

Previsão do índice de Açúcar Total Recuperável (ATR) de Cana-de-Açúcar Usando Redes Neurais Artificiais

Luciano Henrique Alves de Siqueira¹, Anderson Rogério Faia Pinto¹,
Jorge Alberto Achcar¹, Antonio Fernando Crepaldi², José Luís Garcia Hermosilla¹,
Marcelo Seido Nagano³

luciano.has@gmail.com; faiaphd@gmail.com; achcar@fmrp.usp.br;
crepaldi@feb.unesp.br; jlghermosilla@hotmail.com; drnagano@usp.br

¹ Departamento de Engenharia de Produção, Universidade de Araraquara, Av. Carlos Gomes, 1217 – Centro, 14801-340, Araraquara - SP, Brasil.

² Departamento de Engenharia de Produção, Universidade Estadual Paulista, Av. Eng. Luiz Edmundo C. Coube 14-01 – Vargem Limpa, 17033-360, Bauru - SP, Brasil.

³ Departamento de Engenharia de Produção, Universidade de São Paulo, Av. Trabalhador São-carlense, 400 – Centro, 13566-590, São Carlos - SP, Brasil.

DOI: 10.17013/risti.57.85–107

Resumo: A colheita da cana-de-açúcar impacta significativamente a produtividade das usinas sucroalcooleiras e deve ser realizada no ponto ótimo de maturação ou Período de Utilização Industrial (PUI). A estimativa do PUI é obtida por análises laboratoriais que mensuram o índice de Açúcar Total Recuperável (ATR). A questão é que o planejamento das usinas sucroalcooleiras é dependente da previsão de produção (açúcar e etanol), e não há ferramentas capazes de substituir os altos custos das análises laboratoriais do ATR. Logo, este artigo traz um estudo de caso cujo objetivo é propor um modelo computacional para estimar o índice de ATR da produção de uma usina sucroalcooleira do interior do estado de São Paulo. Este modelo é fundamentado nas Artificial Neural Networks (ANN) e foi aplicado às bases de dados das lavouras de 48.151 talhões das safras de 2016/2017 a 2022/2023. As experimentações apresentaram um erro médio absoluto (3,49%) muito inferior às análises laboratoriais (12,17%). A principal contribuição deste estudo está em fornecer um modelo de baixo custo que estima com acurácia muito satisfatória o índice de ATR para o PUI.

Palavras-chave: Usinas; Predição; Produtividade; Inteligência Artificial.

Forecast of the Total Recoverable Sugar (TRS) Index of Sugarcane Using Artificial Neural Networks

Abstract: Sugarcane harvesting significantly impacts the productivity of sugar mills and should be carried out at the optimal maturity point or Period of Industrial Utilization (PIU). The PIU is estimated through laboratory analyses that measure

the Total Recoverable Sugar (TRS) index. The challenge lies in the fact that the planning of sugar mills depends on production forecasts (sugar and ethanol), and there are no tools capable of replacing the high costs of laboratory TRS analyses. Therefore, this article presents a case study aimed at proposing a computational model to estimate the TRS index of the production of a sugar mill in the countryside of São Paulo, Brazil. This model is based Artificial Neural Networks (ANN) and was applied to databases of 48,151 plots from the 2016/2017 to 2022/2023 harvests. The experiments presented a mean absolute error (3.49%) much lower than laboratory analyses (12.17%). The main contribution of this study lies in providing a low-cost model that estimates the ATR index with very high accuracy of the PIU.

Keywords: Sugar Mills; Prediction; Productivity; Artificial Intelligence.

1. Introdução

A cana-de-açúcar é nativa da Nova Guiné e foi introduzida na África e na Europa pelos árabes (Denham, 2024). No Brasil, maior produtor mundial, a indústria sucroalcooleira tem impulsionado a agricultura e viabilizado a matriz energética por meio da biomassa da cana-de-açúcar (Shanthi et al., 2023). Essas indústrias atendem aos mercados nacional e internacional e contribuem significativamente para o Produto Interno Bruto (PIB). A estimativa brasileira referente à safra 2024/25 é de 685,86 milhões de toneladas (Mt) para uma área de 8.673,1 mil hectares (ha) e produtividade média de 79.079 kg/ha (Silva et al., 2023, CONAB, 2024). Um ciclo de produção abrange desde o plantio dos lotes (matéria-prima) até o manejo, colheita e transformação em açúcar e/ou etanol (Florentino et al., 2018; Shanthi et al., 2023). A produção é, portanto, determinada por um conjunto de fatores (climáticos, ambientais e agrônômicos) que envolvem desde a escolha da variedade e o manejo do canavial (idade, número de cortes, fertilização) até o controle de pragas, doenças e ervas daninhas (Monteiro & Sentelhas, 2014; Monteiro & Sentelhas, 2017; Hammer, et al., 2020). A máxima eficiência fotossintética da cana-de-açúcar (metabolismo C₄) é alcançada em condições de alta radiação solar, solo úmido e temperaturas atmosféricas entre 25°C e 35°C (Shanthi et al., 2023; Ali et al., 2024). A otimização da produção envolve a colheita no momento ideal de maturação ou Período de Utilização Industrial (PUI). Essa maturação ocorre em condições de escassez hídrica prolongada, cuja temperatura média mais adequada para o amadurecimento é de 12–14°C (Dias Neto et al., 2022; Shanthi et al., 2023). No PUI, o teor de açúcar nos colmos está entre 12,257% e 16% (Shanthi et al., 2023).

As análises de maturação exigem o gerenciamento de um grande volume de dados e são realizadas por meio da coleta manual de pequenas amostras de cana de cada lote que são transportadas ao laboratório para a dosagem do PUI (Florentino et al., 2018; Souza et al., 2022; Shanthi et al., 2023). Essas medições são comumente realizadas por espectrometria e o PUI é obtido pelo grau Brix (°Bx)¹. Assim, o PUI presente nas amostras reflete a qualidade de cada lote de cana e resulta em um índice definido como Açúcar Total Recuperável (ATR). A maturação também pode ser induzida artificialmente com a aplicação de maturadores químicos (Shanthi et al., 2023). Esses maturadores,

¹ Indicador de sólidos solúveis no caldo considerando pureza, teor de sacarose, açúcares redutores totais e ATR.

ao modificar a morfologia e a fisiologia da planta, melhoram a qualidade da matéria-prima, otimizam os resultados agroindustriais e econômicos e facilitam o planejamento da colheita (Florentino et al., 2018; Dias Neto et al., 2022; Silva et al., 2023). A colheita é o componente com maior custo de produção (25%) e inclui desde processos tecnológicos até os fatores sociais, econômicos e fisiológicos (Dias Neto et al., 2022). Os períodos e o sequenciamento de colheita dos lotes podem variar dependendo da região e do valor agregado ao ATR (Florentino et al., 2018; Shanthi et al., 2023). Além disso, o mercado de açúcar e etanol é caracterizado por preços voláteis e a lucratividade requer altos níveis de eficiência operacional e ATR (Florentino et al., 2018; CNA, 2023). Tanto a gestão logística quanto a produtiva (cronogramas de colheita e operações de processamento), assim como custos, estratégia comercial e negociações de preços, também estão comumente associadas ao ATR (Consecana-SP 2023; Kathirvel & Venkatachalam, 2025). É ante este contexto que os gestores têm de lidar com dois problemas críticos: i) prever a receita por tonelada mediante a estimativa do ATR (Consecana-SP, 2023), ii) prever a produção para planejar as etapas do ciclo de produção (Souza et al., 2022).

A questão é que os modelos atualmente disponíveis para estimar a produtividade e o índice ATR da cana-de-açúcar não apresentam precisão satisfatória (De Oliveira et al., 2017; Canata et al., 2024; Kathirvel & Venkatachalam, 2025). Há ainda incertezas quanto à ordem hierárquica e ao impacto das variáveis climáticas e de manejo (Everingham et al., 2016; Hammer et al., 2020). A integração desses fatores é geralmente realizada por meio de modelos mecanísticos e modelos empíricos baseados em Big Data (Dias & Inman-Bamber, 2020; Hammer et al., 2020; Kathirvel & Venkatachalam, 2025). Os modelos mecanísticos mais utilizados são o Agro-Ecological Zone Model (AZM-FAO), o Decision Support System for Agrotechnology Transfer (DSSAT/CANEGRO) e o Agricultural Production Systems SIMulator – Sugarcane (APSIM-Sugar). Já os modelos empíricos focam em Data Mining (DM), Machine Learning (ML) e Artificial Intelligence (AI). Exemplos são as técnicas Random Forest (RF), Regression Trees (RT), Multiple Linear Regression (MLR), Artificial Neural Network (ANN), Gradient Boosting Tree (GBT), Support Vector Machine (SVM), Boosted Regression Trees (BRTs), Support Vector Regression (SVR) e Gradient Boosting Machine (GBM). Entretanto, esses modelos possuem algumas limitações: i) Modelos Mecanicistas – enfrentam desafios relacionados à senescência e à parametrização de novas cultivares; ii) Modelos Empíricos – são complexos e dependentes das variáveis de entrada e de grandes quantidades de dados de alta qualidade que podem ter desempenho impactado pela linearidade, multicolinearidade e a Spatial Autocorrelation (SA).

Fato é que muitos modelos não são adequados para dados (não normais) com características autocorrelacionadas e que apresentam condições heterogêneas na cana-de-açúcar (De Oliveira et al., 2017; Canata et al., 2024). A maioria dos bancos de dados usados também são de tamanho e diversidade insatisfatórios (Hammer et al., 2020; Cardoso et al., 2025). Além disso, a falta de dados para benchmarking e a ausência de métricas de avaliação limitam a comparação entre os modelos (Ferracioli et al., 2019; Cardoso et al., 2025; Kathirvel & Venkatachalam, 2025;). Estas questões fazem as indústrias sucroalcooleiras ainda optarem por análises laboratoriais caracterizadas como trabalhosas, demoradas e limitadas em escalabilidade (Phuphaphud et al., 2019; Canata et al., 2024; Kathirvel & Venkatachalam, 2025). Tais análises requerem

tempo e geram altos custos relacionados ao uso de uma série de recursos (humanos, logísticos, laboratoriais, etc.). Há, portanto, uma grande demanda por modelos quantitativos capazes de reduzir o tempo e os custos associados à estimativa ou identificação dos fatores que influenciam a variabilidade da produtividade e do índice ATR. É indiscutível que avanços nos modelos de predição propostos pela literatura são imprescindíveis e altamente relevantes para a indústria sucroalcooleira do Brasil. À vista disso, este manuscrito apresenta um estudo de caso cujo objetivo é propor um modelo computacional para estimar o índice de ATR na produção de cana-de-açúcar de uma usina sucroalcooleira do interior do estado de São Paulo. A formulação do modelo é baseada nas ANN e utiliza a linguagem de programação Python® junto com a rede Feed-Forward. Esse modelo é aplicado a um grande banco de dados reais referente as safras de 2016/2017 a 2022/2023. É importante ressaltar que o uso das ANN é apropriado por se provar eficaz para lidar com as complexidades inerentes às predições utilizando grandes bases de dados históricos (Katal & Singh, 2022; Maier et al., 2023). Assim sendo, este estudo não tem como foco realizar análises comparativas entre os modelos propostos pela literatura para estimar o índice de ATR. Todo escopo está na proposição de um modelo ágil e de acurácia satisfatória na predição dos níveis de produtividade e mensuração do índice de ATR. A perspectiva é que uma tecnologia de fácil acesso e baixo custo possa melhorar o planejamento e a tomada de decisões relacionadas aos fatores de produção e financeiros em mais usinas de açúcar e etanol no Brasil.

2. Método de Pesquisa

Esta é uma pesquisa aplicada de natureza quantitativa e caracterizada como um estudo de caso (Matović & Ovesni, 2023). A revisão de literatura foi realizada por meio de análises bibliográficas fundamentadas em Lim et al. (2022). As buscas realizaram-se nas bases de dados Scopus, Google Scholar, IEEE Xplore, Web of Science e ScienceDirect. As palavras-chave usadas nos campos de buscas avançadas foram; “*sugarcane*”, “*maturation*”, “*forecasting*”, “*deep learning*”, “*machine learning*” e “*artificial intelligence*”. Nos conjuntos de palavras correlatas, como as da área de IA, foi utilizado, para encontrar resultados que contenham ao menos uma das palavras-chave, a expressão booleana “OR”. Além disso, para que fossem encontrados resultados que contenham a combinação de palavras-chave obrigatórias foi utilizado a expressão booleana “AND”. No processo de busca, o único critério de data estabelecido foi a limitação de artigos publicados até maio de 2025.

A abordagem de solução do problema em questão se fundamentou no arcabouço teórico, nos dados documentais da usina em estudo e na realidade prática de mensuração do ATR. Portanto, este estudo utilizou dados históricos reais contendo fatores meteorológicos e de produção referente a 48.151 talhões de diferentes regiões do estado de São Paulo referente às safras de 2016/2017 a 2022/2023. Também foi realizada uma análise preliminar (estatísticas descritivas) para a padronização dos dados e a seleção de atributos relevantes, baseada na literatura e em entrevistas com os especialistas da usina do interior de São Paulo. A análise preliminar foi obtida utilizando o software Minitab® e validou um conjunto independente de dados que detalharam a problemática, a função objetivo e a aplicação das ANN. Esses dados foram tratados por meio de ajustes de hiperparâmetros seguidos da modelagem matemática da ANN. A programação

computacional e as experimentações das ANN foram realizadas em linguagem Python® junto com a rede Feed-Forward utilizando o Google Colab (<https://colab.google/>). Todos os dados tratados (produção e meteorológicos) foram utilizados para treinar e ampliar a acurácia das estimativas das ANN. A eficácia das ANN foi avaliada empregando a métrica Mean Absolute Percentual Error (MAPE).

3. Revisão de Literatura

A literatura tem proposto uma série de modelos para prever a produtividade da cana-de-açúcar e estimar o ATR. Madeiro et al. (2006) estimaram indicadores de produtividade e maturação utilizando ANN parametrizadas por Genetic Algorithm (GA). Everingham et al. (2009) evidenciaram que as análises em torno do crescimento e da produção da cana-de-açúcar são limitadas quando utilizado o APSIM-Sugar. Bocca & Rodrigues (2016) propuseram técnicas de DM (RF, RT, BRT, ANN, SVM e Feature Selection (FS) com o RReliefF) para calcular a produtividade da cana-de-açúcar na Usina Alcídia, Odebrecht Agroindustrial (Teodoro Sampaio), São Paulo, Brasil. Everingham et al. (2016) aplicaram o Random Forest Classifier (RFC) e o Random Forest Regressor (RFR) para explicar a variação anual na produtividade da cana-de-açúcar em Tully, Austrália. Dias & Sentelhas (2017) avaliaram AZM-FAO, DSSAT/CANEGRO e APSIM-Sugar para estimar a produtividade da cana-de-açúcar em usinas na região Nordeste do Brasil. Oliveira et al. (2017) identificam as variáveis associadas ao ATR com uma precisão de 90% usando DM (RF, RT e SVR com RReliefF) e dados da mesma planta estudada por Bocca & Rodrigues (2016).

Por conseguinte, Charoen-Ung & Mittrapiyanuruk (2018) propuseram a RF (71,83%) e GBT (71,64%) para estimar com boa precisão o rendimento de um conjunto de lotes de cana-de-açúcar na Tailândia. Charoen-Ung & Mittrapiyanuruk (2019) também aplicaram RF para prever a produtividade de lotes de cana-de-açúcar em uma usina na Tailândia. Coelho et al. (2019) apresentaram ANN para estimar a Pol do caldo de cana-de-açúcar e obtiveram alta precisão ao utilizar °Brix e Wet Cake Weight (WBW). Dias et al. (2019) investigaram o impacto de ajustes ao APSIM-Sugar para estimar a produtividade sob condições específicas de campo no estado do Piauí (Guadalupe), Brasil. Ferracioli et al. (2019) aplicaram ML (RF, BRT e SVR com RReliefF e a avaliação do impacto da SA) para prever a produtividade de quatro áreas de usinas da Odebrecht Agroindustrial no Brasil. Medar & Rajpurohit (2019) utilizaram algoritmos de Long Term Time Series (LTTS), Weather-And-Soil Attributes (WSA), Supervised Machine Learning (SML) e o Normalized Vegetation Index (NDVI), com precisão de 86,31%, para a previsão da produtividade da cana-de-açúcar na região de Karnataka, Índia. Dias et al. (2020) avaliaram as condições de crescimento, desenvolvimento e produtividade de 27 variedades brasileiras de cana-de-açúcar para melhorar a compreensão das interações Genotype × Environment × Management (G×E×M) por meio da aplicação do APSIM-Sugar em São Romão (Minas Gerais) e Guadalupe (Piauí), Brasil. Hammer et al. (2020) classificaram as principais variáveis de produtividade aplicando DM (RF, GBM e SVM) em bancos de dados de várias usinas no estado de São Paulo, Brasil. Oliveira et al. (2020) avaliaram o desempenho de modelos de ML (SVR com radial kernel e RF), bem como DSSAT/CANEGRO, para estimar o ATR usando dados de colheita de três anos fornecidos pela Usina Santa Fé (Nova Europa), São Paulo, Brasil. Shendryk et al. (2020)

utilizaram imagens de satélite e outros dados para estimar (precisão de 91,2%) que o número ideal de dias para a melhor produtividade é de 180. Rahman & Robson (2020) empregaram imagens de satélite e atributos de plantação para estimar a produção e apresentaram uma forte correlação com o Green Normalized Difference Vegetation Index (GNDVI).

Mais recentemente, Canata et al. (2021) mapearam a produtividade da cana-de-açúcar usando RF e MLR integrados com imagens orbitais de séries temporais e ML em um sítio comercial localizado em Botucatu, estado de São Paulo, Brasil. Dias et al. (2021) aplicaram técnicas estatísticas (ANOVA e Time Series Regression Models) e APSIM-Sugar para entender como a interação $G \times E \times M$ influencia a produtividade em variedades de cana-de-açúcar elite no Brasil, usando dados da mesma planta estudada por Dias et al. (2020). Luciano et al. (2021) combinaram imagens de satélite, dados agrônômicos e meteorológicos com técnicas de ML e RF para estimar a produtividade e as variáveis mais importantes foram o número de colheitas e o Índice de Normalized Difference Moisture Index (NDMI). Um modelo de estimativa de produtividade mais preciso que incorpora dados meteorológicos, índices de vegetação e práticas agrícolas foi proposto por Pignède et al. (2021). Pravallika et al. (2021) discutiram como estimar o rendimento da produção em várias culturas agrícolas e como as mais precisas podem ser obtidas por modelos de ANN e Deep Neural Network (DNN). Nihar et al. (2022) usaram imagens de satélite como parâmetros para técnicas de ML para estimar o rendimento agrícola regional em Uttar Pradesh, Índia. Canata et al. (2024) integraram dados multiespectrais com RF, ANN e MLR para estimar o nível de ATR (Brix e Pureza) em dois campos de cana-de-açúcar perto de Jaboticabal, estado de São Paulo, Brasil. É demonstrado que o RF é superior para °Brix e o ANN é melhor para a Pureza. A variável de entrada mais importante no desempenho de RF foi Growing Degree Days (GDD). Colmanetti et al. (2024) propuseram o ECOSystem Model Simulator (ECOSMOS), considerando a interação $G \times E \times M$ para estimar o rendimento de plantações de cana-de-açúcar em diferentes ambientes no Brasil. Amaro et al. (2025) usaram imagens Sentinel-2A e outros dados (climáticos, agrônômicos e de déficit hídrico) para propor um modelo baseado em RF utilizado para prever a produtividade da cana-de-açúcar em três usinas no estado de São Paulo, Brasil. Satpathi et al. (2025) aplicaram RF, MLR, SVR, ANN, Extreme Gradient Boosting (XGB), Elastic Net (ELNET), Ridge Regression (Ridge) e Least Absolute Shrinkage Selection Operator (LASSO) em onze grandes distritos produtores de cana-de-açúcar de Uttar Pradesh (Bahraich, Bareilly, Basti, Bijnor, Deoria, Gonda, Gorakhpur, Lakhimpur, Pilibhit, Rampur e Saharanpur).

4. Formulação do Modelo Proposto

4.1. Problema de pesquisa

A empresa em estudo está localizada no centro-sul e apresenta um aumento da área de produção de 47.660ha na safra 2012/2013 para 53.593ha em 2022/2023. A produção total corresponde a 28,65 Mt para uma área de 359.729ha até 2022/2023. A empresa em questão possui dois grupos de lotes: i) Cana Própria (CP) - todo o ciclo de produção é administrado pela empresa, ii) Cana de Terceiros (CT) - ciclo de produção realizado por fornecedores de matéria-prima, exceto a colheita e transformação (açúcar ou etanol).

Lotes da CT corresponderam a uma produção de 12,25 Mt em 148.873ha (40,6%). Além disso, é impossível recuperar o histórico de práticas operacionais e agronômicas empregadas nos lotes da CT. Esta questão limita a predição dos fatores e interfere no resultado do ATR. A colheita dos lotes ocorre entre os meses de abril e dezembro, seguindo os níveis do ATR. Logo, os roteiros de colheita são precedidos por uma coleta manual de pequenas amostras de cana de cada lote que são levadas ao laboratório para análise do ATR. Todos os parâmetros e amostragem para a determinação do ATR são regulamentados pelo Conselho dos Produtores de Cana de Açúcar, Açúcar e Etanol do Estado de São Paulo (CONSECANA-SP). A área produtiva atual para a mensuração do índice de ATR é ilustrada pela Figura 1.

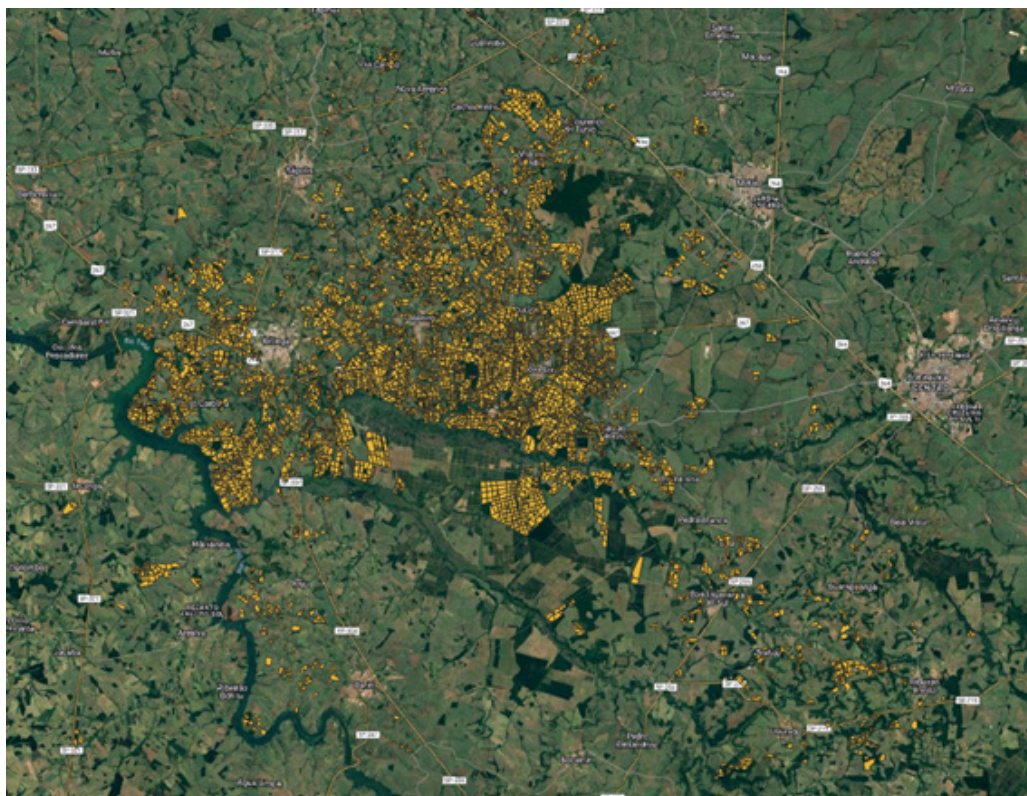


Figura 1 – Área produtiva atual para a mensuração do índice de ATR (Fonte: Empresa)

A Figura 1 expõe uma imagem de satélite cuja cor amarela é a área produtiva para estimar o índice do ATR. As possíveis igualdades ou diferenças entre as médias nos diferentes níveis de cada fator categórico são associadas com a resposta produtividade por lote ou ATR. Este estudo considera vários fatores (covariáveis) que podem estar associados com as respostas de interesse (produtividade e índice ATR): i) Produção e qualidade (produção realizada, pol, pureza, Brix e ATR), ii) Manejo operacional (tipo de

insumo, data de aplicação, quantidade aplicada e modalidade de aplicação), iii) Colheita realizada (data da colheita, duração do ciclo, tipo de colheita realizada, número do corte do lote), iv) Fatores Meteorológicos (precipitação diária, temperaturas máxima, média e mínima, coletados por estação meteorológica da própria empresa), v) Identificação e caracterização dos lotes (código do lote, data de plantio, código da fazenda, variedade plantada, município da fazenda). A base de dados com seus respectivos campos e o que representam é apresentada pela Tabela 1.

Variáveis	Descrição	Base
Operações Após a Colheita (OPC)		B1
OPC_CPD	Controle de Plantas Daninhas (CPD)	
OPC_CPR	Controle de Pragas e Doenças (CPR)	
OPC_AMI	Aplicação de Adubos Minerais (AMI)	
OPC_CSO	Aplicação de Corretivos de Solo (CSO)	
OPC_AVL	Aplicação de Vinhaça Localizada (AVL)	
OPC_ACO	Aplicação de Composto Orgânico (ACO)	
OPC_AVA	Aplicação de Vinhaça por Aspersão (AVA)	
OPC_AAE	Aplicação Aérea de Fertilizantes Foliares (AAE)	
Covariáveis Categóricas Estatísticas (CCE)		B1 e B2
REGIAO	Região onde o lote de cana-de-açúcar foi plantado em CP e CT	
DIAS_CICLO	Dias entre a data de colheita anterior ou plantio e a data da atual colheita	
NRO_CORTE	Número que descreve o estágio atual do lote (quantas vezes já foi colhido)	
TIPO_CICLO	Categoria do talhão (soca, plantada, nunca colhida, colhida anteriormente)	
TIPO_CORTE	Maneira da colheita (manual, mecanizada e com o uso ou não de queimadas)	
COL_ATU_MES	Mês em que ocorreu a colheita no ano atual ou a data do plantio no ano atual	
COL_ANT_MES	Mês em que ocorreu a colheita ou a data do plantio no ano anterior ao respectivo	
Resultado do Balanço Hídrico (RBH)		B1 e B2
BH_Do até BH_D17	Talhão no mesmo decêndio até os 17 decêndios anteriores ao da respectiva colheita	

Tabela 1 – Variáveis associadas a produção de cana-de-açúcar (Fonte: Empresa)

A Tabela 1 mostra que os dados foram separados em duas bases denominadas B1 e B2 e a variável resposta é a previsão do ATR. As variáveis que consideram apenas B1 são referentes às Operações Após a Colheita (OPC). Isso porque B1 contém apenas os registros dos lotes de CP. Em B2, os campos com os dados de manejo foram desprezados porque não é possível determiná-los nos lotes de CT. A base do RBH totaliza 6 meses de dados meteorológicos (BH_Do, BH_D1, BH_D2, BH_D3, BH_D4, BH_D5, BH_D6, BH_D7, BH_D8, BH_D9, BH_D10, BH_D11, BH_D12, BH_D13, BH_D14, BH_D15, BH_D16, BH_D17). Assim, cada campo contém o RBH do talhão no mesmo decêndio (BH_Do) ao decêndio anterior (BH_D1 = D - 1) e sucessivamente até 17 decêndios anteriores ao da respectiva colheita (BH_D17 = D - 17). Todos os dados meteorológicos

foram resumidos com cálculo do RBH por meio do Método de Thornthwaite (Werlang & Werlang, 2021). Algumas variáveis não relevantes foram removidas, enquanto outras foram adicionadas com base na análise da literatura e em relatos do corpo técnico da usina de SP. Após a padronização dos dados, os registros que continham alguma variável com valores nulos foram removidos e também desconsiderados: i) Lotes sem resultados de análise do ATR, ii) Lotes com dias de ciclo fora do período padrão do ciclo de colheita (menor que 180 e maior que 730), iii) Lotes classificados como bisados (não colhidos na safra anterior devido à baixa representatividade de exemplos), iv) Registros nos quais a variável “Aplicado Maturador” é “SIM” (insumos que induzem respostas do índice ATR). Logo, devido à indisponibilidade de informação de operações de manejo empregadas, foram considerados os atributos relevantes em B1 e B2.

4.2.Análise preliminar descritiva

Análise preliminar descritiva demonstra que a média e o desvio-padrão amostral do ATR são dados respectivamente por 132,49 e 18,76 (48.151 talhões). A análise também mostra um valor mínimo (0,100) e um valor máximo (212,0), que indica uma grande variabilidade dos dados (Minitab®). A Figura 2 ilustra o histograma dos dados de cada medida de ATR, enquanto os resultados das covariáveis são apresentados na Tabela 1. Uma análise descritiva preliminar das médias amostrais de ATR em cada nível das covariáveis categóricas regiões, tipo de corte e ciclo de plantio é apresentada nas Tabelas 2 e 3.

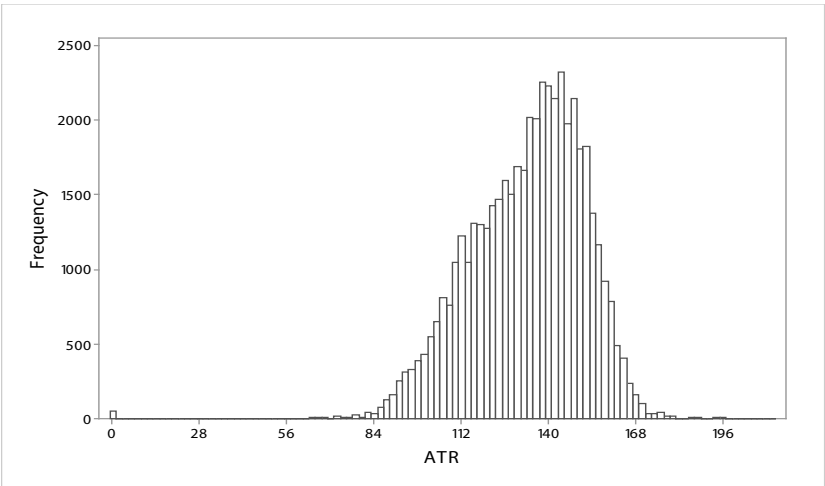


Figura 2 – Histograma das medidas de ATR (Fonte: Elaboração no Minitab®)

Região	N	Média	DP	Mínimo	Mediana	Máximo
AREALVA	169	136,980	10,590	114,000	138,000	168,000
BARIRI	325	142,050	10,970	63,000	144,000	166,000
BOA ESPERANCA DO SUL	2.822	137,140	15,960	0,100	139,000	187,000

Região	N	Média	DP	Mínimo	Mediana	Máximo
BORBOREMA	23	143,350	17,000	99,000	145,000	173,000
DOBRADA	7	149,000	6,860	142,000	146,000	158,000
DOURADO	153	150,230	11,140	125,000	152,000	167,000
GAVIAO PEIXOTO	2.963	133,940	16,860	84,000	136,000	179,000
IACANGA	7	139,710	12,180	122,000	147,000	148,000
IBATE	6	138,330	23,850	92,000	147,000	154,000
IBITINGA	12.908	133,140	17,890	0,100	135,000	201,000
ITAJU	291	138,600	11,620	103,000	138,000	164,000
ITAPOLIS	3.335	131,820	17,720	0,100	135,000	187,000
MATAO	2.883	132,330	17,360	39,000	134,000	177,000
NOVA EUROPA	7.421	127,750	20,370	0,100	130,000	184,000
RIBEIRAO BONITO	357	142,100	12,670	110,000	144,000	174,000
SANTA ERNESTINA	7	155,430	5,910	151,000	152,000	167,000
TABATINGA	13.035	131,560	19,120	0,100	134,000	212,000
TAQUARITINGA	1.071	142,230	12,200	58,000	144,000	166,000
TRABIJU	103	152,000	10,140	117,000	155,000	164,000
Tipo Corte	N	Média	DP	Mínimo	Mediana	Máximo
MANUAL CRUA	138	142,480	13,660	99,000	143,000	177,000
MANUAL QUEIMADA	1	144,000	*	144,000	144,000	144,000
MECANICA CRUA	47.628	132,300	18,350	0,100	135,000	201,000
MECANICA QUEIMADA	384	152,210	21,580	91,000	153,000	212,000
Ciclo Plantio	N	Média	DP	Mínimo	Mediana	Máximo
ANO	5.736	129,990	19,150	0,100	133,000	186,000
ANO E MEIO	30.159	131,520	18,350	0,100	134,000	212,000
INVERNO	12.256	136,050	17,900	0,100	139,000	198,000

Abreviações: N – Número de Observações, DP – Desvio-Padrão Amostral

Tabela 2 – Medidas descritivas para os índices de ATR (Fonte: Elaboração no Minitab®)

A Tabela 2 mostra que as maiores médias de ATR são observadas nas regiões de DOBRADA, DOURADO, SANTA ERNESTINA e TRABIJU. Em relação à covariável tipo de corte, a maior média de ATR é para a MECÂNICA QUEIMADA. Para a covariável CICLO_PLANTIO, é observado que o INVERNO apresenta maior média de ATR. Há também, entre as várias regiões de plantio, uma considerável variação no desvio-padrão de ATR. A Figura 3 demonstra os diagramas de dispersão para as médias de ATR considerando as covariáveis NRO_CORTE, COL_ATU_MES e COL_ANT_MES.

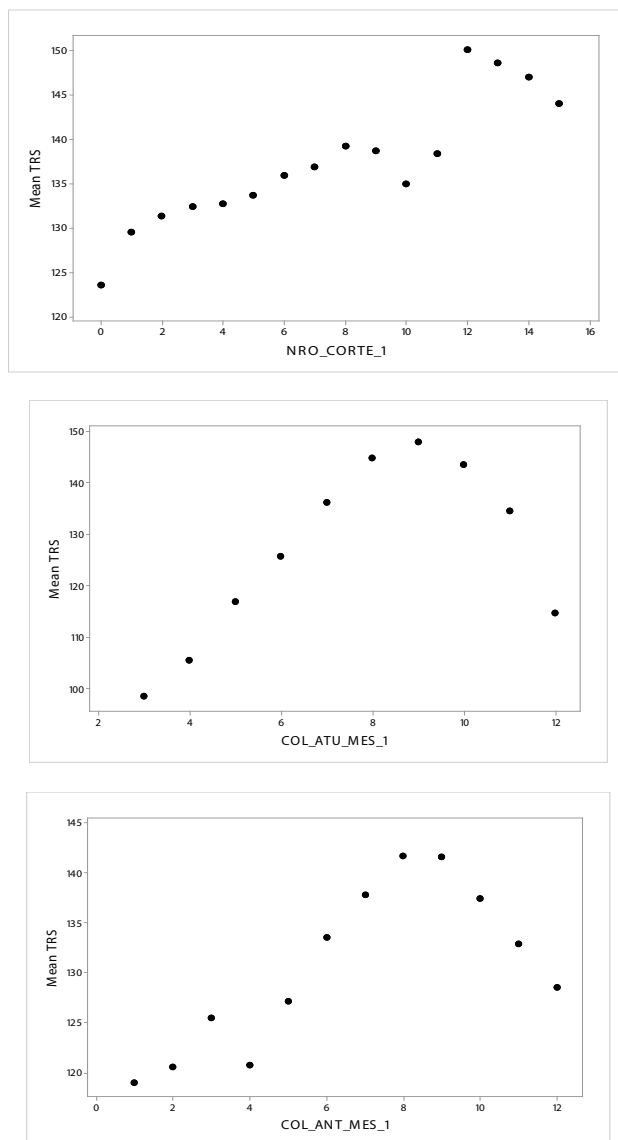


Figura 3 – Diagramas de dispersão para as médias de ATR (Fonte: Elaboração no Minitab®)

A maior média de ATR é encontrada em $NRO = 12$ considerando a covariável NRO_CORTE (número que descreve o estágio atual do lote, quantas vezes ele já foi colhido), em relação à covariável COL_ATU_MES (mês em que ocorreu a colheita no respectivo ano) a maior média de ATR é observada em 9 meses e em relação à covariável COL_ANT_MES 1 (mês em que ocorreu a colheita, ou data do plantio, no ano anterior ao respectivo) tem-se a maior média de ATR entre os meses 8 e 9. A Tabela 3 exibe as

medidas descritivas (médias amostrais, desvios-padrões amostrais, mediana amostral) para os índices de ATR nas covariáveis categóricas de OPC introduzidas na Tabela 1.

Covariáveis	Respostas	N	Média	StDev	Mínimo	Mediana	Máximo
OPC_CPD	0 (Não)	24.106	135,230	16,680	0,100	137,000	187,000
	1 (Sim)	24.045	129,740	19,710	0,100	132,000	212,000
OPC_AMI	0 (Não)	30.926	134,190	17,760	0,100	137,000	198,000
	1 (Sim)	17.225	129,430	19,290	0,100	131,000	212,000
OPC_ACO	0 (Não)	45.211	133,060	17,960	0,100	135,000	212,000
	1 (Sim)	2.940	123,750	23,270	0,100	128,000	198,000
OPC_AVL	0 (Não)	41.695	132,640	18,020	0,100	135,000	212,000
	1 (Sim)	6.456	131,480	21,060	0,100	136,000	198,000
OPC_AVA	0 (Não)	46.795	132,820	18,410	0,100	135,000	212,000
	1 (Sim)	1.356	120,910	16,140	0,100	122,000	157,000
OPC_CSO	0 (Não)	44.859	132,810	18,260	0,100	135,000	212,000
	1 (Sim)	3.292	128,140	20,490	0,100	130,000	212,000
OPC_CPR	0 (Não)	35.355	133,750	17,610	0,100	136,000	206,000
	1 (Sim)	12.796	129,000	20,220	0,100	132,000	212,000
OPC_AAE	0 (Não)	45.517	132,430	18,430	0,100	135,000	212,000
	1 (Sim)	2.634	133,410	18,910	0,100	136,000	195,000

Abreviações: N – Número de Observações

Tabela 3 – Medidas descritivas para os índices de ATR (Fonte: Elaboração no Minitab®)

Note que as maiores médias de ATR estão na resposta “(Não)” ou ausência de OPC, considerando as covariáveis OPC_CPD, OPC_AMI, OPC_ACO, OPC_AVL, OPC_AVA, OPC_CSO, OPC_CPR. Para a covariável OPC_AAE, a maior média de ATR é observada na resposta “1 (Sim)” ou presença de OPC. A Figura 4 ilustra os diagramas de dispersão para as médias de ATR considerando as covariáveis BH_Do até BH_D17. Na sequência, a Figura 5 exibe o diagrama de dispersão para as medidas de ATR considerando a covariável DIAS CICLO.

A Figura 4 mostra os diagramas de dispersão para as médias de ATR considerando as covariáveis BH_Do, BH_D1, BH_D2, BH_D3, BH_D4, BH_D5, BH_D6, BH_D7, BH_D8, BH_D9, BH_D10, BH_D11, BH_D12, BH_D13, BH_D14, BH_D15, BH_D16, BH_D17. Todos os gráficos mostram um decaimento da média de ATR com o aumento do nível de cada covariável (resultado do balanço hídrico do talhão no mesmo decêndio (BH_Do = D - 0), ao decêndio anterior (BH_D1 = D - 1) e sucessivamente até 17 decêndios anteriores ao da respectiva colheita). A Figura 5 mostra o diagrama de dispersão para as medidas de ATR considerando a covariável DIAS CICLO (quantidade de dias entre a data de colheita anterior, ou do plantio, e a data de colheita do respectivo ano). É possível observar que essa covariável aparentemente não afeta o índice de ATR nas plantas de cana-de-açúcar. Logo, a análise descritiva permitiu uma visão preliminar

dos efeitos de diferentes covariáveis na resposta de interesse ATR. A previsão de ATR baseada na informação de todas as covariáveis será realizada usando as ANN.

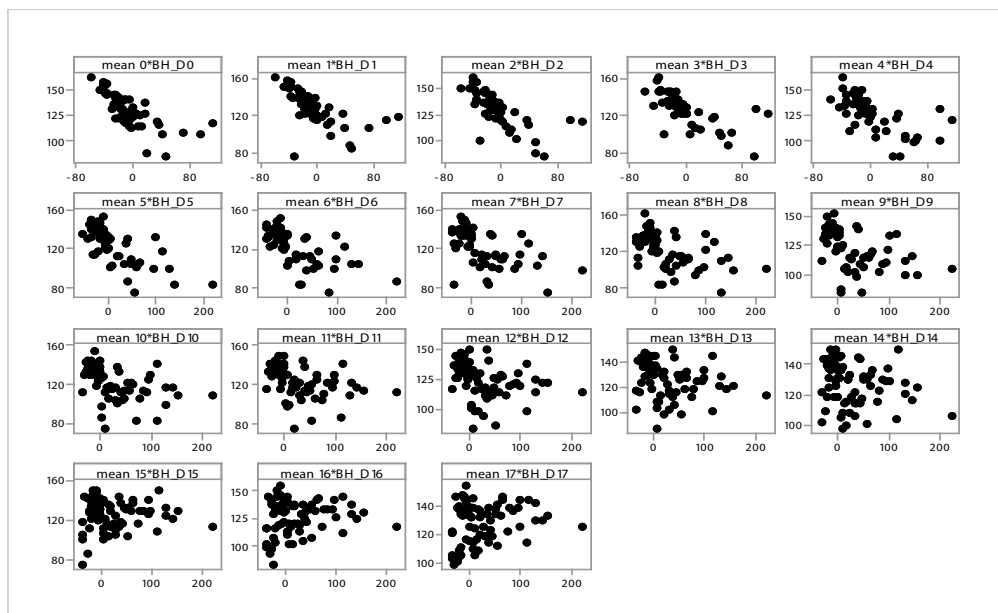


Figura 4 – Diagramas de dispersão para as médias de ATR (BH_Do até BH_D17) (Fonte: Elaboração no Minitab®)

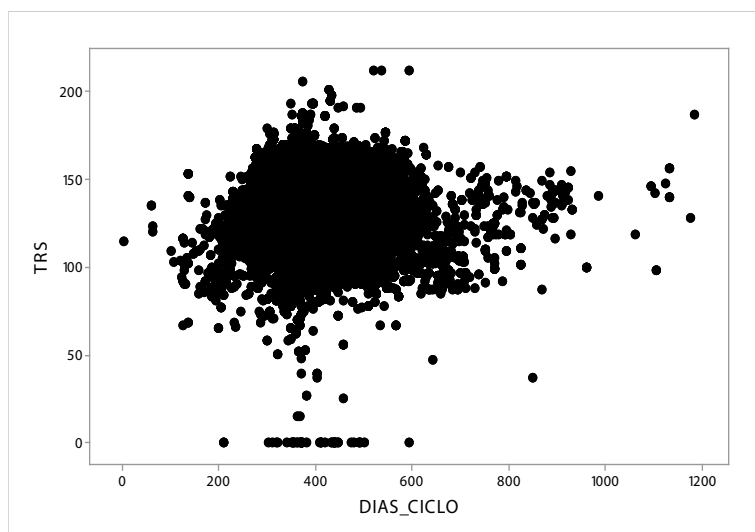


Figura 5 – Diagrama de dispersão para as medidas de ATR (DIAS CICLO) (Fonte: Elaboração no Minitab®)

4.3.Artificial Neural Networks

Em termos técnicos, as ANN são modelos computacionais que processam informações inspiradas no funcionamento do cérebro ou em sinais baseados em neurônios biológicos (Grus, 2016; Katal & Singh, 2022). As ANN são compostas por neurônios artificiais conectados uns aos outros em uma estrutura de camadas (Grus, 2016; Razavi, 2021). Assim, cada neurônio processa informações recebidas de outros neurônios e gera uma nova informação que é enviada para outros neurônios (Grus, 2016). As ANN geralmente possuem três camadas (Joshi & Vyawahare, 2023): i) camada de entrada; ii) camada oculta; iii) camada de saída. Em cada camada, pode haver apenas uma dúzia ou milhões de unidades, ou nós, a depender dos dados e/ou da complexidade do sistema (Joshi & Vyawahare, 2023). A adição de camadas ocorre se as análises e processamentos deparam-se com aumentos nos volumes dos dados e/ou na complexidade da função e são chamadas de redes neurais profundas (Katal & Singh, 2022; Joshi & Vyawahare, 2023). As ANN são amplamente utilizadas para reconhecimento de padrões, predição, controle, otimização, classificação e memória associativa (Grus, 2016; Katal & Singh, 2022; Maier et al., 2023). Todo o aprendizado de uma ANN ocorre mediante a um processo no qual são apresentados um conjunto de dados de entrada e saída cujo objetivo é a calibração do modelo de ANN (Katal & Singh, 2022). É nestes treinamentos que as ANN aprendem a realizar associações para identificar um conjunto de parâmetros que melhor correlacionam à entrada/saída desejada (Grus, 2016; Joshi & Vyawahare, 2023). As topologias das redes utilizadas neste estudo estão descritas na Tabela 4. Na sequência, as configurações de treinamento de cada topologia da rede estão apresentadas pela Tabela 5.

Topologia de Rede	Camadas Escondidas	Neurônios por Camada
Versão 1	5	128 - 64 - 32 - 64 - 32
Versão 2	5	128 - 64 - 32 - 16 - 8

Tabela 4 – Topologias de rede utilizadas nas simulações das ANN
(Fonte: Elaboração no Python®)

As Tabelas 4 e 5 mostram que as versões de configuração das redes foram aplicadas em ambas as bases de dados (B1 e B2). A etapa subsequente consistiu em incorporar o conhecimento heurístico pertinente à modelagem do problema e realizar o pré-processamento de dados das ANN. Portanto, a fim de obter resultados satisfatórios, foram iniciadas as simulações com o objetivo de treinar e executar a ANN. Estas simulações foram realizadas utilizando a rede do modelo Feed-Forward com retropropagação do erro (Grus, 2016). Todos os dados de cada base foram separados em dois conjuntos para a realização das simulações das ANN. As simulações então utilizaram dados de treinamento (70% dos registros) e um conjunto independente de teste (30% dos registros restantes) para validar o aprendizado das ANN.

Número Teste	Função de Ativação	Topologia de Rede	Base de Dados	Épocas de Aprendizizado	Tamanho do Lote de Amostras
1	Tangente Hiperbólica	Versão 1	B1	100	16
2					32
3					64
4				16	
5				200	32
6				64	
7			B2	100	16
8					32
9					64
10				200	16
11				32	
12				64	
13		Versão 2	B1	100	16
14					32
15					64
16				16	
17				200	32
18				64	
19			B2	100	16
20					32
21					64
22				16	
23				200	32
24				64	

Tabela 5 – Configurações de aprendizado das ANN (Fonte: Elaboração no Python®)

5. Experimentações e Análise dos Resultados

As experimentações computacionais avaliaram a viabilidade da aplicação das ANN para a predição do índice de ATR. Estas experimentações foram realizadas para os dados da Tabela 2, enquanto os resultados dos treinamentos estão dispostos na Tabela 6. A divisão dos dados para os treinamentos e as experimentações foi realizada de maneira aleatória pelas próprias ANN. Logo, incluído o período de treinamento, o tempo total de processamento e testes foi de pouco mais de 2 horas. Após serem treinadas, as ANN levam poucos segundos para gerar o valor previsto para as amostras. A métrica MAPE foi utilizada para análise da qualidade dos resultados e a avaliação da eficácia das ANN. A avaliação final foi consubstanciada pelo conjunto dos resultados médios obtidos nos vários experimentos das ANN.

Testes	Tempo (min)	R ²	MAPE
1	6,62	85,03%	3,14%
2	3,41	84,91%	3,22%
3	2,41	83,93%	3,43%
4	14,40	86,58%	2,81%
5	7,64	86,70%	2,79%
6	4,41	85,61%	3,04%
7	4,39	86,43%	2,93%
8	2,02	86,77%	2,94%
9	1,39	80,61%	4,38%
10	7,76	88,26%	2,50%
11	4,39	88,34%	2,51%
12	2,72	87,75%	2,72%
13	6,41	84,73%	3,18%
14	3,21	84,66%	3,33%
15	2,40	81,66%	3,96%
16	13,10	85,86%	3,01%
17	7,41	85,82%	3,06%
18	4,40	85,53%	3,10%
19	3,64	87,05%	2,86%
20	2,39	83,14%	3,74%
21	1,06	-0,04%	12,06%
22	8,39	87,95%	2,71%
23	4,39	86,52%	2,88%
24	2,32	84,62%	3,41%

Tabela 6 – Resultados dos testes do treinamento das ANN (Fonte: Elaboração no Python®)

A Tabela 6 mostra os resultados dos testes de treinamento das ANN e a porcentagem absoluta de erro em relação aos valores reais pela métrica MAPE. Exemplo pertinente como referência das análises é o teste 10. A análise do “Teste 10” evidencia que o modelo apresenta previsões cuja média de diferença entre o valor previsto e o real equivale a 2,50%. Portanto, a variação percentual das previsões, tanto para cima quanto para baixo do valor real, é de 2,50%. Ao considerar todos os resultados, a média foi 3,49% e a mediana 3,05%. A iteração com o melhor resultado foi a obtida pelo “Teste 10” (erro de 2,50%), enquanto a proporcionada pelo “Teste 21” (erro de 12,06%). Para avaliar o ajuste do modelo proposto, a Figura 6 mostra a comparação dos valores reais com os valores médios previstos nas simulações aplicadas na base B1 e na B2. A seguir, uma comparação dos valores reais com os valores previstos nos resultados das pré-análises é ilustrada pela Figura 7.

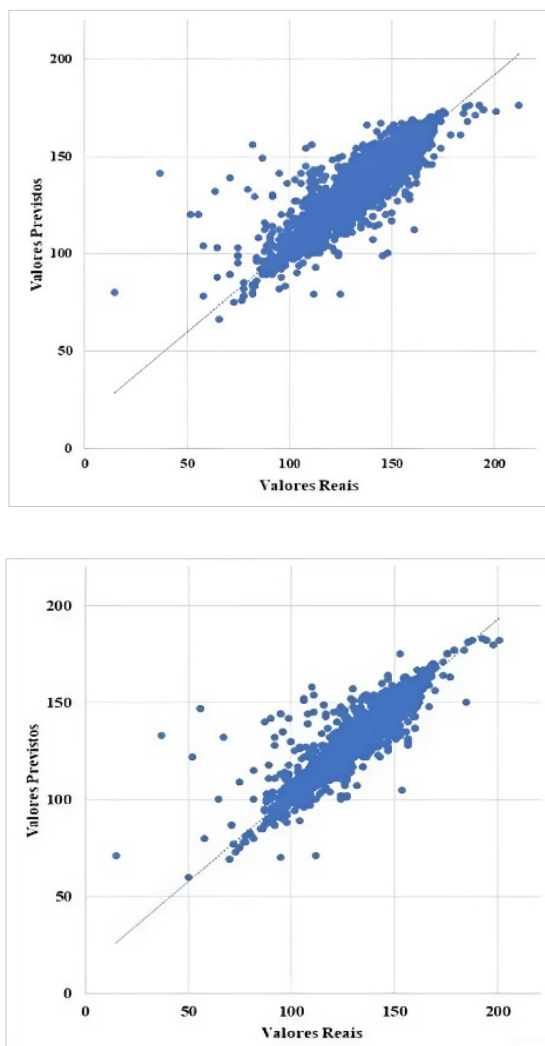


Figura 6 – Comparação dos valores reais com os valores médios das simulações (B1 e B2)
(Fonte: Elaboração no Python®)

As análises das Figuras 6 e 7 demonstram o desempenho satisfatório obtido pelas ANN para lidar com a complexidade da relação entre as variáveis que interferem no índice de ATR. As análises dos resultados e o debate junto ao grupo técnico da empresa em estudo mostraram que o uso das ANN é recomendado para a predição do ATR. As pré-análises do ATR também indicam uma discrepância em relação aos valores reais no momento da colheita se comparados com os resultados das ANN. Além disso, este é um processo muito dispendioso e a capacidade do laboratório da empresa é limitada a uma quantidade de apenas 30 amostras/dia. Tais amostras são também consideradas insuficientes em

termos de representatividade do ATR. Logo, as validações também ocorreram mediante comparações entre as análises laboratoriais e os resultados das ANN. A ocorrência do erro médio das pré-análises (12,17%) foi superior ao maior erro (12,06%) dos testes das ANN. As experimentações evidenciaram um erro médio absoluto (3,49%) muito inferior ao atual método utilizado (12,17%). Esses resultados indicam que o modelo pode servir de apoio às tomadas de decisões operacionais na referida usina do interior de São Paulo.

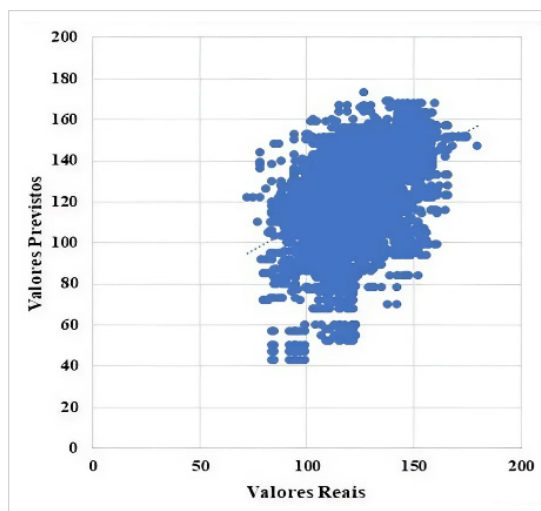


Figura 7 – Comparação dos valores reais com os valores dos resultados das pré-análises
(Fonte: Elaboração no Python®)

6. Considerações Finais

Este artigo apresentou um estudo de caso cujo objetivo foi formular um modelo computacional capaz de estimar o índice de ATR na produção de cana-de-açúcar em uma usina sucroalcooleira no interior do estado de São Paulo. Todo o escopo deste estudo centrou-se na adaptação às exigências do mundo real para propor um modelo ágil capaz de minimizar o tempo e os custos envolvidos no processo de mensuração do índice ATR. Esse modelo foi aplicado a um grande banco de dados reais contendo fatores meteorológicos e de produção referente a 48.151 talhões das safras de 2016/2017 a 2022/2023. A formulação do modelo é baseada nas ANN e utiliza a linguagem de programação Python® junto com a rede Feed-Forward. Também foram realizadas análises preliminares (estatísticas descritivas) para a padronização dos dados e a seleção de atributos relevantes via software Minitab®. Esses dados foram tratados por meio de ajustes de hiperparâmetros seguidos da modelagem matemática das ANN. As experimentações foram realizadas e a eficácia das ANN avaliada por meio da métrica MAPE. As análises apresentaram um erro médio absoluto (3,49%) muito inferior ao método atual utilizado (12,17%). Além disso, as curvas de maturação dos diversos lotes de cana-de-açúcar puderam ser geradas pela ANN. Este estudo é muito relevante para o setor do agronegócio e tem como principal contribuição fornecer um modelo de baixo

custo que estima com acurácia muito satisfatória o índice de ATR. Este modelo pode ser utilizado como suporte para a tomada de decisões mais fundamentadas sobre o momento ideal da colheita ou PIU. A limitação do modelo está no fato de os hiperparâmetros de configuração de aprendizado da rede demandarem grande intervenção manual (sujeito a vieses). Sugestões para futuras pesquisas são: i) Aplicar as ANN às variáveis dinâmicas do ATR; ii) Formular modelos para programar e otimizar o roteiro de colheita; iii) Realizar intensivas experimentações computacionais por meio de estudos comparativos entre outros modelos e as ANN; iv) Formular um modelo estimado de previsão para obter uma equação estimada para fazer previsões; v) Aplicar modelos de ANOVA para verificar a igualdade ou não das médias de ATR nos diferentes níveis de cada covariável categórica (regiões, tipo de corte, ciclo de plantio etc.); vi) Aplicar MLR para verificar conjuntamente os efeitos de todas as covariáveis (categóricas e contínuas) nas respostas de ATR.

Financiamento

Esta pesquisa foi apoiada financeiramente pelos Projetos nº 312585/2021-7 e nº 404819/2023-0 do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e nº 2700441 da Fundação Nacional de Desenvolvimento do Ensino Superior Particular – FUNADESP.

Referências

- Ali, A., Altaf, M. T., Bedir, M., Liaqat, W., Jamil, A., Nadeem, M. A., & Baloch, F. S. (2024). Biotechnological Advancements Toward Sugarcane Crop Improvement. *In Industrial Crop Plants* (pp. 333-352). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-1003-4_13
- Amaro, R. P., Christina, M., Todoroff, P., Le Maire, G., Fiorio, P. R., De Carvalho Pereira, E., & Luciano, A. C. D. S. (2025). Regional model to predict sugarcane yield using sentinel-2 imagery in São Paulo State, Brazil. *Sugar Tech*, 27(1), 108-118. <https://doi.org/10.1007/s12355-024-01468-z>.
- Bocca, F. F., & Rodrigues, L. H. A. (2016). The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and Electronics in Agriculture*, 128, 67-76. <https://doi.org/10.1016/j.compag.2016.08.015>.
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 26(2), 211-252. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Canata, T. F., Júnior, M. R. B., de Oliveira, R. P., Furlani, C. E. A., & da Silva, R. P. (2024). AI-Driven Prediction of Sugarcane Quality Attributes Using Satellite Imagery. *Sugar Tech*, 26(3), 741-751. <https://doi.org/10.1007/s12355-024-01399-9>
- Canata, T. F., Wei, M. C. F., Maldaner, L. F., & Molin, J. P. (2021). Sugarcane yield mapping using high-resolution imagery data and machine learning technique. *Remote Sensing*, 13(2), 232. <https://doi.org/10.3390/rs13020232>

- Cardoso, L. A. S., de Mesquita, B. D. N. R., & Farias, P. R. S. (2025). Use of machine learning algorithms in the context of sugarcane in Brazil: a review. *Iran Journal of Computer Science*, 1-30. <https://doi.org/10.1007/s42044-025-00250-y>
- Charoen-Ung, P., & Mittrapiyanuruk, P. (2018). Sugarcane yield grade prediction using random forest and gradient boosting tree techniques. In *15th International Joint Conference on Computer Science and Software Engineering* (pp. 1-5). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/JCSSE.2018.8457391>
- Charoen-Ung, P., & Mittrapiyanuruk, P. (2019). Sugarcane yield grade prediction using random forest with forward feature selection and hyper-parameter tuning. In *14th International Conference on Computing and Information Technology* (pp. 1-5). Springer Verlag. https://doi.org/10.1007/978-3-319-93692-5_4
- Coelho, A. P., Bettiol, J. V. T., Dalri, A. B., Fischer Filho, J. A., Faria, R. T., & Palaretti, L. F. (2019). Application of artificial neural networks in the prediction of sugarcane juice Pol. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 23(1), 9-15. <https://doi.org/10.1590/1807-1929/agriambi.v23n1p9-15>
- Colmanetti, M. A. A., Cuadra, S. V., Lamparelli, R. A. C., Cabral, O. M. R., de Castro Victoria, D., de Almeida Monteiro, J. E. B., de Freitas, H. C., Galdos, M. V., Marafon, A. C., de Andrade Junior, A. S., dos Anjos e Silva, S. D., Buffon, V. B., Hernandez, T. A. D., & le Maire, G. (2024). Modeling sugarcane development and growth within ECOSMOS biophysical model. *European Journal of Agronomy*, 154, 127061. <https://doi.org/10.1016/j.eja.2023.127061>
- Companhia Nacional de Abastecimento, CONAB. (2024). Boletim - Cana-de-açúcar - 1º levantamento - 2024/25. Tabela de dados - Produção de cana-de-açúcar e subprodutos, 12(1). <https://www.conab.gov.br/info-agro/safras/cana>
- Confederação da Agricultura e Pecuária do Brasil, CNA. (2022). Sumário Executivo I OCDE - FAO Perspectivas Agrícolas 2023-2032. Brasília. <https://www.cnabrazil.org.br/publicacoes/sumario-executivo-i-ocde-fao-perspectivas-agricolas-2023-2032>
- Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Etanol do Estado de São Paulo, CONSECAN-SP. (2023). Regulamento do CONSECAN. São Paulo. <https://www.consecana.com.br/regulamento.asp>
- Denham, T. (2024). Forgetting cane grasses: Switching temporal focus to reveal mosaics of Saccharum diversity. *Plants, People, Planet*, 1–13. <https://doi.org/10.1002/ppp3.10489>
- De Oliveira, M. P. G., Bocca, F. F., & Rodrigues, L. H. A. (2017). From spreadsheets to sugar content modeling: A data mining approach. *Computers and Electronics in Agriculture*, 132, 14-20. <https://doi.org/10.1016/j.compag.2016.11.012>
- De Oliveira, M. P. G., & Rodrigues, L. H. A. (2020). How good are the models available for estimating sugar content in sugarcane?. *European Journal of Agronomy*, 113, 125992. <https://doi.org/10.1016/j.eja.2019.125992>

- Dias, H. B., & Sentelhas, P. C. (2017). Evaluation of three sugarcane simulation models and their ensemble for yield estimation in commercially managed fields. *Field Crops Research*, 213, 174-185. <https://doi.org/10.1016/j.fcr.2017.07.022>
- Dias, H. B., & Inman-Bamber, G. (2020). Sugarcane: Contribution of process-based models for understanding and mitigating impacts of climate variability and change on production. *Systems Modeling*, 217-260. https://doi.org/10.1007/978-981-15-4728-7_8
- Dias, H. B., Inman-Bamber, G., Bermejo, R., Sentelhas, P. C., & Christodoulou, D. (2019). New APSIM-Sugar features and parameters required to account for high sugarcane yields in tropical environments. *Field Crops Research*, 235, 38-53. <https://doi.org/10.1016/j.fcr.2019.02.002>
- Dias, H. B., Inman-Bamber, G., Everingham, Y., Sentelhas, P. C., Bermejo, R., & Christodoulou, D. (2020). Traits for canopy development and light interception by twenty-seven Brazilian sugarcane varieties. *Field Crops Research*, 249, 107716. <https://doi.org/10.1016/j.fcr.2020.107716>
- Dias, H. B., Inman-Bamber, G., Sentelhas, P. C., Everingham, Y., Bermejo, R., & Christodoulou, D. (2021). High-yielding sugarcane in tropical Brazil—Integrating field experimentation and modelling approach for assessing variety performances. *Field Crops Research*, 274, 108323. <https://doi.org/10.1016/j.fcr.2021.108323>
- Dias Neto, A. F., Albiero, D., Rossetto, R., & Biagi, J. D. (2022). Modeling of mechanized sugarcane harvesting to support decision-making on asset management. *Sugar Tech*, 24(3), 798-812. <https://doi.org/10.1007/s12355-022-01126-2>
- Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, 36, 1-9. <https://doi.org/10.1007/s13593-016-0364-z>
- Everingham, Y. L., Smyth, C. W., & Inman-Bamber, N. G. (2009). Ensemble data mining approaches to forecast regional sugarcane crop production. *Agricultural and Forest Meteorology*, 149(3-4), 689-696. <https://doi.org/10.1016/j.agrformet.2008.10.018>
- Ferraciolli, M. A., Bocca, F. F., & Rodrigues, L. H. A. (2019). Neglecting spatial autocorrelation causes underestimation of the error of sugarcane yield models. *Computers and Electronics in Agriculture*, 161, 233-240. <https://doi.org/10.1016/j.compag.2018.09.003>
- Florentino, H. D. O., Irawan, C., Aliano, A. F., Jones, D. F., Cantane, D. R., & Nervis, J. J. (2018). A multiple objective methodology for sugarcane harvest management with varying maturation periods. *Annals of Operations Research*, 267, 153-177. <https://doi.org/10.1007/s10479-017-2568-2>
- Grus, J. (2016). *Data Science do Zero: Primeiras Regras com o Python*. (pp. 213-224). Alta Books

- Hammer, R. G., Sentelhas, P. C., & Mariano, J. C. (2020). Sugarcane yield prediction through data mining and crop simulation models. *Sugar Tech*, 22(2), 216-225. <https://doi.org/10.1007/s12355-019-00776-z>.
- Joshi, M., Bhosale, S., & Vyawahare, V. A. (2023). A survey of fractional calculus applications in artificial neural networks. *Artificial Intelligence Review*, 56, 13897–13950. <https://doi.org/10.1007/s10462-023-10474-8>
- Katal, A., & Singh, N. (2022). Artificial Neural Network: Models, Applications, and Challenges. In: Tomar, R., Hina, M.D., Zitouni, R., Ramdane-Cherif, A. (eds) *Innovative Trends in Computational Intelligence*. (pp. 235-257). EAI/ Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-78284-9_11
- Kathirvel, N., & Venkatachalam, I. (2024). A Detailed Review for Predicting the Quantity of Sugar from Sugarcane using Various Models. In *IEEE Access*, 2025(13), 32122-32146. <https://doi.org/10.1109/ACCESS.2024.3522495>
- Lim, W. M., Kumar, S., & Ali, F. (2022). Advancing knowledge through literature reviews: ‘What’, ‘why’, and ‘how to contribute’. *The Service Industries Journal*, 42(7-8), 481–513. <https://doi.org/10.1080/02642069.2022.2047941>
- Luciano, A. C. S., Picoli, M. C. A., Duft, D. G., Rocha, J. V., Leal, M. R. L., & Maire, G. (2021). Empirical model for forecasting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. *Computers and Electronics in Agriculture*, 184. <https://doi.org/10.1016/j.compag.2021.106063>
- Madeiro, S. S., da Silva Oliveira, F. R., Alexandre, F. B. A., & de Lima Neto, F. B. (2006). Intelligent modeling of sugar-cane maturation. In *Computers in Agriculture and Natural Resources* (p. 642). American Society of Agricultural and Biological Engineers. <https://doi.org/10.13031/2013.21950>
- Matović, N., & Ovesni, K. (2023). Interaction of quantitative and qualitative methodology in mixed methods research: integration and/or combination. *International Journal of Social Research Methodology*, 26(1), 51-65. <https://doi.org/10.1080/13645579.2021.1964857>.
- Medar, R., & Rajpurohit, V. S. (2019). Supervised machine learning techniques for predicting sugarcane yield. *International Journal of Recent Technology and Engineering*, 8(2), 5662-5668. <https://doi.org/10.35940/ijrte.B2612.078219>
- Monteiro, L. A., & Sentelhas, P. C. (2014). Potential and actual sugarcane yields in Southern Brazil as a function of climate conditions and crop management. *Sugar Tech*, 16, 264-276. <https://doi.org/10.1007/s12355-013-0275-0>.
- Monteiro, L. A., & Sentelhas, P. C. (2017). Sugarcane yield gap: can it be determined at national level with a simple agrometeorological model?. *Crop and Pasture Science*, 68(3), 272-284. <https://doi.org/10.1071/CP16334>.
- Nihar, A., Patel, N. R., & Danodia, A. (2022). Machine-learning-based regional yield forecasting for sugarcane crop in Uttar Pradesh, India. *Journal of the Indian Society of Remote Sensing*, 50(8), 1519-1530. <https://doi.org/10.1007/s12524-022-01549-0>

- Pignède, E., Roudier, P., Diedhiou, A., N'Guessan Bi, V. H., Koba, A. T., Konaté, D., & Péné, C. B. (2021). Sugarcane yield forecast in Ivory Coast (West Africa) based on weather and vegetation index data. *Atmosphere*, 12(11), 1-22. <https://doi.org/10.3390/atmos12111459>
- Pravallika, K., Karuna, G., Anuradha, K., & Srilakshmi, V. (2021). Deep neural network model for proficient crop yield prediction. In *International Conference on Design and Manufacturing Aspects for Sustainable Energy* (3rd ed.). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1051/e3sconf/202130901031>
- Phuphaphud, A., Saengprachatanarug, K., Posom, J., Maraphum, K., & Taira, E. (2019). Prediction of the fibre content of sugarcane stalk by direct scanning using visible-shortwave near infrared spectroscopy. *Vibrational Spectroscopy*, 101, 71-80. <https://doi.org/10.1016/j.vibspec.2019.02.005>
- Rahman, M. M., & Robson, A. (2020). Integrating landsat-8 and sentinel-2 time series data for yield prediction of sugarcane crops at the block level. *Remote Sensing*, 12(8). <https://doi.org/10.3390/RS12081313>
- Satpathi, A., Chand, N., Setiya, P., Ranjan, R., Nain, A. S., Vishwakarma, D. K., Saleem, K., Obaidullah, A. J., Yadav, K., K., & Kisi, O. (2025). Evaluating statistical and machine learning techniques for sugarcane yield forecasting in the tarai region of North India. *Computers and Electronics in Agriculture*, 229, 109667. <https://doi.org/10.1016/j.compag.2024.109667>
- Silva, M. D. A., Santos, H. L., Germino, G. H., & Véliz, J. G. E. (2023). Effect of ripeners on the nutrient concentrations of sugarcane leaves. *Journal of Plant Nutrition*, 46(15), 3737-3758. <https://doi.org/10.1080/01904167.2023.2211609>
- Shanthi, R. M., Alarmelu, S., Mahadeva Swamy, H. K., & Lakshmi Pathy, T. (2023). Impact of climate change on sucrose synthesis in sugarcane varieties. In *Agro-industrial perspectives on sugarcane production under environmental stress* (pp. 13-38). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-3955-6_2
- Shendryk, Y., Pan, L., Craigie, M., Stasolla, M., Ticehurst, C., & Thorburn, P. (2020). A satellite-based methodology for harvest date detection and yield prediction in sugarcane. In *International Geoscience and Remote Sensing Symposium* (Virtual, Waikoloa). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/IGARSS39084.2020.9323418>
- Souza, C. C. de, Carneiro Jr., J. B. A., Cristaldo, M. F., Castelhão, R. A., Frainer, D. M., & Viganó, H. H. da G. (2022). Forecast of the production of sugarcane and its derivatives, through ARIMA, in the period from 2020 to 2025 in the Centro-South macroregion of Brazil. *Research, Society and Development*, 11(4), 1,19. <https://doi.org/10.33448/rsd-v11i4.27048>
- Werlang, M. K., & Werlang, E. P. (2021). Water balance and Thornthwaite climatic classification of the camobi neighbourhood, Santa Maria, RS, Brazil. *Revista Geografia Ensino & Pesquisa*, 25(7), 2-26. <https://doi.org/10.5902/2236499440437>