

Aplicação de redes neurais artificiais na previsão do produto interno bruto do Mato Grosso do Sul em função da produção de cana-de-açúcar, açúcar e etanol

A produção de cana-de-açúcar e de seus derivados contribui significativamente para o aumento do Produto Interno Bruto do estado de Mato Grosso do Sul (MS). Gestores públicos e executivos do setor necessitam de informações para que suas decisões relacionadas à produção, tributação, entre outras possam ter melhor qualidade. Para isto, as técnicas de previsão contribuem neste processo. Diante do exposto, este artigo buscou responder qual a capacidade preditiva das Redes Neurais Artificiais (RNAs) para explicar o Produto Interno Bruto de MS em função das variáveis independentes, produção de cana-de-açúcar, açúcar e etanol? Assim, o objetivo geral do trabalho é estimar o Produto Interno Bruto de MS relativo a produção de cana-de-açúcar, de açúcar e de etanol utilizando as RNAs. Como objetivos específicos: verificar as correlações entre as variáveis envolvidas na produção de açúcar e etanol a partir da cana-de-açúcar; construir, treinar e validar um modelo de RNAs com estas variáveis; utilizar o modelo construído e validado para realizar previsões. A relevância da pesquisa é a utilização desta importante ferramenta de previsão para poder prever o fluxo monetário que será demandado com o incremento da produção da cana-de-açúcar, do açúcar e do etanol, podendo planejar investimentos no sentido de propiciar melhores condições de vida à população, não só nos entornos das usinas, mas no Estado que será beneficiado como um todo. Observando os dados coletados e os objetivos propostos, este trabalho é classificado como exploratório descritivo. A série temporal em análise compreende o intervalo de 2002 a 2015. O software utilizado para construir a RNA foi o SPSS - Statistical Package for the Social Sciences - Pacote Estatístico para as Ciências Sociais. Concluiu-se que o modelo predito de RNA foi capaz de explicar o comportamento do PIB do MS em função das variáveis independentes descritas, além disso, o modelo apresentou valores e observações satisfatórias para sua validação. Por fim, utilizando dados hipotéticos para as variáveis independentes, o modelo predito e validado foi utilizado para realizar uma previsão com significativa segurança.

Palavras-chave: Produção de Cana-de-açúcar em MS; Produção de açúcar em MS; Produção de etanol em MS; Predição do PIB de MS com RNA.

Application of artificial neural networks in Mato Grosso do Sul's gross domestic product forecast as a function of sugarcane, sugar and ethanol production

The production of sugar cane and its derivatives contributes significantly to the increase of the Gross Domestic Product of the state of Mato Grosso do Sul (MS). Public managers and industry executives need information so that their decisions related to production, taxation, among others can have better quality. For this, the prediction techniques contribute to this process. Given the above, this paper sought to answer what is the predictive capacity of Artificial Neural Networks (ANNs) to explain the Gross Domestic Product of MS as a function of independent variables, sugarcane production, sugar and ethanol? Thus, the main objective of this work is to estimate the Gross Domestic Product of MS related to sugarcane, sugar and ethanol production using ANNs. Specific objectives: to verify the correlations between the variables involved in sugar and ethanol production from sugarcane; build, train and validate an ANN model with these variables; use the constructed and validated model to make predictions. The relevance of the research is the use of this important prediction tool to predict the cash flow that will be demanded with the increase of sugarcane production, sugar and ethanol, and can plan investments to provide better living conditions. To the population, not only around the mills, but in the state that will benefit as a whole. Observing the collected data and the proposed objectives, this work is classified as exploratory descriptive. The time series under analysis comprises the range from 2002 to 2015. The software used to construct the ANN was the Statistical Package for the Social Sciences (SPSS). It was concluded that the ANN predicted model was able to explain the behavior of the MS GDP as a function of the independent variables described. Moreover, the model presented satisfactory values and observations for its validation. Finally, using hypothetical data for the independent variables, the predicted and validated model was used to make a prediction with significant security.

Keywords: Sugarcane Production in MS; Sugar Production in MS; Ethanol Production in MS; GDP Prediction of MS with RNA.

Topic: **Desenvolvimento, Sustentabilidade e Meio Ambiente**

Received: **10/08/2019**

Approved: **28/09/2019**

Reviewed anonymously in the process of blind peer.

João Bosco Arbués Carneiro Júnior 
Universidade Federal de Mato Grosso, Brasil
<http://lattes.cnpq.br/5073023761658236>
<http://orcid.org/0000-0001-7094-2245>
jbacj@hotmail.com

Celso Correia de Souza 
Universidade Anhuaguera, Brasil
<http://lattes.cnpq.br/2881392515816773>
<http://orcid.org/0000-0002-2689-8264>
csouza939@gmail.com



DOI: 10.6008/CBPC2179-6858.2019.005.0019

Referencing this:

CARNEIRO JÚNIOR, J. B. A.; SOUZA, C. C.. Aplicação de redes neurais artificiais na previsão do produto interno bruto do Mato Grosso do Sul em função da produção de cana-de-açúcar, açúcar e etanol. **Revista Ibero Americana de Ciências Ambientais**, v.10, n.5, p.218-230, 2019.
DOI: <http://doi.org/10.6008/CBPC2179-6858.2019.005.0019>

INTRODUÇÃO

Aproveitando sua vocação para o agronegócio, Mato Grosso do Sul (MS) caracterizado por uma atividade pecuária forte, possui também um papel importante na produção de cana-de-açúcar e seus derivados. Defante et al. (2018) analisaram a expansão da cana-de-açúcar no Brasil e no Mato Grosso do Sul, em que se verificou que a área plantada em MS cresceu 500% entre 2004 e 2015, sendo que a cana-de-açúcar foi plantada em áreas que antes eram ocupadas pela atividade da pecuária. Mais de 75% do total da produção é processada pelas 22 usinas e transformada em açúcar e álcool.

Segundo o UDOP, Mato Grosso do Sul possui a 4ª maior produção de cana-de-açúcar entre os estados do Brasil. A safra 18/19 atingiu 49,5 milhões de toneladas, apresentando um crescimento de 5,4% em relação à safra anterior. Deste total, 84% se destina a produção de álcool e de 16% para produção de açúcar. A atividade canieira está presente em 39 municípios deste estado, gerando 32.191 empregos e gerando R\$834 milhões em salários.

Em virtude da produção do álcool no Brasil ser obtida, preferencialmente, da cana-de-açúcar, tem-se também a opção da produção de açúcar que é destinada, além do consumo interno, para a exportação, gerando altas divisas ao país. Ou seja, através do cultivo da cana-de-açúcar, vários produtos são gerados, de modo que, neste trabalho, foram utilizados, além da própria cana-de-açúcar, o açúcar e o álcool. A previsão de produção de álcool favorece a tomada de decisão no sentido de planejar o volume produção para o consumo ao longo dos períodos subsequentes, principalmente agora, momento de transição da era fóssil para a era renovável.

Diante disso, inúmeras técnicas são empregadas para realizar as predições, pode-se citar desde os modelos tradicionais tais como modelo de regressão linear, séries temporais e dados em painel (modelos econométricos) até o uso de modelos de inteligência artificial, particularmente, neste caso, cita-se especificadamente as Redes Neurais Artificiais.

As Redes Neurais Artificiais (RNAs) vêm sendo empregadas com sucesso na previsão de séries temporais por sua capacidade de aproximar funções não-lineares (FERNANDES et al., 1996; CALÔBA et al., 2002; TORRES et al., 2005). Diante do exposto, questiona-se: Qual a capacidade preditiva das RNAs para explicar o Produto Interno Bruto de Mato Grosso do Sul em função da produção de cana-de-açúcar, açúcar e álcool?.

Dessa forma, o objetivo deste trabalho foi estimar a participação no Produto Interno Bruto de Mato Grosso Sul do volume de produção de cana-de-açúcar, de açúcar e de etanol com a utilização de Redes Neurais Artificiais. Para auxiliar na consecução desse objetivo, pretende-se: verificar as correlações entre as variáveis envolvidas na produção de açúcar e etanol a partir da cana-de-açúcar; construir, treinar e validar um modelo de RNAs com estas variáveis; utilizar o modelo construído e validado para realizar previsões.

Com a utilização desta importante ferramenta de predição do valor do PIB de MS em função da produção de cana-de-açúcar, de açúcar e de etanol, o Estado terá melhores condições de planejar a expansão da área plantada de cana-de-açúcar, como também, os locais dessa expansão, facilitando a proposição de

políticas públicas visando o ordenamento dessa expansão e a proteção do meio ambiente que, de uma maneira ou de outra, vai ser afetado por essa expansão. O Estado também pode prever o fluxo monetário que será demandado com o incremento da produção de açúcar e de etanol, podendo planejar investimentos no sentido de propiciar melhores condições de vida à população, não só no entorno das usinas, mas no Estado que será beneficiado como um todo. No geral, realizar previsões de produção e de gestão sobre a cana-de-açúcar e seus derivados é de grande valia para os atores da sociedade, pois, direciona políticas públicas e auxilia o processo de tomada de decisão de produtores e gestores públicos e privados.

REVISÃO TEÓRICA

As políticas públicas de incentivos à expansão das agroindústrias canavieiras (usinas) no Brasil favoreceram a instalação dessas unidades em determinados locais do país (TSUKADA, 2011), surgindo oportunidades para o desenvolvimento local das regiões onde estão sendo instaladas as usinas. Por outro lado, iniciaram-se discussões pela sociedade com a intenção de garantir, ao mesmo tempo as instalações com desenvolvimento ambiental, social e econômico (CAMPEÃO et al., 2009). Uma dessas garantias, segundo Manzatto et al. (2009), surgiu com o Zoneamento Agroecológico da cana-de-açúcar (ZAE cana), que destinou mais de 60 milhões de hectares para expansão e produção sustentável de cana-de-açúcar no Brasil, sem afetar diretamente áreas destinadas à produção de alimentos, reservas florestais e/ou nativas.

A expansão canavieira se deslocou para a região Centro-Oeste no início desse século. No estado de MS esse fato aconteceu entre os anos 2000 e 2012, em que foram implantadas 16 novas usinas sucroalcooleiras em vários municípios. A implantação dessas unidades produtoras foi consequência de algumas estratégias tais como: da produção de veículos flex fuel no Brasil; o barateamento dos custos logísticos e de armazenagem; região com grandes áreas aptas ao plantio mecanizado; condições climáticas e recursos hídricos (SANT'ANNA et al., 2016), além da busca por energia sustentável; segurança alimentar e elevação do preço da terra nos tradicionais estados produtores (SHIKIDA, 2013).

Existem várias pesquisas em relação à expansão canavieira no Brasil e, em particular, em MS (RUDORFF et al., 2010; COUTINHO et al., 2017), sobre o crescimento econômico motivado pela expansão e, ainda, sobre sustentabilidade canavieira. Inúmeros são os autores que tratam do assunto, no entanto, este artigo procurou determinar a previsão do PIB de MS com o aumento na produção de açúcar e álcool no estado de MS. Este estudo está focado na hipótese de que a evolução do PIB de MS pode ser explicado, em parte, pela produção de açúcar e álcool em MS através de métodos estatísticos de correlação e regressão, bem como, redes neurais artificiais, utilizando variáveis explicativas ligadas à produção desses produtos.

O crescimento da cana-de-açúcar em MS tem ficado aquém das estimativas e expectativas dos agentes econômicos (SOUZA et al., 2010), e das previsões do governo federal no PNA (BRASIL, 2006). Mesmo com os aspectos vantajosos para a expansão da cana nas áreas do cerrado, persistem dificuldades locais para que as metas sejam alcançadas, como apontam Ávila (2009) e Santos (2011), sobre as dificuldades em questões de mudanças da organização produtiva local, dependência de repasses para municípios que apenas

fornece a cana e não recebe as indústrias e perda de dinamismo agrícola a partir da concentração da terra e da renda.

Ainda, Ávila (2009) e Santos (2011) consideram que as dificuldades contratuais, a geração de externalidades (danos ambientais, piora de serviços de saúde e educação) e a sazonalidade da mão de obra inibem a expansão, enquanto as perspectivas de ganho econômico e de saída de endividamento são fatores que atraem a atividade.

Redes Neurais Artificiais (RNA)

As redes neurais artificiais (RNAs) são definidas como sistemas computacionais de processamento paralelo com capacidade de armazenar conhecimento empírico e se assemelha ao cérebro humano por apresentarem dois principais fatores: capacidade de aprendizagem e capacidade de generalização (HAYKIN, 1999).

As redes perceptron de múltiplas camadas (RNA MLP), também denominadas multilayer perceptron, são tipos de redes que são aplicadas a diferentes problemas combinatórios e na solução de diversas tarefas, como por exemplo: o processamento de informações, o reconhecimento de padrões, a previsão do tempo, problemas de classificação, processamento de imagens, dentre outros (VIEIRA et al., 2012; BRAGA et al., 2014).

A RNA MLP possui uma estrutura composta basicamente por três camadas ou mais camadas, sendo, respectivamente, de uma camada de entrada, uma ou mais camadas intermediárias ou ocultas e uma camada de saída. Na camada de entrada são inseridos os parâmetros das variáveis que serão preditoras no processo, ou seja, as variáveis independentes, na camada intermediária ocorrem o processamento dos neurônios e ajustes das funções e pesos sinápticos e, na camada de saída, estão os parâmetros a serem previstos, denominados de variáveis dependentes (Figura 1).

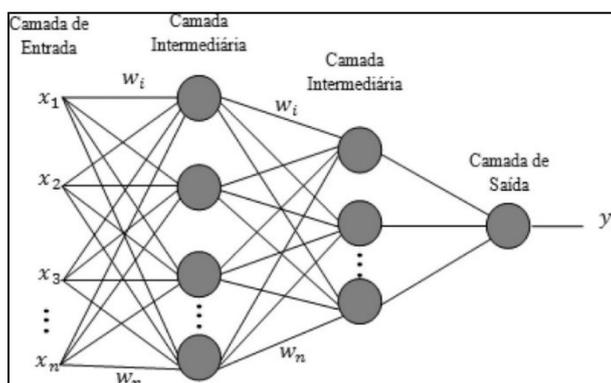


Figura 1: Arquitetura da Rede Neural Artificial MLP. **Fonte:** Adaptado de SILVA et al. (2010).

Observe da Figura 1 que cada um dos neurônios da camada de entrada se conecta a todos os outros neurônios da camada intermediária e estes por sua vez aos neurônios da camada de saída e, dependendo da estimulação gerada pela soma ponderada do produto dos neurônios de entrada por seus pesos sinápticos, a função de transferência e emissão dos sinais de saída serão diferentes (KOVÁCS, 2006; LITTA et al., 2013;

BRAGA et al., 2014).

Entre cada uma das camadas existem pesos denotados por w_i ($i = 1, 2, 3, \dots, n$), que estão associados às entradas dadas por x_i , donde se obtém o produto $w_i x_i$ e, dependendo do estímulo gerado pela soma ponderada das entradas pelos pesos, os neurônios ativam a função de transferência, que emite uma resposta para à próxima camada. A camada de saída, dada por y_i , recebe os estímulos da camada intermediária e gera um padrão de resposta. Portanto, as camadas intermediárias funcionam como extratoras de características dos padrões de entrada, codificadas pelos pesos e permitem que a rede represente o problema com os sinais de saída mais detalhados (HAYKIN, 1999).

Para que a RNA apresente resultados de saída é necessário aplicar o processo de aprendizagem da rede, cuja resposta é resultado do processo obtido do treinamento dessa rede através de um algoritmo. O algoritmo de aprendizagem adotado para a RNA MLP foi o de retropropagação dos erros, ou também denominado de backpropagation, que trata de uma técnica de treinamento supervisionada e realiza o ajuste dos pesos através do cálculo da diferença entre o valor previsto e o valor observado (BUSTAMI et al., 2007; DESHMUK et al., 2010; COUTNHO et al., 2016). Quando o resultado da rede atingir a otimização por meio da minimização dos erros quadráticos, o sinal de saída é validado e se deve então avaliar o desempenho da rede a partir das métricas de avaliação dos dados de saída com relação aos dados observados.

Avaliação de desempenho

A comparação entre os parâmetros previstos e os parâmetros observados requer a obtenção de métricas de avaliação dos dados que permita estimar a probabilidade de acerto das previsões geradas. De acordo com Hair Jr. et al. (2005), para avaliar a qualidade das previsões são calculados coeficientes que quantificam o grau de semelhança entre a previsão e a observação, cujas equações podem ser observadas no Quadro 1.

Quadro 1: Medidas estatísticas de análise de desempenho da RNA MLP a partir dos diferentes termos de eficiência calculados entre os valores previstos e os valores observados.

Métricas	Descrição	Equação
R²	Coeficiente de determinação do grau de ajuste	$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O})^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$
RMSE	Raiz do erro quadrado médio	$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{N}}$
MAPE	Erro percentual absoluto médio	$MAPE = \frac{\sum_{i=1}^n \left \frac{O_i - P_i}{O_i} \right }{N} \times 100$
C	Índice de confiança	$C = r \times \left[1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (P_i - \bar{O} + O_i - \bar{O})^2} \right]$

Onde n é o número total de elementos da amostra; O_i é o valor observado na posição i ; P_i é o valor previsto na posição i ; \bar{O} é a média dos valores observados.

O R^2 fornece a porcentagem da variância da variável resposta com relação às variáveis preditoras. A RMSE mede a amplitude do erro e é um parâmetro de precisão; o MAPE mede a magnitude do erro de previsão e o C é o critério que avalia o desempenho da modelagem de previsão relacionando o grau de dispersão dos dados em torno da média com a exatidão do modelo (HAIR JR. et al., 2005). Os parâmetros R^2

e C devem apresentar resultados ótimos para valores próximos de 1 e resultados ruins para valores próximos de 0. Já os erros MAPE e RMSE são melhores quanto mais próximos de zero.

MATERIAIS E MÉTODOS

Trata-se de pesquisa classificada como exploratória descritiva, pois, teve como finalidade, por meio de dados sobre a produção de açúcar, cana-de-açúcar e etanol, descobrir e observar fenômenos, procurando descrevê-los, classificá-los e interpretá-los, ao mesmo tempo em que buscou explorar, no problema de pesquisa, a tendência do comportamento das produções, se essas produções sofrem variações sazonais. Inicialmente, foi feito um levantamento bibliográfico dos assuntos que envolviam o tema em estudo, depois, foi feita uma revisão bibliográfica desses assuntos para a fixação dos conceitos utilizados, por fim, utilizou-se das RNAs para responder o problema de pesquisa.

O objeto de estudo desta pesquisa refere-se ao setor sucroalcooleiro (sucroenergético) no estado de MS, com delimitação temporal nas safras do período de 2002 a 2015. Os dados propostos para o estudo são dados secundários, e se referem à série histórica da produção anual de açúcar, etanol, cana-de-açúcar e o PIB do estado de Mato Grosso do Sul coletados no período de 2002 a 2015, num total de 14 observações. Esses dados foram obtidos junto ao banco de dados do Instituto de Pesquisas Econômicas Aplicadas (IPEA), do Ministério do Planejamento Orçamento e Gestão (MPOG) e da Secretaria de Estado de Meio Ambiente, Desenvolvimento Econômico, Produção e Agricultura Familiar (SEMAGRO/MS).

Para implementar a RNA se utilizou do Software Statistical Package for the Social Sciences (SPSS). A RNA utilizada foi a do tipo Perceptron Multiestrato (MLP). As variáveis utilizadas no modelo foram: Produto Interno Bruto de Mato Grosso do Sul (PIBMS) como variável dependente, produção de cana-de-açúcar (ProdCana), produção de etanol (ProdEtanol) e produção de açúcar (ProdAcucar), com variáveis independentes. Todas as variáveis são dos tipos escalares.

As variáveis independentes foram tratadas de forma padronizada, ou seja, a média é subtraída do valor observado e o resultado é dividida pelo desvio-padrão. As amostras de testes e de treinamento são estimadas e otimizadas diretamente pelo software, não tendo o pesquisador ação sobre esta seleção. A arquitetura de rede, bem como, a quantidade de camadas ocultas também são selecionadas automaticamente pelo software SPSS. Para determinar a qualidade dos resultados obtidos pela RNA foram levados em consideração o grau de precisão das previsões consideradas, por dois critérios distintos: 1 - Erro percentual médio absoluto (MAPE), dado pela expressão (3); 2 - Erro quadrático médio de previsão (EQMP), expressão (4).

$$MAPE = \frac{1}{h} \sum_{i=1}^h \left| \frac{y_i - y_{pi}}{y_i} \right| \quad (3)$$

$$EQMP = \frac{1}{h} \sum_{i=1}^h (y_i - y_{pi})^2 \quad (4)$$

em que: y_i é o valor real do período i ; y_{pi} é a previsão para o período i e; h é o número de previsões. Por fim, após a construção e validação da RNA se utilizou de dados aleatórios de produção de cana-de-açúcar, etanol e açúcar para prever um valor predito do PIB para o estado de Mato Grosso do Sul com base nesta importante ferramenta de inteligência artificial.

RESULTADOS E DISCUSSÃO

As variáveis trabalhadas nesta pesquisa foram levantadas no período de 2002 a 2015 conforme observa-se na tabela 1. Utilizou-se da correlação linear de Pearson (r) para verificar se existe alguma associação entre as variáveis PIBMS, ProdCana, ProdEtano e ProdAcucar, conforme tabela 2.

Tabela 1: Variáveis PIBMS, ProdCana, ProdAcucar e ProdEtanol, no período de 2002 a 2015

Ano	PIB MS (R\$)	ProdCana (t)	ProdAcucar (t)	ProdEtanol (m³)
2002	48.819.051.364	8.247.056	373.715	418.157
2003	51.998.492.005	8.885.792	402.878	472.793
2004	51.607.223.209	9.475.155	508.783	533.580
2005	52.933.204.633	9.037.918	400.857	495.591
2006	55.954.901.435	11.635.096	574.009	640.843
2007	58.601.664.250	14.869.066	616.196	876.772
2008	61.728.534.407	18.090.388	657.078	1.082.882
2009	62.178.454.387	23.111.237	715.637	1.190.733
2010	69.453.752.752	33.519.668	1.328.546	1.846.197
2011	71.849.268.409	33.859.650	1.581.491	1.589.726
2012	76.158.966.838	37.330.449	1.741.908	1.910.298
2013	81.183.680.932	41.496.041	1.366.308	2.230.545
2014	83.309.239.574	44.683.982	1.331.484	2.402.138
2015	83.082.335.621	47.816.602	1.263.758	2.610.908

Fonte: SEMAGRO (2018); UNICA (2017).

Tabela 2: Correlação linear de Pearson (r) entre as variáveis PIBMS, ProdCana e ProdEtanol, no período de 2002 a 2015.

		ProdCana	ProdAcucar	ProdEtanol	PIBMS
ProdCana	Correlação de Pearson	1	0,904**	0,995**	0,990**
	Sig. (bilateral)		0,000	0,000	0,000
	N	14	14	14	14
ProdAcucar	Correlação de Pearson	0,904**	1	0,871**	0,891**
	Sig. (bilateral)	0,000		0,000	0,000
	N	14	14	14	14
ProdEtanol	Correlação de Pearson	0,995**	0,871**	1	0,989**
	Sig. (bilateral)	0,000	0,000		0,000
	N	14	14	14	14
PIBMS	Correlação de Pearson	0,990**	0,891**	0,989**	1
	Sig. (bilateral)	0,000	0,000	0,000	
	N	14	14	14	14

** A correlação é significativa no nível 0,01 (bilateral).

As variáveis apresentaram alta correlação positiva entre si, valores acima de 0,7. Esse dado já era esperado, uma vez que a produção de cana-de-açúcar, açúcar e etanol são mesmo altamente correlacionadas

entre si, ou seja, a primeira tem relação direta sobre a segunda, que pode ser maior ou menor a depender do mix de produção escolhido pelas unidades produtoras, tal qual, o açúcar ou etanol. Vale aqui mencionar que se a técnica utilizada nesta pesquisa fosse o modelo de regressão linear, o modelo estaria prejudicado, haja vista que as variáveis independentes possuem forte correlação entre si, porém, com o uso das RNAs esta situação não gera problemas à predição.

Após a análise de correlação utilizou-se do software SPSS para construir e validar o modelo de RNAs. A primeira etapa consistiu em determinar o tipo de RNA a ser trabalhado, sendo que foi escolhida a Perceptron Multicamada, com uma camada oculta, As variáveis foram tratadas de forma padronizadas. A tabela 3 apresenta o resumo de processamento de casos da RNA predita.

Tabela 3: Resumo de processamento de casos da RNA.

		N	Porcentagem
Amostra	Treinamento	11	78,6%
	Testes	3	21,4%
Válido		14	100,0%
Excluídos		0	0%
Total		14	100,0%

Como tratado em material e métodos, de maneira aleatória, o próprio software determina a quantidade de variáveis utilizadas como treinamento e também como teste. Nota-se que foi utilizado percentuais distintos do padrão 70% para treinamento e 30% para testes. A tabela 4 apresenta as informações da rede, tratando das características de sua construção.

Tabela 4: Informações de rede.

Camada de entrada	Covariáveis	1	Prod Cana
		2	Prod Açúcar
		3	Prod Etanol
	Número de unidades	3	
Método de reescalonamento para covariáveis		Padronizado	
Camadas ocultas	Número de camadas ocultas		1
	Número de Unidades na Camada Oculta 1		2
	Função de ativação		Tangente hiperbólica
Camada de saída	Variáveis dependentes	1	PIBMS
	Número de unidades		1
	Método de reescalonamento para dependentes de escala		Padronizado
	Função de ativação		Identicidade
	Função de erro		Soma dos Quadrados

A camada de entrada é formada pelas três variáveis independentes, ou covariáveis, produção de cana-de-açúcar, produção de açúcar e produção de etanol. As covariáveis foram reescaladas pelo método padronizado, no qual é subtraído a média e dividido pelo desvio-padrão. A camada oculta contém os nós de rede não observáveis. Neste estudo foi trabalhado com uma camada oculta e sua função de ativação foi a tangente hiperbólica que tem como característica usar argumentos com valores reais e os transforma no intervalo (-1,1). Na tabela 5 se observa este comportamento de associação das variáveis e os valores neste intervalo.

A camada de saída resulta na variável dependente, que neste estudo, foi o PIBMS, e como covariáveis as produções de cana, de açúcar e de etanol. O método de reescalonamento foi o padronizado, a função de

ativação foi a identidade, ou seja, utiliza valores reais e os retorna idênticos. Para medir a qualidade da RNA predita utiliza-se a soma dos quadrados como função de erro. Visualiza-se essas informações na figura 2 do diagrama de rede.

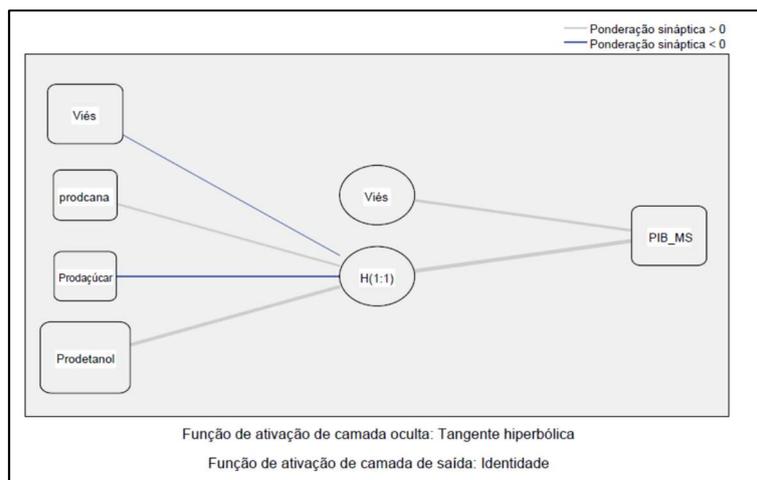


Figura 2: Diagrama de rede.

Na figura 2 nota-se a presença e o comportamento das ponderações sinápticas. As ponderações sinápticas representam a associação entre as variáveis com os ajustes realizados pelas RNAs e por todas as variáveis incluídas no modelo. A tabela 5 traz os valores das ponderações sinápticas do modelo predito.

Tabela 5: Estimativas de parâmetro.

Preditor		Predito	
		Camada oculta 1	Camada de saída
		H(1:1)	PIB_MS
Camada de entrada	(Viés)	-0,017	
	Prod Cana	0,089	
	Prod Açúcar	-0,066	
	Prod Etanol	0,712	
Camada oculta 1	(Viés)		0,146
	H(1:1)		1,760

Os valores preditos na camada de entrada foram tratados de forma padronizada, sendo assim, seus valores oscilam entre -1 e 1, já os valores da camada de saída foram tratados pela função identidade representando assim os valores de associação determinados. Uma das maiores desvantagens das RNAs é a complexidade de entendimento das ponderações sinápticas, pois, diferentes dos modelos tradicionais de previsão, não existe uma fórmula e/ou equação que represente os valores encontrados. Os valores apresentados na tabela 5 foram obtidos mediante o treinamento da RNAs em dado momento. Vale aqui mencionar, que o procedimento de efetivação de uma RNA é único, ou seja, o procedimento de treinamento da rede e os resultados gerados serão consequência da execução naquele momento. Devido a característica de aprendizagem, não se espera que resultados obtidos em execuções futuras sejam iguais, pois cada execução a RNA ‘aprende’ de maneira individual e não sistemática. Para garantir que o modelo construído possa ser utilizado em demais oportunidades, há a possibilidade de salvar e armazenar a RNA ‘treinada’. Em seguida, na tabela 6 é apresentada a sumarização do modelo, no qual se pode observar o erro do modelo, principal medida para a validação das RNAs.

Tabela 6: Sumarização do modelo.

Treinamento	Soma dos erros quadráticos	0,210
	Erro relativo	0,042
	Regra de parada usada	1 passos consecutivos sem diminuição de erros ^a
	Tempo de treinamento	0:00:00.00
Testes	Soma dos erros quadráticos	0,012
	Erro relativo	0,006

Variável Dependente: PIBMS. a. Os cálculos de erro têm como base a amostra de teste.

Como discutido na fundamentação teórica, o erro relativo é a soma dos quadrados dos erros da variável dependente. Na tabela 6 se nota que o erro do modelo predito é 0,006 ou 0,6%, ou seja, um valor relativamente baixo, o que evidencia que a RNA pode ser utilizada para predições com segurança. Em seguida, utilizou-se o cálculo do desvio-padrão da diferença entre o valor predito e o valor real do PIB de MS para medir o desvio encontrado entre estas variáveis. Com o auxílio do software SPSS o valor encontrado para esta medida estatística foi de R\$2.080.405.353,00 em valores absolutos e 3,24% em termos percentuais. Em RNAs esses valores do desvios-padrão são denominados RMSE, Raiz do Erro Quadro Médio.

Em sequência à validação da RNA estimada neste artigo, apresentam-se as figuras 3 e 4, as quais demonstram o comportamento do valor predito por cada valor observado, e o gráfico do valor dos resíduos por cada valor predito da variável dependente, respectivamente. Deseja-se que o comportamento dos valores preditos pelos valores observados deve apresentar linearidade. Na figura 3 se nota que os pontos estão dispersos, com linearidade visível.

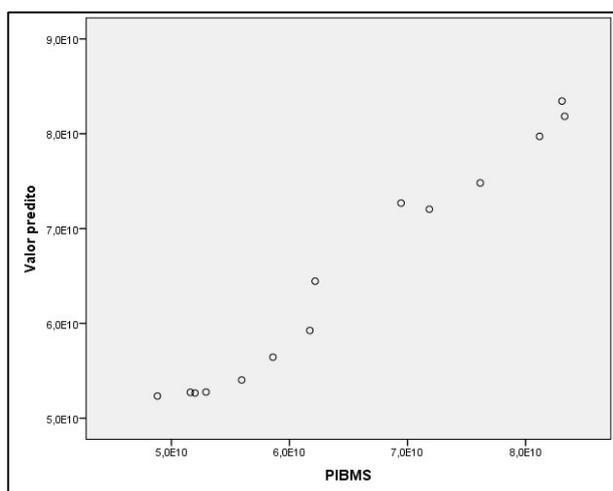


Figura 3: Gráfico do valor predito pelo valor observado.

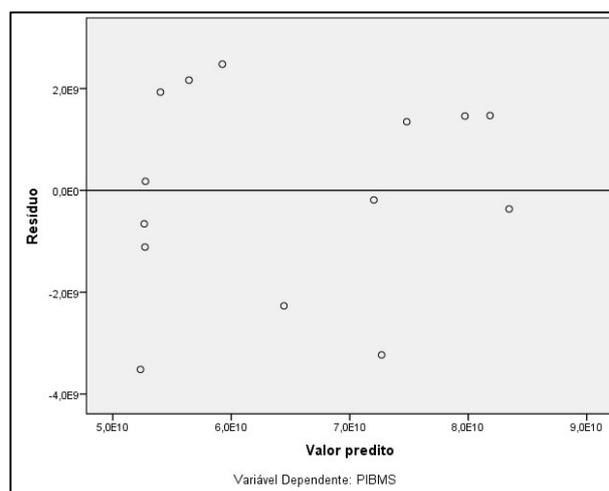


Figura 4: Gráfico dos resíduos por predito.

Na figura 4, a análise dos resíduos demonstra que a hipótese de normalidade foi atendida, pois observa-se no gráfico um comportamento em torno da linha horizontal centrada no zero, sem caracterizar tendência positiva ou negativa. Para finalizar as discussões sobre a validação do modelo predito, é apresentado na figura 5 e na tabela 7 informações referentes análise da importância das variáveis dependentes para a construção da RNA. Está evidenciado na tabela 7 que a variável Prod Cana é a que melhor contribuiu para a predição da RNA, tendo contribuído com 0,875 ou seja 87,5% na estimativa.

Tabela 7: Importância da variável independente.

	Importância	Importância normalizada
Prod Cana	0,875	100,0%
Prod Açúcar	0,096	11,0%
Prod Etanol	0,029	3,3%

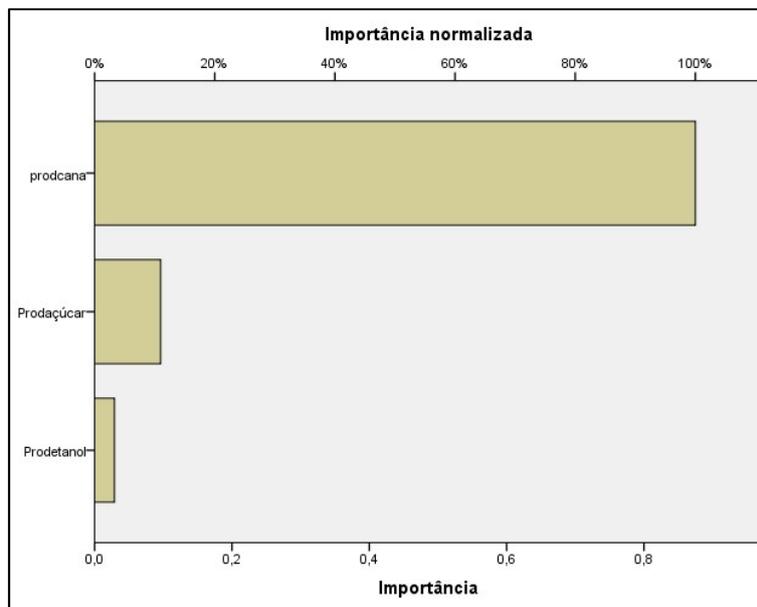


Figura 5: Gráfico da importância da variável independente.

A figura 5 demonstra a importância de cada variável para a RNA. Esta análise é efetuada com base nos valores utilizados para o treinamento e teste, nota-se que a produção de cana-de-açúcar foi utilizada em sua totalidade enquanto que os dados da produção de etanol contribuíram com 3,3%. Ressalta-se que, diferentemente de outros modelos de predição, principalmente os modelos econométricos que nessa situação a teoria sugere que uma variável de baixa importância deveria ser retirada do modelo, nas RNAs isto não é sugerido, pois como já dito, o modelo ‘aprendeu’ com estas três variáveis, em caso de retirada e/ou de acrescentar variáveis, o modelo e o comportamento das ponderações sinápticas mudam sem nenhum controle por parte do pesquisador.

Após a apresentação e a discussão da validação da RNA acredita-se que, com os dados encontrados e analisados, que o modelo possui capacidade de predição significativa, sendo assim, a RNA foi utilizada para predizer e comparar o valor do PIBMS para o ano de 2016, utilizando os valores de produção do ano de 2016 disponibilizados pela SEMAGRO, sendo estes: Prod Cana: 47.033.478 t; Prod Açúcar: 931.215 t e; Prod Etanol: 3.138.533 m³. Assim, na tabela 8 observa-se o funcionamento da RNA treinada com dados reais.

Inicialmente, observa-se na tabela 8, que os valores preditos não possuem relação com uma equação, como nos modelos econométricos, demonstra-se assim a função de aprendizado e as associações sinápticas, pois os resultados preditos pelas RNAs são obtidos por estas ponderações sinápticas através do treinamento na camada oculta. A inexistência de um modelo e/ou equação é considerada uma das principais desvantagens das RNAs, pois não há possibilidade de visualização do modelo, porém, verifica-se que este é sua principal característica, pois o aprendizado não é objetivo.

Observa-se que a maior diferença foi encontrada no ano de 2010, e a menor diferença foi no ano de

2015. O desvio-padrão da diferença foi de 3,24% em termos percentuais e R\$2.080.405.353 em termos monetários, valor relativamente baixo para este tipo de predição e para os valores nominais das variáveis. Nota-se o valor predito de R\$84.711.088.389 para o ano de 2016 obtido com os dados reais de produção para este ano com o intuito de testar a capacidade preditiva da RNA treinada.

Tabela 8: Valores preditos x valores observados.

Ano	PIB MS (R\$)	PIB MS PREV (R\$)	Diferença (R\$)	Dif. Percentual (%)
2002	48.819.051.364	51.565.496.166	-2.746.444.802	-5,63
2003	51.998.492.005	52.172.323.341	-173.831.336	-0,33
2004	51.607.223.209	52.745.880.295	-1.138.657.086	-2,21
2005	52.933.204.633	52.454.534.507	478.670.126	0,90
2006	55.954.901.435	54.194.452.605	1.760.448.830	3,15
2007	58.601.664.250	58.066.915.603	534.748.647	0,91
2008	61.728.534.407	62.053.917.314	-325.382.907	-0,53
2009	62.178.454.387	64.580.998.157	-2.402.543.770	-3,86
2010	69.453.752.752	75.623.418.629	-6.169.665.877	-8,88
2011	71.849.268.409	70.910.911.743	938.356.666	1,31
2012	76.158.966.838	75.981.767.608	177.199.230	0,23
2013	81.183.680.932	80.468.520.666	715.160.266	0,88
2014	83.309.239.574	81.914.267.875	1.394.971.699	1,67
2015	83.082.335.621	83.162.769.722	-80.434.101	-0,10
2016	84.768.405.997	84.711.088.389	57.317.608	0,07

O PIB do MS no ano de 2016 foi de R\$84.768.405.997,00. Portanto, verifica-se a proximidade do valor predito com o valor real, apresentando uma diferença de R\$57.317.608,00 em termos monetários e 0,07% em termos percentuais, ou seja, o modelo de RNA construído, para o ano de 2016, demonstrou uma capacidade preditiva elevada, pois o valor estimado chegou a explicar 99,3%. Vale aqui mencionar que, após o treinamento e validação da RNA, qualquer valor inserido de variáveis independentes poderá ser estimado a variável dependente com o modelo construído, treinado e validado, como foi realizado para o ano de 2016.

CONCLUSÕES

Inicialmente, foi aplicado a correlação de Pearson de medir como as variáveis se relacionavam entre si, donde se concluiu que todas as variáveis possuíam uma forte correlação positiva, o que já era esperado, pois, elas têm relacionamentos produtivos associados. Foi observado que este tipo de associação poderia ser um problema se a técnica de predição não fossem as RNAs, pois, poderia existir problemas de especificação de modelo, autocorrelação, entre outros.

Em seguida, esta pesquisa buscou construir e validar o modelo de RNA predito, então com a entrada do PIB de MS (PIBMS) como variável dependente e a produção de cana-de-açúcar (Prod Cana), Produção de açúcar (Prod Açúcar) e Produção de etanol (Prod Etanol) como covariáveis (variáveis independentes) no software SPSS. Esse objetivo foi alcançado de forma satisfatória, pois, os testes realizados apresentaram valores e observações que creditam a RNA capacidade de predição satisfatória do PIB de MS.

Utilizando dados do ano de 2016 para as variáveis independentes, estimou-se um valor para o PIB de MS. O valor encontrado foi de R\$84.711.088.389. Este valor foi inferior 0,07% do PIB de MS real no ano de 2016, ou seja, um valor insignificante para este tipo de estimação. Desta forma, conclui-se que a RNA

construída possui alta capacidade preditiva após ser estimada, treinada, validada e testada.

A pesquisa demonstra sua relevância a partir do momento que discute uma importante atividade econômica para o estado de Mato Grosso do Sul, sobretudo por utilizar uma técnica de inteligência artificial para prever o PIB desta unidade da federação. Sendo assim, este modelo treinado e validado pode-se utilizar pelos órgãos públicos e também responsáveis pelo setor para tomar decisões de produção, tributárias, comercializam, entre outros. Por fim, sugere-se que novos estudos sejam realizados tais como a inserção de novas variáveis no modelo (população, área plantada, entre outras), ou a utilização dos dados coletados para este artigo em outras técnicas de predição com o intuito de comparar a capacidade preditiva.

REFERÊNCIAS

ÁVILA, S. R. A.. **Efeitos socioeconômicos da expansão da cana de açúcar no vale do São Patrício**. Dissertação (Mestrado) – Universidade de Brasília, Brasília, 2009.

CALÔBA, G. M.; CALÔBA, L. P.; SALIBY, E.. Cooperação entre redes neurais artificiais e técnicas clássicas para previsão de demanda de uma série de vendas de cerveja na Austrália. **Pesquisa Operacional**, v.22, n.3, p.345-358, 2002.

CAMPEÃO, P.; SPROESSER, R. L.; PAULA, N. Q.. Influência do Setor Sucroalcooleiro no Índice de Desenvolvimento Humano Municipal (IDH-M). In: CONGRESSO, SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL, 47. **Anais**. Porto Alegre: SOBER, 2009. p.1-15.

COUTINHO, H. L. C.; TURETTA, A. P. D.; MONTEIRO, J. M. G.; CASTRO, S. S.; PIETRAFESA, J. P.. Participatory Sustainability Assessment for Sugarcane Expansion in Goiás, Brazil. **Sustainability**, v.9, n.9, p.1573, 2017. DOI: <http://doi.org/10.3390/su9091573>

DEFANTE, L. R.; VILPOUX, O.; SAUER, L.. Evolução da produção de cana-de-açúcar no estado do Mato Grosso do Sul. **IGepec**, v.22, n.1, p.150-169, 2018.

FERNANDES, L. G. L.; NAVAUX, P. O. A.; PORTUGAL, M. S.. Previsão de séries de tempo: redes neurais e modelos estruturais. **Pesquisa e Planejamento Econômico**, v.26, n.2, p.253-276, 1996.

HAIR JR., J. F.; BABIN, B. J.; BLACK, W. C.. **Análise Multivariada de Dados**. 5 ed. Porto Alegre: Bookman, 2005.

HAYKIN, S.. **Neural networks, a comprehensive foundation**. New Jersey: Englewood Cliffs, 1999.

MANZATTO, C. V.; JESUS, E. D. A.; BACCA, F. M.; ZARONI, M. J.; PEREIRA, S. E. M.. **Zoneamento agroecológico da cana-**

de-açúcar. Rio de Janeiro: Embrapa Solos, 2009.

RUDORFF, B. F. T.; AGUIAR, D. A.; SILVA, W. F.; SUGAWARA, L. M.; ADAMI, M.; MOREIRA, M. A.. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo state (Brazil) using Landsat Data. **Remote Sensing**, v.2, n.4, p.1057-1076, 2010. DOI: <http://doi.org/10.3390/rs2041057>

SANTOS, G. R.. **Políticas públicas e expansão da agroenergia no Brasil: contradições e desafios à sustentabilidade no ambiente rural em regiões do Cerrado**. Tese (Doutorado) – Universidade de Brasília, Brasília, 2011.

SANT'ANNA, A. C.; GRANCO, G.; BERGTOLD, J. S.; CALDAS, M. M.; XIA, T.; MASI, P.; LINK, T.; LORENZANI, W.. Os desafios da expansão da cana-de-açúcar: a percepção de produtores e arrendatários de terras em Goiás e Mato Grosso do Sul. In: SANTOS, G. R.. **Quarenta anos de etanol em larga escala no Brasil: desafios, crises e perspectivas**. Brasília: IPEA, 2016. p.315.

SHIKIDA, P. F. A.. Expansão canavieira no Centro-Oeste: limites e potencialidades. **Revista de Política Agrícola**, v.22, n.2, p.122-137, 2013.

SOUZA, E. L.; MACEDO, I. C.. **Etanol e bioeletricidade: a cana-de-açúcar no futuro da matriz energética**. São Paulo: Única, 2010.

TORRES, R.; MACHADO, M. A. S.; SOUZA, R. C.. Previsão de séries temporais de falhas em manutenção industrial usando redes neurais. **Engevista**, v.7, n.2, p.23, 2005.

TSUKADA, C. Y. P. S.. **Os efeitos da expansão do setor sucroalcooleiro sobre o desenvolvimento de municípios da região administrativa de Marília**. Dissertação (Mestrado em Geografia) - Universidade Estadual Paulista, São Paulo, 2011.

A CBPC – Companhia Brasileira de Produção Científica (CNPJ: 11.221.422/0001-03) detém os direitos materiais desta publicação. Os direitos referem-se à publicação do trabalho em qualquer parte do mundo, incluindo os direitos às renovações, expansões e disseminações da contribuição, bem como outros direitos subsidiários. Todos os trabalhos publicados eletronicamente poderão posteriormente ser publicados em coletâneas impressas sob coordenação da **Sustenere Publishing**, da Companhia Brasileira de Produção Científica e seus parceiros autorizados. Os (as) autores (as) preservam os direitos autorais, mas não têm permissão para a publicação da contribuição em outro meio, impresso ou digital, em português ou em tradução.