

**UNIVERSIDADE FEDERAL DE ALAGOAS
UNIDADE ACADÊMICA CENTRO DE TECNOLOGIA
CURSO DE ENGENHARIA QUÍMICA**

BRUNO KELVIN VIRGÍNIO DA SILVA

**AVALIAÇÃO DE MÉTODO PARA DETERMINAÇÃO DE AÇÚCARES
REDUTORES TOTAIS EM CANA DE AÇUCAR POR ESPECTRSCOPIA DE
INFRAVERMELHO PROXÍMO (NIR)**

Maceió
2024

BRUNO KELVIN VIRGÍNIO DA SILVA

**AVALIAÇÃO DE MÉTODO PARA DETERMINAÇÃO DE AÇÚCARES
REDUTORES TOTAIS EM CANA DE AÇUCAR POR ESPECTRSCOPIA DE
INFRAVERMELHO PROXÍMO (NIR)**

Trabalho de Conclusão de Curso apresentado
ao curso de Engenharia Química da
Universidade Federal de Alagoas como
requisito para obtenção do título de Bacharel
em Engenharia Química.

Orientador: Prof. Dr. Jorge José de Brito Silva

Coordenador: Prof. Dr. Frede de Oliveira
Carvalho

Maceió
2024

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

S586a Silva, Bruno Kelvin Virgínio da.

Avaliação de método para determinação de açúcares redutores totais em cana de açúcar por espectroscopia de infravermelho próximo (NIR) / Bruno Kelvin Virgínio da Silva. – 2024.

37 f. : il. color.

Orientador: Jorge José de Brito Silva.

Coorientador: Frede de Oliveira Carvalho.

Monografia (Trabalho de Conclusão de Curso em Engenharia Química)
– Universidade Federal de Alagoas. Centro de Tecnologia. Maceió, 2024.

Bibliografia: f. 35-37.

1. Espectroscopia de infravermelho próximo. 2. Açúcares redutores totais (ART). 3. Cana-de-açúcar. 4. *Machine Learning*. I. Título.

CDU: 543.42

Folha de Aprovação

BRUNO KEVIN VIRGÍNIO DA SILVA

AVALIAÇÃO DE MÉTODO PARA DETERMINAÇÃO DE AÇÚCARES REDUTORES TOTAIS EM CANA DE AÇÚCAR POR ESPECTROSCOPIA DE INFRAVERMELHO PROXÍMO (NIR)

Trabalho de Conclusão de Curso apresentado
ao curso de Engenharia Química da
Universidade Federal de Alagoas como
requisito para obtenção do título de Bacharel
em Engenharia Química.

Documento assinado digitalmente
JORGE JOSE DE BRITO SILVA
Data: 13/11/2024 20:35:27-0300
Verifique em <https://validar.it.gov.br>

Orientador (a) - Dr. Jorge José de Brito Silva

Banca examinadora:

Documento assinado digitalmente
YANIA DE LOURDES DAS GRAÇAS TELES
Data: 13/11/2024 15:42:57-0300
Verifique em <https://validar.it.gov.br>

Examinador (a) – Dr. Vania de Lourdes das Graças Teles

Documento assinado digitalmente
ALTAIR MARQUES DA SILVA
Data: 18/11/2024 15:27:17-0300
Verifique em <https://validar.it.gov.br>

Examinador – Dr. Altair Marques da Silva

RESUMO

O presente trabalho teve como objetivo estudar a viabilidade técnica e econômica do uso da espectroscopia de Infravermelho Próximo (NIR) na determinação de açúcares redutores totais (ART) em cana-de-açúcar, comparando-o ao método convencional Concecana. O método tradicional exige várias etapas analíticas, como a amostragem, clarificação e uso de equipamentos específicos, sendo ineficiente nos dias atuais em termos de tempo e recursos. Por outro lado, a técnica NIR apresenta-se como uma alternativa rápida e não destrutiva, com potencial para substituir o procedimento tradicional. Para validar a equivalência dos métodos, foram realizados testes estatísticos, incluindo o teste t pareado e o uso de algoritmos de *Machine Learning*, utilizando redes neurais artificiais para comparar os dados gerados por ambos os métodos. Os resultados mostraram que o NIR é capaz de produzir dados comparáveis aos do método Consecana, com precisão estatisticamente significativa, além de uma expressiva redução de custos operacionais e de tempo. O estudo concluiu que a implementação do método NIR nas usinas de processamento de cana-de-açúcar é viável tanto técnica quanto economicamente, proporcionando uma alternativa mais eficiente e econômica para a determinação de ART, com potencial para melhorar a sustentabilidade e rentabilidade do setor sucroenergético.

Palavras - Chaves: CONSECANA, NIR, *Machine Learning*, ART.

ABSTRACT

This study aimed to evaluate the technical and economic feasibility of using Near-Infrared Spectroscopy (NIR) for the determination of total reducing sugars (TRS) in sugarcane, as an alternative to the traditional Consecana method used in sugar cane. The conventional method involves several analytical steps, such as sampling, clarification, and the use of specific equipment, making it time-consuming and resource-intensive. In contrast, NIR is a faster, non-destructive method with the potential to replace the traditional procedure. To validate the equivalence of both methods, statistical tests, including the paired t-test and *Machine Learning* algorithms using artificial neural networks, were applied to compare data from both techniques. The results demonstrated that NIR produces data statistically comparable to the Consecana method, with significant reductions in operational costs and processing time. The study concluded that implementing the NIR method in sugarcane processing plants is both technically and economically viable, offering a more efficient and cost-effective alternative for TRS determination, with potential benefits for the sustainability and profitability of the sugar-energy sector.

Keywords: CONSECANA, NIR, *Machine Learning*, TRS.

SUMÁRIO

1. INTRODUÇÃO	9
2. OBJETIVOS	11
2.1. GERAL	11
2.2. ESPECIFICO	11
3. REVISÃO IBLIOGRÁFICAS	12
3.1. CANA DE AÇÚCAR	12
3.2. MÉTODO CONVENCIONAL PARA DETERMINAÇÃO DE ART.....	14
3.2.1. Determinação do Peso do Bolo Úmido	15
3.2.2. Determinação de Brix.....	15
3.2.3. Determinação de Pol.....	16
3.2.4. Pureza do caldo.....	17
3.2.5. Açúcares redutores no caldo (AR).....	17
3.2.6. Açúcares redutores totais (ART)	18
3.3. ESPECTROSCOPIA DE INFRAVERMELHOR PRÓXIMO (NIR)	18
3.4. TRATAMENTO DE DADOS E ANÁLISE ESTATÍSTICA	21
3.4.1. Pré - Tratamento de Dados	21
3.4.2. Teste t Pareado.....	22
3.4.3. <i>Machine Learning</i>	23
4. METODOLOGIA.....	24
4.1. TRATAMENTO E PROCESSAMENTO DOS DADOS	24
4.1.1. Tratamento de dados.....	25
4.1.2. Teste t pareado.....	25
4.1.3. Treinamento da <i>Macahine Learning</i>	26
5. RESULTADOS E DISCUSSÕES	28
5.1. TRATAMENTO DE DADOS.....	28

5.2.	TESTE t PAREADO	30
5.3.	TREINAMENTO DA <i>Machine Learning</i>	30
6.	VIABILIDADE ECONÔMICA DE SUBSTITUIÇÃO	34
7.	CONCLUSÃO.....	36
	REFEERÊNCIAS	37

1. INTRODUÇÃO

No Brasil, a cultura da cana de açúcar está diretamente ligada ao desenvolvimento histórico-econômico. Sendo cultivada desde a época da colonização, quando em 1532 com Martin Afonso de Souza, trouxe a primeira muda para a capitania de São Vicente, que ficava nas divisas entre os estados, Minas Gerais, Rio de Janeiro, São Paulo e uma Parte da Região Norte do estado do Paraná, construindo ele mesmo posteriormente o primeiro engenho de açúcar. Tendo uma excelente adaptação as condições edafoclimáticas, propiciou um aumento na cadeia produtiva, sustentada pelo valor de produção (CHEQUIN & GRANDI, 2015).

Considerado um dos mais modernos no mundo, o complexo sucroenergético brasileiro, exerce a liderança como maior produtor mundial em todos os seguimentos: cana de açúcar, açúcar e álcool de cana.

O sistema de pagamento de cana baseado na sua composição foi oficializado em 1978 pelo Instituto de Açúcar e Alcool (IAA) no estado de Alagoas, tendo como base o teor de sacarose para determinar o valor da tonelada, e em 1983 em São Paulo (VALSECHI et al., 1983). Anterior a isso, o pagamento tinha como fator o peso do colmo, sendo o mais deficiente entre os critérios adotados no setor (IAA, 1978).

Em tempos atuais, o pagamento de cana tem como base os açúcares totais recuperáveis (ATR) dado em quilogramas por tonelada de cana (kg/t), composto por sacarose, glicose e frutose, estabelecido pelo Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Alcool (CONSECANA). Este sistema de pagamento é usado apenas no Brasil, sendo evidentemente mais racional, sabido que os açúcares redutores são matéria prima para produção de álcool.

O pagamento de cana em função da qualidade da matéria prima, apresenta uma grande vantagem, pois baseado no peso de açúcares contidos na cana, como acontecia no sistema do antigo IAA (teor de sacarose) e nos dias de hoje, no atual CONSECANA-SP (teor de ATR), tem como objetivo de recompensar o produtor de cana quando o mesmo entrega uma cana mais rica, fazendo assim com o que a indústria tenha capacidade de produzir um açúcar de melhor qualidade.

O ATR é determinado com base em parâmetros analíticos de composição, como fibra, POL (sacarose aparente) e açúcares redutores (AR), que também são utilizados para calcular os açúcares redutores totais (ART). Dessa forma, a determinação da quantidade de açúcares totais recuperáveis na cana é feita diretamente a partir do ART da cana.

A determinação desses parâmetros demanda tempo e mão de obra, resultando em altos custos com reagentes e pessoal para obtenção dos resultados analíticos. Para superar essas dificuldades, surgiu a proposta do uso do espectroscópio de infravermelho próximo (NIR). Embora seja um método de análise sofisticado, o sistema de análises por NIR se destaca por sua elevada robustez, possibilitando a substituição de grande parte das metodologias convencionais de laboratório. Isso reduz drasticamente custos desnecessários, incluindo a determinação de açúcares redutores totais (ART), ao mesmo tempo que mantém a qualidade e as especificidades exigidas para cada análise.

Considerando que uma unidade de processamento de cana-de-açúcar, devido à logística operacional agrícola, processa diversas variedades de cana e enfrenta variações na qualidade da matéria-prima ao longo da safra, surge a necessidade de realizar um estudo de viabilidade técnica e econômica do método, avaliando os resultados obtidos em diferentes condições de matéria-prima.

Dessa forma, nesse trabalho foi realizada a análise estatística para avaliar a viabilidade técnica do método NIR, sucedida pelo estudo de custo de implantação e operação, com base nos dados de custo operacional da Usina Serra Grande.

2. OBJETIVOS

2.1. GERAL

O objetivo desse estudo é otimizar o processo de determinação quantitativa de açúcares redutores totais em cana-de-açúcar, além de avaliar os custos do método convencional em comparação com o método NIR, para que haja a implementação do método instrumental de análise única de infravermelho próximo na usina Serra Grande.

2.2. ESPECIFICO

- Descrever o método atual;
- Descrever o novo método;
- Analisar estatisticamente a viabilidade técnica
- Avaliar a viabilidade financeira da implantação em relação ao custo atual.

3. REVISÃO IBLIOGRÁFICAS

Na revisão bibliográfica será possível compreender a cana como matéria prima e sua composição, com uma breve descrição sobre o método convencional de análise da mesma em laboratórios industriais, bem como uma abordagem sobre o NIR e ferramentas estatísticas adotadas no tratamento de dados.

3.1. CANA DE AÇÚCAR

Oriunda provavelmente da Índia, a cana foi inicialmente levada para costa do mediterrâneo, pelos árabes e pelos chineses, para a costa do Oceano Índico. Durante as batalhas no Extremo Oriente, os cavaleiros das cruzadas, ao retornarem para Europa, acostumados com uso do açúcar, desenvolveram um intenso comércio de açúcar produzido na costa do mediterrâneo, que durou até o início do século XVI (RODRIGUES E ROSS, 2020).

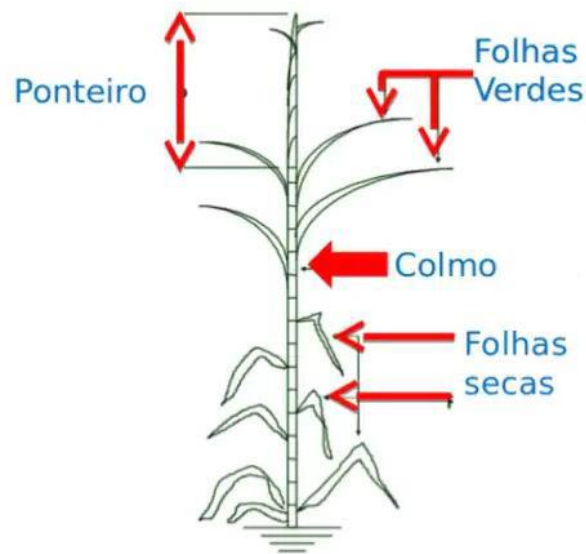
Ainda no século XVI, o monopólio da produção de cana de açúcar era mantido pelos portugueses, que já a cultivavam em ilhas do Atlântico, sendo neste mesmo período, levada a recém descoberta colônia, o Brasil, onde foram encontrados fatores favoráveis para o seu desenvolvimento (RODRIGUES E ROSS, 2020).

O Brasil é o atual líder em produção e exportação de Açúcar, sendo responsável por mais de 30% da produção mundial, apresentando uma receita cambial, aproximada de 8,7 bilhões de dólares em 2020, tendo seu máximo em 2017 quando atingiu 12 bilhões de dólares (Embrapa, 2022).

Alagoas é o maior produtor entre os estados do nordeste do Brasil, com produção na safra 2023-2024 de 19,5 milhões de toneladas de cana de açúcar, sendo aproximadamente 39,5% deste montante referente a cana de fornecedores, produtores de cana de açúcar para comércio com as usinas (Sindaçucar-AL, 2024).

A cana de açúcar é uma espécie de planta que pertence ao grupo de gramíneas perenes altas do gênero *Saccharum*, vindas das regiões tropicais e matéria prima direta para a produção de açúcar e etanol. A Figura 1 apresenta a composição geral da cana, com colmo e folhas.

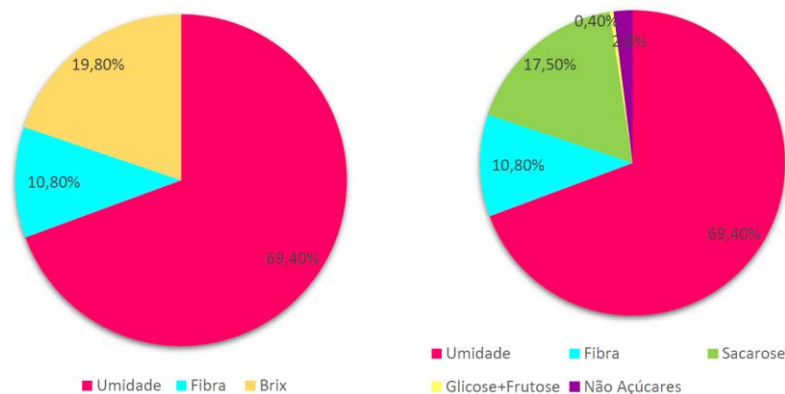
Figura 1: Composição da cana de açúcar



FONTE: RODRIGUES, 2015

Apresentando caules, denominados também como colmo, que se apresentam de forma robusta e fibrosa composto por abundantes quantidades de sacarose (NAANDANJAIN, 2017). O colmo da cana, formado basicamente por caldo e sólidos insolúveis, a Figura 2, mostra a formação em suas devidas proporções.

Figura 2: Composição do colmo da cana de açúcar.



Fonte: CTC, 2014

Os fatores necessários para uma produção de derivados de cana de açúcar são diversos, dentre eles: elevado teor de sacarose; baixo teor de fibra; ausência ou pouco florescimento; elevada resistência a pragas e doenças; pouca exigência em fertilidade e umidade; adequação da variedade ao ambiente (MENEGUETTI et al., 2010).

Na composição da cana de açúcar, encontra-se uma grande variedade de carboidratos, dentre eles os monossacarídeos, glicose e frutose denominados como açúcares redutores e dissacarídeo, sacarose que se apresenta em maior quantidade (HUGOT, 2014).

3.2. MÉTODO CONVENCIONAL PARA DETERMINAÇÃO DE ART

Nos laboratórios das unidades industriais, acontece o acompanhamento da qualidade da matéria prima fornecida, tanto o pagamento da mesma e quanto para os cálculos de eficiência de processo, onde os técnicos são responsáveis pela coleta dos diversos dados durante o procedimento analítico. Dentre esses processos podemos citar as seguintes etapas:

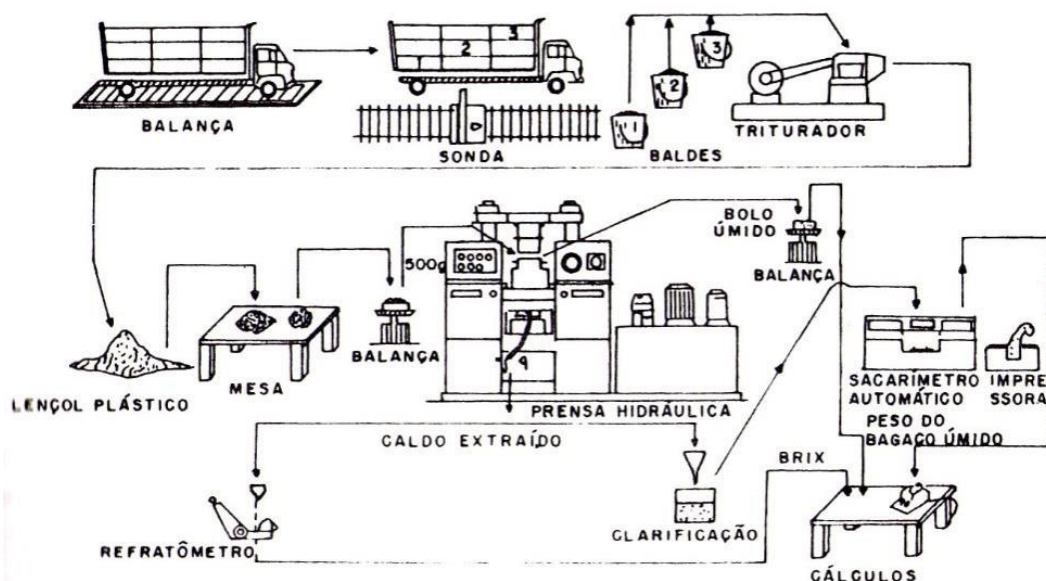
- Amostragem em sonda oblíqua;
- Prensagem da amostra para separação do caldo e bagaço;
- Pesagem do bolo úmido do bagaço;
- Clarificação do caldo para análise de Pol;
- Leitura em equipamento refratômetro quantificação de Brix;
- Leitura em equipamento sacarímetro para quantificação da Pol.

Para a determinação das características da cana de açúcar diversos procedimentos são utilizados (CALDAS, 2012). Para determinar o teor de açúcares redutores totais, existem vários métodos químicos, que fornecem resultados precisos, com alto grau de confiabilidade. Os métodos utilizados no setor podem ser fundamentados na redução de íons de cobre em soluções alcalina (solução de *Fehling*) (CALDAS, 2012) ou sendo na estimativa por equações com base em balanço de massa, que variam de acordo com as regiões do país (CALDAS, 2017).

O acompanhamento da qualidade da matéria-prima ocorre nos laboratórios industriais ou PCTS (Laboratório de Pagamento de Cana por Tonelada de Sacarose), das unidades industriais, onde técnicos e analistas são responsáveis por diversos procedimentos analíticos.

Para a determinação das características da cana de açúcar diversos procedimentos são utilizados. Na Figura 3 pode ser observado brevemente todas as etapas do processo convencional.

Figura 3: Descrição do método convencional



Fonte: IAA/PLANALSUCAR, 2018.

Para que sejam realizadas as análises da cana, se faz necessária uma amostragem mecânica, onde são coletados 10kg de amostra de cada caminhão, através de uma sonda, oblíqua ou horizontal, para que haja uma representatividade, após essa amostragem, a alíquota é desfibrada e segue para o laboratório de análises industriais.

3.2.1. Determinação do Peso do Bolo Úmido

A determinação do peso do bolo úmido (PBU) é dada pela seguinte metodologia, pesar-se 500g $\pm$ 0,5g da amostra desfibrada, coletada pela sonda. Logo após essa amostra é prensada de acordo com as especificações determinadas pela Comissão Internacional para Métodos Uniformes de Análise de Açúcar (ICUMSA) (250kgf/cm² por 1 minuto), O caldo obtido servirá para a determinação dos parâmetros de qualidade e a massa de solido obtida na prensagem será o PBU.

3.2.2. Determinação de Brix

A determinação de Brix (sólidos solúveis totais) do caldo extraído da amostra é realizada em um refratômetro digital. O procedimento consiste em coletar o caldo extraído na prensa hidráulica e posteriormente realizar a leitura no refratômetro. Deve-se verificar se a temperatura do laboratório encontra-se entre 18° e 25°C.

3.2.3. Determinação de Pol

Sendo um composto orgânico pertencente ao grupo dos carboidratos, a sacarose é um dissacarídeo que em sua inversão produz um equivalente de glicose e um equivalente de frutose. Na natureza é possível encontrar uma grande variedade de frutas que possuem sacarose em sua composição, no entanto, somente a cana-de-açúcar e a beterraba produzem sacarose o suficiente para a fabricação de açúcar em forma de cristais (PEREIRA, 2021).

O procedimento tem como base realizar a leitura sacarimétrica do caldo, determinada pelo sacarímetro digital, para isso são adicionados em um bequer, 100ml de caldo e em seguida, mistura-se ao caldo aproximadamente 3,0g de clarificante. Essa etapa é de extrema importância, pois nela acontece a remoção dos não açúcares presentes no caldo, mantendo a sacarose e facilitando a sua quantificação.

Após essa mistura e homogeneização, o caldo é filtrado em papel de filtro qualitativo, para que haja a clarificação. Após clarificado é feita a leitura no sacarímetro ou polarímetro.

Algumas considerações são feitas para os cálculos de terminação de pol do caldo (Pol%Caldo): (1) correção da temperatura da leitura sacarimétrica para 20°C e, (2) uso do clarificante.

Quando o equipamento não possui sistema de compensação automática de temperatura, como acontece com os mais antigos, a correção é feita seguindo a equação 1, definida pela ICUMSA.

$$L_{SAC20^{\circ}} = L_{SACLIDA} + (1 + 0,000255 * (T - 20)) \quad (1)$$

Onde:

$L_{sac\ 20^{\circ}}$: Leitura sacarimétrica corrigida à temperatura de 20°C.

T: Temperatura da Amostra no momento da Leitura (°C)

Em relação ao clarificante, ao ser instituído o método padrão de determinação de pol pela ICUMSA, a clarificação do caldo era feita por subacetato de chumbo (equação 2), que foi proibido por expor os analistas a riscos, devido a toxicidade. Com isso, foram criadas equações para leituras equivalentes.

$$Pol\%caldo = L_{SACPb} * (0,2605 - 0,0009882 * Brix\%caldo) \quad (2)$$

Onde:

L_{SACPb} : Leitura sacarimétrica da amostra clarificada com subacetato de chumbo.

- Clarificação com mistura clarificante à base de alumínio:

$$L_{SACPb} = 1,00621 * L_{Al} + 0,05117 \quad (3)$$

Onde:

L_{Al} : leitura sacarimétrica da amostra clarificada com clarificante à base de alumínio.

Sendo assim, a equação completa para cálculo da pol do caldo com uso desse clarificante passa a ser a seguinte:

$$Pol\%caldo = (1,00621 * L_{Al} + 0,05117) * (0,2605 - 0,0009882 * Brix\%caldo) \quad (4)$$

- Clarificação com octapol (mais usado)

$$L_{SACPb} = 0,99879 * L_{octapol} + 0,47374 \quad (5)$$

Onde:

$L_{octapol}$: leitura sacarimétrica da amostra clarificada com octapol.

Sendo assim, a equação completa para cálculo da pol do caldo com uso desse clarificante passa a ser a seguinte:

$$Pol\%caldo = (0,99879 * L_{octapol} + 0,47374) * (0,2605 - 0,0009882 * Brix\%caldo) \quad (6)$$

3.2.4. Pureza do caldo

A pureza do caldo indica a percentual m/m de sacarose aparente contida em 100g de solúveis (Brix), sendo dada pela equação:

$$Pza\%caldo = \frac{Pol\%caldo}{Brix\%caldo} * 100 \quad (7)$$

3.2.5. Açúcares redutores no caldo (AR)

A determinação dos açúcares redutores no caldo pode ser realizada por vários métodos, por exemplo, o titrimétrico por oxiredução conhecido por Lane- Eynon; por espectrofotometria como o Sommogy-Nelson, titulação potenciométrica automática e ainda por cromatografia líquida ou cromatografia iônica.

Quando consideramos a determinação de AR para o atual sistema de pagamento de cana no Brasil e rapidez nas informações, equações, foram estabelecidas baseadas nas reações naturais da cana e no fato de que quanto maior a pol, por consequência maior a pureza do caldo. Desta forma, a seguir são apresentadas três equações que são usadas atualmente no Brasil.

$$AR\%caldo = 3,3459 - 0,02871 * Pza\%caldo \quad (8)$$

$$AR\%caldo = 6,9539 - 0,0688 * Pza\%caldo \quad (9)$$

$$AR\%caldo = 03,641 - 0,0343 * Pza\%caldo \quad (10)$$

As equações 8, 9 e 10, mostram como são calculados os valores de AR nos estados de Alagoas, Pernambuco e na região sudeste em geral, respectivamente, definidas pelo Consecana de cada um desses estados.

3.2.6. Açúcares redutores totais (ART)

Os açúcares redutores totais, assim como o AR, podem ser determinados tanto analiticamente, como calculado a partir de equações, calculado a partir de transformações estequiométricas, a partir da Pol e do AR.

$$ART\%caldo = \left(\frac{Pol\%caldo}{0,95} \right) + AR\%caldo \quad (11)$$

Onde, 0,95 é o fator de conversação m/m de ART em sacarose, baseado em que: 342kg de ART, equivalem a 360kg de sacarose.

3.3. ESPECTROSCOPIA DE INFRAVERMELHOR PRÓXIMO (NIR)

Métodos espectroscópicos, conseguem ser extremamente vantajosos em relação a outros métodos analíticos, como pequena quantidade de amostra, processo não destrutivo e o principal, a alta frequência de amostragem. Quando inserido no quesito de análise de cana, apresenta uma grande melhoria de processo, pela agilidade da obtenção de resultados.

A espectroscopia é definida como uma técnica utilizada para estudar a interação entre a radiação eletromagnética e a matéria, analisando como a radiação é absorvida, emitida ou dispersa por substâncias para fornecer informações sobre suas propriedades físicas e químicas (PAVIA, 2001).

Espectroscopia de infravermelho (IR) é uma técnica baseada nas vibrações dos átomos em uma molécula. Um espectro de infravermelho é obtido passando radiação infravermelha

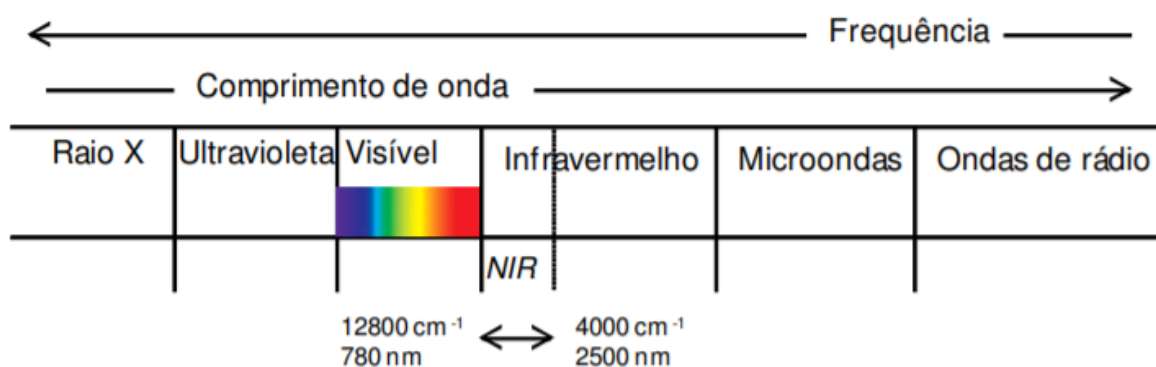
através de uma amostra e determinando que fração da radiação incidente é absorvida em uma energia particular. A energia em que um pico de absorção aparece corresponde à frequência de vibração de parte da molécula da amostra (STUART, 2004).

O NIR (infravermelho próximo) é apresentado como uma técnica de espectroscopia que utiliza a região do espectro eletromagnético com comprimentos de onda entre 700 nm e 2500 nm. Esta técnica é amplamente utilizada para análises qualitativas e quantitativas, especialmente em sistemas biológicos e materiais complexos, devido à sua capacidade de penetrar profundamente na matéria e de analisar amostras com pouca ou nenhuma preparação (PAVIA, 2001).

Os compostos orgânicos ou inorgânicos em sua maioria, apresentam ligações covalentes, ligações essas que são capazes de absorver radiações na região de espectro de infravermelho. No espectro de luz, que vai de raio X à ondas de rádio, o infravermelho (IR, infrared) limita-se entre a região visível e a região de micro-ondas, tendo seu comprimento de ondas compreendendo a faixa entre 12800 a 10 cm^{-1} . É comum caracterizar as bandas de absorção em unidades de números de ondas (ν), que são expressas em centímetros recíprocos (cm^{-1}). Às vezes, comprimentos de onda (λ) medidos em micrômetros (μm) também são utilizados (PEREIRA, 2021).

Na Figura 4, está definida o espectro eletromagnético com as faixas específicas do infravermelho próximo.

Figura 4: Espectro eletromagnético



Fonte: INFOESCOLA, 2006.

O espectro de ondas infravermelho está dividido em três regiões, IR-próximo, IR-médio e IR-distante, sendo cada faixa usada de acordo com o tipo de aplicação e instrumento a ser utilizado, como pode ser visto na Tabela 1.

Tabela 1: Divisão do espectro eletromagnético na região do infravermelho em três partes principais

Região	Comprimento de onda/ Frequência	Característica
Infravermelho Próximo	0,8-2,5 μm (12500-4000 cm^{-1})	Região harmônica/ Combinação
Infravermelho Médio	2,5-50 μm (4000-200 cm^{-1})	Região de Vibração-Rotação
Infravermelho Distante	50-1000 μm (200-10 cm^{-1})	Região de Rotação

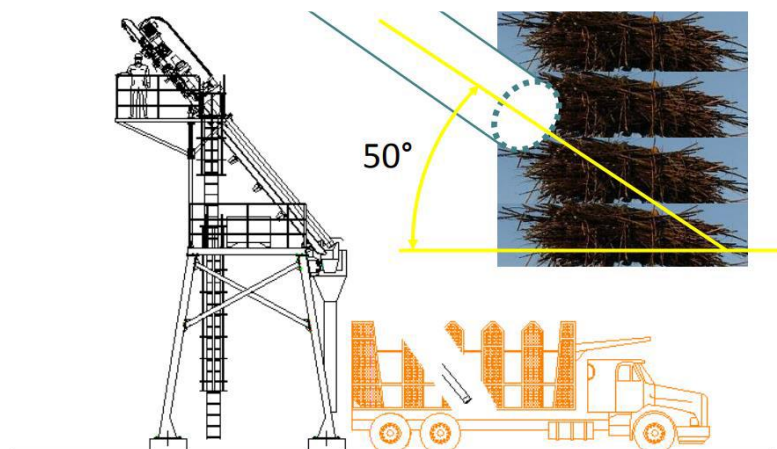
Fonte: LIMA, et al, 2009.

Durante o processo de absorção de radiação infravermelho, a molécula sofre variação do momento dipolo por consequência dos movimentos vibracional ou rotacional da mesma. Por esse motivo temos como exemplo os gases cloro e iodo, que não possuem espectro de IR, pois os momentos dipolo resultante é igual a zero (OLIVEIRA, 2008).

O método convencional, como citado anteriormente, possui diversas etapas, constituindo-se de uma intensa mão de obra e custo operacional, além do risco eminente de prensagem de mãos ao realizar a desfibragem da cana. A operação nos dias hoje em um laboratório de pagamento de cana-de-açúcar acontece após pesagem do caminhão na balança através de uma sonda, seja ela oblíqua, horizontal ou trator para retirada de uma amostragem.

Para o processo no laboratório de pagamento de cana-de-açúcar escolhido para o estudo, se inicia com a amostragem por sonda oblíqua, como indica a Figura 5, devido aos diversos benefícios que ela apresenta, garantindo uma amostragem mais representativa da amostra.

Figura 5: Sonda oblíqua em atuação



Fonte: BIAGI, 2018.

erros grosseiros, ao considerarmos um processo estatístico aplicado a um conjunto de dados e neste processo observam-se pontos que se desviam do padrão. Dessa forma, podemos definir *outliers* como todo objeto que se distancie de forma considerável dos demais dados (FREITAS, 2019).

Podemos definir um *outlier*, como uma anomalia no padrão esperado. Sendo assim, a forma mais inteligente para detecção de *outlier*, seria definir uma região de comportamento normal dos dados. Mesmo aparentando ser uma estratégia de simples execução, apresenta alguns desafios, desde definir a região normal dos dados às ações maliciosas de colaboradores responsáveis pela geração dos dados (CHANDOLA; BANERJEE; KUMAR, 2009). O método $3\sigma_x$ -rule é a abordagem mais simples para *outliers* univariados (SOUZA, ARAUJO & MENDES, 2016),

3.4.2. Teste t Pareado

Para a comparação de dois grupos, uma boa ferramenta a se aplicar é o teste t pareado (para grupos dependentes) (PAES, 2008). O teste t pareado, ou t de *Student* pareado, é uma técnica estatística utilizada para comparar as médias de duas amostras relacionadas. Geralmente, é aplicado quando temos duas medições correspondentes para os mesmos indivíduos ou grupos, como em estudos antes e depois de uma intervenção.

O teste t pareado é usado para determinar se a mudanças estatisticamente significantes entre as medias de dois conjuntos de dados pareados. Aplicando em variáveis pareadas dispostas de forma continua e normalmente distribuídas (MISHRA et al., 2019).

Para aplicação do teste devem ser realizados números iguais de medições, para ambos os métodos analisados. Calcula-se o módulo da média ($\bar{X}_d$) e desvio padrão da média (s_d) da diferença os dois métodos (ALFASSI et al., 2005), sendo o zero tomado como valor de referência, além dessas duas variáveis, para chegar ao valor de t, é necessário o número de observações (n), como mostra a equação 12:

$$t = \frac{|\bar{X}_d|\sqrt{n}}{s_d} \quad (12)$$

Toda via, calcular apenas o valor t, não significa nada, é necessária uma comparação com valores de t crítico ou tabelados, previamente estabelecidos pelo nível de significância. Quando valor t calculado, é maior que o valor t tabelado e determinado pelo número de graus de liberdade, ou p-valor menor que o nível de significância escolhido, há diferença entre os dois métodos e a hipótese nula será rejeitada, como valor t calculado, seja menor que o t tabelado, a

hipótese nula é aceita e é considerado que os métodos são estatisticamente equivalentes (ESTRELLA, 2023).

O teste t pareado é de certa forma um teste qualitativo, baseado na relação entre os valores de t crítico e t calculado para o conjunto de dados que tem seu número de amostra limitados a um intervalo de 10 a 30 pares (COHEN, 2013), visto isso, foi aplicado um método complementar mais robusto estatisticamente, *Machine Learning*, para avaliar as quão similares são, os métodos analíticos, Consecana e NIR, através do cruzamento de dados das duas metodologias partindo da premissa da capacidade da *Machine Learning* em mapeamento multivariável não linear.

3.4.3. *Machine Learning*

Podemos entender *Machine Learning* como a área da inteligência artificial que tem como base modelos empíricos, fazendo com que computadores possam aprender usando conjuntos de dados, sem serem programados diretamente para executar determinada função (RUSSELL E NORVING, 2020).

Desenvolver algoritmos voltados a *Machine learning* tem se tornado cada vez mais comum, pois possibilita o aprendizado de padrões a partir das características de conjunto de dados, na maioria das vezes não lineares (GÉRON, 2019).

A forma de aprendizagem de uma *Machine Learning* varia de acordo com os conjuntos de dados disponíveis, podendo essa aprendizagem acontecer de forma não-supervisionada, quando os dados de saída não estão disponíveis ou definidos, fazendo com que o algoritmo procure por padrões no conjunto de dados de entrada, unicamente; ou de forma supervisionada, onde os dados de entrada usados para o aprendizado da *Machine Learning*, já tem dados de saída correspondentes e conhecidos, sendo esses obtidos de forma experimental ou rotulados por especialistas (GÉRON, 2019; JAMES et al., 2013)

4. METODOLOGIA

Durante o a safra, no setor sucroenergético, são feitas determinações de qualidade e composição da cana recebida pela unidade industrial, para que seja efetuado os pagamentos aos fornecedores, também para quantificar os rendimentos e eficiências industriais.

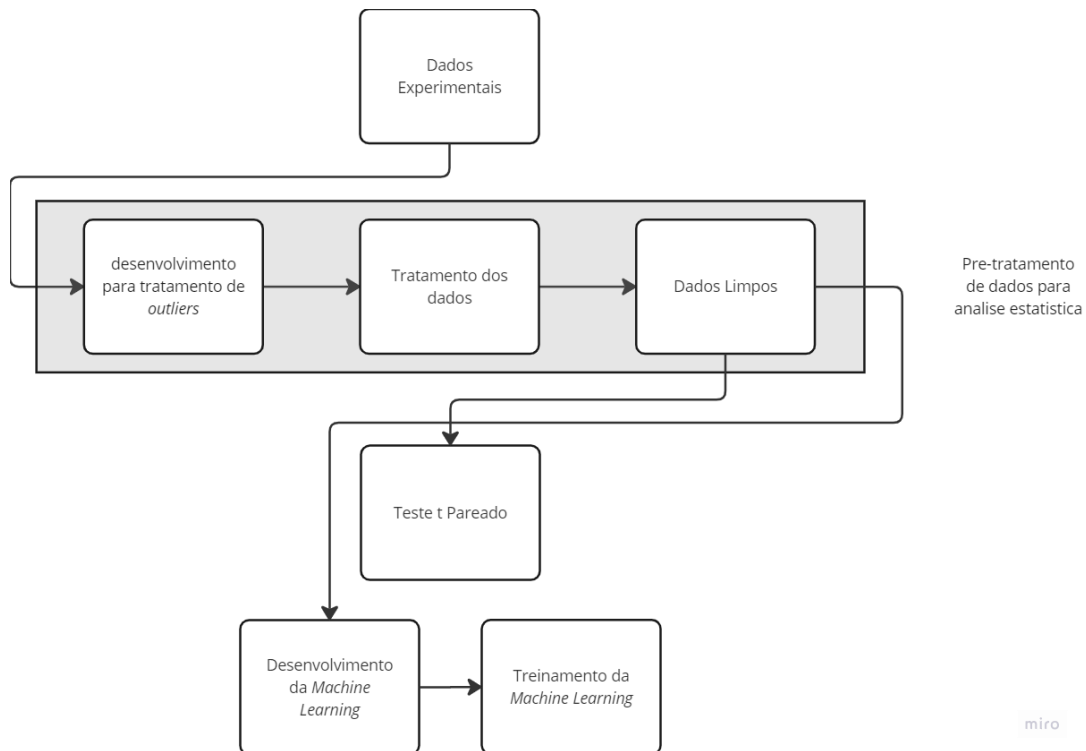
Ao se falar nessas determinações por métodos instrumentais, as unidades do consecana estabelecem alguns padrões para as determinações de paridade das análises, com isso, as empresas durante a safra devem determinar 100 pares de análises, a cada terço do período de moagem (CONSECANA-SP, 2021).

As quantificações são realizadas de modo a amostrar os caminhões que chegam à unidade industrial. Os caminhões que transportam cana de fornecedores são amostrados em sua totalidade, enquanto os caminhões que transportam a cana produzida pela própria usina (Cana Própria) devem ser amostrados em, no mínimo, 30%.

4.1. TRATAMENTO E PROCESSAMENTO DOS DADOS

A Figura 7 mostra um fluxograma da metodologia usada para o tratamento e processamento estatístico dos dados.

Figura 7: Metodologia adotada para a análise de dados do trabalho



Fonte: Autor, 2024

Os dados utilizados para o treinamento das estruturas de *Machine Learning* foram obtidos a partir de análises em laboratórios convencionais e com o equipamento NIR de cana, para processamento e análise em linguagem *Python*.

Com o avanço da tecnologia, a linguagem *Python* vem se tornando cada vez mais popular entre os profissionais de engenharia, tendo como principal vantagem o baixo custo e sua ampla disponibilidade de bibliotecas, graças à grande comunidade de usuários, que busca implantar diversas funcionalidades.

Para a realização deste trabalho, foram utilizadas algumas bibliotecas, incluindo *NumPy*, para matemática computacional e cálculos de álgebra linear; *Pandas*, para leitura e manipulação de dados; e *Scikit-Learn*, para aplicação de *Machine Learning*.

4.1.1. Tratamento de dados

Existem dois tipos de *outliers*: univariados e multivariados (SOUZA, ARAUJO, MENDES, 2016). No caso deste trabalho serão tratados os *Outliers* univariados, que são obtidos da observação de uma distribuição de valores num único espaço dimensional.

A detecção de *outliers* teve como base ferramentas estatísticas. O método $3\sigma_x$ -rule que assume uma distribuição normal dos dados. De acordo com esse método, os valores que x_n que não satisfaçam.

$$|x_n - \bar{x}| > 3\sigma_x, \quad n = 1, \dots, N, \quad (13)$$

O algoritmo para detecção e remoção dos dados inconsistente, foi inscrito em *Python*, com base na equação 13 considerados *outliers*. Representam-se $\bar{x}$ e σ_x como a média e o desvio padrão da variável x (com N amostras). Foi utilizado a blibliotecas *Pandas* e *NumPy*.

4.1.2. Teste t pareado

O teste t pareado é uma técnica estatística utilizada para comparar as médias de duas amostras relacionadas, esse teste é geralmente aplicado quando os dados são coletados, de uma mesa amostra em condições de análises diferentes.

Para execução do teste, foi desenvolvido um algoritmo em *Python*, com o uso da biblioteca *Pandas* e a função *Stats* da biblioteca *SciPy*, onde devido ao grande volume de dados, foram feitas iterações coletando 30 pares de dados de ART de forma randômica, para execução do teste, seguido da avaliação de todos os resultados, sendo rejeitada a validade do método NIR, caso alguma das iterações fosse rejeitada.

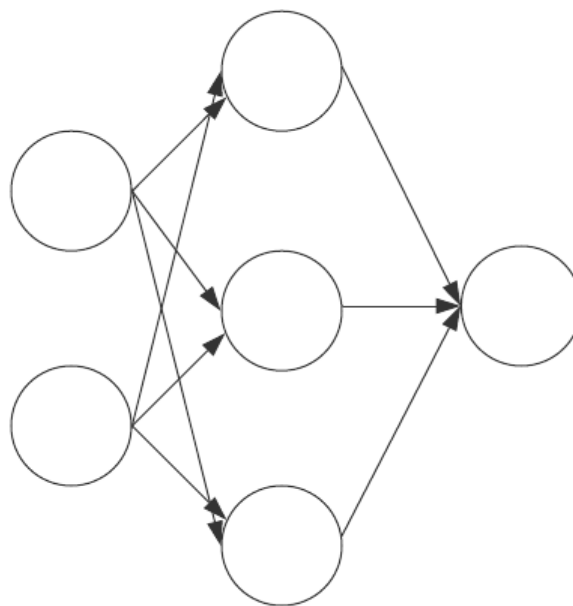
4.1.3. Treinamento da *Machine Learning*

Os dados para treinamento, foram obtidos de forma experimental no Laboratório industrial de uma usina no sudeste do país, a mesma segue as normas do Conselho dos produtores de Cana de Açúcar, Açúcar e Etanol do estado de São Paulo (Consecana – SP), que diz, que ao utilizar o método NIR, a cada terço do período de moagem, deve-se obter um mínimo de 100 pares de dados para envio ao consecana para alimentação da curva padrão vendida por eles e dez pares de dados por turno, para corrigir e aumentar a robustez da curva da própria usina (CONSECANA, 2022).

A *Machine Learning* usada foi uma rede neural artificial (RNA) MLP (*Multilayer Perceptron*), uma RNA multicamadas que tem como objetivo resolver um problema de regressão, entre os dados de POL, Brix e Fibra.

Foram utilizados aproximadamente 5000 pares de dados, sendo 80% para treinamento da RNA e 20% para teste e validação. O treinamento foi feito inicialmente com duas entradas e uma saída com dados de POL, Brix e Fibra, para os dados do método tradicional Consecana e NIR, como mostra a Figura 8. Após o treinamento e validação, a RNA foi treinada com dados cruzados, ou seja, entradas com dados do método tradicional e saída, com dados do método NIR, assim como, entradas com dados do método NIR e saída com dados obtidos pelo método tradicional.

Figura 8: Estrutura de treinamento das células de *Machine Learning*.



Fonte: Autor, 2024

A *Machine Learning* foi desenvolvida usando a linguagem Python, com uso das bibliotecas, Pandas, para leitura e manipulação dos dados e *Sklearn*, para aplicação da MLP, com métricas de desempenho como Score (Coeficiente de determinação R^2) e RMSE (erro médio quadrático), dispostas nas equações 14 e 15, foram calculadas para avaliar a precisão.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

$$RMSE(y, \hat{y}) = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (15)$$

Sendo R^2 o coeficiente de determinação, $\hat{y}_i$ o valor predito da amostra i , y_i o valor real correspondente num total de n .

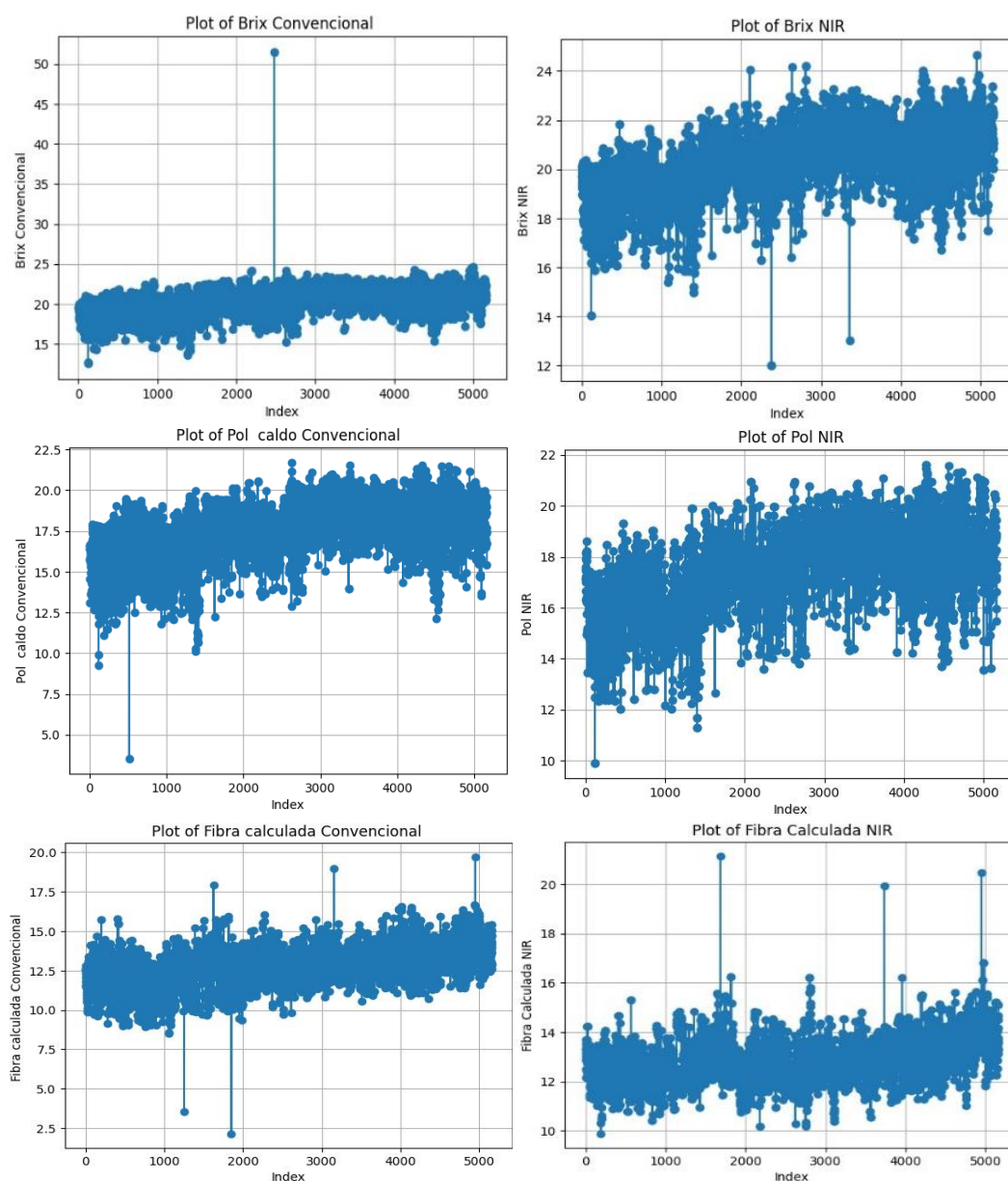
5. RESULTADOS E DISCUSSÕES

São apresentados a seguir os resultados obtidos com base na metodologia adotada no trabalho.

5.1. TRATAMENTO DE DADOS

Os dados foram tratados, segundo a metodologia $3\sigma_x$ -rule, aplicado aos 5164 pares de dados existentes, a Figura 9 a seguir, mostra a o comportamento dos dados, antes da retirada dos valores inconsistentes de POL, Brix e Fibra, determinados experimentalmente pelo método Consecana e método NIR.

Figura 9: Representação gráfica do valor de Brix, POL e Fibra, método convencional e NIR (Sem tratamento).

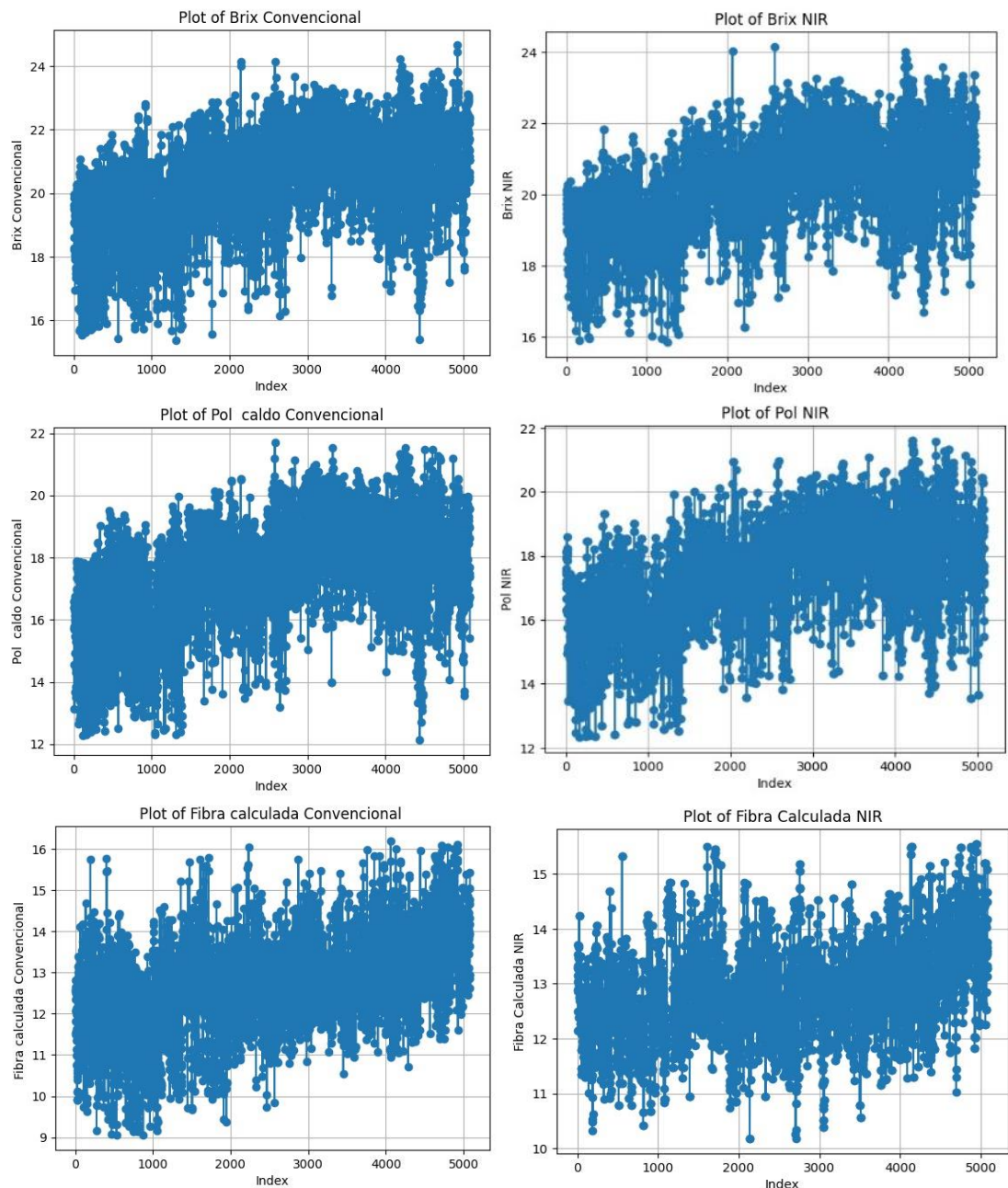


Fonte: Autor, 2024.

Como visto na Figura 9, o conjunto de dados apresenta um número considerável de *outliers*, o que impacta diretamente no comportamento da máquina virtual, impossibilitando uma boa aferição dos tatos em seu treinamento.

A seguir, na Figura 10, é possível avaliar o impacto do tratamento no conjunto de dados de processo, das variáveis, Brix, POL e Fibra.

Figura 10: Representação gráfica do valor de Brix, POL e Fibra, método convencional e NIR (Com tratamento).



Fonte: Autor, 2024.

Ao observar a figura 10, é possível observar a redução nas variações dos dados analíticos, o processamento de dados, serviu para a retirada de 84 pares de cada variável, que

corresponde a 1,63%, reduzindo os pontos de maior inconsistência, pontos esses de possíveis erros operacionais.

5.2. TESTE t PAREADO

A Figura 11 mostra o algoritmo usado para a aplicação do teste t, o mesmo avaliou 5084 pares de dados, em blocos de 30 pares, escolhidos de forma Randômica.

Figura 11: Algoritmo usado para excursão do teste t pareado.

```
import numpy as np
import pandas as pd
from scipy import stats

# Leitura do arquivo
entrada = 'ARTConv_ARTNIR.xlsx'
df = pd.read_excel(entrada)

ART_Convencional = df['ART_Convencional']
ART_NIR = df['ART_NIR']

def teste_t_pareado(ART_Convencional, ART_NIR, indices):
    # Seleção de amostra baseada em índices fornecidos
    before_sampled = ART_Convencional.iloc[indices]
    after_sampled = ART_NIR.iloc[indices]

    # teste t pareado
    t_statistic, p_value = stats.ttest_rel(before_sampled, after_sampled)

    return t_statistic, p_value

# Configurações
num_pares = 30 # Número de Pares
num_iterations = len(ART_Convencional) // num_pares # Número de Iterações

# Shuffle dos índices para garantir que os pares não se repitam
indices = np.random.permutation(len(ART_Convencional))

results = []

for i in range(num_iterations):
    # Seleção de pares consecutivos sem repetição
    subset_indices = indices[i*num_pares:(i+1)*num_pares]
    t_statistic, p_value = teste_t_pareado(ART_Convencional, ART_NIR, subset_indices)
    results.append((t_statistic, p_value))

testet = 0

# Exibir resultados
for i, (t_stat, p_val) in enumerate(results):
    print(f"Iteração {i+1}: Estatística t = {t_stat}, Valor-p = {p_val}")

    if p_val < 0.05:
        testet += 1

print(testet)
```

Fonte: Autor, 2024.

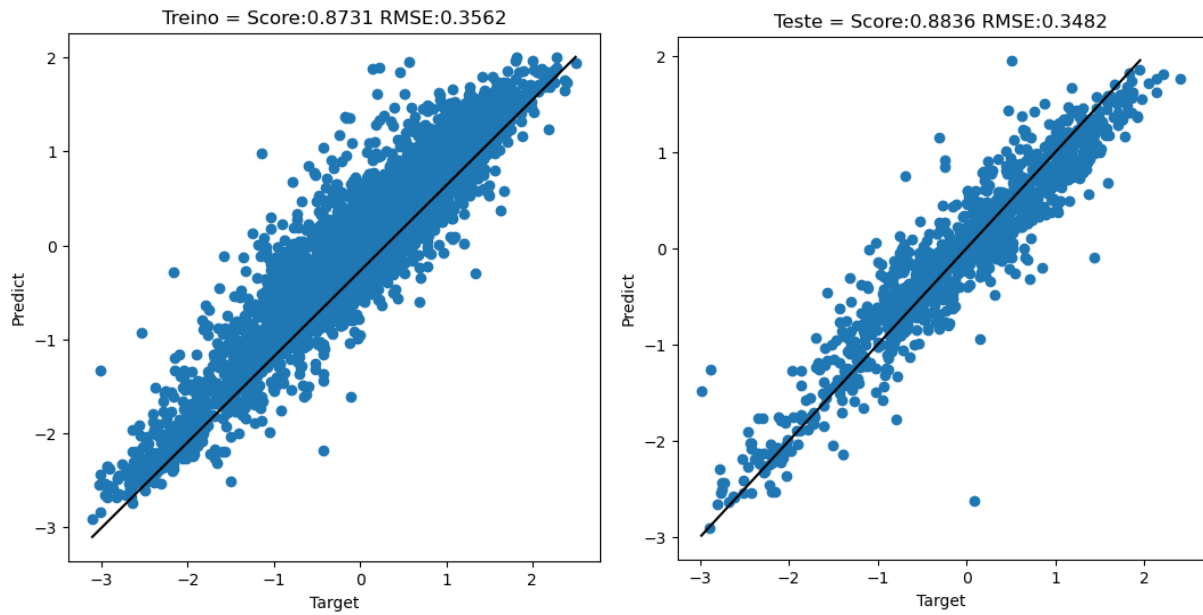
Após a execução do algoritmo, foram feitas 169 iterações, das quais nenhuma apresentou p-valor menor que o definido com nível de significância escolhido de 0,05, significando que não há diferença significativa entre os conjuntos de dados emparelhados.

5.3. TREINAMENTO DA *Machine Learning*

Na etapa de treinamento, foram utilizados 3 neurônios na estrutura de aprendizado, nas Figuras 12 a 15, os dados reais livre de *Outliers* versus os preditos para a estrutura MLP, de forma que, a estrutura treinada, compara os dados de teste, sendo esses Consecana ou NIR, com o padrão assimilado pela *Machine Learning*, tendo como variáveis de entrada a concentração de Brix e de Fibra e, variável de saída a concentração de POL.

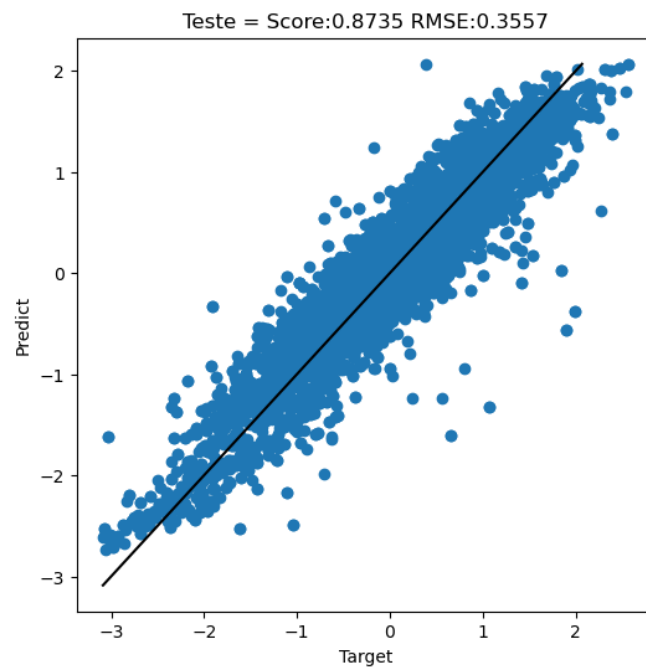
Na figura 12, é possível visualizar a MLP treinada com dados de entrada e saídas do método tradicional Consecana.

Figura 12: Dados de entrada e saída obtidos com o método Tradicional Consecana.



Fonte: Autor, 2024.

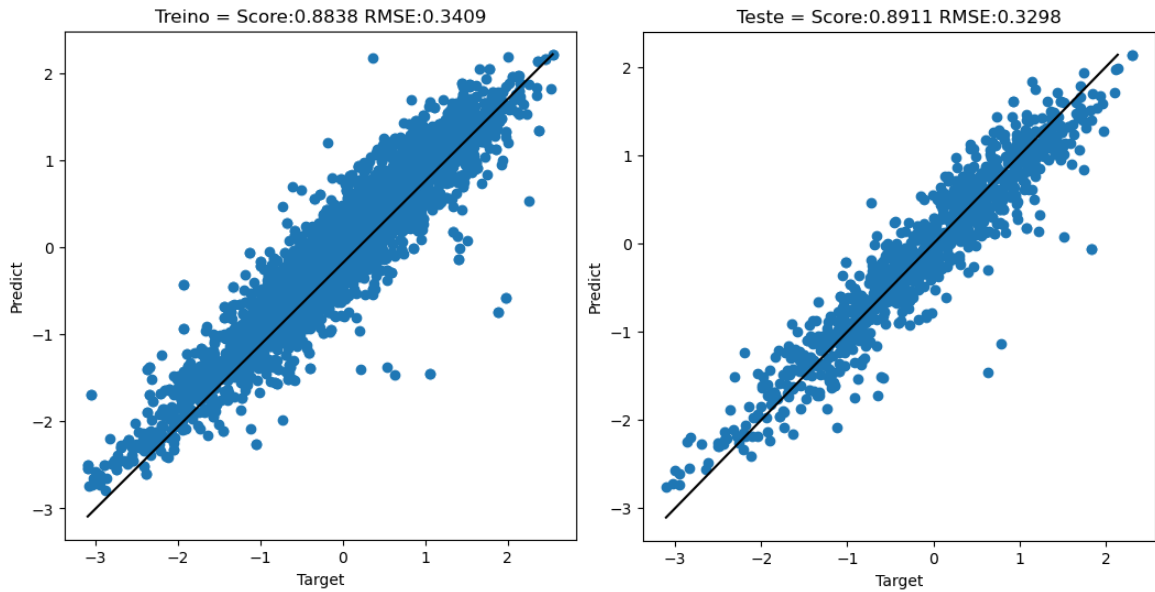
Figura 13: Dados de entrada obtidos com o método Tradicional Consecana e saída método NIR.



Fonte: Autor, 2024.

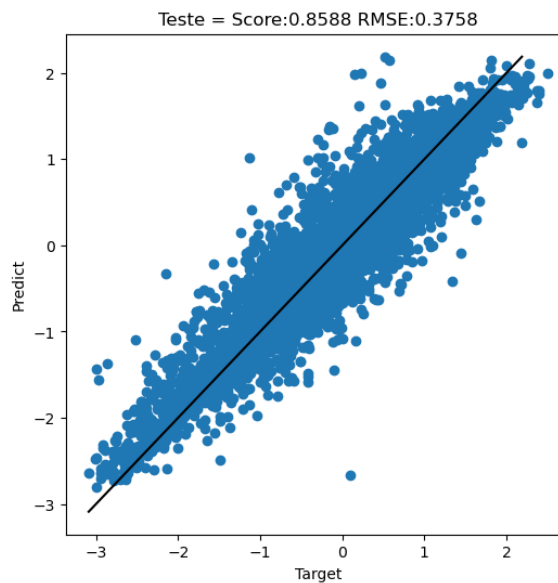
Na Figura 13, é possível visualizar a MLP treinada com os dados de POL e Brix, do método tradicional, correlacionando com os dados de método de análise instrumental NIR.

Figura 14: Dados de entrada e saída obtidos com o método instrumental NIR.



Fonte: Autor, 2024.

Figura 15: Dados de entrada obtidos com o método NIR e saída método Tradicional Concecana.



Fonte: Autor, 2024.

Como era de se esperar, existe uma similaridade estatisticamente significativa, a 95% de confiança, entre os métodos, como é visto pelas métricas de validação. Ao treinar e testar a MLP, com os dados de Brix, Fibra e POL determinados por métodos convencionais Concecana, foi possível observar que as variáveis de validação R^2 e RMSE, ficaram com Valores superiores a 0,87 e inferiores 0,34. O mesmo se repetiu quando os dados de saída a serem comparados com os preditos pela estrutura treinada com dados convencionais, foram os dados de POL do método NIR, validando estatisticamente uma possível substituição.

O mesmo processo foi repetido, desta vez treinando a RNA com dados de Brix, Fibra e POL determinados pelo método instrumental NIR. Durante o treinamento e teste, correlacionando as entradas obtidas pelo NIR com as saídas do método Consecana, foram encontrados valores de R^2 de 0,88 e 0,89. Esses resultados indicam que a RNA, treinada com dados NIR, pode prever com alta precisão os parâmetros Brix, Fibra e POL. Além disso, o modelo foi validado com sucesso ao ser comparado com o método convencional de análise, evidenciando que essa abordagem é uma alternativa viável para substituir ou complementar os métodos tradicionais, oferecendo maior rapidez e eficiência.

Os valores de R^2 e RMSE, devem se aproximar o máximo de 1 e 0, respectivamente, o que foi possível observar nos resultados obtidos, apresentando ainda assim um desvio considerável, que pode justificado, por erros operacionais, durante as determinações e catalogações dos dados e, o fato da MLP desconsidera fatores climáticos que variam durante todo período de safra e atingem diretamente a composição da cana e, as variações quanto as espécies de cana analisadas.

Dessa forma podemos considerar, com base nos valores obtidos que existe um alto grau de correlação entre os conjuntos de dados obtidos pelo método Consecana e NIR, comprovando assim, a viabilidade técnica do método de espectroscopia de infravermelho próximo frente ao tradicional método Consecana, para caracterização da composição da cana de açúcar.

6. VIABILIDADE ECONÔMICA DE SUBSTITUIÇÃO

A comparação dos procedimentos levou em consideração todas as etapas necessárias e seus respectivos tempos para a obtenção dos resultados, para o método convencional consecana e método NIR. A Tabela 2 apresenta os dados de tempo obtidos.

Tabela 2: Comparativo de tempo entre método convencional e método NIR

Método Convencional		Método NIR (Cana)	
Denominações	Custos	Denominações	Custos
Aquisição Equipamentos	R\$ 187.490,00	Aquisição equipamento NIR e Reengenharia de Sonda	R\$ 950.000,00
Mão de Obra	R\$ 202.424,40	Mão de Obra	R\$ 13.494,96
Materiais de Laboratório	R\$ 12.530,00	Materiais de Laboratório	R\$ -
Manutenções	R\$ 16.960,00	Manutenções	R\$ 50.000,00
Custo 1ª Safra	R\$ 402.444,40	Custo 1ª Safra	R\$ 963.494,96
Custo Por Safra	R\$ 231.914,40	Custo Por Safra	R\$ 63.494,96

Fonte: Autor,2024.

Como é possível observar na Tabela 2, o método analítico convencional exige um tempo superior em comparação com os dois métodos possíveis NIR, levando o dobro de tempo do método NIR para caldo de cana e quase 8 vezes quando comparado ao NIR de cana, gerando um ganho operacional, em relação ao método convencional.

Na Tabela 3 podemos observar os custos de aquisição, operação e manutenção, sendo os relativos ao método convencionais com base no custo da usina Serra Grande S/A, mesmo base de custo foi usada para o método NIR, fazendo referência a primeira safra após serem implantados e para a safra consecutiva.

Tabela 3: Custos de implantação, operação e manutenção.

Método Convencional		Método NIR (Cana)	
Etapas do Processo	Tempo (min)	Etapas do Processo	Tempo (min)
Amostragem Sonda Obliqua	5	Amostragem Sonda Oblíqua	3
Desfibragem da Amostra (10kg)	3	Desfibragem da Amostra (10kg)	1
Pesagem de 500g da amostra desfibrada	3	NIR	1
Prensagem da amostra	5		
Peso do Bolo Úmido do bagaço	3		
Clarificação do Caldo	3		
Determinação de °Brix	5		
Determinação de POL	10		
Total	37	Total	5

Fonte: Autor,2024.

O método convencional, além de exigir maior demanda em relação ao tempo, apresenta um custo superior após a implantação dos métodos. É visível que a mão de obra é onde se observa uma maior diferença entre os outros métodos, sendo o NIR de Cana o método que apresenta menos custos, após implantado, apresentando uma viabilidade econômica, reduzindo o custo por safra em até 72,5%, em relação ao método convencional Consecana.

Levando em consideração as informações dispostas na Tabela 3, pode-se admitir que levando em consideração custo por safra de ambos os métodos, a implantação do método NIR, seria paga a partir da sexta safra, gerando uma economia de mais de 160 mil reais por safra.

7. CONCLUSÃO

A análise estatística, incluindo o teste t pareado e o uso de *Machine Learning*, demonstrou que os resultados obtidos pelo método NIR são equivalentes ao método tradicional, validando assim a sua precisão e confiabilidade. O treinamento e o teste com dados cruzados feitos pela RNA mostraram que o método NIR apresenta um alto nível de correção quando tem seus resultados comparados, aos resultados obtidos através do método tradicional Consecana.

Devido a simplicidade operacional, o método NIR mostrou-se mais eficiente na obtenção de resultados para as unidades industriais, por reduzir várias etapas no decorrer do processo de caracterização da matéria prima. Com base na avaliação operacional foi possível observar a redução de tempo para obtenção dos resultados, o que possibilita dessa forma aumentar o número de caminhões amostrados, aliado a isso foi possível observar a redução de custos de mais de 70% em relação ao método tradicional, após a sua implantação diminuindo em mais de 80% os custos com mão de obra e reagente.

Por tanto com base nesses fatores, o estudo mostra que a implementação do método NIR é recomendada, visto que além de garantir resultados tecnicamente precisos para a determinação de ART, apresenta também vantagens econômicas e de operação que justificam sua adoção no Setor Sucroenergético.

REFEERÊNCIAS

- ALFASSI, Z. B.; BOGER, Z.; RONEN, Y. **Statistical Treatment of Analytical Data**. Oxford: Blackwell Publishing, 2005.
- CALDAS, C. **Cálculos Fundamentais Para controle Químico Nas Indústrias de Açúcar e de Álcool**. Capivari, 2017.
- CALDAS, C. **Novo manual para laboratórios sucroalcooleiros**. Piracicaba, 2012.
- CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. **ACM computing surveys (CSUR)**, ACM, v. 41, n. 3, p. 15, 2009.
- CHEQUIN, Bruno Giovani; GRANDI, Guilherme. **O setor sucroalcooleiro brasileiro: origem e desenvolvimento**. Disponível em: . Acesso em: 30 de ago. 2024
- COHEN, Jacob. **Statistical power analysis for the behavioral sciences**. routledge, 2013.
- CONSELHO DOS PRODUTORES DE CANA-DE-AÇÚCAR, AÇÚCAR E ÁLCOOL DO ESTADO DE SÃO PAULO. *Manual do Consecana*. São Paulo: Consecana-SP, 2021. Disponível em: https://edisciplinas.usp.br/pluginfile.php/3116035/mod_resource/content/1/Manual%20do%2000Consecana.pdf. Acesso em: 6 set. 2024.
- CTC, 2015; Disponível em: <http://ctc.com.br/>. Acesso em: 30 de ago. de 2024.
- EMBRAPA. **Visão de Futuro: Trajetória do Agro: Desempenho Recente do Agro – Cana-de-Açúcar**. Campinas, 2020. Disponível em: < <https://www.embrapa.br/visao-de-futuro/trajetoria-do-agro/desempenho-recente-do-agro/cana-de-acucar> >. Acesso em: 01 set. 2024
- ESTRELLA, Bruna Castello Longo. Estudo de casos da aplicabilidade do teste T-pareado e da regressão linear pelo método dos mínimos quadrados: uma abordagem estatística aplicada à Química Analítica. 2023.
- FREITAS, Igor Wescley Silva de. Um estudo comparativo de técnicas de detecção de outliers no contexto de classificação de dados. 2019.

GÉRON, A. **Hands-on Machine Learning with Scikit-Learn , Keras & TensorFlow**. [s.l: s.n.].

HUGOT, Emile. **Handbook of cane sugar engineering**. Elsevier, 2014.

IAA PLANALSUCAR – **Base de dados da pesquisa agropecuária**. Disponível em: [http://www.bdpa.cnptia.embrapa.br/consulta/busca?b=ad&biblioteca=vazio&busca=autoria:"PLANALSUCAR"](http://www.bdpa.cnptia.embrapa.br/consulta/busca?b=ad&biblioteca=vazio&busca=autoria:). Acesso em 05 de set. de 2024.

ICUMSA - International Commission for Uniform Methods of Sugar Analysis. *ICUMSA Methods Book: Methods of Sugar Analysis*. London: ICUMSA, 2022.

INFOESCOLA. Espectro Eletromagnético. Disponível em: <https://www.infoescola.com/fisica/espectro-eletromagnetico/> Acesso em: 29 ago de 2024.

James G, Witten D, Hastie T, Tibishirani R. **An introduction to statistical learning with applications in R**. Amsterdam: Springer; 2013.

MENEGUETTI, C.C, MEZAROBIA, S. GROFF, A. M. Fatores relacionados ao cultivo da cana-de-açúcar. Campo Mourão – PR. 2010.

MISHRA, P.; SINGH, U.; PANDEY, C. M.; MISHRA, P.; PANDEY, G. Application of Student's t-test, Analysis of Variance and Covariance. **Annals of Cardiac Anaesthesia**. EUA, 2019.

NAANDANJAIN. Cana de açúcar. **Naandanjain**. Disponível em: <https://naandanjain.com.br/culturas/cana-de-acucar/>. Acessado em: 30 de ago. 2024.

OLIVEIRA, Ingrid Komorizono. **Aplicação de quimiometria e espectroscopia no infravermelho no controle de qualidade de biodiesel e mistura biodiesel/diesel**. 2008. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2008.

PAVIA, Donald L.; LAMPMAN, George M.; KRIZ, Gary S. **Introduction to Spectroscopy**. 3rd ed. Belmont: Thomson Learning, 2001.

PEREIRA, Beatriz Benassi. **Estudo de viabilidade da espectroscopia de infravermelho próximo para pagamento de cana por teor de sacarose**. 2021. Trabalho de Conclusão de

Curso (Graduação em Química Tecnológica) – Universidade Federal de São Carlos, São Carlos, 2021.

RODRIGUES, Clauber Damas. **Características e Composição Tecnológica da cana-de-açúcar**. Disponível em: <https://pt.slideshare.net/jaojaojaojao/caracteristicas-dacananadeacucar/> Acesso em: 30, ago. de 2024.

RODRIGUES, Gelze Serrat de Souza Campos; ROSS, Jurandyr Luciano Sanches. **A trajetória da cana-de-açúcar no Brasil: perspectivas geográfica, histórica e ambiental**. Edufu, 2020.

Russel S, Norvig P. **Inteligência artificial**. Rio de Janeiro: Elsevier; 2013.

SINDICATO DA INDÚSTRIA DO AÇÚCAR E DO ÁLCOOL NO ESTADO DE ALAGOAS. Informativo, 26 abr. 2024. Disponível em: <https://www.sindacucar-al.com.br/app/uploads/2024/04/260424.01.pdf>. Acesso em: 04 set. 2024.

SOUZA, Francisco AA; ARAÚJO, Rui; MENDES, Jérôme. Review of soft sensor methods for regression applications. **Chemometrics and Intelligent Laboratory Systems**, v. 152, p. 69-79, 2016.