

**UNIVERSIDADE FEDERAL DE ALAGOAS
UNIDADE ACADÊMICA CENTRO DE TECNOLOGIA
CURSO DE ENGENHARIA QUÍMICA**

GISELLY YANARA SILVA CANUTO MENDES COSTA

**APLICAÇÃO DE APRENDIZADO DE MÁQUINA NA PREDIÇÃO DO
RENDIMENTO DA FERMENTAÇÃO DE ALGINASE**

Maceió

2021

GISELLY YANARA SILVA CANUTO MENDES COSTA

**APLICAÇÃO DE APRENDIZADO DE MÁQUINA NA PREDIÇÃO DO
RENDIMENTO DA FERMENTAÇÃO DE ALGINASE**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia Química da Universidade Federal de Alagoas como requisito parcial para obtenção do título de Bacharel em Engenharia Química.

Orientador: Prof. Dr. Carlos Eduardo de Farias Silva.

Coorientador: Prof. Dr. Frede de Oliveira Carvalho.

Maceió

2021

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

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

C837a Costa, Giselly Yanara Silva Canuto Mendes.

Aplicação de aprendizado de máquina na predição do rendimento da fermentação de alginase / Giselly Yanara Silva Canuto Mendes Costa. – 2021.
41 f. il. : figs. ; tabs. color.

Orientador: Carlos Eduardo de Farias Silva.

Coorientador: Frede de Oliveira Carvalho.

Monografia (Trabalho de Conclusão de Curso em Engenharia Química).
Universidade Federal de Alagoas. Centro de Tecnologia. Maceió.

Bibliografia: f. 34-41.

1. Python (Linguagem de programação de computador). 2. Simulação (Computadores). 3. Enzima. 4. Fermentação semi-sólida. I. Título.

CDU: 663.15:004.474

AGRADECIMENTOS

Primeiramente, quero agradecer a Deus por ter me dado forças para seguir até aqui. O caminho não foi nada fácil tampouco rápido, mas estou aqui, graças a Ele, encerrando esse ciclo.

Agradeço também aos meus amigos que sempre me apoiaram, ajudaram e que nunca desistiram de mim, em especial Luana que sempre foi minha irmã de alma desde a infância, Riquinho que mesmo de longe se faz muito presente e Renata, a qual agradeço ao curso por ter me presenteado e que hoje está presente em tudo na minha vida. Amo vocês e agradeço por sempre confiarem em mim, muito mais que eu mesma.

Agradeço aos meus colegas de curso por todos esses longos anos de UFAL, que tornaram tudo um pouco mais leve.

Agradeço aos professores que fizeram parte dessa trajetória em especial ao meu orientador Carlos Eduardo e meu co-orientador Frede, os meus mais sinceros agradecimentos por toda ajuda e paciência. Muito obrigada.

Agradeço ao Elber e ao João por todo auxílio com as programações.

Quero agradecer também as minhas filhas de quatro patas, Nala, Lolla e Nina, que sempre estiveram presentes na minha rotina me dando amor e emanando sensação de paz e companheirismo.

Por fim, quero agradecer a minha família por todo suporte, em especial à minha tia Cada, tia Jane, avó Maria e meu avô Luiz (in memorian). Voinho, sua galega finalmente é engenheira!

RESUMO

Técnicas de inteligência artificial são ferramentas importantes para modelar e otimizar as variáveis que exercem influência em processos fermentativos. O desempenho dos processos de fermentação é afetado por vários fatores, incluindo temperatura, teor de umidade, agitação, nível de inóculo, fontes de carbono e nitrogênio, etc. Para alcançar o melhor desempenho dos processos de fermentação, várias estratégias de otimização de processo foram desenvolvidas. Este trabalho objetiva a aplicação de técnicas de aprendizagem de máquina para prever o rendimento da fermentação semi-sólida usando fungo filamentososo para a produção de alginase utilizando macroalga marrom como substrato. A macroalga escolhida foi do gênero *Sargassum*, e a cepa fúngica o *Cunninghamella echinulata*. No total foram obtidos cerca de 115 dados experimentais independentes a partir da variação dos fatores: tempo de cultivo (1-7 dias), umidade (65-85% m/m), concentração de inóculo ($2 \cdot 10^6$ - $1 \cdot 10^7$ esporos/g_{biomassa}), % de alginato de sódio (indutor, 0-33,33% m/m) e pH da solução nutriente (2,5-8,5), sendo as variáveis de entrada e apenas a atividade enzimática da alginase como variável de saída. Neste trabalho, para a execução dos modelos, simulação e otimização, se utilizou a linguagem de programação em Python utilizando o SciPylibrary. A identificação de relação não linear entre fermentação, fatores e objetivos almejados foi realizado, primeiro, usando as capacidades de aprendizagem de Redes Neurais Artificiais (RNA) e em seguida por Support Vector Machine (SVM), avaliando-se a capacidade preditiva através do coeficiente de determinação (R^2). Para o RNA, com 15 neurônios e utilizando-se uma função de ativação logística foi possível fixar um modelo com $R^2 = 0,877$ enquanto para o SVM com uma função Kernel polinomial obteve-se um modelo com $R^2 = 0,821$, que embora um pouco inferior apresentou regressão semelhante. Em ambos os modelos, as variáveis que exerceram influência foram o tempo de cultivo, umidade, % de alginato de sódio (indutor) e pH da solução nutriente, sendo a concentração de inóculo no intervalo estudado não exercendo influência significativa. Em ambos os métodos, o R^2 máximo encontrado foi considerado satisfatório visto a não-linearidade da fermentação semi-sólida fúngica para excreção enzimática.

Palavras-Chave: Python, simulação, enzima, fermentação semi-sólida.

ABSTRACT

Artificial intelligence techniques are important tools to model and optimize the variables that influence fermentation processes. The performance of fermentation processes is affected by many factors, including temperature, moisture content, agitation, inoculum level, carbon and nitrogen sources, etc. To achieve the best performance of fermentation processes, several process optimization strategies have been developed. This work aims to apply machine learning techniques to predict the yield of semi-solid fermentation using filamentous fungus for alginase production using brown macroalgae as substrate. The macroalgae chosen was the genus *Sargassum*, and the fungal strain was *Cunninghamella echinulata*. In total, about 115 independent experimental data were obtained from the variation of factors: cultivation time (1-7 days), humidity (65-85% m/m), inoculum concentration (2,106-1,107 spores/gbiomass), % of sodium alginate (inducer, 0-33.33% m/m) and pH of the nutrient solution (2.5-8.5), being the input variables and only the alginase enzymatic activity as the output variable. In this work, for the execution of the models, simulation and optimization, the programming language in Python was used using SciPylibrary. The identification of a non-linear relationship between fermentation, factors and desired objectives was carried out, first, using the learning capabilities of Artificial Neural Networks (ANN) and then by Support Vector Macchine (SVM), evaluating the predictive capacity through the coefficient of determination (R^2). For the RNA, with 15 neurons and using a logistic activation function, it was possible to fix a model with $R^2 = 0.877$, while for the SVM with a polynomial Kernel function, a model with $R^2 = 0.821$ was obtained, which, although a little lower, presented similar regression. In both models, the variables that exerted influence were the cultivation time, humidity, % of sodium alginate (inducer) and pH of the nutrient solution, with the inoculum concentration in the studied interval not having a significant influence. In both methods, the maximum R^2 found was considered satisfactory given the non-linearity of the fungal semi-solid fermentation for enzymatic excretion.

Keywords: Python, simulation, enzyme, semi-solid fermentation.

SUMÁRIO

1	INTRODUÇÃO.....	5
2	OBJETIVOS.....	7
2.1	Geral.....	7
2.2	Específicos	7
3	REVISÃO BIBLIOGRÁFICA.....	8
3.1	Fermentação.....	8
3.2	Modelagem de simulação	9
3.2.1	Modelagem clássica fenomenológica.....	12
3.2.2	Modelagem empírica	12
3.2.2.1	Perceptron de Múltiplas Camadas (MLP).....	13
3.2.2.2	Máquina de Suporte de Vetores (SVM).....	16
3.3	Aplicação de aprendizado de máquina em processos fermentativos	19
4	METODOLOGIA.....	21
4.1	Obtenção dos dados experimentais.....	21
4.2	Identificação das máquinas inteligentes.....	22
4.3	Implementação das técnicas.....	22
4.3.1	Treinamento de redes neurais	22
4.3.2	Treinamento de SVM	23
4.4	Normalização e validação dos resultados	23
5	RESULTADOS	25
5.1	RNA.....	26
5.2	SVM	30
6	CONCLUSÃO.....	33
	REFERÊNCIAS	34

1 INTRODUÇÃO

O alginato, um importante polissacarídeo ácido em algas marinhas multicelulares, atraiu atenção como um recurso promissor de biomassa para a produção de produtos químicos médicos e agrícolas e são considerados polissacarídeos estruturais mais abundantes das macroalgas marrons. As macroalgas são utilizadas na alimentação humana, na produção de ficocolóides (ágar, alginato, carragenana) e compostos bioativos, na fabricação de ração humana, fertilizantes agrícolas representando uma importante recurso econômico. O gênero *Sargassum* pertence ao grupo das algas pardas, família *Sargassaceae*, filo *Ochrophyta* (GUIRY, 2014). É fonte de alginato, um ficocoloide amplamente utilizado na indústria de alimentos e de bebidas (TORRES, 2007).

Embora as algas do gênero *Sargassum* não sejam tóxicas, acabam espantando os banhistas quando se depositam na faixa de areia. Com o tempo, começam a se decompor liberando um odor desagradável e podem tornar o banho de mar inviável. Também podem atrapalhar a pesca com rede e o cultivo de mariscos e pescados (TOLEDO, 2017). Uma alternativa para o aproveitamento desse substrato seria a utilização desses resíduos na produção de enzimas por fermentação utilizando fungos filamentosos, pelo fato de serem ótimos microrganismos, que se adaptam às condições diversas.

No entanto, para utilizar o alginato (presente nessa biomassa residual, o *Sargassum*), em aplicações fermentativas, é necessária a sua despolimerização, a qual acontece por ação enzimática da alginase. Além disso, a literatura menciona diversas aplicações da alginase na indústria médica (diagnose e atividade antimicrobiana contra patógenos formados de biofilme, e atividade antitumoral, antihipertensiva, antidiabética, anticoagulante entre outras), fonte de compostos naturais para o processamento de alimentos (produtos da degradação do alginato podem ser usados como crioprotetores, suplementos alimentares, antioxidantes, agentes geleificantes, prebióticos e viscosificantes), na agricultura (oligossacarídeos do alginato podem ser estimuladores vegetais ou biofertilizantes) e em biorefinarias (biocombustíveis e química fina) (MOHAPATRA, 2020).

O termo fermentação em estado sólido, ou fermentação semi-sólida, ou fermentação em meio semi-sólido aplica-se aos processos onde existe crescimento de microrganismos sobre ou dentro de partículas em matriz sólida, onde a quantidade de líquido apresenta um nível de atividade de água que possa garantir o crescimento e metabolismo dos microrganismos, mas não exceda à máxima capacidade de ligação da água com a matriz sólida (PINTO et al., 2006).

Aprendizado de máquina consiste em programar computadores para otimizar o critério de performance de extração de padrões usando dados de exemplo e informações prévias. Isso

pode possibilitar a criação de um modelo preditivo, com experimentos relativamente rápidos, os quais podem resultar na sugestão de experimentos mais promissores a serem executados em laboratório (LARRAÑAGA et al., 2006; FABRIS et al., 2017). Uma das mais importantes aplicações de aprendizado de máquina se dá à criação de sistemas de mineração de dados (KOTSIANTIS et al., 2007). Devido a sua necessidade de casos anteriores para aprender sobre um certo cenário, o algoritmo precisa de uma base de dados para ter resultados bons.

Nos últimos anos, um número limitado de estudos investigou a possibilidade de usar técnicas não estatísticas, como os métodos de inteligência artificial (IA) ou aprendizado de máquina, para o desenvolvimento não linear de modelos empíricos. Os métodos mais comumente usados são redes neurais artificiais (RNAs). Há diversas áreas biológicas onde as técnicas de aprendizado de máquina vêm sendo aplicadas para extração de padrões e informações, que vem crescendo e favorecendo o entendimento de como agrupar dados não lineares, como aqueles obtidos em processos fermentativos (VIEIRA et al., 2020; HECKMANN et al., 2017; AZUAJE, 2016; BARCHI et al., 2016; MALHIS et al., 2015; KANDOI et al., 2015; LARRAÑAGA et al., 2006; HANAI et al., 1999).

Nesse estudo, utilizou-se o aprendizado de máquina para prever o rendimento de alginase produzida por fermentação semi-sólida utilizando biomassa de macroalga como substrato e avaliando-se diversos fatores fermentativos como a umidade do substrato, concentração de inóculo, adição de indutor e pH da solução nutriente.

2 OBJETIVOS

2.1 Geral

O presente trabalho tem como objetivo a aplicação de técnicas de aprendizagem de máquina, visando prever o rendimento da fermentação semi-sólida utilizando fungo filamentosos da produção de alginase e macroalga marrom (*Sargassum*) como substrato.

2.2 Específicos

- Utilização da linguagem *Python* para implementar *machine learning* na fermentação semi-sólida de alginase por fungo filamentosos usando macroalga como substrato;
- Testar técnicas de RNA (Rede Neural Artificial) e SVM (*Support Vector Machine*) na previsão da atividade enzimática;
- Avaliar a convergência dos modelos utilizados na previsão fermentativa de alginase.

3 REVISÃO BIBLIOGRÁFICA

3.1 Fermentação semi-sólida e alginase

A Fermentação Semi Sólida (FSS) é a fermentação de substratos sólidos com baixo teor de água, mas suficiente para o metabolismo e crescimento de microrganismos. Enquanto o teor de água na fermentação clássica submersa (SmF) é $> 95\%$, no SSF varia entre 40 e 80%). O FSS tem sido sugerido como alternativa ao SmF, pois o formador apresenta diversas vantagens, como menor custo de implantação e menor chance de contaminação. Ela desempenha um papel de destaque no aproveitamento de resíduos sólidos, pois consiste em uma técnica de crescimento de micro-organismos sobre e no interior de partículas porosas úmidas na qual o conteúdo de líquido contido na matriz sólida deve ser mantido em valores de atividade de água que assegure o conveniente crescimento e metabolismo celular, mas que não exceda a capacidade máxima de retenção de água na matriz. Dentre os principais produtos da fermentação semi-sólida estão as enzimas, que são produzidas comercialmente principalmente a partir de microrganismos, devido à diversidade dos mesmos, facilidade e controle operacional e maior rendimento em relação aos processos extrativos de tecidos animais e vegetais. As enzimas são aplicadas em diversos setores como indústrias de alimentos, papel, têxtil, animal e farmacêutica, com um mercado em crescente expansão. Além disso, os produtos associados são geralmente mais tolerantes às variações de pH e temperatura, e sua separação não é complexa, via de regra (HAQ; MUKHTAR, 2004; ANDRADE et al, 2015; ARAÚJO et al, 2020).

No entanto, as principais limitações dos bioprocessos de estado sólido incluem o controle dos parâmetros de operação, bem como medições de crescimento e parâmetros metabólicos, como concentração celular, consumo de substrato e produção de biomoléculas. Essas limitações são atribuídas à natureza heterogênea do substrato, que possui estrutura e composição nutricional altamente complexas (PANDEY et al, 2011). Além disso, ao contrário do SmF, o meio FSS apresenta vários gradientes de umidade e temperatura, que podem ter um impacto negativo na eficiência do processo (FARINAS et al, 2011).

A alginase atua como catalizador na despolimerização do alginato através de reações de β -eliminação produzindo oligossacarídeos (PREISS e ASHWELL, 1962; MOHAPATRA, 2020). Essa enzima tem sido isolada e caracterizada a partir de algas marrons, muscos marinhos e bactérias terrestres e marinhas. Além desses organismos, alginase foi encontrada também em fungos (WONG et al., 2000; ZHU e YIN, 2015).

A produção de alginase se dá majoritariamente por fermentação submersa, usando alginato como inductor e aplicando cepas bacterianas, como enfatizado por Sivakumar, Sathisa and Ponmanickam (2015), Beltagy et al. (2016), Li et al. (2017), Shankar et al. (2016), Zhu et

al. (2016), Tavafi et al. (2017), Wang et al. (2017), Chen et al. (2018), Hifney et al. (2018), Li et al. (2019), Sun et al. (2019), Mohapatra (2020), Qiao et al. (2020). Logo, não foram encontrados trabalhos que realizassem a fermentação semi-sólida utilizando fungo filamentosos para a produção de alginase, mostrando que tanto a fermentação realizada quanto essa análise utilizando aprendizado em máquina são inéditos.

Os fungos filamentosos podem ser utilizados para a produção e alginase, alguns trabalhos o utilizaram através de fermentação submersa com as seguintes espécies *Trichoderma reesei* (KONNO et al., 2009), *Aspergillus oryzae* (SINGH et al., 2011) and *Aspergillus fumigatus* (PAPI et al., 2012), *Aspergillus niger*, *Dendryphiella arenaria*, *Eurotium chevalieri*, *Chaetomium funicola*, *Stachybotrys chartarum* and *Emericella nidulans* (HIFNEY et al., 2018, SCHAUMANN e WEIDE, 1990). Portanto, como mencionado anteriormente, mesmo sendo as bactérias a classe de microrganismo mais utilizada, os fungos também podem ser utilizados para esse fim. No entanto, nenhuma publicação foi encontrada para a produção dessa enzima utilizando a espécie *Cunninghamella echinulata*, utilizada nesse trabalho.

A cada ano, a quantidade de dados biológicos disponíveis vem crescendo de forma exponencial, dificultando a extração de informações. Sendo este o maior e atual desafio da biologia computacional, o desenvolvimento de ferramentas e métodos capazes de explorar estes dados se tornam cada vez mais necessários e, nesse contexto, o aprendizado de máquina vem se destacando ao longo dos anos (LARRAÑAGA et al., 2006).

3.2 Modelagem de simulação

A melhoria da qualidade dos produtos, a redução de custos operacionais e a inovação tecnológica são aspectos importantes para sustentar as estratégias das organizações na disputa por mercados. Para assegurarem lucro no médio e longo prazo, necessitam criar e gerenciar conhecimentos, obtendo vantagens reais ao desenvolver os seus sistemas de produção e oferecendo produtos e/ou serviços adequados às necessidades de seus clientes (PEREIRA; COSTA, 2012). Com um mercado cada vez mais competitivo e de mudanças rápidas, a simulação tem-se tornado uma ferramenta útil e poderosa à análise de sistemas complexos que podem envolver situações determinísticas ou estocásticas sob condições específicas determinadas pelo usuário (HARREL; TUMAY, 1995) e à resolução de problemas, indispensáveis aos engenheiros, projetistas e administradores (SHANNON, 1998). Assim, Morabito e Pureza (2010, p.170) definem simulação como “uma importante ferramenta de planejamento que procura emular, por meio de relações lógicas, o funcionamento de sistemas reais, a fim de observar seu comportamento sob diferentes cenários”. Em algumas áreas

organizacionais, considera-se inviável um redesenho dos processos em tempo de execução, seja pelos riscos dos resultados que podem ser apresentados, seja por limitações de orçamento ou mesmo pelo impacto real na vida das pessoas envolvidas no processo. A identificação dessas características torna a simulação uma ferramenta importante a ser utilizada como instrumento de análise e controle dos processos. Com o uso de simulação, torna-se possível definir diversos cenários a serem testados, em diferentes períodos de tempo e com dados distintos de entrada, obtendo-se diferentes valores de saída. Assim, a cada nova execução, a obtenção de diferentes medidas de desempenho suscita a identificação de possíveis causas para esse novo comportamento, sem impacto real no processo (LUCENA et al, 2020).

Bonomi e Schmidell (2001) definem a modelagem matemática de processos fermentativos como a tentativa de representar, através de equações matemáticas, os balanços de massa para cada componente do biorreator, associados às complexas transformações bioquímicas que ocorrem no processo e às velocidades com que essas transformações se processam. Sendo assim, destaca-se o emprego das técnicas de modelagem matemática e de simulação de processos industriais que, além de serem um passo importante para o conhecimento profundo em questão, permitem também a otimização das duas condições operacionais, visando aumentar sua rentabilidade econômica. Estas ferramentas permitem que, uma vez obtido e ajustado um sistema de equações matemáticas que melhor representa o conjunto de fenômenos físicos, químicos e biológicos que regem o processo, obtem-se informações fundamentais para o seu projeto e a sua operação, sem necessidade de recorrer a exaustivos testes de laboratório e/ou planta piloto (ALVES, 2010). Voleski e Votruba (1992) mencionam que, devido à complexidade dos processos reais (que envolve leis físico-químicas, bioquímicas e genéticas), somada às limitações matemáticas, os modelos são baseados na idealidade, e em geral fornecem uma representação fiel de apenas algumas propriedades do processo. Segundo os autores, a formulação do modelo deve possuir um comprometimento entre grau de complexidade e solução economicamente desejável (esforço computacional). Os processos fermentativos possuem características que os distinguem dos processos químicos tradicionais, o que pode ser verificado com as dificuldades na formulação de modelos matemáticos que representam estes processos. Algumas características são as baixas concentrações e baixas velocidades de reação, como resultado da utilização de um meio diluído e problemas de esterilidade, segurança e eventualmente da toxicidade dos processos fermentativos (ALVES, 2010).

Logicamente, a descrição completa de todas as vias e interações metabólicas que ocorrem em um processo biológico seria uma tarefa impossível (OHBA, 1998). Segundo

Sinclair e Kristiansen (1987), em um modelo para fermentação devem ser considerados somente aspectos relevantes nos quais tem-se interesse. Para a elaboração de um modelo matemático é necessário estabelecer as variáveis que caracterizam o estado do sistema em cada instante. Desse modo, com as variáveis de processo definidas, o próximo passo é definir as variáveis independentes que irão afetar o desempenho das respostas e como elas o afetam (BONOMI; SCHMIDELL, 2001). Dessa maneira, um modelo seria uma série de relações entre as variáveis de interesse de um sistema em estudo.

Segundo Pegden, Shannon e Sadowski (1995), a simulação é o processo de projeção de um sistema real a um modelo relativo que, a partir deste, conduz experimentos a fim de entender seu comportamento e/ou avaliar estratégias de operação. Sendo assim, é possível importar a realidade para um ambiente controlado, onde se pode estudar o seu comportamento sob diversas condições, sem riscos físicos e/ou grandes custos envolvidos (FREITAS FILHO, 2001).

Lobão e Porto (1996) e Ören e Yilmaz (2012) afirmam que a simulação dispõe de uma série de inferências sobre as mais diversas atividades nos sistemas de produção, tais como: identificação de problemas; comparação com o desempenho de outros sistemas; estudos sobre a utilização da capacidade instalada, níveis de inventário, lógica de controle, projetos de produto e engenharia, integração, seqüenciamento, gargalos do sistema, melhor arranjo físico e melhor índice de produtividade dos funcionários; treinamento de operadores; análise de sensibilidade; auxílio ao planejamento e a aquisição; e previsão de comportamento e desempenho.

O objetivo da modelagem matemática de um processo fermentativo é, portanto, organizar informações desconexas sobre os eventos em um conjunto coerente, identificar quais sistemas e interações são relevantes em um sistema, descobrir novas estratégias que permitam descrever o comportamento do processo em determinadas condições e entender as características qualitativamente importantes para o processo. A capacidade de predição do modelo é determinada pela correlação entre os dados experimentais e o modelo, ou seja, se os dados previstos pelo modelo estiverem de acordo com os dados experimentais, significa que o modelo descreve os fenômenos que controlam o desempenho do processo (BAILEY, 1998; TORRES; SANTOS, 2015). O resultado mais importante do desenvolvimento de um modelo matemático que representa um sistema é a compreensão que se adquire do que realmente faz o processo funcionar. Os modelos matemáticos podem ser úteis em todas as fases da engenharia, desde pesquisa e desenvolvimento até operações de plantas e em estudos econômicos e de negócios.

3.2.1 Modelagem clássica fenomenológica

Os modelos chamados de fenomenológicos baseiam-se em hipóteses e correlações teóricas. Estes modelos são empregados para explicar os fenômenos e os comportamentos das variáveis envolvidas no processo. Constituem-se de um conjunto de relações matemáticas entre as variáveis alvo do estudo realizado (PIAZZI, 2019). Os modelos matemáticos de processos fermentativos podem ser definidos em três grupos: modelos fenomenológicos, modelos entrada-saída (caixa-preta) e modelos híbridos (caixa-cinza).

Na abordagem fenomenológica, o desenvolvimento do modelo é conduzido pelos aspectos relevantes do processo e pelos chamados princípios fundamentais. Tais modelos tendem a apresentar boa capacidade de extrapolação. Segundo FRACASSO (2011), os modelos fenomenológicos podem ser de parâmetros concentrados ou distribuídos e, geralmente, representam as principais tendências do processo, mesmo fora das condições experimentais. No entanto, o conhecimento necessário para um sistema específico muitas vezes não é disponível. Conseqüentemente, o esforço maior nesse tipo de abordagem é dedicado à identificação correta dos mecanismos relevantes ao processo, o que pode consumir muito tempo (VAN CAN et al., 1997).

Os modelos fenomenológicos para processos fermentativos são constituídos por equações de balanço ou de conservação (de massa, de energia ou de quantidade de movimento, ou seja, os chamados princípios fundamentais), equações de velocidade (como, por exemplo, expressões cinéticas que descrevem a geração ou consumo de espécies dentro do sistema) e equações termodinâmicas, que relacionam propriedades termodinâmicas do sistema (pressão, temperatura, densidade, concentração). As equações cinéticas são denominadas modelos cinéticos (BONOMI; SCHMIDELL, 2001).

3.2.2 Modelagem empírica

Na abordagem empírica, o desenvolvimento do modelo está relacionado com observações do comportamento dos dados medidos do sistema a ser modelado. A principal vantagem dessa estratégia é o fato de ser possível a obtenção, em um período curto de tempo, de um modelo matemático preciso sem que seja necessário um conhecimento detalhado do processo. No entanto, a principal desvantagem dessa abordagem é a impossibilidade de realizar extrapolações, sendo necessário que os experimentos cubram todo o domínio de aplicação do modelo para evitar tal problema. Os modelos empíricos, normalmente, baseiam-se em variáveis externas ao produto, como a temperatura e a umidade relativa do ar de secagem (ADORNO et al, 2013). O principal exemplo de modelos do tipo entrada-saída são as redes neurais artificiais,

porém, será avaliado nesse trabalho também o modelo de suporte de vetores de máquina (VAN CAN et al., 1997). A modelagem matemática vem com o intuito de minimizar perdas, uma vez que ajuda a obter resultados com a redução de custo e tempo, simulando através de modelos matemáticos situações por meio de softwares que executam tais modelos (ADORNO et al, 2013).

3.2.2.1 Redes Neurais Artificiais (RNA) - *Perceptron* de Múltiplas Camadas (MLP)

A aplicação de redes neurais na modelagem e controle de processos vem aumentando muito nos últimos anos. Seu uso é especialmente útil na modelagem de sistemas complexos ou que apresentam alguma incerteza quanto à cinética ou parâmetros do modelo. Em bioprocessos, modelos não lineares são requeridos para controle, otimização e predição do comportamento do mesmo. O desenvolvimento desses modelos é uma tarefa difícil, principalmente quando o processo é muito complexo ou quando o mesmo é pouco compreendido (SIMÕES, 2001).

As redes neurais artificiais são uma descrição genérica para uma classe de modelos computacionais inspiradas na estrutura de neurônios biológicos. Elas podem reconhecer padrões, reorganizar dados e “aprender” comportamentos dinâmicos complexos de sistemas físicos (SILVA et al, 2000). São sistemas computacionais inteligentes de processamento que se aproximam ao cérebro humano por ter uma propensão natural de armazenar conhecimentos de experiências passadas, sendo assim, ela tem a capacidade de criar relações de acordo com experiências passadas, um claro processo de aprendizagem concomitante ao que acontece no cérebro humano.

Uma característica marcante das RNA é a sua total independência do conhecimento da natureza fenomenológica do processo que se vai abordar, de modo que tal abordagem se torna interessante quando as regras de um determinado processo não são determinísticas ou são muito complexas, como por exemplo, problemas onde ocorrem variáveis não-lineares. Tais problemas podem ser resolvidos com o uso de redes neurais artificiais, pois os neurônios artificiais apresentam resposta não linear e são altamente conectados, podendo mapear uma relação multivariável de entrada-saída.

Na teoria, a aplicação de redes neurais para aproximação de funções se tornou possível graças ao teorema de Haaser. Esse teorema afirma que qualquer função contínua definida em um intervalo limitado de números reais pode ser aproximada por um polinômio. Dessa forma se a função de ativação de cada elemento da rede neural for uma função contínua pertencente aos números reais, ela poderá ser aproximada por uma função polinomial, e por consequência, a relação entre a entrada e saída da rede poderá ser aproximada por um polinômio. Isso leva a

conclusão de que será sempre possível definir uma rede neural de múltiplas camadas para atuar como aproximação de um mapeamento não linear específico (SARAIVA, 2017).

Há inúmeros tipos de redes neurais que são estudadas na atualidade, desde algoritmos mais simples como o *perceptron* até a utilização de programas computacionais extremamente complexos. Porém, o maior número de aplicações práticas conhecidas (cerca de 95%) se concentra no tipo redes de multicamadas direta MLP (Perceptron de Múltiplas Camadas).

O Perceptron foi criado em 1958 por Rosenblatt, sendo a forma mais simples da configuração de uma rede neural artificial, cujos pesos e inclinações podem ser treinados para produzir um vetor alvo que quando apresentamos tem que corresponder ao vetor de entrada. Uma vez que é constituída de uma única camada neural e de um único neurônio. O Perceptron de uma única camada é utilizado para dividir duas classes linearmente separáveis. Com o intuito de lidar com os problemas não linearmente separáveis, foram adicionadas camadas de neurônio ocultas no modelo de Rosenblatt, formando então a Rede Neural Artificial *Multilayer Perceptron* (MLP). A complexidade da rede MLP se dá pela quantidade de camadas ocultas que houver e a quantidade de neurônios que essas camadas possuem.

Essas RNAs são compostas de vários elementos computacionais simples (nodos) que interagem localmente. A arquitetura destes modelos é especificada pelas características do nodo, topologia da rede e algoritmo de treinamento.

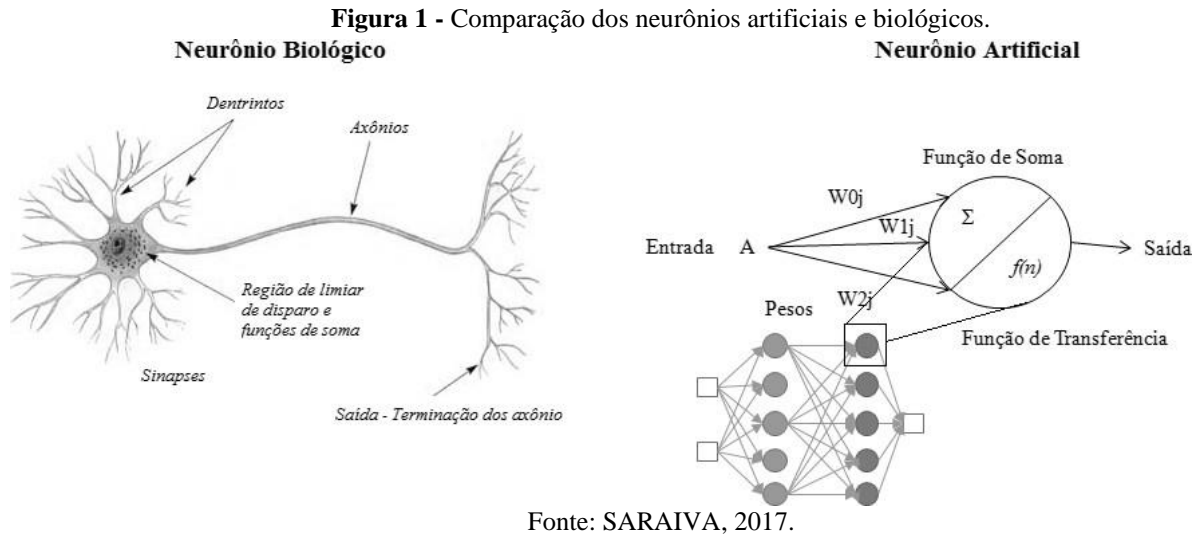
Os nodos em RNA são processadores bastante simples inspirados por seus similares biológicos (neurônios cerebrais). Como enfatizado na relação (**Figura 1**), os neurônios artificiais dispõem de entradas (semelhante biologicamente aos dendritos), pesos (semelhante às sinapses), funções de somas, onde são processados os estímulos captados pela entrada, e a funções de transferências (limiar de disparo do neurônio).

O valor de saída do neurônio artificial, em termos de campo local induzido, é determinado por essa função de transferência, também conhecida por função de ativação, cujo propósito é manter a saída do neurônio artificial dentro do domínio de definição da mesma. Tal função pode assumir várias formas dentre os dois grupos principais, isto é, funções parcialmente diferenciáveis e funções totalmente diferenciáveis.

Esta função funcional pode ser qualquer expressão matemática, mas os matemáticos e cientistas da computação reportam as funções sigmóides logísticas como as mais vantajosas para a maior parte das aplicações.

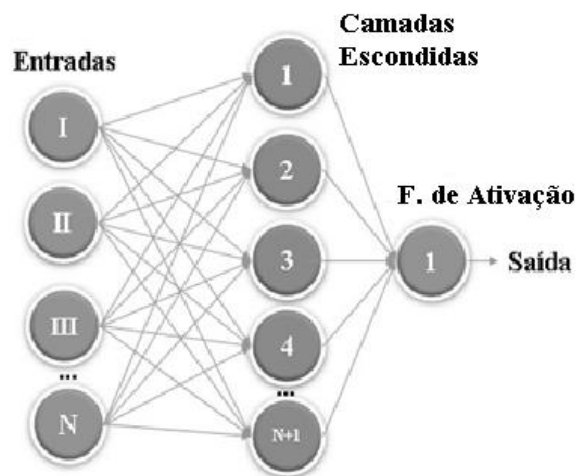
O aprendizado (ou treinamento) de uma rede neural é o processo pelo qual os parâmetros livres de uma rede neural são ajustados por meio de uma forma continuada de estímulo pelo

ambiente externo, sendo o tipo específico de aprendizado definido pela maneira particular como ocorrem os ajustes dos parâmetros livres (BRAGA et al., 2007).



Aprendizado supervisionado é utilizado para treinamento da rede neural quando esta é utilizada como ajustadora de funções multivariáveis e não lineares, caso dos analisadores virtuais. Neste tipo de aprendizado, necessariamente, pressupõe-se a existência de um supervisor, ou professor externo, o qual é responsável por estimular as entradas da rede por meio de padrões de entrada e observar a saída calculada pela mesma, comparando-a com a saída desejada. Como a resposta da rede é função dos valores atuais do seu conjunto de pesos, estes são ajustados através de um processo que aproxima a saída da rede da saída desejada. A arquitetura da RNA básica utilizada é apresentada na **Figura 2**.

Figura 2 - Arquitetura de uma RNA.



onde N representa os números de entrada e o F a função de ativação adequada ao processo.

Fonte: Adaptado Monroy-de-Jesús et al. (2019).

Abaixo se encontra a **Tabela 1** com as principais funções de ativação, onde x são os valores de entrada.

Tabela 1: Principais funções de Ativação

Ativação	Função Correspondente
Linear	$f(x) = ax$
Logística	$f(x) = \frac{1}{(1+e^{-x})}$
Tangente Hiperbólica	$\tanh(x) = \frac{2}{(1+e^{-2x})} - 1$
RELU	$f(x) = \max(0, x)$

Fonte: Adaptado de Data Science Academy (2021).

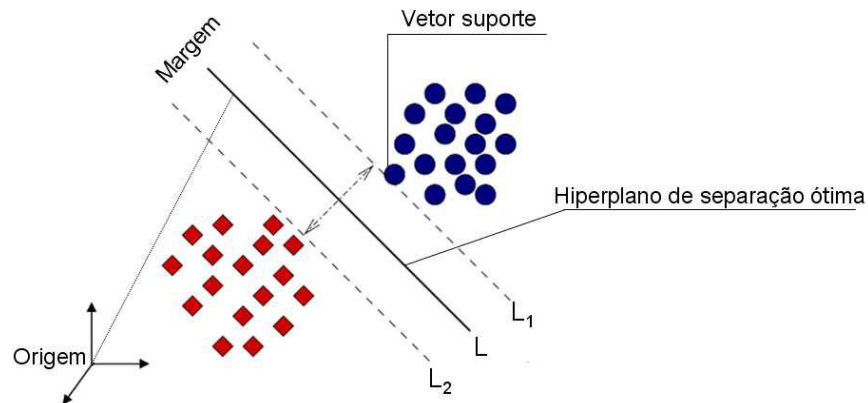
3.2.2.2 Máquina de Suporte de Vetores (SVM)

SVM (*support vector machine*) são sistemas de aprendizagem desenvolvidos por Vapnik (VAPNIK et al, 1999), tais sistemas são baseados na Teoria de Aprendizagem Estatística e treinados com um algoritmo de Otimização Matemática. O SVM possui uma sólida fundamentação teórica e pode alcançar alto desempenho para problemas práticos e complexos, o que o destaca, principalmente problemas ligados à área de controle (IPLIKCI, 2006; SHARMA & SINGH, 2014). Essa técnica pode ser aplicada para extração de característica, reconhecimento de padrões e Regressão.

Outra característica importante do SVM está ligada ao seu treinamento, este ocorre através de uma minimização de uma função quadrática convexa, cujo problema está ligado à otimização matemática. No SVM os poucos parâmetros livres podem ser ajustados pelo usuário e não há uma dependência explícita na dimensão do espaço de entrada, o que pode ser útil em problemas de grandes números de entradas.

A separação ótima entre classes ocorre por meio de um hiperplano condicional (L) (**Figura 3**), tal que este plano é orientado para maximizar a margem (distância entre as bordas, L_1 e L_2) e pelo ponto mais próximo de cada classe.

Figura 3 - Esquema de classificação por meio do SVM.



Fonte: Adaptado de Huang et al. (2002).

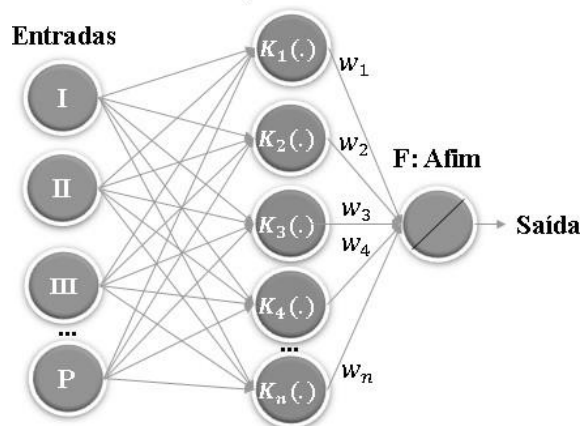
O SVM, quando utilizado para regressão, está diretamente ligado a dois conceitos importantes: A Teoria de Aprendizagem Estatística e a Teoria de Otimização Matemática. O conhecimento dessas teorias é de suma importância para uma boa compreensão da SVM.

A Teoria de Aprendizado Estatísticos em SVM tem como foco o controle da generalização matemática dos dados, ou seja, capacidade de inferência correta de padrões não treinados (SANTOS, 2002). No âmbito científico, existem diversas teorias de aprendizagem que podem ser aplicáveis.

A otimização matemática é um ramo das ciências exatas ligadas a soluções de problemas compostos por funções que devem ser maximizadas ou minimizadas. Essas funções são conhecidas como funções custo ou funções objetivo, e tal procedimento pode ser sujeito às restrições nos valores estimados. No SVM a otimização é utilizada para solução de um PQ (programação quadrática) ou programação quadrática convexa. O algoritmo utilizado no SVM segue conforme a teoria de Lagrange (1797) a qual objetiva caracterizar a solução de um problema de otimização a princípio quando não há restrições de desigualdade. Este método foi desenvolvido por Lagrange em 1797 generalizando um resultado de Fermat de 1692.

Tendo em vista o nível de complexidade que envolve a técnica do SVM, neste trabalho é desenvolvida uma básica explanação do algoritmo de aprendizagem seguindo a Figura 4, com o intuito de atenuar prolixidade da leitura. Para um conhecimento mais profundo do SVM e das teorias que o envolve é sugerido às introduções apresentado nas literaturas (SCHOLKOPF & SMOLA, 2002; WANG, 2005; BEGG & PALANISWAMI, 2006).

Figura 4 - Arquitetura do SVM utilizada.
Função Kernel



Fonte: Adaptado de Liu & Lu (2014).

O conceito básico da regressão SVM é mapear, de forma não linear, os dados originais em um espaço de características dimensional maior e resolver um problema de regressão linear neste espaço.

Algumas das funções Kernel utilizadas na literatura estão presentes na **Tabela 2**. De forma similar a programação do algoritmo RNA, esse modelo, a partir da entrada de dados, gera a melhor configuração de parâmetros para obtenção de saída de dados. O SVM possui quatro funções, sendo elas (i) linear utilizada quando os dados são separáveis linearmente, ou seja, podem ser separados por uma única linha, (ii) sigmoide que é frequentemente usada como uma função de ativação para neurônios artificiais. É interessante notar que um modelo SVM usando uma função de kernel sigmóide é equivalente a uma rede neural perceptron de duas camadas, (iii) polinomial, que representa a semelhança de vetores em um espaço de recurso sobre os polinômios das variáveis originais, permitindo o aprendizado de modelos não lineares e (iv) função de base radial que é o Kernel padrão usado no algoritmo de classificação SVM do sklearn.

Tabela 2: Principais funções de Kernel

Kernel	Função Correspondente	Parâmetros a serem especificados
Linear	$x_i^T x_j$	--
<i>Radial Basis Function</i> (RBF)	$\exp\left(-\gamma \ x_i - x_j\ ^2\right)$	γ
Polinomial	$(x_i^T x_j + k)^p$	k, p
<i>Sigmoid</i>	$\tanh(\gamma x_i^T x_j + r)$	γ, r

onde x_i são os dados de treinamento e x_j são dados de teste.

Fonte: Adaptado de Haykin (1999).

3.3 Aplicação de aprendizado de máquina em processos fermentativos

Com o desenvolvimento da inteligência artificial (IA), técnicas de aprendizado de máquina, como rede neural artificial podem melhorar ainda mais a previsão de modelos fermentativos por meio de sua capacidade de lidar com dados não lineares e complexos, mesmo se esses dados forem ruidosos e impreciso (LEK & GUEGAN, 1999). Com uma precisão comparativamente maior na modelagem e melhor capacidade de generalização, as RNAs são capazes de simular bioprocessos e prever seus resultados, e conseqüentemente, rendimentos.

Geralmente, as vantagens indiscutíveis de ANNs e SVM no processo de otimização da produção biotecnológica são precisão, robustez, facilidade de uso e economia de tempo. ANNs e SVM fazem previsões fortemente baseadas no conjunto de dados de treinamento “bem treinado” e seus programas podem ser executados automaticamente sem muita intervenção humana. A função não linear das RNAs pode até desenvolver um poderoso sistema de predição não linear, que garante a precisão das predições. O princípio do SVM pode garantir fortemente resultados robustos (DENG et al., 2012).

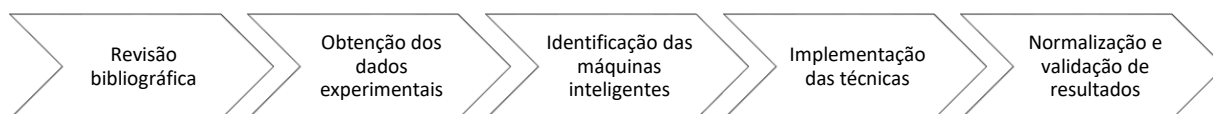
Os mais variados processos fermentativos tem utilizado o aprendizado de máquina para gerar modelos empíricos que predizam seus rendimentos, a citar: Produção de xilitol por *Debaryomyces nepalensis* (Redes Neurais Artificiais acoplada a Algoritmo Genético) (PAPPU e GUMMADI, 2017), Produção de biomassa e β -glucana por *Saccharomyces cerevisiae* NCIM 3458 (Redes Neurais Artificiais) (DESAI et al., 2005), fermentação ruminal de gado leiteiro (CRANINX et al., 2008), fermentação de chá preto (*Support Vector Machine* e *Random Forest*) (ZHU et al. 2019), Produção de dois antitumorais benzoquinonas por *Saccharomyces cerevisiae* (Redes Neurais Artificiais acoplada com Algoritmo Genético) (ZHENG et al., 2017), Fermentação de *Spirulina platensis* com *Lactobacillus platarum* e *Bacillus subtilis* (*Random-centroíde optimization*) (BAO et al., 2018), fermentação de koji (Redes Neurais Artificiais acoplada a Algoritmo Genético) (HANAI et al., 1999), fermentação de ácido γ -aminobutírico por *Lactobacillus fermentum* (Redes Neurais Artificiais e Superfície de Resposta) (RAYOVARAPU et al., 2019), produção de etanol a partir de bagaço de cana (hidrólise e fermentação) (Redes Neurais Artificiais e *Random Forest*) (FISCHER et al., 2017), fermentação de chá (*Random Forest*, *K-Nearest Neighbor*, *Decision Tree*, *Support Vector Machine*, *Linear Determinant Analysis* e *Naive Bayes*) (KIMUTAI et al., 2020), fermentação de grãos de cacau (*K-Nearest Neighbor*) (ANGELIA e LINSANGAN, 2018), lisozima a partir de biomassa bacteriana (*Particle swarm optimization*, *least squares* e *Support Vector*

Macchine) (ZHU e ZHU, 2018) e fermentação de penicilina (*Support Vector Macchine*) (WANG et al., 2010).

4 METODOLOGIA

O presente trabalho seguiu o roteiro conforme a **Figura 5** onde inicialmente foi realizada uma revisão bibliográfica sobre o tema. Os dados experimentais foram obtidos através de experimentos prévios e implementados com o uso de técnicas em máquinas inteligentes, validando os resultados para a predição dos rendimentos.

Figura 5: Fluxograma da metodologia empregada.



4.1 Obtenção dos dados experimentais

A macroalga escolhida foi do gênero *Sargassum*, e a cepa fúngica foi a espécie *Cunninghamella echinulata*. Todas as fermentações ocorreram em frascos vítreos utilizando-se 3g de biomassa macroalgal em temperatura de 30°C, com extrato enzimático obtido em tampão tris-HCl 50 mM pH 7,2-7,5 através de agitação (150 rpm por 30 min), seguida por filtração e centrifugação (3000 rpm por 15 min) para separar a parte sólida do extrato. A solução nutriente (SN) umidificadora foi composta pela seguinte composição: uréia (0,30 g/L), peptona (0,75 g/L), (NH₄)₂SO₄ (1,40 g/L), KH₂PO₄ (2 g/L), MgSO₄.7H₂O (0,30 g/L), CaCl₂.2H₂O (0,40 g/L), ZnSO₄.7H₂O (1,40 mg/L), FeSO₄.7H₂O (5,0 mg/L), CoCl₂.6H₂O (2,0 mg/L) e MnSO₄.4H₂O (1,60mg/L).

Buscou-se otimizar essa solução nutriente em relação aos teores de nitrogênio, e verificou-se um aumento da concentração de peptona (5 g/L) e inclusão de extrato de levedura (2 g/L) (denominada SN₂ sendo utilizada em todas as fermentações posteriores). Estudou-se também a influência da umidade entre 65-85%. A variação da concentração de inóculo entre 2.10⁶-1.10⁷ esporos/g_{biomassa}. Assim como a presença de alginato de sódio como indutor e o efeito do pH da SN₂ no intervalo 2,5-8,5. No total foram obtidos cerca de 115 dados experimentais independentes.

A atividade enzimática foi medida pelo método do DNS (MILLER, 1959), seguindo a metodologia reacional proposta por Sivakumar et al. (2015) com modificações, adicionando o extrato enzimático a uma solução de alginato de sódio 2% (m/m) reagindo em banho termostático (SL-158/A, SOLAB) a 40 °C por 30 min, sendo a reação de DNS posteriormente realizada com leitura em espectrofotômetro a 540 nm (BEL, model SP 2000UV). Esse método

para a medida de ácidos urônicos (resultado da hidrólise promovida pela alginase) é adequado conforme provado por Swift et al. (2014). A atividade enzimática da alginase foi definida como a quantidade de enzima necessária para produzir 1 nmol de monossacarídeos por minuto tendo o alginato como substrato (SIVAKUMAR et al., 2015).

4.2 Identificação das máquinas inteligentes

Os dados a serem utilizados para treinamento das máquinas inteligentes foram obtidos através da variação aleatória das variáveis de entrada do processo, ou seja, pH, concentração do inóculo, concentração de alginato, umidade e tempo. Para a execução dos modelos, simulação e otimização, utilizou a linguagem de programação em Python com o SciPylibrary aplicando as técnicas de redes neurais artificiais (RNA) e *Support Vector Machine* (SVM).

A seleção do melhor conjunto de dados espectrais (original ou qualquer alternativa pré-processada) para garantir o menor erro de predição associado a cada variável de resposta em consideração em ambas as técnicas de correlação foi definida por análise unilateral de variância e, posteriormente, teste de Tukey para múltiplas comparação do conjunto de meios. Esse procedimento foi realizado para detectar diferenças entre as médias dos erros.

4.3 Implementação das técnicas

4.3.1 Treinamento de redes neurais

A questão fundamental na concepção de uma RNA, para uma dada aplicação ou resolução de problema, é a escolha de sua arquitetura. Nessa escolha, define-se a forma como seus neurônios artificiais devem ficar dispostos uns em relação aos outros. Essas disposições são essencialmente estruturadas através do direcionamento das conexões sinápticas dos neurônios artificiais.

O treinamento consiste, a partir de um conjunto de dados de entrada, otimizar os valores de pesos e bias, de maneira que se minimize o erro quadrático entre o valor da variável de estado calculada e o valor experimental. Partindo desse pressuposto, gera uma saída próxima da desejada a partir de quaisquer dados introduzidos em suas entradas. Isto é realizado através de um processo iterativo de ajustes empregados aos seus parâmetros livres (pesos sinápticos e níveis de *bias*). O término do aprendizado ocorre quando a rede neural obtém uma solução generalizada para uma classe de problemas, segundo algum critério pré-estabelecido.

Dentre os algoritmos citados na literatura, pode-se destacar, pela frequência de utilização, dois algoritmos baseados em métodos de otimização diferentes: o de Retropropagação (*Backpropagation*), com a regra do delta generalizado, que é um método de

gradiente descendente bastante utilizado e citado freqüentemente na literatura; e o método de Levenberg-Marquardt, baseado no método quasi-Newton (SARAIVA *et al.*, 2014).

O algoritmo executado para o problema proposto tem um tempo máximo de aprendizado de 5.000 iterações. Cada vez que o programa é rodado, ele fornece uma combinação que foi a mais adequada para a tentativa presente, ou seja, ele faz uma busca automática dos melhores parâmetros. A implementação do método foi realizada na linguagem de programação Python 3.9.6, por meio da biblioteca *Scikit-learn* 0.23.1. A função utilizada da biblioteca foi a *MLPRegressor*.

4.3.2 Treinamento de SVM

O treinamento de SVM envolve a otimização de uma função quadrática convexa, que é um problema de otimização matemática e envolve poucos parâmetros livres que precisam ser ajustados pelo usuário e não há uma dependência, pelos menos de uma forma explícita, na dimensão do espaço de entrada do problema. A partir de um dado conjunto de treinamento de N amostras $\{x_i, y_i\}_{i=1}^N$, a SVM para regressão busca mapear uma função preditora, a partir do ajuste de pesos e bias do hiperplano de separação e com o auxílio de uma transformação não linear que leva o vetor a entrar para uma dimensão maior. A implementação do SVM foi realizada utilizando a função SVR, da biblioteca *Scikit-learn* 0.23.1.

4.4 Normalização e validação dos resultados

O processo de normalização dos dados de entrada e saída são necessários para os modelos da RNA e SVM, a fim de evitar a diferença de magnitude entre os pesos e distâncias, o que fazer com que sejam extraídos padrões não reais e comprometa o treinamento do modelo (KARRI *et al.*, 2018). Nesse sentido, a normalização dos dados foi realizada entre 0 e 1, seguindo a Equação 1, em que y representa o valor normalizado de x_i , enquanto que x_{min} e x_{max} representam os valores mínimo e máximo da variável correspondente.

$$y = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Para as etapas de treinamento e teste dos modelos, as 115 amostras experimentais foram divididas de forma randômica, sendo 80% usada para o treino e 20%, para o teste. Na etapa de teste, foi utilizada a técnica denominada validação por *Holdout*, na qual as previsões dos

modelos são comparadas com a parcela dos dados que não constam no treinamento, aumentando a capacidade de generalização dos modelos e prevenindo o *overfitting*. (GÉRON, 2019).

Em princípio, os RNA's e SVM foram treinados, através do software Phyton. Os critérios utilizados para a avaliação da factibilidade do aprendizado foi o Coeficiente de Determinação (R^2) representado pela equação 2 em que y_i^p representa o valor da atividade enzimática calculado pelo modelo computacional, y_i^o é o valor da atividade calculada experimentalmente e y_m é a atividade experimental média.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i^p - y_i^o)^2}{\sum_{i=1}^N (y_i^p - y_m)^2} \quad (2)$$

5 RESULTADOS E DISCUSSÃO

A investigação da aplicação das máquinas de aprendizagem RNA e SVM para prever o rendimento da fermentação semi-sólida utilizando fungo filamentosos na produção de alginase, foi realizada através de estudo empírico por tentativa e erro das combinações de variáveis de entrada do modelo a ser formado de entrada-saída. As combinações de variáveis podem ser visualizadas na **Tabela 3** abaixo. Quatro modelos foram construídos para as duas máquinas de aprendizado, variando-se para as RNAs, o número de neurônios da camada intermediária, a função de ativação (**função logística**, nesse estudo) e o algoritmo de treinamento. E para SVM, os parâmetros coeficiente 1, ϵ 0,1 e **função kernel polinomial**.

Vale salientar que a saída de todos os modelos foi a atividade enzimática. Na literatura esses parâmetros podem ser encontrados denominados hiperparâmetros já que são parâmetros ajustáveis que permitem controlar o processo de treinamento do modelo. Por exemplo, com redes neurais, você decide o número de camadas ocultas e o número de nós em cada camada. O desempenho do modelo depende muito desses hiperparâmetros (REIS, 2018).

Tabela 3: Variáveis de entrada

ENTRADA	VARIÁVEIS
Modelo 1	pH, tempo e umidade
Modelo 2	pH, tempo, umidade e concentração de alginato
Modelo 3	pH, tempo, umidade, concentração de alginato e concentração do inóculo
Modelo 4	pH, tempo, umidade e concentração do inóculo

Fonte: autora, 2021.

Uma etapa crítica no desenvolvimento de um modelo supervisionado assim como no modelo não supervisionado é avaliar o desempenho no treinamento (supervisionado) e estágios de teste (supervisionados e não supervisionados) para validar que o modelo está resolvendo o problema correto e para verificar se o modelo é robusto. Em outras palavras, que o modelo preditivo é propenso a generalizar para dados invisíveis ou que o agrupamento é completo e consistente (GOODSWEN et al, 2021).

Alguns estudos de processos fermentativos que utilizaram a RNA para prever modelos empíricos de suas fermentações podem ser citados. Em especial, considerando o processo semi-sólido e/ou enzimático: Barchi et al. (2016) estudaram a produção de amilase e protease por

Rhizopus microsporus var. *oligosporus* (fermentação semi-sólida), a partir da mistura de resíduos agroindustriais (farelo de trigo, farelo de soja e bagaço de cana de açúcar), alcançou valores de $R^2 = 0,64-0,97$ com número de neurônios entre 1-100 e melhores funções de ativação como *log-sigmoid* e *hyperbolic tangent sigmoid*, a partir de uma amostragem de 101 pontos experimentais. Hannai et al. (1999) produziu koji (fermentação semi-sólida) por fermentação semi-sólida e aplicou RNA para prever a atividade enzimática no mosto (α -amilase, glucoamilase, proteinase ácida e carboxipeptidase ácida) tendo como variáveis estudadas a umidade, temperatura, tempo de fermentação, tipo do arroz, quantidade fermentada, concentração de inóculo, % de arroz cozido no total de substrato e % *polishing ratio*, e verificaram boas correlações entre cerca de $R^2 = 0,7-0,95$, a partir de 43 pontos experimentais. Vieira et al. (2020) estudou a produção de L-asparaginase por *Penicillium* sp. LAMAI 505, em reatores multiestágios com células imobilizadas por 29 experimentos independentes, avaliando-se como variáveis o pH, tempo de residência, tempo entre os ciclos, concentração de glicose e de L-asparagina (indutor), e verificaram que foram utilizados 7 neurônios, obtendo um $R^2 = 0,85-0,99$ utilizando como funções de ativação a *linear* e *hyperbolic tangent sigmoid functions*.

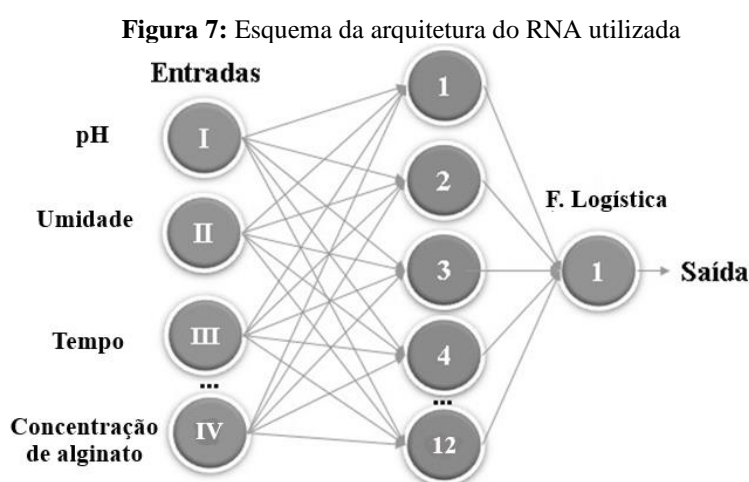
Em todos os casos anteriormente mostrados, existem alguns detalhes, o primeiro fato é que todos utilizam Algoritmo genético acoplado à rede neural artificial, isso permite maior precisão no valor final dos parâmetros que serão utilizados posteriormente, o que não foi o caso desse estudo, que teve por objetivo avaliar a diferença no desempenho ao utilizar diferentes ferramentas de regressão. Outra questão é que nesse estudo a função logística foi a que obteve os melhores resultados nos modelos estudados, diferentemente dos estudos mencionados acima. Além disso, percebeu-se que $R^2 \geq 0,80$ foi considerado como satisfatório nos artigos e forneceram correlações plausíveis visto a não-linearidade dos processos fermentativos. A quantidade de pontos experimentais assim como variáveis estudadas são as mais variadas, mas ressalta-se que nem sempre é possível devido às necessidades de se conduzir diversos experimentos com variações significativas por longos períodos de fermentação, de se obter centenas de dados, e por isso, um $R^2 \geq 0,80$ considerando 5 ou mais variáveis de processo e uma amostragem de até 100 pontos experimentais é um ganho positivo.

5.1 RNA

Para a rede neural observou-se através das simulações que a melhor combinação de entrada-saída foi para o Modelo 2, com as variáveis pH, tempo, umidade e concentração de alginato que obteve o resultado do coeficiente de determinação $R^2 = 0,874$ que segundo a

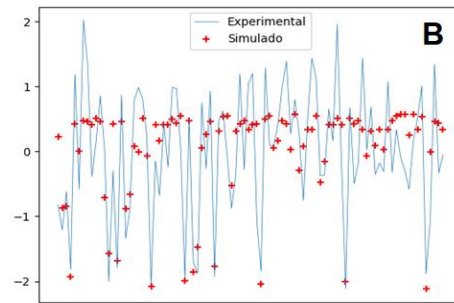
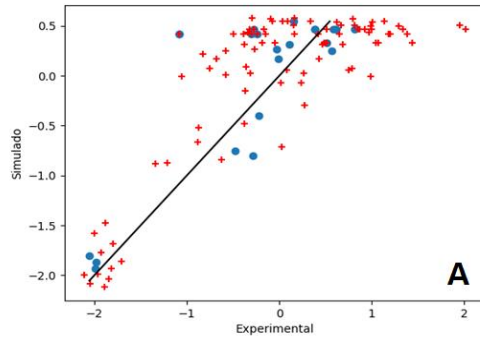
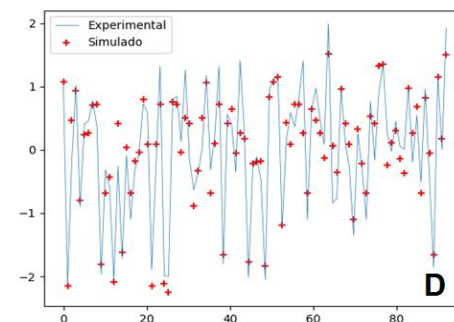
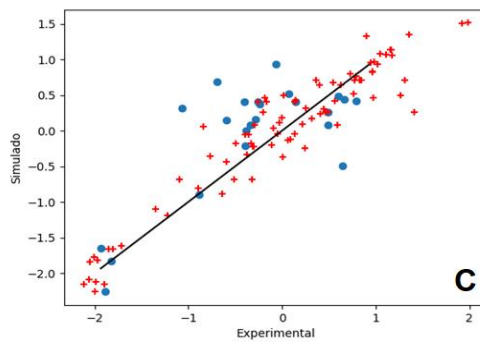
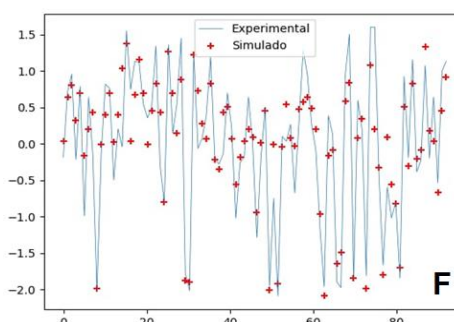
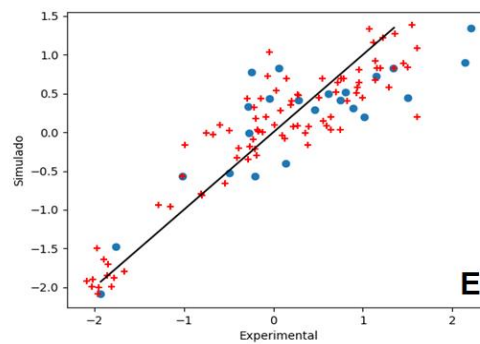
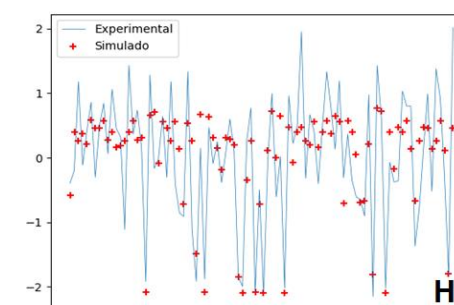
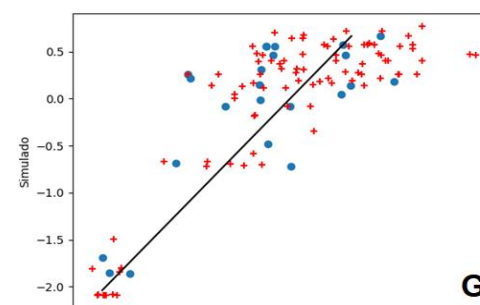
literatura pode ser considerado satisfatório pois é capaz de descrever um sistema complexo (como é o caso da fermentação semi-sólida), e de múltiplas entradas (VIEIRA et al, 2020). Vale salientar que a biblioteca Scikit-Learn do Python faz uma busca automática da rede para a melhor configuração dos parâmetros desta máquina. O algoritmo de treinamento da rede é o Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS), o qual faz parte os métodos de otimização da família quase-Newton, e a **função logística** foi utilizada como função de ativação na camada de saída (SAPTURO; WIDYANINGSIH, 2017). Sendo nesse caso, a combinação de entradas e saída junto com 12 neurônios da camada intermediária, sendo eles parâmetros dessa máquina.

Importante destacar que mesmo diante de um número resumido de dados e com poucas variabilidades no experimento, pôde-se verificar, dentro do domínio de dados, qual a melhor representatividade do modelo de entrada e saída. Assim como se pôde comprovar que a obtenção de dados satisfatórios está condicionada com os parâmetros (variáveis) que foram escolhidas. Baseada na regressão obtida e representada por R^2 , mostrou-se que dentro do intervalo estabelecido, a concentração de inóculo não foi representativa para a aplicação do RNA. Por outro lado, a umidade, o tempo de fermentação, pH da solução nutriente e concentração de alginato (indutor) foram significativos para a aplicação do modelo (representadas na **Figura 7**). Nos gráficos abaixo (**Figura 8**), tem-se os resultados obtidos para os 4 modelos implementados com as redes neurais artificiais, sendo o Modelo 2, com as variáveis descritas acima com o maior coeficiente de regressão, $R^2 = 0,877$.



Fonte: Autor (2021).

Figura 8 – Resultado do treinamento da RNA com os dados da fermentação. Nos gráficos a), c), e) e g) estão os resultados para treinamento (cruz vermelha) e teste (círculos azuis), enquanto b), d), f) e h) representam a diferença entre os pontos experimentais (cruz vermelha) e simulados (linha azul), para os Modelos 1 ($R^2 = 0,614$), 2 ($R^2 = 0,874$), 3 ($R^2 = 0,833$) e 4 ($R^2 = 0,663$), respectivamente.

MODELO 1**MODELO 2****MODELO 3****MODELO 4**

Desse modo, o número de camadas ocultas em um perceptron multicamadas e o número de nós em cada camada podem variar para um determinado problema. Em geral, mais nós oferecem maior sensibilidade ao problema a ser resolvido, mas também ao risco de sobreajuste. Assim como a escolha da arquitetura da rede deve ser escolhida de acordo com a função de ativação. Selecionar funções ou categorizar de uma maneira não trivial, provavelmente não é possível em geral (MURTAGH, 1991).

Esse fato pode ser observado a seguir na **Tabela 4**, onde diversos autores se dedicaram a estudar algum tipo de fermentação, seja ela Fermentação Semi-Sólida (FSS) ou ainda Fermentação Submersa (FS). É possível notar que não há um consenso sobre qual a quantidade de neurônios na camada escondida utilizar para determinar um maior R^2 , independente da quantidade de dados ou da precisão deles. Mas se pode verificar que a maioria utilizou até 25 neurônios, com número de amostragens entre 100-500, como é o caso desse estudo.

Tabela 4: Comparativo de valores do coeficiente de determinação (R^2) entre diversos estudos

Autor	Tipo de fermentação	Método	Função de Ativação	Fermentação	Número amostral	R^2	Neurônios na camada escondida
Costa, 2021 (Autora)	alginase	RNA	logística	FSS	115	0,5 - 0,874	15
Silva et al, 2010	cefalosporina C	RNA	*	FS	300	> 90	17
Vieira et al, 2020	L-asparaginase	RNA e AG	purelin	FSS	29	> 0,85	7
Hannai et al, 1999	a-amilase, glucoamilase, proteinase ácida e carboxipeptidase ácida em koji	RNA e AG	*	FS	43	0,999	19 - 23
Bouaoudat et al, 2012	xylanase, inulinase	RNA e AG	*	FSS	10-300	0,75-0,95	*
Craninx et al, 2008	ácidos graxos do leite	RNA	tanh	FSS	138	-	1 - 16
Desai et al, 2005	β -glucana	RNA	logística	FSS	91	-	1-6
Fischer et al, 2017	etanol	RNA	purelin e logística	FSS e FS	17	0,90 - 0,92	7
Barchi et al, 2016	amilase e protease	RNA e AG	logística e tanh	FSS	101	0,4-0,98	4 - 100

FSS = Fermentação Semi sólida. FS = Fermentação submersa. AG – Algoritmo genético. *Não foi informado

Fonte: Autora (2021).

5.2 SVM

Para o SVM observou-se através das simulações que a melhor combinação de entrada-saída também foi o Modelo 2, com coeficiente de determinação $R^2 = 0,821$, um pouco inferior ao obtido por RNA, que foi de 0,877, mas semelhante.

Vale destacar que a biblioteca Scikit-Learn do Python faz uma busca automática da rede para a melhor configuração dos parâmetros desta máquina. O Scikit-learn expõe uma ampla variedade de algoritmos de aprendizado de máquina, supervisionados e não supervisionados, usando uma interface consistente e orientada a tarefas, permitindo assim uma comparação fácil de métodos para um determinado aplicativo (MICHEL et al., 2012).

Outro ponto a ser analisado é que segundo a literatura, o método SVM estaria mais adequado a trabalhar com um número resumido de dados do que a rede neural, porém nos resultados obtidos nesse trabalho observou-se que essa afirmação não se confirmou, (**Figura 9**), embora os resultados não sejam tão distintos quantitativamente, o que pode ser apenas um resultado mesmo que não definitivo pois com uma maior investigação dos parâmetros do SVM poderia se chegar a uma eficiência melhor do que a RNA. Embora o SVM necessite de menos parâmetros, a RNA é mais difundida (OLIVEIRA *et al.*, 2020).

Porém, em estudos anteriores a RNA possui a melhor relação de tempo de execução e acurácia, a qual pode obter até 99%, embora exija um maior custo computacional em relação ao algoritmo SVM. Contudo, dependendo do tipo de aplicação e necessidade de resposta em menor tempo, pode-se optar pela utilização do SVM. Ambos os algoritmos podem ser eficientes na validação (PALÁCIO et al, 2013). Nesse trabalho, utilizando o modelo 2, obteve-se valores de R^2 próximos para ambas as técnicas.

Uma análise a ser considerada é que as máquinas de aprendizagem, como modelos empíricos, ou seja, onde as características das técnicas não estão diretamente ligadas à natureza dos dados, é que toda a metodologia utilizada nesse trabalho pode ser estendida a mais aplicações, visto que eles dependem da quantidade de dados e não da natureza deles (GOODSWEN et al, 2021).

Não há uma conclusão unânime sobre qual kernel é melhor ou pior para aplicações específicas, alguns autores como realizaram testes para identificar o desempenho do SVM com diferentes kernels, chegando à conclusão geral de que o polinômio e a função RBF gaussiana é a melhor opção para sinais acústicos (ACHIRUL et al, 2018). Por outro lado, Kasnavi et al. (2018) verificaram que o RBF gaussiano e as tangentes hiperbólicas são as melhores para a previsão de um genoma. Para outras aplicações, funções *kernel* menos comuns produzem

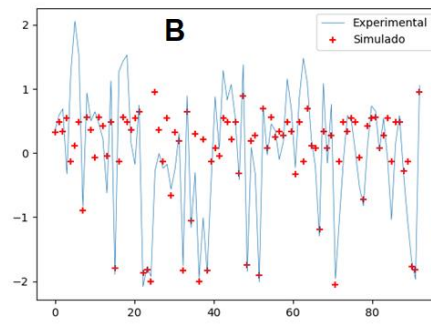
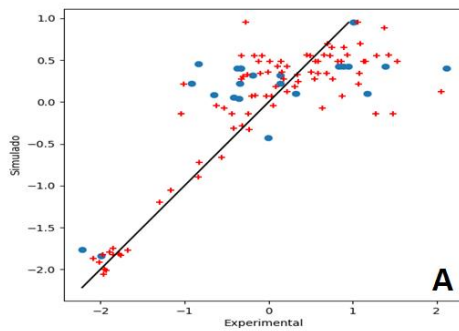
melhores resultados do que o mais popular, por exemplo, Hassan et al. (2016) concluíram que o *kernel* Laplace é ideal para detecção de intrusão. Além disso, *kernels* personalizados podem ser projetados para certas aplicações (CHUI et al, 2019; SAAED et al 2019).

De todo modo, a observação em que todos os autores concordam, é que a seleção da função do kernel deve ser baseada nas características dos dados, e que para obter bons resultados é necessário determinar os parâmetros ideais do kernel usado (CERVANTES et al, 2020). Wang et al. (2010), que estudaram a aplicação de *Support Vector Machine* na predição da fermentação de penicilina, mencionam que as funções de Kernel polinomial, gaussiana ou sigmoidal são as mais usadas em processos fermentativos, e utilizaram a *Gauss Radial Basis Function* (RBF) e *polynomial function*. Zhu e Zhu (2018) produzindo lisozima a partir de biomassa bacteriana também utilizaram a *Gauss Radial Basis Function* (RBF) como função Kernel. Nesse trabalho a função Kernel que obteve melhores resultados foi a polinomial.

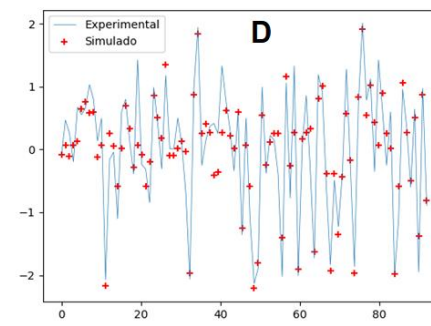
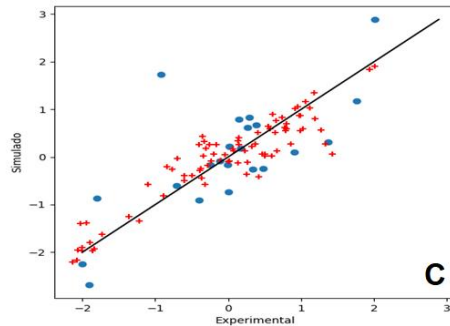
Uma análise a ser considerada é que as máquinas de aprendizagem, como modelos empíricos, ou seja, onde as características das técnicas não estão diretamente ligadas à natureza dos dados, é que toda a metodologia utilizada nesse trabalho pode ser estendida a mais aplicações, visto que eles dependem da quantidade de dados e não da natureza deles (GOODSWEN et al, 2021). Ainda, o uso de *Machine Learning* para sistemas biológicos complexos a partir de pouca quantidade de dados é um desafio e deve ser verificado na identificação de variáveis dominantes que permitam o mapeamento adequado de um bioprocessos e possíveis otimizações que possam advir (SHAIKHINA et al., 2015).

Figura 8 – Resultado do treinamento da SVM com os dados da fermentação. Nos gráficos a), c), e) e g) estão os resultados para treinamento (cruz vermelha) e teste (círculos azuis), enquanto b), d), f) e h) representam a diferença entre os pontos experimentais (cruz vermelha) e simulados (linha azul), para os Modelos 1 ($R^2 = 0,621$), 2 ($R^2 = 0,821$), 3 ($R^2 = 0,778$) e 4 ($R^2 = 0,602$), respectivamente.

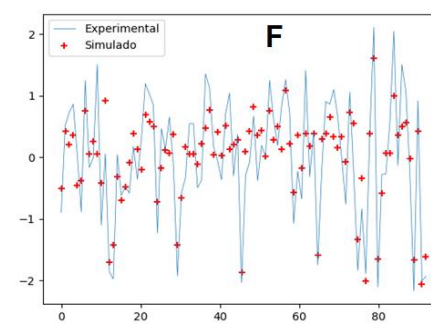
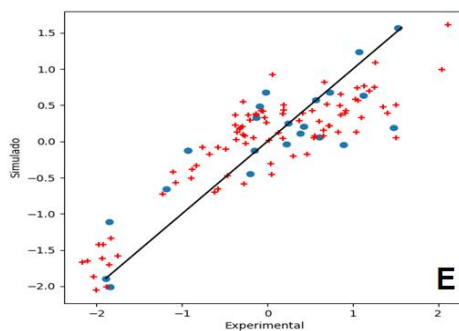
MODELO 1



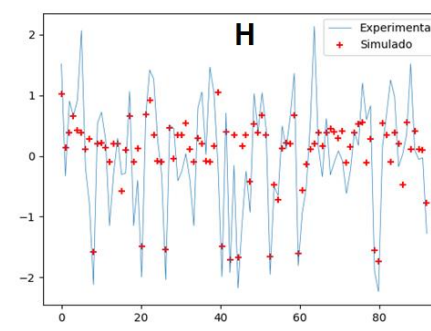
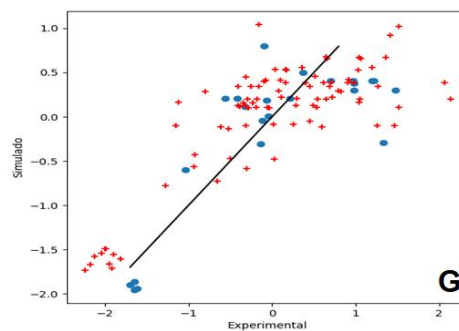
MODELO 2



MODELO 3



MODELO 4



Fonte: Autor (2021).

6 CONCLUSÃO

Ainda que os modelos matemáticos resultantes dessa etapa tenham sido razoáveis na capacidade de simular os processos de fermentação para os quais foram propostos, sabe-se que a complexidade em descrever os bioprocessos com maiores detalhes acaba limitando a eficiência do modelo fenomenológico na tarefa de prever o estado das variáveis de processo. Em virtude dessa dificuldade, Redes Neurais Artificiais e *Support Vector Machine* foram utilizadas como ferramentas de modelagem, construídas para gerar modelos entrada-saída capazes de descrever com qualidade as principais variáveis envolvidas na fermentação semi-sólida de alginase utilizando biomassa de macroalga como substrato.

Este trabalho concluiu que a rede neural artificial forneceu bons ajustes para dados experimentais. A abordagem empírica descrita neste trabalho serve como uma alternativa viável para a modelagem e otimização dos processos de fermentação. Assim como mostra que o acoplamento de redes neurais MLPs e SVM tem boa previsibilidade e precisão na otimização do bioprocessos como multifator. A abordagem proposta aqui também pode ser útil para melhorar a produtividade de outros bioprocessos industriais.

Portanto, o trabalho realizado conseguiu alcançar todos os objetivos propostos, permitindo aprofundar o conhecimento das áreas de Engenharia Bioquímica, otimização, modelagem e simulação através de ferramentas robustas e típicas da Indústria 4.0, as quais raramente são discutidas durante a graduação, além de ter possibilitado o desenvolvimento de habilidades em programação que, cada vez mais, estão sendo exigidas aos novos engenheiros químicos.

REFERÊNCIAS

- ACHIRUL N, et al A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection. **Information**, v. 1, n. 9, p. 5, 2018. doi:10.3390/info9010005
- ADORNO, Wantiê Teles *et al.* Modelagem matemática aplicada á transferência de massa em alimento. **Enciclopédia Biosfera: Centro Científico Conhecer**, Goiânia, v. 9, n. 16, p. 1465-1478, jul. 2013.
- ALVES, Fabiano. **Modelagem e Simulação de Biorreator Operando em Fungos Trametes Versicolor para Produção de Enzima Lacase**. 2010. 91 f. Dissertação (Mestrado) - Curso de Engenharia de Processos Químicos e Bioquímicos, Escola de Engenharia Mauá, Centro Universitário do Instituto Mauá de Tecnologia, São Caetano do Sul, 2010.
- ANDRADE, et al. Produção de proteases por *aspergillus oryzae* em fermentação semi-sólida utilizando farelo de trigo e canola, p. 664-671. In: **Anais do XX Congresso Brasileiro de Engenharia Química - COBEQ 2014. Blucher Chemical Engineering Proceedings**, v.1, n.2. São Paulo, 2015.
- ANGELIA, R. E. e LINSANGAN, N. Classificação do nível de fermentação de grãos de cacau de corte cruzado usando o algoritmo k-NN. In Proceedings of the 2018, 5th International Conference on Bioinformatics Research and Applications. Association for Computing Machinery, p. 64–68, Nova York 2018. doi: <https://doi.org/10.1145/3309129.3309142>
- ARAÚJO, Fernando. et al. Estudo das condições de PH e temperatura para máxima atividade de protease de *Aspergillus oryzae* NRRL 1911. **Brazilian Journal of Development**. V. 6. p.3077-3091, 2020. doi: 10.34117/bjdv6n1-223.
- AZUAJE, F. **Computational models for predicting drug responses in cancer research. Briefings in Bioinformatics**, v. 18, p. 820-829, 2016.
- BAILEY, J. E. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. **Biotechnology Progress**, v. 14, p. 8-20, 1998.
- BAO, J, et al. Mixed Fermentation of Spirulina Platensis with Lactobacillus Plantarum and Bacillus Subtilis by Random-Centroid Optimization. **Food Chemistry**, v. 264, p. 64–72, out. 2018. <https://doi.org/10.1016/j.foodchem.2018.05.027>.
- BARCHI, Augusto César. et al. Artificial intelligence approach based on near-infrared spectral data for monitoring of solid-state fermentation. **Process Biochemistry**, [S.L.], v. 51, n. 10, p. 1338-1347, out. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.procbio.2016.07.017>.
- BEGG, R., PALANISWAMI, M. **Computational Intelligence for Movement Sciences: Neural Networks and Other Emerging Techniques**. ISBN 1-59140-836-9, 2006.
- BELTAGY EA, et al. Purification and characterization of alginate lyase from locally isolated marine *Pseudomonas stutzeri* MSEA04, **Acta Biologica Hungarica**, v. 67, p. 305-317, 2016.

BONOMI, A.; SCHMIDELL, W. Modelagem matemática e simulação de processos fermentativos. In: **Biotecnologia Industrial: engenharia bioquímica**. Editora Edgard Blücher, v. 2. p. 123-178, 2001.

BOUAOUDAT, B. D., et al. Efficient Developments in Modeling and Optimization of Solid State Fermentation. **Biotechnology & Biotechnological Equipment**, v. 6, n. 26, p. 3443–3450, 2012. doi:10.5504/bbeq.2012.0108

BRAGA, A. P. et al. **Redes Neurais Artificiais: Teoria e aplicações**. 2ª ed., Rio de Janeiro: LTC, 2007.

CERVANTES, J., et al. A comprehensive survey on support vector machine classification: applications, challenges and trends. **Neurocomputing**, v. 408, p.189-215, 2020. doi:10.1016/j.neucom.2019.10.118

CHEN P, et al. Purification and characterization of a novel alginate lyase from the marine bacterium *Bacillus sp.* Alg07. **Marine Drugs**, v.86, n. 16, p. 2-13, 2018.

CHUI, K. T., et al. A novel MOGA-SVM multinomial classification for organ inflammation detection. **Applied Sciences**, v. 11, n. 9, p. 2284, 2019.

CRANINX, M., et al. Artificial neural network models of the rumen fermentation pattern in dairy cattle. **Computers and Electronics in Agriculture**, v. 2, n. 60, p. 226–238, 2008. doi:10.1016/j.compag.2007.08.005.

DENG , N. et al. **Máquinas de vetor de suporte: teoria, algoritmos e extensões baseados em otimização**, 1ª ed. , Chapman & Hall / CRC , Nova York, 2012.

DESAI, K. M., et al. Use of an artificial neural network in modeling yeast biomass and yield of β -glucan. **Process Biochemistry**, v. 5, n. 40, p. 1617–1626, 2005. doi:10.1016/j.procbio.2004.06.015

FABRIS, F., et al. **A review of supervised machine learning applied to ageing research**. *Biogerontology*, v. 18, p. 171-188, 2017.

FARINAS C.S. et al, **Modeling the effects of solid state fermentation operating conditions on endoglucanase production using an instrumented bioreactor**, *Ind. Crops Prod.* v. 1, n. 34 p. 1186–1192, 2011.

FISCHER, J., et al. Machine learning techniques applied to lignocellulosic ethanol in simultaneous hydrolysis and fermentation. **Brazilian Journal of Chemical Engineering**, v. 1, n. 34, p. 53–63, 2017. doi:10.1590/0104-6632.20170341s20150475

FRACASSO, A. R. **Cinética de hidratação de soja: estudo comparativo entre soja transgênica e convencional**. 2011. 124 f. Dissertação (Mestrado em Tecnologia de Alimentos) – Universidade Federal do Paraná, Curitiba, 2011.

FREITAS FILHO, P.J. **Introdução à modelagem e simulação de sistemas: com aplicações em Arena**. Florianópolis: Visual Books, 2001.

- GOODSWEN, S. J., et al. Machine learning and applications in microbiology. **FEMS Microbiology Reviews**, p. 1-19, 2021. doi:10.1093/femsre/fuab015
- GUIRY, M, D, 2014. **AlgaeBase. World-wide electronic publication. National University of Ireland, Galway.** Available at: <http://www.algaebase.org>.
- HANNAI, T. et al. **Application of an artificial neural network and genetic algorithm for determination of process orbits in the koji making process. Journal of Bioscience and Bioengineering**, v. 4, n. 87, p. 507–512, 1999. doi:10.1016/s1389-1723(99)80101-7
- HAQ, H. MUKHTAR, Biosynthesis of proteases by *Rhizopus oligosporus* IHS13 in low-cost medium by solid-state fermentation, **J. Basic Microbiol.** v. 4, n. 44, p. 280–287, 2004.
- HARREL, C.H.; TUMAY, K. **Simulation made easy: a manager's guide.** Engineering and Norcross; Management Press, 1995.
- HASSAN, M. A. M., et al. **Performance evaluation of different kernels for support vector machine used in intrusion detection system. International Journal of Computer Networks and Communications**, v. 6, n. 8, p. 39-54, 2016.
- HAYKIN, S. **Neural networks; A comprehensive foundation.** 2nd Ed. Prentice Hall, 1999.
- HECKMANN, D., SCHLÜTER, U., WEBER, A.P.M. **Machine Learning Techniques for Predicting Crop Photosynthetic Capacity from Leaf Reflectance Spectra.** *Molecular Plant*, v. 10, p. 878-890, 2017.
- HIFNEY AF, et al. Upgrading the antioxidant properties of fucoxanthin and alginate from *Cystoseira trinodis* by fungal fermentation or enzymatic pretreatment of the seaweed biomass. **Food Chemistry** v. 269, p. 387-395, 2018.
- HUANG, G.B. Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Trans. Neural Netw.* v.14, p. 274–281, 2002.
- IPLIKCI, S. Support vector machines-based generalized predictive control. **International journal of robust and nonlinear control.** *Int. J. Robust Nonlinear Control* v.16, p. 843–862, 2006.
- KANDOI, G., et al. Prediction of druggable proteins using machine learning and systems biology: A mini-review. **Frontiers in Physiology**, v. 6, p. 366, 2015.
- KARRI, R R, et al.; Optimization and modeling of methyl orange adsorption onto polyaniline nano-adsorbent through response surface methodology and differential evolution embedded neural network, **J. Environ. Manage.** v. 223 p.517–529, 2018. doi:10.1016/j.jenvman.2018.06.027.
- KASNAVI, S. A., et al. O efeito da seleção do kernel na predição ampla do genoma de traços discretos por Support Vector Machine. **Gene Reports**, v. 11, p. 279-282, 2018. doi: 10.1016 / j.genrep.2018.04.006.
- KIMUTAI, G, et al. An Optimum Tea Fermentation Detection Model Based on Deep Convolutional Neural Networks. **Data**, v. 5, n. 2, p. 44, abril 2020. <https://doi.org/10.3390/data5020044>

KONNO N, et al. Crystal structure of polysaccharide lyase Family 20 endo- β -1,4-glucuronan lyase from the filamentous fungus *Trichoderma reesei*. *FEBS Letters*. v. 583, p. 1323-1326, 2009. <https://doi.org/10.1016/j.febslet.2009.03.034>.

KOTSIANTIS, SB, et al. Aprendizado de máquina: uma revisão das técnicas de classificação e combinação. *Artif Intell Ver*, v. 26, p. 159–190 2006. <https://doi.org/10.1007/s10462-007-9052-3>.

LARRAÑAGA, P., et al. **V. Machine learning in bioinformatics**. Briefings in Bioinformatics, v. 7, p. 86-112, 2006.

LEK, S., GUEGAN, J.F. **Artificial neural networks as a tool in ecological modelling, an introduction**. *Ecol. Model.* v.120, 65–73, 1999.

LI S, et al. Combined enzymatic hydrolysis and selective fermentation for green production of alginate oligosaccharides from *Laminaria japonica*. *Bioresource Technology*, v. 281, p. 84-89, 2019.

LI S, et al. Purification and characterization of a new alginate lyase from marine bacterium *Vibrio sp.* SY08. *Marine Drugs*, v. 1, n. 15, p. 1-11, 2017.

LOBÃO, E.C.; PORTO, A.J.V. Proposta de sistematização de estudos de simulação. *Revista de Engenharia e Arquitetura - EESC/USP*, São Carlos, SP, p. 61-69, 1996.

LUCENA, Amarildo J. F. *et al.* Utilização de modelagem e simulação no processo de atendimento de chamados em centrais de operações da polícia militar para otimização do uso de viaturas. *Brazil. J. Of Develop*, Curitiba, v. 6, n. 6, p. 34086-34103, jun. 2020.

MALHIS, N., et al. Computational Identification of MoRFs in Protein Sequences Using Hierarchical Application of Bayes Rule. *PlosOne*, v. 10, p. 1-15, 2015.

MICHEL, V., et al. A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recognition*, v. 6, n. 45, p. 2041–2049, 2012. doi:10.1016/j.patcog.2011.04.006.

MILLER JG .Use of dinitrosalicylic acid reagent for determination of reducing sugars. *Anal. Chem.* v. 3, n. 31, p. 426–428, 1959. <https://doi.org/10.1021/ac60147a030>.

MOHAPATRA, B.R. Biocatalytic characteristics of chitosan nanoparticle-immobilized alginate lyase extracted from a novel *Arthrobacter* species AD-10. *Biocatalysis and Agricultural Biotechnology*, v. 23 n. 101458 p.1-8, 2020.

MONROY-DE-JESÚS, Juan *et al.* Clasificador de plantas medicinales por medio de Deep Learning. *Research In Computing Science*, México, v. 7, n. 148, p. 65-78, mar. 2019.

MORABITO, R.; PUREZA, V. Modelagem e simulação. In: CAUCHICK MIGUEL, P.A.C. et al. Metodologia de pesquisa em engenharia de produção e gestão de operações. Rio de Janeiro: Elsevier, p.165-192, 2010.

MURTAGH, F. **Multilayer perceptrons for classification and regression.** *Neurocomputing*, v. 5-6, n. 2, p.183–197, 1991. doi:10.1016/0925-2312(91)90023-5

NEHA SHARMA & KAILASH SINGH. Neural network and support vector machine predictive control of tert-amyl methyl ether reactive distillation column, **Systems Science & Control Engineering: An Open Access Journal**, v. 2, n.1, p. 512-526, 2014. DOI: 10.1080/21642583.2014.924082.

OHBA, M. S. **Modelagem matemática da produção de glicoamilase por Aspergillus awamori NRRL 3112.** 1998. 213 p. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo, São Paulo, 1998.

OLIVEIRA, Arthur Vitor Matias de *et al.* Estudo Comparativo de Desempenho Entre SVM e MLP no Reconhecimento de Imagens. In: ENCONTRO DE COMPUTAÇÃO DO OESTE POTIGUAR, 4, 2020, Mossoró. **Anais [...]**. Mossoró: Ufersa, 2020. p. 1-8.

ÖREN, T.; YILMAZ, L. Synergies of simulation, agents, and systems engineering. **Expert Systems with Applications**, v.39, p.81-88, 2012.

PALACIOS, R. H. C. et al. **Estimador de Torque em Motores de Indução Trifásicos com Alimentação Desequilibrada Baseado em Redes Neurais Artificiais.** In: **Conferência Brasileira sobre Qualidade da Energia Elétrica (CBQEE)**, 2013, Araxá - MG. Proceedings of Conferência Brasileira sobre Qualidade da Energia Elétrica 2013. p. 1-8.

PANDEY, A. C. et al. **Biofuels: Alternative Feedstocks and Conversion Processes**, Elsevier, San Diego, United States of America, p. 177–201, 2011.

PAPI M, et al. Detection of biofilm-grown *Aspergillus fumigatus* by Means of Atomic Force Spectroscopy: Ultrastructural Effects of Alginate Lyase. **Microscopy and Microanal.** v.5, n. 18, p.1088-1094, 2012. <https://doi.org/10.1017/S1431927612001067>.

PAPPU S M J, GUMMADI S N, Artificial neural network and regression coupled genetic algorithm to optimize parameters for enhanced xylitol production by *Debaryomyces nepalensis* in bioreactor. **Biochem. Eng. J.**, v. 120, p. 136-145, 2017.

PEGDEN, C.D.; SHANNON, R.E.; SADOWSKI, R.P. Introduction to simulation Using SIMAN. **New York: McGraw-Hill**, 1995.

PEREIRA, Carla Roberta; COSTA, Miguel Antonio Bueno da. Um modelo de simulação de sistemas aplicado à programação da produção de um frigorífico de peixe. **Revista Produção Online**, Florianópolis, v. 12, n. 4, p. 972-1001, dez. 2012.

PIAZZI, Ana Carolina Ferreira. **Construção de modelos matemáticos para processos fermentativos – avaliação de efeitos das condições de cultivo.** 2019. 143 f. Dissertação (Mestrado) - Curso de Pós Graduação em Engenharia Química, Universidade Federal de Santa Maria, Santa Maria-Rs, 2019.

PINTO, G. A. S. et al. Fermentação em estado sólido: Uma alternativa para o aproveitamento e valorização de resíduos agroindustriais. **Revista de Química Industrial**, v.74, n.724, p.17-20, 2006.

PREISS, J, ASHWELL G. Alginic acid metabolism in bacteria: I. enzymatic formation of unsaturated oligosaccharides and 4-deoxy-L-erythro-5-hexoseulose uronic acid. **J. Biol. Chem**, n.23, p.309– 316, 1962. [https://doi.org/10.1016/S0021-9258\(18\)93920-7](https://doi.org/10.1016/S0021-9258(18)93920-7). .

QIAO L, et al. Efficient production of ulvan lyase from *Ulva prolifera* by *Catenovulum sp.* LP based on stage-controlled fermentation strategy. **Algal Research**, v. 46, p. 101-112, 2020.

RAYOVARAPU, B.; TALLOPRAGADA, P.; USHA, M.S. Statistical optimization of γ -aminobutyric acid production by response surface methodology and artificial neural network models using *Lactobacillus fermentum* isolated from palm wine. **Biocatalysis and Agricultural Biotechnology**, v. 22, 1010362 p. 1-10, 2019.

REIS, Carlos Henrique. **Otimização de Hiperparâmetros em Redes Neurais Profundas**. 2018. 72 f. TCC (Graduação) - Curso de Ciências da Computação, Unifei, Itajubá, 2018.

SAEED, et al. **Pertanika Journal of Science & Technology**, v. 27, n. 1, p 527-545, jan. 2019.

SANTOS, E. M. **Teoria e Aplicação de Support Vector Machines à aprendizagem e Reconhecimento de Objetos Baseado na Aparência**. Dissertação de Mestrado, Universidade Federal da Paraíba, Campina Grande, Paraíba, Brasil, 2002.

SAPUTRO, DRS & WIDYANINGSIH, P. Método de memória limitada Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) para a estimativa de parâmetros em modelo de regressão logística ordinal ponderada geograficamente (GWOLR). Em **AIP Conference Proceedings** , v. 1868, n. 040009, p 1-9, 2017.

SARAIVA, S, V. **Estudo de um controlador preditivo baseado em sistemas inteligentes**. Trabalho de Conclusão de Curso. Faculdade de Engenharia Química, Universidade Federal de Alagoas, 2017.

SARAIVA, Samuel Vitor *et al.* PROGNÓSTICO DE RADIAÇÃO SOLAR ATRAVÉS MODELOS QUE COMBINAM AS TÉCNICAS WAVELET E REDES NEURAIS. **Revista Brasileira de Geografia Física**, Recife, v. 7, n. 5, p. 808-817, ago. 2014. Bimensal.

SCHAUMANN, K. E WEIDE, G. Enzymatic degradation of alginate by marine fungi, Thirteenth International Seaweed Symposium (SC Lindstrom, PW Gabrielson, eds) **Hydrobiologia**, v. 205, n. 204, p. 589-596, 1990.

SCHOLKOPF, B., SMOLA, A. J. **Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond**. ISBN 0-262-19475-9, 2002.

SHAIKHINA, et al. Machine Learning for Predictive Modelling based on Small. **Data in Biomedical Engineering**. v. 48, p. 469-474, 2015. Doi 10.1016/j.ifacol.2015.10.185.

SHANKAR T, et al. Purification, characterization and immobilization of alginate produced by *Bacillus sp.* associated with *Sargassum wightii*. **Universal Journal of Microbiology Research**, v. 1, n. 4, p.11-22, 2016.

SHANNON, R.E. **Introduction to the art and science of simulation**. In: winter simulation conference, 1998, Texas. **Anais...** Texas: Texas A & M University, 1998.

SILVA, R.G. et al. Ahybrid feedforward neural network model for the cephalosporin C production process. **Brazilian Journal of Chemical Engineering**, São Paulo, v. 17, n. 4-7, p. 587 – 597, 2000.

SIMÕES, P, S. R. **Estudo sobre a viabilidade da simulação e predição de efeitos difusivos em reações de polimerização através de modelos determinísticos e redes neurais**. Tese (Mestrado em Engenharia Química) – Faculdade de Engenharia Química, Universidade Estadual de Campinas, Campinas, 2001.

SINCLAIR, C. G.; KRISTIANSEN, B. **Fermentation kinetics and modeling**. Grã-Bretanha: Open University Press, New York, 1987. p 113.

SINGH RP, et al. Purification and partial characterization of na extracelular alginate lyase from *Aspergillus oryzae* isolated from brown seaweed. **Journal of Applied Phycology**, v. 23, p.755-776, 2011.

SIVAKUMAR T, SATHYA C, SHANKAR, T, PONMANICKAM P. Screening and optimization of alginase producing *Bacillus sp.* From seaweed. **Journal Research** article. v. 2, n. 14, p. 42- 429, 2015.

SUN X, et al. Heterologous expression and purification of a marine alginate lyase in *Escherichia coli*. **Protein expression and purification**, v. 153, p. 97-104, 2019.

SWIFT SM, et al. Characterization of AlgMsp, an alginate lyase from *Microbulbifer sp.* 6532A. **Plos One**. v. 11, n.9, p.1-11, e112939, 2014.

TAVAFI H, et al. Screening and optimization of media compositions for extracelular *alginate lyase* production. **Iranian Biomedical Journal**, v.1, n. 21, p. 48-56, 2017.

TOLEDO, Karina. **Estudo decifra mistério das algas que invadiram o Brasil**. 2017. Elaborada pela repórter da Agência Fapesp. Disponível em: <https://exame.com/ciencia/estudo-decifra-misterio-das-algas-que-invadiram-o-brasil/#:~:text=Embora%20as%20algas%20do%20g%C3%AAnero,o%20banho%20de%20mar%20invi%C3%A1vel..> Acesso em: 07 set. 2021.

TORRES M. R.et al. Extraction and physicochemical characterization of *Sargassum vulgare* alginate from Brazil. **Carbohydrate Research** n. 342, p. 2067-2074, 02 jun 2007.

TORRES, N. V.; SANTOS, G. The (Mathemayical) Modeling Process in Biosciences. **Frontiers in Genetics**, v. 6, p. 1 - 9, 2015.

VAN CAN, H.J.L, et al. Strategy for Dynamic Process Modelling Based on Neural Networks in Macroscopic Balances, **AIChE Journal**, v. 42, n. 12, p. 3403-3418, 1997.

VAPNIK, V. N. **The Nature of Statistical Learning Theory**. Springer Verlag, New York, 2nd edition, 1999.

VIEIRA, W. F., et al. A novel multiple reactor system for the long-term production of L-asparaginase by *Penicillium* sp. LAMAI 505. **Process Biochemistry**, [S.L.], v. 90, p. 23-31, mar. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.procbio.2019.11.012>.

VOLESKY, B.; VOTRUBA, J. Modeling and optimization of fermentation processes. **Amsterdã: Elsevier**, p 266, 1992.

WANG M, et al. Screening of *alginate lyase* – excreting microorganisms from the surface of brown algae. **AMB Express**, v. 74, n. 7, p.1-9, 2017.

WANG, L. **Support Vector Machines: Theory and Applications**. ISSN print edition: 1434-9922, 2005.

WANG, X. *et al.* Modelagem híbrida do processo de fermentação da penicilina com base na máquina de vetor de suporte mínimo. **Elsevier**: Pesquisa e projeto de engenharia química, v. 88, n. 4, p. 415-420, abr. 2010.

WONG TY, et al. Alginate Lyase: Review of major sources and enzyme characteristics, structure-function analysis, biological roles, and applications. **Annual Review Microbiol.** v. 54 p.289- 340, 2000. <https://doi.org/10.1146/annurev.micro.54.1.289>.

ZHENG, Z. Y, et al. Artificial neural network - Genetic algorithm to optimize wheat germ fermentation condition: Application to the production of two anti-tumor benzoquinones. **Food Chemistry**, v. 227, p.264-270, 2017. <https://doi.org/10.1016/j.foodchem.2017.01.077>.

ZHU B, YIN H Alginate lyase: Review of major sources and classification, properties, structure-function analysis and applications. **Bioengineered.** V. 3, n. 36, p. 125-131, 2015. <https://doi.org/10.1080/21655979.2015.1030543>.

ZHU X, ZHU Z. O controle preditivo generalizado da concentração de bactérias no processo de fermentação da lisozima marinha. **Food Sci Nutr**, v.8, n. 6, p. 2459-2465, out 2018. doi: 10.1002 / fsn3.850

ZHU Y, et al. Characterization of na extracelular biofunctional alginate lyase from marine *Microbulbifer* sp. ALW1 and antioxidante activity of enzymatic hydrolysates. **Microbiological Research**, v. 182, p. 49-58, 2016.

ZHU, B, et al. Achieving the Carbon Intensity Target of China: A Least Squares Support Vector Machine with Mixture Kernel Function Approach. **Applied Energy**, v. 233–234, p. 196–207, jan. 2019. doi.org <https://doi.org/10.1016/j.apenergy.2018.10.048>.