

Seleção de hiperparâmetros para uma rede neural Multi-Layer Perceptron aplicada na predição do preço da soja

TITLE: HYPERPARAMETERS SELECTION FOR A MULTI-LAYER PERCEPTRON NEURAL NETWORK APPLIED TO PREDICT THE SOYBEAN PRICE

TÍTULO: SELECCIÓN DE HIPERPARÁMETROS PARA UNA RED NEURONAL MULTI-LAYER PERCEPTRON APLICADA EN LA PREDICCIÓN DEL PRECIO DE LA SOJA

Luiz F. M. Brunetto

luiz.mb@aluno.ifsc.edu.br

LABICON, Instituto Federal De Santa Catarina (IFSC)

Carise E. Schmidt

carise.schmidt@ifsc.edu.br

LABICON, Instituto Federal De Santa Catarina (IFSC)

Alexandre Dalla'Rosa

alexandredr@ifsc.edu.br

Instituto Federal De Santa Catarina (IFSC)

Resumo

O agronegócio brasileiro representa uma parcela significativa do Produto Interno Bruto (PIB), além de ser um setor de produção essencial, por garantir o fornecimento de alimentos para a população. Contudo, o principal desafio enfrentado pelo setor é manter a rentabilidade, ante um cenário de alta dos custos de insumos. Diante disso, demonstra-se importante avaliar ferramentas que possam auxiliar na tomada de decisão, com foco no aumento da rentabilidade. A proposta deste trabalho é selecionar um conjunto de hiperparâmetros para uma Rede Neural Multi-Layer Perceptron (MLP), por meio da análise do Erro Médio Quadrático (EMQ), visando à sua utilização na predição do preço da soja. Para isso, um histórico de dados de preço do grão e da taxa de câmbio do dólar (2006-2023) é utilizado como entrada do modelo matemático. Fazendo uso de simulações computacionais, o preço predito da soja é avaliado, objetivando identificar o conjunto de hiperparâmetros que gera o menor EMQ. Entre as configurações avaliadas, o menor EMQ encontrado quando o modelo é testado na série histórica (2006-2021) tem valor igual a 1,2936. Em um período de validação superior a 21 meses, o erro absoluto máximo observado é de US\$ 0,83, o que representa cerca de 2% do valor real da saca de soja. Esse resultado mostra que a configuração de hiperparâmetros selecionada é capaz de realizar a previsão do preço futuro da soja, acompanhando a tendência de preços do grão, no período em que ele foi avaliado.

Palavras-chave: *Inteligência artificial. Aprendizado de máquina. Séries temporais. Commodities.*

Abstract

Brazilian agribusiness represents a significant portion of the Gross Domestic Product (GDP), in addition to being an essential production sector, as it guarantees the supply of food for the population. The main challenge faced by the sector is to maintain profitability, in the face of a scenario of high input costs. Therefore, it is important to evaluate tools that can help in decision making, focusing on increasing profitability. The goal of this work is to select a set of hyperparameters for a Multi-Layer Perceptron Neural Network (MLP) through the analysis of the Mean Square Error (MSE), aiming for its use in pre-

dicting the price of soybeans. In order to do it, we use historical data (2006-2023) of the grain price and dollar exchange rate as input to the mathematical model. We evaluate the predicted soybean price through computer simulations to identify the set of hyperparameters that generates the lowest MSE. Among the evaluated configurations, the lowest MSE generated is equal to 1.2936. Over a validation period of more than 21 months, the maximum absolute error obtained is US\$ 0.83, which represents approximately 2% of the real price of the sack of soybeans. This result shows that the hyperparameters configuration selected is able to predict the future soybean price in the period in which it is evaluated, following the grain price trend.

Keywords: *artificial intelligence; machine learning; time series; commodities.*

Resumen

El agronegocio brasileño representa una parte significativa del Producto Interno Bruto (PIB), además de ser un sector productivo esencial, ya que garantiza el abastecimiento de alimentos para la población. Sin embargo, el principal desafío al que se enfrenta el sector es mantener la rentabilidad en un escenario de aumento de los costes de los insumos. Ante esto, es importante evaluar herramientas que puedan auxiliar en la toma de decisiones, centrándose en incrementar la rentabilidad. El objetivo de este trabajo fue seleccionar un conjunto de hiperparámetros para una Red Neuronal Multi-Layer Perceptron (MLP), a través del análisis del Error Cuadrático Medio (EMQ), con el fin de utilizarla en la predicción del precio de la soja. Para ello, se utilizaron datos históricos del precio de la soja y de la tasa de cambio del dólar (2006 - 2023) como entrada para el modelo matemático. Mediante simulaciones computacionales, se evaluó el precio predicho de la soja con el objetivo de identificar el conjunto de hiperparámetros que genera el menor EMQ. Entre las configuraciones evaluadas, el EMQ más bajo encontrado cuando se prueba el modelo en la serie histórica (2006-2021) tiene un valor igual a 1,2936. En un período de validación de más de 21 meses, el error absoluto máximo observado fue de US\$ 0,83, lo que representa alrededor del 2% del valor real de una bolsa de soja. Este resultado muestra que la configuración de hiperparámetros seleccionada es capaz de predecir el precio futuro de la soja, siguiendo la tendencia del precio del grano, en el período en que se realizó la evaluación.

Palabras clave: *inteligencia artificial; aprendizaje automático; series de tiempo; commodities.*

Introdução

O agronegócio é um dos setores de destaque na economia brasileira e compreende todas as atividades ligadas à produção de alimentos. A soja tem especial relevância dentro desse setor, já que o Brasil ocupa, atualmente, o posto de maior produtor mundial do grão, tanto em relação à área plantada como quantidade produzida (Embrapa, 2022). Em 2021, o setor do agronegócio alcançou um patamar recorde de participação na composição do Produto Interno Bruto (PIB), correspondente a 27,6%. Contudo, apesar da previsão de uma safra de grãos recorde, o PIB do segmento primário do agronegócio registrou uma queda de 2,48% no primeiro trimestre de 2022, reflexo da forte alta de custos com insumos (CNA, 2022). O aumento nos custos de fertilizantes, sementes, operações de mecanização e defensivos resultou em um crescimento de 67,21% no custeio agrícola da safra de 2022/2023 (IMEA, 2022). Esse custeio é utilizado por grande parte dos produtores, sendo o principal aliado da atividade rural. Toda essa combinação de fatores gera um grande desafio para os produtores de soja, já que ela interfere diretamente na rentabilidade do negócio.

Para o produtor, em especial, é importante dispor de uma previsão do preço de venda do seu produto, como uma métrica para orientar as decisões de produção e comercialização (Fang *et al.*, 2020). Monitorar a volatilidade das variáveis do mercado, tal como o preço de *commodities*, não é uma tarefa fácil, assim como qualquer previsão financeira (Ge *et al.*, 2022). Há uma série de fatores (bolsa de valores, políticas externas, taxa de câmbio, clima, tecnologia, estoques mundiais) que influencia o preço de *commodities* agrícolas, como a soja, dificultando a previsão de preços futuros (Wang *et al.*, 2019). Contudo, entre esses fatores, Rosolen, Araújo e Lyrio (2013) destacam a taxa de câmbio como um elemento que contém informações relevantes para a formação de preço de *commodities*. Tendo em vista que os custos de produção são variáveis ao longo das safras, e fortemente influenciados pelo dólar, é importante encontrar formas de proteção para o produtor, que garantam a sua lucratividade.

Nesse contexto, modelos de previsão são ferramentas importantes, pois podem auxiliar na redução de incertezas relacionadas às oscilações de preços futuros. Na literatura da área, modelos computacionais de inteligência artificial envolvendo algoritmos supervisionados de aprendizagem de máquina têm sido amplamente estudados para previsão de preços de *commodities* (Wang; Gao, 2018; Zhang *et al.*, 2018; Sunarya; Henderi; Tasyriqan, 2019; Ribeiro; Coelho, 2020; Fang *et al.*, 2020; Yu, 2021; Menhaj; Kavooosi-Kalashami, 2022; Wang *et al.*; 2022).

De acordo com o estudo de Ludovico (2020), as Redes Neurais Artificiais (RNA) se destacam nessa área, podendo ser aplicadas isoladamente ou combinadas com outros métodos (Jha; Sinha, 2013; Pinheiro; Senna, 2017; Disconzi, 2018; Spacerski; Santos, 2021). As RNA apresentam características relevantes para tarefas que compreendem previsão e classificação, incluindo o tratamento de não-linearidades e a capacidade de generalização, adaptação, correlação e aprendizado. E essa combinação de características facilita o tratamento de séries temporais.

Contudo, um dos principais desafios de trabalhar com modelos de aprendizagem de máquina, incluindo a RNA, é definir a configuração de hiperparâmetros do modelo que permite otimizar o seu desempenho (Goodfellow; Bengio; Courville, 2016). Para isso, propõe-se esta pesquisa, com o objetivo de selecionar um conjunto de hiperparâmetros para uma Rede Neural *Multi-Layer Perceptron* (MLP), por meio da análise do Erro Médio Quadrático (EMQ), visando à sua utilização na predição do preço da soja.

Material e métodos

A presente pesquisa foi elaborada em quatro etapas: seleção e organização dos dados de entrada; implementação da rede neural utilizada e definição do conjunto de hiperparâmetros de teste; realização de simulações computacionais e obtenção do EMQ na saída da rede para todas as configurações de hiperparâmetros consideradas; seleção do conjunto de hiperparâmetros que com menor EMQ e validação do modelo em um conjunto de dados independente.

Organização dos dados de entrada

Duas séries históricas foram utilizadas como entrada da rede MLP: preço, em dólar, de fechamento diário da saca de soja (60 Kg) no Porto de Paranaguá, e valor diário de fechamento do dólar comercial, em reais, ambos obtidos do Centro de Estudos Avançados em Economia Aplicada (CEPEA, 2022). As séries contemplam o maior período de dados

disponíveis (13/03/2006 a 10/10/2023), representando 4378 dias de negociação. No tratamento dos dados foram retiradas apenas duas dessas informações, correspondentes a datas com algum dado faltante. Após essa seleção, os dados remanescentes foram ordenados por data de ocorrência das operações. A série possui preço máximo da saca de 60 kg, em dólar, igual a \$45,32 e mínimo de \$12,40 (Gráfico 1).

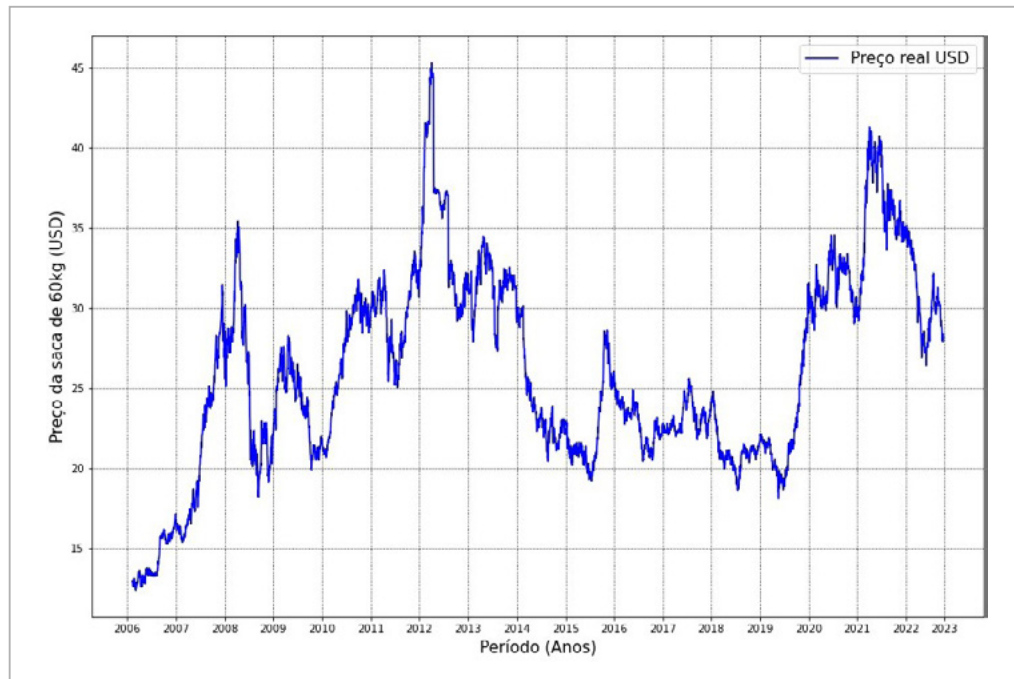


Figura 19 – Série histórica do preço da soja, em dólar

Fonte: elaborado pelos autores, com base nos dados do CEPEA.

Em seguida, os dados de preço da saca de soja e do dólar comercial foram normalizados para que a faixa dinâmica do neurônio da RNA esteja adequada a um intervalo possível de ser previsto por todas as funções de ativação testadas. O normalizador de mínimo e máximo foi aplicado para redimensionar linearmente cada um dos dados ao intervalo [0, 1]. Para avaliar o desempenho da rede MLP, sujeita a diferentes combinações de hiperparâmetros, a primeira parte do conjunto de dados foi subdividida em quatro subconjuntos, sendo o 1º período de 13/03/2006 a 31/12/2009; o 2º período de 01/01/2010 a 31/12/2013; o 3º período de 01/01/2014 a 31/12/2018; e o 4º período de 01/01/2018 a 31/12/2021. Essa subdivisão da série de dados foi realizada com o intuito de avaliar todas as combinações de hiperparâmetros, de forma independente, em cada períodos, através de simulação computacional e cômputo do EMQ na saída da RNA. Por fim, o conjunto de dados restantes, relativos ao período de 03/01/2022 a 10/10/2023, foi utilizado para validação do modelo final.

Nesta pesquisa, foi considerada uma predição após dez dias. Isso significa que, para cada entrada de dados, foi prevista uma saída de dez dias a frente. A estratégia utilizada considera, principalmente, o contexto de produtores que precisam comercializar seu grão para pagamento de insumos a serem utilizados na safra atual.

Arquitetura e topologia da rede neural

Foi utilizada uma rede neural do tipo MPL (HAYKIN, 2008), com topologia *feed-forward* e algoritmo de treinamento *backpropagation*, com pesos atualizados a partir

do método do gradiente descendente. A atualização dos pesos foi realizada por bloco e a dimensão do bloco constituiu um dos hiperparâmetros de teste do modelo.

Na arquitetura, optou-se por utilizar uma única camada oculta, visando não aumentar demasiadamente a complexidade da rede. O número de neurônios dessa camada também constituiu um hiperparâmetro de teste do modelo. Os algoritmos foram implementados em linguagem de programação Python. Todas as simulações computacionais foram conduzidas em um processador AMD Rayzen 7 rodando a 3.4 GHz com 8 GB de memória instalada e sistema operacional Windows.

Hiperparâmetros da rede neural

Com o intuito de melhorar o desempenho da rede, foram definidos alguns hiperparâmetros de teste (Tabela 1), que foram avaliados de forma combinada, testando todas as combinações possíveis e selecionando aquela que resultou em menor Erro Médio Quadrático (EMQ).

Hiperparâmetro	Variações testadas
Taxa de aprendizagem	{0,00001; 0,0001; 0,001; 0,01; 0,1}
Otimizador	Adam; RMSprop
Dimensão do bloco (<i>batch size</i>)	{4, 8, 16}
Função de ativação	Sigmoide; tangente hiperbólica; ReLU
Número de neurônios na camada oculta	{2; 3}

Tabela 3 – Hiperparâmetros da rede explorados nos testes computacionais

O número de épocas foi fixado em 1.000 para possibilitar comparação entre as diferentes configurações de hiperparâmetros testadas no modelo, porém com possibilidade de parada antecipada em 50 épocas. Essa escolha foi empírica, observando a quantidade de dados e considerando velocidade e capacidade de processamento do *hardware*.

Conjuntos de treinamento, teste e validação

Para cada um dos quatro períodos considerados na etapa de seleção dos hiperparâmetros, os subconjuntos de dados utilizados para treinamento, teste e validação foram assim definidos: utilizou-se o próprio período como conjunto de validação, enquanto os demais períodos foram usados para treinamento e teste. Por exemplo, quando o 1º período estava em avaliação, seu subconjunto foi definido como sendo o conjunto de validação, enquanto os demais períodos foram usados para gerar os dados de treinamento e teste (80% para treinamento e 20% para teste). Essa separação foi aleatória, visando aumentar a generalização do modelo durante o treinamento dos dados.

Para validação do modelo final, utilizou-se o período restante (03/01/2022 – 10/10/2023). Nesta etapa, já definidos os hiperparâmetros da rede neural, os dados compreendidos entre 13/03/2006 e 31/12/2021 foram usados para treinamento e teste na proporção de 80% e 20%, respectivamente, enquanto os demais dados da série foram utilizados para validação do modelo.

Seleção do conjunto de hiperparâmetros

Com base na estrutura de rede neural utilizada e nos hiperparâmetros testados (Tabela 1), foram realizadas simulações computacionais para todas as configurações, visando avaliar o desempenho do modelo em cada caso. Esse processo foi repetido para cada um dos quatro períodos considerados. Dessa forma, na primeira fase de seleção foram testadas configurações de hiperparâmetros distintas para cada período. Além disso, cada configuração foi testada 10 vezes, com o intuito de diminuir a aleatoriedade dos dados e reduzir a interferência da sua distribuição na definição do modelo. A cada repetição, os dados presentes nos conjuntos treinamento e teste foram submetidos a uma nova seleção aleatória, mas a proporção (80% - 20%) foi mantida. Para avaliar o erro do modelo na predição do conjunto de teste, foi obtido o EMQ entre as saídas previstas pela rede e as saídas reais, para as 10 repetições. A partir disso, foram selecionados as configurações de hiperparâmetros do modelo com menor EMQ para cada um dos quatro períodos considerados.

As 30 configurações selecionadas passaram então por uma segunda e nova fase de análise, onde cada uma foi reavaliada. Nesta fase, o conjunto de dados de validação foi testado, já que nenhum dos seus dados passaram pelo treinamento da rede, enquanto todos os demais períodos foram considerados dados de treinamento. Para exemplificar, quando o 1º período estava em avaliação, ele próprio foi tratado como conjunto de teste, enquanto o 2º, 3º e 4º períodos foram considerados como dados de treino. A proporção de dados de treinamento e teste manteve-se, respectivamente, em 80% e 20%. Novamente, em cada período avaliado, foi calculado o EMQ para as 10 repetições e então selecionado o conjunto de hiperparâmetros que apresentou a menor EMQ.

Por fim, a terceira e última fase compreendeu a seleção de um único conjunto de hiperparâmetros, entre os quatro selecionados na fase anterior, um de cada período analisado. O objetivo foi identificar qual entre eles possuía a melhor capacidade de generalização, avaliada por meio do EMQ. Para isso, cada uma dessas configurações foi avaliada em parte da série histórica (2006-2021), sendo selecionado o conjunto de hiperparâmetros que apresentou o menor EMQ. O restante da série histórica (2022-2023) foi utilizado para validação do modelo final.

Resultados e Discussão

As configurações de hiperparâmetros do modelo com menor EQM, conforme período, geradas na segunda fase de seleção são apresentadas a seguir (Tabela 2).

Hiperparâmetros de teste	Variações testadas			
	I (1º período)	II (2º período)	III (3º período)	IV (4º período)
Taxa de aprendizagem	0,0001	0,0001	0,0001	0,0001
Otimizador	Adam	MRSprop	Adam	Adam
Dimensão do bloco (<i>batch size</i>)	16	16	16	16
Função de ativação	Tanh	ReLU	Tanh	Sigmoide
Número de neurônios na camada oculta	3	3	2	3
Erro Médio Quadrático	1,5221	1,8580	0,8110	0,7137
Erro Relativo Percentual	3,88%	3,14%	2,79%	2,57%

Tabela 4 – Hiperparâmetros dos modelos com menor média de EMQ para cada período avaliado

Conforme esses resultados, é possível observar que nas configurações selecionadas a taxa de aprendizagem e a dimensão do lote foram as mesmas para todos os períodos. Em relação ao otimizador e os número de neurônios da camada oculta, três dos quatro conjuntos selecionados utilizaram as mesmas opções. A função de ativação foi o hiperparâmetro que teve maior variação entre os modelos selecionados, de forma que, em apenas dois dos quatro modelos ela se repetiu. Em relação ao EMQ, ele variou de 0,7137 até 1,5221 US\$2 entre os períodos considerados. Em termos percentuais, a variação média entre o valor real e o predito por esses conjuntos de hiperparâmetros oscilou entre 2,57 e 3,88%, nos períodos em questão. É possível observar que o erro foi menor no quarto período, que corresponde ao intervalo de 01/01/2018 a 31/12/2021. A configuração que gerou esses resultados possui taxa de aprendizagem igual a 0,0001, otimizador Adam, tamanho de lote igual a 16, função de ativação sigmoide e é composto por 3 neurônios na camada oculta.

Na sequência, são apresentados gráficos comparativos entre os dados reais e os dados previstos pelos conjuntos de hiperparâmetros selecionados em cada período considerado. Eles mostram o preço real da saca de soja e o preço predito pelo modelo, ambos em dólares, além do erro absoluto entre os dois preços, para cada dia de previsão. O Gráfico 2 exibe os dados relativos ao 1º período (2006 a 2010).



Figura 20 – Previsão e erro absoluto da configuração de hiperparâmetros do 1º período em comparação aos dados reais (2006-2010)

Conforme pode ser observado nesse gráfico, a maior diferença entre o preço real e o previsto ocorreu no dia 20/06/2008, quando a saca de soja atingiu o maior valor do período (US\$ 35,46), enquanto o valor predito foi de US\$ 32,40. O erro absoluto entre esses valores é de US\$ 3,06, equivalente a 8,63%. Embora o erro observado nesse ponto seja maior que a média para o período, é necessário observar que a volatilidade média no ano de 2008 também foi superior à volatilidade média do período (1,29%).

Para o 2º período (2010 a 2014), o Gráfico 3 apresenta os resultados do modelo com o conjunto de hiperparâmetros selecionado em comparação aos dados reais, além do erro absoluto.



Figura 21 – Previsão e erro absoluto da configuração de hiperparâmetros do 2º período em comparação aos dados reais (2010-2014)

Os resultados indicaram que a maior diferença ocorreu no dia 10/09/2012, quando o preço real da saca de soja foi de US\$ 44,47 dólares, enquanto o preço previsto pelo modelo foi de US\$ 39,01 dólares. O erro absoluto entre esses pontos é de US\$ 5,46, ou 12,28%. Embora o erro pontual seja desarmônico com o erro observado no período correspondente (0,92%), ele acompanha a volatilidade de 12% que ocorreu nesse dia. Outro momento onde a mesma situação acontece foi no dia 09/01/2013, quando o preço real observado foi de US\$ 37,05 e valor previsto de US\$ 31,85. Para esse dia, o erro absoluto correspondeu a US\$ 5,20, porém a volatilidade diária também foi elevada (14,04%). O Gráfico 4 ilustra os resultados para o 3º período (2014 a 2018).



Figura 22 – Previsão e erro absoluto da configuração de hiperparâmetros do 3º período em comparação aos dados reais (2014-2018)

Para o período em questão, a maior diferença entre o preço real e o predito pelo modelo ocorreu no dia 07/01/2014, quando a saca de soja foi negociada por US\$ 31,05 e o preço predito foi de US\$ 28,50. O erro absoluto nesse dia de negociação correspondeu a US\$ 2,55, ou ainda 8,21%. Assim como nos períodos anteriores, o erro ocorreu em um momento de alta volatilidade dos preços (7,34%), enquanto a média do período foi de 0,95%.

Por fim, para o 4º período (2018 a 2022), o Gráfico 5 ilustra o comportamento dos resultados gerados quando o respectivo conjunto de hiperparâmetros selecionados foi aplicado ao modelo, em relação aos dados reais.

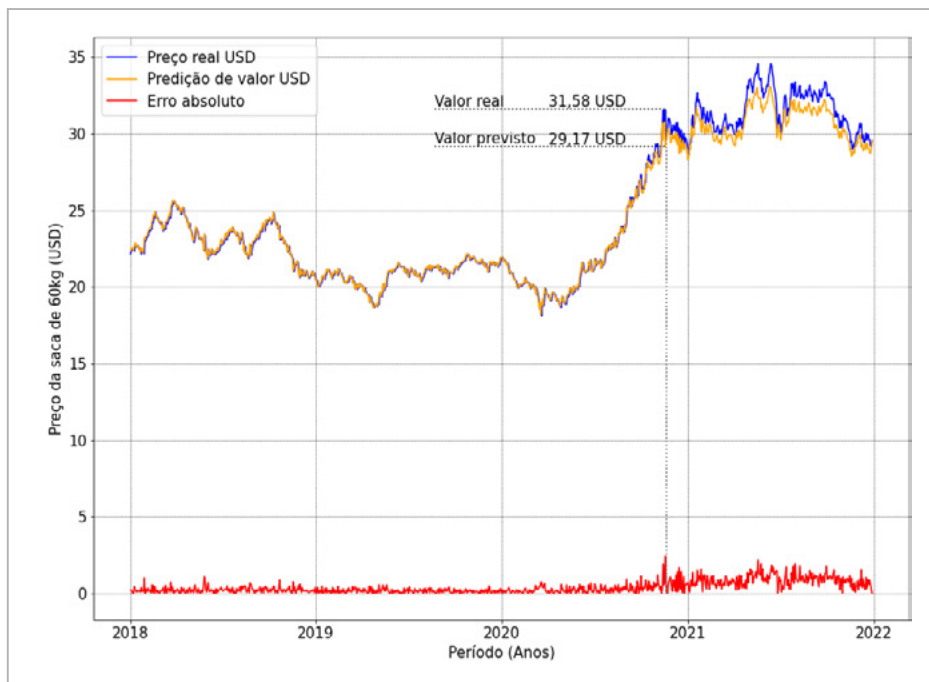


Figura 23 – Previsão e erro absoluto da configuração de hiperparâmetros do 4º período em comparação aos dados reais (2018-2022)

Considerando o último período, o dia de negociação com o maior erro absoluto de previsão foi 27/10/2020, onde o preço real correspondeu a US\$ 31,58 dólares e o valor predito a US\$ 29,17. O erro absoluto nesse dia foi US\$ 2,41 ou 7,63%, quando a volatilidade diária correspondeu a 1,42%, enquanto a volatilidade média do período foi de 0,91%. A configuração de hiperparâmetros do modelo aplicada ao 4º período (2018 a 2022) gerou o menor EMQ. Uma possível causa foi o comportamento dos preços ao longo desse período, que teve a menor volatilidade média (0,91%), apresentando também o menor erro percentual no ponto de maior variação (7,63%).

Para concluir, são apresentados também os resultados obtidos quando os 4 conjuntos de hiperparâmetros (Tabela 3) são testados para a série histórica de dados (2006-2021).

Métricas de erro	Configurações do modelo			
	I (1º período)	II (2º período)	III (3º período)	IV (4º período)
Erro médio quadrático	1,3100	1,3066	1,3085	1,2936
Erro Médio Absoluto	0,8190	0,8176	0,8184	0,8160
Erro Relativo Percentual	3,50%	3,46%	3,47%	3,20%

Tabela 5 – Resultados médios para as métricas de erro em dez repetições, conforme conjunto de hiperparâmetros aplicada ao modelo na fase final, avaliado na série histórica de dados (2006-2021)

A configuração de hiperparâmetros do modelo que apresentou menor EMQ, na análise geral, foi a configuração IV, cujo EMQ foi de 1,2936 US\$2, com erro relativo percentual igual a 3,20%. Pode-se observar, no entanto, que nessa fase final de avaliação, os valores reportados para EQM são bastante similares entre as configurações selecionadas.

O Gráfico 6 ilustra os resultados gerados pela configuração IV, quando o modelo é avaliado no período histórico de 2006 a 2021. Novamente, são apresentados o preço real e o preço predito, além da métrica pontual de erro.

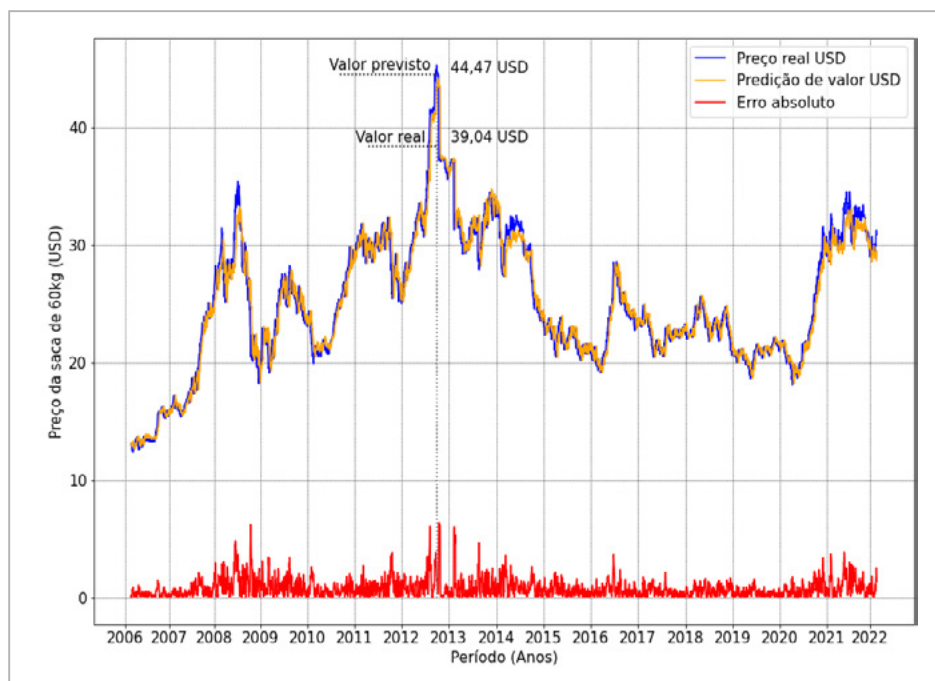


Figura 24 – Previsão e erro absoluto da configuração IV de hiperparâmetros do modelo em comparação aos dados reais da série histórica (2006-2022)

Analisando os resultados, observou-se que o maior erro ocorreu no 2º período, mais especificamente no dia 10/09/2012. Nessa data, o preço da saca de soja foi de US\$ 39,04, enquanto o preço predito é de US\$ 44,47, gerando erro absoluto igual a US\$ 5,43 (13,91%). A volatilidade diária nesse dia correspondeu a 12,10%, sendo maior que a média do período (1,03%). Quando a configuração IV foi aplicada ao modelo e avaliada no período geral, o erro relativo percentual foi igual a 3,20%, superior ao resultado dessa configuração, quando aplicada apenas ao 4º período (2018 a 2022), cujo erro relativo percentual foi 2,57%.

Por fim, os hiperparâmetros selecionados na etapa final foram utilizados para validar o modelo no período final da série histórica. Utilizando taxa de aprendizagem igual a 0,0001; otimizador Adam; dimensão do bloco (*batch size*) de 16 unidades; função de ativação sigmoide e 3 neurônios na camada oculta, a rede neural foi aplicada para previsão do preço da saca de soja no período compreendido entre 03/01/2022 e 10/10/2023, tendo como dados de treinamento e teste o período anterior da série histórica (2006-2021). O Gráfico 7 ilustra os resultados encontrados nesta previsão.



Figura 25 – Previsão e erro absoluto da configuração final de hiperparâmetros do modelo em comparação aos dados reais (03/01/2022 a 10/10/2023)

De acordo com os dados apresentados no Gráfico 7, é possível verificar que o modelo selecionado permite realizar uma boa previsão do preço da saca de soja, com pequena margem de erro, no período em que foi validado. O erro absoluto máximo obtido para o período de 03/01/2022 a 10/10/2023 é de US\$ 0,83, o que representa 2,01% do valor real da saca de soja no dia considerado. O EQM para o período foi de 0,0528 US\$2.

Os resultados encontrados corroboram com a literatura, no que tange ao uso de modelos baseados em redes neurais para precificação do preço de *commodities*. O estudo conduzido por Sobreiro, Araujo e Nagano (2009) compararam o desempenho de uma rede MPL com um modelo ARIMA para precificação da *commodity* etanol. De acordo com os autores, uma rede neural que utiliza duas camadas ocultas com 13 neurônios gerou maior acurácia que o modelo ARIMA, produzindo um erro médio percentual de 4,55%.

Avaliando modelos híbridos que incluem redes neurais, Ribeiro, Sosnoski e Oliveira (2010) analisaram o desempenho dos mesmos na precificação de quatro *commodities*: açúcar, etanol, café e soja. Os resultados indicaram que redes MPL de única camada oculta apresentam um desempenho significativo na precificação desses produtos. Para esses modelos, o erro médio percentual de todas as *commodities* avaliadas foi de 5,77%.

Ferreira *et al.* (2011) também utilizaram redes neurais do tipo MPL como estratégia de previsão preços no contexto do agronegócio, avaliando soja, milho, boi gordo e trigo. Trabalhando com um modelo de camada oculta única, os autores avaliaram o número de neurônios dessa camada. Os resultados indicaram que, usando 5 neurônios, o modelo produziu um erro médio percentual de 4,86% nos dados de validação, na média de todas os produtos considerados.

Com o objetivo de analisar a contribuição de redes neurais MPL para a previsão de preços de *commodities* agrícolas (açúcar, algodão, arroz, café, milho, soja e trigo), Disconzi (2018) também abordou o ajuste das redes a fim de identificar a melhor arquitetura. Na pesquisa, foram avaliados o número de entradas, número de neurônios,

função de ativação e número de camadas intermediárias. Para a *commodity* soja, o coeficiente de determinação para a regressão linear entre o preço real e o preço estimado ficou em 0,834. Segunda a autora, a abordagem aplicada foi capaz de prever o preço diário, no mercado futuro, das *commodities* analisadas, e os resultados foram considerados satisfatórios.

Um tipo especial de redes neurais recorrentes, denominado LSTM (*Long short-term memory*), foi aplicado por Wang e Gao (2018) para prever preços mínimo e máximo da soja. Segundo os autores, o modelo apresentou maior acurácia para prever preços mínimos e máximos quando comparado ao preço de fechamento, além de apresentar acurácia mais elevada quando há alta volatilidade.

Sunarya, Henderi e Tasyriqan (2019) realizaram uma comparação entre uma rede neural do tipo MPL e métodos SMO (*Sequential Minimal Optimization*) na previsão do preço de *commodities*, entre elas a soja. O estudo levou em conta as características dos dados e a precisão dos métodos. Os resultados dos autores apontam que a rede MPL superou os métodos SMO em termos de taxa de precisão, mas gerou um erro percentual absoluto levemente maior.

Para prever o preço de fechamento da soja, Yu (2021) também aplicou um modelo de rede neural do tipo LSTM. O modelo foi avaliado a partir do erro percentual médio absoluto e do coeficiente de correlação. Segundo o autor, o modelo possui excelente capacidade de generalização, apresentando alto efeito de previsão sobre o preço futuro dos produtos.

De acordo com os resultados apresentados nesta seção, foi possível observar como a variação nos hiperparâmetros do modelo pode melhorar sua predição, o que ressalta a importância de avaliar a configuração da rede antes da sua utilização. Além disso, os resultados indicam que a utilização de uma rede MPL de camada única, em conjunto com a seleção de hiperparâmetros, foi capaz de produzir resultados muito bons, gerando um erro máximo absoluto de 2,01% sobre o valor real da saca de soja, para um período de validação superior a 21 meses.

Conclusões

Esta pesquisa avaliou diferentes configurações de hiperparâmetros de uma rede neural MPL, com o intuito de selecionar aquela com a melhor capacidade de previsão do preço da soja. Entre as configurações avaliadas, aquela referenciada como configuração IV produziu o menor erro médio quadrático no cenário estudado. Para esse conjunto, o EMQ obtido foi igual a 1,2936, o que representa, em termos percentuais, 3,20%. Essa configuração é composta por modelo com uma camada oculta com 3 neurônios, função de ativação sigmoide, taxa de aprendizagem de 0,0001 e dimensão do bloco igual a 16. Este modelo, validado em um período independente de tempo, mostrou-se extremamente eficiente, gerando um erro máximo absoluto de US\$ 0,83, o que representa em torno de 2% do valor real.

Os resultados mostram que uma rede neural MPL compacta foi capaz de realizar a previsão do preço futuro da soja, acompanhando a tendência de preços do grão, no período em que este foi avaliado, o que corrobora com a literatura, no que diz respeito à previsão de preços de *commodities* agrícolas. Essas informações podem ser usadas para o futuro desenvolvimento de uma ferramenta computacional, acessível ao produtor rural, para auxiliar na tomada de decisão em relação à projeção do preço de venda do seu produto, visando otimizar sua lucratividade.

Referências

CEPEA (CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA). *Série histórica do preço da soja*. Disponível em: <https://www.cepea.esalq.usp.br/br>. Acesso em: 15 abr. 2022.

CNA (CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL). Aumento dos custos causa queda de 0,8% no PIB do agronegócio neste início de ano. *PIB do Agronegócio: relatório técnico*. Brasília, 2022. Disponível em: [https://www.cepea.esalq.usp.br/upload/kceditor/files/Cepea_CNA_PIB-do-Agronegocio-20junho22\(1\).pdf](https://www.cepea.esalq.usp.br/upload/kceditor/files/Cepea_CNA_PIB-do-Agronegocio-20junho22(1).pdf). Acesso em: 23 ago. 2022.

DISCONZI, C. M. D. G. *Previsão dos preços de commodities agrícolas brasileiras no mercado futuro utilizando redes neurais artificiais*. 2018. Dissertação (Mestrado em Administração) – Universidade Federal do Pampa, Santana do Livramento, 2018.

EMBRAPA (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA). *Soja em números (safra 2020/2021)*. Disponível em: <https://www.embrapa.br/soja/cultivos/soja1/dados-economicos>. Acesso em: 23 ago. 2022.

FANG, Y.; GUAN, B.; WU, S.; HERAVI, S. Optimal forecast combination based on ensemble empirical mode decomposition for agricultural commodity futures prices. *Journal of Forecasting*, v. 39, n. 6, p. 877-886, 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2665> Acesso em: 23 ago. 2022.

FERREIRA, L; MOURA, G.L.; BORENSTEIN, D.; FISCHMANN, A.A. Utilização de redes neurais artificiais como estratégia de previsão de preços no contexto de agronegócio. *Revista de Administração e Inovação*, v. 8, n. 4, p.6-26, 2011. Disponível em: <https://www.revistas.usp.br/rai/article/view/79235>. Acesso em: 23 ago. 2022.

GE, W.; LALBAKHSH, P.; ISAI, L.; LENSKIY, A.; SUOMINEN, H. Neural network-based financial volatility forecasting: a systematic review. *ACM Computing Surveys*, v. 55, n. 1(14), p. 1-30, 2022. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3483596>. Acesso em: 23 ago. 2022.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Massachusetts: The Mit Press, 2016.

HAYKIN, S. *Neural Networks and Learning Machines*. 3 ed. Ontario: Prentice Hall, 2008.

IMEA (INSTITUTO MATO-GROSSENSE DE ECONOMIA AGROPECUÁRIA). Margem Bruta Pressionada. *Boletim Semanal Soja*, n. 273, 22 out. 2022. Disponível em: <https://imea.com.br/imea-site/arquivo-externo?categoria=relatorio-de-mercado&arquivo=bs-soja&numeropublicacao=723>. Acesso em: 15 abr. 2023.

JHA, G. K.; SINHA, K. Agricultural price forecasting using neural network model: An innovative information delivery system. *Agricultural Economics Research Review*, v. 26, n. 2, p.229-239, 2013. Disponível em: <https://ideas.repec.org/a/ags/aerrae/162150.html>. Acesso em: 15 abr. 2023.

- LUDOVICO, S. N. *Previsão de indicadores diários de preço no mercado futuro de commodities agrícolas utilizando aprendizagem de máquina*. 2020. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Alfenas, Alfenas, 2020. Disponível em: <https://bdtd.unifal-mg.edu.br:8443/handle/tede/1762> Acesso em: 15 abr. 2023.
- MENHAJ, M.; KAVOOSI-KALASHAMI, M. Developing a hybrid forecasting system for agricultural commodity prices (case study: Thailand rice free on board price). *Ciência Rural*, Santa Maria, v. 52, n.8, p. 1-11, 2022. Disponível em: <https://www.scielo.br/j/cr/a/mcdxrp7wjNdgBTZGbNztJsh/?lang=en>. Acesso em: 15 abr. 2023.
- PINHEIRO, C.; SENNA, V. Multivariate analysis and neural networks application to price forecasting in the Brazilian agricultural market. *Ciência Rural*, v.47, n.1, p. 1-7, 2017. Disponível em: <https://www.scielo.br/j/cr/a/bcB64kH4hY4Htpc7KhqG6Wg/abstract/?lang=en>. Acesso em: 15 abr. 2023.
- RIBEIRO, C. O.; SOSNOSKI, A. A. K. B.; OLIVEIRA, S. M. Um modelo hierárquico para previsão de preços de commodities agrícolas. *Revista Produção Online*, v.10, n. 4, p. 719-733, 2010. Disponível em: <https://producaoonline.org.br/rpo/article/view/225> Acesso em: 15 abr. 2023.
- RIBEIRO, M. H. D. M.; COELHO, L. S. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, v. 86, 105837, 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1568494619306180> Acesso em: 15 abr. 2023.
- ROSOLEN, D.; ARAÚJO, M. V.; LYRIO, M. Previsão dos preços de commodities por meio das taxas de câmbio. *Estudo Econômico*, vol. 43, n. 4, p. 813-830, 2013. Disponível em: <https://www.scielo.br/j/ee/a/kFF88txxLpqQ88npY8c7snN/?lang=pt> Acesso em: 15 abr. 2023.
- SOBREIRO, V.A.; ARAUJO, P.H.S.L.; NAGANO, M.S. Precificação do etanol utilizando técnicas de redes neurais artificiais. *RAUSP - Revista de Administração*, v. 44, n. 1, p. 46-58, 2009. Disponível em: <http://rausp.usp.br/wp-content/uploads/files/v4401046.pdf> Acesso em: 15 abr. 2023.
- SPANCERSKI, J. S.; SANTOS, J. A. A. Redes neurais aplicadas na predição do preço da soja no estado do Paraná. *Revista Perspectivas Online: Exatas & Engenharia*, v. 11, n. 32, p.19-32, 2021. Disponível em: https://ojs3.perspectivasonline.com.br/exatas_e_engenharia/article/view/2282 Acesso em: 15 abr. 2023.
- SUNARYA, A.; HENDARI; TASYRIQAN, I. The Comparison Between Sequential Minimal Optimization and Multilayer Perceptron Neural Network Methods in Predicting the Commodity Prices. *In: INTERNATIONAL CONFERENCE ON INFORMATICS AND COMPUTING (ICIC)*, 4., 2019, Semarang, Indonésia. *Proceedings [...]*. Piscataway, NJ: IEEE, 2019. p. 1-6. Disponível em: <https://ieeexplore.ieee.org/document/8985679> Acesso em: 15 abr. 2023.
- WANG, C.; GAO, Q. High and Low Prices Prediction of Soybean Futures with LSTM Neural Network. *In: IEEE INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND SERVICE SCIENCE*, 9., 2018, Beijing, China,

2018. *Proceedings* [...]. Piscataway, NJ: IEEE, 2018. p. 140-143. Disponível em: <https://ieeexplore.ieee.org/document/8663896> Acesso em: 15 abr. 2023.

WANG, J.; WANG, Z.; LI, X.; ZHOU, H. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *International Journal of Forecasting*, v. 38, n. 1, p. 21-34, 2022. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0169207019302304> Acesso em: 15 abr. 2023.

YU, Z. Data Analysis and Soybean Price Intelligent Prediction Model Based on LSTM Neural Network. In: Conference on Telecommunications, Optics and Computer Science (TOCS). Shenyang, China, 2021, p. 799-801. Disponível em: <https://ieeexplore.ieee.org/document/9688705> Acesso em: 15 abr. 2023.

ZHANG, D.; ZANG, G.; LI, J.; MA, K.; LIU, H. Prediction of soybean price in China using QR-RBF neural network model. *Computers and Electronic in Agriculture*, v. 154, p. 10-17, 2018. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0168169918302199> Acesso em: 15 abr. 2023.