

## 15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

### ANÁLISE MULTIFATORIAL DO IMPACTO DE INDICADORES ECONÔMICOS NO CONSUMO DE DIFERENTES SETORES DA ATIVIDADE ECONÔMICA UTILIZANDO APRENDIZADO DE MÁQUINA (ML)

FELIPE B. SANTOS<sup>1</sup>, TARDELLI R. C. STEKEL<sup>2</sup>

<sup>1</sup> Graduando em Tecnologia de Análise e Desenvolvimento de Sistemas, Bolsista FAPESP, IFSP, Campus Jacareí, felipe.brigagao@aluno.ifsp.edu.br.

<sup>2</sup> Docente EBTT, IFSP, Campus Jacareí, stekel@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.03.03.04-9 Sistemas de Informação

**RESUMO:** Este estudo explora a aplicação de técnicas de Aprendizado de Máquina (ML) para analisar o impacto de indicadores econômicos no consumo de setores como energia, transporte, indústria, imobiliário e serviços. Diante da complexidade crescente das interações econômicas e das limitações dos modelos lineares tradicionais, o uso de ML, incluindo redes neurais, apresenta-se como uma abordagem promissora para identificar padrões complexos e construir modelos preditivos mais precisos. Foram utilizados dados de instituições como IBGE, FGV, Bacen e CNC, organizados em um arquivo CSV e analisados no Google Colab. Após o pré-processamento, que incluiu conversão de formatos e eliminação de dados faltantes, realizaram-se análises exploratórias para identificar correlações. Modelos de regressão e redes neurais foram utilizados para prever o índice de consumo (ICON) e o índice de confiança do consumidor (ICC). A validação dos modelos indicou um erro médio absoluto (MAE) de 443,72 para o ICON, com uma precisão de aproximadamente 87%, e um MAE de 11,54 para o ICC, com precisão de aproximadamente 90%. Contudo, o modelo apresentou limitações ao prever variações extremas, sugerindo a necessidade de ajustes adicionais. O estudo evidencia o potencial das redes neurais na previsão econômica e aponta caminhos para melhorias futuras.

**PALAVRAS-CHAVE:** aprendizado de máquina; indicadores econômicos; consumo; redes neurais; previsão; análise multifatorial

#### MULTIFACTORIAL ANALYSIS OF THE IMPACT OF ECONOMIC INDICATORS ON CONSUMPTION ACROSS DIFFERENT SECTORS USING MACHINE LEARNING

**ABSTRACT:** This study explores the application of Machine Learning (ML) techniques to analyze the impact of economic indicators on consumption in sectors such as energy, transportation, industry, real estate, and services. Given the increasing complexity of economic interactions and the limitations of traditional linear models, ML, including neural networks, emerges as a promising approach to identify complex patterns and build more accurate predictive models. Data from institutions like IBGE, FGV, Bacen, and CNC were used, organized into a CSV file, and analyzed using Google Colab. After preprocessing steps, including format conversions and removal of missing data, exploratory analyses were conducted to identify correlations. Regression models and neural networks were employed to predict the consumption index (ICON) and the consumer confidence index (ICC). Model validation indicated overall good performance but highlighted challenges in predicting extreme variations, suggesting the need for further adjustments. The study demonstrates the potential of neural networks in economic forecasting and suggests pathways for future improvements.

**KEYWORDS:** machine learning; economic indicators; consumption; neural networks; prediction; multifactorial analysis

#### INTRODUÇÃO

A economia global tem enfrentado desafios crescentes, como crises financeiras, mudanças nas políticas monetárias e oscilações nos preços de mercadorias, que afetam diretamente o comportamento do consumo em diversos setores.

A compreensão do impacto de indicadores econômicos no consumo de diferentes setores é essencial para decisões estratégicas em políticas públicas, investimentos e gestão empresarial. A análise tradicional dessas relações geralmente utiliza modelos lineares e univariados. No entanto, a complexidade crescente da economia moderna, com suas múltiplas interdependências e padrões não lineares, demanda ferramentas analíticas mais avançadas (DETIENNE, DETIENNE e JOSHI, 2003).

Neste contexto, o Aprendizado de Máquina (ML) oferece uma abordagem promissora, capaz de identificar padrões complexos e construir modelos preditivos mais robustos. Por meio de algoritmos que aprendem com dados, o ML pode superar as limitações dos métodos estatísticos tradicionais e fornecer insights mais precisos sobre o impacto de variáveis econômicas no consumo.

De acordo com Ludermir (2021), o uso de ML tem sido essencial para resolver problemas complexos que envolvem grandes volumes de dados. As técnicas de ML, como Redes Neurais Artificiais (RNA) e algoritmos de aprendizado supervisionado e não supervisionado, são particularmente eficazes na modelagem de padrões complexos em dados econômicos.

Este estudo propõe uma análise multifatorial do impacto de indicadores econômicos, como inflação, taxas de juros e confiança do consumidor, no consumo de setores como energia, transporte, indústria, imobiliário e serviços. Utilizando redes neurais e outras técnicas de ML, busca-se capturar a complexidade das relações econômicas e fornecer previsões mais precisas (LUXHOJ, RIIS e STENSBALLE, 1996).

Vários pesquisadores demonstraram a capacidade das redes neurais em lidar com fenômenos de comportamento não-linear (LUXHOJ, RIIS e STENSBALLE, 1996; ZHANG, PATUWO e HU, 1998). Essa habilidade é essencial, pois os problemas em marketing frequentemente envolvem dados como vendas e preços, onde as séries temporais apresentam pontos de inflexão, tendências e não-linearidades (BAETS e VENUGOPAL, 1994). Em algumas situações, os dados podem exibir um comportamento caótico (THALL, 1992; HIBBERT e WILKINSON, 1994) e, mesmo nessas circunstâncias, previsões precisas podem ser feitas utilizando redes neurais (LAPEDES e FARBER, 1987).

Assim, espera-se que o uso de ML permita não apenas uma melhor compreensão da dinâmica econômica, mas também a criação de modelos preditivos capazes de auxiliar na tomada de decisões estratégicas de forma mais eficiente e precisa.

## **MATERIAL E MÉTODOS**

Para analisar o impacto dos indicadores econômicos sobre o consumo nos diversos setores, uma vasta gama de dados foi coletada. Estes dados foram extraídos de instituições renomadas, como o Instituto Brasileiro de Geografia e Estatística (IBGE), Fundação Getulio Vargas (FGV), Banco Central do Brasil (Bacen) e Confederação Nacional do Comércio de Bens, Serviços e Turismo (CNC). Os dados foram organizados em formato CSV e analisados no Google Colab.

O processo de pré-processamento incluiu a conversão de formatos e a remoção de dados faltantes. Em seguida, uma análise exploratória de dados (EDA) foi conduzida, envolvendo a criação de mapas de correlação para visualizar a intensidade e direção das relações entre as variáveis. Esses mapas foram gerados utilizando as bibliotecas Pandas e Matplotlib, sendo fundamentais para identificar as variáveis fortemente correlacionadas que serviram de base para as análises subsequentes.

Modelos de regressão com redes neurais foram então desenvolvidos para prever o índice de consumo (ICON) e o índice de confiança do consumidor (ICC). O modelo de rede neural foi construído utilizando a biblioteca TensorFlow e Keras, consistindo em duas camadas densas (fully connected) de 64 neurônios cada, ativadas por ReLU, e uma camada de saída com um neurônio para prever os valores dos índices. O otimizador RMSprop foi utilizado durante o treinamento, e as métricas de desempenho analisadas incluíram o erro médio absoluto (MAE) e o erro quadrático médio (MSE).

Durante o treinamento do modelo, a estratégia de Early Stopping foi aplicada para evitar overfitting. O modelo foi continuamente avaliado em um conjunto de validação, e após o treinamento, sua performance foi verificada utilizando os dados de teste. As previsões foram então comparadas com os valores reais por meio de gráficos de dispersão, que incluíram linhas de tendência para facilitar a interpretação dos resultados.

## RESULTADOS E DISCUSSÃO

A fim de investigar a relação entre as diferentes variáveis, foi construído um gráfico de correlação utilizando as bibliotecas Pandas e Matplotlib. Um mapa de correlação é uma representação gráfica que mostra a correlação entre diferentes variáveis de um conjunto de dados. A correlação é uma medida estatística que indica a força e a direção do relacionamento linear entre duas variáveis. No contexto de um mapa de correlação, essa relação é visualizada através de uma matriz onde cada célula exibe a correlação entre um par específico de variáveis.

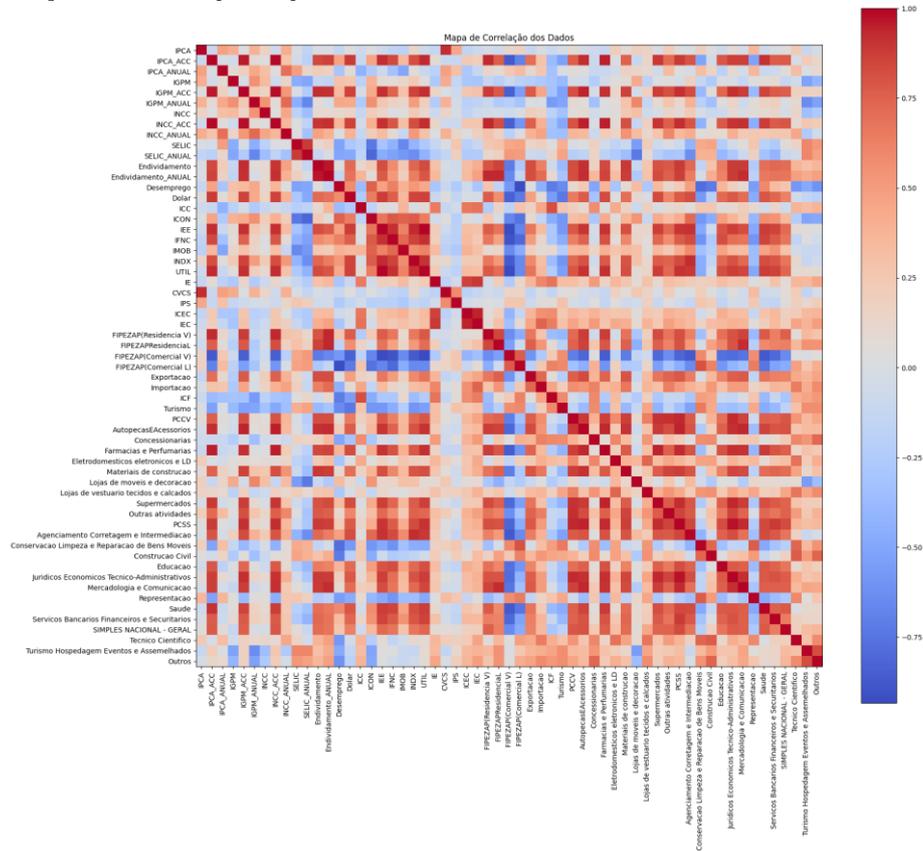


FIGURA 1. Mapa de correlação dos dados utilizados no estudo. As variáveis estão representadas por quadrados, e a força da correlação entre elas é indicada pela cor dos quadrados. As cores mais quentes indicam correlações mais fortes. As cores mais frias indicam correlações mais fracas.

Os resultados obtidos a partir do modelo de regressão com redes neurais demonstraram uma capacidade significativa de previsão dos indicadores. O desempenho do modelo foi avaliado utilizando o erro médio absoluto (MAE) e o erro quadrático médio (MSE), tanto no conjunto de treinamento quanto no conjunto de teste.

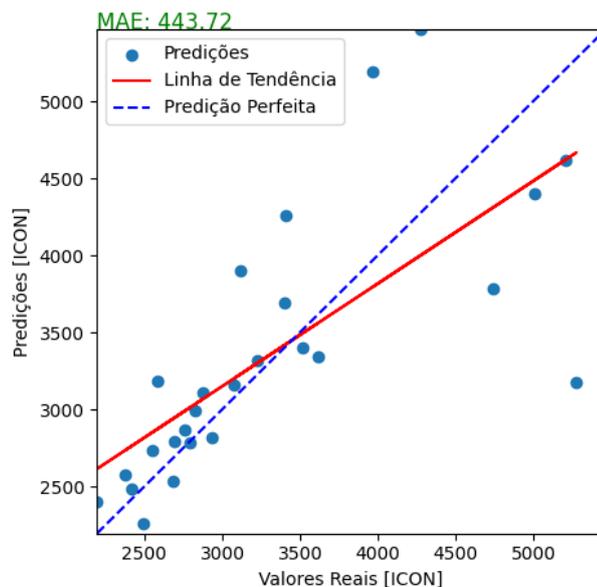


FIGURA 2. Gráfico de dispersão das previsões do modelo comparadas com os valores reais do ICON.

A figura 2 apresenta o gráfico de dispersão que compara os valores previstos pelo modelo com os valores reais do indicador ICON. A linha de tendência (em vermelho) aproxima-se da linha de predição perfeita (linha tracejada azul), indicando uma correlação geral entre as previsões e os valores reais. No entanto, observa-se uma dispersão dos pontos, especialmente em valores mais altos do ICON, sugerindo limitações do modelo em capturar variações extremas. O erro médio absoluto (MAE) é de 443,72, o que evidencia a média das diferenças entre as previsões e os valores reais do ICON. Isso indica que para o ICON, as previsões estão aproximadamente 87% corretas.

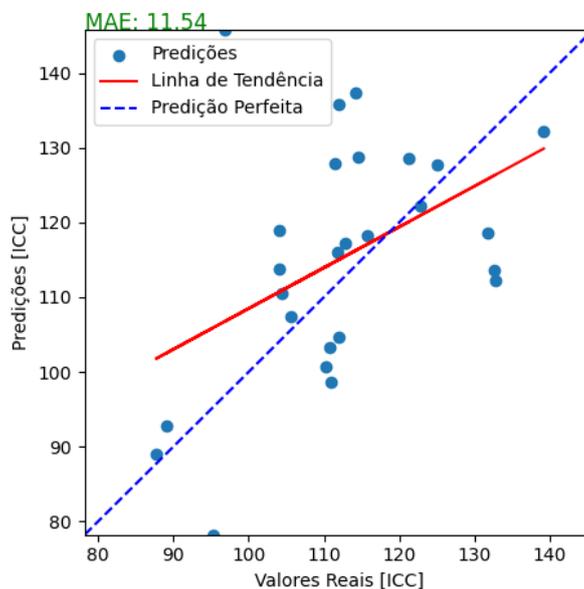


FIGURA 3. Gráfico de dispersão das previsões do modelo comparadas com os valores reais do ICC.

Já a figura 3 apresenta o gráfico de dispersão que compara os valores previstos pelo modelo com os valores reais do indicador ICC. O erro médio absoluto (MAE) é de 11,54, o que evidencia a média das diferenças entre as previsões e os valores reais do ICC. Isso indica que para o ICC, as previsões estão aproximadamente 90% corretas.

Os resultados obtidos com o modelo de regressão de redes neurais mostram que, embora o modelo tenha demonstrado uma boa capacidade de previsão geral, especialmente para valores médios dos indicadores, há limitações em capturar variações extremas, como evidenciado pela dispersão nos gráficos de previsão versus valores reais. A discrepância em valores mais altos do ICON e do ICC sugere que o modelo pode estar sensível a outliers ou a padrões complexos que não foram totalmente capturados pelos dados de treinamento. Para melhorar a precisão, futuras abordagens podem considerar a inclusão de novos dados, ajuste de hiperparâmetros, ou a exploração de arquiteturas de modelos mais complexas que possam capturar melhor as não linearidades presentes nas relações econômicas subjacentes.

## **CONCLUSÕES**

Neste estudo, foi desenvolvido e avaliado um modelo de rede neural para prever o indicador ICON com base em variáveis econômicas relevantes. A análise revelou que, apesar do modelo ter alcançado um desempenho notável na maioria dos casos, há áreas específicas onde melhorias são necessárias.

Primeiramente, o modelo demonstrou um bom nível de precisão ao prever valores médios dos indicadores, o que indica que ele pode ser uma ferramenta útil para análises gerais e para a tomada de decisões em contextos econômicos normais. A capacidade do modelo em identificar padrões e tendências gerais dentro dos dados sugere que a abordagem de redes neurais pode ser promissora para previsões econômicas quando aplicada corretamente.

Entretanto, algumas limitações foram identificadas ao longo do desenvolvimento do modelo. Uma das principais limitações refere-se à dispersão observada em valores mais altos dos indicadores, o que destaca a dificuldade do modelo em prever com precisão esses valores extremos. Isso pode indicar que o modelo não está capturando completamente as complexidades e variações atípicas presentes nos dados econômicos. Outro ponto é que o desempenho do modelo pode ter sido influenciado por características dos dados, como a presença de outliers, a falta de variabilidade suficiente em algumas variáveis e a possível ausência de alguns fatores que impactam os resultados de maneira significativa.

Além disso, o conjunto de dados utilizado pode não ter sido suficientemente amplo ou variado para capturar todas as nuances necessárias para previsões mais robustas, o que representa uma limitação importante na capacidade de generalização do modelo. A arquitetura da rede neural empregada, embora adequada para muitos cenários, pode não ter sido a ideal para lidar com todas as particularidades dos dados econômicos analisados.

Futuras investigações podem explorar a incorporação de técnicas avançadas de pré-processamento de dados, como a remoção ou tratamento de outliers, e a experimentação com diferentes arquiteturas de rede neural para melhorar a capacidade do modelo em lidar com extremos e variações não previstas. Também é recomendável a análise de diferentes conjuntos de dados e fontes adicionais para enriquecer o treinamento e aprimorar a robustez das previsões.

Em resumo, o estudo contribui para a compreensão de como redes neurais podem ser aplicadas na previsão de indicadores econômicos, mas também revela limitações que devem ser abordadas em pesquisas futuras, enfatizando a importância de um desenvolvimento contínuo e da adaptação do modelo às dinâmicas econômicas complexas.

## **CONTRIBUIÇÕES DOS AUTORES**

Felipe Brigagão Santos e Tardelli Ronan Coelho Stekel contribuíram com o desenvolvimento e implementação do modelo de Aprendizado de Máquina, além da coleta e pré-processamento dos dados. Felipe Brigagão Santos esteve diretamente envolvido na concepção e estruturação do artigo, além de realizar a análise exploratória dos dados e a avaliação do desempenho dos modelos. Tardelli Ronan Coelho Stekel contribuiu significativamente na revisão da literatura, além de auxiliar na implementação das redes neurais e na análise dos resultados. Ambos os autores participaram ativamente da redação e revisão final do manuscrito, além de terem contribuído igualmente na discussão e interpretação dos resultados obtidos.

## **AGRADECIMENTOS**

Os autores agradecem Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio concedido por meio da bolsa de iniciação científica, processo 2023/14073-1, fundamental para a realização deste projeto.

## **REFERÊNCIAS**

BAETS, W. R. J.; VENUGOPAL, V. Neural networks and statistical techniques in marketing research: a conceptual comparison. *Marketing Intelligence & Planning*, Bradford, v.12, n.7, p.30-38, 1994.

DETIENNE, K. B.; DETIENNE, D. H.; JOSHI, S. A. Neural networks as statistical tools for business researchers. *Organizational Research Methods*, Thousand Oaks, v.6, n.2, p.236-265, Apr. 2003.

HIBBERT, B.; WILKINSON, I. F. Chaos theory and the dynamics of marketing systems. *Journal of Academy of Marketing Science*, Greenvale, v.22, n.3, p.218-233, Summer 1994.

LAPEDES, A.; FARBER, R. Nonlinear signal processing using neural networks prediction and system modeling. Technical Report LA-UR-87-2662. Los Alamos, National Laboratory, Los Alamos, NM, 1987.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. *Estudos Avançados*, v.35, n.101, 2021.

LUXHOJ, J. T.; RIIS, J. O.; STENSBALLE, B. A hybrid econometric-neural network modeling approach for sales forecasting. *International Journal of Production Economics*, Amsterdam, v.43, n.2-3, p.175-192, 1996.

THALL, N. Neural forecasts: a retail sales booster. *Discount Merchandiser*, v.32, n.10, p.41-42, 1992.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, Amsterdam, v.14, p.35-62, 1998.