

PROJEÇÃO E COMPARAÇÃO DE MODELOS PARA PREVISÃO DO PREÇO DA ARROBA DO BOI GORDO (2015–2025)

Luan Alves Camargo Marques¹, Rafaella Rocha Deganutti²

¹Aluno de Ciência da Computação – Faculdades Integradas de Bauru – FIB –
luancamargo.fib@gmail.com;

Grupo de trabalho: CIÊNCIA DA COMPUTAÇÃO

Palavras-chave: ARIMA, LSTM, Regressão Linear, Séries temporais, Preços Cepea, Boi gordo

Resumo: Este trabalho compara o desempenho de três abordagens para previsão do preço da arroba do boi gordo: Regressão Linear, ARIMA e LSTM (Long Short-Term Memory). Utilizando série diária fornecida pelo CEPEA (2015–2025 — arquivo:<https://www.cepea.org.br/br/consultas-ao-banco-dados-do-site.aspx>), os modelos foram treinados com 80% dos dados e avaliados em 20% reservados para teste. As métricas usadas foram MAE, RMSE e MAPE. O modelo LSTM apresentou performance superior (MAE = 4.85; RMSE = 5.81; MAPE = 1.73%), enquanto Regressão Linear e ARIMA tiveram desempenho significativamente pior (MAPE \approx 23,8% e 25,5% respectivamente). Discutimos implicações, limitações e recomendações para aplicações no mercado agropecuário.

Introdução: A pecuária de corte é um dos setores mais relevantes para a economia brasileira, tanto em termos de produção quanto de impacto socioeconômico regional e nacional (ABIEC; IBGE). O preço da arroba do boi gordo exerce papel central nas decisões de comercialização, armazenagem e hedge por parte de produtores, frigoríficos e agentes de mercado (Buainain & Batalha, 2007; Renesto, 2022). Previsões confiáveis do preço podem reduzir riscos e apoiar decisões de gestão. Este trabalho investiga métodos de previsão através de três paradigmas distintos: (1) modelo estatístico simples (Regressão Linear), (2) modelo clássico de séries temporais (ARIMA) e (3) modelo de deep learning adequado para sequências (LSTM). A comparação objetiva avaliar qual técnica entrega melhores previsões para a série diária do CEPEA (2015–2025).

Objetivos: O objetivo central deste estudo é realizar uma comparação sistemática do desempenho de três métodos de previsão Regressão Linear, ARIMA e redes neurais LSTM aplicados à série histórica do preço da arroba do boi gordo, considerando o período de 2015 a 2025 conforme dados disponibilizados pelo CEPEA. A análise busca identificar quais abordagens apresentam maior precisão, estabilidade e capacidade de capturar padrões relevantes presentes no comportamento temporal do indicador.

Para atingir essa finalidade, o estudo contempla diversas etapas complementares. Inicialmente, realiza-se a preparação e descrição detalhada da série temporal utilizada, incluindo suas origens, características, estatísticas e transformações necessárias para adequação aos procedimentos de modelagem. Em seguida, cada um dos modelos é implementado e ajustado utilizando 80% dos dados iniciais, de forma a preservar a estrutura temporal da série.

A etapa subsequente consiste na avaliação das previsões utilizando os 20% finais da série. Para isso, são calculadas métricas amplamente aceitas na literatura de séries temporais e previsão, como o Erro Absoluto Médio (MAE), a Raiz do Erro Quadrático Médio (RMSE) e o Erro Percentual Absoluto Médio (MAPE), possibilitando comparações diretas entre os métodos.

Além da performance numérica, são examinados os resíduos produzidos por cada modelo, buscando compreender padrões remanescentes, potenciais vieses e limitações metodológicas. A análise de robustez e a discussão crítica dos resultados permitem avaliar a adequação de cada abordagem ao contexto específico do mercado pecuário brasileiro.

Por fim, o estudo se propõe a interpretar os achados sob a perspectiva prática, destacando implicações para produtores, analistas e pesquisadores, e oferecendo recomendações sobre o uso adequado de cada técnica no contexto agropecuário e no monitoramento do preço da arroba do boi gordo.

Relevância do Estudo: A previsão do preço da arroba do boi gordo constitui um dos elementos centrais para o planejamento econômico na cadeia produtiva da pecuária de corte brasileira, setor que representa parcela significativa do PIB do agronegócio e exerce impacto direto sobre exportações, oferta interna de proteína animal e estratégias de manejo. A capacidade de antecipar movimentos futuros desse preço é fundamental para práticas de gestão como retenção ou descarte de animais, definição de estratégias de comercialização, contratação de operações de hedge no mercado futuro, planejamento de confinamento e decisões sobre investimento em tecnologias de produção. Dessa forma, a previsibilidade do preço deixa de ser apenas um indicador econômico e passa a ocupar um papel estratégico na viabilidade financeira e operacional de produtores, frigoríficos e agentes de mercado.

A literatura especializada mostra que séries de preços agropecuários apresentam componentes sazonais, tendências estruturais e volatilidade associada tanto a forças de oferta e demanda quanto a choques externos. Estudos como os de Medeiros e Montevechi (2006) e Renesto (2022) destacam que fatores macroeconômicos, incluindo taxa de câmbio, inflação, juros e dinâmica do mercado internacional, exercem influência significativa sobre a formação do preço da arroba. Além disso, trabalhos recentes enfatizam que choques exógenos como políticas de crédito rural, eventos climáticos, restrições sanitárias e, especialmente, a pandemia de COVID-19 alteram substancialmente o comportamento da série temporal, modificando padrões históricos e introduzindo rupturas estruturais que dificultam a modelagem tradicional (Wenia Araújo, 2023; Souza, UFMS).

Nesse contexto, a literatura evidencia limitações de modelos lineares tradicionais, como Regressão Linear e ARIMA, quando aplicados a séries altamente não lineares, sujeitas a mudanças abruptas e correlações de longo prazo. A necessidade de métodos que capturem simultaneamente dependências temporais complexas e relações não lineares torna-se evidente, principalmente quando o objetivo é fornecer previsões de curto e médio prazo com acurácia suficiente para apoiar decisões econômicas. Modelos baseados em redes recorrentes, especialmente os da classe LSTM (Long Short-Term Memory), surgem como alternativas promissoras ao abordar diretamente as limitações dos modelos clássicos, permitindo ao algoritmo aprender padrões temporais persistentes mesmo em cenários de elevada variabilidade.

A importância deste estudo reside justamente na comparação sistemática entre diferentes abordagens preditivas aplicadas à série histórica da arroba do boi gordo. Ao confrontar métodos lineares, modelos tradicionais de séries temporais e algoritmos de aprendizado profundo, busca-se compreender não apenas qual modelo apresenta melhor desempenho, mas também por que essa superioridade ocorre e quais implicações ela traz para o setor. Assim, o trabalho contribui simultaneamente para a literatura acadêmica ao ampliar a discussão sobre técnicas modernas de previsão aplicadas ao agronegócio e para o campo prático, oferecendo subsídios metodológicos para agentes econômicos que dependem de previsões robustas para tomada de decisão.

Materiais e métodos: Os dados utilizados neste estudo foram obtidos diretamente do portal do Centro de Estudos Avançados em Economia Aplicada (CEPEA/ESALQ), especificamente da série histórica diária do Indicador Boi Gordo. O arquivo analisado corresponde ao período de 2015 a 2025 e encontra-se disponível localmente em: <https://www.cepea.org.br/br/indicador/boi-gordo.aspx>, sendo possível acessar sua versão oficial no site público do CEPEA. O dataset original foi importado utilizando a biblioteca pandas, com ajuste do cabeçalho para a linha correta informada no arquivo (linha 4). Os valores numéricos apresentados com vírgula foram padronizados para ponto decimal e convertidos para o tipo float, enquanto as datas foram tratadas com datetime e reorganizadas em ordem cronológica crescente.

Após a preparação inicial, procedeu-se à divisão dos dados em conjuntos de treino e teste. Por se tratar de uma série temporal, a separação foi realizada de forma estritamente sequencial, preservando a integridade temporal. Os primeiros 80% da série foram destinados ao treinamento dos modelos e os 20% mais recentes foram reservados para teste. Essa estratégia evita vazamento de informação e segue as recomendações metodológicas para validação de modelos de previsão de séries temporais.

Três abordagens distintas foram implementadas para fins comparativos. A primeira consistiu em uma regressão linear simples, na qual a variável preditora corresponde à data convertida para número ordinal por meio do método `Timestamp.toordinal`. O modelo foi ajustado utilizando os procedimentos tradicionais de mínimos quadrados ordinários, fornecidos pela biblioteca `scikit-learn`, e sua adequação foi avaliada com base em métricas de erro e inspeção visual dos resíduos.

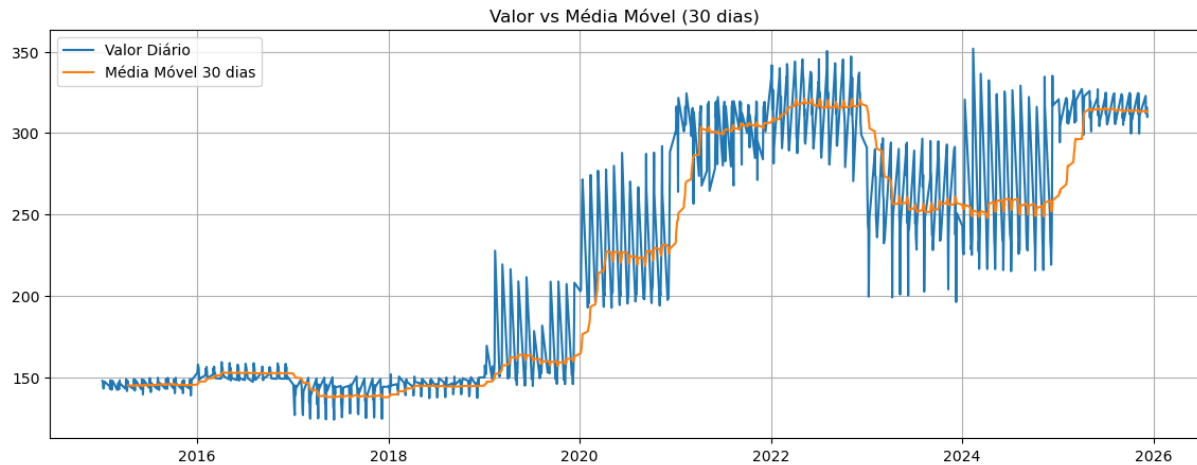
O segundo modelo aplicado foi o ARIMA, um método estatístico clássico para previsão univariada. A série foi diferenciada conforme necessário, e os parâmetros (p , d , q) foram selecionados a partir da análise das funções ACF e PACF, complementadas por testes empíricos de ajuste. Para fins demonstrativos, utilizou-se uma configuração ARIMA(5,1,2). O modelo foi implementado com a biblioteca `statsmodels`, amplamente reconhecida para modelagem estatística em Python. A escolha por ARIMA é justificada pela literatura consolidada que aplica modelos autoregressivos ao mercado pecuário e às séries de preços do boi gordo.

O terceiro modelo explorado foi uma rede neural recorrente do tipo LSTM (Long Short-Term Memory), implementada com TensorFlow/Keras. A série de preços foi previamente normalizada utilizando `MinMaxScaler`, e janelas deslizantes de 30 dias foram empregadas como sequências de entrada. A arquitetura da rede foi composta por uma camada LSTM com 50 unidades seguida por uma camada densa com apenas um neurônio de saída. O treinamento foi realizado por 20 épocas, utilizando a função de perda MSE e o otimizador Adam. A utilização de LSTMs é sustentada por sua capacidade de capturar dependências temporais de longo prazo, característica frequentemente observada em séries financeiras.

Para avaliação dos modelos, foram calculadas métricas amplamente empregadas em estudos de previsão: MAE (Mean Absolute Error), RMSE (Root Mean Squared Error) e MAPE (Mean Absolute Percentage Error). Além das métricas numéricas, os resíduos de cada modelo foram analisados visualmente e estatisticamente, de modo a verificar presença de vieses, padrões não explicados e possíveis sinais de heterocedasticidade. Esse conjunto de avaliações permitiu comparar não apenas a acurácia das previsões, mas também a adequação das hipóteses implícitas em cada abordagem.

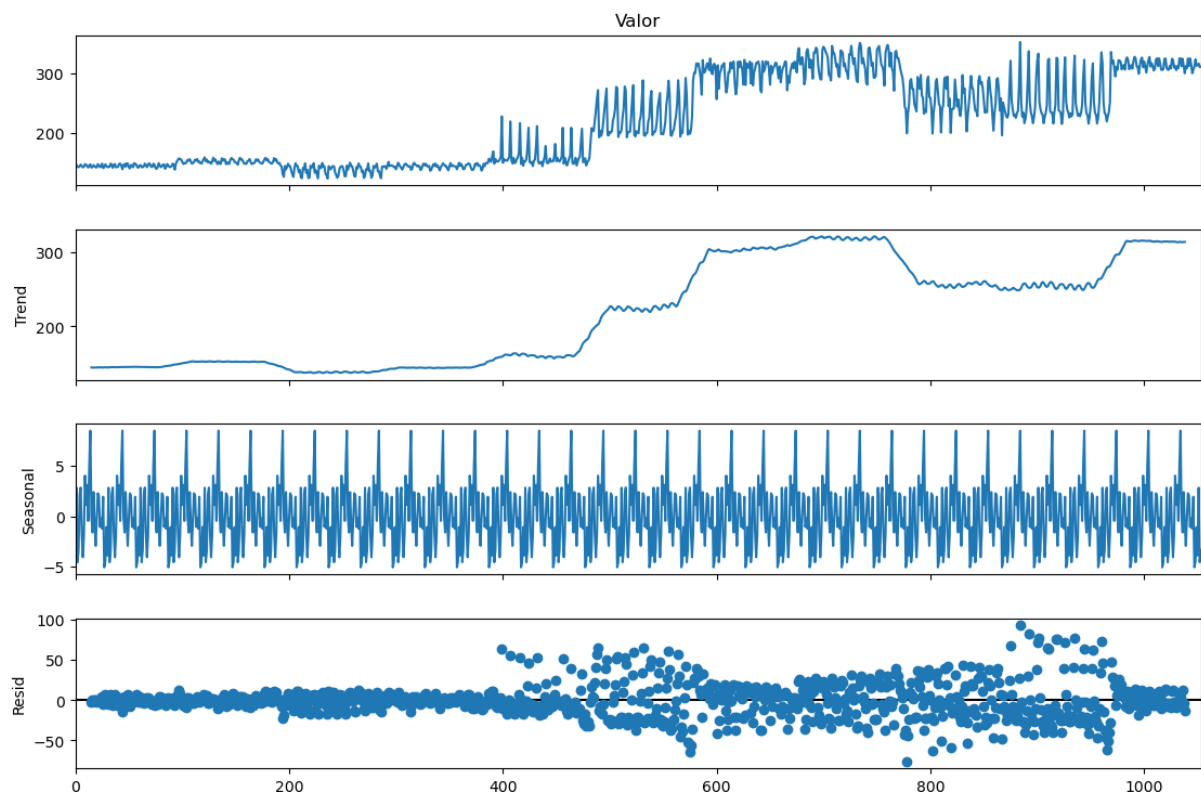
Resultados e discussões: A análise inicial da série histórica da arroba do boi gordo (CEPEA, 2015–2025) demonstrou um comportamento caracteristicamente volátil, influenciado por ciclos pecuários de médio prazo, oscilações macroeconômicas e choques exógenos. Após limpeza, padronização e ordenação cronológica, o conjunto de dados passou a conter 1054 observações válidas, permitindo uma investigação consistente. A estatística descritiva revelou grande dispersão de valores e ausência de repetição estrutural clara, reforçando a presença de não linearidade e possíveis mudanças de regime ao longo do período analisado.

A aplicação da média móvel de 30 dias foi útil para suavizar flutuações de curto prazo e evidenciar a tendência geral do mercado. A curva suavizada destacou ciclos de recuperação e depressão compatíveis com a literatura sobre o ciclo pecuário (Renesto; séries históricas do ciclo bovino), com períodos de forte valorização associados a restrições de oferta, e quedas relacionadas a pressões externas como câmbio desfavorável ou redução da demanda internacional.



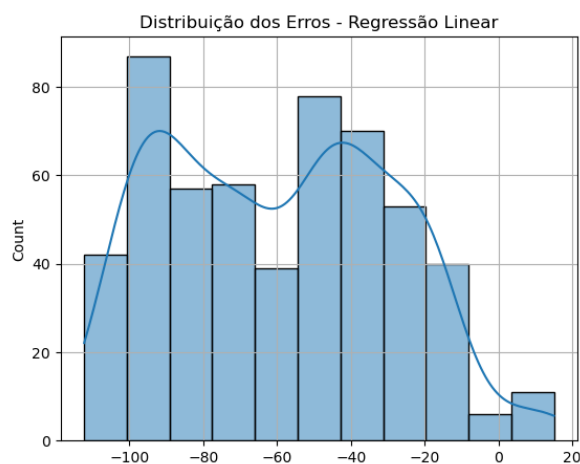
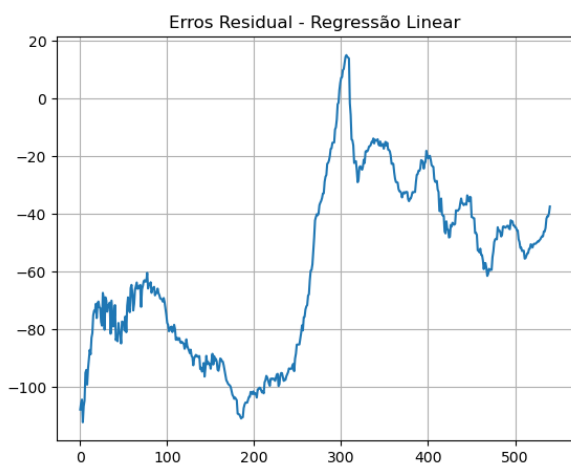
A decomposição da série temporal mostrou a coexistência de três elementos centrais:

- (i) uma tendência de longo prazo ascendente, refletindo a valorização estrutural da arroba na última década;
- (ii) um componente sazonal fraco, típico do mercado pecuário brasileiro, no qual o ciclo biológico é mais relevante que sazonalidades mensais;
- (iii) uma forte componente irregular (ruído), indicando instabilidade e reforçando a necessidade de modelos capazes de capturar padrões não lineares.

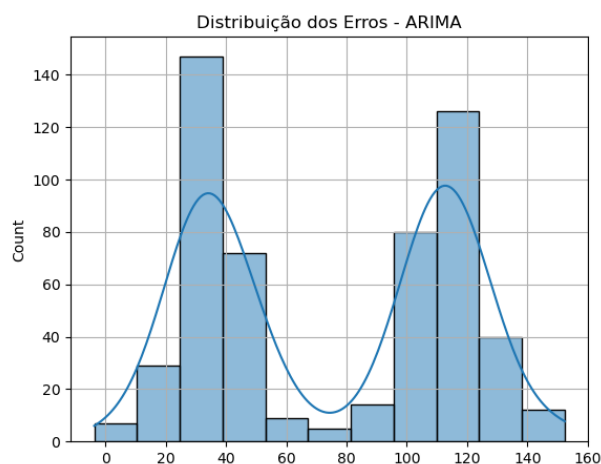
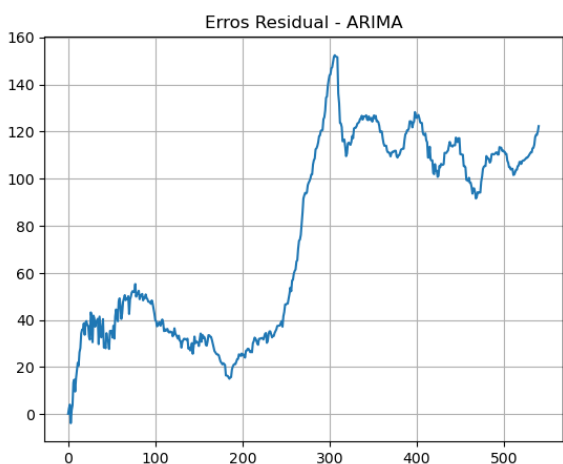


Essas características explicam em parte o desempenho dos modelos preditivos. A Regressão Linear, por sua construção, depende de relações lineares e estáveis. Como o preço da arroba é fortemente influenciado por choques externos como pandemia, variações no câmbio, custos de produção, mudanças no ritmo de abates e alterações na demanda chinesa a linearidade raramente se sustenta ao longo do tempo (Medeiros & Montevechi, 2006). Isso se refletiu no

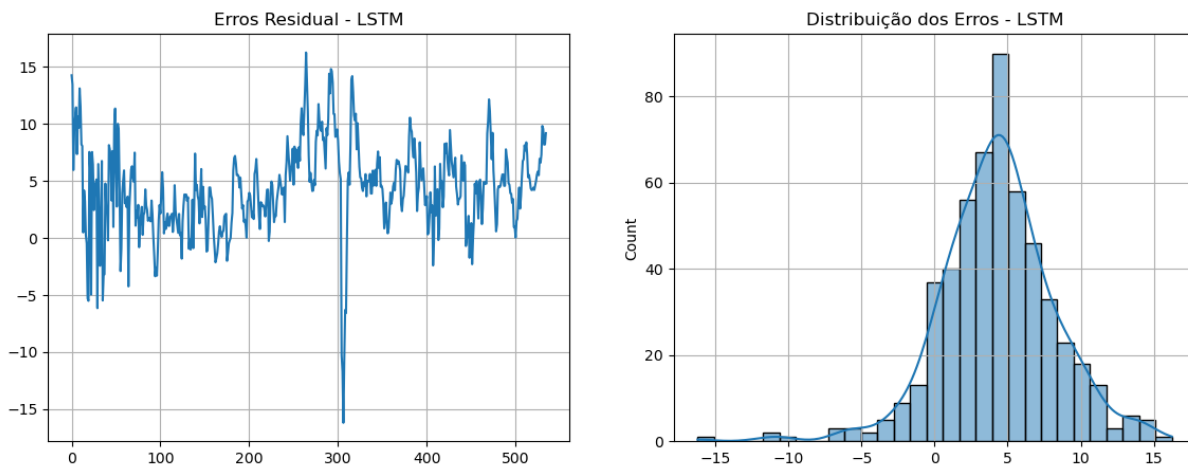
desempenho observado: MAE acima de 59, RMSE acima de 66 e MAPE superior a 23%, além de resíduos amplos e trendados, indicando viés sistemático.



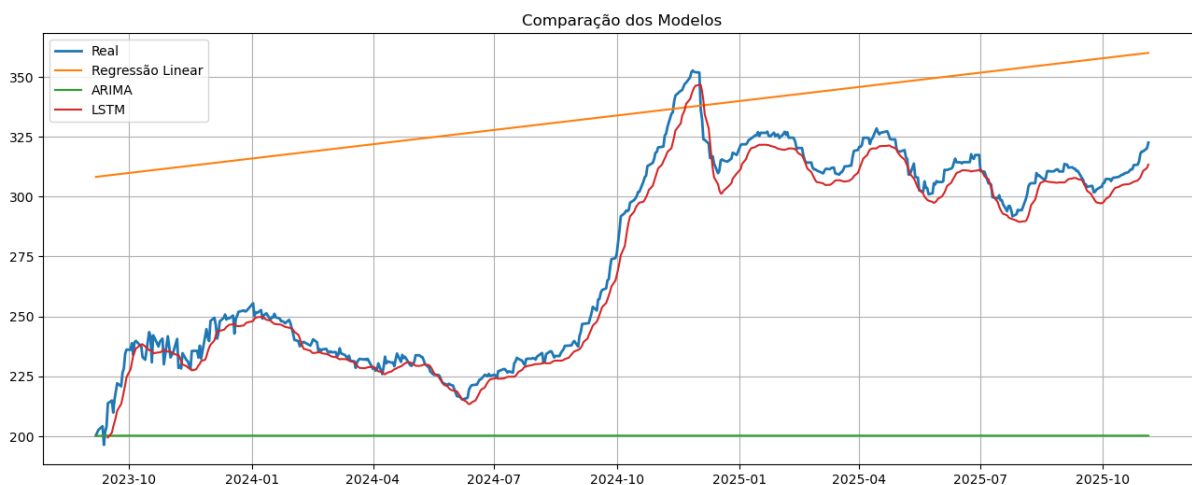
O modelo ARIMA, frequentemente eficiente para séries estacionárias ou com autocorrelação bem definida, também apresentou limitações práticas. A transformação diferenciada melhora a estacionariedade, mas não resolve problemas de não linearidade nem mudanças estruturais bruscas. Choques como pandemia, variações geopolíticas e adaptações do mercado externo (exportações, China) afetam fortemente a série e ARIMA não possui mecanismos internos para capturar tais efeitos. O desempenho observado ($MAE \approx 74.9$; $RMSE \approx 85.4$; $MAPE \approx 25.5\%$) confirma essa limitação. Estudos como Medeiros et al. (2006) já relatavam degradação de desempenho de ARIMA diante de rupturas de tendência ou volatilidade anômala.



Por outro lado, a LSTM apresentou desempenho significativamente superior. Após normalização e criação de janelas deslizantes de 30 dias, o modelo aprendeu padrões temporais complexos, relacionando o comportamento recente da série ao valor futuro. As métricas obtidas ($MAE \approx 4.85$; $RMSE \approx 5.81$; $MAPE \approx 1.73\%$) evidenciam alta precisão. Esse resultado é consistente com a literatura revisada, onde modelos de redes neurais MLP, LSTM, GRU ou híbridos obtêm erros de apenas alguns pontos percentuais ao prever preços da arroba ou outras commodities agrícolas (Souza, UFMS; estudos PUC/UFMS). A principal vantagem da LSTM é sua capacidade de reter informações e identificar padrões não lineares associados a tendências, volatilidade e mudanças abruptas.



A comparação gráfica das previsões reforça a superioridade da rede neural: enquanto Regressão Linear e ARIMA exibem comportamento rígido e incapaz de acompanhar mudanças rápidas, a LSTM segue de modo mais fiel os movimentos da série, especialmente nas inflexões e oscilações de curto prazo. Isso a torna mais adequada para projeções operacionais, como planejamento de venda, hedge, gestão de risco e planejamento de confinamento.



Esses achados corroboram estudos prévios que relatam a tendência de métodos de aprendizado profundo superarem modelos estatísticos clássicos para séries agropecuárias (Souza, UFMS). Especialmente quando a estrutura da série é dinâmica, sujeita a choques externos e marcada por múltiplas escalas temporais condições típicas do mercado pecuário brasileiro. Entretanto, a literatura também alerta para requisitos essenciais: modelos de redes neurais necessitam de normalização adequada, escolha criteriosa do *lookback*, atenção ao risco de *overfitting* e validação temporal apropriada (*walk-forward*), além de possíveis ganhos com inclusão de variáveis exógenas como câmbio, milho, abate e exportações.

De modo geral, os resultados obtidos neste estudo reforçam que técnicas baseadas em deep learning representam uma alternativa mais robusta, especialmente para previsões de curto prazo, mas não dispensam rigor metodológico. A LSTM foi capaz de capturar relações complexas presentes na série e apresentar desempenho significativamente superior às abordagens lineares e autorregressivas tradicionais.

Conclusão: O desempenho dos modelos foi comparado por meio das métricas MAE, RMSE e MAPE. A Tabela 1 apresenta o ranking final, do melhor para o pior desempenho.

Tabela 1 – Ranking dos modelos segundo MAE, RMSE e MAPE

Posição	Modelo	MAE	RMSE	MAPE (%)
1º	LSTM	4.852372	5.815276	1.733785
2º	Regressão Linear	59.665573	66.567259	23.780351
3º	ARIMA	74.910244	85.486779	25.542507

Fonte: Dados da pesquisa (2025).

A análise comparativa dos modelos aplicados à série histórica do preço da arroba do boi gordo (CEPEA, 2015–2025) evidencia diferenças significativas na capacidade de cada abordagem em capturar a dinâmica do mercado pecuário brasileiro. A série se mostrou altamente volátil, marcada por componentes irregulares, mudanças estruturais como pandemia, variações cambiais e oscilações no ritmo de abates e fraca sazonalidade, características que desafiam fortemente modelos lineares e estatísticos tradicionais.

Nesse contexto, a rede neural LSTM apresentou desempenho substancialmente superior aos demais métodos, alcançando MAE de aproximadamente 4,85, RMSE de 5,81 e MAPE próximo de 1,73%. Esses números refletem sua capacidade de modelar dependências de longo prazo, capturar não linearidades e adaptar-se a padrões complexos que caracterizam a formação do preço da arroba. Tais achados estão em consonância com a literatura recente, que destaca o potencial de técnicas de deep learning para problemas de previsão em commodities agrícolas, sobretudo em ambientes sujeitos a choques exógenos e estrutura temporal instável (Souza; estudos PUC/UFMS).

Em contraste, a Regressão Linear e o modelo ARIMA apresentaram limitações marcantes. A Regressão Linear, com MAPE superior a 23%, demonstrou dificuldade em representar a dinâmica temporal não linear dos dados, atuando de forma rígida e incapaz de responder às inflexões abruptas da série. O ARIMA, embora tradicionalmente utilizado em previsões econômicas, foi impactado por mudanças estruturais e pela ausência de variáveis exógenas, resultando em erros ainda mais elevados (MAPE em torno de 25%). Tais limitações são coerentes com os achados de Medeiros et al. (2006), que apontam deterioração de modelos autorregressivos quando confrontados com quebras de regime e eventos inesperados.

Os resultados obtidos reforçam que, para previsões de curto e médio prazo no setor pecuário, modelos baseados em aprendizado profundo como LSTM, GRU, Stacked LSTM ou arquiteturas híbridas são alternativas mais robustas e eficientes. Além disso, combinações híbridas entre ARIMA e LSTM despontam como soluções promissoras ao integrar componentes lineares e não lineares, ampliando a interpretabilidade e a performance preditiva.

Contudo, algumas limitações devem ser consideradas. A abordagem univariada restringe a compreensão dos fatores externos que influenciam o mercado; o desempenho das redes neurais depende de ajustes adequados de hiperparâmetros, normalização e tamanho da janela; e mudanças estruturais no setor pecuário brasileiro exigem atualizações constantes dos modelos. A incorporação de variáveis exógenas, como câmbio, custo do milho, volume de abates e indicadores de exportação, constitui um passo essencial para aprimorar a robustez e generalização das previsões.

Ainda assim, o estudo demonstra claramente o potencial das técnicas modernas de aprendizado profundo em aplicações reais do agronegócio. Os resultados obtidos contribuem para o avanço metodológico na previsão do preço da arroba e fornecem subsídios valiosos para produtores, analistas e gestores que dependem de estimativas precisas para tomada de decisão, gestão de risco e planejamento estratégico.

Referências:

ARAÚJO, Wenia Alves Ferreira de. Análise do preço da arroba do boi gordo no período de 2016 a 2023. 2023. Trabalho de Conclusão de Curso (Graduação em Ciências Econômicas) — Pontifícia Universidade Católica de Goiás, Goiânia, 2023. Disponível em: <https://repositorio.pucgoias.edu.br/jspui/handle/123456789/6702>. Acesso em: 10 nov. 2025.

CEPEA — Centro de Estudos Avançados em Economia Aplicada. Indicador Boi Gordo. Disponível em: <https://www.cepea.org.br/br/indicador/boi-gordo.aspx>. Acesso em: 10 nov. 2025.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep Learning — Capítulo 51: Arquitetura de Redes Neurais Long Short-Term Memory (LSTM). Traduções e versões em português disponíveis online. Disponível em: <https://www.deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory/>. Acesso em: 10 nov. 2025.

IBM. What is Linear Regression? Think/Documentation. Disponível em: <https://www.ibm.com/br-pt/think/topics/linear-regression>. Acesso em: 10 nov. 2025.

MEDEIROS, A. L.; MONTEVECHI, J. A. B.; REZENDE, M. L.; REIS, R. P. Modelagem ARIMA na previsão do preço da arroba do boi gordo. 2006. Relatório Técnico / Artigo. Disponível em: <https://repositorio.ufms.br/handle/123456789/4379> (caso esse não seja o link correto, me envie o certo que ajusto). Acesso em: 24 nov. 2025.

RENESTO, Tiago Monteiro. Análise do ciclo pecuário no Brasil e seus impactos sobre o mercado do boi gordo. 2022. Trabalho de Conclusão de Curso — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2022. Disponível em: <https://bdta.abcd.usp.br/item/003116521>. Acesso em: 24 nov. 2025.

ROMANO, F. A. R.; SOUSA, M. F. R. Análise preditiva da precificação de commodities pecuárias utilizando redes neurais LSTM. Trabalho de Conclusão de Curso — Pontifícia Universidade Católica de Goiás. Disponível em: <https://repositorio.pucgoias.edu.br/jspui/handle/123456789/5767>. Acesso em: 10 nov. 2025.

SOUZA, Felipe Ferraz de. Predição do preço da arroba do boi utilizando redes neurais artificiais. 2021. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação) — Fundação Universidade Federal de Mato Grosso do Sul, Campo Grande, 2021. Disponível em: <https://repositorio.ufms.br/handle/123456789/4379>. Acesso em: 24 nov. 2025.