

REDES NEURAIIS APLICADAS NA PREDIÇÃO DO PREÇO DA SOJA NO ESTADO DO PARANÁ

Jandrei Sartori Spancerski¹ & José Airton Azevedo dos Santos^{1}*

RESUMO

SPANCERSKI, J. S.; SANTOS, J. A. A. Redes neurais aplicadas na predição do preço da soja no estado do Paraná. **Perspectivas Online: Exatas & Engenharias**, v.11, n.32, p.19-32, 2021.

Atualmente, o setor agrícola enfrenta o desafio de crescer, de modo competitivo, para atender a demanda interna e manter o espaço conquistado no mercado externo. Produtores, no mercado competitivo da soja, precisam de ferramentas de previsão de preço. As previsões de preço incorporam informações cruciais no momento da comercialização da safra. Neste contexto, este trabalho tem como objetivo aplicar modelos, baseados em redes neurais artificiais, para previsão do preço da saca de soja no estado do Paraná. A base de dados,

disponibilizada pelo Centro de Estudos Avançados em Economia Aplicada (CEPEA), apresenta uma série de preços mensal compreendido ente janeiro/2000 e agosto/2020. Modelos de previsão, baseados em Redes Neurais LSTM e BLSTM, foram implementados na linguagem Python. Os resultados obtidos, para um horizonte de curto prazo, mostram que os dois modelos de previsão fornecem estimativas confiáveis para o preço da saca de soja.

Palavras-chave: LSTM; BLSTM; previsão de preço.

¹Universidade Tecnológica Federal do Paraná – UTFPR - Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio – PPGTCA - Avenida Brasil, 4232, Medianeira, PR, CEP 85884-000, Brasil.

(*) e-mail: airton@utfpr.edu.br

NEURAL NETWORKS APPLIED IN THE FORECAST OF SOYBEAN PRICE IN PARANÁ STATE

Jandrei Sartori Spancerski¹ & José Airton Azevedo dos Santos^{1}*

ABSTRACT

SPANCERSKI, J. S.; SANTOS, J. A. A. Redes neurais aplicadas na predição do preço da soja no estado do Paraná. **Perspectivas Online: Exatas & Engenharias**, v.11, n.32, p.19-32, 2021.

Currently, the agricultural sector faces the challenge of growing, in a competitive way, to meet domestic demand and maintain the space conquered in the foreign market. Producers, in the competitive soybean market, need price forecasting tools. Price forecasts incorporate crucial information at the time of marketing the harvest. In this context, this work aims to apply models, based on artificial neural networks, to forecast the price of soybeans in the state of Paraná. The database, made available by

the Center for Advanced Studies in Applied Economics (CEPEA), presents a series of monthly prices between January/2000 and April/2020. Prediction models, based on LSTM and BLSTM Neural Network, were implemented in Python. The results obtained, for a short term horizon, show that the two forecast models provide reliable estimates for the price of soya sack.

Keywords: LSTM; BLSTM; price forecast.

¹Federal Technological University of Paraná - UTFPR - Graduate Program in Computational Technologies for Agribusiness - PPGTCA - Avenida Brasil, 4232, Medianeira, PR, Postal Code 85884-000, Brazil.

(*) e-mail: airton@utfpr.edu.br

1. INTRODUÇÃO

Há cerca de cinco mil anos atrás a soja já era cultivada pelos chineses. Espalhou-se pela Ásia, a três mil anos, onde começou a ser utilizada como alimento. Passou a ser cultivada comercialmente, nos Estados Unidos, no início do século XX. Com o desenvolvimento, dos primeiros cultivares comerciais, houve um crescimento muito rápido da produção (EMBRAPA, 2020).

No estado da Bahia, em 1882, iniciou-se o desenvolvimento da soja no Brasil, por meio de materiais genéticos trazidos dos Estados Unidos. A expansão da soja no Brasil começa, nos anos 1970, com o processo de ampliação das indústrias de óleo. Um fator que também contribuiu, para o crescimento da cultura, foi o aumento da demanda internacional pelo grão (APROSOJA, 2020).

Os subprodutos do grão de soja são matérias-primas importantes para diversos setores industriais. Podendo ser transformados em produtos alimentícios. Estes produtos, devido as suas qualidades nutricionais, são considerados alimentos com alto valor nutricional (RUCHS, 2010).

Segundo Feliciane et al. (2015) a previsão futura do preço da soja auxilia o produtor, no processo de tomada de decisão, já que o momento de comercialização da safra ainda é considerado um ponto crítico, na viabilidade e manutenção do seu negócio. As Redes Neurais Artificiais (RNAs) podem ser utilizadas como ferramentas de apoio no processo de tomada de decisão. São ferramentas de grande importância em um sistema de planejamento, pois direcionam a melhor forma de aproveitar todos os recursos disponíveis.

As RNAs são modelos matemáticos inspirados no funcionamento de neurônios biológicos. Podem ser treinadas para aprenderem sobre determinado padrão, sendo capazes de generalizar, ou seja, estimar dados desconhecidos. As redes neurais são capazes de memorizar, analisar e processar um grande número de dados obtidos de um experimento. É uma técnica de modelagem que pode resolver muitos problemas não lineares e complexos (SEBASTIAN, 2016; ABRAHAM, 2019, BASTIANI et al., 2018).

As RNAs têm sido utilizadas com sucesso em tarefas de predição e modelagem de séries temporais. Uma série temporal é qualquer conjunto de observações ordenado no tempo. Um modelo de séries temporais fornece previsões de observações futuras por meio da relação que possui com valores passados (MORETIN; TOLOI, 2004; HAYKIN, 2001).

Diversos trabalhos utilizaram métodos de previsão em aplicações voltadas ao mercado da soja. Dentre eles, podem-se citar os trabalhos de: Fabris et al. (2011) que analisou a produção de soja no Brasil por meio de um Modelo ARIMA. Silva (2002) que realizou uma comparação, entre métodos univariados, na análise do preço médio da soja no Brasil. Castro (2016) que utilizou métodos de previsão univariados e multivariados na previsão de valores de exportação da soja. Feliciane et al. (2015) que utilizou modelos estocásticos para previsão do preço da soja no Brasil. Já Abraham et al. (2019) realizou estimativas, da produção da soja brasileira, por meio de redes neurais artificiais.

Neste contexto, este trabalho tem como objetivo aplicar modelos, baseados em redes neurais artificiais, para previsão do preço da saca de soja no estado do Paraná (US\$/60kg), no período entre Janeiro/2000 e Agosto/2020.

intermediárias, considerando uma sequência *forward* e uma sequência *backward* que são transmitidas para a camada de saída (Figura 2). Com este tipo de rede, a camada de saída pode obter informações de estados passados e futuros simultaneamente. Isto faz com que este tipo de rede possa compreender melhor o contexto.

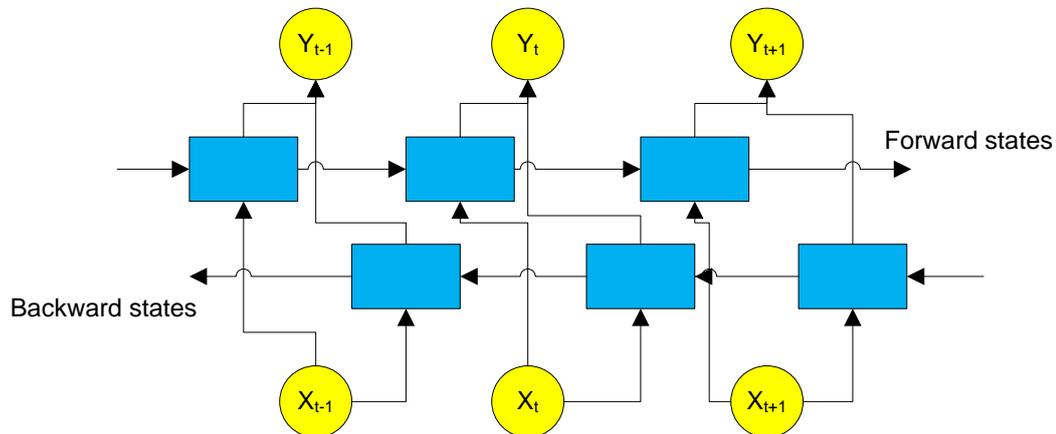


Figura 2: Rede BLSTM. (Fonte: Adaptado de Huang et al., 2015).

1.2. Métricas

Neste trabalho, os modelos de rede neurais são avaliados pelas seguintes métricas (CANKURT; SUBASI, 2015):

Raiz Quadrada do Erro Médio Quadrático (RMSE): Raiz do erro médio quadrático da diferença entre a predição e o valor real. Valor perfeito – RMSE=0.

Erro Médio Absoluto (MAE): Como o RMSE, o MAE possui dimensão igual à dimensão dos valores observados e preditos. Seu valor representa o desvio médio entre observado e predito. Observa-se que o, Erro Médio Absoluto, é muito preciso e menos afetado por *outliers*. Valor perfeito – MAE=0.

2. METODOLOGIA

A metodologia, utilizada neste trabalho, pode ser classificada quanto a natureza, objetivos e a forma de abordar o problema. Quanto a natureza é considerada aplicada, porque gera conhecimentos que podem auxiliar o produtor no momento da comercialização da safra. Quanto aos objetivos descritiva pois descreve as características dos dados do preço da soja e quanto a forma de abordar o problema quantitativa, devido que as informações, relativas ao problema em estudo, são quantificadas através de técnicas estatísticas e modelagens computacionais (GIL, 2017).

2.1. Base de dados

Para previsão do preço da saca de soja (US\$/60kg) utilizou-se uma base de dados com

248 meses (jan/2000 - ago/2020) obtida do Centro de Estudos Avançados em Economia Aplicada (CEPEA, 2020). Os preços relativos aos meses de maio, junho, julho e agosto de 2020 foram retirados do conjunto de dados, para serem utilizados posteriormente para testar os modelos. Os dados obtidos, da base de dados, já estavam limpos e sem a presença de *outliers* (Figura 3).

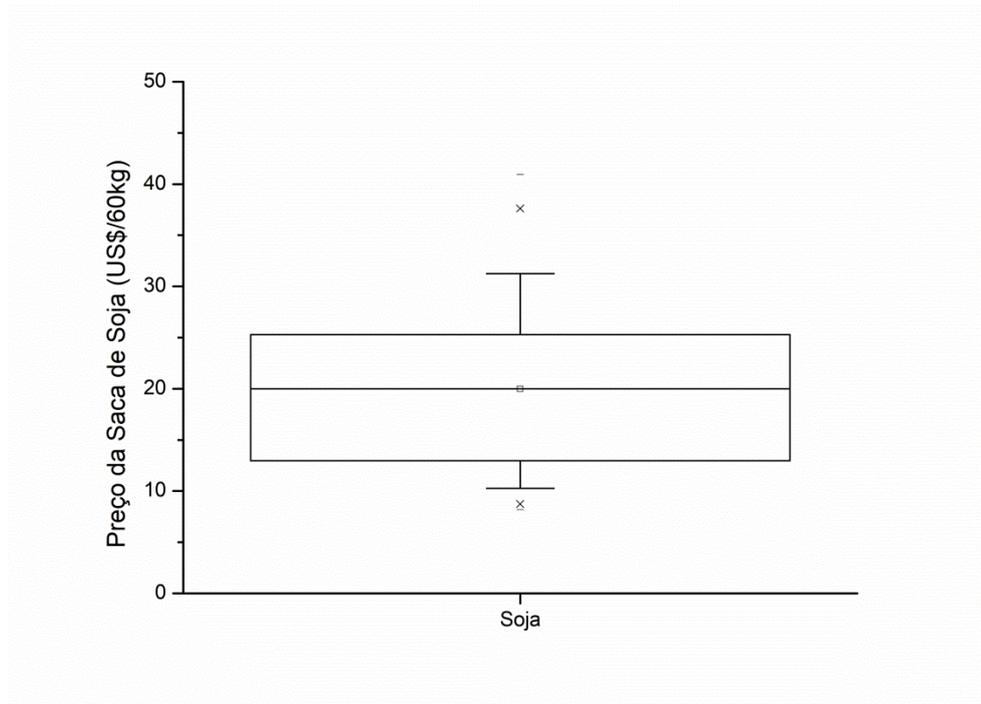


Figura 3: *Boxplot* dos dados.

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

Tabela 1: Dez primeiros registros do arquivo do CEPEA. Fonte: CEPEA (2020)

Data	Soja (US\$/60kg)
2000-1	11,0515
2000-2	10,86619
2000-3	10,4745
2000-4	10,573
2000-5	10,68636
2000-6	10,26905
2000-7	9,796667
2000-8	9,88087
2000-9	10,3515
2000-10	10,12429

A série histórica, obtida do CEPEA, é apresentada na Figura 4.

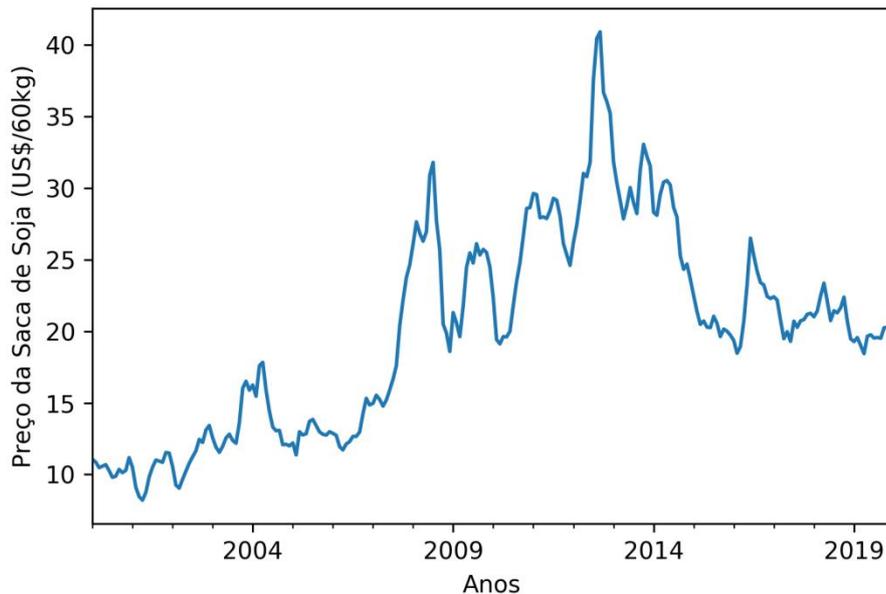


Figura 4: Série - Preço da saca de soja (US\$/60kg). (Fonte: CEPEA., 2020).

A partir da segunda metade de 2006, a soja, assim como outras *commodities* agrícolas, entrou em uma trajetória de crescimento de preço. Obtendo, em 2008, um ápice de seu preço. Desvalorizando-se, logo após, devido a crise financeira mundial.

O preço da soja se recuperou em agosto de 2012, onde atingiu uma nova marca histórica, o que representou uma elevação de 12,41% em relação ao pico anterior (FEE, 2020). Em 2020, nos oito primeiros meses, manteve-se, em média, entorno de 19,48 US\$/60kg (CEPEA, 2020).

2.2. Recursos

Para implementação dos algoritmos foi utilizado o Jupyter Notebook. Um aplicativo interativo *Web*, de código aberto, que permite que você escreva e execute códigos de computador. As bibliotecas de aprendizado de máquina: Scikit-learn, Pandas, Matplotlib e Keras foram também utilizadas. Scikit-learn é uma biblioteca de aprendizado de máquina, para a linguagem Python, que inclui vários algoritmos de classificação, regressão e agrupamento. Matplotlib é a mais famosa biblioteca de visualização de dados do Python. A biblioteca Pandas fornece ferramentas de análise de dados e estrutura de dados de alta performance. Keras é uma biblioteca para rede neural de alto nível, escrita em Python, e roda como *frontend* em TensorFlow ou Theano.

2.3. Etapas do Trabalho

Iniciou-se, este trabalho, pela análise descritiva dos dados obtidos pelo Centro de Estudos Avançados em Economia Aplicada (CEPEA). A análise exploratória ajuda a extrair informações importantes de um conjunto de dados. Na sequência, na etapa de treinamento e teste, foram comparados, por meio das métricas, vários modelos LSTM e BLSTM.

Selecionando-se os modelos com melhor desempenho. Finalmente, na etapa de teste, os modelos LSTM e BLSTM foram testados com os dados que não participaram do processo de treinamento e teste.

3. RESULTADOS E DISCUSSÃO

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

Tabela 2: Análise descritiva de dados.

Média (US\$/60kg)	19,99
Mínimo (US\$/60kg)	8,2
Máximo (US\$/60kg)	40,9
Desvio Padrão (US\$/60kg)	7,15
Coefficiente de Variação (%)	35,76

Pode-se observar, dos dados apresentados na Tabela 2, que o preço, para o período em estudo, ficou em média de 19,99 US\$/60kg. Apresentando, neste período, preços mínimo e máximo de 8,20 US\$/60kg e 40,9 US\$/60kg, respectivamente.

Observa-se também, da Tabela 2, que o coeficiente de variação é 35,76%, considerado alto, o que indica variabilidade dos dados.

Neste trabalho vários modelos de redes neurais, LSTM e BSLTM, foram testados utilizando como indicadores de desempenho: o erro médio quadrático e o erro médio absoluto. As redes neurais foram treinadas com 163 amostras (67%) e validadas com 81 amostras (33%). Os melhores modelos (Tabela 3) utilizam o algoritmo de otimização Adam com os seguintes hiperparâmetros: *cells*=16, *batch*=1, *learning rate*=0.001 e *activate*=relu. Definiu-se, como critério de parada do treinamento, a função *EarlyStopping()* com o parâmetro *patience*=18. O parâmetro *patiente* indica o número de épocas, após a qual nenhuma melhoria foi observada.

Tabela 3: Resultados de Treino e Teste – Modelos LSTM e BLSTM.

Cenário	Treino		Validação		Tempo (s)
	RMSE	MAE	RMSE	MAE	
LSTM	1,22	0,86	1,16	0,95	44,79
BLSTM	1,26	0,88	1,07	0,86	44,46

Pode-se observar, dos dados apresentados na Tabela 3, que os resultados obtidos pelos modelos (RMSE e MAE), para a fase de treino e teste, são similares. Observou-se, também, que o tempo de processamento ficou muito próximo para os dois modelos.

Na sequência, na Figura 5, apresentam-se os gráficos de dispersão, dos valores preditos em relação aos valores observados, para os dois modelos. Pode-se observar, dos gráficos, que

os dados observados (CEPEA) apresentam uma boa correlação com os dados preditos, já que os pontos estão próximos a linha de ajuste.

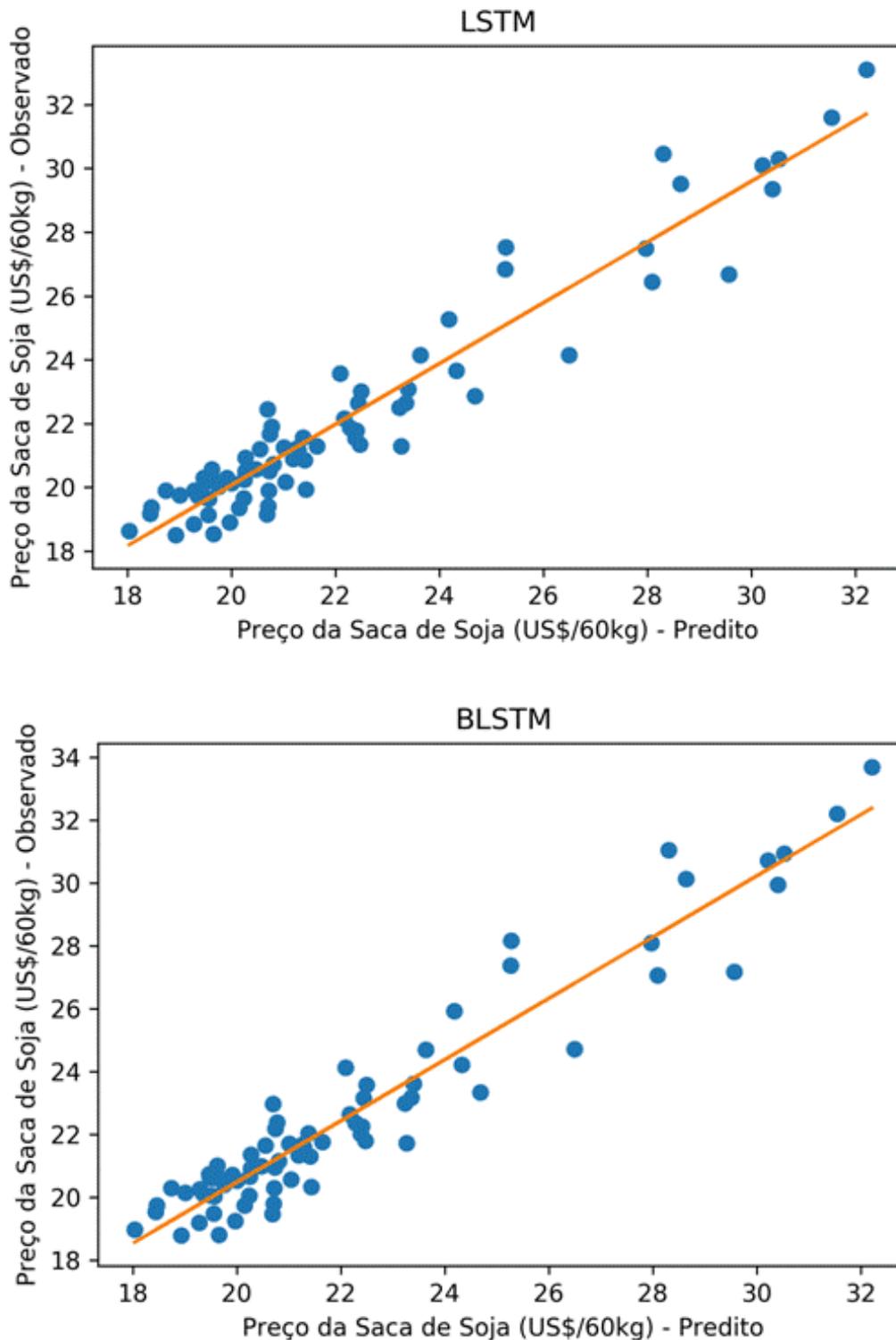


Figura 5: Dados preditos em função dos dados observados.

Pode-se também notar, por meio da Figura 6 apresentada a seguir, a boa aderência dos dados previstos com os dados reais para os dois modelos.

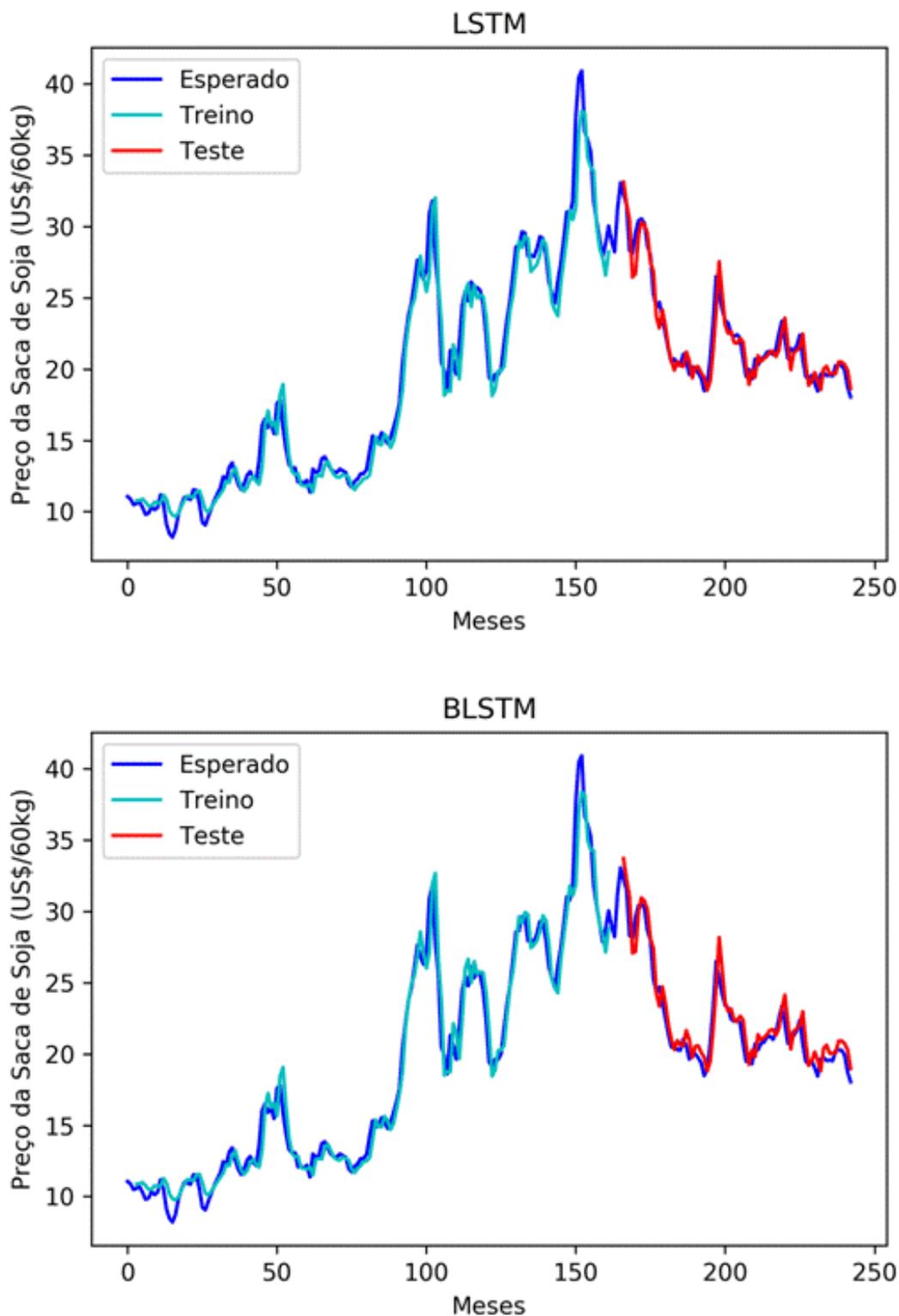


Figura 6: Previões – treino e teste.

3.1. Previsões

Na Tabela 4 apresentam-se os dados, observado, preditos e os Erros Relativos Percentuais (ERP), para os quatro meses que não participaram da etapa de treinamento e teste. O ERP é obtido por meio da equação:

$$ERP = \frac{Observado - Predito}{Observado} \times 100 \quad 1$$

Tabela 4: Preço da Saca de Soja: Observados e Preditos (US\$/60kg) e Erros Percentuais Relativos (%).

Mês	CEPEA	LSTM	BLSTM	ERP-LSTM	ERP-BLSTM
mai/2020	18,33	18,2	18,36	0,709219858	0,163666121
jun/2020	19,85	18,73	20,02	5,64231738	0,856423174
jul/2020	20,73	19,45	20,96	6,174626146	1,109503136
ago/2020	22,43	20,45	21,86	8,827463219	2,541239412
Média				5,338406651	1,16770796

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

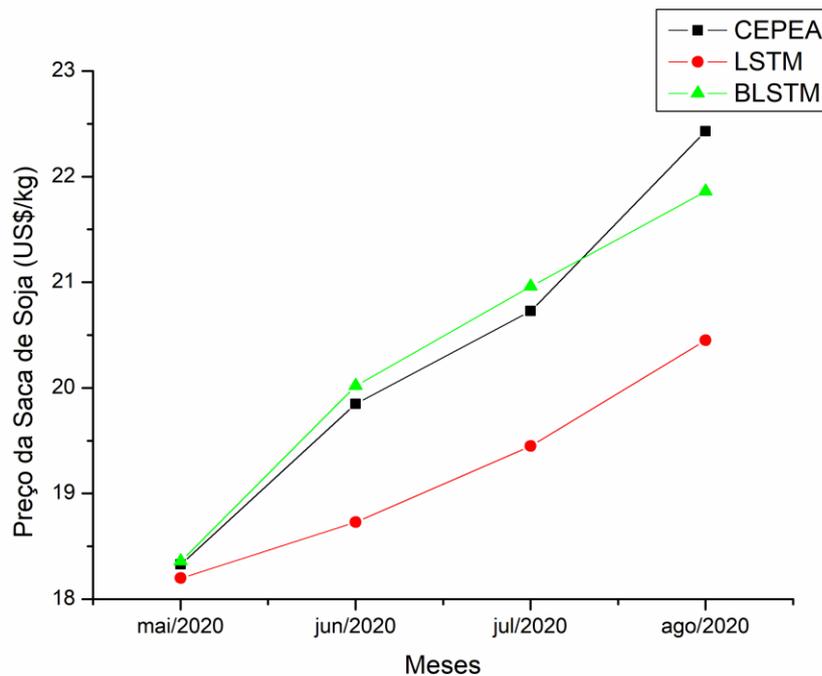


Figura 6: Preços da saca de Soja (CEPEA, LSTM e BLSTM).

Por meio dos resultados apresentados, na Tabela 4 e na Figura 6, conclui-se que os

resultados das previsões, dos dois modelos, estão próximos aos fornecidos pelo CEPEA. Contudo, o modelo BLSTM apresenta um erro percentual relativo médio menor que o modelo LSTM (ERP-LSTM= 5,33% e ERP-BLSTM= 1,16%).

5. CONCLUSÕES

Neste trabalho apresentou-se uma aplicação, de modelos de Redes Neurais Artificiais, para previsão do preço da saca de soja no estado do Paraná. A série de preços da saca de soja, no período de janeiro/2000 a agosto/2020 (248 observações mensais), foi obtida pelo Centro de Estudos Avançados em Economia Aplicada (CEPEA).

Inicialmente, realizou-se uma comparação entre os modelos, BLSTM e LSTM, implementados na linguagem Python. Os modelos passaram pelas fases de preparação de dados, definição da estrutura, estimativas, avaliação dos resultados e validação dos modelos. Os resultados obtidos, dos dois modelos, podem ser considerados relevantes, como demonstrados pela boa aderência dos dados previstos com os dados reais.

Na sequência, observou-se, para os meses que não participaram do treinamento da rede (maio, junho, julho e agosto de 2020), 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, dos modelos implementados neste trabalho. Não obstante, observou-se uma leve superioridade na qualidade do ajuste do modelo BLSTM.

Apesar dos modelos LSTM e BLSTM apresentarem resultados adequados para uma previsão de quatro meses, sugere-se, para outro trabalho de pesquisa, proceder a estimação, utilizando a biblioteca Keras, com outros modelos, tais como: os modelos MLP (*Multilayer Perception*) e GRU (*Gated Recurrent Unit*).

6. REFERÊNCIAS

ABRAHAM, B. **Statistical methods for forecasting**. New York: Wiley & Sons, 2019.

APROSOJA. **Associação Brasileira de Produtores de Soja**. A Soja. Disponível em: <<https://aprosojapr.com.br/a-soja/>>. Acesso em: 2 mar. 2020.

BASTIANI, Miliano et al. Application of data mining algorithms in the management of the broiler production. **Revista GEINTEC-Gestão, Inovação e Tecnologias**, v. 8, n. 4, p. 4574-4587, 2018.

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

CASTRO, Nicole Rennó et al. Modelos univariados e multivariados aplicados à previsão de valores de exportação: uma análise comparativa para o complexo soja. **Revista ESPACIOS**. v. 27, n. 3, 2016.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). **Preço do frango CEPEA/ESALQ - Estado SP**. Disponível em: <<https://www.cepea.esalq.usp.br/br/indicador/frango.aspx>>. Acesso em: 19 mar. 2020.

CHEVALIER, G. **The Long Short-Term Memory (LSTM) cell can process data sequentially and keep its hidden state through time.** CC-BY-SA-4.0, Disponível em: <https://commons.wikimedia.org/wiki/File:The_LSTM_Cell.svg>. Acesso em: 2 fev. 2021.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). **Soja.** Disponível em: <<https://www.embrapa.br/soja>>. Acesso em: 2 mar. 2020.

FABRIS, S. R.; FABRIS, J. P.; DULLIUS, A.I.S. Análise da Produção da Cultura do Soja no Brasil Através dos Modelos ARIMA. **Revista GEINTEC-Gestão, Inovação e Tecnologias**, v. 1, n. 2, p. 49-56, 2011.

FEE. **O preço da soja nos últimos 10 anos.** Disponível em: <<http://panoramainternacional.fee.tche.br/article/o-preco-da-soja-no-ultimo-decenio/>>. Acesso em: 2 jun. 2020.

FELICIANI, A. V.; SOUZA, A. M.; SOUZA, F. M. Modelos estocásticos de previsão dos preços da soja no Brasil. **Custos e agronegócio online**, v. 11, n. 1, p. 263-280, 2015.

GIL, A. C. **Como elaborar projetos de pesquisa.** 6 ed. São Paulo: Atlas, 2017.

GRAVES, A.; JAITLEY, N. Towards end-to-end speech recognition with recurrent neural networks. In: **International conference on machine learning.** PMLR, 2014. p. 1764-1772.

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

HUANG, Zhiheng; XU, Wei; YU, Kai. Bidirectional LSTM-CRF models for sequence tagging. **arXiv preprint arXiv:1508.01991**, 2015.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de series de temporais.** São Paulo: Edgard Blucher, 2004.

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 (IJCNN).** Anchorage, Alaska, 2017.

QINGNAN, S., JANKOVIC, M. V., BALLY, L. **Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network.** arXiv:1809.03817, Cornell University, 2018.

RUCHS, E. L. **Modelo Matemático para tomada de decisões no processo produtivo e de esmagamento da soja.** Dissertação de Mestrado – UNIJUÍ. Ijuí: 2010.

SEBASTIAN, Sumam. Performance evaluation by artificial neural network using WEKA. **International Research Journal of Engineering and Technology**, v. 3, n. 3, p. 1459-1464, 2016.

SILVA, W. Comparação entre os métodos de previsão univariados para o preço médio da soja no Brasil. In: **XXII Encontro Nacional de Engenharia de Produção.** Curitiba, 2002.

TUBINO, D. F. **Planejamento e controle da produção-Teoria e Prática.** São Paulo: Atlas, 2009

ZHAO, Z. et al. LSTM network: a deep learning approach for short-term traffic forecast. **IET Intelligent Transport Systems**, v. 11, n. 2, p. 68-75, 2017.