UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO
INSTITUTO DE QUÍMICA DE ARARAQUARA

Camila Pierroti Sampaio

ESTUDO COMPARATIVO ENTRE ÁRVORES DE DECISÃO E REDES NEURAIS ARTIFICIAIS PARA DETERMINAÇÃO DO TEOR DE SILÍCIO NO FERRO-GUSA EM ALTO-FORNO

ARARAQUARA
2022

CAMILA PIERROTI SAMPAIO

ESTUDO COMPARATIVO ENTRE ÁRVORES DE DECISÃO E REDES NEURAIS ARTIFICIAIS PARA DETERMINAÇÃO DO TEOR DE SILÍCIO NO FERRO-GUSA EM ALTO-FORNO

Trabalho de Conclusão de Curso, apresentado no curso de graduação de Engenharia Química do Instituto de Química da Universidade Paulista Júlio de Mesquita Filho como requisito para obtenção do título de Bacharel em Engenharia Química.

Aprovado em: 02 de Agosto de 2022

BANCA EXAMINADORA

Érica Filletti

Profa. Dra. Érica Regina Filletti Nascimento
Instituto de Química – UNESP, Araraquara.

Gustavo A. Vieira

Prof. Dr. Gustavo Nakamura Alves Vieira
Instituto de Química – UNESP, Araraquara.

Prof. Dr. Elias de Souza Monteiro Filho
Instituto de Química – UNESP, Araraquara.

AGRADECIMENTO

Agradeço primeiramente à minha família, em especial minha mãe, pai e avó, por sempre me apoiarem, tanto na vida profissional e acadêmica, quanto na vida pessoal.

Aos que possibilitaram a realização deste trabalho, Matheus Santos e Tamires Milagres, por todo aprendizado compartilhado durante o estágio que realizei na RedLab.

À Profª Drª Érica Filletti, pelas recomendações e pelas instruções deste trabalho como um todo, mais especialmente no assunto de redes neurais artificiais.

Ao amigo Jordão Natal, por me auxiliar na programação em Python e no assunto de árvores de decisão.

RESUMO

As árvores de decisão e as redes neurais artificiais são algoritmos que fazem parte da tecnologia chamada de aprendizado de máquinas (*machine learning*), utilizada para identificar padrões em bases de dados e realizar predições. As árvores de decisão possuem "nós" que dividem os dados em duas subclasses ("nós filhos"), através de uma pergunta que resulta em duas respostas: sim ou não. Essa pergunta ocorre de tal forma a diminuir o erro quadrático para os casos de regressão (previsões numéricas). Enquanto isso, as redes neurais artificiais se inspiram no modelo biológico, em que a maioria dos modelos realiza uma multiplicação sobre os valores de entrada, análogo à ponderação das conexões de neurônios. Com isso, essas conexões são somadas e passam por uma função de ativação, antes de apresentar o resultado. Essas duas ferramentas receberam a mesma base de dados e treinaram seus respectivos algoritmos em relação aos valores de exemplo fornecidos. Ambos os casos, de árvores e redes neurais, foram utilizadas para predição de quantidade de silício presente em ferro-gusa. O silício é importante para controle térmico do alto-forno siderúrgico. Os dois algoritmos apresentaram valores de erro quadrático MSE e percentual MAPE semelhantes. A maior diferença entre os resultados, entretanto, foi em relação à linearização da reta construída entre valores calculados e reais para o conjunto de validação dos modelos, em que o coeficiente de determinação R^2 foi de 0,363 para árvores e 0,497 para redes neurais.

Palavras-chave: aprendizado de máquina; árvore de decisão; rede neural artificial; quantidade de silício.

ABSTRACT

Decision trees and artificial neural networks are algorithms that compose the technology called machine learning, used to identify patterns in databases and make predictions. Decision trees have "nodes" that divide the data into two subclasses ("child nodes"), by asking a question that results in two answers: yes or no. This question is asked looking for decreasing the squared error of the regression (numerical predictions). On the other hand, artificial neural networks are inspired by the biological model – most models use a multiplication on input values, similar to the weighting of neuron connections. Then, these connections are summed and go through an activation function that will show the result. The same database was used by the two algorithms, that trained their model with the given output. Both tree and artificial networks were used to predict the amount of silicon present in molten iron. Silicon is important because of the thermal control of the blast furnace. The two algorithms showed similar mean squared error MSE and similar mean absolute percentual error MAPE. However, the biggest difference among the results was on the linearization between the calculated data and the actual value for the validation set, with a coefficient of determination R^2 of 0.363 for trees and 0.497 for artificial neural networks.

Keywords: machine learning; decision tree; artificial neural network; silicon amount.

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

A metalurgia é definida como um conjunto de tratamentos, incluindo processos físicos e químicos, para extração de metais a partir de seus minérios. Dessa forma, a siderurgia é utilizada para a obtenção do ferro, um dos constituintes do aço. O aço é uma liga que contém como principais componentes o ferro e o carbono, sendo formado geralmente por 0,002% a 2% em peso de carbono. Outros elementos podem estar presentes, de acordo com a composição final desejada, como o silício[1].

No processamento do aço em usina integrada há diferentes etapas: a redução no alto-forno, o transporte no carro-torpedo, a oxidação no conversor, a correção da composição no refino secundário e a solidificação no lingotamento. Outras etapas podem ser adicionadas, conforme o tipo de aço desejado[1].

O alto-forno é considerado o reator mais complexo da siderurgia. Nele, há coexistência dos três estados da matéria, interagindo entre si: sólido, líquido e gás. O processo em seu interior é acompanhado de alta pressão e de alto gradiente temperatura – as temperaturas variam entre, aproximadamente, 150 ºC na parte superior (saída de gases) e 2000 ºC, na frente das ventaneiras (região que ocorre a combustão do coque)[1].

No alto-forno, são formados o ferro-gusa (liga de ferro e carbono) e a escória (mistura de óxidos indesejáveis) em duas diferentes camadas. O ferro-gusa segue para as etapas seguintes, para obtenção do aço, enquanto a escória pode ser tratada para comercialização. O silício é o elemento de liga mais importante dentro do ferro-gusa em relação ao controle térmico do alto-forno[1].

Além de controle térmico, o silício auxilia na remoção de impurezas do ferro-gusa, através da transferência delas para a escória. Consequentemente, um ferro-gusa de maior qualidade é obtido[2]. Na etapa seguinte, no conversor, o silício contribui no fornecimento de energia e, então, na redução do gasto energético e dos gastos econômicos[6].

O teor de silício é dificilmente identificado em tempo real, graças à complexidade e às altas temperaturas do alto-forno[2]. Neste contexto, a análise de dados para predizer a quantidade de silício são importantes para criar modelos e prever a condição do produto, inferindo também sobre os custos envolvidos no processo.

Para realizar essa análise dos dados, árvores de decisões e redes neurais artificiais foram os modelos de aprendizado de máquina escolhidos neste trabalho, para encontrar padrões e interpretar os parâmetros fornecidos.

Portanto, o objetivo desse trabalho é prever a quantidade de silício presente no ferro gusa por meio de modelos de aprendizado de máquina. Logo, busca-se:

- Construir um modelo de aprendizado de máquina baseado em árvores de decisão;
- Construir um modelo de aprendizado de máquina baseado em redes neurais artificiais;
- Comparação entre os modelos citados acima.

2 REVISÃO BIBLIOGRÁFICA

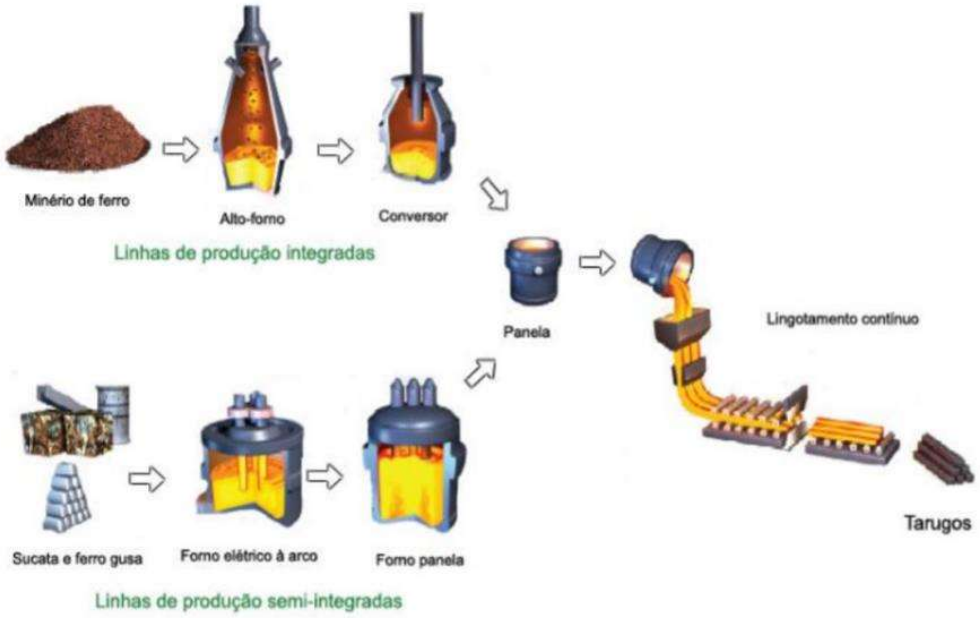
Nesta seção são apresentados os principais conceitos para entendimento da indústria siderúrgica e da consequente importância do silício, além de conceitos para compreensão do aprendizado de máquina, incluindo as etapas preliminares de preparação da base de dados.

2.1 Processamento siderúrgico

As indústrias siderúrgicas realizam a extração de ferro a partir de seu minério. No Brasil, o teor de ferro presente nas reservas é de 64,7% em média, graças ao alto teor de ferro nas hematitas, que variam de 60 a 67%[3]. Durante o processamento do aço, as etapas são realizadas de tal forma a obter a porcentagem de carbono desejada, formando aços de baixo, médio e alto teor do mesmo[1].

No processamento de aço, há dois tipos principais de usinas: as integradas e as semi-integradas. Para as primeiras, as matérias-primas para obtenção de aço incluem o minério de ferro. Para o segundo tipo de usina, também chamadas de recicladoras, o aço é obtido a partir da sucata de aço e, portanto, a etapa de redução do minério não é necessária[1]. A Figura 1 mostra um esquema do processamento do aço para esses dois casos, sendo que neste trabalho há foco nas usinas integradas.

Figura 1 - Comparação entre usina integrada e semi-integrada para produção de aço.



Fonte: MATTIELLO (2011).

O processo nas usinas integradas inicia-se com a seleção de matérias-primas. Após encontrar o minério, há desmonte (fragmentação através de explosivos) e britagem (liberação da hematita). Em seguida, ocorre a classificação da partícula de acordo com sua granulometria, sendo que as menores partículas (abaixo de 8 mm) sofrem processo de concentração de produção de pelotas e sínter[1].

A próxima etapa é o alto-forno, que transforma as matérias-primas em ferro-gusa. O ferro-gusa é uma liga de ferro e carbono produzido através da interação entre o minério de ferro e o carvão, realizando a redução de óxido de ferro III (Fe_2O_3), conforme a reação abaixo. Há pelo menos 92% de ferro e entre 3,5% até 4,5% de carbono[1,4].

$$Fe_2O_{3\,(s)} + 3\,CO_{(g)} \rightarrow 2\,Fe_{(s)} + 3\,CO_{2\,(g)}$$

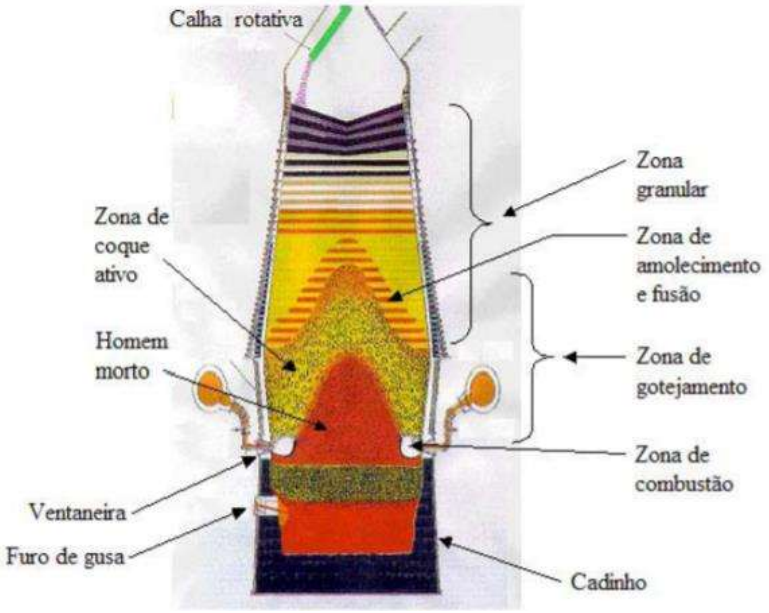
Assim, no alto-forno ocorre a redução das matérias-primas, que são: a carga metálica, (composto de pelotas, sínter e/ou minério granulado), o combustível sólido, geralmente coque ou carvão vegetal, e os fundentes e as injeções auxiliares (auxiliam na combustão, como gás natural, carvão pulverizado e outros) [1].

Conjuntamente com o ferro-gusa, é produzida a escória (mistura de óxidos, silicatos metálicos e metais) que, graças à diferença de densidade, permanece separada do gusa no interior do alto-forno[1].

Dessa forma, o alto-forno é formado por diferentes zonas, identificadas na Figura 2:

- Zona granular: combustível e carga metálica mantêm-se em duas camadas separadas, da mesma forma em que foram carregadas para o interior do alto-forno[1];
- Zona coesiva: camadas de combustível e carga metálica alternadas, em que a temperatura é suficiente para o início do amolecimento e fusão da carga. Por isso, na Figura 2, esta camada é chamada de "Zona de amolecimento e fusão" [1];
- Zona de gotejamento: nos interstícios do combustível sólido, há gotejamento de ferro gusa e escória. Essa zona contém duas subdivisões - homem-morto e zona de coque ativo[1];
- Zona de combustão: região ao lado das ventaneiras e que, portanto, está parcialmente vazia[1];
- Cadinho: possui duas camadas líquidas, gusa e escória. O cadinho, então, possui perfurações (furo de gusa) que possibilitam o vazamento destas duas camadas para a panela, próxima etapa de processamento do aço[1].

Figura 2 – Zonas presentes no alto-forno.



Fonte: VIEIRA (2012).

Para a reação de combustão, o combustível utilizado é, em geral, coque ou carvão, que além de serem fontes de calor, devem formar um leito permeável que suporte a carga. O coque é uma fonte não-renovável de energia, mais poluente que o carvão vegetal e exige etapa de

coqueificação do carvão mineral para ser utilizado. Entretanto, o coque é o mais utilizado e o mais barato, sendo o carvão vegetal usado principalmente em pequenas instalações, mesmo sendo uma fonte renovável[1].

Após o alto-forno, a escória é drenada para potes ou transformada para ser comercializada. O ferro gusa segue para o refino primário no conversor, onde há reações de oxidação. Em seguida, há o refino secundário na panela. Em ambas as etapas de refino, as propriedades são ajustadas e o ferro-gusa é transformado em aço[1].

Por último, ocorre o lingotamento que solidificará o aço líquido na forma desejável. Assim, por se tratar de processos em alta temperatura, o controle térmico das operações para obtenção do aço é de extrema importância e, portanto, o silício é um elemento chave do processo[1].

2.2 O silício do ferro-gusa

O controle térmico do alto-forno é realizado principalmente através do teor de silício e da temperatura do ferro-gusa. Uma temperatura muito alta de ferro-gusa pode significar consumo desnecessário de combustível, além de perda de produtividade. Por outro lado, baixa temperatura pode entupir os furos de gusa e, então, dificultar a saída dos produtos do alto-forno, resultando em maior custo de produção[2].

O silício é considerado o elemento de liga (elementos adicionados para melhorar determinada propriedade do produto) mais importante no ferro-gusa para controle térmico do alto-forno[1]. Um teor suficiente deste elemento significa aquecimento do alto-forno e maior facilidade para remoção de impurezas, como fósforo e enxofre, através de sua passagem para a escória[2].

Por outro lado, havendo silício em excesso, o ferro-gusa se torna rígido e quebradiço, além de indicar desperdício de matéria prima (coque em excesso). O baixo teor de silício também é problemático, visto que significa um resfriamento do alto-forno. Então, o ideal é haver um controle conjunto da temperatura e da composição, com porcentagem de silício de $0{,}4\% \pm 0{,}7\%$ no gusa[2].

A presença de sílica, principal fonte de silício, ocorre principalmente a partir dos minérios, das cinzas de coque, obtidas após sua combustão, e do carvão mineral pulverizado nas ventaneiras. A incorporação do silício no ferro-gusa pode ocorrer de maneira direta ou indireta[5], conforme mostra a Figura 3.

Na maneira direta, a sílica (SiO_2) presente na escória reage com o carbono (C) do ferro-gusa, e o silício (Si) é transferido diretamente da escória ao ferro-gusa, gerando monóxido de carbono (CO), conforme a reação abaixo[1]. Essa reação é altamente dependente da temperatura, sendo endotérmica e, assim, quanto maior a temperatura, maior a incorporação de Si no gusa[5].

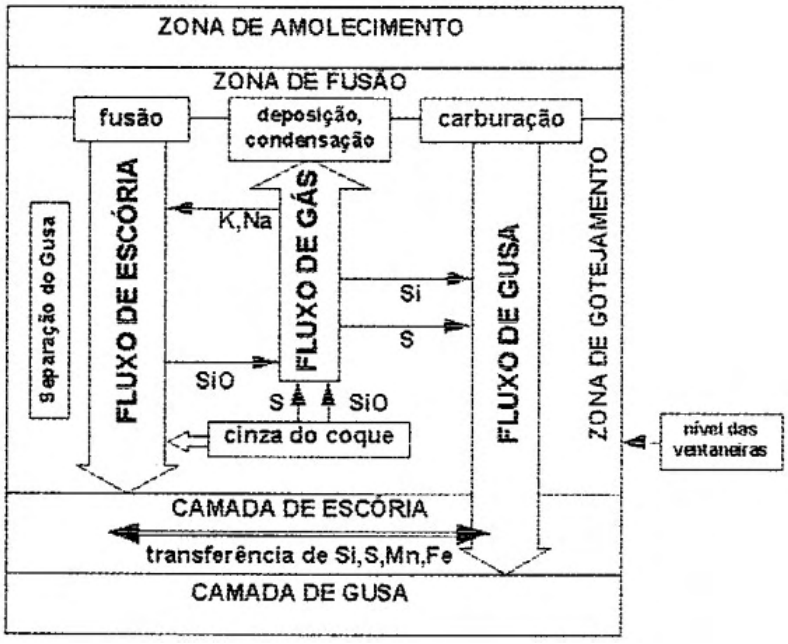
$$SiO_2 + 2\,C \rightarrow Si + 2\,CO$$

No método indireto, o SiO_2 da escória ou cinza é transferido para a fase gasosa, em forma de monóxido de silício (SiO). Em seguida, há reação com o carbono presente no gusa, transferindo o silício para o ferro-gusa. As duas reações para o método indireto estão representadas abaixo. Essa é a principal rota de incorporação do silício, visto que o método direto ocorre de forma mais lenta[1].

$$SiO_2 + \ C \rightarrow SiO + CO$$

$$SiO + C \rightarrow Si + \ CO$$

Figura 3 - Reações de incorporação de elementos de liga no alto-forno



Fonte: MOURÃO et al. (2007).

Dessa forma, o teor de silício presente no ferro-gusa indica três importantes fatores: o estado térmico do alto-forno, a qualidade do produto (interferindo na qualidade do aço) e quão

custoso será o processo de produção[6]. O teor de Si é também um parâmetro importante na escória, pois quanto maior a presença de sílica, maior o volume de escória. Um alto volume pode significar paradas no alto-forno para vazamento de escória, além de aumentar o risco de projeções para fora do conversor, etapa seguinte do processo[5].

Além disso, no conversor, o aporte térmico do ferro-gusa é fundamental para fornecimento de energia, diminuindo o custo da operação. Isso ocorre principalmente graças à presença de silício e sua reação exotérmica de oxidação[6]. Entretanto, silício em excesso, dificulta a conversão do gusa em aço, pois aumenta os gastos para obter o produto[5].

Sendo o alto-forno um reator fechado de altas temperaturas, são poucos os métodos e instrumentos para detectar previamente o índice de silício no ferro-gusa, na zona do cadinho[2]. Portanto, analisar os dados relacionados ao teor de silício, assim como utilizar ferramentas para predizer este parâmetro, são grandes estratégias para a siderurgia, impactando não somente na redução de custo, mas também na qualidade do produto.

2.3 Estruturação da análise e do tratamento de dados

A demanda pela análise de dados pelas empresas é cada vez maior. A análise de dados combinada com conceitos de matemática e estatística pode levar a informações consistentes, confiáveis e estratégicas. Muitas vezes, somente um método de modelização não gera os resultados esperados e, por isso, é necessário observar os dados mais a fundo e verificar qual método se adequa mais às necessidades da empresa[7].

No tratamento de dados, os dados são tratados como variáveis e podem possuir valores numéricos (ou quantitativos) ou não numéricos (ou qualitativos). Dentre os valores numéricos, há valores que variam no intervalo de números reais, chamados contínuos, ou que variam em números inteiros, chamados valores discretos[7]. A Figura 4 mostra as etapas da estatística descritiva, que estão detalhadas a seguir.

Figura 4 - Etapas da estatística descritiva.

Fonte: FERREIRA et al. (2021).

a) Etapas de coleta e organização

Nas duas primeiras etapas, coleta e organização, deve-se ter uma visão clara do problema em questão para, então, definir como serão coletados esses dados e posteriormente organizados. Assim, os objetivos da análise devem estar claros para verificar como e quais dados serão coletados e quais serão as colunas e linhas dessa base de dados[7].

b) Etapa de tratamento

Para a etapa de tratamento, procura-se adequar os dados aos objetivos da análise. Por exemplo, se há muitos dados faltantes, pode-se voltar à etapa inicial e coletar mais dados, para uma base mais completa. Além disso, os dados de uma variável podem ser normalizados, quando deseja-se definir uma escala fixa[7].

Caso haja erros nos valores de uma variável, deve-se realizar uma escolha: deletar estes dados (substituir por nulo), substituir por zero, realizar uma regressão linear para selecionar novos valores ou substituir pela média são algumas das opções[7].

Outro método necessário dentro do tratamento de dados é identificação de valores extremos, ou seja, que estão muito distantes da média dos valores e que aumentam, portanto, o desvio padrão. Esses valores extremos podem ser mantidos, removidos ou transformados. Novamente, a decisão de manter ou não esses valores depende de cada variável e de como eles podem impactar negativamente os resultados[7].

c) Etapa de análise

Na análise de dados, há três categorias principais: a análise univariada, em que cada variável é avaliada separadamente. Na análise multivariada, são selecionadas relações entre duas ou mais variáveis para impactarem uma outra variável e, assim, avaliar seus efeitos. Por último, na análise por correlação há duas variáveis que se comportam de forma sincronizada[7].

d) Etapa de apresentação e interpretação

Na apresentação e interpretação, pode-se optar por representação tabular ou gráfica. As tabelas fornecem informações mais detalhadas sobre os dados e de como eles podem ser tratados. Por outro lado, os gráficos, por serem mais intuitivos e de rápida compreensão, são a ferramenta utilizada para uma visualização mais fácil de valores quantitativos. Ainda, existem diferentes tipos de gráficos a serem escolhidos: barras, linhas, pontos e outros[7].

A apresentação de dados deve permitir visualizar os pontos chave observados durante a análise de dados: por exemplo, se há uma correlação, um gráfico em linhas pode ser escolhido para destacar este comportamento. A interpretação de dados, então, permite escolher qual modelo será aplicado para solucionar o problema definido no início[7].

Muitas vezes, a interpretação de dados não ocorre somente a partir da observação de gráficos ou de modelizações mais simples, como a regressão linear, visto que o modelo escolhido pode não se encaixar aos dados de forma adequada. Nesses casos, pode-se optar por ferramentas de aprendizado de máquina (do inglês *machine learning*)[7].

2.4 Aprendizado de máquina

O aprendizado de máquina é uma das partes da inteligência artificial. O objetivo do *machine learning* é, então, utilizar dados e algoritmos para aprendizado, da mesma forma que humanos, visando melhorar os resultados e sua acurácia. Seu uso geralmente ocorre para predição de valores ou para classificação. Em ambos os casos, utiliza-se um algoritmo que, através da base de dados, criará um padrão[8].

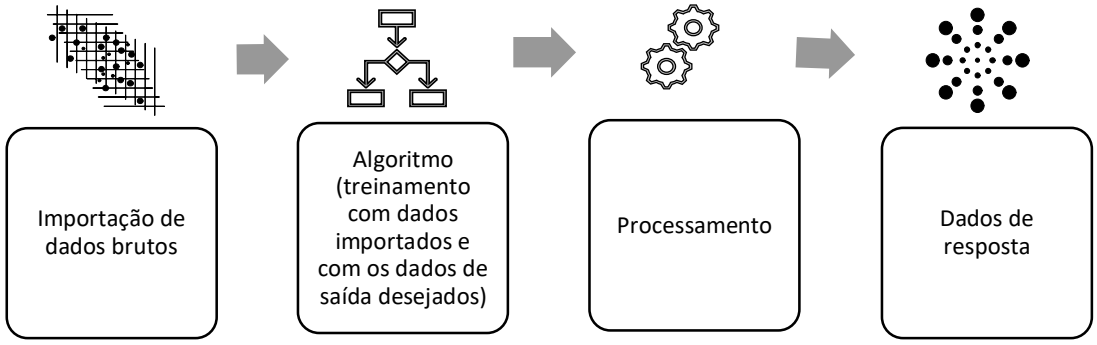
Nos problemas de classificação, há por exemplo a decisão de se um e-mail é spam ou não. Como exemplo de predição, há a previsão de vendas de artigos que permite, então, focar em um público-alvo ou em um produto[8].

Depois desse processo de decisão, em que o algoritmo classificará ou predirá algo, o modelo pode ser avaliado por uma função de erro. Então, essa função compara os resultados obtidos com os reais. Por último, pode ser realizada uma optimização do processo, buscando diminuir os valores de erro[8].

Os métodos de *machine learning* podem ser supervisionados, em que parte dos dados são usados para treinar os algoritmos, contendo já os dados de saída. Por outro lado, no *machine learning* não supervisionado, o algoritmo analisa e descobre padrões nos dados, criando então

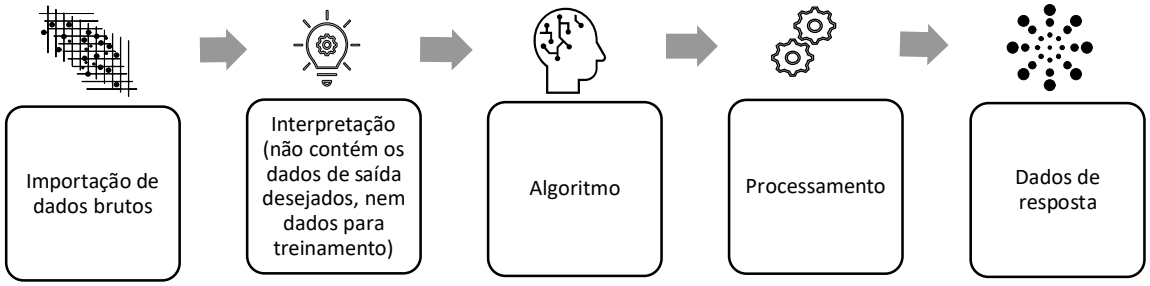
agrupamentos chamados de "*clusters*", não necessitando do fornecimento dos dados de saída[8]. A Figura 5 e a Figura 6 mostram o processo que ocorre, respectivamente, para o método supervisionado e não supervisionado.

Figura 5 - Método de *machine learning* supervisionado.



Fonte: Adaptado de Crayon Data (2018).

Figura 6 - Método de *machine learning* não supervisionado.



Fonte: Adaptado de Crayon Data (2018).

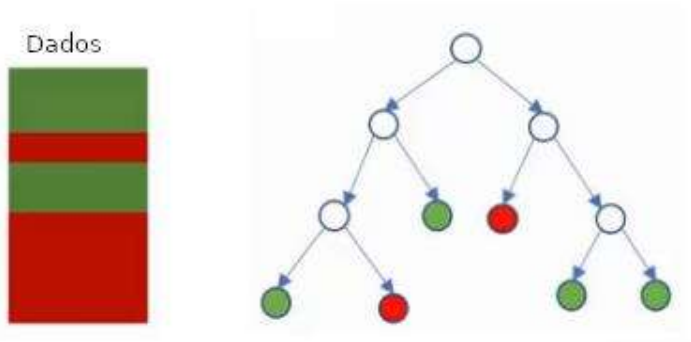
Há, ainda, o semi-supervisionado, que fica entre os dois outros métodos, que é utilizado, geralmente, quando não há dados suficientes – permite que uma base de dados seja treinada através de um conjunto de dados de saída desejado, o qual guiará, em seguida, a classificação de uma base de dados maior (que não possui dados de saída a serem treinados)[8].

O aprendizado de máquina contém diversos tipos de algoritmos e subdivisões. Por exemplo, as redes neurais artificiais são campos dentro do *machine learning* e, mais especificamente, do *deep learning*, que é o aprendizado mais profundo - comparado com o *machine learning*, o *deep learning* permite bases de dados maiores e tenta eliminar ainda mais a intervenção manual humana[8]. Há, ainda, outros métodos de *machine learning*, como as árvores de decisão. Ambos os casos, de redes neurais e de árvores, foram utilizados nesse trabalho para o caso de predição de valores.

2.4.1 Árvores de decisão

As árvores de decisão classificam itens de acordo com suas características, fazendo uma série de perguntas em cada nível. Cada questão forma um nó, o qual irá gerar duas respostas possíveis (sim ou não), formando então duas subclasses. Portanto, o nó raiz (primeiro nó) possui o conjunto completo de dados e, conforme descemos a hierarquia, os dados são classificados de acordo com sua probabilidade de pertencerem à categoria em questão[9]. A Figura 7 mostra um exemplo de árvore aleatória.

Figura 7 - Exemplo de árvore de decisão.



Fonte: ANSELMO (2020).

Dessa forma, cada nó da árvore representa uma equação diferente para separação das subclasses. Diferentes algoritmos podem ser utilizados na construção das árvores, dentre eles o CART (árvores de classificação e regressão ou, do inglês, *Classification and Regression Trees*). Nele, são construídas árvores binárias (cada nó possui dois nós-filhos) cujos nós são produzidos de acordo com o maior ganho de informação[10].

Diferentes critérios podem ser utilizados como parâmetro para ganho de informação. Dentro da biblioteca ScikitLearn da linguagem Python, utilizada neste trabalho, o critério padrão é a minimização do erro $H(Q_m)$ mostrado na Equação (1) [10]. Nela, y é o valor real, $\overline{y_m}$ a média dos preditos pelo modelo e m é o número do nó em questão dentro da base de dados Q_m, a qual contém um total de n_m valores.

$$H(Q_m) = \frac{1}{n_m} \sum_{y \in Q_m} (y - \overline{y_m})^2 \qquad \textbf{(1)}$$

Muitas vezes, árvores de decisão são o modelo de *machine learning* escolhido por sua facilidade de interpretação, visto que apresentam perguntas relativamente simples sobre os

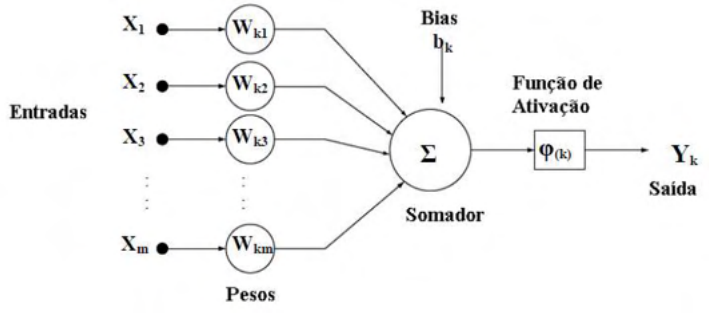
dados em cada nó. Por outro lado, pequenas mudanças nos dados de entrada podem levar a grandes alterações nos resultados das árvores de decisão[9].

Então, para realizar este modelo, a base de dados é dividida em dois conjuntos: treinamento e validação. Para o caso supervisionado, tratado neste trabalho, o objetivo é obter valores utilizando o conjunto de treino, que serão posteriormente validados no outro conjunto. A finalidade é atingir valores preditos (valores de saída) próximos aos valores exemplo fornecidos (neste caso, valores de quantidade de silícios obtidos experimentalmente)[11].

2.4.2 Redes Neurais Artificiais

As redes neurais artificiais (RNA's) possuem características de desempenho em comum com o modelo biológico, inspirando-se neste. A primeira RNA foi criada em 1943, por McCulloch & Pitts, permitindo o desenvolvimento de modelos matemáticos que procuravam analogias biológicas. Assim, o neurônio era conhecido como uma unidade de processamento binária, representado na Figura 8[12].

Figura 8 - Modelo de RNA segundo McCulloch & Pitts.



Fonte: SOUZA et al. (2012).

Conforme mostra a figura acima, os dados (ou vetores de dados) X_m entram na rede e os sinais são propagados através de conexões. Um peso associado w_{km} é atribuído a essas conexões antes da função aditiva (representada por Σ). Ainda, há a aplicação de uma função de ativação $\varphi_{(k)}$, que é geralmente não linear, na entrada da rede. O bias b_k é capaz de regular o efeito da função de ativação. Por último, é gerada uma saída Y_k [12,13,14].

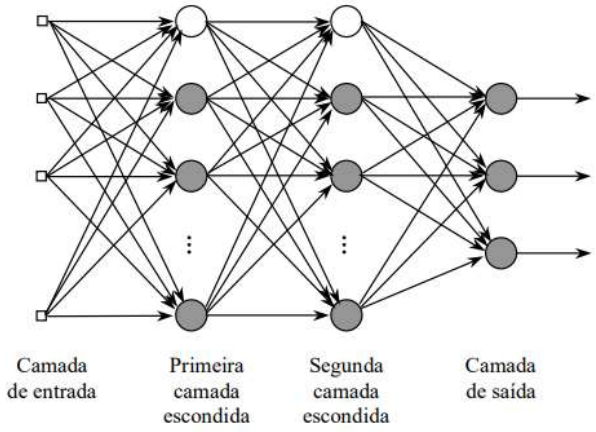
Outros trabalhos foram importantes para a evolução das RNA's, podendo ser citado o de Rosenblatt (1958), baseado em um neurônio biológico. Ele desenvolveu o modelo do *Positron*, em que os dados de entrada são divididos em duas classes por um hiperplano e a

aprendizagem ocorre por ajuste iterativo dos pesos sinápticos. Esse ajuste, em geral, ocorre por minimização dos erros entre o valor real e calculado[14].

A caracterização de uma rede neural é feita através de três parâmetros: a conexão entre suas unidades (arquitetura), a determinação dos pesos das conexões (algoritmo de treinamento) e a função de ativação. As arquiteturas do tipo multi-camadas Perceptron MLP (do inglês *MultiLayer Perceptron*) são as amplamente conhecidas dentro das RNA's[12].

As MLP's são constituídas por uma camada de entrada, uma ou mais camadas ocultas, e uma camada de saída. Então, o sinal de entrada é transmitido por diversas camadas em direção à saída. Um esquema está na Figura 9, mostrando que o neurônio em cada camada da rede está conectado com todos os neurônios da camada anterior[12].

Figura 9 - Representação da arquitetura multi-camadas Perceptron



Fonte: SILVA (1998).

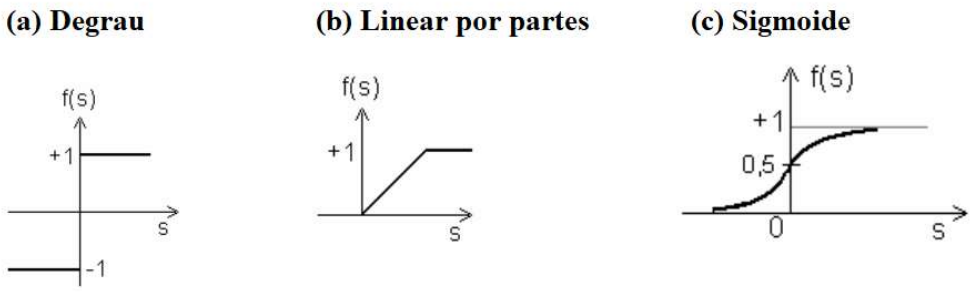
Um dos algoritmos mais utilizados é o de retro propagação (baseado em gradiente descendente), em que primeiro calcula-se o erro de cada saída para, em seguida, calcular o erro propagado no sentido inverso. Com isso, ajusta-se os erros e um novo padrão é apresentado[13]. Outro algoritmo é o de Marquardt-Levenberg, utilizado neste trabalho, que é uma modificação do Método de Gauss-Newton. A diferença é que no método de Marquardt-Levenberg, representado na Equação (2), há a introdução do parâmetro μ, o qual é capaz de fazer a equação tornar-se aquela do método Gauss-Newton quando ele tende a zero[15].

$$\Delta \underline{x} = \left[J^T(\underline{x})\, J(\underline{x}) + \mu I\right]^{-1} J^T(\underline{x})\, \underline{e}(\underline{x}) \qquad \textbf{(2)}$$

sendo J a matriz jacobiana, I a matriz identidade e e(x) o erro associado ao vetor x.

No caso da função de ativação, ela pode ser linear ou não-linear. As funções mais comuns são a função degrau, a linear por partes e a sigmoide, conforme mostra a Figura 10[12].

Figura 10 - Funções de ativação não-lineares.



Fonte: Adaptado de MINUSSI (2008).

As redes neurais possuem a característica de aprenderem e apresentarem bom desempenho para conjuntos de dados incompletos e com ruídos: as RNA's possuem tolerância à falha. Isso ocorre porque, caso um neurônio falhe, há ainda outros neurônios que podem sobrescrever este erro[12].

2.5 Estudos em siderurgia utilizando aprendizado de máquina

Um dos estudos, realizado por FORRER (2012), analisou a predição das características da camada de refratário em siderúrgicas com *machine learning*, mais especificamente na etapa do conversor. Nele foram utilizados métodos supervisionados, lineares e não lineares, mas utilizando uma base de dados relativamente pequena – uma base maior melhoraria a predição dos métodos. Os resultados foram avaliados através de dados estatísticos, com melhores valores para o modelo Gaussiano (comparado com os modelos linear e máquina de vetores de suporte)[16].

Além disso, no artigo de Diniz et al. (2019), o modelo de redes neurais artificiais foi utilizado para predizer a quantidade de silício em ferro-gusa de alto-forno, provando a eficiência de seu algoritmo de poda que, além de diminuir o tamanho da RNA, ainda melhorou o desempenho desta, através da redução da quantidade de variáveis de entrada que, provavelmente, estavam confundindo a rede[6].

Segundo Martins (2019), as tecnologias da indústria 4.0 tendem a aumentar a competitividade e a produtividade das siderúrgicas. Segundo ele, um dos pilares para uma

produção inteligente é a virtualização – criação de uma versão virtual capaz de descrever o mundo físico por modelos matemáticos, o que inclui a inteligência artificial e o aprendizado de máquina[17].

Ainda, O Autor afirma que existem diversos estudos de inteligência artificial realizados por siderúrgicas brasileiras. Entretanto, estes estudos não são publicados por questão de confidencialidade e, pelo mesmo motivo, seu artigo tratou as respostas aos formulários de forma geral, não citando uma siderúrgica em especial.

Portanto, esse campo de pesquisa de teor de silício ligado a aprendizado de máquina ainda pode evoluir muito, sendo uma área de oportunidades para siderúrgicas que desejam aumentar sua eficiência e produtividade, mantendo-se no mercado. Consequentemente, este trabalho de conclusão de curso mostra-se atual e coerente com o contexto da produção de aço.

3 MATERIAIS E MÉTODOS

Nesta seção está a descrição de qual foi a abordagem para analisar os resultados dos dois modelos de aprendizagem de máquina e de como a base de dados foi tratada anteriormente à aplicação dos modelos de árvore de decisão e de redes neurais artificiais.

3.1 Diferentes abordagens para analisar os resultados

Os dados utilizados nesse trabalho foram coletados de um alto-forno durante 12 meses em 2017 e 2018 e 2 meses (janeiro e fevereiro) em 2019, visando realizar um monitoramento do alto-forno e da produção de ferro-gusa, dentro de uma siderúrgica. Esses dados foram tratados e buscou-se desenvolver um modelo adequado de previsões de comportamento do refratário utilizado nesse processo.

Assim, dois diferentes métodos de aprendizado de máquina foram utilizados para a análise de dados: árvore de decisão e rede neural artificial. Para a realização da comparação entre esses dois modelos, a mesma base de dados e as mesmas formas de avaliação foram utilizadas.

Para verificar a performance dos métodos, a regressão linear foi realizada, considerando os valores reais de composição de silício presente no gusa e os valores obtidos a partir do método de aprendizado de máquina, para o conjunto de validação. Com isso, construiu-se um gráfico permitindo obter o coeficiente de determinação R^2, cuja fórmula está representada na Equação (3) [18].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \textbf{(3)}$$

em que y_i é o valor real individual de cada um dos n valores, $\overline{y_i}$ é o valor médio e o $\widehat{y_i}$ valor predito individual da variável.

Então, quanto maior o valor de R^2 (mais próximo de 1), melhor os dados estão ajustados ao modelo e mais próximo os valores obtidos por *machine learning* se aproximam dos valores reais. Entretanto, essa medida sozinha não é capaz de verificar se o modelo é adequado (os valores podem, por exemplo, oscilar entre valores positivos e negativos em torno da linha de tendência e apresentar, ainda, um alto R^2).

Outro parâmetro para comparação de métodos foi o erro quadrático médio MSE (do inglês, *Mean Squared Error*), presente na Equação (4). Essa métrica penaliza as grandes diferenças entre o valor calculado ao valor real, sendo útil para modelos que não toleram

grandes erros. Ainda, foi utilizado o erro percentual absoluto médio MAPE (do inglês, *Mean Absolute Percentage Error*), da Equação (5), que fornece uma porcentagem de afastamento do valor real[18].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad \textbf{(4)}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad \textbf{(5)}$$

Com isso, o MAPE e o MSE foram testados em diferentes conjuntos: sobre o conjunto de treinamento e de validação para os modelos de árvores de decisão e redes neurais artificiais e, ainda, na fase de teste para as redes neurais. Entretanto, anteriormente ao aprendizado de máquina, foi necessário tratar e padronizar os dados que, muitas vezes, estavam fora do formato ideal para utilização.

3.2 Tratamento preliminar dos dados: preparação para o Aprendizado de Máquina

A base de dados veio separada por meses para cada ano, em arquivos Excel. Então, utilizando o programa Jupyter Notebook (Anaconda 3), através da linguagem de programação Python, a base de dados foi preparada para ser utilizada, posteriormente, dentro dos modelos de *machine learning*.

As bibliotecas utilizadas no tratamento de dados foram: Pandas, para criação de tabelas (dataframes), Matplotlib, para realização de gráficos e visualização de dados, e Numpy, que ofereceu ferramentas estatísticas. Então, o primeiro passo foi unir todos os dados de diferentes meses em somente uma base, formando uma "base global de dados".

Nesta base global, as colunas foram renomeadas e os tipos de dados foram alterados (por exemplo, uma coluna numérica que chegou com o tipo "texto" foi modificada para o tipo "número"). Ainda, as colunas contendo somente zeros foram deletadas, resultando na Tabela 1, em que o tipo "datetime64[ns]" é tipo data, "int64" numérico inteiro 64 bits, "object" caracteres (números ou letras) e "float64" numérico ponto flutuante 64 bits.

Tabela 1 - Colunas classificadas pela quantidade de valores não nulos e pelo tipo de dado.

(Continua)

Coluna	Quantidade de não nulos	Tipo de dado
Variável 1	9769	int64
Variável 2	9769	float64
Variável 3	9769	float64
Variável 4	9769	float64
Variável 5	9769	float64
Variável 6	9769	float64
Variável 7	8411	float64
Variável 8	9769	float64
Variável 9	9768	float64
Variável 10	9767	float64
Variável 11	9769	int64
Variável 12	9769	float64
Variável 13	9769	float64
Variável 14	9769	float64
Variável 15	9769	float64
Variável 16	9009	float64
Variável 17	9116	datetime64[ns]
Variável 18	9769	int64
Variável 19	8546	float64
Variável 20	9769	object
Variável 21	9769	object
Variável 22	9769	object
Variável 23	9769	object
Variável 24	9769	object
Variável 25	9769	object
Variável 26	9769	object
Variável 27	9769	object
Variável 28	9769	object
Variável 29	9769	int64
Variável 30	1323	object
Variável 31	3476	object
Variável 32	3164	object
Variável 33	1160	object
Variável 34	259	object
Variável 35	16	object
Variável 36	1350	object
Variável 37	1350	object
Variável 38	9769	object
Variável 39	9704	object
Variável 40	9704	object
Variável 41	9704	object

Tabela 1 - Colunas classificadas pela quantidade de valores não nulos e pelo tipo de dado.

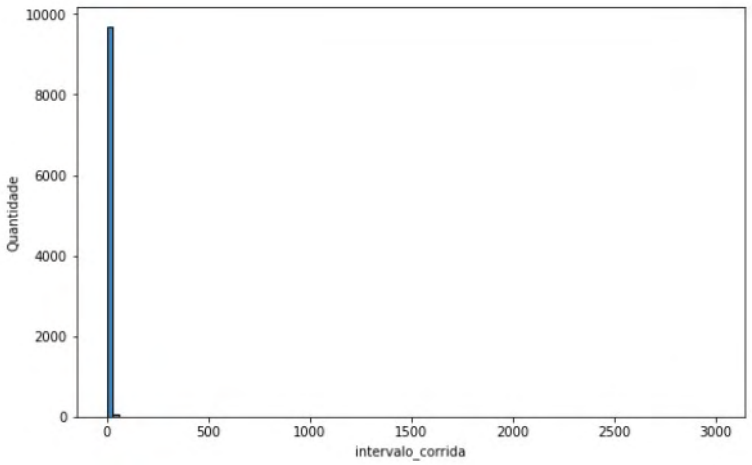
(Conclusão)

Coluna	Quantidade de não nulos	Tipo de dado
Variável 42	9769	object
Variável 43	9769	int64

Fonte: Elaborado pela Autora (2022).

A próxima etapa foi, então, visualizar esses dados em formatos de histograma para todas as colunas numéricas. Um exemplo dessa visualização está na Figura 11.

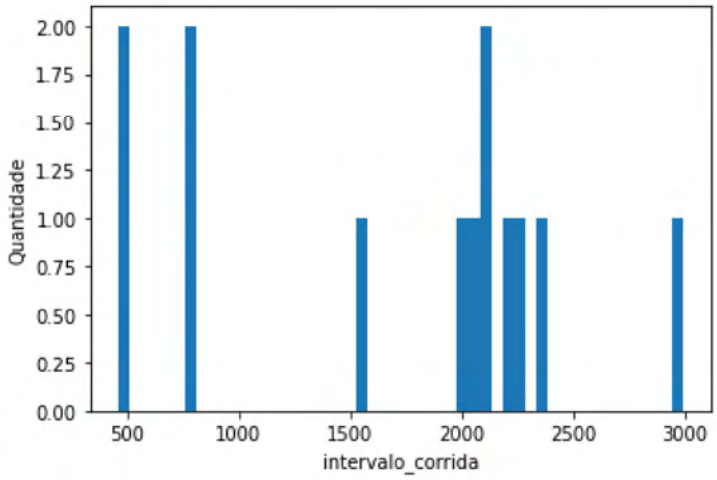
Figura 11 - Visualização gráfica do intervalo de corrida (quantidade de tempo entre uma corrida e outra)



Fonte: Elaborado pela Autora (2022).

A visualização em histogramas permite verificar se os dados estão bem distribuídos: na figura acima, por exemplo, observa-se que há um pico próximo ao valor 0 e que, ainda, ele não está centralizado, pois há valores isolados, distantes dos outros dados. Portanto, foi realizado um "zoom" e verificou-se que havia somente 13 valores de intervalo de corrida acima de 100, conforme mostra a Figura 12.

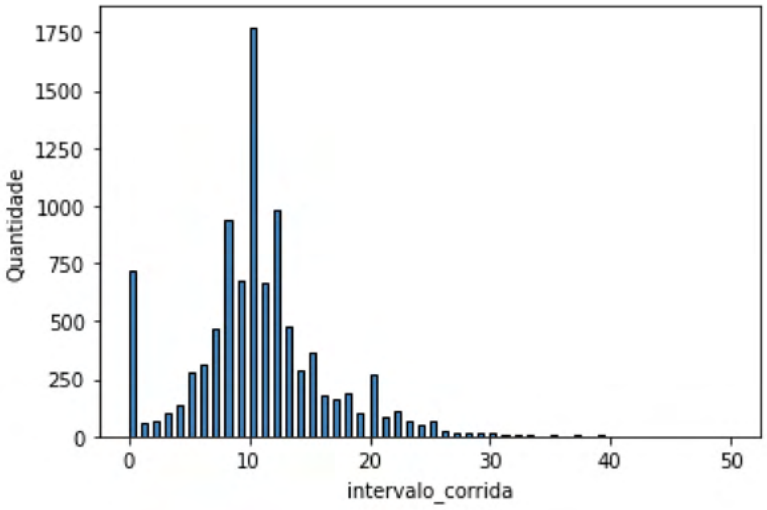
Figura 12 - Histograma com valores de intervalo de corrida acima de 100.



Fonte: Elaborado pela Autora (2022).

Esses 13 valores, representando apenas 0,13% dos dados desta coluna, foram então apagados para obter um comportamento mais bem distribuído, resultando no histograma da Figura 13. Além disso, ao apagar esses dados que estavam fora do intervalo de distribuição médio, a estatística desse atributo foi diretamente modificada, como mostra a Tabela 2: o desvio padrão diminuiu de 68,3 a 5,7.

Figura 13 - Histograma para o intervalo de corrida após apagar dados fora da média de valores.



Fonte: Elaborado pela Autora (2022).

Tabela 2 - Dados estatísticos antes e após apagar valores baseando-se no histograma.

	Antes da modificação	*Após a modificação*
Quantidade de não nulos	9769	9756
Valor máximo	2993	50
Valor mínimo	0	0
Valor médio	12,8	10,6
Desvio padrão	68,3	5,7

Fonte: Elaborado pela Autora (2022).

O mesmo tratamento foi realizado para todas as outras colunas numéricas, observando o histograma e os dados estatísticos, com o objetivo de obter atributos mais bem distribuídos. Entretanto, para realizar este tipo de análise, outro parâmetro teve que ser levado em conta: a veracidade em relação à realidade.

Um exemplo é a coluna de massa final, que apresentava valores iguais a 0. Estes valores foram deletados: não é verossímil, após uma produção de ferro gusa, que a massa deste seja zero. Uma possível explicação é erro humano de digitação manual. Por outro lado, na coluna referente ao intervalo de corrida, os valores iguais a 0 são esperados e explicáveis: estas são corridas que ocorrem uma seguida da outra.

Ainda, todos valores nulos precisaram ser deletados ou substituídos para serem utilizados nos algoritmos de *machine learning*. Neste caso, tratando-se de dados referentes à realidade, não podendo simplesmente supor valores, todas as linhas contendo valores nulos precisaram ser apagadas.

Então, para não perder uma grande quantidade de dados, considerando que há colunas com poucos dados (por exemplo, a variável 35 com somente 16 dados não nulos, como mostrou a Tabela 1), estas foram deletadas.

Então, foi possível obter uma base de dados mais condizente com a realidade, e com o máximo de dados possíveis, incluindo tipos numéricos e textuais. Por último, foi realizado um filtro deletando os valores textuais e binários, que não poderiam ser relacionados numericamente dentro dos algoritmos de *machine learning*. A base de dados final está representada na Tabela 3 - 7196 dados e 16 atributos.

Tabela 3 - Colunas classificadas pela quantidade de valores não nulos e pelo tipo de dados, após tratamento preliminar.

Coluna	Quantidade de não nulos	Tipo de dado
Variável 1	7196	int64
Variável 2	7196	float64
Variável 3	7196	float64
Variável 4	7196	float64
Variável 5	7196	float64
Variável 6	7196	float64
Variável 7	7196	float64
Variável 8	7196	float64
Variável 9	7196	float64
Variável 10	7196	float64
Variável 11	7196	int64
Variável 12	7196	float64
Variável 13	7196	float64
Variável 14	7196	float64
Variável 15	7196	float64
Variável 16	7196	float64

Fonte: Elaborado pela Autora (2022).

4 RESULTADOS E DISCUSSÃO

Após o pré-tratamento de dados, a coluna de teor de silício no ferro-gusa obtida apresentou as características mostradas na Tabela 4.

Tabela 4 - Dados estatísticos sobre a quantidade de silício presente no refratário.

Medida estatística	**Quantidade**
Elementos não nulos	7196
Valor máximo	1,80
Valor mínimo	0,00
Valor médio	0,50
Desvio padrão	0,20

Fonte: Elaborado pela Autora (2022).

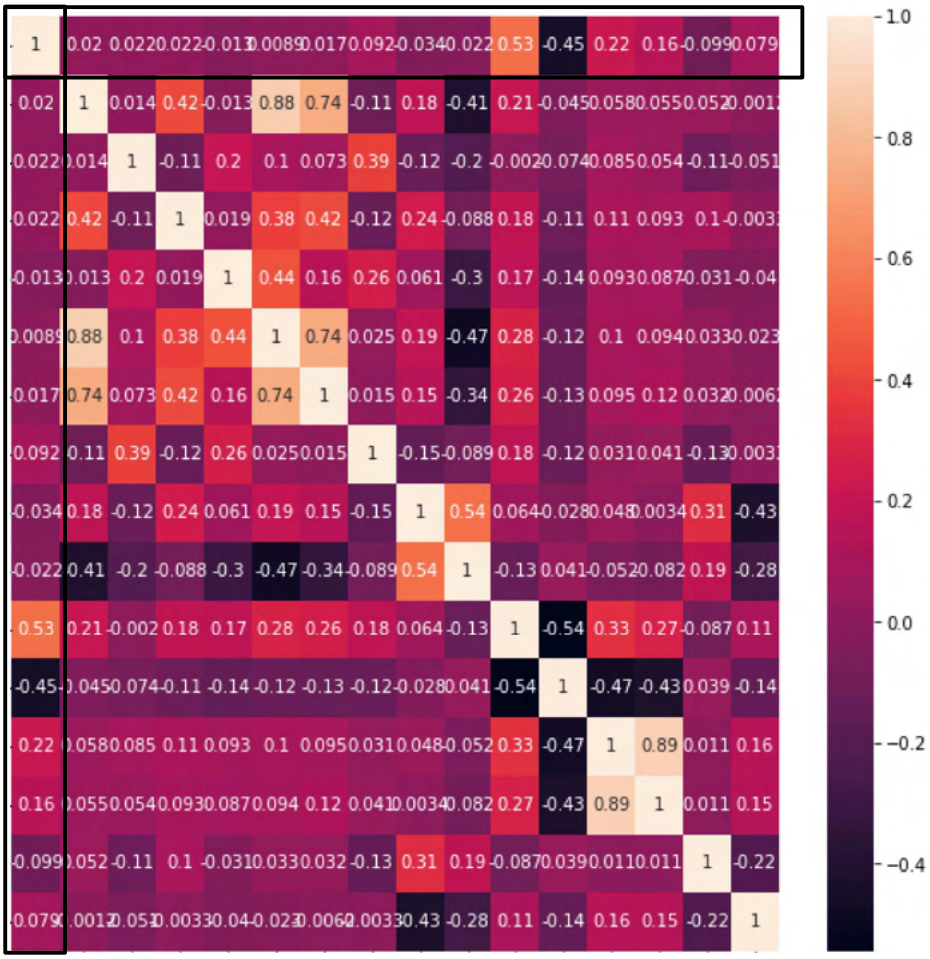
Uma análise preliminar foi realizada para verificação de relação linear entre as variáveis. Para tal, utilizou-se uma matriz de correlação. Em uma análise de correlação de Pearson, variáveis são analisadas em pares, verificando sua relação linear. Então, o cálculo para duas variáveis x e y é feito conforme a Equação (6)[19].

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \quad \textbf{(6)}$$

em que $\bar{x}$ é a média de valores da variável x, e $\bar{y}$ a média para a variável y. Ainda, s_x e s_y são, respectivamente, os desvios padrão para x e y. O valor de ρ pode variar entre -1 (inversamente proporcionais) até +1 (diretamente proporcionais), passando por 0 (sem relação linear entre as variáveis)[19].

Realizando uma matriz de correlações utilizando Python, foi possível obter a Figura 14. Nela, observa-se que a diagonal é igual a 1, ou seja, de linearidade perfeita, visto que a variável é proporcional a ela mesma[19]. A primeira linha e primeira coluna da figura abaixo são da variável estudada neste trabalho, a quantidade de silício. Pode-se observar que ela não possui uma relação linear muito forte com nenhuma outra variável, sendo a maior relação 0,53, com a temperatura do ferro gusa, e -0,45, com a quantidade de enxofre no gusa.

Figura 14 - Matriz de correlações entre as variáveis numéricas da base de dados



Fonte: Elaborado pela Autora (2022).

Portanto, após verificar que as variáveis não possuem correlação linear entre os pares, optou-se por utilizar diferentes tipos de aprendizagem de máquina, todos não-lineares. Então, os métodos utilizados de aprendizado de máquina foram: árvores de decisão e redes neurais artificiais.

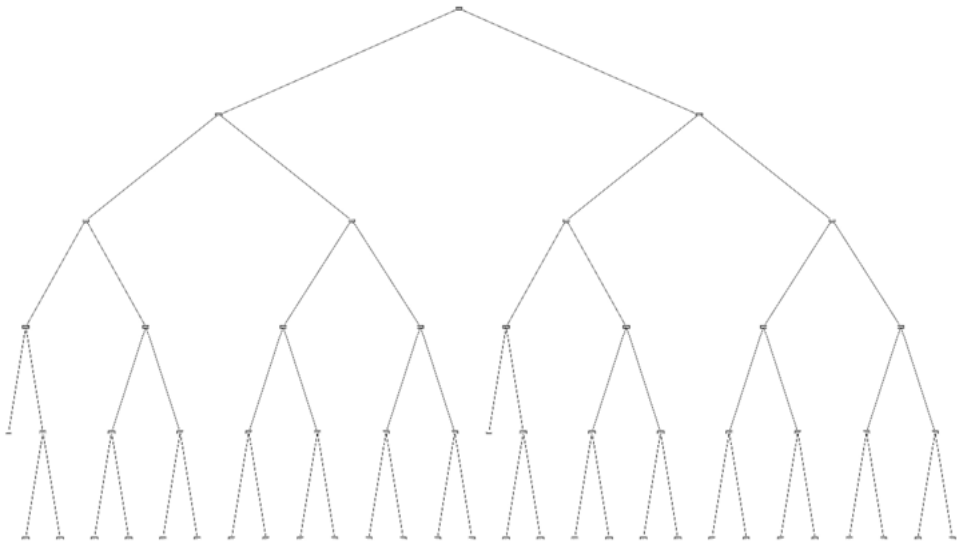
4.1 Árvores de decisão

Para reduzir o consumo de memória e diminuir a complexidade do modelo, alguns parâmetros devem ser definidos anteriormente, principalmente aqueles ligados ao tamanho das árvores, como o "*max_depth*". Esse parâmetro determina a profundidade máxima da árvore, ou seja, a quantidade de níveis[20]. Então, para encontrar esse valor automaticamente, foi utilizada a biblioteca Sklearn em Python, com o método *GridSearchCV*, que otimiza esse parâmetro dentro

de um intervalo fornecido. Para um intervalo de "*max_depth*" fornecido entre 1 e 20, o valor ótimo encontrado foi de 5 níveis.

Ainda, dois outros parâmetros foram encontrados pelo *GridSearchCV*: "*min_samples_leaf"* e "*min_samples_split"*, resultando em 3 e 9, respectivamente. O primeiro parâmetro determina o número mínimo de amostras presentes em um nó final da árvore, sem nós-filhos. Enquanto isso, o "*min_samples_split*" é o número mínimo de amostras para dividir um nó interno, que terá nós-filhos[20]. Um esquema da árvore de decisão optimizada está na Figura 15.

Figura 15 - Árvore de decisão gerada após optimização de parâmetros.



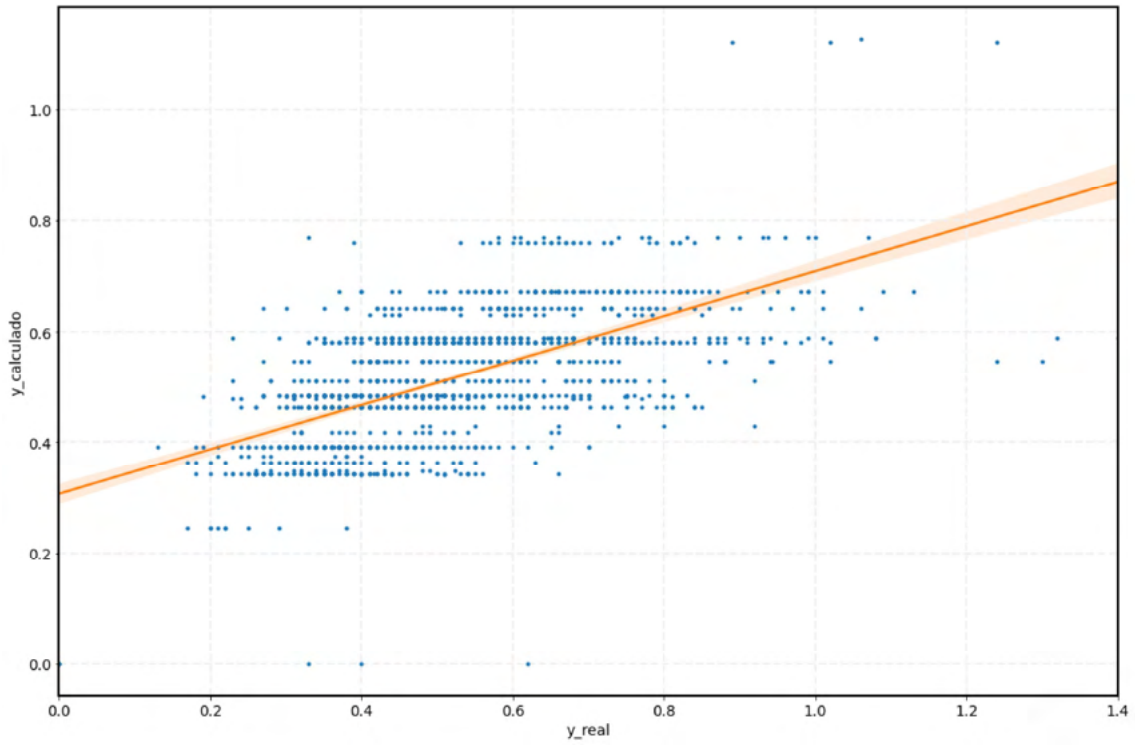
Fonte: Elaborado pela Autora (2022).

Então, o valor de MSE para este caso foi de 0,0205 e de MAPE de 22,3%, separando 80% dos valores da base de dados para treinamento (5756 valores) e 20% para teste (1440 valores). O gráfico da Figura 16 mostra a relação entre os valores de silício reais "y_real" e os calculados pelo modelo no conjunto de validação, "y_calculado".

Ainda nesta figura, foi realizada uma regressão linear. Quanto mais próxima a reta laranja fica da diagonal, mais próximos os valores reais e calculados estão. O valor de R^2 foi de 0,363 e a Equação (7) mostra os valores de coeficiente linear e angular para a reta.

$$y_{calculado} = 0{,}403 * y_{real} + 0{,}306 \quad \textbf{(7)}$$

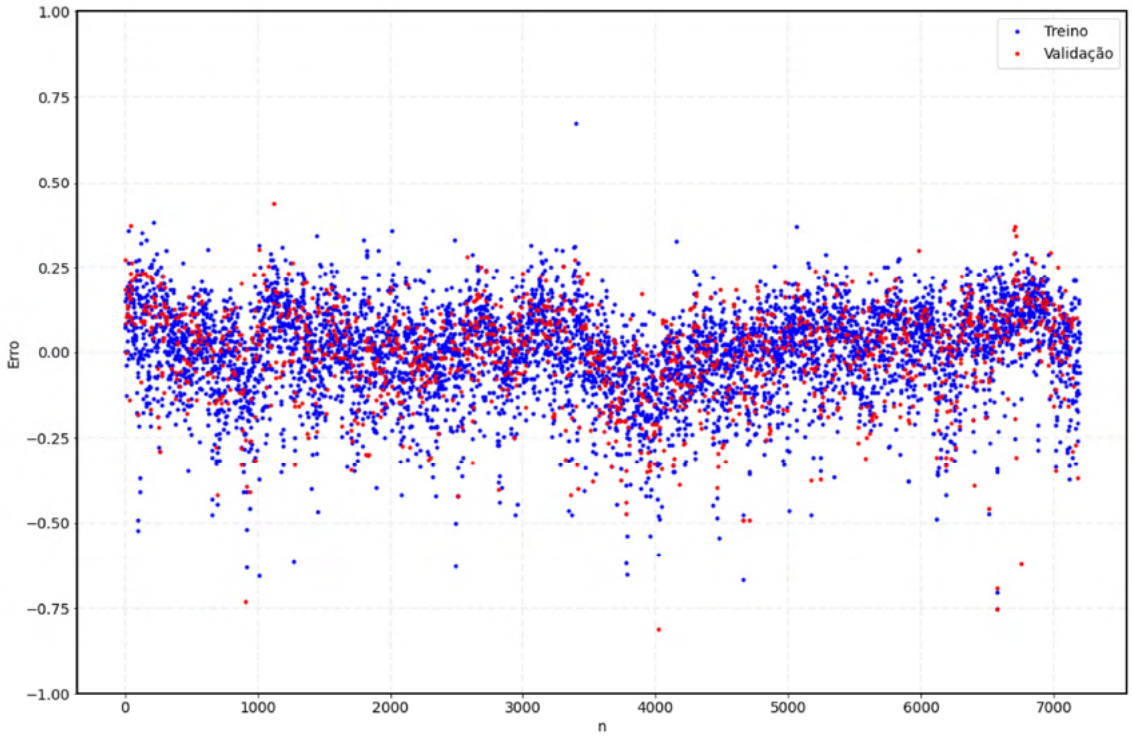
Figura 16 - Valores preditos e reais no mesmo gráfico para Árvore de Decisão (conjunto de validação), assim como a regressão linear.



Fonte: Elaborado pela Autora (2022).

Para todos os dados utilizados em treinamento e em validação, há um erro associado, calculado pela diferença entre o valor experimental e o valor predito. O resultado está presente na Figura 17. Observa-se que o MSE parece pequeno, porém, ao comparar com os dados estatísticos de composição de silício, os valores também são baixos: a média de silício é 0,50 e seu valor máximo é de 1,80. Então, analisando o MAPE conjuntamente com o MSE, percebe-se que há uma variação de 22% em relação ao valor real.

Figura 17 - Erros para conjuntos de treinamento e de validação para árvore de decisão.



Fonte: Elaborado pela Autora (2022).

4.2 Redes neurais artificiais

Diferentes parâmetros foram testados para a construção das redes neurais no Matlab, buscando o valor ótimo (ou seja, aquele de maior coeficiente de determinação R^2 na regressão linear e de menor erro, fornecido pelo histograma de erros) para os três conjuntos de teste, validação e treino.

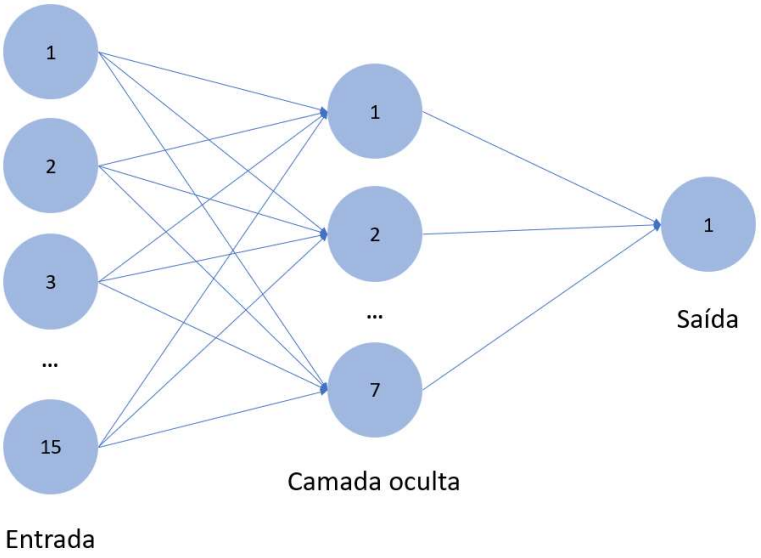
A separação dos dados para treinamento, validação e teste foi feito de tal forma que havia 5038 dos dados para o primeiro caso (70% do total de 7196 valores), 1079 para o segundo (15%) e 1079 para o terceiro (15%). O primeiro parâmetro foi o modelo, testado entre retro propagação em gradiente descendente e modelo de Levenberg-Marquardt, sendo que este último apresentou melhores resultados.

Outro parâmetro foi a função de ativação da primeira camada intermediária, variada em tangente sigmoide e logaritmo sigmoide. Após os testes, a função escolhida foi a logarítmica.

Por último, alterou-se o número de camadas de neurônios entre 1 e 10, obtendo o melhor valor em 7 camadas. Uma representação da rede neural está na Figura 18.

Portanto, foram realizados 10 testes para cada variação de parâmetro: com modelo de Levenberg-Marquardt logarítmico e tangencial e com modelo gradiente descendente, também logarítmico e tangencial, resultando em 40 testes, até encontrar os melhores resultados.

Figura 18 - Modelo da rede neural artificial com parâmetros otimizados.

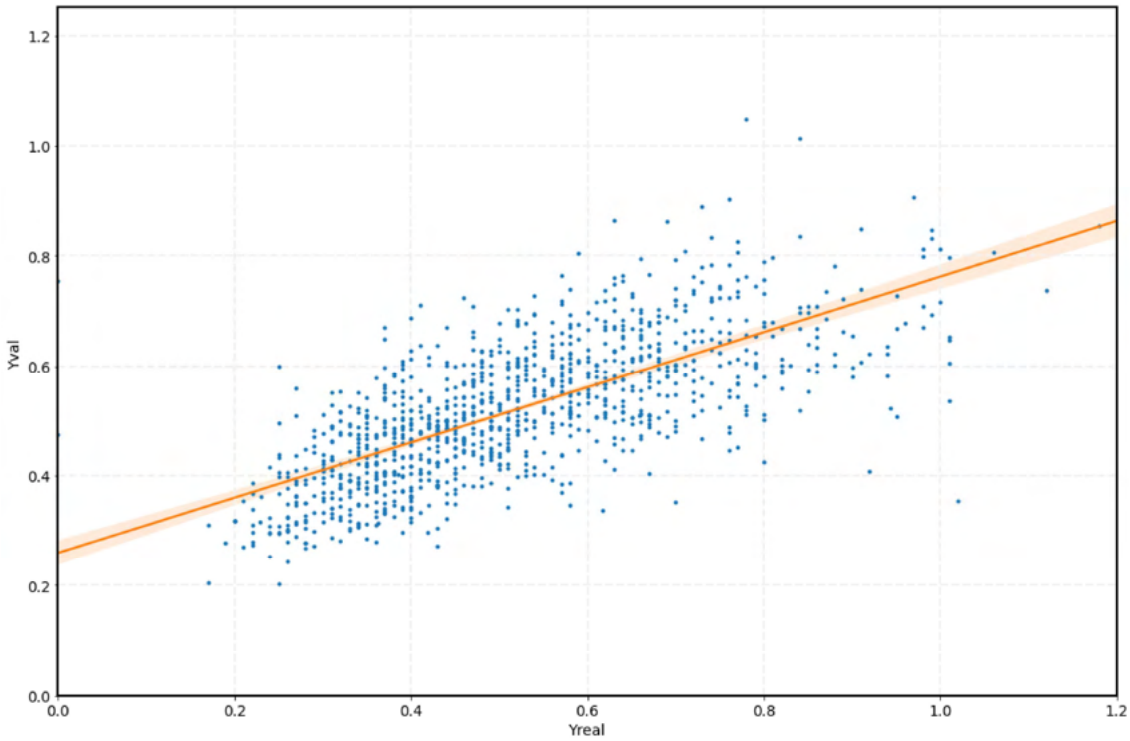


Fonte: Elaborado pela Autora (2022).

Então, para a rede neural em questão, os valores utilizados na etapa de validação estão na Figura 19. Nela, há equação da reta de regressão linear dada pela Equação (8), cujo valor de R^2 foi 0,497.

$$y_{calculado} = 0{,}503 * y_{real} + 0{,}259 \quad \textbf{(8)}$$

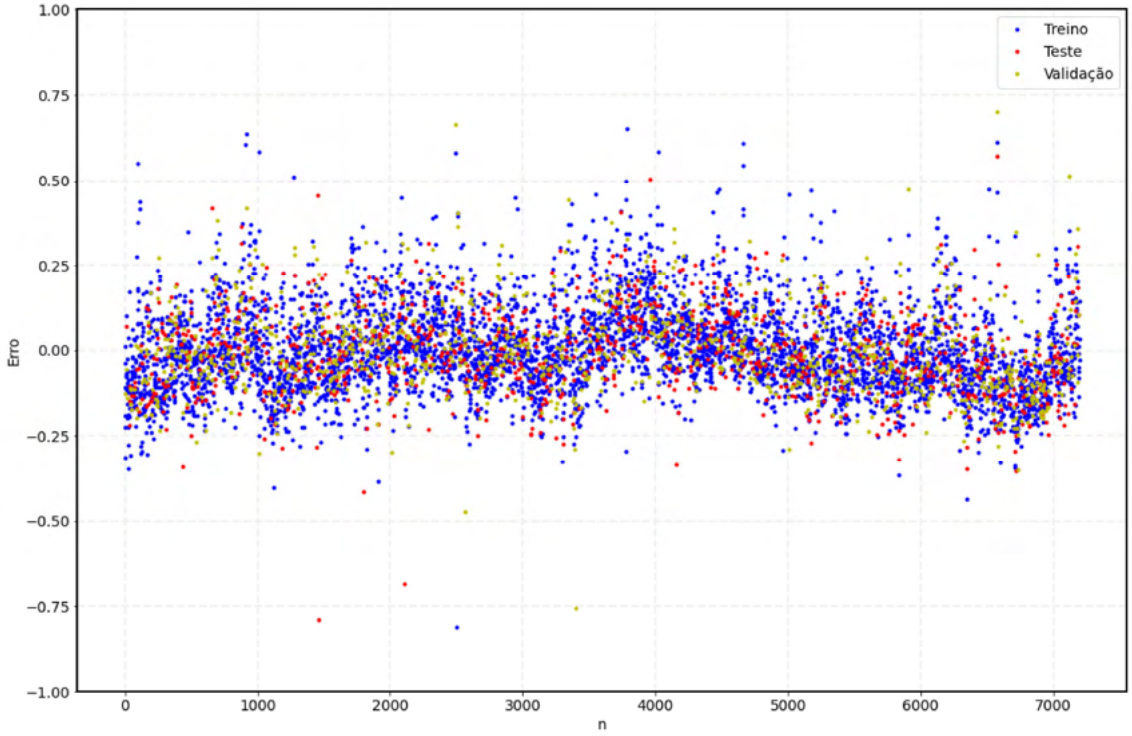
Figura 19 - Valores preditos e reais no mesmo gráfico para Redes Neurais Artificiais (conjunto de validação), assim como a regressão linear.



Fonte: Elaborado pela Autora (2022).

O valor de MSE foi 0,0163 e de MAPE foi de 21,0%. A Figura 20 representa os erros, valor calculado menos o experimental, para os três casos. Novamente, os erros variam em torno de zero, mas deve-se considerar que os valores de quantidade de silício são baixos, todos menores que 1,80.

Figura 20 - Erros para conjuntos de treinamento, teste e validação para redes neurais artificiais.



Fonte: Elaborado pela Autora (2022).

4.3 Comparação entre métodos

Ambos os modelos, árvore de decisão e rede neural, tiveram os dados divididos de tal forma a obter um conjunto de treinamento maior: 80% dos dados foram utilizados para treino na árvore e 70% na RNA. Ter mais exemplos no conjunto de treinamento ajuda o modelo a melhor estimar os valores[11].

A Tabela 5 mostra a comparação de dados entre a árvore de decisão e as redes neurais artificiais. Novamente, o MSE e o MAPE foram calculados para os 7196 pontos (todos valores da base de dados), enquanto a equação da reta e o R^2 foram obtidos para o conjunto de validação.

Observa-se que as duas formas de avaliação de erros, MSE e MAPE, foram melhores para as redes neurais: quanto menor o MSE e o MAPE, mais semelhante está o valor calculado pelo modelo do valor experimental. Entretanto, apesar de valores melhores para a RNA, os valores ainda foram próximos daqueles para a árvore de decisão: respectivamente, 21,0% para 22,3% de MAPE e de 0,0163 para 0,0205 para MSE.

Tabela 5 - Resultados para avaliação dos métodos: árvore de decisão e redes neurais.

Métrica	*Árvore de decisão*	*Redes neurais*
MSE	0,0205	0,0163
MAPE	22,3%	21,0%
Equação da reta	$y = 0{,}403 * y_{real} + 0{,}306$	$y = 0{,}503 * y_{real} + 0{,}259$
R^2	0,363	0,497

Fonte: Elaborado pela Autora (2022).

A maior diferença, entretanto, foi no coeficiente de determinação para a linearização da reta do conjunto de validação, sendo 0,363 para a árvore de decisão e 0,497 para a RNA. Estes valores de R^2, apesar de apresentarem valores baixos, devem ser analisados conforme sua realidade: a base de dados provém de produções reais de ferro-gusa e podem, por exemplo, apresentar erros e ruídos.

Ainda, se os valores calculados e reais fossem iguais, a equação da reta passaria pelo valor 0 em seu coeficiente linear e 1 em seu coeficiente angular (ou seja, $y_{calculado}=y_{real}$). As redes neurais apresentaram o coeficiente linear e angular mais próximos destes valores esperados, em relação às árvores.

Então, verificando os gráficos de erro para ambos os casos, observa-se que a grande maioria dos valores ficou entre -0,5 e +0,5. No caso da árvore de decisão, somente 21 pontos encontraram-se fora deste intervalo, e na RNA 22 pontos. Analisando o index de cada ponto, verificou-se que não se tratava dos mesmos pontos para árvore e RNA – se os pontos fossem os mesmos, os erros poderiam estar associados à base de dados e não ao modelo, ou seja, poderiam ser erros experimentais.

Portanto, as redes neurais mostraram melhor desempenho em geral, com valores menores de erros e valor maior de coeficiente de determinação. Isso pode ser explicado pela grande capacidade das RNA's de realizar um mapeamento não-linear, em bases de dados que as variáveis interagem entre si de forma complexa[21].

Ainda, há a Teoria da Aproximação Universal, que afirma que qualquer função pode ser aproximada por uma RNA contendo ao menos uma camada oculta e utilizando a função de

ativação sigmoide[21], que foi o caso constatado neste trabalho, após 40 diferentes simulações para a construção da rede.

Entretanto, graças aos valores de erros semelhantes, é discutível se é adequado utilizar a RNA, que é um modelo de aprendizado profundo (do inglês, *deep learning*), ao invés das árvores, modelo de aprendizado de máquina. A diferença é que o *deep learning* (subdivisão do *machine learning* capaz de utilizar múltiplas camadas para aprendizagem) necessita menor intervenção humana, porém, exige maior tempo para processamento e mais capacidade de hardware que o *machine learning*[22,23].

5 CONCLUSÃO

Após uma análise preliminar, utilizando a matriz de correlações, para verificar se os valores da base de dados eram linearmente proporcionais entre si, conclui-se que não havia uma relação linear, sendo então necessário recorrer a métodos não-lineares. Assim, foram realizadas simulações utilizando aprendizado de máquina: árvore de decisão e rede neural artificial.

Diante dos valores apresentados para os erros MSE e MAPE, a rede neural e a árvore de decisão apresentaram resultados semelhantes. Entretanto, ao comparar o valor do coeficiente de determinação, a rede neural apresentou melhor desempenho: respectivamente, 0,497 e 0,363 para RNA e árvore de decisão.

Esperava-se que o modelo de RNA apresentasse melhores resultados graças à Teoria de Aproximação Universal, e por ser um modelo pouco sensível a falhas e à falta de dados. Enquanto isso, mudanças nos dados de entrada das árvores de decisão podem levar a grandes alterações nos resultados.

Portanto, simulações realizadas utilizando *machine learning* podem ajudar a solucionar problemas ligados à quantidade de silício em ferro-gusa, identificando padrões presentes nos dados para possivelmente levar a testes laboratoriais. Entretanto, pode-se avaliar o uso das técnicas de *machine learning*, verificando se de fato é necessário utilizar um método de *deep learning*, o qual pode levar a valores semelhantes apesar de demorar mais tempo.

O valor de coeficiente de determinação não foi suficientemente alto em nenhum dos casos para considerar os resultados como excelentes. O ideal seria continuar a coleta de dados, obtendo mais valores, para então verificar se os modelos apresentariam melhorarias.

Uma próxima etapa poderia ser testar um diferente algoritmo de *machine learning*: as árvores aleatórias (do inglês, *Random Trees*), que são um aprofundamento das árvores de decisão, trazendo um conjunto destas. Além disso, podem-se testar diferentes parâmetros nas redes neurais artificiais, alterando tanto o algoritmo, quanto a função de ativação e o número de camadas.

Ainda assim, o desenvolvimento deste tipo de estudo ajuda no processo produtivo, tendendo a melhorar a eficiência das operações na siderurgia e a reduzir, por exemplo, o número de paradas para manutenção e o gasto energético em altos-fornos, a partir do monitoramento da quantidade de silício em ferro-gusa.

6 REFERÊNCIAS

1 MOURÃO, M. B. et al. **Introdução à siderurgia**. São Paulo: Associação Brasileira de Metalurgia e Materiais, 2007.

2 FONTES, D. O. L. **Desenvolvimento de soft sensor para predição do estado térmico do ferro gusa em alto-forno usando fuzzy c-médias e modelo exógeno auto-regressivo não linear.** Dissertação (Mestrado em Engenharia Química) - Centro de Ciências e Tecnologia, Universidade Federal de Campina Grande. Campina Grande, 2020. Disponível em: <http://dspace.sti.ufcg.edu.br:8080/jspui/bitstream/riufcg/15281/1/DIANE%20OTÍLIA%20LIMA%20FONTES%20–%20DISSERTAÇÃO%20%28PPGEQ%29%202020.pdf>. Acesso em 26 jun. 2022.

3 DA FONTE, L. **Panorama nacional da indústria do ferro e aço.** Dissertação (Mestrado em Geociencias) - Instituto de Geociencias, Universidade Estadual de Campinas. Campinas, 2003.

4 PAULA, G. M. **Economia de baixo carbono: avaliação de impactos de restrições e perspectivas tecnológicas**. Produção Independente de Ferro-Gusa ("Guseiros"). Ribeirão Preto, 2014. Disponível em: <http://www.comexresponde.gov.br/portalmdic/arquivos/dwnl_1423738671.pdf>. Acesso em 14 mar. 2022.

5 DINIZ, A. P. M. **Modelos de previsão do conteúdo de silício no ferro gusa usando redes neurais artificiais.** Dissertação (Mestrado em Engenharia Elétrica) – Centro Tecnológico, Universidade Federal do Espírito Santo. Vitória, 2018. Disponível em: <https://repositorio.ufes.br/bitstream/10/9572/1/tese_10911_DISSERTAÇÃO%20-%20ANA%20PAULA%20MIRANDA%20DINIZ.pdf>. Acesso em 26 jun. 2022.

6 DINIZ, A. P. M. et al. Previsão do teor de silício no ferro-gusa utilizando redes neurais artificiais e algoritmo de poda. **Sociedade Brasileira de Automática**: v. 1, n. 1, 2019. Disponível em: <https://www.sba.org.br/open_journal_systems/index.php/cba/article/view/431/394>. Acesso em 24 jun. 2022.

7 FERREIRA, R. G. C. et al. **Preparação e Análise Exploratória de Dados**. Porto Alegre: Sagah, 2021. p. 13-21. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9786556902890/pageid/21>. Acesso em 12 mar. 2022.

8 Machine Learning. **IBM Cloud Education.** 15 de jul. de 2022. Disponível em: <https://www.ibm.com/cloud/learn/machine-learning>. Acesso em 07 abr. 2022.

9 KINGSFORD C., SALZBERG S. L. What are decision trees? **Nature Biotechnology**, v. 26, p. 1011–1013, set. 2008. Disponível em: <https://www.nature.com/articles/nbt0908-1011.pdf>. Acesso em 18 abr. 2022.

10 Decision Trees. **Scikit-Learn**. Disponível em: <https://scikit-learn.org/stable/modules/tree.html#tree>. Acesso em 23 mai. 2022.

11 PEREIRA F., MITCHELL T., BOTVINICK M.. Machine learning classifiers and fMRI: A tutorial overview. **NeuroImage**, v. 45, n. 1, p. S199-S209, 2009. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1053811908012263>. Acesso em 22 mai. 2022.

12 SILVA, L. N. C. **Redes neurais artificiais**. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de

Campinas. Campinas, 1998. Disponível em: <https://www.dca.fee.unicamp.br/~vonzuben/theses/lnunes_mest/cap2.pdf> Acesso em 21 mai. 2022.

13 MINUSSI, C. **Redes neurais – introdução e principais conceitos**. Universidade Estadual Paulista, 2008. Disponível em: <https://www.feis.unesp.br/Home/departamentos/engenhariaeletrica/pos-graduacao/apostila-redes-neurais-anna-diva_minussi.pdf>. Acesso em 25 abr. 2022.

14 MARTINIANO, A. et al. **Utilizando uma rede neural artificial para aproximação da função de evolução do sistema de Lorentz.** Revista Produção e Desenvolvimento, v.2, n.1, p.26-38, jan./abr., 2016.

15 HAGAN, M. T.; MENHAJ, M. B. Training Feedforward Networks with the Marquardt Algorithm. **IEEE Transactions on Neural Networks**, vol. 5, no. 6, 1994.

16 FORRER, M. **Prediction of refractory wear with Machine Learning methods**. Graz University of Technology, 2012.

17 MARTINS, M. S. **Inovações tecnológicas da indústria 4.0: aplicações e implicações para a siderurgia brasileira**. Dissertação (Mestrado em Econômia) - Instituto de Economia e Relações Internacionais, Universidade Federal de Uberlândia, Uberlândia, 2019. Disponível em: <http://repositorio.ufu.br/bitstream/123456789/25113/1/InovaçõesTecnológicasIndústria.pdf>. Acesso em 22 mai. 2022.

18 CHICCO D., WARRENS M. J., JURMAN G. **The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation**. PeerJ Computer Science, 2021. Disponível em: < https://doi.org/10.7717/peerj-cs.623>. Acesso em 21 jun. 2022.

19 Métodos e fórmulas para Correlação. **Suporte ao Minitab©**. Disponível em: <https://support.minitab.com/pt-br/minitab/19/help-and-how-to/statistics/basic-statistics/how-to/correlation/methods-and-formulas/methods-and-formulas/#:~:text=par%20de%20variáveis.-,Fórmula,n%20–%202%20graus%20de%20liberdade.> Acesso em 02 mai. 2022.

20 Sklearn.tree.DecisionTreeRegressor. **Scikit-Learn**. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html>. Acesso em 06 mai. 2022

21 IGNÁCIO, L. V. R. Et al. O uso de inteligência artificial para a previsão do preço do petróleo. **Espacios**, v. 38, n. 24, 2017. Disponível em: <http://ww.revistaespacios.com/a17v38n24/a17v38n24p03.pdf>. Acesso em 22 mai. 2022.

22 LECUN Y., BENGIO Y., HINTON G. Deep learning. **Nature**, v. 521, p. 436–444, 2015. Disponível em: <https://www.nature.com/articles/nature14539>. Acesso em 31 mai. 2022.

23 Simplifying the difference: machine learning vs deep learning. **Singapore Computer Society**, 2020. Disponível em: <https://www.scs.org.sg/articles/machine-learning-vs-deep-learning>. Acesso em 31 mai. 2022.