

UNIVERSIDADE FEDERAL DE ALAGOAS – UFAL
CENTRO DE TECNOLOGIA - CTEC
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA – PPGEQ

TAINÁ DE BORBOREMA HITTA

**DESENVOLVIMENTO DE MODELO DE INFERÊNCIA DE CCL₄ EM UMA
UNIDADE DE PRODUÇÃO DE MVC**

MACEIÓ

2023

Tainá de Borborema Hita

DESENVOLVIMENTO DE MODELO DE INFERÊNCIA DE CCl_4 EM UMA
UNIDADE DE PRODUÇÃO DE MVC

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Alagoas como requisito para obtenção do título de Mestre em Engenharia Química.

Orientador: Prof. Dr. João Inácio Soletti

Coorientador: Prof. Dr. Rodolfo Junqueira Brandão

Maceió

2023

Catálogo na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

H674d Hita, Tainá de Borborema.
Desenvolvimento de modelo de inferência de CCl₄ em uma unidade de produção de MVC / Tainá de Borborema Hita. – 2023.
58 f. : il. color.

Orientador: João Inácio Soletti.
Coorientador: Rodolfo Junqueira Brandão.
Dissertação (Mestrado em Engenharia Química) – Universidade Federal de Alagoas. Centro de Tecnologia. Maceió, 2023.

Bibliografia: f. 54-58.

1. Analisadores virtuais. 2. Tetracloro de carbono. 3. Monocloreto de vinila. I.
Título.

CDU: 66.095.255

AGRADECIMENTOS

Primeiramente, agradeço à Deus, que permitiu a realização desse projeto. Agradeço de forma especial à minha mãe Angela e ao meu pai Luis, por não medirem esforços em toda minha vida e sempre serem exemplos de determinação. Agradeço imensamente ao meu marido João Guilherme, o meu maior incentivador, pela abdicação e apoio principalmente nesse último ano! Ao meu pequeno grande amor, meu filho Miguel, que também abdicou do nosso tempo juntos para a minha dedicação, e me ensina diariamente sobre o extraordinário no “ordinário”. Ao meu irmão Cauê e toda a minha família (Hita, Borborema e Lessa Rocha), por serem sempre lugar de segurança e amor. A todos os meus queridos amigos, que acolheram minhas preocupações e me apoiaram nas dificuldades Bia, Lu, Pati, Caio, aos “os melhores” e “SWAT”. Agradeço à Comunidade dos Viventes, em especial aos membros internos, que são exemplo diários da busca do Amor.

Sou grata ao Prof. Dr. João Inácio Soletti, que me apoia e incentiva desde a graduação em todos os momentos da minha carreira. Professor Johnny, o senhor é o meu exemplo de Engenheiro Químico, muito obrigada por tanto ao longo desses anos! Ao meu coorientador Prof. Dr. Rodolfo Brandão por toda a disponibilidade, contribuições, pelas discussões e apoio, que foram fundamentais. Agradeço aos meus colegas de trabalho pelo ensinamento diário, em especial ao meu líder, Francisco Brasileiro pela orientação. Também agradeço à minha banca de qualificação pelas críticas, recomendações e incentivo para desenvolvimento do trabalho.

RESUMO

Na produção de Monocloreto de Vinila (MVC), o 1,2-Dicloroetano (1,2-DCE) deve passar por um processo de purificação, já que o processo de craqueamento de 1,2-DCE é altamente suscetível à inibição e deposição por traços de impurezas, o mesmo deve ser mantido em níveis elevados de pureza. Uma das impurezas presentes no DCE é o tetracloreto de carbono, que é um promotor importante da reação de craqueamento, devido a pirólise de DCE ser catalisada principalmente pelo cloro presente na molécula de CCl_4 , entretanto, sua presença também acelera a formação de coque o que causa efeitos indesejáveis no processo como a diminuição do ciclo de operação. Assim sendo, o monitoramento da presença de CCl_4 online na saída do processo de purificação do 1,2-DCE possibilita maior controle da reação de craqueamento, evita perdas de produção, possibilita uma otimização do processo, além de gerar um aumento do tempo de campanha dos fornos de craqueamento. A importância dessa análise em tempo real, a necessidade de controle dessa impureza e as dificuldades de manutenção dos instrumentos analíticos motivaram este trabalho para o desenvolvimento de um modelo de inferência da análise de CCl_4 . O modelo estatístico foi desenvolvido pelo método de aprendizado de máquina supervisionado, a partir das ferramentas da Aspen e Python. Os resultados obtidos indicam caminhos de variáveis de entrada para inferência da composição de CCl_4 , além de um modelo viável de analisador virtual.

Palavras-chave: Analisador Virtual, Tetracloreto de Carbono e Monocloreto de Vinila.

ABSTRACT

In the production of Vinyl Monochloride (VMC), 1,2-Dichloroethane (1,2-DCE) must undergo a purification process, as the 1,2-DCE cracking process is highly susceptible to inhibition and deposition for traces of impurities, it must be maintained at high levels of purity. One of the impurities present in DCE is carbon tetrachloride, which is an important promoter of the cracking reaction, due to the pyrolysis of DCE being catalyzed mainly by the chlorine present in the CCl_4 molecule, however, its presence also accelerates the formation of coke, which causes undesirable effects in the process such as the reduction of the operation cycle. Therefore, online monitoring of the presence of CCl_4 at the output of the 1,2-DCE purification process enables greater control of the cracking reaction, prevents production losses, enables process optimization, and generates an increase in campaign time of the cracking furnaces. The importance of this real-time analysis, the need to control this impurity and the difficulties in maintaining the analytical instruments motivated this work to develop an inference model for the analysis of CCl_4 . The statistical model was developed using the supervised machine learning method, using Aspen and Python tools. The results obtained indicate paths of input variables for inferring the composition of CCl_4 , in addition to a viable virtual analyzer model.

Key words: Virtual Analyzer, Soft Sensor, Carbon Tetrachloride and vinyl monochloride.

Lista de Figuras

Figura 1 – Ilustração de uma coluna de destilação	15
Figura 2 - Fluxograma Simplificado da Purificação do DCE	16
Figura 3 - Fornalha industrial de Craqueamento de DCE.....	17
Figura 4 - Balanço de CCl_4 nas correntes de 1,2 DCE na torre de pesados ...	19
Figura 5 - Estrutura básica de um analisador virtual	21
Figura 6 - Pirâmide de Automação	24
Figura 7 - Esquema de um neurônio artificial	27
Figura 8 - Fluxograma para desenvolvimento do analisador virtual	32
Figura 9 - Fluxograma Simplificado da Purificação do DCE com Variáveis Estudadas	34
Figura 10 - Etapa de Construção do Modelo no Aspen IQ.....	38
Figura 11 - Resultado da Correlação de Pearson	41
Figura 12 – Plotagem Dados de Análises de CCl_4 no Laboratório e no Analisador em Linha.....	43
Figura 13 - Predição e Real para Rede Neural.....	46
Figura 14 - Predição (vermelho) e valores reais (verde)	46
Figura 15 - Predição e Real para PLS.....	47
Figura 16 - Predição e Real para PLS.....	48
Figura 17 - Predição e Real para Fuzzy Não Linear.....	48
Figura 18 - Predição e Real para Fuzzy Não Linear.....	49
Figura 19 - Predição e Real para Modelo Híbrido	49
Figura 20 - Predição e Real para Modelo Híbrido	50
Figura 21 - MLP com uma camada oculta.....	50
Figura 22 - Parâmetros MLP no Python	51
Figura 23 - Representações do Modelo Previsto Real em MLP no Python.....	52

Lista de Tabelas

Tabela 1 - Variáveis do Processo de Purificação do MVC.....	33
Tabela 2 - Ferramentas Computacionais Usadas no Python.....	37
Tabela 3 - Resultado Modelo Usando os Dados de Análise de Laboratório	42
Tabela 4 - Resultado Modelo Usando os Dados do Analisador em Linha	44
Tabela 5 - Resultados para Tipos de Modelos estudados com Dados Analisador de Campo	45
Tabela 6 - Hiperparâmetros Utilizados na Construção dos Modelos de RNA..	51

SUMÁRIO

1.	INTRODUÇÃO	10
2.	OBJETIVO	12
2.1.	Objetivo Geral.....	12
2.2.	Objetivos Específicos	12
3.	REVISÃO BIBLIOGRÁFICA.....	13
3.1.	Processo de Produção de DCE e MVC.....	13
3.1.1.	Oxicloração	13
3.1.2.	Purificação de DCE.....	14
3.1.3.	Craqueamento do DCE	16
3.2.	Comportamento e Estudos de CCl₄.....	17
3.2.1	Contexto Geral e Histórico do CCl₄	17
3.2.2	Tetracloroeto de Carbono no Processo de Pirolise.....	18
3.3.2.2.	Automação Industrial	23
3.3.2.3.	Tecnologias Empregadas no Desenvolvimento	24
3.4.	Aprendizagem de Máquina.....	25
3.5.	Etapas para Desenvolvimento de Modelos de Aprendizado de Máquinas	29
3.6.	Métricas para Validação e Avaliação do Modelo.....	30
4.	METODOLOGIA	32
4.1.	Estudo do Problema: Seleção das Variáveis.....	32
4.2.	Coleta de Dados	35
4.6.	Análise dos Resultados.....	37
5.	RESULTADOS	39
5.1.	Seleção das Variáveis.....	39
5.2.	Estudo com Diferentes Dados	42

5.2.1.	Dados de Análises de Laboratório	42
5.2.2.	Dados de Análises de Analisador em Linha	42
5.3.	Tipos de Modelos Estudados.....	44
5.3.1.	Rede Neural Monotônica	45
5.3.2.	Regressão por Mínimos Quadrados Parciais (PLS).....	46
5.3.3.	Fuzzy Não Linear Regressão por Mínimos Quadrados Parciais...	48
5.3.4.	Híbrido: Regressão por Mínimos Quadrados Parciais e Rede neural artificial.....	49
5.3.5.	MLP (Multi-layer Perceptron regressor) no Python.....	50
6.	CONCLUSÃO	53
7.	REFERÊNCIAS	54

1. INTRODUÇÃO

Na indústria química, o processo de purificação de compostos de interesse é um dos mais importantes, já que a remoção adequada das impurezas impacta diretamente na qualidade do produto e na capacidade de produção. Portanto, com o contínuo aumento da competitividade industrial e a possibilidade de aplicação de novas soluções tecnológicas para redução de custos e aumento de produção, a aplicação de ferramentas de controle e otimização nos processos de separação se tornam essenciais nos ambientes industriais.

Uma das ferramentas mais usadas para contínua otimização dos processos de separação são os simuladores de processo. Estes auxiliam no entendimento do processo, bem como, possibilitam o desenvolvimento de controles de multivariável e de analisadores virtuais online (MARCUCCI, 2020).

Os analisadores virtuais online são recursos que tem ganhado força dentro da indústria química, pois, com estudo de bases de dados de processo e uso de aplicações de simulações e análises de dados, pode-se gerar modelos para inferir análises em tempo real (MARCIEL, 2021).

Portanto, a possibilidade de inferência de composições essenciais para o processo reduz os custos analíticos, mas principalmente possibilita o uso dessas novas variáveis para o controle de processo, aumentando a capacidade produtiva ou qualidade da produção, podendo, muitas vezes, ter um ganho operacional expressivo.

Um estudo realizado por Moraes et al (2014), com o objetivo de determinar variáveis de entrada da coluna de destilação de alta pureza de DCE, identificou que as 5 temperaturas mais sensíveis podem ser entradas para compor um modelo de inferência de composição de produtos de base, possibilitando a construção de analisadores virtuais e a aplicação de controle inferencial.

Na produção de monocloreto de vinila (MVC), matéria prima na produção do termoplástico policloreto de vinila (PVC), o 1,2-dicloroetano (DCE) é introduzido na serpentina de um forno de craqueamento sob a forma de líquido subresfriado, onde é aquecido por convecção dos gases de combustão e por radiação das paredes aquecidas do forno, vaporizado e então craqueado gerando o MVC e ácido clorídrico (HCl) (BRAGA, 2003).

Este DCE craqueado no forno deve estar com alta pureza, acima de 99,5 %peso/peso. Para que seja garantido esta pureza, ele passa por processos de destilação, onde é realizado o controle dos maiores contaminantes da reação de craqueamento, como por exemplo, o CCl_4 (BRAGA, 2003).

Devido ao CCl_4 ser uma fonte abundante de radicais de cloro, sendo também um eficiente catalisador da pirólise de DCE, assim o CCl_4 pode ser utilizado como promotor no processo em estudo. No entanto, este composto também atua como promotor na formação de coque, causando consequências prejudiciais ao processo como: aumento na resistência a transferência de calor proveniente da fornalha, queda de pressão ao longo do reator, diminuição na eficiência do processo e um decaimento na seletividade de MVC (BORSA et al., 1999).

Portanto, o monitoramento da quantidade de CCl_4 nas reações de craqueamento são de grande importância. Atualmente são usadas duas formas de acompanhamento nas indústrias: análises constantes laboratoriais e instrumentação analítica.

As análises laboratoriais não conseguem dar resultados em tempo real, além de terem um custo elevado associado aos recursos humanos qualificados, bem como manutenção dos equipamentos e compra de reagentes. Já a instrumentação online analítica é uma boa solução, porém, os equipamentos analíticos além de terem custos elevados, exigem manutenção e calibração contínua e específica.

Deste modo, o desenvolvimento de modelos estatísticos que possam inferir a concentração de CCl_4 no processo possibilitariam o acompanhamento da variável online, e até a aplicação da mesma nos controles dos processos, garantindo ajustes preditivos nos processos. Desta forma garante-se os ajustes preditivos nos processos de separação e controle da qualidade da matéria prima nos fornos, bem como o controle e aumento da conversão da reação de craqueamento.

2. OBJETIVO

2.1. Objetivo Geral

Desenvolver um modelo estatístico para a concentração do CCl_4 , a partir do comportamento no processo de produção de MVC.

2.2. Objetivos Específicos

- i) Capacitação em técnicas de desenvolvimento de modelos estatísticos do tipo supervisionado;
- ii) Tratamento, usando análises estatísticas, dos dados de entrada e saída dos modelos;
- iii) Desenvolvimento de analisadores virtuais de CCl_4 através da ferramenta do Aspen IQ e Python.
- iv) Comparação do desempenho entre os modelos desenvolvidos;
- v) Validação, a partir de indicadores estatísticos, a representação da inferência desenvolvida comparando com dados de análises reais;

3. REVISÃO BIBLIOGRÁFICA

3.1. Processo de Produção de DCE e MVC

O monômero cloreto de vinila (MVC) é de fundamental importância, pois é a matéria prima base para o policloreto de vinila (PVC), sendo o PVC o segundo plástico mundialmente mais utilizado. Cerca de 96% da produção do MVC é utilizada para a produção de PVC (OLIVEIRA et al, 2015).

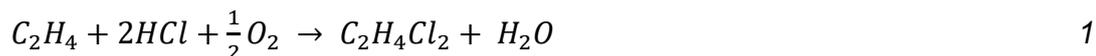
Atualmente, uma das tecnologias mais usadas para produção de MVC é o 1,2-DCE, que é produzido através da cloração direta ou oxicloração de etileno, quando decomposto termicamente produz MVC, HCl e outros subprodutos (OLIVEIRA et al, 2015). Neste tópico 3.1 será apresentado os principais processos na produção de DCE e MVC.

3.1.1. Oxicloração

O processo de oxicloração consiste na combinação do cloreto de metila, de um hidrocarboneto e de oxigênio na presença de um catalisador à base de cloreto de cobre. Rasching desenvolveu o primeiro processo de oxicloração em nível industrial em 1928 para produzir clorobenzeno, que foi, em seguida, hidrolisado a fenol (KIRK-OTHMER, 1999).

O primeiro processo de oxicloração em larga escala para a produção comercial de cloreto de vinila foi colocado em funcionamento em 1958 pela *The Dow Chemical Company*. Esta planta, empregando um reator de leito fixo contendo um catalisador de cloreto cúprico suportado em um material ativo, produziu 1,2-dicloroetano a partir de etileno. A alta temperatura envolvida na reação foi moderada pelo uso de um diluente apropriado. Em uma reação de oxicloração típica, as correntes de gases pré-aquecidas a uma temperatura na faixa de 453 – 473 K são alimentadas em um reator de leito fixo ou fluidizado contendo cloreto de cobre impregnado em alumina ou sílica (KIRK-OTHMER, 1999).

No reator de oxicloração, as correntes gasosas combinadas de etileno (C_2H_4), ácido clorídrico (HCl) e oxigênio do ar fluem para dentro do reator em presença de catalisador de leito fluidizado para produzir DCE de acordo com a seguinte abaixo:



Reações secundárias incluem a oxidação catalítica do etileno para produzir monóxido e dióxido de carbono. São também formados numerosos hidrocarbonetos clorados, incluindo: tetracloreto de carbono (CCl_4), clorofórmio ($CHCl_3$), cloreto de etila (C_2H_5Cl), tricloroetano ($C_2H_3Cl_3$) e outros (FIGUEIREDO, 2009).

A quantidade destes hidrocarbonetos clorados (produtos intermediários) é inferior a 1% da quantidade de DCE produzida, devido ao impacto dos contaminantes na conversão de reação e tempo de operação dos fornos de craqueamento. Portanto, é necessário que o DCE passe por processo de purificação (FIGUEIREDO, 2009).

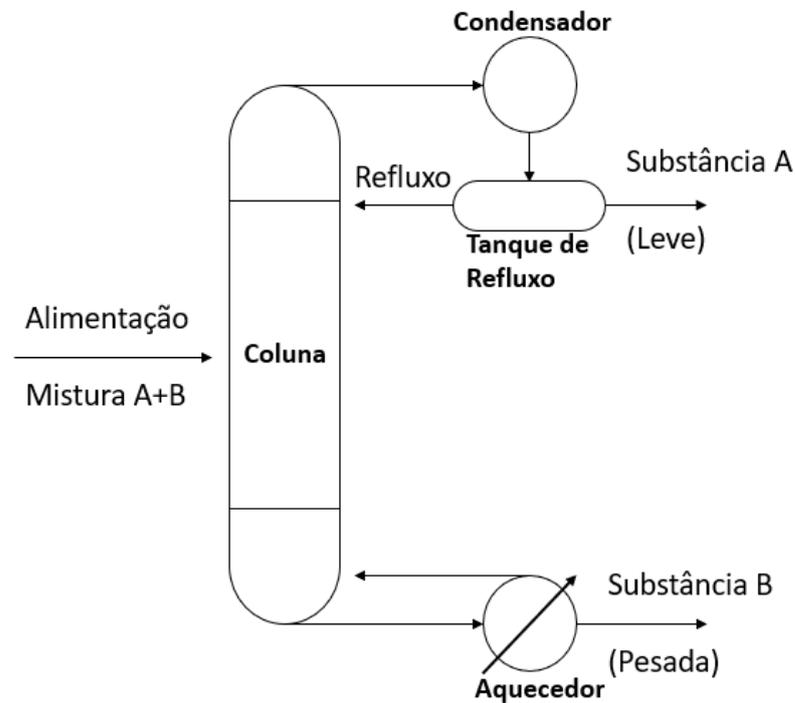
3.1.2. Purificação de DCE

A purificação de DCE consiste no processo de destilação a partir de colunas. A destilação tem como objetivo remover água e hidrocarbonetos clorados de baixo ponto de ebulição (leves) do DCE proveniente da área de oxicloração e remover hidrocarbonetos clorados de alto ponto de ebulição.

Segundo Luyben (1992), a destilação é uma das operações unitárias mais utilizadas na indústria química, a partir dela consegue-se a separação das substâncias que formam uma determinada mistura através da diferença do ponto de ebulição. Portanto, o princípio é o aquecimento da mistura para a ebulição da substância mais volátil, chamada de leve (Silva et al, 2022).

O esquema abaixo, da Figura 1, exemplifica um processo de separação usando coluna de destilação. Onde, para o diagrama abaixo tem-se duas substâncias misturadas chamadas de A e B, onde a substância A é mais volátil que a substância B, portanto, mais leve. O produto é aquecido dentro da coluna, a partir do aquecedor na base da coluna. Os gases formados sobem pela coluna, onde há maior concentração da substância A, portanto, a substância com menor ponto de ebulição. Este produto é condensado e removido, onde parte deste produto é realimentado na coluna. Como a substância A está sendo constantemente eliminada no topo da coluna, a concentração da substância B aumenta no fundo da coluna (LUYBEN, 1992).

Figura 1 – Ilustração de uma coluna de destilação

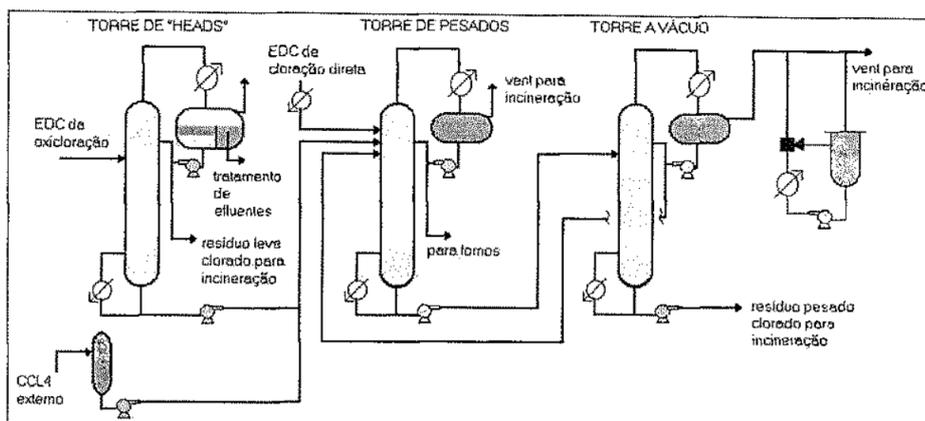


Fonte: Elaborado pelo Autor (2023)

Na tecnologia industrial estudado neste trabalho tem o esquema apresentado na Figura 2, onde o tratamento de DCE possui três colunas de destilação posicionadas em série. A primeira coluna opera com o objetivo de remover os componentes leves da mistura reacional do DCE, portanto, são removidas água e substâncias cloradas de baixo ponto de ebulição como produto de topo do DCE (BRAGA, 2003).

Já na segunda coluna de destilação (torre dos pesados), onde os produtos de alto ponto de ebulição separam-se do DCE que saindo como produto de topo, puro e seco, está especificado para ser craqueado. Nesta torre também é alimentado o DCE não convertido no processo de craqueamento, sendo uma importante corrente de reciclo devido ao seu volume (BRAGA, 2003).

Figura 2 - Fluxograma Simplificado da Purificação do DCE



Fonte: Braga (2003)

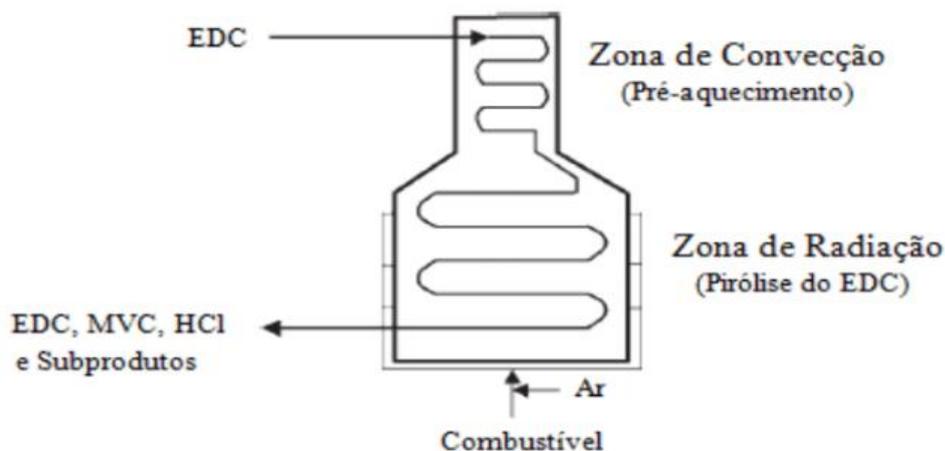
Um estudo realizado por Moraes et al (2014), com o objetivo de determinar variáveis da coluna de destilação de 1,2 -dicloroetano, que afetam as composições dos produtos. E obtiveram que as 5 temperaturas mais sensíveis podem ser entradas para compor um modelo de inferência de composição dos compostos Tetracloreto de Carbono (CCl_4) e Clorofórmio (CHCl_3), possibilitando a construção de analisadores virtuais e a aplicação de controle inferencial (MORAIS et al, 2014).

3.1.3. Craqueamento do DCE

No processo de craqueamento, o DCE puro e seco sofre um processo de craqueamento, ou seja, as moléculas de DCE são submetidas a alta pressão e alta temperatura nas fornalhas, onde irão se “quebrar” formando moléculas menores, que são o MVC e o HCl.

As reações endotérmicas acontecem em uma faixa de temperatura entre 480-550 °C e pressões de 3-30 bar. O dispositivo de reação consiste em um reator tubular inserido em um forno industrial como ilustrado na Figura 3. A primeira parte é a zona de convecção, onde o reagente é pré-aquecido (OLIVEIRA et al, 2014).

A segunda parte consiste na zona de radiação onde ocorrem as reações de craqueamento. O diâmetro do tubo é escolhido de modo a dar uma velocidade superficial do gás entre 10 e 20m/s. O comprimento do reator deve garantir um tempo espacial de 5-30 s (DIMIAN; BILDEA, 2008).

Figura 3 - Fornalha industrial de Craqueamento de DCE

Fonte: Oliveira et al (2014)

O mecanismo de reação da pirólise de DCE em condições industriais é de grande complexidade. Ranzi et al. (1992) propõem um esquema envolvendo mais de 33 reações, com 46 espécies sendo 33 moleculares e 13 radicais. Borsa (1999) desenvolveu um dos mais detalhados modelos cinéticos para pirólise de DCE, envolvendo 71 espécies moleculares, 64 espécies radicais e 818 reações. Já Park et al. (2005) utilizam em seu trabalho um modelo que envolve 108 reações e 48 espécies.

O craqueamento ocorre em uma taxa de conversão entre 50% a 55%, ou seja, cerca de 50% a 55% do DCE alimentado nas fornalhas é convertido em MVC e HCl (BRAGA, 2003).

3.2. Comportamento e Estudos de CCl₄

3.2.1. Contexto Geral e Histórico do CCl₄

A cloração de hidrocarbonetos produz vários produtos químicos, intermediários e solventes clorados que são largamente utilizados no mercado. Dentre os principais hidrocarbonetos clorados, pode-se destacar o clorometano (cloreto de metila), o diclorometano (cloreto de metileno), o triclorometano (clorofórmio) e o tetraclorometano (tetracloroeto de carbono) (KIRK-OTHMER, 1999).

O tetracloroeto de carbono, ou tetraclorometano, CCl₄, a temperatura e pressão ambientes, é um líquido denso, incolor, com odor não irritante e não

inflamável. O produto comercial contém, normalmente, estabilizadores. O tetracloreto de carbono é miscível em muitos líquidos orgânicos, sendo um poderoso solvente para graxas, asfalto e resinas benzênicas (KIRK-OTHMER, 1999).

O CCl_4 foi o primeiro produto químico orgânico produzido industrialmente em larga escala. Em 1890, o processo de fabricação comercial estava sendo investigado pela United Alkali Company, na Inglaterra. No mesmo ano, ele foi produzido na Alemanha e exportado para os Estados Unidos. A produção em larga escala deste composto nos Estados Unidos iniciou-se em 1907 (DANTAS, 2005).

Durante a primeira guerra mundial, a produção dos Estados Unidos aumentou enormemente em virtude da sua utilização na indústria de borracha. Na segunda grande guerra mundial, a demanda requerida pelo mercado estimulou a produção do CCl_4 e iniciou a sua utilização como matéria-prima para a produção dos clorofluorometanos, que constitui, hoje, a sua principal aplicação (DANTAS, 2015).

O tetracloreto de carbono era formalmente usado para desengordurar metais, como fluido de limpeza a seco e fluido de extintores de incêndio, dentre outras aplicações. Durante a década de cinquenta, a demanda desse produto como matéria-prima para a produção dos clorofluorcarbonos aumentou e, como consequência, houve um crescimento contínuo da utilização deste produto no mercado (DANTAS, 2005).

3.2.2. Tetracloreto de Carbono no Processo de Pirolise

No processo de craqueamento do DCE, o CCl_4 tem papel de catalisador, mas também como promotor na formação de coque. A produção do coque é negativa ao processo pois causa o aumento na resistência a transferência de calor proveniente da fornalha, queda de pressão ao longo do reator, diminuição na eficiência do processo e um decaimento na seletividade de MVC (Borsa et al., 1999).

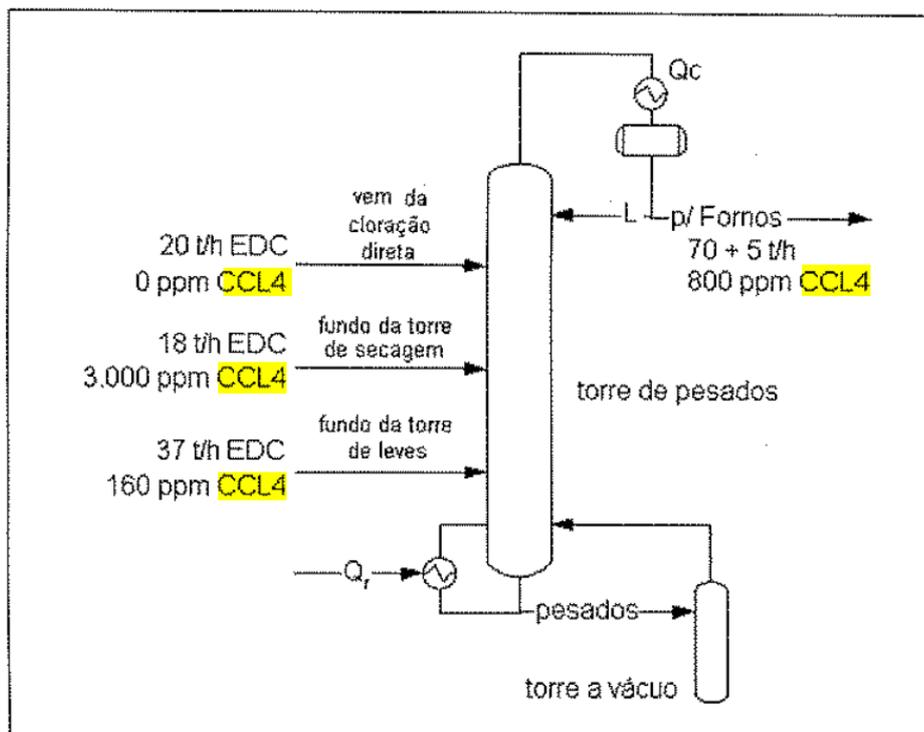
Segundo estudo realizado por Oliveira et al (2015), mostraram que os resultados da adição de pequenas quantidades de CCl_4 em uma determinada faixa de concentração (0-100 ppm), mostraram um aumento da conversão de

DCE no processo. Dessa forma, foi identificado que é possível um aumento de mais de 12% aumentando a concentração de 0 para 100 ppm, o que resulta em um aumento de 13,37% na quantidade de MVC produzido. Porém, por consequência ocorre um aumento da formação de coque, acarretando a redução do tempo de operação dos fornos, cerca de 10 meses a menos quanto operado sem adição de promotor.

Neste estudo de Oliveira et al (2015), a concentração otimizada de CCl_4 no craqueamento do DCE é em torno 50 ppm. Pois, garante uma relação de maior conversão e tempo de campanha do forno. Portanto, mostrando a necessidade da medição para controle da concentração de CCl_4 em processo industrial.

Já de acordo com Braga (2003), no processo de craqueamento de 1,2-DCE, o objetivo final do controle de CCl_4 é obter uma concentração deste componente de cerca de 800 ppm na alimentação dos fornos de pirólise. Um balanço de CCl_4 , na coluna de pesados, que alimenta os fornos, é mostrado na Figura 44 e indica que o produto de base da torre de secagem deveria conter cerca de 3.000 ppm.

Figura 4 - Balanço de CCl_4 nas correntes de 1,2 DCE na torre de pesados



Fonte: Braga (2003)

3.3. Análises Químicas em Processos Industriais

Comumente, na indústria, quando há necessidade de controlar e acompanhar alguma composição química é necessário realização de análises no laboratório por um analista capacitado. Entretanto, desta maneira há um grande tempo de resposta para obter a informação ou impossibilidade de análises com alta frequência, tornando impossível utilizar desta forma de informação para controles automáticos.

Entretanto, duas outras opções são possíveis e consideradas para obter informações de concentração química nos fluídos de processo, são elas: Analisadores em Linha e Analisadores Virtuais.

3.3.1. Analisadores em Linha

Analisadores em linha são instrumentos utilizados para medições de variáveis do processo. Entre suas principais vantagens estão a disponibilização de resultados em tempo real, o que permite a atuação/intervenção rápida no processo, e a minimização de procedimentos de coletas de amostras em campo. Exemplos de analisadores incluem medidores de pH, condutívimetros e cromatógrafos (KROL; ZABIEGALA; NAMIESNIK, 2010).

Os analisadores em linha fornecem uma resposta mais rápida da composição dos produtos em comparação às análises laboratoriais. Porém, estes estão sujeitos a frequentes interferências e contaminações devido às substâncias utilizadas nas correntes reais, que agredem, contaminam e comprometem o funcionamento do equipamento, exigindo manutenções mais frequentes, período em que informações importantes deixam de ser coletadas (MASSA, 2017).

3.3.2. Analisador Virtual

Portanto, quando comparado dos analisadores em linha aos virtuais, segundo Jianxul e Hiuhe (2012), os analisadores virtuais têm maior confiabilidade e precisão nas estimativas, principalmente em faixas extremas de operação.

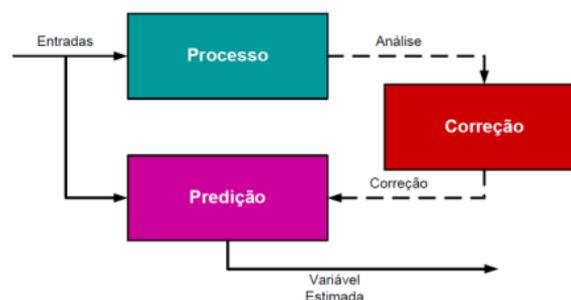
Os analisadores virtuais podem ser definidos como modelos matemáticos desenvolvidos para estimar variáveis de interesse, e podem ser uma alternativa viável aos analisadores on-line e às análises laboratoriais, que demandam alto custo devido à aquisição e manutenção dos equipamentos, bem como a utilização de mão-de-obra qualificada, e maior tempo de análise, a depender da técnica empregada (NAGAI; ARRUDA, 2004).

Sensor virtual ou analisador virtual é o termo usado para definir um sistema que por meio de entradas conhecidas, infere um valor para uma determinada variável de processo que não é mensurada fisicamente. Sendo dessa forma capaz de estimar variáveis que são difíceis ou até que são incapazes de serem medidas em tempo real (GUEDES, 2020).

Segundo Facchin (2021), a Figura 5 mostra os componentes de um analisador virtual:

- a) O bloco Processo consiste nos dados de processo provenientes de instrumentos de medição disponíveis;
- b) O bloco Predição contém o modelo matemático que relaciona as variáveis de processo com a propriedade a ser estimada;
- c) O bloco Correção é composto por uma estratégia de ajuste do modelo a partir dos resultados de laboratório obtidos.

Figura 5 - Estrutura básica de um analisador virtual



Fonte: Extraído de (Facchin, 2021)

Uma das principais motivações para utilização de um sensor virtual na predição da composição de um processo industrial é a econômica. Além

do menor custo de manutenção, quando comparado aos analisadores em linha, também a rapidez de resposta do sensor virtual (WERLE, 2012).

Uma das principais aplicações da construção de analisadores virtuais na indústria é o seu uso em sistemas de controle de processos voltados para a otimização, como os sistemas de controle avançado; que frequentemente necessitam de informações em tempo real sobre propriedades de diversas correntes da planta (CAMPOS et al., 2013).

3.3.3. Condições Importantes no Desenvolvimento dos Sensores Virtuais

Segundo (Campos et al., 2013), a obtenção de um bom desempenho dos analisadores virtuais envolve muitas decisões e escolhas vinculadas a vários aspectos, como:

- Finalidade do analisador virtual;
- Os dados requeridos para uma boa estimativa da propriedade;
- Quais as variáveis que irão compor o modelo;
- Estrutura e características do modelo ou algoritmo a ser utilizado;
- Seleção e validação do modelo escolhido;
- Mecanismos de ajuste;
- Acompanhamento do desempenho e manutenção frequente do algoritmo ou modelo.

Além disso, para um bom desempenho do analisador virtual, deve-se escolher variáveis para composição do modelo obedecendo alguns critérios como variáveis independentes entre si, que influenciem de fato a propriedade que se deseja inferir; conhecimento do processo que se está estimando o comportamento (CAMPOS et al., 2013).

Adicionalmente, Segundo Campos et al (2013) espera-se que as fontes geradoras das variáveis escolhidas para compor o modelo em questão estejam provendo dados coerentes com a realidade. Ou seja, no caso de o modelo ser composto por várias informações de sensores da planta, como sensores de temperatura, pressão, vazão etc., é essencial que eles estejam calibrados, e sua manutenção se dê com frequência, pois em caso de falha ou incoerência de algum deles, o analisador virtual também

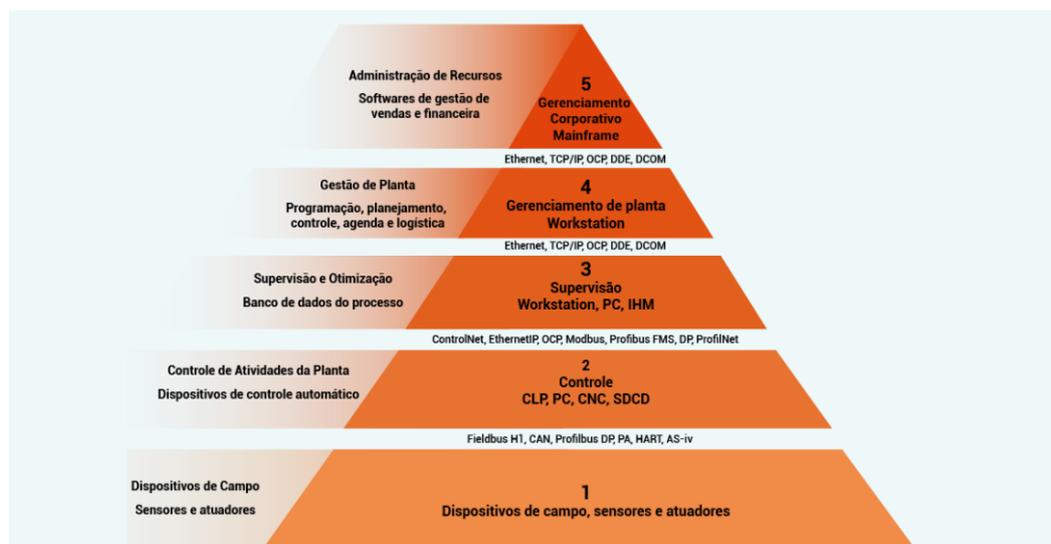
terá seu valor gerado equivocadamente. Portanto, a necessidade de tratar com planos de manutenção diferenciados para os instrumentos utilizados em sensores virtuais online.

Também, sua aplicação só é possível com maior capacidade e infraestrutura de automação industrial, tais como, sistema básico de controle, registros das medições com frequência aceitável a variabilidade e tempo operacional, análises com precisão e exatidão dos dados amostrados, infraestrutura de comunicação e segurança entre a camada de automação de controle básico e supervisão.

3.3.4. Automação Industrial

As redes industriais são formas automatizadas de comunicação usadas para gerenciar os processos industriais. Elas são utilizadas em equipamentos como atuadores, computadores, máquinas, sensores e interfaces que transmitem informações/dados entre si. O controle de informações torna-se muito necessário visto o grande fluxo de atividades realizadas diariamente em uma indústria (CÉSAR, 2018).

Segundo SILVA (2017), a pirâmide de automação industrial, como ilustrado na Figura 6, é uma representação visual dos níveis de automação presentes nas indústrias. Ela visa organizar de uma forma hierárquica os cinco níveis do trabalho presentes no setor industrial e mostrar como todos esses níveis estão relacionados uns com os outros (FROES et al, 2021).

Figura 6 - Pirâmide de Automação

Fonte: Froes et al (2021)

3.3.5. Tecnologias Empregadas no Desenvolvimento

Sobre as tecnologias empregadas para desenvolvimento dos Sensores virtuais, encontram-se na literatura três classes distintas de sensores com modelos empregados, os quais podem ser: modelagem caixa branca ou princípios fundamentais, modelagem caixa preta ou empírica e os modelos híbridos (WERLE, 2012).

Os modelos fenomenológicos, ou caixa branca, são construídos a partir de equações que representam princípios fundamentais como as leis da física, da termodinâmica ou da conservação; sendo, portanto, fundamentados em conceitos teóricos. Já para a obtenção dos modelos empíricos, ou identificação de sistemas, são utilizadas relações matemáticas que buscam representar as relações de causa e efeito entre as variáveis de entrada e a propriedade que se deseja inferir e cujos parâmetros são determinados a partir de dados (RODRIGUES, 2014).

Em um analisador virtual, a modelagem empírica objetiva construir modelos matemáticos de processos dinâmicos a partir de dados experimentais (como medidores de vazão, pressão, temperatura, dentre outros). Estes modelos têm grande aplicabilidade, pois apresentam a vantagem de não demandar conhecimento fenomenológico do processo; interessando-se apenas nas relações entre as entradas e saídas dele

(MASSA, 2017). A forma de modelagem empírica ou híbrida é uma aplicação do que se chama aprendizagem de máquina.

3.4. Aprendizagem de Máquina

A Aprendizagem de Máquina é um subconjunto da Inteligência Artificial (IA), onde é referenciada em inglês como machine learning. Ela se concentra na construção de sistemas que aprendem, ou melhoram o desempenho, com base nos dados que consomem. Com aprendizagem da máquina, é possível identificar padrões a partir de um volume grande de dados, estabelecendo assim modelos de aprendizagem (ESCOVEDO; KOSHIYAMA, 2020).

Em geral, dois tipos principais de algoritmos de machine learning são usados hoje: aprendizado supervisionado e aprendizado não supervisionado (ORACLE, SD).

3.4.1. Aprendizado Não-Supervisionado

Em alguns casos, apesar de existir o objetivo da tarefa desejada, os resultados não são conhecidos e, portanto, não se tem os rótulos para passar ao modelo. Este modelo aprende a executar uma tarefa a partir de dados não-rotulados (sem um resultado conhecido), apenas com base em suas características e padrões semelhantes, ou seja, o modelo deduz estruturas a partir de uma amostra do problema (ORACLE, SD).

Aprendizado não-supervisionado, destacam-se duas técnicas: associação e clusterização (VICERI, 2020). A associação permite o descobrimento de regras e correlações, identificando conjuntos de itens que frequentemente ocorrem juntos. Já na clusterização (ou agrupamento) o conjunto todo em análise sofre segmentações em vários grupos, com base nas semelhanças encontradas. É uma técnica que permite dividir automaticamente um conjunto de dados em grupos de acordo com medidas de similaridade ou de distância. Existem diversas fórmulas para se obter medidas de similaridade, dentre elas podemos destacar o método do cosseno e a correlação de Pearson (VICERI, 2020).

3.4.2. Aprendizado Supervisionado

Nesta abordagem o modelo aprende a executar uma tarefa a partir de exemplos rotulados, ou seja, a partir das respostas corretas, que de alguma forma devem ser passadas para o modelo. O Aprendizado Supervisionado consiste em duas classes de algoritmos: Classificação e Regressão (VICERI, 2020).

Na classificação, o objetivo é identificar a qual categoria pertence uma determinada amostra do problema. Exemplos de técnicas de classificação são: Árvores de Decisão, Algoritmo de Naïve Bayes e redes neurais (VICERI, 2020).

Diferentemente da classificação, na regressão, a ideia é prever um valor numérico; ou em outros termos: identificar uma categoria em escala contínua. Alguns exemplos de técnicas de regressão são: regressão linear, regressão logística e as redes neurais (VICERI, 2020).

3.4.3. Tipos de Algoritmos de Aprendizado

3.4.3.1. Redes Neurais Artificiais

Redes Neurais Artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência (CARVALHO, 2009).

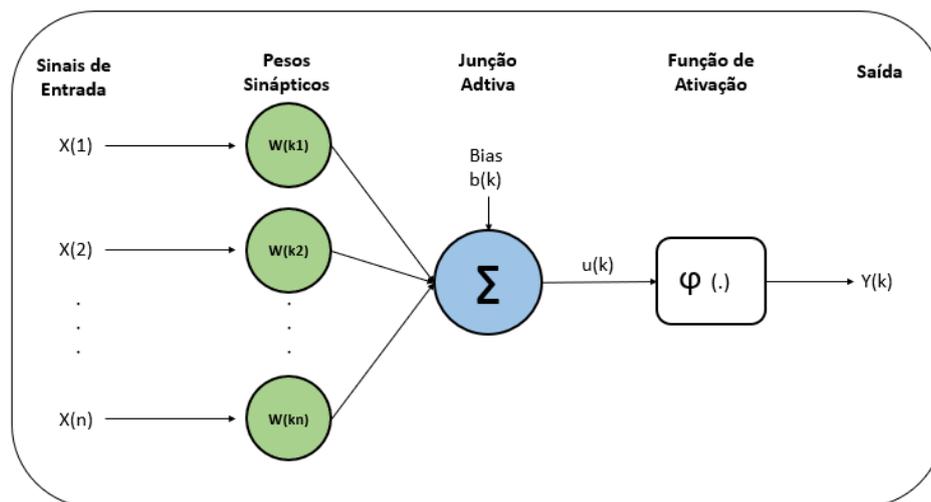
Uma rede neural artificial é composta por várias unidades de processamento, cujo funcionamento é bastante simples. Essas unidades, geralmente são conectadas por canais de comunicação que estão associados a determinado peso (W). As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. O comportamento inteligente de uma Rede Neural Artificial vem das interações entre as unidades de processamento da rede, como pode ser exemplificado na Figura 7 abaixo (CARVALHO, 2009).

A operação de uma unidade de processamento, proposta por McCulloch e Pitts em 1943, pode ser resumida da seguinte maneira:

- Sinais são apresentados à entrada;
- Cada sinal é multiplicado por um número, ou peso, que indica a sua influência na saída da unidade;

- É feita a soma ponderada dos sinais que produz um nível de atividade;
- Se este nível de atividade exceder um certo limite (*threshold*) a unidade produz uma determinada resposta de saída.

Figura 7 - Esquema de um neurônio artificial



Fonte: Adaptado de Heykin (2001)

Na geração de RNA há possibilidade de utilização de Hiperparâmetros. Alguns hiperparâmetros importantes das RNA incluem: a quantidade de camadas, o número de neurônios em cada camada, a função de ativação, o otimizador e a taxa de aprendizagem. A quantidade de neurônios e o número de camadas definem as diferentes arquiteturas da RNA. A função de ativação é aplicada aos neurônios da rede para introduzir não-linearidades no modelo.

3.4.3.2. Regressão Parcial de Mínimos Quadrados (PLS)

Regressão parcial de mínimos quadrados (PLS) é uma técnica que reduz os preditores a um conjunto menor de componentes não correlacionados e efetua regressão de mínimos quadrados para esses componentes no lugar dos dados originais. A regressão PLS é particularmente útil quando os preditores são altamente colineares ou quando existem mais preditores do que observações e a regressão de mínimos quadrados ordinários gera coeficientes com erros padrão altos ou

que falha totalmente. A PLS não assume que os preditores são fixos, ao contrário da regressão múltipla. Isto significa que os preditores podem ser medidos com erro, tornando a PLS mais robusta à incerteza da medição (MINITAB, SD).

Na regressão PLS, a ênfase está em desenvolver modelos preditivos. Por isso ela não é muito usada para filtrar variáveis que não sejam úteis para explicar a resposta.

Ao contrário da regressão de mínimos quadrados, a PLS pode ajustar as múltiplas variáveis de resposta em um único modelo. A regressão PLS ajusta as múltiplas variáveis de resposta em um único modelo. Como a regressão PLS modela as variáveis de resposta de forma multivariada, os resultados podem diferir significativamente daqueles calculados para as variáveis de resposta individual. Deve-se modelar várias respostas separadamente somente se as respostas não estiverem correlacionadas (MINITAB, SD).

3.4.3.3. Fuzzy Não-Linear

O conceito de sistemas *fuzzy*, mais precisamente a lógica *fuzzy*, foi introduzido por Zadeh (1965). Colocados inicialmente como uma extensão da lógica tradicional, ele se baseia em que uma proposição lógica não é necessariamente verdadeira ou falsa, mas possui, de fato, graus de verdade. Na teoria clássica dos conjuntos, o conceito de pertinência de um elemento a um conjunto é bem definido. Sendo assim, para um conjunto A contido em um universo X, os elementos deste universo simplesmente pertencem ou não àquele conjunto.

Um sistema *fuzzy* tem como objetivo modelar o modo aproximado de raciocínio, permitir o desenvolvimento de sistemas que imitam a habilidade humana em tomar decisões racionais em um ambiente de incerteza e imprecisão, capturando informações imprecisas da linguagem natural, convertendo-as para a forma numérica (MENDEL, 1999).

A lógica *fuzzy* pode ser aplicada atualmente em diversas áreas devido sua característica de lidar com problemas reais em um raciocínio próximo do humano. Abaixo segue uma lista de algumas características de sistemas

onde a aplicação da lógica *fuzzy* é necessária ou benéfica McNeil e Thro (1994):

- sistemas complexos que são difíceis ou impossíveis de modelar;
- sistemas controlados por especialistas;
- sistemas com entradas e saídas complexas e contínuas;
- sistemas que se utilizam da observação humana como entradas ou como base para regras;

3.5. Etapas para Desenvolvimento de Modelos de Aprendizado de Máquinas

Segundo Escovedo e Koshiyama (2020) as principais etapas para o desenvolvimento de modelos de machine learning são detalhadas abaixo:

1. Entender o problema e definir objetivos – nesta etapa é importante estudar e entender o problema e objetivo do trabalho, avaliando variáveis de impacto, informações necessárias, modelos matemáticos que já existam, correlações físicas, literatura desenvolvida em outros estudos e se há métodos experimentais;
2. Coletar e analisar os dados – em seguida é necessário coletar e analisar dados adequados para o problema para poder prepará-los, pois, é possível esses dados iniciais terem informações incompletas ou inconsistentes.
3. Pré-processamento – Existem diversos recursos matemáticos para tratamento de dados, como médias móveis, correção de outliers, remoção de períodos não representativos ao que se deseja modelar etc.
4. Modelagem – consiste em elencar e gerar os modelos possíveis e passíveis para cada tipo de problema, estimar os parâmetros que compõem os modelos baseando-se nas variáveis pré-processadas na etapa de tratamento dos dados e avaliar os resultados de cada modelo, usando métricas de comparação.
5. Pós-processamento – deve-se combinar e avaliar os resultados obtidos dos modelos avaliando o resultado com ação real, procurando entender a representatividade do modelo.

6. Implantação do Modelo e Geração de Valor – Implementação do modelo e busca-se levantar a geração de valor da implementação tanto qualitativamente quanto quantitativamente.

3.6. Métricas para Validação e Avaliação do Modelo

As métricas de validação são utilizadas para analisar a qualidade dos modelos de machine learning. Ou seja, traz a informação do desempenho do modelo em dados desconhecidos. Para Modelos Supervisionados de Classificação as métricas mais conhecidas são acurácia, F1 Score e precisão (FILHO, SD).

Acurácia é a métrica mais simples, ela representa o número de previsões corretas do modelo. Trata-se de uma métrica bastante útil quando os dados estão balanceados, vai dar uma visão geral do quanto o modelo está identificando as classes corretamente (FILHO, SD).

O F1 Score é uma média harmônica entre precisão e recall. Ela é muito útil quando você possui um dataset com classes desproporcionais, e o seu modelo não emite probabilidades. Isso não significa que não possa ser usada com modelos que emitem probabilidades, tudo depende do objetivo de sua tarefa de machine learning. Em geral, quanto maior o F1 score, melhor (FILHO, SD).

Número de exemplos classificados como pertencentes a uma classe, que realmente são daquela classe (positivos verdadeiros), dividido pela soma entre este número, e o número de exemplos classificados nesta classe, mas que pertencem a outras (falsos positivos).

Já para Modelos de Regressão as principais usadas são MSE (Mean Squared Error), MAE (Mean Absolute Error) e R-quadrado (AZANK, 2020).

Talvez a MSE seja a mais utilizada, esta função calcula a média dos erros do modelo ao quadrado. Ou seja, diferenças menores têm menos importância, enquanto diferenças maiores recebem mais peso.

O MAE é bastante parecido com MSE, em vez de elevar a diferença entre a previsão do modelo, e o valor real, ao quadrado, ele toma o valor absoluto. Neste caso, em vez de atribuir um peso de acordo com a

magnitude da diferença, ele atribui o mesmo peso a todas as diferenças, de maneira linear. (AZANK, 2020)

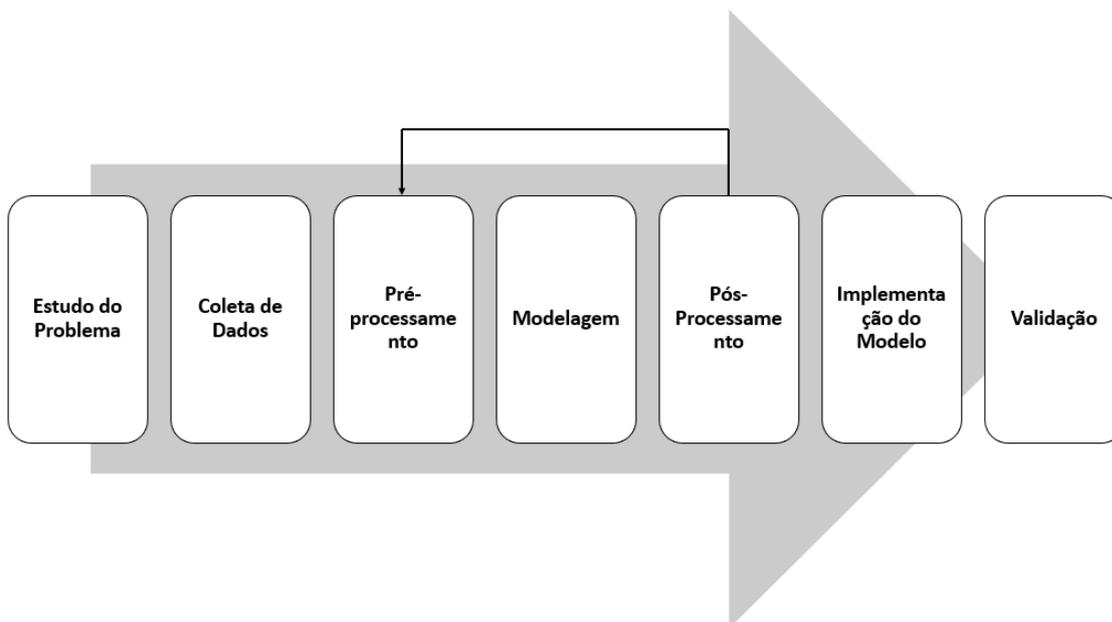
Já o R^2 é muito utilizada em modelos de regressões da área de finanças, o R^2 , ou Coeficiente de Determinação, é uma métrica que visa expressar a quantidade da variância dos dados que é explicada pelo modelo construído. Em outras palavras, essa medida calcula qual a porcentagem da variância que pôde ser prevista pelo modelo de regressão e, portanto, nos diz o quão “próximo” as medidas reais estão do nosso modelo (AZANK, 2020).

O valor do seu R^2 varia de 0 a 1 e geralmente é representado em porcentagem. Por exemplo, um $R^2 = 70\%$ nos dizem que 70% da variância de nossos dados pode ser explicada pelo modelo construído, enquanto os outros 30%, teoricamente, se tratariam de uma variância residual (AZANK, 2020).

4. METODOLOGIA

A metodologia utilizada para desenvolvimento deste trabalho é baseada nos clássicos trabalhos que envolvem a aplicação de *machine learning* (ESCOVETO; KOSHIYAMA, 2020), conforme ilustrado na Figura 8.

Figura 8 - Fluxograma para desenvolvimento do analisador virtual



Fonte: Elaborado pelo Autor (2023)

4.1. Estudo do Problema: Seleção das Variáveis

Foram levantadas possíveis variáveis que podem ter correlação com composição de CCl_4 . Foram levantadas todas as medições existentes nas colunas de destilação (temperatura, nível, vazão e pressão), que podem descrever o perfil das colunas e, portanto, estas podem impactar na presença do tetracloreto de carbono. Segue na Tabela 1 todas as variáveis de processo estudadas dentro deste trabalho:

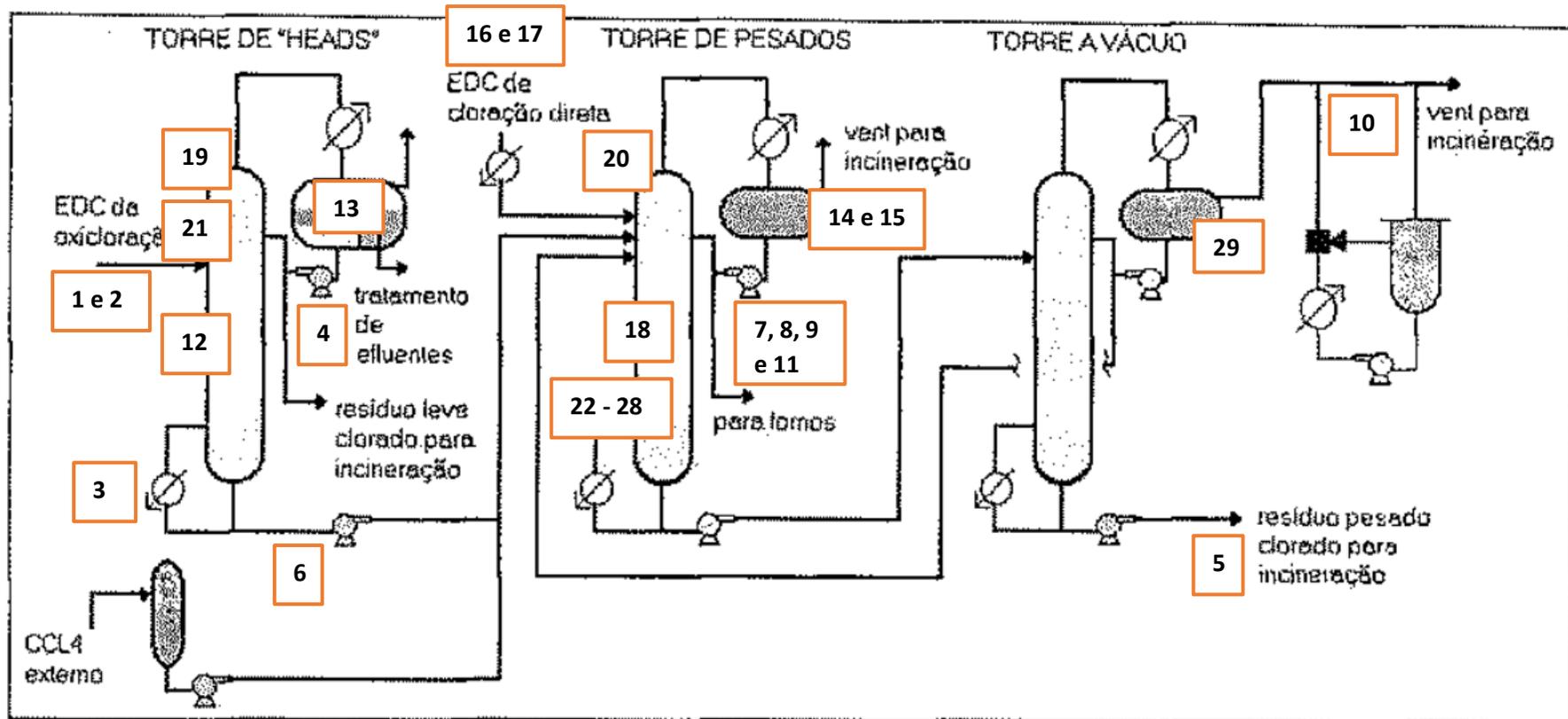
Tabela 1 - Variáveis do Processo de Purificação do MVC

Nº	VARIÁVEIS
1	VAZÃO DE EDC IMPURO
2	VAZÃO DE EDC DA OXYCLORAÇÃO
3	VAZÃO NO PERMUTADOR DA COLUNA 01
4	VAZÃO DE REFLUXO DA COLUNA 01
5	VAZÃO HCU TANQUE
6	VAZÃO DA COLUNA 01 DE FUNDO
7	DESTILADO DA COLUNA 02 – Válvula
8	VAZÃO EDC DA COLUNA 02
9	DESTILADO DA COLUNA 02
10	VAZÃO DESTILADO COLUNA 03 TOPO
11	VÁLVULA PURIFICAÇÃO PARA CRAQUEAMENTO
12	NÍVEL DA COLUNA 01
13	NÍVEL VASO CONDENSADO COLUNA 01
14	NÍVEL VASO CONDENSADO COLUNA 02
15	NÍVEL VASO CONDENSADO COLUNA 02
16	NÍVEL TQ EDC A
17	NÍVEL TQ EDC B
18	DIFERENCIAL DE PRESSÃO COLUNA 02
19	PRESSÃO COLUNA 01 NO TOPO
20	PRESSÃO NO TOPO DA COLUNA 02
21	TEMPERATURA TOPO COLUNA 01
22	TEMPREA13_1 – TEMPERATURA COLUNA 02
23	TEMPREA13_2- TEMPERATURA COLUNA 02
24	TEMPREA13_3 - TEMPERATURA COLUNA 02
25	TEMPREA13_4- TEMPERATURA COLUNA 02
26	TEMPREA13_5- TEMPERATURA COLUNA 02
27	TEMPREA13_6- TEMPERATURA COLUNA 02
28	TEMPREA13_7- TEMPERATURA COLUNA 02
29	NÍVEL VASO CONDENSADO COLUNA 03

Fonte: Elaborado pelo Autor (2023)

As numerações das respectivas variáveis da Tabela 1 estão representando na Figura 9 os pontos de medição do processo de cada uma.

Figura 9 - Fluxograma Simplificado da Purificação do DCE com Variáveis Estudadas



Fonte: Adaptado de Braga (2003)

4.2. Coleta de Dados

A coleta de dados foi realizada a partir de *scripts* desenvolvidos no projeto em SQL, para baixar dados reais do banco já existentes do historiador de processos. Alguns novos dados não historiados no mesmo banco de dados foram migrados para que os dados sejam coletados no mesmo formato.

Para este trabalho, alguns tipos de dados de processo estavam disponíveis:

- a. Um período de 7 anos de dados a cada 1 minuto das variáveis citadas na Tabela 1;
- b. Um período de 7 anos de dados a cada 2 dias das Análises de CCL₄ do Laboratório;
- c. Um período de 7 anos de dados a cada 5 min das análises de CCL₄ no analisador em linha.

Portanto, neste trabalho foram realizados estudos dados de laboratório e dados do analisador, bem como com restrição de períodos para entender quais dados conseguiriam representar estatisticamente o modelo com maior correlação.

4.3. Pré-Processamento

Na etapa de pré-processamento, foi realizada a identificação dos dados espúrios, avaliação da dinâmica de cada variável e impacto na variável de saída.

O tratamento dos dados espúrios consistiu em eliminar dados fora do range de operação de cada variável de entrada, além de eliminar períodos em que as colunas estavam fora de operação. A condição usada para eliminar períodos de planta parada foram as variáveis de vazão de alimentação das colunas e vazão de produção da planta.

O entendimento da dinâmica do processo e impacto das variáveis na entrada é essencial para ser usado no desenvolvimento do modelo no estado dinâmico, mas também para processo de validação do modelo. Sendo possível avaliar se o comportamento dos pesos das variáveis na concentração condiz com a dinâmica do processo.

4.4. Análise de Correlação das Variáveis

Foi conduzida a análise de correlação entre as variáveis, usando método de Pearson. Esta análise pode ser feita com algumas ferramentas como Python, Excel, R e etc. Para esse estudo foi usado o Python.

A correlação consiste na relação estatística entre duas variáveis contínuas. Se a associação entre os elementos não for linear, o coeficiente não será representado adequadamente.

O coeficiente de correlação de Pearson pode ter um intervalo de valores de +1 a -1. Um valor de 0 indica que não há associação entre as duas variáveis. Um valor maior que 0 indica uma associação positiva. Isto é, à medida que o valor de uma variável aumenta, o mesmo acontece com o valor da outra variável. Um valor menor que 0 indica uma associação negativa. Isto é, à medida que o valor de uma variável aumenta, o valor da outra diminui.

Portanto, a partir dessa análise é possível compreender a correlação das variáveis de entrada com a variável de saída. Além disso, é possível avaliar quais variáveis possuem alta correlação, e pode-se sintetizar as variáveis utilizadas na análise do modelo.

4.5. Aplicação para Desenvolvimento do Modelo

Neste trabalho está sendo usado método de análise supervisionada de regressão no desenvolvimento de modelos estatísticos.

Os dados de processo são divididos em dois grupos: o primeiro grupo, denominado de treinamento ou calibração, é utilizado para ajustar os parâmetros do modelo, sendo fornecido tanto as variáveis secundárias como os resultados os quais se deseja estimar. Já o segundo conjunto de dados, denominado de grupo de teste, é utilizado para testar a qualidade de predição do modelo, calculando, a partir das variáveis secundárias e dos parâmetros do modelo, a propriedade a ser estimada (Marciel, 2021). A porcentagem de dados de teste e validação foi 80 e 20, respectivamente. A seleção dos grupos foi definida para todos os modelos de forma randômica.

Para aplicação dos algorítmicos, várias aplicações já foram desenvolvidas como Python, Pavilion, R e Aspen IQ. Para este trabalho foi escolhido o uso de ferramentas da Aspen e Python de *machine learning* como para desenvolvimento dos modelos e carregamento, pois, esta aplicação já é

consolidada e com infraestrutura disponível em ambiente industrial, podendo desta forma ter implementação e aplicação direta em ambiente industrial. Com esta aplicação é possível ser acompanhado o modelo online durante períodos operacionais.

Dentro da aplicação do Aspen IQ existem as seguintes metodologias de desenvolvimento de modelos: Rede Neural Monotônica, PLS, *Fuzzy* Não-Linear e Regressão por Mínimos Quadrados Parciais/Rede Neural Artificial. Neste trabalho foram comparados também os resultados dos modelos para cada técnica utilizada.

Já no Python a etapa de treinamento e validação das máquinas de aprendizado supervisionado, foram utilizadas as bibliotecas Scikit-Learn para implementação e avaliação das métricas de desempenho do modelo de RNA.

A Tabela apresenta as ferramentas e as versões utilizadas na etapa de implementação das rotinas computacionais no Python.

Tabela 2 - Ferramentas Computacionais Usadas no Python

Linguagem de Programação	<i>Python</i>
Bibliotecas Python	<i>Numpy</i>
	<i>Matplotlib</i>
	<i>Seaborn</i>
	<i>Pandas</i>
	<i>Scikit-learn</i>
	<i>Keras</i>
Ambiente de Desenvolvimento	<i>Google Colab</i>

Fonte: Elaborado pelo Autor (2023)

4.6. Análise dos Resultados

Para análises de dados também foi usada a ferramenta de avaliação da Aspen IQ, exemplificado na Figura 10, e Python, além do uso do Microsoft Office Excel para aplicação das métricas e gráficos. Foram definidas o uso das métricas MSE (Mean Squared Error) e R^2 .

Figura 10 - Etapa de Construção do Modelo no Aspen IQ

The screenshot displays the Aspen IQ software interface. On the left, the 'Commands' window shows a workflow for building a model. The 'Build Mode' is selected, and the process follows a sequence: 'Specify Data' → 'Condition Data' → 'Build Sensor Validation' → 'Deploy'. Below this, the 'Model Identification' section includes 'Align Data' → 'Select Variables' → 'Analyze Dynamics' → 'Build Inferential Sensor' → 'Deploy'. The 'Interactive' mode is selected, and a 'Run Batch' button is visible. A 'Model Notes' window is also present.

On the right, the 'Model' dialog box is open, showing the following details:

Variables

Name	Status	Dead Time	Filter	M
P11309.PV	Active	00:05:00	00:01:52	N
T11310_2.PV	Active	00:10:00	00:01:23	N
T11310_3.PV	Active	00:00:00	00:02:11	N
L19418.PV	Active	00:00:00	00:02:34	N

Model Status: Ready Variables: 4 Base Time: 00:01:00

Results

Set	Samples	RSQ	RMS Error	Avg Error	Max Error	Correlation
Trn	56	0.9363	96.37	72.44	246.7	0.9676

Dep: AC1301A.PV Model: Monotonic NN LV: 4 Hidden Units: 4

Model Type: Monotonic Neural Network

Buttons: OK, Cancel, Build, Expert, Help

Fonte: Elaborado pelo Autor (2023)

5. RESULTADOS

5.1. Seleção das Variáveis

Segundo a metodologia da correlação de Pearson, resultado na Figura 11, as variáveis de entrada com alguma correlação com a “Análise LAB” e “Analisador de CCl₄”, pois de acordo com Cohen (1988) correlação de Pearson inferior a $|0,2|$ é considerado correlação desprezível. Portanto, as variáveis com correlação superior a $|0,2|$ foram:

- Item 01 Tabela 01 - Vazão de Alimentação DCE com Impurezas;
- Item 04 Tabela 01 - Vazão de Refluxo da Coluna de Destilação 01;
- Item 10 Tabela 01 - Vazão de Destilado da Coluna de Destilação 3;
- Item 16 e 17 da Tabela 01 - Nível dos Tanques Pulmão A e B de DCE.
- Item 20 da Tabela 01 - Pressão topo Coluna de Destilação 02;
- Itens 22-28 da Tabela 01 - Temperatura da Coluna de Destilação 2;

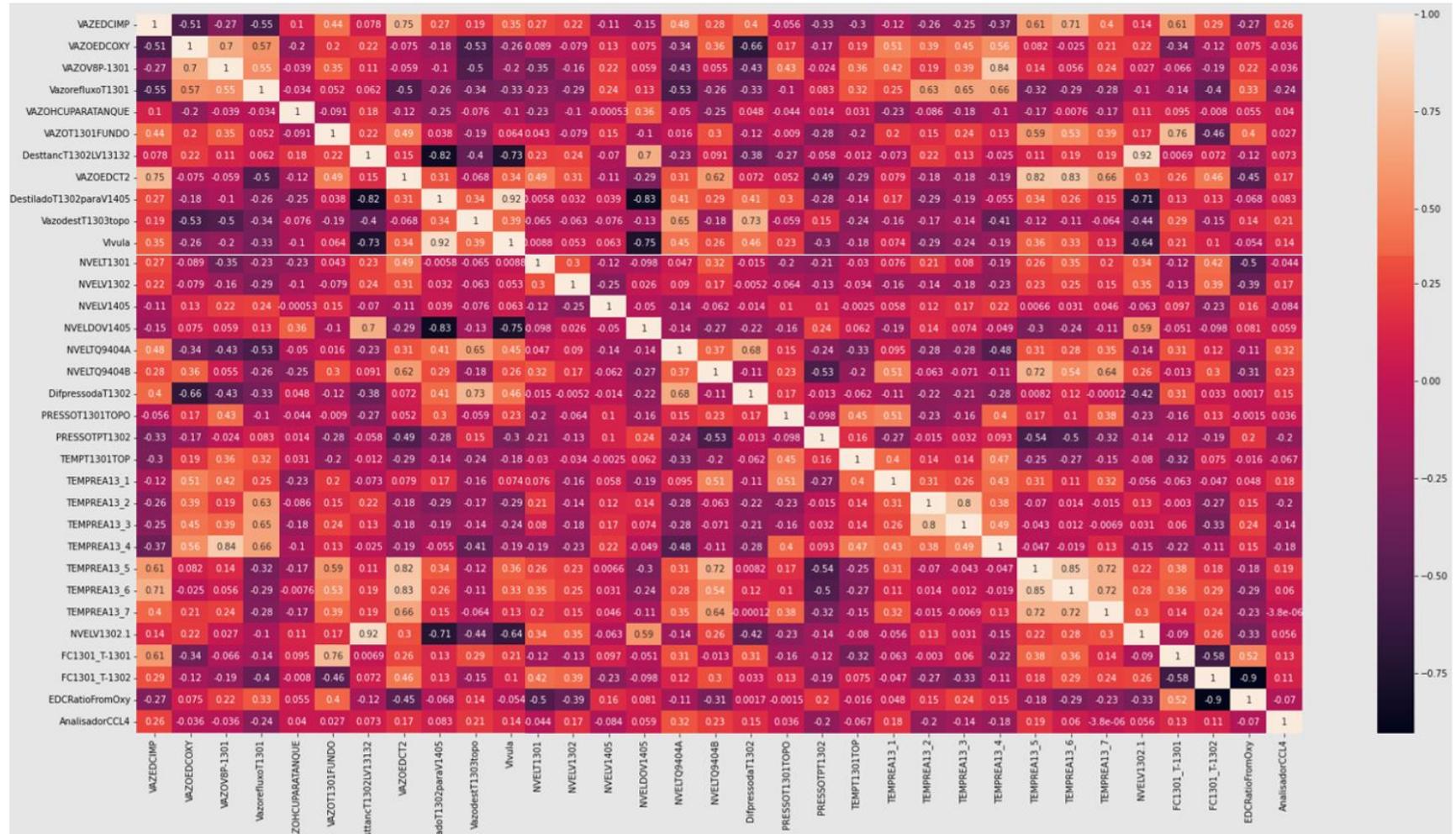
Conforme esperado, a maior correlação dessas variáveis com análise de CCl₄, tem representatividade física significativa, pois, são as variáveis de carga térmica e de massa das colunas. Isto é representativo, pois, o princípio da operação de coluna de destilação é a partir da diferença do ponto de ebulição. Ou seja, comportamento térmico e de massa da coluna.

Principalmente as temperaturas, conforme os resultados obtidos por Moraes et al (2014), onde para ele as temperaturas mais sensíveis podem ser entradas para compor um modelo de inferência de composição de produtos de base, possibilitando a construção de analisadores virtuais e a aplicação de controle inferencial

As variáveis de temperatura apresentaram elevada correlação entre elas. Isto possibilita realizar sintetização dos dados, usando apenas a variável de maior correlação com Análise de CCl₄.

Também, de acordo com Braga (2003), a concentração de CCl_4 está fortemente impactada pelo calor trocado e a separação nos refeedores das colunas 1 e 2, portanto, mesmo que se houvesse uma correlação de Pearson inferior a $|0,2|$, segundo Cohen (1988), as temperaturas e pressões dessas colunas devem ser consideradas na modelagem.

Figura 11 - Resultado da Correlação de Pearson



Fonte: Elaborado pelo Autor (2023)

5.2. Estudo com Diferentes Dados

5.2.1. Dados de Análises de Laboratório

As análises de laboratório são elaboradas uma vez ao dia ou mais, de uma amostra pontual do momento da coleta. Portanto, ao final dos 7 anos, foi possível resgatar 1163 dados de análises realizadas. Porém todas as amostras foram registradas com uma única hora de análise, mostrando um falso registro da hora da coleta da amostra, impactando na avaliação estatística e geração do modelo.

O estudo com esses dados foi realizado a partir da média de 1 hora das variáveis levantadas na metodologia desse trabalho e ferramenta do Aspen IQ, com os seguintes resultados da Tabela 3:

Tabela 3 - Resultado Modelo Usando os Dados de Análise de Laboratório

	Quantidade de Amostras	R ²	Erro Médio Relativo
Treinamento	931	0,207	92,65
Validação	232	0,237	94,96
Todas	1163	0,210	93,11

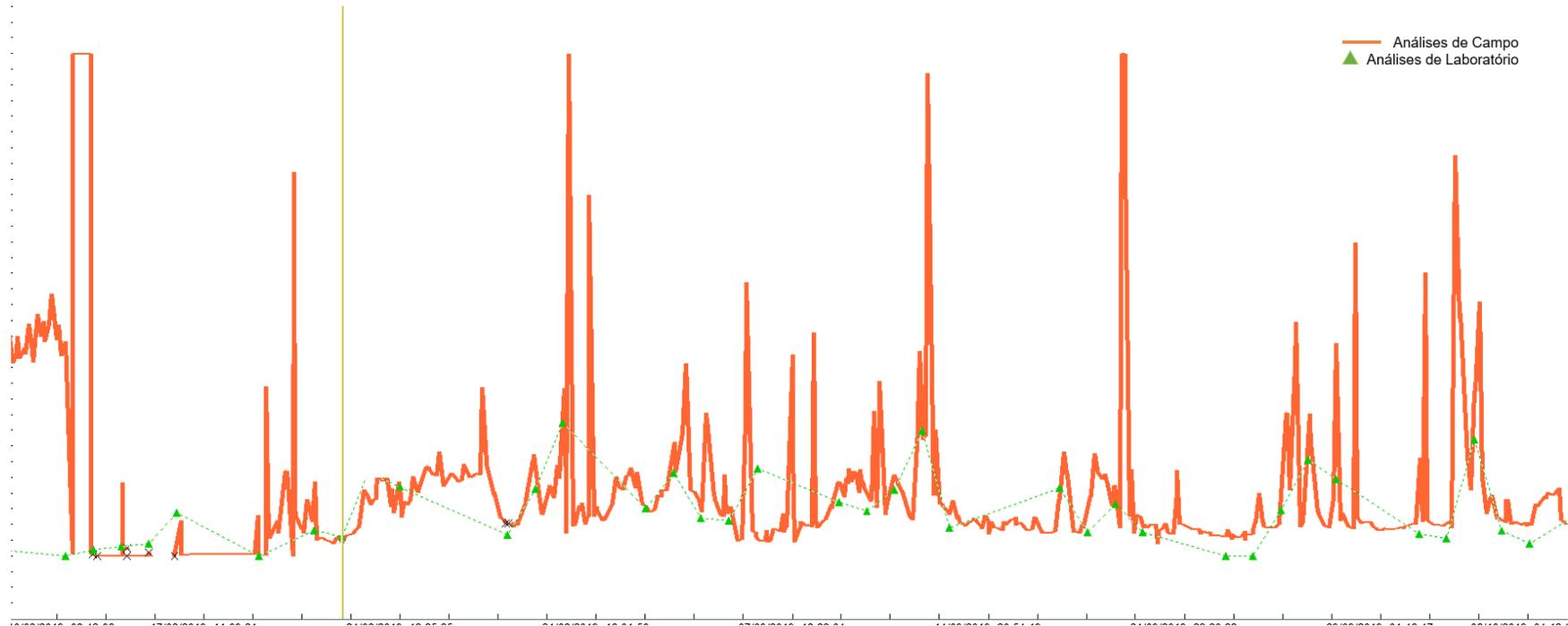
Como o R² está abaixo de 0,70, portanto, há uma fraca correlação do modelo gerado quando comparado com os dados reais. Evidenciando a impossibilidade de uso estatístico desses dados para geração de sensor virtual.

5.2.2. Dados de Análises de Analisador em Linha

Como a diferença entre as análises de laboratório e analisador em campo, em alguns períodos, foi igual ou inferior a 50 ppm, como período de 3 meses apresentados na Figura 122. Então foi realizado o desenvolvimento do modelo também com os dados do analisador de linha. O analisador de linha, ou também como é chamado analisador de campo, tem maior frequência de análise, gerando, portanto, mais dados de análise e maior avaliação do comportamento dinâmico do processo.

Em laranja é o resultado das análises de campo por minuto, e o triangulo verde são os resultados das análises de laboratório para mesmo período.

Figura 12 – Plotagem Dados de Análises de CCl₄ no Laboratório e no Analisador em Linha



Na avaliação inicial para desenvolver o modelo com todos os períodos de dados disponíveis, foram obtidos os seguintes resultados, da Tabela 4, no Aspen IQ.

Tabela 4 - Resultado Modelo Usando os Dados do Analisador em Linha

	Quantidade de Amostras	R ²	Erro Médio Relativo
Treinamento	50437	0,245	293,8
Validação	12609	0,240	293,0
Todas	63046	0,244	293,8

Apesar da maior quantidade de amostras, a correlação do modelo não foi adequada, como pode ser confirmado pelo R² inferior a 0,70. Ou seja, não houve grande mudança de correlação quando estudado todo o período de 7 anos de dados. Essa baixa correlação para esse período se deve as mudanças no processo industrial durante o período, além das variações da matéria prima ao longo dos anos. Portanto, foi avaliado a importância do desenvolvimento do modelo com dados mais recentes, ou seja, com maior representação das condições operacionais atuais.

5.3. Tipos de Modelos Estudados

Após estudo do período dos dados no tópico acima, foi identificado a melhor performance de modelo para as amostras do último ano. Para ter quantidade de dados satisfatória para avaliação, foi então usado os dados do analisador em campo, realizando apenas tratamento dos dados para períodos de paradas das colunas e valores da análise abaixo de 0 ppm, com os resultados da Tabela 5 para cada modelo estudado:

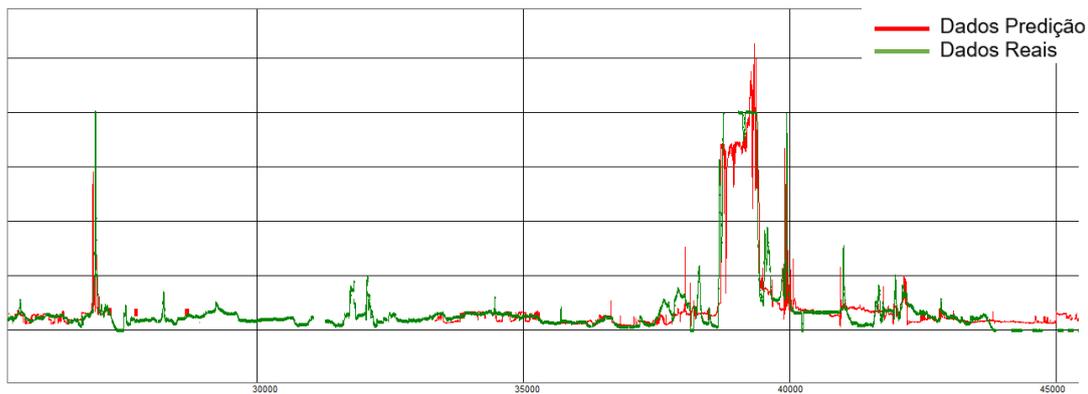
Tabela 5 - Resultados para Tipos de Modelos estudados com Dados Analisador de Campo

	Quant. de Amostras	R ²	Erro Médio Relativo
Híbrido: Regressão por Mínimos Quadrados Parciais e Rede neural artificial (Aspen)	42.602	0,635	109,4
Regressão por Mínimos Quadrados Parciais (Aspen)	42.602	0,520	151,1
<i>Fuzzy</i> Não Linear Regressão por Mínimos Quadrados Parciais (Aspen)	42.602	0,543	165,7
Rede Neural Monotônica (Aspen)	42.602	0,766	106,4
MLP Regressor (Python)	42.602	0,800	91,2

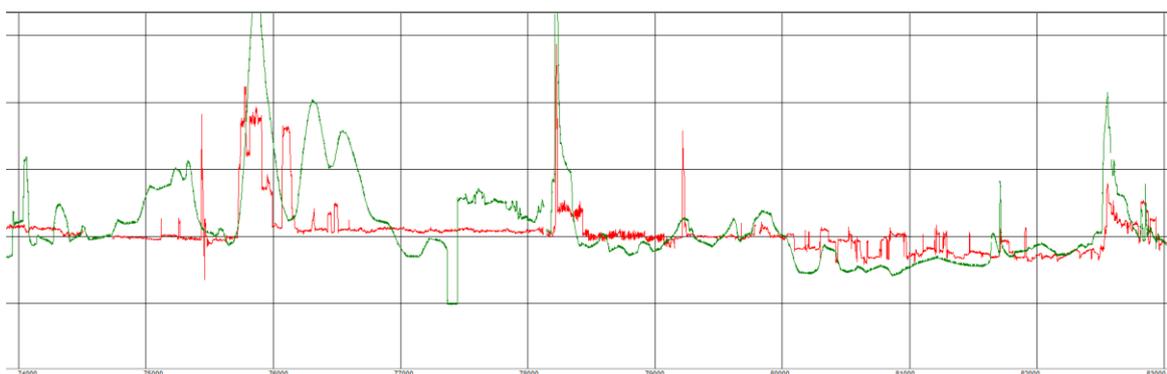
5.3.1. Rede Neural Monotônica

As redes neurais monotônicas fornecem modelos inferenciais com monotonicidade garantida. Isso significa que a direção da resposta da variável inferida a qualquer mudança em uma variável de modelo independente é garantida como consistente em todos os pontos operacionais. Um exemplo de onde isso é importante é no uso de um modelo inferencial para escalonamento de ganho de um controlador; se for utilizada uma rede neural monotônica, é garantido que o ganho do controlador não mudará de sinal em nenhum ponto de operação (ASPENTECH, 2017).

O modelo desenvolvido na Rede Neural Monotônica foi o que obteve maior R², superior a 0,7, e menor erro associado, quando comparado com os modelos desenvolvidos no Aspen. Na Figura 13 foram plotados a predição (vermelho) e o comportamento real (verde), pode-se observar que o modelo consegue identificar alguns patamares mais elevados de CCl₄ no processo, bem como segue a tendência do comportamento. Podendo ser aplicado para validação e refinamento no processo.

Figura 13 - Predição e Real para Rede Neural

Quando se é feita aproximação no gráfico em alguns períodos, pode se notar que a predição consegue acompanhar a tendência dos dados reais e acompanhando picos de análises (período 78000 da Figura 14), porém, algumas variações, como as variações no tempo 75000 o modelo não conseguiu prever, como isso foi notado em todos os modelos estudados, então isso pode acontecer devido a não observação de variações nas variáveis do processo que justificassem essa alteração da concentração de CCl_4 . Portanto, o que pode estar relacionado a alguma variável não medida ou a uma análise falsa do instrumento de campo devido a ações operacionais como purgas para limpeza e desobstruções em linha ou equipamentos.

Figura 14 - Predição (vermelho) e valores reais (verde)

5.3.2. Regressão por Mínimos Quadrados Parciais (PLS)

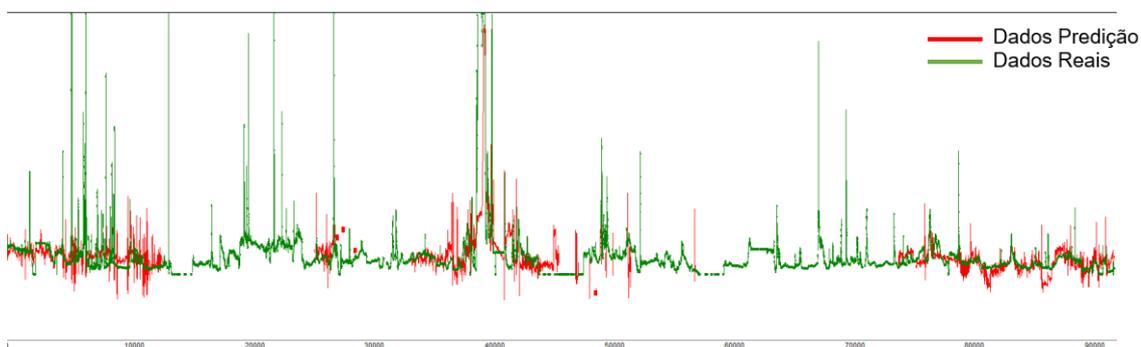
Esse tipo de modelo encontra combinações das variáveis de entrada que são correlacionadas de maneira ideal e capturam a maior quantidade de

variância (ou informação) da variável de saída. Essas combinações sintetizadas de variáveis de entrada são chamadas de variáveis latentes. Em seguida, uma função linear é usada para prever a saída alvo dos valores (geralmente chamados de pontuações) dessas variáveis latentes (ASPENTECH, 2017)

Avaliando também as tendências do modelo predito versus os dados reais de análises, é possível observar, que assim como os algoritmos de rede neural, o modelo consegue prever alguns padrões de subida e descida junto aos dados reais, porém, alguns picos não foram bem representados, como avaliado na Figura 15, além de os dados de predição (vermelho) aparentarem ter maior variação quando comparado aos valores reais (verde).

Esta menor representação do modelo pelo PLS deve-se a não-linearidade do processo, o que o PLS, por ser uma linearização, não tem capacidade de identificar e representar. Também, como o PLS faz essa combinação das variáveis de entrada reduzindo em número menor e mais simplificado, para as variáveis latentes, então essa simplificação do método pode explicar a não capacidade de representar bem todas as variações do processo.

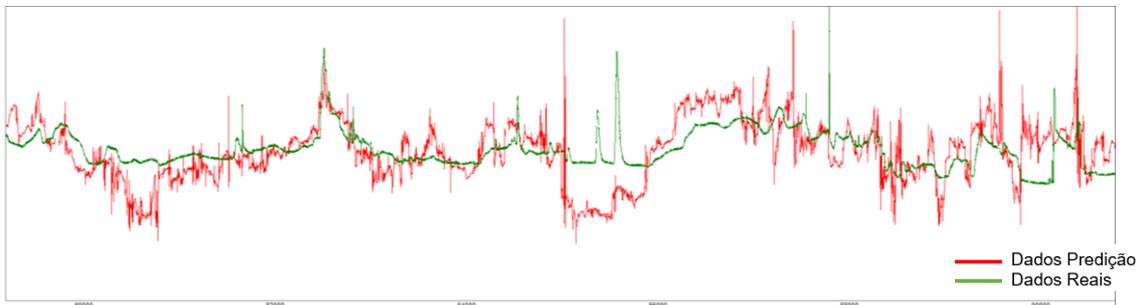
Figura 15 - Predição e Real para PLS



Da mesma forma, quando se é aproximado alguns períodos, pode se notar que a predição consegue acompanhar a tendência dos dados reais e acompanhando alguns picos de análises (período 82000 da Figura 166), porém, algumas variações, como as variações no tempo 85000 o modelo não conseguiu prever. O que pode acontecer devido a não observação de variações nas variáveis do processo que justificassem essa alteração da concentração de CCl_4 . Portanto, pode estar relacionado a alguma variável não medida ou a uma análise

falsa do instrumento de campo devido a ações operacionais como purgas para limpeza e desobstruções em linha ou equipamentos.

Figura 16 - Predição e Real para PLS

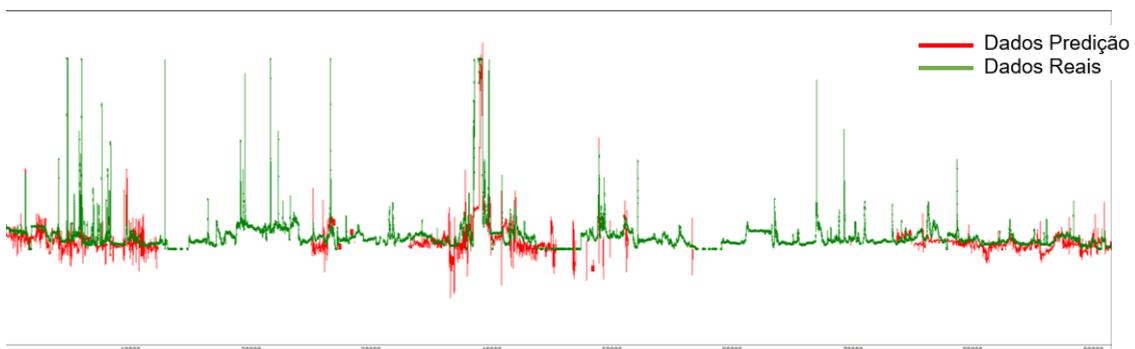


5.3.3. Fuzzy Não Linear Regressão por Mínimos Quadrados Parciais

O algoritmo Fuzzy PLS é uma variação não linear do algoritmo PLS padrão. No algoritmo Fuzzy PLS, separadamente para cada variável latente uma curva não linear é ajustada através das pontuações de entrada (x) e saída (y) do PLS. Entretanto no PLS padrão a curva é uma linha reta (ASPENTECH, 2017).

Avaliando também as tendências do modelo predito versus os dados reais de análises na Figura 17, é possível observar, que o modelo consegue prever mais padrões de subida e descida, quando comparado ao PLS, isso se explica pela possibilidade de avaliação não-linear do Fuzzy, porém, assim como o modelo Rede Neural Monotônica alguns picos não foram bem representados.

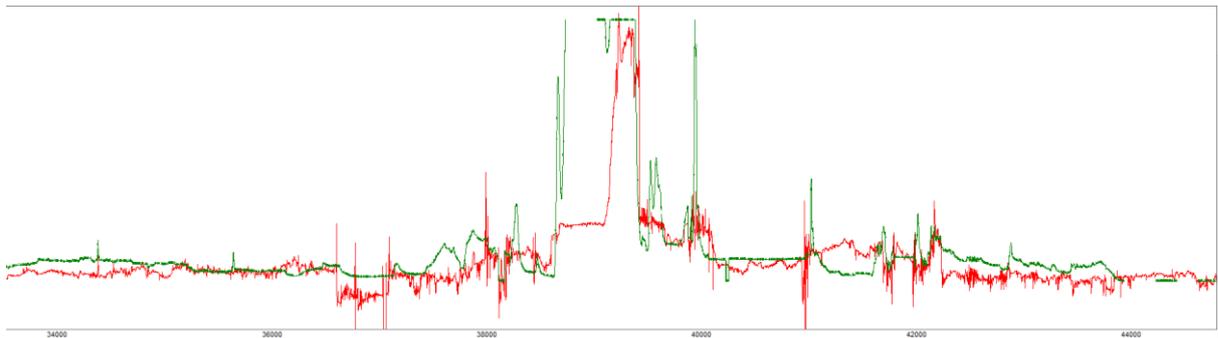
Figura 17 - Predição e Real para Fuzzy Não Linear



Quando se é aproximado para uma dada região do gráfico nos períodos 36000, pode-se notar que a predição consegue acompanhar a tendência dos dados reais e acompanhando alguns picos de análises (período 38000 da Figura

18), porém, mais variações o modelo não conseguiu prever, quando comparado a Rede Neural Monotônica nesses mesmos períodos, portanto mostrando menor representatividade do modelo Fuzzy quando comparado ao Monotônico. Isso pode ser explicado pela redução de variáveis, que assim como o método PLS simples, o Fuzzy-PLS também faz.

Figura 18 - Predição e Real para Fuzzy Não Linear



5.3.4. Híbrido: Regressão por Mínimos Quadrados Parciais e Rede neural artificial

Um modelo PLS híbrido é uma combinação paralela de um modelo PLS linear padrão e uma rede neural. A rede neural modela o desvio entre as previsões do modelo PLS linear para os valores das variáveis dependentes alvo e os valores reais das variáveis dependentes (ASPENTECH, 2017).

Avaliando também as tendências do modelo predito versus os dados reais de análises, é possível observar, que assim como os algoritmos anteriores, o modelo consegue prever alguns padrões de subida e descida junto aos dados reais, porém, alguns picos não foram bem representados. Esta menor representatividade foi observada nos métodos lineares, o que pode explicar a menor correlação do método híbrido. Portanto, o que indica que as correlações e comportamentos do processo é um comportamento não-linear.

Figura 19 - Predição e Real para Modelo Híbrido

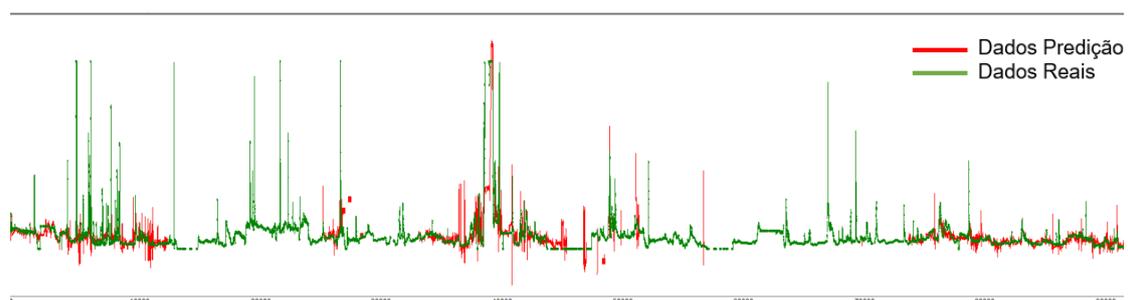
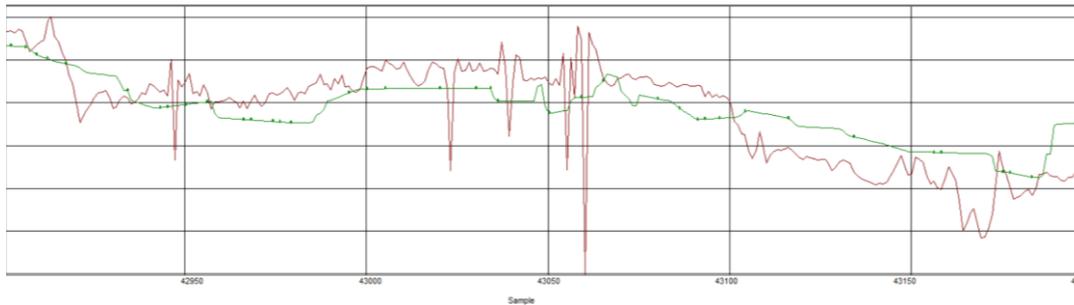
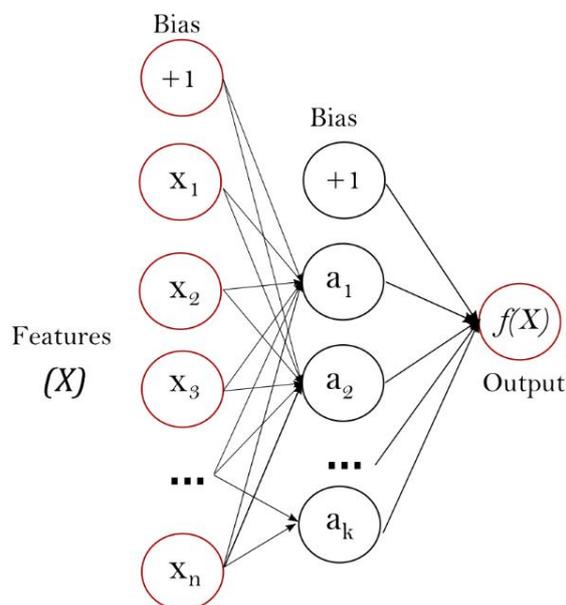


Figura 20 - Predição e Real para Modelo Híbrido

5.3.5. MLP (Multi-layer Perceptron regressor) no Python

Multi-layer Perceptron (MLP) é uma rede neural que aprende uma função $f(\cdot): R^m \rightarrow R^o$ treinando em um conjunto de dados, onde m é o número de dimensões de entrada, e o é o número de dimensões de saída. Dado um conjunto de características $X = x_1, x_2, x_3, \dots, x_m$ e um alvo y , ele pode aprender um aproximador de função não linear para classificação ou regressão. A Figura 1 mostra um MLP de camada oculta com saída escalar (SCIKIT-LEARN, 2023).

Figura 21 - MLP com uma camada oculta

Fonte: SCIKIT-LEARN (2023)

A camada mais à esquerda, conhecida como camada de entrada, consiste em um conjunto de neurônios representando as características de entrada. Cada neurônio na camada oculta transforma os valores da camada anterior com uma soma linear ponderada, seguido por uma função de ativação não linear - como

a função tan hiperbólica. A camada de saída recebe os valores da última camada oculta e os transforma em valores de saída. Portanto, quanto a vantagem do MLP é a capacidade de aprender modelos não-lineares (SCIKIT-LEARN, 2023).

A RNA usando MLP foi implementada utilizando o Python e bibliotecas conforme Tabela 2 na Metodologia desse trabalho. A Tabela 6 e Figura 22 apresenta os conjuntos de hiperparâmetros utilizados na construção dos modelos.

Tabela 6 - Hiperparâmetros Utilizados na Construção dos Modelos de RNA

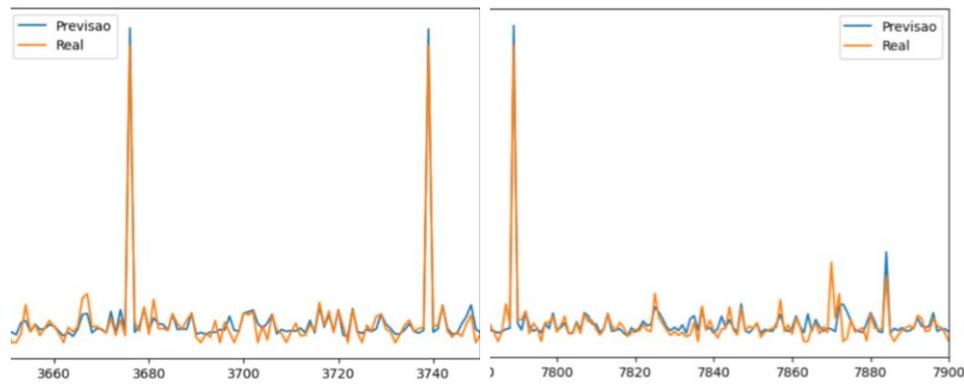
Hiperparâmetro	Valores
Nº de neurônios nas camadas	[100,11,10]
Função de ativação	'tanh'
Otimizador	'SGD'
Taxa de Aprendizagem	"Inviscaling"

Figura 22 - Parâmetros MLP no Python

```
rna = MLPRegressor(hidden_layer_sizes=(100,11,10),
                  max_iter=2000,
                  tol=0.00001,
                  learning_rate_init=0.2,
                  solver="sgd",
                  activation="tanh",
                  learning_rate="inviscaling",
                  verbose=2,
                  )
```

O modelo em MLP, como apresentado na Tabela 5, obteve um R2 de 0,8 e, portanto, superior a 0,7, ou seja, com uma representatividade estatística importante. Também, quando plotado os gráficos dos valores de predição e real, pode ser visto que o modelo foi capaz de prever comportamentos de subida e descida, conforme representado na Figura 23. Também, o erro absoluto em porcentagem desse modelo foi 3,8%, ou seja, inferior a 5%, o que demonstra uma correlação estatística satisfatória para implantação em um analisador virtual online.

Figura 23 - Representações do Modelo Previsto Real em MLP no Python



6. CONCLUSÃO

Apenas os modelos de Rede Neural obtiveram R^2 superior a 0,70, ou seja, com representatividade estatística. Portanto, esses modelos Rede Neural Monotônica do Aspen e MLP do Python foram capazes de representar realizar previsões aceitáveis para aplicação e implantação de analisador virtual online. Isso foi observado, devido a capacidade de modelar comportamentos não-lineares, diferente dos outros modelos com valores de R^2 entre 0,5 e 0,65 que demonstram menor representatividade estatística. Também, quando avaliado os gráficos plotados dados da previsão e dados reais, pode-se confirmar esta informação

Também, quando comparado os dois modelos desenvolvidos por RNA (Aspen IQ e Python), há vantagens e desvantagens associados a viabilidade de implantação. Os algoritmos usados pela Rede Neural Monotônica apesar do R^2 menor e maior erro relativo associado, a aplicação possui uma facilidade na implantação em rede de automação industrial. Já o modelo desenvolvido em MLP é necessário o desenvolvimento da infraestrutura para implantação em área industrial.

O valor de R^2 maior do que 70%, demonstra que a geração da inferência de CCl_4 é possível para esse processo e dados. Também, que as variáveis que mais impactam nas concentrações de CCl_4 , e devem ser usadas para representar na construção de analisador virtual online são a vazão de alimentação DCE com impurezas nas colunas de purificação, vazão de refluxo da coluna de destilação leves, vazão de destilado da coluna de destilação 3, nível dos tanques A e B de DCE, pressão e temperatura topo coluna de destilação dos pesados. Portanto, essas variáveis, como apresentadas no trabalho, representam termodinamicamente as separações nas colunas e, portanto, a concentração de CCl_4 . Adicionalmente, a estas variáveis para serem usadas no analisador virtual online, então a gestão sobre a instrumentação para garantia de confiabilidade da informação do analisar é necessário ser revisado.

7. REFERÊNCIAS

AZANK, F. Como avaliar seu modelo de regressão - As principais métricas para avaliar seus modelos de regressão. 2020. Acessado em Julho de 2021: <https://medium.com/turing-talks/como-avaliar-seu-modelo-de-regress%C3%A3o-c2c8d73dab96>

BORSA, A. G., HERRING, A. M., MCKINNON, J. T., MCCORMICK, R. L., YAMAMOTO, S., TERAOKA, Y., NATORI, Y. Characterization of Coke Formed in Vinyl Chloride Manufacture. *Ind. Eng. Chem. Res.* Vol.38, pp.4259, 1999.

BORSA, A.G., Industrial Plant/Laboratory Investigation and Computer Modeling of 1,2-dichloroethane Pyrolysis. Ph.D. Thesis, Colorado School of Mines. USA, 1999.

BRAGA, L. R. J. Controle de Colunas de Destilação de Alta Pureza um Caso Industrial. Campina Grande – PB, 2003.

CAMPOS, M. M. DE; GOMES, M. V. DE C.; PEREZ, J. M. G. T.. Controle avançado e otimização na indústria do petróleo. 1.ed. Rio de Janeiro: interciência, 2013.

CARVALHO, A.P.L.F. Redes Neurais Artificiais, São Carlos-SP, 2009. Acessado em Julho de 2021: <https://sites.icmc.usp.br/andre/research/neural/>.

COHEN, JACOB. Statistical power analysis for the behavioral sciences. Hillsdale, NJ, Erlbaum, 1988.

CÉSAR, A. Redes Industriais: o que são e para que servem na indústria 4.0. Transformação Digital, 2018. Acessado em Janeiro de 2023 em: <https://transformacaodigital.com/mercado/redes-industriais-o-que-sao-epara-que-servem-na-industria-4-0/>

HAYKIN, S. Redes Neurais, Princípios e Prática. 2ª. ed. Porto Alegre: Bookman, 2001.

JIANXUL, L. HIUHE, S. Soft Sensing Modeling using neuro fuzzy based on Rought set Theory. Proceeding of the American Control Conference, Anchorage: USA. V.1, p. 543-548, 2002.

KIRK-OTHMER Concise Encyclopedia of Chemical Technology, 4th Edition” Wiley-Inter-science, Nova York, 1999

KRÓL, S.; ZABIEGALA, B.; NAMIESNIK, J. Monitoring VOCs in atmospheric air I. Online gas analyzers. Trends in Analytical Chemistry, Amsterdam, v. 29, n. 9, p. 1092-1100, out. 2010.

OLIVEIRA, T.C.L., ALVES, J.J.N., NETO, A.T.P., SANTOS, T.D.K.JUNIOR, A.A.M. Avaliação da Influência da Adição de CCl₄ no Desempenho de Fornalhas Industriais de Pirólise de 1,2-Dicloroetano. Rio de Janeiro, 2015.

ORACLE, SD. O que é Machine Learning?. Acessado em julho de 2021: <https://www.oracle.com/br/data-science/machine-learning/what-is-machine-learning/>

LUYBEN, W.L. (1992), Pratical Distillation Control, editado por W.L. Luyben, Van Nostrand Reinhold, USA.

MARCIEL, L. Analisador Virtual para uma Indústria de Borracha Sintética. Porto Alegre, 2021.

MARCUCCI, S. Simulação de Processos: A Importância da Modelagem na Análise de Viabilidade Técnica, 2020. Acessado em setembro de 2022: <https://propeq.com/simulacao-de-processos/>

MASSA, A. R. C. G. Analizador Virtual para a Determinação do Teor dos Contaminantes Mapd em um Reator Trickle Bed. Salvador – BA, 2017.

MCCULLOCH, W.S. AND PITTS, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. The Bulletin of Mathematical Biophysics, 5, 115-133, 1943.

MENDEL, J. M. Fuzzy Logic Systems for Engineering: A Tutorial. Proceedings of the IEEE. IEEE, v. 83, p. 345–377, 1999.

MINITAB 18. O que é a regressão parcial de mínimos quadrados?. Acessado em dezembro de 2021: <https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/partial-least-squares-regression/what-is-partial-least-squares-regression/>

MORAIS JR, A.A., Rojas, L.O.A., Brito, R.P. Seleção Ótima de Entradas no Estudo de Inferência e Controle de Composições em um Processo de Destilação Industrial de Alta Pureza. Caldas Novas – GO, 2014.

NAGAI, E. Y.; ARRUDA, L. V. R de. Identificação de modelos inferenciais fuzzy para a estimação de medidas de concentração em processos de refino. In: CONGRESSO BRASILEIRO DE PETRÓLEO E GÁS, 3., 2004, Salvador. Anais eletrônicos.. [S.l.: s.n.], 2004. Acessado em: Janeiro 2023 <https://silo.tips/download/identificacao-de-modelos-inferenciais-fuzzy-para-a-estimacao-de-medidas-de-concent>.

PARK, Y., CHOI, B. S., YI, J. Simulation of imbalance reduction between two reactors in an ethylene dichloride cracker, Chemical Engineering Science, vol.60,pp. 1237-1249, 2005.

RANZI, E. , DENTE , M. , ROVAGLIO , M..Pyrolysis and chlorination of small hydrocarbons , Chem. Eng. Commun. ,vol. 117 ,pp. 17 – 39, 1992.

RODRIGUES, B. S. Métodos de Construção de Analisadores Virtuais para Estimação de Teor de Enxofre de Hidrocarbonetos. Belo Horizonte – MG, 2014.

SCIKIT-LEARN DEVELOPERS (BSD LICENSE), Neural network models (supervised), 2007-2023. Acesso em Junho de 2023: https://scikit-learn.org/stable/modules/neural_networks_supervised.html

SILVA, D. Pirâmide de Automação Industrial: o que é? Logique, 2017. Acessado em Janeiro de 2023 em: <https://www.logiquesistemas.com.br/blog/piramide-deautomacao-industrial/>

SILVA, F. G. R.; PITTA, J. V. T.; SIQUEIRA, L. de M.; OLIVEIRA, L. A.; GUARDIEIRO, B. de P.; MOREIRA, E.; PEREIRA, T. P.; SOUZA, D. L. Construction and operation of a distillation column. Research, Society and Development, [S. l.], v. 11, n. 15, p. e532111537635, 2022. DOI: 10.33448/rsd-v11i15.37635. Acesso em: Jan. 2023.: <https://rsdjournal.org/index.php/rsd/article/view/37635>.

VICERI, As classificações dos algoritmos de Machine Learning, 2020. Acessado em Julho de 2021: <https://viceri.com.br/insights/as-classificacoes-dos-algoritmos-de-machine-learning/>

WEBER, L. O. Analisadores Virtuais Baesados em Modelo Neural para Monitoramento e Controle de Colunas de Destilação com Aquecimento Distribuído. Florianópolis-SC, 2012.

ZADECH, L. A. Fuzzy Sets. Information and Control, v. 8, p. 338–353, 1965.