



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Especialização em Ciência de Dados



Predição da eficiência de redução de uma Caldeira de Recuperação usando Rede Neural Artificial

Paula Camila Mendes Gomes Gaspar Botrel

João Monlevade, MG
2024

Paula Camila Mendes Gomes Gaspar Botrel

Predição da eficiência de redução de uma Caldeira de Recuperação usando Rede Neural Artificial

Trabalho de conclusão de curso apresentado ao curso de Ciência de Dados do Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto, como parte dos requisitos necessários para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Dr. Matheus Nohra Haddad

João Monlevade, MG

2024

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

B749p Botrel, Paula Camila Mendes Gomes Gaspar.
Predição da eficiência de redução de uma caldeira de recuperação
usando rede neural artificial. [manuscrito] / Paula Camila Mendes Gomes
Gaspar Botrel. - 2024.
31 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Matheus Nohra Haddad.
Produção Científica (Especialização). Universidade Federal de Ouro
Preto. Departamento de Engenharia de Produção.

1. Caldeiras. 2. Celulose - Eficiência industrial. 3. Previsão
(Estatística). 4. Redes neurais (Computação). I. Haddad, Matheus Nohra.
II. Universidade Federal de Ouro Preto. III. Título.

CDU 004.8:519.2

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



FOLHA DE APROVAÇÃO

Paula Camila Mendes Gomes Gaspar Botrel

Predição da Eficiência de Redução de Uma Caldeira de Recuperação Usando Uma Rede Neural Artificial

Trabalho de conclusão de curso apresentado ao curso de Especialização em Ciência de Dados da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Especialista em Ciência de Dados

Aprovada em 26 de junho de 2024

Membros da banca

Prof. Dr. Matheus Nohra Haddad - Orientador
Dr. Gustavo Matheus de Almeida
Dra. Luciana Cerqueira Souza Solimani

Matheus Nohra Haddad, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 14/08/2024



Documento assinado eletronicamente por **Matheus Nohra Haddad, PROFESSOR DE MAGISTERIO SUPERIOR**, em 14/08/2024, às 16:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0759622** e o código CRC **0E78878B**.

Agradecimentos

Quero agradecer a todas as pessoas e vivências que tornaram essa jornada menos difícil e que colaboraram com a minha formação. Em especial, expresso minha gratidão para a minha família pela paciência e por sempre apoiarem as minhas decisões e a Deus por me dar sabedoria e discernimento.

Agradeço ao professor Matheus, pela orientação durante a execução desse trabalho, e também às professoras e aos professores que contribuíram fortemente com meu desenvolvimento e com a construção de novos conhecimentos durante a pós graduação.

Agradeço aos meus colegas de turma pelo apoio nos momentos de dificuldade, por compartilharem seus entendimentos e auxiliarem na resolução das dúvidas. Em especial, ao Daniel e ao Paulo, pela parceria durante todos os trabalhos que desenvolvemos juntos.

Agradeço aos meus colegas de trabalho por compreenderem minha ausência durante os períodos das aulas e ao Flávio e ao Ronaldo pela mentoria que contribuiu para o aprimoramento desse trabalho.

Agradeço à Cenibra pela oportunidade de realizar essa pós-graduação e pela disponibilização dos dados para realização deste trabalho.

“Dados são nossa mais inesgotável fonte de magia”
(Paráfrase de Alvo Dumbledore, Harry Potter e as Relíquias da Morte)

Resumo

A eficiência de redução é um ponto relevante na operação das caldeiras de recuperação e estando bem ajustada é possível ter uma economia considerável nos custos de produção de celulose. A obtenção desse parâmetro é feita a partir de análises laboratoriais, o que demanda tempo e capital humano, fazendo com que os ajustes na combustão das caldeiras de recuperação aconteçam de forma defasada e lenta. Nesse contexto, o objetivo desse trabalho foi construir um modelo para prever a eficiência de redução da caldeira de recuperação e identificar as variáveis do processo que mais influenciam nas reações de redução do sulfato de sódio a sulfeto. Foi realizado o estudo de uma caldeira de recuperação de uma fábrica de celulose do Brasil e constatado que a vazão do licor preto e a vazão de óleo queimado são as variáveis de maior correlação com a eficiência de redução, sendo que essa relação se dá de forma direta e indireta, respectivamente. Usando essas variáveis e outras dez, como vazão dos níveis de ar, concentração e temperatura do licor preto, foram construídos modelos de redes neurais artificiais do tipo Perceptron Multicamadas com a variável resposta nas formas contínua, sem e com validação cruzada, e categórica, sem e com balanceamento de classes, a fim de identificar qual mais se adéqua a realidade do problema. A partir do *keras tuner*, obteve-se os melhores hiperparâmetros, dentro de um espaço de busca, para treinar as redes. Com os modelos construídos, pode-se verificar que não é trivial prever a eficiência de redução na forma contínua e que trabalhar com classes é mais factível. O melhor modelo construído foi utilizando três classes de eficiência de redução - baixa (<90%), normal (90% a <95%) e alta (>=95%) - e sem balanceá-las. Para este modelo, a acurácia global foi 79,2% e para a classe baixa, de maior interesse, foi 62,5% e o percentual de falsos-positivos para a classe baixa foi 12,8. Dessa forma, a utilização desse modelo de rede neural se mostra viável visto que pode possibilitar à equipe operacional a realização de ajustes mais rápidos e precisos nos parâmetros da caldeira de recuperação, trazendo ganhos em todas as etapas do processo.

Palavras-chaves: Caldeira de recuperação. Eficiência de redução. Predição . Rede neural artificial.

Abstract

Reduction efficiency is a relevant point in the recovery boilers operation and, if it is well adjusted, it is possible to achieve considerable savings in pulp production costs. This parameter is obtained through laboratory analyses, which requires time and human capital, causing adjustments to the combustion of recovery boilers to take place in a delayed and slow manner. In this context, the objective of this work was to build a model to predict the reduction efficiency of the recovery boiler and identify the process variables that most influence the sodium sulfate reduction reactions to sulfide. A study was carried out on a recovery boiler at a Brazilian pulp mill and it was found that the flow rate of black liquor and flow rate of burned oil are the variables with the greatest correlation with reduction efficiency with this relationship occurring directly and indirectly, respectively. Using these variables and ten others, such as flow rate of the air levels, concentration and temperature of burned black liquor, some MultiLayer Perceptron artificial neural network models were built with the response variable in continuous forms, without and with cross-validation, and categorical, imbalanced and balanced classes, in order to identify which one best suits the reality of the problem. Using Keras Tuner, the best hyperparameters were obtained, within a search space, to train the networks. With the models built, it can be seen that it is not trivial to predict the reduction efficiency in continuous form and that working with classes is more feasible. The best model constructed was using three reduction efficiency classes - low (<90%), normal (90% to <95%) and high (>=95%) - with imbalanced data. For this model, the global accuracy was 79.2% and for the low class, of greatest interest, it was 62.5% and the percentage of false positives for the low class was 12.8. Therefore, the use of this model of neural network model proves to be feasible, since it can enable the operational team to make faster and more accurate adjustments to the recovery boiler parameters, bringing gains at all stages of the process.

Keywords: Recovery Boiler. Reduction Efficiency. Prediction. Artificial Neural Network.

Lista de ilustrações

Figura 1 – Representação esquemática de uma Caldeira de Recuperação.	4
Figura 2 – Diagrama da metodologia adotada no trabalho.	9
Figura 3 – Gráficos de dispersão das variáveis de entrada com a eficiência de redução. .	15
Figura 4 – Mapa de calor para o coeficiente de correlação de Pearson para as variáveis de entrada e a variável de saída.	16
Figura 5 – Boxplot do grupo de treinamento e de teste para duas variáveis de entrada. .	17
Figura 6 – Gráfico com o coeficiente de determinação de cada modelo construído. . . .	19
Figura 7 – Gráfico com as estimativas do modelo 7 e os valores reais da eficiência de redução.	19
Figura 8 – Comportamento do erro a partir do gráfico de dispersão e do histograma. . .	20
Figura 9 – Acurácia global e por faixas obtida com os 10 modelos construídos.	20
Figura 10 – Matriz de confusão para modelos contínuos.	21
Figura 11 – Comportamento das redes treinadas por meio da validação cruzada.	22
Figura 12 – Matriz de confusão dos modelos construídos utilizando a validação cruzada.	22
Figura 13 – Acurácia global e por classe obtida com os 10 modelos categóricos construídos	24
Figura 14 – Matriz de confusão do modelo categórico 10	24
Figura 15 – Acurácia global e por classe obtida com os 10 modelos categóricos construídos usando balanceamento de classes	25
Figura 16 – Matriz de confusão dos modelos categóricos construídos usando balanceamento de classes.	26

Lista de tabelas

Tabela 1 – Avaliação dos modelos em função do total de parâmetros usados.	27
Tabela 2 – Avaliação dos modelos em função da acurácia global e para cada classe de eficiência de redução.	27
Tabela 3 – Avaliação dos modelos em função dos falsos positivos para cada classe de eficiência de redução.	27

Lista de quadros

Quadro 3.1 – Variáveis analisadas por cada circuito da Caldeira de Recuperação.	10
Quadro 3.2 – Espaço de busca utilizado no tuner para construção do melhor modelo. . .	12
Quadro 4.1 – Variáveis de entrada utilizadas na construção do modelo para prever a eficiência de redução	14
Quadro 4.2 – Hiperparâmetros do modelo de rede MLP para predição da eficiência de redução.	18
Quadro 4.3 – Hiperparâmetros do modelo de rede MLP para predição das clases da efici- ência de redução	23

Lista de abreviaturas e siglas

Na₂S Sulfeto de sódio

Na₂SO₄ Sulfato de sódio

MLP *Multi Layer Perceptrons*

PIMS *Plant Information Management System*

RNA Rede Neural Artificial

SMOTE *Synthetic Minority Oversampling Technique*

Sumário

1	INTRODUÇÃO	1
1.1	Objetivo geral	1
1.1.1	Objetivos específicos	2
1.2	Justificativas	2
1.3	Organização do Trabalho	3
2	REVISÃO DA LITERATURA	4
2.1	Caldeira de Recuperação Química: recuperação dos compostos inorgânicos	4
2.2	Rede neural artificial	6
3	METODOLOGIA	9
4	RESULTADOS	14
4.1	Pré-Processamento	14
4.2	Modelo de Rede Neural Artificial	18
5	CONSIDERAÇÕES FINAIS	28
	REFERÊNCIAS	30

1 Introdução

A Caldeira de Recuperação é o equipamento complexo e crucial no processo de produção de celulose kraft. Nela o licor preto proveniente do cozimento da madeira é queimado, gerando energia térmica. Diferente das demais caldeiras que são exclusivamente ambientes de oxidação de materiais, na Caldeira de Recuperação existe também uma zona redutora, na qual acontece a conversão do **Sulfato de sódio** (Na_2SO_4) a **Sulfeto de sódio** (Na_2S). A efetividade dessa conversão é conhecida como eficiência de redução e é calculada pela equação (1.1) (VAKKILAINEN *et al.*, 2005). Esse parâmetro é importante para o processo kraft e imprescindível para sua viabilidade econômica (MACCALLUM, 2015).

$$\text{Eficiência de redução} = \frac{\text{Na}_2\text{S}}{\text{Na}_2\text{S} + \text{Na}_2\text{SO}_4} \quad (1.1)$$

Para que as reações de redução aconteçam de forma satisfatória e que a reoxidação, gerando sulfato, seja minimizada, os parâmetros da caldeira de recuperação, como vazão, pressão e temperatura de cada nível de ar e dos combustíveis utilizados na queima, precisam estar bem regulados (MACCALLUM, 2015; DAMASCENO *et al.*, 2020). Porém, como a eficiência de redução é obtida a partir de rotinas laboratoriais morosas, há escassez de informação no momento adequado o que prejudica os ajustes operacionais da caldeira a fim manter uma boa eficiência de redução (<90%).

A fim de melhorar a operação da caldeira, foram criados sensores virtuais para prever em tempo real a eficiência de recuperação a partir de variáveis medidas online. Segundo Souza, Araújo e Mendes (2016), os sensores virtuais já são aplicados em processos industriais importantes, como em fábricas de celulose e papel, para variáveis difíceis de medir. Como há grande disponibilidade de dados reais, neste trabalho os sensores foram construídos a partir de modelos de predição usando redes neurais artificiais. Além disso, foi realizada a identificação dos parâmetros que afetam a eficiência de recuperação da caldeira em estudo e como se dá essa relação, visto que na literatura, existem algumas variáveis, por exemplo, ar secundário, cuja influência não é consenso (DAMASCENO *et al.*, 2020; MACCALLUM, 2015) e ainda algumas variáveis que não foram estudadas.

1.1 Objetivo geral

O objetivo geral deste trabalho é desenvolver um modelo que faça a predição da eficiência de redução da Caldeira de Recuperação Química 3 da Cenibra.

1.1.1 Objetivos específicos

Para cumprimento do objetivo geral é necessário atender aos seguintes objetivos específicos:

- Identificar quais as variáveis mais influenciam na eficiência de redução da Caldeira de Recuperação Química;
- Identificar se há correlação entre as variáveis que influenciam na eficiência de redução da Caldeira de Recuperação Química e como essa influência acontece;
- Avaliar os modelos de redes neurais buscando identificar qual mais se adéqua à realidade empírica do problema da Caldeira de Recuperação 3 da Cenibra.

1.2 Justificativas

Atualmente a eficiência de redução na Caldeira de Recuperação 3 da Cenibra é obtida a partir de análises realizadas no laboratório em amostras de licor verde e com frequência de 3 vezes ao dia. Em vista disso, as ações para corrigir baixas eficiências (<90%) são sempre defasadas e demandam muito tempo para chegar na condição ideal.

A disponibilização, em tempo real, do valor predito para a eficiência de redução a partir de dados do processo possibilitará ao operador um melhor entendimento do comportamento da Caldeira de Recuperação. Com isso, será possível a realização de ajustes mais precisos nos parâmetros citados previamente e com tempo de resposta reduzido, o que proporcionará um melhor controle operacional da caldeira e maior estabilidade em todo processo. Com a manutenção da eficiência de redução em patamares ótimos (>90%) alguns ganhos no processo podem ser atingidos como a redução da carga morta do licor e a redução da geração de dregs (resíduo inorgânico do circuito de recuperação do licor).

A redução da carga morta do licor pode diminuir o processo de formação de incrustações no circuito. Isso acontece porque o decréscimo de Na_2SO_4 aumenta o limite de solubilidade e o ponto de cristalização, dificultando a precipitação dos sais (FREDERICK *et al.*, 2004). Dessa forma, o licor fica com menor grau de sujidade o que melhora o desempenho das plantas de evaporação e pode resultar em maior produção de celulose. Além disso, ao reduzir a carga morta, reduz-se também o consumo de energia que é dispendido para evaporar e aquecer a água associada a carga morta nos digestores, evaporadores e caldeiras de recuperação. Segundo Grace e Tran (2009), o consumo energético de cada kg de carga morta é de 5535 kJ/kg. Uma redução de 1 kg de carga morta por tonelada de celulose produzida geraria uma economia de vapor de 2kg/tSA, o que geraria a redução de consumo de 170 toneladas de óleo por ano.

Por sua vez, a redução da geração de dregs tem um retorno ambiental relevante visto que esse resíduo ainda não tem uma forma de ser reaproveitado em larga escala e ele é majoritariamente disposto em aterro industrial (RODRIGUES *et al.*, 2019). A menor geração de dregs reduziria o custo com o transporte do material e a utilização do aterro industrial. Além disso, há o impacto no processo de caustificação, visto que a menor quantidade de dregs aumenta a capacidade dos equipamentos e reduz o tempo de retenção (CHANDRA; EMPIE, 2007).

Em vista dos impactos expostos em operar com baixa eficiência de recuperação e a baixa frequência da realização da análise que determina esse parâmetro, o presente trabalho tem por finalidade construir um modelo que seja capaz de prever com boa acurácia a eficiência de recuperação a partir de dados de processo mensurados de forma online.

1.3 Organização do Trabalho

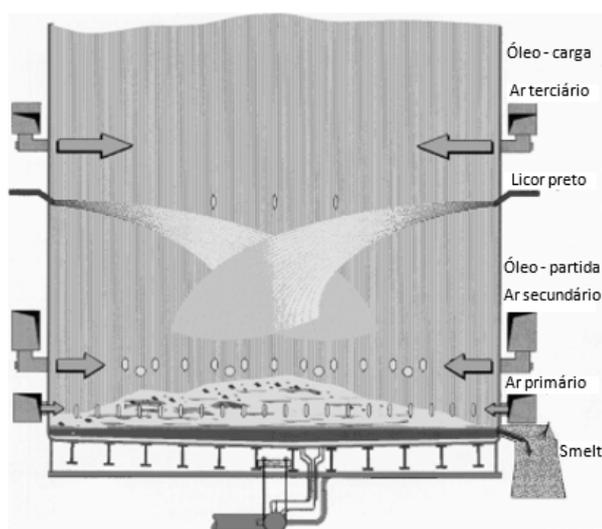
O texto foi estruturado da seguinte forma: no capítulo 2 é realizada uma revisão de literatura sobre caldeiras de recuperação na indústria de celulose com ênfase na eficiência de redução, bem como sobre o método utilizado para a construção dos modelos de predição: a rede neural artificial. O capítulo 3 apresenta a metodologia adotada para atingir os objetivos do trabalho, com os critérios adotados na construção dos modelos. Por sua vez, o capítulo 4 aborda os resultados alcançados, uma comparação entre os modelos construídos e uma avaliação crítica de cada um deles. Finalmente, no capítulo 5 são apresentadas as considerações finais e sugestões de algumas pesquisas futuras a partir do que foi desenvolvido nesse trabalho.

2 Revisão da Literatura

2.1 Caldeira de Recuperação Química: recuperação dos compostos inorgânicos

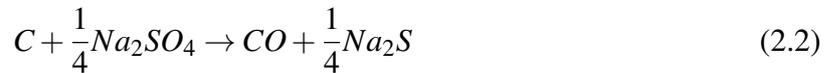
A Caldeira de Recuperação Química pode ser dividida em duas partes: a fornalha, onde acontecem as reações químicas e a seção de transferência de calor, onde a água é transformada em vapor superaquecido. Nesse trabalho, será dado foco à região da fornalha, a qual está representada na Figura 1. Nesta região é onde os combustíveis, licor preto e óleo - em menor proporção - e o ar, em diferentes níveis, são inseridos para que a combustão aconteça. Também é onde acontecem as reações químicas que proporcionam a recuperação do sulfeto de sódio.

Figura 1 – Representação esquemática de uma Caldeira de Recuperação.



Fonte: adaptado de [Vakkilainen et al. \(2005\)](#)

A combustão do licor preto é tipicamente dividida em três etapas: secagem, pirólise e combustão do leito carbonizado ([VAKKILAINEN et al., 2005](#)). O combustível é injetado na caldeira com cerca de 70% de sólidos e a primeira etapa da combustão consiste na evaporação da água das gotas do licor. Na sequência, a pirólise se inicia, promovendo reações de degradação térmica da parte orgânica do licor com a liberação de gases voláteis. Por fim, acontece a combustão do leito carbonizado - também chamado de camada - o qual é uma pilha de material formada por carbono, carbonato de sódio, sulfato e sulfeto de sódio. Durante essa etapa acontecem as reações de redução, nas quais o sulfato de sódio reage com o carbono originando o sulfeto de sódio, conforme as equações (2.1)-(2.3):



A eficiência de redução é a medida da quantidade de sulfato que foi convertida a sulfeto, ela é geralmente realizada a partir de rotinas laboratoriais, mas já estão disponíveis tecnologias capazes de disponibilizar esse parâmetro com frequência elevada seja a partir de dispositivos que fazem a coleta da amostra e as titulações de forma automática e sem intervenção humana ou por tecnologia visual combinada com algoritmos. Quanto maior é esse parâmetro mais ajustado está o processo e menor quantidade de sódio inutilizável chega na etapa de cozimento. As condições operacionais da caldeira de recuperação têm influência nas taxas de redução.

Para promover a combustão do licor preto, o ar é introduzido na caldeira em diferentes elevações, geralmente 3 ou 4. A caldeira de recuperação que foi objeto desse trabalho tem 3 níveis de ar, similar à que está representada na Figura 1. O ar primário é alimentado em todas as paredes e tem o objetivo de ajustar o formato da camada e mantê-la afastada das paredes da fornalha e de fornecer oxigênio para a combustão da parte externa do leito carbonizado, fornecendo energia para as reações de redução - que são endotérmicas - e proporcionando um ambiente redutor. O ar secundário é inserido em duas paredes, geralmente a frontal e a traseira, e determina o tamanho da camada e promove a secagem, combustão do licor e dos gases voláteis resultantes da pirólise. Por fim, o ar terciário tem a função de fornecer oxigênio para garantir a combustão completa do licor e dos gases, além de fazer uma selagem para evitar o arraste de material particulado para a parte alta da caldeira (FERREIRA; CARDOSO; PARK, 2010).

De acordo com MacCallum (2015), o ar primário e secundário são os melhores parâmetros de controle para a eficiência de redução. Damasceno *et al.* (2020) defendem que a configuração da injeção de ar tem grande influência na eficiência de redução. Segundo ele, este parâmetro é significativamente afetado - de maneira inversamente proporcional - pelo vazão de ar primário, porém tem pouca ou nenhuma influência do ar secundário.

As características do licor preto também impactam na eficiência de redução. Segundo Cardoso *et al.* (2012), o aumento da concentração de sólidos do licor preto queimado na caldeira representa uma melhora no processo de recuperação em função do aumento da temperatura da caldeira. Para MacCallum (2015), a viscosidade influencia na eficiência de redução, porém em menor grau. Além disso, eles afirmam que ajustes na pressão e temperatura do licor aumentam a temperatura do leito carbonizado, que é o fator determinante para atingir maior eficiência de redução.

O óleo combustível é usualmente utilizado nos momentos de parada e partida da caldeira e para completar o balanço energético da fábrica, ou seja, para aumentar a produção de vapor da caldeira. Conforme a Figura 1 os queimadores geralmente utilizados na partida da caldeira estão localizados na mesma altura do ar secundário, ou seja, próximos ao leito carbonizado, e quando estão em uso podem interferir na sua forma e altura, o que pode impactar a eficiência de redução, uma vez que a camada é importante para manter o ambiente redutor da fornalha (VAKKILAINEN *et al.*, 2005; MACCALLUM, 2015).

2.2 Rede neural artificial

A Rede Neural Artificial (RNA) é um modelo computacional que tenta reproduzir as capacidades de aprendizado e generalização do cérebro humano. Nas RNA a unidade básica de processamento é o neurônio artificial que possui diversas entradas (x_i). Os valores das entradas, ponderados pelos pesos sinápticos (w_i), são somados e ao resultado dessa operação (R) é aplicada a função de ativação ($f(\cdot)$), que tem o papel de ativar ou não esse resultado e fornecer um valor de saída do neurônio (y). O funcionamento descrito acima pode ser esquematizado pelas equações (2.4) e (2.5) (SILVA; SPATTI; FLAUZINO, 2010):

$$R = \sum_i x_i \cdot w_i \quad (2.4)$$

$$y = f(R) \quad (2.5)$$

A configuração dos neurônios e das conexões entre eles recebe o nome de arquitetura, sendo as mais comuns a *feedforward* e as recorrentes. Neste trabalho, será empregada a rede *Multi Layer Perceptrons* (MLP) que utiliza a arquitetura *feedforward*, na qual o fluxo de informação é unidirecional. Ou seja, inicia nos neurônios da camada de entrada, que recebem as informações do banco de dados, passa para os neurônios das camadas intermediárias, que irão extrair as informações requeridas dos dados e chegam nos neurônios da camada de saída que produzem o resultado final da rede (SILVA; SPATTI; FLAUZINO, 2010).

Para construir uma RNA é necessário definir alguns hiperparâmetros, como número de neurônios e de camadas ocultas, a função de ativação, a função de custo, o otimizador e a taxa de aprendizagem. A quantidade de neurônios em cada camada e o número de camadas é definida de maneira empírica e estão relacionadas com a qualidade da resposta obtida do modelo construído. Quanto mais neurônios e camadas, maior a flexibilidade para ajustar os pesos e maior a probabilidade de obter uma rede que resolva o problema estudado. Por outro lado, a complexidade da rede aumenta muito, gerando um custo computacional grande, e há o risco que a rede fique muito ajustada aos dados utilizados para construí-la, fenômeno conhecido como *overfitting*, e perca a capacidade de generalização (ALMEIDA *et al.*, 2003b).

A função de ativação limita a amplitude da saída de um neurônio a um intervalo de valores finitos, que tipicamente está entre 0 e 1 ou entre -1 e 1. A função sigmoide, cujo gráfico tem forma de S, é a função de ativação mais utilizada na construção das [RNA](#), talvez pelo fato de ter um comportamento adequadamente balanceado entre os comportamentos linear e não-linear e por ser diferenciável. Essa função assume um valor no intervalo contínuo de 0 a 1. A função tangente hiperbólica é do tipo sigmoide e a amplitude varia de -1 a 1, o que a torna capaz de perceber pequenas diferenças e de fazer modelos mais finos e com alta sensibilidade ([HAYKIN, 2007](#)). Existem outras funções que podem ser utilizadas de acordo com as características do problema, como a Softmax e a ReLU. A Softmax é adequada para problemas de classificação e geralmente é utilizada na camada de saída. A função ReLU é um bom ponto de partida na construção de um modelo de redes neurais, pois ela não ativa todos os neurônios ao mesmo tempo, o que reduz o custo computacional, porém deve ser usada apenas nas camadas ocultas.

A função de custo, também conhecida como função de perda, é um direcionador para o aprendizado de máquinas. O objetivo é que o custo seja o menor possível e a função de custo vai ajudar o algoritmo a encontrar uma metodologia para aprender de forma eficiente e precisa. O erro quadrático médio - soma quadrática das diferenças dos valores reais e dos valores preditos pelo modelo - e o erro médio absoluto - soma do módulo da diferença entre os valores reais e os valores preditos pelo modelo - são exemplos de função de custo ([BRAGA; CARVALHO; LUDERMIR, 2007](#)).

O otimizador tem a finalidade de minimizar a função de custo, determinando os pesos sinápticos e *bias* que alcançam esse objetivo. Um dos métodos mais empregados é o gradiente descendente, o qual segue a direção oposta ao do gradiente, que indica a direção da maior inclinação da superfície. Outros métodos como o SGD, ADAM, RMSProp, apresentam variações deste mesmo conceito de otimização ([BRAGA; CARVALHO; LUDERMIR, 2007](#)).

A taxa de aprendizagem é o tamanho do passo que se dá a cada iteração na busca por minimizar a função de custo. Se o passo for muito pequeno, o ajuste dos pesos será suave e o algoritmo precisará de um tempo significativo para encontrar o mínimo da função de custo. Por outro lado, se a taxa de aprendizagem for grande, pode-se não encontrar os parâmetros que minimizam a função de custo e deixar a rede instável ([HAYKIN, 2007](#)).

Definidos os hiperparâmetros, a rede passa pelo processo de treinamento supervisionado, o qual utiliza um algoritmo de aprendizado para promover o ajuste adequado dos pesos sinápticos. O algoritmo *backpropagation* é utilizado nas redes MLP e ele é composto pela iteração em duas etapas: uma para frente (*forward*), da entrada para saída, e a outra no sentido contrário, para trás (*backward*). Na primeira etapa os dados iniciais são processados obtendo a saída, o valor predito. Esse resultado é comparado com o valor real e a diferença entre eles é conhecida como erro. Na fase *backward*, o mecanismo de retropropagação faz o ajuste dos pesos dos neurônios de toda a rede em função dos erros, começando pela camada de saída até a camada de entrada. Esse processo se repete por inúmeras vezes para que a rede possa convergir (HAYKIN, 2007). O critério de parada do treinamento é geralmente o número de épocas, que é a apresentação de todas as amostras do conjunto de treinamento à rede neural, ou quando a função de custo atingir um valor aceitável.

O uso de RNA para modelar caldeiras de recuperação é uma realidade, visto que uma grande quantidade de dados está disponível nas indústrias. Almeida *et al.* (2003a), Almeida *et al.* (2010) e Sainlez e Heyen (2011) utilizaram as redes neurais para prever a produção de vapor, Belisário *et al.* (2020) desenvolveram um sensor virtual para estimar emissões de óxido de enxofre e outro para material particulado e BELISÁRIO *et al.* (2023) construíram um modelo para prever a queda de pressão dos gases de exaustão na bancada. Almeida *et al.* (2003b) se propuseram a avaliar o uso de redes neurais para prever a eficiência de redução, porém o tamanho da massa de dados utilizada no estudo foi um fator limitante.

A predição da eficiência de redução já foi estudada a partir de outras metodologias. Silva *et al.* (2008) utilizaram a minimização da energia livre de Gibbs e desenvolveram um modelo que conseguiu prever a eficiência de redução com um erro máximo de 1,1%. Damasceno *et al.* (2020) utilizaram o software comercial *Aspen Plus* para prever simultaneamente a eficiência de redução e a produção de vapor. O modelo fenomenológico foi validado por dados industriais e da literatura e o erro para a eficiência de redução foi menor que 1,0%.

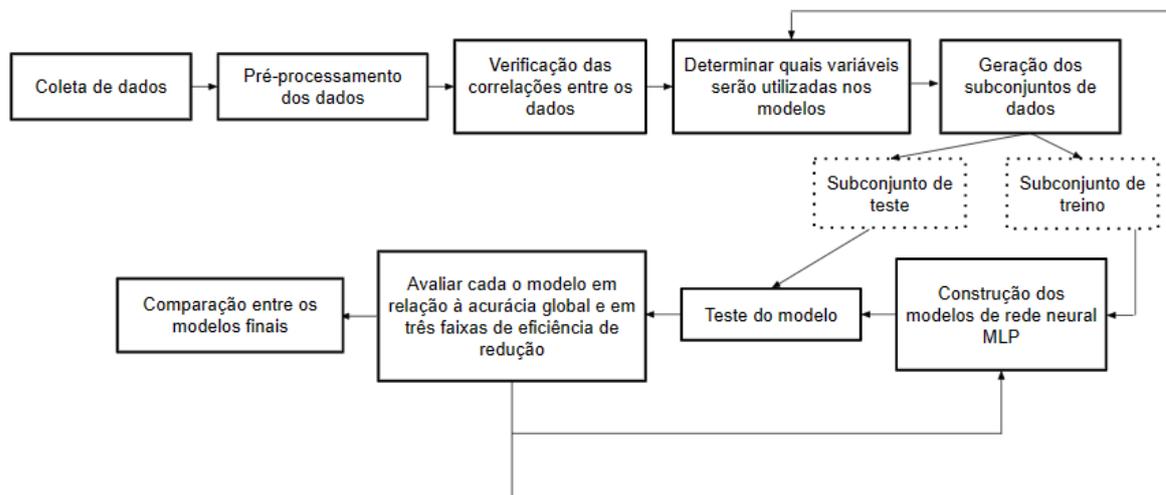
3 Metodologia

O presente estudo classificou-se como aplicado, quantitativo, empírico, descritivo e de modelagem. A pesquisa abordada é aplicada uma vez que os conhecimentos adquiridos serão aplicados de forma imediata na indústria de celulose e tem objetivos específicos (GIL *et al.*, 2002). Os dados foram coletados de instrumentos instalados no processo e de análises de laboratório por meio do historiador de dados e foram analisados de forma estatística o que caracteriza a pesquisa como qualitativa. O estudo é empírico uma vez que o modelo elaborado trabalha com um contexto real a fim de elucidar as correlações entre as variáveis (CAUCHICK-MIGUEL *et al.*, 2018). Ele é descritivo visto que visa buscar correlações entre os parâmetros de operação da Caldeira de Recuperação, como vazão, temperatura e pressão do ar e dos combustíveis alimentados com a variável resposta que é a eficiência de redução (MARCONI; LAKATOS, 2003).

Por fim, o procedimento utilizado na pesquisa será a modelagem, cujo foco é descrever problemas complexos a partir de um modelo computacional, de modo a identificar a dinâmica entre as variáveis e facilitar a tomada de decisão de quem estiver usando o modelo (CAUCHICK-MIGUEL *et al.*, 2018). Esse é exatamente o objetivo do trabalho: prever o que acontece na prática com a eficiência de redução de uma caldeira de recuperação com base nos dados de entrada operacionais do processo de forma que essa informação possa ser utilizada pelos operadores para auxiliar nos ajustes dos parâmetros da caldeira de forma mais rápida e eficaz.

A Figura 2 apresenta, de forma esquemática, a metodologia utilizada que será descrita a seguir.

Figura 2 – Diagrama da metodologia adotada no trabalho.



Fonte: elaborado pela autora.

Os dados de 52 variáveis operacionais da Caldeira de Recuperação 3 da Cenibra, conforme exposto no quadro 3.1, foram coletados utilizando o sistema historiador de dados *Plant Information Management System (PIMS)* através do *software* Excel. Essas variáveis foram escolhidas com base nas informações disponíveis na literatura sobre fatores que influenciam na eficiência de redução (MACCALLUM, 2015; DAMASCENO *et al.*, 2020; CARDOSO *et al.*, 2012; VAKKILAINEN *et al.*, 2005) e com o intuito de avaliar outros parâmetros que no dia a dia da operação aparentam ter alguma influência. O período de coleta de dados foi de outubro de 2023 a janeiro de 2024, referente à campanha (tempo de operação entre as paradas para inspeção previstas na NR13) atual da caldeira. Esse período foi selecionado, visto que representa melhor a condição presente, uma vez que durante a parada mencionada anteriormente alguns instrumentos utilizados na medição das variáveis coletadas passam por verificações e ajustes, quando necessários. Além disso, nesse período, a análise de eficiência de redução foi realizada com uma frequência maior, 6 vezes ao dia.

Quadro 3.1 – Variáveis analisadas por cada circuito da Caldeira de Recuperação.

Circuito	Variáveis analisadas
Ar e gás	Vazão, temperatura e pressão do ar primário, secundário e terciário e oxigênio residual.
Combustíveis	Vazão, temperatura e pressão do licor preto e do óleo e concentração do licor.
Licor verde e branco	Eficiência de redução, sulfidez.
Água e vapor	Vazão de alimentação de água e de produção de vapor e temperatura e pressão do vapor produzido.

Fonte: elaborado pela autora.

Com exceção da eficiência de redução (variável resposta do modelo), que é um dado gerado pelo laboratório, as demais variáveis são medidas a partir dos instrumentos instalados em linha e estão disponíveis em intervalos de 1 em 1 minuto. Ou seja, a frequência dos dados de eficiência de redução é a limitante para este trabalho. Por isso, todos os dados foram coletados considerando a média horária, que é um recurso já disponível no historiador, dos horários de realização de amostragem do licor verde, que no período em estudo foram realizadas 6 vezes ao dia às 1, 5, 9, 13, 18 e 21 horas.

Toda a parte de análise de dados e construção do modelo foi desenvolvida no ambiente Google Colab utilizando a linguagem de programação Python. Primeiro os dados crus passaram por um pré-processamento para eliminar momentos em que a caldeira de recuperação estava em processo de parada, parada ou no início de operação e possíveis falhas dos instrumentos de medição. O pré-processamento de dados é uma etapa importante visto que a qualidade do modelo obtido no trabalho é extremamente dependente da qualidade dos dados utilizados na sua construção. Dados de baixa qualidade podem levar a modelos não precisos e ineficazes (NICKOLAS; SHOBHA, 2022).

Uma correção dos dados de vazão de óleo queimado foi realizada, substituindo os valores negativos por zero. Isso foi necessário porque a vazão de óleo é resultado da diferença da medição de dois instrumentos (vazão de óleo para caldeira e vazão de óleo retornado para o tanque de estocagem) que inevitavelmente apresentam diferenças entre eles, gerando valores negativos e próximos de zero quando não há queima de óleo na caldeira.

Em seguida, foram removidos os dados dos momentos de parada e partida da caldeira e para isso, tomou-se como base os parâmetros do vapor produzido, do licor preto, licor preto queimado e do ar terciário, o qual é alimentado na caldeira apenas quando há queima de licor preto. Levando em consideração os valores normais de operação, foram excluídas todas as linhas de dados em que os valores da temperatura, pressão e vazão do vapor, vazão, concentração e temperatura do licor e vazão de ar terciário estavam menores que 400 °C, 55 kgf/cm², 100 t/h, 50 m³/h, 60%, 120°C e 40 t/h, respectivamente. Também foram excluídas as linhas em que a eficiência de redução era menor que 75% e maior que 98,5%, com o objetivo de eliminar momentos de erros na coleta da amostra, na realização da análise e no lançamento do resultado.

Após o tratamento dos dados foi realizada a verificação da correlação entre as variáveis pelo método de Pearson, a qual foi expressa pelo mapa de calor - função *seaborn.heatmap* utilizando como conjunto de dados a função *pandas.DataFrame.corr(method='pearson')* - e pelo gráfico de dispersão - função *seaborn.scatterplot* - com o objetivo de identificar as variáveis que mais influenciam na eficiência de redução e se há correlação entre as variáveis de entrada.

Como o número de variáveis disponíveis inicialmente é elevado e isto poderia tornar o modelo mais complexo e até provocar uma pobre capacidade de generalização, a partir da avaliação das correlações, definiu-se as variáveis que foram utilizadas na construção dos modelos. Inicialmente foram eliminados os parâmetros com correlação baixa com a variável resposta e variáveis de entrada que tinham alta correlação entre elas, uma vez que apenas uma das duas variáveis já é o suficiente para explicar a outra.

Na sequência, os parâmetros selecionados, foram divididos nos subconjuntos de treino e de teste, de modo que os mesmos conjuntos sempre fossem utilizados nas diferentes técnicas de construção do modelo de predição. Para dividir os dados foi utilizada uma função de divisão aleatória com semente fixa de forma que 75% dos dados fossem utilizados para treinar o modelo e os outros 25% para teste. Apesar da ordem de grandeza dos dados disponíveis para a construção dos modelos estar entre 10⁰ e 10², ou seja, uma variação não muito grande, optou-se por normalizar as variáveis de entrada por meio da função *sklearn.preprocessing.StandardScaler*, na qual o conjunto de dados de cada variável passou a ter desvio padrão igual a 1. Por fim, a variável resposta foi trabalhada de duas formas, sendo elas:

1. contínua, como ela é originalmente; e
2. categórica, a partir do agrupamento dos dados, considerando a divisão: eficiência de redução < 90%, de 90% a < 95% e >= 95%.

Para alguns modelos aplicou-se no subconjunto de treino a técnica de validação cruzada visando melhorar a capacidade de generalização dos modelos. Para isso, esse subconjunto foi dividido em 5 grupos, sendo que 4 eram usados para o treinamento e o grupo restante para validação. Esse procedimento foi repetido por 5 vezes, de modo que todos os 5 grupos sejam usados para validar os modelos. Por fim, o grupo de teste foi utilizado para testar os modelos obtidos.

Os dados para a construção do modelo categórico foram trabalhados de duas formas: sem e com balanceamento das classes. Para fazer o balanceamento utilizou-se a técnica *Synthetic Minority Oversampling Technique (SMOTE)* para criação de observações artificiais das classes minoritárias, visto que essa técnica se mostra robusta e bem sucedida em várias aplicações de diferentes esferas do conhecimento (FERNÁNDEZ *et al.*, 2018).

A partir dos dados pré-processados foram construídos os modelos de Rede Neural Artificial MLP devido ao seu caráter de adaptação à problemas complexos e de qualquer natureza (SHEPHERD, 2012). Para isso, foi utilizada a função *tensorflow.keras.Sequential* do keras tuner com o objetivo de encontrar os melhores hiperparâmetros para o modelo a partir de uma pesquisa aleatória em um espaço de busca pré determinado. O espaço de busca utilizado está descrito no Quadro 3.2. O otimizador escolhido foi o Adam, o qual é amplamente utilizado no aprendizado de máquinas (SHEPHERD, 2012), sendo que a taxa de aprendizagem poderia ser 0,01 ou 0,001. A função de custo utilizada para os modelos contínuos foi o erro quadrado médio - mse - e para os modelos categóricos a entropia cruzada - *sparse_categorical_crossentropy*.

Quadro 3.2 – Espaço de busca utilizado no tuner para construção do melhor modelo.

Camada	Nº de camadas	Nº de neurônios	Função de ativação
Primeira	1	8 - 1024 ¹	tanh e sigmoide
Ocultas	1, 2 e 3	16 - 4096 ¹	ReLU, tanH e sigmoide
Saída	1	1 ou 3 ²	linear ou softmax ²

¹ Intervalo de números múltiplos de 2. ² Para modelos construídos com a variável resposta de forma contínua foi usado um neurônio com a função de ativação linear e para a variável categórica usou-se 3 neurônios e a função de ativação softmax.

Fonte: Elaborado pela autora.

Com os hiperparâmetros definidos, foram treinadas 10 redes MLP usando cada uma das abordagens a seguir:

1. contínua,
2. contínua com validação cruzada,
3. categórica sem balanceamento de classes,

4. categórica com balanceamento de classes

Para os modelos construídos usando a eficiência de redução como uma variável contínua, foi realizada a avaliação do coeficiente de determinação (R^2) entre os valores reais e os valores preditos. Em função da dificuldade encontrada em desenvolver um modelo com boa previsibilidade da eficiência de redução de forma contínua, a resposta do modelo contínuo bem como os valores reais foram convertidos em classes - baixa ($\leq 90\%$), normal (90% a $< 95\%$) e alta ($\geq 95\%$) - e foi elaborada a matriz de confusão da qual extraiu-se os dados de acurácia global por faixa e o percentual de falsos-positivos por faixa a fim de verificar o atendimento do modelo para à faixa baixa de operação, que é a de maior interesse. Seguindo a linha da classificação, trabalhou-se com modelos categóricos, aplicando a segunda etapa de avaliação dos modelos contínuos, ou seja, matriz de confusão, acurácia e falsos-positivos. Por fim, os melhores modelos obtidos utilizando cada uma das abordagens foram comparados usando as técnicas já citadas e também com base na complexidade a fim de determinar qual é o mais adequado para aplicação.

4 Resultados

4.1 Pré-Processamento

A base de dados inicial continha 51 variáveis de entrada, uma variável de saída e 507 observações para cada uma delas. Após aplicar os filtros para fazer a remoção das linhas dos momentos em que a caldeira estava parada ou partindo, restaram 497 observações. Construiu-se a matriz de calor para o coeficiente de correlação de Pearson entre as 52 variáveis (51 de entrada e a variável de saída). Avaliando a correlação entre as variáveis de entrada e a eficiência de redução, esse coeficiente variou entre -0,53 e 0,49, o que evidencia que a eficiência de redução se correlaciona com as variáveis de forma não linear e que um conjunto de variáveis influenciam simultaneamente na eficiência de redução.

Buscando reduzir o número de variáveis de entrada que seriam utilizadas na construção dos modelos, foram eliminadas as variáveis cujo coeficiente da correlação estava entre -0,15 e 0,15, restando ainda 38. Avaliando a correlação entre as variáveis de entrada, verificou-se que algumas delas tinham uma correlação mais alta ($<-0,8$ e $> 0,8$) e com isso, optou-se por eliminá-las, mantendo a variável de maior correlação com a eficiência de redução. O novo banco de dados passou a ter 25 variáveis de entrada. Por fim, fez-se mais um filtro, levando em consideração o que foi levantado na literatura sobre os parâmetros que influenciam na eficiência de redução e o conhecimento acerca da Caldeira de Recuperação 3 da Cenibra e chegou-se em 12 variáveis de entrada (Quadro 4.1).

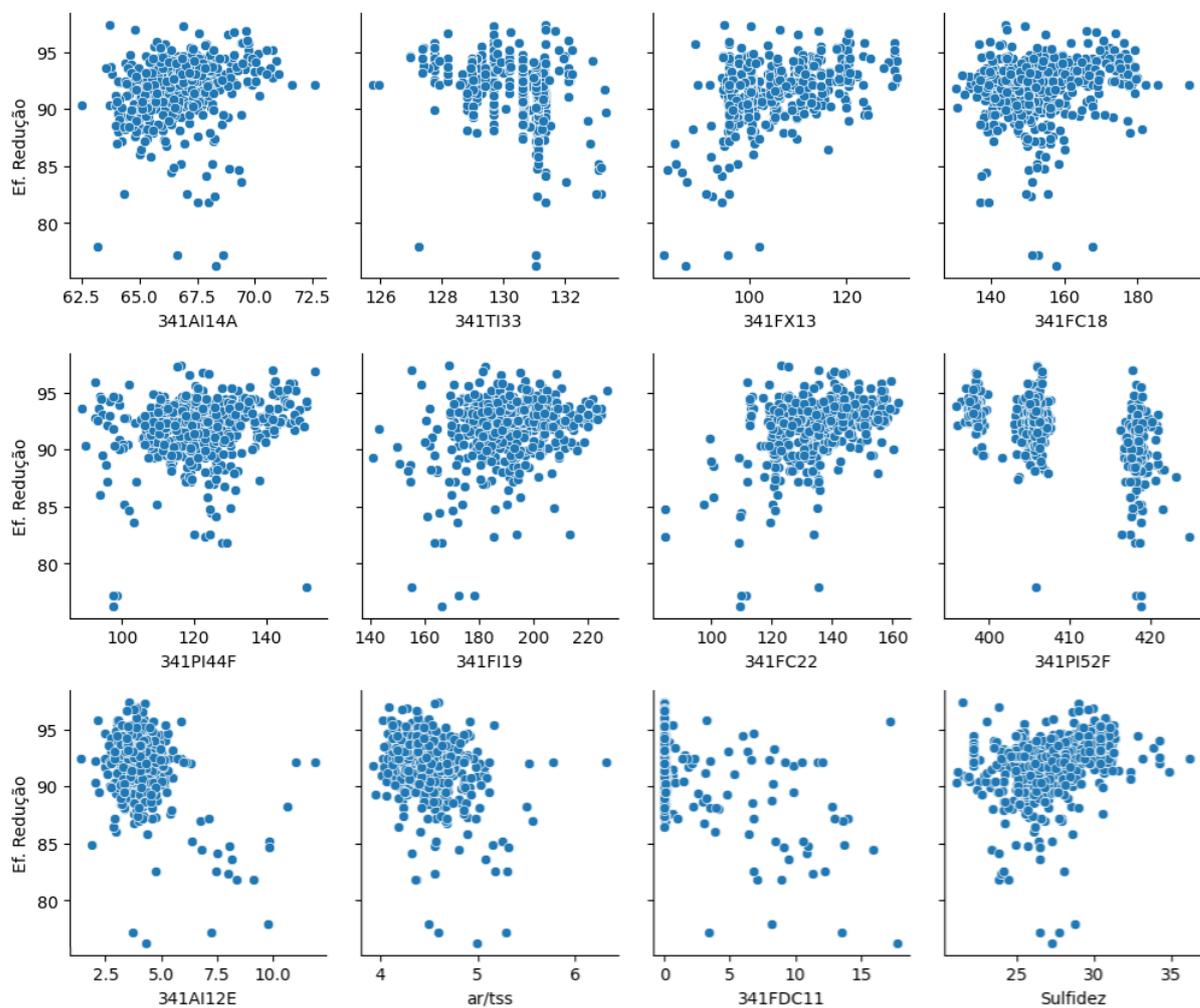
Quadro 4.1 – Variáveis de entrada utilizadas na construção do modelo para prever a eficiência de redução

Código	Variável, unidade
341AI14A	Concentração do licor preto, %
341TI33	Temperatura do licor preto, °C
341FX13	Vazão de licor preto queimado, tss/h
341FC18	Vazão de ar primário, t/h
341PI44F	Pressão de ar primário, mmH ₂ O
341FI19	Vazão de ar secundário, t/h
341FC22	Vazão de ar terciário, t/h
341PI52F	Pressão de ar terciário, mmH ₂ O
341AI12E	Oxigênio residual na fornalha, %
Ar/tss	Vazão total de ar por vazão de licor queimado, t_{ar}/tss
341FDC11	Vazão de óleo queimado, t/h
Sulfidez	Sulfidez do licor branco, %

Fonte: elaborado pela autora.

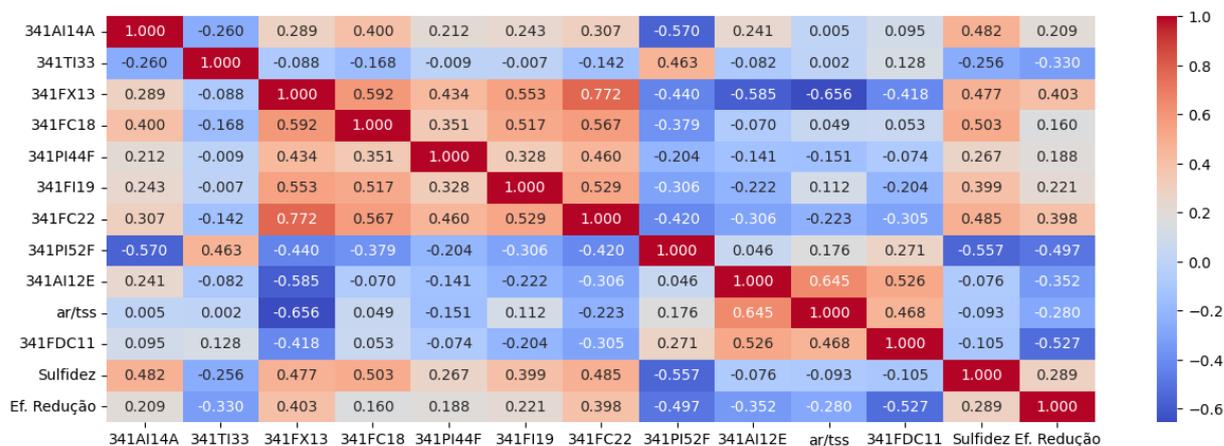
A Figura 3 apresenta como cada uma das 12 variáveis de entrada está dispersa em função da variável de saída. Fazendo uma análise puramente visual é possível inferir que certas variáveis, como vazão e concentração de licor, têm uma relação positiva com a eficiência de redução, já outras variáveis, como a vazão de óleo e p teor de oxigênio residual, estão inversamente relacionadas com a variável de saída, ou seja, a medida que a variável de entrada assume valores mais altos, a tendência é que a eficiência de redução caia. Avaliando os coeficientes de Pearson presentes na Figura 4, verifica-se que algumas variáveis, como a vazão e pressão de ar primário, conclui-se que essas variáveis se relacionam de forma direta com a eficiência de redução, mas com baixa intensidade.

Figura 3 – Gráficos de dispersão das variáveis de entrada com a eficiência de redução.



Fonte: elaborado pela autora.

Figura 4 – Mapa de calor para o coeficiente de correlação de Pearson para as variáveis de entrada e a variável de saída.



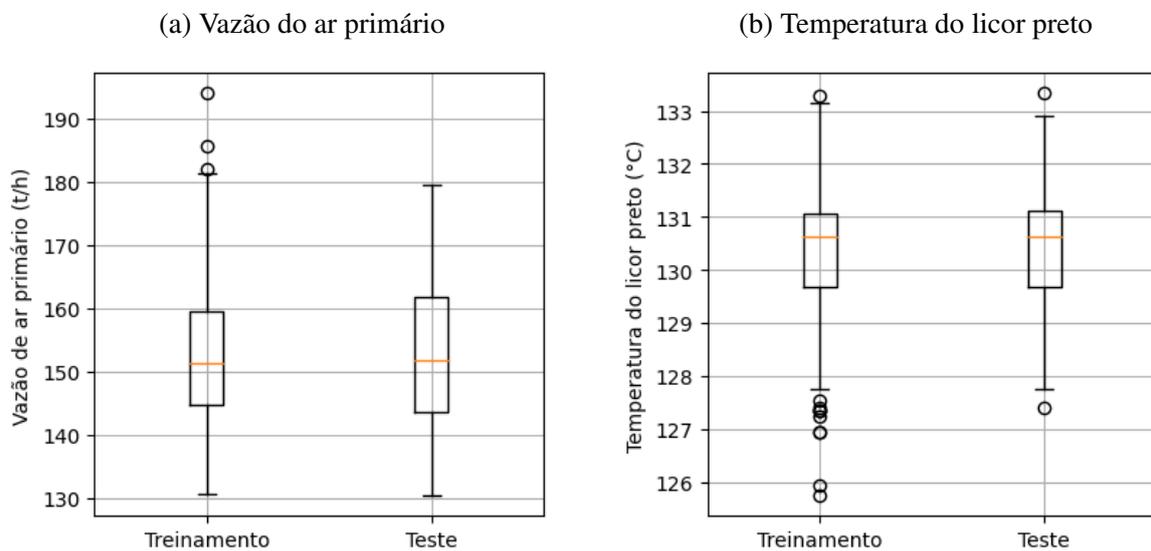
Fonte: elaborado pela autora.

De acordo com a literatura, o fluxo de ar tem uma relação inversa com a eficiência de redução (DAMASCENO *et al.*, 2020), (MACCALLUM, 2015). Os coeficientes de Pearson, considerados de forma crua, para o fluxo de ar de cada um dos três níveis em função da eficiência de redução sugerem uma situação contrária. Porém, ao aprofundar a análise, o aumento do fluxo de ar também tem uma relação positiva com a queima de licor, ou seja, uma maior quantidade de ar está sendo alimentada na cadeira pois o fluxo de licor queimado também é maior. Quando a avaliação é feita para uma quantidade fixa de licor, que é o que a variável ar/tss expressa, observa-se exatamente o que a literatura aborda, o coeficiente de Pearson passa a ser negativo, -0,28 para os dados em estudo.

Outra informação interessante confirmada a partir da avaliação dos dados é a relação da temperatura do licor com a eficiência de redução. A tendência é que temperaturas maiores gerem eficiências de redução mais baixas. Isso pode ser explicado em função do tamanho da gota do licor, a qual tende a diminuir em função do aumento da temperatura devido à mudança da viscosidade. Segundo Vakkilainen *et al.* (2005), ao contrário do que é praticado para muitos combustíveis, cujo tamanho de gota deve ser muito pequeno para maximizar as taxas de combustão, para o licor preto a atomização é feita objetivando gotas maiores, de 2 a 3 mm de diâmetro, de modo que o material não queimado atinja o leito carbonizado. Dessa forma, gotas pequenas são queimadas quando ainda estão em suspensão e a formação da camada fica prejudicada, bem como o ambiente redutor necessário para atingir boas eficiências de redução.

Após o tratamento inicial dos dados e da seleção das variáveis de entrada que participaram da construção dos modelos, os dados foram divididos - usando sempre a semente 1605 - em dois grupos, ficando o de treinamento com 372 observações e o de teste com 125. Foi realizada uma avaliação do conjunto de dados de cada grupo para cada variável. Conforme os exemplos mostrados na Figura 5, a estatística descritiva, neste caso primeiro e terceiro quartis e a mediana, de ambos os grupos é semelhante. O que difere um pouco é que o grupo de treinamento tem mais observações discrepantes (fora dos limites dos boxplots), o que é importante para que depois de treinado o modelo também seja capaz de prever situações extremas, que não estão retratadas no grupo de teste. Na sequência, os dados foram normalizados de forma que o desvio padrão de todas as variáveis de entrada fosse igual a 1.

Figura 5 – Boxplot do grupo de treinamento e de teste para duas variáveis de entrada.



Fonte: elaborado pela autora.

Para a construção do modelo categórico, a variável de saída foi dividida em três classes:

- baixa: para eficiência de redução menor ou igual a 90%;
- normal: para eficiência de redução entre 90 e 95%; e
- alta: para eficiência de redução maior ou igual a 95%.

Para o conjunto de treinamento as 372 observações ficaram divididas em 79, 275 e 18, respectivamente. No grupo de teste, das 125 observações 16 pertencem ao primeiro grupo, 102 ao segundo e 7 ao último. Por fim, também foi realizado o balanceamento de classes para o conjunto de treinamento, utilizando a técnica **SMOTE**, gerando 196 e 257 observações artificiais para as classes baixa e alta, respectivamente, de modo que o número de observações fosse igual ao da classe normal.

4.2 Modelo de Rede Neural Artificial

O primeiro modelo de rede trabalhado se baseia na eficiência de redução na forma contínua. Com a utilização do KerasTuner, o melhor modelo encontrado possui três camadas ocultas, conforme o Quadro 4.2 e taxa de aprendizagem de 0,001.

Quadro 4.2 – Hiperparâmetros do modelo de rede MLP para predição da eficiência de redução.

Camada	Nº de Neurônios	Função de ativação
Primeira	1024	tanh
Ocultas 1	128	tanh
Ocultas 2	2048	relu
Ocultas 3	2048	relu
Saída	1	linear

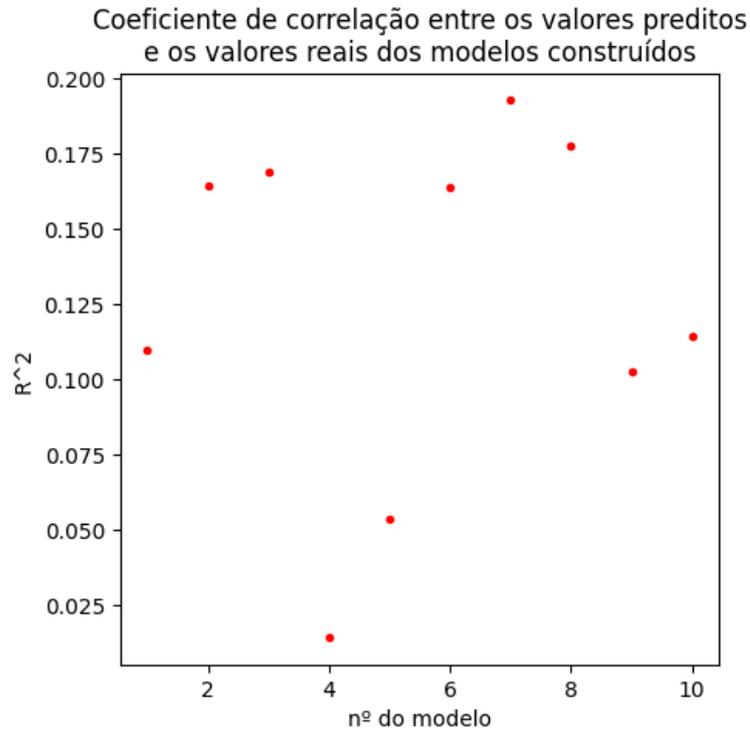
Fonte: elaborado pela autora.

Usando os hiperparâmetros do Quadro 4.2, a rede MLP foi treinada 10 vezes com o conjunto de treino completo e testada com o conjunto de teste. O coeficiente de correlação dos valores reais *versus* os valores preditos de cada um desses modelos está representado na Figura 6. Esse coeficiente expressa quanto da variância da eficiência de redução é explicada pelo modelo, que neste caso está entre 1,4 e 19,3%. A Figura 7 exhibe a comparação dos valores reais e preditos do modelo 7, cujo R^2 foi 0,193, e a Figura 8 mostra como o erro (diferença entre o valor predito e o valor real) do modelo 7 se comporta.

Cerca de 80% dos valores dos erros está entre -3 e 4. Os extremos dessa faixa são erros significativos, visto que a eficiência de redução está entre 75 e 98,5%, ou seja, um erro igual a 4 representa 17% dessa faixa. Além disso, o R^2 obtido revela que a previsibilidade desse modelo é baixa. Em função dessas limitações, decidiu-se por avaliar esse modelo de outra forma, transformando os dados reais e preditos em três faixas, as mesmas descritas na seção anterior que foram usadas na construção do modelo categórico, e calculando a acurácia global e a de cada classe. A Figura 9 apresenta os resultados para cada um dos 10 modelos construídos.

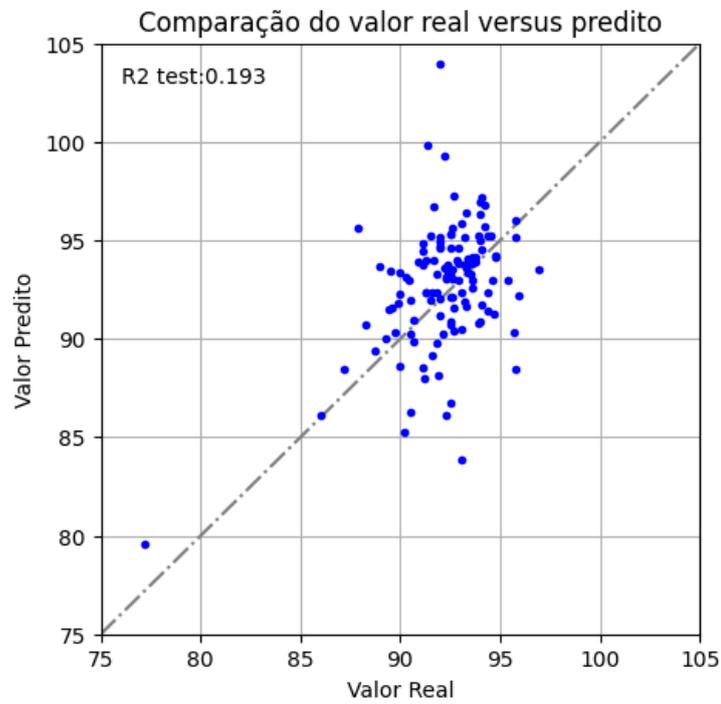
Neste estudo, é muito importante avaliar o comportamento da classe baixa, uma vez que quando opera-se nessa faixa de eficiência de redução são observados prejuízos no processo e consequentemente financeiros. De acordo com os dados da Figura 9 os modelos 1 e 4 foram os que apresentaram a maior acurácia para a classe baixa. Porém a acurácia para a faixa normal foi melhor para o modelo 1. Outro modelo que indicou um resultado interessante e que deve ser melhor explorado é o modelo 8. Por sua vez, o modelo 7, que apresentou o melhor coeficiente de determinação, teve a melhor previsibilidade na classe alta, quando comparado com os demais modelos. Porém ele apresentou um dos menores valores de acurácia para a classe baixa, que é a de maior interesse. Dessa forma, foi feita uma análise comparativa apenas entre os modelos 1 e 8.

Figura 6 – Gráfico com o coeficiente de determinação de cada modelo construído.



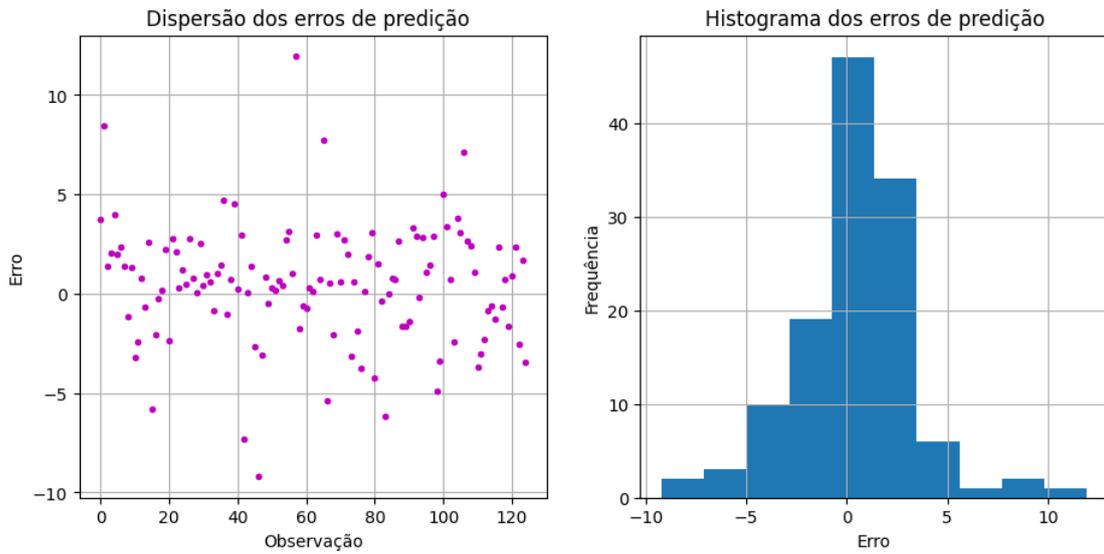
Fonte: elaborado pela autora.

Figura 7 – Gráfico com as estimativas do modelo 7 e os valores reais da eficiência de redução.



Fonte: elaborado pela autora.

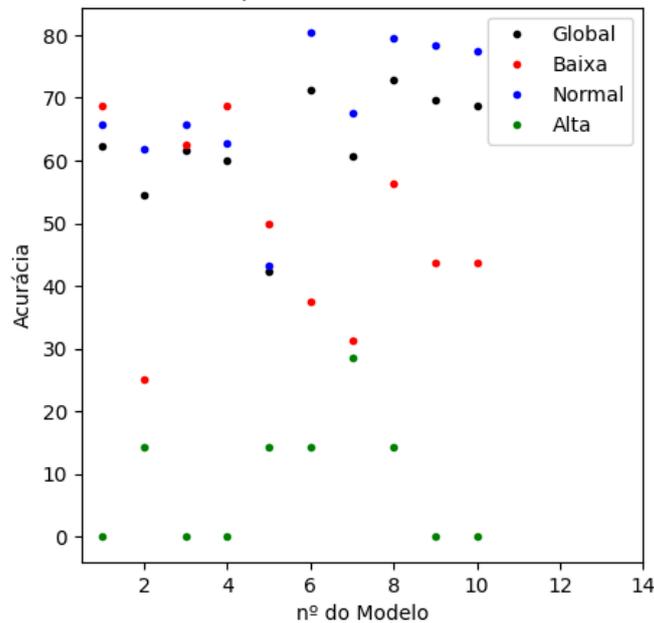
Figura 8 – Comportamento do erro a partir do gráfico de dispersão e do histograma.



Fonte: elaborado pela autora.

Figura 9 – Acurácia global e por faixas obtida com os 10 modelos construídos.

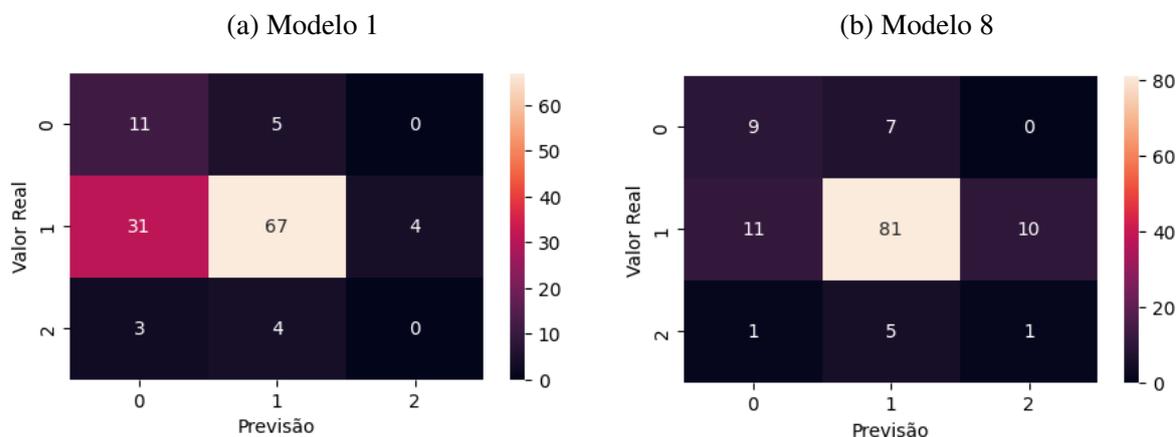
Acurácia do modelo para cada faixa de Eficiência de redução



Fonte: elaborado pela autora.

A Figura 10 exibe a matriz de confusão dos valores reais *versus* preditos para os modelos 1 e 8. Como já havia sido observado na Figura 9, o modelo 1 acertou mais vezes a previsão de classe baixa, 11 contra 9 do modelo 8, para um total de 16 observações. Por outro lado, a quantidade de observações rotuladas como classe baixa quando na verdade pertenciam às classes normal e alta também é maior, 34 contra 11 do modelo 8. Isso é um problema, visto que o modelo cairá em descrédito quando o usuário se deparar com a predição frequente de classe baixa quando os parâmetros da caldeira estiverem bem ajustados. Diante disso, entende-se que o melhor modelo treinado até então é o modelo 8.

Figura 10 – Matriz de confusão para modelos contínuos.



Fonte: elaborado pela autora.

Visando obter melhores resultados, os mesmos hiperparâmetros do Quadro 4.2 foram usados para treinar modelos usando a técnica de validação cruzada. Os dados de treinamento foram divididos em 5 grupos que foram usados para treinar e validar o modelo. Depois o grupo de teste foi utilizado para testar a rede criada e gerar os dados para avaliação da efetividade da rede. Da mesma forma como realizado anteriormente, a análise foi realizada avaliando o coeficiente de determinação e a acurácia global e das classes criadas.

De acordo com a Figura 11a, os coeficientes de correlação para os modelos construídos com validação cruzada ficaram próximos dos modelos exibidos anteriormente, sendo 0,240 para o modelo 6 o maior valor alcançado. Comparando com os demais modelos, o modelo 6 tem elevada acurácia apenas para a classe alta, conforme expresso na Figura 11b. O modelo 10 foi o que retornou o pior coeficiente de determinação, $R^2 = 0,046$, e o que teve a pior acurácia para a faixa de eficiência de redução baixa. Já no modelo 3 a acurácia das duas faixas extremas - baixa e alta - é elevada, 68,75 e 57,14%, respectivamente. Os outros dois modelos que apresentaram, na sequência, melhor acurácia para a classe baixa, 43,75%, são os modelos 4 e 9.

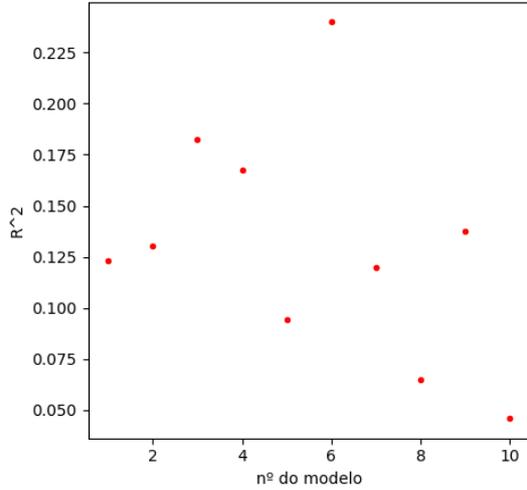
A Figura 12a mostra que, assim como pode ser observado na Figura 11b, a acurácia do modelo 3 para a classe normal é pequena. Além disso, a matriz de confusão revela que 30,3% das observações das classes normal e alta somadas foram preditas como pertencentes à classe baixa, ou seja, um valor significativo de falso positivo - o modelo aponta que tem-se um problema, quando na verdade não se tem -, o que inviabiliza a utilização desse modelo. O desempenho dos modelos 4 e 9, avaliado as matrizes de confusão expostas nas Figuras 12b e 12c, é bastante similar. O modelo 4 teve um acerto a mais que o modelo 9, que foi na classe alta, porém o modelo 9 teve um falso-positivo na faixa baixa a menos que o modelo 4. Com base no coeficiente de determinação, o modelo 4 consegue explicar melhor a variância da eficiência de redução, quando trabalhada de forma contínua, este modelo explica 16,7% enquanto o modelo 9, 13,8%.

Figura 11 – Comportamento das redes treinadas por meio da validação cruzada.

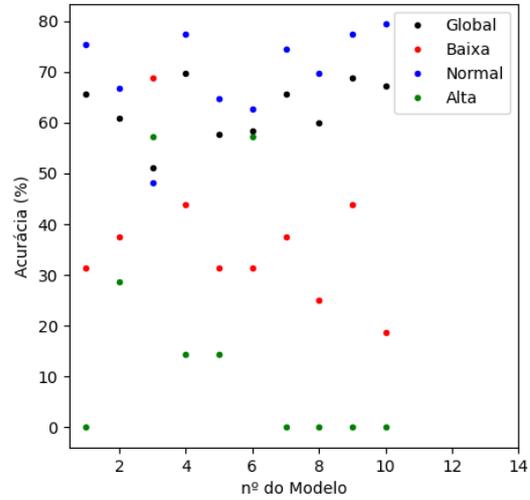
(a) Coeficiente de determinação

(b) Acurácia global e por classe

Coefficiente de correlação entre os valores preditos e os valores reais dos modelos construídos usando validação cruzada



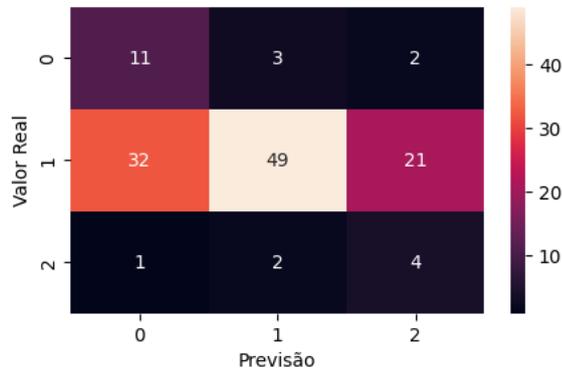
Acurácia do modelo para cada faixa de Eficiência de redução



Fonte: elaborado pela autora.

Figura 12 – Matriz de confusão dos modelos construídos utilizando a validação cruzada.

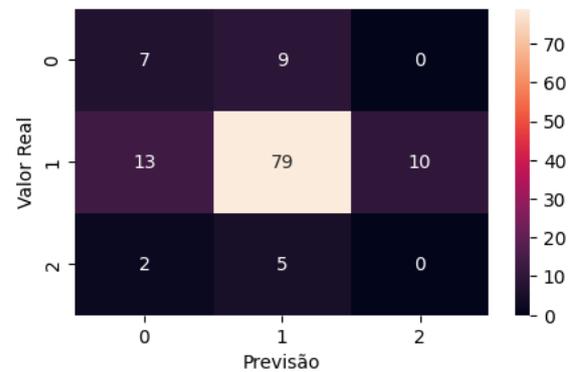
(a) Modelo 3



(b) Modelo 4



(c) Modelo 9



Fonte: elaborado pela autora.

Independente da análise, quando comparados os modelos 4 e 9 obtidos com a aplicação do método de validação cruzada, com o modelo 8, que não foi utilizada essa técnica, o modelo 8 se mostra superior. Em vista disso, os próximos modelos construídos não utilizaram o recurso da validação cruzada. Esses modelos foram trabalhados com a eficiência de redução já na forma categórica, a fim de avaliar se a capacidade de predição aumentaria.

Utilizando o KerasTuner, foi encontrado o melhor modelo para predizer as classes pré-definidas da eficiência de redução. Esse modelo conta com duas camadas ocultas e o treinamento foi realizado utilizando a taxa de aprendizagem de 0,01. O Quadro 4.3 contém os demais hiperparâmetros do modelo. É válido ressaltar que a resposta desse modelo é a probabilidade das variáveis de entrada pertencerem a cada uma das três classes, por isso foram utilizados três neurônios na camada de saída.

Quadro 4.3 – Hiperparâmetros do modelo de rede MLP para predição das clases da eficiência de redução

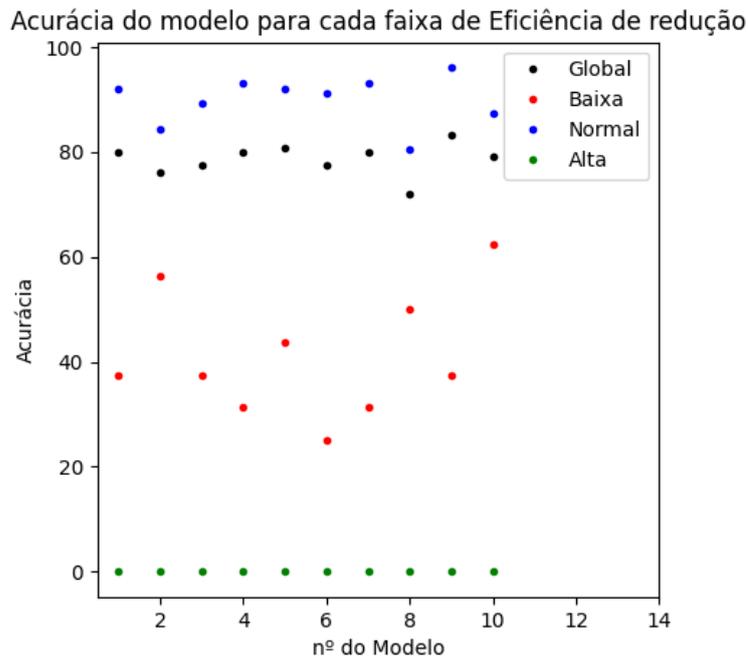
Camada	Nº de Neurônios	Função de ativação
Primeira	1024	sigmoide
Oculto 1	128	relu
Oculto 2	128	relu
Saída	3	softmax

Fonte: elaborado pela autora.

Com os hiperparâmetros definidos, foram treinadas 10 redes MLP, as quais foram testadas com o conjunto de teste e os resultados estão apresentados na Figura 13. Como a saída do modelo é composta por três valores de probabilidade, para fazer a análise da acurácia e da matriz de confusão, considerou-se como valor predito a classe que tinha o maior argumento de probabilidade.

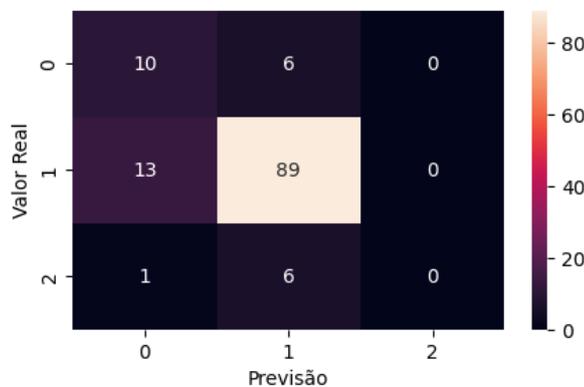
Nenhum dos dez modelos foi capaz de acertar as predições da classe alta. Isso pode ter acontecido, pois apenas 5% dos dados representam essa faixa. Analisando a classe baixa, o modelo 10 é o que teve melhor acurácia e na sequência foram os modelos 2 e 8. A acurácia da classe normal e consequentemente a global desses modelos também são mais baixas que as do modelo 10. Dessa forma, o melhor modelo categórico é o 10 e a matriz de confusão está representada na Figura 14.

Figura 13 – Acurácia global e por classe obtida com os 10 modelos categóricos construídos



Fonte: elaborado pela autora.

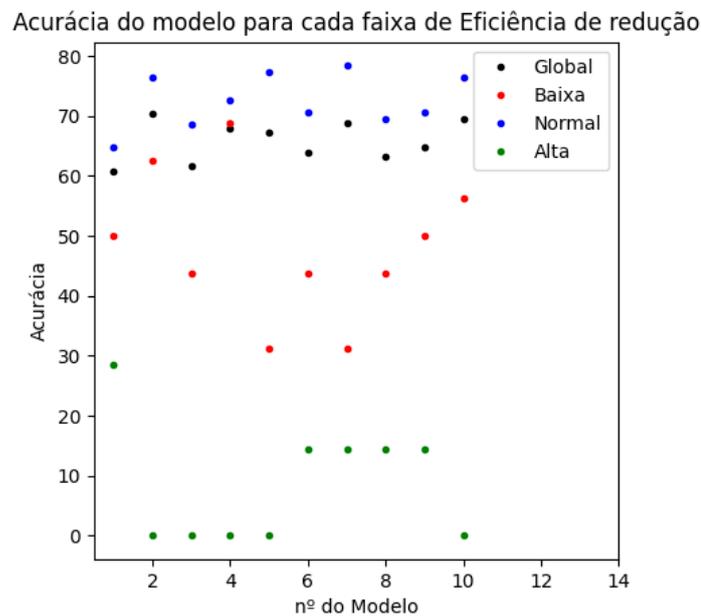
Figura 14 – Matriz de confusão do modelo categórico 10



Fonte: elaborado pela autora.

Em busca de melhorar o modelo categórico e ter previsibilidade também da faixa alta, foi realizado o balanceamento de classes do conjunto de treinamento e usando os hiperparâmetros do Quadro 4.3, mais 10 redes foram treinadas. Os resultados obtidos com cada um dos modelos utilizando o conjunto de teste estão dispostos na Figura 15. Metade dos modelos conseguiram prever uma parcela do comportamento da classe alta, mas ainda muito pequena. Além disso, os modelos com melhor desempenho para a faixa baixa, não conseguiram prever a faixa alta. Os modelos 4, 2 e 10 tiveram, nessa ordem, a maior acurácia para a classe baixa, 68,75%, 62,5% e 56,25%, respectivamente, e também estão entre as 4 melhores acurácias globais.

Figura 15 – Acurácia global e por classe obtida com os 10 modelos categóricos construídos usando balanceamento de classes



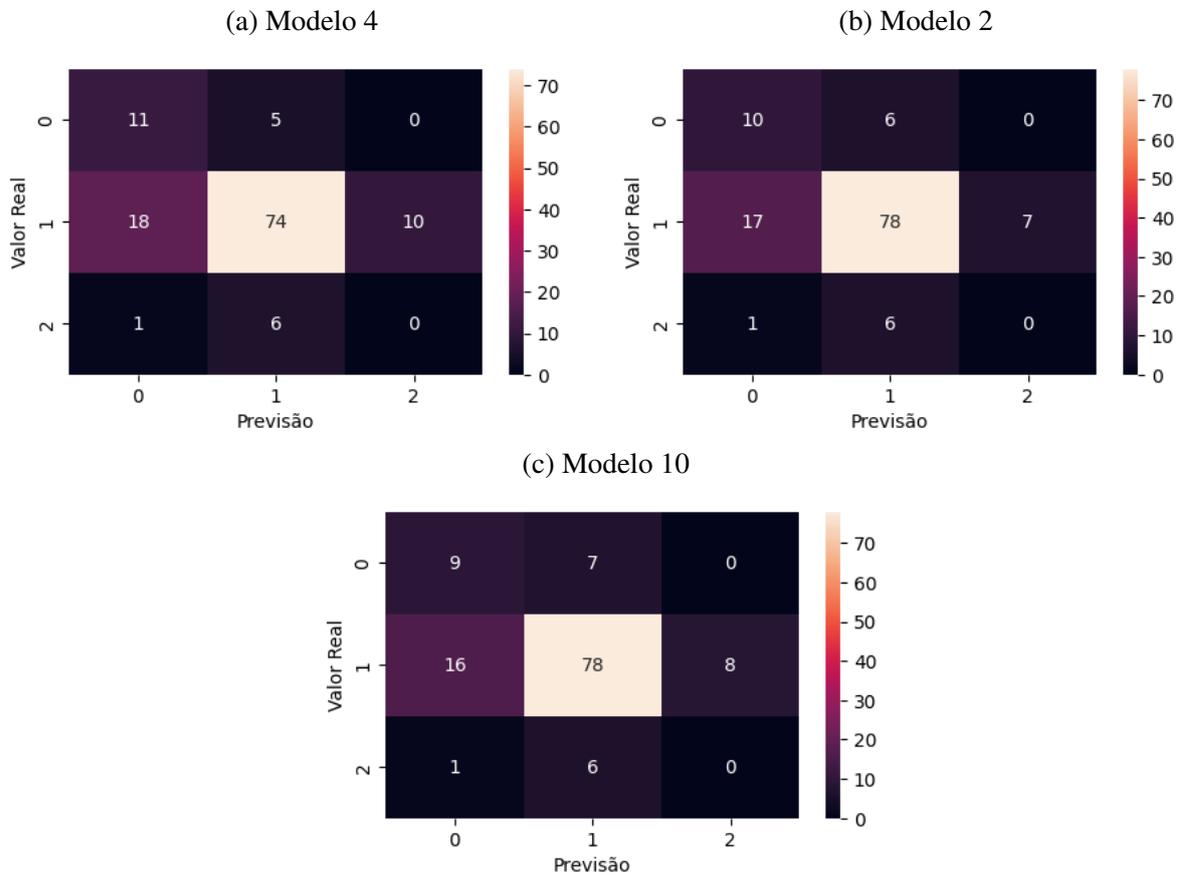
Fonte: elaborado pela autora.

A Figura 16 mostra a matriz de confusão desses modelos. A medida que a acurácia da classe baixa vai aumentando, a quantidade de observações que são da classe normal, mas foram preditas como pertencentes à classe baixa também aumenta. Comparando esses 3 modelos com o modelo 10 que foi obtido sem balanceamento de classe, os modelos 2 e 10 se mostram inferiores, visto que o modelo sem balanceamento de classes tem a mesma acurácia ou maior para a faixa baixa e tem menor percentual de falsos-positivos para essa faixa, 12,8% contra 16,5 e 15,6%, para os modelos 2 e 10 respectivamente. Dessa forma, o melhor modelo treinado com balanceamento de classes é o modelo 4.

Por fim, os três melhores modelos - modelo contínuo 8, modelo categórico e sem balanceamento de classes 10 e modelo categórico com balanceamento de classes 4 - foram comparados com o intuito de identificar o mais adequado. A Tabela 1 apresenta o total de parâmetros de cada modelo e as Tabelas 2 e 3 exibem um resumo com a acurácia global e por classes e com o percentual de falsos-positivos por classes para cada um desses modelos.

Avaliando a quantidade de parâmetros utilizados em cada modelo, a rede contínua 8 é a mais complexa e com isso tem maior custo computacional para predizer a eficiência de redução com base nas variáveis de entrada usadas nesse trabalho.

Figura 16 – Matriz de confusão dos modelos categóricos construídos usando balanceamento de classes.



Fonte: elaborado pela autora.

Em função dos ganhos de processo e financeiros abordados no Capítulo 1, a princípio, não acertar a predição da classe alta, como aconteceu nos modelos 10 e 4, não seria muito crítico, a não ser que ela tenha sido classificada como baixa. Para esses três modelos, uma das sete observações da classe alta foi apontada como baixa. Ou seja, com relação à classe alta, os três modelos estão no mesmo patamar. Avaliando a classe normal a melhor acurácia foi conquistada pelo modelo 10, a qual é quase 10 pontos superior ao modelo 8 e quase 15 pontos maior que a do modelo 4. A acurácia da faixa normal é importante, visto que ela representa a maior parte dos dados e dessa forma confere maior credibilidade ao modelo.

Por fim, analisando a classe baixa, verifica-se que o aumento da acurácia resulta no aumento do percentual de falsos-positivos. O modelo 10, quando comparado ao modelo 8, tem um ganho de mais de 6 pontos percentuais na acurácia e um incremento de apenas 1,83 no percentual de falsos-positivos. Em função disso e dos demais pontos discutidos, o modelo 10 mostra-se superior ao modelo 8. Por sua vez, comparando o modelo 4 com o modelo 10, o modelo 4 apresenta os mesmos 6,25 pontos superiores na acurácia, porém o aumento no percentual de falsos positivos é de 4,59 pontos. A maior quantidade de falsos-positivos faz com que o modelo seja desacreditado pelos usuários e com isso ao longo do tempo eles deixam de usar o modelo. Diante do exposto, o modelo 10, treinado com base em classes sem balanceamento da eficiência de redução, é considerado o mais adequado para auxiliar na operação da Caldeira de Recuperação 3.

Tabela 1 – Avaliação dos modelos em função do total de parâmetros usados.

Modelo, técnica	Total de parâmetros do modelo
8, Contínuo	4.220.113
10, Categórico sem balanceamento de classes	161.411
4, Categórico com balanceamento de classes	161.411

Fonte: elaborado pela autora.

Tabela 2 – Avaliação dos modelos em função da acurácia global e para cada classe de eficiência de redução.

Modelo, técnica	Classe baixa (%)	Classe normal (%)	Classe alta (%)	Global (%)
8, Contínuo	56,25	77,45	14,29	72,80
10, Categórico sem balanceamento de classes	62,50	87,25	0,00	79,20
4, Categórico com balanceamento de classes	68,75	72,55	0,00	68,00

Fonte: elaborado pela autora.

Tabela 3 – Avaliação dos modelos em função dos falsos positivos para cada classe de eficiência de redução.

Modelo, técnica	Classe baixa (%)	Classe normal (%)	Classe alta (%)
8, Contínuo	11,01	52,17	9,32
10, Categórico sem balanceamento de classes	12,84	52,17	0
4, Categórico com balanceamento de classes	17,43	47,83	8,47

Fonte: elaborado pela autora.

5 Considerações Finais

A utilização de RNA para aplicações industriais está em expansão, visto que a robustez dos recursos computacionais aumenta rapidamente ao longo do tempo, o que permite a criação de modelos capazes de prever com elevada taxa de acerto parâmetros cuja medição é dispendiosa. Para construir modelos que representem a realidade e com capacidade de generalização, é de suma importância a qualidade dos dados utilizados e que eles reflitam todos os cenários operacionais. Por isso, é imprescindível o pré-tratamento dos dados, visando identificar e remover valores que não condizem com o processo avaliado.

No presente trabalho, o conjunto de dados foi submetido a alguns filtros, de modo a eliminar condições que não eram relevantes para o modelo e garantir assim a qualidade dos dados. Por meio do método de Pearson foi possível identificar a correlação dos dados de entrada com a eficiência de redução e comparar com as informações disponíveis na literatura. A vazão de licor preto e a vazão de óleo queimado foram as variáveis que apresentaram coeficientes de correlação mais expressivos. Sendo que a primeira tem uma influência positiva na eficiência de redução e a segunda tem relação inversa. A partir da matriz de correlação e das informações disponíveis na literatura sobre as variáveis operacionais que têm influência na eficiência de redução, foram selecionadas 12 variáveis de entrada para construção dos modelos, sendo elas: vazão, concentração e temperatura do licor preto, vazão e pressão do ar primário, vazão do ar secundário, vazão e pressão do ar terciário, oxigênio residual da fornalha, razão entre a vazão total de ar e a vazão de licor queimado, vazão de óleo queimado e sulfidez do licor branco.

Foram construídos modelos utilizando a eficiência de redução de forma contínua, como ela é, sem e com validação cruzada e de forma categórica, com três classes - baixa (<90%), normal (90% a <95%) e alta (>=95%) -, sem e com balanceamento de classes. A definição dos hiperparâmetros dos modelos foi realizada por meio do keras tuner e na sequência os modelos foram treinados com um conjunto de treino e teste. O modelo com maior correlação entre os valores preditos e valores reais foi o desenvolvido com validação cruzada, cujo R^2 é igual a 0,24. Esse valor é considerado baixo, visto que apenas 24% da variância da eficiência de redução era explicada pelo modelo. Em função dessa limitação, a eficiência de redução passou a ser avaliada por classes.

O modelo com melhor capacidade de classificação foi treinando com a variável de saída inicialmente categórica e sem balanceamento de classes. Para este modelo, a acurácia global foi 79,2%. Esse valor cai para a classe baixa, 62,5% e reduz drasticamente para a classe alta, a qual o modelo não foi capaz de prever, que pode ter acontecido devido a existência de poucas observações para essa classe. Apesar dessa deficiência, como a principal vantagem do modelo é informar os momentos em que a eficiência de redução está na faixa baixa, entende-se que a utilização deste no dia a dia pode auxiliar na realização de ajustes dos parâmetros da Caldeira de Recuperação, como regular a vazão de ar e a temperatura do licor, de forma mais rápida, resultando em uma operação mais estável em relação à eficiência de redução.

Como sugestões para trabalhos futuros seria interessante utilizar uma base de dados maior que contenha uma quantidade de dados das classes baixa e alta mais representativa com o objetivo de melhorar a previsibilidade do modelo para essas classes. Outra recomendação é testar a utilização de outras variáveis de entrada disponíveis a fim de encontrar um modelo que consiga prever com elevada precisão a eficiência de recuperação de forma contínua. Por fim, seria proveitoso fazer uma análise de sensibilidade para melhorar a compreensão dos efeitos das variáveis de entrada na eficiência de redução, variável de saída, a fim de direcionar a ação dos operadores da Caldeira de Recuperação para as variáveis de maior impacto.

Referências

- ALMEIDA, G. M.; CARDOSO, M.; OLIVEIRA, E. D.; PARK, S. W. Predição da geração de vapor e da razão: toneladas de vapor gerado por tonelada de sólidos secos em caldeiras de recuperação via modelos de redes neurais. **O Papel**, LXIV, n. 10, p. 99–105, 2003.
- ALMEIDA, G. M. *et al.* Desenvolvimento de metodologia e de modelos para variáveis de performance em caldeiras de recuperação via redes neurais artificiais. Universidade Federal de Minas Gerais, 2003.
- ALMEIDA, G. M. D.; CARDOSO, M.; RENA, D. C.; PARK, S. W. Graphical representation of cause-effect relationships among chemical process variables using a neural network approach. **International Journal of Computational Intelligence and Applications**, World Scientific, v. 9, n. 01, p. 69–86, 2010.
- BELISÁRIO, A. B. *et al.* Análise de emissões em caldeiras de recuperação química de fábricas de celulose kraft: predição e análise de sensibilidade com redes neurais artificiais. Universidade Federal de Minas Gerais, 2020.
- BELISÁRIO, A.; EDBERG, A.; BJÖRK, M.; ALMEIDA, G.; VAKKILAINEN, E. On the diagnosis of a fouling condition in a kraft recovery boiler: combining process knowledge and data-based insights. **Tappi Journal**, v. 22, p. 162–171, 04 2023.
- BRAGA, A. de P.; CARVALHO, A. de L. F.; LUDERMIR, T. **Redes neurais artificiais: teoria e aplicações**. [S.l.]: LTC Editora, 2007. ISBN 9788521615644.
- CARDOSO, M.; COSTA, G. A.; OLIVEIRA, E. de; RAVAGNANI, S. Simulation of eucalyptus kraft black liquor combustion in industrial recovery boilers. **Latin American applied research**, v. 42, n. 2, 2012.
- CAUCHICK-MIGUEL, P. A.; FLEURY, A.; MELLO, C. H. P.; NAKANO, D. N. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. 3. ed. Rio de Janeiro: Elsevier, 2018. 244 p. ISBN 9788535291346. Disponível em: <<https://integrada.minhabiblioteca.com.br/books/9788595153561>>. Acesso em: 8 abr. 2022.
- CHANDRA, Y.; EMPIE, H. J. Impact of deadload reduction in the kraft chemical recovery system. **Tappi Journal**, TECH ASSN PULP PAPER IND INC 15 TECHNOLOGY PARK SOUTH, NORCROSS, GA 30092 USA, v. 6, n. 6, p. 20–25, 2007.
- DAMASCENO, A.; CARNEIRO, L.; ANDRADE, N.; VASCONCELOS, S.; BRITO, R.; BRITO, K. Simultaneous prediction of steam production and reduction efficiency in recovery boilers of pulping process. **Journal of Cleaner Production**, v. 275, DEC 1 2020. ISSN 0959-6526.
- FERNÁNDEZ, A.; GARCIA, S.; HERRERA, F.; CHAWLA, N. V. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. **Journal of artificial intelligence research**, v. 61, p. 863–905, 2018.
- FERREIRA, D. J. d. O.; CARDOSO, M.; PARK, S. W. Gas flow analysis in a kraft recovery boiler. **Fuel Processing Technology**, Elsevier, v. 91, n. 7, p. 789–798, 2010.

- FREDERICK, W. J.; SHI, B.; EUHUS, D. D.; ROUSSEAU, R. W. Crystallization and control of sodium salt scales in black liquor concentrators. **Tappi Journal**, v. 3, n. 6, p. 7–13, 2004.
- GIL, A. C. *et al.* **Como elaborar projetos de pesquisa**. [S.l.]: Atlas São Paulo, 2002. v. 4.
- GRACE, T. M.; TRAN, H. The effect of dead load chemicals in the kraft pulping and recovery system. **Tappi Journal**, v. 8, n. 7, p. 18–24, 2009.
- HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. [S.l.]: Bookman Editora, 2007.
- MACCALLUM, C. First maximize then optimize the reduction efficiency in a kraft recovery boiler. **Pulp & Paper-Canada**, v. 116, n. 4, p. 20+, JUL-AUG 2015. ISSN 0316-4004.
- MARCONI, M. d. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 5. ed. [S.l.]: São Paulo: Atlas, 2003.
- NICKOLAS, S.; SHOBHA, K. Efficient pre-processing techniques for improving classifiers performance. **Journal of Web Engineering**, River Publishers, v. 21, n. 2, p. 203–228, 2022.
- RODRIGUES, L.; RODRIGUES, E.; ALBANI, C.; REIS, A. D.; LOUZADA, D.; SAGRILLO, V. D. Resíduo do processo kraft (dregs) como matéria-prima alternativa para cerâmica vermelha. **Cerâmica**, SciELO Brasil, v. 65, p. 162–169, 2019.
- SAINLEZ, M.; HEYEN, G. Recurrent neural network prediction of steam production in a kraft recovery boiler. In: **Computer Aided Chemical Engineering**. [S.l.]: Elsevier, 2011. v. 29, p. 1784–1788.
- SHEPHERD, A. J. **Second-order methods for neural networks: Fast and reliable training methods for multi-layer perceptrons**. [S.l.]: Springer Science & Business Media, 2012.
- SILVA, I. N. d.; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais para engenharia e ciências aplicadas**. 5. ed. [S.l.]: São Paulo: Artliber, 2010.
- SILVA, W.; RIBEIRO, J.; JR, E. da C.; COSTA, A. D. Reduction efficiency prediction of cenibra's recovery boiler by direct minimization of gibbs free energy. **Brazilian Journal of Chemical Engineering**, SciELO Brasil, v. 25, p. 603–611, 2008.
- SOUZA, F. A.; ARAÚJO, R.; MENDES, J. Review of soft sensor methods for regression applications. **Chemometrics and Intelligent Laboratory Systems**, Elsevier, v. 152, p. 69–79, 2016.
- VAKKILAINEN, E. *et al.* **Kraft recovery boilers—Principles and practice**. [S.l.]: Valopaino Oy, 2005.