10.18605/2175-7275/cereus.v13n2p163-175

ARTIGO ORIGINAL

<< Recebido em: 24/04/2021 Aceito em: 10/06/2021. >>

REVISTA CEREUS ISSN: 2175-7275

Previsão da produtividade de arroz: uma aplicação de redes neurais recorrentes LSTM

Forecasting rice productivity: an application of recurrent LSTM neural networks

Jandrei Sartori Spancerski [1], José Airton Azevedo dos Santos [2]

RESUMO

O arroz, responsável por suprir a população com calorias e proteínas, ocupa uma posição de destaque do ponto de vista social e econômico. É um produto essencial na cesta básica do consumidor brasileiro. Neste contexto, este trabalho propõe um modelo LSTM (*Long Short-Term Memory*) para previsão da produtividade de arroz no estado do Rio Grande do Sul. A base de dados, obtida pelo Instituto Rio Grandense do Arroz (IRGA), apresenta uma série histórica, da produtividade de arroz, das safras compreendidas no período entre 1921 e 2020. O modelo de previsão, baseado em Redes Neurais LSTM, foi implementado por meio da biblioteca de aprendizado de máquina Pytorch. Os resultados obtidos, para as safras 2017/18, 2018/19 e 2019/20, mostram que o modelo de previsão forneceu estimativas confiáveis para a produtividade do arroz no Rio Grande do Sul.

Palavras-chave: Redes neurais artificiais. Produtividade de arroz. Previsão.

ABSTRACT

Rice, responsible for supplying the population with calories and protein, occupies a prominent position from the social and economic point of view. It is an essential product in the basic basket of Brazilian consumers. In this context, this work present an LSTM (Long Short-Term Memory) model for forecasting rice productivity in the state of Rio Grande do Sul. The database, obtained by the Rio Grandense Rice Institute (IRGA), presents a historical series, of rice productivity, of the harvests between 1921 and 2020. The forecasting model, based on LSTM Neural Networks, was implemented through the Pytorch machine learning library. The results obtained for the 2017/18, 2018/19 and 2019/20 harvests show that the forecast model provided reliable estimates for rice productivity in Rio Grande do Sul.

Keywords: Artificial neural networks. Rice productivity. Forecast.

[1] Discente do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio (PPGTCA). Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: jandreisst@gmail.com

[2] Doutor em Engenharia Elétrica. Docente do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio (PPGTCA). Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: airton@utfpr.edu.br

1. INTRODUÇÃO

A orizicultura, cultura do arroz, é uma atividade muito importante do ponto de vista econômico e social. O arroz, devido ao seu valor nutritivo e custos baixos, é produzido principalmente nos países subdesenvolvidos. Sendo, no Brasil, uma das formas principais de alimento da população. É considerado como fonte de energia, devido a alta concentração de amido. Fornecendo também, proteínas, minerais, vitaminas e baixo teor de lipídios (SANTOS; TAVARES, 2018; BRONDANI et al., 2006; WATTO; MUGERA, 2014).

A orizicultura se adapta a várias condições de clima e solo e desempenha, na economia brasileira, um papel estratégico. O cultivo é feito principalmente de forma irrigada e a produção nacional concentra-se nos estados da região sul. O estado, do Brasil, maior produtor de arroz em casca é o Rio Grande do Sul. Na porção sul do estado encontram-se os principais municípios produtores. A produção, destaque na pauta das exportações do Rio Grande do Sul, tem como destino o mercado interno e externo (ATLAS, 2020; WALTER et al., 2008).

A agricultura brasileira vem cada vez mais se desenvolvendo apoiada na utilização de técnicas modernas de plantio e colheita, gerando com isso maiores ganhos de produtividade e competitividade no mercado externo. Modelos de previsão tem um papel importante na gestão da agricultura. Segundo Rohrig (2021) a previsão da produtividade de arroz contribui: na correção da fertilidade do solo, no planejamento da safra subsequente, nas operações de pós-colheita, no planejamento da aquisição dos insumos, na contratação da mão de obra, etc.

Redes Neurais Artificiais (RNAs) são objetos de programação que imitam o funcionamento de redes neurais biológicas. São sistemas compostos por unidades de processamento simples que interagem por meio de conexões com pesos e processam informações devido a estímulos externos. As RNAs têm sido utilizadas com sucesso em tarefas de predição e modelagem de séries temporais. Séries temporais são conjuntos de observações ordenadas no tempo (HAYKIN, 2001; BASTIANI, et al., 2018; PINHEIRO et al., 2020; SANTOS; CHAUKOSKI, 2020).

Diversos trabalhos utilizaram métodos de previsão em aplicações voltadas ao mercado do arroz. Marasca e Souza (2016) investigaram, por meio de modelos ARIMA, a produtividade mundial de arroz. Obtendo resultados, de previsão, que evidenciam a tendência crescente da produtividade ao longo dos anos. Péres e Pire (2018) utilizaram modelos ARIMA na previsão do preço do arroz no mercado internacional. Observaram que

as previsões obtidas, em seu estudo, seguem o comportamento estocástico gerado pela série de preços do arroz. Já Awai e Siddique (2011) empregaram modelos ARIMA na produção de arroz em Bangladesh. Observaram que os modelos foram mais eficientes em previsões de curto prazo.

Embora a cultura do arroz tenha importância mundial poucos são os trabalhos que utilizam redes neurais recorrentes na previsão da produtividade do arroz, principalmente utilizando a biblioteca Pitorch. Em grande parte, são utilizadas, na previsão da produtividade, abordagens tradicionais como os modelos ARIMA.

Neste contexto, este trabalho propõe um modelo LSTM (*Long Short-Term Memory*) para previsão da produtividade de arroz, no estado do Rio Grande do Sul, no período entre 1921 e 2020.

2. MATERIAIS E MÉTODOS

O trabalho foi dividido em quatro etapas:

1. Coleta e Análise dos dados: Esta etapa começou pela coleta dos dados no site do IRGA (Instituto Riograndense do Arroz). Na sequência, com a intenção de obter informações importantes sobre os dados, realizou-se uma análise exploratória dos mesmos.
2. Modelagem: Na etapa de modelagem foram implementados vários modelos de redes neurais recorrentes LSTM. Nesta etapa o modelo com melhor desempenho, no conjunto de validação, é selecionado.
3. Teste: Nesta etapa o modelo LSTM é testado com dados que não participaram do processo de treinamento e validação (Safras 2017/18, 2018/19 e 2019/20).
4. Previsão: Finalmente, nesta etapa, realizou-se a previsão da produtividade do arroz para a safra de 2020/21.

Base de Dados:

Para previsão da produtividade de arroz (kg/ha) utilizou-se uma base de dados, com 99 safras (1921/22 - 2019/20), obtida pelo Instituto Rio Grandense do Arroz (IRGA, 2020). Os dados coletados, do site do IRGA, já estavam limpos e sem a presença de *outliers* (Figura 1).

Os dez primeiros registros do conjunto de dados do IRGA são apresentados na Tabela 1.

Tabela 1. Dez primeiros registros do conjunto de dados.

Data	Produtividade (kg/ha)
1921/22	2190
1922/23	2178
1923/24	2100
1924/25	1994
1925/26	1991
1926/27	2229
1927/28	2160
1928/29	2235
1929/30	2264
1930/31	2208

Fonte: IRGA (2020).

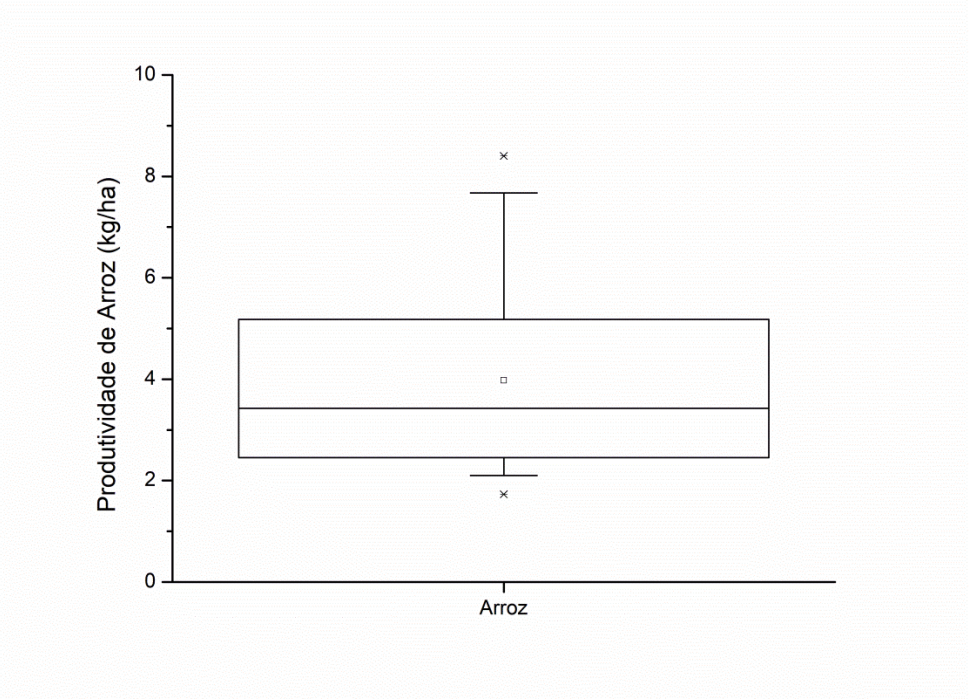

Figura 1. *Boxplot* – produtividade de arroz

Redes Recorrentes:

Redes Neurais Recorrentes (*Recurrent Neural Network* (RNNs)), capazes de agregar ciclos em seus algoritmos, têm sido utilizadas com muito sucesso em tarefas que envolvem entradas sequenciais, como fala, linguagem, séries temporais, entre outras.

As redes neurais recorrentes LSTM tem mostrado excelente habilidade, quando comparadas com outras redes recorrentes, para apreender dependências de longo prazo.

A topologia de um neurônio, de uma LSTM, é baseada em uma célula de memória. Cada célula contém portas que tem a capacidade de adicionar, descartar ou atualizar as informações no tempo, de modo a prever melhor estados futuros. As portas regulam o fluxo de informações (NELSON et al., 2018; GRAVES, 2014).

Um neurônio LSTM (célula) trabalha com uma sequência de entrada x_t (Figura 2) e cada porta (*gate*), dentro de uma célula, usa unidades de ativação para controlar se elas são acionadas ou não, fazendo com que a mudança de estado e a adição de informações fluam através da célula. O parâmetro C_t representa o estado da célula no instante t, este estado representa as informações que chegaram até esse passo em instante de tempos passados. O *gate* de esquecimento f_t determina quais informações devem ser jogadas fora pela célula. O *gate* de entrada i_t determina os valores de entrada para atualizar o estado da célula e o *gate* de saída O_t determina o que produzir com base na entrada e na memória da célula. Todos esses valores acabam sendo concatenados, multiplicados ou somados, conforme mostra o circuito apresentado na Figura 2 (GRAVES, 2014).

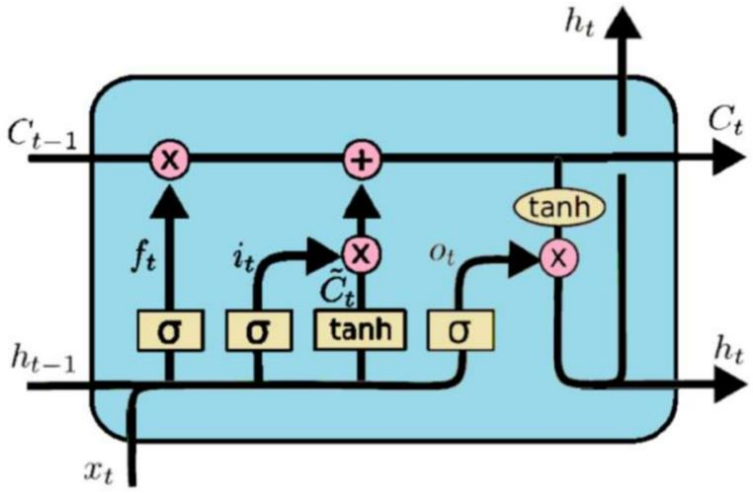

Figura 2. Célula LSTM
Fonte: Graves (2014).

As equações 1, 2, 3 ,4 e 5 descrevem uma rede LSTM.

$$f_t = \sigma_g\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (1)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$c_t = f_t * c_{t-1} + i_t * \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \quad 4$$

$$h_t = o_t * \sigma_h(c_t) \quad 5$$

Onde: σ_g– Função sigmoide, σ_h- Função tangente hiperbólica, W e U- Matrizes de peso e b- bias.

Treinamento, Validação e Teste:

Para criar os subconjuntos de dados, de treinamento, validação e de teste das redes neurais, foram usadas 99 observações da base de dados do IRGA. A tarefa de um modelo de *machine learning* é prever dados que não foram utilizados em sua construção. Portanto, as produtividades relativas as safras de 2017/18, 2018/19 e 2019/20 foram retirados do conjunto de dados, para serem utilizados posteriormente no teste do modelo (conjunto de teste). Neste trabalho utilizou-se o método de fragmentação de *Houldout* onde a base de dados foi dividida com 70% (67) dos dados para treinamento dos algoritmos (Conjunto de Treinamento) e 30% (29) para validação (Conjunto de validação).

Recursos:

Utilizou-se, na implementação do algoritmo LSTM, a linguagem de programação Python e o ambiente de desenvolvimento Jupyter Notebook. As bibliotecas Matplotlib e Pandas foram utilizadas na visualização e análise de dados. O modelo foi implementado por meio da biblioteca Pytorch. Pytorch é uma biblioteca, de *machine learning*, desenvolvida pelo laboratório de inteligência artificial do Facebook (CHUNG, 2020).

Em relação ao *hardware* foi utilizado um notebook com o sistema Windows 7. O notebook conta com um processador Intel(R) Core(TM) i7 e 16 GB de memória RAM.

Métricas:

Neste trabalho utilizaram-se, para avaliar a precisão das previsões, as seguintes métricas (CANKURT; SUBASI, 2012):
MAPE (*Mean Absolute Percentage Error*): O MAPE exprime a porcentagem obtida pela divisão da diferença entre os valores real (y_i) e predito ($\hat{y}_i$) pelo valor real. Quanto menor, o valor do MAPE, mais preciso é o modelo de previsão.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad 6$$

Onde *n* é o número total de previsões.

RMSE (*Root Mean Square Error*): Raiz do erro médio quadrático da diferença entre a predição ($\hat{y}_i$) e o valor real (y_i). Tem sempre valor positivo e quanto mais próximo de zero, maior a qualidade dos valores preditos.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad 7$$

Onde *n* é o número total de previsões.

3. RESULTADOS E DISCUSSÃO

Inicialmente, neste trabalho, realizou-se uma análise descritiva dos dados (Tabela 2).

Tabela 2. Análise descritiva.

Média (kg/ha)	3975,4
Mínimo (kg/ha)	1733
Máximo (kg/ha)	8402
Desvio Padrão (kg/ha)	1841,2
Coeficiente de Variação (%)	46,3

Observa-se, dos dados apresentados na Tabela 2, que a produtividade do arroz, no período em estudo, ficou em média 3975,4 kg/ha. A maior produtividade foi de 8402 kg/ha na safra de 2019/20 enquanto a menor foi de 1733 kg/ha na safra de 1944/45. Pode-se também observar um coeficiente de variação muito alto (46,3%), o que indica variabilidade dos dados (PIMENTEL, 2009).

A série histórica, da produtividade do arroz no estado Rio Grande do Sul, é apresentada na Figura 3. Pode-se notar, por meio desta figura, uma tendência ao aumento, da produtividade de arroz, ao longo dos anos.

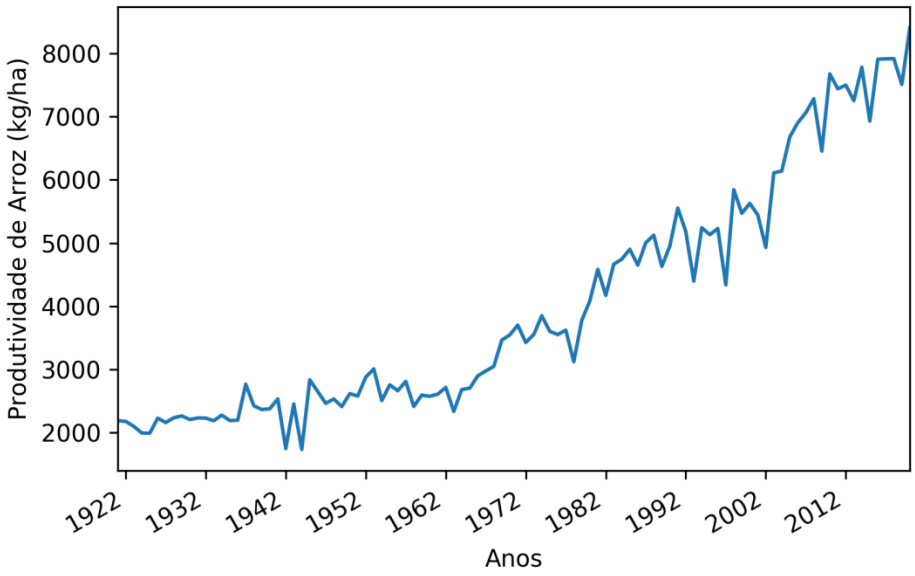

Figura 3. Série: produtividade do arroz

O melhoramento genético com plantas mais resistentes e produtivas e recomendações técnicas, como o controle químico de pragas e doenças e a mecanização nas operações de plantio e colheita, ajudaram no aumento da produtividade de arroz do país (AGROSABER, 2020).

Treinamento e Validação:

Para escolha do melhor modelo, para previsão da produtividade de arroz, vários modelos LSTMs foram treinados. Com base nos resultados das métricas MAPE e RMSE o modelo com melhor desempenho foi selecionado. Este modelo, obtido por meio das métricas, utiliza o algoritmo de otimização Adam (*Adaptive moment*) com os seguintes hiperparâmetros (Tabela 3):

Tabela 3. Hiperparâmetros do modelo.

Hiperparâmetro	Valor do Hiperparâmetro
Activation function of the hidden layer	tanh
Epochs	1516
Dropout	0,01
Hidden layer dimension	8
Number de network layers	2
Activantion function of the output layer	tanh
Learning rate	0,17

Na sequência testou-se o modelo, com os hiperparâmetros da Tabela 3, para os algoritmos de otimização *Stochastic Gradient Descent* (SGD) e *Root Mean Square Propagation* (RMSProp) (Tabela 4).

Tabela 4. Testes dos algoritmos de otimização.

Algoritmo de Otimização	MAPE (%)	RSME
Adam (torch.optim.Adam(model.parameters(),lr)	0,83	71,06
SGD (torch.optim.SGD(model.parameters(),lr)	3,61	248,88
RMSProp (torch.optim.RMSprop(model.parameters(),lr)	4,61	341,64

Pode-se observar, por meio da Tabela 4, que o algoritmo de otimização Adam apresentou um menor valor de MAPE e RSME (0,83% - 71,06), seguido do otimizador SGD (3,61% - 248,88) e RMSProp (4,61% - 341,646).

Na Figura 4 apresenta-se a curva de aprendizagem, do modelo LSTM, utilizando o otimizador Adam. Pode-se observar, nesta figura, a boa estabilidade na convergência das curvas de treino e validação.

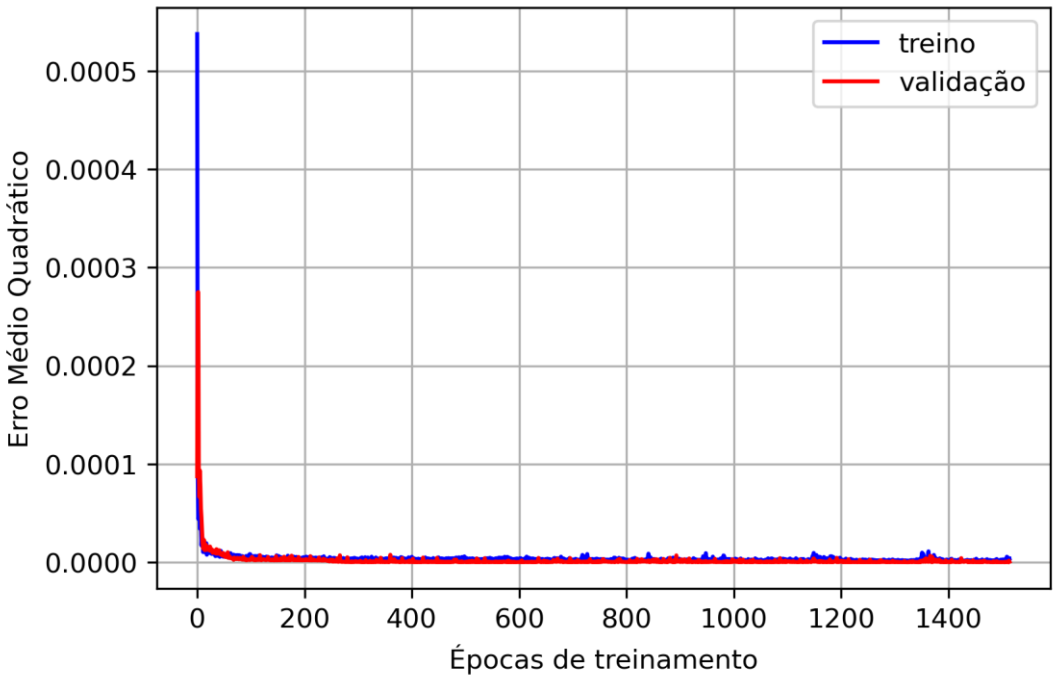

Figura 4. Curvas de aprendizagem de treino e validação

Na Figura 5 apresentam-se os resultados das previsões para o modelo LSTM para o conjunto de validação. Pode-se observar, por meio desta figura, a boa aderência dos dados previstos com os dados reais.

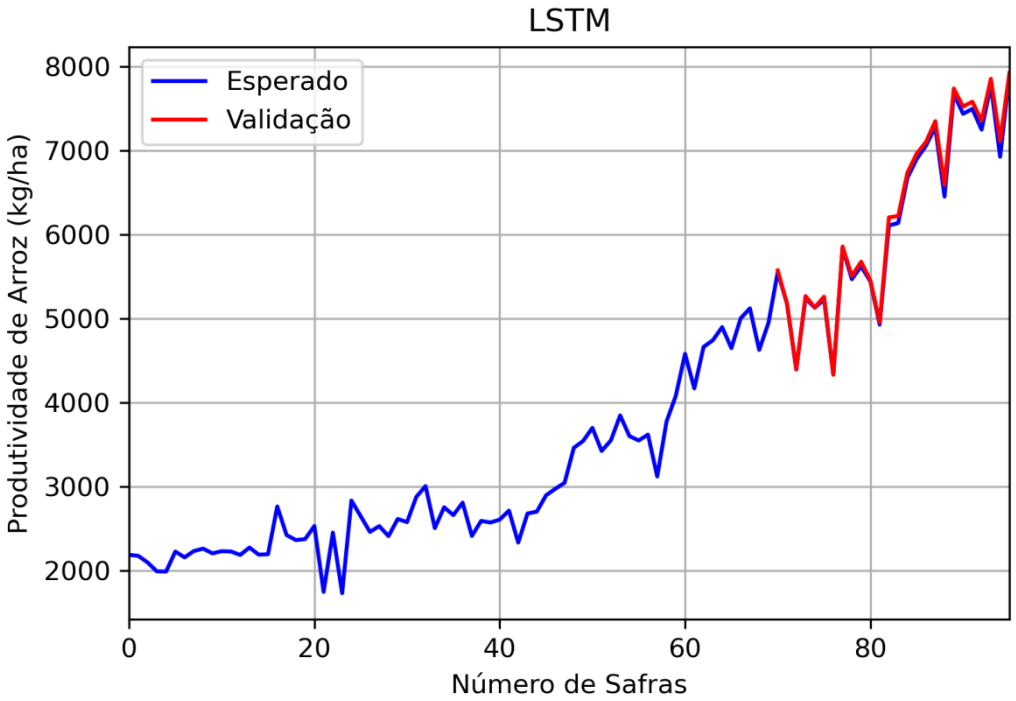

Figura 5. Produtividade do arroz (Conjunto de validação)

Previsões:

Na Tabela 5 apresentam-se os dados, observados, preditos e os Erros Relativos Percentuais (ERP), para as três safras que não participaram da etapa de treinamento e validação. O ERP é obtido por meio da equação:

$$ERP = \left|\frac{y - \hat{y}}{y}\right| \times 100 \qquad 8$$

Onde y é valor observado e $\hat{y}$ o valor predito.

Tabela 5. Resultados: previsão da produtividade (kg/ha) e erros percentuais relativos (%).

Safra	IRGA	LSTM	ERP
2017/18	7917	7961,951	0,567774
2018/19	7508	7469,996	0,506185
2019/20	8400	7961,951	5,214873

Os resultados das previsões, em termos gráficos, são apresentados na Figura 6.

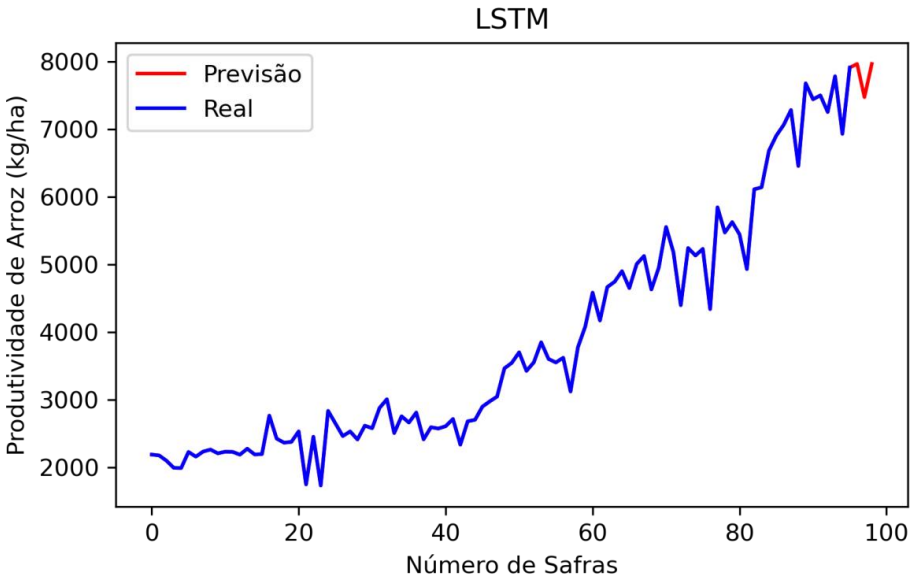

Figura 6. Previsões da Produtividade do arroz (conjunto de teste)

Pode-se notar, por meio dos resultados da Tabela 5, que o modelo LSTM apresentou um ERP baixo, em torno de 0,5% para as safras de 2017/18 e 2018/19. Apresentando um ERP um pouco mais alto para a safra de 2019/20 (5,2%). Portanto, conclui-se, por meio dos resultados apresentados na tabela, que as previsões, do modelo LSTM, estão muito próximas aos fornecidos no site do IRGA.

É importante destacar que o modelo, para a safra 2020/21, previu uma produtividade de arroz de 7420 kg/ha. Segundo o Instituto Riograndense do arroz a produtividade, desta safra, vai ficar um pouco abaixo da safra de 2019/20 (CANAL RURAL, 2021).

4. CONSIDERAÇÕES FINAIS

Neste trabalho, foi proposto um modelo, de redes neurais recorrentes LSTM, para previsão da produtividade de arroz no estado do Rio Grande do Sul. A construção do modelo se baseou em dados, da produtividade de arroz, obtidos pelo Instituto Riograndense do Arroz (IRGA), para as safras de 2021/22 à 2019/2020, totalizando 99 observações.

Inicialmente, baseado nos resultados das métricas, no conjunto de validação, o modelo com melhor desempenho foi selecionado. Observou-se também que o algoritmo de otimização Adam obteve, em comparação com os algoritmos SGD e RMSProp, menores valores de erro (Erro Percentual Médio e Raiz Quadrada do Erro Médio Quadrático). Os resultados obtidos, pelo modelo LSTM, utilizando o algoritmo de otimização Adam, podem

ser considerados relevantes, como demonstrados pela boa aderência dos dados previstos com os dados reais para o conjunto de validação.

Na sequência, observou-se, para as safras que não participaram do processo de treinamento e validação da rede (2017/18, 2018/19 e 2019/20), que as previsões foram bem precisas e as diferenças entre valores reais e preditos foram pequenas. Portanto, a proximidade entre valores preditos e reais demonstram a boa capacidade de generalização, para um horizonte de curto prazo, do modelo LSTM implementado neste trabalho.

Apesar do modelo LSTM apresentar resultados adequados para uma previsão de curto prazo, sugere-se, para outros trabalhos de pesquisa, proceder a estimação, utilizando a biblioteca Pytorch, com outros modelos, tais como: os modelos BLSTM (*Bidirectional Long Short-Term Memory*) e GRU (*Gated Recurrent Unit*).

REFERÊNCIAS

AGROSABER. **Em área 75% menor, Brasil produz 5 vezes mais arroz. Saiba o segredo**. Disponível em: https://agrosaber.com.br/em-area-75-menor-brasil-produz-5-vezes-mais-arroz-saiba-o-segredo/. Acesso em 20 ago. 2020.

ATLAS: Atlas Socioeconômico do Rio Grande do Sul. **O Rio Grande do Sul é o maior produtor de arroz em casca do brasil**. Disponível em: https: //atlassocioeconomico.rs.gov.br/arroz. Acesso em 10 out. 2020.

AWAI, M. A.; SIDDIQUE, M. A. B. Rice production in bangladesh employing by Arima model. **Bangladesh Journal of Agricultural Research**, vol. 36, n. 1, 2011.

BASTIANI, M.; SANTOS, J. A. A.; SCHMIDT, C. A P.; SEPULVEDA, G. P. L. Application of data mining algorithms in the management of the broiler production. **Geintec**, v. 8, 2018.

BRONDANI, G.; VEY, I. H.; MADRUGA, S. R.; TRINDADE, L. L.; VENTURINI, J. C. Diferenciais de custos em culturas de arroz: a experiência do Rio Grande do Sul. **Revista Universo Contábil**, vol. 2, n. 1, 2006.

CANAL RURAL. **Arroz: produtividade no RS está acima da média histórica, diz Irga**. Disponível em: https://www.canalrural.com.br/noticias/agricultura/arroz/arroz-produtividade-rs-safra-21/. Acesso em 5 abr. 2021.

CANKURT, S.; SUBASI, A. Comparasion of linear regression and neural network models forecasting tourist arrivals to turkey. **Eurasian Journal of Science & Engineering**. 2015.

CHUNG K. P. A Robust and Low-cost Computer Method Based on Deep Residual Learning. **Journal of Physics: Conference Series**, vol. 1650, n. 2, 2020.

GRAVES A. **Towards end-to-end speech recognition with recurrent neural networks**. In: 31st International Conference on Machine Learning, Proceedings […]. Beijing: ICML-14, 2014.

HAYKIN, S. **Neural networks: a comprehensive foundation.** New Delhi: Pearson Prentice Hall, 2001.

IRGA: **Instituto Riograndense do Arroz. Área e produção do arroz**. Disponível em: https://irga.rs.gov.br/upload/arquivos/201909/19141756-producao-rs-x-br.pdf. Acesso em 14 abr. 2021.

MARASCA, L.; SOUZA, A. M. Previsão mundial de arroz. **Revista Espacios**, vol. 37, n. 7, 2016.

NELSON, M. Q.; PEREIRA, A. C. M.; OLIVEIRA R. A. **Stock market's price prediction with LSTM neural networks**. In: International Joint Conference of Neural Networks. Proceedings […]. Anchorage: IJCNN , Alaska, 2017.

PÉREZ, S. G. C.; PIRE, R. Estimación del precio internacional del arroz (Oryza sativa L.) bajo el modelo ARIMA. **Revista Mexicana de ciências agrícolas**, vol. 1, n. 1, 2018.

PIMENTEL, F. **Curso de estatística experimental**. Piracicaba: ESALQ, 2009.

PINHEIRO, T. C., SANTOS, J. A. A., PASA, L. A. Gestão da produção de frangos de corte por meio de redes neurais artificiais, **Revista Holos**, 2020.

ROHRIG, B. **Como calcular a estimativa da produtividade de arroz antes da colheita**. Disponível em: https://blog.aegro.com.br/estimativa-da-produtividade-de-arroz/. Acesso em 20 mar. 2021.

SANTOS, I. Z.; TAVARES, M. Eficiência técnica, alocativa e de custos na produção de arroz no Brasil. **Revista Observatório de la Economia Latinoamericana,** vol. 1, n. 1, 2018.

SANTOS, J. A. A.; CHAUKOSKI, Y. Previsão do consumo de energia elétrica na região sudeste: um estudo de caso usando SARIMA e LSTM. **CEREUS**, vol. 12, n. 4, 2020.

WALTER, M.; MARCHEZAN, E.; AVILA, L. A. Arroz: composição e características nutricionais. **Ciência Rural**, vol. 38, n. 4, 2008.

WATTO, M. A.; MUGERA, A. W. Measuring production and irrigation efficiencies of rice farm.: evidence from the Punjab Province, Pakistan. **Asian Economic Journal,** vol. 28, n. 2, 2014.