

# Aplicação da Arquitetura LSTM na Previsão da Inflação

João Emilio Soares Michalski  
IME-UFRGS  
Porto Alegre-RS, Brasil  
joemilio55@gmail.com

João Henrique Ferreira Flores  
IME-UFRGS  
Porto Alegre-RS, Brasil  
joao.flores@ufrgs.br

## Resumo

Economic time series forecasting represents one of the most significant challenges in economics, directly impacting economic policy decisions. This research investigated the applicability of Long Short-Term Memory (LSTM) recurrent neural networks for forecasting Brazilian inflation, measured by the IPCA, from January 2015 to March 2025. The primary objective was to evaluate the predictive capacity of LSTM-based models by comparing different variable configurations and time horizons. Four distinct LSTM models were developed: Model 1 used 22 predictor variables, including 12 IPCA lags and macroeconomic variables (IGP, Selic, unemployment, GDP); Model 2 employed 16 variables, combining extensive IPCA lags with selected economic variables; Model 3 implemented a univariate architecture with 3 IPCA lags; and Model 4 used exclusively 12 consecutive IPCA lags. Different hyperparameters were tested via grid search, such that each of the four models was tested with 84 distinct combinations, totaling 336 different configurations. Variables were temporally standardized, and the data were split into training and testing sets at an 80/20 ratio, maintaining chronological order. The results demonstrated that Model 4, utilizing exclusively 12 IPCA lags in a univariate architecture, achieved the best overall performance with the lowest MAE (0.1834) and MSE (0.0669) values. This finding reinforces the importance of inflationary memory and seasonal patterns in Brazilian price dynamics, providing evidence that inflation persistence can be captured by machine learning models, making them a viable tool to complement traditional econometric models in the formulation of monetary policy.

## Keywords

Inflação, Aprendizagem da Máquina, Modelagem, LSTM

## 1 Introdução

A previsão de séries temporais econômicas representa uma das atividades mais fundamentais e desafiadoras no campo da economia aplicada, impactando diretamente a vida de milhões de pessoas através das decisões de política econômica que dela derivam. A capacidade de antecipar movimentos em variáveis macroeconômicas como inflação, taxa de câmbio, crescimento do PIB e índices de emprego é essencial para a formulação de políticas públicas eficazes, planejamento empresarial estratégico e tomada de decisões de investimento. Quando autoridades monetárias conseguem prever com maior precisão a trajetória da inflação, por exemplo, podem ajustar suas políticas de forma mais tempestiva e adequada, evitando tanto recessões desnecessárias quanto surtos inflacionários que corroem o poder de compra da população. Da mesma forma, empresas e indivíduos se beneficiam dessas previsões para planejar investimentos, definir preços e proteger seus patrimônios contra volatilidades econômicas.

Nas últimas décadas, o campo da inteligência artificial testemunhou avanços extraordinários, particularmente no desenvolvimento de redes neurais artificiais e técnicas de aprendizado profundo. Essas tecnologias revolucionaram diversas áreas do conhecimento, desde reconhecimento de padrões em imagens até processamento de linguagem natural, demonstrando capacidades superiores em capturar relações complexas e não lineares em grandes volumes de dados. Como observado por [1], o aprendizado profundo tem se mostrado particularmente eficaz em tarefas que envolvem a descoberta de representações hierárquicas em dados complexos, características que são especialmente relevantes para séries temporais econômicas. No contexto econômico, as redes neurais apresentam vantagens sobre modelos tradicionais ao conseguirem identificar padrões sutis em séries temporais, adaptarem-se a mudanças estruturais nos dados e processarem múltiplas variáveis simultaneamente sem a necessidade de especificações rígidas sobre as relações funcionais entre elas. Arquiteturas especializadas como as redes neurais recorrentes (RNNs) e suas variações mais sofisticadas, incluindo as redes LSTM (Long Short-Term Memory), foram especificamente desenvolvidas para lidar com a natureza sequencial e temporal dos dados, características intrínsecas das séries econômicas.

Neste contexto, a aplicação de redes neurais à previsão do IPCA emerge como uma abordagem que pode complementar os métodos tradicionais de análise econômica. Este trabalho busca investigar como essas tecnologias podem capturar padrões complexos nos dados inflacionários brasileiros, oferecendo uma ferramenta adicional para economistas e formuladores de política. A exploração dessas técnicas não pretende substituir o conhecimento econômico consolidado, mas sim expandir o arsenal metodológico disponível para o estudo da dinâmica inflacionária, potencialmente contribuindo para previsões mais precisas e, consequentemente, para decisões de política econômica mais bem fundamentadas.

O presente trabalho tem como objetivo principal avaliar a capacidade preditiva de modelos baseados em redes neurais LSTM para previsão da inflação brasileira, medida pelo IPCA, comparando diferentes configurações de variáveis e horizontes temporais. Para tanto, pretende-se desenvolver quatro modelos LSTM distintos utilizando diferentes combinações de variáveis macroeconômicas brasileiras, comparar o desempenho preditivo dos modelos propostos através de métricas estatísticas apropriadas e avaliar a adequação das redes LSTM como ferramenta complementar aos métodos econométricos tradicionais na análise inflacionária.

Esta pesquisa limita-se à análise da inflação brasileira medida pelo IPCA no período de janeiro de 2015 a março de 2025, utilizando dados mensais de variáveis macroeconômicas selecionadas. O período de aproximadamente 10 anos foi escolhido por proporcionar uma amostra que entendemos que seja adequada para o treinamento de modelos de redes neurais LSTM, garantindo a captura de diferentes ciclos econômicos e choques macroeconômicos que

afetaram a economia brasileira, incluindo a recessão de 2015-2016, a pandemia de COVID-19 e seus efeitos inflacionários subsequentes.

As variáveis utilizadas neste trabalho foram selecionadas com base em referências empíricas estabelecidas nos estudos [2-4]. Esses estudos fornecem base teórica sólida e evidência prática sobre a relevância macroeconômica de indicadores como a inflação (IPCA e INPC), o Produto Interno Bruto (PIB), a taxa de desemprego, a taxa de juros (Selic), a expectativa de inflação, e indicadores de atividade econômica. Com exceção do Índice Geral de Preços (IGP), todas as variáveis utilizadas neste trabalho derivam diretamente da estrutura empírica adotada nos referidos artigos. A inclusão do IGP-M justifica-se por sua capacidade de capturar pressões inflacionárias em diferentes estágios da cadeia produtiva, complementando a informação fornecida pelas demais variáveis macroeconômicas selecionadas. O estudo foca exclusivamente em previsões de um passo à frente (one-step ahead), não abordando horizontes de previsão mais longos. Esta escolha metodológica é fundamentada em várias considerações importantes. Primeiro, as previsões de um passo à frente tendem a ser significativamente mais precisas do que previsões de horizontes mais longos, uma vez que os erros de previsão se acumulam ao longo do tempo, criando um efeito cascata onde pequenos erros iniciais podem se amplificar substancialmente em séries temporais econômicas caracterizadas por alta volatilidade. Segundo, para autoridades monetárias e formuladores de política econômica, as previsões de curto prazo são frequentemente as mais relevantes para decisões operacionais e intervenções tempestivas no mercado. Terceiro, ao focar no one-step ahead, o estudo pode concentrar-se na otimização da arquitetura LSTM para essa tarefa específica, permitindo uma análise mais profunda das diferentes configurações do modelo sem as complexidades metodológicas adicionais introduzidas por estratégias multi-step ahead.

## 2 Referencial metodológico

### 2.1 Modelos de Equilíbrio Geral Estocástico Dinâmico (DSGE)

Conforme [5], os modelos dinâmicos estocásticos de equilíbrio geral (DSGE) representam uma evolução natural da teoria do equilíbrio geral, incorporando elementos dinâmicos e estocásticos que permitem uma análise mais robusta dos fenômenos macroeconômicos. Os modelos DSGE fundamentam-se na premissa de que os agentes econômicos - famílias, empresas e governo - otimizam suas decisões intertemporalmente, considerando suas restrições orçamentárias e as informações disponíveis. Essa abordagem busca fornecer microfundamentos sólidos para a análise macroeconômica, conectando o comportamento individual dos agentes com os resultados agregados da economia. A estrutura matemática dos modelos DSGE baseia-se em um sistema de equações que descrevem o comportamento otimizador dos agentes. O sistema típico inclui a equação de Euler para o consumo das famílias é representada por 1:

$$E_t = \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\sigma} (1 + r_{t+1}) \right] = 1. \quad (1)$$

Em que  $E_t$  representa o operador expectativa condicional à informação disponível no período  $t$ ,  $\beta$  é o fator de desconto subjetivo,  $C_t$  denota o consumo no período  $t$ ,  $\sigma$  é o coeficiente de aversão

relativa ao risco, e  $r_{t+1}$  representa a taxa de juros real no período  $t + 1$ .

A função de produção agregada é representada por 2:

$$Y_t = A_t K_t^\alpha L_t^{1-\alpha}. \quad (2)$$

Sendo  $Y_t$  o produto agregado,  $A_t$  representa a produtividade total dos fatores,  $K_t$  é o estoque de capital,  $L_t$  denota o trabalho e  $\alpha$  é a elasticidade do produto em relação ao capital.

O processo estocástico para choques de produtividade segue a especificação de 3:

$$\ln(A_t) = \rho_a \ln(A_{t-1}) + \varepsilon_{a,t}. \quad (3)$$

Onde  $\rho_a$  é o parâmetro de persistência do choque e  $\varepsilon_{a,t}$  é um ruído branco com distribuição normal.

Em [5] é estabelecido que os parâmetros fundamentais incluem  $\beta$  (fator de desconto subjetivo),  $\sigma$  (coeficiente de aversão relativa ao risco),  $\alpha$  (elasticidade do produto em relação ao capital) e os parâmetros de persistência dos choques estocásticos. Estes parâmetros são calibrados ou estimados com base em dados empíricos e evidências econométricas, sendo cruciais para determinar as propriedades dinâmicas do modelo.

Os avanços recentes em técnicas de aprendizado de máquina e a crescente disponibilidade de dados macroeconômicos de alta frequência abrem novas possibilidades metodológicas para a análise inflacionária. Enquanto os modelos DSGE oferecem fundamentação teórica sólida, as redes neurais recorrentes apresentam características complementares, particularmente sua capacidade de capturar relações não-lineares complexas e padrões temporais nos dados sem a necessidade de especificações estruturais predefinidas. Esta flexibilidade metodológica pode ser especialmente valiosa para identificar dinâmicas inflacionárias que emergem diretamente dos dados empíricos.

### 2.2 Previsão de séries temporais

Séries temporais econômicas, como inflação, PIB e taxa de juros, apresentam características complexas como não linearidade e dependências de longo prazo. Tradicionalmente, técnicas como ARIMA, Vetores Autorregressivos (VAR) ou mesmo modelos estruturais como DSGE foram amplamente empregados para a análise e previsão dessas variáveis, como exemplificado pelo modelo SAMBA (Stochastic Analytical Model with a Bayesian Approach) desenvolvido para o Banco Central do Brasil [6], pelos estudos de modelos DSGE regionais para avaliar impactos da política monetária sobre o PIB brasileiro [7] e pelas análises de previsão macroeconômica desenvolvidas pelo Grupo de Conjuntura do IPEA para projeções do PIB e inflação brasileiros [8]. Essas metodologias consolidaram-se na literatura econométrica devido à sua capacidade de modelar estruturas temporais específicas e fornecer interpretações teóricas consistentes com a teoria econômica. No entanto, esses modelos sofrem de certas limitações devido à sua dependência de relações lineares, que podem falhar em capturar a não linearidade e complexidade inerentes dos sistemas econômicos [9]. Além disso, o modelo pode ser sensível a outliers e requer um conhecimento prévio sobre a série temporal para uma modelagem eficaz, o que limita sua eficácia frente à complexidade dos dados macroeconômicos contemporâneos.

### 2.3 Redes Neurais Recorrentes

As redes neurais recorrentes (RNNs) constituem uma classe de modelos de aprendizado profundo voltada para a modelagem de dados sequenciais, onde a ordem temporal da informação é essencial. Diferentemente das redes neurais do tipo feedforward, que processam entradas de forma independente, as RNNs utilizam conexões recorrentes que permitem a retenção de informações ao longo do tempo. Esta capacidade de manter informações temporais habilita o aprendizado de dependências temporais complexas [1]. Essa característica torna as RNNs particularmente eficazes em tarefas como previsão de séries temporais, tradução automática e análise de dados econômicos. Apesar de seu potencial, as RNNs tradicionais enfrentam dificuldades em capturar dependências de longo prazo devido a problemas como o desvanecimento ou explosão do gradiente durante o processo de treinamento, especialmente quando aplicadas a sequências extensas. Devido a isso, outras arquiteturas foram desenvolvidas [1].

### 2.4 Modelos LSTM (Long Short-Term Memory)

Para resolver as limitações das RNNs tradicionais, [10] propuseram a arquitetura Long Short-Term Memory (LSTM), uma variante das RNNs com mecanismos de controle de fluxo de informação. As redes LSTM são apresentadas como uma arquitetura especializada desenvolvida para superar o problema do gradiente desvanecente [1]. Esta arquitetura introduz um mecanismo de memória celular controlado por estruturas denominadas "portas"(gates). A estrutura de uma LSTM é composta por três portas principais que regulam o fluxo de informações. A porta de entrada define quais informações serão armazenadas na célula de memória. A porta de esquecimento elimina informações desnecessárias do estado anterior. A porta de saída decide quais informações serão passadas adiante na sequência, conforme Figura 1.

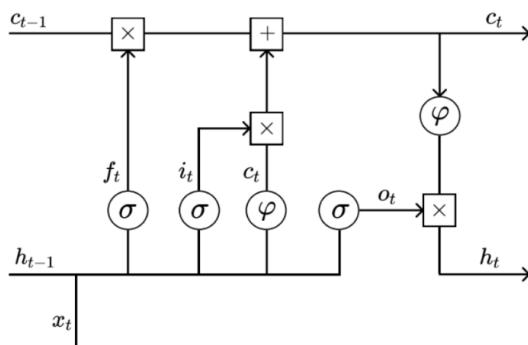


Figura 1: Diagrama de uma LSTM, considerando  $\sigma$  como a função de ativação logit [11]

Especificamente, a arquitetura LSTM many-to-one que será utilizada representa uma configuração específica de redes neurais recorrentes onde uma sequência de múltiplas entradas temporais é processada para produzir uma única saída. Conforme descrito em [1], esta arquitetura é particularmente adequada para tarefas de previsão de séries temporais, classificação de sequências e resumo de informações temporais, onde o objetivo é extrair uma representação condensada ou fazer uma previsão baseada em toda

a sequência de entrada. Esta configuração contrasta com outras arquiteturas de RNNs, como one-to-many ou many-to-many, sendo especificamente otimizada para cenários onde múltiplas observações temporais devem ser sintetizadas em uma única decisão ou previsão. Na configuração many-to-one, a rede processa sequencialmente cada elemento da entrada temporal, mantendo informações relevantes através dos estados ocultos e da memória de célula da LSTM, conforme Figura 2. Esta estrutura permite que a rede capture dependências temporais complexas ao longo de toda a sequência de entrada, aproveitando a capacidade inerente das células LSTM de preservar informações importantes em horizontes temporais extensos. O processamento sequencial garante que cada observação temporal seja contextualizada pelas observações anteriores, criando uma representação cumulativa que incorpora tanto padrões de curto prazo quanto tendências de longo prazo presentes na série temporal. Esta abordagem é especialmente vantajosa para dados econômicos, onde eventos passados podem ter influência duradoura sobre comportamentos futuros, e onde a ordem temporal das observações carrega informação estrutural importante para a qualidade das previsões.

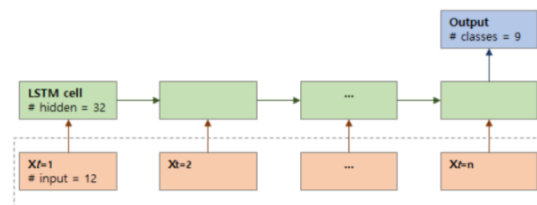


Figura 2: Diagrama de uma LSTM, considerando a arquitetura many-to-one [12]

Durante o processamento sequencial, cada célula recebe como entrada o vetor de características  $x_t$  do tempo atual e o estado oculto  $h_{t-1}$  do passo anterior. A célula LSTM atualiza seus estados internos (estado da célula  $C_t$  e estado oculto  $h_t$ ) através dos mecanismos de gates de esquecimento, entrada e saída, permitindo que informações relevantes sejam preservadas ao longo da sequência e informações irrelevantes sejam descartadas. A característica distintiva desta arquitetura é que apenas o estado oculto final  $h_t$  é utilizado para gerar a saída  $y$ , tipicamente através de uma ou mais camadas densas que mapeiam o estado final para o espaço de saída desejado. Esta abordagem permite que toda a informação temporal da sequência seja condensada em uma única representação vetorial que captura os padrões temporais mais relevantes, funcionando como um mecanismo de compressão inteligente que extrai as características mais preditivas da série histórica para fundamentar a previsão final.

### 2.5 Definições Econométricas

**2.5.1 Inflação (IPCA).** A inflação é o aumento generalizado e contínuo dos preços de bens e serviços em uma economia, resultando na perda do poder de compra da moeda. No Brasil, o principal indicador oficial de inflação é o Índice Nacional de Preços ao Consumidor Amplo (IPCA), calculado mensalmente pelo IBGE. O IPCA mede a variação de preços de uma cesta de bens e serviços consumidos

por famílias com renda de 1 a 40 salários mínimos. Ele é divulgado geralmente na primeira quinzena do mês subsequente ao de referência e serve como referência para o regime de metas de inflação do Banco Central do Brasil.

**2.5.2 Produto Interno Bruto (PIB).** O PIB é a soma de todos os bens e serviços finais produzidos em um país durante um determinado período. No Brasil, é calculado pelo IBGE com base em três óticas distintas: produção, renda e despesa. A ótica da produção corresponde à soma dos valores adicionados setoriais mais impostos líquidos sobre produtos. A ótica da renda representa a soma dos rendimentos pagos aos fatores de produção. A ótica da despesa é expressa por 4:

$$PIB = N + I + G + (X - M). \quad (4)$$

Sendo que N representa o consumo das famílias, I os investimentos, G os gastos do governo, X as exportações e M as importações. O PIB é divulgado trimestralmente e serve como principal medida da atividade econômica nacional.

**2.5.3 Taxa de Desemprego.** A taxa de desemprego representa a proporção da população economicamente ativa que está desocupada, ou seja, não trabalha, mas está disponível e procurando emprego. No Brasil, esse indicador é calculado pelo IBGE (Instituto Brasileiro de Geografia e Estatística), com base na PNAD Contínua, e é divulgado trimestralmente [8]. A metodologia adotada segue padrões internacionais estabelecidos pela Organização Internacional do Trabalho (OIT), garantindo comparabilidade com outros países e permitindo análises consistentes da evolução do mercado de trabalho brasileiro ao longo do tempo.

**2.5.4 Taxa Selic.** A Selic é a taxa básica de juros da economia brasileira, definida pelo Comitê de Política Monetária (Copom) do Banco Central do Brasil [13]. Ela influencia todas as demais taxas do mercado e é o principal instrumento para controle da inflação no regime de metas inflacionárias. As decisões do Copom ocorrem a cada 45 dias, aproximadamente, em reuniões regulares onde são avaliadas as condições econômicas domésticas e internacionais para definir o nível adequado da taxa de juros.

**2.5.5 Índice Geral de Preços – Mercado (IGP-M).** O IGP-M é calculado mensalmente pela FGV IBRE e reflete a variação de preços em diferentes estágios da produção. Ele é composto por três sub-índices com ponderações específicas: IPA-M (60% - preços no atacado), IPC-M (30% - preços ao consumidor), e INCC-M (10% - custo da construção civil). Sua coleta ocorre entre os dias 21 do mês anterior e 20 do mês de referência, sendo amplamente usado em reajustes de contratos, como aluguéis. A inclusão do IGP-M defasado no modelo LSTM busca antecipar tendências de inflação antes que se manifestem no IPCA, aproveitando sua característica de capturar pressões inflacionárias nos estágios iniciais da cadeia produtiva.

## 3 Metodologia

### 3.1 Tratamento dos Dados

As fontes de dados integradas para o estudo incluem o IPCA do IBGE como variável target, complementado por indicadores macroeconômicos relevantes: IGP Mensal da FGV, Taxa Selic do Banco

Central do Brasil, Taxa de Desemprego do IBGE e PIB do IBGE. O processamento específico dos dados envolveu a padronização de formatos de data para garantir consistência temporal, conversão de frequências das variáveis macroeconômicas de trimestral para mensal e alinhamento temporal de todas as séries para um período comum de análise de janeiro/2015 a março/2025. A criação de variáveis defasadas foi estruturada considerando as características específicas de cada indicador econômico e suas propriedades temporais na dinâmica inflacionária. O IPCA recebeu 12 defasagens, representadas por  $IPCA_{(t-1)}$  até  $IPCA_{(t-12)}$ , para capturar completamente a sazonalidade anual característica deste índice de preços. Os demais indicadores receberam estruturas de defasagem mais parcimoniosas, alinhadas com suas respectivas dinâmicas econômicas. O IGP Mensal foi especificado com 3 defasagens,  $IGP_{(t-1)}$  até  $IGP_{(t-3)}$ , para capturar os impactos de curto prazo das variações de preços no atacado e suas transmissões para o varejo. A Taxa Selic recebeu defasagens  $SELIC_{(t-1)}$  até  $SELIC_{(t-3)}$ , refletindo os efeitos defasados da política monetária sobre a economia. A Taxa de Desemprego foi modelada com defasagens  $DESEMPREGO_{(t-1)}$  até  $DESEMPREGO_{(t-3)}$  para capturar a dinâmica do mercado de trabalho e seus impactos sobre a demanda agregada e pressões inflacionárias via canal de custos salariais. O PIB recebeu tratamento diferenciado, sendo especificado apenas com uma defasagem  $PIB_{(t-4)}$  devido ao seu caráter estrutural na economia.

### 3.2 Otimização de hiper parâmetros

Para garantir a seleção ótima da arquitetura LSTM, será implementado um sistema de busca de hiper parâmetros utilizando grid search que testará sistematicamente diferentes combinações de configurações. O espaço de busca incluirá unidades LSTM variando de 20 a 50 em incrementos de 5 unidades, três funções de ativação compreendendo ReLU, Tanh e Sigmoid, além de taxas de dropout iniciando em 0.05 até 0.20 com incrementos de 0.05. O processo de seleção envolverá a execução de um grid search completo testando todas as combinações possíveis, totalizando 84 configurações para cada modelo resultantes da combinação de 3 funções de ativação, 7 opções de unidades LSTM e 4 taxas de dropout, onde cada configuração será treinada por até 50 épocas com implementação de parada antecipada e utilizando o erro quadrático médio no conjunto de validação como métrica de seleção. A taxa de dropout é uma técnica de regularização que desativa aleatoriamente uma proporção específica dos neurônios durante o treinamento, reduzindo o risco de overfitting e melhorando a capacidade de generalização do modelo. As épocas representam o número de vezes que o algoritmo de aprendizado percorre todo o conjunto de dados de treinamento, sendo que cada época consiste em uma passagem completa pelos dados para ajustar os pesos da rede neural através do processo de otimização. A configuração apresentando menor erro quadrático médio será automaticamente selecionada como arquitetura ótima, seguida por uma segunda etapa de otimização da parada antecipada que testará valores de paciência variando de 15 a 35 épocas em incrementos de 5, permitindo a seleção individual do valor ótimo para cada modelo baseado na minimização do erro quadrático médio. Os valores de paciência determinam quantas épocas consecutivas sem melhoria no desempenho do modelo no conjunto de validação o algoritmo deve aguardar antes de interromper o treinamento,

sendo um parâmetro crucial para evitar tanto o término prematuro do aprendizado quanto o overfitting por treinamento excessivo. O modelo final será então treinado utilizando as características selecionadas através do processo de grid search e posteriormente avaliado no conjunto de teste para verificação de sua capacidade de generalização.

### 3.3 Especificações dos Modelos

Todos os modelos utilizaram o otimizador Adam com função de perda do erro quadrático médio, parada antecipada monitorando a loss de validação para evitar o sobre ajuste, e serão treinados por até 50 épocas com lote de tamanho 8. O otimizador Adam (Adaptive Moment Estimation) é um algoritmo de otimização que combina as vantagens do momentum com a adaptação da taxa de aprendizado, calculando médias móveis exponenciais tanto dos gradientes quanto dos seus quadrados para ajustar individualmente as taxas de aprendizado de cada parâmetro, proporcionando convergência mais rápida e estável em comparação com métodos tradicionais como o gradiente descendente estocástico (Kingma & Ba, 2014). Os dados serão normalizados e divididos em conjuntos de treino e teste com proporção 80/20, mantendo rigorosamente a ordem temporal para preservar a estrutura de dependência temporal dos dados.

**3.3.1 Modelo 1:** O Modelo 1 implementará uma arquitetura LSTM multivariada que utilizará múltiplas defasagens de variáveis macroeconômicas para prever o IPCA. A estrutura do modelo incorporará uma ampla gama de variáveis preditoras, incluindo 12 defasagens do próprio IPCA representadas por  $IPCA_{(t-1)}$  até  $IPCA_{(t-12)}$ , 3 defasagens do IGP Mensal de  $IGP_{(t-1)}$  até  $IGP_{(t-3)}$ , 3 defasagens da Taxa Selic variando de  $SELIC_{(t-1)}$  até  $SELIC_{(t-3)}$ , 3 defasagens da Taxa de Desemprego compreendendo  $DESEMPREGO_{(t-1)}$  até  $DESEMPREGO_{(t-3)}$  e 1 defasagem do PIB representada por  $PIB_{(t-1)}$ , totalizando 22 variáveis preditoras.

**3.3.2 Modelo 2:** O Modelo 2 implementará uma arquitetura LSTM que combinará defasagens extensas do IPCA com variáveis macroeconômicas selecionadas para prever o comportamento inflacionário futuro. A estrutura incorporará as defasagens  $IPCA_{(t-1)}$  até  $IPCA_{(t-12)}$  representando os 12 meses anteriores do próprio índice, complementadas por variáveis econômicas com defasagens específicas:  $PIB_{(t-4)}$ ,  $IGP_{(t-1)}$ ,  $SELIC_{(t-1)}$ , e  $DESEMPREGO_{(t-1)}$ , totalizando 16 variáveis preditoras.

**3.3.3 Modelo 3:** O Modelo 3 implementará uma arquitetura LSTM univariada minimalista que utilizará exclusivamente defasagens de curto prazo do próprio IPCA para realizar previsões inflacionárias. A estrutura do modelo incorporará apenas 3 defasagens consecutivas do IPCA representadas por  $IPCA_{(t-1)}$  até  $IPCA_{(t-3)}$  como variáveis preditoras.

**3.3.4 Modelo 4:** O Modelo 4 implementará uma arquitetura LSTM univariada que utilizará 12 defasagens consecutivas do IPCA para capturar padrões sazonais completos e realizar previsões inflacionárias de médio prazo. A estrutura incorporará defasagens de  $IPCA_{(t-1)}$  até  $IPCA_{(t-12)}$ , representando exatamente um ano completo de histórico do índice de preços.

## 4 Resultados

Após a execução do grid search, como descrito anteriormente, os resultados demonstram que a função de ativação tanh apresentou melhor desempenho nos modelos 1 e 2, enquanto a função sigmoide obteve os melhores resultados nos modelos 3 e 4. Observou-se também que valores de dropout intermediários (0.15- 0.20) foram predominantes nas configurações ótimas, indicando a importância da regularização para prevenir o sobre ajuste, conforme pode-se observar na Tabela 1.

**Tabela 1: Modelos finais após grid search**

Modelo	Unidades LSTM	Função ativação	Dropout	MSE
Modelo 1	25	ReLU	0.20	0.009195
Modelo 2	40	tanh	0.15	0.011227
Modelo 3	45	sigmoide	0.15	0.014796
Modelo 4	30	sigmoide	0.15	0.013669

A análise comparativa dos quatro modelos apresentados na Tabela 1 revela características distintas de desempenho e configuração. O Modelo 1 apresentou o melhor resultado em termos de precisão, alcançando o menor erro quadrático médio (0.009195) com uma arquitetura mais compacta de 25 unidades LSTM e função de ativação ReLU, demonstrando eficiência na generalização dos padrões temporais. O Modelo 2, com 40 unidades LSTM e função tanh, obteve um desempenho intermediário (MSE: 0.011227), mantendo um equilíbrio adequado entre complexidade e acurácia. Os Modelos 3 e 4, ambos utilizando função sigmoide, apresentaram erros superiores (0.014796 e 0.013669, respectivamente), sugerindo que esta função de ativação pode não ser a mais adequada para o problema em questão. A configuração de paciência variou significativamente entre os modelos, com o Modelo 3 apresentando maior tolerância ao sobre ajuste (35 épocas) comparado ao Modelo 1 (15 épocas), indicando diferentes estratégias de regularização durante o treinamento. Com os modelos treinados, passou-se a avaliação dos mesmos quanto ao conjunto de teste.

**Tabela 2: Desempenho dos Modelos LSTM com Configurações ótimas no conjunto de teste (one step ahead)**

Modelo	MAE	MAPE	MSE
Modelo 1	0.2054	<b>118.45%</b>	0.0902
Modelo 2	0.1867	127.29%	0.0702
Modelo 3	0.1977	156.50%	0.0790
Modelo 4	<b>0.1834</b>	137.82%	<b>0.0669</b>

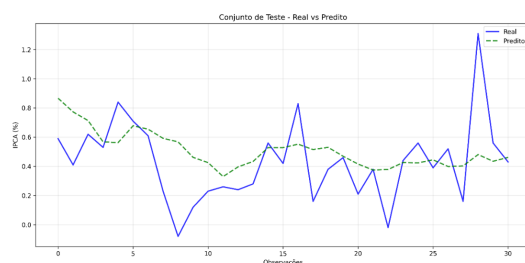
A avaliação dos quatro modelos LSTM propostos foi realizada com métricas estatísticas padronizadas: Erro Absoluto Médio (MAE), Percentual de Erro Absoluto Médio (MAPE) e Erro Quadrático Médio (MSE). Os resultados finais foram obtidos após o treinamento completo utilizando as configurações ótimas identificadas na Tabela 1. Conforme demonstrado na Tabela 2, o Modelo 4 emergiu como o de melhor desempenho geral, apresentando os menores valores de MAE (0.1834) e MSE (0.0669), indicando maior precisão nas previsões. Verificamos na mesma tabela que o Modelo 1, apesar de

ter apresentado os maiores valores de MAE (0.2054) e MSE (0.0902), obteve o menor MAPE (118.45%), sugerindo melhor desempenho relativo em termos percentuais. Seguimos com a análise gráfica do ajuste ao conjunto de teste, de forma a verificar se algum dos modelos apenas prevê a média ou se consegue capturar as flutuações do índice.



**Figura 3: Estimativas do modelo 1 no conjunto de teste**

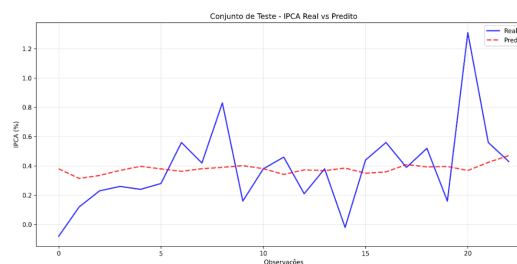
Ao analisarmos a Figura 3 referente ao Modelo 1, notamos limitações significativas na capacidade preditiva do modelo, especialmente no tratamento de movimentos extremos da série temporal do IPCA. As previsões apresentam-se visivelmente mais suaves que os dados reais, concentrando-se predominantemente numa faixa restrita entre 0% e 0.8%, enquanto os valores observados exibem variações substancialmente mais amplas, incluindo picos superiores a 1.2% e períodos próximos à deflação, mostrando que o modelo mantém um padrão conservador. Particularmente notável é a incapacidade do modelo de antecipar o pico inflacionário próximo à observação 28, onde o IPCA real atinge aproximadamente 1.3%, enquanto a previsão permanece estagnada em torno de 0.3%.



**Figura 4: Estimativas do modelo 2 no conjunto de teste**

Já o Modelo 2, conforme a Figura 4, apresenta um desempenho coerente no conjunto de teste, onde as previsões mantêm boa capacidade de seguimento da tendência geral dos dados observados. As previsões exibem comportamento mais estável e suave em relação aos valores reais, mas conseguem capturar adequadamente os movimentos direcionais principais da série. Observa-se boa aderência nas observações iniciais e um ajuste razoável durante o período intermediário. O pico final próximo à observação 28, onde o valor real atinge aproximadamente 1.3%, representa o ponto de maior divergência, com as previsões mantendo-se em níveis mais moderados,

em torno de 0.5-0.6%, ainda assim seguindo a direção ascendente do movimento inflacionário.



**Figura 5: Estimativas do modelo 4 no conjunto de teste**

A Figura 5 confirma certas limitações no conjunto de teste do Modelo 4, onde o modelo mantém um padrão de previsões excessivamente suavizadas, com estimativas concentradas predominantemente na faixa de 0.3% a 0.5%. Esta performance sugere que o modelo desenvolveu uma estratégia de "regressão à média", priorizando estabilidade em detrimento da precisão na captura de variações bruscas, o que compromete sua utilidade prática para previsões econômicas em contextos de alta incerteza e volatilidade macroeconômica. O Modelo 3 não é apresentado, pois se restringe a manter a estimativa junto a média de 0.5%, não apresentando nenhum destaque a ser mencionado.

## 5 Conclusão

Esta pesquisa investigou a aplicabilidade de redes neurais recorrentes do tipo Long Short-Term Memory (LSTM) para a previsão da inflação brasileira, desenvolvendo e comparando quatro arquiteturas distintas que variaram desde configurações multivariadas complexas até abordagens univariadas. Os resultados demonstraram que o Modelo 4, utilizando exclusivamente 12 defasagens do IPCA em uma arquitetura univariada, apresentou o melhor desempenho geral com os menores valores de MAE (0.1834) e MSE (0.0669). Surpreendentemente, os modelos multivariados que incorporaram múltiplas variáveis macroeconômicas não superaram as abordagens mais simples, sugerindo que a informação histórica do próprio índice de preços ao longo de um ciclo sazonal completo contém maior poder preditivo para a inflação brasileira do que a inclusão de variáveis econômicas adicionais. Este achado reforça a importância da memória inflacionária e dos padrões sazonais na dinâmica de preços brasileira, oferecendo evidências de que a persistência inflacionária pode ser efetivamente capturada por modelos de aprendizado de máquina. É fundamental destacar que o objetivo principal desta pesquisa foi a maximização da capacidade preditiva dos modelos LSTM, e não a explicação dos mecanismos causais subjacentes à dinâmica inflacionária brasileira. Diferentemente dos estudos econométricos tradicionais que buscam identificar e quantificar os impactos de variáveis exógenas sobre a inflação, este trabalho adotou uma abordagem puramente preditiva, focada em desenvolver ferramentas de previsão que possam complementar os modelos explicativos existentes. Neste contexto, a superioridade do modelo univariado não deve ser interpretada como evidência de irrelevância das variáveis macroeconômicas para a compreensão

da inflação, mas sim como indicativo de que, para fins preditivos, a informação contida na trajetória histórica do próprio IPCA pode ser suficiente para gerar previsões precisas. Os resultados obtidos, mesmo com foco preditivo, demonstram que redes LSTM constituem uma ferramenta viável e promissora para complementar os modelos econométricos tradicionais utilizados. A capacidade destes modelos de capturar padrões não-lineares complexos na dinâmica inflacionária pode enriquecer o processo de tomada de decisão em política monetária, especialmente considerando que o modelo mais eficaz se baseou exclusivamente na trajetória histórica da inflação. Embora o estudo apresente limitações, e valores elevados de MAPE que indicam desafios na previsão pontual.

## Referências

- [1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT Press, Cambridge, 2016.
- [2] A. A. Palma and M. S. Portugal. Preferences of the central bank of brazil under the inflation targeting regime: estimation using a DSGE model for a small open economy. *Journal of Policy Modeling*, 33(4):567–586, 2011.
- [3] V. G. Machado and M. S. Portugal. Phillips curve in brazil: an unobserved components approach. *Estudos Econômicos*, 44(4):787–814, 2014.
- [4] M. S. Portugal, R. C. Madalozzo, and R. O. Hillbrecht. Inflation targeting in brazil: an empirical evaluation. *Revista de Economia Contemporânea*, 3(1):77–106, 1999.
- [5] M. S. Portugal, A. F. Barcellos, and V. Cerqueira. Modelos DSGE: uma resenha. *Revista Brasileira de Economia*, 63(1):21–48, 2009.
- [6] Marcos R. Castro, Solange N. Gouvea, André Minella, Rafael C. Santos, and Nelson F. Souza-Sobrinho. SAMBA: stochastic analytical model with a bayesian approach. *Brazilian Review of Econometrics*, 35(2):103–170, 2015.
- [7] C. J. Costa Junior, A. M. Teixeira, and M. F. Silva. DSGE para macroeconomia regional: uma aplicação para o estado de goiás. In *Anais do 50º Encontro Nacional de Economia*, Fortaleza, 2022. ANPEC.
- [8] Instituto de Pesquisa Econômica Aplicada. Previsões macroeconômicas e análise de conjuntura. *Carta de Conjuntura*, (58), 2023.
- [9] S. A. Raza, S. T. Jawaid, S. Afshan, and M. Z. A. Karim. Inflation prediction in emerging economies: machine learning and FX reserves integration for enhanced forecasting. *International Review of Economics & Finance*, 94:103704, 2024.
- [10] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [11] Uday Kamath, John Liu, and James Whitaker. *Deep Learning for NLP and Speech Recognition*. Springer Publishing Company, Incorporated, 1st edition, 2019. ISBN 3030145956.
- [12] Seungeun Chung, Jiyouon Lim, Kyoung Noh, Gague Kim, and Hyuntae Jeong. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors*, 19:1716, 2019. doi: 10.3390/s19071716.
- [13] Banco Central do Brasil. Taxa selic, 2024. URL <https://www.bcb.gov.br/controleinflacao/taxaselic>. Acesso em: 14 abr. 2024.