



MINISTÉRIO DA EDUCAÇÃO
FUNDAÇÃO UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
Escola de Administração e Negócios – ESAN
Programa de Pós-Graduação *Stricto Sensu* em Administração



Bruno Matos Porto

Revisão bibliométrica sobre modelagem e previsão do preço da soja: uma comparação entre os modelos ARIMAX, redes neurais e máquina de aprendizado extremo

Campo Grande/MS
2021



MINISTÉRIO DA EDUCAÇÃO
FUNDAÇÃO UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
Escola de Administração e Negócios - ESAN
Programa de Pós-Graduação *Stricto Sensu* em Administração



Bruno Matos Porto

**Revisão bibliométrica sobre modelagem e previsão do preço da
soja: uma comparação entre os modelos ARIMAX, redes
neurais e máquina de aprendizado extremo**

Dissertação entregue ao Programa de Pós Graduação stricto sensu em Administração, curso de Mestrado, como requisito parcial para obtenção do título de mestre em administração, sob orientação do Prof. Dr. Matheus Wemerson Gomes Pereira.

Campo Grande/MS
2021

Bruno Matos Porto

**Revisão bibliométrica sobre modelagem e previsão do preço da soja:
uma comparação entre os modelos ARIMAX, redes neurais e máquina
de aprendizado extremo**

Dissertação apresentada ao Programa de Pós Graduação em Administração da Universidade Federal de Mato Grosso do Sul, como requisito à obtenção do grau de Mestre em Administração.

Campo Grande, MS, 05 de março de 2021.

COMISSÃO EXAMINADORA

Prof. Dr. Michel Ângelo Constantino de Oliveira
Universidade Católica Dom Bosco

Prof. Dr. Renato Luiz Sproesser
Universidade Federal de Mato Grosso do Sul

Prof. Dr. Matheus Wemerson Gomes Pereira
Universidade Federal de Mato Grosso do Sul

RESUMO

PORTO, Bruno Matos. **Revisão bibliométrica sobre modelagem e previsão do preço da soja: uma comparação entre os modelos ARIMAX, redes neurais e máquina de aprendizado extremo**. 2021. 118 f. Dissertação (Mestrado em Administração) – Escola de Administração e Negócios, Universidade Federal de Mato Grosso do Sul, Campo Grande, MS, 2021.

Orientador: Prof. Dr. Matheus Wemerson Gomes Pereira.

Defesa: 05/03/2021

Modelos de previsão são úteis nas predições de preços de *commodities* agrícolas e servem para a tomada de decisão dos produtores rurais, investidores, autoridades agrícolas e participantes do mercado. A lacuna científica sobre previsão do preço de *commodities* agrícolas foi identificada por meio da análise bibliométrica, que gerou resultados de parâmetros quantitativos da produção científica e mostrou as tendências sobre a temática, além disso, forneceu referências para os econometristas do campo da previsão de preços de *commodities* agrícolas. O problema de previsão foi modelar um conjunto de dados de preços mensais da soja em três estados: Paraná, Rio Grande do Sul e Mato Grosso. O objetivo geral foi avaliar os modelos autorregressivos integrados de médias móveis com variável exógena (ARIMAX), rede neural perceptron multicamada (MLP) e máquina de aprendizagem extrema (ELM). O modelo teórico da pesquisa foi a lei do preço único. Os conjuntos de treinamento e teste foram modelados com métodos ARIMAX, MLP e ELM. Posteriormente, são comparadas as previsões nos conjuntos de teste com indicadores de desempenho da previsão. A precisão da previsão dos métodos foram comparadas por meio de avaliações dos gráficos box plots e testes de variâncias dos erros de cada modelo, assim identificando se tiveram diferenças estatísticas significativas entre as métricas de erro geradas na previsão. Os resultados revelam que o ARIMAX é a melhor abordagem para prever os preços do Paraná, enquanto a MLP obteve o desempenho da previsão superior, para prever os preços do Rio Grande do Sul. Por fim, a MLP e ELM com a melhor acurácia entre as comparações das previsões de preços do Mato Grosso. As demais comparações não tiveram diferenças significativas ($p > 0,05$) entre as variâncias dos erros dos modelos de previsão. Desta análise, o estudo concluiu que, para previsão do preço da soja, os modelos de inteligência artificial (IA), em uma parte dos casos, não são superiores em comparação ao modelo estatístico ARIMAX.

Palavras-chave: Desempenho de previsão, Previsão de séries temporais, Previsão de preços agrícolas, *Commodity* agrícola, Modelos de previsão, Precisão da previsão.

ABSTRACT

PORTO, Bruno Matos. **Bibliometric review on modeling and forecasting of soybean price: a comparison between models ARIMAX, neural networks and extreme learning machine**. 2021. 118 f. Dissertation (Master in Management) - Business and Management School, Federal University of Mato Grosso do Sul, Campo Grande, MS, 2021.

Forecasting models are useful in the prediction of prices commodity agriculture and serve for decision-making from rural producers, investors, agricultural authorities, and market participants. The scientific gap on price forecasting commodity agriculture was identified through bibliometric analysis, which generated results in quantitative parameters of scientific production and showed trends on the subject, in addition, provided references for econometrists in the field of agricultural price forecasting commodities agriculture. The forecasting problem was to model a set of monthly soybean price data in three states: Paraná, Rio Grande do Sul, and Mato Grosso. The general objective was to evaluate the models autoregressive integrated moving average with exogenous variable (ARIMAX), multilayer perceptron neural network (MLP), and extreme learning machine (ELM). The theoretical model of the research was the law of the single price. The training and test sets were modeled using methods ARIMAX, MLP, and ELM. Subsequently, the forecasting in the test sets is compared with forecast performance indicators. The forecast accuracy of the methods was compared through evaluations of the box plots and tests of variances of the errors of each model, thus identifying whether there were significant statistical differences between the error metrics generated in the forecast. The results reveal that ARIMAX is the best approach to forecast prices in Paraná, the MLP obtained the performance of the superior forecast, to predict prices from Rio Grande do Sul. Finally, MLP and ELM with the best accuracy among the comparisons from Mato Grosso's price forecasting. The other comparisons did not have significant differences ($p > 0.05$) between the variances of the errors of the forecast models. From this analysis, the study concluded that, for the prediction of soybean prices, artificial intelligence (AI) models, in some cases, are not superior in comparison to the statistical model ARIMAX.

Keywords: Forecasting performance, Time series forecast, Agricultural price forecast, Agricultural commodity, Forecasting models, Forecast accuracy.

LISTA DE ILUSTRAÇÕES

Figura 1 - Processo de seleção dos artigos para a pesquisa.....	25
Figura 2 - Produção científica anual das publicações	27
Figura 3 - Periódicos que mais publicam sobre previsão de preços agrícolas nos últimos dez anos	28
Figura 4 - Afiliações mais relevantes entre 2010 a outubro de 2020	29
Figura 5 - Produção científica do país entre 2010 e outubro de 2020.....	30
Figura 6 - Países mais citados entre 2010 a outubro de 2020	30
Figura 7 - Quantidade de citações por artigos do portfólio	31
Figura 8 - Palavras-chave mais frequentes entre 2010 e outubro de 2020.....	44
Figura 9 - Tópicos de tendência das pesquisas sobre previsão de preços das <i>commodities</i> agrícolas no período de 2010 a outubro de 2020.....	45
Figura 10 - Nuvem de palavras dos resumos e títulos dos artigos	47
Figura 11 - Gráfico de três campos (autores, palavras-chave e periódicos).....	47
Figura 12 - Rede de co-citação de autores entre 1981 até outubro de 2020.....	48
Figura 13 - Os dados dos conjuntos de treinamento são de janeiro de 2011, e janeiro de 2014 a dezembro de 2018, e os dados do conjunto de teste são de janeiro de 2019 a dezembro de 2019.....	72
Figura 14 - Preços da soja no Rio Grande do Sul e os preços da Bolsa de Chicago.....	75
Figura 15 - Preços da soja no Mato Grosso e os preços da Bolsa de Chicago.....	75
Figura 16 - Séries temporais dos preços da Bolsa de Chicago e CEPEA em dólares US\$	76
Figura 17 - Séries de preços da CBOT e Paraná com uma diferenciação	79
Figura 18 - Séries de preços no Rio Grande do Sul e Mato Grosso com uma diferenciação.....	79
Figura 19 - Estimativa do ARIMAX no conjunto de treinamento para o preço do Paraná	80
Figura 20 - Estimativa do ARIMAX no conjunto de treinamento para o preço do Rio Grande do Sul	80
Figura 21 - Estimativa do ARIMAX no conjunto de treinamento para o preço do Mato Grosso.....	81

Figura 22 – Gráfico da ACF sobre os resíduos para identificar se os pressupostos do modelo ARIMAX são atendidos no conjunto de treinamento Paraná	82
Figura 23 – Gráfico da ACF sobre os resíduos para identificar se os pressupostos do modelo ARIMAX são atendidos no conjunto de treinamento Rio Grande do Sul.....	83
Figura 24 – Gráfico da ACF sobre os resíduos para identificar se os pressupostos do modelo ARIMAX são atendidos no conjunto de treinamento Mato Grosso.....	83
Figura 25 - Previsões ARIMAX do conjunto de teste Paraná de janeiro de 2019 até dezembro de 2019.....	85
Figura 26 - Previsões ARIMAX do conjunto de teste do Rio Grande do Sul de janeiro de 2019 até dezembro de 2019	85
Figura 27 - Previsões ARIMAX do conjunto de teste do Mato Grosso, de janeiro a dezembro de 2019.....	86
Figura 28 - Arquiteturas da MLP e ELM com base no conjunto de treinamento Paraná	88
Figura 29 - Arquiteturas da MLP e ELM com base no conjunto de treinamento Rio Grande do Sul	88
Figura 30 - Arquiteturas da MLP e ELM com base no conjunto de treinamento Mato Grosso.....	89
Figura 31 - Treinamento MLP e ELM dos preços CEPEA de janeiro de 2011 até dezembro de 2018.....	90
Figura 32 - Treinamento MLP e ELM dos preços da Conab do Rio Grande do Sul, de janeiro de 2014 até dezembro de 2018	90
Figura 33 - Treinamento MLP e ELM dos preços da Conab do Mato Grosso de janeiro de 2014 até dezembro de 2018	91
Figura 34 - Previsões MLP e ELM dos preços do Paraná de janeiro a dezembro de 2019	91
Figura 35 - Previsões MLP e ELM dos preços do Rio Grande do Sul de janeiro de 2019 até dezembro de 2019.....	91
Figura 36 - Previsões MLP e ELM dos preços do Mato Grosso de janeiro de 2019 até dezembro de 2019.....	92
Figura 37 - Boxplot dos indicadores de desempenho para o conjunto de teste Paraná.	96
Figura 38 - Boxplot dos indicadores de desempenho para o conjunto de teste Rio Grande do Sul	98

Figura 39 - Boxplot dos indicadores de desempenho para o conjunto de teste Mato Grosso..... 98

LISTA DE TABELAS

Tabela 1 - Séries temporais de preços mensais da soja	71
Tabela 2 - Número de diferenças necessárias para tornar estacionárias as séries, testes de raiz unitária ADF e KPSS nos conjuntos de treinamento	78
Tabela 3 - Parâmetros e métricas do ARIMAX nos conjuntos de treinamento.....	81
Tabela 4 - Teste de hipótese Ljung-Box.....	84
Tabela 5 - Parâmetros das redes neurais e máquina de aprendizagem extrema para o conjunto de treinamento	87
Tabela 6 - Indicadores de desempenho dos modelos ARIMAX, MLP e ELM do conjunto de teste Paraná	93
Tabela 7 - Indicadores de desempenho das previsões do conjunto de teste Rio Grande do Sul.....	94
Tabela 8 - Indicadores de desempenho das previsões do conjunto de teste Mato Grosso	95
Tabela 9 - Variâncias entre as medidas de erro dos modelos ARIMAX, MLP e ELM	100

LISTA DE QUADROS

Quadro 1 - Artigos científicos identificados em base de dados	26
Quadro 2 - Indicadores de desempenho mais utilizados nas publicações.....	68
Quadro 3 - Lista de pacotes e funções em linguagem de programação R	73
Quadro 4 - Aplicação prática em linguagem de programação R da rede neural perceptron multicamada (MLP).....	117
Quadro 5 - Aplicação prática em linguagem de programação R da máquina de aprendizagem extrema (ELM).....	118

LISTA DE SIGLAS

ABC	<i>Artificial Bee Colony Algorithm</i>
ABIOVE	Associação Brasileira das Indústrias de Óleos Vegetais
ACF	Função de autocorrelação
ACF1	Autofunção de correlação no atraso 1
ADF	<i>Augmented Dickey-Fuller</i>
AES	Análise espectral singular
AESM	Análise multivariada singular espectral
AIC	<i>Akaike information criteria</i>
AICc	<i>Akaike information criteria correction</i>
ANN	<i>Artificial Neural Network</i>
ARFIMA	<i>Auto Regressive Fractionally Integrated Moving Average</i>
ARCH	<i>Autoregressive Conditional Heterocedasticity</i>
ARIMA	Modelos Autorregressivos Integrados de Médias Móveis
ARIMAX	Modelos Autorregressivos Integrados de Médias Móveis com Variável Exógena
ASCAR	Associação Sulina de Crédito e Assistência Rural
APE	<i>Absolute percentage errors</i>
BIC	<i>Bayesian information criterion</i>
BMA	Modelo bayesiano de média
BM&F	Bolsa de Mercadorias e Futuros
BPNN	<i>Back propagation neural network</i>
CBOT	<i>Chicago Board of Trade</i>
CEPEA	Centro de Estudos Avançados em Economia Aplicada
CNA	Confederação da Agricultura e Pecuária do Brasil
CRMS	Quadrado médio da raiz cumulativa
CONAB	Companhia Nacional de Abastecimento
DMA	<i>Dynamic model averaging</i>
ELM	<i>Extreme Learning Machine</i> - Máquina de Aprendizado Extremo
EMATER-RS	Associação Riograndense de Empreendimentos de Assistência

	Técnica e Extensão Rural
EMD	<i>Empirical mode decomposition</i>
EQA	Erro Quadrático Acumulado
EqCM	Correção de equilíbrio
EQM	Erro quadrado médio da previsão
ER	Erro relativo
ESALQ	Escola Superior de Agricultura Luiz de Queiroz
FAO	<i>Food and Agriculture Organization of the United Nations</i>
FMI	Fundo Monetário Internacional
FPM	Modelo de preços futuros
GA	Algoritmo genético
GARCH	<i>Generalized autoregressive conditional heteroskedastic</i>
GBM	<i>Gradiente boosting machine</i>
GDGA	<i>Gradient-based optimization algorithm</i>
GRU	Gated recurrent neural network
GSO	<i>Glowworm Swarm Optimization</i>
HW	<i>Holt-Winters</i>
KNN	<i>K-nearest neighbors</i>
KPSS	Kwiatkowski, Phillips, Schmidt, and Shin
IA	Inteligência Artificial
IEA	Instituto de Economia Agrícola
IPEA	Instituto de Pesquisa Economicamente Aplicada
ITD	<i>Intrinsic time-scale decomposition</i>
LIC	Limites de confiança inferior
LSC	Limites de confiança superior
LPU	Lei do Preço Único
LSTM	Rede neural de memória de longo prazo
MA	Média Móvel
MAD	Desvio Médio Absoluto
MAE	Erro absoluto médio
MASE	Erro médio absoluto em escala
MAPA	Ministério da Agricultura, Pecuária e Abastecimento

MAPE	<i>Mean Absolute Percentage Error</i>
ME	Erro Médio
MLP	<i>Multilayer perceptron neural network</i>
MLP	Perceptron Multicamada
MPE	<i>Mean Percentage Error</i>
MSE	Mean Square Error
MSSA	<i>Multivariate Singular Spectrum Analysis</i>
MT	Mato Grosso
NAÏVE	Previsão ingênua
PR	Paraná
QR-RBF	<i>Quantile regression-radial basis function</i>
RF	<i>Random forests</i>
RMSE	Raiz do Erro Quadrático Médio
RNA	Rede Neural Artificial
RNN	Rede neural recorrente
RS	Rio Grande do Sul
SARIMA	Autorregressivo Integrado de Média Móvel Sazonal
SSE	<i>Sum of square for forecast error</i>
SVR	Support vector machine for regression
STACK	<i>Stacking</i>
PSO	<i>Particle swarm optimization</i>
TDNN	Rede neural de atraso de alimentação
Theil	Estatística U de Theil
VMD	<i>Variational mode decomposition</i>
WNN	<i>Wavelet Neural Networks</i>
WPT	<i>Wavelet packet transform</i>
XGB	<i>Extreme gradient boosting machine</i>

SUMÁRIO

RESUMO.....	5
CAPÍTULO I	16
1.1 Introdução	16
1.1.1 Objetivos	22
1.1.2 Objetivo geral	22
1.1.3 Objetivos específicos	22
1.1.4 Teoria e objeto de estudo	22
CAPÍTULO II.....	23
2.1 Revisão bibliométrica:	23
2.1.1 Introdução.....	23
2.1.2 Procedimentos metodológicos	24
2.1.3 Processo de pesquisa	25
2.1.4 Critérios de inclusão e exclusão	25
2.2 Resultados.....	26
2.2.1 Descrição dos principais artigos encontrados na revisão sobre previsão de preços agrícolas	32
2.2.2 Discussão	43
2.2.3 Observações gerais	43
2.2.4 Principais tendências do campo da previsão de preços das <i>commodities</i> agrícolas.....	44
2.3 Considerações e apontamentos da revisão.....	50
CAPÍTULO III	52
3.1 Revisão de literatura.....	52
3.1.1 Lei do Preço Único (LPU).....	52
3.1.2 Comercialização da soja	54
3.1.3 Modelos de previsão e tomada de decisão	56
3.2 Modelos autorregressivos integrados de médias móveis com variável exógena (ARIMAX)	59
3.3 Redes Neurais.....	61
3.3.1 Algoritmo de <i>Perceptron</i> multicamada (MLP)	62
3.4 Algoritmo de máquina de aprendizagem extrema (ELM)	64
3.5 Indicadores de Desempenho	65

CAPÍTULO IV	67
4.1 Procedimentos Metodológicos	67
4.1.1 Classificação da pesquisa	67
4.1.2 Coleta de dados.....	69
4.1.3 Pacotes em R <i>studio</i>	72
4.2 Resultados	74
4.2.1 Modelagem ARIMAX	74
4.2.2 Treinamento e teste dos modelos de redes neurais (MLP) e máquina de aprendizado extremo (ELM).....	87
4.3 Discussão dos resultados (Comparações empíricas de previsões)	92
4.3.1 Boxplots e testes de variâncias de erros das abordagens de previsão	96
CAPÍTULO V	103
5.1 Principais desdobramentos da pesquisa	103
5.2 Considerações finais	104
5.2.1 Limitações da pesquisa e sugestões para estudos futuros.....	106
Referências	107
Apêndice 1 - Aplicação empírica em R (scripts)	117

CAPÍTULO I

1.1 Introdução

As previsões de preços de *commodities* agrícolas são úteis para agricultores, mercado, indústrias do agronegócio, comunidades agrícolas e governos (ALLEN, 1994; BRANDT; BESSLER, 1983; DAS et al., 2019; JHA; SINHA, 2012). Porém, o mercado agrícola possui uma característica de difícil previsibilidade dos preços (CERETTA; RIGHI; SCHLENDER, 2010; DENG; SÜN, 2019).

Essa característica do mercado é decorrente de fatores externos (bolsa de valores, políticas externas, taxa de câmbio, clima, tecnologia e estoques mundiais) que influenciam o preço das *commodities* agrícolas (JUNIOR et al., 2018). Neste contexto, o risco é maior no setor agrícola, em virtude de efeitos da oferta e demanda e os fatores climáticos que afetam a produção. Fenômenos climáticos como El Nino e La Nina afetam a produção da soja, gerando oscilações na produtividade por causa do aumento nas chuvas e fortes estiagens (FELICIANI; SOUZA; SOUZA, 2015). Essa complexidade do mercado agrícola torna a previsão precisa dos preços futuros de *commodities* agrícolas cada vez mais difíceis (WANG et al., 2019).

As condições de mercado sofrem variações com o passar do tempo, o que gera uma necessidade dos produtores rurais obterem informações confiáveis sobre mudanças nos preços. Neste sentido, os métodos que podem ser usados nesse processo de antever ocorrências futuras dos preços para incorporar essa informação ao processo de decisão são os modelos de previsão (MENDES; JUNIOR, 2007; SABU; KUMAR, 2020). As predições dos preços tem sido objeto de estudo, tanto dos profissionais do setor agrícola quanto dos pesquisadores. Por meio da previsão é possível reduzir a incerteza nas decisões das pessoas que realizam transações (PINHEIRO; SENNA, 2016, 2017a, 2017b; RIBEIRO; SOSNOSKI; OLIVEIRA, 2010).

A previsão de *commodities* auxiliam os *stakeholders* no planejamento de estoque a atender o consumo interno, exportações, investimentos em tecnologias, tamanho da área plantada, aumento da produtividade e organizações públicas nas necessidades de investimento no setor agrícola (NOMELINI et al., 2017). Os dados de preços da agricultura possibilitam aplicar modelos de previsão amplamente utilizados para apoiar as decisões de planejadores agrícolas, a fim de reduzir o risco de mercado.

Para os participantes do mercado, tais como os compradores e vendedores, é relevante a análise dos preços do mercado agrícola, uma vez que esta análise os ajudam na

diminuição de incertezas quanto às oscilações dos preços (PINHEIRO; SENNA, 2015, 2016; SABU; KUMAR, 2020). Neste sentido, Mendes e Junior afirmam que “o preço agropecuário é uma variável decisória muito importante para o produtor rural e para o setor agropecuário” (p. 249, 2007).

O processo de planejamento decisório da comercialização ocorre por meio da análise de séries temporais de preços (MENDES; JUNIOR, 2007). Os modelos de previsão apoiam a tomada de decisão dos agentes que participam de atividades que precisam de planejamento, avaliar as políticas e reduzir as incertezas (BRESSAN, 2004; BRESSAN; LIMA, 2002).

De acordo com Bressan (2004), uma das finalidades das predições econômicas é a redução de incertezas que, por sua vez, tem relevância na agropecuária, estando o setor sujeito à volatilidade dos preços. A pesquisa pode apoiar a tomada de decisão dos agentes que participam dos mercados de soja e reduzir as incertezas dos preços das *commodities* agrícolas, uma vez que na tomada de decisão econômica as predições são uma ferramenta indispensável (MEDEL, 2013).

O setor agrícola é relevante para o desempenho da economia brasileira (AIUBE; FERREIRA; LEVY, 2020; CASAS; BACHA; CARVALHO, 2016; SCHNEIDER; FERREIRA; ALVES, 2013). De acordo com a (CNA, 2020), o PIB do agronegócio brasileiro cresceu 3,81% em 2019, frente a 2018. Com esse desempenho, em 2019 o PIB do agronegócio representou 21,4% do PIB brasileiro total. O Brasil mantém a liderança nas exportações de soja, além de ser o segundo maior produtor da *commodity* e estar entre os quatro maiores países que mais esmagam a oleaginosa no mundo (SOUZA; BACCHI; ALVES, 2019).

O Brasil deve elevar o cultivo da oleaginosa em função das demandas regional por alimentos e também global (FAO NO BRASIL, 2018). Assim, o país sul-americano vai desempenhar um papel importante como fornecedor de alimentos para o mundo (FAO NO BRASIL, 2018). A FAO projeta que o Brasil deve ultrapassar os Estados Unidos até 2026 e ser o maior produtor de soja do mundo. A previsão é de que as exportações de soja em 2026 sejam dominadas pelo Brasil e os Estados Unidos; ambos os países representarão aproximadamente 80% das exportações mundiais do produto (FAO, 2017). Embora tenha a estimativa para 2026 da FAO, segundo a carta de conjuntura do Ipea para a safra 2019-2020, o Brasil é o maior produtor mundial de soja, com 126 milhões de toneladas (IPEA, 2020).

O país destaca-se como um grande produtor e exportador da *commodity* (AIUBE; FERREIRA; LEVY, 2020). A escolha do preço da soja foi motivada pelo crescimento dos volumes de exportações nos últimos anos e pelo papel significativo na pauta das exportações

brasileiras. A previsão do preço deste grão tem um papel importante nos países em desenvolvimento, especialmente aqueles que são seus produtores. No Brasil, a soja representou cerca de 25% das exportações de *commodities* agrícolas no valor exportado em 2020. Considera-se significativa que a proporção da balança comercial brasileira dependa da soja (MAPA, 2020).

O preço da oleaginosa foi objeto de previsão de muitas pesquisas anteriores (DAVID et al., 2017; DRACHAL, 2019; KITWORAWUT; RUNGREUNGANUN, 2019; PINHEIRO; SENNA, 2017b; PUCHALSKY et al., 2018; RIBEIRO; COELHO, 2020; SOUZA et al., 2018; WANG et al., 2017, 2019; ZHANG et al., 2018).

Tendo como base o estudo bibliométrico apresentado no capítulo II, o presente trabalho visa preencher a seguinte lacuna de pesquisa: do portfólio de artigos avaliados, nenhum comparou simultaneamente o desempenho de previsão das redes neurais *perceptron* multicamada (MLP), *extreme learning machine* (ELM) e modelos autorregressivos integrados de médias móveis com variável exógena (ARIMAX) nos preços da soja no Brasil. Portanto, a pesquisa atual é uma tentativa de preencher essa lacuna na literatura empírica.

No estudo bibliométrico realizado nas bases de dados *Web of Science* e *SCOPUS* não foi encontrada uma revisão bibliométrica ou sistemática sobre previsão de preços de *commodities* agrícolas. Os estudos dedicados à revisar este tema ainda são poucos, como as revisões de literatura (BRANDT; BESSLER, 1983; WANG et al., 2020).

Os modelos abordados na análise bibliométrica realizada podem fornecer referência para os pesquisadores do campo da previsão de preços de produtos agrícolas, onde a pesquisa aponta as tendências sobre a temática propostas no capítulo II desta dissertação, tendências essas que podem servir de base para os econometristas. Devido às características de complexidade e não linearidade de uma série temporal de preços de *commodities* agrícolas, dois modelos de inteligência artificial (IA) e o modelo estatístico ARIMAX são usados na previsão de preços da soja.

Segundo Ribeiro (2014, p. 12), “a definição do problema envolve a compreensão de como a previsão é utilizada, aliada com as expectativas dos futuros utilizadores”. Neste processo devem ser indicados o modelo, o tamanho da janela de previsão e qual é o nível de acuracidade das predições (RIBEIRO, 2014). A previsão não é um fim em si mesmo, contudo é um meio de produzir informações para apoiar os objetivos de determinada tomada de decisão dos agricultores (MARCHEZAN; SOUZA, 2010).

Por ser a principal *commodity* agrícola brasileira, o processo de prever o preço da soja possui implicações importantes, assim como a comparação entre métodos para descobrir

o que deve acontecer com os preços tem sua relevância, dado que indica aos participantes do mercado agrícola o modelo de previsão mais preciso que pode auxiliar os agentes a tomar as decisões relacionadas aos preços futuros.

A característica principal da comercialização de *commodities* é o alto grau de incerteza relacionado ao preço futuro do produto; por esta razão, a previsão dos preços tem um papel fundamental para os agentes do mercado. Mediante as predições, os atores do mercado podem decidir de forma mais assertiva sobre o momento de vender e de comprar, incorporando as informações geradas pela previsão na gestão de risco (RIBEIRO, 2014).

O Brasil tem forte participação no mercado de soja. Isto acarreta para o produtor de soja muita exposição às demandas do mercado externo e as variações de preços da *commodity* (AIUBE; FERREIRA; LEVY, 2020). A previsão dos preços pode ajudar a cadeia de abastecimento agrícola na tomada de decisões necessárias para minimizar o risco de flutuações de preços (SABU; KUMAR, 2020).

Acerca de preços das *commodities*, Aguiar e Borestein (2012), afirmam que há espaço para estudos que comparam o desempenho de modelos de previsão. Fang et al., (2020); Ribeiro e Coelho (2020); e Zhang et al., (2020), reconhecem a importância da acurácia dos modelos de previsão na tomada de decisão agrícola, tanto na esfera pública para as autoridades agrícolas quanto ao mercado agrícola para os gestores e executivos do setor.

Os preços das *commodities* agrícolas estão sujeitos à alta volatilidade, sendo relevante comparar a acuracidade da previsão para os tomadores de decisão (BRANDT; BESSLER, 1983; SABU; KUMAR, 2020). Por esta razão, é extremamente importante avaliar a eficácia de um método de previsão (KOHZADI et al., 1996; MORETTIN; SINGER, 2020; PETROPOULOS et al., 2020).

A pesquisa de Lima et al. (2010) testou os modelos ARIMA-GARCH e de redes neurais artificiais (RNA) aplicados em uma série de preços da soja. Na comparação entre os modelos a RNA superou a combinação dos modelos ARIMA-GARCH.

Jha e Sinha (2013) abordaram as previsões de preços da soja e mostarda de colza por meio do ARIMA, rede neural de atraso de alimentação (TDNN) e o híbrido ARIMA-TDNN com o objetivo de comparar estes modelos. Os resultados demonstram de modo geral que o modelo TDNN obteve melhor desempenho de predição em comparação ao ARIMA.

Feliciani, Souza e Souza (2015) compararam a capacidade preditiva dos modelos combinados ARIMA-ARCH e SARIMA-ARCH nos preços de tonelada da soja. Os resultados mostraram que o método SARIMA adaptou-se melhor a série em comparação ao modelo ARIMA-ARCH, mostrando, portanto, um desempenho de previsão superior.

Pinheiro e Senna (2017a) previram os preços da soja e açúcar por explorar a combinação de dois modelos: análise espectral singular (AES) e RNA. As principais constatações foram que, em comparação aos modelos individuais (Holt-Winters, RNA, SARIMA, AES), a combinação (AES-RNA) teve melhor desempenho de previsão.

Souza et al. (2018) usaram os preços da soja, saca de 60 kg, do estado de São Paulo, para gerar as previsões com o método ARIMA e constataram que o ARIMA (0, 1, 5) foi eficaz nas predições do preço da soja de São Paulo.

Kitworawut e Rungreunganun (2019) compararam a previsão de preço do milho da Tailândia dos modelos univariados ARIMA, Holt-Winters sazonal multiplicativo e aditivo. Os resultados evidenciaram que o modelo ARIMA (2,0,0) superou ambos os modelos Holt-Winters sazonal aditivo e multiplicativo, gerando os menores erros de previsão.

Ribeiro e Coelho (2020) estudaram a previsão dos preços da soja e trigo com a proposta de comparar o desempenho da previsão de conjuntos de regressão GBM, XGB, RF e STACK, bem como os modelos individuais MLP, SVR e KNN. Os principais resultados foram que os modelos baseados em abordagem de conjuntos apresentaram desempenho relevante (altamente precisos) no curto prazo, enquanto os modelos XGB e STACK tiveram as melhores acurácias.

Existe uma enorme diversidade de modelos de previsão, aplicados no preço da soja, conforme apresentado nos estudos anteriores. Embora os modelos de inteligência artificial sejam mais sofisticados, é necessário comparar a precisão da previsão com métodos de predição tradicionais, como o modelo ARIMAX e, a partir disso, escolher a abordagem que obteve os menores erros de previsão, com base nos indicadores de desempenho do erro da previsão.

Diante das constatações dos estudos anteriores não existe um único modelo de previsão para os preços da soja que tenha sido superior em todos os horizontes. Em determinados horizontes de previsão, tem modelos que até se destacam em comparação a outras abordagens de predição, mas nenhum método chega a ser o superior em todas as janelas de previsão. Este fato possibilita que modelos de previsão sejam testados e comparados, ou seja, mais pesquisas a fim de contribuir para encontrar modelos mais precisos para previsão de preços da soja.

O problema da pesquisa foi modelar os preços da soja para os estados do Paraná, Rio Grande do Sul e Mato Grosso. Este problema é importante, pois ajuda os agentes de mercado a lidar com a diminuição de incertezas relacionadas ao preço futuro da soja; aponta aos pesquisadores e participantes do mercado os métodos mais precisos na previsão de preço da

soja; amplia a discussão sobre previsão de preços de *commodities* agrícolas, ao comparar abordagens de previsão que aumentam a popularidade do modelo mais preciso entre econométristas. Conforme descrito na introdução, os preços da soja sofrem alta volatilidade, o que dificulta a tomada de decisão agrícola. Uma das características do mercado de soja é sua difícil previsibilidade dos preços; sendo assim, os modelos de previsão podem apoiar o processo de tomada de decisão dos produtores de soja na comercialização da *commodity*.

Com base na importância da precisão da previsão de preços da soja, questiona-se: qual método de previsão tem melhor precisão em relação ao preço da soja? Entre os modelos de predição testados, qual gerou as previsões mais precisas para os preços da soja? Então, qual é a melhor abordagem de previsão para os preços da *commodity*? A presente dissertação considera responder as questões norteadoras.

Visando atender as questões de pesquisa delineadas neste estudo, avalia-se e identifica-se os modelos de previsão com a melhor acurácia entre modelos autorregressivos integrados de médias móveis com variável exógena (ARIMAX), máquina de aprendizagem extrema (ELM) e redes neurais *perceptron* multicamada (MLP).

A primeira avaliação realizada para responder as questões norteadoras foi comparar as métricas de erro, mostrando que o ARIMAX obteve um melhor ajuste para dentro e fora da amostra e teve melhor capacidade preditiva para os preços da soja do Paraná. De modo geral, a rede neural teve o melhor desempenho de previsão do preço da soja do Rio Grande do Sul. Para previsão de preços do Mato Grosso, a rede neural gerou valores menores em todos os indicadores de desempenho.

De forma complementar, para melhorar as comparações desenvolvidas, as métricas de erros dos modelos de previsão são plotadas e comparadas em gráficos de Boxplots. Esta segunda avaliação permite concluir que o modelo ARIMAX foi superior em comparação às abordagens de IA para os preços do Paraná, no entanto, teve maior variabilidade dos erros de previsão para os preços do Mato Grosso. O método de rede neural (MLP) foi mais preciso em relação aos modelos concorrentes (ARIMAX e ELM), tanto na dispersão quanto a mediana dos erros de predições de preços do Rio Grande do Sul e Mato Grosso.

Em segundo lugar, os resultados dos testes de variâncias dos modelos comprovam ter diferenças estatísticas significativas entre as abordagens de previsão: ARIMAX comparado ao modelo de rede neural (MLP) para os preços do Paraná; MLP em comparação a máquina de aprendizado (ELM) para os preços do Rio Grande do Sul; MLP em relação ao modelo ARIMAX para os preços do Mato Grosso e a variância dos erros da ELM comparada com erros do modelo ARIMAX, também nos preços do Mato Grosso. Os resultados das demais

comparações não tiveram diferenças significativas ($p > 0,05$) entre as variâncias de erros dos modelos de previsão.

1.1.1 Objetivos

1.1.2 Objetivo geral

Avaliar os modelos de previsão de preço da soja comparando os métodos ARIMAX, redes neurais e máquina de aprendizado extremo.

1.1.3 Objetivos específicos

- a) Identificar as lacunas de pesquisa sobre previsão do preço de *commodities* agrícolas;
- b) Mostrar as principais tendências na previsão do preço de *commodities* agrícolas por meio da análise bibliométrica;
- c) Gerar previsões dos modelos ARIMAX, redes neurais *perceptron* multicamada e máquina de aprendizagem extrema nos preços de soja;
- d) Comparar o desempenho da previsão dos modelos ARIMAX, MLP e ELM por meio de métricas de erro.

1.1.4 Teoria e objeto de estudo

Nesta pesquisa, foi utilizado o modelo teórico da lei do preço único. O objeto de estudo foi a previsão de preço da soja. Esta dissertação fundamenta-se da mesma perspectiva dos estudos avaliados na revisão bibliométrica do capítulo II que mostram os modelos de previsão do campo da predição de preço de *commodity* agrícola com foco no preço da soja.

CAPÍTULO II

Este capítulo apresenta a revisão bibliométrica para identificar as lacunas de pesquisa sobre previsão de preços de *commodities* agrícolas e justificar a modelagem realizada. Este estudo de revisão está dividido em onze seções. A seção 2.1.1 apresenta o objetivo da revisão; as seções 2.1.2 até 2.1.4 mostram como foi realizada a coleta de dados com base nos artigos e os critérios adotados da análise bibliométrica. As seções seguintes apresentam os resultados, discussão e comentários finais.

Dois dos objetivos específicos desta pesquisa foi identificar as lacunas científicas sobre previsão do preço de *commodities* agrícolas e mostrar as principais tendências na previsão do preço de *commodities* agrícolas por meio da análise bibliométrica. Foi realizada a análise bibliométrica sobre a temática. Este capítulo responde ao primeiro e segundo objetivos específicos.

2.1 Revisão bibliométrica:

Previsão de preços das *commodities* agrícolas: uma revisão bibliométrica sobre modelos

2.1.1 Introdução

Este estudo é uma pesquisa bibliométrica com bases de dados internacionais, abrangendo o tema de modelagem dos preços das *commodities* agrícolas. Um estudo bibliométrico trata-se de um método quantitativo para medir os diversos índices de produção de um campo de pesquisa (QUEVEDO-SILVA et al., 2016).

Seu principal objetivo é quantificar a produção científica, apresentar tendências de pesquisa, apontar oportunidades de estudo e destacar os trabalhos mais relevantes de determinada área do conhecimento (QUEVEDO-SILVA et al., 2016).

O método da revisão bibliométrica, segundo Silva, Hayashi e Hayashi (2011), abrange dois tipos: pode ser avaliativa e ou descritiva. A ênfase avaliativa está relacionada com a contagem de referências e citações em trabalhos, já o enfoque descritivo referem-se à produtividade de autores, periódicos, países e abordagens das pesquisas. Nesta análise bibliométrica a abordagem é descritiva, na medida em que é quantificada a produção sobre o campo de pesquisa.

A primeira revisão de literatura sobre previsão de preços das *commodities* agrícolas foi publicada em 1983 por Brandt e Bessler (1983). Após tanto tempo, em abril de 2020, foi publicada uma nova revisão por Wang et al. (2020), apontando as tendências e avanços para o campo de pesquisa. Vale salientar que não se trata de um estudo bibliométrico, mas sim de uma revisão de literatura; falta uma revisão bibliométrica sobre este importante assunto que envolve diferentes métodos para a problemática de previsão de preço de *commodity* agrícola.

Em publicação recente, Wang et al. (2020) mostram a relevância de uma revisão sobre modelos de previsão aplicados nos preços de produtos agrícolas, revelando a oportunidade de pesquisa de uma análise bibliométrica acerca do campo de estudo. Modelos de previsão são úteis nas predições de preços de produtos agrícolas e servem para a tomada de decisão dos produtores rurais, investidores, autoridades agrícolas e participantes do mercado.

O objetivo da revisão bibliométrica foi identificar a lacuna de pesquisa sobre modelos de previsão aplicados nos preços das *commodities* agrícolas. Para alcançar este objetivo foi selecionado um portfólio de artigos, em encontro com o tema: previsão de preços agrícolas. Esta revisão resume os exemplos de métodos de previsão no campo da predição de preço de *commodity* agrícola com foco no preço da oleaginosa.

Essa revisão é relevante, pois existe uma competição na literatura empírica entre os métodos de predição mais precisos para previsão de preços agrícolas. Foram evidenciados os principais artigos, autores e periódicos mediante a análise bibliométrica que permitiu também encontrar a lacuna científica para o desenvolvimento de novas pesquisas. Como delimitação da pesquisa, a busca pelos artigos ocorreu nas bases de dados: *Web of Science* e *SCOPUS*.

2.1.2 Procedimentos metodológicos

Foi estabelecido como critério de inclusão somente artigos científicos publicados em periódicos, com fator de impacto nas bases de dados *Web of Science* e *SCOPUS*. Assim, teses, dissertações e artigos de congressos, simpósios e encontros científicos foram desconsiderados nesta revisão.

A revisão bibliométrica foi realizada com auxílio da interface *biblioshiny*, do pacote *bibliometrix* de Aria e Cuccurullo (2017) da linguagem de programação R., para retirar os artigos em duplicidades, ler os resumos e textos completos e exclusão de artigos com base nos critérios.

2.1.3 Processo de pesquisa

Foram selecionados artigos científicos publicados em periódicos nos idiomas português e inglês. O processo de procura consistiu em uma busca manual de artigos publicados em duas bases de dados para o período entre 01 de janeiro de 2010 até 25 de outubro de 2020. A pesquisa foi realizada utilizando as bases *SCOPUS* na opção título do artigo, resumo, palavras-chave e *Web of Science* coleção principal na opção tópico. Este recorte temporal corresponde a 85% do total de publicações sobre o campo de estudo nas duas bases de dados.

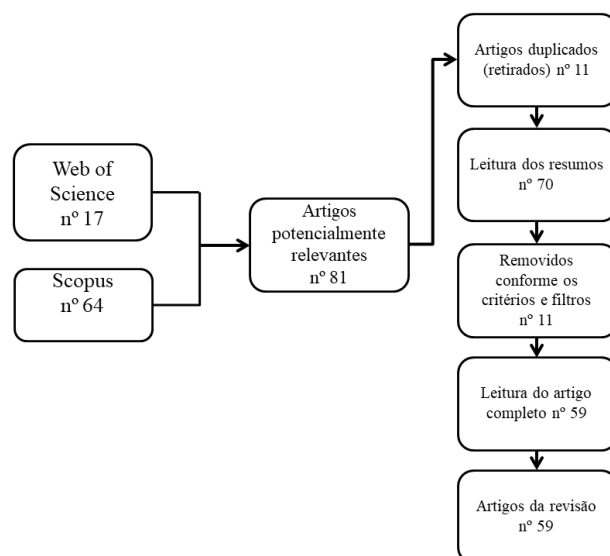
Para as duas bases, a busca foi realizada utilizando as seguintes palavras-chave com os respectivos operadores booleanos: (previsão AND preço da soja AND série temporal) OR (*forecast* OR "*forecasting models*" OR *forecasting*) AND ("*soybean price*" OR "preço da soja") AND ("*time series*"). O *download* dos artigos para leitura do resumo foi realizado nas duas bases de dados de forma manual.

2.1.4 Critérios de inclusão e exclusão

Nesta pesquisa, os principais critérios de inclusão foram: (i) ser um artigo de periódico; e (ii) modelos de previsão aplicados no preço de *commodity* agrícola. A busca nas bases *SCOPUS* e *Web of Science* foi realizada em vinte e cinco de outubro de dois mil e vinte, resultando em 81 artigos. Foram gerados os metadados no formato *BibTeX* nas duas bases e importados para o R, sendo identificados os artigos científicos duplicados e eliminados.

Como critério de exclusão, por meio da leitura dos títulos e dos resumos dos 70 artigos científicos, foram excluídos os artigos que não se enquadraram no objetivo proposto pelo estudo. A partir disso, foi verificado se tinham pesquisas em duplicidade, onde se constatou onze textos nesta situação, sendo descartada a possibilidade de usá-los como portfólio de artigos neste trabalho.

Figura 1 - Processo de seleção dos artigos para a pesquisa



Fonte: Elaborada pelo autor com base na revisão bibliométrica

Foram excluídos os artigos cujo conteúdo não tinham relação com previsão de preço de grãos e também os que não faziam parte do contexto da pesquisa de modelagem do preço. Apresenta-se no Quadro 1 a quantidade de artigos encontrados no período em cada base de dados.

Quadro 1 - Artigos científicos identificados em base de dados

Base de dados	Editora	URL	Artigos	Período
SCOPUS	Elsevier	https://www.elsevier.com/pt-br/solutions/scopus	42	01/2010 até 10/2020
Web of Science	Clarivate Analytics	https://clarivate.com/webofsciencengroup/solutions/web-of-science-core-collection/	17	01/2010 até 10/2020

Fonte: Elaborada pelo autor com base nos artigos do portfólio selecionado

O número foi refinado para 59 artigos científicos. Destes, 17 é da *Web of Science* e 42 artigos da base de dados *SCOPUS*, restando um total de 59 artigos, definindo a amostra para análise da revisão, conforme ilustrado no Quadro 1. Para contemplar as pesquisas de periódicos nacionais não indexados nas bases internacionais foram selecionados artigos do *Google Scholar*, pois são estudos relevantes tratando da temática.

2.2 Resultados

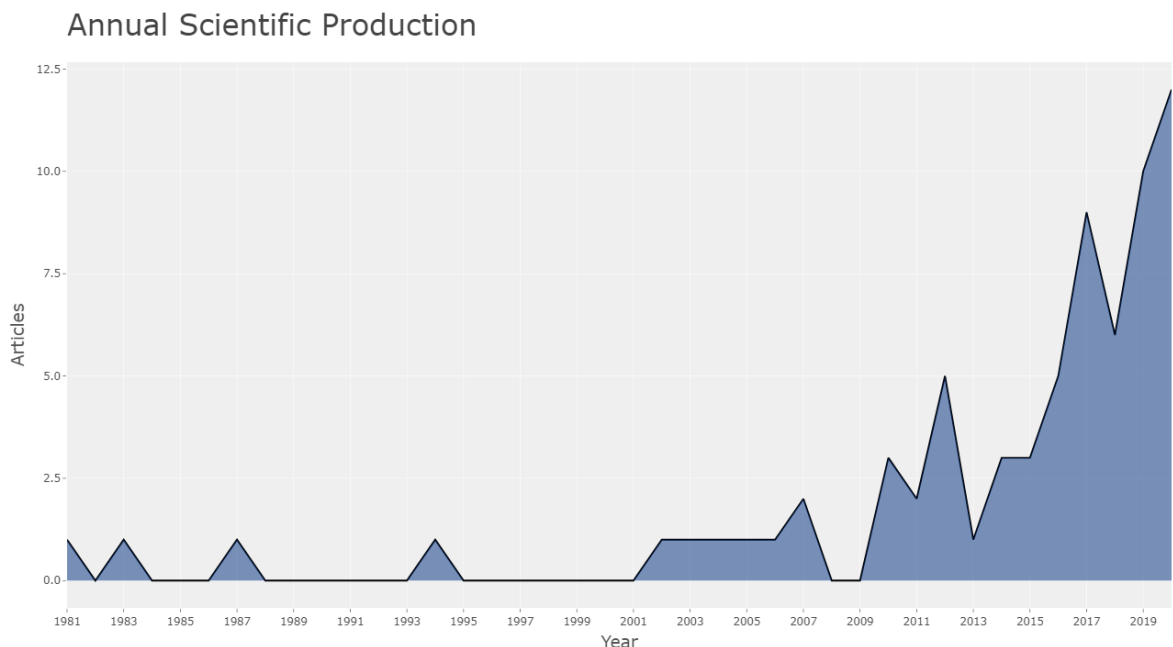
A partir do enfoque da pesquisa, este subcapítulo apresenta os resultados quantitativos da análise bibliométrica para identificar: (i) a produção científica anual das

publicações; (ii) periódicos que mais publicam sobre o tema; (iii) produção científica do país; e (iv) os países e autores mais citados. Para alcançar a proposta do estudo, os artigos encontrados foram tabulados em gráficos, a fim de ter um panorama em termos da quantidade de artigos publicados no período.

Demonstra-se na Figura 2 a evolução das publicações no decorrer dos anos; um total de 70 artigos foram encontrados com base na estratégia de busca. Nos últimos quatro anos o interesse por estudos sobre previsão de preços agrícolas tem aumentado significativamente. Com base nos artigos coletados das bases de dados *SCOPUS* e *Web of Science*, a produção científica anual aumentou exponencialmente a uma taxa de 13,23% de crescimento anual (Figura 2).

No intervalo entre 1981 e 2006, a quantidade de publicação não passou de um artigo publicado por ano. As publicações estão concentradas no período de 2010 a outubro de 2020. Neste intervalo foram encontrados 59 artigos, sendo que num período maior, entre 1981 até outubro de 2020, obteve-se 70 artigos, mostrando que 85% das publicações estão concentradas nos últimos dez anos.

Figura 2 - Produção científica anual das publicações



Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

Segundo Deng e Sün (2019) e Triantafyllou, Dotsis e Sarris (2015) a volatilidade, sem precedentes nos preços mundiais de *commodities* agrícolas, geram muita incerteza e risco para todos os participantes do mercado e torna o planejamento a curto e longo prazos mais

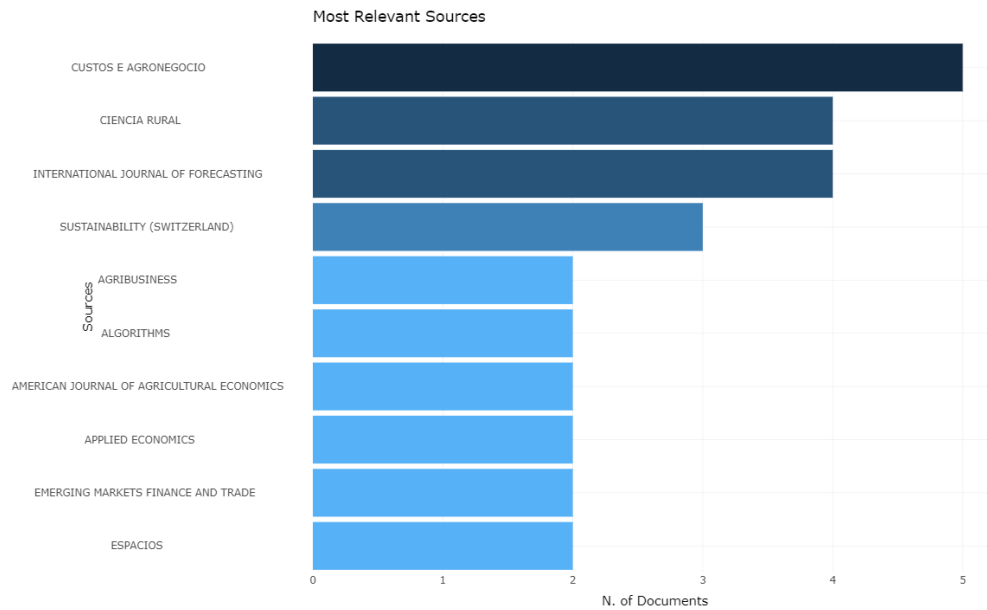
difícil. Para Drachal (2019) a partir da crise mundial de 2008 foi observado um aumento expressivo na variabilidade dos preços das *commodities* agrícolas. Sendo que as *commodities* agrícolas passaram por enormes oscilações de preços na última década (DEGIANNAKIS et al., 2020).

A Organização das Nações Unidas para a alimentação e a agricultura (FAO) afirma que os preços dos alimentos experimentaram alta volatilidade a partir de 2008 (FAO, 2010). Este fato de alta volatilidade, com início em 2008, é considerado como a principal razão do crescimento das publicações em 2010 (Figura 2). A quantidade de estudos sobre a previsão de preços das *commodities* agrícolas torna-se significativa em 2010. Posteriormente, as pesquisas sobre a temática aumentam claramente, mostrando que o interesse por previsões de preços agrícolas ocorreu principalmente em função das altas oscilações dos preços nesse período.

A volatilidade dos preços é um problema para o produtor da *commodity* agrícola, pois precisam tomar a decisão do melhor momento de comercialização do produto (Sabu e Kumar, 2020). Neste contexto, diminuir as incertezas com relação ao preço futuro da soja pode apoiar o processo decisório.

Foi gerado um gráfico com a quantidade de artigos por periódico. A Figura 3 apresenta as fontes mais produtivas nos últimos dez anos entre 2010 e outubro de 2020. Dentre as revistas científicas com maior número de publicações sobre previsão de preços agrícolas, se destacaram duas fontes brasileiras: Custos e @gronegócio online e Ciência Rural. Destacaram-se também dois periódicos: *International Journal of Forecasting* e *Sustainability*; as demais fontes publicaram dois artigos cada uma.

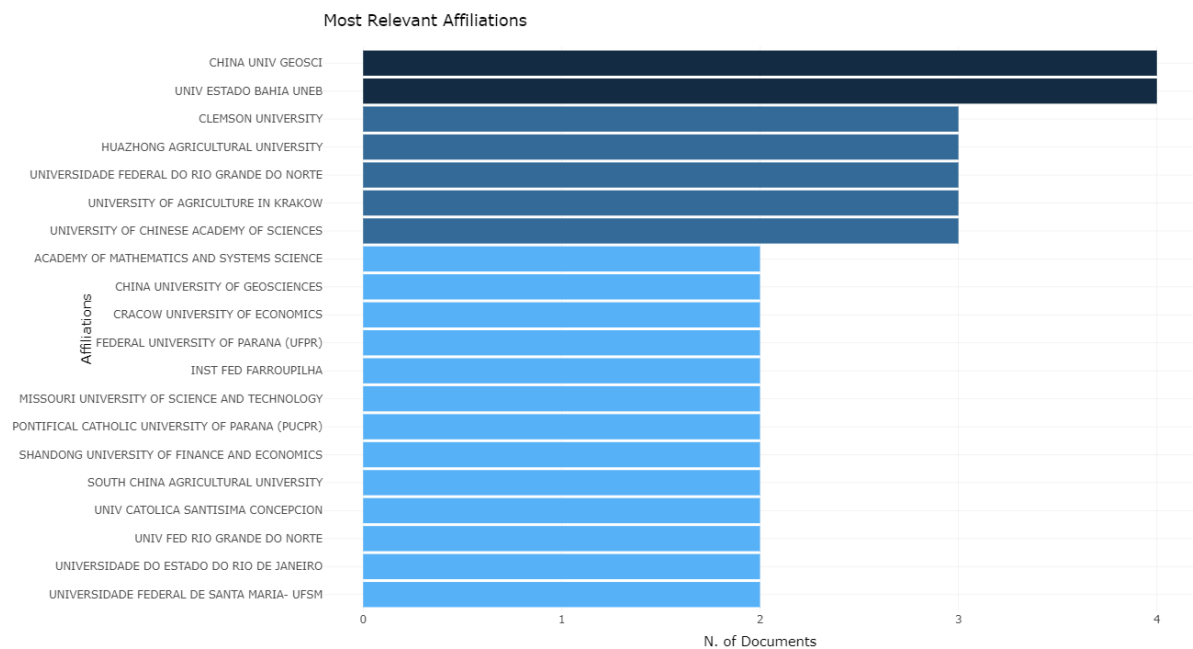
Figura 3 - Periódicos que mais publicam sobre previsão de preços agrícolas nos últimos dez anos



Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

As vinte instituições mais importantes, considerando o número de estudos gerados entre 2010 a outubro de 2020, são apresentadas na Figura 4. As universidades brasileiras que se destacaram na produção sobre a temática foram a Universidade do Estado da Bahia (UNEB) e Universidade Federal do Rio Grande do Norte (UFRN).

Figura 4 - Afiliações mais relevantes entre 2010 a outubro de 2020



Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

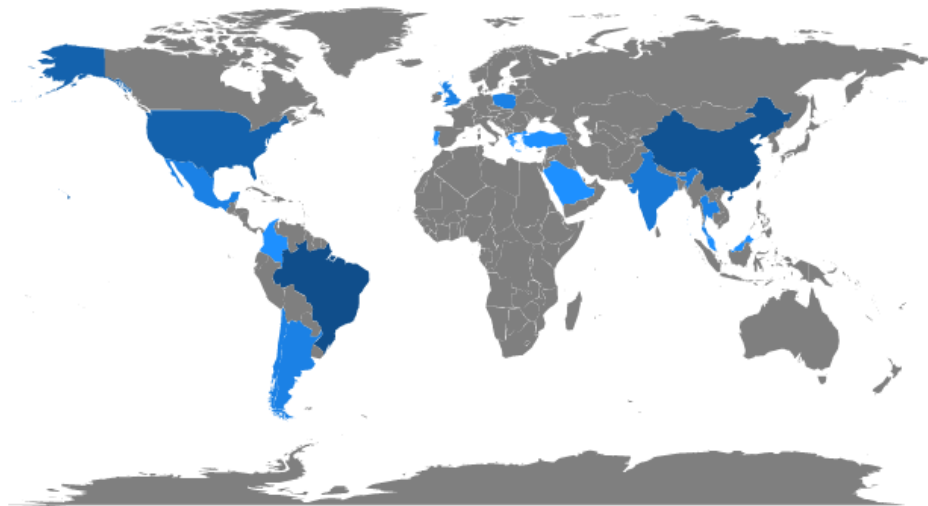
As instituições chinesas mais produtivas são *China University of Geosciences* e *Huazhong Agricultural University*. A maior parte, 75% das instituições que geram conhecimento neste campo de pesquisa, são chinesas e brasileiras.

A Figura 5 mostra os países mais produtivos entre 2010 a outubro de 2020; quanto mais escura a cor azul, maior foi o número de artigos publicados pelo país. As áreas cinzas indicam que a região ou país não publicou um artigo.

A produção mundial sobre o tema está concentrada em três países: Brasil, China e Estados Unidos da América. Estes são os principais países que geram conhecimento sobre previsão de preços das *commodities* agrícolas.

Figura 5 - Produção científica do país entre 2010 e outubro de 2020

Country Scientific Production

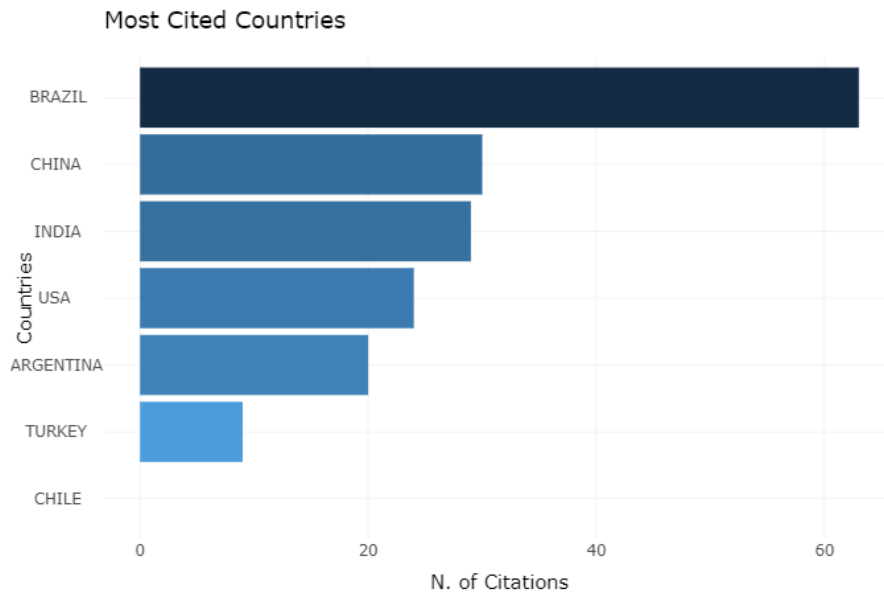


Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

O Brasil foi o país que mais contribuiu em termos de quantidade de artigos publicados para o avanço da pesquisa, liderando, dentre todos os países, com vinte e quatro artigos no período de (2010 a outubro de 2020). O segundo país mais produtivo foi a China, com dezenove artigos publicados no mesmo período. Os USA geraram nove estudos; Argentina, Chile, México, Polônia, Tailândia e Reino Unido produziram dois artigos cada.

Além do total de publicações destacados anteriormente, são mostrados a seguir os países mais citados durante o período analisado, como pode ser observada na figura 6:

Figura 6 - Países mais citados entre 2010 a outubro de 2020

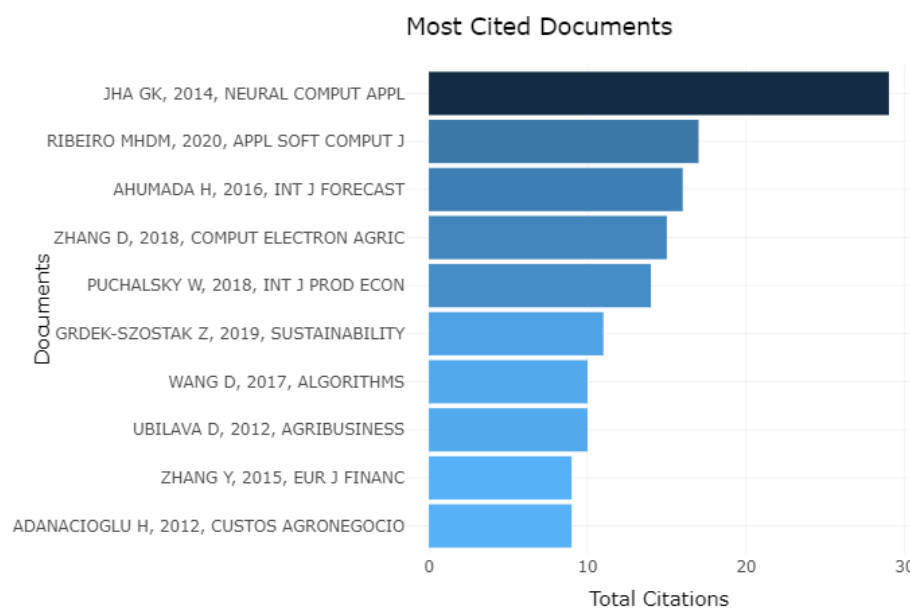


Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

O Brasil tem mais que o dobro de citações (63) em relação a China, que é o segundo país mais citado no período, com 30 citações. A Índia e USA foram citados 29 e 24 vezes, respectivamente. China e Índia tiveram praticamente o mesmo número de citações. Dentre todos os países destaca-se o Brasil como o mais citado do mundo nos últimos dez anos.

Para identificar os estudos mais relevantes em termos de total de citações, a figura 7 apresenta os dez artigos mais citados nas bases de dados *Web of Science* e *SCOPUS*, no intervalo de 2010 a outubro de 2020.

Figura 7 - Quantidade de citações por artigos do portfólio



Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

Observa-se na figura 7 que os quatros estudos mais citados foram Jha e Sinha (2012) com 29 citações; Ribeiro e Coelho (2020) obteve 17 citações; Ahumada e Cornejo (2016a) com 16 citações e Zhang et al., (2018) com 15 citações. Com estes estudos, mostra-se a importância dessas pesquisas no campo da previsão de preços das *commodities* agrícolas.

A partir da sumarização dos estudos foi realizada uma descrição das pesquisas: (i) os objetivos; (ii) as amostras com a periodicidade das séries de preços agrícolas; (iii) os intervalos de previsão que os autores decidiram; (iv) os indicadores de desempenho adotados por cada estudo, para avaliar a acurácia dos modelos de previsão; e (v) os principais resultados, as sugestões para pesquisas futuras e as limitações.

2.2.1 Descrição dos principais artigos encontrados na revisão sobre previsão de preços agrícolas

A literatura sobre modelagem e previsão de preços das *commodities* agrícolas experimentou uma expansão crescente nos últimos anos. Apesar desse crescente interesse pela predição de preço de produto da agricultura, pouco tem sido feito em termos de sistematização das contribuições de artigos que remetam ao tema. Assim, considera-se importante descrever os estudos.

A pesquisa de Marchezan e Souza (2010) teve o objetivo de realizar previsões para o ano de 2007 nas cotações de preço do milho, soja e arroz do Rio Grande do Sul (RS). Para isto, aplicou-se os modelos ARIMA e SARIMA e utilizaram-se os preços de cotações mensais no período de janeiro de 1995 até dezembro de 2006, coletados da Emater/RS-ASCAR.

Os resultados revelaram que o ARIMA (2,1,1)(0,1,0) foi eficiente para prever a série de preço de milho, ARIMA (0,1,2)(0,1,0) para preços da soja e o método sazonal SARIMA (1,1,0)(1,0,1) para o arroz, constatando-se que as previsões realizadas pelos modelos citados foram representativas da realidade. Para verificar quais modelos ajustou-se melhor aos preços foi usado o erro quadrado médio (MARCHEZAN; SOUZA, 2010).

O estudo teve como objetivo explorar os modelos ARIMA-GARCH e de redes neurais artificiais (RNA), aplicados em uma série de preços da soja. A pesquisa utiliza a decomposição da série através de uma ondaleta. A série preços da soja foi de julho de 1997 a março de 2008; a fonte da coleta dos dados foi o Cepea. Os modelos foram comparados por meio dos indicadores de desempenho MAPE, coeficiente de desigualdade de Theil e

correlação. Os resultados apontaram que o modelo que melhor se ajustou aos dados foi ARIMA-GARCH, com o uso de ondaletas. Já com o uso da decomposição de ondaletas, o modelo que gerou menor MAPE (1,1537%) foi a rede neural recorrente (LIMA et al., 2010).

O uso de ondaletas melhorou ambos os modelos de previsão. Os pesquisadores consideraram que os resultados estavam alinhados com uma pesquisa anterior sobre previsão de *commodities* agrícolas. Concluíram que as predições da RNA, com a decomposição de ondaletas, possibilitou uma ferramenta alternativa para prever série de preços da soja, assim as RNA superou o modelo ARIMA-GARCH. Também consideraram que o uso de RNA se torne difícil por não existir um método para determinar a arquitetura ideal para a série escolhida (LIMA et al., 2010).

Com o objetivo de propor um modelo de previsão para preços à vista de *commodities* agrícolas, o modelo hierárquico, rede neural *feedforward*, suavização exponencial e média móvel foram aplicados aos preços de soja, café arábica, açúcar e etanol. A previsão realizada foi de curto prazo. Para comparar os modelos, foram geradas as medidas de desempenho MAE, MSE, MAPE e variância do erro (RIBEIRO; SOSNOSKI; OLIVEIRA, 2010).

Os resultados indicaram que Suavização Exponencial geraram previsões melhores do que o modelo de Médias Móveis, mas a rede neural obteve desempenho de previsão superior em comparação aos modelos testados na pesquisa. Nas considerações finais pode-se dizer que o modelo hierárquico foi útil para prever preços de *commodities* e para identificar a dependência destes preços em relação a variáveis exógenas que estejam associadas a características micro ou macroeconômicas da *commodity* (RIBEIRO; SOSNOSKI; DE OLIVEIRA, 2010).

O artigo abordou as previsões de redes neurais *perception* multicamadas nos preços das *commodities* agrícolas soja, boi gordo, milho e trigo, com um enfoque de previsões estratégicas para tomada de decisão, onde um modelo de previsão pode ser um método estratégico. O objetivo dos autores foi validar as redes neurais por meio da previsão dos preços dos produtos citados anteriormente. A fonte dos preços foi a Emater-RS; a amostra foi em periodicidade mensal de janeiro de 1992 a dezembro de 2006. Para avaliar o modelo foram utilizados os critérios RMSE, coeficiente de desigualdade de Theil, R^2 e o MAPE (FERREIRA et al., 2011).

Os principais resultados mostraram que, para a amostra de preço da soja, o modelo gerou um RMSE (0,004906), Theil (0,02571), R^2 (0,91097) e o MAPE de 4,37%. No entanto, na amostra de preço do milho as previsões foram menos precisas RMSE (0,007656), Theil (0,040514), o R^2 foi de 0,772965 e o MAPE foi de 6,29%. Com base nos resultados de

previsão do Ferreira et al. (2011) afirmam que “o modelo proposto é uma boa alternativa para predição de preços futuros e para apoio da decisão no contexto do agronegócio brasileiro”, ou seja, a rede neural perception multicamada foi considerada uma alternativa como ferramenta para os gestores do agronegócio nacional (FERREIRA et al., 2011).

Para investigações futuras sugeriram-se empregar os métodos de elaboração de cenários por meio da internet. As previsões geradas pelo modelo foram relevantes, sendo altamente precisas, com excelentes níveis de acurácia para os preços de produtos agrícolas. Porém, não comparam as previsões da pesquisa com resultados de previsões de estudos anteriores sobre previsão de *commodities* agrícolas (FERREIRA et al., 2011).

Na previsão de preços de oleaginosas no atacado comercializadas na Índia utilizaram os modelos: rede neural de atraso de alimentação (TDNN) e ARIMA. Neste contexto, o objetivo foi avaliar a adequação da rede neural em comparação com ARIMA nas séries mensais de preços das principais culturas da Índia, a saber, soja, amendoim, colza e mostarda. Os últimos 12 meses foram utilizados para os testes de previsão. Para avaliar a capacidade de previsão dos modelos foram adotadas as medidas de desempenho RMSE e MAD (JHA; SINHA, 2012).

Os resultados de previsão apresentaram que, nos horizontes 6 e 12 passos à frente, o modelo TDNN obteve um melhor desempenho; entretanto, nos curtos prazos 1 e 3 passos à frente, o ARIMA foi mais preciso em comparação com TDNN. De modo geral, considerando todas as séries de preços de oleaginosas e todos os intervalos de previsão, os resultados sugerem um desempenho superior do modelo TDNN sobre o ARIMA. Sendo, que as conclusões foram o desempenho da TDNN foi melhor que o ARIMA para todos os horizontes, exceto 1 mês à frente. Os modelos de redes neurais são preferíveis em relação aos modelos lineares, no contexto de previsão de preços agrícolas (JHA; SINHA, 2012).

O estudo abordou as previsões mensais de preços no atacado de soja e mostarda de colza por meio da RNA, como suporte às decisões relacionadas à segurança alimentar na Índia. Neste sentido, o objetivo da pesquisa foi comparar a previsão dos modelos ARIMA, TDNN e a metodologia híbrida ARIMA-TDNN nos preços de soja e mostarda colza comercializadas neste país. O horizonte de previsão foi de até um ano à frente, assim incluíram os resultados de um mês, três meses, seis meses e doze meses à frente. A capacidade de modelar os preços foi avaliada com base nos critérios de desempenho: RMSE e MAD (JHA; SINHA, 2013).

Os resultados demonstram de modo geral que o modelo TDNN obteve melhor desempenho de predição em comparação ao ARIMA, exceto para um horizonte de previsão

de um mês à frente no qual o ARIMA foi mais preciso do que TDNN. Em geral, o modelo TDNN teve melhor desempenho de previsão em 6 e 12 meses à frente, enquanto os modelos ARIMA dominam nas previsões de 1 e 3 meses. Para a série de preço de soja, o modelo híbrido ARIMA-TDNN forneceu previsões superiores em comparação com ARIMA e TDNN. Entretanto, no caso da série de preços da mostarda de colza, o método híbrido superou ambos os modelos em todos os horizontes de previsão. Os achados empíricos sugeriram que o modelo híbrido gerou melhores previsões (JHA; SINHA, 2013).

Concluiu-se com o estudo que os modelos não lineares podem ser preferíveis aos modelos lineares para previsão de preços agrícolas. As previsões preços agrícolas são informações cruciais para a produção em todos os níveis, assim como a previsão deve ser precisa e útil para as decisões locais de agricultores, comerciantes e formuladores de políticas. Este cenário requer um sistema de suporte de apoio à decisão baseado em inteligência artificial (IA) e métodos estatísticos clássicos com uso software de computação que forneça aconselhamento personalizado a cada agricultor, tendo em vista as suas condições locais (JHA; SINHA, 2013).

Na pesquisa aplicaram-se os modelos ARIMA e Alisamento Exponencial de Holt-Winters na série temporal do preço médio mensal do milho no Rio Grande do Sul. Os resultados revelaram que a previsão do modelo Holt-Winters versão aditiva foi por uma pequena diferença mais precisa, julgado pelos critérios MAD e SSE. O autor considerou que ambos os modelos resultaram previsões satisfatórias e podem ser aplicados para previsão do preço do milho e outras culturas (TIBULO; CARLI, 2014).

O estudo teve como proposta prever e descrever mediante os modelos mistos ARIMA-ARCH e SARIMA-ARCH uma série de preços de tonelada de soja produzidos no Brasil no período de 2000 a 2011, num único horizonte de seis meses à frente. A série de preços foi coletada da Associação Brasileira das Indústrias de Óleos Vegetais (ABIOVE). Para mensurar a acurácia do modelo foram utilizados os critérios de desempenho MAPE e EQM (FELICIANI; SOUZA; SOUZA, 2015).

Os resultados mostraram que o método SARIMA (1,0,0) (1,1,0)₁₂ - ARCH (1), adaptou-se melhor a série em comparação ao modelo ARIMA-ARCH, tendo um desempenho de previsão superior, com a precisão do MAPE (0,090) e EQM (0,132). Concluiu-se que SARIMA (1,0,0)(1,1,0)₁₂ - ARCH (1), foi um bom modelo de previsão. Para pesquisas futuras sugere-se aplicar os modelos Threshold (ARCH) (TARCH) (FELICIANI; SOUZA; SOUZA, 2015).

O estudo verificou empiricamente a previsão de preços fora da amostra em uma estrutura multivariada para prever o preço da soja. A pesquisa testa cinco modelos multivariados, a saber, o modelo de preços futuros (FPM), FPM aumentado, modelo de fundamentos e caminhada aleatória. Os períodos de previsão selecionados são de três, seis e nove meses à frente e as medidas de erro utilizadas para comparar as estruturas multivariadas são MAPE e RMSE (AHUMADA; CORNEJO, 2016).

Nas comparações empíricas das previsões ficou constatado que o FPM aumentado e modelo de fundamentos mostraram ganhos de previsão. No curto prazo o modelo de fundamentos pode ser melhor que os outros métodos. Concluiu-se que a descoberta de preços da soja é uma função central dos mercados futuros. Neste sentido, fornecer modelos alternativos para previsões de preços futuros mais confiáveis são úteis para produtores e consumidores de produtos agrícolas. Prever preços de soja fora da amostra em uma estrutura multivariada é uma abordagem diferente do que é mais utilizado na literatura. Dependendo do horizonte, os modelos alternativos são uma opção para aquelas previsões que acontecem fora da amostra (AHUMADA; CORNEJO, 2016).

A pesquisa de Ahumada e Cornejo (2016a) realizou a previsão de preços do milho, soja e trigo no curto prazo 1 e 4 passos à frente, através de modelos econométricos de correlação entre os preços e uma estrutura multivariada com variáveis explicativas de *commodities*. Os modelos usados na pesquisa foram uma correção de equilíbrio (EqCM) e caminhada aleatória. O objetivo foi analisar as precisões de previsão dos modelos nos preços de alimentos considerando a dependência cruzada dos preços dos produtos agrícolas.

Na avaliação do desempenho dos modelos de previsão foram utilizados os indicadores de desempenho denominados RMSE e MAPE. Os resultados mostraram que, no caso dos preços do milho e soja, o EqCM foi melhor previsor. Por grande parte do período fora da amostra os achados indicaram que os modelos EqCM apresentaram desempenho superior do que a caminhada aleatória. Concluíram que melhorias nas previsões foram alcançadas usando modelos que levam em consideração as interações dos preços, isto é, EqCMs conjuntos (AHUMADA; CORNEJO, 2016a).

Na previsão de curto prazo de preços dos contratos futuros de café, etanol, bovinos, milho e soja negociados na BM&FBOVESPA foi abordado uma combinação das RNA com análise multivariada singular espectral (AESM). Para comparar esse modelo utilizou-se os modelos clássicos e redes neurais. O objetivo foi comparar os resultados de previsão do modelo híbrido (AESM-RNA) com os Holt-Winters, versões aditiva e multiplicativa, e SARIMA. Para avaliar as previsões dos modelos adotaram-se as medidas de erro Erro

quadrado médio (MSE) e Quadrado médio da raiz cumulativa (CRMS) (PINHEIRO; SENNA, 2016).

Os achados mostraram que a previsão dos modelos HW teve desempenho de previsão semelhante. O modelo AESM-RNA superou todos os outros modelos testados, exceto para a série de preços de bovinos na qual o SARIMA gerou um melhor desempenho de predição. Nas considerações finais foi ressaltada a importância da predição de preços de contratos futuros agrícolas para aos agentes do mercado e especuladores, pois a tomada de decisão presume conhecimento dos preços futuros. A contribuição empírica foi constituir um modelo para apoiar na formulação e implementação de políticas de gerenciamento de riscos, mediante a previsão precisa dos preços dos contratos futuros agrícolas.

Na previsão de preço das *commodities* agrícolas: etanol, boi gordo, milho, café e soja. A proposta do artigo foi avaliar a contribuição do modelo AESM na previsão dos preços de *commodities* agropecuárias. As séries temporais são do banco de dados da ESALQ e correspondem ao período de novembro de 2008 até dezembro de 2013, com periodicidade semanal. Os horizontes de previsão foram de 3, 6, 9 e 12 semanas à frente, fora da amostra. O estudo usou os modelos HW versões sazonal aditiva e multiplicativa, SARIMA, ARIMA-GARCH e RNA. Para medir a acuracidade dos modelos foram adotados as medidas de erro: Erro Quadrático Médio (EQM) e Erro Quadrático Acumulado (EQA). Os resultados de previsão mostraram que os modelos HW tiveram desempenhos semelhantes, mas o modelo AESM superou todos os modelos comparados na previsão, assim o modelo AESM agrega evidências favoráveis quando aplicado na previsão dos preços das *commodities* utilizadas na pesquisa (PINHEIRO; SENNA; MATSUMOTO, 2016).

Concluiu-se que existe grande relevância em prever preços das *commodities* agrícolas para os participantes do mercado e foram comparados os resultados da pesquisa com os achados de estudos anteriores. Os resultados contribuíram empiricamente para o mercado agropecuário, por auxiliar na formulação e na aplicação de políticas direcionadas ao setor agropecuário, por meio da previsão como um instrumento para o planejamento e análise do comportamento dos preços agrícolas. Para as investigações futuras, recomenda-se usar mais de uma base de dados e incluir mais *commodities* num horizonte de previsão mais amplo (PINHEIRO; SENNA; MATSUMOTO, 2016).

Foi aplicada a metodologia proposta por Pinheiro e Senna (2015) nos preços dos produtos açúcar, algodão, milho, café e soja. A periodicidade das séries de preços foi semanal e o período de previsão foi de doze semanas fora da amostra. A comparação dos modelos ANN-MSSA e RNA foram por meio do critério de avaliação erro médio quadrático (MSE).

Após a utilização dos testes de normalidade nas séries de preços e intervenção supervisionada na RNA, os autores decidiram a quantidade de camadas e neurônios.

O estudo combinou o modelo MSSA para decompor a série temporal com o modelo da RNA. Os resultados mostram que ANN-MSSA obteve menores erros de previsão. As previsões também foram comparados com as pesquisas de previsão de preços agrícolas, mostrando que os resultados encontrados estavam alinhados com achados de pesquisas anteriores. Os autores consideraram que o modelo ANN-MSSA pode ser uma opção que contribui na previsão de preços de produtos agrícolas (PINHEIRO; SENNA, 2017b).

A pesquisa dos autores previu os preços da soja e açúcar no Brasil num horizonte de doze semanas à frente, explorando a combinação de dois modelos: análise espectral singular (AES) e rede neural artificial (RNA). As séries de preços foram coletadas da ESALQ e os indicadores de desempenho adotados foram erro quadrático médio (EQM) e erro quadrático acumulado (EQA). As principais constatações foram que, em comparação aos modelos individuais (HW sazonal aditiva e multiplicativa, RNA, SARIMA, AES), a combinação (AES-RNA) teve melhor desempenho de previsão (PINHEIRO; SENNA, 2017a).

A proposta apresentada é a combinação de um modelo e demonstração da superioridade na acuracidade da previsão do mesmo (AES-RNA). Concluiu-se que é importante pesquisar sobre a combinação de modelos e evidências científicas de que este (AES-RNA) pode ser adotado pelos profissionais do setor para prever preços de *commodities* agrícolas. Por fim, sugere-se usar maior número de séries preços com *commodities* agrícolas diferentes e mais horizontes de previsão (PINHEIRO; SENNA, 2017a).

A pesquisa foi um experimento que utilizou os preços de trigo, milho e soja com objetivo de enriquecer a literatura de pesquisa e oferecer uma nova forma de pensar sobre a previsão de preços de *commodities* agrícolas. Foi proposto quatro modelos híbridos baseados na rede neural de propagação traseira (BPNN) otimizados pelo algoritmo de otimização de enxame de partículas (PSO) e métodos de decomposição: decomposição do modo empírico (EMD), transformação de pacotes de wavelet (WPT), escala de tempo intrínseca decomposição (ITD) e decomposição de modo variacional (VMD) (WANG et al., 2017).

Os principais resultados mostraram que (I) todos os modelos híbridos combinados com a técnica de decomposição tem um desempenho melhor que o modelo único PSO–BPNN; (II) VMD contribui mais para melhorar a capacidade de previsão do modelo PSO–BPNN, enquanto o WPT classifica em segundo; (III) ITD tem melhor desempenho que EMD em ambos os casos de milho e soja. Chegou-se à conclusão que os modelos propostos tiveram

bom desempenho na previsão de preços futuros de *commodities* agrícolas (WANG et al., 2017).

Mediante os modelos de Média Móvel Fracionária Integrada (ARFIMA) e ARIMA, foi previsto quatro séries de preços de produtos agrícolas, sendo eles soja, milho, café e açúcar. As amostras são de janeiro de 2009 a dezembro de 2013; já o período de previsão fora da amostra foi de janeiro a dezembro de 2014. A fonte de coleta dos dados foi o CEPEA. O objetivo da pesquisa foi examinar as séries preços dos produtos agrícolas. Adotou-se os indicadores de desempenho para avaliar as predições. No caso, foram medida de Entropy-Theil (E), MAE, autofunção de correlação no atraso 1 (ACF1) e MASE (DAVID et al., 2017).

Os principais achados na comparação entre os modelos aplicados mostraram que ARFIMA foi o melhor previsor nos preços de açúcar, café e soja, mas para o preço de milho o modelo ARIMA foi melhor. Na avaliação dos indicadores MASE, o modelo ARIMA gerou melhores previsões. As previsões de longos prazos feitas pelo ARFIMA foram significativamente melhores em comparação com ARIMA. O modelo ARFIMA gerou um desempenho de previsão melhor em comparação ao ARIMA, para os preços de *commodities* agrícolas (DAVID et al., 2017).

Zhang et al. (2018) criaram um algoritmo híbrido (GDGA) de previsão para o modelo de rede neural com função de regressão radial quantílica (QR-RBF), para prever o preço mensal da soja na China. Os dados usados são de janeiro de 2010 a dezembro de 2015, em um total de oito variáveis como principais determinantes sobre o preço.

O período de previsão foi de oito meses à frente. Para mensurar a precisão dos modelos, os autores usaram o erro relativo (ER) e MAPE; também compararam a precisão com o algoritmo genético (GA) e regressão linear multivariada. Os principais achados demonstraram que o maior valor absoluto de ER foi de 1,79% e o MAPE foi de 1,11%, para o modelo (QR-RBF) com o algoritmo (GDGA), o que os pesquisadores afirmaram ser uma precisão satisfatória e eficaz, em contraste com os resultados de previsão dos modelos GA e regressão linear multivariada, com MAPEs 1,77% e 5,86%, respectivamente (ZHANG et al., 2018).

Esses valores indicaram que QR-RBF apresentou desempenho superior na previsão. Foi realizado o confronto dos resultados da pesquisa com a literatura pertinente ao tema. O resultado foi que o GDGA forneceu previsões melhores na comparação com os dois modelos e o QR-RBF foi significativamente melhorado pelo algoritmo híbrido GDGA (ZHANG et al., 2018).

No estudo, o objetivo foi compreender a variável preço no contexto do planejamento e nas tomadas de decisão na comercialização da soja no estado de São Paulo, por meio dos preços do grão, saco de 60 kg. Os dados foram coletados no Instituto de Economia Agrícola, com periodicidade mensal. A amostra foi de janeiro 2000 a dezembro de 2016. Analisando a variação percentual dos preços, o autor evidenciou que os preços da soja tem alta variação dentro de um mesmo ano como ao longo dos dezesseis anos avaliados (SOUZA et al., 2018).

Para gerar as previsões do ano de 2017 foi utilizado o método ARIMA (0, 1, 5). As previsões do modelo ARIMA foram feitas dentro e fora da amostra. A partir disso, determinou-se os limites de confiança inferior (LIC) e superior (LSC) das previsões dos preços da soja ao longo do ano de 2017. A fim de verificar a precisão do modelo ARIMA (0, 1, 5), foram encontrados os erros R^2 (0,924), RMSE (4,042), MAPE (4,314) e MAE (3,086) para o modelo. Verificou-se que o preço da soja de São Paulo sofreu quedas e aumentos durante determinados períodos do ano, assim como foi explicado quais são os motivos desses movimentos sazonais. Foi constatado que o ARIMA (0, 1, 5) foi eficaz para predições do saco de soja de São Paulo (SOUZA et al., 2018).

Na previsão de médio prazo de preço do milho no Brasil foram aplicados os modelos ARIMA, com o objetivo de avaliar a potencialidade da utilização do modelo na previsão de preços do milho. As observações analisadas durante a pesquisa foram de março de 2004 a março de 2016. Foi usada a estatística U de Theil para mensurar a acurácia, onde os resultados mostraram um coeficiente U-Theil (0,0206) com base nas predições. A pesquisa considerou que antecipar preços do milho gera vantagens aos interessados no produto, pois é um mercado competitivo. O modelo ARIMA (1,2) foi o melhor para prever a série de preço do milho analisada. Recomenda-se usar os modelos ARMAX e ARCH nas pesquisas futuras (CAS, 2018).

A pesquisa combinou cinco modelos de otimização com redes neurais da Wavelet (WNNs) para melhorar a qualidade do WNN. As previsões do estudo consideraram duas séries temporais: preço da soja e demanda de produto alimentício. As previsões foram avaliadas no curto e longo prazos de 1 mês para soja e 12 meses para demanda de produtos. Os indicadores de desempenho usados para avaliar os modelos combinados foram MAPE e MSE. Realizadas as simulações e previsões do preço da soja constatou-se que, dentre todos os modelos testados, o ELM e otimização de enxame de vermes (GSO) foram mais precisos no curto prazo, gerando erros de MSE (0.17999), MAPE (0.76450) e MAPE (1.71919), (PUCHALSKY et al., 2018).

Concluiu-se que o ELM supera a maioria dos modelos durante os procedimentos de validação. No entanto, os algoritmos usados para combinar com o modelo WNN superaram os resultados do ELM em quase todos os casos. Também foi constatado que, nos casos em que a série temporal tem poucas observações disponíveis, o WNN parece ser uma abordagem consistente, fornecendo uma previsão razoável. Para pesquisas futuras, recomenda-se análises mais aprofundadas nos melhores algoritmos, otimização para obter parâmetros que aprimoram as previsões, além de testar séries temporais com o componente sazonal presente na série (PUCHALSKY et al., 2018).

No preço do milho da Tailândia compararam-se modelos univariados ARIMA, Holt-Winters sazonal multiplicativo e Holt-Winters sazonal aditivo, com objetivo de modelar os fatores que afetam o preço do milho e estabelecer um modelo de previsão de preço deste grão. Para comparar a acuracidade dos três modelos no horizonte de 12 pontos à frente, foram usadas as medidas de erro RMSE, MAPE, MSE e MAE. Os resultados evidenciaram que o modelos ARIMA (2,0,0) superou os modelos Holt-Winters sazonal aditivo e multiplicativo, gerando os menores erros de previsão RMSE (0,8678), MAPE (12,1009) e MAE (4.7592). Concluiu-se que o ARIMA foi o modelo que gerou os menores erros de previsão entre os modelos testados na pesquisa (KITWORAWUT; RUNGREUNGANUN, 2019).

A pesquisa objetivou demonstrar os benefícios de técnicas de combinação e estratégias baseadas no algoritmo artificial de colônia de abelhas (ABC) para previsão de preços de *commodities* agrícolas. São aplicados os seguintes métodos ARIMA: regressão de vetores de suporte (SVR), rede neural recorrente (RNN), retorno fechado, rede neural de aluguel (GRU) e rede neural de memória de longo prazo (LSTM). Os autores usaram o método de combinação de previsão em todos os modelos, para comparar os desempenhos dos métodos de previsão híbrida e avaliar a eficácia da estratégia de combinação com base em uma melhoria do algoritmo ABC. As séries de preços de soja e milho são diárias de *Chicago Board of Trade's* (CBOT). O desempenho dos modelos de previsão é avaliado em termos de quatro critérios de desempenho: RMSE, MAE, MAPE e Dstat (WANG et al., 2019).

Os resultados mostraram que, dos modelos híbridos, o modelo ARIMA-EMD produz as melhores previsões em séries de milho e soja e, entre os modelos únicos, o ARIMA gerou previsões mais precisas dentre os cinco modelos. Além disso, a combinação baseada em ABC foi uma estratégia superior a outros modelos de combinação para os preços de milho e soja. Esses resultados indicam que a combinação de previsão baseada em ABC ajuda a reduzir erros para previsão e melhorar a precisão de previsão. Assim, a estratégia de combinação ABC possui uma forte robustez, garantindo uma alta precisão e melhorando efetivamente o

desempenho da previsão. Concluiu-se que a nova abordagem de combinação obteve melhorias de acurácia e robustez da previsão. O estudo empírico deve contribuir para previsão de *commodities* agrícolas, fornecendo solução para os riscos do mercado e orientações aos agentes governamentais dos setores agrícolas com relação ao planejamento orçamentário e tomada de decisão (WANG et al., 2019).

O estudo realizou a previsão de preços do milho, soja e trigo por meio da média dinâmica do modelo DMA, modelo de probabilidade mediana, modelo bayesiano de média (BMA), NAÏVE e ARIMA, com o objetivo de aplicar novos esquemas de combinação de modelos bayesianos para prever os preços spot das *commodities* de grãos. Foi um total de 17 variáveis coletadas para utilizar como amostras dos modelos de previsão. As amostras tinham 120 observações mensais e as previsões foram avaliadas fora da amostra mediante o erro quadrático médio da raiz (RMSE), erro absoluto médio (MAE) e erro médio absoluto em escala (MASE) (DRACHAL, 2019).

Os principais resultados indicaram que, no caso do preço do trigo, o modelo DMA produziu previsões significativamente mais precisas do que a maioria dos modelos alternativos. No preço do milho o modelo BMA foi significativamente mais preciso do que a maioria das abordagens concorrentes, mesmo assim DMA gerou resultados de previsão muito próxima do BMA. No caso do preço da soja o melhor modelo foi DMA, sendo que o ARIMA também gerou previsões semelhantes. O resultado foi que as previsões da combinação de modelos bayesianos foram mais precisas do que modelos econométricos convencionais, exceto para o preço da soja (DRACHAL, 2019).

Na previsão de curto prazo dos preços da soja e trigo pagos ao produtor do Paraná, o objetivo foi o de comparar o desempenho da previsão de conjuntos de regressão GBM, XGB, RF e STACK com os modelos individuais MLP, SVR e KNN, para prever preços de soja e trigo. Para medir a precisão dos modelos foram adotados os indicadores MAPE, RMSE, MAE e MSE (RIBEIRO; COELHO, 2020).

Os principais resultados foram que os modelos baseados em abordagem do conjunto apresentaram bom desempenho no curto prazo. Esses resultados demonstram a generalização da capacidade e robustez da abordagem XGB quando aplicados a previsões um mês antes do preço pago ao produtor de soja. Além disso, os modelos XGB e STACK são os melhores, com desempenhos semelhantes. Os métodos com melhor desempenho para prever preços de *commodities* estruturais, um mês à frente, são XGB para preço da soja e RF para o preço do trigo (RIBEIRO; COELHO, 2020).

Concluiu-se que o método baseado em conjuntos foi altamente preciso tanto no preço da soja quanto do trigo, gerando MAPEs menores que 1%. Os resultados foram relevantes para auxiliar na tomada de decisões dos produtores e cooperativas rurais no estado do Paraná, permitindo a estes escolherem o melhor momento para comercializar os produtos agrícolas ao atacadista, bem como desenhar estratégias de mercado que maximizem seus lucros. Segundo os autores, a limitação da pesquisa foi desconsiderar fatores como clima e institucionais do agronegócio, que também são determinantes na volatilidade dos preços agrícolas (RIBEIRO; COELHO, 2020).

2.2.2 Discussão

2.2.3 Observações gerais

Abordagens estatísticas tradicionais, como o modelo de média móvel integrada autorregressiva ARIMA, têm sido utilizadas para prever o preço de *commodities* agrícolas na última década Cas (2018); Castro et al., (2016); Ceretta, Righi e Schlender (2010); Feliciani, Souza e Souza (2015); Lima et al., (2010); Marchezan e Souza (2010); Nomelini et al., (2017); Pinheiro e Senna (2016); Santos Felipe et al., (2012); Souza et al., (2018); Tibulo e Carli (2014).

Os algoritmos de inteligência artificial (IA) estão ganhando cada vez mais atenção na previsão de preços, visando superar as limitações dos modelos estatísticos tradicionais. Por exemplo, eles não conseguem capturar facilmente padrões não lineares (FAN; WANG; LI, 2016; WANG et al., 2017). Nos últimos anos, abordagens de inteligência artificial, como as redes neurais artificiais, têm gerado previsões mais precisas para os preços de grãos (Bressan ,2004); Ceretta, Righi e Schlender (2010); Ferreira et al., (2011); Jha e Sinha (2012, 2013); Lima et al., (2010); Pinheiro e Senna (2016, 2017a); Puchalsky et al., (2018); Ribeiro, Sosnoski e Oliveira (2010).

No Brasil, as pesquisas mais relevantes utilizando modelos de redes neurais aplicadas no preço da soja são de Bressan (2004); Ceretta, Righi e Schlender (2010); Ferreira et al., (2011); Lima et al., (2010); Pinheiro e Senna (2016, 2017a, 2017b); Puchalsky et al., (2018); Ribeiro, Sosnoski e Oliveira (2010).

Considerando o portfólio de artigos avaliados, em termos de aplicação de modelos o método ARIMA foi o mais utilizado, mostrando que tem gerado previsões úteis e se

consolidado ao longo dos anos. Desta forma, é o modelo de referência para pesquisas que tenham por finalidade comparar métodos de previsão do preço de *commodities* agrícolas.

O modelo *seasonal autoregressive integrated moving average* (SARIMA) foi proposto em alguns estudos como uma alternativa de modelo de previsão Feliciani, Souza e Souza (2015); Marchezan e Souza (2010); Pinheiro e Senna (2017a); e Pinheiro, Senna e Matsumoto (2016).

Em concorrência com os modelos estatísticos as abordagens de inteligência artificial (IA), tais como as redes neurais artificiais de diversas tipologias, tem sido amplamente utilizadas e gerado previsões precisas de preços agrícolas.

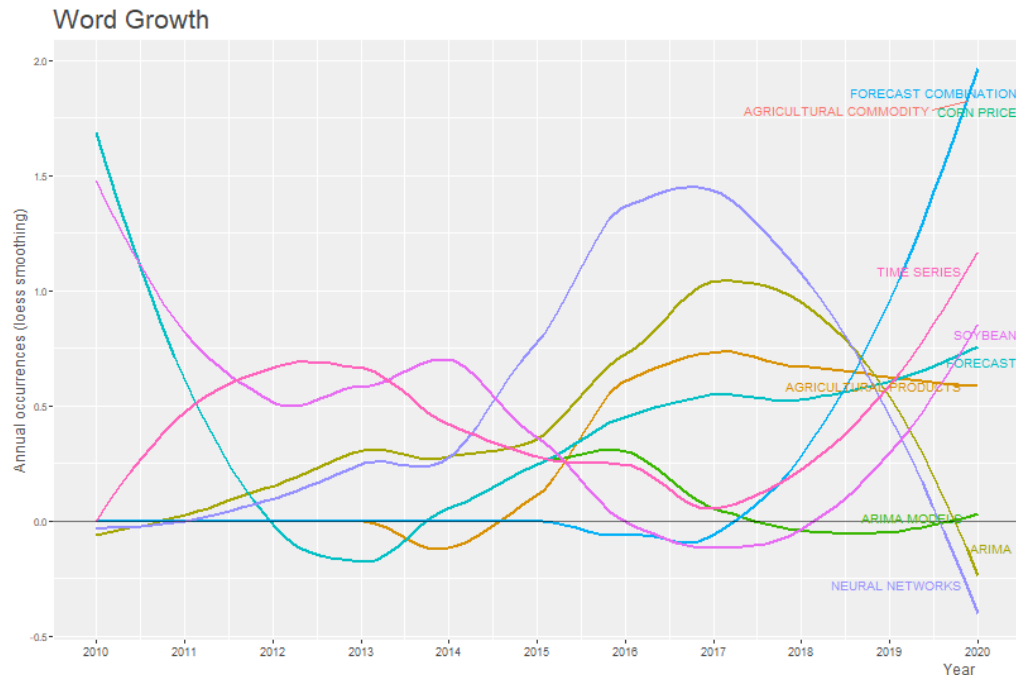
2.2.4 Principais tendências do campo da previsão de preços das *commodities* agrícolas

Na maior parte dos estudos, as redes neurais produzem previsões superiores em comparação com os modelos ARIMA, SARIMA, ARIMA-GARCH, ARCH, ARIMA-GARCH e *Holt-Winters*. Foram identificados muitos artigos que utilizam as redes neurais de diversas tipologias para prever o preço. No entanto, foi possível observar que pesquisas comparando o desempenho dos modelos de inteligência artificial (IA), tais como redes neurais e máquina de aprendizado, são escassas.

Uma abordagem de previsão que se consolidou para prever o preço de *commodity* são os métodos híbridos de previsão. Trata-se da combinação de modelos individuais para melhorar a acuracidade das previsões (PETROPOULOS et al., 2020; WANG et al., 2019; ZOU et al., 2007).

A Figura 8 apresenta as palavras-chave mais frequentes entre 2010 e outubro de 2020. O gráfico mostra em 2017 a predominância da previsão com as redes neurais e ARIMA. A partir de 2019, a previsão de preços das *commodities* por meio dos modelos redes neurais e ARIMA começam diminuir.

Figura 8 - Palavras-chave mais frequentes entre 2010 e outubro de 2020



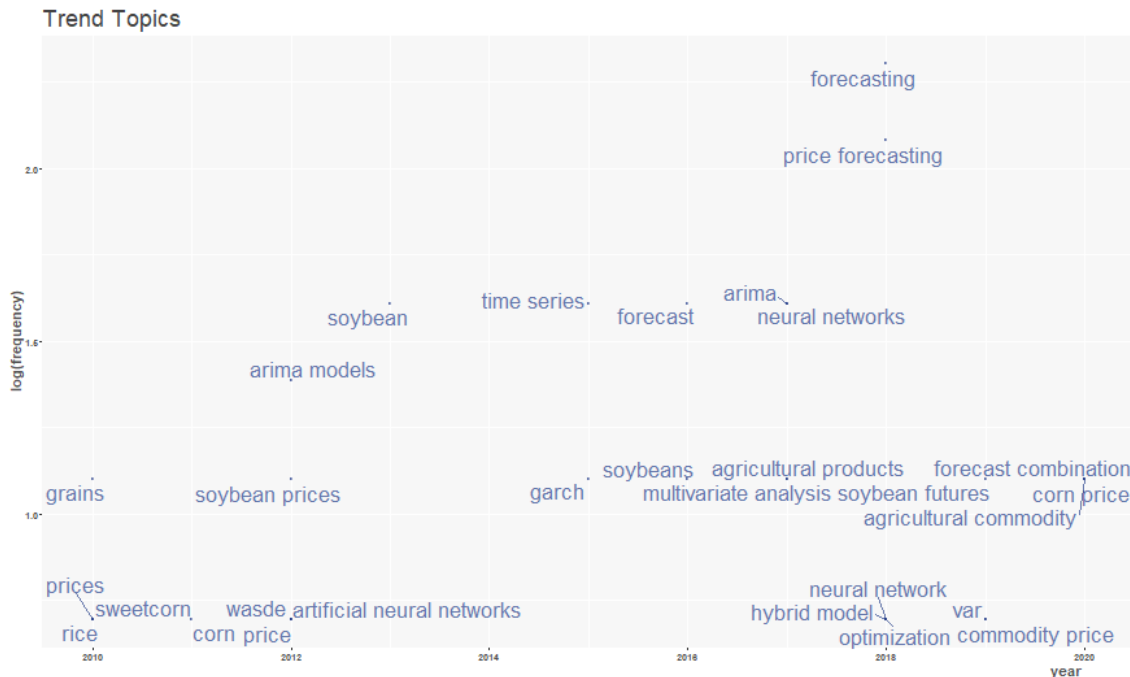
Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

Em contrapartida, os métodos híbridos de previsão desde 2018 parecem ser os modelos mais promissores para previsão de preços de *commodity* agrícola. Avaliando o ano de 2020, os modelos híbridos se destacam como método mais utilizado, enquanto no mesmo ano o uso dos modelos ARIMA e redes neurais parecem estar em declínio.

É mostrado na figura 9 um gráfico com os tópicos de tendência para o portfólio de artigos avaliados no período de 2010 a outubro de 2020. Identificou-se que os artigos analisados se enquadram em três categorias principais. A primeira abordagem são as previsões dos preços agrícolas com base em modelos univariados, principalmente os modelos econométricos; o modelo de referência ARIMA e modelos de estrutura multivariada, ou seja, utilizou mais de uma variável para realizar as previsões.

A segunda abordagem amplamente utilizada para o problema de previsão de preços são os modelos de inteligência artificial (IA) e computacional, principalmente as redes neurais. A terceira abordagem e a mais relevante são modelos híbridos, tanto econométricos quanto combinações de modelos de inteligência artificial (IA).

Figura 9 - Tópicos de tendência das pesquisas sobre previsão de preços das *commodities* agrícolas no período de 2010 a outubro de 2020



Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

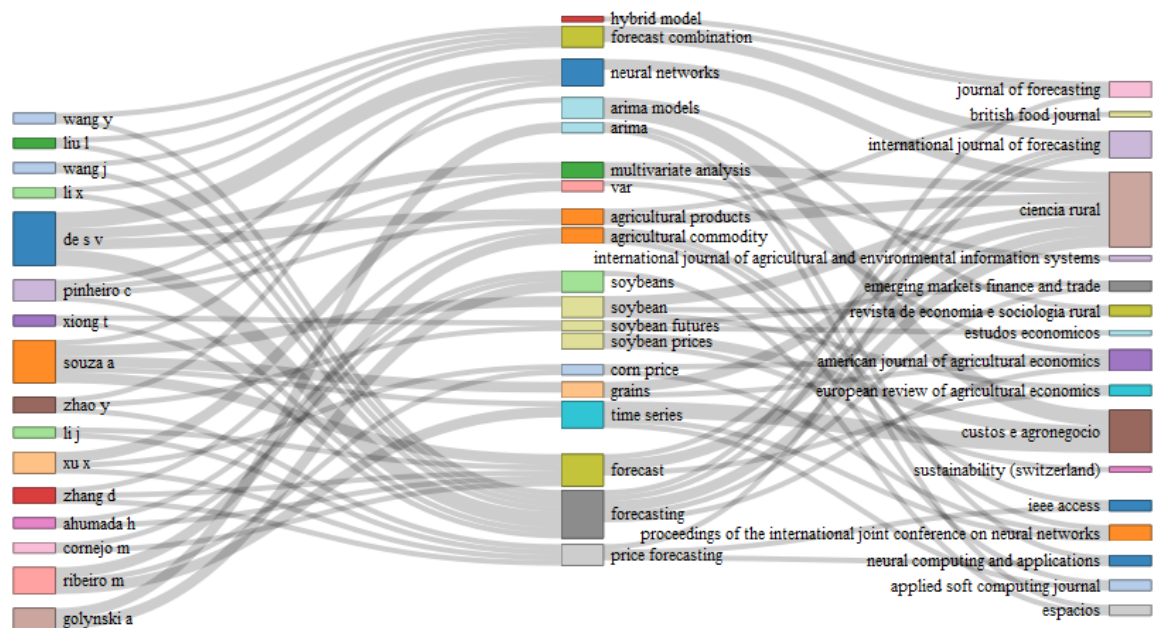
Para a tarefa de modelagem dos preços agrícolas, os métodos híbridos são as abordagens mais promissoras, relacionadas ao desempenho de previsão dos modelos. A tendência parece mostrar que os estudos futuros serão previsões de preço de *commodity* com os modelos combinados. Petropoulos et al. (2020) e Wang et al., (2019) afirmam que a combinação de previsões de diferentes métodos tem se mostrado um grande sucesso.

Os pesquisadores estão direcionados a modelar os preços criando novos algoritmos de modelos híbridos, uma vez que é o estado da arte da problemática da previsão de preços da soja, e dado que estes modelos são os mais precisos em comparação às abordagens concorrentes, amplamente adotados em 2020.

Para identificar os principais temas abordados foi gerado um levantamento das principais palavras que constituem o título e o conteúdo dos resumos dos artigos (Figura 10), conforme Camargo e Justo (2013) a nuvem de palavras é uma técnica básica de visualização de texto que organiza e agrupa as palavras em função da sua frequência e possibilita rápida identificação das palavras-chave de um corpus textual.

O banco de dados dos resumos e títulos foi organizado e preparado num arquivo txt, para gerar a nuvem de palavras, por meio da linguagem de programação R e mediante ao pacote *wordcloud2* de Lang e Chien (2018). Assim, o *script* foi rodado no R *studio*.

Para a construção da nuvem de palavras adotaram-se os termos em inglês. Observa-se na Figura 10 que as palavras mais frequentes são as mesmas utilizadas anteriormente para buscar os artigos nas bases de dados: *SCOPUS*, *Web of Science* e *Google Scholar*.

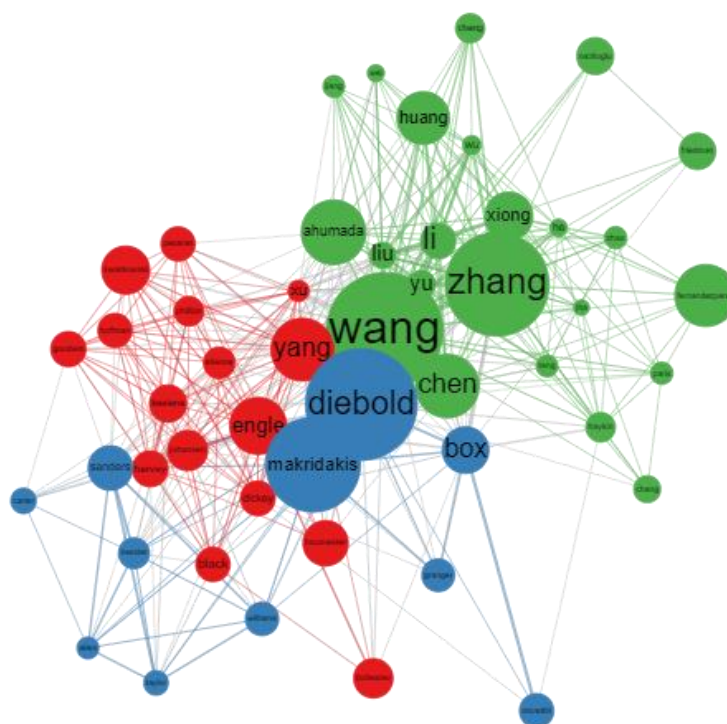


Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

Os periódicos *journal of forecasting* e *international journal of forecasting* são os meios de divulgação científicos que mais tem publicado as pesquisas com abordagens híbridas e combinação de previsão. O gráfico de três campos também revela que os termos “*hybrid model*” e “*forecast combination*” está relacionado com os autores Wang, Liu, Wang e Li.

A rede de co-citação entre autores no período de 1981 a outubro de 2020, é apresentada na Figura 12. Esta rede mostra as comunidades de autores principais em grupos de cores diferentes que podem ser úteis na identificação da co-citação. O tamanho do círculo indica a força da co-citação do autor. As cores e linhas indicam os respectivos *clusters* (Figura 12).

Figura 12 - Rede de co-citação de autores entre 1981 até outubro de 2020



Fonte: Elaborado pelo autor, por meio do *biblioshiny* com base no pacote *bibliometrix*.

A rede de co-citação (Figura 12) foi obtida por meio do *bibliometrix*, pacote em R (ARIA; CUCCURULLO, 2017). Seu *layout* foi gerado usando o algoritmo de *Leading Eigenvalues*. Os autores mais citados são representados com círculos maiores, pois sua contagem de citações é alta. Os círculos menores são autores com baixa contagem de citações.

A rede de co-citação está formada em três grandes *clusters* que abrangem as cores. Os principais autores que juntos compõem o *clusters* em verde com maior influência na rede são Wang, Zhang, Chen, Liu e Ahumada. Em relação aos dois grandes grupos (azul e vermelho) são formados principalmente por Diebold, Box-Jenkins, Makridakis, Yang e Engle. Estes dois *clusters* estão mais relacionados pela proximidade e nós da rede.

As principais constatações do estudo bibliométrico foram: (i) o ARIMA e redes neurais são os modelos mais utilizados na previsão de preços de *commodities* agrícolas. No entanto, o modelo ARIMA não tem gerado previsões superiores em comparação aos modelos de inteligência artificial (IA) e modelos híbridos.

Esta primeira tendência dos estudos sobre a temática, desdobrada pela análise bibliométrica, foi consistente com o que afirmam Mitra e Paul (2017) e Drachal (2019) para a pesquisa de previsão de preços agrícolas, o ARIMA tem sido o modelo de referência, sendo um método amplamente utilizado nas últimas décadas para esse fim.

(II) as redes neurais são mais precisas para prever preços de *commodities* agrícolas do que os modelos econométricos; (iii) os modelos híbridos de IA geram previsões com melhores níveis de acuracidade em comparação aos métodos estatísticos tradicionais ARIMA, modelos individuais e redes neurais em que o desempenho de previsão dos modelos híbridos são melhores do que os dos modelos únicos;

(IV) nenhum modelo de previsão é superior em termos de acuracidade em todos os horizontes de previsão, portanto, de modo geral, os modelos híbridos de IA são mais precisos, mas não é verdade para todos os intervalos de previsão: curto, médio e longo prazos; e (v) nenhuma pesquisa do portfólio de artigos avaliados avança na relação entre os constructos modelos de previsão e tomada de decisão, ou seja não especificam no que os modelos de previsão apoiam e auxiliam os tomadores de decisão.

2.3 Considerações e apontamentos da revisão

A utilização de modelos de previsão se tornou uma atividade importante de pesquisa para o mercado agrícola. As características de volatilidade dos preços de soja implicam desafios para o processo de tomada de decisão dos participantes do mercado, sejam planejadores agrícolas, produtor rural, especuladores ou interessados nos preços, o que torna os modelos de previsão métodos indispensáveis para auxiliar e apoiar as decisões. As lacunas científicas foram identificadas por meio da análise bibliométrica. Foram selecionadas publicações nas bases de dados *Google Scholar*, *Web of Science* e *SCOPUS*, e avaliada uma amostra de 59 artigos.

Os principais achados da análise bibliométrica sobre a temática foram: (i) são os principais periódicos *Custos e Agronegócio online*, *Ciência Rural*, *International Journal of Forecasting* e *Sustainability*; (ii) os pesquisadores destaque são Pinheiro; Senna; Jha; Sinha; Ahumada e Cornejo; e (iii) o Brasil se destaca com o maior número de publicações e as universidades *China University of Geosciences*, *Huazhong Agricultural University*, Universidade do Estado da Bahia (UNEB) e Universidade Federal do Rio Grande do Norte (UFRN).

Os estudos dedicados à revisão dos métodos de previsão de preços das *commodities* agrícolas ainda são poucos. A revisão permitiu identificar como oportunidade de pesquisa comparar a acuracidade dos modelos híbridos de inteligência artificial (IA), redes neurais e *machine learning*.

Chegou-se à conclusão que os resultados da pesquisa podem contribuir com a discussão sobre modelagem e previsão de preços de *commodities* agrícolas. Os modelos abordados neste estudo bibliométrico podem fornecer referência para os econométricos do campo da previsão de preços de produtos agrícolas, e a pesquisa aponta as tendências sobre a temáticas propostas no final deste estudo. Portanto, forneceu direções de pesquisa para pesquisadores.

Para estudos posteriores sugere-se as lacunas científicas constatadas na revisão bibliométrica. A contribuição do trabalho de pesquisa foi o estudo bibliométrico, que contribui para discussões de estudos empíricos sobre previsão de preços de *commodities* agrícolas. A principal limitação da revisão foi não analisar o conteúdo dos artigos selecionados. No entanto, a descrição e abordagem dos principais estudos foram realizadas.

CAPÍTULO III

Neste capítulo são abordados os conceitos fundamentais para se entender a lei do preço único e os modelos de previsão. Apresenta-se nas seções 3.2.1 a comercialização da soja; as seções 3.3.1. Aborda os modelos de previsão e tomada de decisão; já a 3.4 explica as redes neurais, *perceptron* multicamada; a 3.5, o algoritmo de aprendizado máquina de aprendizagem extrema e a 3.6, as medidas de erro. Este capítulo está dividido em oito partes principais.

3.1 Revisão de literatura

3.1.1 Lei do Preço Único (LPU)

A teoria do preço único foi desenvolvida por Mundlak e Larson (1992). É um modelo teórico para entender a formação de preços de produtos homogêneos em diversos mercados (FILHO et al., 2012). A teoria propõe que, em situações iguais de concorrência e ausência de custos de transporte, os produtos com alto grau de padronização devem ser comercializados em preços parecidos em diferentes regiões, isto quando os preços dos produtos são cotados na mesma moeda (FILHO et al., 2012; SOUSA; CAMPOS, 2009).

A LPU está diretamente relacionada ao conceito de arbitragem, onde a longo prazo os preços tendem a convergirem quando convertidos na mesma moeda corrente (BARBOSA; MARGARIDO; JUNIOR, 2002). Mediante a arbitragem os mercados que tenham os maiores preços tendem à redução, por causa do aumento da oferta (BARBOSA; MARGARIDO; JUNIOR, 2002). Em contrapartida, os mercados com preços baixos serão levados ao aumento de preços, em razão do aumento na demanda. Esse processo ocorre até que os preços tornem-se iguais em diferentes países (BARBOSA; MARGARIDO; JUNIOR, 2002).

A teoria apresenta como as variações dos preços internacionais afetam as variações dos preços domésticos (BARBOSA; MARGARIDO; JUNIOR, 2002; GIEMBINSKY, 2006; MARGARIDO; TUROLLA; FERNANDES, 2001). Segundo Krugman e Obstfeld (2006, p. 398),

A lei do preço único estabelece que, em mercados competitivos em que os custos de transporte não são levados em consideração e não existem barreiras oficiais ao

comércio, como tarifas, os produtos idênticos vendidos em diferentes os países devem ter o mesmo preço, quando expresso em termos da mesma moeda.

Conforme Krugman e Obstfeld (2006), o preço doméstico do produto pode ser definido como função de preços domésticos, internacionais e a taxa de câmbio nominal. Assim, a lei do preço único é definida da seguinte maneira:

$$P_{it} = \alpha + \beta P_{jt} + u_t \quad (1)$$

Onde P_{it} e P_{jt} são preços de uma *commodity* nos mercados de duas regiões i e j , para um dado período de tempo t ; α é uma constante; u_t é o termo aleatório ou termo de erro; e o coeficiente β é a elasticidade de transmissão de preço entre essas regiões observadas, de maneira que as variações de preço na região j são transmitidas para a região i se o valor de β for igual a um. Variações de preço na região j não afetam a região i se o valor de β for igual à zero. Espera-se que o valor de β seja encontrado entre zero e um, dependendo da política comercial utilizada pelo país e o grau de integração do produto com o mercado (BARBOSA; MARGARIDO; JUNIOR, 2002).

Os estudos empíricos que tratou da transmissão de preços de soja no Brasil foram os trabalhos de Giombinsky (2006) e Pino e Rocha (1994). As pesquisas constataram uma alta relação entre os preços na Bolsa de Chicago (CBOT) e os preços recebidos pelo produtor no mercado nacional. Os autores concluíram que o preço de grão no Brasil é afetado principalmente pelo preço de grão da CBOT.

Alguns estudos empíricos sobre a transmissão de preços de soja no Brasil comprovaram os pressupostos teóricos da lei do preço único. A pesquisa Libera e Waquil (2009) teve o objetivo de verificar se havia integração e quais são as características da transmissão de preços entre o mercado físico para o produtor no Brasil e o mercado futuro para a *commodity* soja. Os resultados indicaram que havia transmissão de preços entre mercados físico e futuro. Os autores concluíram que 70% das variações no preço futuro de soja eram transmitidas para o preço do mercado físico de soja (LIBERA; WAQUIL, 2009).

A pesquisa empírica de Tonin e Barczsz (2008), também constatou a validade da teoria do preço único para a formação dos valores da soja no Brasil. Esse estudo comprovou a existência de transmissão de preços entre o Brasil e os preços internacionais da CBOT.

3.1.2 Comercialização da soja

O Complexo da soja é caracterizado por três produtos principais, sendo eles: soja *in natura*, farelo de soja e o óleo bruto (LIBERA; WAQUIL, 2009). A soja é o principal grão oleaginoso cultivado e uma das principais *commodities* comercializadas no mundo (AGNOL et al., 2007; DUARTE, 2007; MARGARIDO; TUROLLA; FERNANDES, 2001; SOUSA; CAMPOS, 2009).

Nos últimos anos, a soja estabeleceu-se como a principal *commodity* agrícola para exportação no Brasil (CERETTA; RIGHI; SCHLENDER, 2010). O valor da produção agrícola do país foi de R\$ 319,6 bilhões em 2017, sendo a soja responsável por 35,1% desse total. De acordo com a (CNA, 2020) o PIB do agronegócio brasileiro, cresceu 3,81% em 2019, frente a 2018. Com esse desempenho, em 2019 o PIB do agronegócio representou 21,4% do PIB brasileiro total.

Miquelluti (2015, p. 17) afirma que “políticas públicas e decisões que irão afetar o setor como um todo são tomadas baseadas nas produções das principais culturas, soja, milho, arroz, trigo, café, laranja, para citar algumas”. As safras da agricultura empreendem muita influência no comportamento do mercado agrícola (MIQUELLUTI, 2015).

O problema de riscos de mercado está associado à instabilidade dos preços das *commodities*. Como os preços são dependentes de forças da oferta e procura, existem fatores que influenciam a oferta e demandas agrícolas. Esses fatores são cercados de risco e incerteza (MENDES; JUNIOR, 2007). Neste contexto, o risco é maior no setor agrícola em virtude de efeitos da oferta e demanda e os fatores climáticos que afetam a produção.

Segundo Junior et al., (2018) após a produção e colheita da soja os grãos são enviados para as cooperativas, cerealistas, indústrias de esmagamento ou armazenamento. Na propriedade rural essa prática de armazenar é comum entre os produtores e serve para comercializar a soja no mercado em momento que os preços finais estão em alta.

Há um déficit da capacidade de armazenagem dos produtores de soja, por isso são forçados a vender a produção depois da colheita, obtendo preços mais baixos e custos de transporte mais altos, por não possuir armazéns (JUNIOR et al., 2018).

Para Agnol et al. (2007) a comercialização da soja no Brasil era dominada pelas quatro maiores empresas privadas e processadoras de soja: ADM, Bunge, Cargill e Dreifuss que comproum, juntas, aproximadamente 70% da produção. As cooperativas correspondem

por, aproximadamente, 30% da compra de soja. Os principais meios de comercialização da soja são por meio da Bolsa de Mercadorias e Futuros (BM&F) e da *Chicago Board of Trade* (FORIGO; ZAGO; ABREU, 2010).

Segundo Junior et al. (2018) os contratos de soja são negociados em dólar pelos operadores. Por meio das bolsas também são utilizadas ferramentas e software (SAP, *Oracle*, BI – *Business Intelligence*) para acompanhar as cotações de compra e venda. Assim, “Isso possibilita o acompanhamento de várias cotações de diversas *commodities* ao mesmo tempo, bem como a construção de gráficos de acompanhamento e evoluções históricas” (JUNIOR et al., 2018, p. 75).

Neste contexto, a competitividade da sojicultura exige dos gestores rurais uma atenção voltada ao preço do produto. Devido às variações de preço da *commodity* no mercado, isso tem gerado um efeito negativo nas estratégias empresariais das empresas do setor (JUNIOR et al., 2018). Os produtores rurais passaram a utilizar meios de comercialização mais flexíveis e menos arriscados. Essas formas de comercializar necessitam de acesso às informações de mercado para auxiliar as negociações (JUNIOR et al., 2018).

Além disso, Junior et al. (2018, p. 78) afirmam que prever o preço da soja no curto, médio e longo prazos é uma ferramenta de diferenciação competitiva:

A forte disputa entre as empresas do agronegócio fez com que as mesmas se voltassem à busca de estratégias que as diferenciasssem e as tornassem competitivas. Deste modo, conhecer os fatores que influenciam o mercado e a utilização de estratégias que consistem em estimar preços futuros de curto, médio e longo prazo, por meio dos canais de comercialização disponíveis, tornaram-se mecanismos fundamentais de diferenciação frente à concorrência.

De acordo com Junior et al. (2018), o preço da soja é o principal fator determinante no momento da compra ou venda. O autor ainda ressalta que, por ser uma *commodity* e o seu preço ser regulado por muitos fatores externos como bolsa de valores, políticas externas, câmbio, clima, tecnologia e outros, este cenário gera forte concorrência, tornando-se a diferença de centavos no preço da saca de soja, o que acaba se tornando significativo, dependendo do volume negociado.

Diante das constatações é importante conhecer cotações futuras de preços, pois isso pode afetar o planejamento econômico de pequenos, médios e grandes produtores (RIBEIRO; COELHO, 2020). A alta volatilidade dos preços mundiais da soja causou incerteza e vulnerabilidade, principalmente nos países em desenvolvimento (DEGIANNAKIS et al., 2020; DENG; SÜN, 2019).

As previsões dos preços agrícolas consistem em uma das principais alternativas de apoio à execução do planejamento e avaliação das atividades da agricultura, em razão de ser uma variável chave para decisões de compra e venda das *commodities* agrícolas (PINHEIRO; SENNA, 2017a).

3.1.3 Modelos de previsão e tomada de decisão

De acordo com Balestrin (2002) os estudos de Simon percorreram por muitas áreas do conhecimento, sendo voltados à compreensão das soluções de problemas nas organizações e o processo decisório. A teoria de Simon criticou alguns dos principais fundamentos da economia neoclássica e da administração clássica (BALESTRIN, 2002).

O gestor tem que tomar as decisões da melhor forma possível, possibilitando que todos na organização tomem decisões mais assertivas (BALESTRIN, 2002). Ainda de acordo com Balestrin (2002, pág. 4):

“processo decisório envolve a seleção consciente ou inconsciente de determinadas ações entre aquelas que são fisicamente possíveis para o agente e para aquelas pessoas sobre as quais ele exerce influência e autoridade.”

Segundo Simon (1987) os métodos de tomada de decisão tem sido amplamente melhorados pela utilização de técnicas de pesquisa operacional, ciência de gerenciamento e tecnologias de sistemas especialistas. Neste sentido, (Bressan, 2004, p. 3) afirma que:

Diversas técnicas de previsão auxiliam a tomada de decisões por parte dos agentes envolvidos em atividades que necessitam de planejamento, avaliação de políticas e redução da incerteza. Um dos objetivos das previsões econômicas é a redução da incerteza, que é de especial importância dentro do setor agropecuário, constantemente sujeito a distúrbios irregulares.

A pesquisa de Ferreira et al (2011) menciona os modelos de previsão como um diferencial competitivo para as empresas do agronegócio. Os autores abordam as previsões de redes neurais *perception* multicamadas nos preços das *commodities* agrícolas soja, boi gordo, milho e trigo com um enfoque de previsões estratégicas para tomada de decisão, ou seja, um modelo de previsão pode ser um método estratégico (FERREIRA et al., 2011).

A teoria de base do estudo do Ferreira et al. (2011) foi a teoria da decisão. A teoria é relevante para entender no que a modelagem de preço de *commodity* agrícola auxilia e apoia o processo de tomada de decisão. Analisar os preços agrícolas, como a variável determinante para as transações, tem sua relevância principalmente para os interessados no agronegócio,

sendo eles vendedores, compradores e participantes do mercado que buscam proteção contra as variações de preços (PINHEIRO; SENNA, 2016, 2017a).

As fontes de dados secundários disponíveis publicamente deu origem a uma riqueza de novas necessidades e oportunidades de previsões. A quantidade cada vez maior de dados da agricultura possibilita aplicar modelos de previsão amplamente utilizados para apoiar as decisões de planejadores agrícolas, a fim de reduzir risco dos níveis de produção e atender as demandas.

As previsões de preços dos produtos da agricultura são úteis para agricultores, mercado, indústrias do agronegócio, comunidades agrícolas e governos (ALLEN, 1994; DAS et al., 2019; JHA; SINHA, 2012). Ainda sobre o aspecto dos modelos de previsão, melhorar a acurácia das predições dos preços agrícolas é relevante para os investidores, os próprios produtores agrícolas e formuladores de políticas (FANG et al., 2020).

O produtor de soja no Brasil é um tomador de preço, desta forma a principal estratégia competitiva do segmento ocorrerá em função do preço da *commodity* (CAMPEÃO; SANCHES; MACIEL, 2020; GIEMBINSKY, 2006). Neste mercado de concorrência perfeita, gerar informações para o produtor, mediante a previsão do preço futuro da oleaginosa com um erro de predição baixo, pode ser um método relevante para o processo decisório do produtor de soja em relação à comercialização da *commodity* agrícola.

Os estudos de Bressan (2004); Ceretta; Righi e Schlender (2010); Feliciani; Souza e Souza (2015); Mendes e Junior (2007) e Pinheiro e Senna (2017b) apontam que a volatilidade dos preços da soja é um dos problemas para o produtor da *commodity* na decisão do melhor momento de comercialização do produto. Neste contexto, diminuir as incertezas com relação ao preço futuro da soja pode apoiar o processo decisório.

Nas transações abrangendo a comercialização da soja do produtor rural, o risco inerente às oscilações de preços da *commodity* é um item fundamental. Desta forma, previsões precisas tornam-se um importante método na tomada de decisão do produtor de soja de modo a avaliar o potencial dos modelos como instrumento de suporte ao processo decisório.

Compreender e antecipar a dinâmica do ambiente global do agronegócio será cada vez mais crítico para a gestão e o desempenho das propriedades (CARVALHO; PRÉVOT; MACHADO, 2014, p. 507). Nesta conjuntura, os modelos de previsão contribuem para diminuir as incertezas do processo de tomada de decisão do produtor rural (SABU; KUMAR, 2020).

Os modelos de previsão de séries temporais auxiliam no processo de tomada de decisão econômica e na redução das incertezas (BRESSAN, 2004). As previsões de série de

tempo tem despertado interesse dos pesquisadores pela sua relevância nas atividades de tomada de decisão em muitas áreas do conhecimento (SALLES et al., 2017).

Estes modelos univariados possuem uma característica básica. O modelo faz a previsão com base na variável em si, isto significa utilizar a série temporal dentro da amostra para fazer a previsão num período fora da amostra. Um dos principais autores no campo da previsão de séries temporais afirma que os modelos univariados são modelos de previsão que dependem apenas do comportamento dos dados passados para prever os valores futuros da série temporal (CHATFIELD, 1978; PETROPOULOS et al., 2020).

As previsões servem de apoio à tomada de decisões no que se refere ao planejamento de produção nos níveis operacional, tático e a alta administração (SAGAERT et al., 2017). Ribeiro e Coelho (2020) argumentam que as predições de séries de preços de *commodities* agrícolas exercem papel relevante na conjuntura econômica, assim como a acuracidade das predições é relevante tanto quanto os preços realizados, sendo que dependente do modelo de previsão adotado. A importância dos modelos passa por várias áreas do conhecimento. As previsões podem ser usadas para apoiar a revisão ou implementação de políticas públicas, desenvolvimento de planejamento estratégico nas empresas e decisões no mundo corporativo.

Desta forma, Zhang et al. (2020, p. 28197) também afirmam a importância de previsões precisas de preços agrícolas e como é relevante para as tomadas de decisões de autoridades agrícolas:

Para a agricultura, grande flutuação de preços aumentará a incerteza na produção, aumentando, assim, o número de riscos que devem ser gerenciados. Consequentemente, uma previsão precisa do preço de *commodities* agrícolas é vital para as autoridades agrícolas tomar decisões científicas e garantir um ambiente favorável para operação da economia social.

Neste contexto, o método ARIMA se consolidou para prever preços de *commodities* agrícolas, auxiliando nos processos de tomada de decisão dos participantes do mercado agrícola. Além disso, foi considerado um modelo referência para rotina de previsão de séries de preços agrícolas, em outras palavras, o modelo ARIMA é um modelo de *benchmark* (DRACHAL, 2019).

Um exemplo da utilização do modelo pode ser visto na pesquisa de Santana, Russo e Figueroa (2012), onde aplicaram os modelos ARIMA nos dados da produção de cana de açúcar do Instituto de Pesquisa Econômica Aplicada (IPEA). Foi realizada uma estatística descritiva e teste de normalidade na série temporal, constatando-se que a série não possuía distribuição normal.

Verificou-se que o modelo ARIMA (2,1,0) é o melhor modelo para realizar previsões de longo prazo da produção de cana de açúcar, por meio do critério de avaliação MAPE. O modelo ARIMA (2,1,0) gerou um percentual de (10,73%) comparado com o ARIMA (1,1,0), com um MAPE igual a 18,93%.

3.2 Modelos autorregressivos integrados de médias móveis com variável exógena (ARIMAX)

O modelo univariado utilizado geralmente para prever uma série de tempo, é o método ARIMA (CHERDCHOONGAM; RUNGREUNGANUN, 2016; GUTIERREZ, 2003; PETROPOULOS et al., 2020). De acordo com Arêdes e Pereira (2008), Bundchen e Werner (2016); e Adli (2020) o modelo ARIMA (autorregressivo integrado de médias móveis) foi proposto por Box-Jenkins em 1976. Este método usa uma combinação de três componentes que são autorregressivo (AR), filtro de integração (I) e médias móveis (MA). Segundo Constantino, Mendes e Moreira (2019, p. 251),

Os modelos autorregressivos desenvolvem pressupostos e procedimentos para construção e estimativa. A maioria dos procedimentos de análise de séries temporais parte do pressuposto que estas sejam estacionárias, será preciso transformá-las caso ainda não sejam.

O método de previsão ARIMAX foi desenvolvido por Box e Tiao (1975), a qual é uma extensão do modelo ARIMA. A metodologia ARIMAX é considerada multivariada, pois acrescenta as variáveis exógenas (CAMELO et al., 2017, 2018a, 2018b).

O modelo ARIMAX é a aplicação do método ARIMA, adicionando a variável independente (X) (ADLI, 2020; CHERDCHOONGAM; RUNGREUNGANUN, 2016). A diferença entre o método ARIMA e o modelo ARIMAX é que o segundo possui a variável explicativa (X), sendo um modelo multivariado que insere um componente linear como variáveis exógenas (CAMELO et al., 2017, 2018a; CHERDCHOONGAM; RUNGREUNGANUN, 2016).

O modelo ARIMAX utilizado nesta pesquisa é uma versão automática chamada na função em R de *auto.arima*, onde o valor de d é escolhido com base em repetidos testes de raiz unitária KPSS do Kwiatkowski et al., (1992). Em seguida, são selecionados os valores de p e q, minimizando o Critério de Informação Akaike (AIC) para determinar os valores,

$$\phi(B)(1 - B^d)y_t = c + \theta(B)\varepsilon_t \quad (2)$$

Onde, ε_t é um processo de ruído branco com zero médio e variância σ^2 ; B é o operador de mudança de turno e $\phi(Z)$ e $\theta(Z)$ são polinômios da ordem p e q, respectivamente. Para garantir causalidade e invertibilidade, presume-se que $\phi(Z)$ e $\theta(Z)$ não têm raízes para $|Z| < 1$ (BROCKWELL; DAVIS, 1991). Se $c \neq 0$, existe um polinômio implícito de ordem d na função de previsão (HYNDMAN; KHANDAKAR, 2008). O modelo de previsão automático ARIMA seleciona a ordem de p, q e d por meio do critério de Informação conforme o AIC, em que:

$$AIC = -2 \log(L) + 2(p + q + P + Q + k) \quad (3)$$

$k = 1$ se $c \neq 0$ e 0 de outra forma, e L é a probabilidade máxima do modelo fixado para dados diferidos $(1 - B^m)^D (1 - B)^d y_t$. A probabilidade de o modelo completo para y_t não é realmente definida e, portanto, o valor da AIC para diferentes níveis de difusão não é comparável (HYNDMAN; KHANDAKAR, 2008).

O algoritmo Hyndman-Khandakar seleciona o modelo que apresenta o menor valor AIC de: ARIMA (2, d, 2), ARIMA (0, d, 0), ARIMA (1, d, 0) e ARIMA (0, d, 1) dentre os quatros modelos (HASSANI et al., 2017). Para ler os detalhes de funcionamento da modelagem ARIMA basta ver a seção 8.7 em (HYNDMAN; ATHANASOPOULOS, 2018a).

Utilizar o modelo ARIMA junto com variáveis explicativas (ARIMAX) pertence à classe das regressões, que engloba uma grande variedade de modelos, incluindo modelos clássicos de regressão múltipla, onde as variáveis de entrada têm um efeito instantâneo na variável de saída (ADLI, 2020; CAMELO et al., 2018a; CHERDCHOONGAM; RUNGREUNGANUN, 2016). Para representar o modelo ARIMAX, utiliza-se a equação (4).

$$\phi(B)y_t = c + \beta x_t + \theta(B)\varepsilon_t \quad (4)$$

Onde, x_t é uma variável explicativa no tempo t. É possível ter mais de uma variável explicativa junto com o efeito instantâneo. O efeito das variáveis explicativas na variável dependente pode ser atrasado em algumas situações (ADLI, 2020; CAMELO et al., 2017, 2018a, 2018b; CHERDCHOONGAM; RUNGREUNGANUN, 2016).

A escolha das variáveis exógenas (X) utilizadas pelo modelo ARIMAX são os preços do FMI de contrato futuro (primeiro contrato a termo) da Bolsa de Chicago (FMI, 2020). Representa as variáveis exógenas, que foram incluídas no modelo de duas formas: 96

observações para predição de preço da soja do Paraná e 60 períodos para prever cada série de preços do Rio Grande do Sul e Mato Grosso.

Esta variável foi escolhida, pois a teoria do preço único, explicada na seção 3.1.1, considera que a formação de preço doméstico da *commodity* agrícola pode ser definida como função de preços domésticos, internacionais e a taxa de câmbio nominal.

A variável da Bolsa de Chicago (CBOT) foi escolhida para verificar a validade teórica dos pressupostos teóricos da lei do preço único, assim testar empiricamente com o método ARIMAX se o modelo teórico funciona para os preços da soja.

Considerando que existem evidências de estudos empíricos, Giembinsky (2006); Libera e Waquil (2009); Pino e Rocha (1994) e Tonin e Barczysz (2008), que trataram da transmissão de preços de soja no Brasil, nas quais se constata uma alta relação entre os preços na CBOT e os preços recebidos pelo produtor no mercado nacional. Estes estudos concluíram que o preço da soja no Brasil é afetado principalmente pelo preço do grão da CBOT (GIEMBINSKY, 2006; LIBERA; WAQUIL, 2009; PINO; ROCHA, 1994; TONIN; BARCZSZ, 2008).

As variáveis usadas no modelo ARIMAX (2, 0, 1)(1, 0, 0)[12] são os preços do Paraná Cepea (2020), com preços de soja da Bolsa de Chicago FMI (2020) como variáveis exógenas. Para o método ARIMAX (0, 1, 1) foram os preços do Rio Grande do Sul Conab (2020) como variável preditora, com o preço do grão de soja da CBOT, como variável regressora do modelo. Por fim, utiliza-se o preço doméstico da *commodity* agrícola do Mato Grosso Conab (2020) e o preço internacional FMI (2020), como variáveis dependentes e exógenas do modelo ARIMAX (1, 0, 1).

3.3 Redes Neurais

Uma rede neural é um método de previsão que modela a capacidade de aprendizado de um neurônio de um cérebro humano (LANDAU; TAYLOR, 1998). Funciona como um conjunto de elementos de processamento de informações que pode ser ensinado a classificar padrões (LANDAU; TAYLOR, 1998). Cada elemento é denominado neurônio, que envia um sinal para os neurônios da rede (LANDAU; TAYLOR, 1998). Os sinais são recebidos por meio de nós de conexão enviados para camadas ocultas da rede (LANDAU; TAYLOR, 1998). Constantemente, tem-se observado a necessidade de pesquisar sobre previsões de rede neurais em economia, finanças e negócios (HAIDER; HANIF, 2009).

As redes neurais são muito aplicadas como uma opção potencial para previsão de série tempo (FARAWAY; CHATFIELD, 1998; ZHANG; BERARDI, 2001), sendo muito usadas e obtendo sucesso que determinou um papel importante na predição de séries cronológicas (ZHANG; BERARDI, 2001).

As previsões servem de apoio à tomada de decisões no que se refere ao planejamento de produção nos níveis operacional, tático e a alta administração (SAGAERT et al., 2017). Empresas em diversas indústrias enfrentam um problema de demanda muito flutuante e lidam com essa questão gerando previsões de demanda com base nas informações disponíveis para gerar previsões precisas (MIYAOKA; HAUSMAN, 2008).

A rede neural de *Feed-forward* constitui-se em camadas: uma de entrada, uma de saída e, de modo geral, uma ou mais camadas escondidas. A capacidade de aprendizagem em reconhecer padrões desta tipologia de rede neural possibilita ser utilizada como um modelo de previsão aplicado nos negócios (LAW, 2000). As entradas constituem a camada inferior e as saídas formam a camada superior; é possível ter camadas ocultas (HYNDMAN; ATHANASOPOULOS, 2018b).

Diante da flexibilidade das redes neurais, elas não têm um processo sistêmico para criação de modelos (PALMER; JOSÉ MONTAÑO; SESÉ, 2006). A obtenção de um modelo de rede neural robusto ocorre mediante a escolha de um grande número de parâmetros que são obtidos experimentalmente, por meio de tentativa e erro (PALMER; JOSÉ MONTAÑO; SESÉ, 2006). Este método não será utilizado neste estudo.

Para selecionar as arquiteturas dos algoritmos *perceptron* multicamada e máquina de aprendizado extremo foram sem intervenção dos usuários, ou seja, o pacote (nnfor) no R determinou a arquitetura que se ajustou melhor as séries de preços da soja. Os detalhes sobre o funcionamento do pacote nnfor¹ pode ser visto em (KOURENTZES, 2019; KOURENTZES; BARROW; CRONE, 2014).

3.3.1 Algoritmo de *Perceptron* multicamada (MLP)

Devido ao crescente avanço na computação e disponibilidade de dados, as redes neurais são cada vez mais usadas na previsão de diferentes campos de pesquisa

¹ nnfor: é um pacote da *The Comprehensive R Archive Network CRAN* da linguagem de programação R, da autoria de Nikolaos Kourentzes, intitulado *Time Series Forecasting with Neural Networks*. Descrição do pacote: Modelagem automática de séries temporais com redes neurais. Permite especificação de redes totalmente automática, semi-manual ou totalmente manual. Para detalhes da metodologia de especificação, consulte: (i) Crone e Kourentzes (2010) <doi:10.1016/j.neucom.2010.01.017>; e (ii) Kourentzes et al. (2014) <doi:10.1016/j.eswa.2013.12.011>.

(KOURENTZES; BARROW; CRONE, 2014). Uma rede neural constitui-se em camadas: uma de entrada, uma de saída e, de modo geral uma ou mais camadas escondidas (KAASTRA; BOYD, 1996; LAW, 2000; PALMER; JOSÉ MONTAÑO; SESÉ, 2006).

As redes neurais classificam-se em modelos de aprendizagem supervisionada e aprendizagem não supervisionada (CLAVERIA; TORRA, 2014). Os métodos de redes neurais são, por vezes, aplicados pela capacidade de mapear todas as funções lineares ou não lineares (CLAVERIA; TORRA, 2014; FAN; WANG; LI, 2016; WHITE, 1989; ZHANG; BERARDI, 2001).

O algoritmo de rede neural *perceptron* multicamada (MLP) é a rede neural mais utilizada na previsão de séries temporais (KAASTRA; BOYD, 1996; ZHANG; BERARDI, 2001; CLAVERIA; TORRA, 2014; FAN, WANG; LI, 2016). Além disso, é eficiente na predição de série de tempo (DE OLIVEIRA; NOBRE; ZÁRATE, 2013). No tipo de arquitetura MLP as observações (dados) são transmitidos da camada de entrada em direção à camada de saída (PALMER, MONTANO; SESÉ, 2006).

Os neurônios dentro de cada camada possuem interligação por meio de forças de conexão (pesos) que, junto com a arquitetura da rede neural, acumula o aprendizado de uma rede treinada (KAASTRA; BOYD, 1996). Além de neurônios de processamento existe um neurônio de polarização que faz a conexão de cada unidade de processamento nas camadas ocultas e de saída (KAASTRA; BOYD, 1996). Este neurônio tem um valor positivo e tem a função semelhante à interceptação em métodos de regressão (KAASTRA; BOYD, 1996).

De acordo com Kaastra e Boyd (1996) as formas dos neurônios e do viés são organizadas em camadas: uma de entrada, uma de saída e, geralmente uma ou mais camadas ocultas. Neste sentido, “O número de camadas escondidas e neurônios dentro de cada camada pode variar de acordo com o tamanho e a natureza do conjunto de dados” (KAASTRA; BOYD, 1996, p. 217). O modelo de MLP utilizado neste estudo é totalmente automático e foi proposto por Kourentzes, Barrow e Crone (2014) para cumprir o desafio de ter a capacidade de utilizar a rede neural MLP nos dados de séries cronológicas, sem a intervenção de um especialista (CRONE; KOURENTZES, 2010).

O método de rede neural MLP é supervisionado com base em *perceptron* simples original com a programação traseira que treina a rede. O funcionamento básico geralmente ocorre em uma camada de entrada de nós de origem, uma camada de saída e uma ou mais camadas escondidas de neurônios que aumentam a capacidade de aprendizagem do MLP. O número de neurônios ocultos determina o poder de aprender da rede neural (CLAVERIA; TORRA, 2014; FAN, WANG; LI, 2016).

O método de MLP é fundamentado em torno de um filtro neural iterativo que combina uma ferramenta gráfica simples de análise da distância euclidiana em parcelas sazonais ano a ano (CRONE; KOURENTZES, 2010). A forma funcional do modelo é exibida na sequência.

$$\hat{y}_{t+1} = \beta_0 + \sum_{h=1}^H \beta_h g \left(\gamma_{0i} + \sum_{i=1}^I \gamma_{hi} p_i \right). \quad (5)$$

Onde, $\mathbf{w} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ ambos são os pesos do modelo MLP com $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_H]$, $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_{11}, \dots, \boldsymbol{\gamma}_{HI}]$ para as camadas escondidas e a saída; nessa ordem, $\boldsymbol{\beta}_0$ e $\boldsymbol{\gamma}_{0i}$ são os vieses de cada neurônio, que para cada neurônio agem de forma semelhante a interceptação em uma regressão; H é a quantidade de nós escondidos na MLP e $g(\cdot)$ é uma função de transferência não-linear, que em geral é o sigmoide logístico ou a tangente hiperbólica (KOURENTZES, BARROW; CRONE, 2014). A rede pode modelar interações de entradas. As saídas dos nós escondidos estão em conexão a um nó de saída que gera a predição (KOURENTZES, BARROW; CRONE, 2014).

3.4 Algoritmo de máquina de aprendizagem extrema (ELM)

Huang, Zhu e Siew (2006) propõem um novo algoritmo de aprendizagem para as redes neurais *Feedforward*, chamado de máquina de aprendizagem extrema. Este algoritmo possui como vantagens: tem uma tendência em alcançar o menor erro de treinamento; obtém a menor norma de pesos e o bom desempenho de generalização; e a velocidade de aprendizado pode ser muito mais rápida (HUANG; ZHU; SIEW, 2004).

Para N amostras arbitrárias distintas $(\mathbf{X}_i, t_i)$, onde $\mathbf{X}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathcal{R}^n$ e $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathcal{R}^m$, SLFN padrão com $\tilde{N}$ nós ocultos e função de ativação $g(x)$ são modelados matematicamente como

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(\mathbf{x}_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = o_j, \quad j = 1, \dots, N, \quad (6)$$

Onde $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ é o vetor de peso que liga o i th nó escondido e os nós de entrada, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ é o vetor de peso que liga o i th nó escondido e os nós de

saída, e b_i é o limiar do nó oculto i . $\mathbf{w}_i \cdot \mathbf{x}_j$ denota o produto interno de $\mathbf{w}_i$ e $\mathbf{x}_j$. Que SLFN padrão com $\tilde{N}$ nós ocultos com função de ativação $g(x)$ pode aproximar estas N amostras com zero erro significa que $\sum_{j=1}^{\tilde{N}} \|\mathbf{o}_j - t_j\| = 0$, isto é, existe β_i , $\mathbf{w}_i$ e b_i de tal modo que

$$\sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = t_j, \quad j = 1, \dots, N. \quad (7)$$

Este algoritmo da máquina de aprendizagem extrema (ELM) foi utilizado com a rede neural *perceptron* multicamada (MLP). Os detalhes sobre o funcionamento do ELM podem ser vistos em (HUANG; ZHU; SIEW, 2006).

3.5 Indicadores de Desempenho

Medidas de erro são métodos de se obter informações sobre a robustez do modelo de previsão dos valores reais, tanto para o ajuste do modelo aos dados quanto para a mensuração de valores fora da amostra (MAKRIDAKIS, 1993; MORETTIN; SINGER, 2020; PETROPOULOS et al., 2020).

O *mean absolute error* (MAE) constitui-se numa das primeiras medidas de erro, sendo singela de calcular e simples de compreender (BONETT; SEIER, 2003). O erro percentual absoluto médio (MAPE) e MAE são indicadores de desempenho que geram a diferença entre o valor real e o previsto, neste sentido, quanto menor os valores gerados pelos indicadores, melhor é a acuracidade das previsões (SAFARI; DAVALLOU, 2018).

O MAPE é uma medida de erro estatística que gera informações de erro de previsão em porcentagem, sendo simples de calcular e simples de compreender (DAS; MISHRA; ROUT, 2017). É muito utilizado pela sua facilidade de interpretação (MYTTENAERE et al., 2016). MAPE é um dos critérios mais conhecidos de acuracidade da predição (KIM; KIM, 2016). MAE é classificado na previsão de série de tempo como um critério básico de erro de previsão (DAS; MISHRA; ROUT, 2017).

$$MAE = T^{-1} \sum_{t=1}^T |Y_{(t)} - \hat{Y}_{(t)}| \quad (8)$$

$$MSE = \sum_{t=1}^T \left(\frac{Y_{(t)} - \hat{Y}_{(t)}}{T - 1} \right)^2 \quad (9)$$

$$MAPE = T^{-1} \sum_{t=1}^T \left| \frac{Y_{(t)} - \hat{Y}_{(t)}}{Y_{(t)}} \right| \quad (10)$$

$$RMSE = \left(T^{-1} \sum_{t=1}^T (Y_{(t)} - \hat{Y}_{(t)})^2 \right)^{\frac{1}{2}} \quad (11)$$

Fonte: Safari e Davallou (2018)

Onde, $Y_{(t)}$ é os valores reais no tempo t ; $\hat{Y}_{(t)}$ é os valores de previsão no tempo t ; e T é o tamanho da amostra (SAFARI; DAVALLOU, 2018).

O *Mean Squared Error* (MSE) mensura o erro quadrático; os resultados do MSE, quanto mais próximo de zero, melhor. Esta medida também apresenta a diferença entre o valor realizado e o valor previsto e geralmente é utilizada para verificar como o modelo de previsão se ajustou às observações (DAS; MISHRA; ROUT, 2017). A raiz do erro quadrático médio (simbolizado por RMSE) representa as diferenças quadráticas entre as séries de preços em relação às observações e ajustes.

Da mesma forma, o RMSE na Equação (11), pode ser interpretado da seguinte forma: quando o valor dessa métrica é alto significa que a margem de erro nas variáveis ajustadas também é alta, enquanto valores próximos de zero indicam um ajuste muito preciso (CAMELO et al., 2018b).

CAPÍTULO IV

Neste capítulo apresenta-se a classificação metodológica do estudo. A seção 4.1.2 mostra a coleta dos dados. Em seguida, nas seções 4.1.3 até 4.1.6 são apresentados o *software* (linguagem de programação R) e os pacotes específicos do R, para rodar os modelos de previsão utilizados.

Na seção 4.2 são apresentados os principais resultados da pesquisa e a discussão dos mesmos, sendo: as comparações empíricas de previsões e a comparação das previsões com a literatura de previsão do preço de *commodities* agrícolas.

Este capítulo responde ao terceiro e quarto objetivos específicos: c) gerar previsões dos modelos ARIMAX, redes neurais perceptron multicamada e máquina de aprendizagem extrema nos preços de soja; e d) comparar o desempenho da previsão dos modelos ARIMAX, MLP e ELM por meio de métricas de erro.

4.1 Procedimentos Metodológicos

4.1.1 Classificação da pesquisa

Segundo Major (2017) a pesquisa positivista é um dos paradigmas dominantes nas ciências que considera leis universais. “A formulação dessas leis deve obedecer às regras do método científico preconizadas para as ciências naturais, de forma que, por meio da explicação dos fenômenos observados, seja possível prever sua ocorrência no futuro” (MAJOR, 2017, p. 173).

Este estudo está embasado na perspectiva do positivismo, uma vez que parte da hipótese de que a aplicação dos modelos de previsão (métodos quantitativos) para prever o preço da soja gera resultados importantes em termos precisão das previsões. Os métodos quantitativos consistem em quantificar as informações coletadas e fazer um tratamento mediante métodos estatísticos (RICHARDSON, 2009).

Os métodos de previsões são classificados em duas categorias: modelos qualitativos ou modelos quantitativos (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998). Os métodos quantitativos de previsão dividiram-se em dois grandes grupos, a saber: modelos de série de tempo e métodos explicativos. No caso dos modelos univariados, as previsões são geradas a partir das observações registradas (séries temporais) (PETROPOULOS et al., 2020; RIBEIRO, 2014).

Os métodos de previsão com base em séries temporais são mais fáceis de usar, assim possuindo vantagem sobre os modelos explicativos. Assumem que as previsões dependem somente do comportamento passado dos dados e os padrões existentes nas séries temporais, desconsiderando as variáveis exógenas que influenciam a série temporal (RIBEIRO, 2014).

A pesquisa tem uma abordagem quantitativa, dado que os modelos testados baseiam-se na inteligência artificial e estatística para predições do comportamento dos preços agrícolas. A análise quantitativa dos dados segue o que ponderam Bettis et al. (2014) sobre o seu uso na pesquisa em administração quanto à previsão de ocorrências futuras versus dados empíricos.

A estratégia adotada foram previsões com os intervalos de predição definidos previamente. Estas foram realizadas com base nos preços históricos do comportamento da soja. A coleta dos dados foi com base nos dados secundários de preços da soja em três fontes de dados.

As amostras classificam-se como coleta de dados longitudinais, pois é realizado um recorte temporal nas amostras, além de terem sido coletadas uma única vez nas respectivas bases de dados. A pesquisa tem um caráter quantitativo quanto à coleta de informações, análise dos dados e os métodos para alcançar os resultados.

Os indicadores de desempenho de previsão *mean absolute percentage error* (MAPE), *mean absolute error* (MAE), *mean squared error* (MSE) e *root mean square error* (RMSE), que foram utilizados nessa pesquisa, são medidas de erro consolidadas na literatura de previsão de séries temporais. Em se tratando de predição de *comoditties*, esses indicadores são os mais aceitos para medir o desempenho de modelos de previsão.

As métricas de erros, explicadas matematicamente na seção 3.4.1 e posteriormente avaliadas na discussão de resultados 4.3, são apresentadas no Quadro 2. Este quadro mostra os indicadores de desempenho, mais utilizados na literatura empírica e autores que adotaram as medidas de erro para mesurar a capacidade de previsão dos métodos.

Quadro 2 - Indicadores de desempenho mais utilizados nas publicações

Métricas de erros	Autores
<i>Mean absolute error</i> (MAE)	Zou et al., (2007), Ribeiro, Sosnoski e Oliveira (2010), Jha e Sinha (2012), Carrasco-Gutierrez e Almeida (2013), David et al., (2017), Wang et al., (2017), Souza et al. (2018), Drachal (2019) e Ribeiro e Coelho (2020).

<i>Mean absolute percentage error</i> (MAPE)	Ribeiro, Sosnoski e Oliveira (2010), Ahumada e Cornejo (2016), Ahumada e Cornejo (2016a), Feliciani, Souza e Souza (2015), Wang et al., (2017), Puchalsky et al., (2018), Zhang et al., (2018), Kitworawut e Rungreunganun (2019), Wang et al., (2019), Ribeiro e Coelho (2020) e Zhang et al., (2020).
<i>Mean squared error</i> (MSE)	Allen (1994), Zou et al., (2007), Ribeiro, Sosnoski e Oliveira (2010), Pinheiro e Senna (2015), (2016); Puchalsky et al., (2018) e Ribeiro e Coelho (2020).
<i>Root mean square error</i> (RMSE)	Allen (1994), Ferreira et al., (2011), Ahumada e Cornejo (2016), Jha e Sinha (2012), Jha e Sinha (2013), Ahumada e Cornejo (2016b), Wang et al., (2017), Drachal (2019), Kitworawut e Rungreunganun (2019), Ribeiro e Coelho (2020) e Sabu e Kumar (2020).

Fonte: Elaborada pelo autor.

As métricas foram usadas em pesquisas desenvolvidas no Brasil sobre previsão de preços de *commodities* agrícolas Feliciani; Souza e Souza (2015); Nomelini et al., (2017); Pinheiro e Senna, (2017c); Souza et al., (2018); e Tibulo e Carli (2014); o mesmo ocorreu em pesquisas internacionais (AHUMADA; CORNEJO, 2016a; ALLEN, 1994; PUCHALSKY et al., 2018). Este levantamento dos indicadores, mais utilizadas serviu para justificar a escolha dos indicadores de precisão para medir a acurácia de cada método de previsão.

Os modelos de previsão de redes neurais tem mostrado ser capazes de gerar previsão precisas para os preços da soja e milho em muitas pesquisas (BRESSAN, 2004; BRESSAN; LIMA, 2002; CERETTA; RIGHI; SCHLENDER, 2010; FERREIRA et al., 2011; PINHEIRO; SENNA, 2016, 2017b, 2017a; PUCHALSKY et al., 2018; RIBEIRO; SOSNOSKI; OLIVEIRA, 2010).

Conforme consultado em estudos anteriores, o modelo de rede neural proposto na pesquisa tem gerado previsões do preço das *commodities* agrícolas com desempenhos de precisão eficientes, o que reforça a escolha do método de predição.

4.1.2 Coleta de dados

As séries temporais são formadas por valores observados em vários períodos ordenados e subsequentes (KAZMIER, 1982; PUCHALSKY et al., 2018; SHUMWAY; STOFFE, 2010; STEVENSON, 1981). Segundo a Embrapa soja (2020), os três primeiros

estados com maior nível de produção da safra 2019/2020 foram Mato Grosso, Paraná e Rio Grande do Sul. Por esta razão, os preços destes três estados foram escolhidos para serem analisados neste estudo. Além disso, a soma da produção de soja dos três estados representou aproximadamente 59% da produção nacional em 2020 (IBGE, 2021).

Os preços foram coletados nas bases de dados do Centro de Estudos Avançados em Economia Aplicada (Cepea)² da Escola Superior de Agricultura Luiz de Queiroz (Esalq), Fundo Monetário Internacional (FMI)³ e a Companhia Nacional de Abastecimento (CONAB)⁴.

Os preços da soja em dólares são fornecidos pelo Fundo Monetário Internacional (FMI), Centro de Estudos Avançados em Economia Aplicada (CEPEA) e os preços da CONAB são em reais; todos os preços tem periodicidade mensal. Tanto os preços do FMI quanto do CEPEA e da CONAB são referentes à saca de 60 quilos de soja (CEPEA, 2020; CONAB, 2020; FMI, 2020).

As séries de preços da soja do CEPEA são do grão comercializado no porto do Paraná e a do FMI são os valores do contrato futuro (primeiro contrato a termo) da Bolsa de Chicago (CBOT). Os preços de soja do Cepea são nacionais e os da Bolsa de Chicago (CBOT), internacionais.

Em relação à escolha da variável exógena usada no modelo ARIMAX, é importante ressaltar que o preço da CBOT está relacionado com o modelo teórico de Mundlak e Larson (1992). Além disso, Giombinsky (2006); Pino e Rocha (1994); e Tonin e Barczysz (2008), comprovam a transmissão de preço por meio dos modelos econométricos, mostrando que o preço da CBOT é a principal variável determinante na formação do preço doméstico.

Para prever e validar os modelos, o procedimento mais realizado é a separação da série temporal em duas partes: uma estima os parâmetros dos métodos, denominada conjunto de treinamento, e a outra serve para medir a acurácia da previsão realizada para o conjunto de teste (HERRERA et al., 2019; HYNDMAN; ATHANASOPOULOS, 2018b; MANCUSO; WERNER, 2019; MORETTIN; SINGER, 2020; WANG et al., 2019).

Herrera et al. (2019); Wang et al., (2019); e Morettin e Singer (2020) ponderam que os estudos geralmente utilizaram a maior parte da série para treinar o método, sendo uma parte menor da série separada para avaliar as previsões. No presente estudo adota-se o mesmo procedimento para os preços da soja.

² Disponível em: <https://www.cepea.esalq.usp.br/br/consultas-ao-banco-de-dados-do-site.aspx>

³ Disponível em: <https://www.imf.org/en/Research/commodity-prices>

⁴ Disponível em: <http://sisdep.conab.gov.br/precosiagroweb/>

Os conjuntos de treinamento são nos períodos de janeiro de 2011 até dezembro de 2018, atingindo 96 observações mensais de amostra para cada série de preços. Foi elaborada a Tabela 1 para resumir as séries temporais de preços da soja e demonstrar a quantidade de observações de cada série temporal.

Tabela 1 - Séries temporais de preços mensais da soja

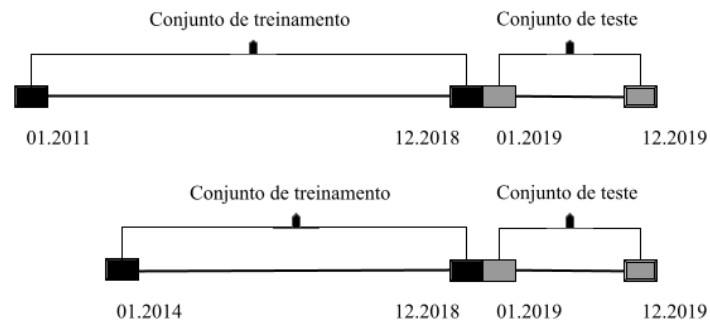
Base de dados	Preços da soja	Data da coleta	Número de observações
CEPEA	Conjunto de treinamento PR	18/09/2020	96
FMI	Conjunto de treinamento CBOT	18/09/2020	96
FMI	Conjunto de treinamento CBOT	18/09/2020	60
CONAB	Conjunto de treinamento RS	02/10/2020	60
CONAB	Conjunto de treinamento MT	02/10/2020	60

Fonte: elaborada pelo autor.

Os conjuntos de teste dos preços compreendem janeiro a dezembro de 2019; são doze observações para comparar com as previsões dos modelos no mesmo horizonte de previsão, doze meses à frente.

Para cada série de preços da soja simulada neste estudo são realizadas três modelagens (ARIMAX, MLP e ELM) propostas com 96 e 60 períodos (treinamentos), reservando-se as últimas 12 observações para testes de precisão. A Figura 13 mostra os conjuntos de dados e seus respectivos períodos de treinamentos e teste dos métodos de previsão.

Figura 13 - Os dados dos conjuntos de treinamento são de janeiro de 2011, e janeiro de 2014 a dezembro de 2018, e os dados do conjunto de teste são de janeiro de 2019 a dezembro de 2019



Fonte: Elaborada pelo autor.

As séries temporais são divididas em dois conjuntos de treinamento e um conjunto de teste, conforme mostrado na Figura 13. O desempenho de previsões de 12 meses à frente são avaliadas sobre os dados de teste.

4.1.3 Pacotes em R studio

As abordagens de rede neural *perceptron* multicamada (MLP) e máquina de aprendizagem extrema (ELM) são métodos da inteligência computacional no campo da inteligência artificial, respectivamente. Para realizar as previsões e comparar a robustez dos modelos ARIMAX, MLP e ELM foi adotado o apoio da linguagem de programação R.

Os procedimentos de análises de predição e acuracidade das previsões ocorreram no R studio, mediante os pacotes *forecast* e *nnfor* do R (HYNDMAN et al., 2020; KOURENTZES, 2019). Os modelos de rede neural (MLP) e máquina de aprendizagem (ELM) que são usados nesta pesquisa pertencem ao manual do pacote *nnfor* (KOURENTZES, 2019). O quadro 3 mostra uma lista indicativa de pacotes e funções empregadas neste estudo.

Quadro 3 - Lista de pacotes e funções em linguagem de programação R

Seção relacionada	Pacotes	Funções	Finalidade
2.1 Revisão bibliométrica	bibliometrix	interface biblioshiny	Análise bibliométrica
4.2.1 e 4.2.2	ggplot2	autoplot	Para visualização de dados, ajuste e previsões de modelos
Testes de raiz unitária	urca	ur.kpss, ur.df	Testes Kwiatkowski-Phillips-Schmidt-Shin (KPSS) e ADF
Série Temporal	ts	ggtsdisplay	Visualização das séries estacionárias
Série Temporal	forecast	diff; ndiffs	Estima o número de diferenças necessárias para tornar a série estacionária
4.2.1. Modelagem ARIMAX	methods	summary	Visualização dos parâmetros do modelo ARIMAX
4.2.1. Modelagem ARIMAX	forecast	auto.arima	Função para ajuste e previsão com modelos ARIMAX.
3.7 Indicadores de Desempenho	forecast	accuracy	Medidas de precisão para um modelo de previsão.
4.2.1. Modelagem ARIMAX	forecast	checkresiduals	Verifique se os resíduos de um modelo de série temporal parecem ruído branco e teste de hipótese Ljung-Box
4.2.2 Treinamento e teste de MLP ELM	nnfor	mlp; elm	Previsão de séries temporais com redes neurais e máquina de aprendizado
4.3.1 Boxplots e testes de variâncias	stats	var.test	Executa um teste F para comparar as variâncias de duas amostras de populações normais.
4.3.1 Boxplots e testes de variâncias	ggplot2	geom_boxplot	Análise dos gráficos de Box plots dos erros

Fonte: Elaborado pelo autor.

Para analisar a condição de estacionariedade de séries de preços da soja foram usadas as funções *ndiffs* e *diff*. Para delimitar especificamente os meios de aplicação dos modelos de previsão, as funções em R utilizadas para prever são *auto.arima*, *mlp* e *elm*. Já para comparar a acurácia dos métodos de predição foi usada a função *accuracy*.

No quadro 3 são apresentados os pacotes e funções necessárias para aplicar as abordagens de previsão. Neste procedimento empírico para os pesquisadores que tiverem a necessidade de replicar as abordagens de predição da pesquisa em outros dados de (séries temporais) foram disponibilizadas as execuções dos *scripts* em linguagem de programação R, com todos os códigos abertos para rodar os modelos de previsão e as instruções em cada etapa. No apêndice 1, quadros 4 e 5, são mostrados os procedimentos práticos para replicabilidade dos métodos.

4.2 Resultados

4.2.1 Modelagem ARIMAX

De acordo com Herrera et al. (2019); Mancuso e Werner (2019); e Morettin e Singer (2020) para desenvolver a modelagem de dados via ARIMAX é necessário dividir a série temporal em dois conjuntos de dados: teste e treino. Neste contexto, foi realizado esse procedimento com os preços divididos em duas partes, a saber, os conjuntos de treinamento e teste.

Segundo Anggraeni et al. (2017) o método ARIMAX não tem sido muito aplicado como modelo de previsão na área econômica. Ao mesmo tempo, tem gerado resultados de previsão promissores quando prevê variáveis que são afetadas por outras variáveis. Camelo et al., (2017); Camelo et al., (2018a); e (2018b) utilizaram o modelo ARIMAX para previsão de velocidade do vento, e seus resultados apresentaram previsões precisas e eficazes das variáveis modeladas.

O modelo ARIMAX tem mostrado ser um bom método para modelar o preço do arroz (Anggraeni et al., 2017), o preço da borracha natural na Tailândia (Cherdchoongam e Rungreunganun, 2016), os preços do aço na Turquia (Adli, 2020) e previsão de velocidade do vento (CAMELO et al., 2017, 2018a, 2018b).

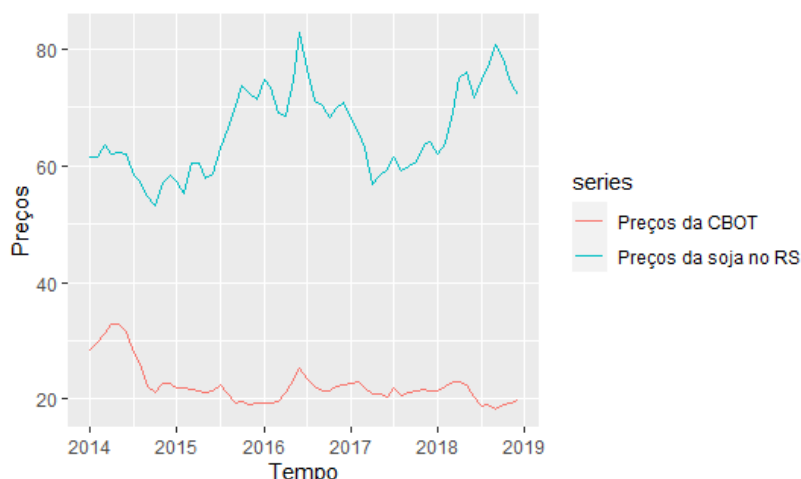
O conjunto de treinamento foi de janeiro de 2011 até dezembro de 2018 (96 observações) e o conjunto de teste foi de janeiro de 2019 até dezembro de 2019 (12 observações). Para os preços da soja no Rio Grande do Sul e Mato Grosso, os conjuntos de treinamento⁵ tiveram um recorte temporal diferente, nos períodos citados.

⁵ Os períodos de conjuntos de treinamento para Rio Grande do Sul e Mato Grosso correspondem ao período de 2014 até 2018, porque a Conab não disponibilizou os preços que antecedem o ano de 2014. Por este motivo, os dados disponíveis para coleta são a partir de 2014.

O comportamento dos preços de soja da Bolsa de Chicago (CBOT) e da soja do Cepea estão na Figura 15. Os preços do grão no Rio Grande do Sul e Mato Grosso, junto aos preços da CBOT, estão nas Figuras 14 e 15. São selecionadas essas variáveis para previsão de preços da soja dos modelos ARIMAX, pois o modelo teórico da lei do preço único pressupõe que a formação de preço agrícola ocorre principalmente em função de preços internacionais e domésticos (MUNDLAK; LARSON, 1992).

A visualização do comportamento dos dados é recomendação de Hyndman e Athanasopoulos (2018b), que orientam mostrar os dados em gráfico de série temporal, representados nas Figuras 14, 15 e 16.

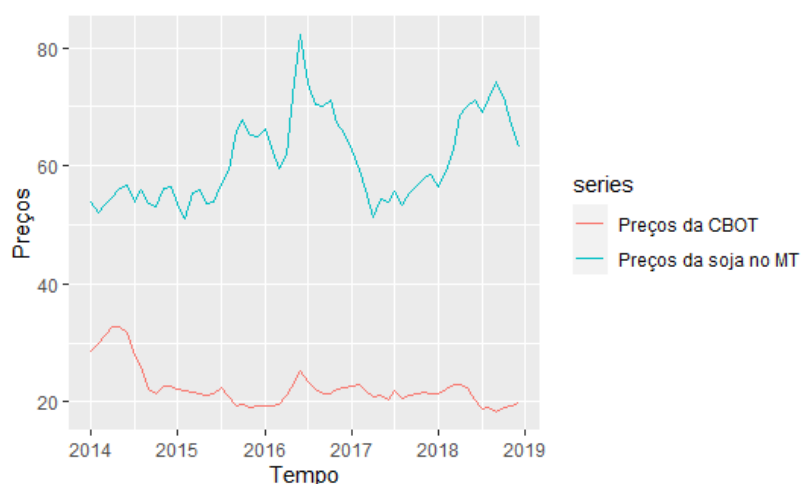
Figura 14 - Preços da soja no Rio Grande do Sul e os preços da Bolsa de Chicago



Fonte: Elaborado pelo autor, por meio do R studio e com base em dados do FMI e CONAB

A Figura 14 mostra um comparativo entre os preços da soja no Rio Grande do Sul e os preços da Bolsa de Chicago. Já a Figura 15 apresenta os preços do grão no Mato Grosso, em comparação à Bolsa referida anteriormente. Todos os valores do Rio Grande do Sul e Mato Grosso estão em reais.

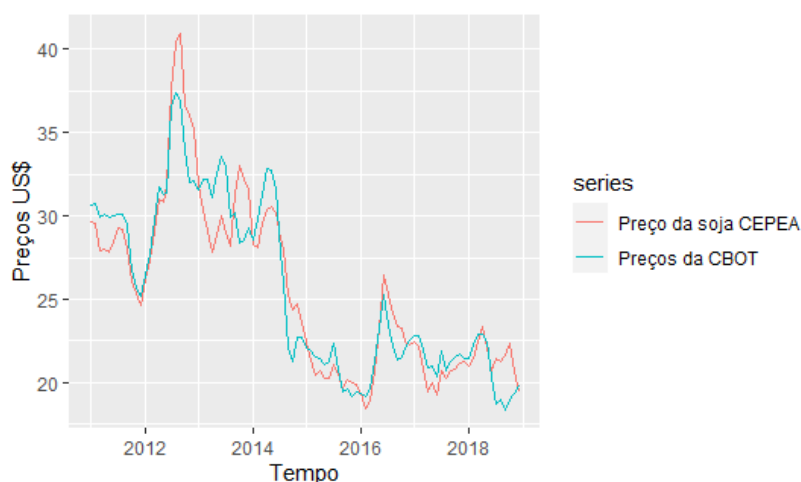
Figura 15 - Preços da soja no Mato Grosso e os preços da Bolsa de Chicago



Fonte: Elaborado pelo autor, por meio do R studio e com base em dados do FMI e CONAB

A Figura 16 mostra os preços no Paraná em dólar na cor vermelha e os preços da bolsa de Chicago na cor azul. Ambas as variáveis estão na mesma moeda, é possível identificar que os preços domésticos acompanham o comportamento dos preços internacionais da CBOT. No entanto, os preços domésticos são mais voláteis do que o preço da Bolsa de Chicago. Para Giombinsky (2006) e Pino e Rocha (1994), o preço doméstico da *commodity* sofre influências de preços internacionais que afetam a volatilidade do preço interno.

Figura 16 - Séries temporais dos preços da Bolsa de Chicago e CEPEA em dólares US\$



Fonte: Elaborado pelo autor, por meio do R studio e com base em dados do FMI e Cepea

Os preços da tonelada de soja da CBOT foram convertidos no valor da saca de 60 kg. Para descobrir o valor da saca foi preciso dividir o preço da tonelada por 16,67 (número de sacas que equivalem a uma tonelada). Foi realizada a conversão, pois o modelo ARIMAX não

se ajustou bem aos preços da tonelada como variáveis exógenas. Ambos os preços correspondem ao período de janeiro de 2011 até dezembro de 2018. Este intervalo representa o conjunto de treinamento do ARIMAX para encontrar os parâmetros que melhor representam os preços e verificar como o modelo se comporta.

Segundo Constantino, Mendes e Moreira (2019) os métodos autorregressivos possuem pressupostos para estimar e desenvolver o modelo, sendo que a modelagem de séries temporais considera as séries estacionárias. Nos casos em que as séries não são estacionárias, devem ser transformadas (ADLI, 2020; ANGGRAENI et al., 2017; CARRASCO-GUTIERREZ; ALMEIDA, 2013; CONSTANTINO; MENDES; MOREIRA, 2019).

Posteriormente, os preços foram submetidos ao teste de hipótese, que apresenta o número de diferenças necessárias para tornar uma série estacionária. Este teste foi empregado mediante a função *ndiffs* do pacote *forecast* (HYNDMAN et al., 2020). O respectivo teste apresentou uma única diferenciação para todas as séries de preços da soja, conforme mostra a Tabela 2.

Depois que os dados são diferenciados, é verificado se as séries temporais são estacionárias (ANGGRAENI et al., 2017). Um método de determinar objetivamente se as séries são estacionárias é um teste de raiz unitária (ANGGRAENI et al., 2017; HYNDMAN; ATHANASOPOULOS, 2018c). A partir do número de diferenciações recomendadas pela função *ndiffs*, as séries foram diferenciadas uma vez, por meio da função *diff*, conforme mostrado na Tabela 2.

As séries de preços da soja, com uma diferença, foram submetidas aos testes de raiz unitária⁶ *Augmented Dickey-Fuller* (ADF), conforme desenvolvido em Dickey e Fuller (1979) e KPSS Kwiatkowski et al. (1992), para verificar se os dados das séries temporais são estacionários ou não. Estes testes diferem quanto à especificação da hipótese nula, sendo a hipótese nula do teste ADF a presença de uma raiz unitária na série; já o teste KPSS tem a hipótese nula de estacionariedade. De acordo com Bueno (2011), os testes KPSS e ADF são complementares.

Realizado o processo de diferenciação para tornar as séries estacionárias, é necessário comprovar adotando-se os testes de raiz unitária KPSS e ADF, por meio das funções *ur.kpss*, *ur.df* ambas do pacote *urca* (PFAFF, 2008) em R. O teste KPSS foi escolhido, pois tem maior eficácia que o teste de *Dickey Fuller* Aumentado (ADF). Segundo Bueno (2011), o KPSS foi desenvolvido para superar as limitações do teste ADF.

⁶ Maiores detalhes sobre o procedimento operacional para a execução dos testes de raiz unitária dos tipos, Dickey-Fuller Aumentado (ADF) e KPSS, podem ser encontrados em Bueno (2011) e Wolff e Souza (2013).

No teste de raiz unitária ADF, a interpretação é a seguinte:

H_0 : Tem raiz unitária. (Não é estacionária)

H_1 : Não tem raiz unitária. (É estacionária)

Tomar decisão: Espera-se que a estatística t do teste, seja negativa e menor do que o valor crítico para que a hipótese nula (H_0) seja rejeitada e concluir que a série é estacionária. A Tabela 2 exhibe os resultados dos testes de raiz unitária.

Tabela 2 - Número de diferenças necessárias para tornar estacionárias as séries, testes de raiz unitária ADF e KPSS nos conjuntos de treinamento

Conjunto de treinamento	Série Temporal	Diferenças	KPSS	ADF (t)	p-valor	Valor Crítico (1%)
PR	Preço da soja no PR	1	0.0695	-6.32	1.14	-2.6
CBOT	Preço da soja no CBOT	1	0.0602	-5.90	7.22	-2.6
RS	Preço da soja no RS	1	0.0709	-6.59	4.51	-2.6
MT	Preço da soja no MT	1	0.0675	-6.20	2.49	-2.6

Nota: Estatística (t), o valor da estatística (t) é inferior ao valor crítico da estatística *Dickey-Fuller* ao nível de 1% para as séries com diferença.

Fonte: Elaborado pelo autor por meio do R *studio* com base no pacote *urca*.

Neste teste de raiz unitária KPSS, a interpretação é a seguinte:

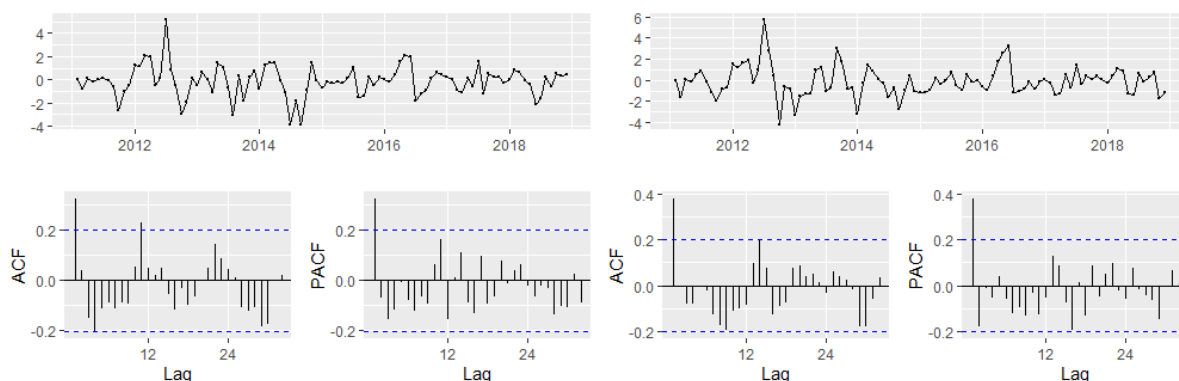
H_0 : Os dados são estacionários.

H_1 : Os dados não são estacionários.

Tomar decisão: se o p-valor for maior do que 0,05, ou seja, $p > 0,05$ (não rejeita H_0) (HYNDMAN; ATHANASOPOULOS, 2018c). Observa-se que depois da primeira diferença o p-valor de todas as séries foram maiores do que 0,05, evidenciando a não rejeição da hipótese nula. Portanto, concluiu-se que as séries diferenciadas são estacionárias.

O teste ADF também indicou que as quatro séries de preços da soja são estacionárias, com rejeição da H_0 , ou seja, a hipótese nula de raiz unitária é rejeitada, indicando que as séries de preços da soja são estacionárias. Os resultados dos testes KPSS e ADF foram apresentados na Tabela 2. As séries de preços da soja, com uma diferenciação são mostradas nas Figuras 17 e 18.

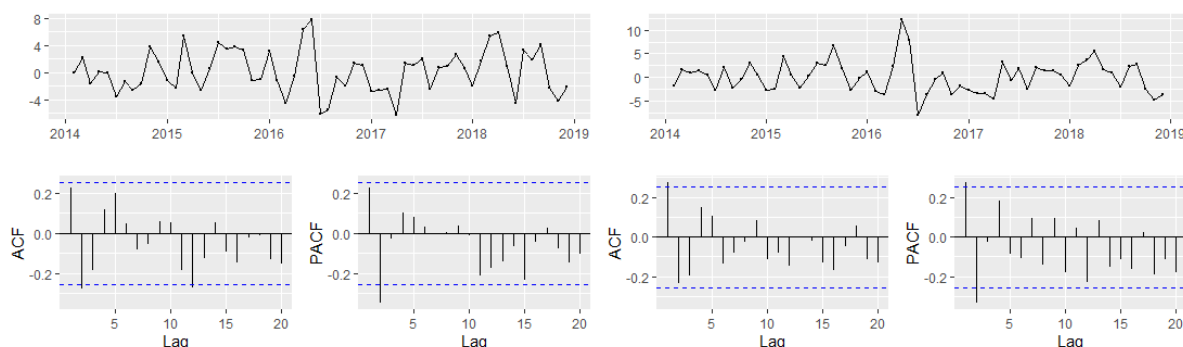
Figura 17 - Séries de preços da CBOT e Paraná com uma diferenciação



Fonte: Elaborado pelo autor por meio da *ggtsdisplay* com base em dados do FMI e Cepea

Apresentam-se os preços estacionários em gráficos de séries temporais nas Figuras 17 e 18 para visualização da transformação realizada nos preços da soja, pelos procedimentos anteriormente expostos.

Figura 18 - Séries de preços no Rio Grande do Sul e Mato Grosso com uma diferenciação



Fonte: Elaborado pelo autor, por meio da *ggtsdisplay* com base em preços da CONAB.

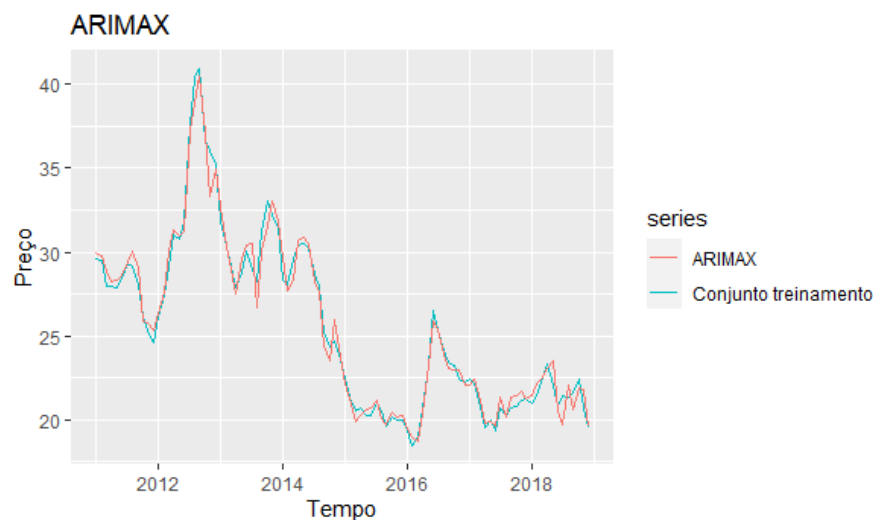
A partir dos conjuntos de treinamento, o modelo ARIMAX foi estimado dentro deste conjunto para verificar como o método se ajustou. Os parâmetros do ARIMAX são encontrados por meio da função *auto.arima*, que faz parte do pacote *forecast* (HYNDMAN; KHANDAKAR, 2008).

A escolha dos parâmetros da função: *auto.arima* leva em consideração o critério Akaike que identifica o menor valor para o coeficiente do critério de informação Akaike, na qual resulta no melhor modelo do ponto de vista estatístico (HYNDMAN; ATHANASOPOULOS, 2018d).

Os ajustes dos modelos ARIMAX nos conjuntos de treinamento são gerados para observar como o ARIMAX acompanhou os preços da soja e mostrar os melhores parâmetros

para os preços. As Figuras 19, 20 e 21 apresentam o comportamento das estimativas dos conjuntos de treino.

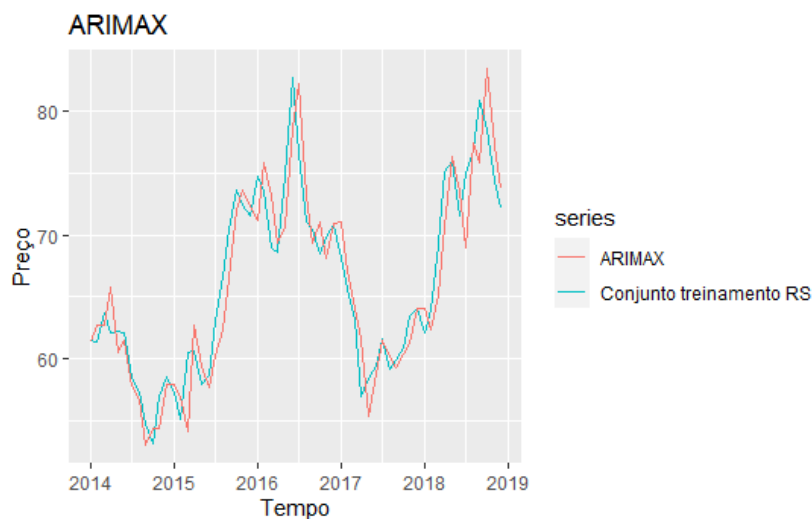
Figura 19 - Estimativa do ARIMAX no conjunto de treinamento para o preço do Paraná



Fonte: Elaborado pelo autor, por meio do R *studio* com base nos pacotes *forecast* e *ggplot2*.

Em azul são as observações (séries em nível), em vermelho os ajustes do ARIMAX aos preços da soja (conjuntos de treinamento). Considerando a Figura 19, a modelagem parece ter se ajustado bem aos preços da soja. Foi utilizado o ARIMAX (0, 1, 1) para a estimativa de preços do Rio Grando do Sul, comparando com os preços realizados que mostra ter um ajuste entre o preço da soja e o modelo, conforme a Figura 20.

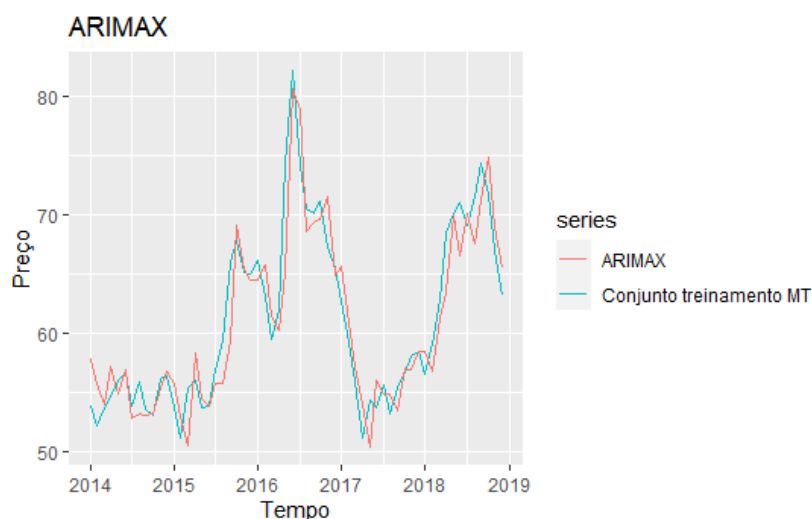
Figura 20 - Estimativa do ARIMAX no conjunto de treinamento para o preço do Rio Grande do Sul



Fonte: Elaborado pelo autor, por meio do R *studio* com base nos pacotes *forecast* e *ggplot2*.

Mediante a Figura 21 foi possível identificar que o modelo consegue acompanhar o comportamento dos preços da soja.

Figura 21 - Estimativa do ARIMAX no conjunto de treinamento para o preço do Mato Grosso



Fonte: Elaborado pelo autor, por meio do R *studio* com base nos pacotes *forecast* e *ggplot2*.

Os resultados de estimativa do método ARIMAX para todos os preços da soja mostraram que o modelo acompanhou bem próximo dos preços reais, no entanto para afirmar que o método de previsão se ajustou bem aos dados, é necessário julgar os modelos de predição através das medidas de erro (PETROPOULOS et al., 2020).

O resultado da estimativa dos modelos são ARIMAX (2, 0, 1)(1, 0, 0)[12] para o conjunto de treinamento Paraná, ARIMAX (0, 1, 1) para o conjunto de treinamento Rio Grande do Sul e ARIMAX (1, 0, 1). A acurácia das estimativas do ARIMAX, obtidas mediante os indicadores de desempenho dos conjuntos de treinamento para as séries de preços do Paraná, Rio Grande do Sul e Mato Grosso, pode ser vista na Tabela 3.

Tabela 3 - Parâmetros e métricas do ARIMAX nos conjuntos de treinamento

Preço	MAPE	RMSE	MASE	AIC	AICc	BIC	XREG	P, D, Q, R
PR	2.04	0.69	0.12	221.19	222.14	236.58	0.0037	(2, 0, 1)(1, 0, 0)[12]
RS	3.18	2.69	0.20	291.63	292.06	297.86	0.32	(0, 1, 1)
MT	3.39	2.71	0.21	303.08	304.19	313.55	0.34	(1, 0, 1)

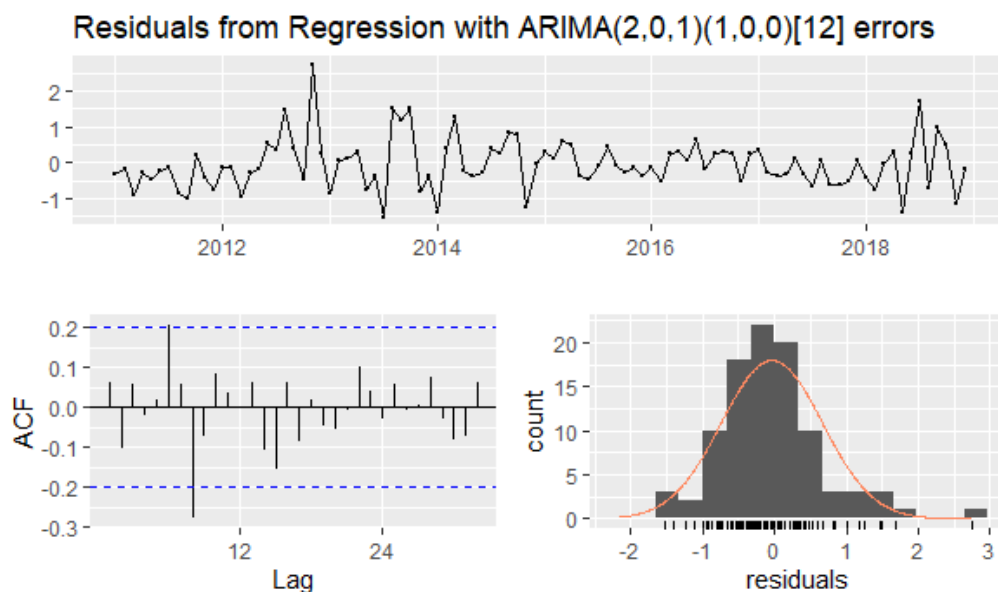
Fonte: Elaborado pelo autor com base em preços do FMI, Cepea e Conab.

As métricas de erro dos conjuntos de treinamento são dados na Tabela 3, sendo que a interpretação ocorre da seguinte forma: quanto menor o valor, melhor foi o ajuste do modelo (PETROPOULOS et al., 2020; SAFARI; DAVALLOU, 2018). A partir disso, avalia-se os indicadores MAPE (2,04%) e RMSE (0,69). No caso do MAPE, o modelo acertou aproximadamente 98% dos preços realizados. Ambos os erros foram muito baixos, o que indica um bom ajuste do ARIMAX (2,0,1)(1,0,0)[12].

Em uma modelagem adequada na abordagem de predição, os resíduos do modelo ajustado devem ser ruído branco (MORETTIN; SINGER, 2020). O ruído branco é quando as séries não apresentam autocorrelação (HYNDMAN; ATHANASOPOULOS, 2018e; MORETTIN; SINGER, 2020). Para verificar se os resíduos são correlacionados, aplicou-se os gráficos da função de autocorrelação ACF apresentados nas Figuras 22, 23 e 24.

Para checar os resíduos dos modelos ARIMAX (2,0,1)(1,0,0)[12], (0, 1, 1) e (1, 0, 1) foi empregada a função em R: *checkresiduals* do (HYNDMAN et al., 2020). A Figura 22 apresenta o gráfico ACF dos resíduos do ARIMAX (2,0,1)(1,0,0)[12]. Esse gráfico mostra que quase todas as autocorrelações estão dentro dos limites, com exceção de um *lag* 12, assim, sugere-se que os resíduos são ruído branco.

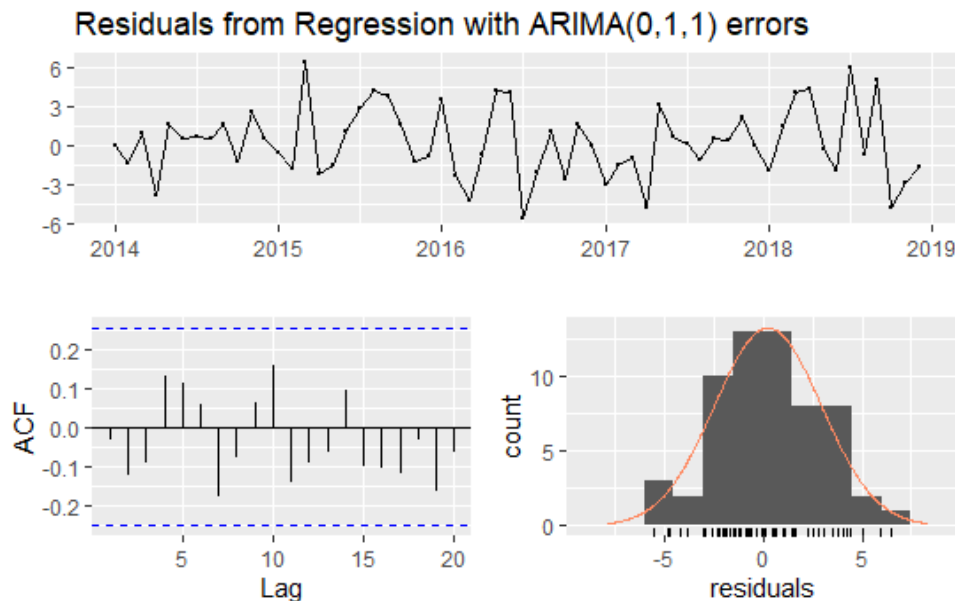
Figura 22 – Gráfico da ACF sobre os resíduos para identificar se os pressupostos do modelo ARIMAX são atendidos no conjunto de treinamento Paraná



Fonte: Elaborado pelo autor, por meio do R *studio* com base no pacote *forecast*.

Os gráficos da função de autocorrelação ACF (Figuras 23 e 24) dos resíduos dos modelos ARIMAX (0,1,1) e ARIMAX (1,0,1) mostram que todas as autocorrelações estão dentro dos limites. Essa é a primeira evidência de que os resíduos são ruídos brancos.

Figura 23 – Gráfico da ACF sobre os resíduos para identificar se os pressupostos do modelo ARIMAX são atendidos no conjunto de treinamento Rio Grande do Sul

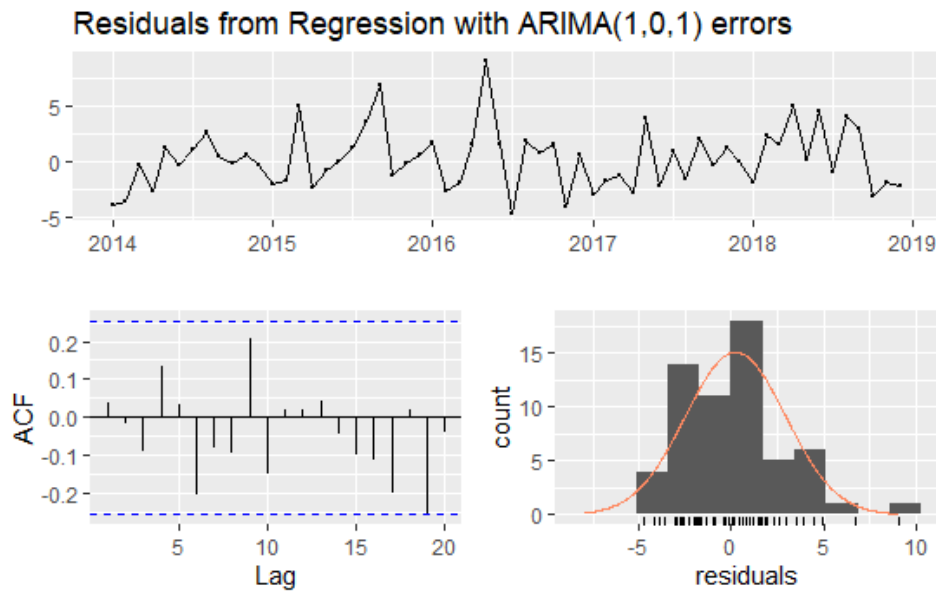


Fonte: Elaborado pelo autor por meio do R *studio* com base no pacote *forecast*.

A avaliação dos resíduos, feita por meio da diferença entre os preços e o ajuste feito pelo ARIMAX, é relevante para identificar os pressupostos do modelo ARIMAX. De acordo com os pressupostos da abordagem ARIMAX, os resíduos do modelo deve se comportar como ruído branco (HYNDMAN; ATHANASOPOULOS, 2018a; SHUMWAY; STOFFE, 2010).

Assim, é importante realizar os testes sobre os resíduos para identificar se os pressupostos dos modelos ARIMAX são atendidos. Observe que cada autocorrelação está próxima de zero (Figuras 23 e 24). Além disso, nos gráficos da ACF os picos estão dentro dos limites (as linhas tracejadas nas cores azuis).

Figura 24 – Gráfico da ACF sobre os resíduos para identificar se os pressupostos do modelo ARIMAX são atendidos no conjunto de treinamento Mato Grosso



Fonte: Elaborado pelo autor, por meio do R *studio* com base no pacote *forecast*.

O teste de Ljung e Box (1978), é usado para testar a qualidade do ajuste de um modelo de série temporal. Para confirmar que os resíduos são ruído branco, utiliza-se o teste de hipótese proposto por Ljung e Box (1978), teste esse que leva o nome dos autores, e implementado no pacote *forecast* do (HYNDMAN et al., 2020). Proceda-se a realização do teste Ljung-Box, no qual a interpretação ocorre mediante as seguintes hipóteses:

H_0 : Os resíduos são ruído branco.

H_1 : Os resíduos não são ruído branco.

Tomar decisão: se o p-valor for maior do que 0,05, ou seja, $p > 0,05$ (não rejeita H_0).

A Tabela 4 apresenta os resultados dos testes Ljung-Box, que serviram para confirmar os ruídos brancos dos métodos ARIMAX.

Tabela 4 - Teste de hipótese Ljung-Box

ARIMAX	Q*	DF	Modelo DF	Total de lags	Ljung-Box (p-valor)
(2, 0, 1)(1, 0, 0)[12]	22.877	14	5	19	0.06229
(0, 1, 1)	10.736	10	2	12	0.3785
(1, 0, 1)	10,688	8	4	12	0.22

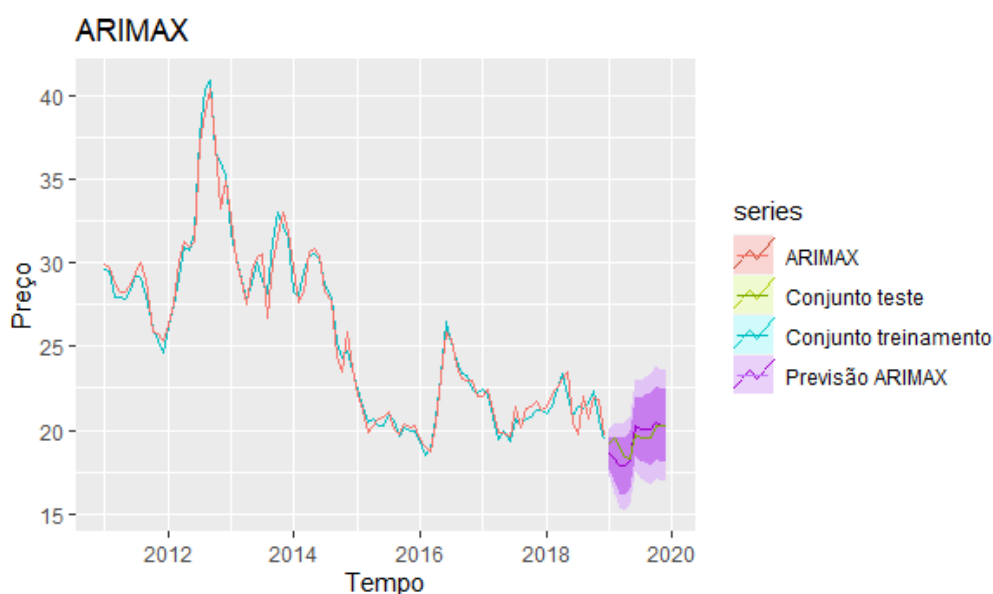
Fonte: Elaborado pelo autor por meio do pacote *forecast* com base na função *checkresiduals*.

Os p-valores foram: 0.06229, 0.3785 e 0.22, novamente indicando que os resíduos dos modelos ARIMAX podem ser considerados ruído branco, portanto, não rejeita H_0 . Os resíduos são ruídos brancos, atendendo os pressupostos do modelo ARIMAX. Em que a

hipótese nula, não é rejeitada para os resíduos nos três modelos ARIMAX estimados para cada séries de preços da soja.

Posteriormente, foi realizada a previsão do conjunto de teste, que corresponde ao período entre janeiro a dezembro de 2019. A Figura 25 apresenta as previsões em que o modelo ARIMAX (2,0,1)(1,0,0)[12] se ajustou bem aos preços dentro da amostra e previu igualmente fora da amostra.

Figura 25 - Previsões ARIMAX do conjunto de teste Paraná de janeiro de 2019 até dezembro de 2019

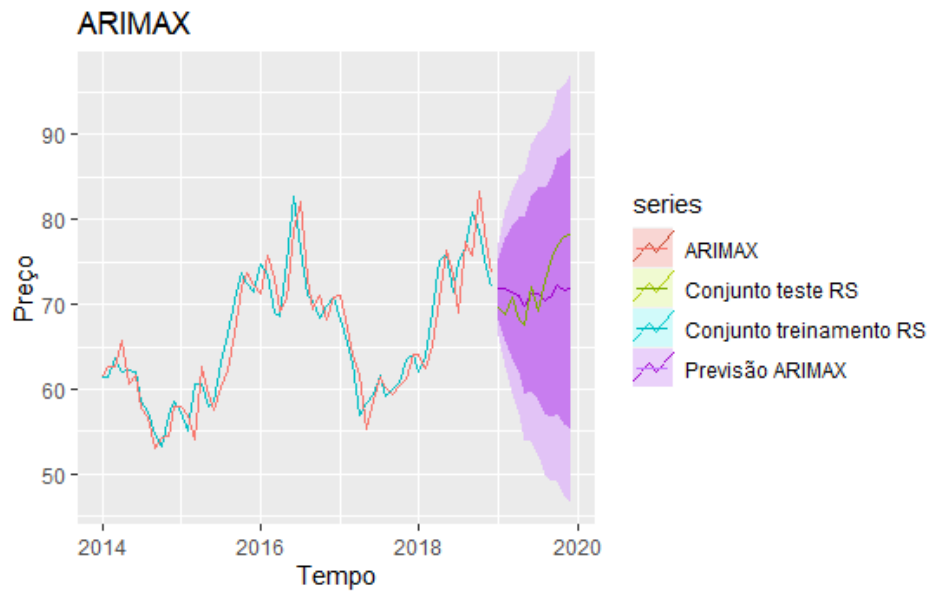


Fonte: Elaborado pelo autor, por meio do R studio com base nos pacotes *forecast* e *ggplot2*.

Em vermelho pode-se visualizar o comportamento (ajustes) dos modelos ARIMAX nos conjuntos de treinamento; a cor azul representa os conjuntos de treinamento; em verde são os conjuntos de teste; e as previsões de janeiro até dezembro de 2019 estão em roxo.

Os resultados de previsão mostram que o modelo ARIMAX (2,0,1)(1,0,0)[12] pode prever o preço da soja do Paraná com MAPE 2,62%, que mostra alta precisão. Este primeiro resultado indica que as variáveis apresentadas pelo modelo teórico para determinar a formação de preço são viáveis para gerar previsões via ARIMAX, devido à acurácia que o modelo analítico gerou usando as variáveis selecionadas.

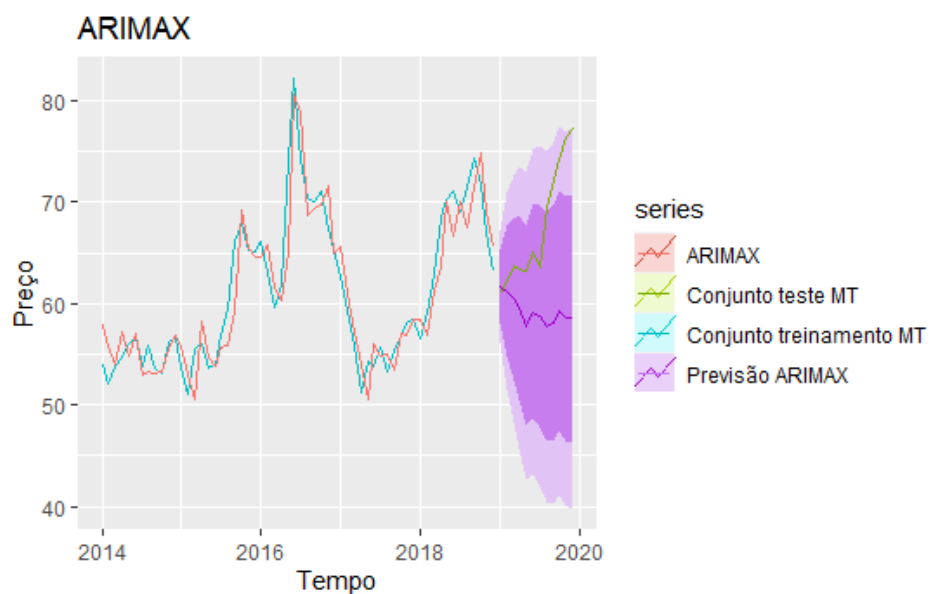
Figura 26 - Previsões ARIMAX do conjunto de teste do Rio Grande do Sul de janeiro de 2019 até dezembro de 2019



Fonte: Elaborado pelo autor por meio do R *studio*, com base nos pacotes *forecast* e *ggplot2*.

Com as variáveis indicadas pelo modelo teórico da lei do preço único Mundlak e Larson (1992), o modelo analítico, mostrou bom ajuste aos preços dentro da amostra e previu igualmente o período do conjunto de teste. O modelo ARIMAX (0,1,1) na previsão foi capaz de capturar o comportamento do preço para o conjunto de teste, conforme indicado pela Figura 26.

Figura 27 - Previsões ARIMAX do conjunto de teste do Mato Grosso, de janeiro a dezembro de 2019



Fonte: Elaborado pelo autor por meio do R *studio* com base nos pacotes *forecast* e *ggplot2*.

Os resultados de predição mostram que o método ARIMAX (1,0,1) para fora da amostra do Mato Grosso não foi preciso com MAPE de 11,92%. Foi possível observar uma diferença grande entre a previsão e o realizado no período, assim foi a previsão do ARIMAX com pior desempenho de precisão.

4.2.2 Treinamento e teste dos modelos de redes neurais (MLP) e máquina de aprendizado extremo (ELM)

A modelagem e previsão via métodos desenvolvidos por Crone e Kourentzes (2010), Kourentzes, Barrow e Crone (2014) e publicados no pacote *nnfor* Kourentzes (2019) está amplamente adotado em vários campos de estudo, por exemplo, na previsão do preço das *commodities* de energia Herrera et al., (2019); produção de petróleo Porto e Philippi (2018a); (2018b); demanda turística Porto, Philippi e Leite (2019); preço da soja Puchalsky et al. (2018); consumo de água Karamaziotis et al., (2020) e previsão de demanda industrial Kück e Freitag (2021).

As séries de preços não foram submetidas aos testes de estacionariedade para a previsão dos métodos de inteligência artificial (IA), pois as funções *mlp* e *elm* do pacote (*nnfor*) fazem a transformação automática de vetor de entrada de uma rede neural. A metodologia empregada é um filtro neural interativo baseado na combinação da estimativa da distância euclidiana e MLPs. Os detalhes sobre o funcionamento da metodologia de transformação automática de séries temporais podem ser vistos na pesquisa (CRONE; KOURENTZES, 2010).

Os modelos de IA são treinados no conjunto de treinamento para capturar os padrões dos preços da soja. Os parâmetros das redes neurais (MLP) e a máquina de aprendizado (ELM) foram gerados onde: nós ocultos é o número de nós na camada escondida; repetições é o número de treinamento da MLP e ELM; e inputs é o número de entradas da arquitetura da rede neural (KOURENTZES, 2019). Todos os parâmetros são ilustrados na Tabela 5.

Tabela 5 - Parâmetros das redes neurais e máquina de aprendizagem extrema para o conjunto de treinamento

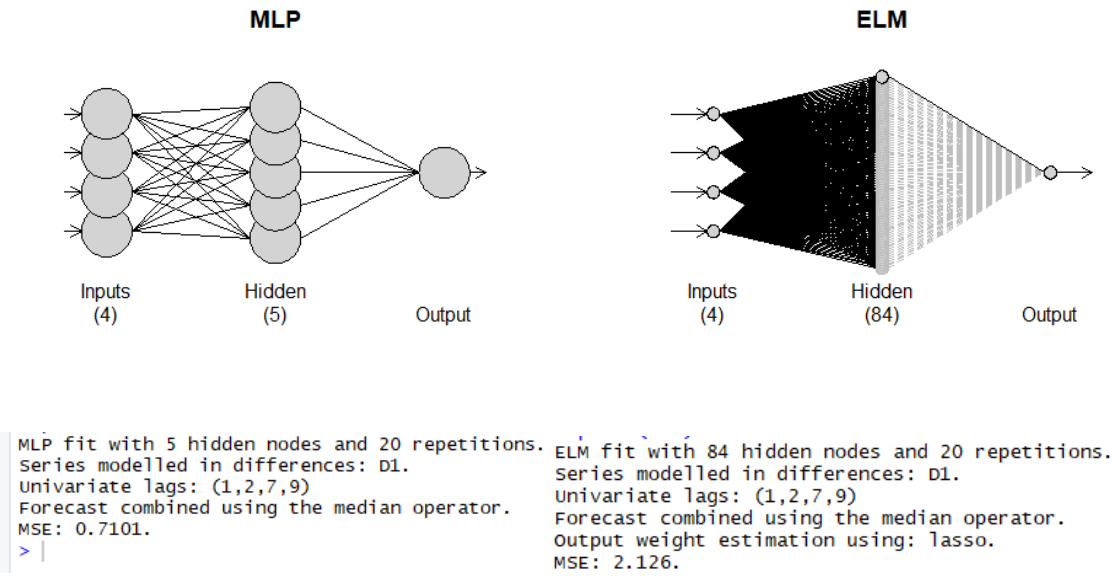
Modelos	Série Temporal	Nós ocultos	Repetições	Raios univariados	Inputs
MLP	Preço da soja PR	5	20	(1, 2, 7, 9)	4
ELM	Preço da soja PR	84	20	(1, 2, 7, 9)	4
MLP	Preço da soja RS	5	20	(1,2,12)	3
ELM	Preço da soja RS	45	20	(1,2,12)	3
MLP	Preço da soja MT	5	20	(1,2,10,12)	4

ELM	Preço da soja MT	45	20	(1,2,10,12)	4
-----	------------------	----	----	-------------	---

Fonte: Elaborado pelo autor por meio do R *studio* com base no pacote *nnfor*.

As arquiteturas de ambos os modelos são mostradas. Essas respectivas arquiteturas apresentam os parâmetros que foram selecionados para treinar os algoritmos utilizando os conjuntos de treinamento. As variáveis de entrada (*inputs*) foram os preços da soja com objetivo de prever o preço da variável resposta (*outputs*). Os resultados de treinamento das arquiteturas da MLP e ELM estão representados nas Figuras 28, 29 e 30. O *perceptron* multicamada (MLP) funciona de forma semelhante aos neurônios do cérebro humano. Assim, a arquitetura da rede neural é composta por camadas.

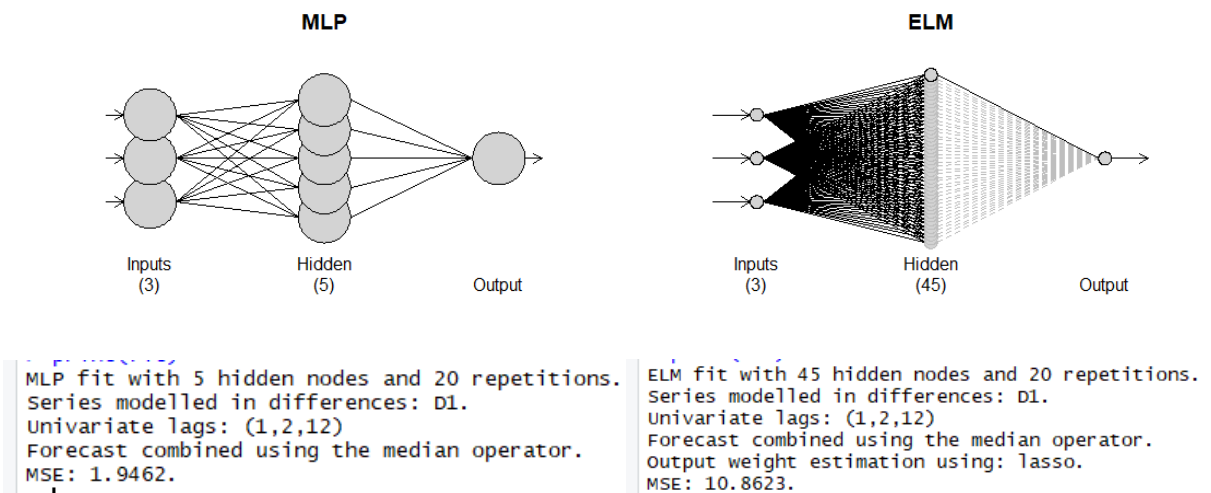
Figura 28 - Arquiteturas da MLP e ELM com base no conjunto de treinamento Paraná



Fonte: Elaborado pelo autor por meio do R *studio* e com base nos preços (CEPEA, 2020a).

Observa-se na Figura 28 que a MLP apresentou um número menor de nós ocultos em sua estrutura, enquanto a ELM teve uma estrutura um pouco mais complexa com maior número de nós na camada escondida. Na primeira arquitetura (MLP) a estrutura foi mais simples em comparação à segunda.

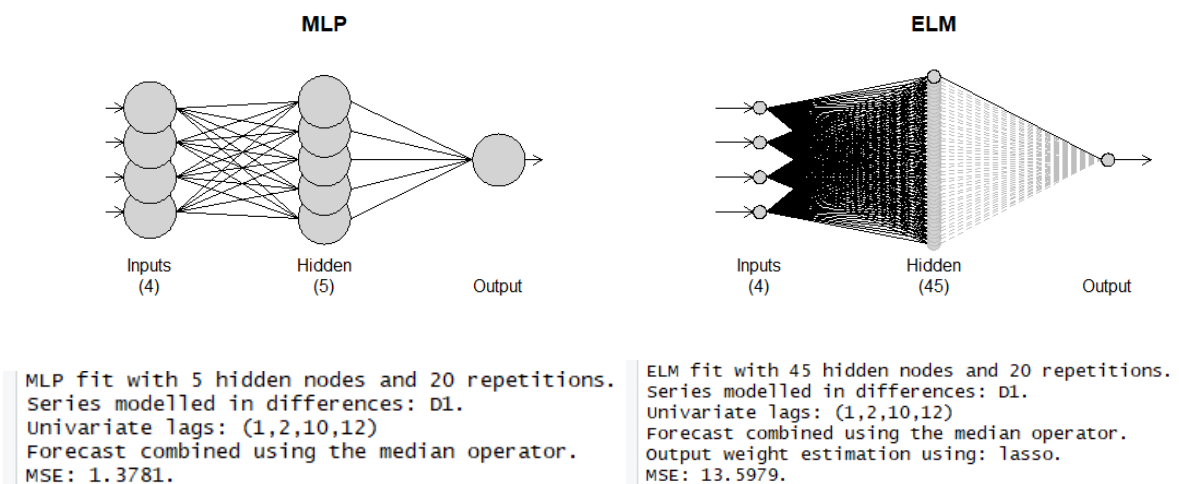
Figura 29 - Arquiteturas da MLP e ELM com base no conjunto de treinamento Rio Grande do Sul



Fonte: Elaborado pelo autor por meio do R *studio* e com base nos preços da CONAB.

A mesma interpretação vale para as arquiteturas da MLP e ELM nos conjuntos de treinamento do Rio Grande do Sul e Mato Grosso. As arquiteturas mostram mais números de nós na camada escondida para a máquina de aprendizado extremo (ELM).

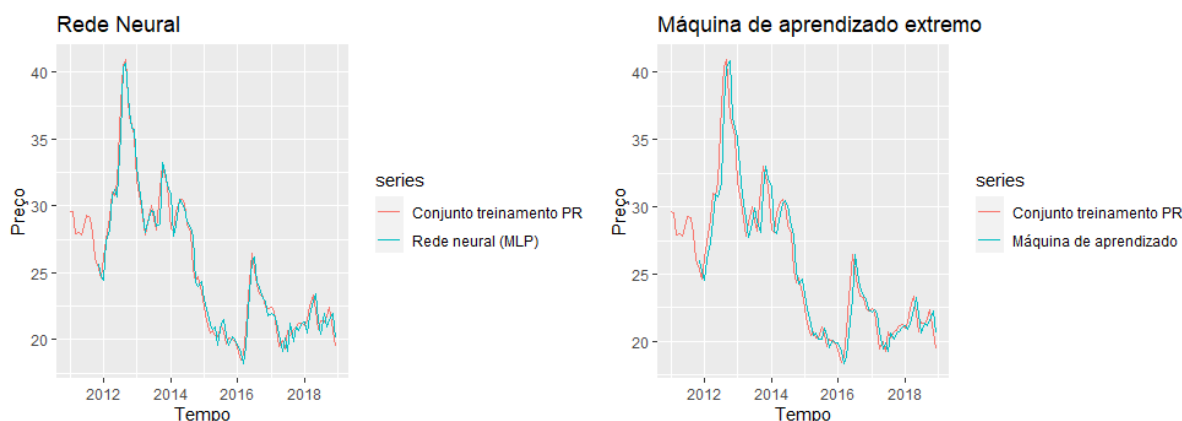
Figura 30 - Arquiteturas da MLP e ELM com base no conjunto de treinamento Mato Grosso



Fonte: Elaborado pelo autor por meio do R *studio* e com base nos preços da CONAB.

Os conjuntos de treinamento foram usados para todas as séries de preços, a fim de ajustar as redes neurais e a máquina de aprendizado. Observa-se como os modelos aprenderam com os preços e acompanham o seu comportamento de forma precisa, mostrando um ajuste que pode ser observado os gráficos das Figuras 31 a 33.

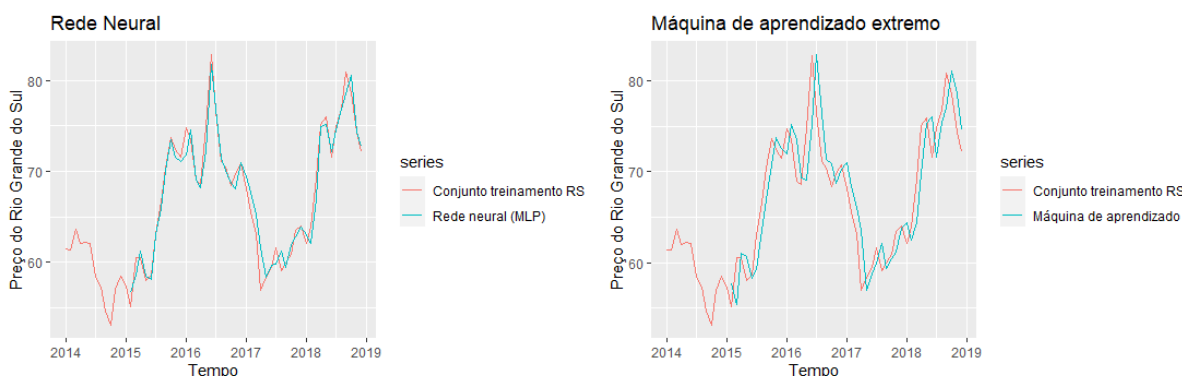
Figura 31 - Treinamento MLP e ELM dos preços CEPEA de janeiro de 2011 até dezembro de 2018



Fonte: Elaborado pelo autor por meio do *R studio* e com base nos preços (CEPEA, 2020a).

Na fase de treinamento dos dois modelos, a rede neural aprendeu melhor com os preços da soja (Figuras 31 e 33). Pode-se observar que os ajustes do modelo de rede neural aos dados foram mais precisos em comparação aos ajustes da abordagem de máquina de aprendizado (ELM).

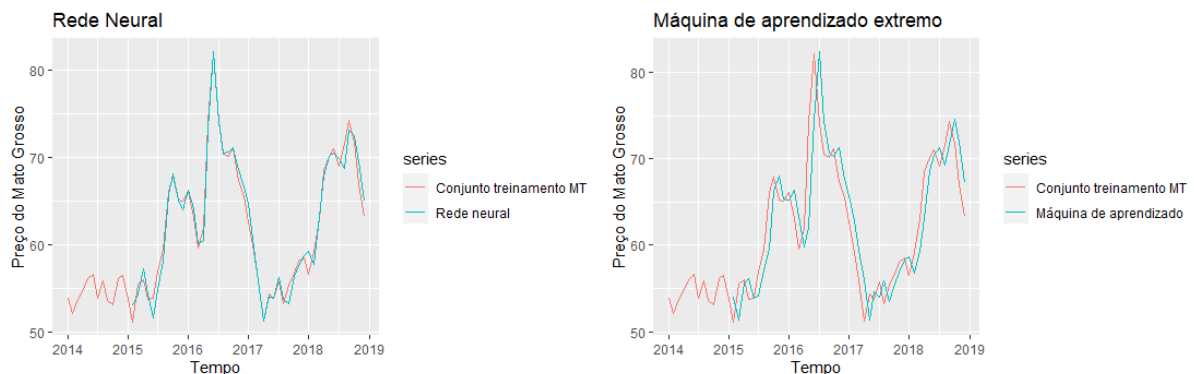
Figura 32 - Treinamento MLP e ELM dos preços da Conab do Rio Grande do Sul, de janeiro de 2014 até dezembro de 2018



Fonte: Elaborado pelo autor por meio do *R studio* com base nos preços preços da CONAB.

Avaliando os indicadores de desempenho para a fase de treinamento dos modelos, é evidente que, entre os métodos de inteligência artificial, o melhor ajuste aos preços da soja foi da abordagem MLP com MAPEs de 2,51%, 1,41% e 1,32%, enquanto a máquina de aprendizado estimou os preços da soja com MAPEs de 4,10%, 4,48% e 5,05%.

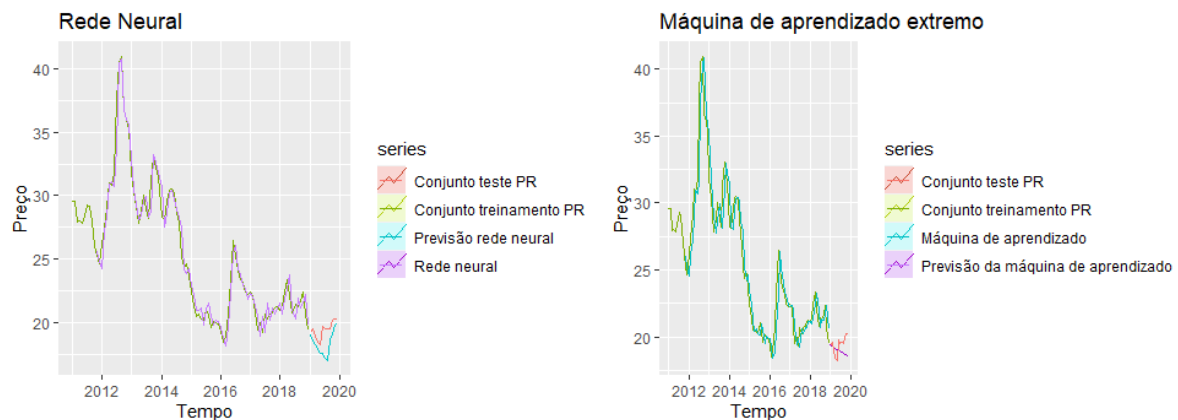
Figura 33 - Treinamento MLP e ELM dos preços da Conab do Mato Grosso de janeiro de 2014 até dezembro de 2018



Fonte: Elaborado pelo autor por meio do R *studio* com base nos preços da CONAB.

Por fim, as Figuras 34 a 35 comparam os preços dados reais de 2019 com previsões de 12 passos à frente das abordagens MLP e ELM para cada preço da soja nos conjuntos de treinamento e teste.

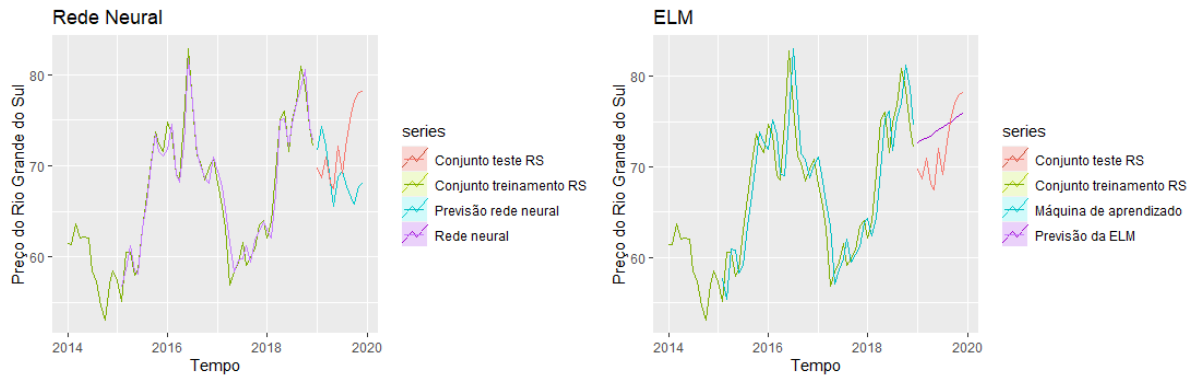
Figura 34 - Previsões MLP e ELM dos preços do Paraná de janeiro a dezembro de 2019



Fonte: Elaborado pelo autor por meio do R *studio* com base no pacote nnfor.

Posteriormente, foi realizada a previsão do conjunto de teste, que corresponde ao período entre janeiro a dezembro de 2019. A Figura 34 apresenta as previsões em que o modelo de rede neural se ajustou bem aos preços do Paraná dentro e fora da amostra. A previsão do modelo ELM não foi muito precisa com MAPE de 4,18%.

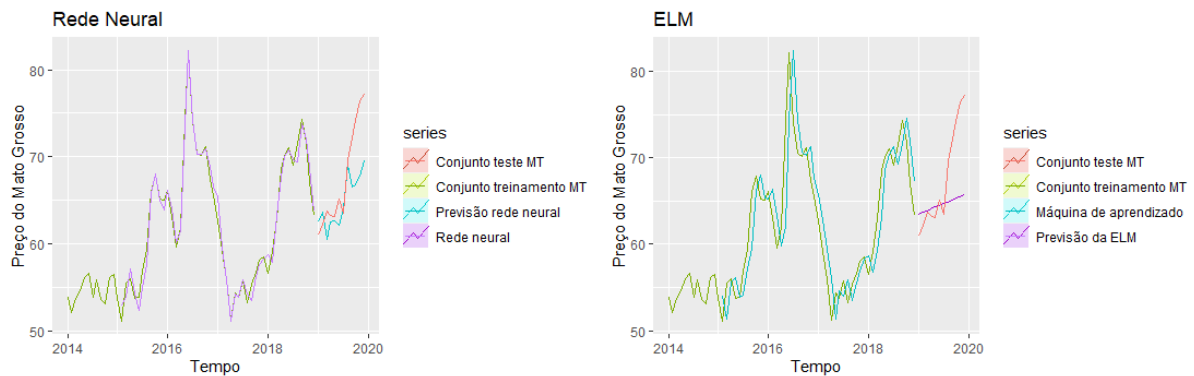
Figura 35 - Previsões MLP e ELM dos preços do Rio Grande do Sul de janeiro de 2019 até dezembro de 2019



Fonte: Elaborado pelo autor por meio do R *studio* com base no pacote nnfor.

Para dentro e fora das amostras (Figuras 35 e 36), o modelo de rede neural mostra um melhor ajuste do que a máquina de aprendizado, mostrando que nestes casos pode ser uma opção de apoio, mais precisa para escolher nos processos decisórios de agentes do mercado agrícola.

Figura 36 - Previsões MLP e ELM dos preços do Mato Grosso de janeiro de 2019 até dezembro de 2019



Fonte: Elaborado pelo autor por meio do R *studio* com base no pacote nnfor.

Os gráficos mostram que a rede neural é superior, comparada a máquina de aprendizado para preços de soja. Em termos de MAPE e RMSE, o modelo de rede neural teve o desempenho de previsão 2,86%, 2,49, 4,79% e 4,72 em relação à máquina de aprendizado extremo 4,31%, 3,46, 5,85% e 5,87 nos conjuntos de teste.

4.3 Discussão dos resultados (Comparações empíricas de previsões)

Esta seção tem por objetivo avaliar os resultados empíricos e a análise do desempenho de previsão dos modelos testados. Foram geradas as métricas de erro dos modelos para todas as séries de preços da soja. As métricas de desempenho apresentam a precisão dos métodos em comparação com os dados realizados.

Para validação das abordagens de predição do ARIMAX, MLP e ELM são gerados os erros em termos de indicadores de desempenho dos métodos. Nesta fase de validação, os achados são mostrados nas Tabelas 6, 7 e 8, com os resultados de melhor desempenho relatados em negrito. Os desempenhos de previsão são avaliados em termos de quatro medidas de erro de previsão.

Os critérios de avaliação ME, RMSE, MAE e MAPE, entre outros, apresentam a acurácia do ARIMAX, MLP e ELM em comparação com os preços realizados. Uma característica na Tabela 6, é que as previsões do modelo ARIMAX foram precisas, pois os valores são baixos. Outra informação importante extraída dos indicadores de desempenho foi que as previsões apresentaram um MAPE inferior a 10%.

De acordo com Lewis (1982) o percentual abaixo de 10% pode ser interpretado pela classificação do autor como previsão altamente precisa. Com base nas Tabelas 6 e 7, pode-se afirmar que não teve diferenças significativas entre os erros do ARIMAX na fase de treinamento em comparação aos erros da fase de teste. Porém, a mesma interpretação não ocorre para os conjuntos de treinamento e teste do Mato Grosso (Tabela 8).

Tabela 6 - Indicadores de desempenho dos modelos ARIMAX, MLP e ELM do conjunto de teste Paraná

Modelos	Conjuntos	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil
ARIMAX	Treinamento	-0,03	0,69	0,52	-0,23	2,04	0,12	0,06	NA
ARIMAX	Teste	0,11	0,63	0,50	0,60	2,62	0,12	0,81	1,07
MLP	Treinamento	0,007	0,89	0,63	0,07	2,51	0,60	-0,3	NA
MLP	Teste	1,25	1,44	1,25	6,36	6,36	1,20	NA	NA
ELM	Treinamento	8,67	1,45	1,06	0,04	4,10	0,23	0,37	NA
ELM	Teste	5,02	0,97	0,82	2,44	4,18	0,19	0,69	1,68

Nota: Os menores valores de todas as métricas dos conjuntos de treinamento e teste estão em negrito.

Fonte: Elaborado pelo autor, por meio do R studio com base no pacote *forecast* e função *accuracy*

Na modelagem e previsão dos conjuntos treino e teste do Paraná, o método ARIMAX apresentou o menor RMSE (0,63% e 0,69%), enquanto a ELM obteve o maior valor RMSE (1,45%) no conjunto de treinamento. O ARIMAX demonstrou um melhor ajuste

para dentro e fora da amostra na previsão do preço da soja e teve melhor capacidade preditiva fora da amostra para os preços da soja.

Este resultado mostra que a Lei do Preço Único é válida no mercado brasileiro de soja, pois com as variáveis de preços internacionais e domésticos, indicadas pela teoria do preço único, a abordagem ARIMAX previu o preço da soja com precisão.

O modelo que melhor previu o preço foi o ARIMAX, pois em quase todos os indicadores gerou valores menores, mostrando ser um modelo de previsão superior com relação ao preço do Paraná. Os estudos de Ahumada e Cornejo (2016a), (2016b); Ribeiro, Sosnoski e Oliveira (2010) alcançaram MAPEs de 6,69%, 8,15% e 5,77%, respectivamente, no curto horizonte de previsão, enquanto o modelo ARIMAX do presente estudo gerou em sua previsão o menor MAPE de 2,62%, conforme pode ser visto na Tabela 6.

Na comparação entre os modelos de previsão para os preços no Rio Grande do Sul não houve diferenças de erro significativas em termos de medidas de erro, no entanto, os erros da rede neural (MLP) foi o modelo que gerou os menores valores. De modo geral, a rede neural da tipologia *perceptron* multicamada (MLP) teve o melhor desempenho de previsão, neste caso.

Tabela 7 - Indicadores de desempenho das previsões do conjunto de teste Rio Grande do Sul

Modelos	Conjuntos	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil
ARIMAX	Treinamento	0,24	2,69	2,14	0,34	3,18	0,20	-0,03	NA
ARIMAX	Teste	1,00	3,68	3,19	1,14	4,32	0,30	0,67	1,42
MLP	Treinamento	-0,02	1,22	0,92	-0,001	1,41	0,09	0,03	NA
MLP	Teste	-0,32	2,49	2,06	-0,57	2,86	0,19	0,15	0,98
ELM	Treinamento	6,04	3,38	2,80	0,05	4,48	0,28	0,22	NA
ELM	Teste	-1,88	3,46	3,04	-2,81	4,31	0,29	0,59	1,39

Nota: Os menores valores de todas as métricas dos conjuntos de treinamento e teste estão em negrito.

Fonte: Elaborado pelo autor, por meio do R studio com base no pacote *forecast* e função *accuracy*

A MLP gerou as melhores previsões para o preço da soja do Rio Grande do Sul e ajustou-se bem aos dados, mostrando ser uma alternativa no apoio às decisões dos produtores rurais sobre o que ocorrerá com os preços da soja no futuro.

Comparando com as previsões de curto prazo da pesquisa de Puchalsky et al. (2018), o algoritmo “otimização de enxame de vermes (GSO)” gerou a melhor previsão entre todos os modelos comparados no estudo, produzindo um MAPE de 0,587%. Os métodos validados na

dissertação alcançaram níveis de acurácia muito próximos dos achados empíricos (PUCHALSKY et al., 2018).

Nas previsões de preços do Mato Grosso, a rede neural (MLP) foi superior em comparação com ARIMAX; em termos de RMSE, MAPE e MAE as diferenças foram consideráveis. Não houve grandes diferenças nos indicadores de precisão entre MLP e ELM. Se considerar os indicadores de desempenho (Tabela 8), a rede neural (MLP) gerou valores menores em todas as métricas.

Tabela 8 - Indicadores de desempenho das previsões do conjunto de teste Mato Grosso

Modelos	Conjuntos	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil
ARIMAX	Treinamento	0,21	2,71	2,10	0,21	3,39	0,21	0,03	NA
ARIMAX	Teste	8,39	10,58	8,50	11,73	11,91	0,85	0,73	4,19
MLP	Treinamento	-0,00	1,05	0,78	0,02	1,32	0,08	0,09	NA
MLP	Teste	2,84	4,52	3,42	3,84	4,79	0,34	0,59	1,76
ELM	Treinamento	-1,96	3,68	2,91	0,06	5,05	0,31	0,29	NA
ELM	Teste	3,03	5,87	4,24	3,91	5,85	0,42	0,73	2,25

Nota: Os menores valores de todas as métricas dos conjuntos de treinamento e teste estão em negrito.

Fonte: Elaborado pelo autor, por meio do R studio com base no pacote *forecast* e função *accuracy*

A previsão do modelo de rede neural gerou erros inferiores, comparando a acurácia do método testado no preço da soja do Mato Grosso com os estudos de Ferreira et al. (2011), Kitworawut e Rungreunganun, (2019); e Souza et al., (2018) onde realizaram previsões de longo prazo e a acuracidade das previsões geraram MAPEs de 4,86%, 12,10%, e 4,31%, nesta ordem.

No entanto, o desempenho de previsão dos modelos empregados na dissertação não gerou previsões de longo prazo mais precisas do que o estudo de Lima et al. (2010), em que as redes neurais recorrentes com o uso de ondaletas geraram um MAPE de 1,15%. Essa acurácia foi para previsão de dez passos à frente.

Em todos os casos, as três abordagens de previsão apresentaram bom desempenho com MAPE abaixo de 10% nas três séries de preços da soja modeladas. Os modelos de previsão foram comparados, sendo possível selecionar o modelo com melhor desempenho de precisão para cada caso: preços do Paraná, Rio Grande do Sul e Mato Grosso.

Os resultados da previsão mostraram quais modelos foram mais precisos. Este desdobramento indica quais métodos servirá de apoio à tomada de decisão na agricultura e no planejamento da produção agrícola. Esta seção avaliou as comparações entre modelos

ARIMAX, MLP e ELM, por meio de medidas de acurácia. Estas comparações serviram para validação dos modelos como capazes de fornecerem precisões nas previsões de preços da soja.

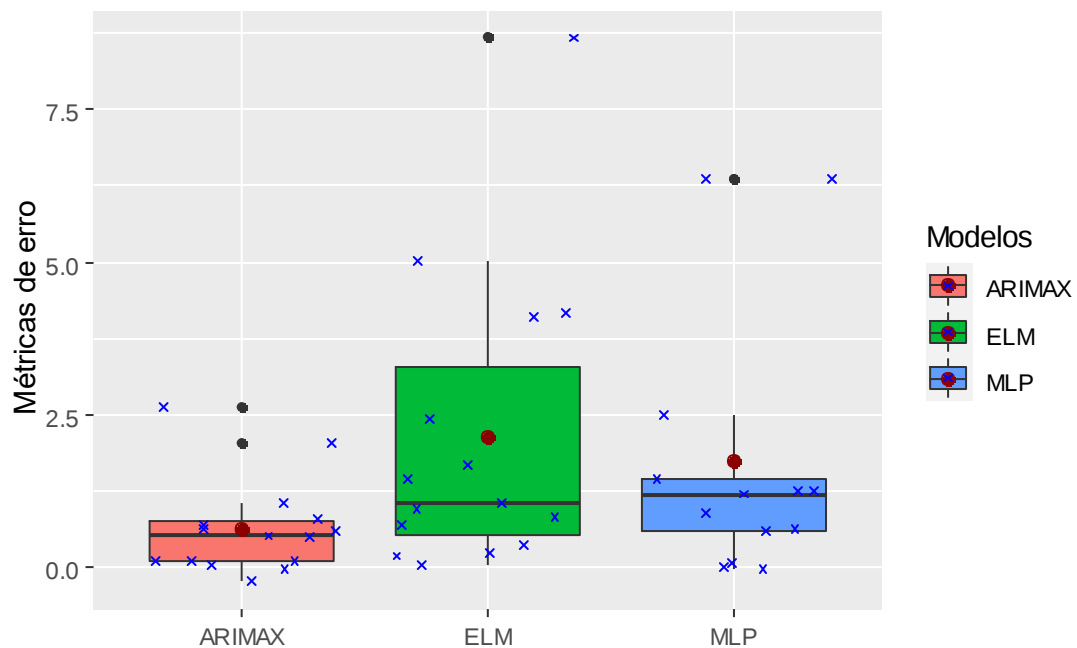
4.3.1 Boxplots e testes de variâncias de erros das abordagens de previsão

Nesta seção são avaliados erros dos modelos nos gráficos de boxplots e comparados através de testes de variâncias para responder as questões norteadoras. Uma hipótese foi levantada como base nos resultados das comparações de testes de variâncias. Um gráfico de boxplot é realizado com base nos quartis, sendo uma opção ao histograma; tem como finalidade resumir a distribuição dos dados (MORETTIN; SINGER, 2020). Também é utilizado para comparar vários conjuntos de dados (MORETTIN; SINGER, 2020).

Na Figura 37 é apresentado o Boxplot das abordagens de previsão para o preço da soja do Paraná. A Figura mostra que o comportamento dos erros de modelos MLP e ELM são muito semelhantes para os preços do Paraná. Essa característica pode ser observada comparando os limites da amplitude interquadrática.

Para o conjunto de teste do Paraná a mediana do ARIMAX foi inferior, em comparação aos modelos de inteligência artificial. Os resultados da Figura 37, indicam que o modelo ARIMAX geram as previsões mais precisas dos três modelos. Além disso, a dispersão dos erros do ARIMAX também foi menor mostrando ser superior para prever o preço da soja do Paraná.

Figura 37 - Boxplot dos indicadores de desempenho para o conjunto de teste Paraná



Fonte: Elaborado pelo autor, por meio do R *studio* com base no pacote *ggplot2*.

A Figura 37 apresenta a comparação das métricas de erro entre os modelos estimados a partir de gráfico boxplot, as medianas identificadas pelas linhas que divide as caixas, são os pontos de análise e divide os grupos inferiores e superiores dos resultados das métricas dos erros. A mediana do grupo de dados das métricas de erro do modelo ARIMAX obteve a menor variabilidade e com isso ele apresenta evidências de melhor resultado entre os modelos estimados.

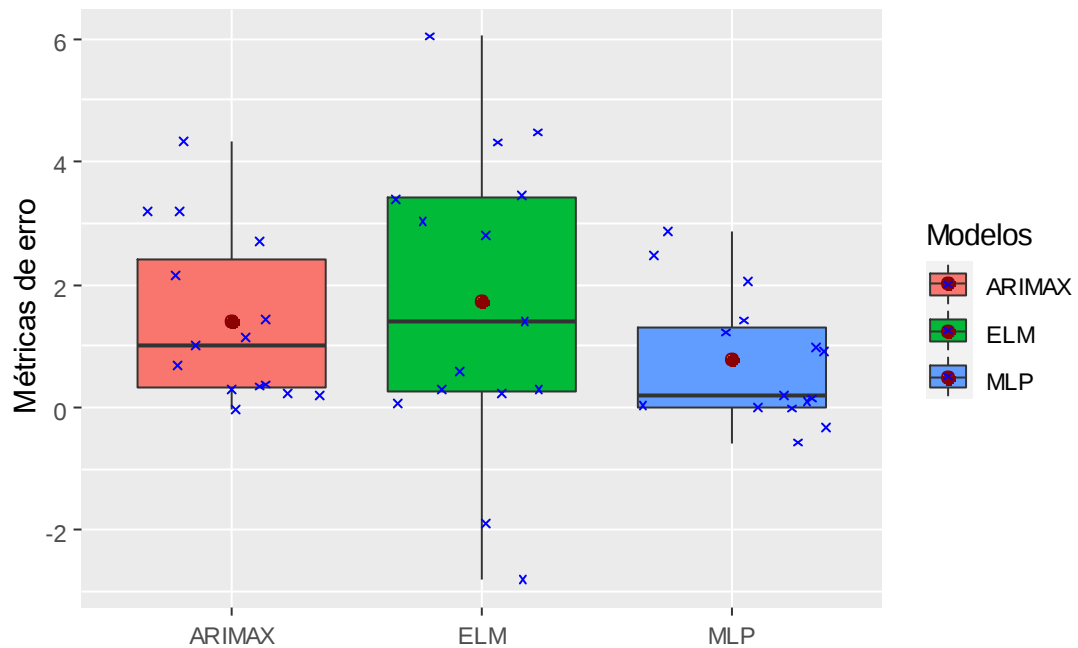
Esse gráfico permite identificar que a posição de 50% dos erros de previsão do modelo ARIMAX estão dentro de 0 e 1 (entre o primeiro e terceiro quartis). Os erros dos três modelos tiveram *outliers* superiores. Os resultados mostram que a previsão com ARIMAX é mais precisa dentre todos os métodos aplicados no preço do Paraná. Pode-se constatar que na média os erros do modelo ARIMAX mostraram menor variabilidade do que as duas abordagens de IA.

Através da análise dos gráficos de Box Plot (Figuras 38 e 39) foi possível verificar a distribuição dos erros de cada modelo de previsão. Para comparar as dispersões e avaliar se foram significativas as diferenças entre os erros dos três modelos, foi elaborado o gráfico Box Plot com os erros dos três modelos do conjunto de teste do Rio Grande do Sul (Figura 38).

Para o conjunto de teste do Rio Grande do Sul a menor variabilidade dos erros foi obtida pela rede neural (MLP), sendo que a mediana dos erros da MLP é menor do que o

limite inferior dos erros do modelo ARIMAX. Além disso, o valor da mediana dos erros de previsão é próximo de 0%, o que representa a tendência simétrica das previsões (Figura 38).

Figura 38 - Boxplot dos indicadores de desempenho para o conjunto de teste Rio Grande do Sul

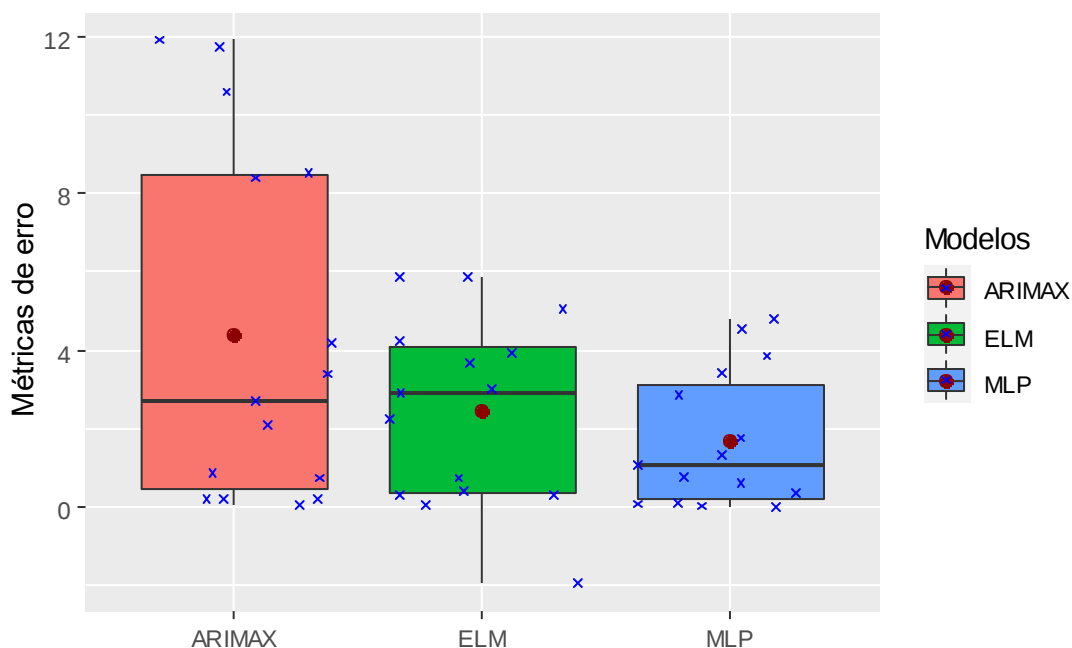


Fonte: Elaborado pelo autor, por meio do R studio com base no pacote *ggplot2*.

A maior variabilidade de erros foi da máquina de aprendizado. Nenhum dos métodos gerou erros atípicos (*outliers* superiores ou inferiores) e 50% dos erros de previsão do modelo de rede neural (MLP) estão entre 0 e 2. Portanto, as redes neurais tiveram o melhor desempenho de previsão em comparação aos modelos para os preços do Rio Grande do Sul, indicando que a rede neural é a abordagem mais precisa para prever os preços do Rio Grande do Sul.

Na previsão do conjunto de teste do Mato Grosso, a maior variabilidade dos erros de previsão foi do ARIMAX (Figura 39), enquanto os erros da máquina de aprendizado não tiveram muita variabilidade. A rede neural novamente gerou a previsão mais precisa entre os métodos empregados. Essa característica pode ser observada comparando a mediana da MLP, sendo a menor em relação às abordagens concorrentes.

Figura 39 - Boxplot dos indicadores de desempenho para o conjunto de teste Mato Grosso



Fonte: Elaborado pelo autor por meio do R *studio* com base no pacote *ggplot2*.

Neste caso, o modelo ARIMAX apresentou os erros com a maior variabilidade, mostrando ser a modelagem menos precisa para os preços de Mato Grosso. Entre os métodos de previsão, por outro lado, a rede neural foi o modelo que obteve a menor variabilidade dos erros de previsão (Figura 39). Os indicadores de previsão para os preços do Mato Grosso mostram que a rede neural gerou o melhor desempenho da previsão em comparação a ELM e ARIMAX. Os resultados mostram que MLP e ELM apresentam melhor desempenho quando comparados ao modelo ARIMAX (Figura 39).

O estudo avalia a acuracidade dos modelos de previsão considerando oito métricas de erros o que reforça o rigor metodológico da modelagem realizada, para as constatações relacionadas à precisão dos métodos. As previsões de preços da soja de estudos anteriores Ahumada e Cornejo (2016b); Drachal (2019); Puchalsky et al., (2018); e Zhang et al., (2018), avaliaram duas ou no máximo três medidas de erro para analisar a acurácia de abordagens da predição. Está foi à primeira vantagem da presente dissertação em relação as modelagens de trabalhos relacionados.

Por meio da análise gráfica Boxplots de erros tem-se evidências de quais foram os melhores desempenhos de previsão em cada conjunto de teste, mas para não ter dúvidas se existe ou não diferenças estatísticas significativas entre a previsão das abordagens, é preciso comparar as variâncias dos erros de cada modelo para os conjuntos.

O objetivo do teste de variância foi verificar se tem diferença significativa nos erros de previsão de cada modelo. Para tanto, foi feita a comparação dos erros do ARIMAX em relação a MLP, MLP em relação a ELM e por fim ELM e ARIMAX, sendo todas as combinações possíveis para cada série de preços. Para testar se as diferenças entre os erros dos modelos são estatisticamente significativas foi empregado um teste de variância.

A análise dos erros dos modelos foi conduzida usando a função em R: *var.test* do pacote *stats* (R Core Team, 2020), para checar se tem diferenças significativas entre as variâncias de erros dos modelos.

H_0 : As amostras não são iguais.

H_1 : As amostras são iguais.

Tomar decisão: se o p-valor for menor do que 0,05, ou seja, $p < 0,05$ (não rejeita H_0).

Os resultados dos testes de variância dos modelos foram gerados através da linguagem de programação R. Este procedimento é implementado no pacote *stats* (R Core Team, 2020) em R, na função *var.test*. Esta função usada para fazer um teste F^7 para diferenças na variância dos erros entre os três modelos de previsão. Os testes F podem testar as diferenças entre pares de populações. Os p-valores gerados pela função podem ser vistos na Tabela 9, onde são organizados estes resultados.

Tabela 9 - Variâncias entre as medidas de erro dos modelos ARIMAX, MLP e ELM

Modelos em comparação	Conjuntos de teste	Comparar duas variâncias (p-valor)	As diferenças foram significativas?
ARIMAX ≠ MLP	Paraná	0,001772	Sim
MLP = ELM	Paraná	0,1341	Não
ELM = ARIMAX	Paraná	0,07461	Não
ARIMAX = MLP	Rio Grande do Sul	0,2507	Não
MLP ≠ ELM	Rio Grande do Sul	0,009864	Sim
ELM = ARIMAX	Rio Grande do Sul	0,1318	Não
ARIMAX ≠ MLP	Mato Grosso	0,0008934	Sim
MLP = ELM	Mato Grosso	0,4143	Não
ELM ≠ ARIMAX	Mato Grosso	0,008848	Sim

Fonte: Elaborado pelo autor, por meio do R studio com base no pacote stats.

⁷ Os detalhes sobre o cálculo operacional para obter o teste T do teste de variância, podem ser encontrados em Pace (2012a) e (2012b).

Nota: Os resultados significativos ($p < 0,05$) estão em negrito. Os modelos que foram superiores, com diferenças significativas estão em negrito. Nível de significância de 5%.

Testes F de dois lados (usando a função *var.test* em R) foram aplicados para comparar a variabilidade dos erros de três modelos de previsão. O resultado foi considerado significativo em $p < 0,05$, em quatro comparações dos métodos. Nove testes foram realizados no total.

A partir do resultado dos testes de variância constata-se que teve diferenças estatísticas significativas entre os modelos comparados, em quatro situações: na comparação entre ARIMAX e MLP para os preços do Paraná; MLP e ELM para os preços do Rio Grande do Sul; ARIMAX e MLP para os preços do Mato Grosso e ELM e ARIMAX para os preços do Mato Grosso. Portanto, as previsões são superiores com base em métricas de erros, uma vez que são estatisticamente diferentes em termos de precisão com base nos testes de variâncias, podem ser consideradas previsões estatisticamente superiores.

As demais comparações não tiveram diferenças significativas ($p > 0,05$) entre as variâncias dos erros dos modelos de previsão. A Tabela 9 também mostra a rejeição da hipótese de que os níveis de acurácia entre ARIMA, MLP e ELM são os mesmos. Os valores na maior parte das comparações mostram que as previsões não são estatisticamente diferentes.

É importante ressaltar que, não foram observadas diferenças estatísticas significativas com relação a maior parte das comparações de variâncias dos erros realizadas entre os modelos de previsão. A Tabela 9 ilustra esses resultados. Os resultados observados a partir da Tabela 9 indicam a primeira hipótese levantada neste estudo.

Hipótese 1: A precisão de previsão de métodos *benchmarks*, como o ARIMAX, não estão muito longe da acurácia dos métodos de IA para os preços da soja.

A precisão das previsões dos modelos que obtiveram diferenças estatísticas significativas retoma a discussão de qual método empregar para os preços da soja, apesar de estudos anteriores Jha e Sinha (2012) e (2013); Kohzadi et al., (1996); Lima et al., (2010); e Ribeiro, Sosnoski e Oliveira (2010), afirmarem categoricamente que a rede neural é um modelo superior em comparação à abordagem ARIMA.

Neste estudo não foi constatado os mesmos resultados, pois em uma parte dos testes realizados, os modelos ARIMAX, MLP e ELM não tiveram diferenças estatísticas significativas, mostrando que o modelo ARIMAX não é inferior em relação a MLP e ELM,

exceto para os preços do Mato Grosso. Em que a rede neural, tende a produzir os melhores desempenhos, conforme mostrado na Figura 39.

A partir dos resultados de testes de variâncias, não rejeita-se a hipótese 1 levantada neste estudo, pois os testes de variâncias (a maior parte do total de testes), o resultado foi não significativo comprovando parcialmente a hipótese. No entanto, quatro comparações entre modelos de previsão, o resultado foi considerado significativo com $p < 0,05$. Embora uma das diferenças estatísticas significativas, mostrou que o ARIMAX teve desempenho superior em comparação às abordagens de IA. Os conjuntos (séries de preços da soja) tem um número limitado para confirmar a hipótese levantada, então é preciso realizar mais comparações entre modelos com número maior de preços de *commodities* agrícolas.

Este estudo propõe os modelos de previsão de preços da soja como alternativa para auxiliar a tomada de decisão dos agentes do mercado do grão, pois diante da acurácia dos modelos podem contribuir para minimizar os riscos de oscilações de preços.

Uma comparação dos modelos ARIMAX, MLP e ELM é feita com base em métricas de erro, boxplots e testes de variâncias que mostram os melhores métodos para prever os preços da soja do Paraná, Rio Grande do Sul e Mato Grosso.

As comparações realizadas nesta pesquisa são relevantes porque uma das características do mercado de soja é sua difícil previsibilidade dos preços. Os preços da soja sofrem alta volatilidade, o que dificulta a tomada de decisão agrícola. Os modelos podem ajudar os agentes de mercado para diminuir as incertezas com relação ao preço futuro da soja.

Aponta aos pesquisadores e participantes do mercado os métodos mais precisos na previsão de preço da oleaginosa; amplia a discussão sobre previsão de preços de *commodities* agrícolas, ao comparar abordagens de previsão que aumentam a popularidade do modelo mais preciso entre econometristas; e os modelos de previsão podem apoiar o processo de tomada de decisão dos produtores de soja na comercialização da *commodity*.

CAPÍTULO V

5.1 Principais desdobramentos da pesquisa

A análise bibliométrica, possibilitou identificar a lacuna de pesquisa e gerou resultados dos parâmetros quantitativos da produção científica e tendências sobre a temática. A primeira contribuição deste trabalho foi identificar os principais periódicos: Custos e Agronegócio online, Ciência Rural, *International Journal of Forecasting* e *Sustainability*.

Os pesquisadores destaque são Pinheiro, Senna, Jha, Sinha, Ahumada e Cornejo. Destacam-se o Brasil com o maior número de publicações e as universidades *China University of Geosciences*, *Huazhong Agricultural University*, Universidade do Estado da Bahia (UNEB) e Universidade Federal do Rio Grande do Norte (UFRN).

As abordagens ARIMA e redes neurais são os modelos mais utilizados na previsão de preços de *commodities* agrícolas. No entanto, o modelo ARIMA não tem gerado previsões superiores em comparação aos modelos de inteligência artificial (IA) e modelos híbridos. As redes neurais são mais precisas para prever preços de *commodities* agrícolas do que os modelos econométricos.

Os modelos híbridos de IA geram predições com melhores níveis de acuracidade em comparação aos métodos estatísticos tradicionais ARIMA, modelos individuais e redes neurais em que o desempenho de previsão dos modelos híbridos são melhores do que os dos modelos únicos.

Nenhum método de previsão é superior em termos de acuracidade em todos os horizontes de previsão. Portanto, de modo geral, os modelos híbridos de IA são mais precisos, mas não é verdade para todos os intervalos de previsão de curto, médio e longo prazos. É uma tendência a abordagem de modelos híbridos para prever preços de *commodities* agrícolas em pesquisas futuras.

A segunda contribuição advém da avaliação com base nos indicadores de desempenho de previsão, boxplots e variâncias de erros dos modelos em que ocorreram os seguintes desdobramentos. O modelo ARIMAX mostrou ter a melhor precisão da previsão do preço do Paraná para dentro e fora da amostra.

No caso dos preços do Rio Grande do Sul destacou-se a previsão da rede neural (MLP). Para os preços de Mato Grosso, a rede neural (MLP) e máquina de aprendizado (ELM) tiveram desempenho superior em comparação ao ARIMAX. Em todos os casos, as três abordagens de previsão apresentaram bom desempenho, com MAPE abaixo de 10% nas três

séries de preços da soja modeladas.

5.2 Considerações finais

Ao comparar abordagens de previsão, as comparações são geralmente feitas para descobrir o método mais preciso, mostrando que o método é superior com relação a um ou vários desses modelos, em comparação com os métodos existentes que aumentam a popularidade da abordagem mais precisa entre econométristas. O objetivo do estudo foi comparar o desempenho de previsão dos modelos.

Os métodos de previsão foram selecionados e empregados a partir dos resultados da revisão bibliométrica. Os resultados do estudo bibliométrico mostram a produção científica das publicações, periódicos e países mais produtivos, os principais autores, países e periódicos mais citados, palavras-chave mais frequentes, tópicos de tendência, nuvem de palavras, gráfico de três campos e rede de co-citação. Também descreveu as abordagens dos principais estudos, forneceu referências para os econométristas do campo da previsão de preços de *commodities* agrícolas e apontou as tendências sobre a temática.

Este estudo utilizou três fontes de preços da soja: Cepea, FMI e CONAB. Posteriormente, os preços são divididos em conjuntos de treinamento e teste para validar e comparar as precisões de modelos autorregressivos, integrados de médias móveis com variável exógena (ARIMAX), redes neurais perceptron multicamada (MLP) e máquina de aprendizagem extrema (ELM).

Neste estudo, três abordagens de previsão são avaliadas em três diferentes séries de preços da soja, correspondentes aos três principais estados produtores de soja no Brasil, a fim de comparar as predições do preço da soja e selecionar os melhores métodos de previsão para cada estado.

Para modelagem ARIMAX, as séries de preços foram diferenciadas e empregou-se os testes de raiz unitária KPSS e ADF. Em seguida, a análise dos gráficos da função de autocorrelação ACF dos resíduos e o teste de hipótese Ljung-Box comprovam que foram atendidos os pressupostos do modelo ARIMAX. Os resultados da modelagem via ARIMAX mostram que a Lei do Preço Único é válida no mercado brasileiro de soja, pois o método se ajustou bem aos preços domésticos e internacionais da soja.

Neste trabalho, os preços mensais da soja do Paraná, Rio Grande do Sul e Mato Grosso são estimados usando modelos de rede neural e máquina de aprendizado extremo. Os

modelos de inteligência artificial (IA) foram treinados e validados, e seu desempenho foi avaliado com base nos indicadores de desempenho nos conjuntos de preços.

A rede neural apresentou as melhores métricas entre os modelos de IA. Os resultados das modelagens mostraram que a MLP foi a única abordagem que gerou previsões com diferenças estatísticas significativas em duas comparações aos métodos concorrentes. Esses resultados de previsão mostram que o modelo de rede neural foi considerado o melhor método que se ajusta aos preços da soja do Rio Grande do Sul e Mato Grosso.

Três métodos de previsão foram comparados no objeto de estudo de preço da soja. Entre três modelos concorrentes foi encontrado o mais preciso para cada série de preço da soja modelada. A modelagem desenvolvida neste estudo é promissora e pode apoiar a tomada de decisão na previsão do preço da soja. Este estudo responde as questões de pesquisa: Qual método de previsão tem melhor precisão em relação ao preço da soja? Entre os modelos de predição testados, qual gerou as previsões mais precisas para os preços da soja? Então, qual é a melhor abordagem de previsão para os preços da *commodity*?

A primeira avaliação realizada para responder as questões de pesquisa foi avaliar de forma comparativa as métricas de erro na seção 4.3, mostrando que o ARIMAX obteve um melhor ajuste para dentro e fora da amostra e teve melhor capacidade preditiva para os preços da soja do Paraná. De modo geral, a rede neural teve o melhor desempenho de previsão do preço da soja do Rio Grande do Sul. Para previsão de preços do Mato Grosso, a rede neural gerou valores menores em todas as métricas.

De forma complementar, para melhorar as comparações desenvolvidas, as métricas de erros dos modelos de previsão são plotadas e comparadas em gráficos de *Boxplots*. Esta segunda avaliação permite concluir que o modelo ARIMAX foi superior em comparação às abordagens de IA para os preços do Paraná. No entanto, teve maior variabilidade dos erros de previsão para os preços do Mato Grosso. O método de rede neural (MLP) foi mais preciso em relação aos modelos concorrentes (ARIMAX e ELM), tanto na dispersão quanto a mediana dos erros de predições de preços do Rio Grande do Sul e Mato Grosso.

Em segundo lugar, os resultados dos testes de variância dos modelos, mostrados na Tabela 9, comprovam ter diferenças estatísticas significativas entre as abordagens de previsão. ARIMAX comparado ao modelo de rede neural (MLP) para os preços do Paraná; MLP em comparação a máquina de aprendizado (ELM) para os preços do Rio Grande do Sul; MLP em relação ao modelo ARIMAX para os preços do Mato Grosso e a variância dos erros da ELM comparada com erros do modelo ARIMAX, também nos preços do Mato Grosso. Os

resultados das demais comparações não tiveram diferenças significativas ($p > 0,05$) entre as variâncias de erros dos modelos de previsão.

A partir da discussão dos resultados, avaliações de indicadores de desempenho, *boxplots* e testes de variâncias tem evidências estatísticas para concluir que as abordagens de previsão ARIMAX e rede neural são precisas para prever preços do Paraná, Rio Grande do Sul e Mato Grosso, assim estes métodos de previsão podem ser indicados para auxiliar na tomada de decisão dos agentes do mercado agrícola.

5.2.1 Limitações da pesquisa e sugestões para estudos futuros

As limitações desta dissertação devem ser reconhecidas e são as seguintes: (i) os preços utilizados para realizar as previsões foram somente de soja, quando poderiam ser de mais *commodities* agrícolas, por exemplo, o preço do milho. (ii) Apesar de terem sido modelos promissores em termos de precisão da previsão, aplicar somente dois modelos de inteligência artificial e um estatístico também foi uma limitação, considerando que as pesquisas desse campo de previsão de séries temporais geralmente utilizam uma diversidade de modelos para aumentar o nível de comparação das previsões, assim evidenciar os modelos com maior nível de acuracidade. Pesquisas futuras devem comparar principalmente modelos híbridos e adicionar variáveis exógenas para melhorar a acurácia dos métodos combinados.

Referências

- ADLI, K. A. Forecasting steel prices using ARIMAX model: A case study of Turkey. **The International Journal of Business Management and Technology**, v. 4, n. 5, p. 62–68, 2020.
- AGNOL, A. D. et al. **O complexo agroindustrial da soja brasileira** Londrina-PR Circular Técnica 43 Embrapa Soja, , 2007.
- AGUIAR, S. DA C. G. DO E. S.; BORESTEIN, D. Redes bayesianas: uma ferramenta na previsão de preço de commodity. **Revista de Administração e Negócios da Amazônia**, v. 4, n. 3, p. 237–253, 2012.
- AHUMADA, H.; CORNEJO, M. Forecasting food prices: The case of corn, soybeans and wheat. **International Journal of Forecasting**, v. 32, n. 3, p. 838–848, 2016a.
- AHUMADA, H.; CORNEJO, M. Out-of-sample testing price discovery in commodity markets: the case of soybeans. **Agricultural Economics**, v. 47, p. 709–718, 2016b.
- AIUBE, F. A. L.; FERREIRA, B. C. F.; LEVY, A. Modelo de fatores para Commodities e cenários de preços no curto prazo: O caso da soja. **Estudos Economicos**, v. 50, n. 1, p. 159–182, 2020.
- ALLEN, P. G. Economic forecasting in agriculture. **International Journal of Forecasting**, v. 10, p. 81–135, 1994.
- ANGGRAENI, W. et al. The Performance of ARIMAX Model and Vector Autoregressive (VAR) Model in Forecasting Strategic Commodity Price in Indonesia. **Procedia Computer Science**, v. 124, p. 189–196, 2017.
- ARÊDES, A. F. DE; PEREIRA, M. W. G. Potencialidade da utilização de modelos de séries temporais na previsão do preço do trigo no Estado do Paraná. **Revista de Economia Agrícola**, v. 55, n. 1, p. 63–76, 2008.
- ARIA, M.; CUCCURULLO, C. bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v. 11, n. 4, p. 959–975, 2017.
- BALESTRIN, A. Uma análise da contribuição de Herbert Simon para as teorias organizacionais. **Revista Eletrônica de Administração**, v. 8, n. 4, p. 2–17, 2002.
- BARBOSA, M. Z.; MARGARIDO, M. A.; JUNIOR, S. N. Análise da elasticidade de transmissão de preços no mercado brasileiro de algodão. **Nova Economia**, v. 12, n. 2, p. 80–108, 2002.
- BETTIS, R. et al. Quantitative empirical analysis in strategic management. **Strategic Management Journal**, v. 35, n. 7, p. 949–953, 2014.
- BONETT, D. G.; SEIER, E. Statistical Inference for a Ratio of Dispersions Using Paired Samples. **Journal of Educational and Behavioral Statistics**, v. 28, n. 1, p. 21–30, 2003.
- BOX, G. E. P.; TIAO, G. C. Intervention analysis with applications to economic and environmental problems. **Journal of the American Statistical Association**, v. 70, n. 349, p. 70–79, 1975.

- BRANDT, J. A.; BESSLER, D. A. Price forecasting and evaluation: An application in agriculture. **Journal of Forecasting**, v. 2, n. 3, p. 237–248, 1983.
- BRESSAN, A. A. Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais. **RAE-eletrônica**, v. 3, n. 1, p. 1–20, 2004.
- BRESSAN, A. A.; LIMA, J. E. DE. Modelos de previsão de preços aplicados aos contratos futuros de boi gordo na BM&F. **Nova Economia**, v. 12, n. 1, p. 117–140, 2002.
- BROCKWELL, P. J.; DAVIS, R. A. **Time series: theory and methods**. [s.l.] Springer-Verlag, 1991.
- BUENO, R. DE L. DA S. **Econometria de séries temporais**. 2. ed. São Paulo-SP: Cengage Learning, 2011.
- BUNDCHEN, C.; WERNER, L. Comparação da Acurácia de Previsões de Demanda Turística em Sedes Olímpicas. **Turismo em Análise**, v. 27, n. 1, p. 85–107, 2016.
- CAMARGO, B. V.; JUSTO, A. M. IRAMUTEQ: Um software gratuito para análise de dados textuais. **Temas em Psicologia**, v. 21, n. 2, p. 513–518, 2013.
- CAMELO, H. DO N. et al. Previsão de velocidade do vento no nordeste brasileiro através do modelo ARIMAX. **Revista Ciência e Tecnologia**, v. 20, n. 36, p. 9–17, 2017.
- CAMELO, H. DO N. et al. Proposta para Previsão de Velocidade do Vento Através de Modelagem Híbrida Elaborada a Partir dos Modelos ARIMAX e RNA. **Revista Brasileira de Meteorologia**, v. 33, n. 1, p. 115–129, 2018a.
- CAMELO, H. DO N. et al. A hybrid model based on time series models and neural network for forecasting wind speed in the Brazilian northeast region. **Sustainable Energy Technologies and Assessments**, v. 28, p. 65–72, 2018b.
- CAMPEÃO, P.; SANCHES, A. C.; MACIEL, W. R. E. Mercado Internacional de Commodities Uma Análise da Participação do Brasil no Mercado Mundial de Soja entre 2008 e 2019. **Desenvolvimento em Questão**, v. 18, n. 51, p. 76–92, 2020.
- CARRASCO-GUTIERREZ, C. E.; ALMEIDA, F. M. DE M. Modelagem e Previsão do Preço do Café Brasileiro. **Revista de Economia**, v. 39, n. 2, p. 7–27, 2013.
- CARVALHO, D. M. DE; PRÉVOT, F.; MACHADO, J. A. D. O uso da teoria da visão baseada em recursos em propriedades rurais: uma revisão sistemática da literatura. **Revista de Administração**, v. 49, n. 3, p. 506–518, 2014.
- CAS, C. G. Aplicação do modelo arima para previsão do preço da commodity milho. **GEPROS Gestão da Produção, Operações e Sistemas**, v. 13, n. 1, p. 263–279, 2018.
- CASAS, A. L. LAS; BACHA, R. A. F.; CARVALHO, C. M. E. O agronegócio e o marketing rural no estado de Mato Grosso do Sul. **História e Perspectivas**, n. 55, p. 271–289, 2016.
- CASTRO, N. R. et al. Modelos univariados e multivariados aplicados à previsão de valores de exportação: uma análise comparativa para o complexo soja. **Revista ESPACIOS**, v. 37, n. 3, p. 1–13, 2016.
- CEPEA. **Indicador da soja ESALQ/BM&FBOVESPA - Paranaguá**. Disponível em: <<https://bit.ly/2Oh7Uas>>. Acesso em: 22 abr. 2019.

CERETTA, P. S.; RIGHI, M. B.; SCHLENDER, S. G. Previsão do preço da soja: uma comparação entre os modelos ARIMA e redes neurais artificiais. **Informações Econômicas**, v. 40, n. 9, p. 16–27, 2010.

CHATFIELD, C. The Holt-Winters Forecasting Procedure. **Journal of the Royal Statistical Society**, v. 27, n. 3, p. 264–279, 1978.

CHERDCHOONGAM, S.; RUNGREUNGANUN, V. Forecasting the Price of Natural Rubber in Thailand Using the ARIMA Model. **International Journal of Applied Science and Technology**, v. 9, n. 4, p. 271–277, 2016.

CLAVERIA, O.; TORRA, S. Forecasting tourism demand to Catalonia: Neural networks vs. time series models. **Economic Modelling**, v. 36, p. 220–228, 2014.

CNA. **PIB do agronegócio cresce 3,81% em 2019** Confederação da Agricultura e Pecuária do Brasil, , 2020. Disponível em: <https://www.cnabrazil.org.br/assets/arquivos/boletins/sut.pib_dez_2020.5mar2020vf.pdf>

CONAB. **Preços médios mensais - produtor**. Disponível em: <<http://sisdep.conab.gov.br/precosiagroweb/>>. Acesso em: 20 maio. 2019.

CONSTANTINO, M.; MENDES, D. R. F.; MOREIRA, T. B. S. Analysis and forecasting of inflation rate in Brazil: an approach to the SARIMA Model. **Revista Científica Hermes**, v. 24, p. 244–257, 2019.

CRONE, S. F.; KOURENTZES, N. Feature selection for time series prediction - A combined filter and wrapper approach for neural networks. **Neurocomputing**, v. 73, n. 10–12, p. 1923–1936, 2010.

DAS, P. et al. **Price Forecasting using EMD-SVR hybrid model: An Ensemble learning Paradigm**. Eighth International Conference on Agricultural Statistics. **Anais...**New Delhi, India: ICAS, 2019Disponível em: <10.13140/RG.2.2.33669.70886>

DAS, S. R.; MISHRA, D.; ROUT, M. A hybridized ELM-Jaya forecasting model for currency exchange prediction. **Journal of King Saud University - Computer and Information Sciences**, 2017.

DAVID, S. A. et al. Dynamics of Commodities Prices : Integer and Fractional Models. **Fundamenta Informaticae**, v. 151, p. 389–408, 2017.

DE MYTTENAERE, A. et al. Mean Absolute Percentage Error for regression models. **Neurocomputing**, v. 192, p. 38–48, 2016.

DE OLIVEIRA, F. A.; NOBRE, C. N.; ZÁRATE, L. E. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index - Case study of PETR4, Petrobras, Brazil. **Expert Systems with Applications**, v. 40, n. 18, p. 7596–7606, 2013.

DE SOUZA, C. C. et al. Análise do padrão estacional e previsão dos preços da soja aos produtores do estado de São Paulo, Brasil. **Extensão Rural**, v. 25, n. 1, p. 78–92, 2018.

DEGIANNAKIS, S. et al. Forecasting realized volatility of agricultural commodities. **International Journal of Forecasting**, p. 2–23, 2020.

DENG, H. L.; SŪN, Y. Q. Soybean Price Pattern Discovery Via Toeplitz Inverse Covariance-Based Clustering. **International Journal of Agricultural and Environmental Information Systems**, v. 10, n. 4, p. 1–17, 2019.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, v. 74, n. 366, p. 427–431, 1979.

DRACHAL, K. Analysis of Agricultural Commodities Prices with New Bayesian Model Combination Schemes. **Sustainability**, v. 11, n. 5305, p. 2–23, 2019.

DUARTE, M. T. **Um estudo sobre modelos de previsão de preços no mercado de grão de soja**UberlândiaInstituto de Economia e Relações Internacionais, , 2007. Disponível em: <<https://repositorio.ufu.br/handle/123456789/13476>>

EMBRAPA SOJA. **Soja em números (safra 2019/20)**. Disponível em: <<https://www.embrapa.br/soja/cultivos/soja1/dados-economicos>>. Acesso em: 20 set. 2020.

FAN, X.; WANG, L.; LI, S. Predicting chaotic coal prices using a multi-layer perceptron network model. **Resources Policy**, v. 50, p. 86–92, 2016.

FANG, Y. et al. Optimal forecast combination based on ensemble empirical mode decomposition for agricultural commodity futures prices. **Journal of Forecasting**, p. 1–10, 2020.

FAO. **Price volatility in agricultural markets: Evidence, impact on food security and policy responses**Rome, ItalyEconomic and Social Development Department, Food and Agriculture Organization of the United Nations (FAO), , 2010. Disponível em: <<http://www.fao.org/docrep/013/am053e/am053e00.pdf>>

FAO. **Perspectivas Agrícolas OCDE-FAO: Brasil vai ultrapassar os Estados Unidos como o maior produtor de soja até 2026**. Disponível em: <<http://www.fao.org/americas/noticias/ver/pt/c/992188/>>. Acesso em: 3 ago. 2020.

FAO NO BRASIL. **OCDE e FAO preveem aumento de 17% na produção agrícola e pesqueira na América Latina e no Caribe até 2027**. Disponível em: <<http://www.fao.org/brasil/noticias/detail-events/pt/c/1144775/>>. Acesso em: 3 ago. 2020.

FARAWAY, J.; CHATFIELD, C. Time Series Forecasting with Neural Networks: A Comparative Study Using the Airline Data. **Journal of the Royal Statistical Society**, v. 47, n. 2, p. 231–250, 1998.

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

FERREIRA, L. et al. Utilização de Redes Neurais Artificiais como Estratégia de Previsão de Preços no Contexto de Agronegócio. **RAI – Revista de Administração e Inovação**, v. 8, n. 4, p. 6–26, 2011.

FILHO, R. B. et al. Impactos da abertura comercial brasileira na transmissão de preços de soja em grãos no mercado internacional. **Revista Brasileira de Economia de Empresas**, v. 12, n. 1, p. 63–79, 2012.

FMI. **Preços de commodities primárias do FMI**. Disponível em:

<<https://www.imf.org/en/Research/commodity-prices>>. Acesso em: 18 set. 2020.

FORIGO, E. M.; ZAGO, C. A.; ABREU, L. F. **O hedge como ferramenta para diminuição dos riscos nas operações financeiras de uma cooperativa de agroalimentos**. VI Congresso Nacional de Excelencia em Gestão. **Anais...**Niterói: Congresso Nacional de Excelencia em Gestão, 2010Disponível em: <http://www.inovarse.org/sites/default/files/T10_0229_1249_3.pdf>

GIEMBINSKY, R. C. **Competitividade e preços do complexo soja no Brasil**. [s.l.] Universidade Federal de Uberlândia (UFU), 2006.

GUTIERREZ, J. L. C. **Monitoring of the Corumbá - i Dam Instrumentation Neural Networks and the Box & Jenkins null Models**. [s.l.] Pontifícia Universidade Católica do Rio de Janeiro, 2003.

HAIDER, A.; HANIF, M. N. Inflation Forecasting in Pakistan Using Artificial Neural Networks. **Pakistan Economic and Social Review**, v. 47, n. 1, p. 123–138, 2009.

HASSANI, H. et al. Forecasting accuracy evaluation of tourist arrivals. **Annals of Tourism Research**, v. 63, p. 112–127, 2017.

HERRERA, G. P. et al. Long-term forecast of energy commodities price using machine learning. **Energy**, v. 179, p. 214–221, 2019.

HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: Theory and applications. **Neurocomputing**, v. 70, p. 489–501, 2006.

HUANG, G.-B.; ZHU, Q.; SIEW, C.-K. **Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks**. (2, Ed.)IEEE International Joint Conference on Neural Networks. **Anais...**Budapest: IEEE, 2004Disponível em: <<https://ieeexplore.ieee.org/document/1380068>>

HYNDMAN, R. et al. **forecast: Forecasting functions for time series and linear models**CRAN, , 2020. Disponível em: <<https://pkg.robjhyndman.com/forecast/>>

HYNDMAN, R. J.; ATHANASOPOULOS, G. ARIMA modelling in R. In: **Forecasting: principles and practice**. Melbourne, Australia: OTexts, 2018a.

HYNDMAN, R. J.; ATHANASOPOULOS, G. Time series graphics. In: **Forecasting: principles and practice**. Melbourne, Australia: OTexts, 2018b.

HYNDMAN, R. J.; ATHANASOPOULOS, G. 8.1 Stationarity and differencing. In: **Forecasting: principles and practice**. Melbourne, Australia: OTexts, 2018c.

HYNDMAN, R. J.; ATHANASOPOULOS, G. 8.6 Estimation and order selection. In: **Forecasting: principles and practice**. Melbourne, Australia: OTexts, 2018d.

HYNDMAN, R. J.; ATHANASOPOULOS, G. White noise. In: **Forecasting: principles and practice**. Melbourne, Australia: OTexts, 2018e.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: The forecast package for R. **Journal Of Statistical Software**, v. 27, n. 3, p. 2–22, 2008.

IPEA. **Conjuntura Agrícola Brasileira**BrasíliaInstituto de Pesquisa Econômica Aplicada, , 2020. Disponível em:

- <https://www.ipea.gov.br/%0Aportal/images/stories/PDFs/conjuntura/201124_cc_49_nota_21_agropecuaria.pdf>. Acesso em: 14 dez. 2020
- JHA, G. K.; SINHA, K. Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. **Neural Computing and Applications**, v. 24, n. 3, p. 1–9, 2012.
- 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.
- JUNIOR, E. F. DE O. et al. EXPORTAÇÃO DE SOJA NO ESTADO DE MATO GROSSO DO SUL: CARACTERÍSTICAS DA COMERCIALIZAÇÃO. **Revista em Agronegócio e Meio Ambiente**, v. 11, n. 1, p. 71–97, 2018.
- KAASTRA, I.; BOYD, M. Designing a neural network for forecasting financial and economic time series. **Neurocomputing**, v. 10, n. 3, p. 215–236, 1996.
- KARAMAZIOTIS, P. I. et al. An empirical investigation of water consumption forecasting methods. **International Journal of Forecasting**, v. 36, n. 2, p. 588–606, 2020.
- KAZMIER, L. J. **Estatística aplicada à economia e administração**. São Paulo: McGraw-Hill do Brasil, 1982.
- KIM, S.; KIM, H. A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, v. 32, n. 3, p. 669–679, 2016.
- KITWORAWUT, P.; RUNGREUNGANUN, V. Corn Price Modeling and Forecasting Using Box-Jenkins Model. **Applied Science and Engineering Progress**, v. 12, n. 4, p. 264–276, 2019.
- KOHZADI, N. et al. A comparison of artificial neural network and time series models for forecasting commodity prices. **Neurocomputing**, v. 10, n. 2, p. 169–181, 1996.
- KOURENTZES, N. **nnfor: Time Series Forecasting with Neural Networks** CRAN, , 2019. Disponível em: <<https://cran.r-project.org/package=nnfor>>
- KOURENTZES, N.; BARROW, D. K.; CRONE, S. F. Neural network ensemble operators for time series forecasting. **Expert Systems with Applications**, v. 41, n. 9, p. 4235–4244, 2014.
- KRUGMAN, P. R.; OBSTFELD, M. **Economía Internacional Teoría y Política**. 7. ed. Madrid: Pearson Educación, 2006.
- KÜCK, M.; FREITAG, M. Forecasting of customer demands for production planning by local k-nearest neighbor models. **International Journal of Production Economics**, v. 231, p. 2–22, 2021.
- KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? **Journal of Econometrics**, v. 54, n. 1–3, p. 159–178, 1992.
- LANDAU, L. J.; TAYLOR, J. G. **Concepts for Neural Networks a survey**. London: Springer-Verlag, 1998.

LANG, D.; CHIEN, G. **Package wordcloud2: Create Word Cloud by “htmlwidget”** Comprehensive R Archive Network CRAN, , 2018. Disponível em: <<https://github.com/lchiffon/wordcloud2>>

LAW, R. Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. **Tourism Management**, v. 21, n. 4, p. 331–340, 2000.

LEWIS, C. D. Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. **Butterworth-Heinemann**, 1982.

LIBERA, A. A. D.; WAQUIL, P. D. Transmissão de preços entre mercados de soja. **Revista de Política Agrícola**, v. 18, n. 3, p. 58–69, 2009.

LIMA, F. G. et al. Previsão de preços de commodities com modelos ARIMA-GARCH e redes neurais com ondaletas: velhas tecnologias – novos resultados. **Revista de Administração**, v. 45, n. 2, p. 188–202, 2010.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297–303, 1978.

MAJOR, M. J. O positivismo e a pesquisa ‘alternativa’ em Contabilidade. **Revista Contabilidade & Finanças**, v. 28, n. 74, p. 173–178, 2017.

MAKRIDAKIS, S. Accuracy concerns measures: theoretical and practical concerns. **International Journal of Forecasting**, v. 9, n. 4, p. 527–529, 1993.

MAKRIDAKIS, S. G.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting: Methods and Applications**. 3. ed. New York: John Wiley & Sons, 1998.

MANCUSO, A. C. B.; WERNER, L. A comparative study on combinations of forecasts and their individual forecasts by means of simulated series. **Acta Scientiarum. Technology**, v. 41, n. 1, p. 2–9, 2019.

MAPA. **Balança Comercial do Agronegócio - 2020**. Disponível em: <<https://www.gov.br/agricultura/pt-br/assuntos/noticias/exportacoes-do-agronegocio-somam-quase-us-9-bilhoes-em-agosto/NotaaimpresaAgosto2020.pdf>>. Acesso em: 14 dez. 2020.

MARCHEZAN, A.; SOUZA, A. M. Previsão do preço dos principais grãos produzidos no Rio Grande do Sul. **Ciência Rural**, p. 1–7, 2010.

MARGARIDO, M. A.; TUROLLA, F. A.; FERNANDES, J. M. Análise da elasticidade de transmissão de preços no mercado internacional de soja. **Pesquisa & Debate**, v. 12, n. 2, p. 5–40, 2001.

MEDEL, C. A. How Informative are In-Sample Information Criteria to Forecasting? The Case of Chilean GDP. **Latin American Journal of Economics**, v. 50, n. 1, p. 133–161, 2013.

MENDES, J. T. G.; JUNIOR, J. B. P. **Agronegócio: uma abordagem econômica**. São Paulo: Pearson Prentice Hall, 2007.

MIQUELLUTI, D. L. **Métodos alternativos de previsão de safras agrícolas**. [s.l.] Universidade de São Paulo, 2015.

MITRA, D.; PAUL, R. K. Hybrid time-series models for forecasting agricultural commodity prices. **Model Assisted Statistics and Applications**, v. 12, n. 3, p. 255–264, 2017.

MIYAOKA, J.; HAUSMAN, W. H. How Improved Forecasts Can Degrade Decentralized Supply Chains. **Manufacturing & Service Operations Management**, v. 10, n. 3, p. 547–562, 2008.

MORETTIN, P. A.; SINGER, J. M. **Introdução à ciência de dados fundamentos e aplicações**. São Paulo-SP: Departamento de Estatística da Universidade de São Paulo, IME-USP, 2020.

MUNDLAK, Y.; LARSON, D. F. On the transmission of world agricultural prices. **World Bank Economic Review**, v. 6, n. 3, p. 399–422, 1992.

NOMELINI, Q. S. S. et al. Uso de modelagem univariada e multivariada com séries temporais como ferramenta de gestão do agronegócio na cultura de soja do Brasil. **Revista ESPACIOS**, v. 38, n. 8, p. 1, 2017.

PACE, L. Performing t Tests. In: **Beginning R**. Berkeley, CA: Apress, 2012a. p. 125–138.

PACE, L. One-Way Analysis of Variance. In: **Beginning R**. Berkeley, CA: Apress, 2012b. p. 139–147.

PALMER, A.; JOSÉ MONTAÑO, J.; SESÉ, A. Designing an artificial neural network for forecasting tourism time series. **Tourism Management**, v. 27, n. 5, p. 781–790, 2006.

PETROPOULOS, F. et al. **Forecasting: theory and practice**. 1. ed. New York: preprint arXiv:2012.03854 Cornell University, 2020.

PFAFF, B. **Analysis of Integrated and Cointegrated Time Series with R**. 2. ed. New York: Springer-Verlag, 2008.

PINHEIRO, C. A. O.; SENNA, V. DE. **Previsão de preços de commodities agrícolas através do modelo de análise espectral singular multivariada**. XLVII SBPO. Anais...Pernambuco-PE: Simpósio Brasileiro de Pesquisa Operacional, 2015Disponível em: <<http://cdsid.org.br/sbpo2015/wp-content/uploads/2015/08/141120.pdf>>

PINHEIRO, C. A. O.; SENNA, V. DE. Previsão de Preços Através da Análise Espectral Multivariada: Evidências para Commodities da BM&Fbovespa. **Brazilian Business Review**, v. 13, n. 5, p. 133–162, 2016.

PINHEIRO, C. A. O.; SENNA, V. DE. Previsão de preços através de redes neurais e análise espectral: evidências para o mercado futuro das commodities açúcar e soja. **Custos e @gronegócio online**, v. 13, n. 4, p. 103–128, 2017a.

PINHEIRO, C. A. O.; SENNA, V. DE. Aplicacao de analise multivariada e redes neurais para previsao de precos no mercado agricola brasileiro. **Ciência Rural**, v. 47, n. 1, p. 2–7, 2017b.

PINHEIRO, C. A. O.; SENNA, V. DE. Forecasting of prices through neural networks and spectral analysis: evidence for the future market commodity sugar and soy. **Custos e @gronegócio online**, v. 13, n. 4, p. 103–128, 2017c.

PINHEIRO, C. A. O.; SENNA, V. DE; MATSUMOTO, A. S. Price forecasting for future contracts on agribusiness through neural network and multivariate spectral analysis. **Revista de Gestão, Finanças e Contabilidade**, v. 6, n. 3, p. 98–124, 2016.

PINO, F. A.; ROCHA, M. B. Transmissão de preços de soja no brasil. **Revista de Economia**

e **Sociologia Rural**, v. 32, n. 4, p. 345–361, 1994.

PORTO, B. M.; PHILIPPI, D. A. **Previsão de séries temporais por meio de redes neurais**. Anais do VI Simpósio de Engenharia de Produção - SIMEP. **Anais...**Salvador: even3, 2018aDisponível em: <<https://www.even3.com.br/anais/6simep/79506-previsao-de-series-temporais-por-meio-de-redes-neurais/>>

PORTO, B. M.; PHILIPPI, D. A. **Previsão de séries temporais de petróleo por meio de redes neurais**. Encontro da Associação Nacional de Pós Graduação e Pesquisa em Administração (EnANPAD). **Anais...**Curitiba/PR: XLII Encontro da ANPAD, 2018bDisponível em: <<https://tinyurl.com/y9l6y59d>>

PORTO, B. M.; PHILIPPI, D. A.; LEITE, V. A. W. Previsão dos modelos univariados e rede neural da demanda turística do estado de Mato Grosso do Sul. **Caderno Virtual de Turismo**, v. 19, n. 3, 2019.

PUCHALSKY, W. et al. Agribusiness time series forecasting using Wavelet neural networks and metaheuristic optimization: An analysis of the soybean sack price and perishable products demand. **International Journal of Production Economics**, v. 203, p. 174–189, 2018.

QUEVEDO-SILVA, F. et al. Bibliometric study: Guidelines on its application. **Revista Brasileira de Marketing**, v. 15, n. 2, p. 246–262, 2016.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**Vienna, AustriaR Foundation for Statistical Computing, , 2020. Disponível em: <<https://www.r-project.org/>>

RIBEIRO, C. D. O.; SOSNOSKI, A. A. K. B.; DE OLIVEIRA, S. M. Um modelo hierárquico para previsão de preços de commodities agrícolas. **Revista Produção Online**, v. 10, n. 4, p. 719–733, 2010.

RIBEIRO, C. D. O.; SOSNOSKI, A. A. K. B.; OLIVEIRA, S. M. DE. Um modelo hierárquico para previsão de preços de commodities agrícolas. **Revista Produção Online**, v. 10, n. 4, p. 719–733, 2010.

RIBEIRO, M. H. D. M.; COELHO, L. DOS S. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. **Applied Soft Computing Journal**, v. 86, p. 2–17, 2020.

RIBEIRO, R. J. P. **Modelos de previsão de preços de commodities**, 2014. Disponível em: <<http://hdl.handle.net/10362/13782>>

RICHARDSON, R. J. **Pesquisa social: métodos e técnicas**. 3. ed. São Paulo: Atlas, 2009.

SABU, K. M.; KUMAR, T. K. M. Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala. **Procedia Computer Science**, v. 171, p. 699–708, 2020.

SAFARI, A.; DAVALLOU, M. Oil price forecasting using a hybrid model. **Energy**, v. 148, p. 49–58, 2018.

SAGAERT, Y. R. et al. Temporal Big Data for Tactical Sales Forecasting. **Interfaces**, 2017.

SALLES, R. et al. **A Framework for Benchmarking Machine Learning Methods Using Linear Models for Univariate Time Series Prediction**. International Joint Conference on

Neural Networks (IJCNN). **Anais...** Anchorage: 2017 Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7966139&isnumber=7965814>>

SANTANA, L. V. R.; RUSSO, S. L.; FIGUEROA, M. L. **Aplicação da Metodologia Box-Jenkins na Previsão da Produção da Cana-de-Açúcar** Natal/RN Anais do Congresso de Matemática Aplicada e Computacional, , 2012. Disponível em: <<https://goo.gl/At4ipS>>

SANTOS FELIPE, I. J. DOS et al. Application of ARIMA models in soybean series of prices in the north of Paraná. **Custos e @gronegócio online**, v. 8, n. SPL, p. 78–91, 2012.

SCHNEIDER, S.; FERREIRA, B.; ALVES, F. **Aspectos multidimensionais da agricultura brasileira: diferentes visões do Censo Agropecuário 2006**. Brasília: Ipea, 2013.

SHUMWAY, R. H.; STOFFE, D. S. ARIMA Models. In: **Time series analysis and its applications: with R examples**. 3. ed. New York: Springer, 2010. p. 77–156.

SILVA, M. R. DA; HAYASHI, C. R. M.; HAYASHI, M. C. P. I. Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. **InCID: Revista de Ciência da Informação e Documentação**, v. 2, n. 1, p. 110–129, 2011.

SIMON, H. A. Making management decisions: The role of intuition and emotion. **Academy of Management**, v. 1, n. 1, p. 57–64, 1987.

SOUSA, E. P. DE; CAMPOS, A. C. Transmissão Inter-Regional de Preços no Mercado Brasileiro de Soja. **Revista Análise Econômica**, v. 27, n. 51, p. 83–105, 2009.

SOUZA, C. C. DE et al. Análise do padrão estacional e previsão dos preços da soja aos produtores do estado de São Paulo, Brasil. **Extensão Rural**, v. 25, n. 1, p. 78–92, 2018.

SOUZA, M. O. DE; BACCHI, M. R. P.; ALVES, L. R. A. Análise de fatores que influenciam o processamento de soja no Brasil. **Revista de Economia e Agronegócio**, v. 17, n. 3, p. 486–506, 2019.

STEVENSON, W. J. **Estatística aplicada à administração**. São Paulo: Harper & Row do Brasil, 1981.

TIBULO, C.; CARLI, V. Previsão do preço do milho, através de séries temporais. **Scientia Plena**, v. 10, n. 10, p. 1–10, 2014.

TONIN, J. M.; BARCZSZ, S. S. **Transmissão de preços da soja entre os mercados externo e interno: uma abordagem para a região de maringá**. XLVI Congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural. **Anais...** Rio Branco/Acre: Sociedade Brasileira de Economia, Administração e Sociologia Rural, 2008

TRANTAFYLLOU, A.; DOTSI, G.; SARRIS, A. H. Volatility Forecasting and Time-varying Variance Risk Premiums in Grains Commodity Markets. **Journal of Agricultural Economics**, v. 66, n. 2, p. 329–357, 2015.

WANG, D. et al. Performance Analysis of Four Decomposition-Ensemble Models for One-Day-Ahead Agricultural Commodity Futures Price Forecasting. **Algorithms**, v. 10, n. 108, p. 2–25, 2017.

WANG, J. et al. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. **International Journal of Forecasting**, 2019.

WANG, L. et al. Agricultural product price forecasting methods: research advances and trend. **British Food Journal**, v. 122, n. 7, p. 2121–2138, 2020.

WHITE, H. Learning in Artificial Neural Networks: A Statistical Perspective. **Neural Computation**, v. 1, p. 425–464, 1989.

WOLFF, L.; SOUZA, A. M. Estudo sobre os preços do açúcar e alcoóis e a sua influência de longo prazo no volume dos estoques públicos no Brasil. **Revista Produção Online**, v. 13, n. 1, p. 162–179, 2013.

ZHANG, D. et al. Prediction of soybean price in China using QR-RBF neural network model. **Computers and Electronics in Agriculture**, v. 154, p. 10–17, 2018.

ZHANG, D. et al. Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons. **IEEE Access**, v. 8, p. 28197–28209, 2020.

ZHANG, G.; BERARDI, V. Time series forecasting with neural network ensembles: an application for exchange rate prediction. **The Journal of the Operational Research Society**, v. 52, n. 6, p. 652–664, 2001.

ZOU, H. F. et al. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. **Neurocomputing**, v. 70, n. 16–18, p. 2913–2923, 2007.

Apêndice 1 - Aplicação empírica em R (scripts)

Quadro 4 - Aplicação prática em linguagem de programação R da rede neural *perceptron* multicamada (MLP)

```
#### Pacote “nnfor” Previsão de séries temporais com redes neurais MLP ####
library(forecast) # carregar pacote dependente
library(nnfor) # carregar pacote utilizado
library(nnfor)

## Treinamento do modelo rede neural
fit <- mlp(Conjunto_treinamento_MT)
print(fit)
plot(fit)

## Testes do modelo no conjunto
autoplot(Conjunto_treinamento_MT, series = "Conjunto treinamento MT", xlab = "Tempo",
ylab = "Preço do Mato Grosso", main = "Rede Neural") + autolayer(fit$fitted, series =
"Rede neural")
accuracy(Conjunto_treinamento_MT, fit$fitted)
```

```
## Previsão da rede neural no conjunto teste (fora da amostra)
frc <- forecast(fit,h=12)

autoplot(Conjunto_treinamento_MT, series = "Conjunto treinamento MT", xlab ="Tempo",
ylab = "Preço do Mato Grosso", main = "Rede Neural") + autolayer(fit$fitted, series =
"Rede neural") + autolayer(frc, series = "Previsão rede neural") +
autolayer(Conjunto_teste_MT, series = "Conjunto teste MT")

accuracy(frc, x=Conjunto_teste_MT)
```

Fonte: Elaborado pelo autor.

Quadro 5 - Aplicação prática em linguagem de programação R da máquina de aprendizagem extrema (ELM)

```
#### Pacote “nnfor” algoritmo ELM para previsão de séries temporais ####
library(forecast) # carregar pacote dependente
library(nnfor) # carregar pacote utilizado
fit <- elm(Conjunto_treinamento_MT)
print(fit)
plot(fit)

autoplot(Conjunto_treinamento_MT, series = "Conjunto treinamento MT", xlab ="Tempo",
ylab = "Preço do Mato Grosso", main = "Máquina de aprendizado extremo") +
autolayer(fit$fitted, series = "Máquina de aprendizado")

accuracy(Conjunto_treinamento_MT, fit$fitted)

frc <- forecast(fit, h=12)

autoplot(Conjunto_treinamento_MT, series = "Conjunto treinamento MT", xlab ="Tempo",
ylab = "Preço do Mato Grosso", main = "ELM") + autolayer(fit$fitted, series = "Máquina
de aprendizado") + autolayer(frc, series = "Previsão da ELM") +
autolayer(Conjunto_teste_MT, series = "Conjunto teste MT")

accuracy(frc, x=Conjunto_teste_MT)
```

Fonte: Elaborado pelo autor.