

PREVISÃO DO PREÇO DO AÇÚCAR: UM ESTUDO DE CASO USANDO O SOFTWARE WEKA

Jandreí Sartori Spancerski (UTFPR) E-mail: jandreiss@gmail.com
José Airton Azevedo dos Santos (UTFPR) E-mail: airton@utfpr.edu.br

Resumo: O açúcar, desde a colonização, sempre foi um produto de destaque no nosso país. A previsão de preços futuros do produto é muito importante, no processo de tomada de decisão, para produtores, analistas de mercado, entre outros. Neste contexto, este trabalho tem como objetivo avaliar a eficácia de modelos, fornecidos pelo software WEKA, para previsão do preço do açúcar. A base de dados apresenta uma série histórica de preço mensal do açúcar no período entre Janeiro/2006 e Julho/2020, totalizando 175 observações. Modelos de previsão *Multilayer Perception* e *SMOreg* foram utilizados na previsão do preço do açúcar. Resultados obtidos, dos dois modelos, foram comparados. Verificou-se, para um horizonte de curto prazo, que os dois modelos de previsão fornecem estimativas confiáveis para o preço do açúcar.

Palavras-chave: *Multilayer perception*; *SMOreg*; Açúcar.

SUGAR PRICE FORECAST: A CASE STUDY USING WEKA SOFTWARE

Abstract: Sugar, since colonization, has always been a prominent product in our country. The forecast of future product prices is very important, in the decision-making process, for producers, market analysts, among others. In this context, this work aims to evaluate the effectiveness of models, provided by WEKA software, for forecasting the price of sugar. The database presents a historical series of monthly sugar prices in the period between January/2006 and July/2020, totaling 175 observations. *Multilayer Perception* and *SMOreg* forecast models were used to forecast sugar prices. Results obtained from the two models were compared. It was found, for a short term horizon, that the two forecast models provide reliable estimates for the price of sugar.

Keywords: *Multilayer perception*; *SMOreg*; Sugar.

1. INTRODUÇÃO

O agronegócio, devido a sua capacidade de geração de renda e empregos, exerce um papel importante no desenvolvimento da economia brasileira. O açúcar é uma *commodity* produzida em várias partes do mundo. É produzido a partir da cana-de-açúcar e da beterraba, sendo que a cana-de-açúcar é responsável por mais de 70% da produção mundial total de açúcar (ADVFN, 2020). O Brasil, devido a sua competitividade na produção de cana de açúcar, está entre os principais países produtores, consumidores e exportadores de açúcar (SATOLO; BACCHI, 2009).

Estudo do comportamento de preços, de produtos agropecuários, pode gerar parâmetros de interesse para produtores, analistas de mercado, entre outros. Segundo Nicola et al. (2008) a atividade açucareira, devido a flutuação dos preços do produto, é uma atividade de alto risco financeiro. Portanto, essa atividade geralmente necessita de ferramentas que minimizem os riscos e as incertezas no processo de decisão (BACCHI; ALVES, 2004).

Modelos de previsões, de séries temporais, podem ser utilizados na previsão de preços de ativos financeiros. Segundo Moore et al. (2006), séries temporais são medidas de determinadas variáveis (preço, produção, etc) tomadas em intervalos regulares de tempo. Um modelo de séries temporais fornece previsões de observações futuras por meio da relação que possui com valores passados (MORETTIN; TOLOI, 2004).

O *Waikato Environment for Knowledge Analysis* (WEKA) é um software livre que contém diversas técnicas de mineração de dados. O WEKA possui também um ambiente dedicado a análise de séries temporais. Este ambiente, denominado de *Forecast*, contém os algoritmos de previsão *Multilayer Perception* (MLP) e *SMOreg*.

Diversos trabalhos utilizaram métodos de previsão em aplicações voltadas ao setor açucareiro. Dentre eles, podem-se citar os trabalhos de: Nicola et al. (2008) que realizaram a previsão do preço do açúcar utilizando modelos da família ARCH. Soares et al. (2019) que aplicaram a metodologia Box-Jekins na previsão da demanda do bagaço da cana de açúcar. Bader Essam (2017) que realizou a previsão da produção e do consumo do açúcar no Egito e Mehmood et al. (2019) realizaram, por meio da metodologia ARIMA, a previsão da produção de cana de açúcar no Pasquitão.

Dentro deste contexto, este trabalho tem como objetivo avaliar a eficácia de modelos, fornecidos pelo software WEKA, para previsão do preço do açúcar no período entre Janeiro/2006 e Julho/2020.

Este trabalho está organizado da seguinte maneira. Na seção Materiais e Métodos são descritas as metodologias empregadas na previsão do preço do açúcar, sendo também apresentado, nesta seção, um breve resumo dos algoritmos MLP e SMOReg. Na Seção Resultados e Discussão são apresentados os resultados obtidos da aplicação, dos algoritmos, em observações mensais do preço do açúcar. Comentários finais e conclusões finalizam o trabalho.

2. MATERIAIS E MÉTODOS

Base de dados:

Para previsão do preço do açúcar utilizou-se uma base de dados com 175 meses (Jan/2006 - Jul/2020) (INDEXMUNDI, 2020). Os dados obtidos, da base de dados, já estavam limpos e sem a presença de *outliers*. Na Figura 1 apresenta-se o *boxplot* dos dados.

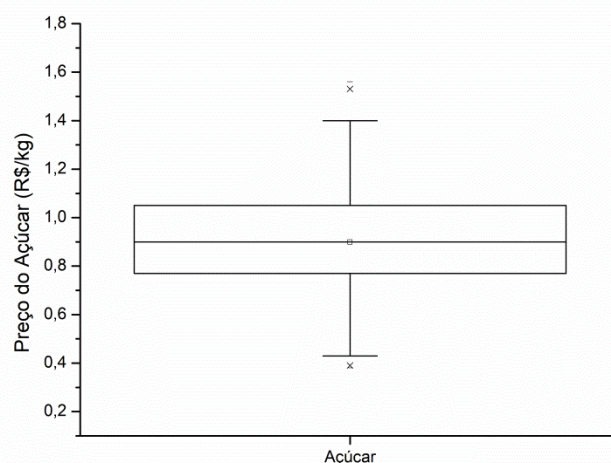


Figura 1 – *Boxplot* – Preço do açúcar.
Fonte – INDEXMUNDI (2020).

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

Tabela 1 – Dez primeiros registros do conjunto de dados.

Data	Preço do Açúcar (R\$/kg)
2006-1	0,8
2006-2	0,87
2006-3	0,82
2006-4	0,83
2006-5	0,78
2006-6	0,77
2006-7	0,77
2006-8	0,65
2006-9	0,58
2006-10	0,56

Fonte – INDEXMUNDI (2020).

2. O gráfico da série temporal, utilizada na estimação dos modelos, é apresentado na Figura

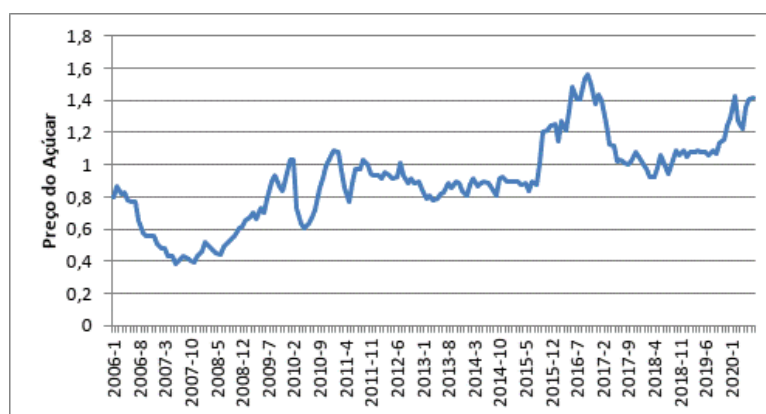


Figura 2 – Gráfico ilustrativo da série temporal do preço do açúcar (R\$/kg).

Fonte – INDEXMUNDI (2020).

Aprendizado de máquina:

Segundo Rolim et al. (2017), o Aprendizado de Máquina (*Machine Learning*), subcampo da Inteligência Artificial, trabalha com a ideia de que as máquinas são capazes de adquirir conhecimento a partir de dados. Possui aplicações em diversos campos do conhecimento. O Aprendizado de Máquina pode ser dividido em aprendizado supervisionado e não supervisionado. No aprendizado supervisionado o modelo apreende a partir de resultados pré-definidos. No aprendizado não supervisionado não existem resultados pré-definidos para o modelo utilizar como referência na aprendizagem. Este tipo de aprendizagem tem como objetivo agrupar elementos com características similares (PELLUCCI et al., 2011).

O software WEKA:

O software WEKA é uma coleção de algoritmos de *Machine Learning* implementados na linguagem Java (RUSSEL & MARKOV, 2017). O software está disponível para Windows, Linux e outras plataformas. Ele contém ferramentas para pré-processamento de dados,

classificação, regressão, agrupamento, regras de associação, visualização e previsão. O WEKA trabalha com arquivos no formato ARFF (*Attribute-Relation File Format*). O ARFF é um arquivo texto contendo um conjunto de observações, precedido por um pequeno cabeçalho (Figura 3). O cabeçalho é utilizado para fornecer informações a respeito dos campos que compõem o conjunto de observações.

```
%% Preço do acucar
%% 2006-2019.

@relation acucar1
@attribute Açucar_Preço numeric
@attribute Date date 'yyyy-MM'

@data

0.8,2006-01
0.87,2006-02
0.82,2006-03
```

Figura 3 – Arquivo ARFF.
Fonte - Autoria própria (2020).

Algoritmos:

Os algoritmos, do software WEKA, utilizados neste trabalho, são:

Sequential Minimal Optimization (SMOreg): O algoritmo SMOreg implementa a máquina de vetor de suporte para regressão (SVR). O SVR é uma extensão da Máquina de Vetores de Suporte (SVMs). SVMs são algoritmos de aprendizado de máquinas baseados na teoria de aprendizagem estatística. O SMOreg utiliza o algoritmo de otimização SMO e está disponível no WEKA (SHEVADE et al., 2000; SANTOS, 2016).

Multilayer Perception (MLP): As redes MLPs são muito utilizadas em mineração de dados. As redes neurais são compostas por unidades de processamento simples chamadas neurônios. Os neurônios são dispostos em camadas interligadas por conexões. As conexões são associadas a pesos. Seu treinamento é supervisionado e utiliza o algoritmo *backpropagation*. O *backpropagation* é o algoritmo responsável pelo aprendizado da rede. As redes MLPs podem realizar tanto regressão quanto classificação (PINHEIRO, 2020; HAYKIN, 2001).

Métricas:

Neste trabalho, os modelos obtidos foram avaliados pelas seguintes métricas (CANKURT; SUBASI, 2015):

- MSE (*Mean Squared Error*): É uma forma de avaliar a diferença entre um estimador e o verdadeiro valor da quantidade estimada.
- RSME (*Root Mean Squared Error*): Penaliza mais os erros maiores, permite avaliar a qualidade de um previsor em relação aos dados.
- MAPE (*Mean Absolute Percent Error*): O MAPE mede o tamanho do erro em termos percentuais.

Na Tabela 2 apresentam-se as equações das métricas MSE, RMSE e MAPE.

Tabela 2 – Equações – MSE, RMSE e MAPE.

Sigla	MSE	RMSE	MAPE
Equação	$\frac{1}{n} \sum_{i=1}^n ((y_i - \hat{y}_i)^2)$	$\sqrt{\frac{1}{n} \sum_{i=1}^n ((y_i - \hat{y}_i)^2)}$	$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)/y_i \times 100$

Fonte - RAMEDANI (2012).

Onde: y_i é o valor real do período i , \hat{y}_i é a previsão para o período i e n é o número de observações.

Parâmetros:

Os parâmetros de configuração, dos dois algoritmos do software WEKA, são apresentados na Figura 4.

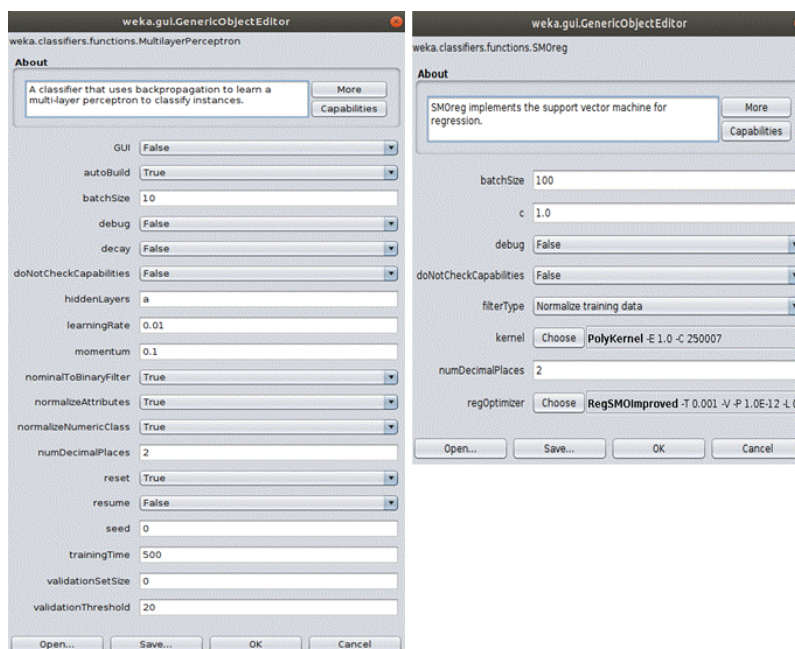


Figura 4 – Parâmetros de Configuração.

Fonte - Autoria própria (2020).

3. RESULTADOS E DISCUSSÃO

Inicialmente, neste trabalho, realizou-se uma análise descritiva dos dados (Tabela 3).

Tabela 3 – Análise descritiva dos dados.

Parâmetro analisado	Preço do açúcar
Média (R\$/kg)	0,907
Mínimo (R\$/kg)	0,38
Máximo (R\$/kg)	1,56
Desvio Padrão (R\$/kg)	0,265
Coefficiente de Variação (%)	29,2

Fonte - Autoria própria (2020).

Pode-se observar, dos dados apresentados na Tabela 3, que o preço do açúcar ficou, para o período em estudo, em média de 0,907 R\$/kg. Apresentando, neste período, preços mínimo e máximo de 0,38 R\$/kg e 1,56 R\$/kg. Observa-se também um coeficiente de variação dos dados de 29,2%, considerado alto de acordo com Pimentel (2009). Observa-se que quanto maior o coeficiente de variação maior a variação nos dados. Esta variação, nos preços do açúcar, pode ser observada na série apresentada na Figura 2.

Resultados – treinamento e teste:

Na Tabela 4 apresentam-se os resultados, para quatro passos a frente (4-Step-Ahead (4-S-A)), obtidos das métricas para os dois algoritmos.

Tabela 4 – Métricas – MAPE, RMSE e MSE (Conjunto de Teste).

	MLP				SMOreg			
Métricas	1-S-A	2-S-A	3-S-A	4-S-A	1-S-A	2-S-A	3-S-A	4-S-A
MAPE	5,85	8,72	11,47	13,91	5,37	8,19	10,63	12,96
RMSE	0,07	0,11	0,14	0,16	0,07	0,10	0,13	0,16
MSE	0,005	0,012	0,020	0,027	0,005	0,010	0,018	0,025

Fonte - Autoria própria (2020).

Observa-se, da Tabela 4, que os resultados obtidos, pelos dois modelos, são muito similares para o conjunto de teste. Neste trabalho, utilizou-se 80% do conjunto de dados para treinamento e 20% para teste. Nas Figuras 5 e 6 apresentam-se as curvas dos dados reais e preditos para os conjuntos de treinamento e de teste para os dois modelos.

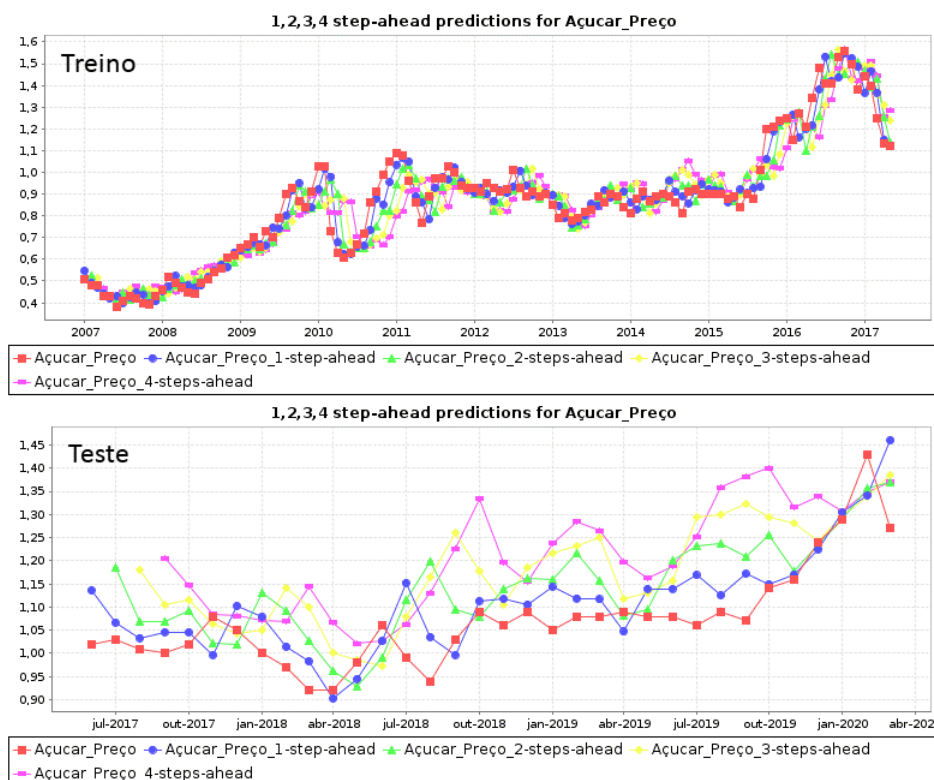


Figura 5 – Preço do Açúcar – SMOreg (Treino e Teste).

Fonte - Autoria própria (2020).

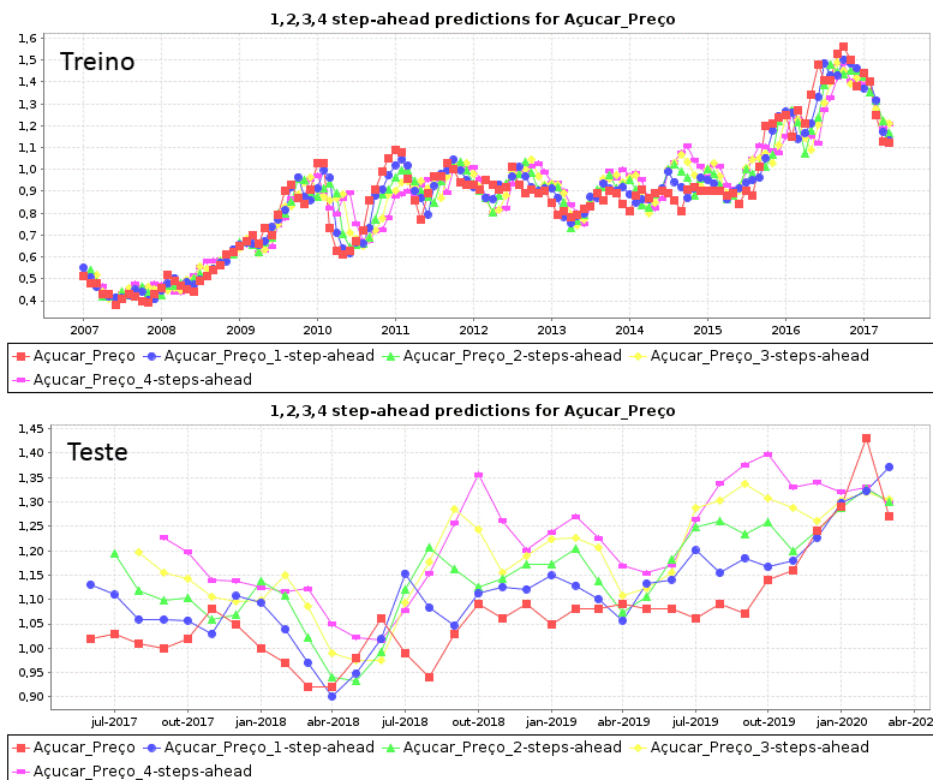


Figura 6 – Preço do Açúcar – MLP (Treino e Teste).
 Fonte - Autoria própria (2020).

Previsões:

Na Tabela 5 apresentam-se os dados, observados, preditos e os Erros Relativos Percentuais (ERP), para os quatro meses que não participaram da etapa de treino e teste. O ERP é obtido por meio da equação:

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

Tabela 5 – Previsões do preço do açúcar (R\$/kg) e Erros percentuais relativos (%).

Mês	IndexMundi	MLP	SMOReg	ERP - MLP	ERP - SMOReg
abr/20	1,22	1,25	1,27	2,45901639	4,098360656
mai/20	1,36	1,29	1,23	5,14705882	9,558823529
jun/20	1,41	1,34	1,26	4,96453901	10,63829787
jul/20	1,42	1,43	1,46	0,70422535	2,816901408
Média=				3,31870989	6,778095866

Fonte - Autoria própria (2020).

Os resultados das previsões e dos erros percentuais, em termos gráficos, são apresentados nas Figuras 7 e 8.

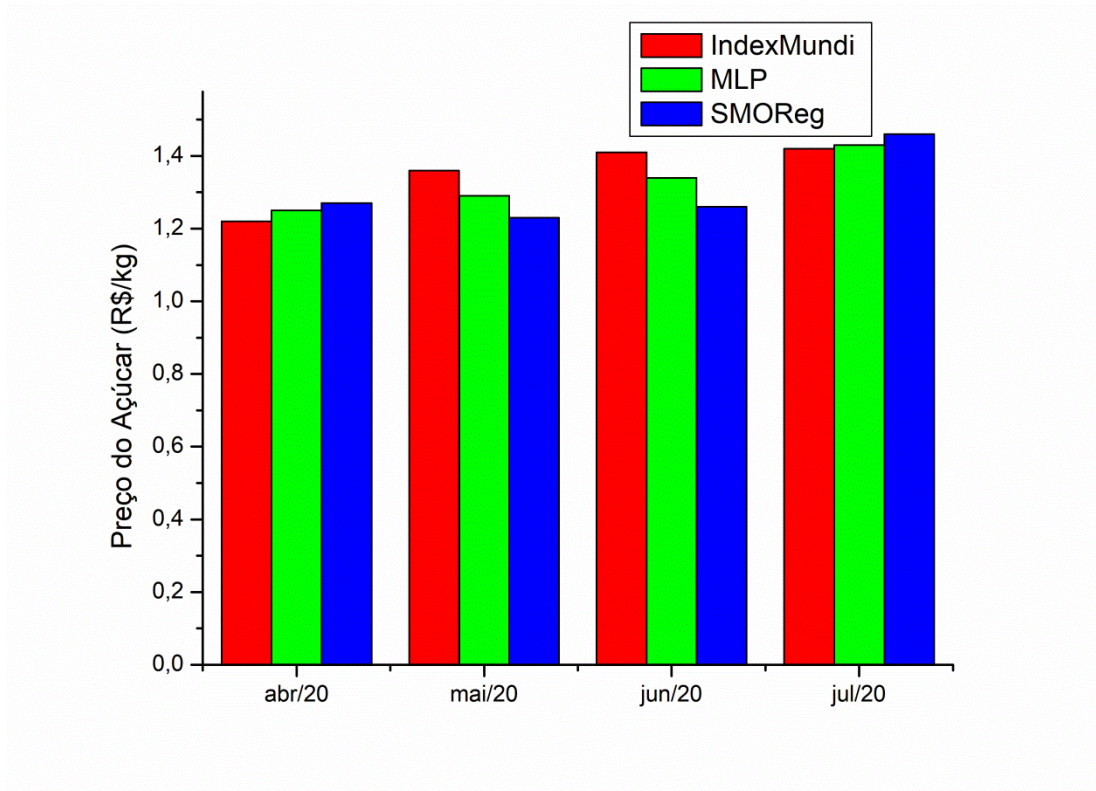


Figura 7 – Preço do Açúcar – Previsões.
 Fonte - Autoria própria (2020).

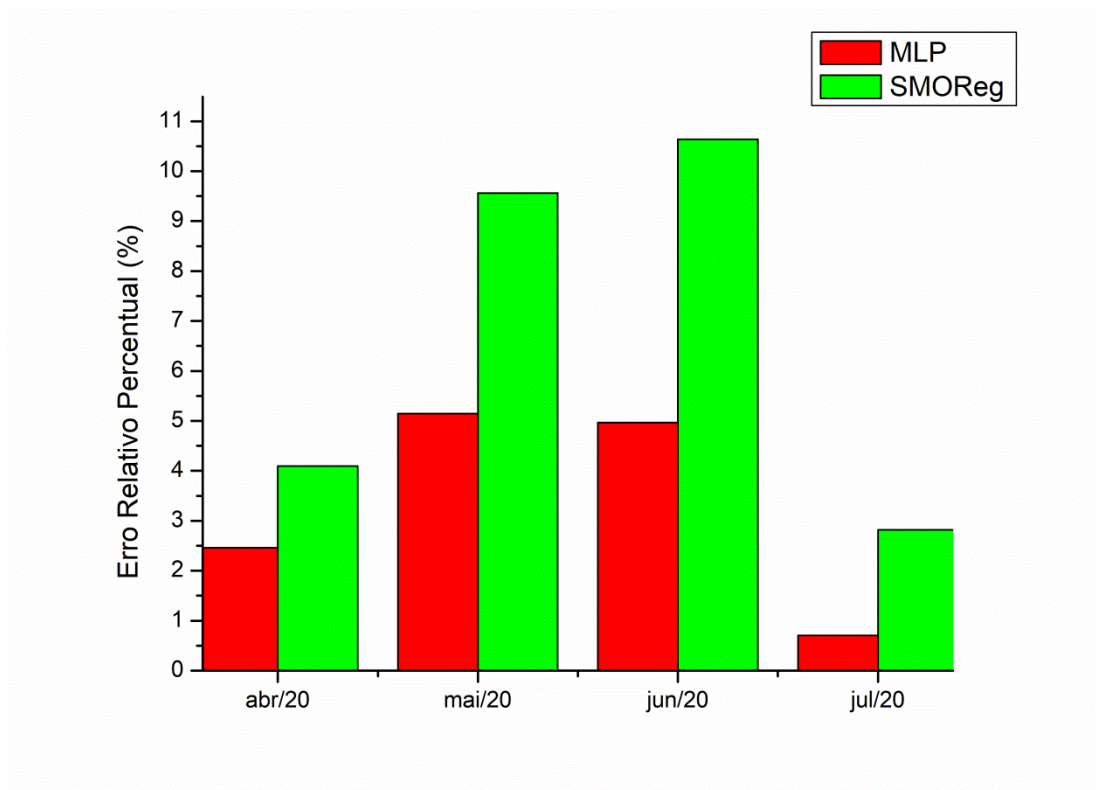


Figura 8 – Preço do Açúcar – Erros percentuais Relativos.
 Fonte - Autoria própria (2020).

Por meio dos resultados apresentados, na Tabela 5 e nas Figuras 7 e 8, conclui-se que os resultados das previsões, dos dois modelos, estão próximos aos fornecidos pelo IndexMundi. Contudo, o modelo MLP apresenta um erro percentual relativo médio menor que o modelo SMOReg (ERP-MLP= 3,318% e ERP-SMOReg= 6,778%).

4. CONCLUSÕES

Neste trabalho abordou-se a aplicação de dois algoritmos, de previsão de séries temporais do software WEKA, na previsão futura do preço do açúcar. Os modelos passaram pelas fases de: preparação de dados, definição das estruturas, estimativas, avaliação dos resultados e validação.

Inicialmente, realizou-se uma comparação entre o modelo MLP e o modelo SMOReg. O desempenho dessas duas técnicas foi avaliado usando medidas de erro estatístico padrão MSE, MAPE e RMSE. Os resultados mostraram que os modelos MLP e SMOReg apresentaram resultados similares.

Na sequência, observou-se, para os meses que não participaram do treinamento da rede (abril, maio, junho e julho de 2020), que as previsões foram bem precisas e as diferenças entre valores reais e preditos foram pequenas. Portanto, a proximidade entre valores preditos e reais demonstram a boa capacidade de generalização, para um horizonte de curto prazo, dos modelos utilizados neste trabalho. Não obstante, observou-se uma superioridade na qualidade do ajuste do modelo MLP.

Por fim, pode-se observar que ambos os modelos são válidos e podem auxiliar na minimização dos riscos e incertezas do processo de tomada de decisão do setor açucareiro.

REFERÊNCIAS

- ADVFN *Introdução ao açúcar*. Disponível em: <https://br.advfn.com/investimentos/commodities/acucar/introducao>>. Acesso em 30/9/2020.
- BACCHI, M. R. P.; ALVES, L. R. A. Formação de preço do açúcar cristal empacotado ao varejo da região centro-sul do Brasil. *Agric. São Paulo*, 51(1), 2004.
- BADER ESAM, A. Economic Modelling and Forecasting of Sugar Production and Consumption in Egypt. *International Journal of Agricultural Economics*, 2(4), 2017.
- HAYKIN, S. *Neural networks: a comprehensive foundation*. New Delhi: Pearson Prentice Hall, 2001.
- INDEXMUNDI *Preço do Açúcar*. Disponível em: <https://www.indexmundi.com/pt/preços-de-mercado/?mercadoria=açúcar&meses=180&moeda=brl>. Acesso em 30/9/2020.
- MEHMOOD, Q.; SIAL, M.; RIAZ, M.; SHAHEEN, N. Forecasting the production of sugarcane crop of Pakistan for the 2018-2030, using Box-Jenkins Methodology. *Journal of Animal and Plant Sciences*, 2019
- MOORE, D. S.; MCCABE, G. P.; DUCKWORTH, W. M.; SCLOVE, S. L. A. *A prática da estatística experimental: como usar dados para tomar decisões*. Rio de Janeiro: LTC, 2006.

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

NICOLA, D. S.; FREITAS, C. A.; PAZ, M. V. Previsão de preços do açúcar e análise de sua volatilidade no mercado futuro brasileiro (2003-2007): Uma aplicação da família ARCH. *Research in Agricultural & Applied Economics*, 2008.

PELLUCCI, P. R. S.; PAULA, R. R.; SILVA, W. B. O.; LADEIRA, A. P. Utilização de técnicas de aprendizado de máquinas no reconhecimento de entidades nomeadas no português. *E-xacta*, 2011.

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

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

RAMEDANI Z. A method based on neural networks for generating solar radiation map. *International Journal of Energy and Environment*, 3(5), 2012.

ROLIM, V. B.; MELLO, R. F. L.; Costa E. B. Utilização de técnicas de aprendizado de máquinas para acompanhamento de fóruns educacionais. *Revista Brasileira de Informática na Educação*, 2017.

RUSSEL, I.; MARKOV, Z. An Introduction to the Weka Data Mining System, *ACM SIGCSE*, 2017.

SANTOS, C. B.; PILATTI, L. A.; PEDROSO, B. CARVALHO, D. R.; GUIMARÃES, A. M. Previsão do Índice de Desenvolvimento Humano e da expectativa de vida nos países da América Latina por meio de técnicas de mineração de dados. *Ciência & Saúde Coletiva*, 23(11), 2018.

SATOLO, L. F.; BACCHI, H. R. Dinâmica econômica das flutuações na produção de cana-de-açúcar. *Economia Aplicada*, 13(3), 2009.

SHEVADE, S. K. Improvements to the SMO algorithm for SVM regression, *Neural Networks*, *IEEE Transactions on Neural Networks*, 11(5), 2000.

SOARES, L. D. S.; LUNKES, S. A.; SPLANCERSKI, J. S.; JÚNIOR, E. S., SANTOS, J. A. A.; GIRARDI, J. C. Aplicação do método de Box-Jenkins na previsão da demanda do bagaço da cana de açúcar. *Revista Brasileira de Energias Renováveis*, 2019.