

APLICAÇÃO DE REDES NEURAIIS RECORRENTES NA PREVISÃO DE PREÇOS DE FRUTAS NO SETOR AGRÍCOLA BAIANO

 <https://doi.org/10.56238/arev7n1-073>

Data de submissão: 06/12/2024

Data de publicação: 06/01/2025

Iara Fábica Castro Viana Silva

Graduada em Gestão da Tecnologia da Informação (IFBAIANO)
Instituto Federal Baiano (IFBAIANO)
E-mail: zziarafcastro@gmail.com

Caio Henrique Rodrigues Carvalho

Mestre em Engenharia Elétrica (UFPI)
Instituto Federal Baiano (IFBAIANO)
E-mail: caio.carvalho@ifbaiano.edu.br

Tiago do Carmo Nogueira

Doutor em Engenharia Elétrica e de Computação (UFG)
Instituto Federal Baiano (IFBAIANO)
E-mail: tiago.nogueira@ifbaiano.edu.br

Danyelee de Oliveira Santana

Mestranda em Ciência da Computação (UEFS)
Instituto Federal Baiano (IFBAIANO)
E-mail: danyelee.santana@ifbaiano.edu.br

Gelson da Cruz Junior

Doutor em Engenharia Elétrica (Unicamp)
Universidade Federal de Goiás (UFG)
E-mail: gcruzjr@ufg.br

Cássio Dener Noronha Vinhal

Doutor em Engenharia Elétrica (Unicamp)
Universidade Federal de Goiás (UFG)
E-mail: vinhal@ufg.br

Matheus Rudolfo Diedrich Ullmann

Doutorando em Engenharia Elétrica e de Computação (UFG)
Instituto Federal da Bahia (IFBA)
E-mail: matheusullmann@ifba.edu.br

RESUMO

Este estudo investiga o uso de Redes Neurais Recorrentes (RNN), especificamente LSTM e GRU, na previsão de preços de frutas cultivadas pela agricultura familiar na Bahia, com base em 1.883 registros de banana nanica, banana prata e mamão formosa. O modelo LSTM apresentou o melhor desempenho, com resultados de RMSE variando entre 0,186 e 0,606, MAE entre 0,142 e 0,483, MAPE entre 7,286 e 16,624 e MSE entre 0,035 e 0,367 para as frutas analisadas. Destaca-se o potencial das RNN no apoio a tomadas de decisões no setor agrícola, com propostas de trabalhos futuros que incluem a incorporação de variáveis exógenas e o desenvolvimento de uma plataforma gratuita para pequenos produtores.

Palavras-chave: Agricultura. Séries Temporais. Previsão de Preços. Redes Neurais Recorrentes.

1 INTRODUÇÃO

A Bahia é o segundo maior estado produtor de frutas frescas do Brasil, com 664 mil hectares dedicados ao cultivo, segundo dados da Federação da Agricultura e Pecuária da Bahia - FAEB (2023). Nesse contexto, o estado se destaca como o 2º maior produtor de bananas do país, com uma produção de 866 mil toneladas (SEAGRI, 2023). A banana é cultivada em todas as regiões da Bahia, sendo a cidade de Bom Jesus da Lapa um dos principais polos produtores. Outro destaque é o mamão, que embora seja cultivado em todo o território nacional, tem mais de 50\% de sua produção concentrada na Bahia nas últimas décadas, com a cidade de São Félix do Coribe como principal centro produtor (LANDAU; SILVA, 2020).

A agricultura familiar é uma parte essencial da produção agrícola baiana, respondendo por cerca de 298 mil toneladas de produtos (CODEVASF, 2024). Caracterizada pelo cultivo em pequenas propriedades rurais, essa modalidade não só fornece alimento para os próprios agricultores, mas também contribui significativamente para a economia local. No entanto, os agricultores familiares enfrentam desafios como fragilidades socioeconômicas e baixa tecnologia produtiva, o que pode impactar negativamente na produção de frutas como a banana e o mamão (AQUINO; ALVES; VIDAL, 2020).

Para que a produção agrícola seja comercializada de maneira viável, é crucial que os preços de venda cubram os custos e evitem prejuízos. Nesse cenário, a contabilidade desempenha um papel fundamental, ajudando os produtores rurais a tomar decisões informadas que aumentem os resultados econômicos (ASSIS et al., 2021). A falta de um tratamento contábil adequado pode levar à desinformação sobre os custos, impactando negativamente a projeção de lucros (DUTRA, 2020).

A instabilidade dos preços dos produtos rurais, frequentemente causada pela oferta sazonal ao longo do ano, gera desafios significativos na gestão financeira dos produtores familiares. A sazonalidade, dividida em longo e curto prazo, é um fator essencial na previsão e gestão agrícola (YOO; OH, 2020). Enquanto a sazonalidade de longo prazo está relacionada a mudanças permanentes na oferta e demanda, a de curto prazo envolve alterações temporárias, como variações na temperatura e nas condições de precipitação (CHU et al., 2020).

Diante desses desafios, informações precisas e oportunas sobre a dinâmica sazonal são cada vez mais necessárias para otimizar a gestão dos sistemas de cultivo e detectar anomalias sazonais (BOSCHETTI et al., 2015). A sazonalidade resulta em variações de preços que, muitas vezes, causam prejuízos econômicos (RODRIGUES et al., 2021). Portanto, é fundamental que os produtores tenham controles financeiros eficazes para avaliar e decidir sobre os preços de venda, mesmo que estes sejam influenciados pelo mercado (PEREIRA; SANTOS, 2022).

Além disso, comerciantes também enfrentam dificuldades na formulação dos preços agrícolas devido às variações na oferta e demanda, às mudanças climáticas e ao processo de cultivo (OKTOVIANY; KNOBLOCH; KORN, 2021). Nesse contexto, a demanda por métodos de previsão da sazonalidade de produtos agrícolas tem crescido (LIU et al., 2021; SUN et al., 2023b). Nessa circunstância, a Agricultura Digital, que utiliza dados e abordagens tecnológicas para melhorar a produtividade, tem se mostrado uma ferramenta promissora (KURUMATANI, 2018). Ela processa dados gerados no campo agrícola (KULBA; MEDENNIKOV, 2020), essenciais para o funcionamento dos modelos de previsão (YUAN; LING, 2020), permitindo avanços significativos na compreensão e solução de problemas no setor.

A previsão de séries temporais, que consiste em analisar sequências de observações ao longo do tempo, como dias, meses, trimestres ou anos (RAHMAN et al., 2023) para realizar prognósticos, utilizando métodos estatísticos e modelagem (XU; HSU, 2022), tem se destacado como uma técnica poderosa na detecção de anomalias sazonais. Deste modo, as Redes Neurais Artificiais (RNA) têm atraído grande atenção dos pesquisadores nessa área de previsão de séries temporais (LIU et al., 2021), sendo aplicadas para prever os preços dos produtos no mercado agrícola, o que contribui para a estabilização da oferta e demanda (YOO; OH, 2020). As RNAs oferecem uma ampla gama de possibilidades, superando métodos estáticos tradicionais (YUAN; LING, 2020), e são particularmente eficazes em capturar padrões não lineares, que podem existir sem uma razão aparente (OKTOVIANY; KNOBLOCH; KORN, 2021; KURUMATANI, 2018).

Entretanto, quando se trata de previsão com fatores sazonais, as Redes Neurais Recorrentes (RNN) são geralmente preferidas. Sua capacidade de “memória” na transmissão de dados permite capturar padrões sequenciais, como a sazonalidade, de maneira mais eficaz, superando as limitações de outros modelos (LIU et al., 2021). Assim, as RNNs se destacam como uma ferramenta essencial para a previsão eficiente de preços agrícolas, facilitando tanto o treinamento de dados quanto a tomada de decisões baseadas em conjuntos de dados complexos.

A previsão precisa dos preços pode capacitar a rede de produção agrícola a tomar decisões estratégicas para lidar com as variações de valor (REDDY et al., 2022). Com base nisso, este trabalho propõe o uso de modelos baseados em Redes Neurais Recorrentes para desenvolver uma metodologia eficaz de previsão de preços para algumas das principais frutas cultivadas no estado da Bahia, como a banana e o mamão. O intuito é oferecer suporte aos pequenos produtores agrícolas na tomada de decisões econômicas e na otimização de suas operações, auxiliando na gestão financeira e na redução de riscos.

Para alcançar esse objetivo, os modelos *Long Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU) foram selecionados e avaliados usando diversas métricas, como o Erro Médio Absoluto (MAE), Erro Absoluto Percentual Médio (MAPE), Erro Quadrático Médio (MSE) e Raiz do Erro Quadrático Médio (RMSE), em uma base de dados disponível no repositório online do Centro de Estudos Avançados em Economia Aplicada (CEPEA) da Universidade de São Paulo (USP). Essa abordagem visa identificar o modelo mais adequado para o cenário agrícola baiano.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados sobre técnicas de aprendizado de máquina na previsão de preços de produtos agrícolas, a Seção 3 apresenta a metodologia utilizada neste trabalho, a Seção 4 apresenta os resultados e discussões e a Seção 5 aborda as conclusões deste estudo e os trabalhos futuros.

2 TRABALHOS RELACIONADOS

Esta seção apresenta trabalhos do estado da arte, dos últimos 4 anos, que utilizaram Redes Neurais Artificiais em previsões de preços agrícolas para a descoberta de informações úteis. Os trabalhos foram escolhidos mediante a *string* de busca: (“*agriculture*” or “*agricultural time series analysis*”) and (“*time series data*”) and (“*price prediction*”) and (“*seasonality*”), nas bases: IEEE Xplore, ScienceDirect, Google Acadêmico, ACM Digital Library e SpringerLink. Como resultado, foram coletados 9 artigos de autores de diferentes países, fato que mostrou que a previsão de preços na área agrícola é um assunto discutido em âmbito global.

A previsão de preços é um método que requer a taxa de amostra de um período. Pesquisas recentes nessa área podem ajudar uma empresa/organização a alcançar seu objetivo e cumprir sua meta. Conseqüentemente, Santos, Maciel e Ballini (2020) discutem como o aumento dos preços resulta numa maior volatilidade nos mercados agrícolas, o que gera um risco maior de ganhar ou perder dinheiro ao negociar essas commodities (produtos primários com baixo teor de industrialização). Essas abordagens utilizam diferentes métodos para realizar a previsão, além de utilizar critérios de avaliação para a precisão geral, conservação de características de estatísticas e comparação entre algoritmos (KURUMATANI, 2020).

Kurumatani (2020) comparou os modelos LSTM, GRU e Rede Neural Recorrente Simples (SRNN), em conjunto com os métodos *time-alignment of time point forecast* (TATP) e *direct future time series forecast* (DFTS), para a previsão de preço do repolho, tomate e alface da China. A base utilizada continha cerca de 470 valores semanais de preços, demanda, preferência do consumidor e clima. Os modelos foram avaliados pelas métricas de erro, MAPE e RMSPE (sendo 'P' o indicativo de probabilidade). A LSTM obteve o melhor desempenho com RMSPE de 0,313 e MAPE de 0,240.

Yuan e Ling (2020) desenvolveram uma aplicação para previsão de preços com os modelos, ARIMA, LSTM, SVR, Prophet e XGBoos, sendo avaliados pelo métrica MSE. Foram utilizados dados semanais de preços do tomate, frango e pimenta, juntamente com fatores de temperatura, umidade e preço do petróleo bruto. O LSTM obteve o melhor resultado com MSE de 0,304.

Sabu e Kumar (2020) buscaram avaliar as previsões de preço das nozes de Querala, Índia. Foram analisados cerca de 210 dados mensais, comparando-se os métodos SARIMA, Holt-Winter e LSTM, e utilizando a RMSE para avaliação. O LSTM obteve o melhor resultado para a previsão dos preços das nozes, atingindo RMSE de 7,2780.

Chu et al. (2020) avaliaram diferentes métodos de aprendizado de máquina para lidar com o desafio de previsões de preços da uva chinesa em um cenário de dados escassos. Os modelos comparados incluíram *Ensemble Empirical Mode Decomposition* (EEMD), *Least Squares Support Vector Machine* (LSSVM), *Empirical Mode Decomposition-Reconstruction-Addition* (EEMD-R-ADD), *Support Vector Machine* (SVM), *Backpropagation Neural Network* (BPNN) e *Extreme Learning Machine* (ELM). Para avaliar a eficácia desses modelos, foram utilizados critérios de erro como MAE, RMSE. Os autores destacaram que, devido às peculiaridades da uva, que apresenta picos de preço causados por sua baixa resistência ao armazenamento, transporte de longa distância e características específicas de cultivo, o método EEMD-R-ADD obteve o melhor desempenho, com MAE de 0,089, RMSE de 0,114.

Zheng et al. (2020) propuseram um sistema de previsão de preços de mercado combinando o método *Wavelet* com uma rede neural *Multi Layer Perceptron* (MLP), com o objetivo de mitigar as flutuações anormais nos preços agrícolas chineses. Usando dados mensais de espinafre, repolho, tomate, pimentão e batata, a precisão do modelo foi avaliada pela MAE, MAPE e RMSE. Embora os desvios entre os preços previstos e reais fossem pequenos, os autores observaram que a frequência e amplitude da volatilidade dos preços podem influenciar as previsões do modelo. A previsão do preço da batata apresentou os melhores resultados, com MAE de 0,083, MAPE de 4,54 e RMSE de 0,10.

Oktoviany, Knobloch e Korn (2021) desenvolveram um modelo de previsão de preços do milho para apoiar a avaliação de riscos em sistemas de gestão e auxiliar na tomada de decisões. Utilizando dados semanais de preços, temperatura, produção, importação, consumo e exportação nos Estados Unidos, Brasil e Argentina, compararam os modelos *K-Nearest Neighbors* (K-NN) e *Random Forest*. Os resultados foram calibrados com o modelo de Monte Carlo para simular cenários de preços, avaliando a precisão com MAE, RMSE e MAPE. O modelo KNN apresentou a melhor performance, com MAE de 0,0779, RMSE de 0,1041 e MAPE de 1,5192.

Ozden (2023) investigou o método mais adequado para prever múltiplas medidas dos preços de cebola, batata e alho na Turquia, utilizando dados diários de preço e volume comercial desses produtos. Foram avaliados os modelos *Convolutional Neural Network* (CNN), LSTM e *Random Forest*, com os resultados sendo analisados pelas métricas MAE e RMSE. A CNN destacou-se apresentando o melhor desempenho, com MAE de 0,047 e RMSE de 0,070.

Sun et al. (2023a) propuseram um modelo de previsão híbrido, baseado na Decomposição de Modo Variacional (VMD), EEMD e LSTM, para lidar com os significativos erros de previsão causados pelas flutuações nos preços da carne suína, cebolinha chinesa, cogumelos shitake e couve-flor. Foram utilizados dados semanais de mercados atacadistas na China. O método, denominado VMD–EEMD–LSTM, foi comparado com outros modelos, como *Random Forest*, ELM, LSTM, e EEMD–LSTM, sendo avaliados pelas métricas MAE, MAPE e RMSE. O VMD–EEMD–LSTM apresentou o melhor desempenho com MAE de 0,2263, MAPE de 0,0584 e RMSE de 0,835, superando os demais modelos avaliados.

Visando melhorar a previsão do preço da lentilha na Índia, Ray et al. (2023), projetaram um modelo híbrido ARIMA-LSTM. O modelo ARIMA foi utilizado para estimar o efeito *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC), critérios que quantificam o quão bem o modelo se ajusta aos dados. O LSTM foi empregado para o treinamento. Para o cenário de comparação, foram escolhidos os modelos ARIMA, GARCH, LSTM, aplicando-se as métricas de avaliação RMSE e MAPE. O modelo ARIMA-LSTM os melhores resultados, com RMSE de 6,290 e MAPE de 2,886.

A Tabela 1 apresenta um resumo das principais informações sobre os trabalhos revisados no estado da arte. Esses estudos revelam que técnicas de Aprendizado de Máquina, mais especialmente Redes Neurais Artificiais, têm sido amplamente aplicadas na previsão de preços em diversos cenários agrícolas. Com base nesses avanços, a próxima seção abordará a metodologia proposta neste trabalho, detalhando os modelos e técnicas utilizados para desenvolver uma abordagem eficaz na previsão de preços das frutas cultivadas no estado da Bahia.

Tabela 1. Resumo dos Trabalhos Relacionados.

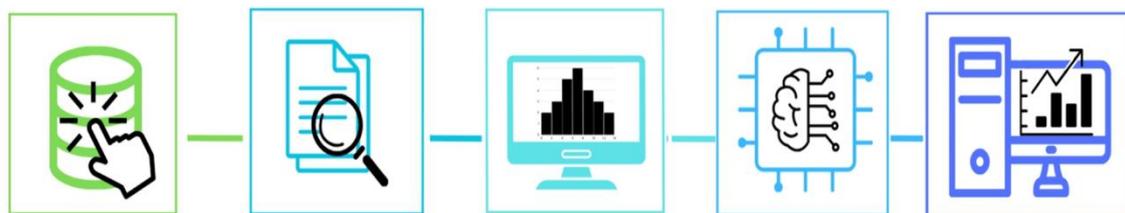
Trabalho	Previsão de Preço	Técnica	Métrica(s)
Kurumatani (2020)	Repolho, tomate e alface	LSTM	RMSPE (0,313); MAPE (0,240)
Yuan e Ling (2020)	Tomate, frango e pimenta	LSTM	MSE (0,304)
Sabu e Kumar (2020)	Nozes	LSTM	RMSE (7,2780)

Chu et al. (2020)	Uva	EEMD-R-ADD	MAE (0,089); RMSE (0,114)
Zheng et al. (2020)	Espinafre, repolho, tomate, pimentão e batata	MLP	MAE (0,083); MAPE (4,54); RMSE (0,10)
Oktoviany, Knobloch e Korn (2021)	Milho	KNN	MAE (0,0779); RMSE (0,1041); MAPE (1,5192)
Ozden (2023)	Cebola, batata e alho	CNN	MAE (0,047); RMSE (0,070)
Sun et al. (2023a)	Carne suína, cebolinha chinesa, cogumelos shiitake e couve-flor	VMD-EEMD-LSTM	MAE (0,2263); MAPE (0,0584); RMSE (0,835)
Ray et al. (2023)	Lentilha	ARIMA-LSTM	RMSE (6,290); MAPE (2,886)

3 METODOLOGIA

metodologia desta pesquisa está dividida em cinco partes: A subseção 3.1 apresenta como os dados foram obtidos. Na 3.2, são descritas as arquiteturas escolhidas. A subseção 3.3 aborda o pré-processamento dos dados. A 3.4 detalha os modelos baseados em RNN. Finalmente, a 3.5 apresenta a análise dos resultados. A Figura 1 ilustra essa organização.

Figura 1. Etapas desta Metodologia.



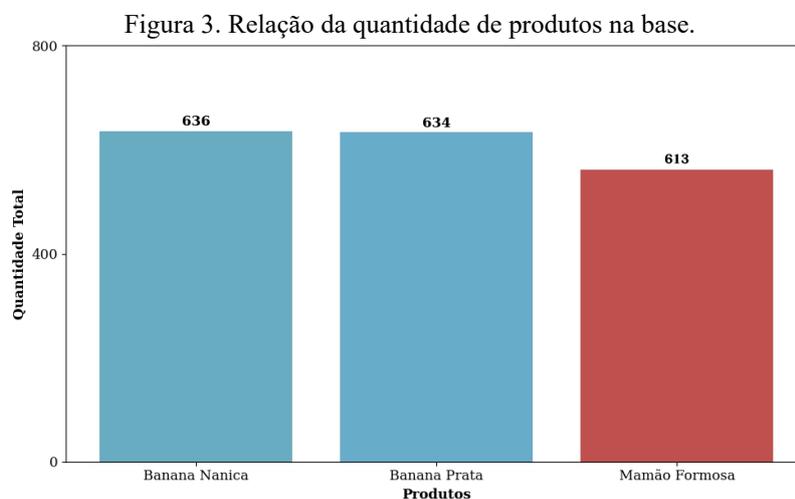
3.1 Aquisição dos Dados 3.2 Seleção das Arquiteturas 3.3 Pré-processamento 3.4 Modelagem Preditiva 3.5 Análise dos Resultados

3.1 AQUISIÇÃO DOS DADOS

Para o desenvolvimento deste trabalho foi utilizado um conjunto de dados com informações diárias sobre os preços de três frutas típicas da região baiana: banana nanica, banana prata e mamão formosa, disponibilizado no repositório online do Centro de Estudos Avançados em Economia Aplicada (CEPEA, 2024), da Universidade de São Paulo (USP).

O banco de dados contém 1270 registros de preços para bananas (nanica e prata) e 613 registros para mamão formosa, no período de 2012 a 2023. Organizado em formato de tabela, o mesmo contém

cinco colunas: Produto (tipo de fruta), Região, Data, Unidade (quilograma, caixa de 22 Kg e caixa de 20 Kg) e Preço. A Figura 2 ilustra os produtos e suas respectivas quantidades no banco de dados.



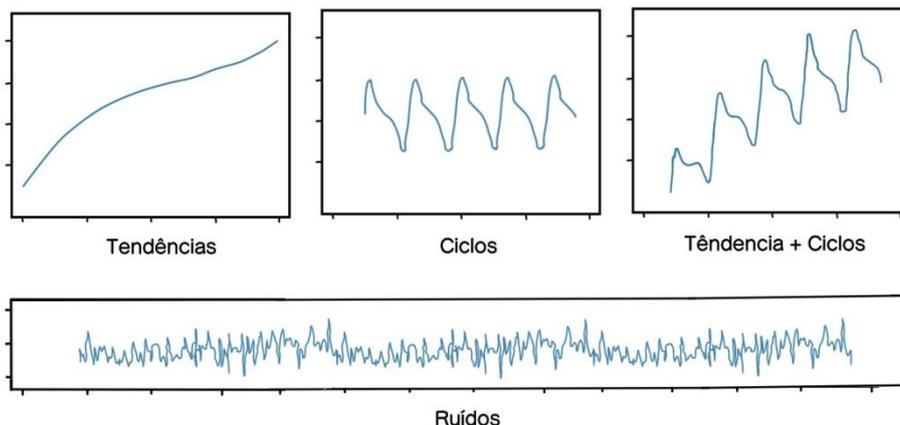
3.2 SELEÇÃO DAS ARQUITETURAS

Uma série temporal consiste em uma sequência de observações bem definidas, obtidas por meio de medições repetidas ao longo do tempo, como dias, meses, trimestres ou anos (RAHMAN et al., 2023). A previsão dessas séries envolve a análise dos dados para realizar um prognóstico, utilizando métodos estatísticos e modelagem (XU; HSU, 2022). A análise de séries temporais é uma técnica poderosa para detectar anomalias sazonais. Parte-se do pressuposto de que os valores observados apresentam padrões de repetição ao longo do tempo. O modelo geral para representar esses padrões é dado pela equação:

$$Y_t = T + S + \epsilon$$

Y_t representa o valor da série temporal em um instante t , T corresponde à tendência, S à sazonalidade, e ϵ ao erro ou ruído aleatório que pode interferir na análise (SUN et al., 2023b). A figura 3 ilustra como os componentes podem aparecer graficamente.

Figura 4. Tendência, sazonalidade (ciclos) e ruído em séries temporais.



Fonte: Adaptado de Seber e Wild (2004).

O levantamento do estado da arte teve como objetivo identificar as arquiteturas mais utilizadas na previsão de preços de produtos agrícolas. A Tabela 2 resume as principais informações dos estudos relacionados. Observou-se uma clara predominância das RNNs nas tarefas de previsão de séries temporais, evidenciando sua eficácia e popularidade nessa área.

Tabela 2. Arquiteturas utilizadas nos trabalhos relacionados.

Trabalhos	Arquiteturas
Kurumatani (2020)	SRNN, LSTM e GRU
Yuan e Ling (2020)	ARIMA, LSTM, SVR, Prophet e XGBoos
Sabu e Kumar (2020)	SARIMA, Holt-Winter e LSTM
Chu et al. (2020)	EEMD, LSSVM, EEMD-R-ADD, SVM, BPNN e ELM
Zheng et al. (2020)	MLP
Oktoviany, Knobloch e Korn (2021)	K-NN e Random Forest
Ozden (2023)	CNN, LSTM e Random Forest
Sun et al. (2023a)	Random Forest, ELM, LSTM, EEMD-LSTM, VMD-LSTM
Ray et al. (2023)	ARIMA, GARCH, LSTM e ARIMA-LSTM

Entre os modelos analisados, o *Long Short-Term Memory* (LSTM) se destacou como o mais frequentemente utilizado nos trabalhos relacionados. Sua escolha para este estudo deve-se à sua eficácia no processamento de séries temporais longas e à sua capacidade de solucionar o problema do desaparecimento do gradiente durante o treinamento (ZHANG et al., 2019; LIU; CHEN; WU, 2020).

O LSTM destaca-se por seu mecanismo de portas que permite à rede manter informações relevantes por períodos prolongados, sendo ideal para a tarefa de previsão de preços agrícolas.

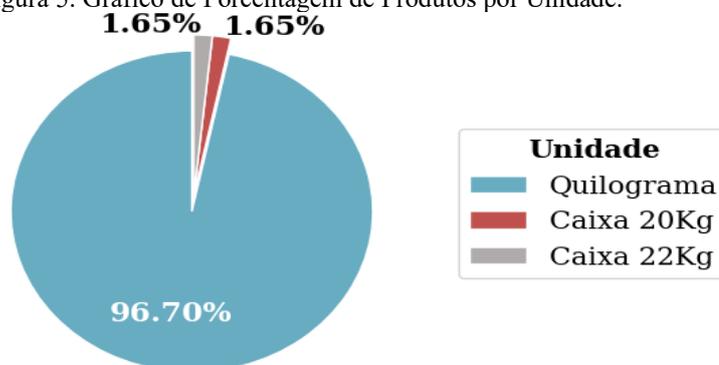
Em complemento, optou-se também pelo modelo *Gated Recurrent Unit* (GRU) devido à sua similaridade com o LSTM, mas com uma estrutura mais simplificada e menor quantidade de parâmetros. Embora compartilhe a capacidade de processar sequências através de portas de entrada, saída e esquecimento, o GRU transmite apenas um estado oculto entre as células, o que pode levar a uma redução no tempo de treinamento, sem comprometer significativamente a performance (KURUMATANI, 2020; FOROUTAN; LAHMIRI, 2024). Portanto, os modelos LSTM e GRU foram selecionados para a previsão dos preços no cenário agrícola baiano.

3.3 PRÉ-PROCESSAMENTO

As etapas de pré-processamento envolveram a limpeza de dados, remoção de valores indesejados, seleção de características relevantes e dimensionamento dos dados (QULMATOVA; KARIMOV; AZIMOV, 2022).

Primeiramente, foi verificada a presença de valores nulos/inválidos no conjunto de dados, e os registros identificados foram removidos. Além disso, observou-se que os produtos estavam classificados em três categorias com base em suas unidades de massa, conforme ilustrado na Figura 4. As categorias “caixa” representam produtos vendidos em maiores quantidades, o que resulta em um preço médio superior ao preço por quilograma. Como as entradas associadas à “caixa” correspondem a apenas 3,3% do total do conjunto de dados, decidiu-se por descartá-las para evitar distorções na análise.

Figura 5. Gráfico de Porcentagem de Produtos por Unidade.



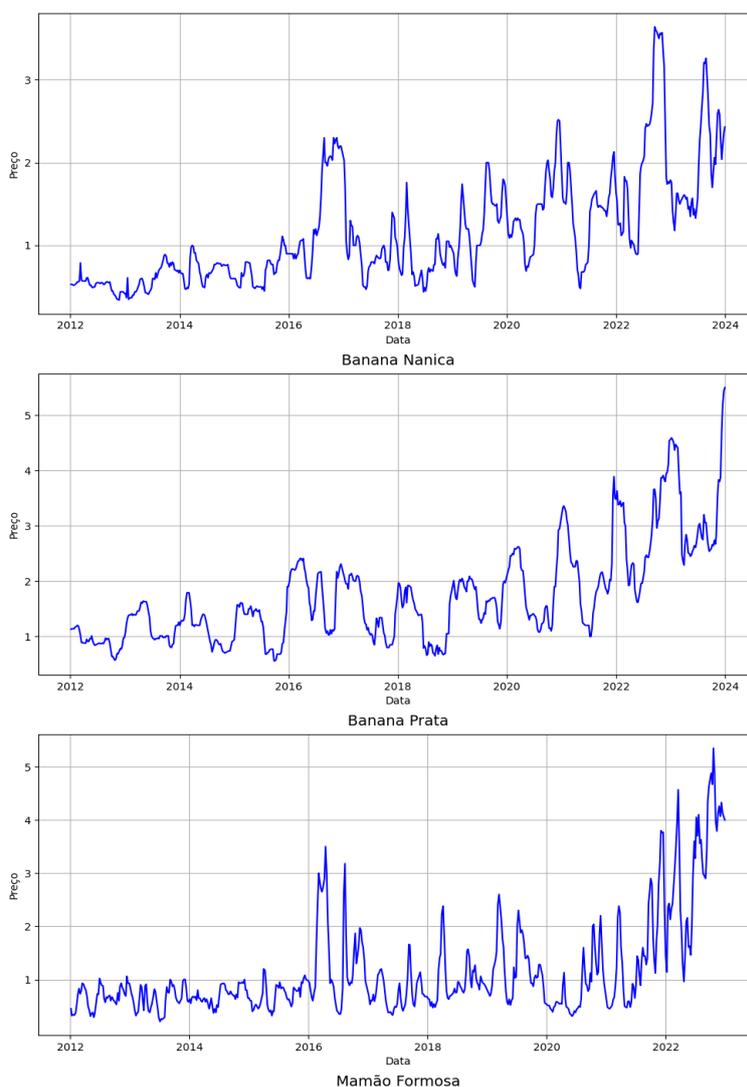
Em seguida, os dados foram normalizados para se ajustarem a um intervalo entre 0 e 1. Esse processo de normalização é essencial porque ajuda a padronizar a escala dos dados, facilitando a convergência dos modelos de aprendizado de máquina durante o treinamento. Ao manter todos os

valores dentro do mesmo intervalo, os modelos podem aprender de maneira mais eficiente e evitar que variáveis com escalas maiores dominem o processo de otimização (SEVERO, 2023). Com o pré-processamento concluído, foi possível avançar para a próxima etapa da metodologia.

3.4 MODELAGEM PREDITIVA

Para realizar a previsão, seguindo as práticas do estado da arte, a base de dados processada foi dividida em duas partes: os dados de janeiro de 2012 a dezembro de 2022 foram usados para o treinamento, e o ano de 2023 foi reservado para o teste. Os dados de treinamento são utilizados para o modelo aprender a identificar padrões nas informações, enquanto os dados de teste permitem analisar a eficiência do modelo. Essa divisão permite avaliar a capacidade do modelo de fazer previsões com dados novos. Os dados foram transformados em série temporal com frequência diária.

Figura 6. Série temporal.



A Figura 5 demonstra que o preço dos produtos da base exibem tendência, sazonalidade e ruídos, confirmando que os dados são de séries temporais. Conforme indicado no trabalho de Yuan, San e Leong (2020), é evidente que os preços apresentam instabilidade, com tendências claras e padrões sazonais. Para lidar com esses dados, a escolha de um modelo que consiga captar tanto as características das séries temporais quanto a sazonalidade é a mais apropriada.

Os modelos aplicados nesse estudo utiliza-se de RNNs, que são eficazes para análise de séries temporais devido à sua capacidade de memória de curto prazo (LIU et al., 2021). As RNNs são compostas por neurônios artificiais que simulam a estrutura de uma rede neural biológica (LIU; CHEN; WU, 2020), possuindo uma arquitetura composta por camadas de entrada, ocultas e de saída (KURUMATANI, 2020). A camada de entrada recebe e processa as informações iniciais por meio de funções de ativação. Em seguida, a camada oculta realiza o processamento e a análise dos dados, gerando previsões. Finalmente, a camada de saída fornece o resultado.

O gradiente descendente é utilizado durante o processo de treinamento e tenta encontrar coeficientes que melhor se ajustem aos dados, até achar os melhores pesos para o modelo (CHENG; WEI; CHENG, 2020). Os pesos são utilizados para ponderar a influência de cada neurônio, sendo ajustado durante o treinamento, para melhorar o desempenho da previsão. Sua fórmula matemática é a seguinte:

$$y_t = g(VA_t)$$

$$A_t = f(Ux_t + WA_{t-1})$$

A partir da fórmula da RNN, pode-se ver que no tempo t : o valor de entrada, o valor oculto e o valor de saída da rede são: X_t , A_t , Y_t (LIU et al., 2021). Contudo, as RNNs sofrem com a perda de gradiente quando lidam com longas sequências de dados (YUAN; LING, 2020), o valor de erro dos gradientes podem se tornar muito pequenos, dificultando o aprendizado da rede (MÁRQUEZ et al., 2021). Este problema é superado utilizando o modelo de Memória de longo e curto prazo - LSTM (SABU; KUMAR, 2020).

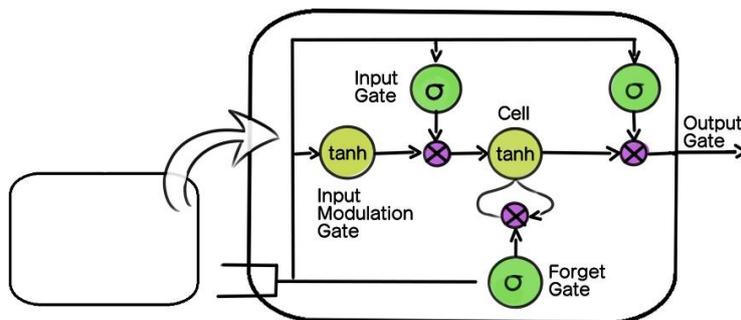
A LSTM é um tipo especial de RNN que resolve o problema do gradiente evanescente (YUAN; LING, 2020). Ela possui uma porta de entrada (*Input Gate*), porta de saída (*Output Gate*) e porta do esquecimento (*Forget Gate*). As ativações são calculadas em cada camada e comparadas com a saída desejada. A porta de esquecimento e a porta de entrada são usadas para controlar a posição da unidade, e a saída controla as saídas geradas pela LSTM (RAY et al., 2023). Dessa forma, a LSTM tem a capacidade de analisar as séries temporais com longo alcance de previsão, permitindo a exploração

das mudanças dinâmicas das sequências de entradas s (YUAN; LING, 2020). A formula de matemática da LSTM é a seguinte:

$$\begin{aligned}
 f_t &= \sigma(W_{fA}A_{t-1} + W_{fX}X_t + b_f) \\
 i_t &= \sigma(W_{iA}A_{t-1} + W_{iX}X_t + b_i) \\
 \tilde{c}_t &= \tanh(W_{cA}A_{t-1} + W_{cX}X_t + b_c) \\
 c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\
 Y_t &= \sigma(W_{YA}A_{t-1} + W_{YX}X_t + b_Y) \\
 A_t &= Y_t \circ \tanh(c_t)
 \end{aligned}$$

Na célula LSTM, f_t representa a porta de esquecimento, oque deve ser descartado da memória. A porta de entrada, i_t determina o vai ser adicionado à memória, enquanto $\sim ct$ gera o novo dado a ser potencialmente armazenado. O estado da memória, ct atualiza o estado da memória com base na porta de esquecimento e na porta de entrada. A porta de saída, Y_t , decide o valor que será usado como saída da célula, e o estado oculto final, A_t é gerado com base no estado da memória e na porta de saída. A Figura 6 mostra uma representação gráfica da célula LSTM.

Figura 6. Representação gráfica da LSTM



Fonte: Adaptado de Abdelwahed, Letaifa e Ksouri (2023).

O GRU, assim como a LSTM é um tipo especial de RNN, porém com uma estrutura mais simples (MIGLIATO, 2021). Ele possui apenas dois portões, um portão de redefinição (reset gate) e um portão de atualização (*update date*), que decidem quais informações devem ser passadas para a saída (KURUMATANI, 2018). As equações da célula GRU podem ser definidas como:

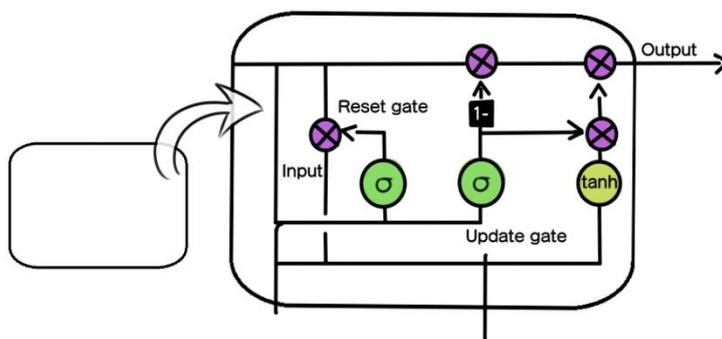
$$\begin{aligned}
 z_t &= \sigma(W_x X_t + U_z h_{t-1} + b_z) \\
 r_t &= \sigma(W_r X_t + U_r h_{t-1} + b_r)
 \end{aligned}$$

$$\hat{h}_t = \tanh(W_h X_t + U_h(r_t \odot h_{t-1}) + b_h)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t$$

Na célula GRU, X_t representa o valor de entrada que contém a informação a ser processada. O vetor do update gate z_t controla quais informações antigas deve ser mantida e quais novas devem ser incorporada. O vetor do reset gate, r_t decide quanto da informação anterior deve ser esquecida. O vetor de saída, h_t mostra o estado oculto da célula no tempo atual, e \hat{h}_t sugere o novo estado antes das decisões das portas de controle serem aplicadas. A Figura 7 mostra uma representação da célula de GRU.

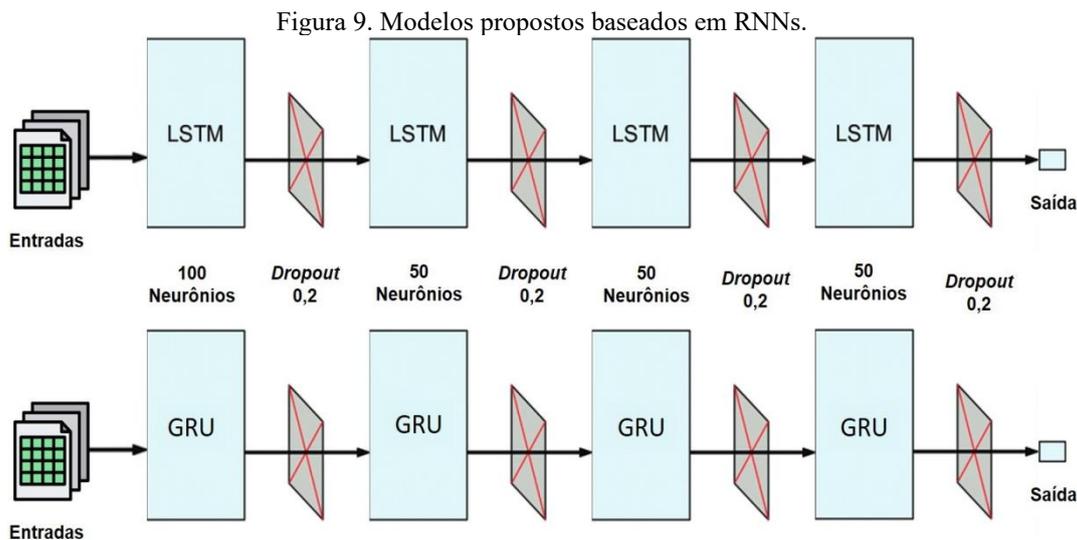
Figura 7. Representação gráfica do GRU.



Fonte: Adaptado de Wang et al. (2024).

Enquanto a LSTM carrega a memória de longo e curto prazo, através do estado da célula e do estado oculto, as GRUs têm apenas um estado oculto, que carrega a memória de curto prazo, embora os dois modelos sejam superiores a um modelo RNN mais simples (SILVA, 2020).

Para a construção dos modelos LSTM e GRU empregados neste estudo, os dados de entrada foram preparados em janelas de 30 dias *lags*. Após análise empírica, os modelos foram configurados com quatro camadas, sendo a primeira com 100 neurônios e as três camadas subsequentes com 50 neurônios cada. Cada camada é seguida por uma camada de Dropout com taxa de 20% para reduzir o risco de sobreajuste, situação onde o modelo não consegue generalizar e, em vez disso, se adequa muito estritamente ao conjunto de dados de treinamento. Além disso, uma camada densa final é adicionada, contendo uma unidade de saída com ativação linear, responsável pela previsão do preço. A Figura 8 ilustra a arquitetura dos modelos aplicados.



3.5 ANÁLISE DOS RESULTADOS

Para avaliar a capacidade dos modelos de previsão de preço, diversos autores recorrem a métricas de erro para validar os resultados das simulações (ZHANG et al., 2019; LIU; CHEN; WU, 2020; QULMATOVA et al., 2022; DWIVEDI et al., 2021). Essas métricas comparam os valores previstos com os reais, focando na diferença entre eles. As principais métricas de erro incluem RMSE, MAE, MAPE e MSE, onde Y_t representa o valor real no tempo i , \hat{Y}_t é o valor previsto para o mesmo tempo, e n é o número total de pontos de dados (SUN et al., 2023a).

O MAE calcula a média dos erros absolutos entre os valores previstos e os reais, medindo diretamente a diferença entre o valor real e o previsto. Os erros absolutos são expressos nas unidades reais dos dados. Um MAE mais próximo de 0 indica um modelo mais preciso, como ilustrado pela fórmula abaixo:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

O MAPE, por sua vez, calcula o erro absoluto em termos percentuais, oferecendo uma visão relativa do erro. No entanto, não há um padrão universal para um MAPE considerado excelente, pois variações nos preços podem dificultar a previsão precisa (QULMATOVA et al., 2022):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

O MSE é uma métrica comum na avaliação da precisão de séries temporais (DWIVEDI et al., 2021). O MSE calcula a média dos quadrados das diferenças entre os valores previstos e reais. Valores altos de MSE indicam que o modelo teve um desempenho ruim nas previsões:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

O RMSE, por sua vez, fornece a raiz quadrada da média dos quadrados das diferenças entre os valores previstos e reais (AGARWAL; RAY; TRIPATHI, 2023). A utilização da raiz quadrada ajuda a manter as unidades consistentes e facilita a interpretação dos resultados. Um RMSE mais baixo indica um modelo mais eficaz:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

4 RESULTADOS E DISCUSSÕES

Os hiperparâmetros utilizados no treinamento dos modelos foram definidos com base em uma abordagem empírica, onde diferentes combinações foram testadas e ajustadas para obter o melhor desempenho possível em termos de precisão e capacidade de generalização. A Tabela 3 resume os hiperparâmetros escolhidos, que incluem tanto aspectos da arquitetura dos modelos quanto parâmetros de treinamento.

Tabela 3. Hiper-parâmetros utilizados no treinamento dos modelos.

Tipo	Hiper-parâmetro
Arquitetura	LSTM E GRU
<i>Dropout</i>	0,2
Função de ativação na camada de saída	Linear
Otimizador	RMSprop
Taxa de Aprendizagem	0,001
Função de custo do treinamento	MSE
Métrica de avaliação do treinamento	MAE

Batch Size	8, 16 e 32
Épocas treinamento	50, 100, 150 e 200

Os modelos foram implementados utilizando camadas LSTM e GRU, escolhidas pela sua capacidade de capturar dependências temporais em dados sequenciais. Ambas as arquiteturas foram experimentadas devido à sua eficácia comprovada em tarefas de séries temporais, como a previsão de preços agrícolas, que é o foco deste trabalho. A técnica de *Dropout* foi aplicada com uma taxa de 0,2 para reduzir o risco de sobreajuste (*overfitting*), garantindo assim uma generalização mais eficaz durante o treinamento.

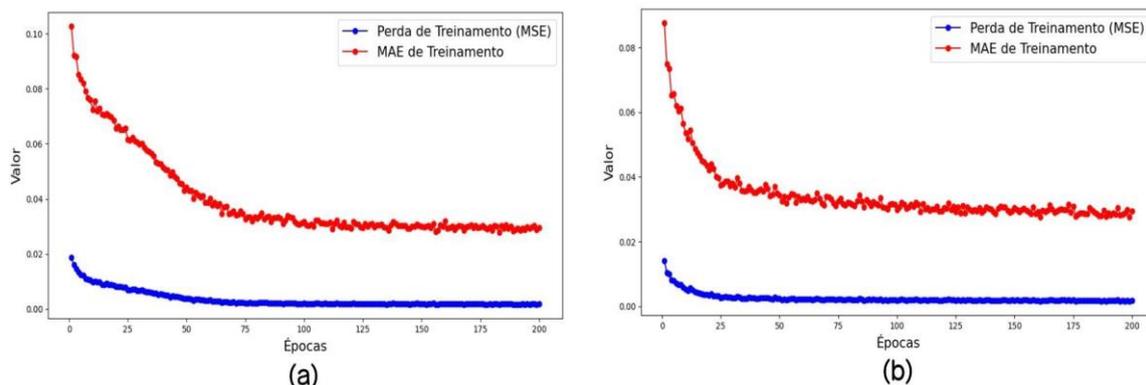
A função de ativação linear foi adotada na camada de saída devido à natureza do problema de regressão, onde o objetivo é prever valores contínuos, como preços. A ativação linear é a escolha mais natural para este tipo de tarefa, pois permite que a rede produza uma saída não limitada por intervalos fixos. O otimizador RMSprop foi escolhido pela sua eficácia em problemas de séries temporais, onde as variâncias dos gradientes podem ser muito diferentes ao longo das iterações. RMSprop ajusta dinamicamente a taxa de aprendizagem para cada parâmetro, contribuindo para um treinamento mais estável e eficiente (DIJINGA, 2024). Uma taxa de aprendizagem de 0,001 foi selecionada com base em testes preliminares. Esse valor é comumente utilizado como ponto de partida para o otimizador RMSprop e demonstrou ser adequado para garantir uma convergência estável, evitando lentidão no aprendizado.

A função de custo *Mean Squared Error* (MSE) foi empregada por ser muito utilizada em problemas de regressão. O MSE penaliza desvios maiores de forma mais severa, o que ajuda o modelo a se concentrar na minimização dos erros de maior magnitude. O erro absoluto médio (MAE) foi escolhido como métrica de avaliação, pois fornece uma interpretação direta do erro médio em termos das unidades de saída. Ao complementar a função de custo MSE, o MAE facilita a análise do desempenho do modelo ao longo do treinamento, especialmente em termos de previsibilidade (JUNIOR, 2021).

Os tamanhos de *batch* 8, 16 e 32 foram testados para determinar o melhor equilíbrio entre eficiência computacional e estabilidade do gradiente. Esses valores são comumente usados em tarefas de aprendizado profundo e oferecem diferentes compensações entre precisão e tempo de treinamento. Por fim, o número de épocas foi ajustado para 50, 100, 150 e 200, permitindo observar o comportamento do modelo em diferentes etapas de treinamento. Esse conjunto de valores possibilitou avaliar se o modelo estava melhorando com o aumento da quantidade de épocas.

Como observado na figura 10, do experimento da banana nanica com os hiperparâmetros de 8 *Batch Size* e 200 Épocas, é possível notar que o MSE do modelo LSTM é significativamente menor do que a do modelo GRU durante o treinamento, o que indica que a LSTM se adequou melhor a série temporal do produto.

Figura 10. (a) Exemplo do Treinamento do LSTM e (b) Exemplo do Treinamento do GRU.



A avaliação dos modelos foi realizada utilizando as métricas apresentadas na subseção 3.5. A Tabela 4 apresenta o desempenho do modelo LSTM no conjunto de teste sob diferentes configurações de hiperparâmetros, para cada tipo de fruta, comparando-se os valores de RMSE, MAE, MAPE e MSE. Os melhores resultados foram obtidos com: 200 épocas e *batch size* de 8 para a banana nanica, banana prata 100 épocas e *batch size* de 8 enquanto para mamão formosa 100 épocas e *batch size* de 16.

Tabela 4. Melhores Resultados utilizando LSTM.

Produto	Épocas	Batch Size	RMSE	MAE	MAPE	MSE
Banana Nanica	200	8	0,186	0,142	7,286	0,035
Banana Prata	100	8	0,278	0,196	5,853	0,077
Mamão Formosa	100	16	0,606	0,483	16,624	0,367

Na tabela 5, foi feita a comparação com os mesmo hiperparâmetros, mas aplicados no modelo GRU. O melhor resultado para a banana nanica foi 200 épocas, 8 *Batch Size*, banana prata 50 épocas, 8 *Batch Size* e para o mamão formosa 100 épocas, 32 *Batch Size*.

Tabela 5. Melhores Resultados utilizando GRU.

Produto	Épocas	Batch Size	RMSE	MAE	MAPE	MSE
---------	--------	------------	------	-----	------	-----

Banana Nanica	200	8	0,192	0,145	7,391	0,037
Banana Prata	50	8	0,283	0,193	5,938	0,080
Mamão Formosa	100	32	0,639	0,503	17,830	0,409

A Figura 11 mostra a comparação entre os preços gerados pelos modelos selecionados nesta pesquisa (LSTM) e o preço real de cada produto. A linha sólida em vermelho representa valores reais e a linha azul representa valores previstos.

Figura 11. Comparativo entre Preço Real x Preço Previsto das Frutas no modelo.



5 CONCLUSÕES E TRABALHOS FUTUROS

A previsão de preços é um assunto discutido globalmente devido aos prejuízos causados pela falta de um controle financeiro. Para auxiliar na criação de estratégias de tomada de decisões, ou evitar os danos causados pela flutuação de preços, pesquisadores de diferentes países buscam métodos que possam prever esses preços de maneira automática, trazendo a menor taxa de erro possível, a análise preditiva baseada em RNNs tem mostrando-se uma tecnologia promissora.

Nessa pesquisa, à análise preditiva, foi utilizada como recurso para prever o preço das 3 principais frutas, produzida no território baiano pelos produtores familiares. Os modelos foram aplicados em um base de dados contendo, ao todo, 1.883 registro de preços das frutas, banana nanica, banana prata e mamão formosa. Por meio dos resultados obtidos, e após a comparação entre os modelos, conclui-se que o modelo LSTM aplicado nesse trabalho, apresentou um desempenho satisfatório no contexto da academia, atingindo o objetivo da pesquisa. Vale ressaltar que a análise preditiva pode auxiliar na gestão financeira e de risco, identificando os períodos onde o preço tem mais oscilação, criando uma maneira do produtor ajustar o seu preço de revenda de forma gradativa, aumentando a lucratividade. Deste modo, espera-se que esse modelo possa auxiliar o produtor na criação de medidas que facilitem o controle de gestão.

Como proposta para trabalhos futuros, pretende-se aprimorar o modelo apresentado variáveis exógenas, pois os preços dos produtos agrícolas são influenciados por vários fatores, como flutuações de temperatura, feriados e níveis de consumo, a junção dos preços com esses fatores pode ser um caminho valioso para melhorar sua precisão. Adicionalmente, construir uma plataforma online, onde os pequenos produtores poderão consultar, de forma gratuita e intuitiva, as projeções de preços, além de analisar outras culturas com focos em mais regiões do Brasil, visando fornecer informações independentes e confiáveis ao setor agrícola.

REFERÊNCIAS

- ABDELWAHED, N.; LETAIFA, A. B.; KSOURI, A. Predicting sentiment analysis for web users with a deep learning approach. In: 2023 IEEE Tenth International Conference on Communications and Networking (ComNet). [S.l.: s.n.], 2023. p. 1–11.
- AGARWAL, N.; RAY, S.; TRIPATHI, K. Time series forecasting of agriculture yield of cotton with regression model implementation. In: IEEE. 2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON). [S.l.], 2023. p. 1–6.
- AQUINO, J. R. de; ALVES, M. O.; VIDAL, M. de F. Agricultura familiar no nordeste do brasil: um retrato atualizado a partir dos dados do censo agropecuário 2017. *Revista Econômica do Nordeste*, v. 51, n. Suplemento Especial, p. 31–54, 2020. 2
- ASSIS, B. H. et al. A importância da contabilidade e do direito no agronegócio. *Revista Projetos Extensionistas*, v. 1, n. 1, p. 195–208, 2021.
- BOSCHETTI, M. et al. Assimilating seasonality information derived from satellite data time series in crop modelling for rice yield estimation. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). [S.l.: s.n.], 2015. p. 157–160.
- CEPEA. Banco de dados Cepea. 2024. Acesso em: 22 mai. 2024. Disponível em: (https://www.hfbrasil.org.br/br/banco-de-dados-precos-medios-dos-hortifruticolas.aspx?produto=4®iao%5B%5D=6&periodicidade=anual&ano_inicial=2020&ano_final=2024&pagina=1).
- CHENG, W.; WEI, S.; CHENG, F. Grey system correlation-based feature selection for time series forecasting. In: IEEE. 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS). [S.l.], 2020. p. 837–841.
- CHU, X. et al. A novel machine learning-based strategy for agricultural time series analyzing and forecasting: a case study in china's table grape price. In: IEEE. 2020 IEEE 8th International Conference on Computer Science and Network Technology (ICCSNT). [S.l.], 2020. p. 75–80.
- CODEVASF. Cerca de 298,3 mil toneladas de itens agrícolas foram produzidas em 2023 nos projetos de irrigação da Codevasf no oeste da Bahia. 2024. Accessed: 2024-07-28. Disponível em: (<https://www.codevasf.gov.br/noticias/2024>). 2
- DIJ KINGA, F. J. The RMSprop optimizer. 2024. Acesso em: 2024-06-23. Disponível em: (<https://medium.com/@fernando.dijkinga/the-rmsprop-optimizer-78f02efb63e9>).
- DUTRA, T. R. A aplicação da contabilidade na propriedade rural: As ferramentas contábeis como método de gestão. *Repositório de Trabalhos de Conclusão de Curso*, 2020.
- DWIVEDI, S. A. et al. Analysis and forecasting of time-series data using sarima, cnn and lstm. In: IEEE. 2021 international conference on computing, communication, and intelligent systems (icccis). [S.l.], 2021. p. 131–136.

ECONOMICA, B. Bahia é segundo estado no ranking de produção de banana do país. 2023. <<https://bahiaeconomica.com.br/wp/2023/09/22/bahia-e-segundo-estado-no-ranking-de-producao-de-banana-do-pais/>>. Acessado em: 2 out. 2024.

FAEB, F. da Agricultura e Pecuária do Estado da B. Doce Bahia: Estado é potência nacional na fruticultura. 2023. Accessed: 2024-07-28. Disponível em: <<https://sistemafaeb.org.br/doce-bahia-estado-e-potencia-nacional-na-fruticultura>>. 1

FOROUTAN, P.; LAHMIRI, S. Deep learning systems for forecasting the prices of crude oil and precious metals. *Financial Innovation*, Springer, v. 10, n. 1, p. 111, 2024.

INBOX, O. Banana Prata 1kg. 2024. <<https://hmg.organicosinbox.com.br/produto/banana-prata-1kg/>>. Acessado em: 2 out. 2024.

JUNIOR, C. de O. Métodos para regressão. 2021. Acesso em: 2024-02-07. Disponível em: <<https://medium.com/data-hackers/prevendo-n%C3%BAmoros-entendendo-m%C3%A9tricas-de-regress%C3%A3o-35545e011e70>>.

KULBA, V.; MEDENNIKOV, V. A model of the russia's agriculture readiness for digital transformation. In: IEEE. 2020 13th International Conference "Management of large-scale system development" (MLSD). [S.l.], 2020. p. 1–5.

KURUMATANI, K. Time series prediction of agricultural products price based on time alignment of recurrent neural networks. In: IEEE. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). [S.l.], 2018. p. 81–88.

KURUMATANI, K. Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. *SN Applied Sciences*, Springer, v. 2, n. 8, p. 1434, 2020.

LANDAU, E. C.; SILVA, G. A. da. Evolução da produção de mamão (carica papaya, caricaceae). In: LANDAU, EC; SILVA, GA da; MOURA, L.; HIRSCH, A.; GUIMARAES, DP (Ed . . . , 2020).

LIU, D.; CHEN, A.; WU, J. Research on stock price prediction method based on deep learning. In: IEEE. 2020 2nd International Conference on Information Technology and Computer Application (ITCA). [S.l.], 2020. p. 69–72. 8, 11, 13

LIU, Z. et al. Forecast methods for time series data: a survey. *Ieee Access*, IEEE, v. 9, p. 91896–91912, 2021.

MÁRQUEZ, J. P. et al. Ethanol fuel demand forecasting in brazil using a lstm recurrent neural network approach. *IEEE Latin America Transactions*, IEEE, v. 19, n. 4, p. 551–558, 2021.

MIGLIATO, A. L. T. Detecção de Outliers em Dados não Vistos de Séries Temporais por meio de Erros de Predição com SARIMA e Redes Neurais Recorrentes LSTM e GRU. Tese (Doutorado) — Universidade de São Paulo, 2021.

OKTOVIANY, P.; KNOBLOCH, R.; KORN, R. A machine learning-based price state prediction model for agricultural commodities using external factors. *Decisions in Economics and Finance*, Springer, v. 44, n. 2, p. 1063–1085, 2021.

OZDEN, C. Comparative analysis of cnn, lstm and random forest for multivariate agricultural price forecasting. *Black Sea Journal of Agriculture*, Hasan ONDER, v. 6, n. 4, p. 422–426, 2023.

PEREIRA, J. L. S. R.; SANTOS, N. S. dos. Controle financeiro na agricultura familiar: Uma investigação sobre sua utilização e relevância. *Epitaya E-books*, v. 1, n. 9, p. 132–143, 2022.

PROATIVA. Mamão Formosa. 2024. (<https://www.proativaalimentos.com.br/mamao-formosa>). Acessado em: 2 out. 2024.

QULMATOVA, S. et al. Crop production under different climatic conditions by analyzing agricultural data using multiple linear regression, winter holt, and artificial intelligence. In: *Proceedings of the 6th International Conference on Future Networks & Distributed Systems*. [S.l.: s.n.], 2022. p. 242–252.

QULMATOVA, S.; KARIMOV, B.; AZIMOV, D. Data analysis and forecasting in agricultural enterprises. In: *Proceedings of the 6th International Conference on Future Networks & Distributed Systems*. [S.l.: s.n.], 2022. p. 536–541.

RAHMAN, M. T. et al. Time series forecasting of agricultural products sale using deep learning. In: *IEEE. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. [S.l.], 2023. p. 1–6. 3, 7

RAY, S. et al. An arima-lstm model for predicting volatile agricultural price series with random forest technique. *Applied Soft Computing*, Elsevier, v. 149, p. 110939, 2023.

REDDY, P. C. S. et al. Data analytics in farming: Rice price prediction in andhra pradesh. In: *2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*. [S.l.: s.n.], 2022. p. 1–5.

RODRIGUES, J. L. et al. Desafios para a sustentabilidade da cadeia produtiva do abacaxi em itaberaba, bahia. *Revista Metropolitana de Sustentabilidade (ISSN 2318-3233)*, v. 11, n. 3, 2021.

SABU, K. M.; KUMAR, T. M. Predictive analytics in agriculture: Forecasting prices of arecanuts in kerala. *Procedia Computer Science*, Elsevier, v. 171, p. 699–708, 2020.

SANTOS, V. F. dos; MACIEL, L. dos S.; BALLINI, R. Efeito das operações de hedge e especulação sobre a volatilidade dos preços de commodities agrícolas nos eua. *Economía Aplicada*, v. 24, n. 3, p. 343–366, 2020.

SEAGRI. Dia da Banana: Segundo maior produtor do Brasil, estado da Bahia investe em. 2023. Acesso em: 14 jul. 2024. Disponível em: (<http://www.seagri.ba.gov.br/noticias/2023/09/22/dia-da-banana-segundo-maior-produtor-do-brasil-estado-da-bahia-investe-em>).

SEBER, G. A. F.; WILD, C. J. *Encontros Com O Acaso: Primeiro Curso De Análise De Dados E Inferência*. 1. ed. Português: Editora da Universidade, 2004.

SEVERO, J. Normalização de dados contínuos. 2023. Acesso em: 22024-05- 01. Disponível em: <https://pt.linkedin.com/pulse/normaliza%C3%A7%C3%A3o-de-dados-johanes-severo>.

SILVA, E. F. d. Alocação das reservas internacionais com base em previsão de fatores econômicos e financeiros utilizando redes neurais recorrentes e modelo nelson-siegel dinâmico. 2020.

SUN, C. et al. A study on agricultural commodity price prediction model based on secondary decomposition and long short-term memory network. *Agriculture, MDPI*, v. 14, n. 1, p. 60, 2023.

SUN, F. et al. Agricultural product price forecasting methods: A review. *Agriculture, MDPI*, v. 13, n. 9, p. 1671, 2023.

WANG, T. et al. Application of a multi-model fusion forecasting approach in runoff prediction: A case study of the yangtze river source region. *Sustainability, MDPI*, v. 16, n. 14, p. 5964, 2024.

XU, J.-L.; HSU, Y.-L. Analysis of agricultural exports based on deep learning and text mining. *The Journal of Supercomputing, Springer*, v. 78, n. 8, p. 10876–10892, 2022.

YOO, T.-W.; OH, I.-S. Time series forecasting of agricultural products' sales volumes based on seasonal long short-term memory. *Applied sciences, MDPI*, v. 10, n. 22, p. 8169, 2020.

UAN, C. Z.; LING, S. K. Long short-term memory model based agriculture commodity price prediction application. In: *Proceedings of the 2020 2nd international conference on information technology and computer communications*. [S.l.: s.n.], 2020. p. 43–49.

YUAN, C. Z.; SAN, W. W.; LEONG, T. W. Determining optimal lag time selection function with novel machine learning strategies for better agricultural commodity prices forecasting in malaysia. In: *Proceedings of the 2020 2nd international conference on information technology and computer communications*. [S.l.: s.n.], 2020. p. 37–42.

ZHANG, S. et al. Optimizing time-series prediction on china's green trade economy. In: *IEEE. 2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. [S.l.], 2019. p. 1579–1584.

ZHENG, G. et al. The research on agricultural product price forecasting service based on combination model. In: *2020 IEEE 13th International Conference on Cloud Computing (CLOUD)*. [S.l.: s.n.], 2020. p. 4–9.