



ISSN:1984-2295

Revista Brasileira de Geografia Física

Homepage: <https://periodicos.ufpe.br/revistas/rbgfe>



Aplicação de inteligência artificial na previsão de produções de açúcar e etanol, no período de 2020 até 2025, nas macrorregiões Centro-Sul e Norte-Nordeste do Brasil

Celso Correia de Souza¹, Joao Bosco Arbués Carneiro Junior², Márcia Ferreira Cristaldo³, Raul Asseff Castelhão⁴, Daniel Massen Frainer⁵

¹Pesquisador Produtividade do CNPq, Campo Grande, MS, csouza939@gmail.com

²Universidade Federal de Rondonópolis (UFR), Rondonópolis, MT, jbacj@hotmail.com

³Instituto Federal de Mato Grosso do Sul (IFMS), Aquidauana, MS, marcia.cristaldo@ifms.edu.br

⁴Universidade Anhanguera Uniderp, Campo Grande, MS, raulassefcastelao@gmail.com

⁵Universidade Estadual do Mato Grosso do Sul (UEMS), Campo Grande, MS, danielfrainer@gamil.com

Artigo recebido em 15/02/2022 e aceito em 13/06/2022

RESUMO

A produção de cana-de-açúcar tem um papel muito importante na economia brasileira, desde o período colonial, com a produção de açúcar e etanol, produtos amplamente comercializados pelo Brasil. O objetivo deste trabalho foi usar recursos de redes neurais artificiais (RNA), implementado no software Weka, Version 3.6.4, para fazer a previsão da produção de açúcar e etanol nas macrorregiões Centro-Sul e Norte-Nordeste brasileiras, no período de 2020 à 2025. Utilizou-se para isso séries históricas das produções desses dois produtos no período de 1980 a 2019, com quarenta observações. Os resultados podem ser considerados satisfatórios, com erros médios absolutos (MAE) relativamente baixos para as duas previsões feitas, do açúcar e do etanol anidro. Com os dados das produções de etanol anidro e açúcar já consolidados para o ano de 2020, foi possível determinar as diferenças entre as produções reais e as previsões realizadas, com diferença de 27,2% para o etanol anidro e de 0,95% para o açúcar. O motivo dessa diferença maior para o etanol anidro depende de novos estudos.

Palavras-chaves: Desenvolvimento regional, previsão de açúcar, previsão de etanol anidro, produção de cana-de-açúcar.

Application of Artificial Intelligence in the Forecast of the Productions of Sugar and Anhydro Ethanol in the Period from 2020 to 2025, in the Centro-South and North-Northeast Macroregions of Brazil

ABSTRACT

The production of sugarcane has played a very important role in the Brazilian economy, since the colonial period, with the production of sugar and ethanol, products widely traded in Brazil. The objective of this work was to use artificial neural networks (ANN) resources, implemented in the Weka software, Version 3.6.4, to forecast sugar and ethanol production in the Center-South and North-Northeast macro-regions of Brazil, in the period 2020. to 2025. For this purpose, historical series of the productions of these two products were used in the period from 1980 to 2019, with forty observations. The results can be considered satisfactory, with relatively low mean absolute errors (MAE) for the two predictions made, sugar and anhydrous ethane. With the data on the production of anhydrous ethanol and sugar already consolidated for the year 2020, it was possible to determine the differences between the actual productions and the forecasts made, with a difference of 27.2% for anhydrous ethanol and 0.95% for the sugar. The reason for this greater difference for anhydrous ethanol depends on further studies.

Keywords: Regional development, anhydrous ethanol forecast, sugar forecast, sugarcane production.

Introdução

A cultura da cana-de-açúcar teve origem na Polinésia, feita pelos árabes e disseminou no mundo, com essa cultura se adaptando muito bem nas Américas, com clima favorável ao seu desenvolvimento. A cana-de-açúcar é uma planta perene, podendo produzir por um período médio de

quatro a seis anos, e o seu manejo é relativamente fácil. Oficialmente, foi Martim Affonso de Souza, que em 1532 trouxe a primeira muda de cana-de-açúcar ao Brasil e iniciou o seu cultivo na Capitania de São Vicente. Nessa mesma Capitania foi construído o primeiro engenho de açúcar do Brasil (UDOP, 2003).

Atualmente, o cultivo da cana-de-açúcar se concentra nas regiões Centro-Sul e Norte-Nordeste do Brasil, sendo o estado de São Paulo da macrorregião Centro-Sul do Brasil o maior produtor de cana-de-açúcar, e que tem a maior área plantada da cultura, consequentemente, o estado de São Paulo é o maior produtor brasileiro de açúcar e etanol etílico (Ronquim e Fonseca, 2018; Antunes et al., 2017; Wissmann et al., 2014).

O álcool etílico, também chamado de etanol, é encontrado em bebidas como cerveja, vinho, aguardente, na indústria de perfumaria e em produtos de limpeza. Na produção do etanol é possível obter dois compostos, o etanol anidro e o etanol hidratado. Segundo Souza et al. (2022), o etanol anidro, que é misturado à gasolina é caracterizado pelo teor alcoólico mínimo de 99,3%, determinado pelo Instituto Nacional de Pesos e Medidas (INPM). O etanol hidratado, utilizado como combustível, é uma mistura de etanol e água com teor alcoólico mínimo de 92,6° (INPM).

Na década de 1970, em decorrência da crise do petróleo, houve a iniciativa brasileira da produção de etanol hidratado para reduzir a dependência do país do petróleo importado (Carvalho et al., 2013). Em 1975 o governo lança o Programa Nacional do Álcool (PROÁLCOOL), o principal objetivo era aumentar a produção de etanol hidratado e incentivar a produção, no Brasil, de veículos *flex*, veículos movidos a etanol hidratado ou gasolina, ou ainda, uma mistura dos dois (PROÁLCOOL, 2008).

Como o valor do petróleo oscila muito, dependendo da conjuntura mundial de produção e consumo, logo após a criação do PROÁLCOOL o preço do barril de petróleo diminuiu muito, de modo que a importância do programa reduziu significativamente, voltando a ter uma razoável importância apenas em 1979, surgindo daí segunda fase do programa (PROÁLCOOL, 2008; Dias e Sentelhas, 2018).

Apesar dos altos e baixos do programa PROÁLCOOL, o cultivo da cana-de-açúcar vem contribuindo para a melhoria do cenário social e industrial brasileiro, propiciando o desenvolvimento do agronegócio do país e creditando ao etanol um importante espaço no setor automobilístico nacional pelo fato de ser um combustível ambientalmente correto e plenamente sustentável (Araújo e Sobrinho, 2020; Rodrigues, 2010).

A participação do Brasil no mercado internacional de açúcar é elevada, aproximadamente 40% das exportações mundiais são brasileiras, além do que, o mercado interno brasileiro, consumidor de açúcar, também é muito relevante. Já, o etanol produzido no país tem o seu

consumo quase que exclusivamente no mercado interno brasileiro. O país é o segundo maior produtor mundial de etanol, em que aparece em primeiro lugar os EUA, que utiliza o milho como matéria prima na sua produção, o que se torna um gargalo na produção do biocombustível, pois compete com a produção de alimentos (FAO, 2019; Bader Esam, 2017).

Segundo Viegas (2010), enquanto uma tonelada de cana produz apenas 89,5 litros de etanol, uma tonelada de milho pode produzir até 407 litros de etanol. Por outro lado, no Brasil, a produtividade média de milho por hectare é de aproximadamente 5 toneladas, enquanto a produtividade média de cana é de 70 toneladas. Daí, em um hectare pode-se produzir em média, aproximadamente, 2035 litros de etanol de milho ou 6265 litros de etanol de cana. Além do mais, em termos financeiros, quando se compara o preço de produção do etanol brasileiro, que é de R\$1,16/L, o etanol americano, que é produzido a partir do milho, tem o seu custo a R\$1,85/L, portanto, altamente competitivo o custo do etanol brasileiro.

Santos et al. (2019), em seu artigo “Produção de álcool anidro mais limpa” aborda os efeitos causados pela emissão de alguns gases poluentes oriundos dos combustíveis fósseis causadores do efeito estufa, responsáveis diretos pelo aquecimento global. Daí, a procura expressiva para a substituição de tais combustíveis por fontes de energia mais limpa que possuam sustentabilidade.

De acordo com Toledo (2012), o Brasil se destaca pela produção do álcool combustível, especificamente o álcool anidro, que é conhecido como um combustível renovável e tem sua produção realizada para substituir o chumbo que é um metal venenoso adicionado à gasolina. A forma de produção do etanol anidro é aparentemente mais cara, porém, rentável a longo prazo garantindo um *playback* bastante rápido. Daí conclui-se que esse tipo de produção além de ser mais limpa enquadra-se em melhoria econômica e ambiental.

Segundo Pacheco (2011), a produção de etanol não impacta seriamente o meio ambiente, aliás, constitui uma alternativa para diminuir problemas ambientais e energéticos no Brasil e no mundo, em vista da escassez e da alta dos preços dos combustíveis fósseis, além da poluição por eles causada. A posição do Brasil no mundo, quando se refere à produção de etanol, é bastante vantajosa, pelo uso de tecnologias avançadas no cultivo da cana-de-açúcar e na sua produção. O aumento do cultivo da área plantada de cana-de-açúcar não exige nenhum desmatamento, como também, não concorre com áreas destinadas à produção de alimentos. O autor também enfatiza a

sustentabilidade da matriz energética brasileira que possui 46,8% das fontes de energias renováveis, enquanto a média mundial é de apenas 14%.

O Brasil tem uma importante indústria automobilística, com destaque de produção de carros flex, que utilizam tanto combustíveis fósseis quanto etanol, colaborando sobremaneira para uma menor poluição atmosférica, reduzindo a emissão de CO₂. Em termos financeiros, quando se compara o preço de produção do etanol brasileiro, que é de US\$0,22/L, o etanol americano, que é produzido a partir do milho, tem o seu custo a US\$0,35/L, portanto, altamente competitivo o custo do etanol brasileiro (Viegas, 2010).

Estudos e pesquisas desenvolvidas por NovaCana sobre o etanol propiciaram novas tecnologias de cultivo e aumento de sua produção que, aliadas à sustentabilidade e à viabilidade econômica propiciou a produção do etanol de várias gerações. A produção de etanol de primeira geração envolve a aplicação de melhores práticas agrônomicas de produção, com o desenvolvimento de melhores cultivares, com variedades de cana-de-açúcar mais produtivas e resistentes. Também, envolve inovações na industrialização da cana-de-açúcar, com o desenvolvimento de tecnologia de fermentação a baixa pressão e extração a vácuo do etanol, que mantém a produtividade, diminui os custos de destilação e da quantidade de vinhaça produzida e melhora o nível de reutilização de água, que na década de 1990 era de 62,7%, elevando-se para 87,8% em 2005 (NovaCana, 2008).

Segundo Lorenzoni (2019), é preciso ficar atento à evolução da comercialização do etanol para um mercado consumidor cuja tendência é crescer, tanto nacional quanto internacionalmente devido às legislações ambientais que tende a priorizar os biocombustíveis em detrimento dos combustíveis fósseis podendo, com isso, estabelecer uma concorrência entre o uso da terra para a produção de cana-de-açúcar e o seu uso para a produção de alimentos, principalmente, naquelas regiões que já não dispõem de terras férteis para essas produções. Daí, é necessário investir no etanol de segunda geração, também conhecido como etanol celulósico, que é uma nova alternativa para o uso energético da biomassa, que traz vantagens ambientais e econômicas, mas cujas tecnologias ainda estão em fase de aperfeiçoamento.

Segundo Figueiredo e Cesar (2021) e Lorenzi et al. (2019), o Brasil apresenta vantagens competitivas na produção de etanol de segunda geração devido à grande disponibilidade das duas matérias primas que entram na produção desse tipo de etanol devido à melhor eficiências das caldeiras

que resulta em excedentes de bagaço de cana, e a proibição da queima da cana para a colheita, resultando em excedente de palha. Esses dois subprodutos da cana-de-açúcar são matérias primas para a produção de etanol de segunda geração. Essa nova tecnologia de produção de etanol utiliza a mesma planta de produção do etanol comum.

Carvalho et al. (2013) em seu artigo “Cana-de-açúcar e álcool combustível: histórico, sustentabilidade e segurança energética”, afirmam que a cana-de-açúcar tem importância desde os primórdios na história do Brasil, desde o seu descobrimento, e que essa cultura tem papel importante na economia brasileira, sendo um dos principais produtos comercializados. Na década de 1970 com o advento da crise econômica, geopolítica e da possibilidade de esgotamento de petróleo, os países dependentes desse combustível buscaram novas alternativas energéticas. Em 1975 criou-se o Programa Nacional do Alcool (PROALCOOL), que teve alguns anos de ascensão, pelo aumento de destilarias e comercialização de carros movidos com mistura de etanol. Em decorrência da diminuição do preço do barril de petróleo a importância do programa reduziu significativamente, voltando ao auge apenas em 1979, ou seja, a segunda fase do programa. Concebido como um dos vetores da resposta nacional à crise dos preços do petróleo dos anos 1970, o programa persistiu em momentos em ascensão em outros não, alcançando por mais de três décadas.

O Brasil é o segundo maior produtor de etanol, perdendo apenas para o EUA, em que usa como matéria prima o milho, o que torna um gargalo a produção de biodiesel, pois compete com a produção de alimentos. Novas tecnologias desenvolvidas para aumentar a produção de etanol, aliadas à sustentabilidade e viabilidade econômica vêm sendo realizadas, a principal é o etanol de segunda geração, conhecido como etanol celulósico, além do etanol de terceira e quarta geração (Leite e Cortez, 2020; Pacheco, 2011).

Oliveira et al. (2010), no trabalho “Aplicação de redes neurais artificiais na previsão da produção de álcool” descreve a aplicação de Redes Neurais Artificiais na tarefa de previsão da produção de álcool no Brasil no ano de 2006, a partir de dados de produção anteriores. É também apresentada uma comparação entre os resultados obtidos por meio da Rede Neural com os obtidos utilizando técnicas de séries temporais, sendo que a Rede Neural obteve melhores resultados. A RNA implementada obteve uma previsão melhor para a produção de álcool do que aquela obtida por técnica de tratamento da série no domínio do tempo (séries temporais). Assim, foi possível verificar a

eficiência da RNA no tratamento de padrões não lineares e a sua aplicação na tarefa de previsão. Para a obtenção de melhores resultados, modificações na topologia e treinamento da RNA podem ser aplicadas tais como: variação da quantidade de neurônios, aumento do tempo de treinamento, teste com outros parâmetros de treinamento.

Penedo et al. (2007) no artigo *Previsão de Preços do Açúcar Utilizando Redes Neurais Artificiais*, enfatiza que para os países latino-americanos, a produção e exportação de produtos agrícolas são de grande importância no equilíbrio de suas contas comerciais. Portanto, conhecer o comportamento comercial desses produtos no mundo é vantajoso para o desenvolvimento de projetos de comercialização e tomadas de decisão. Como a participação do Brasil no mercado internacional de açúcar é elevada, aproximadamente 40% das exportações mundiais são brasileiras, que também possui um forte mercado interno, é esperada a relação do mercado doméstico brasileiro do açúcar com os preços internacionais. Desse modo, em seu trabalho foi analisado o impacto dos preços do contrato futuro de açúcar bruto nº11 da Csce/Nybot, taxa de câmbio comercial, barril de petróleo WTI Spot (mercado norte-americano) e Europe Brent Spot (mercado europeu) nos preços do açúcar cristal acondicionado em sacas de 50 kg. Nessa análise foi utilizado o conceito de Redes Neurais Artificiais (RNA's) para explorar a influência dos parâmetros: número de camadas e números de neurônios intermediários sobre o desempenho das RNA's. A capacidade de estabelecer relações entre diversas variáveis, é uma característica de grande destaque das RNAs que, uma vez construídas, não dependem mais da base de dados de origem, pois o conhecimento está nas conexões entre os neurônios. Conclui-se que a rede utilizada para resolver o problema de previsão de preços do açúcar é uma ferramenta útil para uma tomada de decisão.

Carneiro Junior et al. (2019), em seu artigo "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" afirmam que 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 predição 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 à 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 predição 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. A série temporal em análise compreendeu o intervalo de 2002 a 2015. 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.

Conhecer o comportamento comercial dos derivados da cana-de-açúcar no mercado internacional é vantajoso para o desenvolvimento de projetos de produção e comercialização e para as tomadas de decisão. Como o Brasil tem uma participação importante no mercado internacional de açúcar, um forte mercado interno desse produto, além de produzir um elevado volume de etanol (anidro e hidratado), é esperada a previsão das produções dessas *commodities* tanto para a exportação quanto para suprir o mercado interno.

Um modelo matemático baseado em conceitos de redes neurais artificiais (RNA), implementada no software Weka, Version 3.6.4, aparece com destaque para solucionar o problema proposto, pois, relaciona várias variáveis sem imposição de condições e, uma vez construído o modelo matemático, não depende mais da base de dados de origem, pois, o relacionamento entre as variáveis permanece nas conexões entre os neurônios (Spancersk e Santos, 2021).

Diante do exposto pode-se levantar a seguinte hipótese de pesquisa deste estudo: a previsão de produção de cana-de-açúcar e dos seus derivados (açúcar e etanol), ao longo do tempo, pode ser realizada de forma eficiente com a

utilização de RNA implementada no software Weka, Version 3.6.4.

Sendo assim, este artigo teve como objetivo aplicar conceitos de Redes Neurais Artificiais (RNA), implementada no software Weka, Version 3.6.4, nas previsões das produções de cana-de-açúcar e dos seus derivados (açúcar e etanol), no período de 2020 a 2025, a partir de dados de produções anteriores.

Material e métodos

O objeto deste trabalho foi o setor da agricultura brasileira ligada ao cultivo da cana-de-açúcar, matéria prima na produção de açúcar e etanol anidro e hidratado, que deu origem ao objetivo central do trabalho, sobre a previsão da produção de etanol anidro no Brasil, no período de 2020 a 2025, nas duas principais macrorregiões produtoras, a Centro-Sul e a Norte-Nordeste, com base em dados da série histórica das suas produções, no período de 1980 a 2019.

Em relação aos procedimentos metodológicos, este trabalho classifica-se como pesquisa documental, com a utilização de dados secundários, extraídos de relatórios gerenciais históricos de órgãos públicos e autarquias brasileiras. Este trabalho também pode ser caracterizado como uma pesquisa exploratória-explicativa, que trata de explorar e aprofundar conhecimentos sobre um tema muito relevante ao Brasil, e explicativa, pois ambas possibilitam comprovar, aprofundar, explicar e identificar fatores que explicam a ocorrência de fatores que expliquem a relevância e importância do tema proposto.

Inicialmente, foi realizada uma pesquisa bibliográfica para descrever, com ênfase e precisão, as informações sobre a cana-de-açúcar, a produção de etanol, tanto anidro quanto hidratado.

O Banco de Dados (BD) utilizado foi o do histórico da produção anual de açúcar, etanol anidro e etanol hidratado no Brasil, de janeiro de 1980 até 2019, no qual se encontrava volume de

cana-de-açúcar moída, o volume de açúcar produzido, em (mil t), volume do etanol anidro, em (m^3) e volume de etanol hidratado, em (m^3). O banco de dados estava dividido em duas macrorregiões principais produtoras de cana-de-açúcar, que são a macrorregião Centro-Sul do Brasil e a macrorregião Norte e Nordeste.

Para criar o banco de dados na etapa de mineração, transformou esses dados no formato aceito pelo software WEKA, desenvolvido na universidade de Waikato, Nova Zelândia, no qual utiliza-se o formato de Arquivo de Atributo-Relação (ARFF) para carregar os dados, onde é possível definir o tipo de dados que estão sendo carregados e, então, fornecer seus próprios dados. No arquivo, definiu-se cada coluna e o que cada coluna contém (Amit e Malika, 2014; Witten et al., 2011; Arun et al., 2007).

Com o propósito de encontrar o conjunto de atributos que melhor se adequa ao algoritmo de aprendizado, aplicou-se o algoritmo *Selection*, baseado no método de correlação (*Correlation-based Feature Selection - CFS*), tendo como objetivo a seleção de subconjuntos de recursos, identificando e removendo a quantidade de informações irrelevantes e redundantes possíveis, pois, o algoritmo CFS classifica os subconjuntos gerados no processamento de acordo com uma função de correlação com base em uma recompensa heurística de avaliação (Eibe et al., 2004; Hall, 1999).

A configuração do ambiente foi modificada para contemplar outra forma de dividir a base de dados, *Train/Test Percentage Split (data randomized)*, em que essa opção utiliza um processo randômico para selecionar os exemplos dos conjuntos de treinamento e teste. O padrão da ferramenta é utilizar 70% dos registros para o conjunto de treinamento e 30% para teste (Bouckaert et al., 2010). O Quadro 1 mostra as configurações dos algoritmos CFS e SVM, com relação ao etanol anidro.

Quadro 1. Configurações dos algoritmos CFS e SVM.

Algoritmo	Configuração etanol anidro
CFS	Evaluator: weka.attributeSelection.CfsSubsetEval -P 1 -E 1 Search: weka.attributeSelection.BestFirst -D 1 -N 5
SVM	Weka.classifiers.functions.SupportVector.Puk -S 2.977280920220959 - O 0.7991929798422421

O algoritmo CFS, Quadro 1, avalia a importância de um subconjunto de atributos em função da previsão individual de cada atributo e o grau de correlação entre eles, selecionando um subconjunto de recursos de forma explícita. O algoritmo CFS avalia cada recurso individualmente e atribui uma pontuação com base em sua importância (Koprinska et al., 2015; Zdravko e Ingrid, 2006).

Com o objetivo de realizar a previsão de produção de açúcar e etanol foi utilizado o

As medidas de avaliações utilizadas para avaliar o desempenho do modelo utilizado são as seguintes:

- 1 Correlation Coefficient - CC (Coeficiente de Correlação): indica a força e a direção do relacionamento linear entre as duas ou várias variáveis a serem testadas. O seu valor, denotado por r , varia de -1 a 1 e, quanto mais próximo de 1 ou -1, maior o grau de associação entre as variáveis. Esse coeficiente indica se as amostras estão corretamente classificadas;
- 2 Mean absolute error – MAE (Erro Médio Absoluto): Ele expressa o erro, em uma maneira de fácil compreensão. Medida que indica a média do afastamento de todos os valores fornecidos pelos classificadores e o seu real valor;
- 3 Root mean squared error - RMSE (Erro Médio da Raiz Quadrada): Medida calculada pela média da raiz quadrada da diferença entre o valor calculado e o valor correto. É a raiz quadrada do Erro Médio Absoluto. Ela estima a qualidade de um classificador
- 4 Relative absolute error – RAE (Erro Relativo Absoluto): É a medida em porcentagem que corresponde ao erro total absoluto. Assim como o coeficiente de correlação, essa medida implica diretamente no resultado da previsão. Quanto mais alto o coeficiente de correlação, mais baixa será a média de erro absoluto, e vice versa.
- 5 Root relative squared error – RRSE (Erro Relativo da Raiz Quadrada): Medida que reduz o quadrado do erro relativo na mesma dimensão na quantidade sendo predita, incluindo raiz quadrada. Assim como a raiz quadrada do erro significativo (RMSE), esse exagera nos casos em que o erro da previsão foi significativamente maior do que o erro significativo.

algoritmo support vector machine (SVM), Quadro 1, que é um algoritmo de aprendizado de máquina supervisionado que pode ser usado para desafios de classificação ou regressão. Seu foco maior é no treinamento e classificação de um dataset. Nesse algoritmo foi plotado cada item de dados como um ponto no espaço n -dimensional, com o valor de cada recurso sendo o valor de uma determinada coordenada (Ignacio et al., 2017).

Resultados e discussão

Visando realizar a previsão dos volumes de produções de açúcar, etanol anidro e etanol hidratado, no período de 2020 a 2025, foi utilizado o aplicativo computacional Weka Versão 3-6-4 (Bouckaert et al, 2010), como emulador de Redes Neurais Artificiais. Segundo Carneiro Junior et al. (2019), Carvalho et al. (2013) e Penedo et al. (2007), as RNA são capazes de estabelecer relações entre diversas variáveis, sem imposição de domínio e restrições em seus campos de definição.

Lembrando que, segundo Rodrigues (2010), a cana-de-açúcar é a matéria prima comum na produção de açúcar, etanol anidro e etanol hidratado, desse modo, inicialmente, foram coletados os dados desses produtos via internet, com periodicidade anual, no período de 1980 a 2019, totalizando 40 observações.

O Quadro 2 estão apresentadas as coletas realizadas das produções de cana-de-açúcar, açúcar, etanol anidro e etanol hidratado da macrorregião Centro-Sul do Brasil, no período de 1980 a 2019.

Dos quatro atributos iniciais o CFS selecionou três atributos para previsão de produção de etanol anidro, sendo: ano, produção de cana-de-açúcar e de açúcar. Já, para previsão de açúcar foram selecionados os atributos: ano, produção de cana-de-açúcar e de etanol anidro. Segundo Santos et al. (2019) e Pacheco (2011), a produção de etanol anidro é mais vantajosa, pois, é mais lucrativa, além de impactar menos o meio ambiente, razão pela qual, neste estudo, só foram feitas as previsões de açúcar e de etanol anidro.

Na Tabela 1 pode-se verificar os resultados do algoritmo SVM mostrando suas principais métricas, um dos valores mais importantes produzidos pela Weka, no qual mostra uma ótima classificação com valores próximos de 1

Quadro 2. Produções de cana-de-açúcar, açúcar, etanol anidro e etanol hidratado da macrorregião Centro-Sul do Brasil, no período de 1980 a 2019.

Ano	Cana-de-Açúcar (mil t)	Açúcar (mil t)	Etanol Anidro (mil m³)	Etanol Hidratado (mil m³)
1980	83432	5253	1823	1233
1981	58387	5146	1014	2400
1982	65320	5613	3003	1632
1983	66664	5512	2124	4610
1984	145780	5282	1827	5761
1985	163374	4834	2813	6986
1986	156608	4809	1759	6542
1987	169795	5027	1788	7880
1988	166487	5253	1513	8451
1989	162436	4140	1303	8598
1990	170195	4509	1088	8620
1991	179031	5761	1802	9125
1992	176295	6118	1938	8120
1993	173264	7304	2379	8003
1994	196037	9407	2577	8555
1995	204383	10185	2588	8256
1996	231558	11618	3821	8257
1997	248730	11354	4762	8489
1998	269749	15160	4804	7412
1999	263894	16900	5370	6234
2000	207099	12643	4802	4262
2001	244218	15972	5746	4430
2002	270407	18778	6270	4882
2003	298772	20426	8029	4984
2004	327806	22149	7298	6265
2005	336783	22015	7275	7036
2006	372754	25796	7066	9000
2007	431114	26201	7143	13191
2008	505116	26750	8247	16868
2009	541962	28645	6206	17479
2010	556945	33501	7413	17971
2011	493159	31304	7454	13089
2012	532758	34097	8730	12632
2013	597061	34295	11008	14568
2014	573145	32011	10740	15492
2015	617709	31221	10643	17581
2016	607137	35628	10655	14996
2017	596330	36060	10419	15670
2018	573169	26510	9141	21812
2019	590361	26761	9946	23313

Fonte: Adaptado de UNICA (2021).

O Quadro 3 apresenta as coletas realizadas das produções de cana-de-açúcar, açúcar, etanol

anidro e etanol hidratado da macrorregião Norte-Nordeste do Brasil, no período de 1980 a 2019.

Quadro 3. Produções de cana-de-açúcar, açúcar, etanol anidro e etanol hidratado da macrorregião Norte-Nordeste do Brasil, no período de 1980 a 2019.

Ano	Cana-de-Açúcar (mil t)	Açúcar (mil t)	Etanol Anidro (mil m³)	Etanol Hidratado (mil m³)
1980	40248	3001	283	368
1981	32694	2789	439	387
1982	36661	3244	546	642
1983	38485	3574	345	784
1984	57088	3536	275	1328
1985	59804	3199	383	1647
1986	71238	3348	404	1800
1987	54703	3158	195	1596
1988	53617	2817	204	1477
1989	60466	3074	150	1870
1990	52235	2857	199	1608
1991	50191	2770	185	1611
1992	47164	3146	279	1393
1993	33271	1858	144	767
1994	44811	3245	298	1321
1995	44547	3328	421	1346
1996	57236	3211	795	1522
1997	53883	3533	921	1243
1998	46730	2863	863	797
1999	43119	2487	742	638
2000	49718	3554	818	709
2001	48824	3246	719	641
2002	50243	3789	746	725
2003	59990	4493	847	877
2004	57393	4536	978	848
2005	48345	3808	791	718
2006	54904	4192	965	814
2007	64610	4826	1051	1143
2008	64100	4299	1089	1322
2009	60231	4312	859	1146
2010	63464	4505	910	1082
2011	66056	4621	1127	1012
2012	55720	4149	1115	750
2013	54233	3299	1179	721
2014	60782	3560	1339	909
2015	49115	2616	1017	991
2016	44704	3107	934	669
2017	44872	2550	947	828
2018	47707	2530	779	1382
2019	52316	2844	939	1398

Fonte: Adaptado de UNICA (2021).

Tabela 1. Resultados do algoritmo SVM para as previsões de produções de etanol anidro e de açúcar, nas regiões Centro-Sul e Norte-Nordeste, no período de 2020 a 2025.

Métricas*	Macrorregião Centro-Sul		Macrorregião Norte-Nordeste	
	Acurácia prod. etanol anidro	Acurácia prod. açúcar	Acurácia prod. etanol anidro	Acurácia prod. açúcar
CC	97%	96%	95%	91%
RAE	32,74%	17,88%	20,08%	35,27%
RRSE	25,57%	17,14%	28,19%	40,25%
MAE	0,111	0,148	0,69	0,192
RMSE	0,134	0,185	0,110	0,273

*CC: Coeficiente de correlação; RAE: Erro absoluto relativo; RRSE: Erro relativo da raiz quadrada; MAE: Erro médio absoluto e; RMSE: Erro médio da raiz quadrada.

Fonte: Autores (2021).

As regiões Centro-Sul e Norte-Nordeste apresentaram coeficientes de correlação (CC) acima de 90% tendo, com isso, amostras corretamente classificadas. O MAE indicou a média dos afastamentos de todos os valores fornecidos pelos classificadores e o seu valor real. O RMSE se mostrou bem baixo, indicando a qualidade do classificador.

Observando os resultados verifica-se uma alta correlação entre os atributos e uma baixa média

de erro absoluto, indicando uma ótima previsão nos resultados. Para a obtenção desse resultado utilizou-se todo o banco de dados disponível para avaliar sua correlação entre as variáveis e validação para previsão futura das duas macrorregiões, no período de 2020 a 2025. Após verificar a correlação realizou-se a previsão futura das produções de açúcar e de etanol anidro utilizando um modelo de RNA support vector machine. Os resultados estão na Tabela 2.

Tabela 2. Parâmetros da configuração da RNA para a previsão futura das produções de açúcar e de etanol anidro, no período de 2020 a 2025.

Camada de entrada	Configuração do algoritmo SVM	Saída
Produção de açúcar Produção de etanol anidro	Classes = Ausentes Atributos = 2 Interfaces = 4 Instâncias = 1	Previsão de produção de açúcar Previsão de produção de etanol anidro

Fonte: Autores (2021).

Da Tabela 2 observa-se que foi usado o algoritmo SVM, com as suas configurações características, para realizar as previsões previstas. Na Tabela 3 verifica-se uma relação de previsões,

de 2020 até 2025, para a produção de açúcar em relação à previsão de produção de etanol anidro por macrorregião administrativa brasileira.

Tabela 3. Previsão de produção de açúcar em relação à previsão de produção de etanol anidro, do ano de 2020 até 2025, por macrorregião administrativa brasileira.

Ano	Macrorregião Centro-Sul		Macrorregião Norte-Nordeste	
	Produção de açúcar (mil t)	Produção de etanol anidro (mil m ³)	Produção de açúcar (mil t)	Produção de etanol anidro (mil m ³)
2020	27014	12201	3080	1048
2021	27500	12365	3065	1068
2022	27971	12523	3051	1088
2023	28429	12676	3036	1107
2024	28872	12824	3021	1127
2025	29301	12967	3007	1146

Fonte: Autores (2021).

Da Tabela 3 é possível notar que as previsões seguem uma ordem crescente, a menos da produção de açúcar em que aconteceu um pequeno recuo na produção. Nessas previsões expostas na Tabela 3 podem existir erros de previsão percentuais relativamente altos, como

indicados na Tabela 1. Quando da elaboração desta pesquisa, já existiam dados consolidados da produção de açúcar e de etanol anidro, na safra de 2020, primeiro ano após o final da série histórica utilizada neste trabalho, que foi de 1980 a 2019, e que se encontram na Tabela 4.

Tabela 4. Valores das produções de açúcar e etanol anidro da safra 2020, por macrorregião produtora brasileira.

Ano	Macrorregião Centro-Sul		Macrorregião Norte-Nordeste	
	Produção de açúcar (mil t)	Produção de etanol anidro (mil m ³)	Produção de açúcar (mil t)	Produção de etanol anidro (mil m ³)
2020	26761	9946	2845	939

Fonte: Unica (2021).

Comparando os valores das produções de açúcar e etanol anidro das regiões Centro-Sul e Norte-Nordeste, respectivamente, relativas às Tabelas 3 e 4, no ano de 2020, em que na Tabela 3 são as previsões de produções encontradas pelo software e, na Tabela 4, as produções reais, percebe-se, para a macrorregião Centro-Sul, uma diferença de previsão de produção para a produção real do açúcar de 0,95%, e para produção de etanol anidro a diferença foi de 22,7%.

Já, para a macrorregião Norte-Nordeste as diferenças foram de 8,3% e 11,6%, respectivamente, para as diferenças entre as previsões de produções e as produções reais de açúcar e etanol anidro. Estas diferenças entre previsões e produções reais estão abaixo dos erros encontrados, na Tabela 1.

Observe, ainda, das Tabelas 3 e 4, que as diferenças para o etanol anidro são maiores, indicando que podem haver outros fatores influenciando as tomadas de decisão, pois, como afirma NovaCana (2008), a evolução na produção de etanol é evidente, com o surgimento de novas tecnologias de produção, tanto para o etanol de primeira geração, com a otimização de produção, quanto para o etanol de segunda e terceira geração, introduzidas recentemente nessa produção. Não se pode ignorar que as previsões de produções dessas commodities dependem muito também de decisões governamentais.

Na Figura 1 verifica-se de forma gráfica a previsão de produção de açúcar em relação à previsão de produção de etanol anidro, de 2020 até 2025, para a macrorregião Centro-Sul.

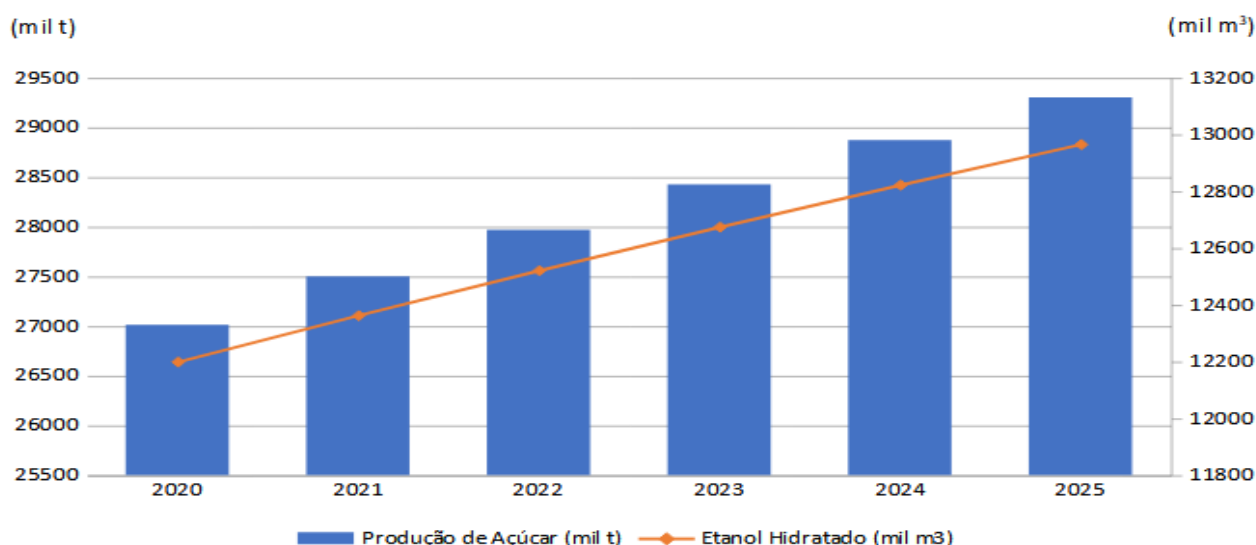


Figura 1. Previsão de produção de açúcar e de etanol anidro, de 2020 até 2025, para a macrorregião Centro-Sul. Fonte: Autores (2021).

Observa-se, da Figura 1, que as duas previsões de produções, de açúcar e de etanol anidro, na macrorregião Centro-Sul, estão altamente correlacionadas de forma positiva, com coeficiente igual a um ($r = +1$). Na Figura 2

verifica-se de forma gráfica a previsão de produção de açúcar em relação à previsão de produção de etanol anidro, de 2020 até 2025, para a macrorregião Norte-Nordeste.

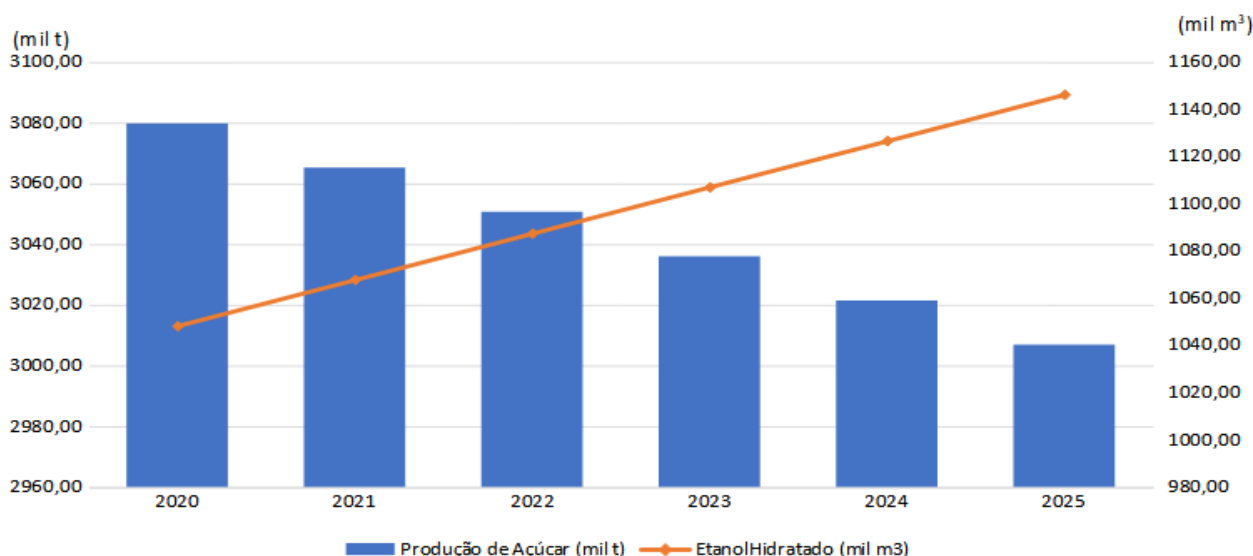


Figura 2. Previsão de Produção de açúcar em relação à previsão da produção de etanol anidro, de 2020 até 2025, para macrorregião Norte-Nordeste.

Fonte: Autores (2021).

Observa-se, da Figura 2, que as duas previsões de produções, de açúcar e de etanol anidro, na macrorregião Norte-Nordeste, estão altamente correlacionadas de forma negativa, com coeficiente igual a menos um ($r = -1$). Por outro

lado, observa-se das Tabelas 2 e 3 e das Figuras 1 e 2 que os valores das previsões de produções de açúcar e etanol anidro para a macrorregião Centro-Sul estão bem acima dos valores da macrorregião Norte-Nordeste.

Conclusão

Neste estudo ficou evidenciado que o Brasil apresenta uma participação elevada no mercado internacional de açúcar, aproximadamente 40% das exportações mundiais são brasileiras, o que, cada vez mais, justifica estudos dessa natureza. A escolha das variáveis produção de açúcar e produção de etanol anidro para a realização das previsões se deve ao fato de que o açúcar tem uma pauta elevada nas exportações brasileiras e o etanol anidro é utilizado como combustível no mercado interno brasileiro, com a sua adição à gasolina.

Os resultados deste estudo podem ser usados pelos órgãos públicos e também responsáveis pelo setor canavieiro para as tomadas de decisão de produção, tributária, comerciais, entre outros. Sugere-se que novos estudos sejam realizados na previsão de produção de etanol hidratado e mesmo as produções de açúcar e etanol anidro com a inserção de novas variáveis no modelo como área plantada, preços de vendas, entre outras, com o intuito de se obter previsões mais ligadas à realidade.

A capacidade de uma RNA de estabelecer relações entre diversas variáveis, é muito importante, pois, uma vez iniciado o processo, com as entradas das variáveis, o resultado final não depende mais da base de dados original, pois, com o aprendizado da rede, o conhecimento permanece nas conexões entre os neurônios.

Utilizando uma série de observações, de 1980 a 2019, relativas as produções de cana-de-açúcar, produção de açúcar, de etanol anidro e hidratado, foi possível fazer previsões das produções de açúcar e etanol anidro com seis anos de antecedência, isto é, de 2020 a 2025, o que mostra a eficácia do software utilizado, o Weka, Version 3.6.4, que dispõe do algoritmo SVM, tem ótimo potencial para realizar previsões. Os erros médios absolutos (MAE) foram relativamente baixos para as duas previsões feitas, do açúcar e do etanol anidro.

Quanto à eficiência do software, notou-se que, utilizando uma série de observações relativamente pequena, as diferenças entre previsões de produções e produções reais, para o

ano de 2020, que já se dispõe dos dados, as diferenças foram pequenas, com 0,95% para o açúcar e 22,7% para o etanol anidro. Diferenças

maiores para o etanol anidro, cujos motivos dependem de novas pesquisas.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de Produtividade em Pesquisa.

Referências

- Amit, K.Y.H.; Malik, S.S.C., 2014. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. *Renewable and Sustainable Energy Reviews* 31, 509-519.
- Antunes, W.R., Schöffel, E.R., Silva, S.D.A., Härter, A., Campos, A.D.S., C.R.S. Montero, C.R.S., 2017. Productive performance of early sugarcane genotypes in Rio Grande do Sul state, *Revista Scientia Agraria* 18(2), 136-142. <https://doi.org/10.5380/rsa.v18i2.50533>.
- Araújo, D. F. C., Sobrinho, F. L. A., 2020. A cultura agrícola da cana-de-açúcar no Brasil: contribuição ao estudo dos territórios rurais e suas contradições e conflitos. *Revista Geopauta* 4(1), 162-183. <<http://periodicos2.uesb.br/index.php/geo>>.
- Arun, K.S., Helga, J., Rafn, B., 2007. Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. *Patient Education and Counseling* 67(1), 21-31.
- Bader Esam, A., 2017. Economic Modelling and Forecasting of Sugar Production and Consumption in Egypt. *International Journal of Agricultural Economics*, 2(4), 96-109.
- Bouckaert, R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., Scuse, D., 2010. WEKA Manual for Version 3-6-4. Disponível em: <http://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/WekaManual%20oficial.pdf>. Acesso em: 22.ago.2021.
- Carneiro Junior., J.B.A., Souza, C.C., 2019. 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 (RICA)* 10(5).
- Carvalho, L.C., Bueno, R.C.O.F., Carvalho, M.M., Favoreto, A.L., Godoy, A.F., 2013. Cana-de-açúcar e álcool combustível: histórico, sustentabilidade e segurança energética. *Enciclopédia Biosfera. Centro Científico Conhecer. Goiânia* 9(16), p. 530.
- Dias, H.B., Sentelhas, P.C., 2018. Sugarcane yield gap analysis in Brazil – A multi-model approach for determining magnitudes and causes. *Science of the total environment* 638, 1127–1136.
- Eibe, F., Mark, H., Len, T., G., Holmes, I. H. W., 2004. Data mining in bioinformatics using Weka. *Bioinformatics* 20(15), 2479–2481. DOI. <<https://doi.org/10.1093/bioinformatics/bth261>>.
- FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS 2019. Food Outlook: biannual report on global food markets. Rome / Italy, FAO.
- Figueiredo, B.B., Cesar, F.I.G., 2021. Estudo comparativo da eficiência na produção de álcool de segunda geração. *Revista Científica Acertte*, ISSN 2763-8928, 1(6), e1643. <https://doi.org/10.47820/acertte.v1i6.43>.
- Hall, M.A., 1999. Correlation-based feature selection for machine learning. P. 198. Thesis (Doctor of Philosophy) - Department of Computer Science. University of Waikato. Hamilton (New Zealand). Disponível em: <<https://www.cs.waikato.ac.nz/~mhall/thesis.pdf>> Acesso em: 23.ago.2021.
- Ignacio, A., Carreras, V., Kaynig, C., Rueden, K.W., Eliceiri, J., Schindelin, A.C.H., Sebastian, S., 2017. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* 33(15), p. 2424–2426. DOI: <https://doi.org/10.1093/bioinformatics/btx180>.
- Koprinska, I., Rana, M., Agelidis, V.G., 2015. Correlation and instance based feature selection for electricity load forecasting. *Knowledge-Based Systems*, v.82. p.29-40.
- Leite, R.S., Cortez, L.A.B., 2020. O Etanol Combustível no Brasil. Disponível em: <http://alunosonline.uol.com.br/geografia/prod-cao-etanol-no-brasil.html>. Acesso em: 03.jul. 2021.
- Lorenzi, B.R., Andrade, T.N., 2019. O etanol de segunda geração no Brasil: políticas e redes sociotécnicas. *Revista Brasileira de Ciências Sociais* 34(100).

- NovaCana, 2008. Uso da água na produção de etanol de cana-de-açúcar. Campinas: Proamb. Engenharia. Projeto Programa de Pesquisa em Políticas Públicas. Disponível em: <<https://www.novacana.com/cana/uso-agua-producao-cana-etanol>>. Acesso em: 22.ago.2021.
- Oliveira, A.C.S., Souza, A.A., Lacerda, W.S., Gonçalves, L.R., 2010. Aplicação de redes neurais artificiais na previsão da produção de álcool. *Ciências e Agrotecnologia* 34 (2).
- Pacheco, T.F., 2011. Produção de Etanol: primeira ou segunda geração? Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/32985/1/CITE-04.pdf>>. Acesso em 20 ago.2021.
- Penedo, A.S.T., Pacagnella Júnior, A.C., Oliveira, M.M.B., 2007. Previsão de preços do açúcar utilizando redes neurais artificiais. *Núcleos* 4(1-2).
- PROÁLCOOL Programa Brasileiro de Álcool, 2008. Proálcool: Lição de Vida. Disponível em: <http://www.biodieselbr.com/proalcool/proalcool.htm>. Acesso em: 01 set..2021.
- Rodrigues, L.D., 2010. A cana-de-açúcar como matéria-prima para a produção de biocombustíveis: impactos ambientais e o zoneamento agroecológico como ferramenta para mitigação. Disponível em:< <http://atividadarural.com.br/artigos/5601927a79cad.pdf>>. Acesso em: 01 ago.2021.
- Ronquim, C. C.; Fonseca, M. F., 2018. Avanço das áreas de cana-de-açúcar e alterações em áreas de agropecuária no interior paulista. Campinas: Embrapa Territorial.
- Santos, L.M.A., Marques, H.L., Sousa, V.C. Rabesco, G.C., Ferreira, R.M., Marques, H.P.G., Oliveira, D.A., Vieira, P.T., 2019. Produção de álcool anidro mais limpo. *Revista Científica Multidisciplinar - Núcleo do Conhecimento* 03, 67- 80.
- Souza, C.C., Carneiro Junior, J.B.A., Cristaldo, M.F., Castelhão, R.A., Frainer, D.M., Viganó, H.H.G., 2022. Previsão de produção de cana-de-açúcar e seus derivados, através de ARIMA, no período de 2020 a 2025, na macrorregião Centro-Sul do Brasil. *Research, Society and Development* 11(4). DOI: <http://dx.doi.org/10.33448/rsd-v11i4.27048>.
- Spancersk, A.S., Santos, J.A.A., 2021. Previsão do preço do açúcar: um estudo de caso usando o software Weka. *Revista de Engenharia e Tecnologia* 13(1).
- Toledo, K., 2012. Etanol de segunda geração com biogás. FAPESP. Fundação de Amparo à Pesquisa de São Paulo
- UDOP. União Nacional de Bioenergia, 2003. A História da Cana-de-açúcar. Da Antiguidade aos Dias Atuais. Disponível em <https://www.udop.com.br/noticia/2003/01/01/a-historia-da-cana-de-acucar-da-antiguidade-aos-dias-atuais.html>. Acesso em 10 ago.2021.
- UNICA. Observatório da Cana, 2021. Históricos de moagem de cana-de-açúcar, área plantada e colhida, produção de açúcar e etanol no Brasil. Disponível em: <https://observatoriodacana.com.br/>. Acesso em 10 ago.2021.
- Viegas, T., 2010. Competitividade internacional do etanol brasileiro: oportunidades e ameaças. Rio de Janeiro: Blog Infopetro, Instituto de Economia: Universidade Federal do Rio de Janeiro.
- Wissmann, M.A., Oyamada, G.C., Wesendonck, C.C., Shikida, P.F.A., 2014. Evolução do cultivo da cana-de-açúcar na região Centro-Oeste do Brasil. *Revista Brasileira de Desenvolvimento Regional* 2(1), 95-117.
- Witten, I.H., Frank, E., Hall, M., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers46-Referencias Inc. ISBN 0123748569, 9780123748560.
- Zdravko, M., Ingrid, R., 2006. An introduction to the WEKA data mining system. *ACM SIGCSE Bulletin* 38(3), pp. 367–368. DOI: <https://doi.org/10.1145/1140123.1140127>.