

**UNIVERSIDADE FEDERAL DO PAMPA
ENGENHARIA QUÍMICA**

KAMILLA VERA SAPATA

**AVALIAÇÃO DE TÉCNICAS DE *MACHINE LEARNING* NA DETECÇÃO E
DIAGNÓSTICO DE FALHAS EM VÁLVULAS DE CONTROLE**

**Bagé
2021**

KAMILLA VERA SAPATA

**AVALIAÇÃO DE TÉCNICAS DE *MACHINE LEARNING* NA DETECÇÃO E
DIAGNÓSTICO DE FALHAS EM VÁLVULAS DE CONTROLE**

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

Orientador: Prof. Dr. Alexandre Denes Arruda

Coorientadora: Dra. Andressa Apio

Coorientador: Dr. Jônathan William Vergani Dambros

**Bagé
2021**

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

S241a Sapata, Kamilla Vera

Avaliação de técnicas de Machine Learning na detecção e diagnóstico de falhas em válvulas de controle / Kamilla Vera Sapata.

83 p.

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

"Orientação: Alexandre Denes Arruda".

1. Machine Learning. 2. Falhas. 3. Válvulas de controle. 4. Python. 5. Análise de dados. I. Título.

KAMILLA VERA SAPATA

AVALIAÇÃO DE TÉCNICAS DE MACHINE LEARNING NA DETECÇÃO E DIAGNÓSTICO DE FALHAS EM VÁLVULAS DE CONTROLE

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

Dissertação defendida e aprovada em: 08 de outubro de 2021.

Banca examinadora:

Prof. Dr. Alexandre Denes Arruda
Orientador
UNIPAMPA

Dr. Andressa Apio
Co orientador
CEO LATOS
CPF: 027.695.280-41

Dr. Jônathan William Vergani Dambros
Co orientador
CEO LATOS
CPF: 018.098.130-77

Prof. Dr. Rodolfo Rodrigues
UFSM



Assinado eletronicamente por **ALEXANDRE DENES ARRUDA, PROFESSOR DO MAGISTERIO SUPERIOR**, em 13/10/2021, às 13:19, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **Jônathan William Vergani Dambeos, Usuário Externo**, em 13/10/2021, às 19:43, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **Andressa Apio, Usuário Externo**, em 13/10/2021, às 20:07, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **Rodolfo Rodrigues, Usuário Externo**, em 14/10/2021, às 10:43, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



A autenticidade deste documento pode ser conferida no site https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0637496** e o código CRC **9919D18D**.

Dedico este trabalho a minha família e amigos.

AGRADECIMENTOS

À Universidade Federal do Pampa, pela qualidade do ensino e estrutura proporcionada para a minha formação profissional.

A minha família, principalmente meu pai José Pedro, minha mãe Silvia Simone e meu irmão Fellipe por todo apoio e exemplo de união, amor e perseverança.

Ao meu orientador Prof. Dr. Alexandre Arruda, pela confiança, paciência e apoio na busca de novos conhecimentos.

Aos meus amigos, que estiveram comigo durante esta jornada, principalmente a Nathália Fialho, por todo companheirismo, paciência e apoio.

Aos professores do curso de Engenharia Química da UNIPAMPA, pelos ensinamentos compartilhados.

“Devemos promover a coragem onde há medo, promover o acordo onde existe conflito, e inspirar esperança onde há desespero.”

Nelson Mandela

RESUMO

A automação dos processos é uma das características da Indústria 4.0 que gera uma maior disponibilidade de dados históricos provenientes do extenso monitoramento de equipamentos, processos e produtos. Dessa forma, surge a necessidade de desenvolver meios que possibilitem a análise dessas informações em busca de soluções e oportunidades para otimizar a produção. As metodologias baseadas em sistemas inteligentes estão sendo cada vez mais aplicadas e estudadas para esta finalidade. As técnicas de *Machine Learning* estudam a elaboração de algoritmos computacionais que permitam o desenvolvimento de modelos a partir da análise de dados fornecidos e, com isso, a generalização da informação aprendida na forma de previsões. Na indústria química a área de Detecção e Diagnóstico de falhas, utiliza estes dados para minimizar as falhas e monitorar o comportamento do processo frente a estas condições atípicas. Dessa forma, o objetivo deste trabalho é avaliar diferentes técnicas de *Machine Learning* na Detecção e Diagnóstico de falhas em válvulas de controle. O processo químico estudado é um vaso de acúmulo e separação de misturas, desenvolvido no *software* UniSim® Design. A linguagem Python será utilizada para o processamento dos dados e para o desenvolvimento dos algoritmos. A metodologia HAZOP será aplicada para determinar a válvula de controle que apresenta maior risco ao processo. Nesta válvula, serão aplicados dois tipos de falhas, agarramento e supressão de pressão. Para Detecção e Diagnóstico das falhas serão utilizados os modelos de Árvores de Decisão, Redes Neurais e Florestas Randômicas. O UniSim® Design proporcionou a simulação e modelagem do processo químico, assim como o dimensionamento dos equipamentos e instrumentos utilizados no sistema. O diâmetro obtido para o vaso de separação foi de 0,76 m, e o comprimento foi de 2,27 m. Através da metodologia HAZOP, foi possível identificar as situações de risco e suas consequências o processo químico. E dessa forma, determinou-se que a válvula de controle que apresenta maior risco em caso de falhas é a válvula VLV-001. A análise exploratória dos dados foi necessária para o entendimento do processo químico simulado, e resposta das variáveis frente a modificações do sistema. A técnica de Floresta Randômica foi o modelo que apresentou os melhores resultados na detecção e diagnóstico das falhas na válvula de controle com acurácia de 82% nos testes. Este resultado pode ser otimizado após

o ajuste dos hiperparâmetros, obtendo-se uma acurácia de 84%. Mediante o exposto, conclui-se que as técnicas de *Machine Learning* são capazes de realizar a detecção e diagnóstico de falhas em válvulas de controle.

Palavras-chave: *Machine Learning*. Falhas. Válvulas de controle. Python.

ABSTRACT

Process automation is one of the characteristics of Industry 4.0 that generates greater availability of historical data from extensive monitoring of equipment, processes and products. Thus, there is a need to develop means that enable the analysis of this information in search of solutions and opportunities to optimize production. The methodologies based on intelligent systems are being increasingly applied and studied for this purpose. Machine Learning techniques study the development of computational algorithms that allow the development of models based on the analysis of provided data and, therefore, the generalization of the information learned in the form of predictions. In the chemical industry, the Failure Detection and Diagnosis area uses this data to minimize failures and monitor the behavior of the process under these unusual conditions. Thus, the objective of this work is to evaluate different Machine Learning techniques in the Detection and Diagnosis of failures in control valves. The chemical process studied is an accumulation and separation vessel for mixtures, developed in the UniSim® Design software. The Python language will be used for data processing and algorithm development. The HAZOP methodology will be applied to determine the control valve that poses the greatest risk to the process. In this valve, two types of faults will be applied, gripping and pressure suppression. For Detection and Diagnosis of failures, models of Decision Trees, Neural Networks and Random Forests will be used. UniSim® Design provided the simulation and modeling of the chemical process, as well as the dimensioning of the equipment and instruments used in the system. The diameter obtained for the separation vessel was 0.76 m, and the length was 2.27 m. Through the HAZOP methodology, it was possible to identify the risk situations and their consequences in the chemical process. Thus, it was determined that the control valve that presents the greatest risk in case of failures is the VLV-001 valve. Exploratory data analysis was necessary to understand the simulated chemical process and the response of the variables to system modifications. The Random Forest technique was the model that presented the best results in the detection and diagnosis of failures in the control valve, with an accuracy of 82% in the tests. This result can be optimized after adjusting the hyperparameters, obtaining an accuracy of 84%. Based on the above, it is concluded that Machine Learning techniques are capable of detecting and diagnosing failures in control valves.

Keywords: Machine Learning. Failures. Control valves. Python.

LISTA DE FIGURAS

Figura 1 – Vaso de separação bifásico utilizado no tratamento primário do petróleo.....	20
Figura 2 – Principais algoritmos de aprendizagem supervisionada.....	24
Figura 3 – Esquema representativo de uma Árvore de Decisão com nós internos, externos e ramos.....	25 26
Figura 4 – Estrutura de uma Floresta Randômica.....	27
Figura 5 – Modelo de um <i>perceptron</i> ou neurônio artificial.....	27
Figura 6 – Rede neural <i>feedforward</i> com duas camadas.....	28
Figura 7 – Matriz de classificação de riscos com as categorias frequência e severidade.....	31
Figura 8 – Legenda da Matriz de classificação de riscos.....	31
Figura 9 – Exemplo de planilha utilizada no HAZOP.....	32
Figura 10 – Comportamento da válvula com agarramento.....	34
Figura 11 – a) Ilustração de um atuador pneumático, b) Ilustração de válvula de controle do tipo globo.....	35
Figura 12 – Módulos de falhas em válvulas de controle no <i>software</i> UniSim® Design.....	37
Figura 13 – Tarefas para a detecção e diagnóstico das falhas.....	43
Figura 14 – Simulação do processo contendo os equipamentos e instrumentos utilizados.....	46
Figura 15 – Gráfico para o dimensionamento do vaso de separação.....	48
Figura 16 – Curva característica da bomba B-001.....	49
Figura 17 – Conjunto de dados obtidos na simulação do processo.....	51
Figura 18 – Estatísticas descritivas do conjunto de dados.....	51
Figura 19 – Matriz de correlação das variáveis.....	52
Figura 20 – Gráfico do <i>setpoint</i> do controlador LIC-SP.....	53
Figura 21 – Resposta da variável LIC-PV frente a alterações LIC-SP.....	54
Figura 22 – Resposta da abertura das válvulas frente a alterações no <i>setpoint</i> do nível do vaso.....	55
Figura 23 – Resposta da variável PIC-PV frente a alterações no <i>setpoint</i> do nível do vaso.....	56

Figura 24 – Resposta da variável FIC-PV frente a alterações no <i>setpoint</i> do nível do vaso.....	56
Figura 25 – Distribuição dos dados na série temporal.....	57
Figura 26 – Distribuição dos dados com <i>features</i>	58
Figura 27 – Distribuição do conjunto de dados.....	59
Figura 28 – Importância das variáveis na classificação dos modelos.....	61
Figura 29 – Classificação da Floresta Randômica para a série temporal.....	62
Figura 30 – Classificação da Floresta Randômica para as <i>features</i>	63
Figura 31 – Classificação da Floresta Randômica com o ajuste dos hiperparâmetros.....	65
Figura 32 – Detecção e Diagnóstico das falhas com o modelo de Floresta Randômica.....	66

LISTA DE QUADROS

Quadro 1 – Planilha utilizada na metodologia HAZOP.....	42
---	----

LISTA DE TABELAS

Tabela 1 – Parâmetros de processo e palavras-guia mais utilizadas.....	30
Tabela 2 – Componentes utilizados na simulação do processo.....	38
Tabela 3 – Composição da corrente de alimentação do processo simulado.....	39
Tabela 4 – Valores dos parâmetros x e y.....	40
Tabela 5 – Valores propostos para a razão L/D.....	40
Tabela 6 – Equipamentos e instrumentos utilizados na simulação.....	46
Tabela 7 – Parâmetros da sintonia dos controladores utilizados no sistema.....	47
Tabela 8 – <i>Setpoint</i> dos controladores.....	48
Tabela 9 – Dimensionamento do vaso horizontal de separação V-001.....	49
Tabela 10 – Dimensionamento das Válvulas de controle do tipo globo.....	49
Tabela 11 – Resultados da técnica de Árvores de Decisão para a Detecção e Diagnóstico das falhas.....	59
Tabela 12 – Resultados da técnica de Redes Neurais Artificiais para a Detecção e Diagnóstico das falhas.....	60
Tabela 13 – Resultados da técnica de Floresta Randômica para a Detecção e Diagnóstico das falhas.....	61
Tabela 14 – Hiperparâmetros utilizados nas Florestas Randômicas.....	64
Tabela 15 – Ajuste dos Hiperparâmetros utilizados nas Florestas Randômicas...	64
Tabela 16 – Resultado da classificação da Floresta Randômica com os Hiperparâmetros ajustados.....	65

LISTA DE ABREVIATURAS E SIGLAS

Csv – *Comma-separated values*

Cv – Coeficiente de vazão

DDF – Detecção e Diagnóstico de Falhas

HAZOP – *Hazards and Operability Analysis*

IoT – *Internet of Things*

PID – Proporcional Integral Derivativo

P&ID – *Piping and Instrumentation Diagram/Drawing*

PSF – *Python Software Foundation*

RNA – Rede Neural Artificial

TI – Tecnologia da Informação

TCC – Trabalho de Conclusão de Curso

UNIPAMPA – Universidade Federal do Pampa

LISTA DE SÍMBOLOS

A_v – Área transversal do vaso

A_t – Área do vaso

D – Diâmetro do vaso

D_l – Diâmetro do vaso que contém a parte líquida

D_v – Diâmetro do vaso que contém a parte vapor

H – Altura do vaso

Q_l – Vazão de líquido

Q_v – Vazão de vapor

t_r – tempo de residência

V_p – Velocidade de projeto

x_i – Entradas dos neurônios

y_1 – Saídas dos Neurônios

w_i – Pesos

Σ – Somatório

ϕ – Função de ativação

α – Valor verdadeiro da amostra

β – valor previsto

n_{exemplos} – número total de amostras

SUMÁRIO

1 INTRODUÇÃO.....	17
2 OBJETIVOS.....	19
2.1 Objetivo geral.....	19
2.2 Objetivos específicos.....	19
3 REVISÃO BIBLIOGRÁFICA.....	20
3.1 Vasos de separação.....	20
3.2 Indústria 4.0.....	21
3.3 <i>Machine Learning</i>	22
3.3.1 Aprendizado supervisionado.....	23
3.3.1.1 Árvores de Decisão.....	24
3.3.1.2 <i>Florestas Randômicas</i>	25
3.3.1.3 Redes Neurais Artificiais.....	27
3.4 Python.....	29
3.5 HAZOP.....	29
3.6 Detecção e Diagnóstico de Falhas em processos químicos.....	32
3.6.1 Falhas em válvulas de controle.....	33
3.6.1.1 Agarramento.....	33
3.6.1.2 Supressão de pressão.....	34
3.7 Simulação e modelagem de processos químicos.....	35
3.7.1 Simulador UniSim® Design.....	36
4 MATERIAIS E MÉTODOS.....	38
4.1 Metodologia para a simulação do processo.....	38
4.2 Metodologia para dimensionamento dos equipamentos.....	39
4.2.1 Dimensionamento do vaso de separação.....	39
4.2.2 Dimensionamento das válvulas de controle.....	41
4.2.3 Dimensionamento da bomba.....	41
4.3 Metodologia para realização do HAZOP.....	41
4.4 Metodologia para a coleta dos dados.....	42
4.5 Metodologia para a análise exploratória de dados.....	43
4.6 Metodologia para Detecção e Diagnóstico das Falhas.....	43
4.6.1 Aplicação das <i>features</i>	44

4.6.2 Marcação dos dados.....	44
4.6.3 Separação dos dados.....	44
4.6.4 Escalonamento dos dados.....	45
4.6.5 Treinamento dos algoritmos.....	45
4.6.5.1 Avaliação dos algoritmos de <i>Machine Learning</i>	45
4.6.6 Ajuste dos hiperparâmetros.....	45
5 RESULTADOS E DISCUSSÃO.....	46
5.1 Simulação do processo químico.....	46
5.2 Dimensionamento dos equipamentos.....	48
5.3 Metodologia HAZOP.....	50
5.4 Análise exploratória dos dados obtidos.....	50
5.5 Detecção e Diagnóstico das falhas.....	57
5.5.1 Pré-processamento dos dados.....	57
5.5.2 Avaliação das técnicas de <i>Machine Learning</i>	59
5.5.2.1 Árvores de decisão.....	59
5.5.2.2 Redes Neurais Artificiais.....	60
5.5.2.3 Floresta Randômica.....	60
5.5.2.4 Detecção e Diagnóstico das falhas com a Floresta Randômica.....	61
5.5.2.4.1 Ajuste dos hiperparâmetros da Floresta Randômica.....	63
6 CONSIDERAÇÕES FINAIS.....	67
REFERÊNCIAS.....	69
ANEXOS.....	75
ANEXO A.....	75

1 INTRODUÇÃO

A indústria química está em constante evolução, com produção diária de dados a uma taxa sem precedentes, incluindo dados históricos do processo químico e dados contextuais de sistemas de negócios e manufatura. As plantas industriais modernas geram inúmeros dados a partir dos sensores aplicados no processo e este volume se expande a cada dia exponencialmente. Surge assim, uma nova era industrial que utiliza de sistemas ciberfísicos, Internet das coisas e armazenamento em nuvem. A indústria 4.0 baseia-se na tendência atual de aumento na comunicação entre os diversos setores e equipamentos da indústria e busca explorar as vantagens de um mundo cada vez mais conectado. A crescente capacidade de processamento e instrumentação dos processos químicos permite que grandes quantidades de variáveis e dados monitorados sejam armazenadas (SILVA, 2020).

Surge, dessa forma, uma demanda por ferramentas que permitam trabalhar com grandes volumes de dados e que sejam capazes de extrair informações do comportamento do processo, fornecendo soluções para otimização da produção. Assim como, metodologias que permitam aumentar a segurança e monitoramento dos processos químicos. Para processar esses dados e alcançar um nível mais amplo de compreensão das plantas industriais, são utilizadas técnicas como estatística multivariada e Machine Learning (SOARES, 2017).

Machine Learning é uma abordagem baseada em dados, que contrasta com a abordagem tradicional baseada em modelos. Esta técnica fundamenta-se na detecção automática de padrões relevantes dentro de um conjunto de dados. As ferramentas computacionais são utilizadas para suprir a complexidade do desenvolvimento de modelos fenomenológicos e criar modelos empíricos que permitem prever o comportamento de determinado processo químico (BAUGHMAN & LIU, 1995). A aplicação dos algoritmos de *Machine Learning* na indústria química aumentou nos últimos anos, principalmente na área de Detecção e Diagnóstico de falhas (VENKATASUBRAMANIAN, 2008).

A detecção e diagnóstico de falhas baseada em dados históricos explora as informações obtidas através do extenso monitoramento de equipamentos, variáveis e produtos a fim de encontrar sinais de desvios nos padrões do processo. O monitoramento eficiente das plantas industriais proporciona uma maior segurança na

produção, além de melhorar a qualidade dos produtos e facilitar a otimização do processo (CHIANG *et al.*, 2001). Falhas em válvulas de controle e atuadores normalmente resultam em parada da produção gerando custos e perdas não programados, assim como possíveis riscos a segurança dos trabalhadores (MOUTINHO, 2016). Dessa forma, ter meios para monitorar e controlar estas anomalias tornam as indústrias economicamente eficientes e ajudam a evitar prejuízos.

Baseado no exposto, o objetivo deste trabalho é avaliar diferentes algoritmos de *Machine Learning* na Detecção e Diagnóstico de Falhas em válvulas de controle. Realizando a simulação de um processo químico no software UniSim® Design e a análise e processamento dos dados obtidos na linguagem Python. Será realizado o dimensionamento do vaso de separação, válvulas de controle, instrumentos e bomba presentes no sistema. E a aplicação da metodologia HAZOP, para avaliação e identificação dos riscos do processo.

Este Trabalho de Conclusão de Curso é dividido em seis partes. Primeiro têm-se a introdução, em que é exposto em linhas gerais o tema, objetivos e motivações que levaram ao estudo. O objetivo geral e objetivos específicos são delimitados no segundo capítulo. O terceiro capítulo apresenta a revisão bibliográfica que embasou a escrita do trabalho. O quarto capítulo traz os materiais e métodos que foram utilizados para o desenvolvimento das técnicas de *Machine Learning* e para a simulação do processo. No quinto capítulo é apresentado e discutido os resultados obtidos no estudo. E o sexto capítulo apresenta as considerações finais e sugestões para trabalhos futuros.

2 OBJETIVOS

Neste capítulo serão apresentados o objetivo geral e os objetivos específicos deste trabalho.

2.1 Objetivo Geral

Avaliar métodos de *Machine Learning* na detecção e diagnóstico de falhas em válvulas de controle em um processo químico simulado computacionalmente.

2.2 Objetivos Específicos

- Realizar a simulação do processo químico escolhido no software UniSim® Design;
- Realizar o dimensionamento dos equipamentos;
- Realizar a metodologia HAZOP no sistema estudado;
- Determinar através do HAZOP qual válvula de controle será analisada;
- Obter dados do processo em operação normal e com falhas na válvula de controle;
- Realizar a análise exploratória dos dados obtidos em Python;
- Avaliar qual a melhor técnica de *Machine Learning*, entre os métodos de classificação selecionados, para realizar a detecção e o diagnóstico das falhas.

3 REVISÃO BIBLIOGRÁFICA

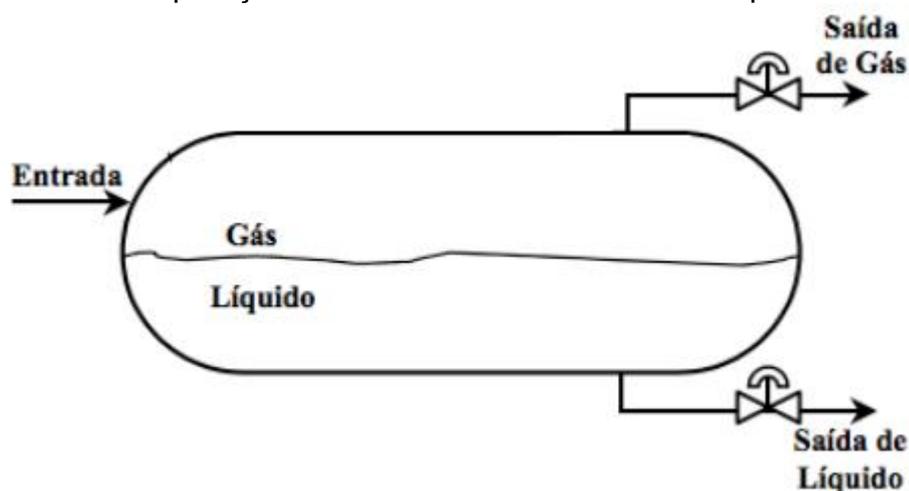
Este capítulo apresenta uma breve revisão bibliográfica sobre vasos de separação, a Indústria 4.0, detecção e diagnóstico de falhas e as principais técnicas de *Machine Learning*. São apresentados também os conceitos da linguagem Python e suas bibliotecas, simulação de processos químicos e a metodologia HAZOP.

3.1 Vasos de separação

Os vasos de separação são equipamentos amplamente utilizados na indústria química, como por exemplo, em usinas de açúcar, etanol e usinas petroquímicas. Estes vasos são fabricados a fim de resistir a pressões internas e externas e também a agentes externos que podem danificá-los. Na indústria petroquímica os vasos de separação são utilizados no tratamento primário do petróleo para reduzir o teor de vapor de água e outros contaminantes presentes na mistura.

A Figura 1 apresenta um vaso de separação bifásico utilizado na separação do gás livre da água e/ou óleo. Nesta representação a mistura líquido/gás entra no vaso, onde, através da força da gravidade ocorre a separação das gotículas mais pesadas da mistura que se acumulam no fundo do vaso, onde o líquido é coletado. Enquanto o líquido acumula no fundo do vaso, o gás é liberado da mistura e vai para a parte superior do separador. (ANJOS, 2015).

Figura 1 – Vaso de separação bifásico utilizado no tratamento primário do petróleo



Fonte: Adaptado de Petrobras (2007)

Os separadores horizontais são mais comumente utilizados em campos que demandam grandes volumes totais de líquidos. Estes possuem uma maior área superficial para o contato entre os fluidos o que facilita a separação das fases. Este fator também facilita a decantação das gotas de óleo presentes na corrente de gás e a separação de bolhas de gás provenientes de espuma. Porém, esta conformação exige disponibilidade de um espaço maior para sua instalação (ANJOS, 2015).

3.2 Indústria 4.0

A industrialização ao longo dos anos passou por diversos períodos que foram transformadores para a sociedade e para o setor fabril. A primeira revolução industrial trouxe inovações no modo como os produtos eram fabricados, enquanto a segunda revolução industrial difundiu a utilização de energia elétrica. A terceira revolução industrial, caracterizou-se pela automação dos processos e pela introdução da tecnologia da informação (TI) na linha de produção. Por fim, a quarta revolução industrial ou indústria 4.0, surge para aprimorar os avanços alcançados e otimizar o processo de manufatura e negócio. Tornando-se uma nova era industrial baseada no conceito de *smart factories* com elevada automação, eficiência e integração (KAGERMANN; WAHLSTER; HELBIG, 2013).

Atualmente, o setor industrial busca dinamizar a sua produção adaptando-se à crescente competitividade do mercado e às novas necessidades dos clientes que demandam produtos personalizados de alta qualidade e com um tempo de produção reduzido. Para isso, é necessário unir o mundo real e o virtual, impulsionando uma transformação digital representada por uma mudança na forma como as empresas gerenciam a cadeia de valor. Dessa forma, um novo modelo de indústria surge com o objetivo de transformar o ramo industrial através da digitalização e da exploração do potencial das novas tecnologias que emergem, denominando-se Indústria 4.0 (ROJKO, 2017).

Fazendo uso de sistemas *ciber* físicos, internet das coisas (IoT) e *Cloud Computing* a indústria 4.0 torna-se hiperconectada e capaz de extrair e armazenar uma grande quantidade de dados através de diversos sensores. Estes dados, atualmente, são um dos ativos mais importantes para qualquer negócio, independente do porte ou setor da indústria. Assim, o *Big Data* torna-se uma ferramenta indispensável para esse novo modelo de indústria, visto que através

deste recurso é possível tratar os dados obtidos e aproveitá-los para gerar oportunidades de negócio (CUADRADO, 2019).

A Indústria 4.0 engloba tecnologias capazes de interagir com o ambiente em que se encontram e resolver os possíveis problemas que surgem, buscando soluções e tomando decisões de forma autônoma. A aplicação de algoritmos de *Machine Learning*, ou aprendizagem de máquinas, é necessária para compreender a complexidade e os potenciais dessa nova revolução industrial, assim como criar modelos que valorizem os dados e facilitem a tomada de decisão. As aplicações e pesquisas voltadas à área de *Machine Learning* são cada vez mais numerosas e desafiadoras, uma vez que o poder computacional aumenta e os hardwares tornam-se mais acessíveis (CUADRADO, 2019).

3.3 *Machine Learning*

Machine Learning ou aprendizagem de máquina é uma subárea da inteligência artificial, que se baseia na detecção autônoma de padrões relevantes em um determinado conjunto de dados e na generalização de informações. Para isso, são utilizados algoritmos estatísticos que são treinados a partir de um modelo responsável por realizar previsões sem o conhecimento das leis físico-químicas ajustando-se aos dados da melhor forma que os represente (SEGARAN, 2007).

Nos últimos anos, *Machine Learning* tornou-se uma ferramenta muito comum devido ao aumento exponencial da área do *Big Data* relacionada à interpretação e análise de grandes volumes de dados demandados na indústria 4.0. Esta técnica vem sendo utilizada para solucionar problemas tradicionais de estatística multivariada como clusterização, regressão e redução de dimensionalidade (MONTAVON; ORR; MÜLLER, 2012; SHALEV-SHWARTZ; BEN-DAVID, 2014).

É necessário garantir dados variados em grande quantidade e com qualidade para se alcançar uma elevada eficiência nos algoritmos de *Machine Learning*. Visto que, todo o conhecimento do algoritmo provém dos dados utilizados no treinamento do mesmo, se faz necessário muitos exemplos para a relação dos dados ser aprendida com o menor erro possível. Assim, torna-se comum descartar dados redundantes que são comuns e facilmente classificáveis e duplicar os dados raros que são difíceis de serem perfeitamente classificados (HASTIE *et al.*, 2008).

Para o treinamento do algoritmo de *Machine Learning*, a base de dados é separada em dois grupos: dados de treino e dados de teste. Através dos dados de treino é realizada a criação e treinamento do modelo utilizado no algoritmo. Os dados de teste são utilizados para simular previsões reais que o modelo apresentará. Dentre as formas mais comuns de realizar a validação têm-se as amostragens randômicas e k-fold (HASTIE *et al.*, 2008).

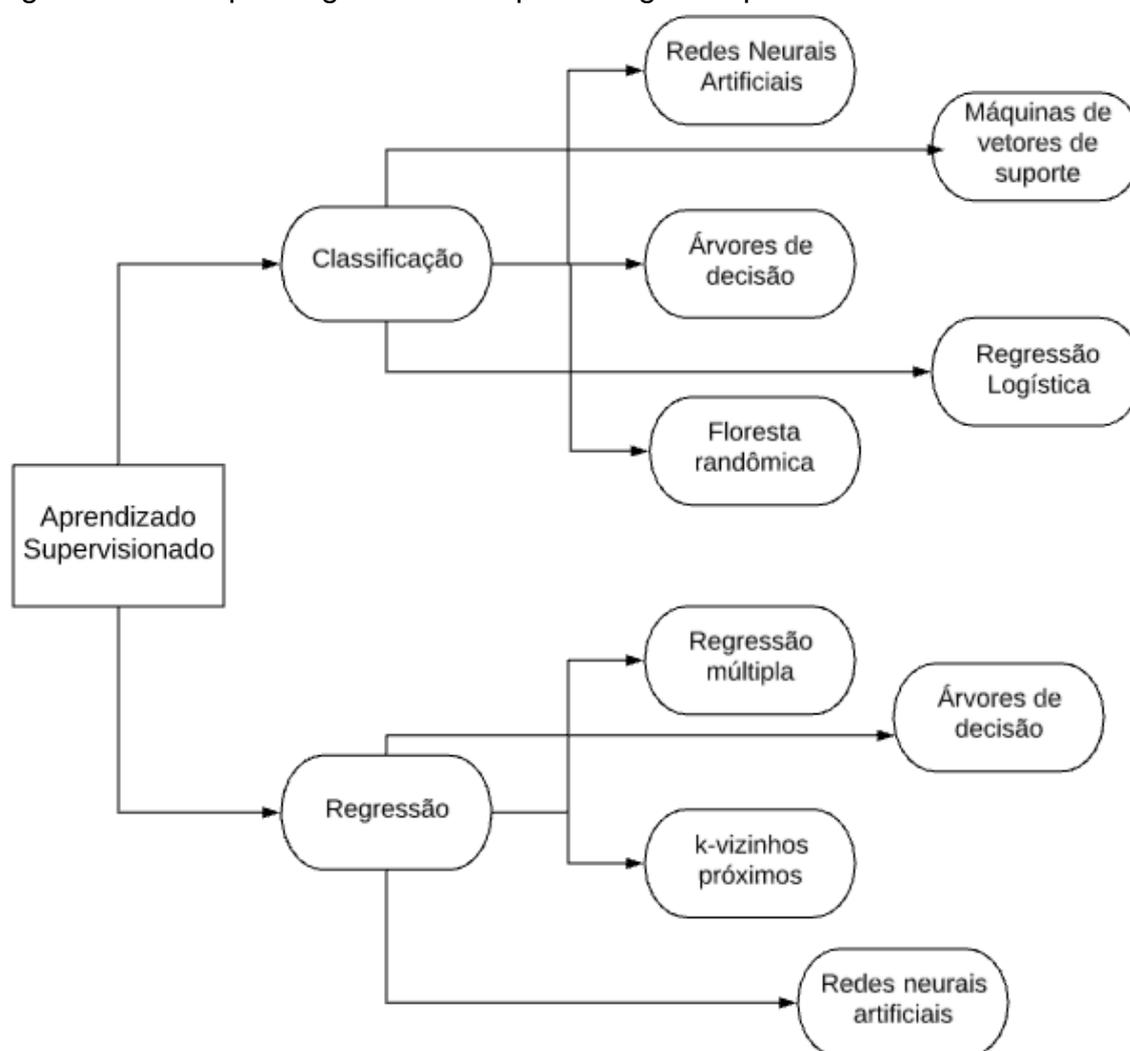
As técnicas de *Machine Learning* apresentam muitas vantagens quando comparadas a métodos estatísticos tradicionais, uma vez que os algoritmos descartam o uso de suposições sobre a estrutura do espaço de estados, ideias erradas ou modelagem imprecisa do processo. Eles ainda são capazes de detectar as não linearidades do processo e padrões que abordagens tradicionais não são capazes de descobrir. Além disso, não são necessárias as análises das dinâmicas internas do processo ou estimação de parâmetros das equações, que geralmente são características desconhecidas no procedimento (SOARES, 2017).

Os métodos de *Machine Learning* são divididos em três classes de aprendizado: aprendizado não supervisionado, aprendizado supervisionado e aprendizagem por reforço.

3.3.1 Aprendizado supervisionado

Os algoritmos de aprendizado supervisionado são os mais utilizados no mundo. Este modelo relaciona, a partir dos dados fornecidos pelo programador, uma entrada com uma saída. Não é necessário uma formalização matemática rígida ou conhecimentos profundos de estatística. O aprendizado supervisionado se divide em duas classes: problemas de regressão, em que a saída é um valor numérico e de classificação, onde a saída é um booleano (SOARES, 2017). A Figura 2 apresenta um esquema com os principais algoritmos de aprendizagem supervisionada.

Figura 2 – Principais algoritmos de aprendizagem supervisionada



Fonte: Adaptado de Beck *et al.* (2016)

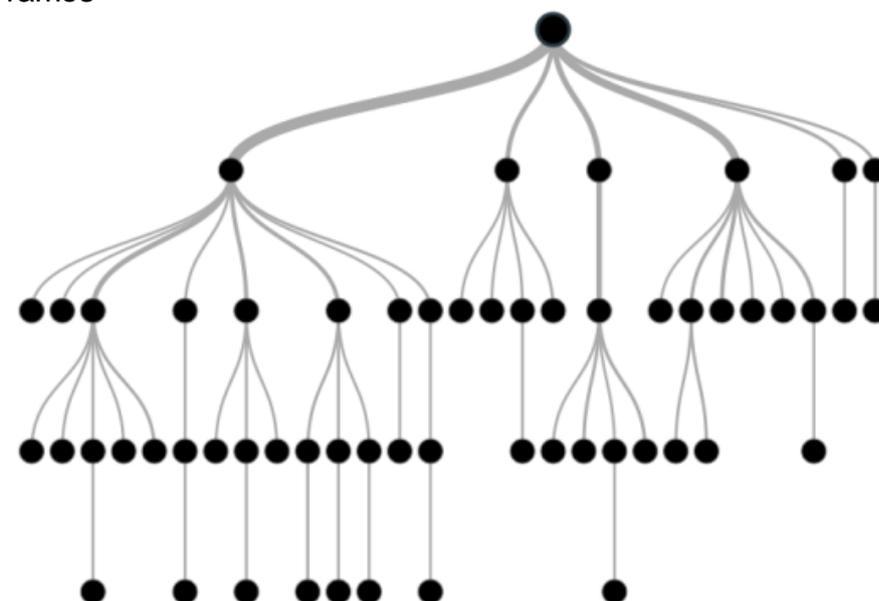
3.3.1.1 Árvores de Decisão

Árvores de Decisão podem trabalhar com dados categóricos e numéricos, o que garante ampla utilização dessa classe de algoritmo em *Machine Learning* e análise multivariada. Elas se assemelham a fluxogramas, onde o caminho de informação é unidirecional e no formato de uma árvore. O mecanismo da árvore de decisão é simples sendo possível seguir a estrutura da mesma para entender facilmente como a decisão é feita (COSTA, 2006).

Sua estrutura, como pode ser observada na Figura 3, consiste em nós internos e externos que são unidos por ramos. O nó interno é uma unidade de processamento que após realizar testes sobre um determinado atributo toma a decisão que determina o próximo nó a ser visitado. E os nós externos, também

chamados de folha ou nó terminal, são os possíveis resultados do teste. Este teste, em problemas de classificação, pode ser o próprio atributo no caso de dados categóricos ou alguma expressão booleana para dados numéricos (BURROWS, 1997).

Figura 3 – Esquema representativo de uma Árvore de Decisão com nós internos, externos e ramos



Fonte: Analytics Vidhya (2016)

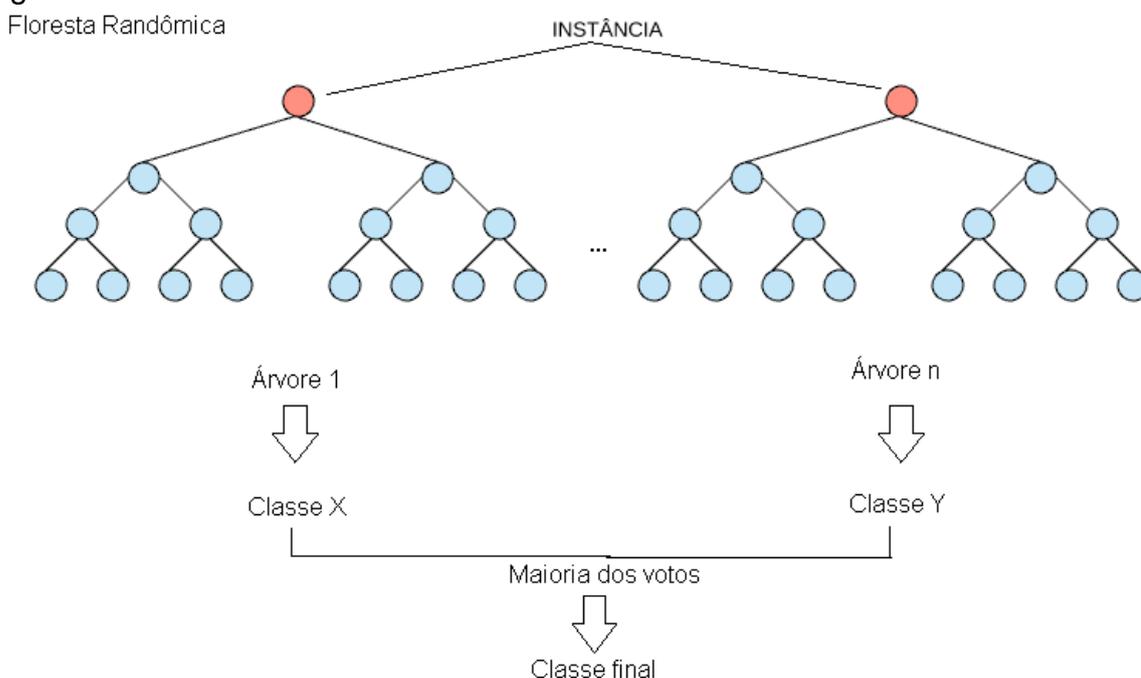
No processo de treinamento, os dados são apresentados ao nó inicial (nó raiz) da árvore de decisão, onde o algoritmo escolherá o teste a ser realizado sobre o atributo. O melhor teste que será atribuído ao nó é calculado por uma métrica que indicará a qualidade com que cada atributo separa um conjunto de dados de acordo com as suas respectivas classes. Desse nó, surgiram todas as ramificações possíveis desse teste sendo que a cada nó uma nova métrica será calculada. Este processo é repetido até que o nó terminal seja obtido, e a legenda ou valor atribuído sejam satisfatórios para os dados inseridos na árvore (COSTA, 2006).

3.3.1.2 Florestas Randômicas

Florestas Randômicas ou *Random Forest* é um algoritmo de classificação ou regressão, que consiste em treinar um grupo de árvores de decisão geradas dentro de um mesmo conjunto. Cada árvore de decisão corresponde a um domínio diferente dos dados de entrada, como apresentado na Figura 4. Esta técnica

considera cada domínio individualmente, escolhendo aquele que obter maior aproximação da previsão selecionada. A construção das árvores ocorre na etapa de treinamento onde são geradas classes (classificação) ou uma previsão média dos valores (regressão) em cada uma (ANALITICS VIDHYA, 2016).

Figura 4 – Estrutura de uma Floresta Randômica



Fonte: Adaptado de Analytics Vidhya (2020)

A baixa correlação entre as árvores de decisão que compõem a floresta randômica trazem à esta técnica um melhor desempenho do que os modelos individuais. Os modelos não correlacionados podem produzir previsões mais precisas, visto que as próprias árvores protegem umas às outras de seus erros individuais. Dessa forma, o fluxo de escolha na ramificação da floresta randômica será na direção dos modelos com as melhores previsões. A floresta randômica permite que cada árvore individualmente realize uma amostra aleatória dos dados de entrada, resultando em árvores totalmente diferentes, e que apenas um subconjunto destes dados seja escolhido para cada membro. Gerando uma maior diversificação das árvores e uma menor correlação entre elas (BREIMAN, 2001).

Este modelo é considerado um método de redução de dimensionalidade, identificando as variáveis mais significativas dentro de uma amostra. A floresta randômica é capaz de estimar dados perdidos e mantém a precisão de resultados mesmo quando há dados faltantes. Possui também métodos para equilibrar erros em

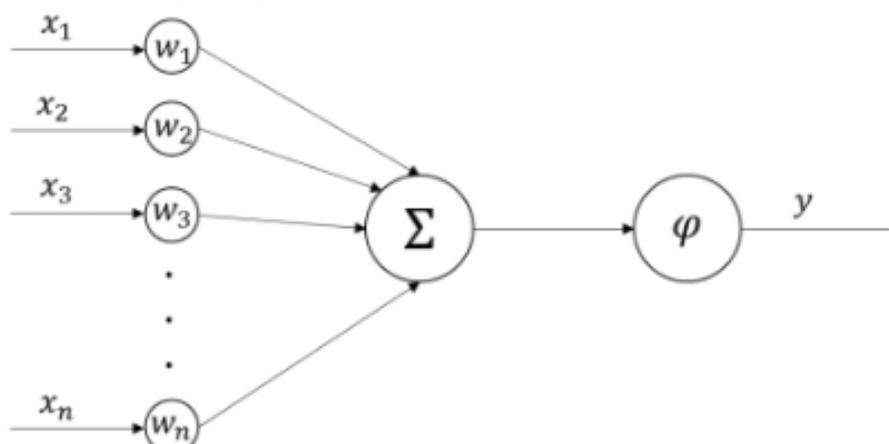
determinados conjuntos de dados onde as classes estão desequilibradas. Assim como, para dados não classificados, gerando um agrupamento não supervisionado com visualização de dados e detecção de *outliers* (ANALITICS VIDHYA, 2016).

3.3.1.3 Redes Neurais Artificiais

As redes Neurais Artificiais (RNAs) são uma das técnicas mais utilizadas de *Machine Learning*, e se baseiam em algoritmos que representam o funcionamento do cérebro humano usando vários neurônios. Estes neurônios ou *perceptrons* são conectados entre si e realizam uma soma ponderada de várias entradas. É realizado um ajuste de parâmetros, comparando o resultado das entradas com um limite interno de ativação, chamado de pesos ou *bias*. Após essa comparação, os resultados passam por uma função de ativação que relaciona as entradas e saídas dos dados fornecidos realizando o ajuste das *bias* (HAYKIN, 2008).

A Figura 5 apresenta o modelo de um neurônio em que as entradas são representadas pelos x_i , as saídas pelo y e os pesos pelo w_i . O Σ é o elemento somatório e φ representa a função de ativação que é responsável por definir o nível de intensidade da resposta.

Figura 5 – Modelo de um *perceptron* ou neurônio artificial



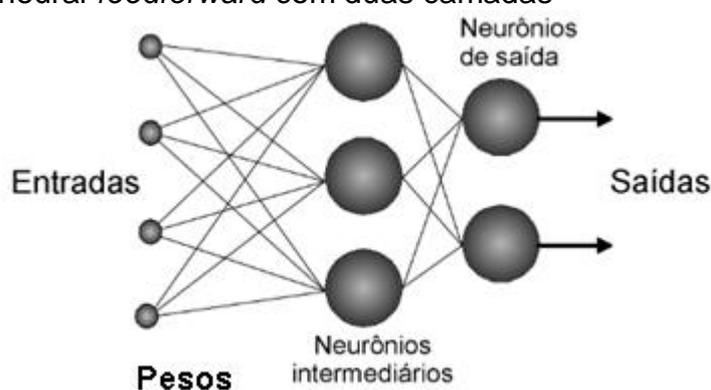
Fonte: Silva (2020, p. 14)

Geralmente as RNAs consistem em diversas camadas de neurônios arranjadas uma em frente à outra. Os dados entram na primeira camada, conhecida como camada de entrada, que usualmente contém neurônios com a função de ativação identidade. Após a entrada, os dados vão ativando camada por camada até chegarem à última camada. Esta, chamada de camada de saída, normalmente

possui a função de ativação, que pode ser linear, limiar, rampa ou sigmoide (BAUGHMAN; LIU, 1995).

A arquitetura de uma rede neural é constituída pelo número de entradas e saídas, número de camadas, número de neurônios por camada, a forma como os pesos estão associados e quais neurônios serão corrigidos. Existem três categorias básicas de arquitetura para as RNAs: *feedforward*, *feedback* e recorrentes (ZANINI, 2004). A Figura 6 apresenta um exemplo de rede *feedforward* com duas camadas de neurônios.

Figura 6 – Rede neural *feedforward* com duas camadas



Fonte: Adaptado de Tafner (1998)

As RNAs são consideradas aproximações universais de funções, devido a sua não linearidade, adaptabilidade e por gerar mapeamento entrada-saída, o que as torna devidamente adequadas para modelagem de sistemas dinâmicos (SARAIVA *et al.*, 2020). Porém, as RNAs impõem muitos desafios, pois dependem da quantidade de dados utilizados no treinamento sem garantia de uma boa generalização, ou de alcançar a melhor rede possível (BAUGHMAN; LIU, 1995).

A exploração destas técnicas para soluções de problemas voltados à indústria química já vem sendo aplicada há alguns anos. Como por exemplo, Soares (2017) que utiliza um modelo de RNA para detecção e diagnóstico de falhas em processos químicos. Attolini (2014) estudou e propôs uma metodologia para detectar avarias em válvulas de controle, partindo de simulação computacional. E Dambros (2016) que utiliza *Machine Learning* para detecção de agarramento em válvulas de controle para sinais de referência variável e para casos com presença de ruído.

3.4 Python

A linguagem Python foi desenvolvida em 1991 pelo holandês Guido Von Rossum. É classificada como uma linguagem de programação de propósito geral, possuindo uma sintaxe elegante e tipagem dinâmica e forte, interpretada e interativa. Python é um software de código aberto e a especificação da linguagem é mantida pela Python Software Foundation (PSF). Atualmente, é uma das linguagens mais utilizadas e populares do mundo principalmente em aplicações envolvendo análise de dados, *Machine Learning*, inteligência artificial, desenvolvimento de jogos e pesquisas acadêmicas. O destaque na área de *Machine Learning* é devido ao desenvolvimento crescente de novas bibliotecas voltadas para aplicação dessa técnica (PYTHON, 2021).

Numpy é uma biblioteca que faz o processamento de matrizes e vetores, auxiliando na organização de um enorme fluxo de dados, na sua manipulação e elaborações estatísticas. Matplotlib e Seaborn são bibliotecas abrangentes para a criação de gráficos estatísticos, visualizações animadas e interativas em Python. Pandas é uma biblioteca utilizada para tratamento de bancos de dados, oferecendo algoritmos para operar e manipular tabelas numéricas e séries temporais. A Scikit-learn é uma biblioteca desenvolvida especificamente para aplicação prática de *Machine Learning*, ela oferece ferramentas simples e eficientes para análise preditiva dos dados (PYTHON.ORG, 2021).

3.5 HAZOP

Dentre as metodologias de análise de riscos existentes o *Hazards and Operability Analysis* (HAZOP) é o método mais completo e rigoroso, sendo o favorito das empresas. O objetivo do HAZOP é identificar os potenciais riscos nas plantas industriais e documentá-los. Este método baseia-se em estudar o processo, a operação, a disposição dos equipamentos e os fatores externos e humanos de forma metódica e sistemática, evidenciando as prováveis situações de risco. O HAZOP pode ser aplicado em qualquer tipo de instalação, sejam novas ou já em operação, e sua duração depende da complexidade da planta analisada (FREEDMAN, 2003).

A qualidade da metodologia HAZOP está intrinsicamente ligada à qualidade e quantidade de informações utilizadas para o estudo. Dentre as documentações necessárias para realizar o método têm-se: os diagramas P&ID's, diagramas de processo, descrição do processo, memoriais descritivos entre outros. O HAZOP é um trabalho em equipe, desenvolvido por um grupo multidisciplinar que através de suas experiências podem criar discussões e chegar à identificação dos riscos. Após a identificação, a equipe responsável deve determinar a probabilidade e gravidade da ocorrência e recomendar propostas para mitigar ou eliminar as situações perigosas (FREEDMAN, 2003).

O procedimento de estudo HAZOP através de uma descrição geral do processo cria questionamentos sobre cada parte deste, para detectar possíveis desvios de operação e decidir quais destes desvios podem constituir perigo. Este questionamento abrange toda a instalação e operação, sendo submetidos a certo número de questões desenvolvidas a partir de “palavras-guia”. Estas “palavras-guia” são utilizadas para garantir que as questões explorem todas as formas possíveis às quais o processo possa desviar das intenções de projeto. O objetivo de aplicar estas “palavras-guia” aos parâmetros do processo é gerar um número de desvios teóricos para descobrir as consequências e motivos de possíveis falhas na operação (KLETZ, 1999). A Tabela 1 apresenta os parâmetros de processo e palavras-guia mais utilizadas no procedimento de realização do HAZOP.

Tabela 1 – Parâmetros de processo e palavras-guia mais utilizadas

Parâmetros do processo	Palavras-guia
Fluxo	Menos/mais
Pressão	Menos/mais
Temperatura	Nenhum (a)
Agitação	Reverso/Oposto
Nível	Menos/mais

Fonte: Autora (2021)

O HAZOP utiliza como ferramenta a matriz de classificação de riscos para determinar o nível dos riscos e prioridades para programar recomendações que surgirem no estudo (FREEDMAN, 2003). Na Figura 7, tem-se um exemplo desta matriz, que relaciona as categorias severidade, que indica o grau de gravidade dos riscos, e a frequência com que estes riscos ocorrem. A Figura 8 apresenta a legenda da Matriz de classificação de riscos apresentada anteriormente.

Figura 7 – Matriz de classificação de riscos com as categorias frequência e severidade

		Frequência				
		A	B	C	D	E
SEVERIDADE	IV	2	3	4	5	5
	III	1	2	3	4	5
	II	1	1	2	3	4
	I	1	1	1	2	3

Fonte: Adaptado de Aguiar (2008)

Figura 8 – Legenda da Matriz de classificação de riscos

Severidade		Frequência		Risco	
I	Desprezível	A	Extremamente Remota	1	Desprezível
II	Marginal	B	Remota	2	Menor
III	Crítica	C	Improvável	3	Moderado
IV	catastrófica	D	Provável	4	Sério
		E	Frequente	5	Crítico

Fonte: Adaptado de Aguiar (2008)

Para a realização do HAZOP são utilizadas planilhas, como a apresentada na Figura 9, que podem variar de acordo com o planejamento da equipe que irá realizar a técnica. O cabeçalho desta planilha deve indicar o sistema que esta sendo analisado, unidade, localização do nó escolhido e a equipe responsável (AGUIAR, 2008).

Figura 9 – Exemplo de planilha utilizada no HAZOP

Análise de Perigos e Operabilidade					
Unidade					
Sistema:			Equipe:		Data:
Localização do Nó:					Página:
Item	Desvio	Causas	Consequências	Salvaguardas	Observações

Fonte: Adaptado de Aguiar (2008)

Os métodos de análise de risco, como o HAZOP, são uma excelente ferramenta para aumentar significativamente o padrão de segurança das indústrias. Assim como, para melhorar as condições de operabilidade da planta ao apresentar *insights* para melhorias dos controles do processo. O HAZOP possibilita a identificação dos possíveis cenários de risco no processo, aumentando as oportunidades de identificar antecipadamente os panoramas de falhas e organização de planos de contingência (KLETZ, 1999).

3.6 Detecção e diagnóstico de falhas em processos químicos

Falha de processo é a indicação de mudanças físicas no processo em relação à sua faixa normal de operação. A causa dessas falhas podem ser desvios no nível do fluido, no fluxo, temperatura e pressão ou também se referem às próprias mudanças físicas ocorridas no processo industrial, como vazamentos, entupimentos, incrustações, corrosão e desgastes. Os tipos de falhas mais comuns na indústria química são: escalonamento de trocadores de calor, envenenamento de catalisador e vazamentos (HIMMELBLAU, 1983 *apud* TEIXEIRA, 2000).

O procedimento de Detecção e Diagnóstico de Falha (DDF) é essencial dentro da indústria química de uma forma geral. A partir do bom controle de uma planta industrial é possível garantir a segurança e perfeito funcionamento da mesma. Esse processo se baseia em saber o momento em que ocorre uma falha na planta e qual falha ocorreu respectivamente. Visto que, os defeitos podem ocorrer em qualquer ponto da planta industrial, trazendo prejuízos que vão desde diminuir a eficiência do processo até causar a perda de vidas humanas (TEIXEIRA, 2000).

A DDF mostra-se cada vez mais relevante devido às mudanças que tornaram mais rigorosas a legislação ambiental e trabalhista e também devido ao aumento e

complexidade das plantas industriais. Dessa forma o principal objetivo de aplicar esta técnica é garantir uma operação que satisfaça as especificações de desempenho, incluindo restrições ambientais, de segurança e econômicas. Quando aplicada corretamente, a DDF pode reduzir os custos de operação e o tempo ocioso da planta e ainda melhorar a qualidade dos produtos e segurança dos trabalhadores (CHIANG *et al.*, 2001).

Os métodos de análise estatística multivariada ou modelagem matemática são os mais utilizados para DDF. Porém, nos últimos anos o emprego de algoritmos de *Machine Learning* para essa área aumentou. A DDF é um problema de classificação, e a utilização de algoritmos que recebem entradas numéricas e têm saídas categóricas tornaram-se mais popular e com previsões muito coerentes. Estes modelos mostram-se capazes de aprender padrões inesperados em desafios de âmbito computacional e análise estatística (VENKATASUBRAMANIAN, 2008).

3.6.1 Falhas em válvulas de controle

Processos industriais demandam um número elevado de malhas de controle, onde cada uma tem sua importância na eficiência e qualidade da produção. Seu objetivo é manter automaticamente as variáveis de processo dentro de seus limites de operação. Para isso utiliza-se um sensor que mede essas variáveis e um controlador que determinará a ação de controle para manter essa variável dentro dos limites desejados. Essa ação é direcionada a um atuador ou elemento final de controle, que geralmente é uma válvula de controle. Dessa forma, é necessário manter as válvulas de controle em perfeito funcionamento para garantir um fluxo estável do processo. As falhas em válvulas de controle mais comuns, que serão estudadas neste trabalho, são o agarramento e a supressão de pressão (MOUTINHO, 2016).

3.6.1.1 Agarramento

Agarramento é definido como a resistência a iniciar o movimento, normalmente determinado como a diferença entre o valor necessário para superar o atrito estático e o atrito dinâmico. Esta falha causa uma não linearidade na válvula de controle, causando oscilações em forma periódica, chamados de ciclos limites. A

variabilidade faz com que as variáveis que estão sendo controladas excedam seus limites de operação, eficiência e segurança. O agarramento é a avaria mais comum em válvulas e também a perturbação que mais causa distúrbios à malha de controle. Diversos trabalhos vêm sendo publicados e muito se tem pesquisado sobre como detectar, quantificar e compensar essa falha (CUADROS, 2011; MOUTINHO, 2016). O comportamento da válvula com agarramento pode ser observado nas regiões apresentadas na Figura 10.

Figura 10 – Comportamento da válvula com agarramento



Fonte: Adaptado de Gerchman (2013)

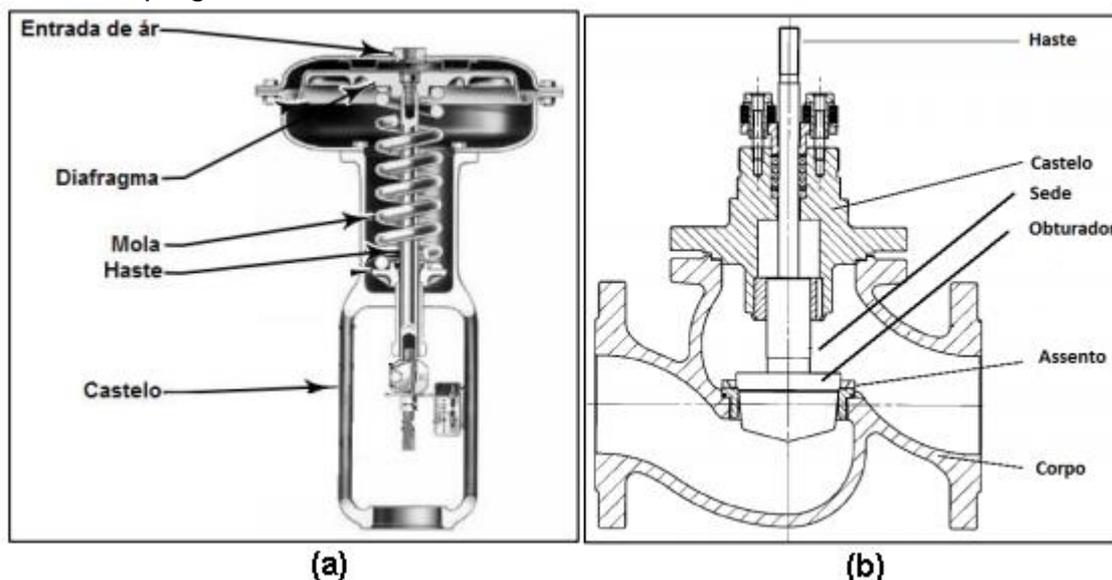
3.6.1.2 Supressão de pressão

A falha de supressão de pressão influencia diretamente o volume de ar que é enviado para o atuador da válvula de controle, representados na Figura 11. Essa falha surge do bloqueio ou limitação das linhas de fornecimento de ar ao diafragma do atuador, Figura 11 a, prejudicando o desempenho do sistema. Dependendo da causa desta falha o grau de avaria é diferente, o que pode resultar em diferentes níveis máximos de abertura para a válvula de controle. Em casos de bloqueio total de ar de instrumento para o atuador, este não poderá se mover por toda sua faixa de operação (MOUTINHO, 2016).

Dependendo da posição de falha da válvula, Falha Fecha ou Falha Abre, esta assumirá uma determinada condição no caso de ocorrência da avaria. Quando houver falta ou suprimento indevido de ar na válvula, e esta for Falha Fecha, o instrumento passará a condições de completamente fechada (0%). Se a posição da

válvula for Falha Abre, o instrumento passa a assumir a condição de completamente aberta (100%). A escolha da ação das válvulas é de fundamental importância para a segurança do processo e para a escolha dos parâmetros do controlador (Mathias, 2008).

Figura 11 – a) Ilustração de um atuador pneumático, b) Ilustração de válvula de controle do tipo globo



Fonte: Adaptado de Fisher (2005)

3.7 Simulação e modelagem de processos químicos

A simulação e modelagem de processos químicos são ferramentas amplamente utilizadas na Engenharia Química, pois permitem a previsão de parâmetros e dados sem que sejam feitas alterações no processo real. Esta abordagem ainda possibilita a redução de custos e o desenvolvimento de projetos industriais. A simulação de processos e modelagem de equipamentos, objetiva simular computacionalmente condições operacionais e realizar testes para encontrar os parâmetros mais adequados para determinado processo químico. Estas ferramentas são muito úteis para o desenvolvimento de novos processos ou até mesmo avaliação de mudanças nos processos em operação (SOUZA; COSTA, 2018).

Os simuladores computacionais possuem uma variedade de ferramentas que permitem aplicar modelos matemáticos complexos que podem representar diversos fenômenos químicos. Possuem uma interface gráfica dinâmica e amigável e uma

extensa biblioteca de equipamentos, pacotes termodinâmicos e compostos químicos. Rotinas matemáticas para solução de sistemas de equações algébricas e equações diferenciais também podem ser obtidas através desses *softwares* (SOUZA; COSTA, 2018).

Os avanços tecnológicos e novas necessidades dos clientes resultam em uma transformação do mercado de simulação e modelagem de processos. Os modelos obtidos nos softwares, que tradicionalmente eram ferramentas apenas de análise e tomada de decisão de *desktop*, passam a ser aproveitados em novas soluções de alto valor como o monitoramento operacional “conectado” (HONEYWELL, 2020).

Entre os simuladores, sejam eles licenciados comercialmente ou elaborados a partir de projetos de pesquisa de universidades de vários países do mundo, os mais utilizados acadêmica e industrialmente são o Aspen HYSYS®, AVEVA PRO-II®, gPROMS®, Petro-SIM® e UniSim® Design (SOUZA; COSTA, 2018).

3.7.1 Simulador UniSim® Design

O UniSim® Design Suite da Honeywell apresenta um ambiente totalmente integrado que possibilita a utilização de vários equipamentos típicos da indústria química e simulação de processos químicos reais em estado dinâmico e estacionário. Este software ajuda engenheiros e pesquisadores na modelagem de processos, no monitoramento do desempenho da planta industrial, na solução de problemas, no planejamento de negócios e gerenciamento de ativos. Ele também proporciona uma visão das variáveis mais importantes do processo, sem que seja necessário preocupar-se com despesas de capital (HONEYWELL, 2020).

O UniSim® Design possui ambiente de modelagem intuitivo com interface gráfica, análise gráfica de resultados, pacotes termodinâmicos e operações unitárias básicas. Ele calcula com elevada precisão propriedades físicas, de transporte, de separação e de cinética de reações químicas. Dessa forma este software torna-se líder de mercado, proporcionando resultados realistas de simulação de processos em um ambiente estável e dinâmico, podendo ser utilizado em modo off-line e online. Este software também apresenta um módulo, apresentado na Figura 12, onde é possível aplicar falhas em válvulas de controle, como por exemplo, agarramento, vazamentos, falhas no atuador, etc. (HONEYWELL, 2020).

Figura 12 – Módulos de falhas em válvulas de controle no *software* UniSim® Design

The screenshot displays the 'Dynamics' configuration window for valve failures in UniSim Design. The interface is organized into several sections:

- Left Navigation Panel:** Lists various simulation modules including 'Dynamics', 'Specs', 'Pipe', 'Holdup', 'Actuator', 'Act. (Station)', 'Limit Switch', 'Flow Limits', 'Limits (Bypass)', 'Malfunction', 'Malfunction (LS)', and 'Strip Chart'. The 'Dynamics' module is currently selected.
- Upset Types:** A list of failure modes with radio buttons for selection:
 - Actuator Failure
 - Positioner Offset
 - Leakage
 - Plugging Factor
 - Stem Inertia
 - Sticky Stem
 - Positioner Hunting
- Use Malfunction:** Use Malfunction
- Use as Plugging Restriction Only:** Use as Plugging Restriction Only
- Description:** A text area for providing details about the upset.
- Status:** A dropdown menu showing 'Valve Stem Status' and 'Normal'.
- Upset Details:** A section for configuring the upset's behavior:
 - Fail Position:** A dropdown menu set to 'Fail To Trip S'.
 - Duration:** A dropdown menu set to 'Hold'.
 - Delay Time:** A numeric field set to '000:00:00.00'.
 - ESD Trip State:** A dropdown menu set to 'Fail Shut'.
 - Enable Upset:** A button to activate the configured upset.

At the bottom of the window, there are tabs for 'Design', 'Rating', 'Worksheet', 'Dynamics', and 'Cost'. Below the tabs are three buttons: 'Delete', 'OK' (highlighted in green), and 'Ignored'.

Fonte: Autora (2021)

4 MATERIAIS E MÉTODOS

Neste capítulo serão descritos os materiais e métodos utilizados nas etapas de simulação e coleta de dados do processo químico. Da mesma forma, serão expostos os materiais e técnicas para realização da metodologia HAZOP e também a descrição dos algoritmos de *Machine Learning* que serão aplicados na Detecção e Diagnóstico das falhas na válvula de controle.

4.1 Metodologia para a simulação do processo químico

A elaboração do modelo inicial do processo segue a metodologia descrita por Garcia (2006), nos seguintes passos:

- Definição da base da simulação;
- Caracterização da carga (alimentação) do sistema;
- Definição das operações unitárias envolvidas no processo;
- Especificações dos graus de liberdade do sistema.

A simulação do processo será realizada no *software* UniSim® Design R461. O sistema escolhido para realização do estudo é um vaso de separação de misturas bifásico, visto a ampla utilização e importância deste processo em diversas áreas industriais. Para a definição da base da simulação, serão adicionados os componentes químicos utilizados, segundo a Tabela 2. O pacote termodinâmico escolhido é o UNIQUAC, visto que este modelo descreve melhor o equilíbrio de fases líquido/vapor e representa adequadamente as propriedades termodinâmicas dos compostos escolhidos (PARANHOS, 2014).

Tabela 2 – Componentes utilizados na simulação do processo

Componente	Fórmula molecular
Benzeno	C_6H_6
Metano	CH_4

Fonte: Autora (2021)

A corrente de alimentação do processo tem vazão molar de 100,0 kgmol/h, temperatura de 25 °C e pressão de 250,0 kPa. A composição dos componentes na alimentação está descrita na Tabela 3.

Tabela 3 – Composição da corrente de alimentação do processo simulado

Componente	Fração molar
Benzeno	0,70
Metano	0,30

Fonte: Autora (2021)

4.2 Metodologia para dimensionamento dos equipamentos

4.2.1 Dimensionamento do vaso de separação

O dimensionamento do vaso será realizado segundo Novo (1982), para um tambor horizontal de separação líquido/vapor. Nesta metodologia o cálculo do diâmetro do vaso é realizado iterativamente até que se satisfaça as dimensões requeridas pelas duas fases. Para isso é necessário variar os valores de x , e respectivos y , e da razão comprimento/diâmetro, L/D , dentro das faixas indicadas por Novo (1982) e apresentadas nas Tabelas 4 e Tabela 5, de forma a se obter um valor único ideal para o diâmetro, D . Novo (1982) apresenta x e y como apresentado na Equação 1 e Equação 2, respectivamente.

$$x = \frac{H}{D} \quad (1)$$

$$y = \frac{A_v}{A_t} \quad (2)$$

Onde:

H = Altura do vaso em m.

D = Diâmetro do vaso em m;

A_v = Área transversal do vaso em m²;

A_t = Área do vaso em m².

Tabela 4 – Valores dos parâmetros x e y

x	y
0,20	0,142378
0,25	0,195501
0,30	0,252315
0,35	0,311918
0,40	0,373530
0,45	0,436445
0,50	0,50

Fonte: Adaptado de Novo (1982)

Tabela 5 – Valores propostos para a razão L/D

L/D
2,5
3,0
3,5
4,0
4,5
5,5
5,5

Fonte: Adaptado de Novo (1982)

Após, calcula-se os diâmetros requeridos pelas duas fases, líquido e vapor, de acordo com a Equação 3 e Equação 4, respectivamente.

$$D_v^2 = \frac{4}{\pi \left(\frac{L}{D}\right)} \frac{Q_v}{V_p} \frac{x}{y} \quad (3)$$

$$D_l^3 = \frac{4}{\pi \left(\frac{L}{D}\right)} \frac{Q_l \times t_r}{(1 - y)} \quad (4)$$

Onde:

D_v = Diâmetro do vaso que contém a parte vapor em m;

Q_v = Vazão de vapor em m³/s;

V_p = Velocidade de projeto em m/s;

D_l = Diâmetro do vaso que contém a parte líquida em m;

Q_1 = Vazão de líquido em m^3/s ;

t_r = tempo de residência em s.

Para cada valor de (x, y) , em um valor fixo de L/D , calcula-se os diâmetros requeridos e plota-se um gráfico (x, D) , onde o ponto de intersecção das curvas D_v e D_1 corresponde ao diâmetro do vaso.

4.2.2 Dimensionamento das válvulas de controle

As válvulas de controle serão dimensionadas pelo UniSim® Design quanto ao seu coeficiente de vazão, cv . O valor da perda de carga nas válvulas será de 9 psi, conforme heurística apresentada por Turton (2001).

4.2.3 Dimensionamento da bomba

A bomba B-001 será dimensionada pelo UniSim® Design, com os valores de vazão (m^3/h), da corrente de entrada da bomba, *Head* (m) e eficiência para gerar a curva de operação da bomba.

4.3 Metodologia para realização do HAZOP

A realização da metodologia HAZOP seguirá à proposta apresentada por Aguiar (2008), segundo os passos:

- Determinar os pontos do subsistema a serem analisados, chamados de nós;
- Criar a planilha para realização da técnica;
- Aplicar as “palavras-guia” nos nós;
- Verificar os desvios e realizar o levantamento das causas;
- Verificar a consequência dos desvios e possíveis formas de minimizá-los.

Neste TCC, o HAZOP será aplicado no sistema de estudo para determinar qual das válvulas de controle apresenta maior risco ao processo. A válvula de controle selecionada nesta etapa será utilizada posteriormente para aplicação das falhas e avaliação das técnicas de *Machine Learning*. O Quadro 1 apresenta a planilha para realização da metodologia HAZOP no sistema proposto.

Quadro 1 – Planilha utilizada na metodologia HAZOP

Análise de Perigos e Operabilidade									
Unidade:									
Sistema:					Equipe:			Data:	
Localização do nó:								Página:	
Desvio	Causas	Deteção	Salvaguardas	Efeitos	Cat. Freq	Cat. Sev	Cat. Risco	Obs	nº

Fonte: Autora (2021)

4.4 Metodologia para a coleta dos dados

A coleta de dados será realizada utilizando o modo dinâmico do UniSim® Design R461. Os dados serão obtidos através do monitoramento e controle dos instrumentos aplicados na simulação do sistema. As variáveis de processo que serão monitoradas são o nível e pressão do vaso e a vazão da corrente de descarga da bomba B-001. Para isso devem ser definidos os valores mínimos e máximos das variáveis de processo, assim como o *set point* de cada uma. Deve-se ajustar também a ação dos controladores, em reversa ou direta, e os parâmetros da sintonia Proporcional Integral (PI). Os dados simulados e gerados no software UniSim são salvos em arquivos do tipo csv (Comma-separated values) que possibilitam a manipulação destes na linguagem Python.

Os dados coletados serão referentes à operação do sistema sem falhas e com a aplicação de duas falhas na válvula de controle selecionada na metodologia HAZOP. Neste TCC foram selecionadas as falhas do tipo agarramento e supressão de pressão, com ação Falha-Fecha, para realizar o estudo proposto. Para que seja possível analisar a resposta do sistema frente às duas situações, será feita uma alteração no *setpoint* do controlador do nível do vaso.

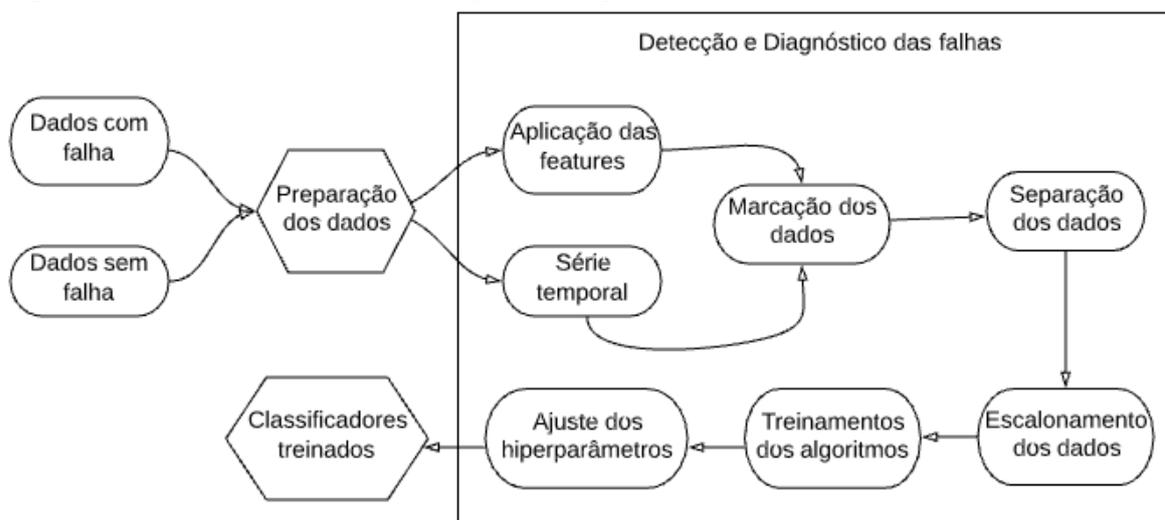
4.5 Metodologia para a análise exploratória de dados

A partir do conjunto de dados obtido, serão construídos gráficos temporais com a finalidade de explorar as variáveis e o comportamento do processo durante a operação normal e com a ocorrência das falhas. A análise exploratória dos dados será realizada utilizando a linguagem Python e suas bibliotecas e será desenvolvida no ambiente computacional do Jupyter Notebook. Esta ferramenta é um aplicativo *web open-source* gratuito destinado à criação de códigos em diversas linguagens de programação.

4.6 Metodologia para Detecção e Diagnóstico das Falhas

O método selecionado para realizar a DDF é uma adaptação da metodologia da classificação supervisionada proposta por Leal (2019). Neste método os algoritmos supervisionados de classificação são treinados para isolar uma situação de falha a partir dos dados normais e anormais. O procedimento consiste em cinco etapas, conforme a Figura 13.

Figura 13 – Tarefas para a detecção e diagnóstico das falhas



Fonte: Adaptado de Leal (2019)

Serão testadas duas metodologias para o treinamento dos algoritmos na Detecção e Diagnóstico das falhas. Primeiramente utilizando os dados coletados como série temporal e após realizando a extração de características das variáveis, compondo um novo grupo de dados denominado *features*. A série temporal consiste

em dados obtidos sequencialmente ao longo do tempo, ou seja, uma sequência de pontos em ordem sucessiva que normalmente ocorre em intervalos de tempo uniformes.

4.6.1 Aplicação das *features*

Nesta etapa é realizado um processo de transformação dos dados obtidos, onde cada coluna com as variáveis do processo são convertidas em novos parâmetros como, por exemplo, média, valores máximos e mínimos, desvio padrão, mediana dos valores e etc. Para isso, é necessário que o banco de dados esteja com uma marcação temporal, *timestamp*, e que seja escolhido o intervalo de tempo para aplicação das *features*.

4.6.2 Marcação dos dados

Como os dados, obtidos no UniSim® Design não possuem marcação indicando se o comportamento é normal ou anormal é necessário realizar uma etapa de marcação dos dados para categorizar cada comportamento. A marcação dos dados será realizada utilizando a função `get_dummies` do Python, que irá converter as variáveis categóricas, agarramento, sem falha e falha fecha em variáveis *dummy* indicadoras, para realizar a multiclassificação dos dados.

4.6.3 Separação dos dados

Os dados devem ser separados em conjuntos menores para realizar o treinamento e avaliação dos algoritmos de *Machine Learning*. Para isso são divididos em grupos de dados de treinamento e grupo de dados de teste. Os dados de treinamento serão utilizados para que o algoritmo possa aprender as características do conjunto. Após o treinamento, os dados de teste são inseridos no algoritmo de forma a avaliar o desempenho da técnica aplicada. Os primeiros 80% dos dados serão utilizados para o treinamento e os 20% restantes para o teste.

4.6.4 Escalonamento dos dados

Nesta etapa é realizada uma transformação do conjunto original de dados de modo que cada valor passe a apresentar média zero e variância unitária.

4.6.5 Treinamento dos algoritmos

Após as etapas anteriores o próximo passo será realizar o treinamento dos algoritmos. Neste TCC serão utilizados os modelos de Árvores de Decisão, Redes Neurais Artificiais e Florestas Randômicas, para a detecção e diagnóstico das falhas. Será utilizada a biblioteca *scikit-learn* na importação de funções e métodos para desenvolver os algoritmos de *Machine Learning*.

4.6.5.1 Avaliação dos algoritmos de *Machine Learning*

A avaliação dos algoritmos de Machine Learning será realizada através da biblioteca *scikit-learn* com função *score()* da respectiva biblioteca. Esta função fornece o quão bem o modelo utilizado é capaz de generalizar os dados. Verificando o número que o modelo previu e comparando com o valor esperado do conjunto de treino. O resultado desta função é um valor entre 0 e 1, em que, quanto mais próximo de 1 melhor a acurácia do modelo. A Equação 5 apresenta matematicamente o cálculo realizado por esta função.

$$\text{Acurácia } (\alpha, \beta) = \frac{1}{n_{\text{exemplos}}} \sum_{i=0}^{n_{\text{exemplos}}-1} 1(\beta_i = \alpha_i) \quad (5)$$

Onde:

α = Valor verdadeiro da amostra

β = valor previsto

n_{exemplos} = número total de amostras

4.6.6 Ajuste dos hiperparâmetros

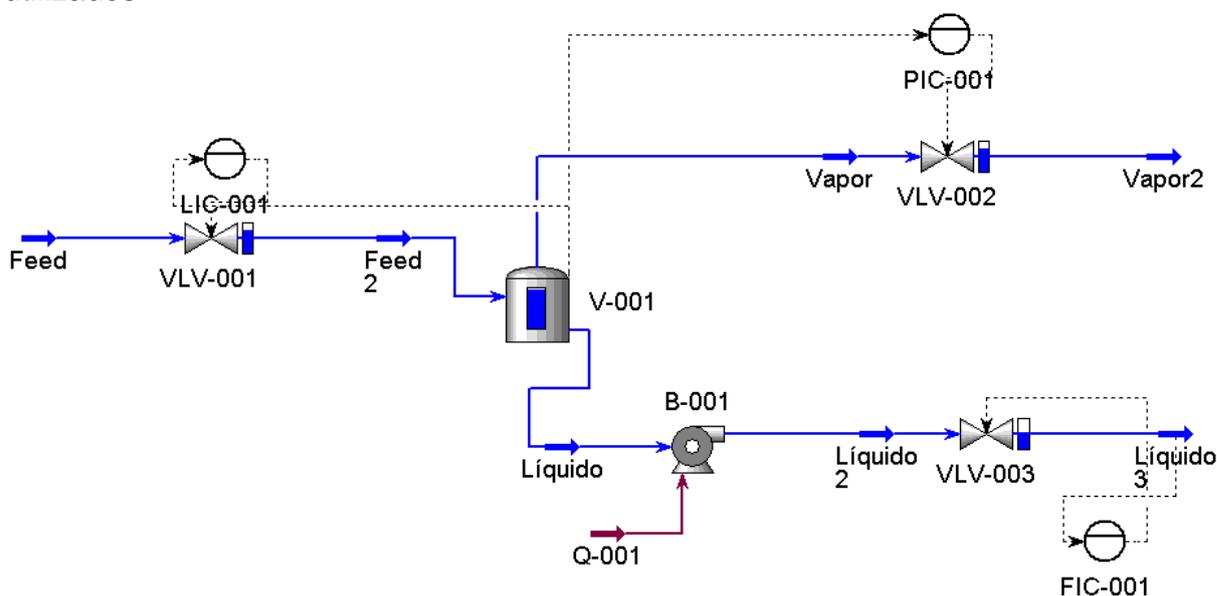
Após avaliar qual método realiza as melhores previsões, recomenda-se realizar a otimização dos hiperparâmetros do modelo. Os hiperparâmetros são os parâmetros ajustáveis que possibilitam controlar a etapa de treinamento dos modelos de *Machine Learning*. Neste processo, o objetivo é buscar a configuração dos hiperparâmetros que resulte no melhor desempenho do algoritmo. Esta etapa será realizada apenas com a técnica que obtiver melhor acurácia na etapa de testes.

5 RESULTADOS E DISCUSSÃO

5.1 Simulação do processo químico

A Figura 14 apresenta a simulação do processo químico obtido no software UniSim® Desgin R461, com os equipamentos e instrumentos utilizados.

Figura 14 – Simulação do processo contendo os equipamentos e instrumentos utilizados



Fonte: Autora (2021)

A Tabela 6 apresenta os equipamentos e instrumentos utilizados, com as respectivas *tags*.

Tabela 6 – Equipamentos e instrumentos utilizados na simulação

Equipamento/Instrumento	Tag
Bomba centrífuga	B-001
Controlador Indicador de Nível	LIC-001
Controlador Indicador de Pressão	PIC-001
Controlador Indicador de Vazão	FIC-001
Vaso de separação trifásico	V-001
Válvula de controle	VLV-001, VLV-002, VLV-003

Fonte: Autora (2021)

O controlador LIC-001 atua na válvula VLV-001 e controlará o nível do vaso V-001. O instrumento PIC-001 atua na válvula VL-002, monitorando a pressão do vaso. E o instrumento FIC-001 será responsável por monitorar a vazão da corrente de descarga da bomba B-001 e atua na válvula VLV-003.

Os controladores são do tipo PI (Proporcional Integral), e foram sintonizados pelo UniSim® Design, quanto ao seu K_c e τ_i e sua ação em reversa ou direta. Os parâmetros obtidos estão apresentados na Tabela 7.

Tabela 7 – Parâmetros da sintonia dos controladores utilizados no sistema

Controlador	K_c	τ_i	Ação
LIC-001	1,57	1,29	Reversa
PIC-001	0,57	1,0	Direta
FIC-001	0,32	0,5	Reversa

Fonte: Autora (2021)

O controlador PIC-001 apresenta ação direta, pois quando ocorre um aumento de sinal da variável controlada, pressão do vaso, ocorre também um aumento no seu sinal de saída, fazendo com que a válvula VLV-002 abra. Em contraponto, os controladores LIC-001 e FIC-001, tem ação reversa uma vez que quando a variável controlada, nível e vazão, aumentam o sinal de saída do equipamento diminui. Neste caso as válvulas de controle VLV-001 e VLV-002 fecham, para que o vaso não transborde em relação ao nível, ou que a bomba cavite em relação a vazão.

Estes controladores são responsáveis por manter as variáveis manipuladas no *setpoint*. No sistema proposto estas variáveis são o nível e a pressão do vaso de separação, e a vazão da corrente de descarga da bomba, e seus *setpoint* foram definidos com base nas dimensões do sistema. A Tabela 8 apresenta estes valores para os respectivos controladores.

Tabela 8 – *Setpoint* dos controladores

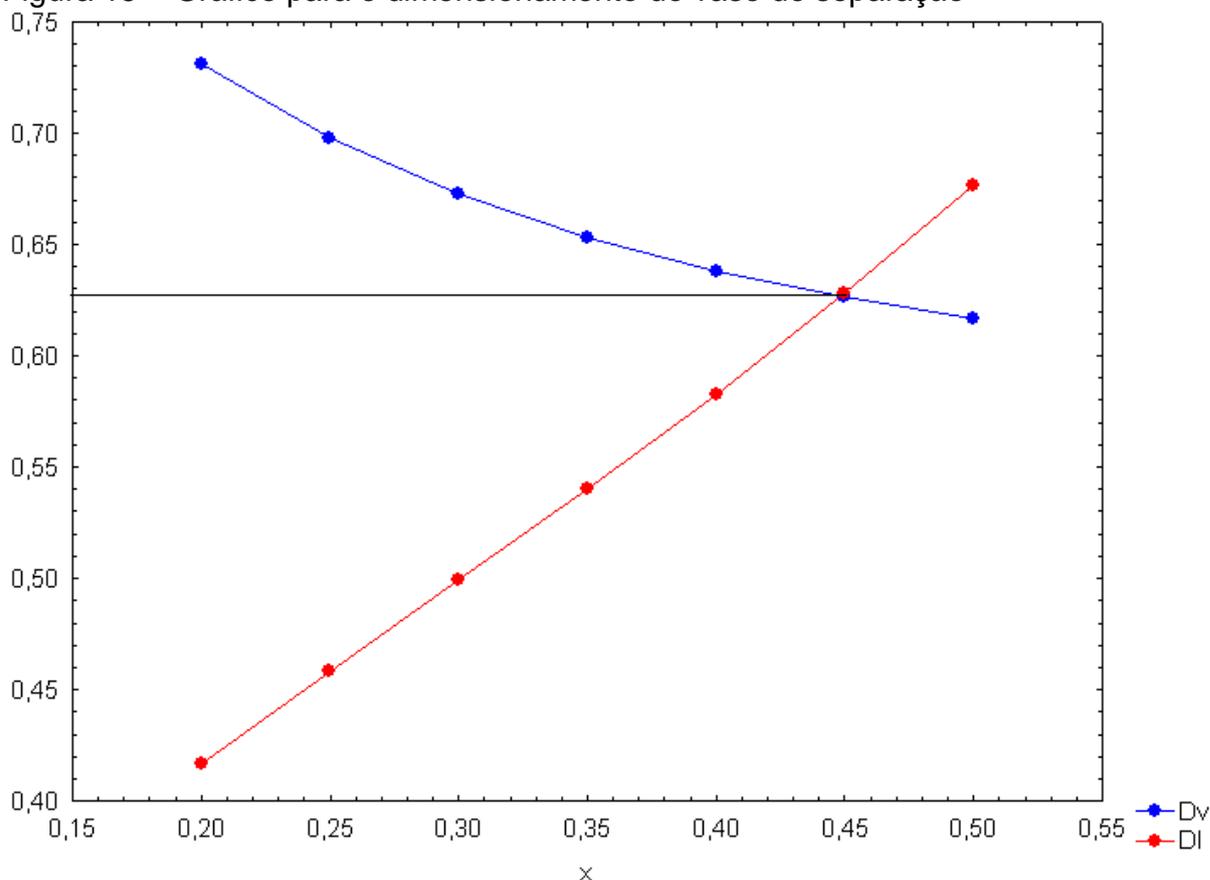
Controlador	<i>Setpoint</i>
LIC-001	0,7 m
PIC-001	170,0 kPa
FIC-001	4,0 m ³ /h

Fonte: Autora (2021)

5.2 Dimensionamento dos equipamentos

A Figura 15 apresenta o gráfico (x, D) , onde foi possível encontrar o valor ideal para o diâmetro do vaso de separação.

Figura 15 – Gráfico para o dimensionamento do vaso de separação



Fonte: Autora (2021)

O diâmetro resultante da intersecção das curvas foi de 0,63m. Foi adicionado um fator de segurança de 20% no diâmetro encontrado. Estes resultados são referentes à relação $L/D = 3$, com tempo de residência de 5 min. A Tabela 9 apresenta os valores obtidos para o dimensionamento do vaso de separação V-001.

Tabela 9 – Dimensionamento do vaso horizontal de separação V-001

Dimensão	Valor
Diâmetro (m)	0,76
Comprimento (m)	2,27

Fonte: Autora (2021)

Anjos (2015) encontrou resultados semelhantes para um vaso horizontal de separação bifásica, utilizado para separação primária do petróleo. O autor encontrou diâmetro de 0,76 m, comprimento de 2,28, razão $L/D = 3$ para um tempo de residência de 10 min. O sistema proposto pelo autor foi dimensionado com base em uma metodologia diferente da utilizada neste TCC.

A Tabela 10 apresenta o dimensionamento das válvulas de controle do tipo globo utilizadas na simulação quanto ao coeficiente de vazão (C_v) calculado de cada uma.

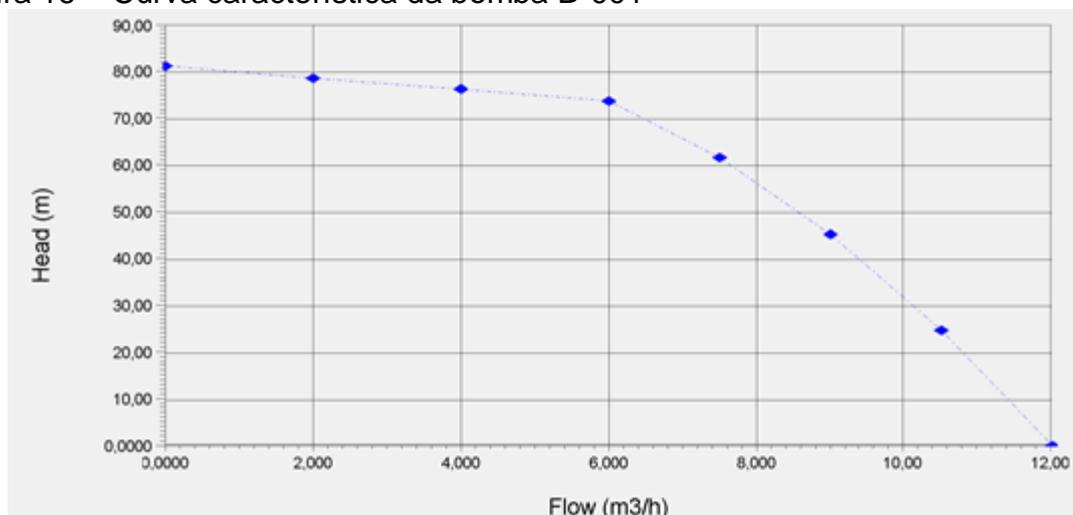
Tabela 10 – Dimensionamento das Válvulas de controle do tipo globo

Válvula de controle	C_v (USGPM)
VLV-001	46,66
VLV-002	27,19
VLV-003	7,93

Fonte: Autora (2021)

A curva característica da bomba centrífuga B-001 é apresentada na Figura 16.

Figura 16 – Curva característica da bomba B-001



Fonte: Autora (2021)

Este gráfico relaciona a altura manométrica alcançada pela bomba, com a vazão de recalque. A altura máxima de sucção da bomba é de 54 m, com potência de 0,95kW, perda de carga de 638 kPa e eficiência de 75%, especificados pelo UniSim® Design. A curva apresentada na Figura 16 representa as condições hidráulicas operacionais da bomba B-001.

5.3 Metodologia HAZOP

Segundo o quadro apresentado no Anexo A, a válvula de controle que apresentará maior risco ao processo em caso de falha é a VLV-001, visto que após a análise de riscos do vaso de separação as situações de falha que envolvem a válvula VLV-001 apresentaram risco sério e menor segundo a Figura 7. Portanto, nesta válvula que serão aplicadas as falhas do tipo agarramento e supressão de pressão.

5.4 Análise exploratória dos dados obtidos

A simulação percorreu 42 horas, com alterações no *setpoint* do controlador LIC-001 de 0,7 para 0,5 metros no nível do vaso de separação, e aplicação das falhas na válvula de controle VLV-001. O conjunto de dados obtidos, conta com 8 colunas e 2570 linhas, com as variáveis tempo, *setpoint*, a variável manipulada dos controladores LIC-001, PIC-001 e FIC-001 e a abertura das válvulas VLV-001, VLV-002 e VLV-003, conforme a Figura 17.

Figura 17 – Conjunto de dados obtidos na simulação do processo

	Time	LIC-SP	LIC-PV	PIC-PV	FIC-PV	VLV-1	VLV-2	VLV-3
0	0.5	0.5	0.499996	170.0010	4.00000	7.485890e+01	75.3473	49.7241
1	60.5	0.5	0.499996	170.0010	4.00000	7.485950e+01	75.3478	49.7240
2	120.5	0.5	0.499997	170.0020	4.00000	7.486000e+01	75.3483	49.7240
3	180.5	0.5	0.499998	170.0010	4.00000	7.486030e+01	75.3487	49.7240
4	240.5	0.5	0.499999	170.0010	4.00000	7.486040e+01	75.3490	49.7240
...
2566	153961.0	0.7	0.237959	72.5577	3.97769	5.502290e-39	0.0000	56.3682
2567	154021.0	0.7	0.208194	70.1422	3.97812	1.363880e-41	0.0000	56.6023
2568	154081.0	0.7	0.176765	67.8226	3.97849	3.380720e-44	0.0000	56.8324
2569	154141.0	0.7	0.142679	65.5924	3.97879	8.379980e-47	0.0000	57.0595
2570	154201.0	0.7	0.104774	63.4398	3.97928	2.077190e-49	0.0000	57.2812

Fonte: Autora (2021)

A Figura 18 apresenta o conjunto de estatísticas descritivas das variáveis do conjunto de dados, como a quantidade de valores (count), a média (mean), o desvio padrão (std), o valor mínimo (min), os quartis da distribuição (25%, 50% e 75%) e o valor máximo (max).

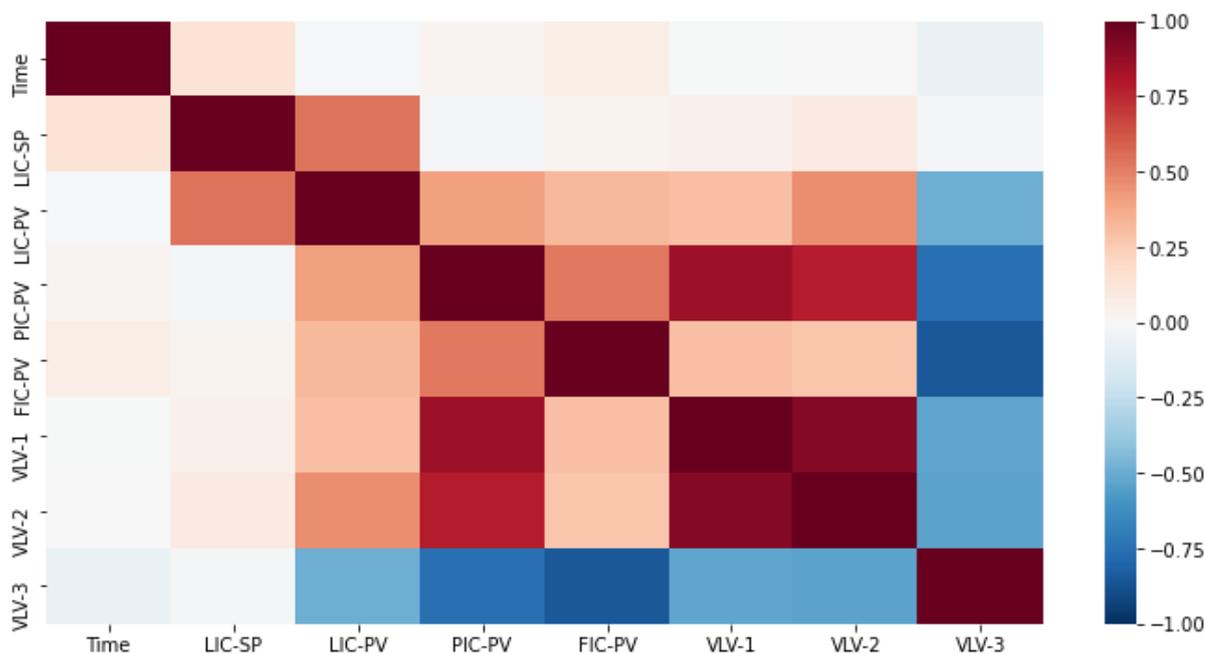
Figura 18 – Estatísticas descritivas do conjunto de dados

	Time	LIC-SP	LIC-PV	PIC-PV	FIC-PV	VLV-1	VLV-2	VLV-3
count	2571.000000	2571.000000	2571.000000	2571.000000	2571.000000	2.571000e+03	2571.000000	2571.000000
mean	77100.675807	0.581602	0.532540	160.878470	3.953744	6.862381e+01	68.026961	50.938825
std	44539.883151	0.098312	0.160056	33.443998	0.467349	2.639435e+01	25.246140	6.141575
min	0.500000	0.500000	0.000000	0.010000	-0.394994	4.920960e-90	0.000000	46.347500
25%	38550.500000	0.500000	0.491918	166.187000	3.991555	7.209910e+01	72.133800	49.608300
50%	77100.500000	0.500000	0.505034	169.981000	3.999840	7.485560e+01	75.320200	49.691300
75%	115651.000000	0.700000	0.699109	171.575500	4.002045	7.721345e+01	76.632400	49.935050
max	154201.000000	0.700000	0.712625	193.670000	6.192180	1.000000e+02	100.000000	100.000000

Fonte: Autora (2021)

A matriz de correlação dos dados mostra a correlação entre as variáveis do processo, sendo apresentada na Figura 19. Neste gráfico a relação positiva dos pares representa que o aumento de uma variável é acompanhado pelo aumento da outra. Em contraponto, a correlação negativa significa que a diminuição de uma é acompanhada do aumento da outra. E quando os pares de variáveis não apresentarem uma correlação linear, o valor apresentado será zero.

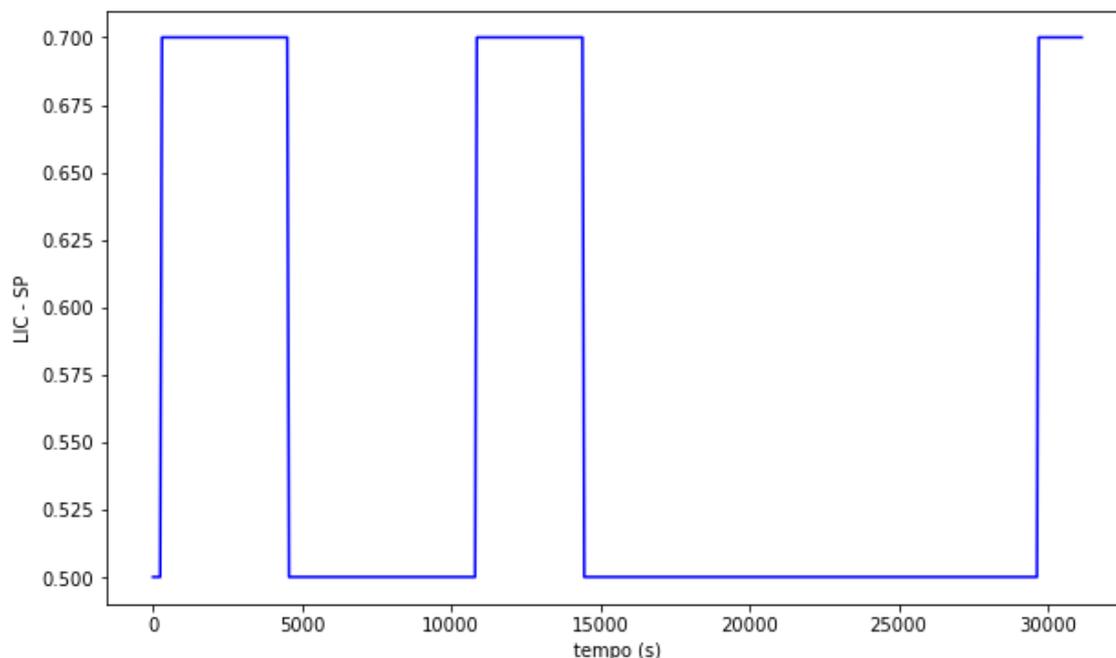
Figura 19 – Matriz de correlação das variáveis



Fonte: Autora (2021)

Em relação à abertura das válvulas pode-se observar a relação positiva da VLV-001 com a VLV-002, visto que quando aumentar o nível do tanque o controlador PIC-001 responderá abrindo também a válvula VLV-002 para aliviar a pressão no vaso. A VLV-003 apresenta correlação negativa com as outras duas, pois se a vazão na descarga da bomba aumentar, consequentemente o nível do tanque cai assim como a pressão sendo necessário fechar a outra válvula para manter o sistema em equilíbrio. Esta situação ocorre quando o *setpoint* for alterado de 0,7 para 0,5. Observa-se de uma forma geral, que as variáveis do sistema apresentam correlação positiva entre si, com exceção da abertura da válvula VLV-003.

O gráfico da Figura 20 apresenta as variações feitas no *setpoint* do controlador LIC-001 de 0,7 para 0,5 metros, e vice-versa.

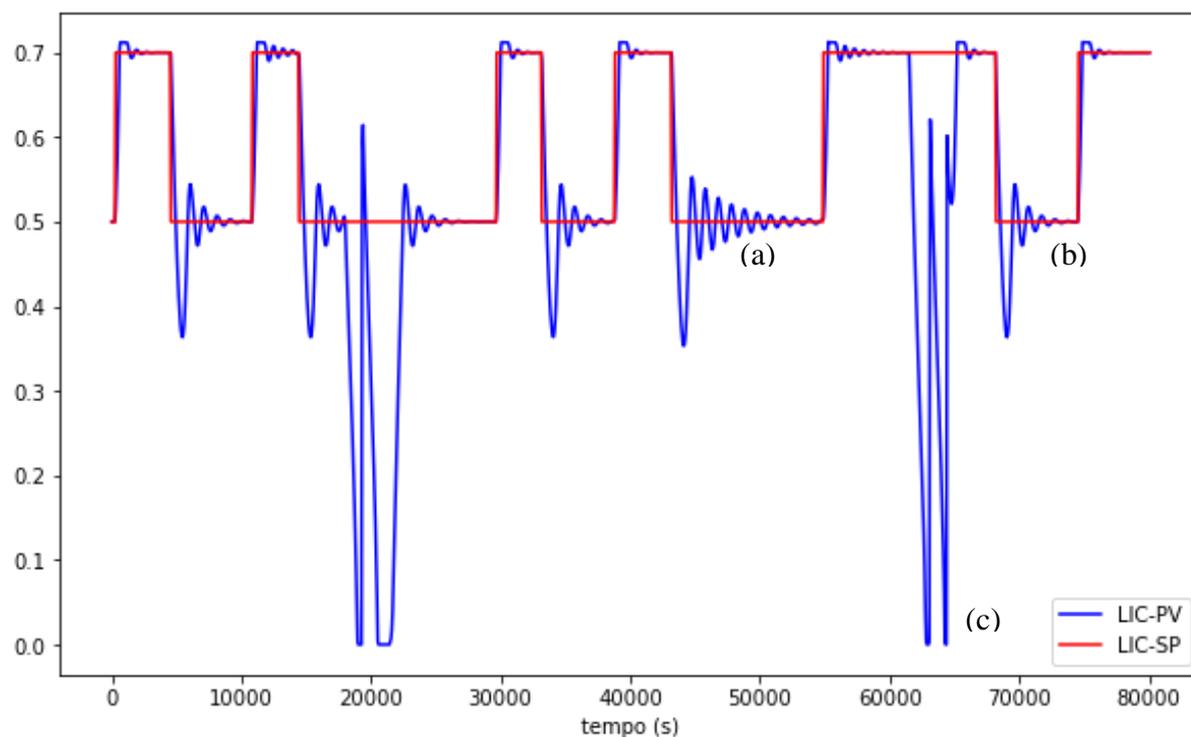
Figura 20 – Gráfico do *setpoint* do controlador LIC-SP

Fonte: Autora (2021)

As alterações com magnitude de 0,2 m no *setpoint* foram realizadas para observar a resposta do sistema na presença de falhas na válvula VLV-001. O valor do *setpoint* foi mantido em determinado valor até que as variáveis monitoradas estabilizassem, para a falha de agarramento e modo sem falha. Para a falha de supressão de pressão, a variação de *setpoint* foi realizada quando ocorreu à cavitação da bomba, visto o nível do vaso ir a zero. Dessa forma, essas alterações não ocorrem de forma periódica, como observado na Figura 20.

O comportamento da variável manipulada, LIC-PV, frente às modificações na variável, LIC-SP, esta representado na Figura 21.

Figura 21 – Resposta da variável LIC-PV frente a alterações LIC-SP

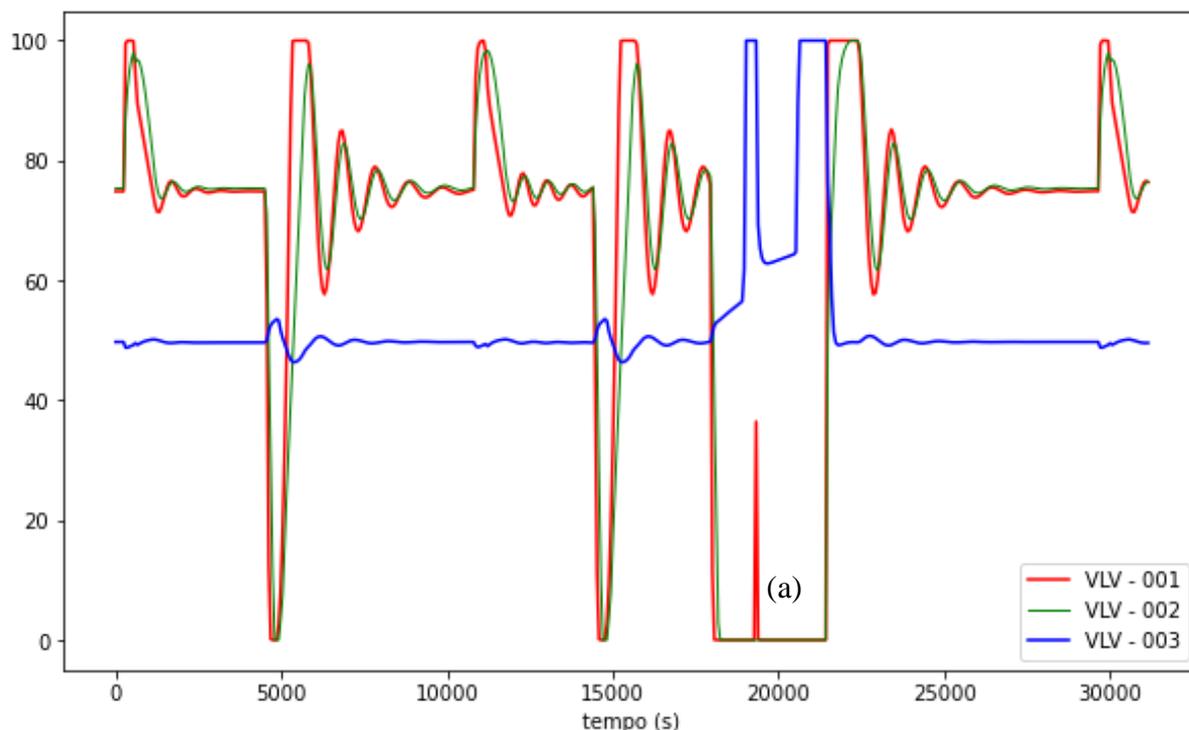


Fonte: Autora (2021)

Neste gráfico é possível observar os pontos em que ocorrem as falhas de agarramento e de supressão de pressão. O agarramento causa uma resposta mais lenta na busca do *setpoint* da variável, ponto a, em comparação a resposta da válvula sem falhas, apresentado no ponto b. Enquanto, no ponto c, observa-se o nível do vaso indo à zero na ocorrência da falha de supressão de pressão, devido à mesma apresentar ação Falha-Fecha.

O gráfico representado da Figura 22 confirma o constatado no mapa de correlação da Figura 19 para a abertura das válvulas, com range de 0 a 100%. Visto que, quando as válvulas VLV-001 e VLV-002 abrem, a resposta da VLV-003 é contrária. O pico da abertura da válvula VLV-001 apresentado no ponto a, representa uma alteração de falha de supressão de pressão para sem falha, e logo após a mudança para falha fecha novamente.

Figura 22 – Resposta da abertura das válvulas frente a alterações no *setpoint* do nível do vaso

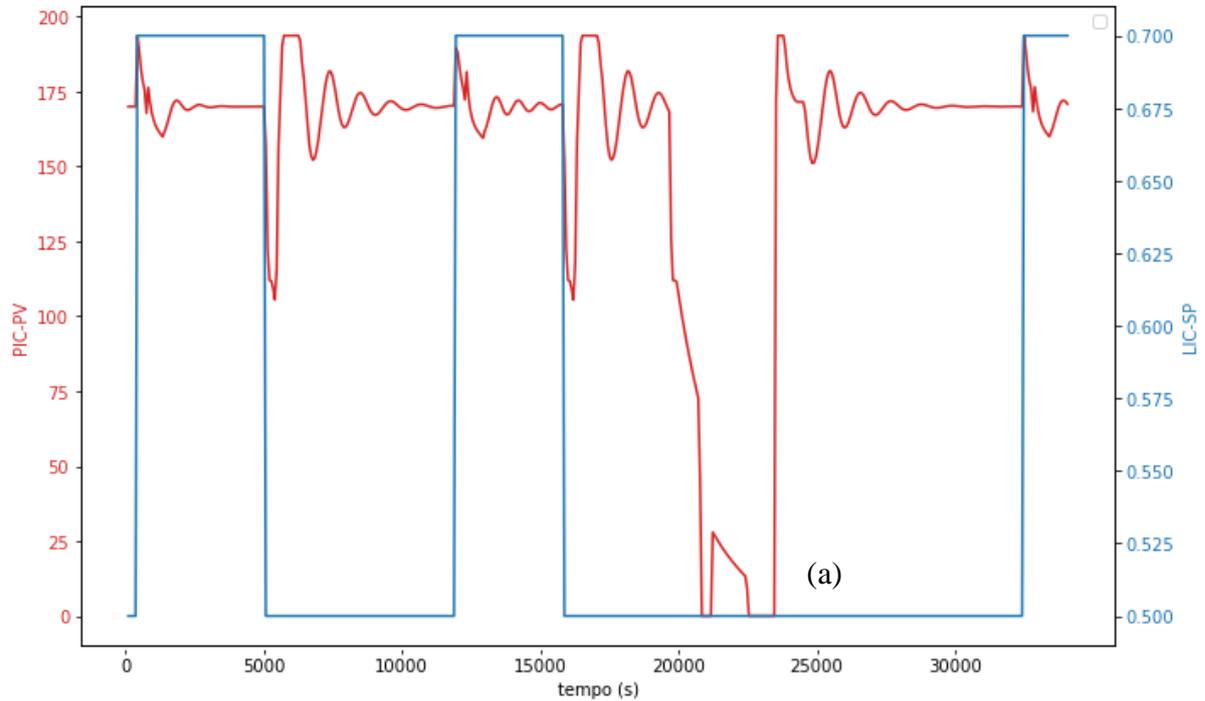


Fonte: Autora (2021)

A variável manipulada, PIC-PV, sendo esta a pressão do vaso de separação, apresenta valores máximos de aproximadamente 200 kPa, quando o nível do vaso é alterado para 0,7 m. E, valores de 110 kPa quando o nível é modificado para 0,5 m. Logo em seguida a cada caso, é possível observar a resposta do controlador na busca do *setpoint* desta variável, apresentado na Figura 23. A pressão é zero, quando se aplica a falha de supressão de pressão, e o sistema entra em colapso com o nível caindo e ocorrendo a cavitação da bomba. O ponto a, representa uma mudança de falha de supressão de pressão para sistema sem falha, porém sem alterações no *setpoint* do nível.

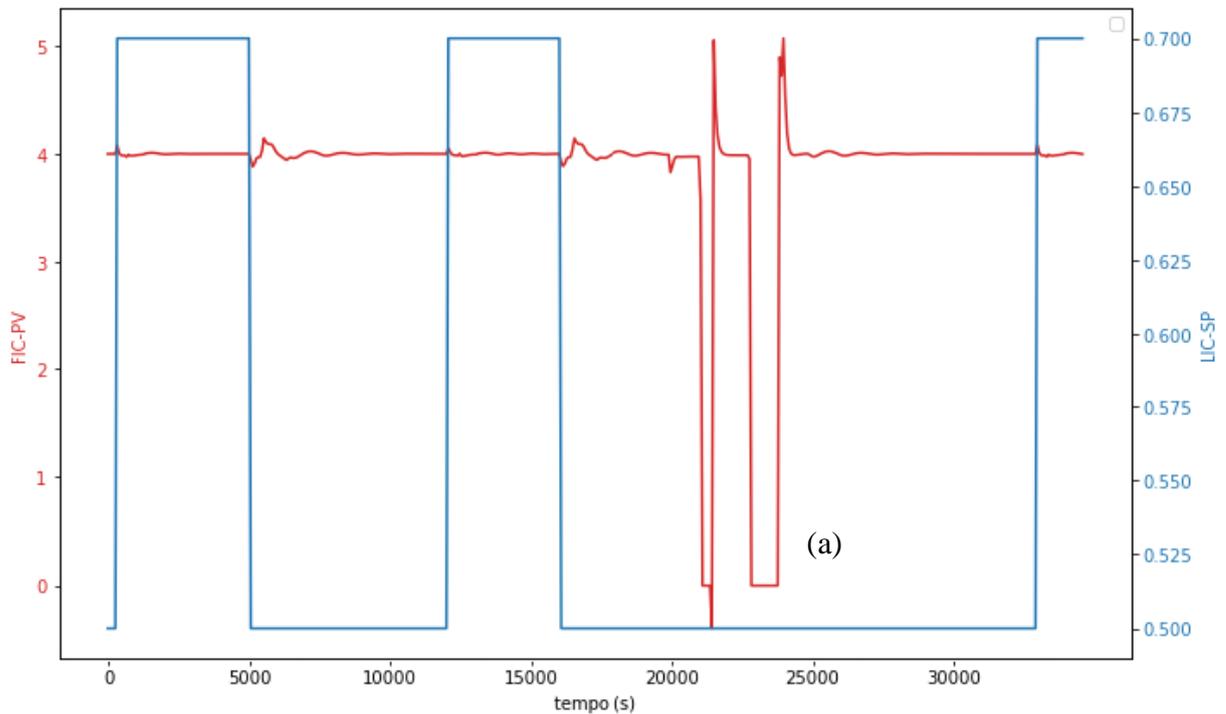
O mesmo pode ser observado na Figura 24, para a vazão da corrente de descarga da bomba, FIC-PV. Sendo esta variável, negativamente correlacionada com a abertura da válvula VLV-003, e positivamente relacionada com a abertura da válvula VLV-001, como pode ser observado na ocorrência da falha de supressão de pressão no ponto a. A vazão será nula, mesmo com a válvula VLV-003 totalmente aberta, pois a VLV-001 estará totalmente fechada e o nível do vaso será zero.

Figura 23 – Resposta da variável PIC-PV frente a alterações no *setpoint* do nível do vaso



Fonte: Autora (2021)

Figura 24 – Resposta da variável FIC-PV frente a alterações no *setpoint* do nível do vaso



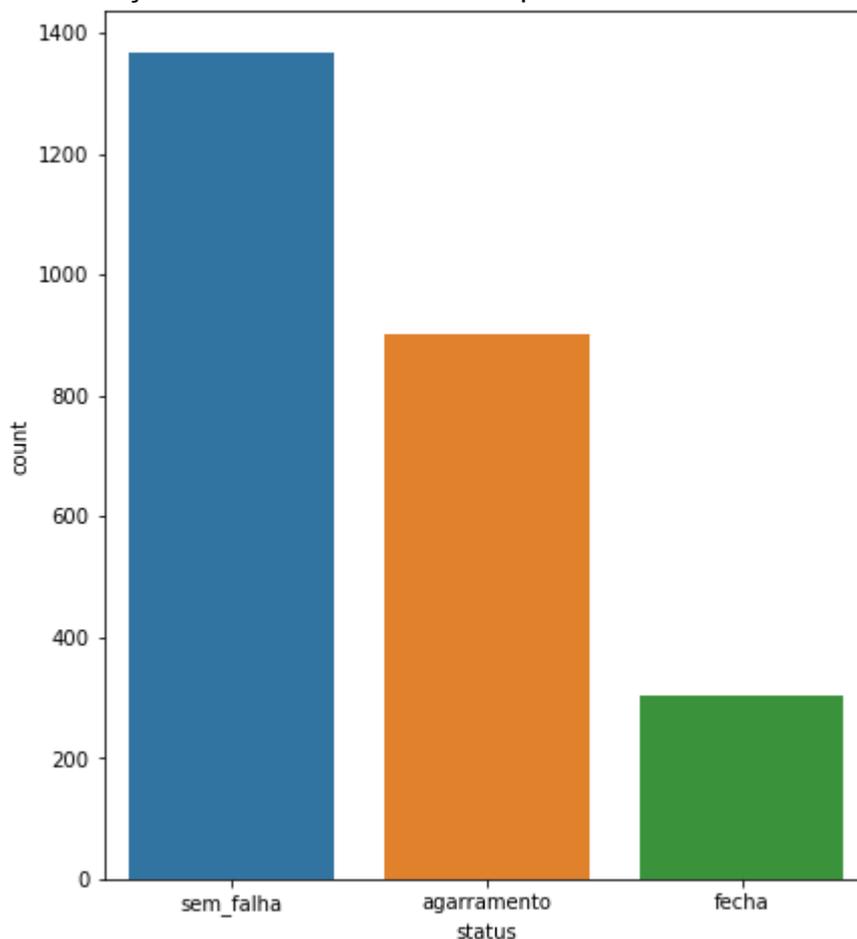
Fonte: Autora (2021)

5.5 Detecção e Diagnóstico das falhas

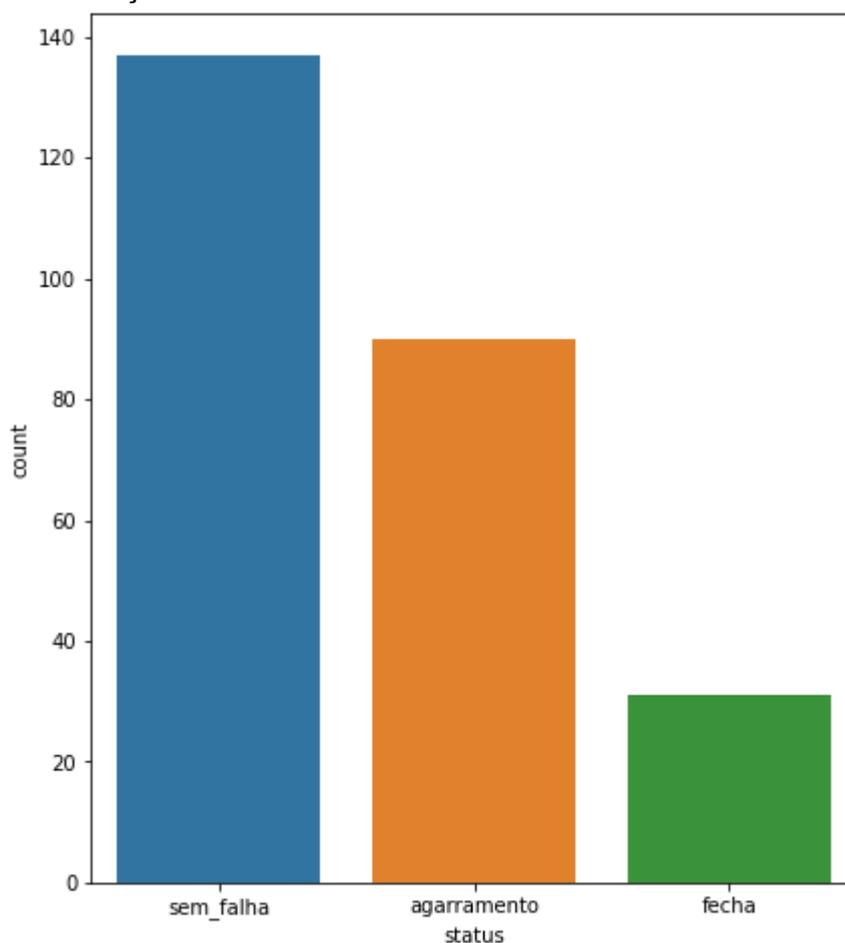
5.5.1 Pré-processamento dos dados

O conjunto de dados, para a série temporal, contém 2571 registros, conforme a Figura 25. Sendo 1368 dados sem falha, 901 com falha de agarramento e 302 com a falha de supressão de pressão. Enquanto que o conjunto referente às *features* têm 258 pontos, devido à extração de características, sendo representado na Figura 26. Destes pontos, 137 são sem falha, 90 com falha de agarramento e 30 pontos com a falha de supressão de pressão.

Figura 25 – Distribuição dos dados na série temporal



Fonte: Autora (2021)

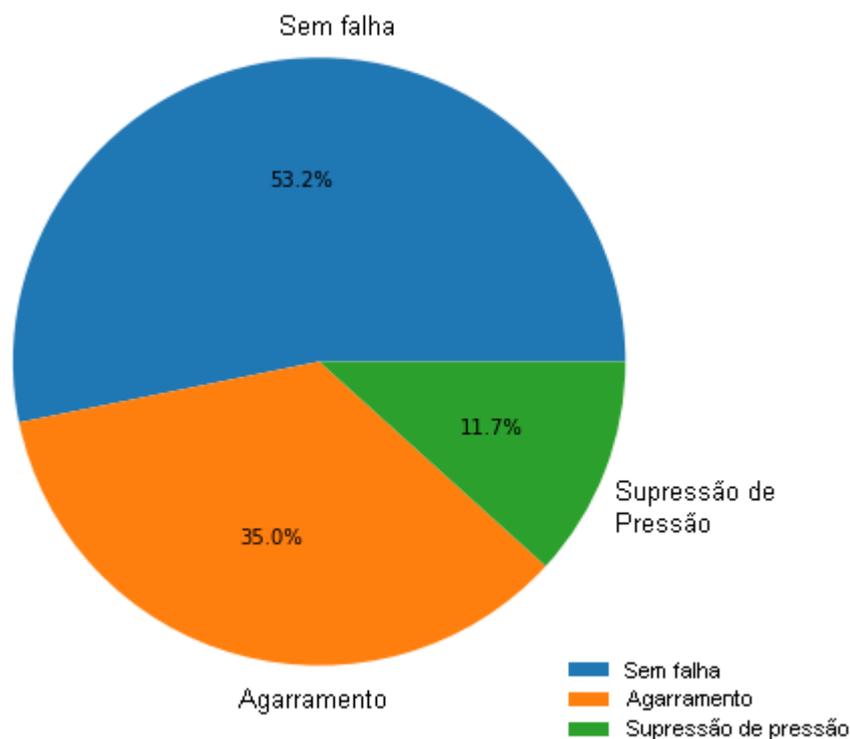
Figura 26 – Distribuição dos dados com *features*

Fonte: Autora (2021)

As características utilizadas nas *features* foram, a média aritmética, os valores máximos e mínimos, a mediana, e os quartis, de 25% e 75%, com amostragem de 10 minutos. Resultando em um novo conjunto de dados com 258 linhas e 43 colunas.

A Figura 27 apresenta a distribuição percentual dos dados dentro do conjunto original, sendo a mesma para as duas metodologias aplicadas. Destes 53,2% são referentes aos dados sem falha, 35% são dados são com agarramento e 11,7 % dos dados tem presença de falha de supressão de pressão.

Figura 27 – Distribuição do conjunto de dados



Fonte: Autora (2021)

5.5.2 Avaliação das técnicas de *Machine Learning*

5.5.2.1 Árvores de decisão

A Tabela 11 apresenta os resultados obtidos para a Detecção e Diagnóstico das falhas na válvula VLV-01 com a técnica de Árvores de Decisão.

Tabela 11 – Resultados da técnica de Árvores de Decisão para a Detecção e Diagnóstico das falhas

	Série temporal		<i>Features</i>	
Árvores de decisão	Treino	Teste	Treino	Teste
	0,68	0,53	0,83	0,69

Fonte: Autora (2021)

A acurácia obtida para a técnica de Árvores de Decisão foi superior a 50% nos dados com a série temporal e foi de 69% para os dados com *features*, na fase

de testes do algoritmo. Obteve-se uma melhora de 16% quando o modelo utilizou a metodologia com as *features*. Leal (2019) obteve acurácia de 72% para detecção de casos isolados de falhas em equipamentos de refrigeração, utilizando um classificador do tipo *Adaboost*. Este tipo de classificador cria uma união de algoritmos para formar um novo com melhor capacidade de aprendizagem.

5.5.2.2 Redes Neurais Artificiais

Os resultados obtidos na Detecção e Diagnóstico de falhas na válvula de controle VLV-01, para a técnica de Redes Neurais Artificiais estão apresentados na Tabela 12.

Tabela 12 – Resultados da técnica de Redes Neurais Artificiais para a Detecção e Diagnóstico das falhas

Redes Neurais Artificiais	Série temporal		<i>Features</i>	
	Treino	Teste	Treino	Teste
	0,71	0,56	0,84	0,62

Fonte: Autora (2021)

A Detecção e Diagnóstico de falhas com o modelo de Redes Neurais Artificiais resultou em 56% de acurácia na série temporal e 62% nos dados com *features*, para a fase de teste do algoritmo. O modelo com *features* apresentou uma melhora de 6% nos resultados para o treinamento do algoritmo. Foi utilizada uma rede neural com duas camadas escondidas de 400 e 200 neurônios, respectivamente. Rojas e De Souza (2018), obtiveram resultados acima de 80% de acurácia para Detecção e Diagnóstico de falhas em processos químicos. O autor testou várias arquiteturas de redes para realizar a classificação dos dados, o que possibilitou resultados melhores em comparação aos apresentados neste TCC.

5.5.2.3 Floresta Randômica

A Tabela 13 apresenta os resultados obtidos para a técnica de Floresta Randômica na Detecção e Diagnóstico de falhas na válvula de controle VLV-01.

Tabela 13 – Resultados da técnica de Floresta Randômica para a Detecção e Diagnóstico das falhas

Floresta Randômica	Série temporal		Features	
	Treino	Teste	Treino	Teste
	0,97	0,79	0,95	0,82

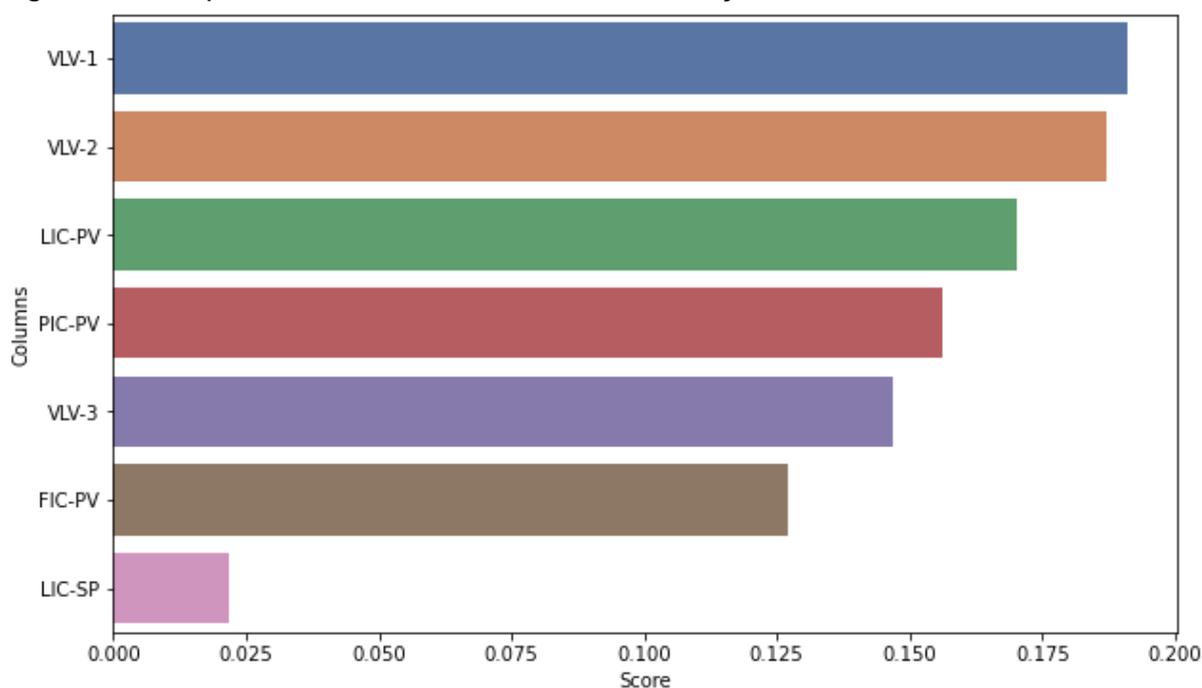
Fonte: Autora (2021)

A Floresta Randômica foi a técnica de Machine Learning que apresentou os melhores resultados na Detecção e Diagnóstico das falhas na válvula VLV-001. Com acurácia de 79% na série temporal e 82% nas *features*, apresentando melhora de 3% na aplicação da extração de características. Reis (2021) obteve resultados similares para o reconhecimento de falhas no maquinário industrial a partir de análises de som. Com acurácia na faixa de 59% a 71%, para as quatro válvulas de controle analisadas pela autora.

5.5.2.4 Detecção e Diagnóstico das falhas com a Floresta Randômica

A Figura 28 apresenta o gráfico com a importância das variáveis do processo na classificação obtida pelos modelos.

Figura 28 – Importância das variáveis na classificação dos modelos

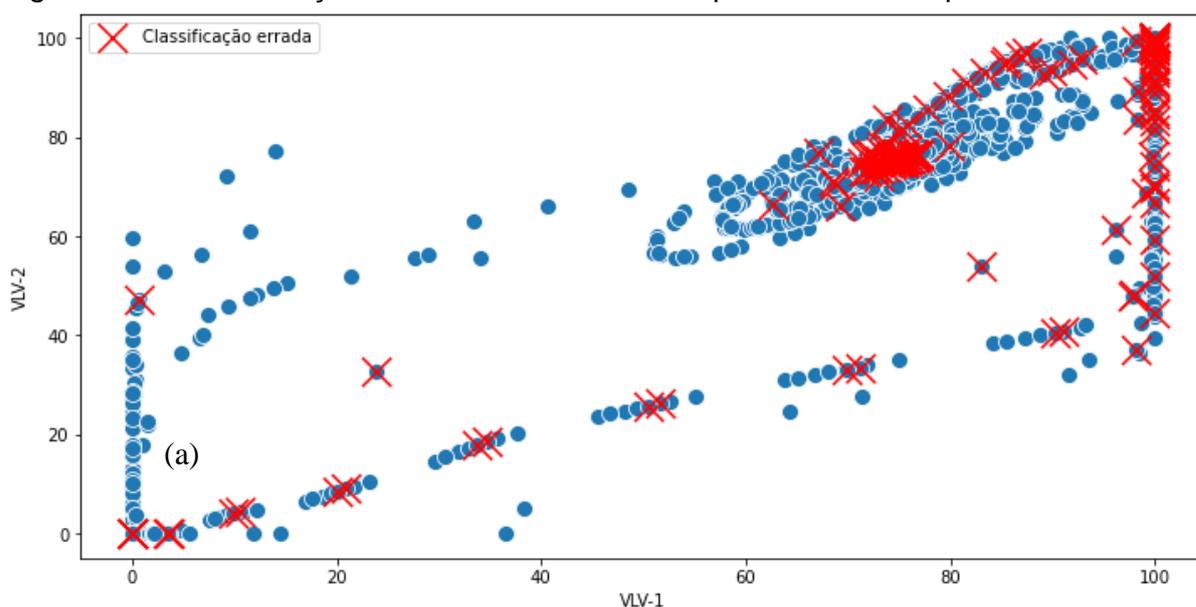


Fonte: Autora (2021)

As variáveis de maior importância foram a abertura da válvula VLV-001 e a abertura da válvula VLV-002. Sendo estas, as primeiras variáveis do sistema a responder as mudanças no *setpoint* do controlador LIC-001, e por consequência influenciam o comportamento das outras variáveis. Dessa forma, foram consideradas pelos algoritmos, com maior relevância ao detectar os episódios de falhas, e neste caso diagnosticar o tipo de falha que ocorreu.

A Figura 29 expõe o resultado da classificação da Floresta Randômica em relação aos dados em série temporal, com as variáveis de maior importância apresentadas na Figura 28. E a Figura 30 apresenta o resultado para os dados com *features* da classificação do mesmo modelo. Os pontos com X nos gráficos são referentes aos dados em que a técnica errou a previsão do comportamento do sistema.

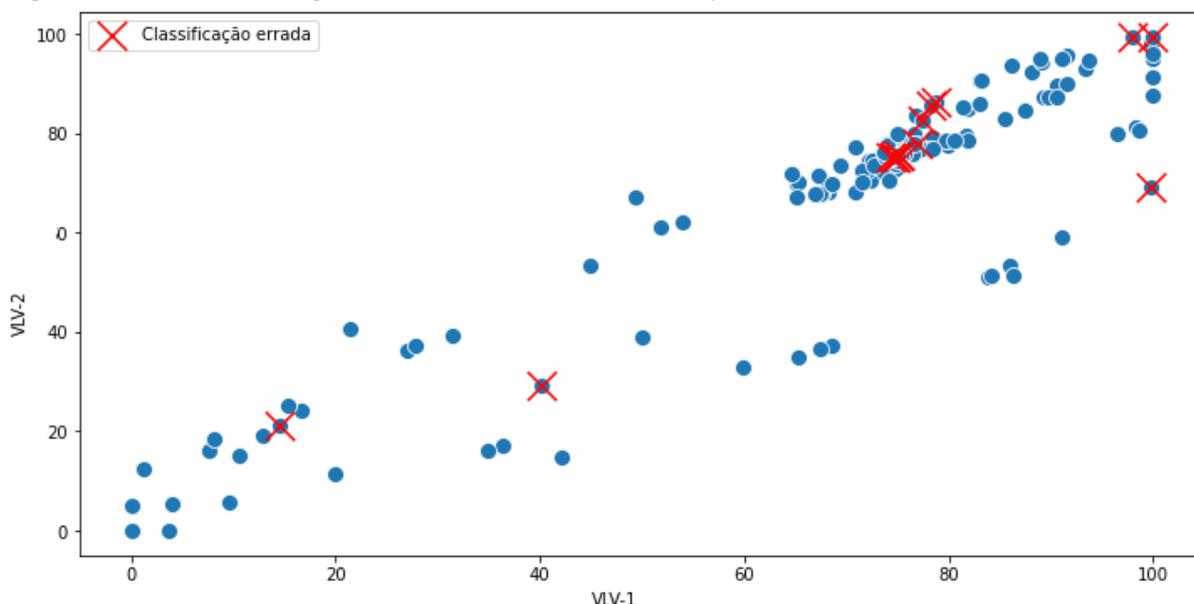
Figura 29 – Classificação da Floresta Randômica para a série temporal



Fonte: Autora (2021)

O comportamento apresentado na Figura 29 é característico de válvulas com agarramento, visto que a válvula apresenta uma resistência ao iniciar o movimento. Quando o equipamento consegue vencer o atrito estático, a abertura da válvula dá um salto, conforme o ponto a, e causa as oscilações características desta falha.

Figura 30 – Classificação da Floresta Randômica para as *features*



Fonte: Autora (2021)

Nota-se a diferença significativa na acurácia das duas metodologias, com uma taxa menor de erro no algoritmo para os dados com *features*. Os pontos em que ocorreu uma maior concentração de erros na classificação com a série temporal foram simplificados devido a extração de características, com intervalo de 10 minutos, no modelo com as *features*. Diminuindo as previsões erradas e tornando o modelo melhor para a detecção e diagnóstico das falhas.

5.5.2.4.1 Ajuste dos hiperparâmetros da Floresta Randômica

A Tabela 14 apresenta os parâmetros aplicados inicialmente na Floresta Randômica das duas metodologias utilizadas no TCC. Os parâmetros analisados foram: 1) Número de árvores na Floresta Randômica; 2) Número mínimo de amostras para dividir um nó interno; 3) Número mínimo de amostras para cada folha nos nós; 4) O número de características para procurar uma divisão no nó; 5) A profundidade máxima de cada árvore da Floresta Randômica; 6) Número de amostras a serem utilizadas na construção das árvores.

Tabela 14 – Hiperparâmetros utilizados nas Florestas Randômicas

Parâmetro	Série Temporal	<i>Features</i>
1	5	5
2	2	2
3	1	1
4	auto	auto
5	-	-
6	True	True

Fonte: Autora (2021)

A Tabela 15 apresenta os hiperparâmetros da Floresta Randômica após o ajuste dos mesmos.

Tabela 15 – Ajuste dos Hiperparâmetros utilizados nas Florestas Randômicas

Parâmetro	Série Temporal	<i>Features</i>
1	200	2000
2	8	3
3	2	1
4	Auto	Sqrt
5	60	50
6	<i>False</i>	<i>False</i>

Fonte: Autora (2021)

Comparando as Tabelas 12 e 13, observa-se o aumento de árvores e a profundidade das mesmas em cada Floresta Randômica, assim como o número de amostras para dividir um nó interno. O parâmetro 6 foi ajustado para *False*, nos dois modelos, fazendo com que a técnica utilize todo o conjunto de dados para construir a floresta.

A Tabela 16 apresenta o resultado da classificação da Floresta Randômica com os hiperparâmetros ajustados, na fase de teste do algoritmo.

Tabela 16 – Resultado da classificação da Floresta Randômica com os Hiperparâmetros ajustados

Série temporal	<i>features</i>
0,856	0,846

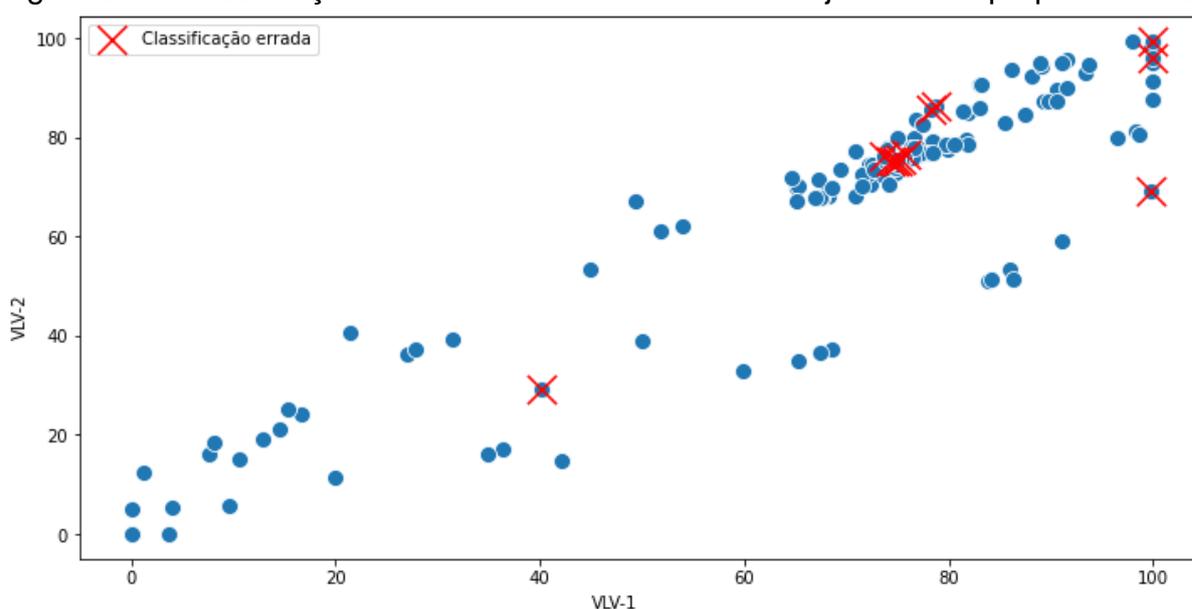
Fonte: Autora (2021)

Observa-se uma melhora no desempenho do modelo em ambos os casos, com série temporal e *features*. Visto que, após o ajuste dos hiperpâmetros obtêm-se Florestas Randômicas mais robustas e que irão analisar um número maior de características do conjunto de dados em cada árvore de decisão dentro do modelo.

A metodologia de aplicação das *features* é mais coerente em relação à série temporal, pois não é vantajoso analisar ponto a ponto da série de dados em busca de falhas para realizar as predições. Por isso utiliza-se a extração de características, em uma determinada faixa de tempo, e cria-se novas variáveis que irão auxiliar na classificação do algoritmo.

A Figura 31 apresenta o resultado da classificação da Floresta Randômica, utilizando as variáveis com maior significância, na Detecção e Diagnóstico das falhas na válvula de controle.

Figura 31 – Classificação da Floresta Randômica com o ajuste dos hiperparâmetros



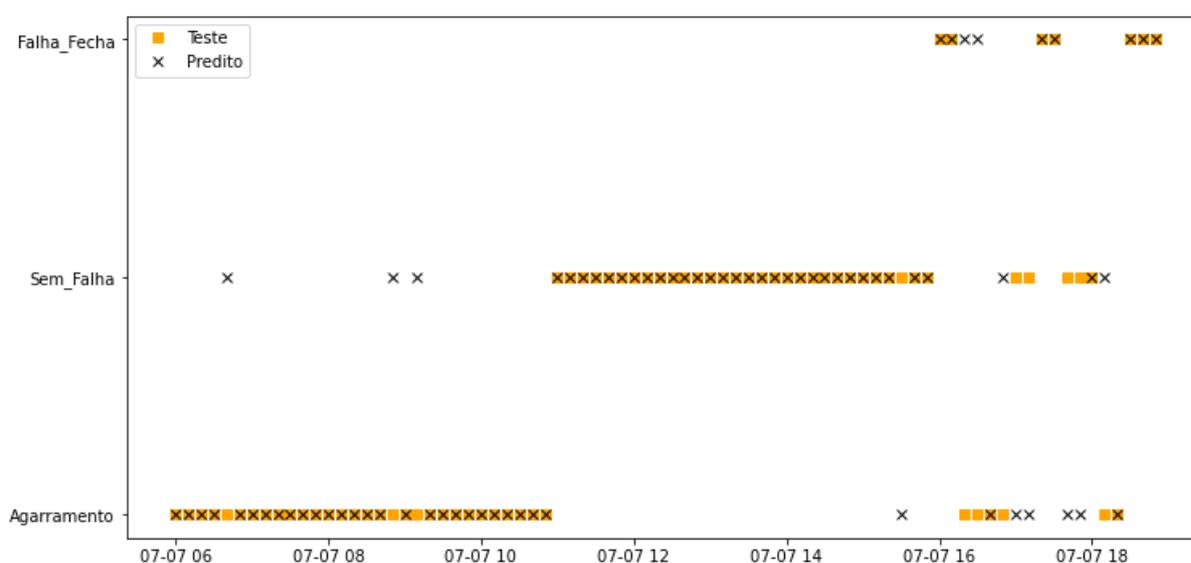
Fonte: Autora (2021)

Em comparação com a Figura 30, é possível notar pontos em que o modelo acertou os resultados que anteriormente tinham sido classificados errados. Os

hiperparâmetros mostram-se importantes para o desempenho do modelo preditivo e caso não seja realizado seu ajuste a Detecção e Diagnóstico de falhas pode ser comprometida.

A Figura 32 apresenta a classificação da Floresta Randômica com o ajuste dos hiperparâmetros em relação aos episódios de falha, agarramento e supressão de pressão, e sem falha. O gráfico relaciona a classificação das falhas pelo modelo com o tempo em , dia, mês e hora.

Figura 32 – Detecção e Diagnóstico das falhas com o modelo de Floresta Randômica



Fonte: Autora (2021)

A Floresta Randômica conseguiu detectar todos os pontos com falha-fecha corretamente, visto que esta falha apresentava o comportamento mais distinto das três classes. Com resposta crítica das variáveis do processo e cavitação da bomba. A falha de agarramento apresenta comportamento similar aos dados sem falha, porém o sistema leva mais tempo para estabilizar. Dessa forma, a classificação destes dois casos torna-se mais complexo e sensível aos modelos o que pode ser observado na Figura 32. O algoritmo errou classificando alguns pontos de agarramento como dados normais do processo, não sendo possível detectar esta falha nestes pontos da operação. Em relação ao diagnóstico das falhas, apenas dois episódios em que ocorreu o agarramento foram classificados como Falha-Fecha (supressão de pressão), demonstrando uma boa previsão do modelo.

6 CONSIDERAÇÕES FINAIS

Neste Trabalho de Conclusão de Curso foram avaliadas técnicas de *Machine Learning* na detecção e diagnóstico de falhas em válvulas de controle, a partir de uma simulação no software UniSim® Design R461. Buscou-se gerar os próprios dados com o software escolhido e o desenvolvimento de gráficos e algoritmos para diferenciar três situações distintas: sistema sem falhas, falha de agarramento e falha de supressão de pressão em uma válvula de controle. E desenvolver a análise e processamento dos dados com a linguagem de programação Python e suas bibliotecas.

A simulação no UniSim® Design R461 promoveu uma experiência positiva no desenvolvimento da simulação e modelagem do processo químico escolhido. Sendo uma ferramenta muito útil para a simulação de falhas em equipamentos industriais, com um módulo que possibilita a aplicação de diversas avarias nas válvulas de controle. Além de proporcionar o dimensionamento dos instrumentos utilizados no controle e monitoramento do vaso de separação e da bomba centrífuga através da obtenção da curva característica da mesma. O dimensionamento do vaso de separação horizontal foi baseado em Novo (1982), obtendo-se um diâmetro de 0,76 m e comprimento de 2,27 m.

A metodologia HAZOP possibilitou identificar as situações de perigo no vaso de separação utilizado na simulação. E identificar a válvula de controle que apresenta maior risco ao processo frente a presença de falhas. Esta válvula foi a VLV-001, visto que as situações de risco envolvendo este equipamento foram mais graves que as restantes.

A análise exploratória dos dados, desenvolvida com a linguagem Python, foi necessária para o entendimento do processo químico simulado computacionalmente, e resposta das variáveis frente a modificações do sistema. A partir dos gráficos obtidos, observa-se a correlação positiva das variáveis do processo, exceto com a abertura da válvula presente na descarga da bomba B-001. A linguagem Python se mostrou uma poderosa ferramenta, com potencial para atender as demandas dentro da análise e processamento de dados. Com bibliotecas robustas para o desenvolvimento dos algoritmos e uma linguagem de fácil entendimento lógico.

O desempenho das técnicas de *Machine Learning* na detecção e diagnóstico das falhas foi significativamente positiva. Todos os modelos utilizados apresentaram boas previsões no processo de treinamento dos algoritmos, com acurácia acima de 50%. Dentre as duas metodologias aplicadas no TCC, os dados onde foi realizada a extração de características apresentou maior sucesso em alcançar o objetivo proposto. A Floresta Randômica foi a técnica que melhor se ajustou aos dados, com ótimas previsões para os cenários de falhas com a metodologia das *features*. O modelo obteve acurácia de 82% na Detecção e Diagnóstico das falhas de agarramento e supressão de pressão na válvula VLV-001. Os resultados obtidos puderam ser otimizados após o ajuste dos hiperparâmetros da Floresta Randômica com uma taxa de acertos de 84%.

Deste modo, o presente TCC atingiu os objetivos propostos, sendo possível consolidar os conhecimentos adquiridos durante a graduação nas áreas de modelagem e simulação de processos químicos, dimensionamento de equipamentos industriais e análises de riscos. Além disso, foram obtidos novos conhecimentos nas áreas de Python, análise de dados e *Machine Learning*, que não são muito explorados durante a graduação.

Em trabalhos futuros recomenda-se estudar a possibilidade de implementação dos algoritmos no controle digital de processos químicos, para a detecção e diagnóstico de falhas em tempo real, com um sistema de monitoramento da operação industrial. Para o modelo de Redes Neurais, sugere-se realizar um estudo mais detalhado deste algoritmo, com a proposta de uma rede mais complexa e robusta, buscando uma metodologia que melhor interprete as situações de falha. Para as Árvores de Decisão propõem-se estudar a aplicação de outros classificadores, como o *Adaboost*, ou ajuste de hiperparâmetros para melhorar as previsões do modelo.

REFERÊNCIAS

AGUIAR, L. A. **Metodologias de análise de riscos APP e HAZOP**. Apostila. Rio de Janeiro, 2008. Disponível em: http://files.visaosegura.webnode.com/200000056-584dc5947a/APP_e_HAZOP.pdf. Acesso em: 20 abr. 2021.

ANALYTICS VIDHYA. **Tree based algorithms**: a complete tutorial from scratch (in R & Python). Gurgaon: Analytics Vidhya, 2016. Disponível em: <https://www.analyticsvidhya.com/blog/2016/04/tree-based-algorithms-complete-tutorial-scratch-in-python/>. Acesso em: 26 mar. 2021.

ANALYTICS VIDHYA. **Let's open de the black box of Random Forests**. Gurgaon: Analytics Vidhya, 2020. Disponível em: <https://www.analyticsvidhya.com/blog/2020/12/lets-open-the-black-box-of-random-forests/>. Acesso em: 11 abr. 2021.

ANJOS, R. C. N. S. **Dimensionamento de vaso separador bifásico horizontal**. 2015. Trabalho de Conclusão de Curso (Graduação em Engenharia de Petróleo) – Universidade Federal do Rio Grande do Norte, Natal, 2015. Disponível em: http://webcache.googleusercontent.com/search?q=cache:gPeZ0cCZgcwJ:arquivos.info.ufrn.br/arquivos/201624908391a03163878ffb4ea4b1f81/TCC_-_2015.2_-_Raquel_Cavalcante_das_Neves_Soares_dos_Anjos.pdf+&cd=1&hl=pt-BR&ct=clnk&gl=br: Acesso em: 26 mar. 2021.

ATTOLINI, M. **Deteção e identificação de falhas em válvulas de controle**. 2014. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014. Disponível em: <https://www.lume.ufrgs.br/handle/10183/109715>. Acesso em: 21 set. 2021.

BARRETO, L. C. *et al.* Aplicação de support vector machine na deteção de falhas em sensores de uma coluna debutanizadora. *In: CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA EM INICIAÇÃO CIENTÍFICA*, 2019, São Paulo. **Anais [...]**. São Paulo: Blucher, 2019. p. 3003-3009. Disponível em: <https://www.proceedings.blucher.com.br/article-details/aplicao-de-support-vector-machine-na-deteco-de-falhas-em-sensores-de-uma-coluna-debutanizadora-31184>. Acesso em: 1 abr. 2021.

BAUGHMAN, D. R.; LIU, Y. A. **Neural networks in bioprocessing and chemical engineering**. San Diego: Editora Academic Press, 1995.

BECK, D. A. C. *et al.* Data Science: acceleration innovation and discovery in chemical engineering. **AIChE Journal**, New York, v. 62, n. 5, p. 1402-1416, 2016. Disponível em: http://depts.washington.edu/maple/pubs/59_Beck_et_al-2016-AIChE_Journal.pdf. Acesso em: 9 mar. 2021.

BREIMAN, L. Random Forest. **Klower Academic Publishers**, Amsterdam, v. 45, n. 1, p. 5-32, 2001. Disponível em: <https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf>. Acesso em: 11 abr. 2021.

BURROWS, W. R. CART regression models for predicting UV radiation at the ground in the presence of cloud and other environmental factors. **Journal of Applied Meteorology and Climatology**, Toronto, v. 36, n. 5, p. 531-544, 1997. Disponível em: https://journals.ametsoc.org/view/journals/apme/36/5/1520-0450_1997_036_0531_crmfpu_2.0.co_2.xml. Acesso em: 1 abr. 2021.

CHIANG, L. H. *et al.* **Fault detection and diagnosis in industrial systems**. London: Editora Springer, 2001.

COSTA, H. R. N. **Aplicação de técnicas de inteligência artificial em processos de fabricação de vidros**. 2006. Tese (Doutorado em Engenharia) – Universidade de São Paulo. Escola Politécnica, São Paulo, 2006. Disponível em: <https://www.teses.usp.br/teses/disponiveis/3/3139/tde-09032007-171929/publico/teserevisada.pdf>. Acesso em: 29 mar. 2021.

CUADRADO, D. A. M. **Utilización de machine learning em la indústria 4.0**. 2019. Dissertação (Mestrado em Engenharia Industrial) – Universidad de Valladolid. Escuela de Ingenierías Industriales, Valladolid, 2019. Disponível em: <https://core.ac.uk/download/pdf/228074134.pdf>. Acesso em: 9 mar. 2021.

CUADROS, M. A. S. L. **Quantificação e compensação de atrito em válvulas de controle pneumáticas**. 2011. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal do Espírito Santo. Programa de Pós-Graduação em Engenharia Elétrica, Vitória, 2011. Disponível em: https://repositorio.ufes.br/bitstream/10/4123/1/tese_2819_TeseDoutoradoMarcoAntoniodeSouzaLeiteCuadros.pdf. Acesso em: 10 mar. 2021.

DAMBROS, J. W. V. **Detecção de agarramento em válvulas de controle para sinais com referência variável**. 2016. Dissertação (Mestrado em Engenharia) – Universidade Federal do Rio Grande do Sul. Programa de Pós Graduação em Engenharia Química, Porto Alegre, 2016. Disponível em: <https://www.lume.ufrgs.br/handle/10183/149795>. Acesso em: 17 abr. 2021.

FERNANDES, A. A. T. *et al.* Leia este artigo de você quiser aprender sobre regressão logística. **Revista de Sociologia e Política**, Curitiba, v. 28, n. 74, p. 1-20, 2020. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-44782020000200207&lng=pt&nrm=iso&tlng=pt. Acesso em: 29 mar. 2021.

FIGUEIRAS, P. R. **Regressão por vetores de suporte aplicado na determinação de propriedades físico-químicas de petróleo e biocombustíveis**. 2014. Tese (Doutorado em Ciências) – Universidade Estadual de Campinas. Instituto de Química, Campinas, 2014. Disponível em: http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/249325/1/Filgueiras_PauloRoberto_D.pdf. Acesso em: 29 mar. 2021.

FISHER. **Control Valve Handbook**. 4. ed., Marshalltown: Emerson Automation Solutions, 2005.

FREEDMAN, P. HAZOP como metodologia de análise de riscos. **Revista Petrotecnica**, Buenos Aires, n. 2, p. 60-64, 2003. Disponível em: <http://biblioteca.iapg.org.ar/archivosadjuntos/petrotecnica/2003-2/hazop.pdf>. Acesso em: 26 mar. 2021.

GARCIA, C. J. **Desenvolvimento de estratégias de integração de processos em tempo real aplicados à produção de GLP**. 2006. Tese (Doutorado em Engenharia química) – Universidade Estadual de Campinas. Programa de Pós-Graduação em Engenharia Química, Campinas, 2006. Disponível em: <http://repositorio.unicamp.br/jspui/handle/REPOSIP/266718>. Acesso em : 29 mar. 2021.

GERCHMAN, M. **Comparação e proposição de métodos de detecção de agarramento em válvulas de controle**. 2013. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/108482/000946346.pdf?sequence=1>. Acesso em : 27 abr. 2021.

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

HASTIE, T. *et al.* **The elements of statistical learning: data mining, inference, and prediction**. 2. ed., Stanford: Editora Springer, 2008.

HAYKIN, S. **Neural Network and learning machines**. London: Editora Pearson, 2008.

HONEYWELL. **Honeywell Process**. New Jersey: Honeywell International. 2020. Disponível em: https://www.honeywellprocess.com/en-US/online_campaigns/connected_plant/Pages/process-simulation.html. Acesso em: 10 mar. 2021.

KAGERMANN, H.; WAHLSTER, W.; HELBIG, J. **Recommendations for implementing the strategic initiative industrie 4.0**. Frankfurt: Acatech, 2013. Disponível em: <https://www.din.de/blob/76902/e8cac883f42bf28536e7e8165993f1fd/recommendations-for-implementing-industry-4-0-data.pdf>. Acesso em: 9 mar. 2021.

KLETZ, T. **Hazop & Hazan: Identifying and assessing process industry**. 4. ed., Boca Raton: CRC Press, 1999.

LEAL, E. S. **Predição de falhas em equipamentos de refrigeração com técnicas de aprendizado de máquina**. 2019. Dissertação (Mestrado em Engenharia de Automação e Sistemas) – Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Automação e Sistemas, Florianópolis, 2019. Disponível em: <https://repositorio.ufsc.br/handle/123456789/204399>. Acesso em: 21 abr. 2021.

MATHIAS, A. C. **Válvulas Industriais, segurança e controle**. 2.ed., São Paulo: Editora Artliber, 2014.

MONTAVON, G.; ORR, G.; MÜLLER, K. R. **Neural Networks: tricks of the trade**. 2. ed., Berlin: Editora Springer, 2012.

MOUTINHO, M. V. B. **Detecção de falhas em internos de válvulas de controle atuando em múltiplas regiões de operação**. 2016. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Espírito Santo. Programa de Pós-Graduação em Engenharia Elétrica, Vitória, 2016. Disponível em: https://repositorio.ufes.br/bitstream/10/9550/1/tese_10163_disserta%C3%A7%C3%A3o_Marcus_Final.pdf. Acesso em: 10 mar. 2021.

NOVO, L. F. **Vasos de pressão: Dimensionamento de tambores/projeto mecânico**. Apostila. Rio de Janeiro, 1982. Disponível em: <https://docero.com.br/doc/8xs15vn>. Acesso em: 29 set. 2021.

PARANHOS, J. F. **Métodos dos modelos termodinâmicos simplificados (MMTS): Uma abordagem eficiente para descrever o equilíbrio líquido – vapor**. 2014. Dissertação (Mestrado em Engenharia Química) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Engenharia Química, Porto Alegre, 2014. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/101174/000929903.pdf?sequence=1>. Acesso em: 17 abr. 2021.

PYTHON. **Python**. Delaware: Python Software Foundation, 2021. Disponível em: <https://www.python.org/about/>. Acesso em: 26 mar. 2021.

PYTHON.ORG. **Python ORG**. Python Package Index. Delaware: Python Software Foundation, 2021. Disponível em: <https://pypi.org/search/?q=bibliotecas>. Acesso em: 26 mar. 2021.

REIS, L. G. **Detecção de defeitos em plantas industriais através de características espectrais do som**. 2021. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Rio de Janeiro. Programa de Pós-Graduação em Engenharia Elétrica, Rio de Janeiro, 2021. Disponível em: <https://www.gta.ufrj.br/ftp/gta/TechReports/Luana21/Luana21.pdf>. Acesso em: 29 set. 2021.

ROJAS, F. D. S; DE SOUZA, M. B. JR. Aplicação de Machine Learning a detecção e diagnóstico de falhas em processos químicos. *In*: CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA, 2018, São Paulo. **Anais [...]**. São Paulo: Blucher, 2018. p. 1213-1216. Disponível em: <https://www.proceedings.blucher.com.br/article-details/aplicao-de-machine-learning-a-deteco-e-diagnostico-de-falhas-em-processo-quimico-28710>. Acesso em: 29 set. 2021.

ROJKO, A. Industry 4.0 concept: background and overview. **International Journal of Interactive Mobile Technologies**, Kassel, v. 11, n. 5, p. 77-90, 2017. Disponível em: <https://online-journals.org/index.php/i-jim/article/view/7072/4532>. Acesso em: 9 mar. 2021.

SALAS, M. V. La Regressión Logística. Una aplicación a la demanda de estúdios universitarios. **Estadística Española**, Madrid, v. 38, n. 141, p. 193-217, 1996. Disponível em: http://metodos-avanzados.sociales.uba.ar/wp-content/uploads/sites/216/2014/03/Regres_log_SALAS-VELASCO.pdf. Acesso em: 11 abr. 2021.

SARAIVA, S. V. **estudo de um controlador preditivo baseado em sistemas inteligentes**. 2017. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) – Universidade Federal de Alagoas, Maceió, 2017. Disponível em: http://repositorio.unicamp.br/jspui/static/aceso_restrito.jsp. Acesso em: 9 mar. 2021.

SEGARAN, T. **Programming collective intelligence: building smart web 2.0**. Sebastopol: Editora O'Reilly, 2007.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning: from theory to algorithms**. New York: Cambridge University Press, 2014. Disponível em: https://content.schweitzer-online.de/static/catalog_manager/live/media_files/representation/zd_std_orig__zd_schw_orig/004/894/425/9781107057135_foreword_pdf_1.pdf. Acesso em: 9 mar. 2021.

SILVA, G. D. **Machine learning no controle preditivo de processos**. 2020. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) – Universidade Federal de Alagoas, Maceió, 2020. Disponível em: <http://www.repositorio.ufal.br/bitstream/riufal/7518/1/Machine%20learning%20no%20controle%20preditivo%20de%20processos.pdf>. Acesso em: 6 mar. 2021.

SOARES, F. D. R. **Técnicas de machine learning aplicadas a inferência e detecção e diagnóstico de falhas em processos químicos industriais em contexto big data**. 2017. Dissertação (Mestrado em Ciências) – Universidade Federal do Rio de Janeiro. Programa de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos, Rio de Janeiro, 2017. Disponível em: <http://186.202.79.107/download/diagnostico-de-falhas-de-processos-quimicos-industriais-em-contexto-big-data.pdf>. Acesso em: 9 mar. 2021.

SOUZA, C. M.; COSTA, M. S. M. **Simulação e otimização de processos contínuos de produção de biodiesel**. 2018. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) – Universidade Federal Fluminense, Niterói, 2018. Disponível em: <https://app.uff.br/riuff/bitstream/1/6993/1/TCC%20Clarissa%20e%20Matheus.pdf>. Acesso em: 10 mar. 2021.

SOUZA, L. A. S. **Noções de Processamento Primário de Petróleo**. Petrobras. Programa Alta Competência. Rio de Janeiro: Petrobras, 2010. Disponível em: <https://www.passeidireto.com/arquivo/29524540/nocoes-de-processamento-primario-de-petroleo>. Acesso em: 20 abr. 2021.

TAFNER, M. A. As redes neurais artificiais: aprendizado e plasticidade. **Revista Cérebro & Mente**, Campinas, v. 2, n. 5, março/maio, 1998. Disponível em: <https://cerebromente.org.br/n05/tecnologia/plasticidade2.html>. Acesso em: 16 de mar. 2021.

TEIXEIRA, A. C. **Detecção e diagnóstico de falhas em sistemas de processos químicos**. Importância do conhecimento de estados intermediários de processos dinâmicos. Desenvolvimento de uma metodologia baseada em redes neurais. 2000. Tese (Doutorado em Engenharia Química) – Universidade Estadual de Campinas. Faculdade de Engenharia Química, Campinas, 2000. Disponível em: <http://repositorio.unicamp.br/jspui/handle/REPOSIP/267640>. Acesso em: 5 mar. 2021.

TURTON, R. *et al.* **Analysis, Synthesis, and Design of Chemical Processes**. 4. ed., Ann Arbor: Editora Pearson, 2001.

VENKATASUBRAMANIAN, V. Drowning in data: informatics and modeling challenges in a data-rich networked world. **AIChE Journal**, New York, v. 55, n. 1, p. 2-8, 2008. Disponível em: <https://aiche.onlinelibrary.wiley.com/doi/epdf/10.1002/aic.11756>. Acesso em: 10 mar. 2021.

ZANINI, A. **Regulação econômica no setor elétrico brasileiro**: uma metodologia para definição de fronteiras de eficiência e cálculo do fator x para empresas distribuidoras de energia elétrica. 2004. Tese (Doutorado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro. Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica, Rio de Janeiro, 2004. Disponível em: <https://www.maxwell.vrac.puc-rio.br/colecao.php?strSecao=resultado&nrSeq=5566@1>. Acesso em: 19 mar. 2021.

ANEXOS

ANEXO A

Análise de Perigos e Operabilidade									
Unidade: Alimentação do processo									
Sistema: Vaso de separação V-001						Equipe: Kamilla, Alexandre, Andressa e Jônathan		Data: 29/09/2021	
Localização do nó: Alimentação a montante do vaso V-001, Vaso V-001, bomba B-001 e vazão de alimentação a jusante do vaso V-001.								Página:1	
Desvio	Causas	Detecção	Salvaguardas	Efeitos	Cat. Freq	Cat. Sev	Cat. Risco	Obs	nº
vazão maior	Válvula VLV-001trancada aberta	LIC-001, LG	By-pass da VLV-001	Aumento do nível no vasoV-001	D	I	Menor	R1: configurar sistema de intertravamento para nível muito altocom parada da bomba de alimentação a montante do vaso V-001	1
			PSV do vaso V-001	Contaminação de líquidona corrente de vapor					
			Alarme de nívelalto	Abertura da PSV					
Vazão menor	Diminuição da vazão de alimentação parao vaso V-001	LIC-001, LG	LIC-001	Diminuição do nível novaso V-001	D	I	Menor	-	2
			FIC-001	Diminuição da pressão no vaso V-001					
			PIC-001	Diminuição da alimentação a jusante doV-001					

Nenhuma vazão	Válvula VLV-001 trancada fechada	LIC-001, LG	By-pass da VLV-001	Diminuição do nível no vaso V-001	D	III	Sério	R2: Configurar sistema de intertravamento para nível muito baixo com parada da bomba B-001	3
			Alarme de nível baixo	Cavitação da bomba B-001					
			FIC-001	Fuga de líquido inflamável com risco de explosão					
Pressão maior	Válvula VLV-002 trancada fechada	PIC-001	By-pass da VLV-002	Aumento da pressão no vaso V-001	D	I	Menor	R3: Instalar sistema de inertização	4
			PSV do vaso V-001	Perda de produto pela PSV					
			PI local						
Pressão menor	Válvula VLV-002 trancada aberta	PIC-001	By-pass da VLV-002	Diminuição da pressão no vaso V-001	D	II	Moderado	R3: Instalar sistema de inertização	5
			PI local	Possível cavitação da bomba B-001					

Nível maior	Fechamento da válvula VLV-003	FIC-001	By-pass da VLV-003	Aumento do nível no vaso V-001	D	I	Menor	R1: configurar sistema de intertravamento para nível muito alto com parada da bomba de alimentação a montante do vaso V-001	6
	Válvula VLV-001 trancada aberta	LIC-001	Alarme de nível alto	Contaminação de líquido na corrente de vapor					
	Parada da bomba B-001		Bomba reserva						
Nível menor	Válvula VLV-001 trancada fechada	LIC-001, LG	By-pass da VLV-001	Diminuição do nível no vaso V-001	D	III	Sério	R2: Configurar sistema de intertravamento para nível muito baixo com parada da bomba B-001	7
			Alarme de nível baixo	Cavitação da bomba B-001					
			FIC-001	Fuga de líquido inflamável com risco de explosão					
Temperatura Maior	Aumento da temperatura da alimentação	Sensor de temperatura local	Malha de pressão da PIC-001	Aumento de pressão no vaso V-001	C	I	Desprezível	R4: Configurar indicação de temperatura na sala de controle	8
				Abertura da PSV					
				Perda de produto na linha do vapor				R3: Instalar sistema de inertização	

Temperatura menor	Diminuição da temperatura da alimentação	Sensor de temperatura local	Malha de pressão da PIC-001	Diminuição da pressão do vaso V-001	C	I	Desprezível	R4: Configurar indicação de temperatura na sala de controle	9
								R3: Instalar sistema de inertização	