REVISTA Mundi
Engenharia, Tecnologia e Gestão
INSTITUTO FEDERAL Paraná

APLICAÇÃO DE MODELOS DE SÉRIES TEMPORAIS NA PREVISÃO DO PREÇO DO FARELO DE SOJA

APPLICATION OF TIME SERIES MODELS TO FORECAST SOYBEAN MEAL PRICE

José Airton Azevedo dos Santos[1]

Resumo: O mercado da soja tem como uma de suas características a flutuação do preço do produto. Tal característica decorre de fatores que estão fora do controle do produtor, como variações na oferta e na demanda, intempéries climáticas, etc. Neste contexto, este trabalho tem como objetivo avaliar a eficácia de modelos de séries temporais, na sua forma univariada, na previsão do preço do farelo de soja no estado do Paraná. A base de dados, disponibilizada pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), apresenta uma série histórica, do preço do farelo de soja, no período entre 2011 e 2020, totalizando 111 observações. Modelos de previsão, baseados em Redes Neurais LSTM (*Long Short-Term Memory*) e ARIMA (*Auto-Regressive Integrated Moving Average*), foram implementados na linguagem Python. Resultados obtidos, dos dois modelos, foram comparados. Verificou-se, para um horizonte de curto prazo, que os dois modelos de previsão fornecem estimativas confiáveis para o preço do farelo de soja.

Palavras-chave: LSTM. ARIMA. Farelo de soja.

Abstract: One of the characteristics of the soybean market is the price fluctuation of the product. This characteristic stems from factors that are outside the control of the producer, such as variations in supply and demand, weather conditions, etc. In this context, this work aims to evaluate the effectiveness of time series models, in their univariate form, in forecasting the price of soybean meal in the state of Paraná. The database, made available by the Brazilian Agricultural Research Corporation (EMBRAPA), presents a historical series, of the price of soybean meal, in the period between 2011 and 2020, totaling 111 observations. Prediction models, based on Neural Networks LSTM (Long Short-Term Memory) and ARIMA (Auto-Regressive Integrated Moving Average), were implemented in the Python language. Results obtained from the two models were compared. We found, for a short-term horizon, that the two forecast models provide reliable estimates for the price of soybean meal.

Keywords: LSTM. ARIMA. Soybean meal.

[1] Doutor em Engenharia Elétrica, PPGTCA, UTFPR, Medianeira, airton@utfr.edu.br.

1 INTRODUÇÃO

A soja é um dos grãos mais produzidos no mundo, junto com o milho, o trigo e o arroz. Seus grãos são muito utilizados pelas agroindústrias para produção de óleo vegetal e ração animal.

O farelo de soja é o principal subproduto resultante do esmagamento da soja para a retirada do óleo. Aproximadamente 90% dos grãos consumidos são direcionados ao processo de esmagamento. O farelo de soja, junto com o milho, constitui a matéria-prima essencial para fabricação de rações. O farelo é considerado um suplemento rico em proteínas para criação de animais. Segundo Hirukuri e Lazzaroto (2014), a demanda por soja em grão e seu principal produto derivado é dependente do mercado de carnes.

Oscilações constantes, no preço da soja, representam um fator de insegurança no controle dos custos de produção de suínos e frangos de corte (BORDIN, 2012). Portanto, projeções do preço do farelo de soja, são muito importantes, para os produtores de aves e suínos, já que influenciam diretamente nos custos de produção.

Métodos de previsão de séries temporais, como RNAs (Redes Neurais Artificiais) e ARIMA (Auto Regressivo Integrado de Média Móvel), têm sido utilizados com muito sucesso em tarefas de predição (CRISTALDO, 2018). Segundo Morettin e Toloi (2004), uma série temporal é um conjunto de observações ordenadas no tempo.

O modelo ARIMA, desenvolvido nos anos 1970 por George Box e Gwilym Jenkins, é um modelo muito utilizado em previsões de séries temporais. Baseia-se no ajuste dos valores observados, visando reduzir para próximo de zero a diferença dos valores produzidos no modelo e os valores observados (SATO, 2013).

As redes neurais artificiais (RNAs), também conhecidas como sistema de processamento paralelo distribuído, tem seu surgimento marcado na década de 80. São modelos computacionais inspirados no funcionamento do cérebro humano. São capazes de memorizar, analisar e processar um grande número de dados obtidos de um experimento (SEBASTIAN, 2016; ABRAHAM et al., 2019, BASTIANI et al., 2018; HAYKIN, 2001).

Diversos trabalhos utilizaram métodos de previsão de preços para *commodities* agrícolas. Dentre eles, podem-se citar os trabalhos de: Tibulo (2014) que realizou uma comparação entre modelos, de séries temporais, aplicados a uma série histórica do preço médio mensal do milho no Rio Grande do Sul. Darekare e Reddy (2017) que realizaram a previsão do preço da soja, na Índia, por meio de modelos ARIMA. Silva (2018) que utilizou a metodologia Box-Jenkins para previsão do preço da commodity café arábica. Já Zhang et al. (2018) realizaram previsões do preço da soja, na China, usando redes neurais artificiais.

Neste contexto, este trabalho tem como objetivo avaliar a eficácia de modelos de séries temporais, na sua forma univariada, na previsão do preço do farelo de soja no estado do Paraná, no período entre 2011 e 2020

2 MODELOS ARIMA E LSTM

2.1 ARIMA

O modelo ARIMA, desenvolvido por George Box e Gwilym Jenkins, é um modelo muito utilizado na previsão de variáveis econômicas, mercadológicas e sociais (CERETTA et al., 2010). Em certos casos, o nome ARIMA e Box-Jenkins são utilizados como sinônimos. O algoritmo ARIMA é aplicável aos dados com correlação alta e estável e tem um bom desempenho para previsões simples e de curto prazo (SATO, 2013).

Modelos ARIMA, geralmente denotados como ARIMA(p,d,q), utilizam uma combinação de *p* valores passados da variável dependente *y* para prever o valor seguinte, combinados com *q* erros observados nas previsões anteriores. Segundo Arêdes (2008), a maioria das séries econômicas são não estacionárias, a aplicação ARIMA(p,d,q) exige a transformação das mesmas por *d* diferenças para torná-las estacionárias. A expressão matemática do modelo ARIMA é dada por:

$$\Delta y_t = \emptyset + \sum_{i=1;p} \beta_i \Delta_{t-i} + u_t + \sum_{i=1;q} \alpha_i u_{t-i} \qquad 1$$

Onde: $\Delta_{(t-i)}$ é o operador de diferenças, u o termo de erro e $\emptyset$, β e α são parâmetros do modelo.

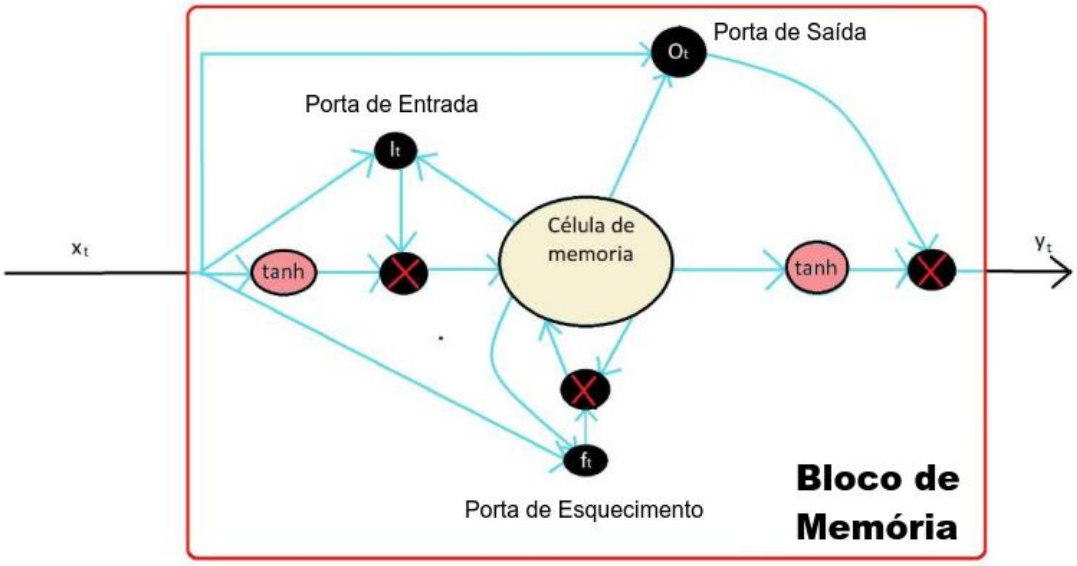
2.2 LSTM

As redes neurais recorrentes, *Recurrent Neural Network* (RNN), são usadas em tarefas que envolvem entradas sequenciais, como fala, linguagem, séries temporais, entre outras. LSTMs (*Long Short-Term Memory*) são redes neurais recorrentes capazes de aprender dependências de longo prazo.

A topologia de um neurônio, de uma rede LSTM, é baseada em um bloco de memória. Cada bloco contém três tipos de portas que gerenciam o fluxo de dados da rede neural. Essas portas determinam quais dados devem entrar, quais devem sair e quais devem ser esquecidos (GRAVES, 2014; NELSON et al., 2017).

Na Figura 1 apresenta-se um bloco de memória. Este bloco contém uma célula de memória e três portas, uma de entrada, uma de saída e uma de esquecimento, identificadas por um ”X”. Estas portas cuidam do fluxo de informações que entram e saem da célula de memória.

Figura 1 – Bloco de memória



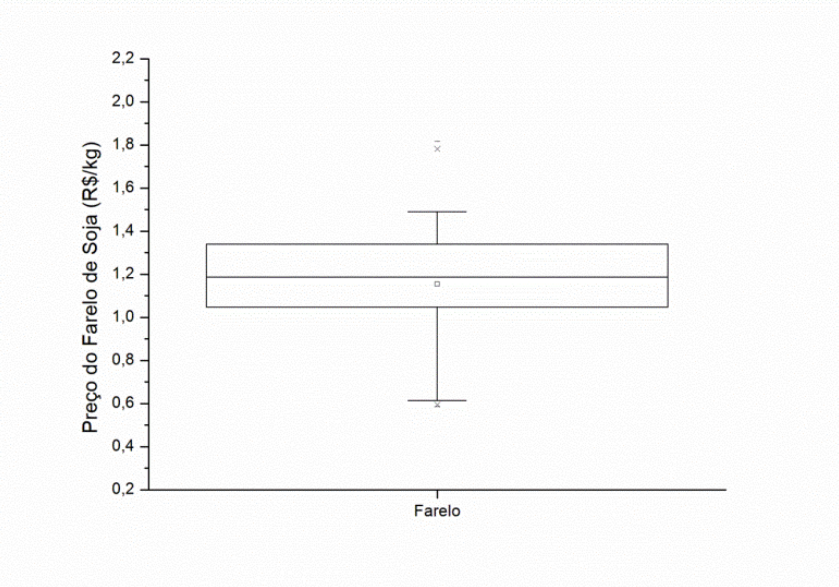
Fonte: Adaptado de Lopéz (2018).

3 MATERIAIS E MÉTODOS

3.1 Coleta de Informações - Base de dados

Para previsão, do preço do farelo, utilizou-se uma base de dados, com 111 meses (Jan/2011 - Mar/2020), obtida da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA, 2020). Os dados obtidos, da base de dados, já estavam limpos e sem a presença de *outliers*. Na Figura 2 apresenta-se o *boxplot* dos dados.

Figura 2 – *Boxplot*



Fonte: O Autor (2020).

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

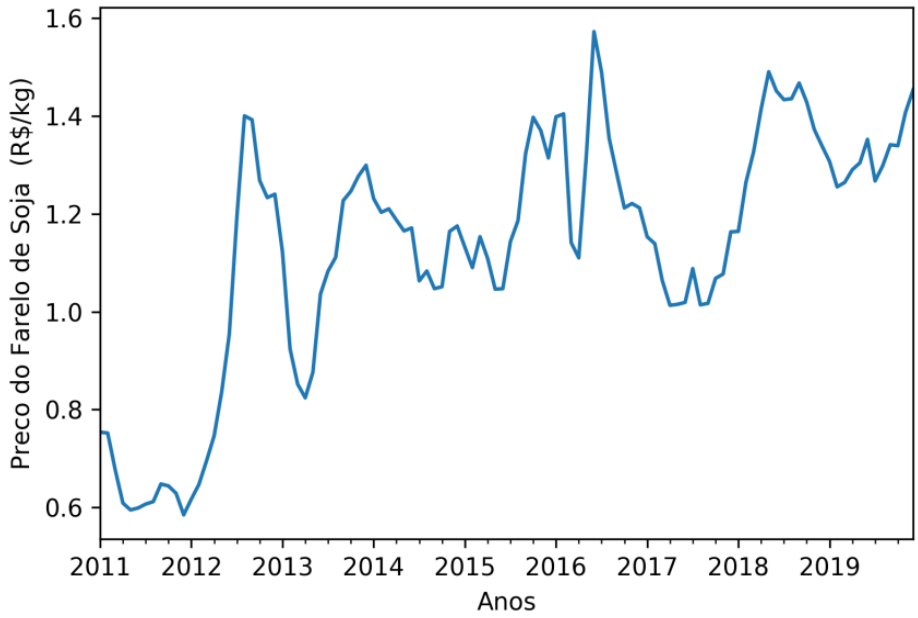
Tabela 1 – Dez primeiros registros do arquivo da EMBRAPA

Data	Farelo de Soja (R$/kg)
2011-1	0,754
2011-2	0,752
2011-3	0,676
2011-4	0,609
2011-5	0,595
2011-6	0,599
2011-7	0,607
2011-8	0,612
2011-9	0,648
2011-10	0,644

Fonte: EMBRAPA (2020).

O gráfico da série histórica é apresentado na Figura 3.

Figura 3 – Gráfico ilustrativo da série temporal do preço do farelo de soja



Fonte: O Autor (2020).

3.2 Método de Fragmentação e critério de parada – Redes Neurais

Para criar os subconjuntos de dados, de treinamento e teste, foram usados 108 observações da base de dados da EMBRAPA. Observa-se que os preços relativos aos meses de Janeiro, Fevereiro e Março de 2020 foram retirados do conjunto de dados, para serem utilizados posteriormente para testar os modelos. Neste trabalho utilizou-se o método de fragmentação de *Houldout* onde a base de dados foi dividida com 67% (72) dos dados para treinamento dos algoritmos e 33% (36) para validação.

Como critério de parada, para a rede LSTM, utilizou-se o método conhecido como Método de Parada Antecipada (*Earling Stopping*). Segundo Silva (2018), este método ajuda a projetar uma rede neural com bom poder de generalização. Neste contexto, definiu-se neste trabalho, como critério de parada do treinamento, a função *EarlyStopping()* com o parâmetro

patience=50. O parâmetro *patience* indica o número de épocas, após a qual nenhuma melhoria foi observada.

3.3 Métricas

Os modelos implementados, neste trabalho, foram avaliados pelas métricas apresentadas no Quadro 1 (CANKURT; SUBASI, 2015; PINHEIRO, 2020).

Quadro 1 – Métricas

Coeficiente de Determinação (r^2)
A qualidade de ajuste de um modelo pode ser avaliada pelo coeficiente de determinação. Este coeficiente indica quanto o modelo foi capaz de explicar os dados coletados.
Raiz Quadrada do Erro Médio Quadrático (RMSE)
Raiz do erro médio quadrático da diferença entre a predição e o valor real. Tem sempre valor positivo e quanto mais próximo de zero, maior a qualidade dos valores preditos.
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.

Fonte: Cankurt e Subasi (2015).

3.3 Etapas de previsão

Segundo Cruz et al. (2016), as etapas importantes para realizar uma previsão são:

1. Definição do problema;
2. Coleta de informações;
3. Análise preliminar dos dados;
4. Escolha e ajuste de modelos;
5. Uso e avaliação do modelo.

Inicialmente, neste trabalho, definiu-se o objeto de estudo e realizou-se a coleta de dados. Na sequência, na próxima seção, serão apresentados: a análise exploratória dos dados, os ajustes dos modelos e suas avaliações.

4 RESULTADOS E DISCUSSÃO

4.1 Análise dos dados

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

Tabela 2 – Análise descritiva dos dados

Resumo Descritivo	Valores
Média (R$/kg)	0,503
Mínimo (R$/kg)	0,343
Máximo (R$/kg)	0,833
Desvio Padrão (R$/kg))	0,123
Coeficiente de Variação (%)	24,45

Fonte: O Autor (2020).

Pode-se observar, dos dados apresentados na Tabela 2, que o preço, para o período em estudo, ficou em média 0,503 R$/kg. Apresentando, neste período, preços mínimo e máximo de 0,343 R$/kg e 0,833 (R$/kg), respectivamente.

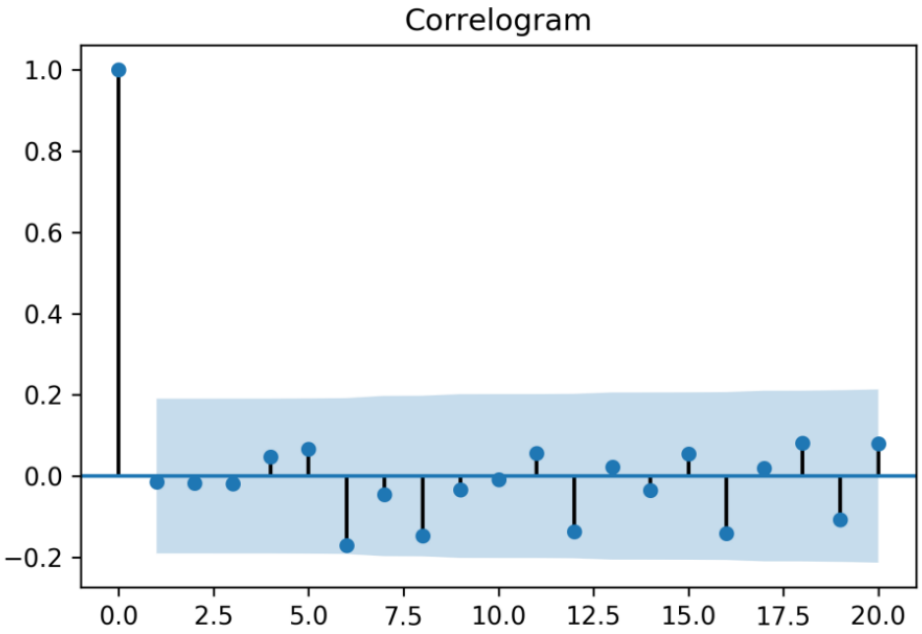
Observa-se também, da Tabela 2, que o coeficiente de variação é 24,25%. Considerado alto, segundo Pimentel (2009), o que indica variabilidade dos dados.

4.1 Escolha e ajuste dos modelos

ARIMA:

Inicialmente, identificou-se o modelo ARIMA(2,1,2), que obteve, dos modelos testados, o menor valor do critério de AKAIKE (AIC). Na sequência, verificou-se a normalidade e a autocorrelação dos resíduos (Figura 4). Obteve-se, do teste de normalidade (*Jarque-Bera normality test*), um p-valor de 0,1, o que revela a não rejeição da hipótese nula de normalidade dos resíduos. Observa-se, do correlograma, que os resíduos não são autocorrelacionados, pois os coeficientes de autocorrelação são estatisticamente iguais a zero, isto é não ultrapassam os limites de confiança.

Figura 4 – Correlograma dos resíduos



Fonte: O Autor (2020).

LSTM:

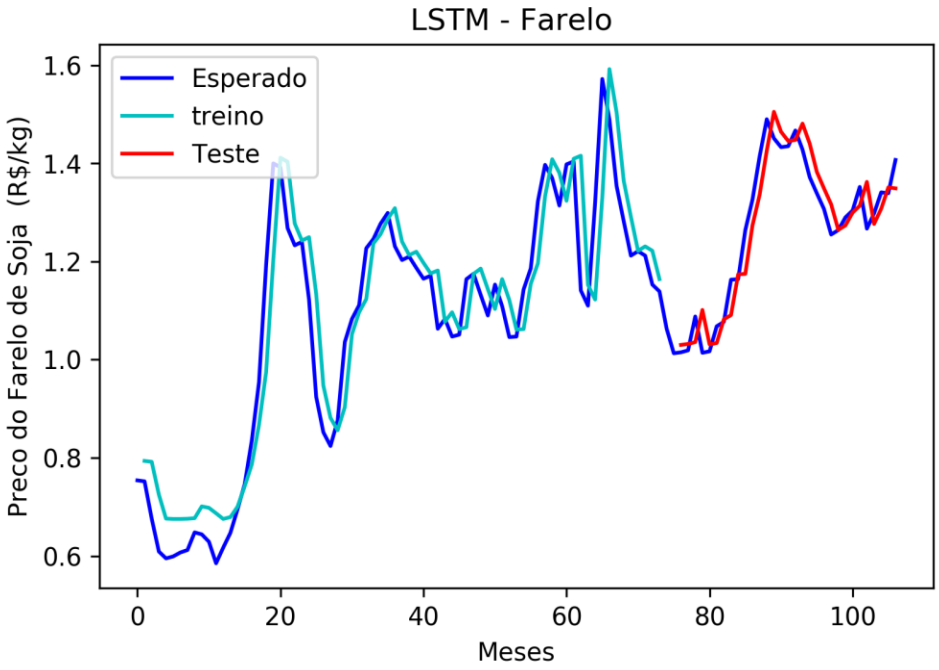
Os modelos de redes neurais LSTM foram implementados por meio da biblioteca Keras, rodando como *frontend* em TensorFlow. Neste trabalho, para obter a melhor arquitetura de rede, variou-se os seguintes parâmetros: *LSTM cells* (4, 8, 12, 16, 20); *batch* (1, 10, 20, 30, 40); *learning rate* (0.1, 0.01, 0.001), *optimizer* (SGD, Adam, Adamax, RMSprop) e *active* (relu, sigmoid, tanh, softmax).

O melhor modelo LSTM utilizou o algoritmo de otimização Adam com os seguintes hiperparâmetros: *LSTM cells* = 16, *batch* = 1, *learning rate* = 0.001 e *activate* = relu.

Para o conjunto de teste o modelo apresentou as seguintes métricas r^2= 0,85, RMSE= 0,059 e MAE= 0,045.

Na Figura 5 apresenta-se o gráfico das previsões para o modelo LSTM.

Figura 5 – Previsão de treino e teste – LSTM



Fonte: O Autor (2020).

4.2 Avaliações dos modelos

Na sequência, realizaram-se previsões, do preço do farelo, para os meses, de Janeiro, Fevereiro e Março de 2020, que não participaram das etapas de treinamento e teste (Tabela 3).

Tabela 3 – Previsões Janeiro/Fevereiro e Março de 2020 (R$/kg)

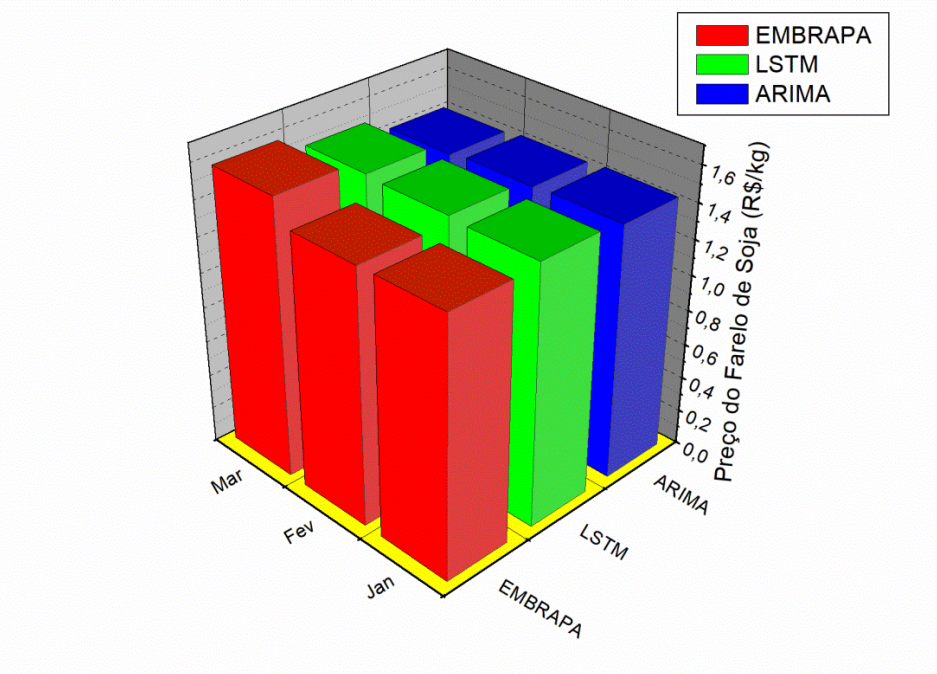
Mês	EMBRAPA	LSTM	ARIMA
jan/20	1,448	1,466	1,44
fev/20	1,443	1,48	1,43
mar/20	1,577	1,495	1,41

Fonte: O Autor (2020).

Por meio dos resultados apresentados, na Tabela 3, conclui-se que os resultados das previsões, dos dois modelos, estão muito próximos aos fornecidos pela EMBRAPA.

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

Figura 6 – Gráfico ilustrativo das previsões: Janeiro/Fevereiro e Março de 2020



Fonte: O Autor (2020).

5 CONSIDERAÇÕES FINAIS

Neste trabalho apresentou-se uma aplicação, de modelos de séries temporais, para previsão do preço do farelo de soja. A série de preços do farelo de soja, no período entre 2011 e 2020, foi fornecida pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). Os modelos passaram pelas fases de: preparação de dados, definição das estruturas, estimativas, avaliação dos resultados e validação.

Observou-se, do gráfico de treinamento e teste, que os dados preditos, pela rede neural, confirmam a tendência apresentadas pelas variáveis reais. A partir dos modelos ARIMA e LSTM, foram estimados os valores referentes aos meses que não participaram da etapa de treinamento e teste (Janeiro, fevereiro e Março de 2020). Observou-se 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.

REFERÊNCIAS

ARÊDES, A. F.; PEREIRA, M. W. G. Potencialidade da utilização de modelos de séries temporais na previsão do preço do trigo. **Revista de Econômia Agrícola**, v. 55, n. 1, p. 63-76, Jan./jun. 2008.

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

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.

BORDIN, R. A. (2012) **Nutrição animal: a relação entre preços de insumos e produção de carne.** Disponível em: https://www.aviculturaindustrial.com.br/imprensa/nutricao-animal-a-relacao-entre-precos-de-insumos-e-producao-de-carne-por-roberto-bordin/20121024-135056-r586. Acesso em: 08 ago. 2020.

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

CERETTA, P. S.; RIGHI, M. B.; SCHELENDER, S. G. Previsão do preço da soja: uma comparação entre os modelos ARIMA e redes neurais artificiais. **Informações Econômica**, v. 40, n. 9, set. 2010.

CRISTALDO, M. F. **Aplicação de inteligência artificial à previsão de cheias de pequenas bacias.** 2018. Tese (Doutorado em Meio Ambiente e Desenvolvimento Regional) – Programa de Pós-Graduação em Meio Ambiente e Desenvolvimento Regional, Universidade Anhanguera-UNIDER, Campo Grande, 2018.

CRUZ, K. S.; BATTISTI, J. F.; JUNIOR, G. L.; WEISE, A. D. Previsão da produção brasileira de biodiesel por meio de modelos de previsão. **Revista Espacios**, v. 37, 2016.

DAREKARE, A.; REDDY A. A. Predicting market price of soybean in major India studies through ARIMA model. **Journal of Food Legumes**, v. 30, n. 2, p. 73-76, Dez. 2017.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). **Soja.** Disponível em: <https://www.embrapa.br/soja>. Acesso em: 02 mar. 2020.

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

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

HIRUKURI, M. H.; LAZZAROTO, J. J. **O agronegócio da soja no contexto mundial e brasileiro.** Londrina: Embrapa, 2014.

LÓPEZ, M. G. **Aplicación de modelos de redes neuronales al modelado y predición del precio de la electricidad em Espana.** Madri: UPM, 2018.

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. **Proceedings…** International Joint Conference of Neural Networks (IJCNN), 2017.

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.

PIMENTEL, F. Curso de estatística experimental. Piracicaba: ESALQ, 2009.
SATO, R. C. Gerenciamento de doenças utilizando séries temporais com o modelo ARIMA. **Einstein**, v. 11, n. 1, jan./mar. 2013.

SEBASTIAN, S. Performance evaluation by artificial neural network using WEKA. **International Research Journal of Engineering and Technology**, v. 3, 2016.

SILVA, C. A. G. Previsão do preço da commodity café arábica: uma aplicação da metodologia Box_jekins. **Revista Espacios**, v 39, n. 4, 2018.

TIBULO, C. Previsão do preço do milho, através de series temporais. **Scientia Plena**, v. 10, n. 10, 2014.

ZHANG, D.; ZANG, G.; LI, J.; MA, KA; LIU, H. Predction of soubean price in china using QR-RBF neural network model. **Computers and Electronic in Agriculture**, v. 154, p. 10-17, Nov. 2018.

Enviado em: 12 nov. 2020

Aceito em: 13 jul. 2021

Editores responsáveis: Bianca Neves Machado / Mateus das Neves Gomes