

Previsão do consumo de cerveja utilizando redes neurais

Luiz Augusto da Silva

IFMG - Campus Bambuí
Bambuí - MG
lluiz.augusto409@gmail.com

Sandy Cristina Barros

IFMG - Campus Bambuí
Bambuí - MG
sandy.cristina211@gmail.com

Resumo—Este artigo apresenta um estudo sobre a previsão do consumo de cerveja utilizando uma arquitetura de rede neural profunda. Inicialmente, é realizada uma análise exploratória dos dados para identificar as variáveis relevantes para a predição. Em seguida, é proposta e implementada a rede neural profunda, utilizando a biblioteca *TensorFlow* em *Python*.

A rede neural é treinada e avaliada com base em um conjunto de dados históricos, considerando diversas características, como temperatura, precipitação, feriados, entre outras. Os resultados são analisados e discutidos, destacando a precisão da previsão e o potencial impacto dessa abordagem no setor cervejeiro. Concluindo assim que as redes neurais profundas são ferramentas promissoras para a previsão do consumo de cerveja, oferecendo *insights* valiosos para a indústria e possibilitando uma melhor gestão da produção e distribuição.

Palavras chaves—cerveja, redes neurais, consumo

I. INTRODUÇÃO

A previsão do consumo de cerveja é um tópico relevante e de interesse em diversos setores, como a indústria de bebidas e a economia nacional. Com o avanço das técnicas de aprendizado de máquina, a aplicação de modelos de redes neurais profundas DNN (*Deep Neural Networks*) tem se mostrado promissora na análise e previsão de diferentes tipos de dados [1].

Neste artigo, é proposto o uso de uma DNN para prever o consumo de cerveja com base em variáveis relevantes, como temperatura, precipitação, dia da semana e feriados. A previsão precisa do consumo de cerveja é de grande importância para as empresas do setor, permitindo um planejamento mais eficiente da produção e distribuição, além de contribuir para a tomada de decisões estratégicas.

Para isso, será utilizado a biblioteca *TensorFlow* em *Python*, que oferece uma ampla gama de ferramentas e recursos para a construção e treinamento de redes neurais profundas. Também será feito uso de uma arquitetura de rede neural adequada, com camadas densas e função de ativação apropriada, a fim de capturar as complexidades e relações entre as variáveis de entrada.

Ao longo deste artigo, será apresentado o processo de preparação dos dados, a construção da DNN, o treinamento do modelo e a avaliação da sua performance. Além disso, será discutido os resultados obtidos, destacando a eficácia da abordagem proposta na previsão do consumo de cerveja.

II. CARACTERIZAÇÃO DO PROBLEMA

O problema em questão consiste na previsão do consumo de cerveja com a utilização de redes neurais, o objetivo é usar um modelo DNN para prever o consumo de cerveja a partir de um *dataset*, buscando compreender o desempenho que um modelo de rede neural profunda tem para prever o consumo de cerveja baseado em fatores que influenciam no consumo da bebida.

Prever acontecimentos futuros é de interesse nas organizações, pois possibilita tomadas de decisões embasadas nas demandas, comportamentos e fatores que responderam diretamente ao consumo de um determinado produto. Essas informações podem ser utilizadas para ajustar o planejamento e o controle da produção para atender de maneira mais precisa às necessidades do mercado. Por meio da antecipação, é possível otimizar os processos de logística, produção e estoque, proporcionando uma gestão mais eficiente. [2]

Ainda de acordo com [2] os erros de previsão tem importância já que em previsões muito discrepantes de valores reais, há um risco, e identificar esses erros é crucial para minimizar os impactos negativos nos processos de produção.

A previsão no mercado de bebidas com a utilização de redes neurais pode ser importante, pois são ferramentas que permitem reconhecer padrões que influenciam no consumo de cerveja, de acordo com [3] esses fatores podem ser sensoriais que envolvem o paladar, visão e que isso molda como a cerveja é consumida. Ainda consoante a [4] esses fatores podem ser sazonais, ou seja, estão relacionados a fatores climáticos e culturais ao longo do ano, outro fator de interesse no consumo de cerveja é se o dia é dia letivo ou se é final de semana, de fato, em um estudo foi constatado que a cerveja é mais consumida nos finais de semana que nos dias da semana [5]. Outro estudo aponta que a produção de cerveja se intensifica entre outubro a janeiro, devido ao verão [6].

Portanto, com o uso de redes neurais essas tendências podem ser reconhecidas e identificadas por um modelo de aprendizado que permita prever o consumo a partir de dados históricos e auxiliar para que a indústria cervejeira se prepare adequadamente para atender à demanda dos consumidores e isso é de extrema relevância já que a cerveja é a terceira bebida mais consumida do mundo de acordo com [7].

III. TRABALHOS RELACIONADOS

Um trabalho de interesse é o [6] que apresentou comparações entre modelos estatísticos e redes neurais para prever a produção semanal de uma cervejaria. Neste estudo, foi abordada a capacidade das redes neurais em relação a outros métodos estatísticos para prever e aprender padrões de interesse para a indústria de cerveja utilizando a arquitetura MLP(*multi layer perceptron*).

O artigo de [8] apresenta redes neurais profundas no desenvolvimento para gerenciar o consumo de carga elétrica a curto Prazo que permite entender melhor como as DNN's podem ser aplicadas para a previsão que envolva o consumo e o aprendizado profundo e como prever fatores de influência no consumo de maneira eficiente.

Outro trabalho de interesse é o de [9], pois apresenta o estudo voltado para o consumo de energia, o autor buscou apresentar comparações entre soluções de aprendizado gerais e tradicionais e o aprendizado profundo.

Por fim, Oliveira [10] apresenta modelos de redes neurais profundas usando *Tensorflow* e *Python* na previsão do consumo de energia, buscando avaliar e calibrar as redes para trabalhar em cenários de previsão com eficiência.

IV. FUNDAMENTAÇÃO TEÓRICA

Nos próximos capítulos será apresentado o método de previsão que será utilizado para prever o consumo de cerveja, o algoritmo de otimização bem como as principais técnicas utilizadas para o aprendizado e bom desempenho de uma rede neural.

A. Redes Neurais

Redes neurais são modelos computacionais inspirados no cérebro humano que permitem o reconhecimento de padrões. São compostas por camadas e neurônios interconectados que formam arquiteturas que processam os valores de entradas aplicados a uma função de ativação que permite que a arquitetura aprenda comportamentos e características e obtenha uma representação significativa na saída [11].

De acordo com [12] dentre as vastas aplicações das redes neurais está a classificação, previsão, otimização dentre outras. As redes neurais são indicadas para a previsão, pois são modelos que podem ser utilizados tanto para prever a longo ou a curto prazo, geralmente as previsões são feitas a partir de valores históricos. Outro fator importante é que em determinados casos as redes neurais se saem melhor comparada a modelos estatísticos convencionais [12].

O treinamento de uma rede neural, consiste em encontrar os pesos perfeitos para um bom funcionamento e o processo de aprendizado consiste em ser: supervisionado isto é, para cada amostra dos sinais de entrada se tem a respectiva saída (resposta) desejada, por reforço em que a rede aprende por interação com um ambiente e aprendizado não-supervisionado que consiste no aprendizado em que a rede aprende a partir de dados de entrada sem a que a saída tenha uma resposta desejada [12].

A DNN(*Deep Neural Network*) é uma arquitetura de rede neural profunda utilizada para resolver problemas não lineares e é composta por entradas, camadas ocultas e saídas [1]. O aprendizado profundo tem se tornado eficiente devido à capacidade dessas redes de aprender representações complexas nos dados [13].

De acordo com [8] a DNN é de aprendizado supervisionado já que a partir de um conjunto de entradas e de pesos aleatórios definidos e uma saída esperada, a rede consegue ajustar os pesos, minimizando o erro produzido entre a saída esperada e a saída obtida.

Geralmente redes neurais profundas, utilizam funções de ativação como, sigmoide, tangente hiperbólica(tanh) ou ReLU que permitem gerar uma saída uma vez que dita como os neurônios vão responder às entradas ponderadas[8].

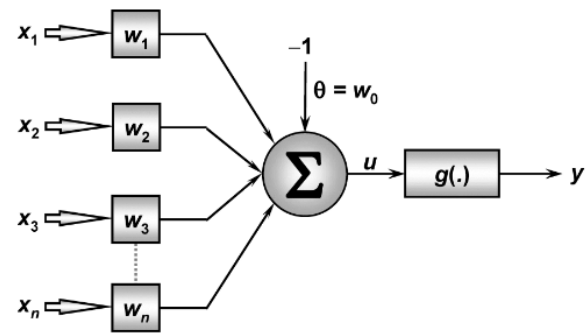


Figura 1. Arquitetura de rede neural simples [11]

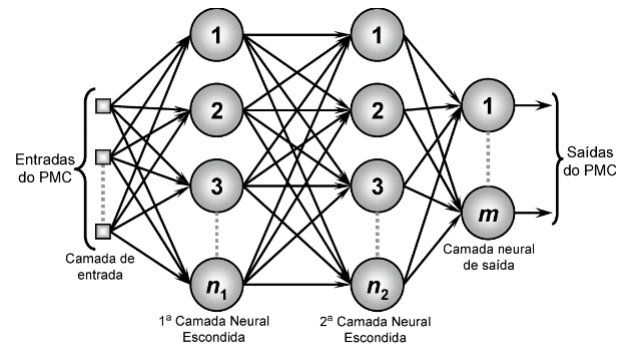


Figura 2. Arquitetura MLP[11]

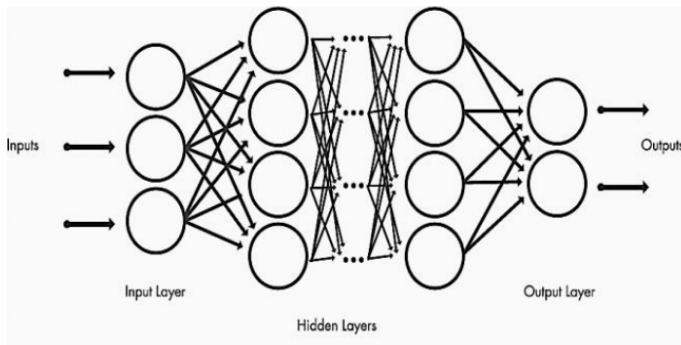


Figura 3. Arquitetura DNN

[8]

A diferença das arquiteturas apresentadas nas figuras 1, 2 e 3 é que as redes neurais simples (figura 1) são compostas por uma camada única de neurônio [11], enquanto as DNN permitem aumentar a quantidade de camadas ocultas da rede, capacitando que a rede aprenda problemas mais complexos nos dados enquanto as informações são processadas pela rede [1]. Ainda para problemas não lineares existe a arquitetura MLP, porém é uma arquitetura com apenas duas camadas ocultas [11].

B. Função de ativação

A função de ativação é um modelo matemático usado na saída de um neurônio para dar forma a separação dos dados de entrada da rede, e geralmente ficam presentes nas camadas ocultas. Uma função amplamente utilizada para a ativação em redes neurais é a tangente hiperbólica que transforma os valores de maneira não linear, a \tanh assume 1 ou -1 e permitindo que a rede aprenda a não linearidade dos dados [8].

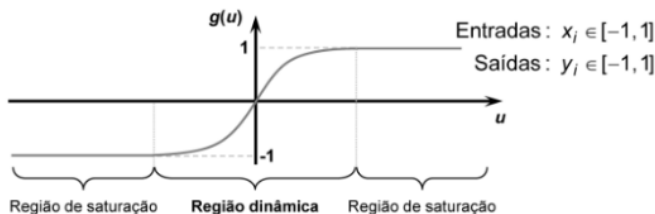


Figura 4. Tangente hiperbólica

É importante definir a função de ativação pois é essencial na construção e no treinamento de um modelo de rede neural. Porém nem sempre existe uma regra para a escolha da função correta, portanto é importante observar a saída da rede e como esta se comportando devidamente com determinada função [14].

C. Algoritmos de otimização e treinamento

Em redes neurais para o ajuste dos pesos das conexões entre os neurônios utiliza-se algoritmos de treinamento que permite encontrar os melhores valores para os pesos das redes em que

consiste numa rede treinada. É utilizado também para algoritmos de otimização que permitem o ajuste e a modificação nos parâmetros da rede [12], geralmente, os algoritmos utilizados para otimização e treinamento de redes neurais são o Adam e o *Backpropagation*, usados no treinamento de redes profundas para ajustar os pesos e minimizar o erro [15].

Mas mais devidamente voltado para o aprendizado profundo, os algoritmos de otimização mais utilizados de acordo com [13] em aprendizado profundo incluem SGD (gradiente descendente), Adam entre outros.

O Adam é um algoritmo que permite adaptar a taxa de aprendizagem para do modelo durante o treinamento e isso é possível a partir do SGD, em que o ajuste utiliza o primeiro e segundo momento do gradiente a partir da médias e variâncias para cada iteração do algoritmo para garantir uma convergência estável [16].

Algumas vantagens do Adam além de adaptar as taxas de aprendizado de forma precisa, dinâmica e automática durante o treinamento, é que o algoritmo permite que as atualizações dos parâmetros não excedam de tamanho para que não haja problemas de convergência ou instabilidade [16].

De acordo com [15], o algoritmo permite trabalhar de forma eficiente mesmo com uma quantidade menor de dados e também responde bem a quando aplicado a conjuntos de dados grandes. Ainda de acordo com [16] poder ser utilizado para desafios complexos e principalmente que envolvam o aprendizado profundo.

O *backpropagation* é um algoritmo de treinamento do tipo *feedforward* que permite o ajuste dos pesos das conexões sinápticas dos neurônios a partir do erro calculado na saída da rede de forma a minimizar esse erro. Consequentemente quanto menor o erro melhor a rede mapeará o problema [11].

Portanto, enquanto o *backpropagation* é responsável por calcular os gradientes dos pesos da rede neural em relação ao erro durante o processo de retropropagação [11] o Adam ajusta a taxa de aprendizado para cada parâmetro esperado com base nos gradientes anteriores [16].

V. METODOLOGIA

Para o desenvolvimento deste trabalho foi utilizado um notebook Acer modelo *Aspire*, com o programa VsCode configurado para trabalhar com *jupyter notebook* na linguagem de programação *Python*. As bibliotecas utilizadas foram: *Random* para inicializar parâmetros de maneira aleatória, foi usada também *pandas*, *numpy* e *tensorflow* e para a visualização e interpretação dos dados: *matplotlib*, *sklearn* e *seaborn*.

Neste capítulo, será apresentado a construção e implementação de uma modelo usando a arquitetura de rede neural profunda para a previsão do consumo de cerveja, será apresentado considerações os dados, sua natureza, tratamento e o treinamento do modelo.

A. Datasets

O dataset *Beer Consumption* - São Paulo¹ consiste em uma amostra de dados coletados em uma área universitária

¹<https://www.kaggle.com/datasets/dongearge/beer-consumption-sao-paulo>

da cidade de São Paulo, com uma faixa etária entre 18 e 27 anos. O conjunto de dados registra o consumo diário de cerveja em litros ao longo do ano de 2015. Além das informações sobre o consumo, o *dataset* também inclui dados climáticos correspondentes a cada dia, como temperatura, precipitação e um indicador se o dia é um final de semana ou não. Apesar deste *dataset* ter sido utilizado em diversos trabalhos, não foi possível encontrar mais informações sobre sua origem. Com base no *dataset* apresentado o objetivo foi prever o consumo em litros utilizando uma rede neural profunda e o algoritmo de otimização de Adam.

B. Análise dos dados

Antes de alimentar a rede neural, realizou-se uma análise dos dados que incluiu a plotagem de gráficos para melhor entender o conjunto de dados. O *Box Plot* na Figura 5 permitiu visualizar o consumo de cerveja ao longo dos dias da semana e identificar um aumento significativo nos finais de semana e também uma baixa quantidade de *outliers* e uma distribuição adequada dos dados.

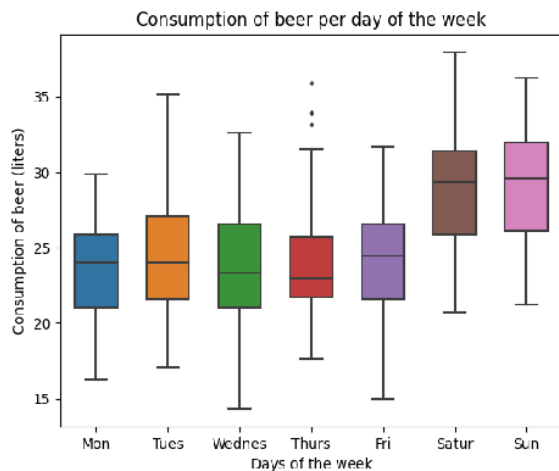


Figura 5. Box Plot consumo de cerveja

A figura 6 mostra o histograma do consumo de cerveja que apresenta que houve mais de 70 dias que o consumo de cerveja foi em média 25 litros. O histograma permitiu visualizar como os dados estão distribuídos, em que foi útil para identificar os tratamentos adequados para os dados.

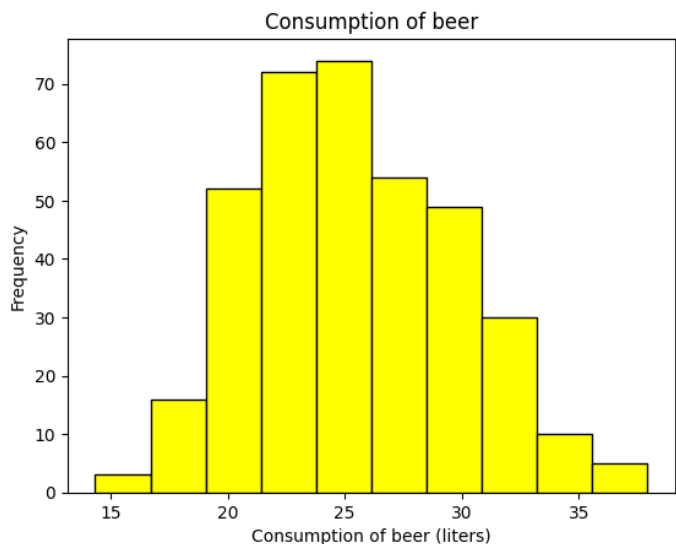


Figura 6. Histograma da media de consumo

Na figura 7, o histograma apresenta a temperatura média, máxima e mínima em graus celsius, em relação à frequência do consumo, permitindo compreender a ocorrência e a predominância de diferentes faixas de temperatura.

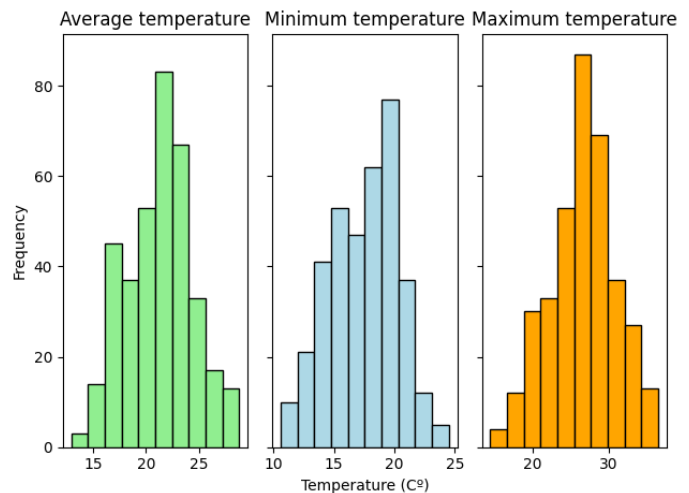


Figura 7. Histograma da temperatura média, mínima e máxima

Na figura 8, foi plotada a precipitação ao longo do ano, evidenciando a relação entre o consumo de cerveja e os períodos de chuva, representada pelos picos e quedas observadas.

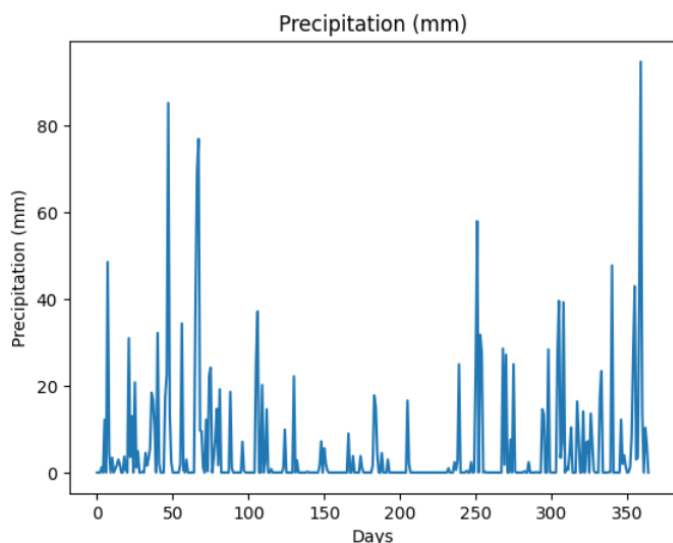


Figura 8. Gráfico da precipitação em milímetros

C. Pré processamento

A etapa de preparação dos dados envolveu a análise dos registros de consumo de cerveja, dados climáticos e os indicadores de dias da semana. Também foi aplicada técnica de padronização nos dados, garantindo a qualidade e consistência necessárias para o treinamento e a avaliação da rede neural.

Os *outliers* não foram removidos já que na análise de dados foram identificados poucos *outliers* na quinta-feira que nesse contexto esses dados podem significar pequenas variações como feriados que podem ocorrer no meio da semana.

Ainda na preparação dos dados, foram divididos 20 % dos dados para teste e 80 % dos dados para treinamento em que 10 % desses dados foram utilizados para verificar o desempenho da rede para fazer a previsão.

Foi excluído a tabela "data" do *dataset*, pois não é necessário para o aprendizado da rede e também foram substituídas as vírgulas por pontos para que o *Python* reconheça os valores decimais.

D. Construção do modelo

Foi definido a arquitetura de redes neurais profundas (DNN) para resolver o problema, pois, permite uma melhor flexibilidade para ajustar o número de camadas ocultas e a quantidade de neurônios, ou seja, uma melhor capacidade de aprendizado. Foi selecionada também pelo fato de conseguir lidar com a não linearidade dos dados, apresentado grande potencial de identificar padrões relevantes no consumo de cerveja.

A função de ativação escolhida para identificar a não linearidade do consumo foi a Tangente Hiperbólica, foi selecionado também o algoritmo de otimização Adam para ajustar a taxa de aprendizado.

Foram definidas as camadas de entrada e saída e três camadas ocultas com 16, 32 e 16 neurônios, respectivamente. Foi escolhido o algoritmo de otimização Adam, para ser

possível ajustar a taxa de aprendizagem e isso permita diminuir o erro ao longo do treinamento.

Foi definido um *DropOut* de 0,1 para evitar *overfitting* na rede. Além disso, para minimizar o erro, utilizou-se o erro quadrático médio no modelo, e estabeleceu-se um limite máximo de 1000 épocas para a convergência.

E. Processo de treinamento

A rede utilizada foi de aprendizado supervisionado e o processo de treinamento foi realizado de maneira que o algoritmo selecionasse 4 linhas do conjunto de dados por vez. Cada linha continha colunas de temperatura média, temperatura mínima, temperatura máxima, precipitação e se o dia era final de semana, a saída era composta pela quantidade consumida em litros.

Para calcular o erro da rede e analisar o quanto ela é capaz de prever, foi necessário converter os valores para a escala normal, pois até então a rede estava reconhecendo os valores da escala padronizada entre -1 e 1. Por fim, foi calculado o erro médio absoluto para verificar o erro entre as previsões do modelo gerado com os valores reais do *dataset*.

VI. RESULTADOS

A taxa de aprendizado ajustada pelo algoritmo Adam foi de 0,00001, percebeu-se que ao utilizá-lo conseguiu-se um bom desempenho na velocidade de convergência. Ao final usaram-se 782 épocas para que o modelo convergisse, ou seja, a partir da época 782 o erro não mudou.

A figura a seguir apresenta o erro em relação às épocas, e como que ele diminuiu ao decorrer do aprendizado.

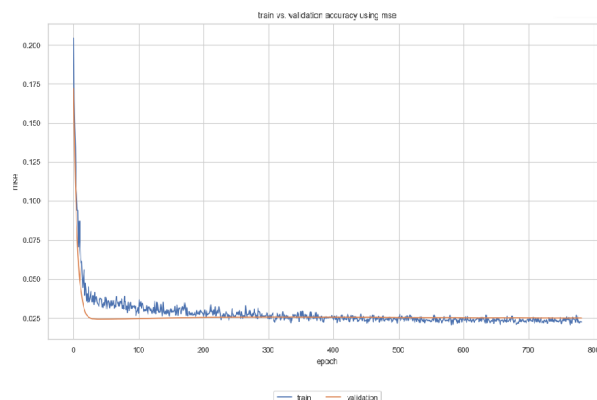


Figura 9. Gráfico de convergência

A rede apresentou um erro médio absoluto de 2,61, o que significa que suas previsões variam em até 2,61 litros para mais ou para menos em relação aos valores reais.

O resultado é satisfatório já que a taxa de erro foi pequena. Isso comprova que as redes neurais profundas é uma ferramenta viável para prever o consumo de cerveja, e que isso pode ter um impacto significativo no setor cervejeiro. Com base nesse trabalho, é possível empregar as abordagens apresentadas para previsão semanal, utilizando dados atuais para prever picos de consumo.

REFERÊNCIAS

- [1] V. Sze e et al., “Efficient processing of deep neural networks”, *Synthesis Lectures on Computer Architecture*, v. 15, n. 2, pp. 1–341, 2020.
- [2] D. F. Tubino, *Planejamento e controle da produção: teoria e prática*. São Paulo: Atlas, 2017, Acesso em: 13 jun. 2023. endereço: <https://www.uniasselvi.com.br/extranet/layout/request/trilha/materiais/livro/livro.php?codigo=29879>.
- [3] N. B. Carvalho, “Cerveja artesanal: pesquisa mercadológica e aceitabilidade sensorial”, Tese (Doutorado em Ciência e Tecnologia de Alimentos), Tese de Doutorado, Universidade Federal de Viçosa, Viçosa, 2015, p. 156.
- [4] G. M. Calôba, L. P. Calôba e E. Saliby, “Cooperação entre redes neurais artificiais e técnicas ‘clássicas’ para previsão de demanda de uma série de vendas de cerveja na Austrália”, *Pesquisa Operacional*, v. 22, pp. 345–358, 2002, Acesso em: 21 jun. 2023. endereço: <https://www.scielo.br/j/pope/a/GvmZQVNPkXxtMBBpkGsSjyH/abstract/?lang=pt#>.
- [5] L. S. Monteiro e et al., “Consumo alimentar segundo os dias da semana–Inquérito Nacional de Alimentação, 2008-2009”, *Revista de Saúde Pública*, v. 51, p. 93, 2017.
- [6] M. M. Cunha, “Previsão de séries temporais por meio de métodos estatísticos e neurais: aplicação em uma indústria de bebidas”, Trabalho de Conclusão de Curso, Universidade Tecnológica Federal do Paraná, 2019.
- [7] M. Marcelino, “A cerveja é a terceira bebida mais consumida no mundo, depois da água e do chá”, 2023, Acesso em: 17 jun. 2023. endereço: <https://empresariodigital.com.br/a-cerveja-e-a-terceira-bebida-mais-consumida-no-mundo-depois-da-agua-e-do-cha/>.
- [8] F. Mohammad, K. B. Lee e Y.-C. Kim, “Previsão de carga de curto prazo usando redes neurais profundas”, *arXiv preprint arXiv:1811.03242*, 2018, Acesso em: 20 jun. 2023. endereço: <https://arxiv.org/abs/1811.03242>.
- [9] R. d. S. D. S. Bártolo, “Previsão Inteligente do consumo de energia com base em modelos Deep Learning”, Acesso em: 20 jun. 2023, Tese de Doutorado, Instituto Politécnico do Porto (Portugal), 2018. endereço: <https://www.proquest.com/openview/c0466f26a68f6defdbce96da12770772/1?pq-origsite=gscholar&cbl=2026366&diss=y>.
- [10] A. G. d. Oliveira Junior, *Calibração de modelos de rede neural para predição de consumo de energia*, <https://repositorio.unifesp.br/handle/11600/61627>, Acesso em: 27 jun. 2023, 2021.
- [11] I. N. d. Silva, D. H. Spatti e R. A. Flauzino, *Redes neurais artificiais para engenharia e ciências aplicadas*, 2ª ed. 2016, Acesso em: 20 jun. 2023. endereço: https://artliber.com.br/amostra/redes_neurais.pdf.
- [12] M. I. V. Furtado, *Redes neurais artificiais: uma abordagem para sala de aula*. Ponta Grossa, PR: Atena Editora, 2019, p. 19, Acesso em: 20 jun. 2023. endereço: <https://cdn.atenaeditora.com.br/documentos/ebook/201908/b8757f5a245a8ae59d75efa8d7cd40c0e2b03766.pdf>.
- [13] Y. Bengio, I. Goodfellow e A. Courville, *Aprendizagem profunda*. Cambridge, MA, EUA: Imprensa do MIT, 2017, Acesso em: 27 jun. 2023.
- [14] L. B. S. e. a. Moraes, *Previsão de carga de curto prazo utilizando redes neurais artificiais*, <https://repositorio.unifei.edu.br/jspui/handle/123456789/3166>, Acesso em: 20 jun. 2023, 2022.
- [15] M. F. Bouzon, *Estudo de algoritmos de otimização inspirados na natureza aplicados ao treinamento de redes neurais artificiais*, <https://repositorio.fei.edu.br/items/9f8061e0-b8f8-469f-9d10-992a3a4ba831>, Acesso em: 21 jun. 2023, 2021.
- [16] D. P. Kingma e J. Ba, “Adam: A method for stochastic optimization”, *arXiv preprint arXiv:1412.6980*, 2014, Accessed on 27th June 2023. endereço: <https://arxiv.org/abs/1412.6980>.