

CAFEICULTURA DE PRECISÃO: PREDIÇÃO DA PRODUTIVIDADE DO CAFÉ A PARTIR DE APRENDIZADO DE MÁQUINAS

RAFAEL DE OLIVEIRA FARIA¹, GABRIEL ARAÚJO E SILVA FERRAZ², FÁBIO MOREIRA DA SILVA³, MICHELE DUARTE DE MENEZES⁴, LUCAS SANTOS SANTANA⁵, ALDIR CARPES MARQUES FILHO⁶.

¹ Professor Doutor, Departamento de Engenharia Agrícola, Universidade Federal de Lavras, rafael.faria@ufla.br

² Professor Doutor, Departamento de Engenharia Agrícola, Universidade Federal de Lavras, gabriel.ferraz@ufla.br

³ Professor Doutor, Departamento de Engenharia Agrícola, Universidade Federal de Lavras, famsilva@ufla.br

⁴ Professor Doutor, Departamento de Ciência do Solo, Universidade Federal de Lavras, michele.menezes@ufla.br

⁵ Doutor, Departamento de Engenharia Agrícola, Universidade Federal de Lavras, lucas.unemat@hotmail.com

⁶ Professor Doutor, Departamento de Engenharia Agrícola, Universidade Federal de Lavras, aldir@ufla.br

Apresentado no
LII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2023
18 a 21 de outubro de 2023 – Ribeirão Preto - SP, Brasil

RESUMO: A Cafeicultura de Precisão (CP) é um conceito emergente que utiliza técnicas avançadas de agricultura de precisão para otimizar o cultivo do café. Estudos têm sido realizados para avaliar a qualidade das malhas amostrais utilizadas nesse contexto. A aplicação da CP permite a identificação de variações espaciais significativas produtividade, relacionadas a diversos fatores como atributos do solo, manejo, pragas e doenças. O presente estudo teve como objetivo explorar a aplicação da Agricultura de Precisão na predição da produtividade do café, utilizando aprendizado de máquinas e atributos químicos do solo e das folhas como variáveis explanatórias. Foram utilizados diferentes algoritmos de aprendizado de máquinas: *adaptive boosting* (AdaBoost), *random forest*, regressão linear, árvores de decisão, *K-nearest neighbor* (kNN) e *support vector machine* (SVM). A produtividade foi predita adequadamente, com maior acurácia encontrada a partir do AdaBoost (MSE de 64,28, RMSE de 8,02, MAE de 6,20 em L planta⁻¹ e coeficiente de determinação R² de 0,85), sugerindo que o impulsionamento (boosting) contido neste algoritmo foi efetivo para a predição da produtividade do café.

PALAVRAS-CHAVE: fertilidade do solo, análises foliares, inteligência artificial.

PRECISION COFFEE CULTIVATION: PREDICTION OF COFFEE PRODUCTIVITY USING MACHINE LEARNING

ABSTRACT: A Precision Coffee Cultivation (PCC) is an emerging concept that employs advanced precision agriculture techniques to optimize coffee cultivation. Studies have been conducted to assess the quality of the sampling grids used in this context. PCC allows for the identification of significant spatial variations in productivity, related to various factors such as soil attributes, management practices, pests, and diseases. This study aimed to explore the application of Precision Agriculture in predicting coffee productivity, using machine learning and soil and leaf chemical attributes as explanatory variables. Different machine learning algorithms were utilized: Adaptive Boosting (AdaBoost), Random Forest, Linear Regression, Decision Trees, K-Nearest Neighbor (kNN), and Support Vector Machine (SVM). Productivity was adequately predicted, with the highest accuracy achieved using AdaBoost (MSE of 64.28, RMSE of 8.02, MAE of 6.20 in L plant⁻¹, and coefficient of determination R² of 0.85), suggesting that the boosting mechanism in this algorithm effectively contributed to coffee productivity prediction.

KEYWORDS: soil fertility, foliar analysis, artificial intelligence.

INTRODUÇÃO: A Cafeicultura de Precisão é um conceito em ascensão que utiliza técnicas avançadas de agricultura de precisão para otimizar o cultivo do café. Com o objetivo de maximizar a rentabilidade das colheitas, a agricultura de precisão oferece tecnologias que auxiliam os produtores rurais na identificação de estratégias para melhorar a eficiência da gestão agrícola. Essa abordagem busca tornar o agronegócio mais competitivo e impulsionar o sucesso da atividade agrícola como um todo, sendo respaldada por diversos estudos e pesquisas. (Alves, Queiroz e Pinto, 2007; Jacintho, 2016; Silva e Alves, 2013).

Ao mapear a produtividade da lavoura, é possível identificar variações espaciais significativas, relacionadas a diversos fatores como atributos do solo, manejo, pragas e doenças. Essas investigações têm como objetivo compreender os padrões de produtividade na cafeicultura e desenvolver estratégias de manejo mais eficientes para otimizar a produção e a rentabilidade nas áreas de cultivo de café. (Faulin e Molin, 2007; Ferraz et al., 2012, Molin et al., 2007; Molin, Faulin e Stanislavski, 2009; Silva e Alves, 2013).

Essa abordagem oferece a possibilidade de explorar padrões espaciais mais complexos e capturar informações relevantes para a previsão da produtividade do cafeeiro (Kittichotsawat et al., 2022; Oliveira, 2022).

O objetivo deste trabalho foi comparar diferentes algoritmos de aprendizado de máquinas para a predição produtividade do café, utilizando atributos químicos do solo e de folhas como variáveis explicativas. O foco principal foi avaliar o desempenho desses modelos na estimativa da produtividade agrícola, buscando identificar qual método apresenta maior.

MATERIAL E MÉTODOS: O experimento foi conduzido na Fazenda Samambaia, localizada no município de Santo Antônio do Amparo, na região dos Campos das Vertentes, Minas Gerais, Brasil. As coordenadas geográficas médias da fazenda são 20°58'19.92" de latitude Sul e 44°53'24.18" de longitude Oeste. A área experimental abrange um total de 54,65 hectares, onde foram cultivados cafeeiros da espécie *Coffea arabica* L., cultivar Acaiaí IAC 474-19.

Foi utilizado uma malha amostral com dois pontos por hectare. Essa abordagem baseia-se em estudos anteriores que investigaram o índice de malha ótima para a cafeicultura de precisão, como os estudos conduzidos por Carvalho et al. (2017a), Ferraz et al. (2017) e Figueiredo et al. (2018). No total, foram estabelecidos 111 pontos amostrais georreferenciados, distribuídos nas diferentes glebas da área experimental. Nestes pontos foram retiradas as amostras de solo, coletadas as folhas e realizado o levantamento de produtividade das plantas (Figura 1).

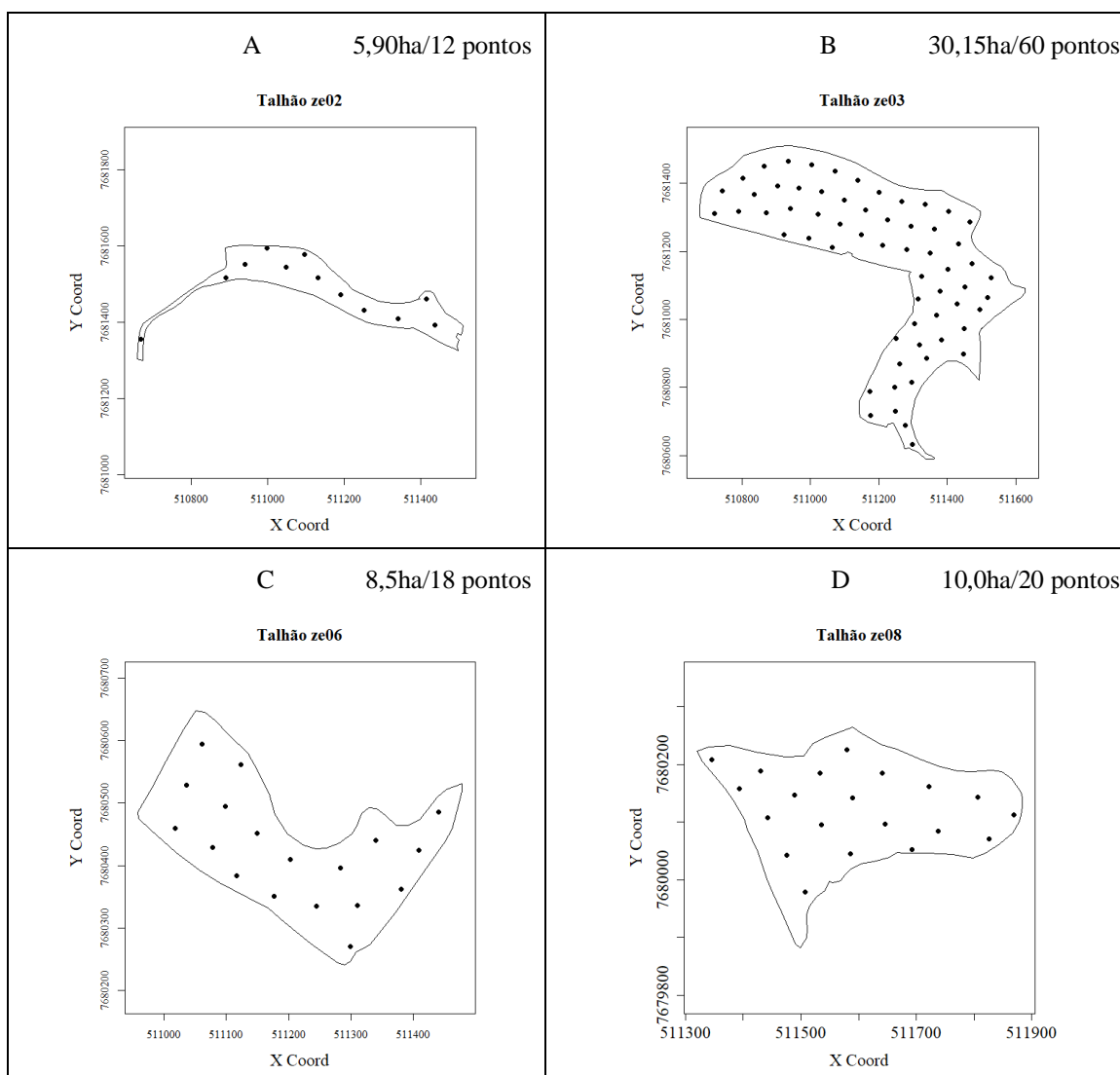


Figura 1 - Croquis das malhas de amostragem nas lavouras ZE02 (A), ZE03(B), ZE06(C) E ZE08(D), Y coord é Latitude, X Coord é longitude em metros.

A coleta das amostras de solo e de folha foram realizadas no mês de maio nos anos de 2017 e 2018. Em cada ponto amostral, foram coletadas 12 subamostras de solo (Figura 2B) que foram encaminhadas para laboratório para as análises de químicas, utilizando a metodologia ICP-OES (Inductively Coupled Plasma Optical Emission Spectrometry). Os atributos químicos do solo analisados foram P, K, Ca, Mg, Al, H + Al, SB, T, pH em água, V%, e MO. As amostras de nutrição foliar foram coletadas simultaneamente às amostras de solo, na mesma data e nas mesmas plantas demarcadas (Figura 2B). A coleta das folhas foi realizada para análise dos teores dos macronutrientes N, P, K, Ca, Mg e S, utilizando o metodologia da Embrapa 2019.

A produtividade de café por planta ($L \text{ planta}^{-1}$) foi determinada por meio da colheita manual sobre panos das quatro plantas mencionadas anteriormente. A colheita foi conduzida por trabalhadores rurais experientes na colheita de café, sendo acompanhados por um técnico responsável pela navegação e anotação das medições (Figura 2A).

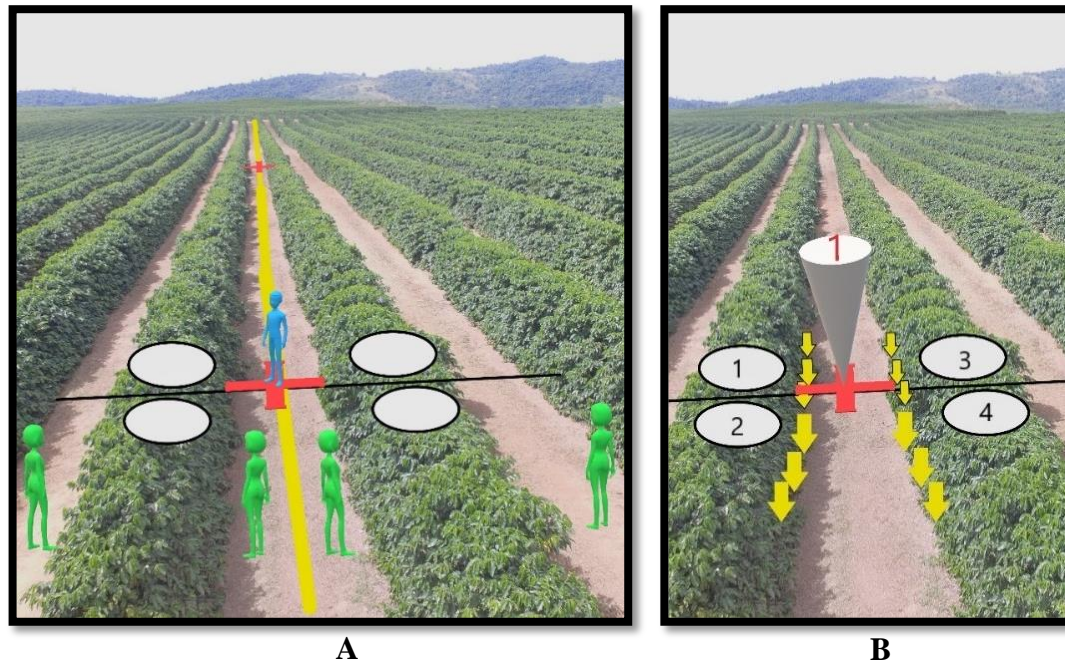


Figura 2 – A- Arranjo dos trabalhadores durante a colheita manual do café nas quatro plantas demarcadas. B – Esquema para a realização da amostragem foliar e de solo, seguindo as mesmas plantas utilizadas durante a colheita.

O software Orange Canvas 3.30.0, baseado na linguagem Python, foi utilizado para o processamento dos dados por meio de redes neurais. Uma gama de algoritmos classificadores de aprendizado de máquinas foi utilizada, como random forest, KNN, regressão linear, árvores de decisão, SVM, redes neurais artificiais e AdaBoost. Para atestar a confiabilidade dos modelos preditivos, as seguintes métricas de acurácia foram calculadas a partir de validação cruzada (10 *fold*). Os modelos foram avaliados utilizando medidas estatísticas, incluindo o coeficiente de determinação (r^2), que indica a proporção da variabilidade da produtividade explicada pelas variáveis de entrada. Além disso, foram utilizadas medidas como a raiz do erro quadrático médio (RMSE), que avalia a diferença entre os valores preditos e os valores reais da produtividade, considerando a escala da variável analisada. O erro absoluto médio (MAE) foi calculado como a média das diferenças absolutas entre os valores preditos e os valores reais, sendo uma medida precisa e robusta da habilidade do modelo em reproduzir a realidade. Por fim, o erro quadrático médio (MSE) foi usado para avaliar as diferenças individuais elevadas ao quadrado, sendo mais sensível a erros grandes. Quanto mais próximo o MSE estiver de zero, melhor será a simulação realizada pelos modelos.

RESULTADOS E DISCUSSÃO: A Tabela 1 exhibe os resultados dos testes de predição da produtividade do café, utilizando diversos métodos. Os indicadores de desempenho, como MSE, RMSE, MAE e R^2 , são utilizados para avaliar a precisão das previsões. A estratégia de validação cruzada de 10 *fold* dividiu o conjunto de dados em partes iguais para treinamento e teste. Esses resultados fornecem *insights* sobre a acurácia e a capacidade de previsão dos métodos avaliados, auxiliando na seleção do método mais adequado para a predição da produtividade do café.

TABELA 1. Métricas de acurácia da predição da produtividade do café obtida a partir dos diferentes algoritmos de aprendizado de máquinas.

Algoritmo	MSE	RMSE	MAE	R ²
AdaBoost	64,28	8,02	6,20	0,85
Random forest	76,88	8,77	6,72	0,83
Linear fegression	92,93	9,64	7,77	0,79
Árvore de decisão	118,60	10,89	8,39	0,73
kNN	203,84	14,28	11,52	0,54
SVM	213,93	14,63	12,55	0,51

Legenda: MSE (Erro quadrático médio), RMSE (Raiz do erro médio ao quadrado), MAE (Erro absoluto médio), R² (Coeficiente de determinação), *adaptative boosting* (AdaBoost), *K-nearest neighbor* (kNN) e *suport vector machine* (SVM).

O método AdaBoost demonstrou os melhores resultados em relação às métricas de acurácia. O MSE foi de 64,28 e o MAE foi de 6,20. Além disso, a raiz do erro médio ao quadrado (RMSE) foi de 8,02. Esses valores indicam um desempenho superior na predição da produtividade do café em comparação com os outros métodos avaliados.

O método random forest também apresentou resultados promissores, com MSE de 76,88, RMSE de 8,77 e MAE de 6,72. Esses valores estão próximos aos obtidos pelo AdaBoost, demonstrando uma boa capacidade de predição do modelo.

Por outro lado, os métodos de regressão linear, árvore de decisão, kNN e SVM apresentaram um desempenho inferior em relação aos dois primeiros métodos mencionados. Esses modelos mostraram valores mais elevados de MSE, RMSE e MAE, indicando uma maior discrepância entre as previsões e os valores observados da produtividade do café. Em relação ao ajuste aos dados observados, o método AdaBoost novamente se destacou com R² de 0,85 (85% da variabilidade dos dados pode ser explicada pelo modelo).

Na Figura 3, é possível observar a comparação entre a produtividade colhida manualmente (representada em azul) e a produtividade predita pelo AdaBoost (representada em laranja). Uma análise detalhada revela discrepâncias na predição do AdaBoost para produtividades baixas, especialmente entre os pontos 1 e 15, bem como entre os pontos 141 e 148, onde a correlação entre os valores preditos e os valores reais não é significativa. No entanto, para a maioria dos pontos de comparação, os resultados são satisfatórios, indicando que o AdaBoost é capaz de fazer previsões precisas para a maioria dos níveis de produtividade. É importante observar que, apesar das discrepâncias mencionadas, o modelo geral do AdaBoost demonstra uma boa capacidade de capturar os padrões de produtividade e fazer previsões próximas aos valores reais. Essa análise ressalta a importância de considerar as limitações do modelo, especialmente em relação a faixas específicas de produtividade, e reforça a necessidade contínua de aprimorar e refinar o algoritmo para alcançar resultados ainda mais precisos.

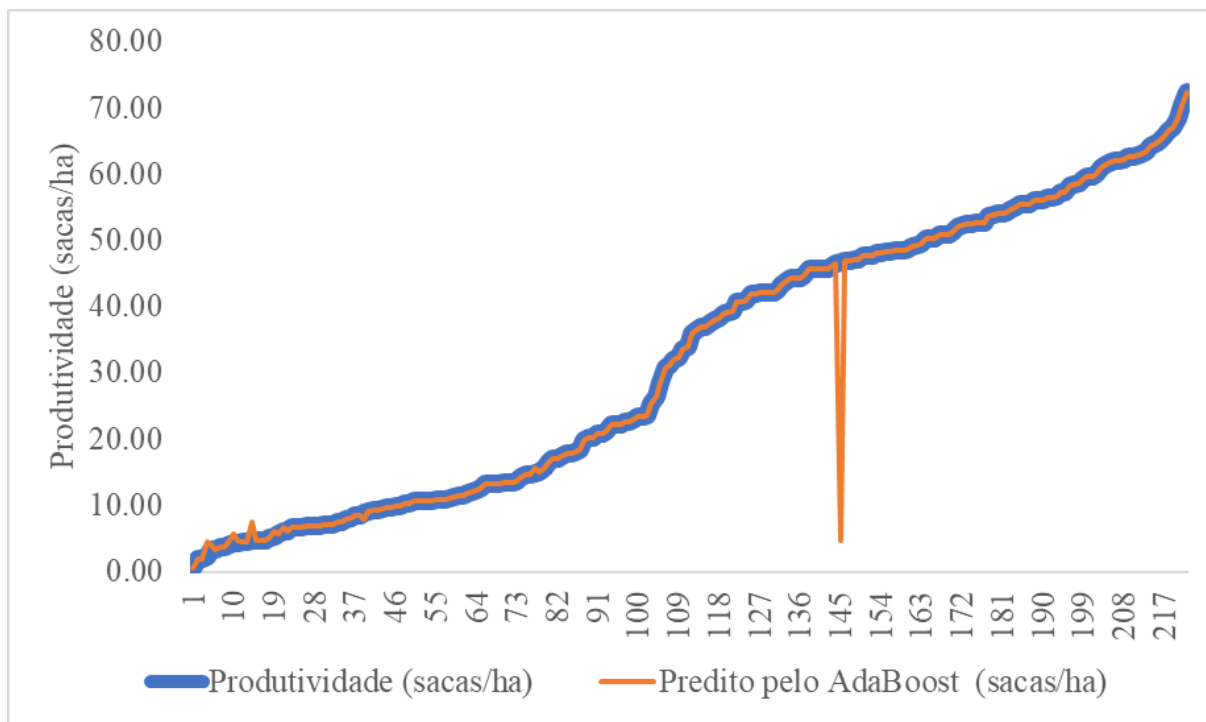


Figura 3 - Comparação entre a produtividade colhida manualmente (linha azul) e a produtividade predita pelo modelo AdaBoost (linha laranja)

Os resultados obtidos destacam a eficácia do algoritmo AdaBoost na previsão da produtividade do cafeeiro ao levar em conta diferentes variáveis. A combinação de dados de múltiplas lavouras e anos de avaliação, juntamente com a seleção criteriosa de atributos, contribui para o desenvolvimento de modelos robustos e confiáveis, auxiliando os produtores na tomada de decisões agrícolas.

No estudo conduzido por Aravind (2021), foram analisados diversos parâmetros climáticos como variáveis de entrada para avaliar qual regressor seria mais adequado na previsão das condições climáticas e locais ideais para o cultivo das dez principais culturas, incluindo inhame, mandioca, trigo, batatas, arroz, milho, batatas-doces, sorgo, soja e plátanos (ou bananas-da-terra). Os resultados apontaram que a combinação do regressor da árvore de decisão e do regressor AdaBoost apresentou maior precisão na previsão da produtividade dessas culturas, corroborando com os dados apresentados neste trabalho.

CONCLUSÕES: O método AdaBoost destacou-se como o mais eficiente na previsão da produtividade do café, utilizando atributos químicos do solo e das folhas como variáveis de entrada. Esses resultados indicam que algoritmos de aprendizado de máquinas pode ser uma escolha promissora para o manejo da lavoura cafeeira, auxiliando os produtores na tomada de decisões e no aumento da eficiência agrícola.

AGRADECIMENTOS: Consórcio Pesquisa Café, UFLA, FAPEMIG, CNPQ, Fazenda Samambaia, Fazenda Bom Jardim.

REFERÊNCIAS:

ALVES, E. A.; QUEIROZ, D. M. DE; PINTO, F. DE A. DE C. Cafeicultura de Precisão. Em: ZAMBOLIM, L. (Ed.). **Boas Práticas Agrícolas na Produção de Café**. Viçosa MG: UFV, 2007. p. 234.

ARAVIND, T. Review of Machine learning models for Crop Yield Prediction. In: **Proceedings of the First International Conference on Combinatorial and Optimization, ICCAP 2021, December 7-8 2021, Chennai, India**. 2021.

CARVALHO, LUIS CARLOS CIRILO *et al.* Comparação entre amostragem foliar convencional e de precisão para análise de micronutrientes na cafeicultura. **Coffee Science**, v. 12, n. 2, p. 272–281, 2017a.

CARVALHO, LUIS C. C.; SILVA, F. M. DA; FERRAZ, G. A. S.; *et al.* Geostatistical analysis of Arabic coffee yield in two crop seasons. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, n. 6, p. 410–414, 2017b.

SILVA, F.C. da. (Ed.). Manual de análises químicas de solos, plantas e fertilizantes 2.ed. Brasília: **Embrapa Informação Tecnológica**; Rio de Janeiro: Embrapa Solos, 2009. 627p.

FAULIN, G. D. C.; MOLIN, J. P. Utilização dos conceitos da agricultura de precisão na cultura do café (*Coffea arabica* L.). **5 simposio de pesquisas dos cafes do Brasil**, n. 1998, p. 70–75, 2007.

FERRAZ, G. A. S.; SILVA, F. M.; *et al.* Geostatistical analysis of fruit yield and detachment force in coffee. **Precision Agriculture**, v. 13, n. 1, p. 76–89, 2012.

FERRAZ, G. A. S.; SILVA, F. M. DA; *et al.* Variabilidade espacial e temporal do fósforo, potássio e da produtividade de uma lavoura cafeeira. **Engenharia Agrícola**, v. 32, n. 1, p. 140–150, 2012.

FERRAZ, G. A. S. *et al.* Methodology to determine the soil sampling grid for precision agriculture in a coffee field. **Dyna**, v. 84, n. 200, p. 316–325, 2017.

FIGUEIREDO, V. C. *et al.* Development of a methodology to determine the best grid sampling in precision coffee growing. **Coffee Science**, v. 13, n. 3, p. 312–323, 2018.

JACINTHO, J. L. **Zonas de manejo na cafeicultura de precisão**. [s.l.] UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO, 2016.

KITTICHOTSATSAWAT, Y.; TIPPAYAWONG, N.; TIPPAYAWONG, K. Y. Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques. **Scientific Reports**, v. 12, n. 1, p. 14488, 25 ago. 2022.

KLOMPENBURG, T. VAN; KASSAHUN, A.; CATAL, C. Crop yield prediction using machine learning: A systematic literature review. **Computers and Electronics in Agriculture**, v. 177, p. 105709, out. 2020.

MOLIN, J. P. *et al.* Variação Espacial Na Produtividade De Milho Safrinha Devido Aos Macronutrientes E À População De Plantas. **Revista Brasileira de Milho e Sorgo**, v. 6, n. 3, p. 309–324, 2007.

MOLIN, J. P.; FAULIN, G. D. C.; STANISLAVSKI, W. M. Yield mapping and variable rate of fertilizers for coffee in brazil. **Acta Horticulturae**, n. 824, p. 261–266, abr. 2009.

OLIVEIRA, M. D. S. **Redes Neurais artificiais para ajuste de modelo de predição da produtividade em clones de Café Canéfora da Região Amazônica**. [s.l: s.n.].

SILVA, F. M. DA; ALVES, M. DE C. **Cafeicultura de Precisão**. Lavras: Editora UFLA, 2013.