

# Revisão Sistemática do estado da arte da aplicação de *Machine Learning* na área de Engenharia Química

*Systematic review of the state-of-the-art application of Machine Learning in Chemical Engineering*

**Betânia Marins**  
*Graduando*  
Engenharia Química  
[betaniamarins@ucl.br](mailto:betaniamarins@ucl.br)

**Maryeen Rodrigues Leão**  
*Graduando*  
Engenharia Química  
[maryeen@ucl.br](mailto:maryeen@ucl.br)

**Marlon Ferrari**  
*Orientador*  
Faculdade UCL  
[marlonferrari@ucl.br](mailto:marlonferrari@ucl.br)

A lista completa com informações de autoria está no final do artigo

## RESUMO

**Objetivo:** Executar revisão da literatura existente sobre a utilização de técnicas de aprendizado de máquina (*Machine Learning*) aplicadas à Engenharia Química para a previsão de falhas nos processos industriais, a fim de apresentar a contribuição de tais técnicas para a inovação de processos industriais obtida em estudos anteriores e potenciais aplicações.

**Método:** Revisão sistemática da literatura utilizando a metodologia prisma, selecionando os estudos a partir dos termos de busca "Machine Learning", "Engenharia Química", "Processos industriais", "Previsão de falhas" e critérios de exclusão e inclusão, por meio da base de dados *Google Scholar*.

**Resultado:** Determinou-se através da análise criteriosa dos artigos selecionados que a maioria dos estudos utilizaram como metodologia o Aprendizado de Máquina Tradicional Supervisionado e que foi possível atender às expectativas quanto a sua eficiência em realizar classificação de variáveis, reconhecer padrões e comportamento para uma melhor compreensão do processo. Porém, em alguns estudos analisados, não foi possível obter resultados representativos por não possuírem a quantidade de dados necessários, tendo assim, a necessidade de obter uma maior quantidade de dados para resultados apropriados.

**Conclusões:** Os resultados obtidos na presente revisão sistemática permitem concluir que a aplicação de *Machine Learning* em problemas da indústria química tem sido de avanços consistentes, de boas taxas de sucesso. Apesar dos avanços, nota-se a dificuldade de produção e coleta de grandes quantidades de dados. É conclusivo também a preferência pela aplicação de tipos supervisionados de aprendizagem.

**PALAVRAS-CHAVE:** Revisão sistemática. Literatura de revisão. Aprendizado de Máquina. Processos industriais.

## ABSTRACT

**Objective:** Do a review of the existing literature about the use of machine learning techniques applied to Chemical Engineering for the prediction of failures in industrial processes, in order to provide the contribution of such techniques to the innovation of industrial processes in potential studies and applications.

**Methods:** Systematic review of the literature using the prism methodology, selecting studies from the search terms "Machine Learning", "Engenharia Química", "Processos Industriais", "Predição de falhas" and exclusion and inclusion criteria, through the database of *Google Scholar* data.

**Results:** It was determined through careful analysis of the selected articles that most of the studies used Supervised Traditional Machine Learning as methodology and that it was possible to meet expectations regarding its efficiency in classifying variables, recognizing patterns and behavior for a better understanding of the process. However, in some analyzed studies, it was not possible to obtain representative results because they did not have the necessary amount of data, thus requiring a greater amount of data for appropriate results.

**Conclusions:** The results obtained in this systematic review allow us to conclude that the application of Machine Learning in problems of the chemical industry has been consistent with advances, with good success rates. Despite the advances, it is noted the difficulty of producing and collecting large amounts of data. The preference for the application of supervised types of learning is also conclusive.

**KEYWORDS:** Systematic review. Review literature. Machine Learning. Industrial processes.

# 1 INTRODUÇÃO

A revolução industrial mais atual, chamada de Indústria 4.0, se mostra conduzida pelos recentes avanços tecnológicos. Uma das grandes oportunidades nessa era é aproveitar o grande volume de dados gerados nas mais diversas áreas para a otimização dos processos industriais. De acordo com a reportagem da revista ABEQ (2017, p. 9):

“As indústrias químicas têm buscado um gerenciamento cada vez mais em tempo real de seus indicadores de desempenho, otimização da cadeia de suprimentos, controle e monitoramento remoto, manutenção preditiva e sistemas inteligentes de gestão de consumo de energia.”

Conforme ABEQ (2017), a grande quantidade de dados gerados pelas diferentes fontes traz a oportunidade de gerar informações de grande utilidade e importância para a tomada de decisões dentro das indústrias, como por exemplo, menores níveis de consumo de energia e manutenção mais efetiva, levando à melhoria de processo. Segundo Raschka e Mirjalili (2019), nesta era da tecnologia moderna, há uma grande quantidade de dados estruturados e não estruturados. A profusão de dados nas indústrias químicas faz com que os métodos do Aprendizado de Máquina tornem-se extremamente interessantes, já que seria humanamente impossível a compreensão de todas as dimensões e padrões complexos existentes nos dados.

A engenharia química vem se transformando a partir das mudanças na indústria. O estudo feito pela *The National Academies of Sciences, Engineering, Medicine* (2022) mostra que “o desenvolvimento do controle de processos e design de processos coincide com o desenvolvimento do uso de computadores na engenharia química”. O estudo ainda conclui que os profissionais atuais e os futuros, da área da engenharia química, precisarão entender o mundo natural e os dados que o descrevem, assim como usar as ferramentas que transformam esses dados em informações que podem ser úteis.

Uma situação que pode ser evitada durante o processo de produção em uma indústria é a manutenção inadequada, a falta de ações preventivas pode acarretar diversas consequências para a empresa, como por exemplo, o alto custo gerado, já que ocorreria a parada da linha de produção. Diante disso, encontra-se na bibliografia diversos estudos realizados a respeito da utilização de uma ferramenta da Inteligência Artificial denominada Aprendizado de Máquina ou *Machine Learning* (ML), utilizada para a predição de falhas em equipamentos durante o processo de produção,

classificação de variáveis, reconhecimento de padrões e comportamento para que, desta forma, haja uma melhor compreensão do processo produtivo evitando, assim, eventos inesperados.

Na segunda metade do século XX, o Aprendizado de Máquina evoluiu como um subcampo da inteligência artificial (IA) envolvendo algoritmos de autoaprendizagem que derivam do conhecimento dos dados para fazer previsões (RASCHKA; MIRJALILI, 2019, p.1). De acordo com Raschka e Mirjalili (2019), o Aprendizado de Máquina proporciona uma abordagem mais conveniente para capturar os padrões que se encontram nos dados para melhorar de forma progressiva o desempenho de modelos preditivos e desta forma, ser possível tomar decisões baseadas em dados de forma que não exija que os humanos construam manualmente as regras e modelos a partir da análise de grande quantidade de dados gerados. A relevância do tema na era atual torna importante o conhecimento deste na área de engenharia química. Esse trabalho tem então como objetivo executar revisão sistemática da literatura existente sobre a utilização dos métodos de Aprendizado de Máquina (*Machine Learning*) aplicadas a Engenharia Química no cenário nacional, para a previsão de falhas nos processos industriais a fim de apresentar a contribuição de tais métodos para a inovação de processos industriais obtida em estudos anteriores e potenciais aplicações. Nos estudos encontrados na bibliografia, buscou-se identificar os métodos de Aprendizado de Máquina utilizados, as áreas da engenharia química onde os estudos eram aplicados, ano de produção dos estudos, tipo da publicação e a universidade de pesquisa, para um melhor entendimento de como a utilização desses métodos tem influenciado em processos produtivos das indústrias químicas nacionais, compreendendo suas atuais e possíveis contribuições futuras. O artigo foi dividido em cinco seções, sendo a primeira seção referente à introdução. Em seguida, a seção 2 apresenta a fundamentação teórica. Na seção 3, apresenta-se a metodologia utilizada. Já na seção 4, será abordada a discussão e o resultado encontrado. Por fim, na última seção, tem-se a conclusão a respeito do estudo realizado.

## 2. Revisão da Literatura

Na indústria atual é possível encontrar vários desafios relacionados a alguns conceitos relacionados à indústria 4.0 (Ustundag, 2018). Um desses conceitos é o de *big data*. O Aprendizado de Máquina busca fornecer métodos mais otimizados, necessários para suprir o crescimento massivo dos dados (Murphy, 2012).

### 2.1. Aprendizado de Máquina

Goodfellow, Bengio e Courville (2017) conceituam o algoritmo de Aprendizado de Máquina como “um algoritmo capaz de aprender com dados”. Murphy (2012) define o Aprendizado de Máquina como “um conjunto de métodos que podem detectar automaticamente padrões nos dados e, em seguida, usar os padrões descobertos para prever o futuro dados, ou para realizar outros tipos de tomada de decisão sob incerteza”. Rashka (2019) considera o Aprendizado de Máquina como “A aplicação e ciência dos algoritmos que dão sentido aos dados”. Nesse estudo, o conceito de Aprendizado de Máquina foi dividido em duas partes: Aprendizado de Máquina tradicional e Redes Neurais Artificiais.

As entradas comumente são referidas na literatura como variáveis independentes ou preditores, enquanto as saídas são chamadas de variáveis dependentes ou respostas (Hastie, Tibshirani e Friedman, 2008). As entradas podem ser qualitativas ou quantitativas, e alguns métodos performam bem nesses dois tipos de entrada, e alguns performam melhor em apenas um dos tipos (Hastie, Tibshirani e Friedman, 2008). As variáveis qualitativas, para serem utilizadas nas equações, geralmente são transformadas em alguma forma de código numérico. A literatura normalmente define “*x*” como variável de entrada e “*y*” como variável de saída.

Murphy (2012) explica que métodos de Aprendizado de Máquina inicialmente criados possuem uma arquitetura simples, com poucos parâmetros, chamados de Aprendizado de Máquina Tradicional, ou clássico. Para Bishop (2006) apesar das propriedades analíticas e computacionais do Aprendizado Tradicional serem úteis, quando os problemas envolvem dimensões extremamente grandes, esses modelos possuem limitações. Esse problema é chamado em muitas literaturas de *curse of dimensionality* (Bishop, 2006, p. 36, apud Bellman, 1961). Buscando melhorar o desempenho destes algoritmos, foi desenvolvido o Aprendizado Profundo, ou *Deep Learning*, que geralmente possui uma quantidade muito maior de parâmetros se comparado ao Aprendizado Tradicional (Murphy, 2012). A Aprendizagem de Máquina Tradicional possui as vantagens de ser mais fácil de ser aplicada e não

precisar de recursos computacionais muito robustos, diferente da *Deep Learning* (Xu, Sekula e Ding, 2021).

Xu, Sekula e Ding (2021) citam alguns desafios da implementação do Aprendizado de Máquina atualmente, como a falta de uma quantidade suficiente de dados para um bom treinamento dos algoritmos e a limitação de alguns algoritmos atualmente em relação à sua precisão.

De acordo com Goodfellow, Bengio e Courville (2017) a principal dificuldade dos algoritmos de Aprendizado de Máquina ocorre ao resolver problemas matemáticos quando as saídas são atualizadas a partir de processos iterativos, e em funções que possuem números reais. Goodfellow, Bengio e Courville (2017) destacam que: “A dificuldade fundamental em realizar matemática contínua em um computador digital é que precisamos representar infinitos números reais com um número finito de padrões de bits.”. Isso faz que, quando a fórmula contém números reais, pode ocorrer aproximações feitas pelo computador que executa o algoritmo, causando erros no resultado que podem ser um problema quando o algoritmo possuir muitas operações e não for projetado para minimizar estes erros (Goodfellow, Bengio e Courville, 2017). Alguns erros comuns são o erro chamado de *Underflow*, quando números próximos de zero são arredondados para zero, e o chamado de *Overflow*, quando números extremamente grandes são aproximados à  $\infty$  ou  $-\infty$ .

Um dos maiores desafios do Aprendizado de Máquina atualmente são os chamados *Underfitting*, que é quando o algoritmo não garante um erro de treinamento baixo, e *Overfitting*, que acontece quando a diferença entre o erro de treinamento e o erro de teste é muito alta. (Goodfellow, Bengio e Courville, 2017). Eles definem que, para um Aprendizado de Máquina ser eficiente, é importante evitar esses dois problemas, e para isso, é necessário controlar a capacidade do modelo de se adequar a várias funções.

### **2.1.1. Aprendizado de Máquina Tradicional**

Os métodos de Aprendizado de Máquina Tradicionais são geralmente divididos em Aprendizado Supervisionado e Aprendizado não Supervisionado (Murphy, 2012). Os modelos de Aprendizado Supervisionado utilizam valores de entrada para prever os valores de saída (Hastie, Tibshirani e Friedman, 2008). Modelos de regressão são um exemplo de Aprendizado Supervisionado, como a regressão linear, muitas vezes servem de base para modelos mais sofisticados (Bishop, 2006). Existem também os modelos de classificação, que tem por objetivo atribuir os valores de entrada a uma classe “k” pertencentes a um conjunto de classes, onde cada valor normalmente pertence a apenas uma classe e função que atribui os valores de entradas a uma classe é chamada de

discriminante (Bishop, 2006). Um exemplo de modelo de aprendizado de máquina tradicional de classificação é o chamado de *K- nearest neighbors* (Hastie, Tibshirani e Friedman, 2008).

Um modelo de aprendizado supervisionado pode ser tanto para classificação quanto para regressão, ou até para ambos, como é o caso da floresta aleatória, ou *random forest*, fato que faz com que esse método seja relativamente popular, além do fato deste modelo ser simples de treinar e de ajustar (Hastie, Tibshirani e Friedman, 2008).

Já os modelos de aprendizado não supervisionado buscam identificar padrões nos dados (Murphy, 2012). Segundo Murphy (2012), diferente do aprendizado supervisionado, o modelo possui apenas saídas, sem nenhuma entrada, e a saída desejada é desconhecida.

### **2.1.2. Redes Neurais**

As chamadas Redes Neurais são um grupo de métodos de aprendizado de máquina que possuem o objetivo de, a partir de entradas com combinações lineares modelar funções não lineares que resultam em aprendizados de máquina que podem ser amplamente aplicados e com resultados altamente satisfatórios (Hastie, Tibshirani e Friedman, 2008).

Esses modelos são chamados de “redes neurais” pois elas foram desenvolvidas inicialmente baseadas no cérebro humano, onde as unidades representam os neurônios e as conexões representam as sinapses (Hastie, Tibshirani e Friedman, 2008).

As operações envolvidas para modelar funções não lineares a partir de entradas com combinações lineares geram uma grande quantidade de modelos, portanto essa classe de modelos consegue aproximar muito bem qualquer função contínua no domínio de números reais, e por isso é comumente chamada de aproximador universal (Hastie, Tibshirani e Friedman, 2008). Porém, interpretar o modelo geralmente é muito difícil, fazendo que esse tipo de modelo seja pouco útil quando o objetivo é um modelo mais compreensível. Como exemplo de redes neurais pode-se citar *Multilayer Perceptron* e *Self Organizing Map* (SOM), sendo este último não supervisionado.

### 3. METODOLOGIA

A metodologia escolhida para esse artigo foi a Revisão Sistemática da Literatura. Para a pesquisa dos artigos foi utilizado como fonte de dados o *Google Scholar*. Antes da busca foram definidos alguns critérios de exclusão e inclusão, para que os artigos selecionados estivessem de acordo com o tema proposto. Os critérios definidos são: artigos do tipo *open access*, para que fosse possível acessar o artigo completo sem nenhuma restrição, artigos nacionais, a fim de compreender o cenário do estudo no Brasil, pesquisas que abordassem o tema *Machine Learning*, pesquisas na área de Engenharia Química e Processos Industriais. O quadro abaixo mostra os critérios de inclusão e exclusão:

**Quadro 1: critérios de inclusão e exclusão escolhidos:**

Critérios de Inclusão	Critérios de Exclusão
<i>Open Access</i>	Não Aborda <i>Machine Learning</i>
Artigos Nacionais	Não é da área da Engenharia Química
Informa o tipo de algoritmo utilizado	

Fonte: Elaborado pelos autores (2023)

Para buscar os trabalhos na fonte de dados foram utilizadas as seguintes palavras chaves: “Machine learning aprendizado de máquina engenharia química previsão de falhas processo produtivo industrial -educação -ensino” onde os termos “-educação” e “-ensino” foram utilizados para a exclusão de artigos que abordassem esses temas, que fogem da temática do estudo. Dessa pesquisa, retornaram como resultado 786 artigos. A estratégia utilizada para a seleção dos artigos foi baseada na metodologia *PRISMA*, como mostra a imagem a seguir:

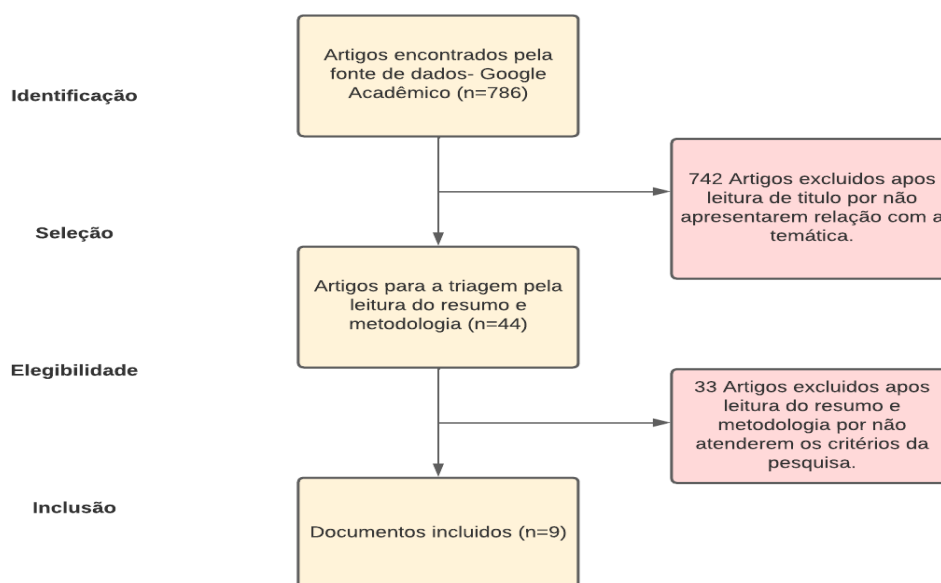


Figura 1. Metodologia PRISMA adaptada para o estudo. Fonte: Elaborado pelos autores (2023)

## 4. Resultados e discussões

Após a análise criteriosa dos estudos obtidos através da triagem realizada na etapa anterior, foram obtidos um total de 9 artigos que se enquadram nos critérios. Dividindo por áreas de atuação na Engenharia Química, se obtém os seguintes resultados:

**Tabela 1 - Quantidade de publicações por área**

Área	Quantidade de Publicações
Indústria de Celulose	4 (44,4%)
Indústria Petrolífera	1 (11,11%)
Indústria de Polímero	1 (11,11%)
Indústria de Alimentícia	1 (11,11%)
Indústria Petroquímica	1 (11,11%)
Outras Indústrias Químicas	1 (11,11%)
Total	9 (100%)

Fonte: Elaborado pelos autores (2023)

É possível perceber através da tabela acima, que as publicações na área da indústria de celulose utilizando os métodos de *Machine Learning* são predominantes, possuindo 4 publicações.

O quadro a seguir mostra a abordagem dos estudos, o ano em que foi publicado, a universidade e o tipo de publicação:



**Quadro 2 - Estudos selecionados**

#	Artigo	Abordagem	Ano	Universidade	Publicação
1	Indústria de Celulose	Aprendizado de Máquina Tradicional	2017	Universidade Federal do Rio Grande do Sul	Monografia
2	Indústria de Celulose	Redes Neurais	2018	Universidade Federal do Rio de Janeiro	Periódico
3	Indústria de Celulose	Redes Neurais e Aprendizado de Máquina Tradicional	2020	Universidade de Minas Gerais	Dissertação
4	Outras Indústrias Químicas	Aprendizado de Máquina Tradicional	2014	Universidade Federal dos Vales do Jequitinhonha e Mucuri, Instituto de Ciências e Tecnologia e Universidade Federal de Uberlândia (COBEQ)	Periódico
5	Indústria Petroquímica	Redes Neurais	2022	Universidade Federal do Rio de Janeiro	Monografia
6	Indústria Petroquímica	Aprendizado de Máquina Tradicional	2021	Universidade Federal de São Carlos	Monografia
7	Indústria Alimentícia	Redes Neurais e Aprendizado de Máquina Tradicional	2021	Universidade Tecnológica Federal do Paraná	Monografia
8	Indústria de Celulose	Aprendizado de Máquina Tradicional	2021	Universidade Federal de Pampa	Monografia
9	Indústria de Polímero	Redes Neurais	2002	Universidade Estadual de Campinas	Dissertação

Fonte: Elaborado pelos autores (2023)

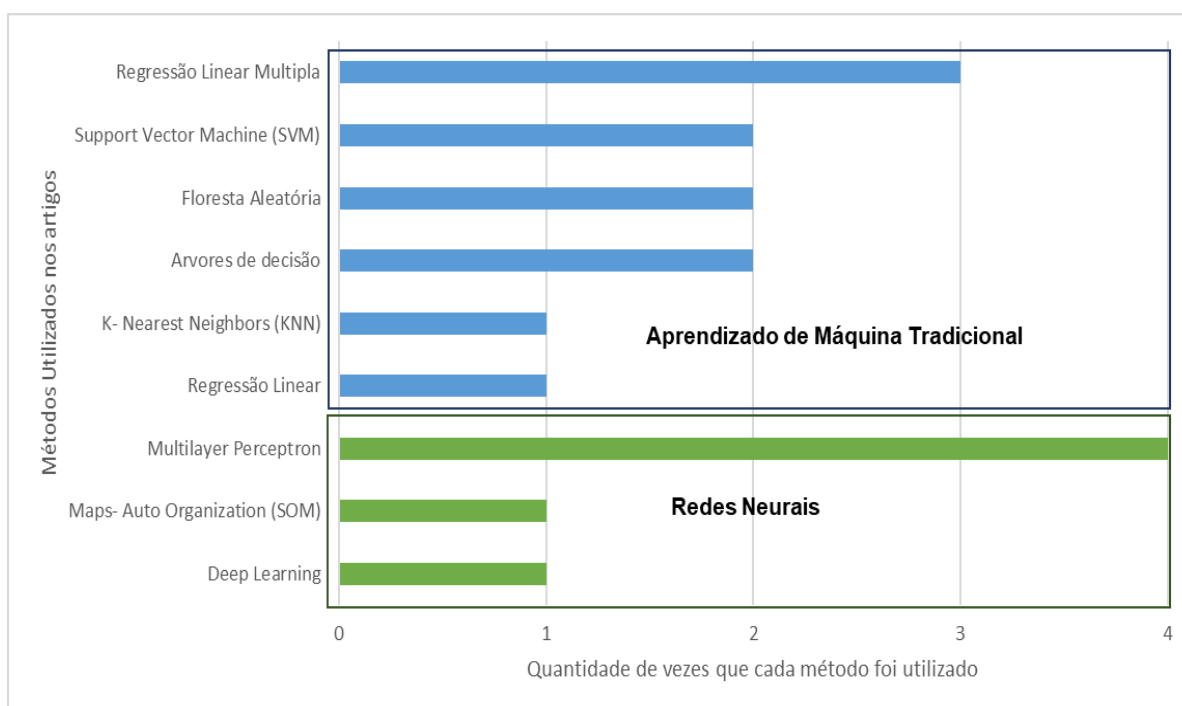
É possível verificar uma maior quantidade de publicações utilizando o Aprendizado de Máquina Tradicional como método em vez das Redes Neurais. Analisando o ano de publicação dos

estudos, identifica-se que, nos últimos dez anos, foram publicados 8 dos 9 estudos, ou seja, 89%, mostrando um aumento recente das contribuições do tema *Machine Learning* na Engenharia Química.

Grande parte dos estudos são monografias, representando 5 das 9 publicações. Logo em seguida, tem-se 2 dissertações e 2 periódicos. Em relação às instituições de produção, a Universidade Federal do Rio de Janeiro contribuiu com 2 estudos, enquanto as demais Instituições apresentadas na tabela contribuíram com 1 estudo cada.

Em seguida, na figura 1, foram observados os métodos utilizados nesses artigos selecionados a fim de compreendermos quais métodos vêm sendo mais utilizados para a otimização de processos produtivos nas indústrias.

**Figura 1 - Quantidade de estudos por métodos utilizados**



Fonte: Elaborado pelos autores (2023)

É possível identificar que o método de *Multilayer Perceptron*, pertencente às Redes Neurais é o mais adotado, sendo utilizado em 4 estudos. Em seguida, os métodos de Regressão Linear Múltipla, Árvores de Decisão e Florestas Aleatórias, todas pertencentes ao Aprendizado de Máquina Tradicional.

A tabela 3 detalha por publicação, área e método utilizado:

**Quadro 3 - Métodos utilizados por área**

<b>Publicação</b>	<b>Área</b>	<b>Método</b>
<b>Modelagem caixa-preta e otimização de uma caldeira de recuperação Kraft (COSTA, 2017)</b>	Indústria de Celulose	<i>Multilayer Perceptron e Regressão Linear Múltipla</i>
<b>Análise de emissões em caldeiras de recuperação química de fábricas de celulose Kraft: predição e análise de sensibilidade com redes neurais artificiais (BELISÁRIO, 2020)</b>	Indústria de Celulose	<i>Multilayer perceptron e Regressão Linear Múltipla</i>
<b>Monitoramento de partículas em caldeira de recuperação kraft por Machine Learning (CARMO et al., 2018)</b>	Indústria de Celulose	Mapas Auto-Organizáveis (SOM)
<b>Aplicação de metodologias estatísticas e machine learning para validação de analisadores contínuos para indústria de celulose (GOMES; VIEIRA, 2021)</b>	Indústria de Celulose	Regressão Linear, Árvores de decisão e Florestas Aleatórias
<b>Aplicação de Redes Neurais no Processo de Filament Winding (CONTANT, 2002)</b>	Indústria de Polímero	<i>Multilayer Perceptron</i>
<b>Utilização de aprendizado supervisionado na predição da demanda de energia no processo de produção de cumeno (VIEIRA; VITOR, 2021)</b>	Indústria Petroquímica	Regressão Linear Múltipla, Árvore de Decisão e Floresta Aleatória
<b>Abordagem baseada em aprendizado de máquina para modelagem e controle preditivo de poços de petróleo com gás lift contínuo (SIMÕES; REY, 2022)</b>	Indústria Petrolífera	<i>Deep Learning</i>
<b>Uso de algoritmos de aprendizado de máquina para predição e classificação de variáveis de interesse em processo industrial de abate de frango (BORTOLETI; BALIEIRO, 2021)</b>	Indústria Alimentícia	<i>K - Nearest Neighbors (KNN), Support Vector Machine (SVM) e Multilayer Perceptron</i>
<b>Desenvolvimento de sistema de detecção de falhas baseado em aprendizado estatístico de máquinas de vetores de suporte (GRANZOTTO; LOPES; 2014)</b>	Outras Indústrias Químicas	<i>Support Vector Machine (SVM)</i>

Fonte: Elaborado pelos autores (2023)

Analisando os resultados obtidos na tabela 3, através dos 9 estudos selecionados, é possível identificar que a Indústria de Celulose possui uma maior quantidade de técnicas utilizadas em seu estudo, tendo os métodos Regressão Linear e *Multilayer Perceptron* com um maior destaque para essa área, logo em seguida, destacam-se as Indústrias Petroquímica e Alimentícia. Vale destacar, também, que todos os métodos utilizados são aprendizados supervisionados com exceção do método de Mapas Auto-Organizáveis (SOM) que é um aprendizado de Rede Neural não supervisionado.

Na tabela 4, verificaram-se os estudos, seus objetivos e resultados alcançados a fim de conhecer o propósito que desejam alcançar e se foi possível alcançá-lo.

**Quadro 4 - Objetivo e resultados alcançados nos estudos**

Publicação	Objetivo	Resultados
<b>Modelagem caixa-preta e otimização de uma caldeira de recuperação Kraft (COSTA, 2017)</b>	Desenvolver um modelo estático para a produção de vapor em uma caldeira de recuperação Kraft e a partir dele, propor as condições ótimas de operação para maximizar a sua produção.	A metodologia aplicada mostrou-se eficiente. O problema de otimização indicou que, a partir de uma redistribuição de ar na caldeira, há um potencial de ganho médio de 5% na eficiência de queima do licor negro quando a caldeira está operando com carga reduzida e de 3,5% quando está operando nas condições nominais.
<b>Aplicação de Redes Neurais no Processo de <i>Filament Winding</i> (CONTANT, 2002)</b>	Relacionar os vários parâmetros do sistema com a qualidade do produto final e com a eficiência do processo.	Os resultados encontrados mostraram a eficiência da aprendizagem das redes em todos os casos estudados. A aplicação de redes neurais na modelagem do processo de <i>filament winding</i> pode levar a uma melhor compreensão do processo, redução do tempo de desenvolvimento de novos produtos e do custo.
<b>Monitoramento de partículas em caldeira de recuperação kraft por Machine Learning (CARMO et al., 2018)</b>	Monitorar a formação de material particulado em uma caldeira de recuperação.	Foi obtida uma rede neural artificial de topologia 20x5 em um mapa auto-organizável (SOM) que permite a identificação de regiões de operação normal, regiões de transição e regiões de maior probabilidade de operação anormal. Os resultados poderiam ser melhorados caso se dispusesse de uma distribuição uniforme dos dados nas classes, pois o número de exemplos de operação normal é muito maior do que os de falhas, o que prejudica o treinamento da rede neural.

Publicação	Objetivo	Resultados
<b>Aplicação de metodologias estatísticas e machine learning para validação de analisadores contínuos para indústria de celulose (GOMES; VIEIRA, 2021)</b>	Validar instrumentos de medição on-line para obtenção das medições laboratoriais através dos dados dos instrumentos.	O analisador virtual construído, apresentou resultados de predição pouco representativos, com $R^2$ de 62,3% após otimização, necessitando de maior quantidade de dados para resultados estatisticamente satisfatórios.
<b>Utilização de aprendizado supervisionado na predição da demanda de energia no processo de produção de cumeno (VIEIRA; VITOR, 2021)</b>	Previsão da energia total e da energia específica (energia por massa de cumeno produzido) de uma planta produtora de cumeno através alquilação do benzeno com propeno.	Os modelos conseguiram prever melhor a energia total da planta do que a energia específica. Em relação à energia total da planta, mesmo com toda a complexidade de um processo químico, os modelos conseguiram prever a energia requerida com erros de aproximadamente $\pm 10\%$ .
<b>Desenvolvimento de sistema de detecção de falhas baseado em aprendizado estatístico de máquinas de vetores de suporte (GRANZOTTO; LOPES; 2014)</b>	Apresentar um sistema de detecção que utiliza as máquinas de vetores de suporte para detectar falhas em processos e realizar o diagnóstico das mesmas.	Os resultados para a detecção e diagnóstico de falhas indicam que as máquinas de vetores de suporte (SVM) com análises de classificação e regressão utilizadas possuem uma maior confiabilidade e uma detecção rápida e eficiente, com a utilização de poucas informações do processo. Os métodos de SVM utilizados para classificação de falhas parecem entregar resultados condizentes e robustos para os cenários investigados.
<b>Abordagem baseada em aprendizado de máquina para modelagem e controle preditivo de poços de petróleo com gás lift contínuo (SIMÕES; REY, 2022)</b>	Predição da vazão de final de óleo a partir de diferentes conjuntos de dados como entrada.	O comportamento do poço foi satisfatoriamente replicado pelas redes treinadas e os testes aos quais os controladores foram submetidos apresentaram boa capacidade de adequação à curva de referência (setpoint) e capacidade de estabilização da vazão de produção em zonas de golfadas.
<b>Uso de algoritmos de aprendizado de máquina para predição e classificação de variáveis de interesse em processo industrial de abate de frango (BORTOLETI; BALIEIRO, 2021)</b>	Predição dos valores de absorção de água na carcaça ao final do processo de abate.	As Redes Neurais previram com eficiência as variáveis: a vazão na bomba, a temperatura no chiller, e o tempo de imersão, que corrobora com o que é apresentado pela literatura. Portanto, comparando- os três algoritmos, o que apresentou o melhor desempenho (considerando apenas o valor do coeficiente de importância da variável) foi o algoritmo SMOReg, sendo este então, o mais recomendado para se realizar a classificação das variáveis do que o algoritmo das RNA's e K-Star. Isto pode ser explicado pela facilidade deste algoritmo em se trabalhar com soluções simultâneas de subproblemas da função objetivo, aplicados a um conjunto de dados com diversas variáveis de entrada.

Publicação	Objetivo	Resultados
<b>Análise de emissões em caldeiras de recuperação química de fábricas de celulose Kraft: predição e análise de sensibilidade com redes neurais artificiais</b> (BELISÁRIO, 2020)	Construir sensores virtuais para emissões de gases e material particulado em fábricas de celulose, e então analisar quais variáveis de processo apresentam maior influência sobre tais emissões.	A rede neural MLP ( <i>Multi-Layer Perceptrons</i> ) melhor se ajustou aos dados e alcançou coeficiente de correlação linear ( $r$ ) significativamente superior àquele da regressão linear múltiplas $r_{mlr} = 0.764$ e $r_{mlp} = 0.939$ . As variáveis mais influentes foram vazão do ar terciário, vazão do ar primário, vazão do licor preto e temperatura do ar primário.

Fonte: Elaborado pelos autores (2023)

Observando os resultados obtidos, os métodos de *Machine Learning* utilizados atenderam às expectativas quanto a sua eficiência em realizar classificação de variáveis, reconhecer padrões e comportamento para uma melhor compreensão do processo. Porém, em 2 dos 9 estudos analisados, não foi possível obter resultados representativos por não possuírem a quantidade de dados necessários, tendo assim, a necessidade de obter uma maior quantidade de dados para resultados apropriados.

## 5. Conclusão

Os resultados obtidos na presente revisão sistemática permitem concluir que a aplicação de *Machine Learning* em problemas da indústria química tem sido de avanços consistentes, de boas taxas de sucesso, demonstrando um potencial de maior adesão e de melhorias dos processos, além do grande poder de compreensão de padrões dos dados envolvidos na indústria, ainda poucos explorados, que podem criar novas oportunidades.

Apesar dos avanços, nota-se a dificuldade de produção e coleta de grandes quantidades de dados - indústria 4.0 e Internet das Coisas - que poderiam acionar modelos mais complexos de Redes Neurais, que seriam capazes de compreender padrões mais intrincados de dados e assim nortear decisões mais completas. Tal necessidade se nota a partir da ampla adoção de modelos lineares tradicionais e/ou Redes Neurais de arquiteturas mais simplificadas, com poucas camadas, presentes nos trabalhos levantados, evidenciando a escassez de dados, sobretudo de plantas e equipamentos em tempo real.

Por fim, é conclusivo também a preferência da aplicação de tipos supervisionados de aprendizagem, demonstrando uma necessidade de verificação e comparação entre o modelo operado nas áreas da indústria química (decisão humana/analógica) e uma possível implantação de um modelo automatizado de decisão. Tal conclusão leva ao fato de ainda haver resistências para a adoção de modelos autônomos, orientados a dados nas indústrias, apesar do alto grau de mecanização já presentes. A tabela abaixo consolida a visão da abordagem de *Machine Learning* frente aos modelos utilizados e seu tipo de aprendizado.

**Quadro 5 - Abordagem, método e tipo de aprendizado utilizados nos estudos**

Área	Abordagem	Método	Tipo de Aprendizado
Indústria de Celulose	Redes Neurais e Aprendizado de Máquina Tradicional	<i>Multilayer Perceptron</i> e Regressão Linear Múltipla	Aprendizado Supervisionado
Indústria de Polímero	Redes Neurais	<i>Multilayer Perceptron</i>	Aprendizado Supervisionado
Indústria Petroquímica	Aprendizado de Máquina Tradicional	Regressão Linear Múltipla, Árvore de Decisão e Floresta Aleatória	Aprendizado Supervisionado
Outras Indústrias Químicas	Aprendizado de Máquina Tradicional	<i>Support Vector Machine (SVM)</i>	Aprendizado Supervisionado
Indústria Petrolífera	<i>Deep Learning</i>	<i>Deep Learning</i>	Aprendizado Supervisionado

Área	Abordagem	Método	Tipo de Aprendizado
Indústria Alimentícia	Aprendizado de Máquina Tradicional e Redes Neurais	<i>K - Nearest Neighbors (KNN), Support Vector Machine (SVM) e Multilayer Perceptron</i>	Aprendizado Supervisionado
Indústria de Celulose	Redes Neurais e Aprendizado de Máquina Tradicional	<i>Multilayer Perceptron e Regressão Linear Múltipla</i>	Aprendizado Supervisionado

Fonte: Elaborado pelos autores (2023)



## REFERÊNCIAS

- BELISÁRIO, Ana Brandão. **Análise de emissões em caldeiras de recuperação química de fábricas de celulose Kraft: predição e análise de sensibilidade com redes neurais artificiais**. Dissertação (Mestrado em Engenharia Química) - Universidade Federal de Minas Gerais, Minas Gerais, 2020.
- BORTOLETI, Claiton Zanini. **Uso de algoritmos de aprendizado de máquina para predição e classificação de variáveis de interesse em processo industrial de abate de frango**. Trabalho de conclusão de curso (Bacharelado em Engenharia Química) - Universidade Tecnológica Federal do Paraná, Francisco Beltrão, 2021.
- COSTA, Liege Pilling Baptista da. **Modelagem caixa-preta e otimização de uma caldeira de recuperação Kraft**. 2017. Trabalho de conclusão de curso (Bacharelado em Engenharia Química) – Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2017.
- BISHOP, Christopher M. **Pattern Recognition and Machine Learning**. 2. ed. Berkeley, California: Springer, 2006.
- Contant, Sheila. **Aplicação de redes neurais no processo de filament winding**. 2002. Dissertação (Mestrado em Engenharia Química)- Universidade Estadual de Campinas, São Paulo, 2002.
- CARMO, E.C. & Parente, Andréa & de Souza Jr, Maurício & Waltz, Flavio. Monitoramento de partículas em caldeira de recuperação kraft por machine learning. **Revista O Papel**, v. 79, n. 3, p. 83-89, jun. 2018.
- GRANZOTTO, M. H.; OLIVEIRA-LOPES, L. C.; Desenvolvimento de sistema de detecção de falhas baseado em aprendizado estatístico de máquinas de vetores de suporte, p. 11819-11828 . In: **Anais do XX Congresso Brasileiro de Engenharia Química - COBEQ 2014** [= Blucher Chemical Engineering Proceedings, v.1, n.2]. São Paulo: Blucher, 2015.
- GOMES, Larissa Vieira. **Aplicação de metodologias estatísticas e machine learning para validação de analisadores contínuos para indústria de celulose**. 2021. Trabalho de conclusão de curso (Bacharelado em Engenharia Química)- Universidade Federal do Pampa, Campus Bagé, Bagé, 2021.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge: MIT Press, 2016.
- HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning**. 2. ed. California: Springer, 2008.
- MURPHY, Kevin P. **Machine learning : a probabilistic perspective**. Cambridge: Massachusetts Institute of Technology, 2012.
- National Academies of Sciences, Engineering, and Medicine. **New Directions for Chemical Engineering**. Washington, DC: The National Academies Press, 2022. Disponível em: <https://doi.org/10.17226/26342>. Acesso em 11 de junho de 2023

RASCHKA, Sebastian, e VAHID Mirjalili. **Python Machine Learning**. 3. ed. Birmingham: Packt Publishing. 2019.

SIMEONI, Osvaldo. **A Brief Introduction to Machine Learning for Engineers**. London: King's College London, 2018. *E-book*. Disponível em: <https://arxiv.org/pdf/1709.02840.pdf>. Acesso em: 11 de junho de 2023.

SOUZA Júnior, Maurício Bezerra de. **Abordagem baseada em aprendizado de máquina para modelagem e controle preditivo de poços de petróleo com gás lift contínuo**. 2022. Trabalho de conclusão de curso (Bacharelado em Engenharia Química) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2022.

VIEIRA, Mateus Vitor. **Utilização de aprendizado supervisionado na predição da demanda de energia no processo de produção de cumeno**. 2021. Trabalho de conclusão de curso (Bacharelado em Engenharia Química) - Universidade Federal de São Carlos, São Carlos, 2021.

USTUNDAG, Alp; Cevickan, Emre. **Industry 4.0. Managing the digital transformation**. Istanbul: 2018. *E- book*. Disponível em: <https://link.springer.com/book/10.1007/978-3-319-57870-5>. Acesso em: 11 de junho de 2023.

XU, Yayin; ZHOU, Ying; SEKULA, Przemyslaw; DING, Lieyun. Machine learning in construction: From shallow to deep learning, Developments in the Built Environment. **Revista Elsevier**. V. 6, Maio 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666165921000041>. Acesso em 12 de junho de 2023.

## **NOTAS**

### **AGRADECIMENTOS**

Agradeço a Deus, a minha família, meus amigos, meus professores, por toda ajuda e apoio, e a todos que, de alguma forma, contribuíram para o desenvolvimento desse trabalho.

### **CONTRIBUIÇÃO E AUTORIA**

Não se aplica.

### **FINANCIAMENTO**

Não se aplica.

### **CONSENTIMENTO DE USO DE IMAGEM**

Não se aplica.

### **CURSO**

Engenharia Química

### **COORDENADOR DO CURSO**

André Luiz Perazzo Amaral

### **DATA DE ENTREGA**

19/06/2023

### **BANCA AVALIADORA**

### **DECLARAÇÃO DE INEXISTÊNCIA DE PLÁGIO**

Declaro que o trabalho não contém plágio ou autoplágio total ou parcial. Todo o conteúdo utilizado como citação direta ou indireta foi indicado e referenciado.

### **LICENÇA DE USO**

Os autores do artigo cedem o direito à divulgação e publicação do material para comunidade acadêmica através de portal da Biblioteca e repositório institucional. Esta autorização permite sua utilização como base para novas pesquisas, caso haja adaptação do conteúdo é necessário atribuir o devido crédito de autoria.