

Apenas através do relatório de classificação, é possível que não tenha ocorrido melhora na classificação, pois algumas falhas tiveram a sua precisão reduzida e em particular a precisão, a sensibilidade e a medida F1 da falha 4 foram zeradas. Isso é um indicativo que algumas das variáveis removidas eram essenciais para a identificação desta falha pelo algoritmo. Para entender melhor foi obtida a matriz de confusão presente na Figura 4.9.

Tabela 4.7: Relatório de Classificação gerado para RF com Redução de Variáveis no Cenário de Incrustação de 13 Falhas.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,94 | 0,99 | 0,96 |
| Falha 1 | 1,00 | 0,98 | 0,99 |
| Falha 2 | 0,65 | 0,42 | 0,51 |
| Falha 3 | 0,31 | 0,44 | 0,36 |
| Falha 4 | 0,00 | 0,00 | 0,00 |
| Falha 5 | 0,88 | 0,67 | 0,76 |
| Falha 6 | 0,42 | 0,16 | 0,23 |
| Falha 7 | 0,00 | 0,01 | 0,00 |
| Falha 8 | 0,02 | 0,00 | 0,01 |
| Falha 9 | 0,59 | 0,66 | 0,62 |
| Falha 10 | 0,96 | 0,84 | 0,90 |
| Falha 11 | 0,46 | 0,06 | 0,11 |
| Falha 12 | 0,72 | 0,89 | 0,80 |
| Falha 13 | 0,71 | 0,81 | 0,76 |

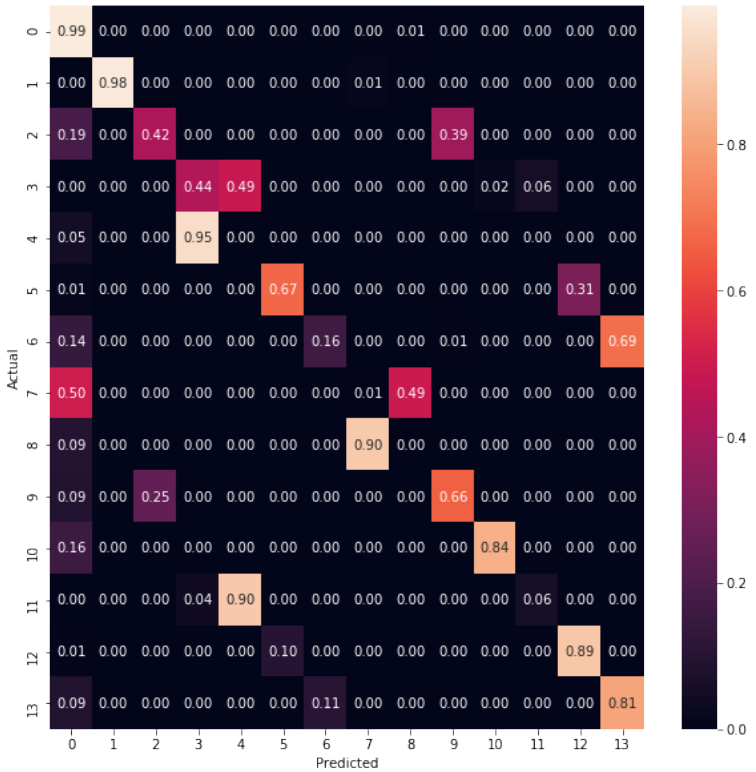


Figura 4.9: Matriz de Confusão obtida por RF com Redução de Variáveis para o Cenário de Incrustação de 13 Falhas.

Comparando a matriz de confusão antes e após a redução de variáveis, é possível chegar a conclusão que a melhora de algumas variáveis parece surgir ao custo da perda de acurácia em outras. Por exemplo, as falhas 5, 6 e 9 tiveram melhora na acurácia através da redução, contudo o algoritmo perdeu a capacidade de identificar as falhas 3, 4 e 11.

Em particular, é observado que o algoritmo RF permanece incapaz de discernir entre as falhas 7 e 8. Esse efeito é percebido para a combinação de diversas variáveis o que se concluiu ser de fato regiões de falha muito parecidas, com impacto muito semelhante nas variáveis de processo, e, ao mesmo tempo, muito próxima da operação normal.

A VI após a redução de variáveis está representada na Figura 4.10. A partir da qual se concluiu que para suprir a falta de outras variáveis, o nível do fundo da coluna C2 com atraso (C2SL(t-1)) passou a ser uma variável de maior importância que SW8T(t) e SW8T(t-1).

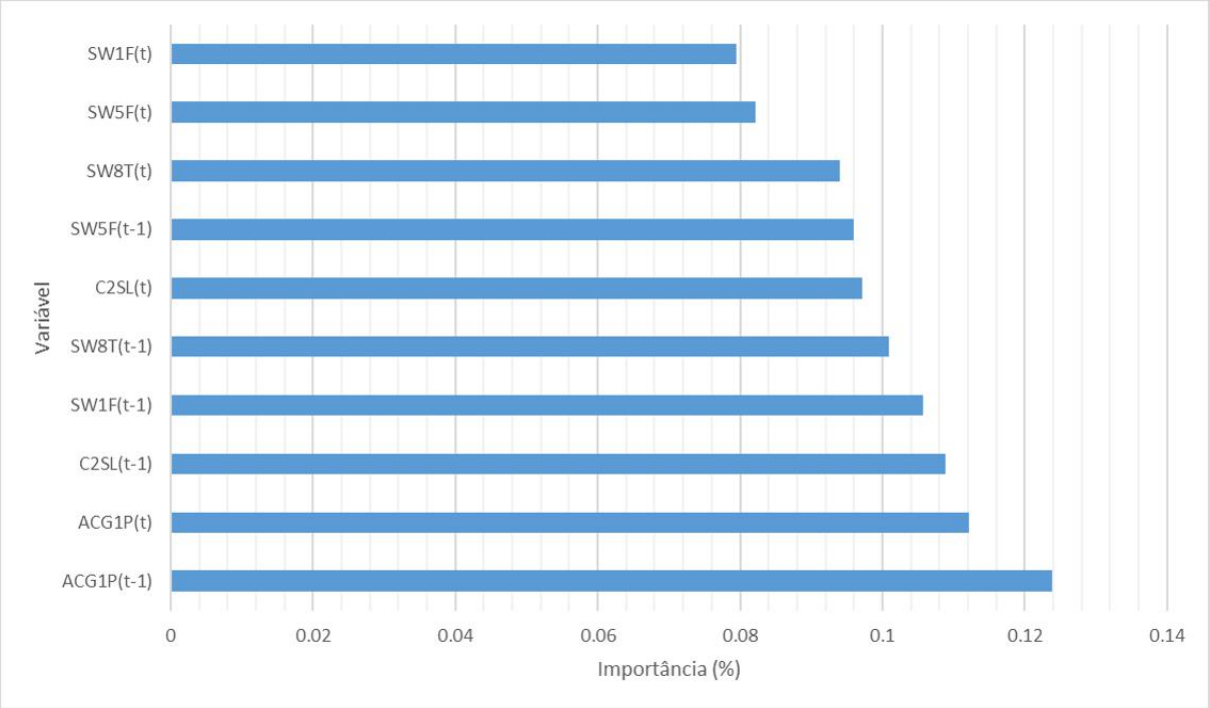


Figura 4.10: Importância de Variável obtida por RF com Redução de Variáveis para o Cenário de Incrustação de 13 Falhas.

#### 4.2.1.3 Máquinas de Vetores de Suporte Linear

Foi então aplicado o algoritmo de SVM Linear nos conjuntos de dados do cenário de incrustação. Após ser realizado o treinamento e o teste, obteve-se a acurácia de, em média, 81,84% para esses dados. Esse resultado foi consideravelmente inferior ao obtido em Nogueira (2021), porém levemente superior ao resultado de RF para esse mesmo cenário. Foi gerado então, o relatório de classificação listado na Tabela 4.8.

É observado então uma melhora significativa nos valores de precisão, sensibilidade e

Tabela 4.8: Relatório de Classificação gerado para SVM Linear no Cenário de Incrustação de 13 Falhas.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,91 | 1,00 | 0,95 |
| Falha 1 | 0,99 | 0,99 | 0,99 |
| Falha 2 | 0,68 | 0,75 | 0,71 |
| Falha 3 | 0,93 | 0,85 | 0,88 |
| Falha 4 | 0,54 | 0,98 | 0,70 |
| Falha 5 | 0,99 | 0,38 | 0,55 |
| Falha 6 | 0,42 | 0,42 | 0,42 |
| Falha 7 | 0,02 | 0,04 | 0,03 |
| Falha 8 | 0,00 | 0,00 | 0,00 |
| Falha 9 | 0,90 | 0,46 | 0,61 |
| Falha 10 | 0,87 | 0,79 | 0,83 |
| Falha 11 | 1,00 | 0,00 | 0,01 |
| Falha 12 | 0,58 | 0,97 | 0,72 |
| Falha 13 | 0,74 | 0,65 | 0,69 |

medida F1, salvo poucas exceções. Espera-se que o modelo de SVM Linear tenha melhor capacidade de identificar as falhas que o modelo RF. Para ilustrar isso, é apresentada a matriz de confusão na Figura 4.11.

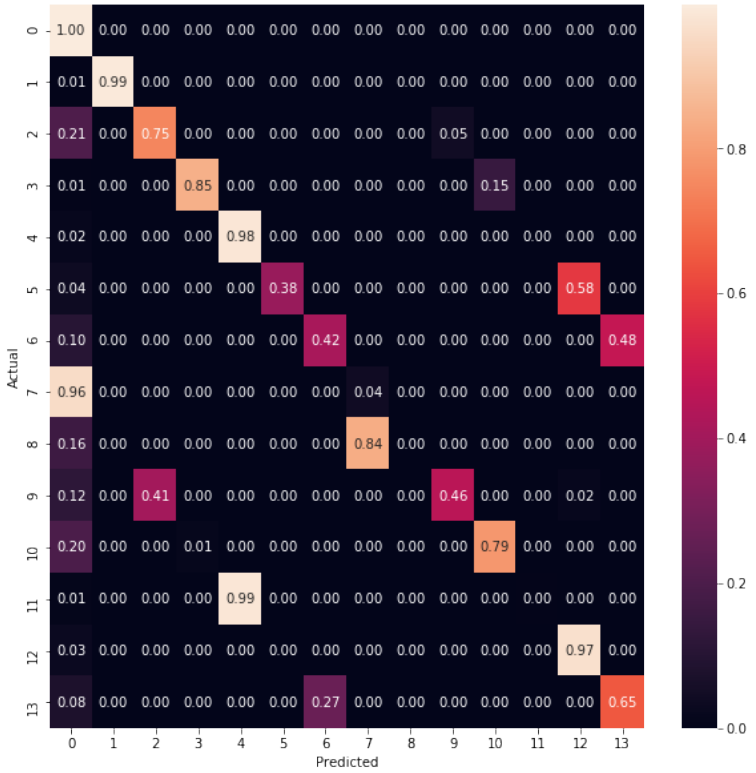


Figura 4.11: Matriz de Confusão obtida por SVM Linear para o Cenário de Incrustação de 13 Falhas.

Torna-se que este algoritmo teve resultados melhores que o RF. SVM Linear foi capaz de distinguir as falhas e, ainda, se elas possuiam ou não incrustação. É evidente, porém que, as falhas 5, 6 e 9 não foram corretamente classificadas para a maior parte das amostras. Especificamente a falha 11 (falha 4 com incrustação) foi quase que totalmente classificada como falha 4.

Além disso, novamente o algoritmo não foi capaz de diferenciar as falha 7 e 8. A falha 7 foi classificada como operação normal quase que completamente, enquanto a falha 8 foi classificada como falha 7. Novamente, se repete o indício que essas regiões são muito semelhantes entre si e a operação normal.

#### 4.2.1.4 Máquina de Vetores de Suporte Gaussiano

O algoritmo de SVM Gaussiano foi aplicado nos conjuntos de dados do cenário de incrustação. Após ser realizado o treinamento e o teste, obteve-se a acurácia de, em média, 80,96% para esses dados. Esse resultado foi consideravelmente inferior ao obtido em Nogueira (2021), e levemente abaixo do valor obtido no SVM Linear. Foi gerado então, o relatório de classificação na Tabela 4.9.

Comparando seus valores com SVM Linear, o relatório de classificação não demonstra melhoras significativas. Para compreensão da diferença entre os resultados do modelo foi gerado a matriz de confusão presente na Figura 4.12.

Tabela 4.9: Relatório de Classificação gerado para SVM Gaussiano no Cenário de Incrustação de 13 Falhas.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,89 | 1,00 | 0,94 |
| Falha 1 | 0,98 | 0,99 | 0,99 |
| Falha 2 | 0,64 | 0,70 | 0,67 |
| Falha 3 | 0,97 | 0,87 | 0,92 |
| Falha 4 | 0,56 | 0,89 | 0,68 |
| Falha 5 | 1,00 | 0,18 | 0,31 |
| Falha 6 | 0,57 | 0,37 | 0,45 |
| Falha 7 | 0,00 | 0,01 | 0,01 |
| Falha 8 | 0,72 | 0,02 | 0,03 |
| Falha 9 | 0,99 | 0,39 | 0,56 |
| Falha 10 | 0,92 | 0,78 | 0,84 |
| Falha 11 | 1,00 | 0,01 | 0,02 |
| Falha 12 | 0,54 | 0,93 | 0,68 |
| Falha 13 | 0,77 | 0,77 | 0,77 |

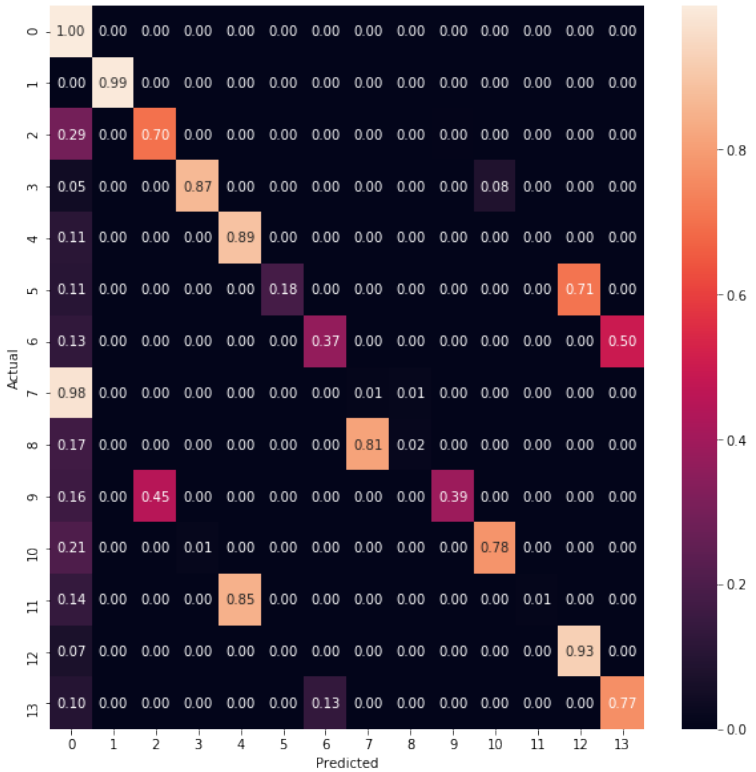


Figura 4.12: Matriz de Confusão obtida por SVM Gaussiano para o Cenário de Incrustação de 13 Falhas.

Primeiramente, nota-se que este algoritmo, em relação aos demais, teve maior dificuldade em discernir as falhas da operação normal. De maneira geral, o resultado desse algoritmo foi muito semelhante ao SVM Linear, entretanto, apresentado uma pior acurácia.

Concluiu-se então que para os três algoritmos apresentados o que obteve melhor resultado foi SVM Linear que alcançou 81,84% na classficiação de classes no cenário de incrustação em H3 para 13 falhas. Porém, a acurácia alcançada foi abaixo do desejado.

### 4.2.2 Cenário de Incrustação com 12 Falhas

A redução no valor da acurácia foi causada em grande parte em decorrência das falhas 7 e 8, que nenhum dos modelos propostos foi capaz de identificar satisfatoriamente. Logo, é assumido que a hipótese em Nogueira (2021) de que a operação com distúrbios leves na presença de incrustação em H3 tenha região de falha mais parecida com falha 1 demonstra-se correta. Foi decidido então, realizar uma nova alteração nos conjuntos de dados de modo a classificar a operação com distúrbios leves na presença de incrustação em H3 como falha 1 com incrustação. Espera-se que, após essa alteração no banco de dados, os modelos sejam capazes de classificar mais corretamente as outras classes também.

#### 4.2.2.1 Banco de Dados para o Cenário de Incrustação

Dessa maneira, os conjuntos de dados passaram a ter 13 classes, compostas pela operação normal e 12 falhas. Sendo necessária alteração nas classes dos conjuntos esclarecida na Tabela 4.10. A distribuição das amostras passou a ser a apresentada na Tabela 4.11.

Tabela 4.10: Descrição das classes do conjunto de treino e teste de Nogueira (2021) modificadas para o cenário de incrustação.

| Classe | Falha Simulada | Incrustação |
|---|---|---|
| Operação Normal | Distúrbios leves | Não |
| Falha 1 | Falha 1 | Não |
| Falha 2 | Falha 2 | Não |
| Falha 3 | Falha 3 | Não |
| Falha 4 | Falha 4 | Não |
| Falha 5 | Falha 5 | Não |
| Falha 6 | Falha 6 | Não |
| Falha 7 | Distúrbios leves e Falha 1 | Sim |
| Falha 8 | Falha 2 | Sim |
| Falha 9 | Falha 3 | Sim |
| Falha 10 | Falha 4 | Sim |
| Falha 11 | Falha 5 | Sim |
| Falha 12 | Falha 6 | Sim |

Tabela 4.11: Distribuição das 12 classes das amostras entre os conjuntos treino e teste do banco de dados de Nogueira (2021) modificado para o cenário de incrustação.

| | **Treino** | | **Teste** | |
|---|---|---|---|---|
| Classe | Número de Amostras | % | Número de Amostras | % |
| Operação Normal | 22244 | 59,29% | 14657 | 65,07% |
| Falha 1 | 562 | 1,50% | 541 | 2,40% |
| Falha 2 | 2747 | 7,32% | 1420 | 6,30% |
| Falha 3 | 504 | 1,34% | 259 | 1,15% |
| Falha 4 | 1124 | 3,00% | 256 | 1,14% |
| Falha 5 | 594 | 1,58% | 378 | 1,68% |
| Falha 6 | 530 | 1,41% | 197 | 0,87% |
| Falha 7 | 2700 | 7,20% | 2201 | 9,77% |
| Falha 8 | 2602 | 6,94% | 1239 | 5,50% |
| Falha 9 | 888 | 2,37% | 404 | 1,79% |
| Falha 10 | 751 | 2,00% | 214 | 0,95% |
| Falha 11 | 1615 | 4,30% | 341 | 1,51% |
| Falha 12 | 658 | 1,75% | 418 | 1,86% |
| **Total** | **37519** | **100%** | **22525** | **100%** |

#### 4.2.2.2 *Random Forest*

Novamente, então, foi aplicado o algoritmo de RF. A acurácia alcançada dessa vez foi, em média, 86,91%. Esse aumento significativo é a consequência de melhora na classificação dos resultados, entretanto ainda distante do obtido em Nogueira (2021). O relatório de classificação é mostrado na Tabela 4.12 e a matriz de confusão na Figura 4.13.

Tabela 4.12: Relatório de Classificação gerado para RF no Cenário de Incrustação de 12 falhas.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,93 | 0,99 | 0,96 |
| Falha 1 | 0,99 | 0,99 | 0,99 |
| Falha 2 | 0,55 | 0,44 | 0,49 |
| Falha 3 | 0,88 | 0,81 | 0,85 |
| Falha 4 | 0,51 | 0,26 | 0,34 |
| Falha 5 | 0,90 | 0,38 | 0,54 |
| Falha 6 | 0,64 | 0,14 | 0,23 |
| Falha 7 | 0,92 | 0,81 | 0,86 |
| Falha 8 | 0,53 | 0,45 | 0,49 |
| Falha 9 | 0,87 | 0,81 | 0,84 |
| Falha 10 | 0,44 | 0,71 | 0,54 |
| Falha 11 | 0,58 | 0,94 | 0,72 |
| Falha 12 | 0,72 | 0,88 | 0,79 |

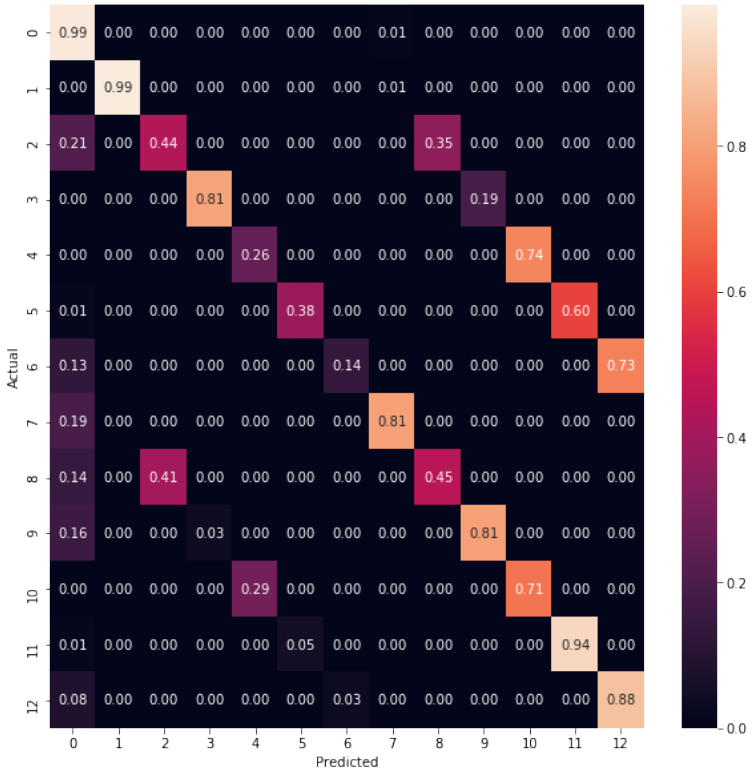


Figura 4.13: Matriz de Confusão obtida por RF para o Cenário de Incrustação de 12 falhas.

O agrupamento na falha 7 foi, relativamente, bem sucedido para essa abordagem. É possível identificar através da comparação do relatório e da matriz do cenário com 12 e com 13 falhas, entretanto, que há poucas melhoras na identificação das classes devido ao agrupamento feito na falha 7. Foi avaliado, então a VI para esse cenário na Figura 4.14.

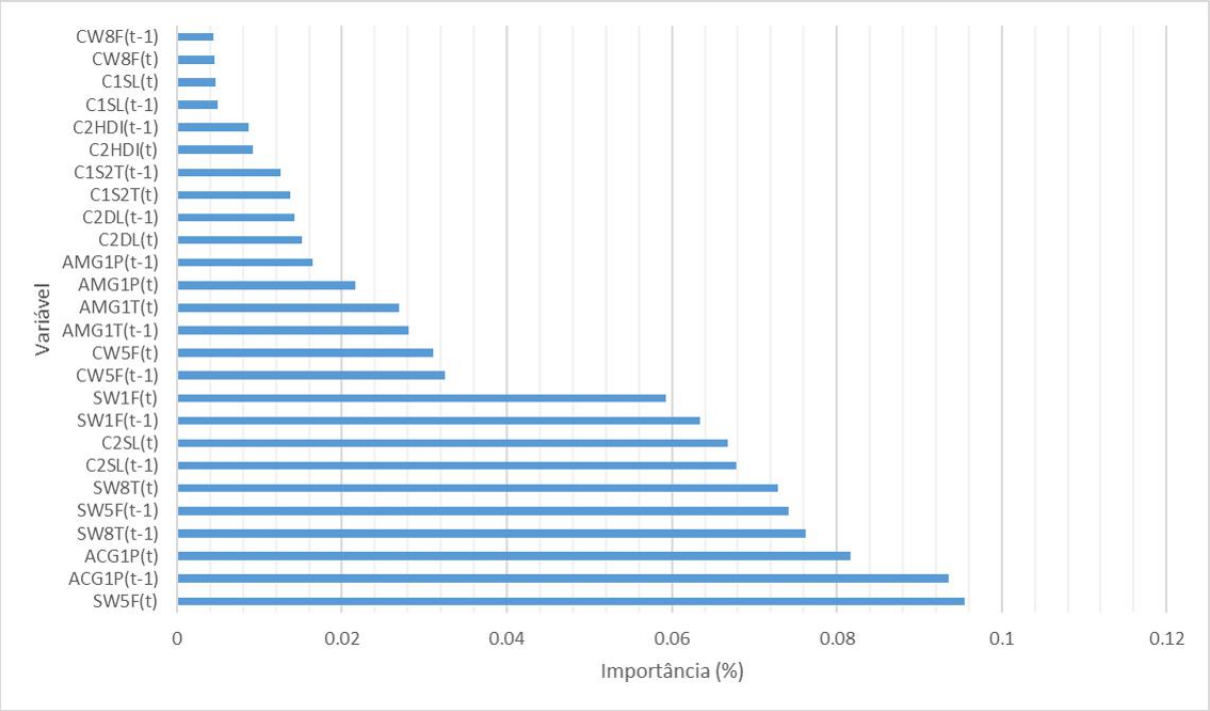


Figura 4.14: Importância de Variável obtida por RF para o Cenário de Incrustação de 12 falhas.

A VI revela que a variável ACG1P(t-1) passou a ter uma relevância mais próxima de SW5F(t). A redução de variável foi realizada novamente removendo as variáveis abaixo de 4,00%, pois, após alguns testes, essa redução se mostrou a que apresentou melhores resultados. O relatório de classificação e a matriz de confusão são apresentadas 4.13 e a matriz de confusão na Figura 4.15, respectivamente.

Tabela 4.13: Relatório de Classificação gerado para RF no Cenário de Incrustação de 12 falhas após a redução de variáveis.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,94 | 0,99 | 0,96 |
| Falha 1 | 1,00 | 0,98 | 0,99 |
| Falha 2 | 0,63 | 0,43 | 0,51 |
| Falha 3 | 0,31 | 0,44 | 0,36 |
| Falha 4 | 0,00 | 0,00 | 0,00 |
| Falha 5 | 0,87 | 0,69 | 0,77 |
| Falha 6 | 0,44 | 0,18 | 0,25 |
| Falha 7 | 0,93 | 0,82 | 0,87 |
| Falha 8 | 0,59 | 0,63 | 0,61 |
| Falha 9 | 0,97 | 0,85 | 0,90 |
| Falha 10 | 0,45 | 0,05 | 0,08 |
| Falha 11 | 0,73 | 0,88 | 0,80 |
| Falha 12 | 0,72 | 0,81 | 0,76 |

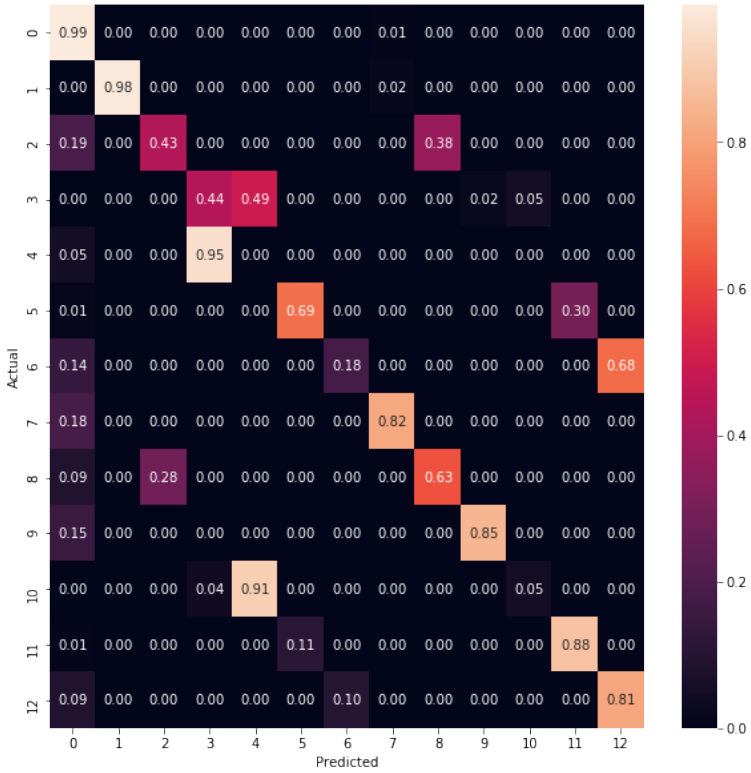


Figura 4.15: Matriz de Confusão obtida por RF para o Cenário de Incrustação de falhas.

A redução de variáveis permitiu a melhora na acurácia de muitas falhas, principalmente as falhas sem incrustação. O valor da acurácia do algoritmo passou a ser 87,02%, em média. As falhas 2, 3 e 4 obtiveram resultados acima de 70%. A falha 6 foi melhor identificada, apesar de sua classificação continuar não sendo satisfatória. Entretanto, a perda de informação prejudicou a identificação de algumas falhas. As falhas 11, principalmente, e 12 perderam significativamente sua sensibilidade. Após uma série de combinações de variáveis, concluiu-se que a modelagem dessas classes não está baseada em uma variável em específico, mas sim no conjunto delas. Além disso, a inclusão de mais variáveis leva na perda de acurácia das primeiras 6 falhas, retornando a um cenário semelhante á antes da redução de variáveis. Verificou-se também a VI após a redução das variáveis (Figura 4.16).

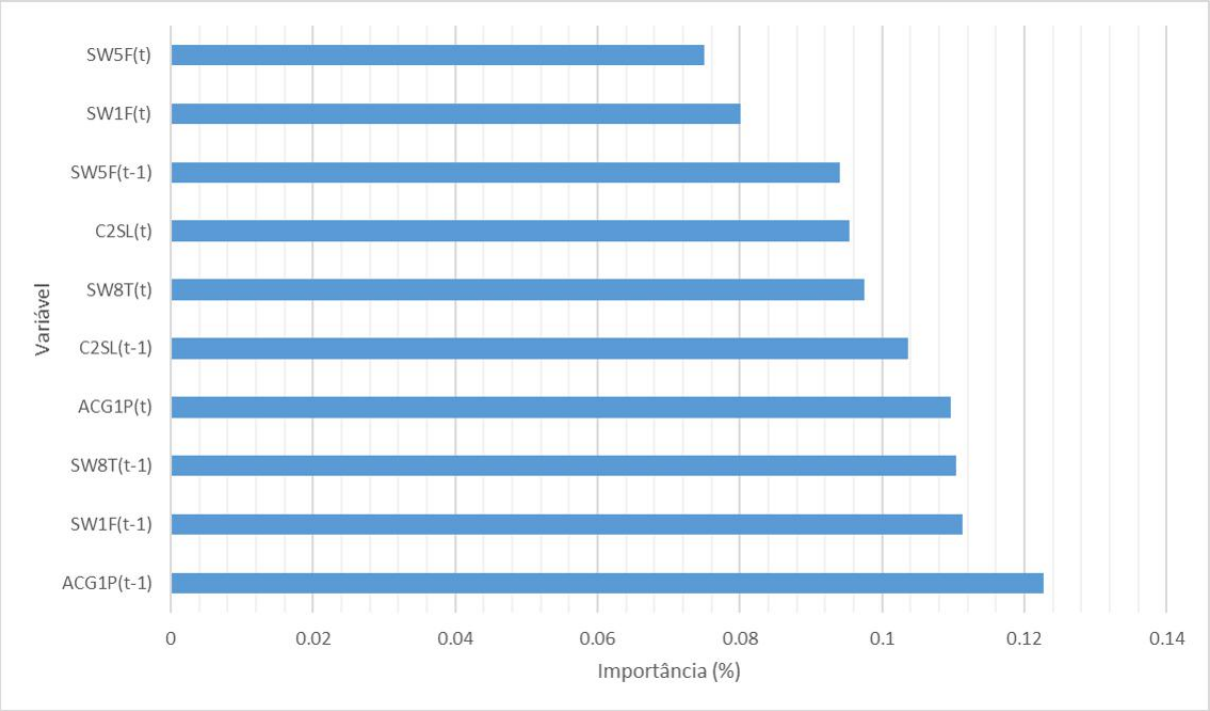


Figura 4.16: Importância de Variável obtida por RF para o Cenário de Incrustação de 12 falhas após a redução das variáveis.

#### 4.2.2.3 Máquinas de Vetores de Suporte Linear

Foi então aplicado o algoritmo de SVM Linear nos conjuntos de dados do cenário de incrustação. Após ser realizado o treinamento e o teste, obteve-se a acurácia de, em média, 88,45% para esses dados. O melhor resultado para o cenário de incrustação. Porém, ainda assim, significativamente abaixo dos resultados de Nogueira (2021). Foram gerados, então, o relatório de classificação (Tabela 4.14) e a matriz de confusão (Figura 4.17).

Tabela 4.14: Relatório de Classificação gerado para SVM Linear no Cenário de Incrustação com 12 falhas.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,92 | 1,00 | 0,96 |
| Falha 1 | 0,99 | 0,99 | 0,99 |
| Falha 2 | 0,68 | 0,75 | 0,71 |
| Falha 3 | 0,93 | 0,85 | 0,88 |
| Falha 4 | 0,54 | 0,98 | 0,70 |
| Falha 5 | 0,99 | 0,38 | 0,55 |
| Falha 6 | 0,42 | 0,42 | 0,42 |
| Falha 7 | 0,99 | 0,69 | 0,82 |
| Falha 8 | 0,90 | 0,46 | 0,61 |
| Falha 9 | 0,87 | 0,79 | 0,83 |
| Falha 10 | 1,00 | 0,00 | 0,01 |
| Falha 11 | 0,58 | 0,97 | 0,72 |
| Falha 12 | 0,74 | 0,65 | 0,69 |

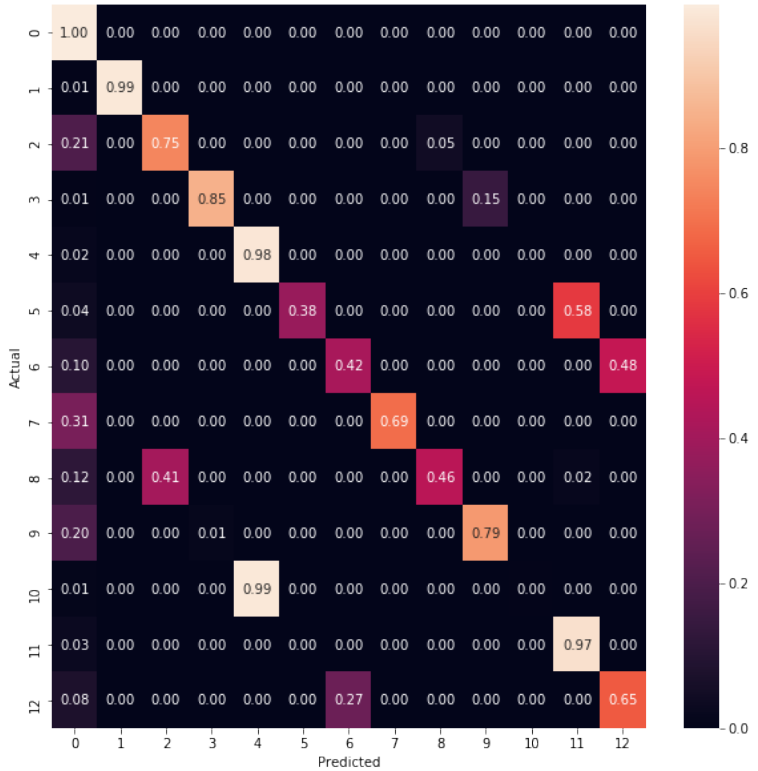


Figura 4.17: Matriz de Confusão obtida por SVM Linear para o Cenário de Incrustação com 12 falhas.

Nota-se que não houveram alterações no resultado do algoritmo exceto pelo agrupamento das antigas falha 7 e 8. O SVM Linear foi capaz de reconhecer essas duas falhas como uma única classe, ratificando a hipótese. Porém, a forma como os modelos SVM realizam a classificação não permitiu que outras melhorias fossem feitas após a alteração dos conjuntos de dados.

#### 4.2.2.4 Máquinas de Vetores de Suporte Gaussiano

O algoritmo de SVM Gaussiano foi aplicado nos conjuntos de dados do cenário de incrustação. Após ser realizado o treinamento e o teste, obteve-se a acurácia de, em média, 87,47% para esses dados. Essa acurácia ficou levemente abaixo do valor para SVM Linear. Foram gerados, então, o relatório de classificação (Tabela 4.15) e a matriz de confusão (Figura 4.18).

Tabela 4.15: Relatório de Classificação gerado para SVM Gaussiano no Cenário de Incrustação.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação Normal | 0,90 | 1,00 | 0,95 |
| Falha 1 | 0,98 | 0,99 | 0,99 |
| Falha 2 | 0,64 | 0,70 | 0,67 |
| Falha 3 | 0,97 | 0,87 | 0,92 |
| Falha 4 | 0,56 | 0,89 | 0,68 |
| Falha 5 | 1,00 | 0,18 | 0,31 |
| Falha 6 | 0,57 | 0,37 | 0,45 |
| Falha 7 | 0,99 | 0,69 | 0,81 |
| Falha 8 | 0,99 | 0,39 | 0,56 |
| Falha 9 | 0,92 | 0,78 | 0,84 |
| Falha 10 | 1,00 | 0,01 | 0,02 |
| Falha 11 | 0,54 | 0,93 | 0,68 |
| Falha 12 | 0,77 | 0,77 | 0,77 |

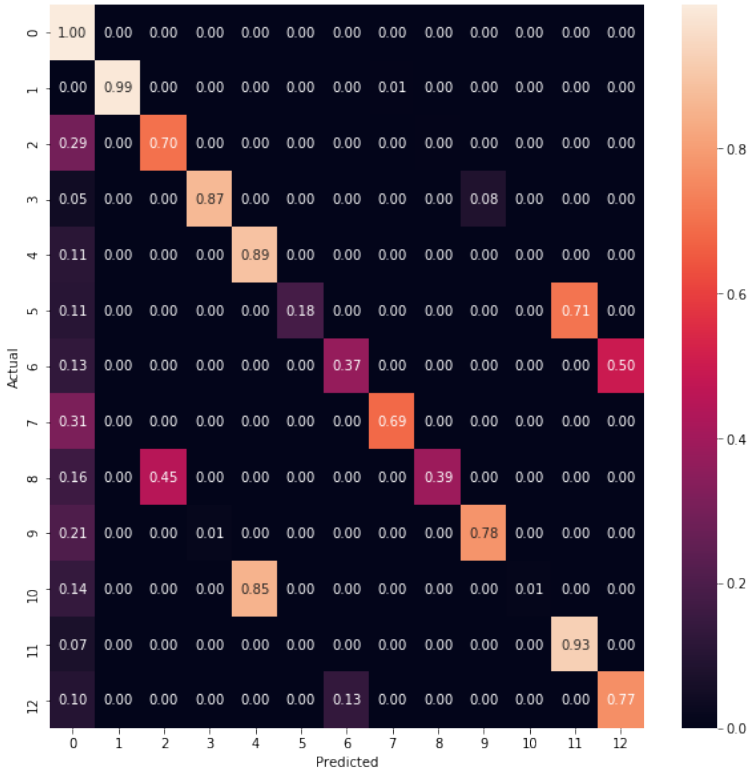


Figura 4.18: Matriz de Confusão obtida por SVM Gaussiano para o Cenário de Incrustação com 12 Falhas.

Assim como em SVM Linear, pode ser observado tanto pelo relatório quanto pela matriz que pouco foi alterado nos resultados. Na verdade, apenas a alteração da falha 7 obteve melhor acurácia. O algoritmo foi capaz de identificar o agrupamento das falhas como uma única classe, ratificando a hipótese.

### 4.2.3 Análise Geral dos Cenários de Incrustação

Foi verificado que a distribuição das amostras em 13 classes era ineficiente. Os valores de acurácia foram 79,42% para RF, 81,84% para SVM Linear e 80,96% para SVM Gaussiano. A simulação da operação normal com distúrbios leves e incrustação em H3 tem comportamento dinâmico muito similar ao aumento significativo na vazão mássica de SW1 e incrustação em H3. Isso confirma essa hipótese de Nogueira (2021).

Para contornar esse problema, as falhas 7 e 8 foram agrupadas, e o novo banco de dados utilizado para treinar novos modelos. O algoritmo que melhor classificou as classes propostas foi SVM Linear. Isso pode ser observado pela sua acurácia de 88,45% frente ao dos demais modelos, 87,47% para SVM Gaussiano e 87,02% para RF após a redução de variáveis (Tabela 4.16).

Em relação a VI, se tornou claro que as variáveis ACG1P(t-1), ACG1P(t), C2SL(t-1), C2SL(t), SW5F(t-1), SW5F(t) SW8T(t-1), SW8T(t), SW1F(t-1) e SW1F(t) foram os mais importantes para a classificação das falhas. As oito primeiras são os pares variável (t)

Tabela 4.16: Acurácia dos métodos aplicados para 12 e 13 falhas do cenário de incrustação.

| Método | 13 Falhas | 12 Falhas |
|---|---|---|
| RF com RV | 79,42% | 87,02% |
| SVM Linear | 81,84% | 88,47% |
| SVM Gaussiano | 80,96% | 87,47% |

e variável com atraso (t-1) que primeiro saem do limite controlável de operação descritos na Tabela 3.4. A última demonstrou ser uma variável de significativa relevância, em especial, após a redução de variável para ambas abordagens, pois a maioria das falhas simuladas é gerada através de um distúrbio na corrente SW1, como foi visto na Tabela 3.2.

Apesar disso, para as falhas relacionadas a pertubações na concentração de contaminantes na corrente SW1, o modelo RF se mostrou bastante ineficiente. Houve uma clara diferença entre o modelo RF e os modelos SVM nas falhas 2 e 3, onde o modelo SVM as identificou com uma sensibilidade muito superior. Observando a matriz de confusão para RF (Figura 4.15), concluiu-se que a redução de variáveis foi prejudicial para elas, indicando a perda de informações importantes para a distinção das classes.

Outro prejuízo acarretado pela redução de variáveis no RF foi a eliminação dos dados da pressão do gás amoniacal. Isso levou a uma perda significativa na sensibilidade e na medida F1 das falhas relacionadas aos sensores de pressão deste gás (falha 4 e falha 10).

Os modelos SVM se destacaram pela sua maior acurácia sem a perda de informações relevantes. Houve poucas diferenças entre a classificação dos algoritmos SVM Linear e Gaussiano. Sendo de forma geral compatíveis os resultados, embora, como já enunciado, o modelo SVM Linear tenha alcançado acurácia superior.

Nenhum modelo foi capaz de diferenciar, entretanto, se a unidade apresentava incrustação na condição de falha no sensor do gás amoniacal. Isso pode ser observado pela falha 10 ter sido classificada quase totalmente como falha 4, apresentando a menor sensibilidade nos relatórios de classificação, não ultrapassando 5%. Esse resultado não foi verificado em nenhuma outra falha, onde a classificação foi dividida entre as condições de falha correspondentes com e sem incrutação, porém não alcançando valores acima de 97% como visto nesse caso. Concluiu-se que, devido ao trocador H3 não alimentar diretamente a coluna 2, a diferença nas variáveis de processo são sutis a ponto de não ser evidente que haja uma distinção no processo. Assim, tornando essa condição inadequada para verificar a presença de incrustação em H3.

O mesmo não acontece com a condição de falha no sensor de pressão do gás ácido, onde verifica-se boa classificação, dessa condição na presença de incrustação (falha 11), com cerca de 97% de classificações corretas. É notável a correlação bem mais direta da temperatura da corrente de entrada em C1 na identificação de falhas no sensor de pressão

da corrente de topo da coluna. Entretanto, a classificação dessa condição sem incrustação ficou dividida em relação a presença de incrustação. Um resultado semelhante ocorreu para o superaquecimento de C1 com incrustação (falha 12), outra condição na qual a temperatura na alimentação tem uma influência significativa.

O oposto ocorreu com a condição de aumento significativo na concentração de $H_2S$ em SW1. A classe sem incrustação foi melhor classificada do que essa mesma condição com incrustação, sendo os resultados desta última divididos entre essas duas classes (falhas 2 e 8).

A condição de aumento da vazão mássica em SW1 (falhas 1 e 7) e o aumento da concentração de $NH_3$ em SW1 (falhas 3 e 9) foram as classes em que foram obtidos os melhores resultados. Apesar de uma leve perda de sensibilidade devido à semelhança operacional da falha 7 com a região de operação normal, ambas as falhas se aproximaram de alcançar um resultado confiável na classificação da incrustação. Ambas as falhas se distinguem por apresentarem pertubações na corrente SW1 exceto pela concentração de $H_2S$. Dessa forma, é esperado que a correlação entre as falhas 2 e 8 e a temperatura da corrente SW8 seja menos significativa do que as falhas 3 e 9 e essa mesma variável de processo.

Conforme esperado, concluiu-se que há uma correlação relevante entre as variáveis de processo manipuladas para obter esses cenários e as variáveis de processo relacionadas à eficiência de troca térmica em H3 e que o algoritmo foi tão eficiente quanto maior a influência entre elas.

Dessa maneira, o resultado obtido apresentou, de maneira geral, uma boa acurácia para esse cenário, possibilitando, hipoteticamente, que com a implementação desse sistema o operador fosse capaz de indicar com uma razoável precisão se há a presença ou não de incrustação na unidade.

## 4.3 Cenários de Pré-Falha

### 4.3.1 Banco de Dados para os Cenários de Pré-Falha

Os conjuntos de dados, conforme já descrito, são obtidos através da aplicação do método de AI escolhido e, então, ambos os conjuntos, treino e teste, são reclassificados. Entretanto, existem duas abordagens: a de uma única região de pré-falha e a de que cada falha possui a sua região de pré-falha. Dessa maneira foi feito para RF, SVM Linear e SVM Gaussiano. É importante destacar que, como observado na Tabela 4.1, a falha quatro não apresenta região de pré-falha, portanto nos resultados para múltiplas pré-falhas a quantidade de pré-falhas é uma unidade menor que a quantidade de falhas.

### 4.3.2 *Random Forest*

Os conjuntos de treino e teste alterados para o cenário de uma região de pré-falha foram alimentados em um algoritmo de RF. Os resultado obtidos estão apresentados na forma do relatório de classificação (Tabela 4.17), matriz de confusão (Figura 4.19) e na VI (Figura 4.20).

Tabela 4.17: Relatório de Classificação gerado por RF no Cenário de uma Região de Pré-Falha.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,94 | 0,99 | 0,96 |
| Falha 1 | 0,94 | 0,99 | 0,97 |
| Falha 2 | 0,98 | 0,99 | 0,99 |
| Falha 3 | 0,96 | 1,00 | 0,98 |
| Falha 4 | 0,99 | 1,00 | 0,99 |
| Falha 5 | 1,00 | 1,00 | 1,00 |
| Falha 6 | 1,00 | 1,00 | 1,00 |
| Pré-Falha | 0,00 | 0,00 | 0,00 |

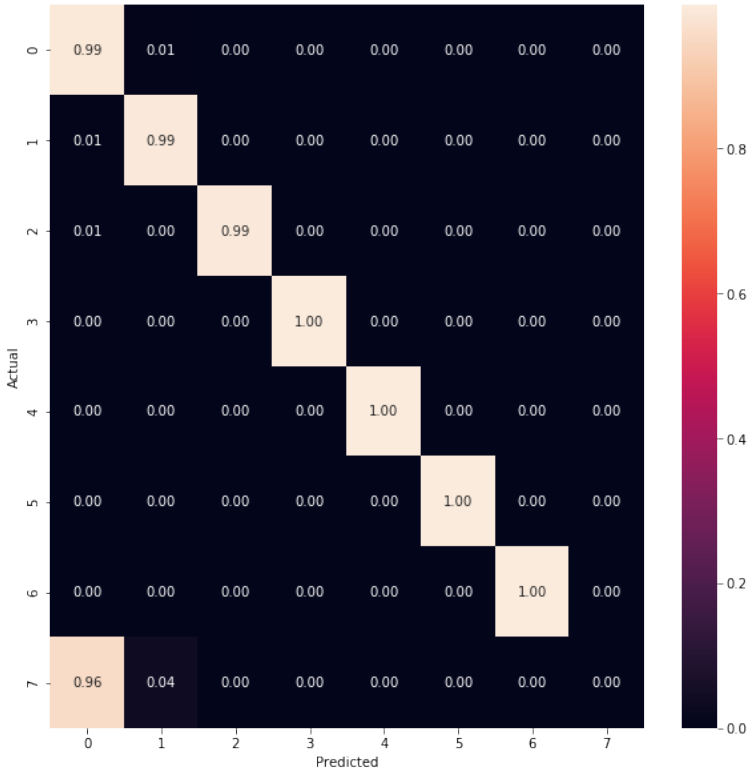


Figura 4.19: Matriz de Confusão obtida por RF para o Cenário de uma Região de Pré-Falha.

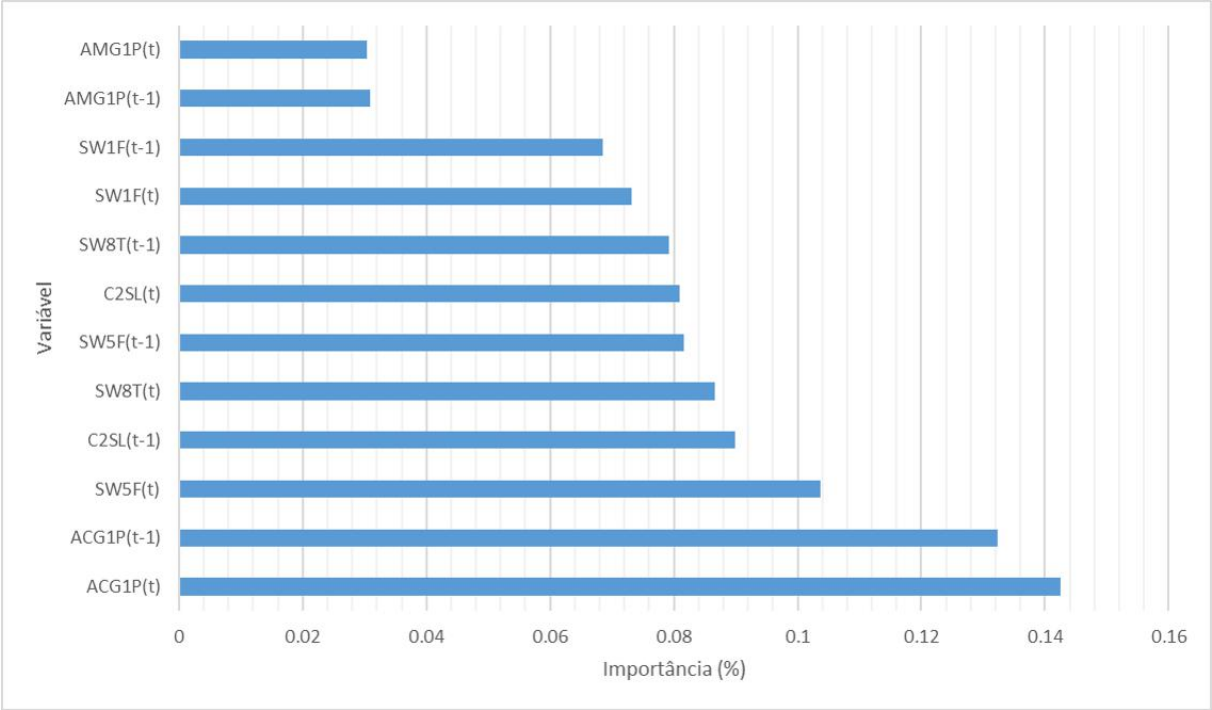


Figura 4.20: Importância de Variável obtida por RF para o Cenário de uma Região de Pré-Falha.

A acurácia para esse cenário foi de 95,09%. Porém, esse resultado não se traduz em

uma correta classificação da região de pré-falha. Tanto pelo relatório de classificação, quanto pela matriz de confusão, é evidente a incapacidade do modelo em separar corretamente a região de pré-falha do comportamento dinâmico com distúrbios leves.

Para o cenário de 6 regiões de pré-falha, os resultados obtidos por RF são representados pelo relatório de classificação (Tabela 4.18), matriz de confusão (Figura 4.21) e VI (Figura 4.22).

Tabela 4.18: Relatório de Classificação gerado por RF no Cenário de Cinco Regiões de Pré-Falha.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,93 | 0,99 | 0,96 |
| Falha 1 | 0,92 | 0,99 | 0,96 |
| Falha 2 | 0,95 | 0,99 | 0,97 |
| Falha 3 | 0,97 | 1,00 | 0,98 |
| Falha 4 | 0,98 | 1,00 | 0,99 |
| Falha 5 | 1,00 | 1,00 | 1,00 |
| Falha 6 | 0,99 | 1,00 | 1,00 |
| Pré-Falha 1 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 2 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 3 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 5 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 6 | 0,00 | 0,00 | 0,00 |

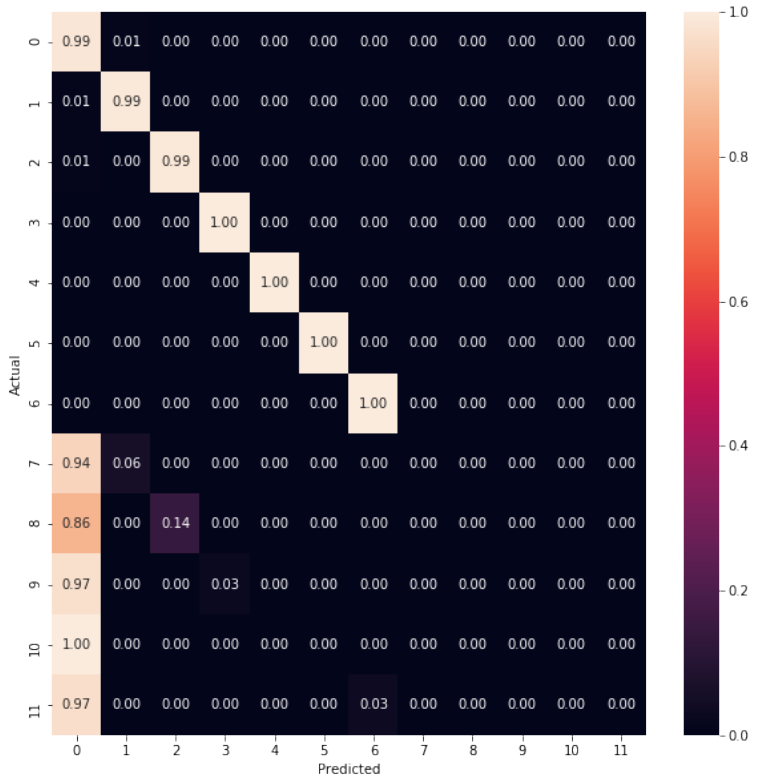


Figura 4.21: Matriz de Confusão obtida por RF para o Cenário de Cinco Regiões de Pré-Falha.

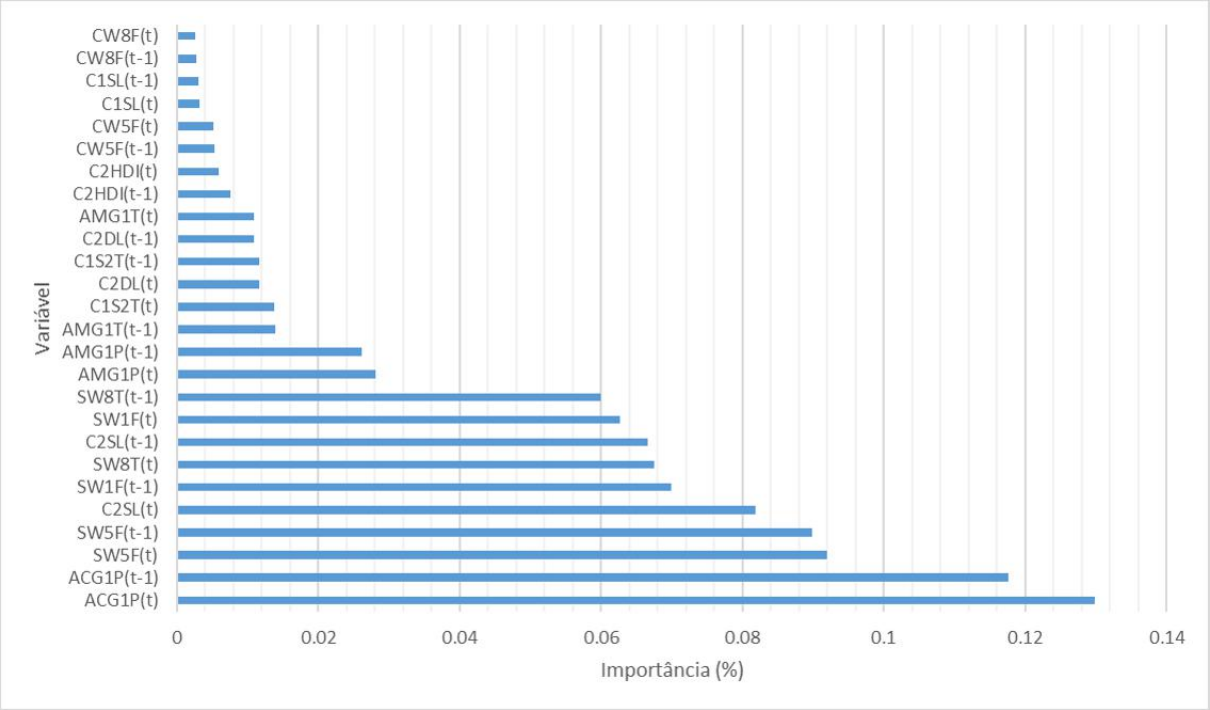


Figura 4.22: Importância de Variável obtida por RF para o Cenário de Cinco Regiões de Pré-Falha.

A acurácia do método foi 93,39%. Porém, novamente, esse valor é muito influenciado

pelas classes de 0 a 6. As classes de pré-falha foram classficadas como operação normal e poucas amostras foram expressadas por sua falha correspondente.

Esses resultados demonstram que as regiões de pré-falha são tão próximas da operação normal que o algoritmo não é capaz de distinguir entre os dois comportamentos. Para contornar essa fraqueza do algoritmo, foi proposta outra maneira de classificar os dados. Para evitar que haja confusão do algoritmo em classificar corretamente essas classes, foi avaliada a classificação em duas etapas: separação da região de operação normal da região anormal e, em seguida, classificação da região de falha e pré-falha. Assim, primeiro, verificou se RF era capaz de distinguir a operação normal das falhas e pré-falhas. Foram criados novos conjuntos de dados com duas classes, a primeira é a operação normal já descrita e a segunda agrupa todos as falhas e pré-falhas apresentadas. Esses conjuntos serão dados como entrada em um novo algoritmo RF. Em seguida, os conjuntos de dados foram trabalhados sem a classe de operação normal, que já teria sido separada na etapa anterior, de modo que RF teria menos dificuldade em classificar a região de pré-falha.

Destaca-se que esse procedimento não foi feito em cascata, esse estudo se baseia na análise preliminar dos dados na qual as etapas 1 e 2 foram feitas de forma independente. Esse método visa avaliar a capacidade dos modelos em separar a classe de Falha & Pré-Falha da operação normal e, separadamente, identificar as classes de falha e pré-falha individualmente na ausência da operação normal. Por isso, não é apropriado calcular a acurácia global desse método. Para todos os métodos foram testadas as abordagens de uma única região de pré-falha e de cinco pré-falhas.

Os resultados da primeira etapa desse novo prodecimento são apresentados através do relatório de classificação (Tabela 4.19), da matriz de confusão (Figura 4.23) e da VI (Figura 4.24). A acurácia encontrada foi de 94,77%.

Tabela 4.19: Relatório de Classificação gerado por são apresentados no Cenário de única Região de Falhas e Pré-Falhas.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,94 | 0,99 | 0,96 |
| Falhas & Pré-Falhas | 0,97 | 0,88 | 0,92 |

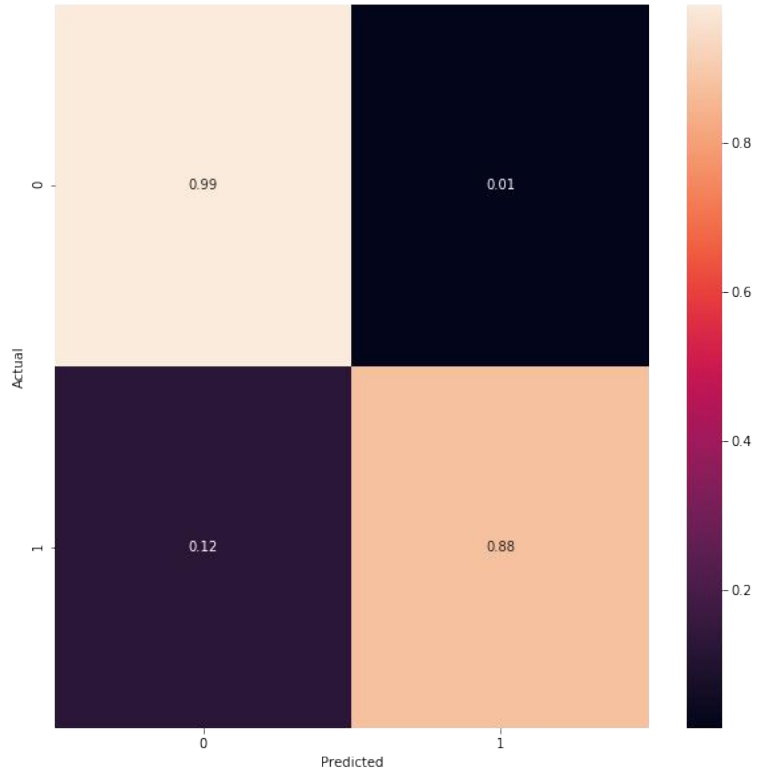


Figura 4.23: Matriz de Confusão obtida por RF no Cenário de de uma única região de Falhas e Pré-Falhas (Etapa 1).

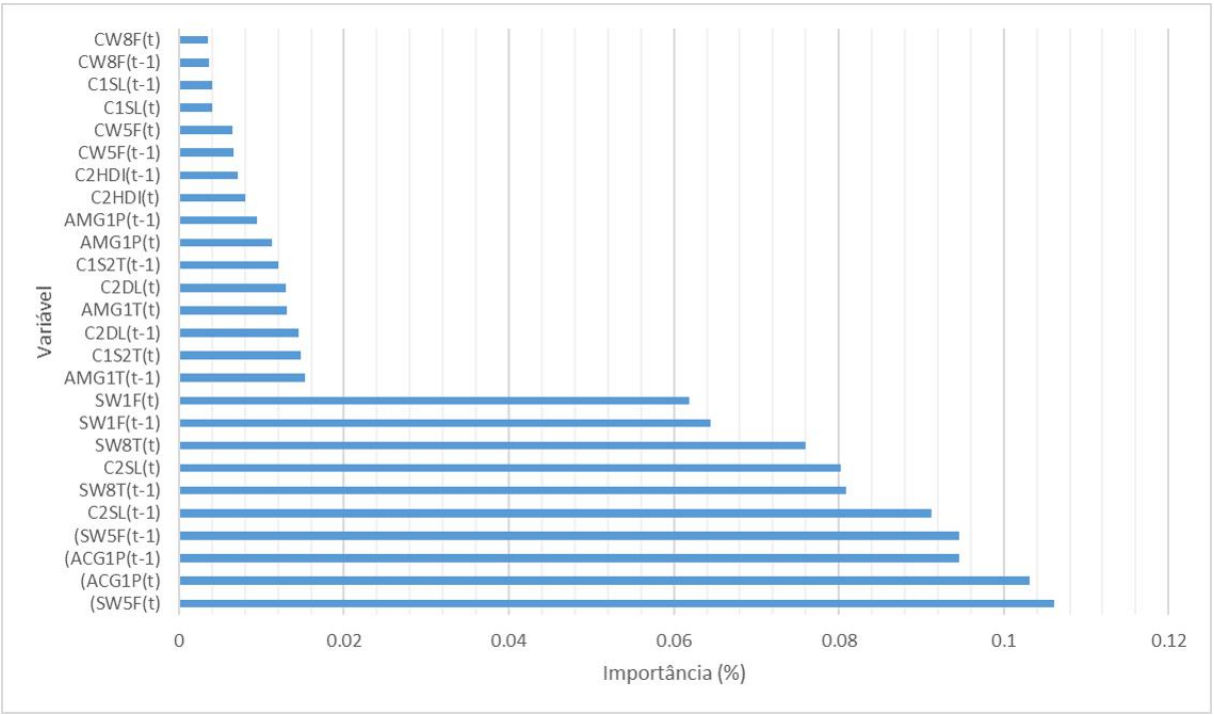


Figura 4.24: Importância de Variável obtida por RF no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1).

A primeira etapa foi realizada com sucesso alcançando 88% de classificações corretas

da região de Falhas & Pré-Falhas, conferindo ao algoritmo uma boa separação entre ela e a região de operação normal que é o principal foco e propósito dessa etapa.

Para os dados sem a operação normal, primeiro o algoritmo foi aplicado para uma única região de pré-falha, isto é, uma única região que agrupa todas as regiões. A acurácia foi de 97,98%. Os resultados estão expressos na Tabela 4.20 e nas Figuras 4.25 e 4.26.

Tabela 4.20: Relatório de Classificação gerado para RF no Cenário de única Região de Pré-Falha Sem Operação Normal.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Falha 1 | 0,98 | 0,99 | 0,99 |
| Falha 2 | 0,97 | 0,99 | 0,98 |
| Falha 3 | 0,99 | 0,99 | 0,99 |
| Falha 4 | 1,00 | 1,00 | 1,00 |
| Falha 5 | 1,00 | 1,00 | 1,00 |
| Falha 6 | 1,00 | 0,99 | 0,99 |
| Pré-Falha | 0,96 | 0,90 | 0,93 |

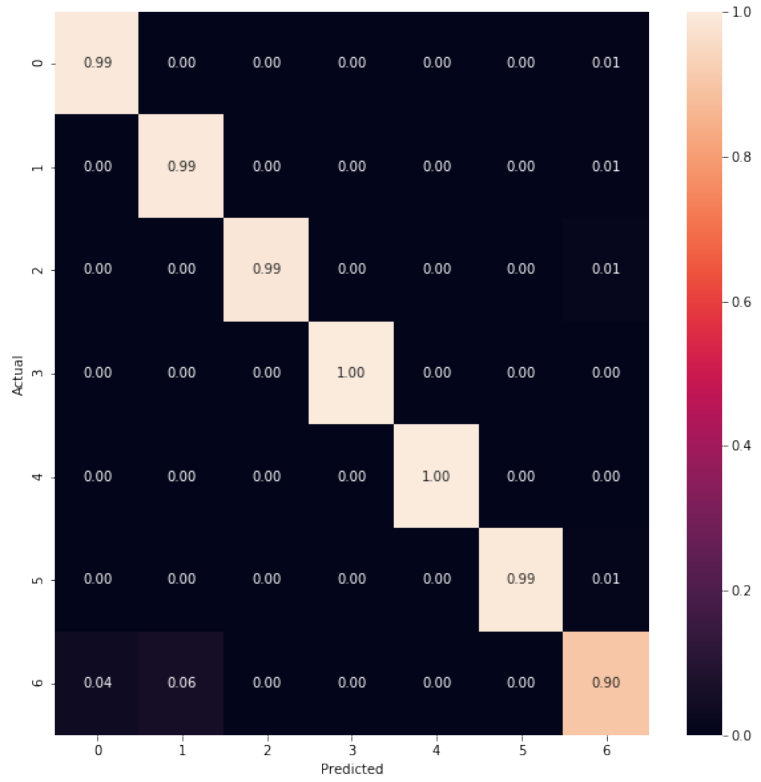


Figura 4.25: Matriz de Confusão obtida por RF no Cenário de única Região de Pré-Falha Sem Operação Normal.

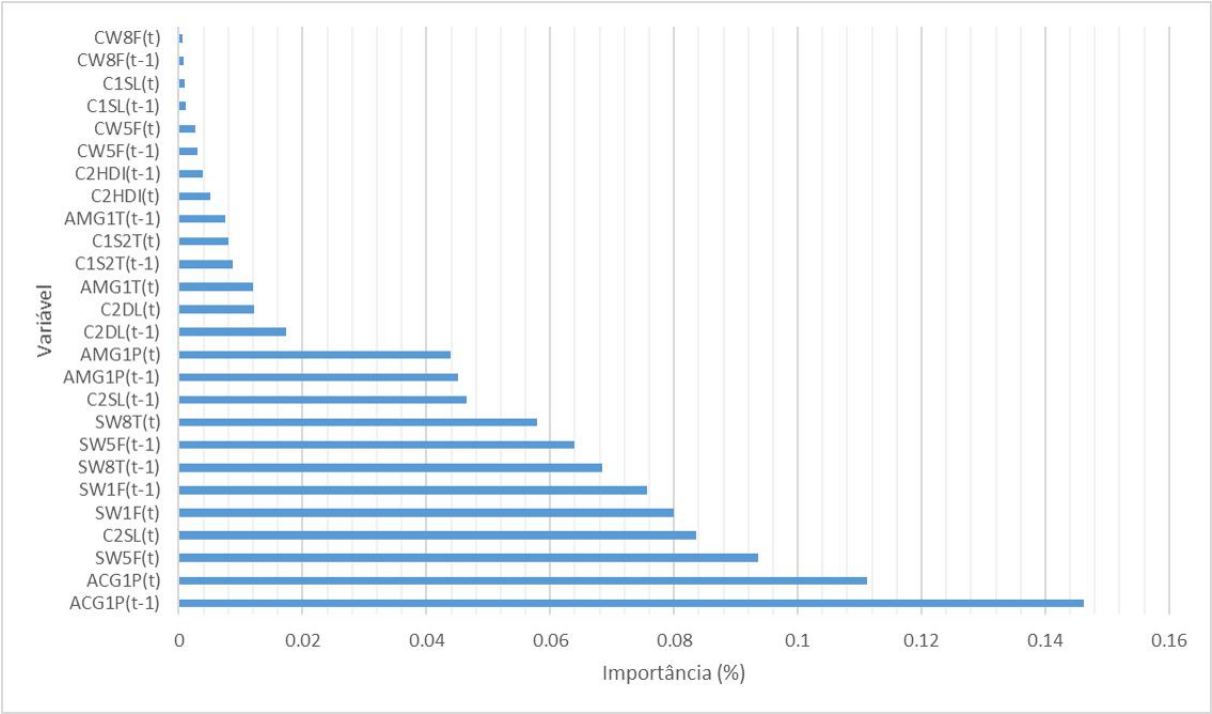


Figura 4.26: Importância de Variável obtida por RF no Cenário de única Região de Pré-Falha Sem Operação Normal.

Então, foi aplicado RF para múltiplas regiões de pré-falhas. A acurácia registrada foi

de 84,93%. Seus resultados são apresentados na Tabela 4.21 e Figuras 4.27 e 4.28.

Tabela 4.21: Relatório de Classificação gerado para RF no Cenário de Cinco Regiões de Pré-Falha Sem Operação Normal.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Falha 1 | 0,79 | 1,00 | 0,88 |
| Falha 2 | 0,82 | 1,00 | 0,90 |
| Falha 3 | 0,95 | 1,00 | 0,98 |
| Falha 4 | 1,00 | 1,00 | 1,00 |
| Falha 5 | 1,00 | 1,00 | 1,00 |
| Falha 6 | 0,95 | 1,00 | 0,97 |
| Pré-Falha 1 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 2 | 0,57 | 0,03 | 0,06 |
| Pré-Falha 3 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 5 | 0,13 | 0,56 | 0,21 |
| Pré-Falha 6 | 1,00 | 0,47 | 0,64 |

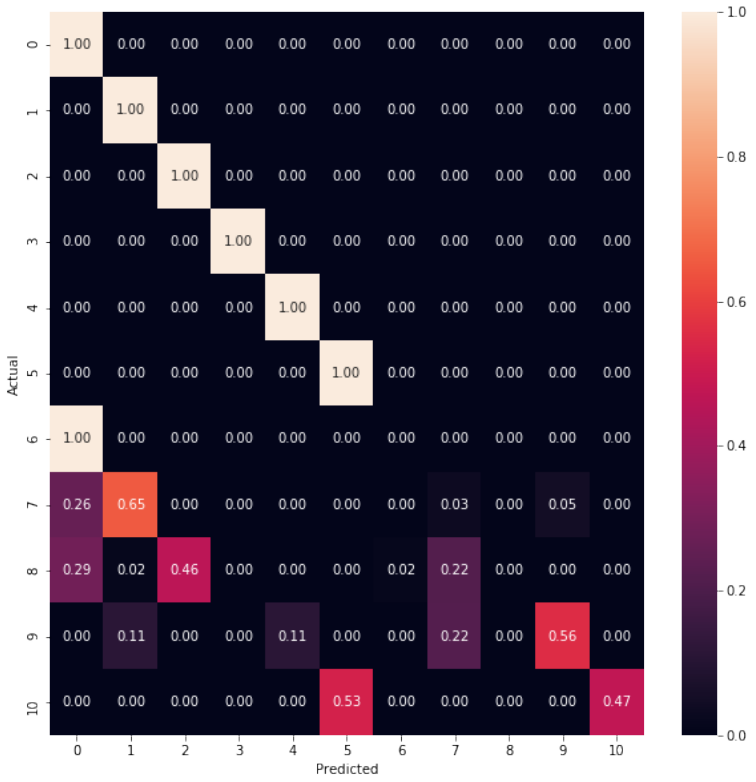


Figura 4.27: Matriz de Confusão obtida por RF no Cenário de Cinco Regiões de Pré-Falha Sem Operação Normal.

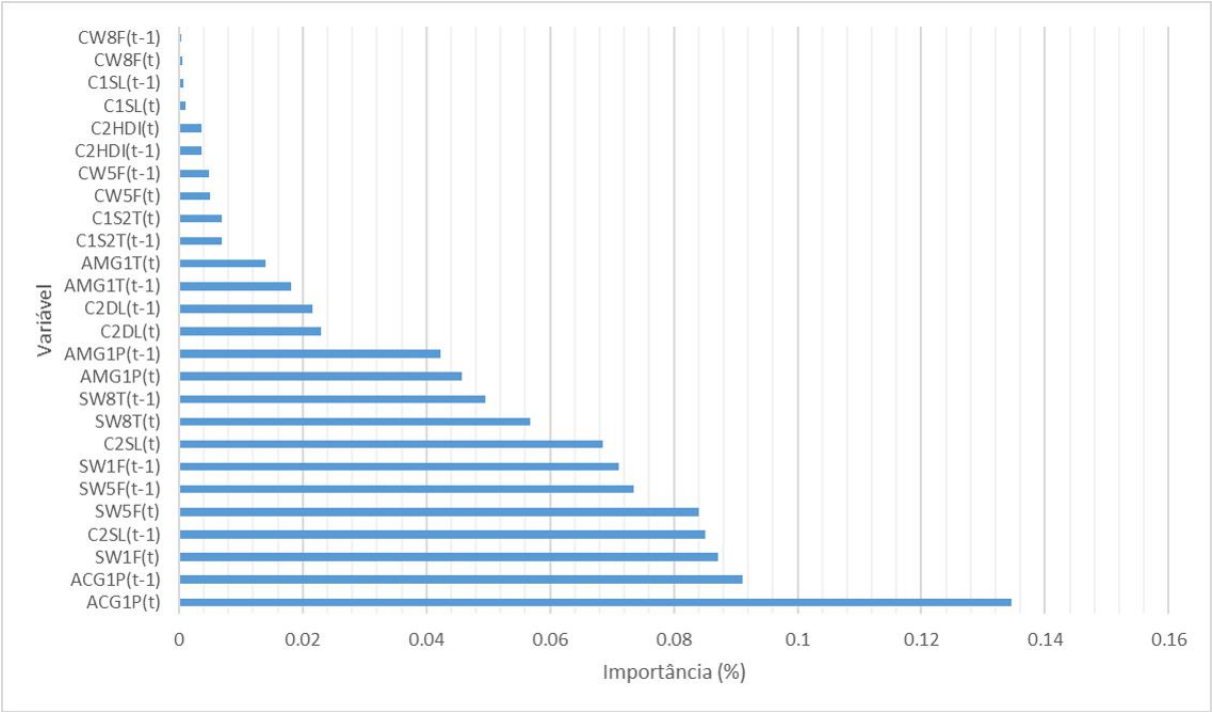


Figura 4.28: Importância de Variável obtida por RF no Cenário de Cinco Regiões de Pré-Falha Sem Operação Normal.

Os resultados para a segunda etapa com o cenário de 6 regiões de pré-falha demons-

traram uma grande dificuldade do algoritmo em classificar corretamente as classes. A sensibilidade e medida F1 foram severamente reduzidas na transição da classicação da região de falha e das regiões de pré-falha, implicando em uma péssima classificação dessa última. Dessa forma, o modelo RF não foi capaz de classificar esse cenário.

### 4.3.3 Máquinas de Vetores de Suporte Linear

A mesma metodologia se seguiu para SVM Linear. Primeiramente, foi avaliado se o algoritmo conseguiria classificar as pré-falhas em uma única classe. Seguem os resultados apresentados na Tabela 4.22 e na Figura 4.29. A acurácia para esse cenário foi igual a 93,84%. Em seguida, foi avaliado o cenário de seis classes de pré-falha. Os seus resultados estão expressos na Tabela 4.23 e na Figura 4.30.

Tabela 4.22: Relatório de Classificação gerado por SVM Linear no Cenário de uma única região de pré-falha.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,92 | 0,99 | 0,96 |
| Falha 1 | 0,98 | 0,97 | 0,98 |
| Falha 2 | 0,97 | 0,99 | 0,98 |
| Falha 3 | 0,95 | 0,98 | 0,96 |
| Falha 4 | 0,98 | 1,00 | 0,99 |
| Falha 5 | 1,00 | 1,00 | 1,00 |
| Falha 6 | 1,00 | 0,99 | 0,99 |
| Pré-Falha | 0,00 | 0,00 | 0,00 |

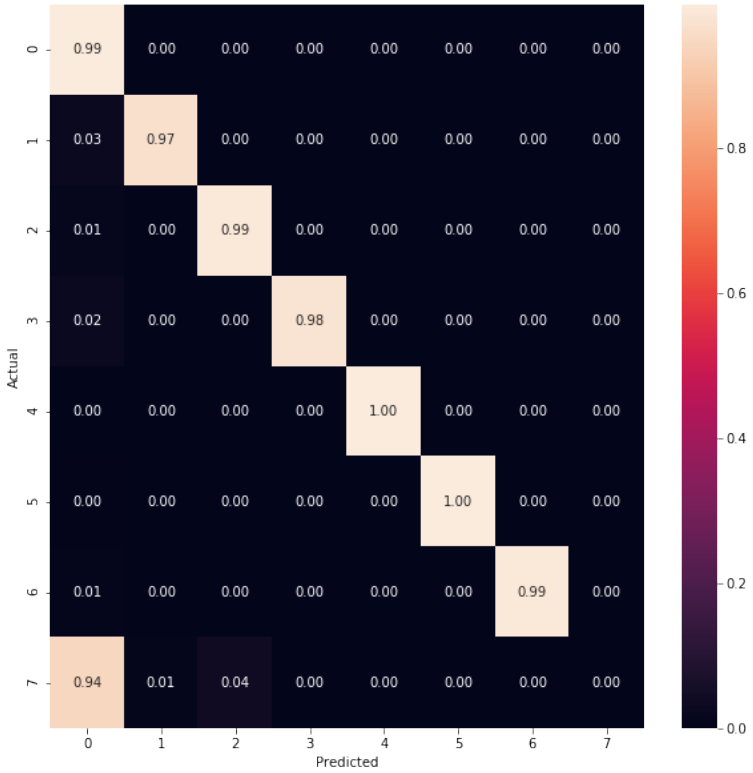


Figura 4.29: Matriz de Confusão obtida por SVM Linear no Cenário de uma única região de pré-falha.

Tabela 4.23: Relatório de Classificação gerado por SVM Linear no Cenário de seis regiões de pré-falha.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,92 | 1,00 | 0,96 |
| Falha 1 | 0,98 | 0,97 | 0,98 |
| Falha 2 | 0,98 | 0,98 | 0,98 |
| Falha 3 | 0,95 | 0,98 | 0,96 |
| Falha 4 | 0,98 | 1,00 | 0,99 |
| Falha 5 | 1,00 | 0,99 | 1,00 |
| Falha 6 | 1,00 | 0,99 | 0,99 |
| Pré-Falha 1 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 2 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 3 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 4 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 5 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 6 | 0,00 | 0,00 | 0,00 |

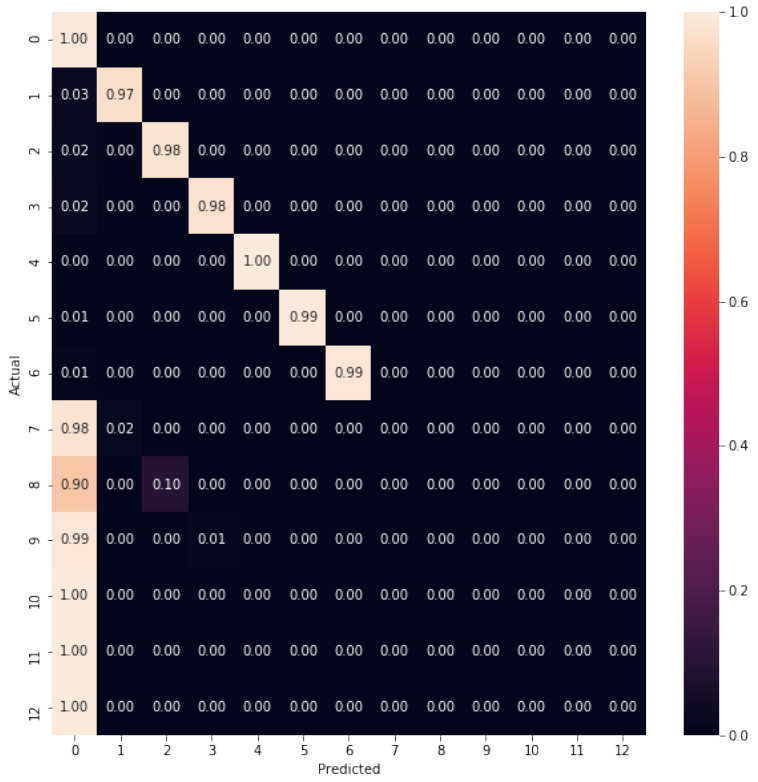


Figura 4.30: Matriz de Confusão obtida por SVM Linear no Cenário de seis regiões de pré-falha.

Assim como para RF, o algoritmo não foi capaz de discernir entre operação normal e as regiões de pré-falha, sejam agrupadas ou separadas. A partir daí, se seguiu o mesmo procedimento feito em RF. O conjunto de dados foi reclassificado em operação normal e falhas & pré-falha e os cenários para uma e múltiplas regiões de pré-falhas.

Assim, a etapa 1 foi realizada. A acurácia foi de 89,66%. Seus resultados são apresentados a seguir (Tabela 4.24 e Figura 4.31).

Tabela 4.24: Relatório de Classificação gerado por SVM Linear no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1).

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,88 | 0,97 | 0,92 |
| Falhas & Pré-Falhas | 0,93 | 0,76 | 0,84 |

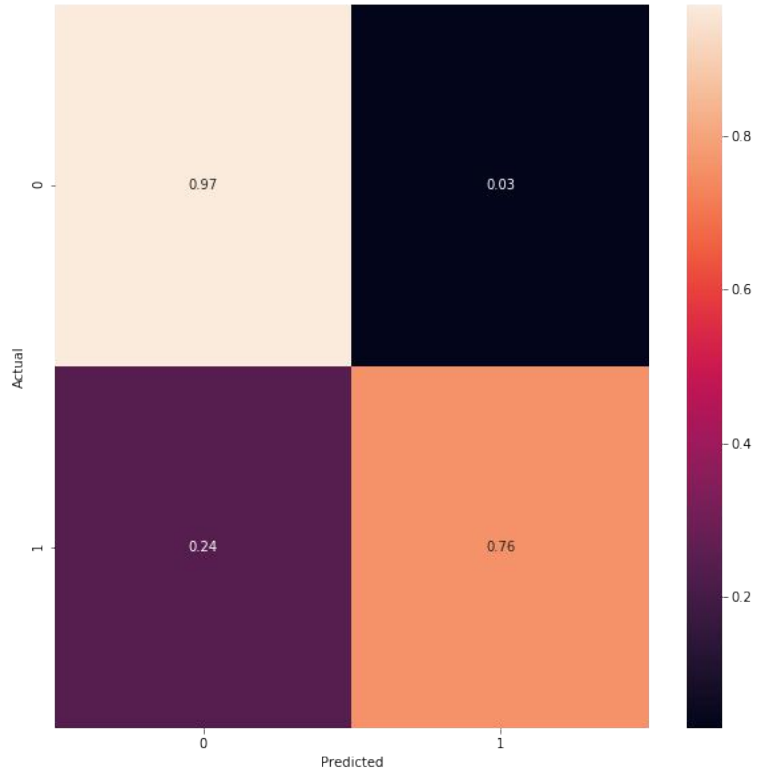


Figura 4.31: Matriz de Confusão obtida por SVM Linear no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1).

Após a remoção da operação normal dos conjuntos de dados, o algoritmo foi treinado e testado para uma única região de pré-falha. A acurácia alcançou 98,20%. Seus resultados são exibidos em Tabela 4.25 e Figura 4.32.

Tabela 4.25: Relatório de Classificação gerado por SVM Linear no Cenário de uma única região de pré-falha sem operação normal.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Falha 1 | 0,99 | 1,00 | 0,99 |
| Falha 2 | 0,97 | 0,99 | 0,98 |
| Falha 3 | 1,00 | 0,99 | 0,99 |
| Falha 4 | 1,00 | 1,00 | 1,00 |
| Falha 5 | 1,00 | 1,00 | 1,00 |
| Falha 6 | 1,00 | 0,99 | 0,99 |
| Pré-Falha | 0,96 | 0,92 | 0,94 |

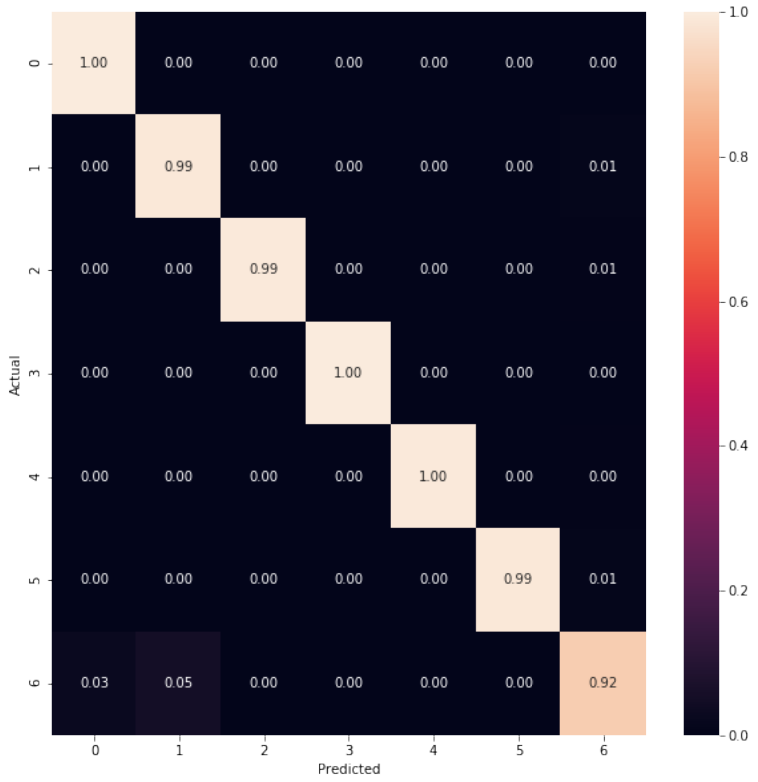


Figura 4.32: Matriz de Confusão obtida por SVM Linear no Cenário de uma única região de pré-falha sem operação normal.

Então, foi realizado a mesma análise para o cenário de múltiplas pré-falhas. A acurácia do método foi 93,96%. Os resultados são exibidos na Tabela 4.26 e na Figura 4.33.

Tabela 4.26: Relatório de Classificação gerado por SVM Linear no Cenário de seis regiões de pré-falha sem operação normal.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Falha 1 | 0,99 | 1,00 | 0,99 |
| Falha 2 | 0,97 | 0,98 | 0,98 |
| Falha 3 | 1,00 | 0,99 | 0,99 |
| Falha 4 | 1,00 | 1,00 | 1,00 |
| Falha 5 | 1,00 | 1,00 | 1,00 |
| Falha 6 | 1,00 | 0,99 | 0,99 |
| Pré-Falha 1 | 0,74 | 0,95 | 0,83 |
| Pré-Falha 2 | 0,63 | 0,22 | 0,32 |
| Pré-Falha 3 | 0,83 | 0,70 | 0,76 |
| Pré-Falha 4 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 5 | 0,09 | 0,42 | 0,14 |
| Pré-Falha 6 | 0,88 | 1,00 | 0,93 |

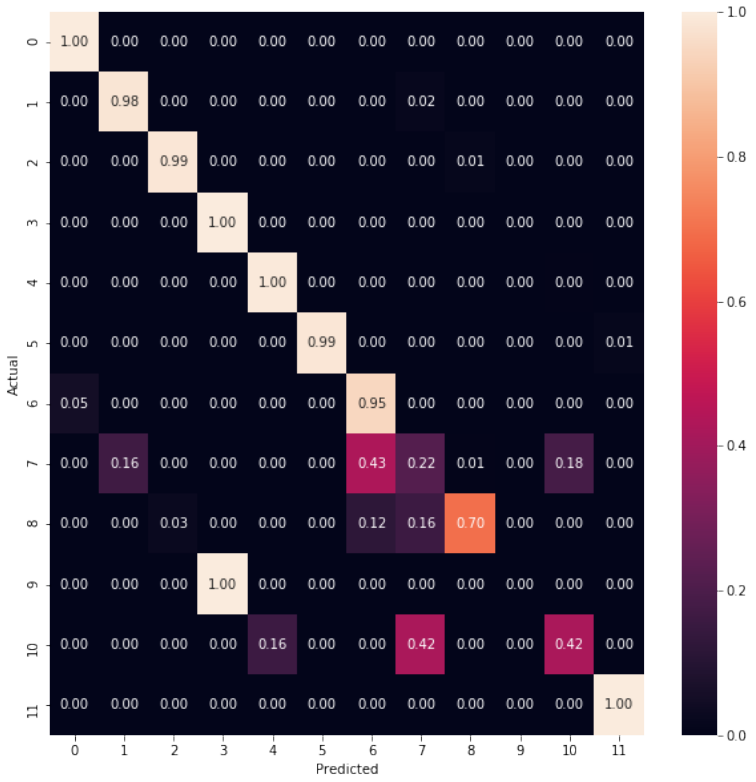


Figura 4.33: Matriz de Confusão obtida por SVM Linear no Cenário de seis regiões de pré-falha sem operação normal.

Em comparação com os resultados para esse mesmo cenário do modelo RF, o algoritmo SVM Linear apresentou uma melhora, se aproximando mais do perfil desejado de classificação. Em particular, houve um grande incremento na sensibilidade das pré-falhas 1, 3 e 6. Entretanto, de maneira geral, permaneceu uma diferença grande entre os valores de sensibilidade e medida F1 das falhas e das pré-falhas.

### 4.3.4 Máquinas de Vetores de Suporte Gaussiano

De forma análoga para SVM Gaussiano, foi avaliado se o algoritmo era capaz de classificar as pré-falhas em uma única classe. A acurácia para esse cenário foi igual a 93,83%. Seguem os resultados apresentados na Tabela 4.27 e na Figura 4.34.

Tabela 4.27: Relatório de Classificação gerado por SVM Gaussiano no Cenário de uma única região de pré-falha.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,92 | 0,99 | 0,96 |
| Falha 1 | 0,98 | 0,97 | 0,98 |
| Falha 2 | 0,97 | 0,99 | 0,98 |
| Falha 3 | 0,95 | 0,98 | 0,96 |
| Falha 4 | 0,98 | 1,00 | 0,99 |
| Falha 5 | 1,00 | 1,00 | 1,00 |
| Falha 6 | 1,00 | 0,99 | 0,99 |
| Pré-Falha | 0,00 | 0,00 | 0,00 |

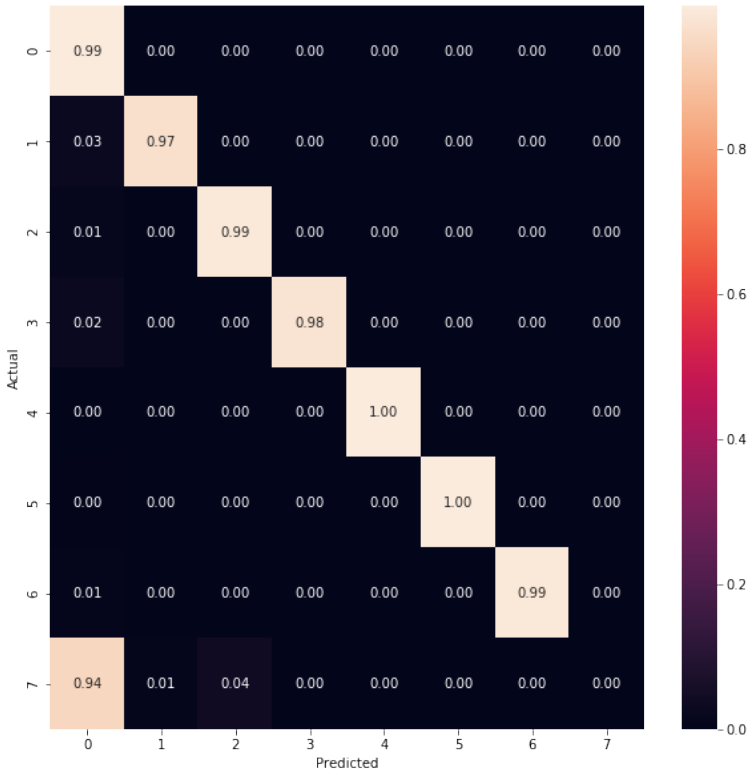


Figura 4.34: Matriz de Confusão obtida por SVM Gaussiano no Cenário de uma única região de pré-falha.

Em seguida, foi avaliado o cenário de seis classes de pré-falha. A acurácia para esse cenário foi igual a 93,75%. Os seus resultados estão expressos na Tabela 4.28 e na Figura 4.35.

Tabela 4.28: Relatório de Classificação gerado por SVM Gaussiano no Cenário de seis regiões de pré-falha.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,92 | 1,00 | 0,96 |
| Falha 1 | 0,98 | 0,97 | 0,98 |
| Falha 2 | 0,98 | 0,98 | 0,98 |
| Falha 3 | 0,95 | 0,98 | 0,96 |
| Falha 4 | 0,98 | 1,00 | 0,99 |
| Falha 5 | 1,00 | 0,99 | 1,00 |
| Falha 6 | 1,00 | 0,99 | 0,99 |
| Pré-Falha 1 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 2 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 3 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 4 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 5 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 6 | 0,00 | 0,00 | 0,00 |

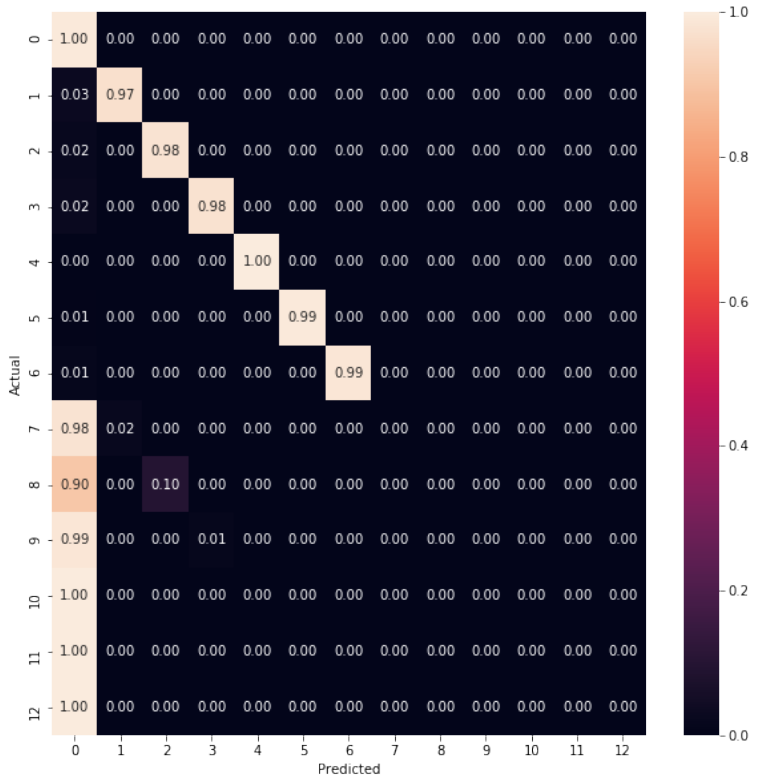


Figura 4.35: Matriz de Confusão obtida por SVM Gaussiano no Cenário de seis regiões de pré-falha.

Assim como para RF, o algoritmo não foi capaz de discernir entre operação normal e pré-falha. A partir daí, se seguiu o mesmo procedimento feito em RF e SVM Linear. O conjunto de dados foi reclassificado em operação normal e falhas & pré-falha e os cenários para uma e múltiplas pré-falhas.

Assim, a etapa 1 foi realizada. A acurácia foi de 93,14%. Seus resultados são apresentados a seguir (Tabela 4.29 e Figura 4.36).

Tabela 4.29: Relatório de Classificação gerado por SVM Gaussiano no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1).

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Operação normal | 0,91 | 1,00 | 0,95 |
| Falhas & Pré-Falhas | 0,99 | 0,81 | 0,89 |

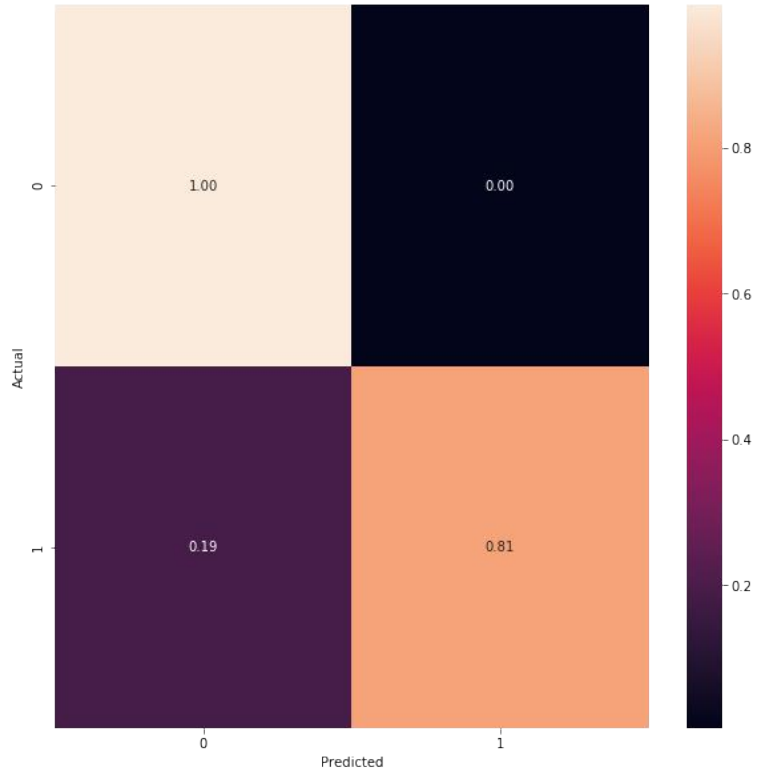


Figura 4.36: Matriz de Confusão obtida por SVM Gaussiano no Cenário de uma única região de Falhas e Pré-Falhas (Etapa 1).

Após a remoção da operação normal dos conjuntos de dados, o algoritmo foi treinado e testado para uma única região de pré-falha. A acurácia alcançou 97,41%. Seus resultados são exibidos na Tabela 4.30 e na Figura 4.37.

Tabela 4.30: Relatório de Classificação gerado por SVM Gaussiano no Cenário de uma única região de pré-falha sem operação normal.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Falha 1 | 0,98 | 0,98 | 0,98 |
| Falha 2 | 0,98 | 0,98 | 0,98 |
| Falha 3 | 0,99 | 0,99 | 0,99 |
| Falha 4 | 1,00 | 1,00 | 1,00 |
| Falha 5 | 1,00 | 0,97 | 0,99 |
| Falha 6 | 1,00 | 0,99 | 0,99 |
| Pré-Falha | 0,91 | 0,92 | 0,92 |

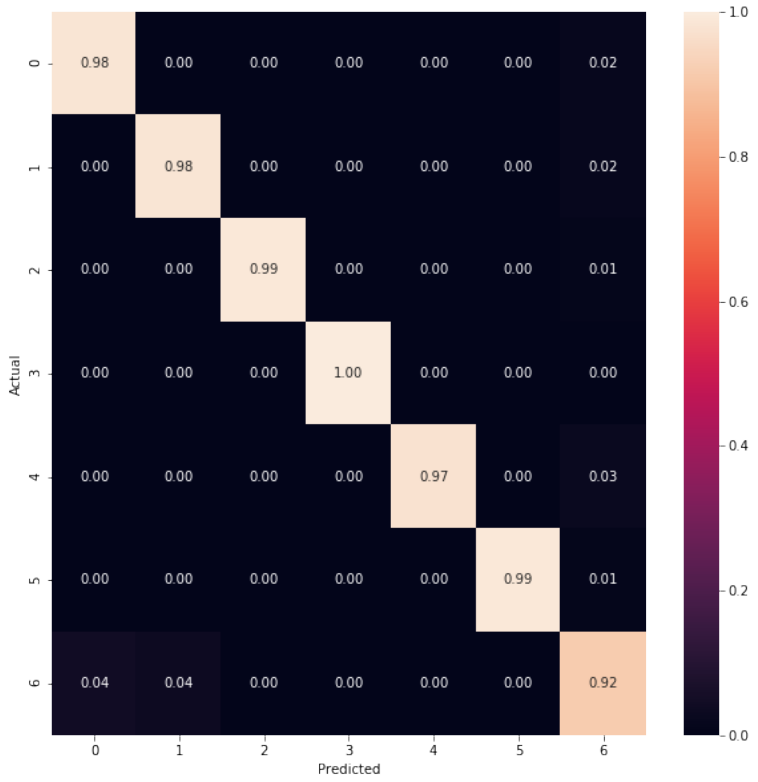


Figura 4.37: Matriz de Confusão obtida por SVM Gaussiano no Cenário de uma única região de pré-falha sem operação normal.

Então, foi realizado a mesma análise para o cenário de múltiplas pré-falhas. A acurácia do método foi 95,59%. Por fim, os resultados são exibidos na Tabela 4.31 e na Figura 4.38.

Tabela 4.31: Relatório de Classificação gerado por SVM Gaussiano no Cenário de seis regiões de pré-falha sem operação normal.

| Classe | Precisão | Sensibilidade | Medida F1 |
|---|---|---|---|
| Falha 1 | 0,99 | 0,96 | 0,97 |
| Falha 2 | 0,97 | 1,00 | 0,99 |
| Falha 3 | 0,99 | 1,00 | 0,99 |
| Falha 4 | 0,94 | 1,00 | 0,97 |
| Falha 5 | 0,98 | 1,00 | 0,99 |
| Falha 6 | 0,99 | 1,00 | 1,00 |
| Pré-Falha 1 | 0,82 | 0,96 | 0,88 |
| Pré-Falha 2 | 0,85 | 0,81 | 0,83 |
| Pré-Falha 3 | 1,00 | 0,45 | 0,62 |
| Pré-Falha 4 | 1,00 | 0,06 | 0,12 |
| Pré-Falha 5 | 0,00 | 0,00 | 0,00 |
| Pré-Falha 6 | 1,00 | 0,94 | 0,97 |

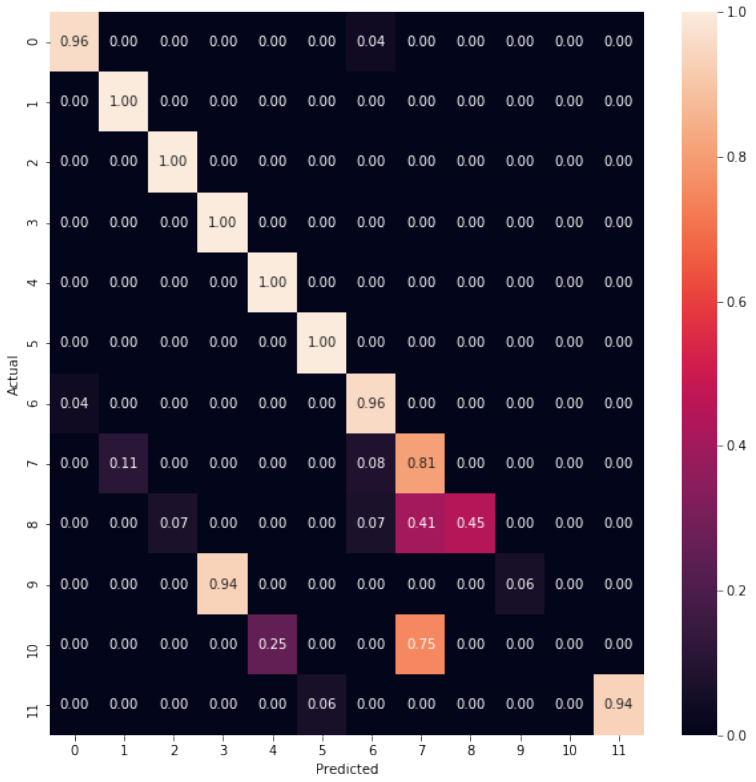


Figura 4.38: Matriz de Confusão obtida por SVM Gaussiano no Cenário de seis regiões de pré-falha sem operação normal.

A análise do cenário encontrado por SVM Gaussiano é análogo ao que foi apresentado para os resultados de SVM Linear (Seção 4.3.4).

### 4.3.5 Análise Geral dos Cenários de Pré-Falha

Os métodos tiveram comportamentos muito semelhantes entre si. Foi observado que a região de pré-falha se encontra muito próxima dos parâmetros da operação normal, implicando em uma baixa sensibilidade da classificação de pré-falha. A alternativa proposta para solucionar esse obstáculo foi classificá-la em etapas, primeiro separando a operação de todas as outras condições de falha e pré-falha e, então, realizando a classificação das falhas e da região de pré-falha. Foram propostos dois cenários, conforme a Figura **??**: a imposição de uma única região de pré-falha e a classificação em regiões de pré-falha distintas para cada falha.

Na primeira etapa, RF alcançou a acurácia de 94,77%, SVM Linear 89,66% e SVM Gaussiano 93,14%. Já para a segunda etapa para uma única região de pré-falha, a acurácia foi de 97,98% no RF, 98,20% no SVM Linear e de 97,41% no SVM Gaussiano. Para a segunda etapa com múltiplas regiões de pré-falha, a acurácia foi de 84,93% no RF, 93,95% no SVM Linear e de 95,59% no SVM Gaussiano (Tabela 4.32).

Para a primeira etapa, o melhor método foi o RF, com acurácia de 94,77%, também com a maior sensibilidade e medida F1 dentre os métodos para a classe de Falhas &

Tabela 4.32: Acurácia dos métodos aplicados para as etapas 1 e 2 do cenário de pré-falha.

| Etapa | RF | SVM Linear | SVM Gaussiano |
|---|---|---|---|
| 1 | 94,77% | 89,66% | 93,14% |
| 2 (uma única região de pré-falha) | 97,98% | 98,20% | 97,41% |
| 2 (múltiplas regiões de pré-falha) | 84,93% | 93,95% | 95,59% |

Pré-Falhas. Para a segunda etapa no cenário de uma única região de pré-falha, a maior acurácia foi de 98,20% no SVM Linear. Apesar da sensibildiade ter sido empatada com o SVM Gaussiano, a precisão e a medida F1 foram superiores para esse método no cenário de uma classe de pré-falha. Por fim, para a segunda etapa com múltiplas regiões de pré-falha, a melhor acurácia foi do modelo SVM Gaussiano chegando a 95,59%.

Dessa forma, todos os três algoritmos aplicados foram capazes de alcançar grande acurácia para o cenário de uma região de pré-falha. Foi verificado também que a região de Pré-Falha é muito difícil de ser classificada em si mesma, isto é, identificar a qual das falhas a pré-falha se refere é uma tarefa que o algoritmo não foi capaz de resolver.

Em relação à VI, não foram observadas mudanças significativas entre a posição relativa das variáveis mais importantes. As mesmas variáveis que eram importantes para o problema original permaneceram em ordem de importância semelhante.

Com isso, pode ser concluído que a região de pré-falha é de fato estreitamente próxima da operação normal, sendo muito difícil distingui-las mesmo utilizando métodos atuais de AI. Porém, acredita-se que, através do procedimento em etapas apresentado nessa seção, foi possível alcançar com um alto grau de sensibilidade na classificação da região de pré-falha. Essa é uma contribuição relevante para a operação da unidade. Isso permite que, caso esse algoritmo fosse aplicado na indústria, o operador fosse alertado que o sistema está em vias de sair da região de operação normal, sendo possível que este seja capaz de tomar medidas antes da falha acontecer. Ainda sim, não seria possível identificar qual a origem ou causa da falha e, com isso, não seria possível auxiliar o operador no processo de tomada de decisão para a mitigação da mesma.

# Capítulo 5

# Conclusões e Considerações Finais

Este trabalho analisou os cenários de incrustação e pré-falha da Unidade de Tratamento de Águas Ácidas através de métodos de Inteligência Artificial. O banco de dados de Nogueira (2021) foi modificado para que as classes pertinentes aos cenários fossem identificadas. Os conjuntos de dados foram implementados em *Random Forest* e Máquinas de Vetores de Suporte (Linear e Gaussiano). Os seus resultados foram analisados por meio de métricas estatísticas apropriadas para problemas de classificação em Aprendizado de Máquinas e da funcionalidade Importância de Variável da biblioteca *scikit-learn* para *Random Forest*.

Para o cenário de incrustação no trocador de calor H3, os métodos apresentaram acurácias baixas em relação à classificação do trabalho original. Isso se deve especialmente à falha 7, caracterizada como distúrbios leves com incrustação. A falha 7 é facilmente confundida com operação normal e com falha 8 (falha 1 com incrustação) em todos os métodos aplicados. A partir dessa observação, cunhou-se a hipótese de que a falha 7 tenha comportamento dinâmico similar ao comportamento da falha 8. Essa hipótese foi confirmada com o aumento da acurácia em todos os métodos diante do agrupamento das falhas 7 e 8 em uma única classe.

Para esse cenário, o algoritmo que obteve melhor acurácia foi SVM Linear com 88,45%. Apesar disso, nenhum método foi capaz de classificar plenamente todas as variáveis. Destaca-se ainda a análise da importância de variável para *Random Forest* que identificou que as variáveis mais importantes para o conjunto de dados original permaneceram com esse *status*. Porém, a variável SW8T(t) e SW8(t-1) tiveram um aumento de importância em relação às demais variáveis. Esse comportamente era esperado pois essa variável é afetada pelo efeito da incrustação no trocador H3.

O cenário de pré-falha é um dos cenários de mais difícil detecção. O comportamento das variáveis se encontra entre os limites controláveis e a região de falha. Dessa forma, para ambas as abordagens as amostras de pré-falha foram classificadas pelos modelos de Inteligência Artificial como distúrbios leves de processo (operação normal). Devido a

similaridade de comportamento das variáveis dessas classes, procurou-se reduzir o impacto da operação normal na classificação. Essa hipótese também foi confirmada. Após a separação da operação normal das demais classes, foi possível classificar as amostras de pré-falha em uma única classe para todos os métodos de Inteligência Artificial. Dentre os métodos, o melhor para a primeira etapa foi RF (acurácia de 94,77%). Para a segunda etapa, SVM Linear foi o melhor método para a classificação do cenário de uma única região de pré-falha, alcançando a acurácia de 98,20%. SVM Gaussiano foi o melhor para o cenário de múltiplas regiões de pré-falha, com acurácia de 95,59%.

Assim, confirmou-se a hipótese da presença da região de pré-falha nessa simulação da Unidade de Tratamento de Águas Ácidas. A classificação da região de pré-falha oferece à equipe de operação da planta a informação do afastamento da operação normal, para uma região de pré-falha. A partir desse dado, é possível que medidas sejam tomadas para evitar a ocorrência de falhas na unidade. Entretanto, não foi possível identificar qual pré-falha é responsável por cada falha específica. Essa informação seria de alto valor para a indústria, pois seria possível prever o acontecimento de uma falha antes dela de fato ocorrer.

A respeito da Importância de Variável para *Random Forest*, as variáveis mais importantes não sofreram grandes alterações no decorrer das abordagens, concluiu-se que as variáveis mais importantes para a identificação do cenário de pré-falha são as mesmas para o cenário das falhas em si.

Apesar de os métodos utilizados terem apresentado comportamento semelhante na classificação das variáveis, SVM Linear se sobressaiu como a melhor escolha para ambos cenários analisados. Foi obtido êxito na classificação da maioria das falhas no cenário de incrustação, após o agrupamento das falhas 7 e 8. No cenário de pré-falha, a remoção da operação normal foi determinante para a classificação correta da classe de uma única pré-falha.

Recomenda-se para trabalhos futuros, o estudo do cenário incrustação em particular da condição de distúrbios leves com incrustação, dado que espera-se que esse seja um cenário possível e, frequente em uma planta industrial. Para o cenário de pré-falha, o estudo da região de pré-falha utilizando outros métodos de Inteligência Artificial e Aprendizado de Máquinas pode esclarecer a dificuldade na sua classificação. A estruturação de uma metodologia em cascata na qual apenas as amostras classificadas corretamente como Falhas & Pré-Falhas seja utilizada na segunda etapa acrescentaria de forma relevante ao tema. Além disso, dada a importância de sistemas preditivos de falhas para a indústria, a pesquisa da classificação da região de pré-falha para cada falha seria capaz de grande contribuição para esse tópico.

# Referências Bibliográficas


Agência Brasil. *ANP interdita refinaria de Paulínia para evitar novos acidentes*. 2018. Acessado em 16 de Fevereiro de 2022. Disponível em: ⟨https://agenciabrasil.ebc.com.br/economia/noticia/2018-08/anp-interdita-refinaria-de-paulinia-para-evitar-novos-acidentes⟩.

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.

ARUNTHAVANATHAN, R. et al. Fault detection and diagnosis in process system using artificial intelligence-based cognitive technique. *Computers & Chemical Engineering*, Elsevier, v. 134, p. 106697, 2020.

BABATUNDE, A. O.; RAY, W. H. Process dynamics, modeling and control. *Oxford University Press*, 1994.

BARROS, D. J. S. Investigação do efeito de variáveis de processo na eficiência de remoção de $H_2S$ em unidade de tratamento de águas ácidas de duas torres. UFPR, 2016.

BELATO, A.; LIMA, R.; ODDONE, R. Hydrocracking–a way to produce high quality low sulphur middle distillates. In: ONEPETRO. *17th World Petroleum Congress*. [S.l.], 2002.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

CHIANG, L.; LU, B.; CASTILLO, I. Big data analytics in chemical engineering. *Annual review of chemical and biomolecular engineering*, Annual Reviews, v. 8, p. 63–85, 2017.

CHIANG, L.; RUSSELL, E.; BRAATZ, R. *Fault detection and diagnosis in industrial systems*. [S.l.]: IOP Publishing, 2001.

CHOLLET, F. *Deep learning with Python*. [S.l.]: Simon and Schuster, 2021.

COLBECK, I.; MACKENZIE, A. R. Air pollution by photochemical oxidants. 1994.

CONAMA. *RESOLUÇÃO No 436, DE 22 DE DEZEMBRO DE 2011*. 2011. Acessado em 13 de Outubro de 2021. Disponível em: ⟨https://www.udop.com.br/download/legislacao/meio/institucional_site_juridico/Conama_Resolucao\%20436.pdf⟩.

CONWAY, D. *The Data Science Venn Diagram*. 2015. Acessado em 18 de

Outubro de 2021. Disponível em: ⟨http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram⟩.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.

EAP. *United States Environment Protection Agency: What is Acid Rain?* 2021. Acessado em 13 de Outubro de 2021. Disponível em: ⟨https://www.epa.gov/acidrain/what-acid-rain⟩.

FAN, S.-K. S. et al. Data-driven approach for fault detection and diagnostic in semiconductor manufacturing. *IEEE Transactions on Automation Science and Engineering*, IEEE, v. 17, n. 4, p. 1925–1936, 2020.

G1 Rio. *Vazamento de gás intoxica funcionários na Reduc, em Caxias*. 2013. Acessado em 16 de Fevereiro de 2022. Disponível em: ⟨https://g1.globo.com/rio-de-janeiro/noticia/2013/07/vazamento-de-gas-intoxica-funcionarios-na-reduc-em-caxias.html⟩.

GE, Z.; SONG, Z.; GAO, F. Review of recent research on data-based process monitoring. *Industrial & Engineering Chemistry Research*, ACS Publications, v. 52, n. 10, p. 3543–3562, 2013.

HATCHER, N.; WEILAND, R. Reliable design of sour water strippers. *Petroleum technology quarterly*, v. 17, n. 4, 2012.

KENSELL, W.; QUINLAN, M. Chapter 11.1 the mw kellogg company refinery sulfur management. *Handbook of Petroleum Refining Process*, 1996.

KNUST, C. M. Análise de superfícies de respostas para projeto de unidades de tratamento de Águas Ácidas. UFRJ, 2013.

KU, W.; STORER, R. H.; GEORGAKIS, C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, Elsevier, v. 30, n. 1, p. 179–196, 1995.

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, W. et al. Transfer learning for process fault diagnosis: Knowledge transfer from simulation to physical processes. *Computers & Chemical Engineering*, Elsevier, v. 139, p. 106904, 2020.

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.

MITROVIC, J. *Heat Exchangers: Basics Design Applications*. [S.l.]: BoD–Books on Demand, 2012.

MORADO, H. P. M. C. Minimização de emissões de unidades de águas ácidas: Modelos substitutos para controle de carga térmica. UFRJ, 2019.

NOGUEIRA, J. N. P. Fault detection and diagnosis of a two column sour water treatment unit based on artificial intelligence algorithms. UFRJ, 2021.

NOGUEIRA, J. N. P.; de SOUZA Jr, M. B.; MELO, P. A. Detecção e diagnóstico de falhas em unidade de tratamento de águas ácidas com random forest. In: *ANAIS DO 23° CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA, Gramado*. [S.l.: s.n.], 2021.

PARK, Y.-J.; FAN, S.-K. S.; HSU, C.-Y. A review on fault detection and process diagnostics in industrial processes. *Processes*, Multidisciplinary Digital Publishing Institute, v. 8, n. 9, p. 1123, 2020.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, JMLR. org, v. 12, p. 2825–2830, 2011.

POE, W. A.; MOKHATAB, S. *Modeling, control, and optimization of natural gas processing plants*. [S.l.]: gulf professional publishing, 2016.

QUINLAN, M.; HATI, A. *Processing $NH_3$ acid gas in a sulphur recovery unit*. 2010.

RAMLI, N. M. Advanced process control. In: *Advanced Applications for Artificial Neural Networks*. [S.l.]: IntechOpen, 2017.

RICH, E. *Artificial Intelligence, Company, New York, New York*. [S.l.]: Mcgraw-Hill Book, 1983.

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.

SARTORI, I. et al. Detecção, diagnóstico e correção de falhas: Uma proposição consistente de definições e terminologias. *Ciência e Engenharia*, v. 21, p. 41–53, 2012.

SHU, Y. et al. Abnormal situation management: Challenges and opportunities in the big data era. *Computers & Chemical Engineering*, Elsevier, v. 91, p. 104–113, 2016.

SILVA, I.; ALENCAR, J.; DANIELSKI, L. Influência de variáveis de processo na simulação de unidades de águas ácidas de refinaria. In: *CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA–(COBEQ)*. [S.l.: s.n.], 2014. v. 10.

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. UFRJ, 2017.

THAKUR, N. *The differences between Data Science, Artificial Intelligence, Machine Learning, and Deep Learning*. 2020. Acessado em 18 de Outubro de 2021. Disponível em: ⟨https://ai.plainenglish.io/data-science-vs-artificial-intelligence-vs-machine-learning-vs-deep-learning-50d3718d51e5⟩.

VENKATASUBRAMANIAN, V. et al. A review of process fault detection and diagnosis:

Part i: Quantitative model-based methods. *Computers & chemical engineering*, Elsevier, v. 27, n. 3, p. 293–311, 2003.

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

WOJSZNIS, W. K.; MEHTA, A.; THIELE, D. *Robust process model identification in model based control techniques*. [S.l.]: Google Patents, 2010. US Patent 7,840,287.

# Apêndice A

# Fluxograma da Simulação Dinâmica

A Figura A.1 é uma versão ampliada do fluxograma da simulação dinâmica apresentada na seção 3.1.

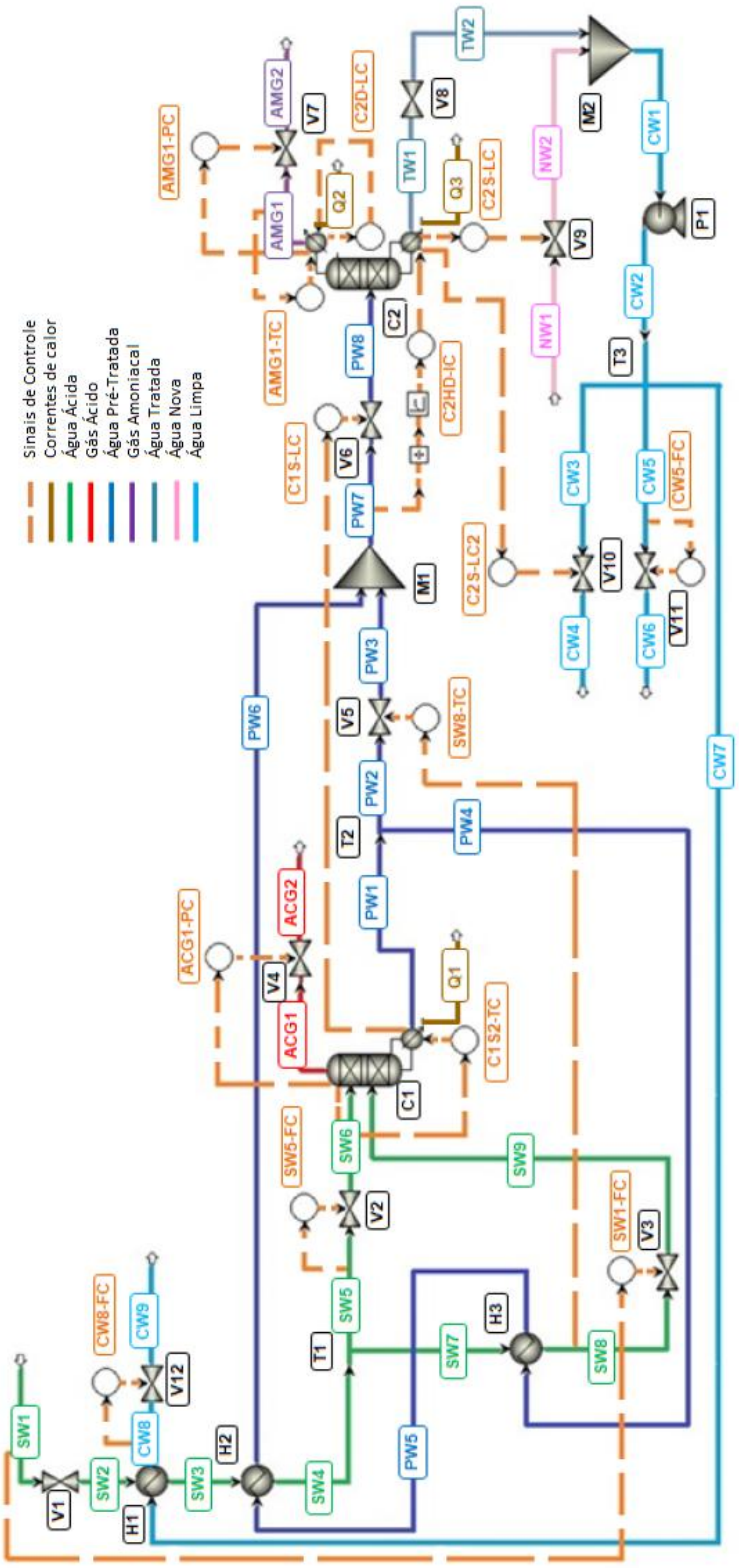


Figura A.1: Versão ampliada do Fluxograma da simulação dinâmica da Unidade de Tratamento de Águas Ácidas com duas torres. Fonte: Nogueira (2021) (legenda traduzida).