

UNIVERSIDADE FEDERAL DE VIÇOSA

RAYANNE OLIVEIRA TEIXEIRA

**MONITORAMENTO NUTRICIONAL DE PALMA DE ÓLEO (*Elaeis guineensis*)
UTILIZANDO MÉTODOS DE APRENDIZAGEM DE MÁQUINA E DADOS
ESPECTRAIS**

**VIÇOSA – MINAS GERAIS
2022**

RAYANNE OLIVEIRA TEIXEIRA

**MONITORAMENTO NUTRICIONAL DE PALMA DE ÓLEO (*Elaeis guineensis*)
UTILIZANDO MÉTODOS DE APRENDIZAGEM DE MÁQUINA E DADOS
ESPECTRAIS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Doctor Scientiae*.

Orientador: Prof. Dr. Júlio César Lima Neves

Coorientador: Prof. Dr. Elpídio Inácio
Fernandes Filho

**VIÇOSA – MINAS GERAIS
2022**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

T266m
2022
Teixeira, Rayanne Oliveira, 1990-
Monitoramento nutricional de palma de óleo (*Elaeis guianenses*) utilizando métodos de aprendizagem de máquina e dados espectrais / Rayanne Oliveira Teixeira. – Viçosa, MG, 2022.

1 tese eletrônica (111 f.): il. (algumas color.).

Orientador: Júlio César Lima Neves.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Solos, 2022.

Referências bibliográficas: f. 106-111.

DOI: <https://doi.org/10.47328/ufvbbt.2024.108>

Modo de acesso: World Wide Web.

1. *Elaeis guianenses*. 2. Dendzeiro - Nutrição.
3. Dendzeiro - Produtividade. 4. Análise foliar. 5. Amazônia.
6. Solos - Análise. 7. Aprendizado do computador.
8. Algoritmos. I. Neves, Júlio César Lima, 1955-.
II. Universidade Federal de Viçosa. Departamento de Solos.
Programa de Pós-Graduação em Solos e Nutrição de Plantas.
III. Título.

CDD 22. ed. 633.851


RAYANNE OLIVEIRA TEIXEIRA

**MONITORAMENTO NUTRICIONAL DE PALMA DE ÓLEO (*Elaeis guineensis*)
UTILIZANDO MÉTODOS DE APRENDIZAGEM DE MÁQUINA E DADOS
ESPECTRAIS**


Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Doctor Scientiae*.

APROVADA: 18 de maio de 2022.

Assentimento:

Documento assinado digitalmente
 RAYANNE OLIVEIRA TEIXEIRA
Data: 07/03/2024 18:45:50-0300
Verifique em <https://validar.iti.gov.br>

Rayanne Oliveira Teixeira
Autor

Documento assinado digitalmente
 JULIO CESAR LIMA NEVES
Data: 07/03/2024 19:59:59-0300
Verifique em <https://validar.iti.gov.br>

Júlio César Lima Neves
Orientador

AGRADECIMENTOS

A Deus por todas as oportunidades que me concede, força e proteção.

À Universidade Federal de Viçosa e ao Departamento de Solos, pela oportunidade de realização do curso.

À Empresa Agropalma, por acreditar e investir na nossa pesquisa e disponibilizar o banco de dados, tornando possível a realização deste trabalho.

Ao meu orientador, professor Júlio César Lima Neves, pelas oportunidades, paciência, ensinamentos e conselhos.

Ao meu coorientador, professor Elpídio Inácio Fernandes Filho, pelo auxílio imprescindível, e ensinamentos que me desperta cada vez mais o interesse nesta linha de pesquisa.

A banca examinadora pelas ricas considerações: Hélio Garcia Leite, João Carlos Ker, Lafayete Gonçalves Campelo Martins.

Ao Laboratório de Geoprocessamento do Departamento de Solos (LabGeo) que foi fundamental para o processamento e análise dos dados e a todos os colegas do LabGeo, pela acolhida e partilha de conhecimento.

Aos professores do Departamento de Solos por todos os ensinamentos adquiridos nas disciplinas.

Aos funcionários do Departamento de Solos, sempre dispostos a nos ajudar.

Aos meus amigos que desde o início me acompanham nesta jornada: Marcel Thomas, Nathalie Sena, Francis Henrique, Prímula Campos, Jaqueline Alexandre, Klever Silveira, Fernanda Zeidan, Raiza Moniz e a todos os colegas da pós-graduação. Agradeço especialmente a minha amiga Patricia Barros, foi muito especial viver a pós-graduação com você. Obrigada por tornar tudo mais leve.

Aos meus estagiários/amigos por toda troca de conhecimento que tivemos e por contribuírem de forma muito positiva no processamento dos dados: Bernardo Sousa e André Fernandes.

Ao Doutor Gustavo Velozo (Ganso), pelas ricas contribuições e ensinamentos em modelagem.

À minha família, minha fonte inesgotável de amor e força, em especial minha mãe Eliene Oliveira e meu irmão Rafael Oliveira.

A todos que contribuíram com essa importante etapa da minha vida.

Muito obrigada!

RESUMO

TEIXEIRA, Rayanne Oliveira, D.Sc., Universidade Federal de Viçosa, maio de 2022. **Monitoramento Nutricional de Palma de Óleo (*Elaeis Guineensis*) Utilizando Métodos de Aprendizagem de Máquina e Dados Espectrais.** Orientador: Júlio César Lima Neves. Coorientador: Elpídio Inácio Fernandes Filho.

A palma de óleo (*Elaeis guineenses*), é uma oleaginosa que fornece dois principais produtos, o óleo de palma e o óleo de palmiste, extraídos da polpa do fruto e semente, respectivamente. Os países Indonésia e Malásia, são os principais produtores mundiais, e o Brasil ocupa o 10º lugar do ranking, com grande maciço concentrados no Estado do Pará. Monitorar a condição nutricional destas lavouras permite ter mais informações sobre as taxas de extração e remobilização de nutrientes e conseqüentemente melhor compreensão sobre a nutrição e sua relação com as produtividades de cachos de frutos. O fato de muitos materiais genéticos terem produções mensais, exige uma forma de monitoramento nutricional que seja mais rápido, assertivo e com informações capazes de subsidiar as estratégias de manejo de fertilização da palma de óleo. Neste contexto, foi proposto utilizar ferramentas de programação e imagens do satélite Sentinel 2A, para compreender o comportamento espectral da cultura através dos índices de vegetação e relacionar estas variáveis com os resultados de tecido foliar. Com base neste banco de dados, foi possível calcular diversos índices de vegetação, treinar e validar diferentes algoritmos capazes de prever a condição nutricional de macronutrientes, micronutrientes e produtividade dos plantios de forma eficaz. Além do tradicional NDVI, foi possível selecionar índices que apresentaram melhor performance para avaliação nutricional da palma, destacando-se o GRNDVI, MSAVI, ARI, VARI e NDWI. Quando se trata de produtividade, os índices Red_edge_NDVI, RGVI, GRNDI e LAI foram os que melhor se ajustaram na base de dados deste estudo. Os algoritmos, Cubist, Ranger e Random Forest mostram-se eficientes para prever a produtividade. Os resultados obtidos neste trabalho evidenciam o grande potencial desta ferramenta para monitoramento dos plantios. Quanto maior o detalhamento e estratificação das informações por material genético e idade, a tendência é obter melhores ajustes dos modelos e seleção dos índices cada vez mais apropriados para monitorar a condição nutricional e produtividade dos plantios de palma de óleo.

Palavras-chave: Machine learning; Diagnose foliar; Amazonia; Índices de vegetação.

ABSTRACT

TEIXEIRA, Rayanne Oliveira, D.Sc., Federal University of Viçosa, May 202. **Nutritional Monitoring of Oil Palm (*Elaeis guineensis*) Using Machine Learning Methods and Spectral Data**. Advisor: Júlio César Lima Neves. Co-advisor: Elpídio Inácio Fernandes Filho.

Oil palm (*Elaeis guineensis*) is an oleaginous plant that provides two main products, palm oil and palm kernel oil, extracted from the pulp of the fruit and seed, respectively. The countries Indonesia and Malaysia are the main world producers, and Brazil occupies the 10th place in the ranking, with plantations concentrated in the State of Pará. Monitoring the nutritional condition of these crops allows for more information on the rates of extraction and remobilization of nutrients and, consequently, a better understanding of nutrition and its relationship with the productivity of fruit bunches. The fact that many genetic materials have monthly productions requires a form of nutritional monitoring that is faster, more assertive and with information capable of subsidizing oil palm fertilization management strategies. In this context, it was proposed to use programming tools and Sentinel 2A satellite images, to understand the spectral behavior of the crop through vegetation indices and relate these variables with the results of leaf tissue. Based on this database, it was possible to calculate different vegetation indices, train and validate different algorithms capable of predicting the nutritional condition of macronutrients, micronutrients and productivity of plantations in an efficient way. In addition to the traditional NDVI, it was possible to select indices that presented the best performance for the nutritional assessment of cactus pear, with emphasis on GRNDVI, MSAVI, ARI, VARI and NDWI. When it comes to productivity, the Red_edge_NDVI, RGVI, GRNDI and LAI indices were the best fit in the database of this study. The Cubist, Ranger and Random Forest algorithms are efficient to predict productivity. The results obtained in this work show the great potential of this tool for monitoring plantations. The greater the detailing and stratification of information by genetic material and age, the tendency is to obtain better adjustments of the models and selection of increasingly appropriate indices to monitor the nutritional condition and productivity of oil palm plantations.

Keywords: Machine learning; Leaf diagnosis; Amazonia; Vegetation indices.

LISTA DE ILUSTRAÇÕES

Quadro 1 - Descrição dos índices de vegetação utilizados neste estudo.	20
Figura 1 - Mapa de localização dos departamentos da empresa Agropalma.....	16
Figura 2 - Fluxograma metodológico utilizado na modelagem por aprendizado de máquina para seleção dos melhores algoritmos para prever a condição nutricional e produtividade de palma de óleo.	37
Figura 3 - Estimativa dos teores de Nitrogênio para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 4 - Valores de R ² validação dos algoritmos utilizados na predição de teor foliar de Nitrogênio na cultura palma de óleo.	41
Figura 5 - Estimativa dos teores de Fósforo para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 6 - Valores de R ² validação dos algoritmos utilizados na predição de teor foliar de Fósforo na cultura palma de óleo.	43
Figura 7 - Estimativa dos teores de Potássio para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 8 - Valores de R ² validação dos algoritmos utilizados na predição de teor foliar de Potássio na cultura palma de óleo.....	45
Figura 9 - Estimativa dos teores de Cálcio para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 10 - Valores de R ² validação dos algoritmos utilizados na predição de teor foliar de Cálcio na cultura palma de óleo.	47
Figura 11 - Estimativa dos teores de Magnésio para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 12 - Valores de R ² validação dos algoritmos utilizados na predição de teor foliar de Magnésio na cultura palma de óleo.	49

Figura 13 - Estimativa dos teores de Enxofre para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 14 - Valores de R ² validação dos algoritmos utilizados na predição de teor foliar de Enxofre na cultura palma de óleo.	51
Figura 15 - Valores de R ² validação dos algoritmos utilizados na predição de teor foliar dos micronutrientes na cultura palma de óleo.	52
Figura 16 - Estimativa dos teores de Cobre para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 17 - Estimativa dos teores de Ferro para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 18 - Estimativa dos teores de Zinco para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 19 - Estimativa dos teores de Manganês para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	Erro! Indicador não definido.
Figura 20 - Estimativa dos teores de Boro para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	57
Figura 21 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento nitrogênio, modelo <i>Cubist</i>	68
Figura 22 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento nitrogênio, modelo <i>Ranger</i>	69
Figura 23 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento nitrogênio, modelo <i>Earth</i>	70
Figura 24 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento nitrogênio, modelo <i>Random Forest</i>	71
Figura 25 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento fósforo, modelo GBM.	72
Figura 26 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento fósforo, modelo <i>Random Forest</i>	73

Figura 27 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento potássio, modelo <i>Cubist</i>	74
Figura 28 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento potássio, modelo <i>Ranger</i>	75
Figura 29 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento potássio, modelo <i>Random Forest</i>	76
Figura 30 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cálcio, modelo <i>Cubist</i>	77
Figura 31 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cálcio, modelo <i>Ranger</i>	78
Figura 32 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento magnésio, modelo <i>Cubist</i>	79
Figura 33 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento magnésio, modelo <i>Random Forest</i>	80
Figura 34 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Enxofre, modelo <i>Cubist</i>	81
Figura 35 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Enxofre, modelo <i>Linear Model</i>	82
Figura 36 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cobre, modelo GBM.	83
Figura 37 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cobre, modelo <i>Random Forest</i>	84
Figura 38 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cobre, modelo <i>Earth</i>	85
Figura 39 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento ferro, modelo <i>Cubist</i>	86
Figura 40 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento ferro, modelo <i>Linear Model</i>	87
Figura 41- Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento manganês, modelo <i>Cubist</i>	88
Figura 42 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento manganês, modelo <i>Random Forest</i>	89
Figura 43 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento zinco, modelo <i>Earth</i>	90

Figura 44 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Boro, modelo <i>Random Forest</i>	91
Figura 45 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Boro, modelo <i>Linear Model</i>	92
Figura 46 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Boro, modelo <i>Earth</i>	93
Figura 47 - Estimativa de produtividade nas idades de 15 a 25 anos para a cultura da palma de óleo, de acordo com os algoritmos <i>GradientBoostingMachine</i> (GBM), <i>Random Forest</i> (RF), <i>Ranger</i> , Regressão Linear (lm), <i>Cubist</i> e <i>Earth</i>	95
Figura 48 - Boxplot dos valores de R ² validação dos algoritmos utilizados na predição de produtividade de cachos de frutos frescos (CFF) na cultura palma de óleo.	96
Figura 49 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo <i>Cubist</i>	99
Figura 50 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo <i>Earth</i>	100
Figura 51 - - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo GBM.	101
Figura 52 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo <i>Linear Model</i>	102
Figura 53 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo <i>Ranger</i>	103
Figura 54 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo <i>Random Forest</i>	104

LISTA DE TABELAS

Tabela 1 - Especificações das Imagens Sentinel-2.....	18
Tabela 2 - Pacotes de programação utilizados no programa estatístico R para processamento dos dados.	19
Tabela 3 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de nitrogênio foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	59
Tabela 4 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de fósforo foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	60
Tabela 5 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de potássio foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	60
Tabela 6 - - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de cálcio foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	61
Tabela 7 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de enxofre foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	62
Tabela 8 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de zinco foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	63
Tabela 9 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de cobre foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	63
Tabela 10 - - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de ferro foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	64
Tabela 11 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de manganês foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	65
Tabela 12 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de boro foliar em palma de óleo e seus respectivos valores de R ² e RMSE de treinamento e validação.	65

Tabela 13 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição de produtividade em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação. 94

SUMÁRIO

1 INTRODUÇÃO	13
2 OBJETIVOS..	15
3 MATERIAL E MÉTODOS	16
3.1 Caracterização da Área de Estudo e Organização do Banco de Dados	16
3.2. Processamento das Imagens Espectrais e Cálculos dos Índices de Vegetação.....	17
3.3 Treinamento e Validação	30
3.3.1 <i>Cubist</i>	30
3.3.2 <i>Earth</i>	31
3.3.3 <i>Linear Model (LM)</i>	32
3.3.4 <i>Ranger</i>	32
3.3.5 <i>Random Forest (RF)</i>	34
3.3.6 <i>Gradient Boosting Machine (GBM)</i>	35
3.3.7. <i>Support Vector Machine Linear (SVMLinear)</i>	Erro! Indicador não definido.
4 RESULTADOS E DISCUSSÃO	38
4.1 Monitoramento Nutricional.....	38
4.2 Produtividade.....	94
5 CONCLUSÃO.....	105
REFERÊNCIAS	106

1 INTRODUÇÃO

A palma de óleo (*Elaeis guianenses* Jacq), também conhecida como dendê ou dendezeiro, é uma palmeira pertencente à família Arecaceae, nativa da Costa Ocidental da África, foi introduzida no Brasil no século XVII apresentando excelente adaptação no litoral da Bahia. Se destaca entre as oleaginosas por produzir dois tipos de óleos - óleo de palma - retirado da polpa, utilizado na indústria de alimentos. E - óleo de palmiste - extraído da amêndoa e aplicada a linha de cosméticos. Possuem características químicas diferentes, que podem ser utilizados inclusive para obtenção de biocombustíveis.

No ano de 2022, a produção mundial de óleo de palma foi de 77.215 Mt (megatoneladas), sendo o continente asiático o principal fornecedor, com os seguintes percentuais de produção: Indonésia - 59%, Malásia - 24% e Tailândia- 4 %. O Brasil ocupou o 10º lugar no ranking mundial com participação de 1%, correspondente a 570 Mt, sendo que 82% deste valor foram produzidos no Estado do Pará e 16% no Estado da Bahia, de acordo com os dados da Foreign Agriculture Service, USDA (2023).

A produção do Pará, advém dos mais de 200 mil hectares de plantios com palma de óleo, concentrados principalmente no nordeste paraense, o qual a produtividade média fica em torno de 3,0 t ha⁻¹ ano⁻¹. Porém, de acordo com a Secretária do Estado do Pará de Desenvolvimento Agropecuário e da Pesca (SEDAP, 2020), esta produção não supri a demanda interna e reflete no cenário nacional, levando o país a importação.

No entanto, é importante destacar que a Amazônia oriental tem grande potencial para expandir os plantios nas áreas já desmatadas. As pesquisas que já foram e continuam sendo desenvolvidas, são importantes norteadores no momento de considerar as diferentes condições de solos, clima, performance dos materiais genéticos em campo, definições das estratégias de manejo, exportação de nutrientes pela planta, épocas de produção dos frutos que influenciam direta e indiretamente na produtividade da lavoura.

Há grandes lacunas de conhecimento sobre a cultura que precisam ser mais bem compreendidos, mas no geral, a partir do terceiro ano de idade iniciam as produções de frutos na palma de óleo, algumas variedades são rentáveis até o 25º ano e essa intensa formação de cachos de frutos ao longo do ciclo econômico produtivo exige grande demanda por absorção de nutrientes. Em muitos casos, este rendimento pode ser comprometido devido ao inadequado balanço nutricional, reforçando a necessidade de acompanhamento frequente da condição química dos plantios para corrigir os desvios nutricionais que podem limitar a produtividade. Atualmente, o método mais usual é o monitoramento através da coleta de tecido foliar para

posterior análise em laboratório, entretanto, o custo é elevado e o tempo para realizar as coletas e obter os resultados é demorado.

Neste contexto, as imagens de satélite são cada vez mais amplamente utilizadas no monitoramento de atividades agrícolas, devido a rapidez de acesso as informações, inclusive nas áreas de difícil acesso. Quanto mais conhecimento sobre a lavoura, melhor a condição para identificar as regiões com maior e menor grau de estresses nutricionais, detectar materiais genéticos mais responsivos a fertilização ou mais sensíveis às intempéries climáticas, identificar as zonas da propriedade com maior potencial produtivo, entre outros indicadores que podem ser extraídas da imagem espectral/índices de vegetação para recomendar quantidade e onde realizar as fertilizações e ter a expectativa de produtividade para alinhamentos estratégicos de manejo, colheita, preços dos produtos, ou seja, é um importante aliado para subsidiar as tomadas de decisões. Entre as diversas aplicações do sensoriamento remoto, pode-se destacar a forma rápida e eficiente na análise indireta da condição foliar através das características espectrais das folhas (GÖKKAYA *et al.*, 2015; OLIVEIRA *et al.*, 2017).

Diversos estudos vêm sendo desenvolvidos em culturas agrícolas com o uso de imagens de satélite (PEDRALI *et al.*, 2022; OSCO *et al.*, 2020; OLIVEIRA *et al.*, 2019), mas para a cultura da palma de óleo ainda há uma grande lacuna, principalmente quando se trata de programação de modelos preditivos para monitoramento nutricional e produtividade. Pesquisas realizadas por Guzmán (2014), obteve as normas gerais e específicas de interpretação dos teores nutricionais para plantios de dendê no Pará. Sousa (2020) definiu os níveis críticos e ótimos de nutrientes foliares para produção de palma de óleo no Pará. Teixeira (2018), complementou os trabalhos já realizados no mesmo campo de estudos e relacionou as faixas de teores nutricionais com as faixas do índice de vegetação (IV) NDVI (Normalized Difference Vegetation Index) através do método da chance matemática relativa, para estimar os valores de macronutrientes e micronutrientes presentes nas folhas de dendê. É notado que para estudar-se e aplicar-se metodologias de sensoriamento remoto é importante ter informações consistentes para calibrar e aplicar modelos. Desta forma, o processamento tende a ter melhores resultados quando unisse os dados e incorporasse na análise as condições reais da área de estudo.

Os índices são equações matemáticas desenvolvidas com base no comprimento de onda do espectro eletromagnético e permitem a avaliação do comportamento espectral da vegetação, muitos foram melhorados para detectar com maior nível de detalhes os teores de clorofila, antocianinas, carotenoides, teor de água, entre outros compostos presentes nas folhas. No processo de aprendizagem de máquina esses resultados são relacionados com os teores nutricionais e aplica-se diferentes algoritmos para selecionar apenas os modelos que estão mais

correlacionados entre si para prever o estado nutricional e produtividade do plantio. Estudos como este são de grande aplicabilidade e importância na gestão de informações, redução de custos e estratégias de manejo.

2 OBJETIVOS

Analisar o potencial de uso das imagens do satélite Sentinel 2A para monitoramento nutricional e produtividade dos plantios de palma de óleo;

Determinar qual ou quais os índices de vegetação são mais adequados para monitorar os teores foliares de macronutrientes, micronutrientes e produtividade nos plantios de palma de óleo;

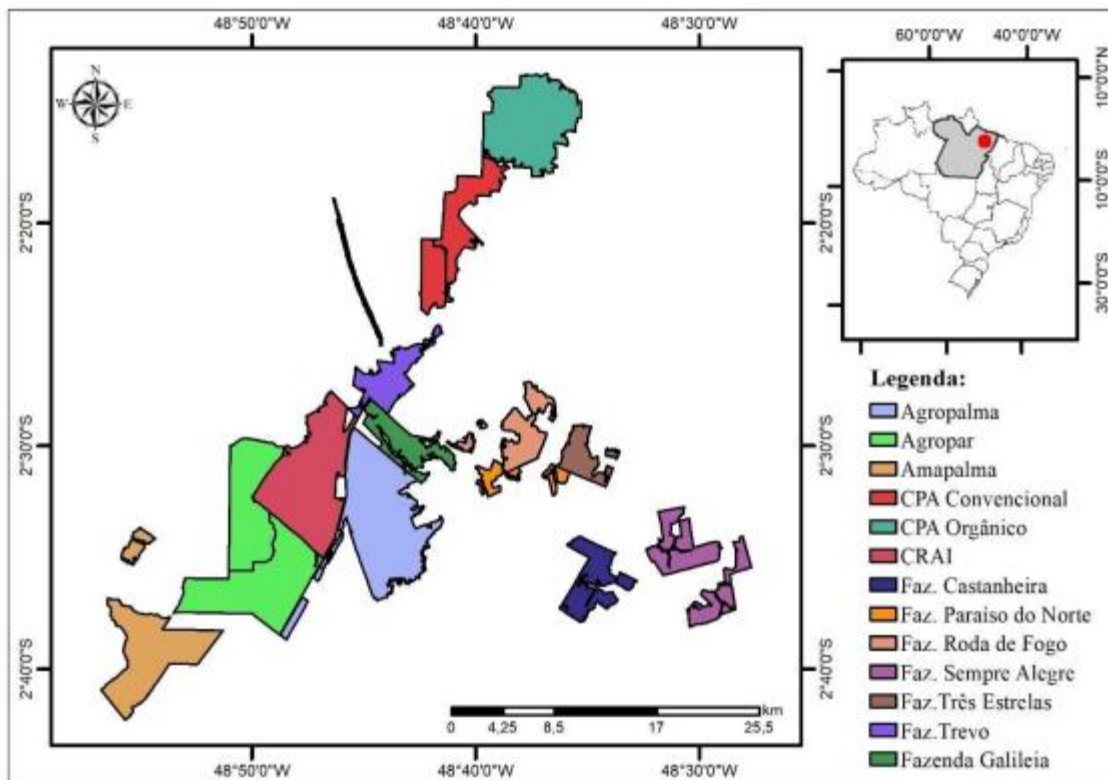
Determinar qual ou quais algoritmos possuem maior acurácia para prever o estado nutricional e produtividade dos plantios de palma de óleo.

3 MATERIAL E MÉTODOS

3.1 Caracterização da Área de Estudo e Organização do Banco de Dados

A pesquisa foi desenvolvida a partir do banco de dados dos plantios de palma de óleo (*Elaeis guineenses* Jacq.), cultivados em áreas de propriedade do Grupo Agropalma, localizada no município de Tailândia, Estado do Pará. A Empresa possui 107 mil hectares de terras, sendo 64 mil hectares destinados a reservas florestais e aproximadamente 42 mil hectares de plantios comerciais, representados por oito departamentos (Figura 1).

Figura 1 - Mapa de localização dos departamentos da empresa Agropalma.



Fonte: Elaborado pela autora, 2022.

Segundo a classificação de Köppen e Geiger, o clima é considerado Am. A cidade de Tailândia tem um histórico de temperatura em torno de 27.0 °C, com média anual de pluviosidade de 2.399 mm (INPE, 2022). Nos dados históricos de precipitação, o mês de fevereiro destaca-se como o mais chuvoso, apresentando média de 435 mm, em contrapartida

o mês de agosto é o mais seco, acumulando média de 37 mm. Ao longo dos meses as temperaturas médias variam 0.9 °C, setembro sobressai como o mês mais quente.

A altitude média da região é de 50 metros. O material de origem é de sedimentos argilosos do Grupo Barreiras, as áreas são planas e o solo dominante na área da empresa são os Argissolos Amarelos distrocoesos abrupticos, com dureza muito acentuada já no topo do horizonte B, que, entretanto, não se manifesta de forma mais clara em razão da elevada precipitação pluviométrica que mantém o solo úmido na maior parte de ano (KER & NEVES, 2010).

Atualmente estão plantados 17 materiais genéticos, a densidade de plantio nas parcelas são 131, 143 (a mais frequente) e 160 plantas/ha, com espaçamento de 7 a 7,8 m entre linhas e 9 m entre plantas.

O banco de dados foi organizado considerando a série anual de 2015 a 2020, na idade de 3 a 32 anos, em que foram selecionadas as seguintes informações: nome dos departamentos, identificação das parcelas de cada departamento, material genético, ano de plantio, ano da colheita, mês da colheita, área da parcela, produção de cachos de frutos frescos (CFF) por mês. Foram considerados a produção nos 12 meses e calculada a produtividade anual (CFF em t/ha/ano) de cada parcela monitorada. Com todas as informações contidas no banco de dados, realizou-se a análise de consistência para eliminar dados discrepantes, parcelas sem identificações e áreas menores que um (1) hectare. Em seguida, foram criados comandos no programa estatístico R, que agrupavam todas as informações da mesma parcela e somavam os valores de produção de CFF (t/ha) de cada material genético para cada ano.

3.2. Processamento das Imagens Espectrais e Cálculos dos Índices de Vegetação

Inicialmente foram identificadas as datas de coletas, os pontos de amostragem de tecido foliar e solos ao longo dos anos, com base nessas informações selecionou-se as imagens pelo site do Serviço Geológico dos Estados Unidos (USGS), obtidas de forma gratuita.

Foram selecionadas imagens do satélite Sentinel 2A o mais próximo possível aos dias de coleta das amostras. Através da base cartográfica disponibilizado pela empresa, foi visualizado no *Google Earth* as delimitações de reserva legal (RL) e áreas de preservação permanente (APP), para serem diferenciadas das parcelas de plantios comerciais e não ocorrer influências nos posteriores cálculos dos índices de vegetação.

O Sentinel 2A executa uma missão multiespectral imageadora administrada pela Agência Espacial Europeia para observação da terra, realizando coleta de dados sobre

vegetação, solos, rios e áreas costeiras. Possui câmera multiespectral regular, com resolução espacial de 10 metros nas bandas azul, verde, vermelho e infravermelho próximo (IVP), e resolução espacial de 20 metros em quatro bandas centradas no intervalo do vermelho limítrofe (Tabela 1).

O satélite apresenta alta capacidade de revisita, a cada 5 dias, com resolução radiométrica de 12 bits por pixel e largura da faixa imageadora de 290 km. Os dados obtidos estavam no Level-1C devido a disponibilidade de cenas para a região de estudo, desta forma as refletâncias obtidas estavam calibradas no topo da atmosfera (TOA), para diminuir os efeitos de absorção e espalhamento da coluna de ar entre o sensor e o alvo. Porém, a presença da atmosfera entre o satélite e a superfície terrestre pode ocasionar efeitos indesejados nas imagens espectrais. Portanto, visando diminuir tais incoerências de valores, as imagens foram processadas para o Level-2A, realizou-se a calibração na base da superfície terrestre (BOA) e retirou-se os efeitos dos gases da atmosfera, com auxílio do *Software Sentinel Application Platform* (SNAP) desenvolvido pela Agência Espacial Europeia (ESA).

Tabela 1 - Especificações das Imagens Sentinel-2

Resolução Espacial (m)	Número de Bandas	Nome da Banda	Comprimento de Onda (nm)
10	B02	Blue	490
	B03	Green	560
	B04	Red	665
	B08	NIR	842
20	B05	Red Edge 1	705
	B06	Red Edge 2	740
	B07	Red Edge 3	783
	B08A	Red Edge 4	865
	B11	SWIR 1	1610
	B12	SWIR 2	2190
60	B01	Aerossol	443
	B09	Water Vapor	940
	B10	Cirrus	1375

Fonte: Elaborado pela autora, 2022.

Após o processamento, todas as imagens foram analisadas no programa estatístico R (Tabela 2), delimitando-se as áreas de interesse, recortando todas as cenas utilizando o *shapefile* da empresa e realizando a reamostragem de *pixel* para colocar todas as bandas na

mesma resolução espacial e com posterior cálculo dos índices de vegetação para cada ano estudado. Em seguida, criou-se uma máscara específica para cada ano de coleta utilizando-se como base o índice NDVI, onde valores menores que 0.4 foram removidos visando diminuir os efeitos do solo exposto. Esta máscara foi utilizada como referência para eliminar os *pixels* que pudessem contaminar os demais índices de vegetação.

Para evitar os efeitos de bordas das parcelas, utilizou-se a ferramenta *Buffer* do *Software* QGis 3.24.0 para reduzir em 20 metros as extremidades e eliminar os *pixels* externos que poderiam causar ruídos nos resultados. Com base nos números de *pixel* dos *raster* (bandas e índices de vegetação), foram obtidas as estatísticas das imagens. Os índices de vegetação considerados nas análises estão contidos no Quadro 1.

Tabela 2 - Pacotes de programação utilizados no programa estatístico R para processamento dos dados.

	Gráficos de Correlação	BoxPlot	Processamento Modelos	Cálculos dos Índices de Vegetação	Extração de Informações dos Rasters (Índices Bandas e Covariáveis)
Pacote no Software R	corrplot	Readr	Lattice	raster	Sf
	RColorBrewer	Dplyr	Labgeo	terra	Dplyr
	dplyr	Tidyr	Caret	RColorBrewer	Terra
	labgeo	Purrr	ggplot2	dplyr	Tidyverse
		Plyr	Dplyr	doParallel	Readr
		ggplot2	Tydir	labgeo	Labgeo
		ggpubr			

Fonte: Elaborado pela autora, 2022

Quadro 1 - Descrição dos índices de vegetação utilizados neste estudo.

Índice	Abreviação	Equação	Observações sobre os índices	Referência
<i>Anthocyanin Reflectance Index</i>	ARI 1	$\left(\frac{1}{G}\right) - \left(\frac{1}{RedEdge1}\right)$	Este índice usa medições de refletância no espectro visível para aproveitar as assinaturas de absorção de pigmentos relacionados ao estresse. As antocianinas são pigmentos solúveis em água abundantes em folhas recém-formadas e naquelas em senescência. A vegetação enfraquecida contém maiores concentrações de antocianinas, então esse índice é uma medida de vegetação estressada. Aumentos em ARI1 indicam mudanças de dossel na folhagem por meio de novo crescimento ou morte.	Gitelson <i>et al.</i> (2001)
<i>Anthocyanin Reflectance Index</i>	ARI 2	$NIR * \left(\frac{1}{G}\right) - \left(\frac{1}{RedEdge1}\right)$	Este índice é uma modificação do ARI1 que detecta maiores concentrações de antocianinas na vegetação.	Gitelson <i>et al.</i> (2001)
<i>Atmospherically Resistant Vegetation Index</i>	ARVI	$\frac{NIR - (2 * R - B)}{NIR + (2 * R - B)}$	A resistência do ARVI aos efeitos atmosféricos (em comparação com o NDVI) é realizada por um processo de autocorreção para o efeito atmosférico no canal vermelho, usando a diferença na radiância entre os canais azul e vermelho para corrigir a radiância em o canal vermelho. Simulações usando cálculos de transferência radiativa em espectros de superfície aritmética e natural, para várias condições atmosféricas, mostram que o ARVI tem uma faixa dinâmica semelhante ao NDVI, mas é, em média, quatro vezes menos sensível aos efeitos atmosféricos do que o NDVI.	Kaufman e Tanre (1992)
<i>Burn Area Index</i>	BAI	$\frac{1}{(0.1 - R)^2 + (0.06 - NIR)^2}$	Índice elaborado para destacar a resposta espectral de alvos carbonizados em imagens pós-fogo. Baseado nas bandas vermelha e infravermelho próximo.	Chuvieco, Martin e Palacios (2002)
<i>Blue Normalized Difference Vegetation Index</i>	BNDVI	$\frac{NIR - B}{NIR + B}$	Índice de vegetação utilizado para estimar índice de área foliar. Ele utiliza a banda do infravermelho próximo associada a banda do azul.	Wang Fu-ming <i>et al.</i> (2007)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Chlorophyll Red-Edge</i>	CHL_RED_EDGE	$\frac{RedEdge1}{NIR}$	Este índice é uma variação do NDVI que utiliza a banda de red-edge.	Gitelson <i>et al.</i> (2003)
<i>Canopy Index</i>	CI	$SWIR - G$	Este índice caracteriza a distribuição espacial de folhas de um dossel de vegetação. Determina a transferência de radiação do dossel, fotossíntese e processos hidrológicos.	Vescovo e Gianelle (2008)
<i>Color Index of Vegetation Extraction</i>	CIVE	$0.441 * R - 0.811 * G + 0.385 * B + 18.78745$	Este método foi proposto para separar as plantas verdes do fundo do solo, a fim de avaliar o estado de crescimento das culturas.	Kataoka <i>et al.</i> (2003)
<i>Carotenoid Reflectance Index 1</i>	CRI1	$\left(\frac{1}{B}\right) - \left(\frac{1}{G}\right)$	Os carotenoides funcionam nos processos de absorção de luz nas plantas, bem como na proteção das plantas dos efeitos nocivos do excesso de luz. A vegetação enfraquecida contém maiores concentrações de carotenoides, então esse índice é uma medida de vegetação estressada. Valores mais altos de CRI1 significam maior concentração de carotenoides em relação à clorofila	Gitelson <i>et al.</i> (2002)
<i>Carotenoid Reflectance Index 2</i>	CRI2	$\left(\frac{1}{B}\right) - \left(\frac{1}{RedEdge1}\right)$	Este índice é uma modificação do CRI1 que proporciona melhores resultados em áreas de alta concentração de carotenoides. Valores mais altos de CRI2 significam maior concentração de carotenoides em relação à clorofila	Gitelson <i>et al.</i> (2002)
<i>Difference Vegetation Index</i>	DVI	$NIR - R$	Índice obtido pela diferença entre a refletância do vermelho e a refletância do infravermelho próximo. É mais simples que o NDVI, mas é propenso a erros de medição nas bandas de infravermelho próximo (NIR) e vermelho (R) porque não é normalizado por sua soma.	Tucker (1980)
<i>Enhanced Red-Green-Blue Vegetation Index</i>	ERGBVE	$\pi * \frac{(G^2 - (R * G))}{(G^2 + (R * G))}$	Este índice é utilizado para identificar a vegetação em nuvens de pontos.	Themistocleous (2019)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Enhanced Vegetation Index</i>	EVI	$\frac{(2.5 * (NIR - R))}{(NIR + 6 * R - 7.5 * B + 1)}$	É um índice de vegetação otimizado que leva em conta as influências atmosféricas e o sinal de fundo da vegetação. Aplicado em regiões de dossel denso onde o índice de área foliar é alto. Desenvolvido para o sensor MODIS.	Jiang <i>et al.</i> (2008)
<i>Enhanced Vegetation Index 2</i>	EVI2	$\frac{(2.5 * (NIR - R))}{(NIR + 2.4 * R + 1)}$	Este índice é sensível a variação na estrutura do dossel, incluindo o índice de área foliar, fisionomia da planta e a arquitetura do dossel. Desenvolvido para o sensor MODIS.	Jiang <i>et al.</i> (2008)
<i>Excess of Green</i>	ExG	$2 * G - R - B$	Índice utilizado para separar automaticamente culturas do solo.	Woebbecke <i>et al.</i> (1995)
<i>Excess Green minus Excess Red</i>	ExGexR	$(2 * G - R - B) - (1.4 * R - G)$	Combinação de dois índices de cores, o Excess Green Index (ExG) e Excess Red Index (ExR). Esses métodos foram aplicados simultaneamente para separar as plantas do solo e do resíduo.	Camargo Neto (2004)
<i>Green-Blue Normalized Difference Vegetation Index</i>	GBNDVI	$\frac{(NIR - (G + B))}{(NIR + (G + B))}$	Este índice corresponde a uma substituição da faixa vermelha do NDVI por uma combinação das faixas verde e azul. Utilizado para a previsão do índice de área foliar da vegetação.	Wang Fu-ming <i>et al.</i> , 2007
<i>Green Chlorophyll Index</i>	GCI	$\frac{NIR}{G} - 1$	Usado para estimar o teor de clorofila nas folhas das plantas com base em bandas de infravermelho próximo e verde. Em geral, o valor de clorofila reflete diretamente a vegetação	Gitelson, Gritz e Merzlyak (2003)
<i>Green Difference Vegetation Index</i>	GDVI	$NIR - G$	Índice semelhante ao NDVI, no entanto, ele mede o espectro verde na faixa de 0,54 a 0,57 microns. É um indicador da atividade fotossintética da cobertura vegetal. E é frequentemente usado na avaliação do teor de umidade e concentração de nitrogênio nas folhas das plantas.	Wu (2014)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Green Leaf Index</i>	GLI	$\frac{2 * G - R - B}{2 * G + R + B}$	Este índice gera uma imagem em tons de cinza determinando se o valor médio do comprimento de onda vermelho e azul, onde os valores da banda são maiores que o valor do comprimento de onda verde, e realiza um processo de normalização para que o valor do pixel da imagem resultante fique entre -1 e 1. Valores negativos geralmente representam solo e áreas sem vegetação, e áreas positivas são vegetação.	Louhaichi, Borman e Johnson (2001)
<i>Green Normalized Vegetation Index</i>	GNDVI	$\frac{(NIR - G)}{(NIR + G)}$	Este índice é mais sensível à variação da clorofila na cultura do que o NDVI e possui um ponto de saturação mais alto. Ele pode ser usado em cultivos com copas densas ou em estágios mais avançados de desenvolvimento.	Gitelson <i>et al.</i> 1996
<i>Green Optimized Soil Adjusted Vegetation Index</i>	GOSAVI	$\frac{(NIR - G)}{(NIR + G + 0.16)}$	Este índice foi baseado no índice de vegetação ajustado ao solo otimizado (OSAVI), que foi uma melhoria do índice de vegetação ajustado ao solo (SAVI) para minimizar a influência do fundo do solo.	Rondeaux <i>et al.</i> 1996
<i>Green and Red Normalized Vegetation Index</i>	GRNDVI	$\frac{(NIR - (G + R))}{(NIR + (G + R))}$	Este índice corresponde a uma substituição da faixa vermelha do NDVI por uma combinação das faixas azul e vermelho. Utilizado para a previsão do índice de área foliar da vegetação.	Wang Fu-ming <i>et al.</i> 2007
<i>Green-Red Vegetation Index</i>	GRVI	$\frac{(G - R)}{(G + R)}$	Este índice é usado como um indicador de fenologia e tem potencial para estimativa de biomassa. Ele explora a alta refletância das plantas na banda do verde e a absorção na parte vermelha e azul do espectro visível através de clorofilas de plantas.	Tucker, 1979
				Falkowski <i>et al.</i> 2005
<i>InfraRed Percentage Vegetation Index</i>	IPVI	$\frac{(NIR)}{(R + NIR)}$	Este índice atua da mesma forma que o NDVI, mas é computacionalmente mais rápido. Os valores variam de 0 a 1.	Crippen (1990)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Inverted Red-Edge Chlorophyll Index</i>	IRECI	$(RedEdge3 - R) * \frac{RedEdge2}{RedEdge1}$	Este índice usa a refletância em quatro bandas para estimar o conteúdo de clorofila do dossel. Neste índice de vegetação, borda vermelha tem sido usada como um indicador de senescência e estresse da vegetação.	Baret e Guyot (1991)
<i>Leaf Area Index</i>	LAI	$3.618 * \frac{(2.5 * (NIR - R))}{(NIR + 6 * R - 7.5 * B + 1)} - 0.118$	Este índice é projetado para analisar a superfície das folhas e estima a quantidade de folhas em uma região específica. LAI é uma medida sem unidade que é calculada como a razão entre a área de folhagem unilateral (iluminada) e a superfície do solo que ela pode cobrir. Este índice de vegetação é importante para monitorar a saúde das culturas e florestas, o meio ambiente e as condições climáticas.	Boegh <i>et al.</i> (2002)
<i>Moisture Adjusted Vegetation Index</i>	MAVI	$\frac{NIR - R}{NIR + R + SWIR}$	Índice de vegetação ajustado à umidade, proposto usando a refletância do vermelho, infravermelho próximo e infravermelho de ondas curtas (SWIR)	Zhu <i>et al.</i> 2014
<i>Modified Chlorophyll Absorption in Reflectance Index 1</i>	MCARI1	$1.2 * [2.5 * (NIR - R) - 1.3 * (NIR - G)]$	Avalia a profundidade da banda de absorção da clorofila em 670nm por meio de uma relação matemática formulada a partir de valores de reflectância obtidos em 550nm e 700nm.	Daughtry <i>et al.</i> 2000
<i>Modified Chlorophyll Absorption in Reflectance Index 2</i>	MCARI2	$1.5 * \frac{[2.5 * (NIR - R) - 1.3 * (NIR - G)]}{[(2 * NIR - 1)^2 - (6 * NIR - 5 * R0.5) - 0.5]0.5}$	Este índice é uma versão otimizada do MCARI que apresenta maior sensibilidade ao índice de área foliar verde.	Zarco <i>et al.</i> 2001
<i>Modified Green Red Vegetation Index</i>	MGRVI	$\frac{G^2 - R^2}{G^2 + R^2}$	Este índice é uma modificação do GRVI, e é definido como a diferença normalizada da refletância verde quadrada e a refletância vermelha quadrada.	Bendig <i>et al.</i> (2015)
<i>Modified Photochemical Reflectance Index</i>	MPRI	$\frac{G - R}{G + R}$	Este índice foi desenvolvido dada a razão entre os comprimentos de onda do verde e vermelho. Ele é capaz de refletir a eficiência da atividade fotossintetizante da planta dada a sua sensibilidade a alterações no ciclo das xantofilas.	Yang <i>et al.</i> (2008)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Modified Soil Adjusted Vegetation Index 1</i>	MSAVI	$\frac{NIR - R}{NIR + R + 0.5} (1 + 0.5)$	Este índice é proveniente de uma modificação do SAVI e procura abordar algumas das limitações do NDVI quando aplicado a áreas com alto grau de superfície do solo exposta.	Qi <i>et al.</i> (1994)
<i>Modified Soil Adjusted Vegetation Index 2</i>	MSAVI2	$(NIR + 1) - 0.5 * \sqrt{((2 * NIR -)^2 + 8 * R)}$	A maior diferença entre o índice SAVI e o MSAVI2, é que no índice SAVI necessita-se de um conhecimento prévio da vegetação para a definição do valor do fator L, enquanto que no MSAVI2 esse fator é auto ajustável.	Qi <i>et al.</i> (1994)
<i>Modified Simple Ratio</i>	MSR	$\frac{\frac{NIR}{R - 1}}{\sqrt{\frac{NIR}{R} + 1}}$	Este índice é aplicado na recuperação de parâmetros biofísicos de florestas boreais usando dados de sensoriamento remoto. Baseia-se em uma avaliação de vários índices de vegetação de duas bandas.	Chen (1996)
<i>MERIS Terrestrial Chlorophyll Index</i>	MTCI	$\frac{RedEdge2 - RedEdge1}{RedEdge1 - R}$	Este índice apresenta-se como o mais adequado para a estimativa de teor de clorofila com dados MERIS.	Dash e Curran (2007)
<i>Modified Triangular Vegetation Index 1</i>	MTVI1	$1.2 * [1.2 * (NIR - G) - 2.5 * (R - G)]$	Este índice torna o TVI adequado para estimativas de LAI ao substituir o comprimento de onda de 750 nm por 800 nm, cuja refletância é influenciada por mudanças nas estruturas das folhas e do dossel.	Haboudane <i>et al.</i> (2004)
<i>Modified Triangular Vegetation Index 2</i>	MTVI2	$\frac{1.5 * [1.2 * (NIR - G) - 2.5 * (R - G)]}{\sqrt{[(2 * NIR + 1)^2 - (6 * NIR - 5 \sqrt{R})] - 0.5}}$	Este índice é semelhante ao MTVI, mas é considerado um melhor preditor de LAI verde. É responsável pela assinatura de fundo dos solos, preservando a sensibilidade ao LAI e a resistência à influência da clorofila.	Haboudane <i>et al.</i> (2004)
<i>Normalized Burn Ratio</i>	NBR	$\frac{NIR - SWIR2}{NIR + SWIR2}$	Índice usada para identificar áreas queimadas e fornecer uma medida da gravidade da queimada. É calculado como uma razão entre os valores NIR e SWIR da maneira tradicional.	Key e Benson (2006)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Normalized Canopy Index</i>	NCI	$\frac{SWIR - G}{SWIR + G}$	Índice desenvolvido para estimar as quantidades relativas de nitrogênio contidas no dossel da vegetação. A refletância em 1510 nm é amplamente determinada pela concentração de nitrogênio das folhas, bem como pela biomassa total da folhagem do dossel.	Vescovo e Gianelle (2008)
<i>Normalized Difference Index (Bandas 4 e 5 - Satellite Sentinel)</i>	NDI45	$\frac{(RedEdge1 - R)}{(RedEdge1 + R)}$	Este índice é aplicado para obtenção de informações de área foliar verde e clorofila do dossel. índice otimizado foi calculado a partir de bandas espectrais centradas em torno de 665 e 705 nm.	Frampton <i>et al.</i> (2013)
<i>Normalized Difference Infrared Index (Banda 11)</i>	NDII11	$\frac{(NIR - SWIR1)}{(NIR + SWIR1)}$	Este índice é utilizado para detectar estresse hídrico de plantas. É uma adaptação do NDVI que utiliza a refletância infravermelha de ondas curtas (SWIR).	Hardisky, Klemas e Smart (1983)
<i>Normalized Difference Vegetation Index</i>	NDVI	$\frac{(NIR - R)}{(NIR + R)}$	Este índice permite identificar a presença de vegetação verde sadia. O índice é calculado por meio da diferença normalizada entre a reflectância da banda do infravermelho próximo (NIR) e a reflectância da banda do vermelho e os seus resultados estão compreendidos no intervalo de -1 a 1, tendo os valores próximos de 1 as melhores condições de vegetação.	Rouse <i>et al.</i> (1974)
<i>Normalized Difference Water Index (SWIR1)</i>	NDWI_SWIR1	$\frac{(NIR - SWIR1)}{(NIR + SWIR1)}$	Este índice usa as bandas NIR e SWIR1 (Banda 11) para obter informações do teor de umidade nas plantas e no solo.	Gao (1996)
<i>Normalized Difference Water Index (SWIR2)</i>	NDWI_SWIR2	$\frac{(NIR - SWIR2)}{(NIR + SWIR2)}$	Este índice usa as bandas NIR e SWIR2 (Banda 12) para obter informações do teor de umidade nas plantas e no solo.	Gao (1996)
<i>Normalized Green-Red Difference Index</i>	NGRDI	$\frac{G - R}{G + R}$	Este índice utiliza a razão normalizada da diferença entre a faixa verde e a faixa vermelha para eliminar a influência de diferentes irradiâncias nas características espectrais da vegetação	Tucker (1979)
<i>Non-linear Vegetation Index</i>	NLI	$\frac{(NIR^2 - R)}{(NIR^2 + R)}$	Este índice visa linearizar as relações com parâmetros de superfície que tendem a ser não lineares.	Goel e Quin (1994)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Optimized Soil Adjusted Vegetation Index</i>	OSAVI	$\frac{(NIR - R)}{(NIR + R + 0.5)}$	Este índice é uma versão otimizada do SAVI de formulação simples.	Rondeaux <i>et al.</i> (1996)
<i>Plant Senescence Reflectance Index</i>	PSRI	$\frac{(R - B)}{(RedEdge2)}$	Este índice é sensível as mudanças de reflectância induzidas por senescência.	Merzlyak <i>et al.</i> (1999)
<i>Pigment Specific Simple Ratio</i>	PSSR	$\frac{NIR}{R}$	Este índice têm as relações mais fortes com a concentração por unidade de área de pigmentos individuais na escala do dossel.	Blackburn (1998)
<i>Perpendicular Vegetation Index</i>	PVI	$\sqrt{((0.355 * NIR - 0.149 * R)^2 + (0.355 * R - 0.852 * NIR)^2)}$	Índice projetado para assumir que o espaço perpendicular do pixel da linha do solo depende linearmente da área de vegetação. Ele tenta eliminar a diferença de fundo do solo, aplicável para regiões áridas e semiáridas. Os valores mais altos do índice PVI indicam o efeito de fundo mais claro do solo, em detrimento da vegetação incompleta.	Richardson e Wiegand (1977)
<i>Red-Blue Normalized Vegetation Index</i>	RBNDVI	$\frac{NIR - (R + B)}{NIR + (R + B)}$	Índice de vegetação utilizado para estimar índice de área foliar. Utiliza a banda do infravermelho próximo associada as bandas do verde e azul.	Wang Fu-ming <i>et al.</i> (2007)
<i>Renormalized Difference Vegetation Index</i>	RDVI	$\frac{\sqrt{(NIR - R)}}{(NIR + R)}$	Este índice é uma combinação dos pontos positivos dos índices NDVI e DVI, para baixas e altas cobertas vegetais.	Roujean e Breon (1995)
<i>Red Edge Normalized Difference Vegetation Index / Normalized Difference Red Edge</i>	RED_EDGE_NDVI (NDRE)	$\frac{(NIR - RedEdge1)}{(NIR + RedEdge1)}$	Este índice é uma versão do NDVI que utiliza a banda Red Edge no lugar da banda do vermelho.	Fernández-Manso <i>et al.</i> (2016)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Red-Green Index</i>	RGI	$\frac{R}{G}$	Este índice identifica mudanças na pigmentação de antocianinas.	Gamon e Surfus, 1999 Coops <i>et al.</i> (2006)
<i>Red Green Blue Vegetation Index</i>	RGVBI	$\frac{G^2 - B * R}{G^2 + B * R}$	Índice que estima a partir da altura das plantas a biomassa foliar.	Bendig <i>et al.</i> (2015)
<i>Rice Growth Vegetation Index</i>	RGVI	$1 - \frac{B - R}{NIR + SWIR1 + SWIR2}$	Índice de vegetação desenvolvido para avaliar o crescimento da cultura do arroz.	Nuarsa, Nishio e Hongo (2011)
<i>Sentinel-2 Red-Edge Position</i>	S2REP	$705 + \frac{35 * (0.5 * (RedEdge3 + R) - RedEdge1)}{(RedEdge2 - RedEdge1)}$	Este índice que mostra a diferença entre a absorção máxima na porção vermelha e a maior refletância na porção NIR. Utilizado para quantificação de teores de clorofila.	Frampton <i>et al.</i> (2013)
<i>Soil Adjusted Vegetation Index</i>	SAVI	$1.5 * \frac{(NIR - R)}{(NIR + R + 0,5)}$	Este índice foi desenvolvido para mitigar o impacto da luminosidade do solo. Para isso, foi adicionado fator de ajuste do solo L à equação dos índices NDVI para corrigir os efeitos do ruído do solo (cor do solo, umidade do solo, variabilidade do solo entre regiões, etc.), que tendem a afetar os resultados.	Huete (1988)
<i>Specific Leaf Area Vegetation Index</i>	SLAVI	$\frac{NIR}{R + SWIR}$	Neste índice a banda SWIR foi incluída para complementar a relação entre vermelho e NIR que formam o princípio subjacente da maioria dos índices de vegetação.	Lymburner <i>et al.</i> (2000)
<i>Simple Ratio or Ration Vegetation Index</i>	SR or RVI	$\frac{NIR}{R}$	Índice baseado na razão entre a banda do infravermelho próximo e o vermelho, produzindo um parâmetro que é altamente sensível à presença de vegetação.	Jordan (1969)
<i>Triangular Vegetation Index</i>	TVI	$0.5 (120 * (NIR - G)) - 200 * (R - G)$	Este índice é calculado como a área de um triângulo hipotético no espaço espectral que conecta refletância do pico verde, a absorção mínima de clorofila e o ombro NIR. Quando a absorção de clorofila causa uma diminuição da refletância vermelha, e a abundância de tecido foliar causa um aumento na refletância NIR, a área total do triângulo aumenta.	Broge e Leblanc (2000)

Continua na próxima página

Continuação

Nome	Abreviação	Equação	Observações sobre os índices	Referência
<i>Visible Atmospherically Resistant Index</i>	VARI	$\frac{(G - R)}{(G + R - B)}$	Este índice apresenta baixa sensibilidade aos efeitos atmosféricos e resultados satisfatórios na estimativa da fração da vegetação. O VARI utiliza em sua formulação as faixas espectrais do azul, verde e vermelho, com detalhe quanto à subtração da banda azul no denominador, proposta para reduzir os efeitos atmosféricos	Gitelson (2003)
<i>Vegetativen</i>	VEG	$\frac{G}{R0.667 * B0.333}$	Índice capaz de separar pixels de plantas de pixels de solo. Essa transformação demonstra um bom contraste entre a planta e o solo.	Hague <i>et al.</i> (2006)

Fonte: Elaborado pela autora, 2022.

Com base nesse processamento, reuniu-se no banco de dados todos os índices de vegetação de acordo com as respectivas parcelas através de um número identificador próprio contendo os valores: máximo, mínimo, mediana, desvio padrão, coeficiente de variação, quartis (Q1, Q2, Q3, Q4) e variação interquartil.

Do total de 64 índices acrescentados no modelo, foi realizada a análise de importância para ranquear apenas os índices de vegetação que mais influenciaram em cada um dos algoritmos. A partir destes dados, realizou-se a análise de correlação de Spearman ($p < 0,05$) para avaliar o grau de correlação entre os diferentes índices capazes de prever a condição nutricional dos plantios. Todos os gráficos foram elaborados com auxílio do pacote ggplot2 no programa estatístico R (Tabela 2).

3.3 Treinamento e Validação

Para a fase de treinamento foram utilizados 70% dos registros do banco de dados, que foram submetidos a seis algoritmos. As variáveis utilizadas como variáveis explicativas foram: departamento, material genético e os teores nutricionais. Os demais registros do banco de dados (30%) foram utilizados apenas na validação. Ao processar os dados com os modelos obtidos na fase de treinamento, sabe-se qual deles apresenta resultados mais próximos dos observados. A avaliação dos modelos de predição foi realizada com base nos valores de R^2 e no Root Mean Squared Error (RMSE), que consiste na diferença quadrática média entre o valor observado e o valor predito pelo modelo.

3.3.1 *Cubist*

O algoritmo *Cubist* é uma técnica de aprendizado de máquina baseada em regras que combina características do algoritmo de árvore de decisão com técnicas de regressão. Utiliza um processo de construção iterativo para gerar regras de decisão que descrevem as relações entre as variáveis de entrada e a variável de saída. Essas regras fornecem insights sobre como as variáveis influenciam o resultado. A técnica é particularmente útil em problemas de regressão, onde o objetivo é prever um valor numérico contínuo.

1. Geração de Regras: O *Cubist* utiliza uma abordagem bottom-up, onde começa com uma única regra que abrange todos os dados de treinamento. Em seguida, o algoritmo busca dividir as regras existentes ou gerar novas regras para melhorar a precisão das previsões.

2. Seleção de Regras: O *Cubist* utiliza critérios de seleção para determinar quais regras devem ser usadas para fazer previsões em uma instância de teste. Esses critérios consideram a cobertura da regra (quantos exemplos são abrangidos) e a precisão da regra (quão bem a regra se ajusta aos dados).

3. Combinação de Regras: Quando várias regras se aplicam a uma instância de teste, o *Cubist* combina as previsões das regras ponderando-as de acordo com sua cobertura e precisão.

4. Refinamento: O *Cubist* realiza um processo de refinamento para melhorar a precisão das previsões. Isso envolve a realização de ajustes nas regras existentes e a geração de novas regras com base nos erros de previsão.

Uma das principais vantagens do algoritmo *Cubist* é sua capacidade de lidar com dados complexos e identificar interações entre variáveis.

3.3.2 *Earth*

O algoritmo EARTH (Exhaustive Algorithm for Regression Tree) é uma técnica de aprendizado de máquina baseada em árvores de regressão. É projetado para modelar relações complexas entre variáveis independentes e dependentes em problemas de regressão. Ele combina as vantagens das árvores de decisão e das funções de base radial (RBF, na sigla em inglês) para capturar interações não lineares, heterogeneidade e efeitos de alta ordem entre as variáveis.

A principal característica do algoritmo EARTH é sua abordagem exaustiva de busca de modelos. Ele realiza uma busca exaustiva entre as possíveis divisões e combinações de variáveis e seus cortes de valores para criar um conjunto completo de modelos de regressão. Cada modelo é representado como uma combinação de funções de base radial e coeficientes, permitindo a modelagem de relacionamentos não lineares.

O processo de construção do modelo no EARTH envolve as seguintes etapas:

1. Busca Exaustiva: O algoritmo examina todas as possíveis divisões e combinações de variáveis independentes para criar modelos de regressão.

2. Seleção do Modelo: O EARTH utiliza critérios de avaliação, como o erro médio quadrático ou o coeficiente de determinação (R^2), para selecionar o melhor modelo a ser utilizado.

3. Refinamento do Modelo: Após a seleção do modelo, o EARTH realiza uma etapa de refinamento, ajustando os cortes de valores e a estrutura das árvores de regressão para melhorar a precisão do modelo.

4. Simplificação do Modelo: Por fim, o EARTH realiza uma etapa de simplificação, removendo termos e ajustando coeficientes para obter um modelo final mais conciso e interpretável.

O algoritmo EARTH é conhecido por sua capacidade de modelar relações não lineares complexas e identificar interações entre variáveis. Ele é amplamente utilizado em áreas como previsão de séries temporais, análise de dados ambientais, entre outros. A interpretabilidade dos modelos gerados pelo EARTH também é uma vantagem, pois permite compreender como as variáveis influenciam as previsões.

3.3.3 *Linear Model* (LM)

O algoritmo *Linear Model* (modelo linear) é uma técnica de aprendizado de máquina que faz uso de modelos lineares para realizar a regressão ou classificação de dados.

Na regressão linear, o algoritmo *Linear Model* estima os coeficientes que multiplicam as variáveis independentes para obter a melhor linha reta que se ajusta aos dados de treinamento. A técnica mais comum é a regressão linear simples, onde há apenas uma variável independente. Porém, é possível expandir para a regressão linear múltipla, que lida com múltiplas variáveis independentes. O algoritmo *Linear Model* tem sido amplamente utilizado devido à sua simplicidade e interpretabilidade. Além disso, é computacionalmente eficiente e adequado para conjuntos de dados de tamanho moderado a grande.

Em resumo, o algoritmo *Linear Model* é uma técnica de aprendizado de máquina que utiliza modelos lineares para realizar regressão ou classificação de dados, com base na suposição de uma relação linear entre as variáveis independentes e a variável de saída.

3.3.4 *Ranger*

O algoritmo *Ranger* é uma implementação rápida e eficiente de algoritmos de floresta aleatória (*random forest*). Ele é projetado para lidar com conjuntos de dados grandes e complexos, oferecendo um desempenho aprimorado em comparação com outras implementações tradicionais.

O *Ranger* é escrito em linguagem de programação C++ e R, e foi desenvolvido para ser executado em sistemas distribuídos. Ele aproveita a paralelização e a distribuição de cálculos em vários núcleos de processamento ou em um cluster de computadores, permitindo uma execução rápida e eficiente de algoritmos de floresta aleatória.

As florestas aleatórias são um conjunto de árvores de decisão, em que cada árvore é treinada em uma amostra aleatória dos dados de treinamento, utilizando amostragem de bootstrap e seleção aleatória de variáveis. O algoritmo *Ranger* implementa essa abordagem, fornecendo um conjunto de parâmetros ajustáveis para controle do treinamento das árvores e ajuste do modelo.

Algumas características e vantagens do algoritmo *Ranger* incluem:

1. Velocidade e escalabilidade: O *Ranger* é conhecido por sua alta velocidade de treinamento, especialmente em grandes conjuntos de dados. Ele foi projetado para aproveitar o paralelismo e a distribuição de tarefas, permitindo um processamento rápido em sistemas distribuídos.

2. Seleção automática de hiperparâmetros: O *Ranger* possui um mecanismo interno para seleção automática de hiperparâmetros, como o número de árvores, a profundidade máxima das árvores

e o número de variáveis selecionadas para cada divisão. Isso facilita a configuração e otimização do modelo.

3. Importância de variáveis: O *Ranger* fornece uma estimativa da importância das variáveis na construção do modelo. Essa informação pode ser usada para selecionar as variáveis mais relevantes, simplificar o modelo ou identificar características importantes nos dados.

4. Uso eficiente de memória: O *Ranger* foi projetado para otimizar o uso de memória, permitindo o treinamento e a inferência de modelos em conjuntos de dados de tamanho considerável.

Em resumo, o algoritmo *Ranger* é uma implementação rápida e escalável de algoritmos de floresta aleatória. Ele oferece um desempenho aprimorado em comparação com outras implementações, sendo especialmente útil em conjuntos de dados grandes e complexos. O *Ranger* é amplamente utilizado em tarefas de aprendizado de máquina, como classificação, regressão e seleção de características.

3.3.5 *Random Forest* (RF)

O algoritmo *Random Forest* (floresta aleatória) é uma técnica de aprendizado de máquina que utiliza um conjunto de árvores de decisão para realizar classificação ou regressão. É uma abordagem popular e eficaz que combina várias árvores de decisão independentes para fazer previsões mais precisas e robustas.

A ideia central do *Random Forest* é construir um conjunto de árvores de decisão independentes e combinar suas previsões para obter um resultado final. Cada árvore é treinada em uma amostra aleatória dos dados de treinamento (amostragem de bootstrap), e apenas um subconjunto aleatório das variáveis é considerado para cada divisão na árvore. Essas duas fontes de aleatoriedade ajudam a reduzir a correlação entre as árvores e promovem a diversidade do conjunto.

Durante o processo de treinamento, cada árvore de decisão é crescente até atingir uma profundidade máxima ou até que o critério de parada seja atingido. Geralmente, a árvore é expandida até que cada folha contenha apenas um número mínimo de exemplos ou não haja mais ganho de informação significativo nas divisões.

Para fazer uma previsão com o *Random Forest*, as previsões de todas as árvores individuais são combinadas. No caso de classificação, o resultado final é determinado por votação majoritária, ou seja, a classe mais frequente entre as árvores é escolhida. Para regressão, a média das previsões das árvores é tomada como o valor final.

As principais vantagens do algoritmo *Random Forest* incluem:

1. Alta precisão: O *Random Forest* é conhecido por sua alta precisão preditiva em comparação com muitos outros algoritmos de aprendizado de máquina. Ele é capaz de capturar relacionamentos complexos entre as variáveis de entrada e a variável de saída.
2. Robustez: O conjunto de árvores independentes ajuda a reduzir o impacto de outliers e ruídos nos dados, tornando o modelo mais robusto.
3. Lidar com variáveis de entrada de diferentes tipos: O *Random Forest* pode lidar com dados numéricos e categóricos, sem exigir uma etapa adicional de pré-processamento para converter variáveis categóricas em numéricas.
4. Estimativa de importância de variáveis: O algoritmo *Random Forest* fornece uma estimativa da importância relativa das variáveis na construção do modelo, permitindo a seleção de características relevantes.

O algoritmo *Random Forest* é amplamente utilizado em problemas de classificação, regressão e detecção de anomalias. Sua versatilidade, precisão e capacidade de lidar com dados complexos o tornam uma escolha popular em várias aplicações de aprendizado de máquina.

3.3.6 *Gradient Boosting Machine (GBM)*

O algoritmo GBM (*Gradient Boosting Machine*) é uma técnica de aprendizado de máquina que combina um conjunto de modelos de aprendizado fracos para criar um modelo de aprendizado forte. Ao contrário do algoritmo *Random Forest*, que combina várias árvores de decisão independentes, o GBM utiliza um conjunto de modelos em sequência, onde cada modelo é treinado para corrigir os erros do modelo anterior.

O GBM é baseado no princípio do aumento de gradiente, que envolve o ajuste iterativo dos modelos para minimizar uma função de perda. Durante o treinamento, o GBM realiza as seguintes etapas:

1. Treinamento do primeiro modelo: O primeiro modelo é treinado em um conjunto de dados de treinamento. Esses modelos iniciais são geralmente simples, como árvores de decisão rasas.
2. Cálculo dos resíduos: O GBM calcula os resíduos, que são as diferenças entre os valores reais e as previsões do modelo atual.
3. Treinamento do próximo modelo: O próximo modelo é treinado nos resíduos do modelo anterior, com o objetivo de prever os resíduos restantes.
4. Combinação dos modelos: As previsões de todos os modelos são combinadas para obter a previsão final. No GBM, essa combinação é feita somando as previsões de todos os modelos.
5. Iteração: Essas etapas são repetidas por várias iterações, permitindo que cada modelo corrija os erros do modelo anterior. O número de iterações é um hiperparâmetro que precisa ser definido antes do treinamento.

O GBM possui algumas características e vantagens notáveis:

1. Flexibilidade: O GBM pode ser aplicado tanto a problemas de regressão quanto de classificação. Além disso, pode utilizar diferentes funções de perda e algoritmos de otimização para se adaptar a diferentes cenários.

2. Pode lidar com dados complexos: O GBM é capaz de capturar relações não lineares e interações complexas entre as variáveis de entrada, tornando-o adequado para conjuntos de dados complexos.

3. Alta precisão: O GBM é conhecido por sua alta precisão preditiva, especialmente quando é treinado com um número suficiente de iterações e parâmetros otimizados.

4. Tratamento automático de valores ausentes: O GBM é capaz de lidar com valores ausentes nos dados de forma automática, sem exigir um tratamento prévio desses valores.

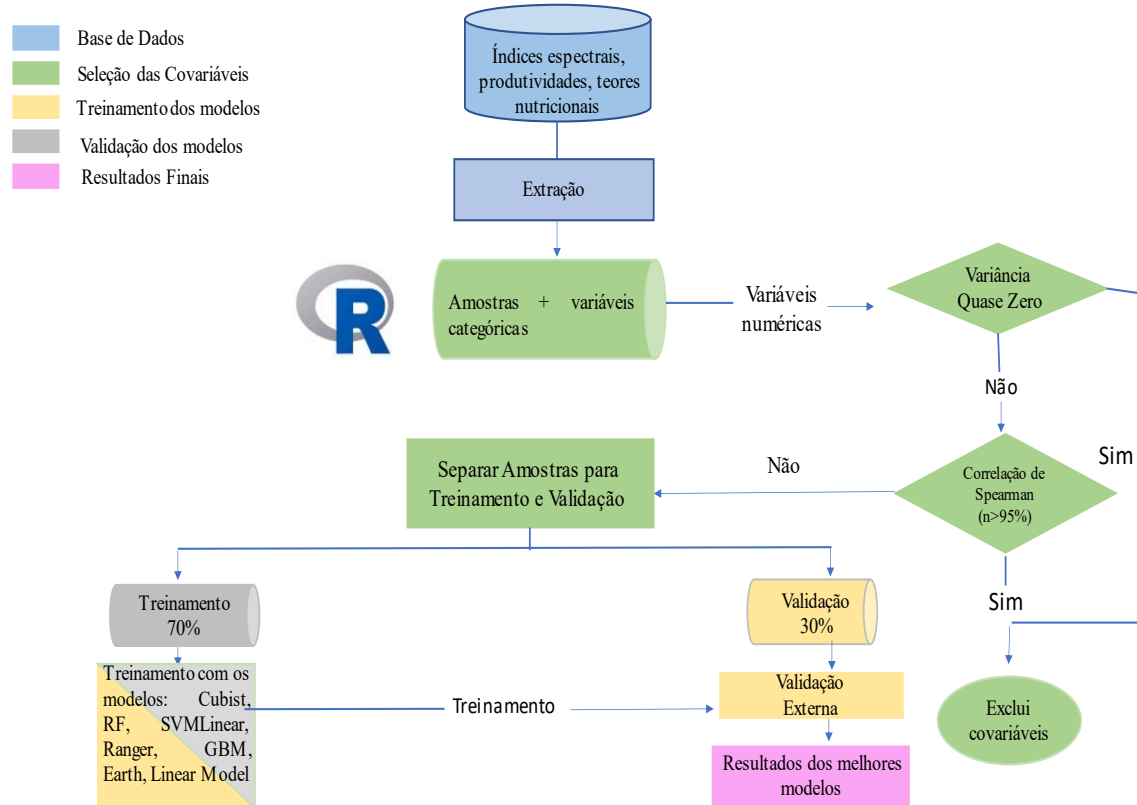
No entanto, o GBM também apresenta algumas considerações importantes:

1. Sensibilidade a outliers: Assim como outros algoritmos baseados em gradientes, o GBM pode ser sensível a outliers nos dados, o que pode afetar sua precisão.

2. Necessidade de ajuste de hiperparâmetros: O GBM possui vários hiperparâmetros que precisam ser ajustados corretamente, como a taxa de aprendizado, o número de iterações e a profundidade dos modelos.

Em resumo, o algoritmo GBM é uma técnica poderosa de aprendizado de máquina que combina um conjunto de modelos fracos para criar um modelo forte e preciso. É amplamente utilizado em problemas de regressão e classificação, oferecendo flexibilidade e alta precisão preditiva.

Figura 2 - Fluxograma metodológico utilizado na modelagem por aprendizado de máquina para seleção dos melhores algoritmos para prever a condição nutricional e produtividade de palma de óleo.



Fonte: Elaborado pela autora, 2022.

4 RESULTADOS E DISCUSSÃO

4.1 Monitoramento Nutricional

Os resultados apresentados na Figura 3, mostram o comportamento dos dados quando submetidos aos diferentes modelos e seus respectivos valores de R^2 . De todos os macronutrientes avaliados, os que apresentaram maior dispersão foram o N, P e K, isso ocorreu devido aos altos teores destes nutrientes em algumas parcelas que se concentram na faixa de suficiência excessiva. Estes resultados refletem diretamente nos valores de índices de vegetação, o que reforça a dispersão de alguns elementos no gráfico.

No geral, todos os algoritmos apresentaram significativos valores de R^2 , porém, é importante observar que a menor variação interquartil ocorreu com os algoritmos *Earth*, *Ranger* e *Random Forest*, indicando boa acurácia para predição dos elementos.

Na Figura 3 é possível analisar o comportamento dos dados em cada um dos algoritmos utilizados no estudo. No geral, todos os modelos apresentaram valores de R^2 acima de 50%, demonstrando o potencial de predição do estado nutricional a partir da extração de informações das imagens de satélite.

A dispersão dos valores no boxplot são indicativos de que mesmo com interessantes resultados na validação, alguns algoritmos tendem a ser mais assertivos devido a menor variabilidade de dados.

Ao analisar os macronutrientes (Figura 3), os elementos nitrogênio, potássio e fósforo apresentaram comportamentos distintos nos modelos *Cubist*, GBM e LM, com grande variação nos dados e valores elevados de mediana e isso traz certa imprevisibilidade, pois ora se tem valores muito altos, ora se tem valores mais baixos e esse comportamento tende a resultados menos assertivos. No entanto, para os algoritmos *Earth*, *Ranger* e RF a amplitude de variação é bem menor quando comparado aos citados anteriormente e mantendo o padrão para todos os macronutrientes.

Em relação ao nitrogênio, potássio e fósforo onde houve maior dispersão dos dados, percebe-se que a mediana está posicionada na parte superior, ou seja, a média e o valor mediano são mais elevados. Isso indica que os resultados de predição para N no modelo *cubist* e GBM, o P para GBM, LM e K para LM possuem tendência a apresentar resultados menos assertivos.

O boxplot dos nutrientes cálcio, magnésio e enxofre apresentam baixa variabilidade e desvio padrão, além de poucos ou inexistentes pontos de outliers. Em todos os modelos esses

elementos apresentam comportamentos muito semelhantes, sendo o enxofre o único que manteve valores de mediana próximos do primeiro quartil.

Os algoritmos Earth, *Ranger* e RF mantiveram o comportamento de dispersão com baixa variabilidade para todos os macronutrientes, demonstrando o potencial destes modelos para prever o estado nutricional dos plantios de dendê.

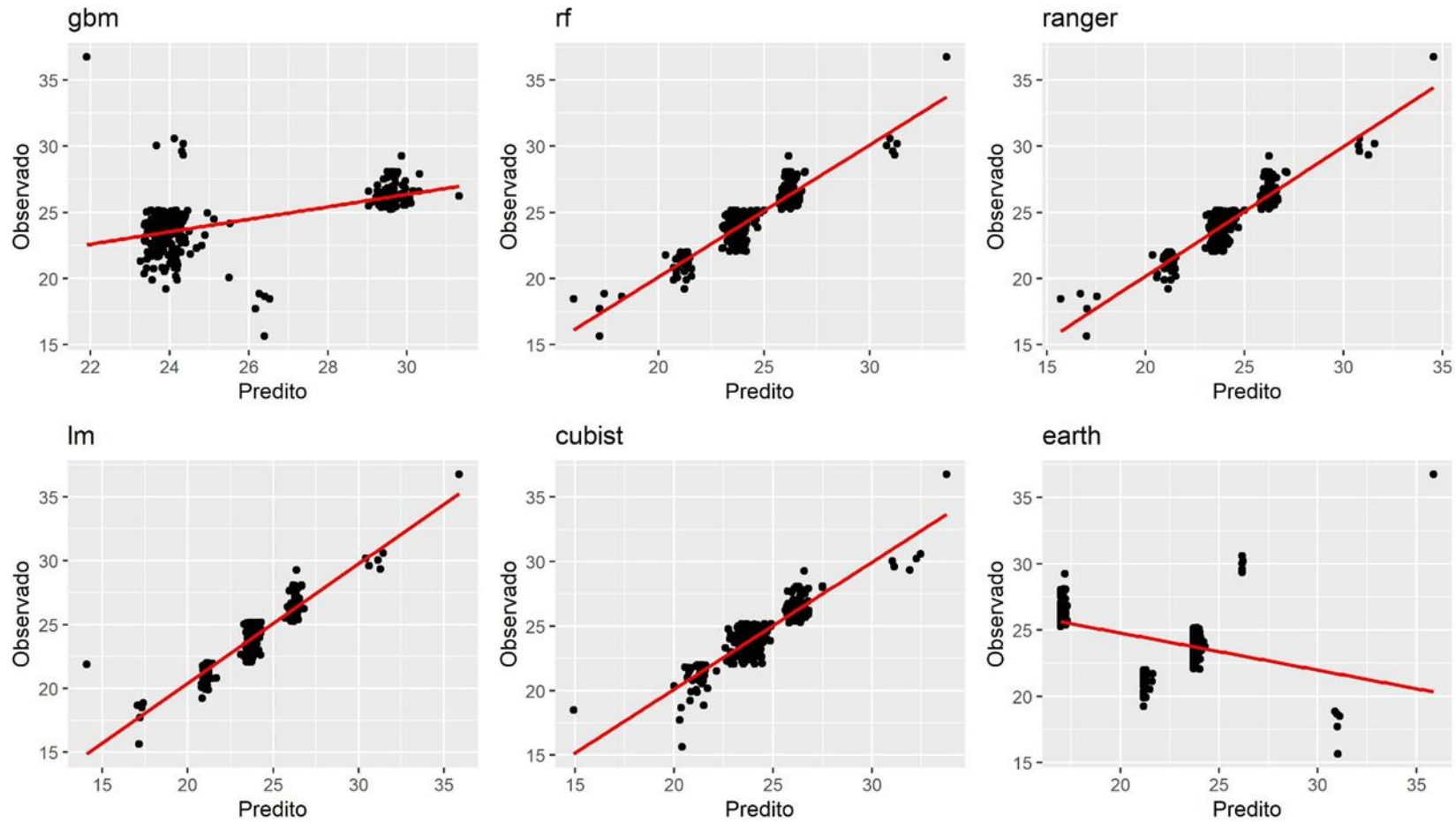
Na Figura 4, observa-se os resultados para os micronutrientes, sendo o cobre o elemento com maior variabilidade, com valores de mediana elevados. Esta performance ocorre novamente nos modelos *Cubist*, GBM e LM. O zinco também apresentou certa amplitude de variação com resultados próximos do terceiro quartil.

Em relação ao manganês, este micronutriente foi o que apresentou comportamento mais distinto entre os modelos, principalmente em relação aos valores da mediana que oscilaram as posições de primeiro e terceiro quartil, certamente isso ocorre devido a variação que há de valores no banco de dados. Pedrali *et al.*, (2022) observou o mesmo comportamento para este elemento, expressando um desvio padrão de 135,90 mg kg⁻¹ em dados de monitoramento nutricional de eucalipto aos 18 meses de idade. Essa variabilidade entre os resultados pode ser explicada pelo fato de o Mn apresentar ligações fracas com os ligantes orgânicos, sendo facilmente oxidado com a mobilidade deste elemento na planta e isso ocorre principalmente durante o processo de senescência, quando parte dos nutrientes dos tecidos foliares se deslocam para outras partes da planta (KIRKBY; RÖMHELD, 2007). Este efeito deve ser considerado como ponto de atenção pois indica que os índices de vegetação utilizados na análise apresentam certa limitação em correlacionar as características espectrais com os sintomas nutricionais de manganês.

Entre todos os elementos, o ferro foi o único com valores de validação abaixo de 30% e assim como o manganês, necessita ser processado com outros modelos e calculados outros índices para avaliar o potencial de resposta nestas novas condições. O B foi um elemento com pouca variação, com exceção do modelo LM onde apresentou a maior dispersão de dados, porém, esse comportamento do LM foi padrão entre todos os micronutrientes.

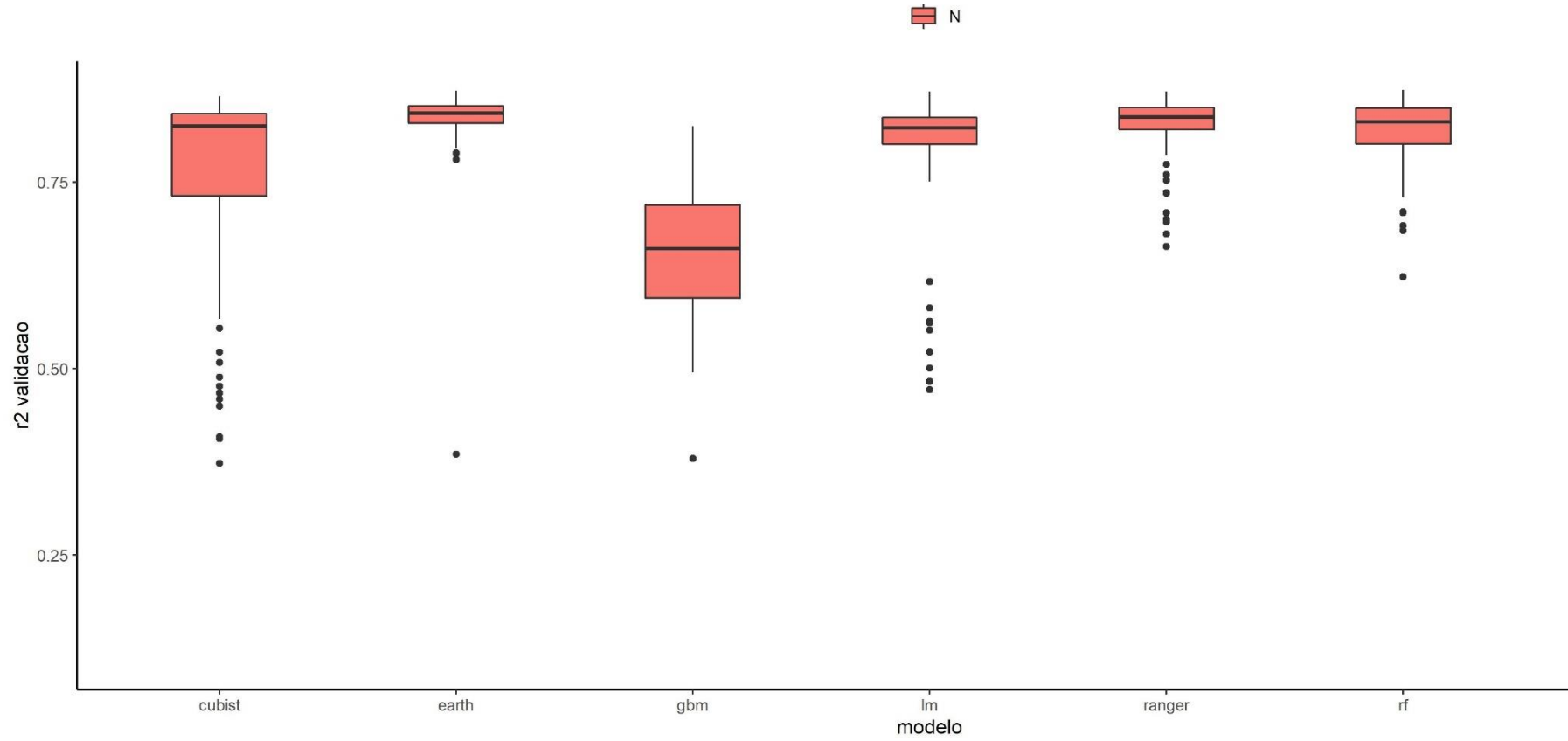
Selecionar os algoritmos que melhor se ajustam ao banco de dados é uma importante etapa da modelagem e com base nesses resultados foi possível realizar a análise de importância para identificar quais as variáveis que mais influenciaram na capacidade de predição do estado nutricional.

Figura 3 - Estimativa dos teores de Nitrogênio para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



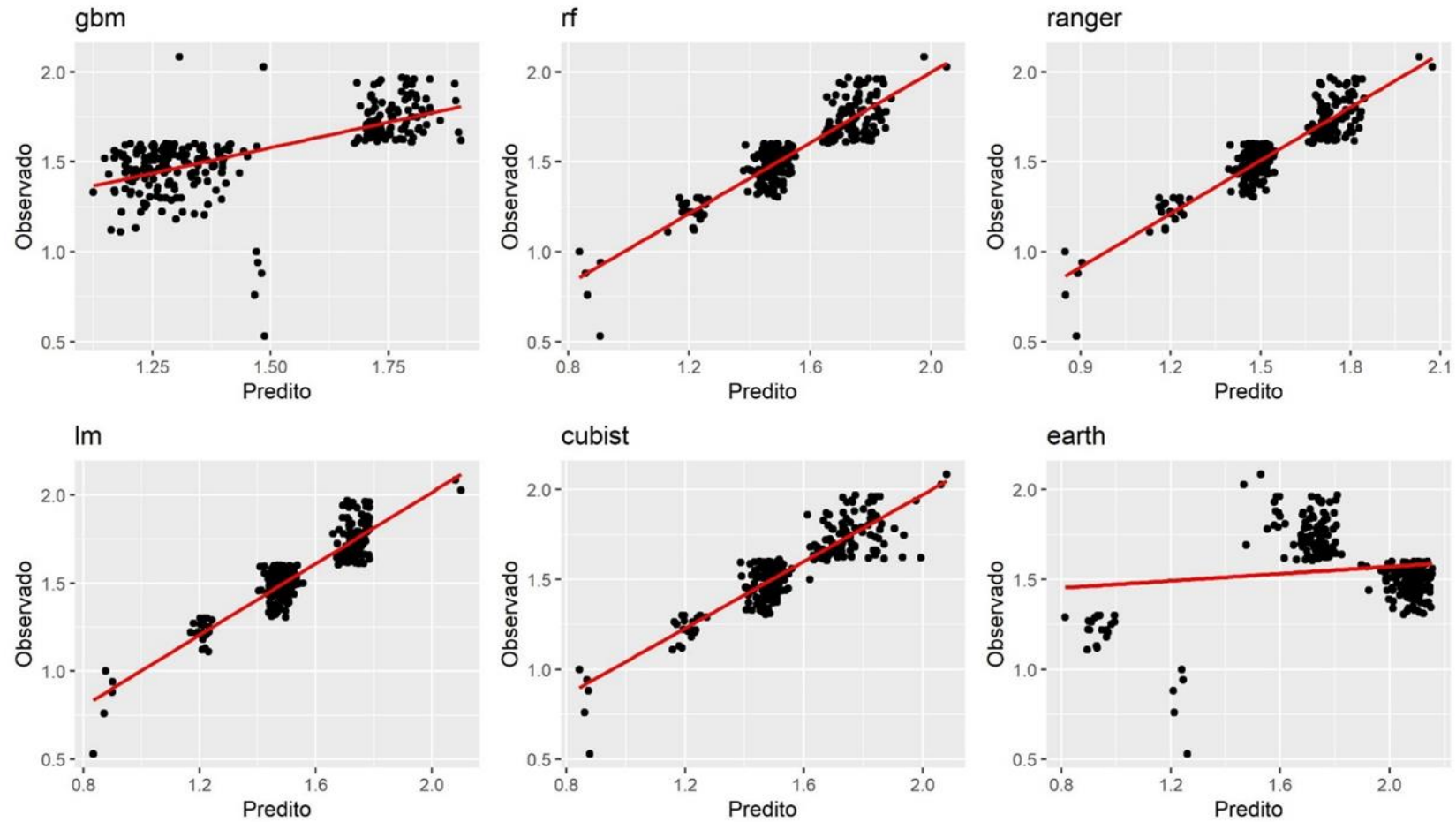
Fonte: Elaborado pela autora, 2022.

Figura 4 - Valores de R^2 validação dos algoritmos utilizados na predição de teor foliar de Nitrogênio na cultura palma de óleo.



Fonte: Elaborado pela autora, 2022.

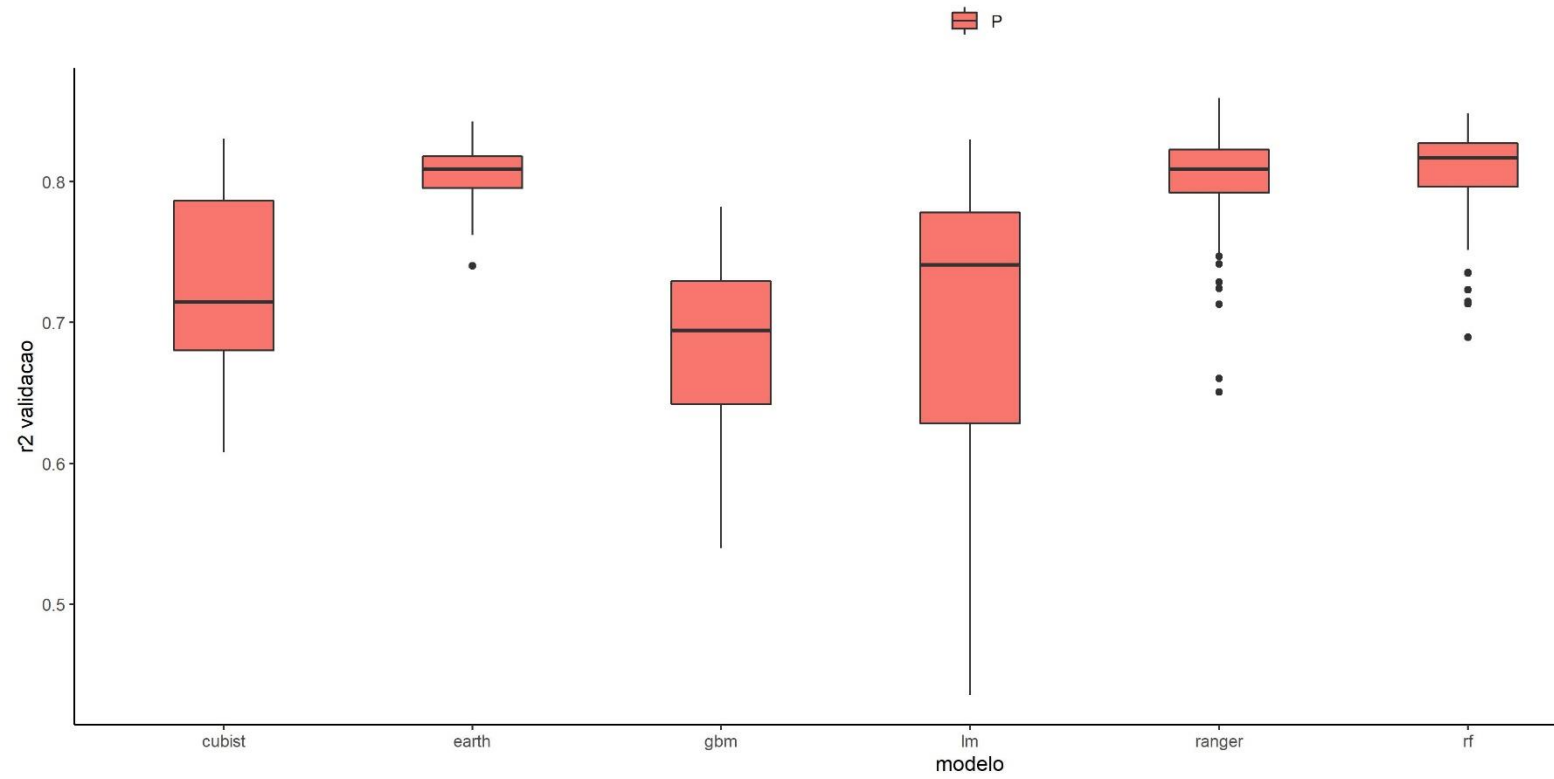
Figura 5 - Estimativa dos teores de Fósforo para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



P

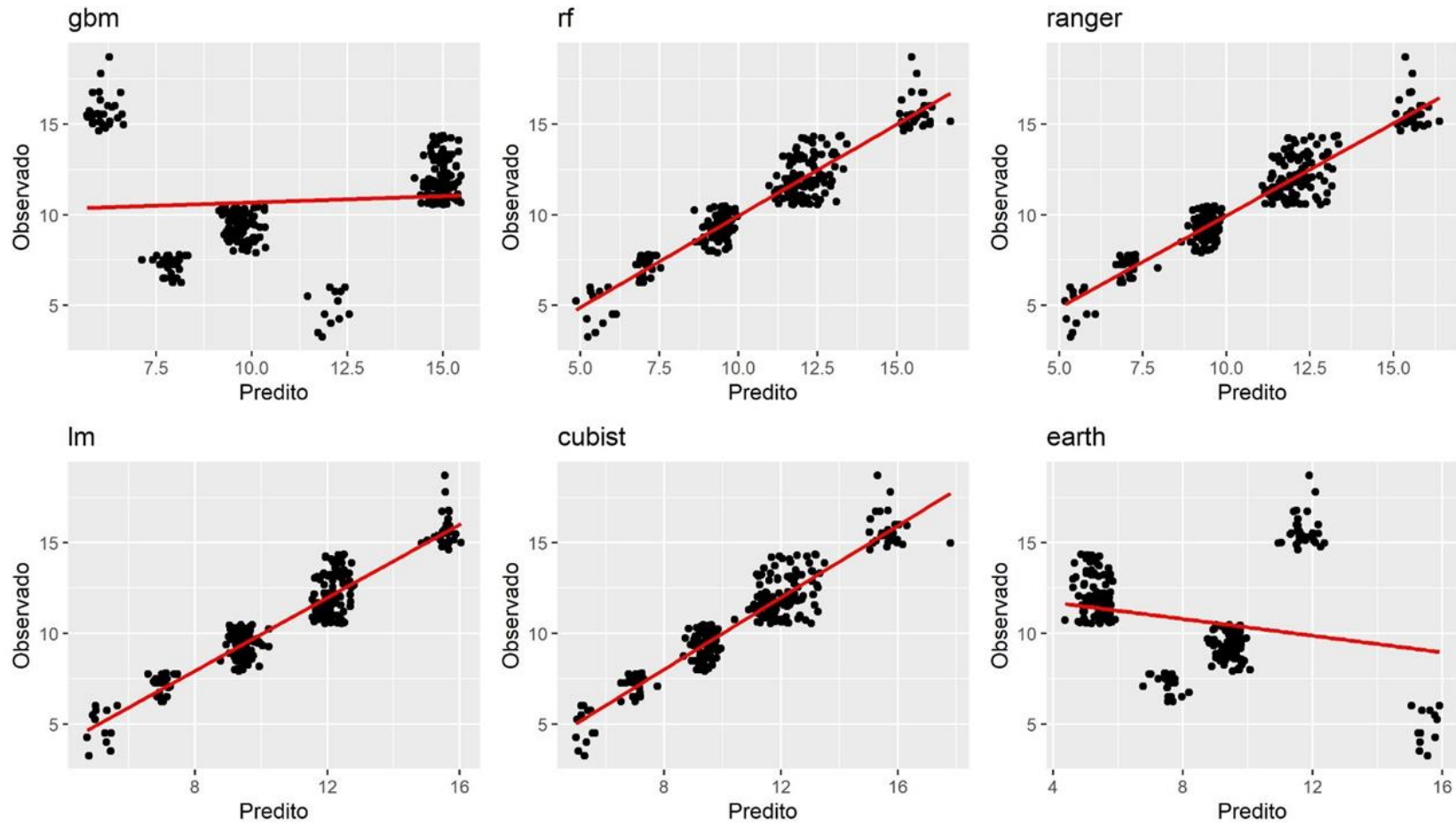
Fonte: Elaborado pela autora, 2022.

Figura 6 - Valores de R^2 validação dos algoritmos utilizados na predição de teor foliar de Fósforo na cultura palma de óleo.



Fonte: Elaborado pela autora, 2022.

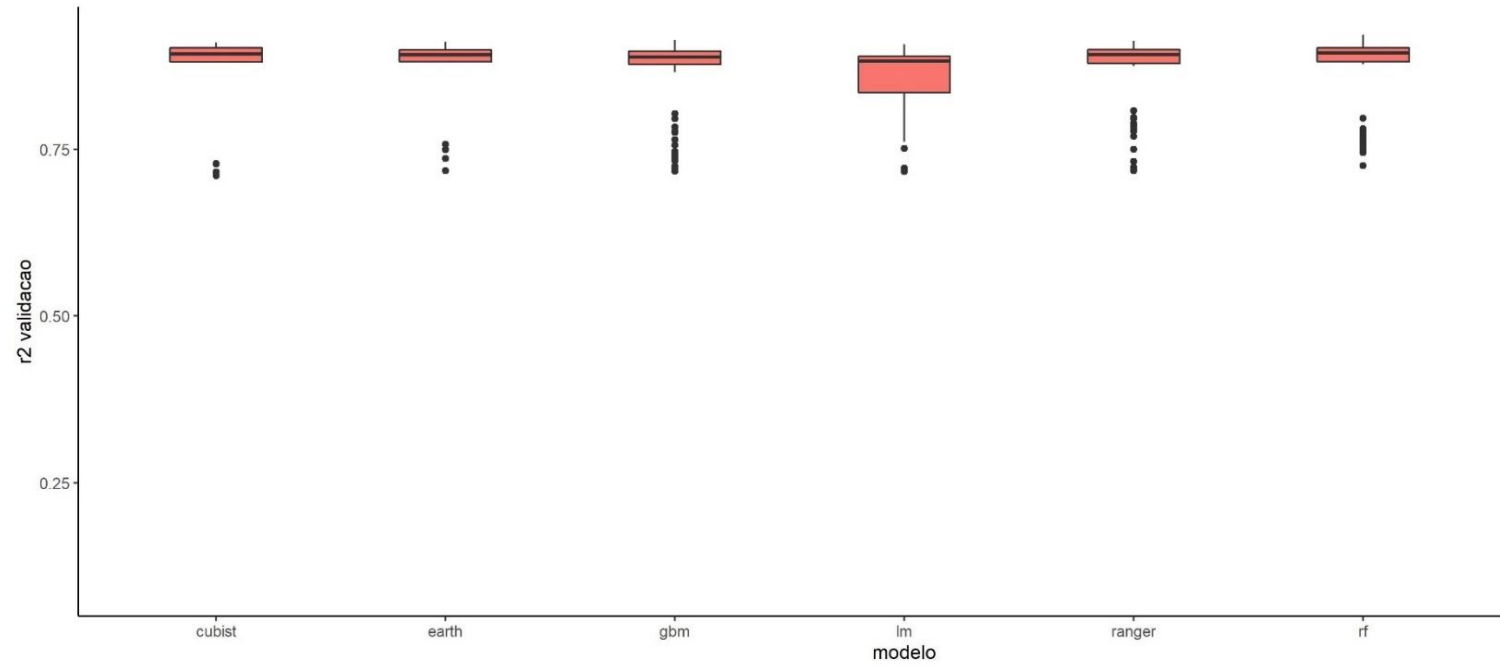
Figura 7 - Estimativa dos teores de Potássio para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



K

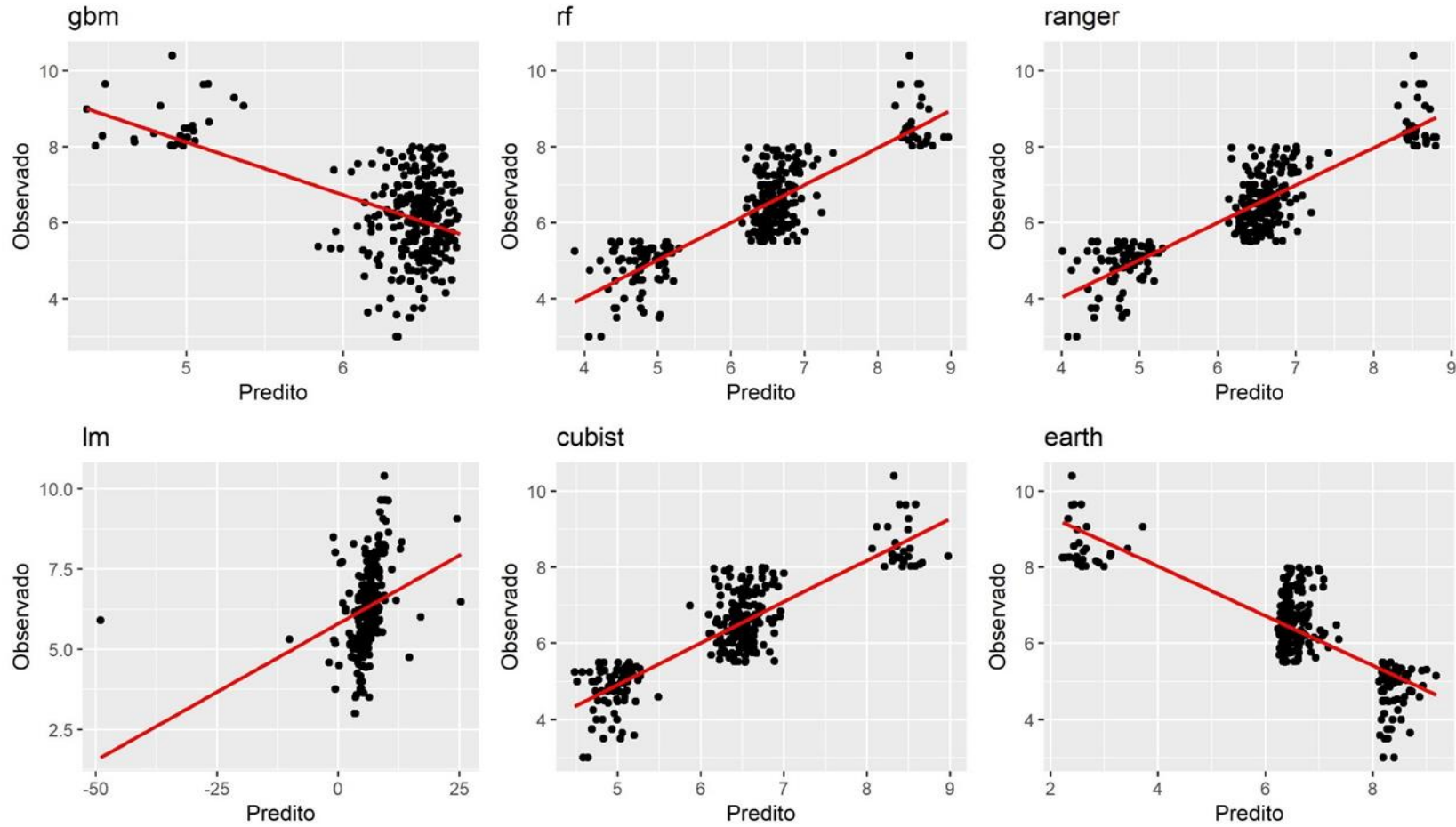
Fonte: Elaborado pela autora, 2022.

Figura 8 - Valores de R^2 validação dos algoritmos utilizados na predição de teor foliar de Potássio na cultura palma de óleo.



Fonte: Elaborado pela autora, 2022.

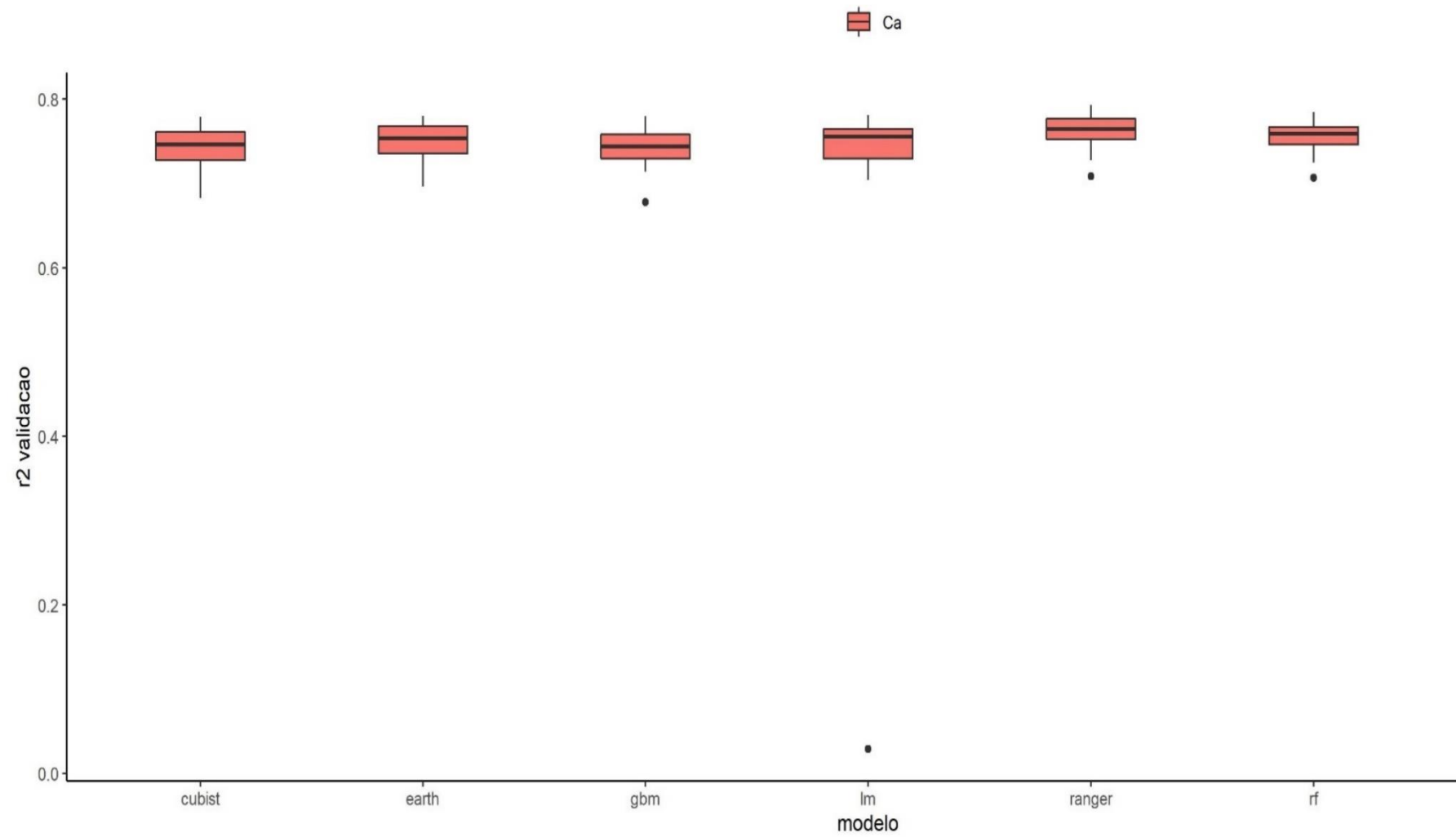
Figura 9 - Estimativa dos teores de Cálcio para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



Ca

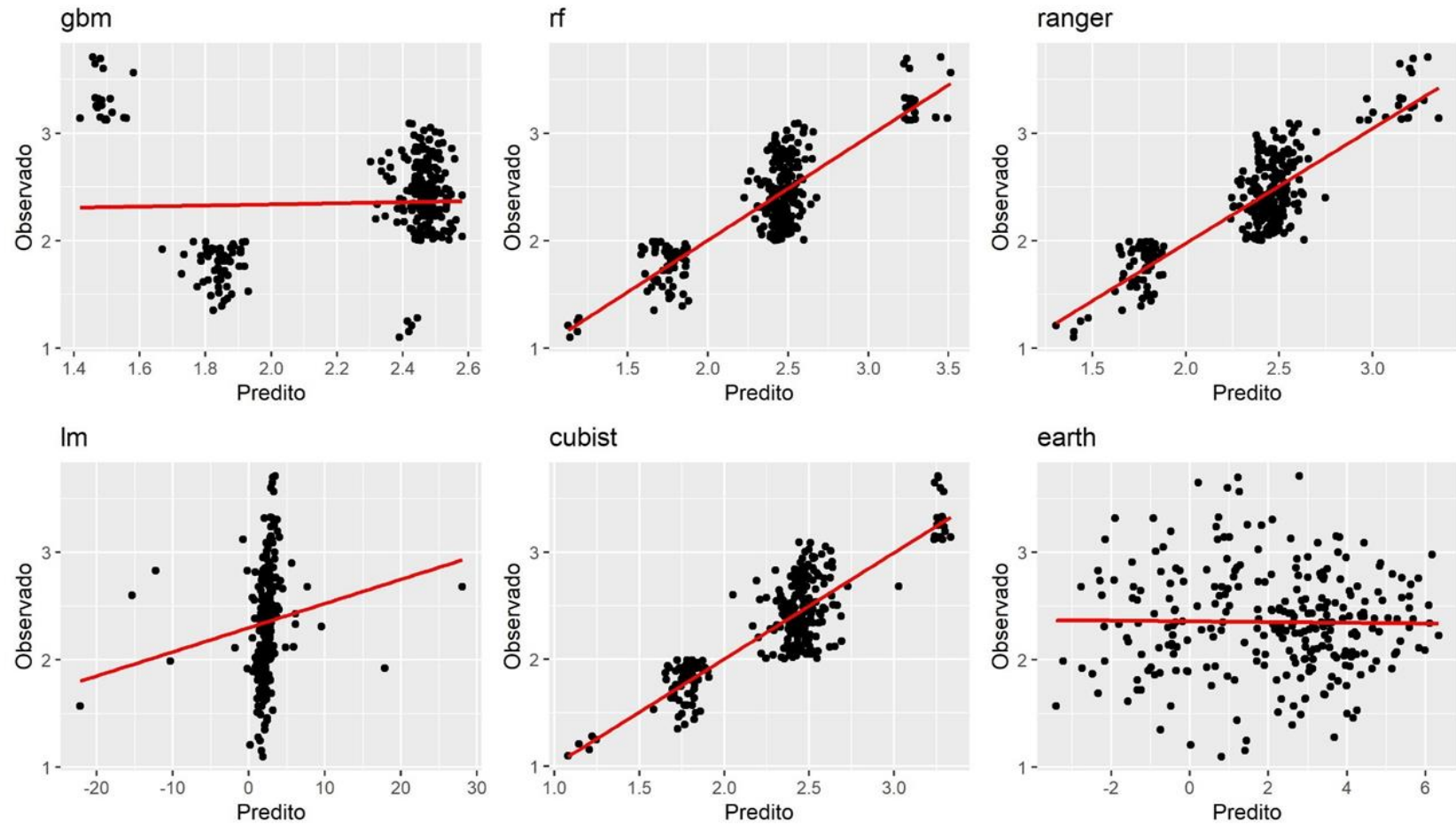
Fonte: Elaborado pela autora, 2022.

Figura 10 - Valores de R^2 validação dos algoritmos utilizados na predição de teor foliar de Cálcio na cultura palma de óleo.

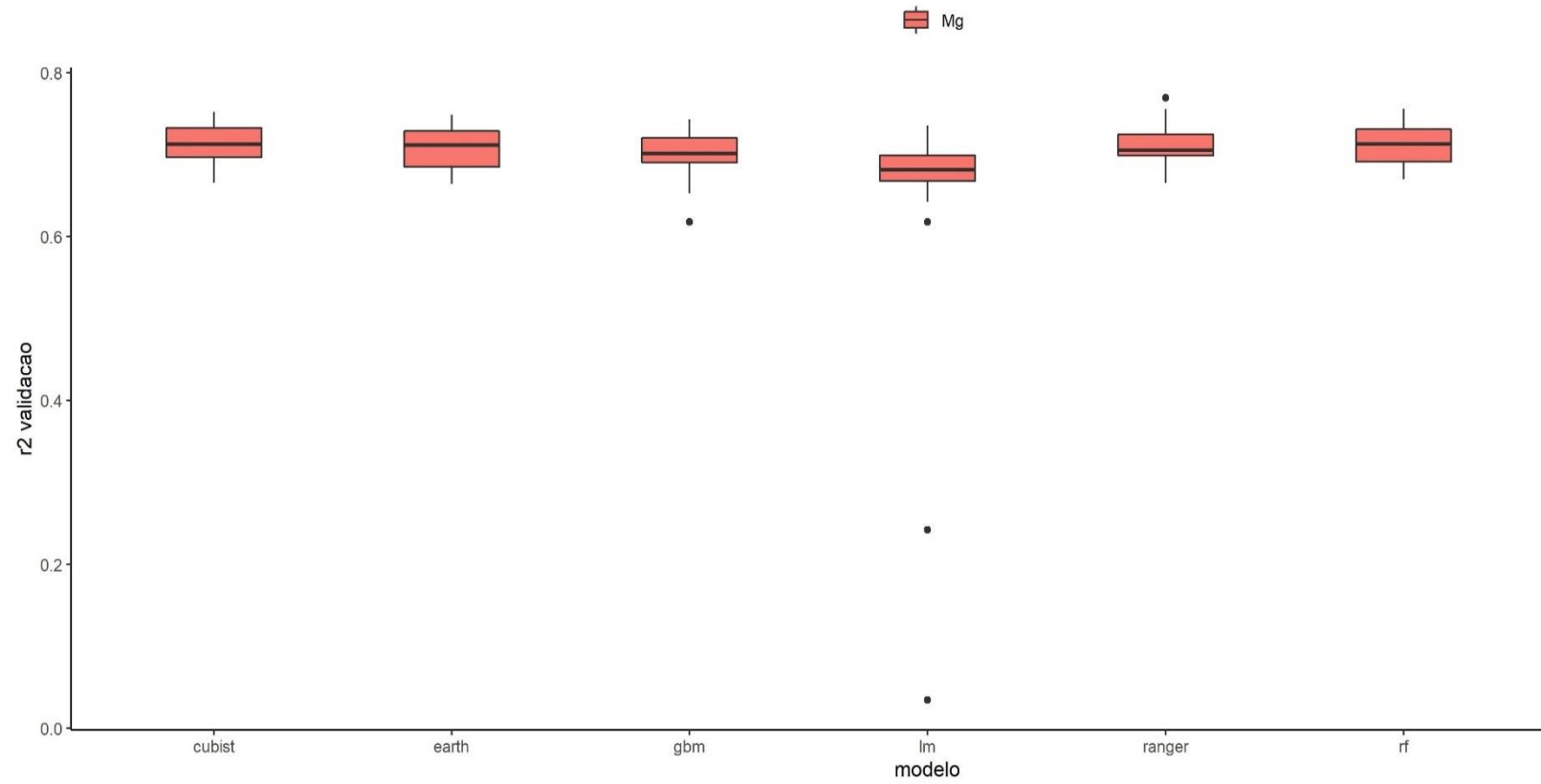


Fonte: Elaborado pela autora, 2022.

Figura 11 - Estimativa dos teores de Magnésio para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.

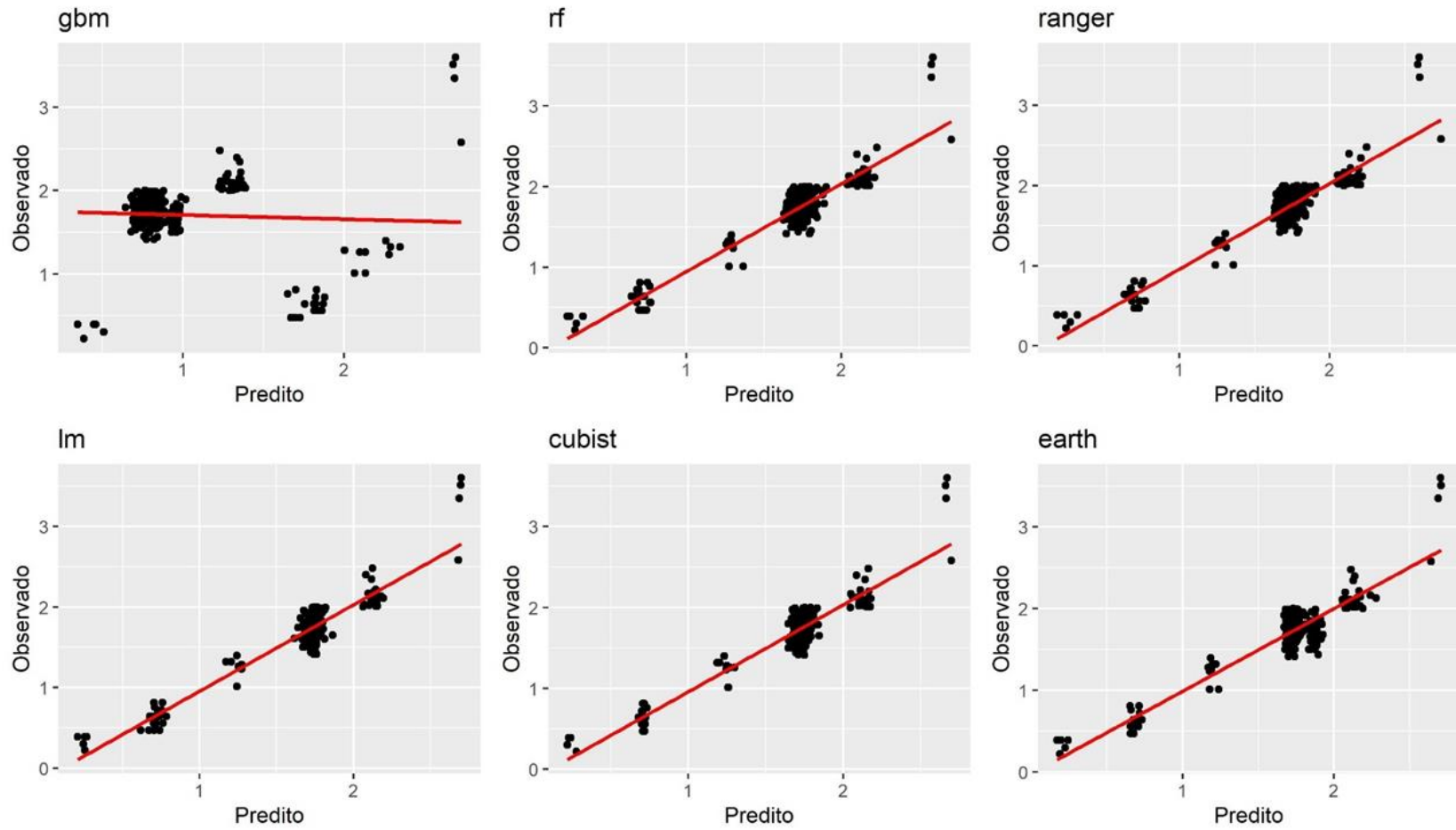


Fonte: Elaborado pela autora, 2022.

Figura 12 - Valores de R^2 validação dos algoritmos utilizados na predição de teor foliar de Magnésio na cultura palma de óleo.

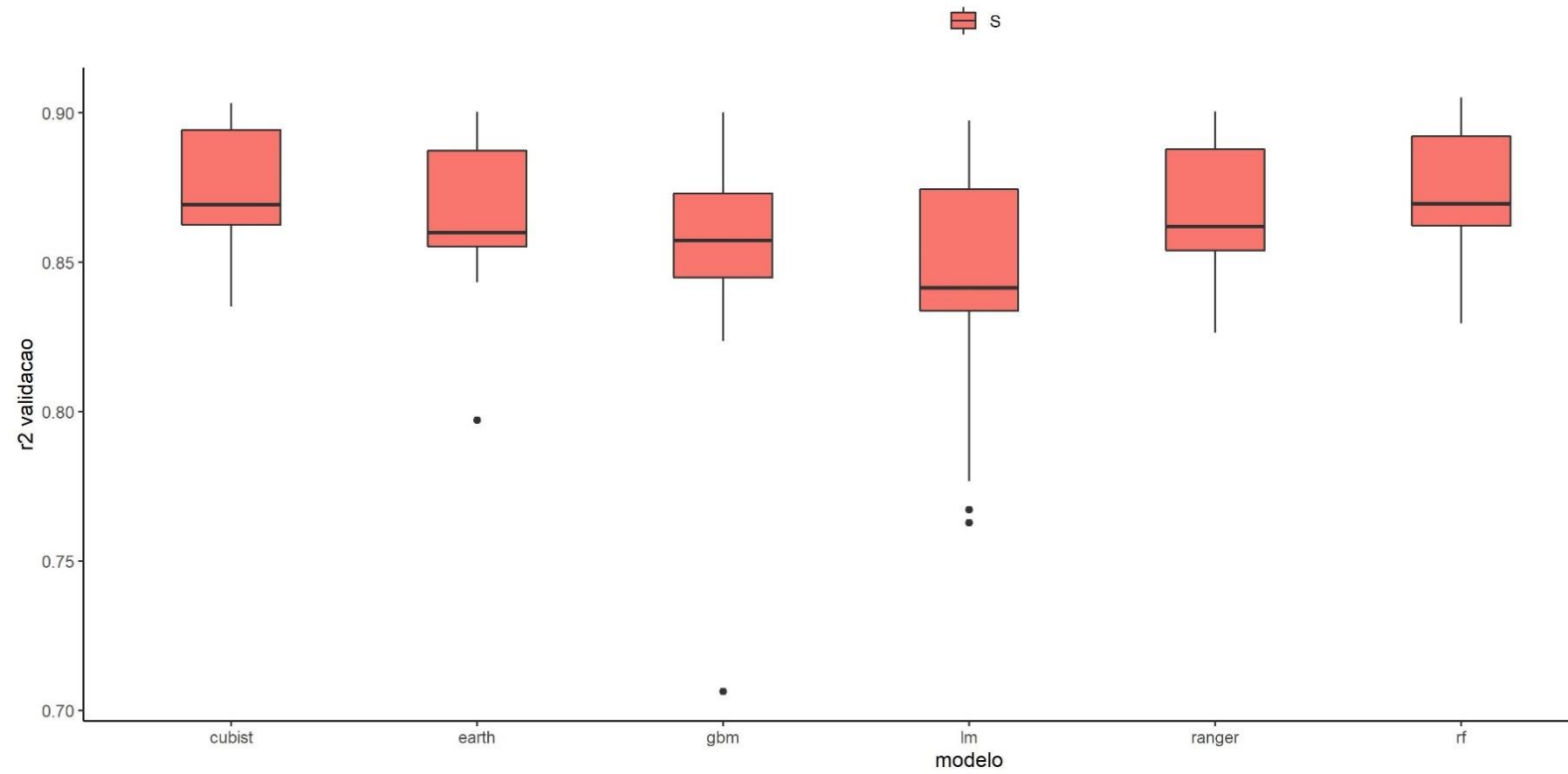
Fonte: Elaborado pela autora, 2022.

Figura 13 - Estimativa dos teores de Enxofre para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.

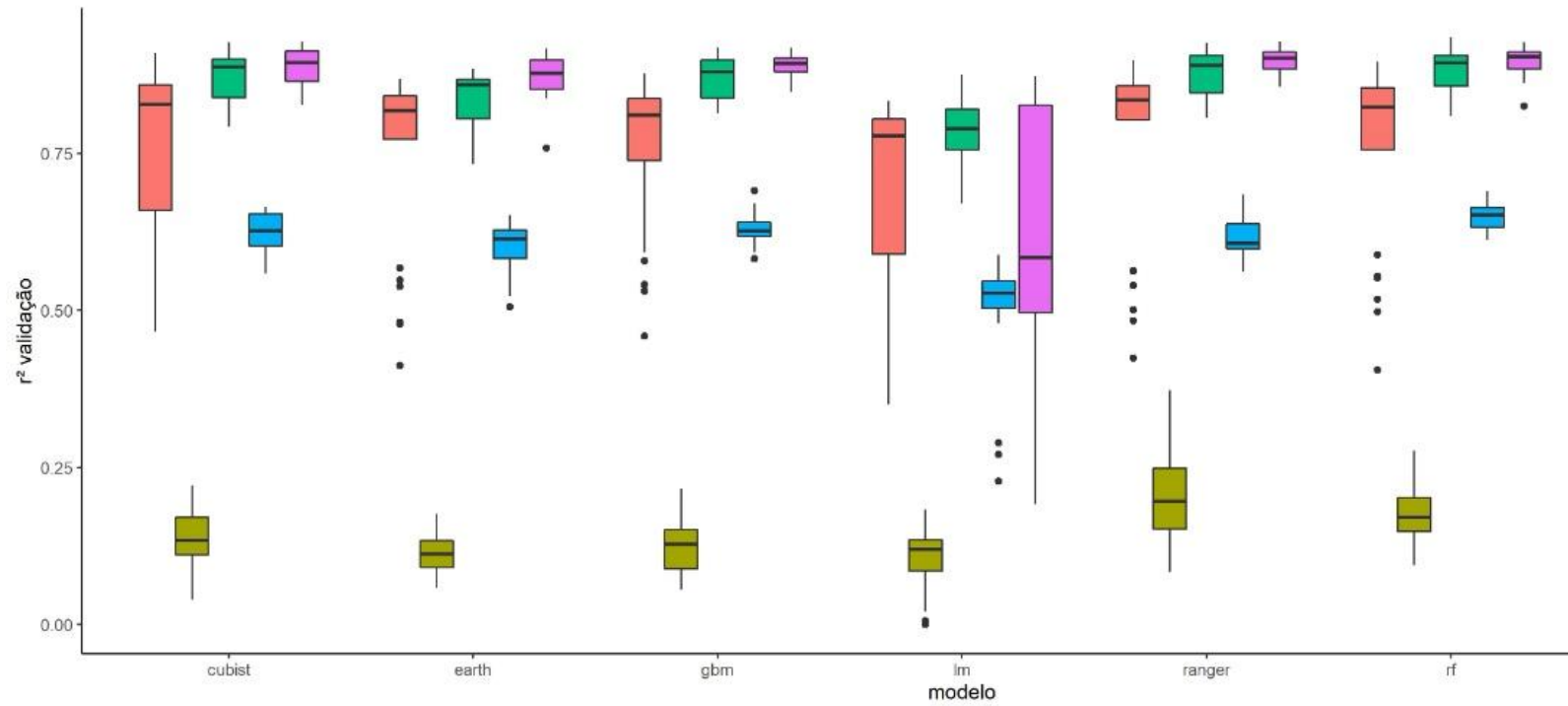


S

Fonte: Elaborado pela autora, 2022.

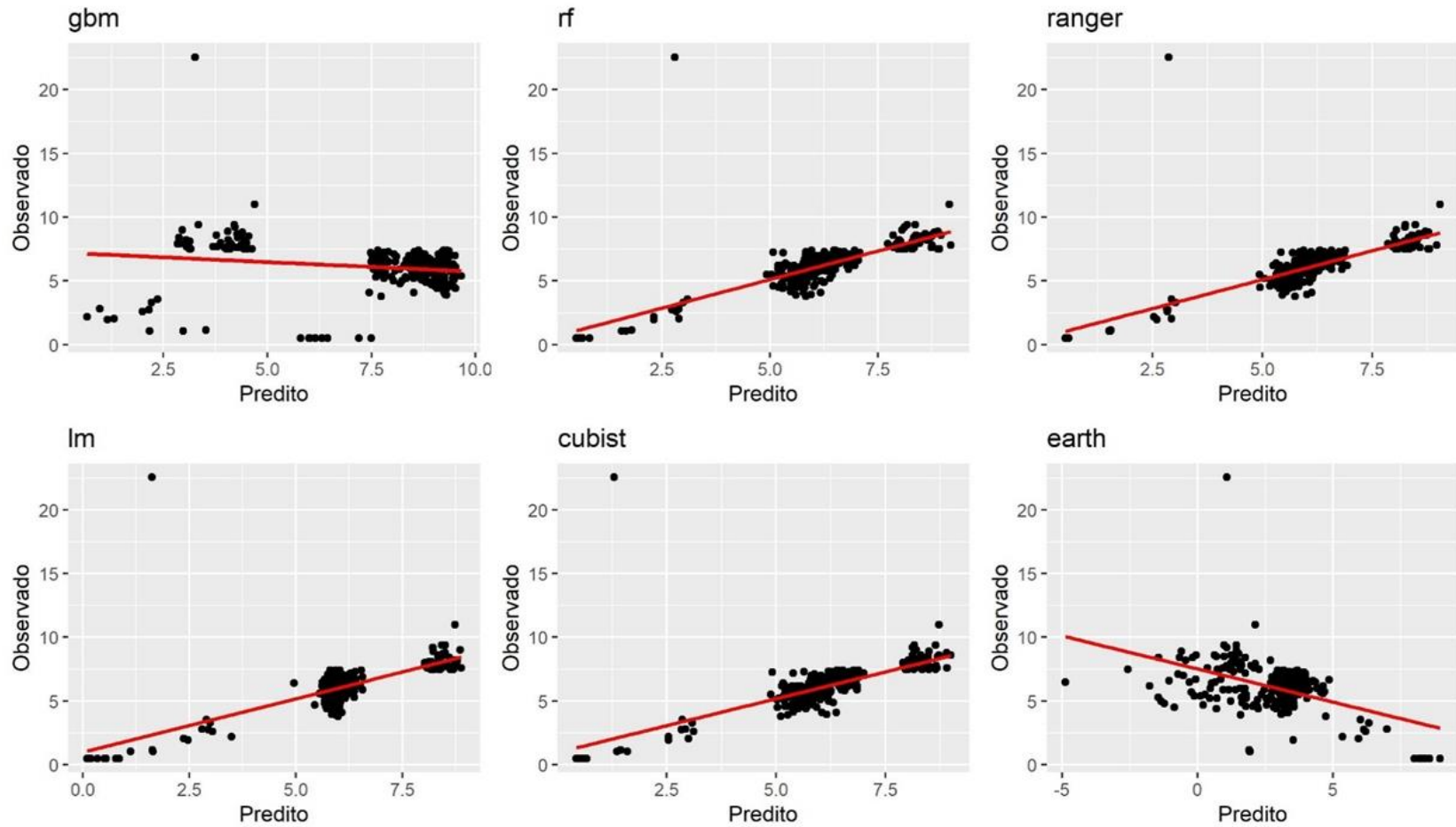
Figura 14 - Valores de R^2 validação dos algoritmos utilizados na previsão de teor foliar de Enxofre na cultura palma de óleo.

Fonte: Elaborado pela autora, 2022.

Figura 15 - Valores de R^2 validação dos algoritmos utilizados na predição de teor foliar dos micronutrientes na cultura palma de óleo.

Fonte: Elaborado pela autora, 2022.

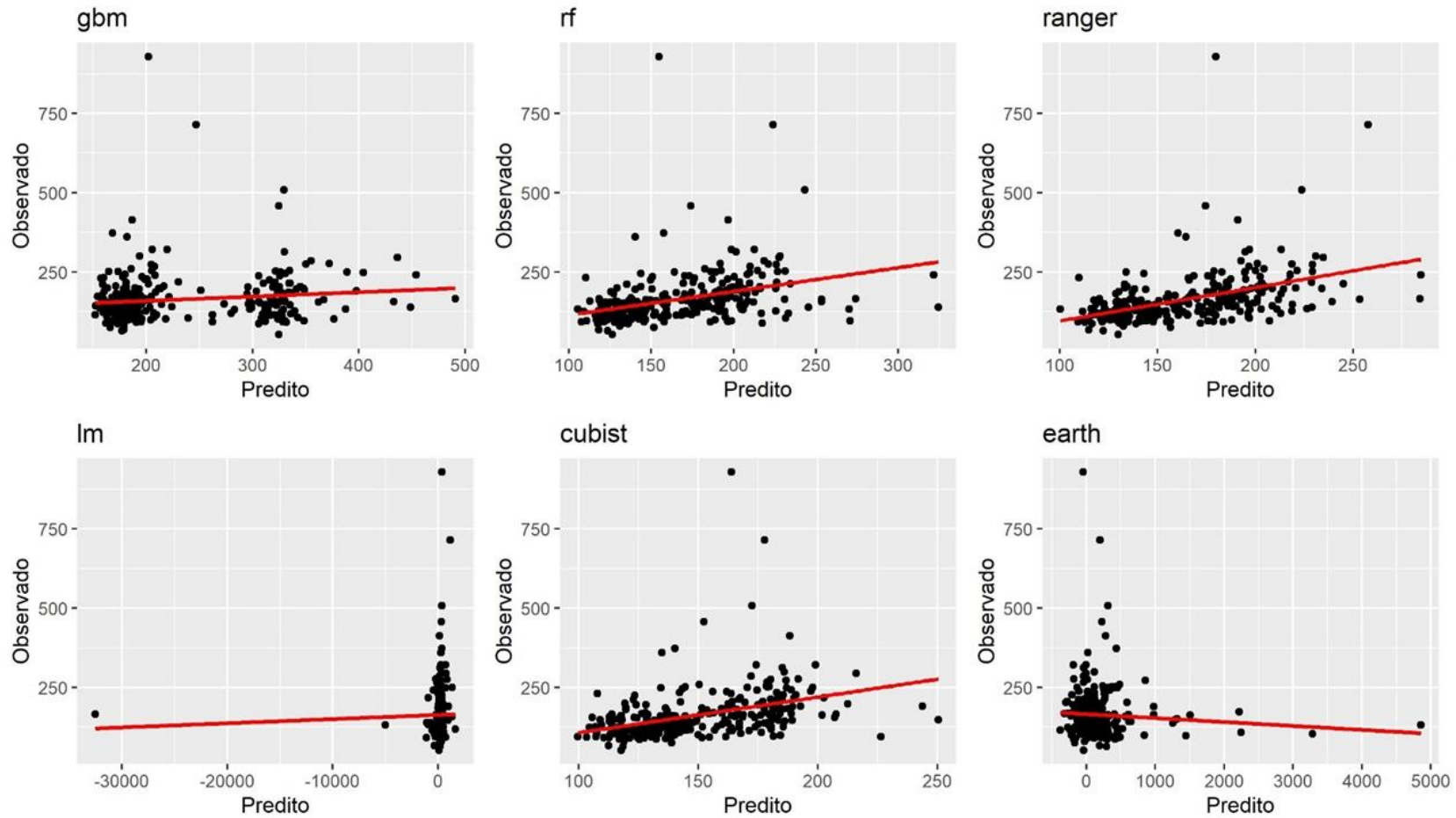
Figura 16 - Estimativa dos teores de Cobre para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



Cu

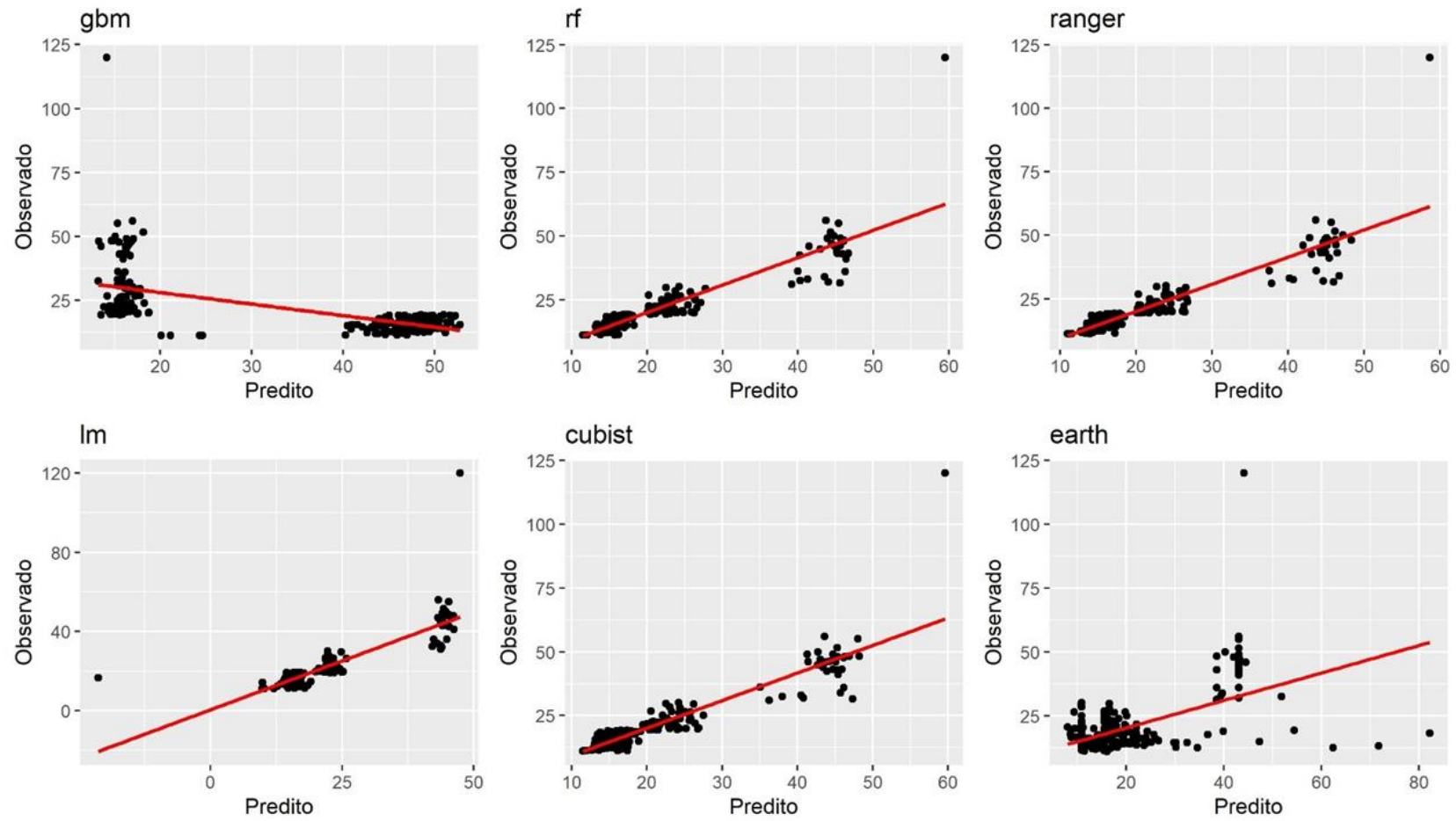
Fonte: Elaborado pela autora, 2022.

Figura 17 - - Estimativa dos teores de Ferro para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



Fonte: Elaborado pela autora, 2022.

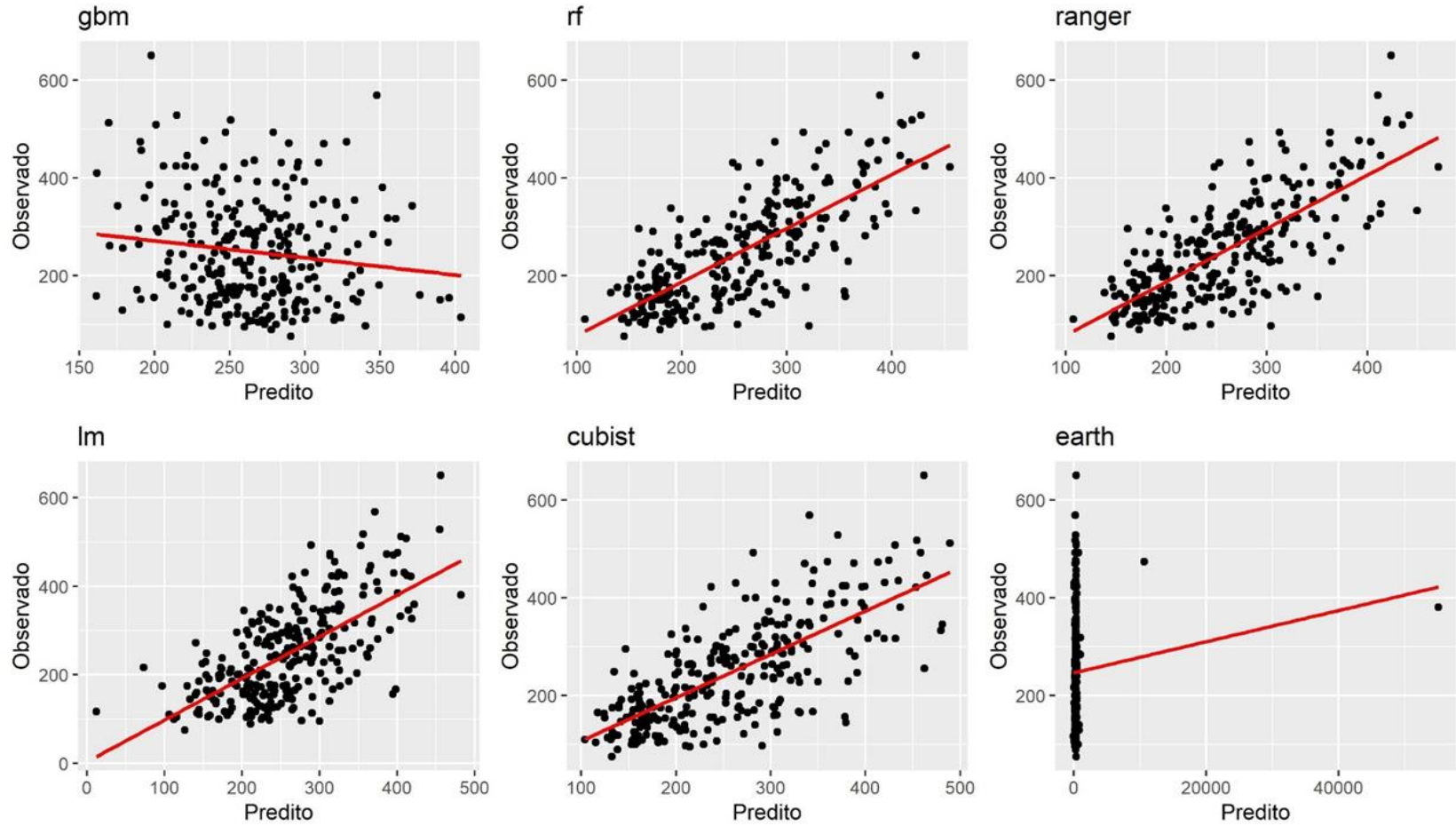
Figura 18 - Estimativa dos teores de Zinco para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



Zn

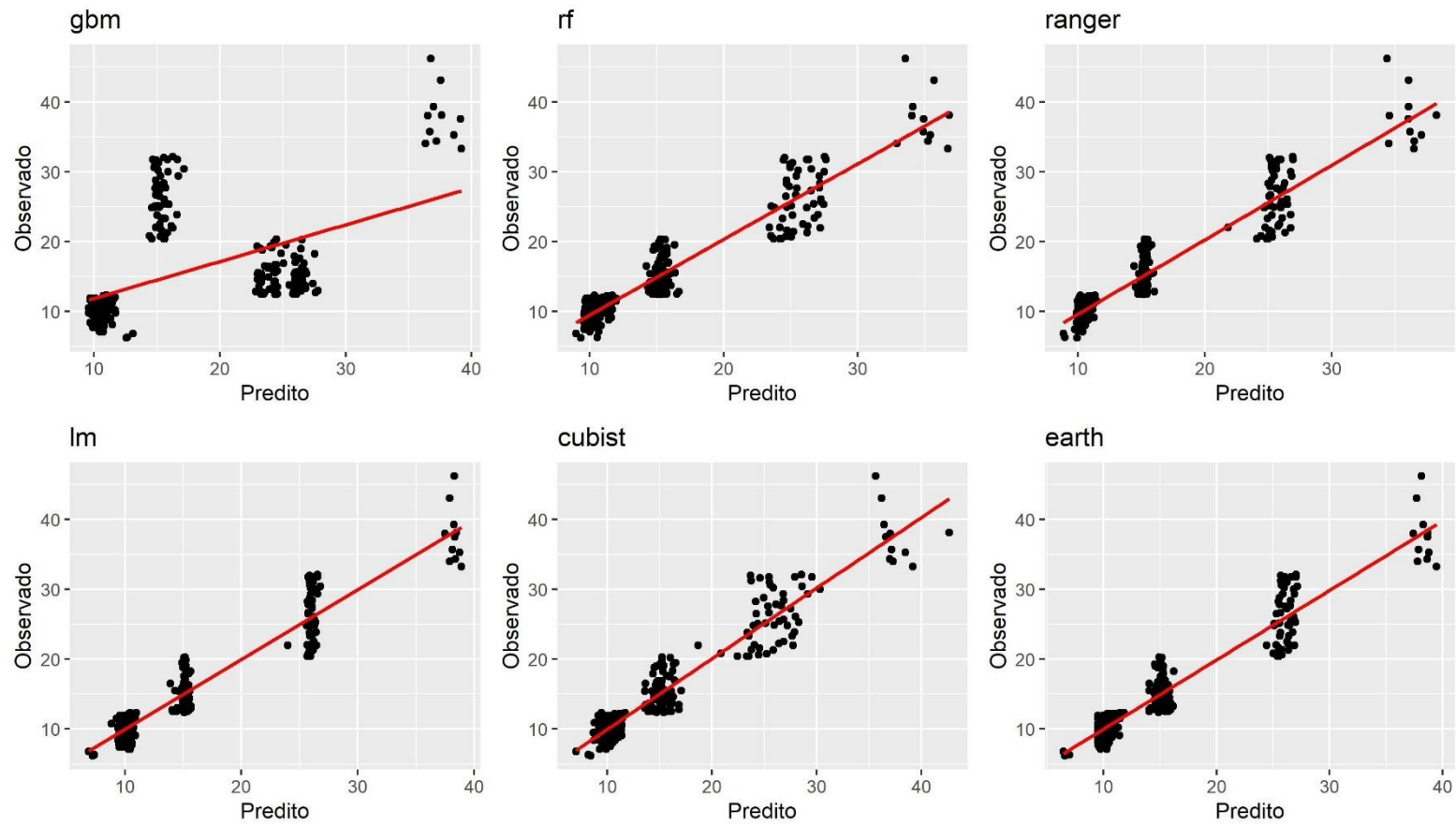
Fonte: Elaborado pela autora, 2022.

Figura 19 - Estimativa dos teores de Manganês para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



Fonte: Elaborado pela autora, 2022.

Figura 20 - Estimativa dos teores de Boro para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.

**B**

Fonte: Elaborado pela autora, 2022.

Nas Tabelas de 3 a 13, encontram-se os índices de vegetação com maior grau de importância, ou seja, aqueles que mais influenciam nos resultados de validação (R^2) e RMSE - Root Mean Squared Error - medida que calcula a raiz quadrática média dos erros entre valores observados (reais) e predições (hipóteses).

O N é o elemento mineral que as plantas requerem em maiores quantidades, cerca de 1 a 5% da matéria seca total da planta consiste em N. Ele é essencial e constituinte de muitos componentes celulares vegetais, incluindo clorofila, aminoácidos e ácidos nucleicos. Por isso, quando ocorre a deficiência de N, rapidamente inibe o crescimento das plantas, neste contexto, ao analisar os índices que se destacaram como promissores para monitorar teores de N, percebe-se que a grande maioria são capazes de detectar mudanças no teor de clorofila. É importante destacar que embora o NDVI seja o índice de vegetação mais popular dentre os demais, há um conjunto de índices que apresentam diferentes respostas à concentração de clorofila e foram mais responsivos.

Muitos destes índices foram aprimorados para capturar as menores variações de reflectância ou absorção, utilizando os intervalos de comprimento de onda estreitos, podendo proporcionar uma maior sensibilidade às alterações nutricionais das plantas. Diversos estudos realizados em culturas agrícolas buscam comparar a eficiência de diferentes índices para detectar o teor de N em diferentes estádios fenológicos, os resultados obtidos por Amaral *et al.* (2020) mostram que além do NDVI, o MTCI, CI e MCARI demonstraram grande potencial para serem utilizados no monitoramento de N na cultura do milho. Ladeia (2020), buscou estimar os teores de N a partir de sensores ativos e passivos na cultura do café, e obteve correlação positiva com os índices VARI, GLI e GRVI.

Os resultados (Tabela 3), mostram que os índices VARI, MCARI, SLAVI, NDI, ARVI, SLAVI e NDWI_SWIR como promissores na detecção de teor de N-foliar, pois revelam informações que não podemos detectar a olho nu, as fórmulas empregadas nessas análises permitem determinar diferentes níveis de clorofila na lavoura, perceber de forma antecipada possíveis estresse na vegetação, minimizar os efeitos atmosféricos que podem influenciar na qualidade das imagens de satélites. Há grandes avanços em estudos utilizando espectrorradiometria para análise nutricional em plantios de eucalipto, estudos feitos por Pedrali *et al.*, 2022 obtiverem resultados de correlação acima de 70% para os elementos N e P, indicando os índices reNDVI, CHLRED-EDE, DSWI e MCARI1 como satisfatórios para monitorar N e o índice ND para o P.

Neste caso, observa-se que os índices de vegetação relacionados à borda do vermelho possibilitaram obter maior correlação com o N. Isso foi observado em outros estudos

com as culturas Capim Tanzânia e Eucalipto, o qual os melhores índices de vegetação utilizados para discriminar os teores de N foram aqueles resistente à atmosfera na região do visível, sendo, portanto, utilizado a banda de transição do vermelho ao infravermelho próximo o que resultou em estimativas mais precisas de nitrogênio foliar (OLIVEIRA *et al.*, 2017; ABRAHÃO *et al.*, 2009).

Em relação ao elemento P, seus sintomas de deficiência não são tão marcantes como os outros macronutrientes, e as características mais evidentes são a redução da expansão foliar e o número de folhas, que reflete em redução do crescimento como um todo. Diferentemente da deficiência de N, a mudança de cor das folhas não está relacionada à clorose. No caso do P, as folhas mais velhas ficam com coloração verde escura, porque a expansão das folhas é muito mais inibida do que a formação da clorofila, e em alguns casos pode ocorrer de ficarem com cores púrpuras ou avermelhadas em consequência da acumulação de antocianina. Ao analisar a tabela 4, é possível verificar que os índices ARI e CRI se destacaram na análise de importância, ambos utilizam medições de reflectância no espectro visível para avaliar as características espectrais de pigmentos relacionados ao estresse vegetativo, como os carotenoides e antocianinas.

Além destes, também se mostraram promissores os índices que monitoram o vigor da vegetação e índice de área foliar, como o MTCI, MCARI, MGRI e LAI.

Tabela 3 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de nitrogênio foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Nitrogênio	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	MCARI1	ARVI	VARI	B03	NDWI_SWIR2	NBR
2	CI2	B02	ARVI	DVI	NBR	NDWI_SWIR2
3	GRVI	CI1	ARI1	MCARI2	ARVI	NDI
4	MGVRI	RGVI	B12	B08	VARI	NCI1
5	MCARI1		ARI2	B8A	SLAVI2	ARVI
6	MSR		B12	B8A		
7	MGVRI					
8	NDI					
R^2 Trein.	0,78	0,84	0,67	0,82	0,83	0,83
RMSE Trein.	6,04	4,27	1,17	1,90	5,21	2,12
R^2 Valid.	0,76	0,84	0,65	0,79	0,83	0,82
RMSE Valid.	4,72	6,62	1,22	3,45	1,81	3,04

Fonte: Elaborado pela autora, 2022.

Tabela 4 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de fósforo foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Fósforo	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	ARI1	ARI1	ARI1	DVI	ARI1	NDI45
2	LAI	LAI	MTCI	MCARI2	ARI1	ARI1
3	NDI45	ExG	NCI2	CRI2	ARI1	ARI1
4	ARI1	CHL_RE D _EDGE	CI1	GOSAVI	CRI1	ExGexR
5	NDI45		ARI1	MGVRI	ARI1	EVI
6	ARI1		ARI1	ARVI	B03	ARI1
7	CRI1		PSSR			
8			ExG			
R^2 Trein.	0,74	0,82	0,67	0,72	0,81	0,82
RMSE Trein.	0,09	0,08	0,11	0,10	0,08	0,08
R^2 Valid.	0,73	0,81	0,68	0,70	0,80	0,81
RMSE Valid.	0,10	0,09	0,11	0,11	0,09	0,09

Fonte: Elaborado pela autora, 2022.

Tabela 5 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de potássio foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Potássio	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	NBR	NBR	ARI1	ARI1	ARI1	ARI1
2	NDWI _SWIR2	NDWI _SWIR2	CRI1	B03	CRI1	CRI1
3	NCI2	NCI2	B02	CRI1	ARI1	B03
4	SLAVI2	SLAVI2			B03	
5	VARI	VARI				
6	B04	B04				
7	CI1	CI1				
R^2 Trein.	0,85	0,86	0,87	0,73	0,87	0,88
RMSE Trein.	7,22	7,94	6,68	15,25	9,48	8,70
R^2 Valid.	0,86	0,86	0,86	0,70	0,87	0,87
RMSE Valid.	3,08	2,42	2,59	13,39	1,23	1,33

Fonte: Elaborado pela autora, 2022.

O K^+ é o cátion mais abundante na planta e pode ser deslocado facilmente nos tecidos e células do vegetal. Tem importante função no estado energético, na translocação, armazenamento de assimilados e manutenção da água nos tecidos vegetais. Dentre os macronutrientes, o K foi o que apresentou menor número de índices na análise de importância (Tabela 5) e assim como o P, tiveram melhor desempenho o ARI e CRI, com exceção do algoritmo *Earth* que selecionou o NDWI_SWIR, SLAVI e VARI com bons indicadores para

monitorar este elemento na cultura da palma. Gökkaya *et al.* (2015) utilizaram lidar e modelarem nutrientes no nível de dossel em floresta do Canadá a partir de dados hiperespectrais e observaram inferior qualidade no ajuste dos modelos para K.

Os macronutrientes secundários foram os que apresentaram maior diversidade de índices (Tabelas 6, 7 e 8), com destaque para o NDVI e suas variações, como o RED_EDGE_NDVI, GRNDVI, NDI e GBNDVI que conseguem captar com maior sensibilidade às variações de clorofila e vigor das lavouras. Este resultado está de acordo com as principais funções destes elementos, como no caso do Mg, responsável por formar a molécula de clorofila, que são porfirinas magnesianas, onde o Mg corresponde a 2,7% do peso molecular destas e representa cerca de 10% do teor total de Mg na folha. Outro ponto importante é que a falta de Mg inibe o metabolismo de N, reduz o teor de N-proteico, aumentando a proporção de N-não proteico, ou seja, a carência de Mg interfere na síntese de proteína. Pedreli *et al.*, (2022), ao relacionar 24 índices de vegetação com os resultados nutricionais para a cultura do eucalipto, também obteve o NDVI com o melhor índice para estimar o teor de Ca na folha, obtendo estimativa de 0,78 (valor de R^2), resultado de validação muito similar com o que encontramos neste estudo.

Tabela 6 - - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de cálcio foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Cálcio	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	SLAVI	ARI1	ARI1	EVI2	ARVI	ARI1
2	MCARI1	VARI	RED_EDGE_NDVI	GRNDVI	CRI1	ARVI
3	NBR	NCI1	B04	MCARI1	VARI	CRI1
4	DVI	CI2	C12	SLAVI	NBR	MTCI
5	NDI45	BAI	ARVI	VARI	NDWI_SWI R2	BAI
6	MTCI	CRI2	VARI	ARI1	ARI	RED_EDGE_NDVI
7	MSAVI		CRI	ARI2	RED_EDGE_NDVI	
8			RGVI	ARVI	ARI	
R^2 Trein.	0,73	0,75	0,73	0,74	0,75	0,73
RMSE Trein.	0,66	0,62	0,65	0,64	0,62	0,63
R^2 Valid.	0,74	0,75	0,74	0,72	0,76	0,75
RMSE Valid.	0,66	0,64	0,65	3,54	0,63	0,63

Fonte: Elaborado pela autora, 2022.

Tabela 7 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de enxofre foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Enxofre	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	GNDVI	ARI1	CI2	GLI	BAI	MTCI
2	ARI2	ARVI	GBNDVI	SLAVI1	CI1	NCI1
3	GRNDVI	CI1	IPVI	ARI1	MSAVI	EVI
4	MSAVI	MSAVI2	NBR	ARVI	NCI1	NDI45
5	NDI45		RED_EDGE _NDVI	BNDVI	ARI1	RED_EDGE _NDVI
6				GRNDVI	GNDVI	
7				MSAVI	MCARI2	
R^2 Trein.	0,87	0,87	0,87	0,84	0,86	0,87
RMSE Trein.	0,13	0,13	0,18	0,15	0,13	0,13
R^2 Valid.	0,87	0,87	0,85	0,84	0,86	0,87
RMSE Valid.	0,13	0,14	0,15	0,15	0,14	0,14

Fonte: Elaborado pela autora, 2022.

Em relação ao S (Tabela 8), os resultados de R^2 validação foram muito promissores e com estimativas melhores quando comparadas ao do Ca e Mg. Pedrali *et al.*, 2022 ao avaliar as melhores estimativas para determinar teor de S em folha de *Eucalyptus saligna*, encontrou o máximo valor de 0,76 de R^2 . É importante destacar que o S também está intimamente relacionado com N, sendo inclusive utilizado na relação N/S da planta para avaliar seu estado nutricional. Muitos dos sintomas de deficiência de S são similares aos de N, incluindo a clorose, redução do crescimento e acumulação de antocianinas, isso ocorre devido ambos serem constituintes de proteínas e um ponto importante nesta análise é perceber a similaridade dos índices para monitorar a condição dos teores de N, Ca, Mg e S.

Entre os micronutrientes, percebe-se um predomínio do índice NDVI e suas variações como importantes preditores para condição nutricional, com destaque para Zn (Tabela 9), Cu (Tabela 10) e B (Tabela 13), pois apresentaram os melhores resultados de validação, com valor máximo de 0,89 para Zn, 0,79 para Cu e 0,90 para B. Resultados similares também foram obtidos por Pedrali *et al.* (2022), onde os dados estimados para o Zn, foram de R^2 0,91 e do Cu R^2 0,85.

Tabela 8 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de zinco foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Zinco	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	ARI	VARI	ARI	GLI	VARI	ARI
2	BAI	ARI	BAI	EVI	ARI	NCI
3	CRI1		SLAVI	GRNDVI	ARVI	
4	NCI		MTCI	MCARI	GCI	
5			GRNDVI			
6			NCI			
7						
8						
R^2 Trein.	0,88	0,86	0,86	0,79	0,88	0,89
RMSE Trein.	3,51	3,38	3,04	3,85	3,16	3,12
R^2 Valid.	0,87	0,83	0,87	0,78	0,87	0,88
RMSE Valid.	3,57	4,14	3,28	4,50	3,35	3,12

Fonte: Elaborado pela autora, 2022.

Tabela 9 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de cobre foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Cobre	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	EVI	ARI	EVI	GLI	EVI	SLAVI
2	MGVRI	LAI	SLAVI	MTCI	LAI	LAI
3	BNDVI	GCI	ARVI	ARVI	PSRI	EVI
4	MAVI		CRI	GDVI	CRI	ARI
5	MTCI		RED_EDGE _NDVI	NCI	VARI	MTCI
6	SLAVI			NDI	ARVI	
7						
8						
R^2 Trein.	0,77	0,77	0,76	0,70	0,79	0,79
RMSE Trein.	1,27	0,84	4,03	1,93	0,79	0,79
R^2 Valid.	0,76	0,75	0,76	0,69	0,76	0,76
RMSE Valid.	3,31	2,43	2,79	4,02	3,44	3,41

Fonte: Elaborado pela autora, 2022.

Tabela 10 - - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de ferro foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Ferro	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	ARI	ARI	BNDVI	ARVI	ARI	ARI
2	BNDVI	BNDVI	ARI	BAI	BNDVI	EVI
3	EVI	ARVI	EVI	MCARI	CRI	LAI
4	NCI		NCI2	SLAVI	NCI	RED_EDGE _NDVI
5	NDI45		VARI	ARI	GLI	PSRI
6				BNDVI	LAI	
				CRI		
				DVI		
				EVI		
7						
R^2 Trein.	0,19	0,19	0,18	0,14	0,25	0,24
RMSE Trein.	5,95	5,70	6,71	7,02	5,92	6,46
R^2 Valid.	0,13	0,11	0,12	0,11	0,20	0,17
RMSE Valid.	7,70	7,78	7,42	4,33	6,81	7,20

Fonte: Elaborado pela autora, 2022.

O nutriente Fe apresentou os menores resultados preditores (Tabela 11), indicando que para este micronutriente a partir destes índices e modelos ainda não é possível fazer o monitoramento com base em dados espectrais. O Mn obteve comportamento similar (Tabela 12), mostrando a necessidade de melhor compreensão dos índices e amplitudes das bandas espectrais podem ser mais bem trabalhadas para monitorar o comportamento nutricional destes elementos na planta.

Tabela 11 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de manganês foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Manganês	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	NDI45	NCI	MTCI	ARI	MTCI	MTCI
2	PSRI	ARI	PSRI	CI1	PSRI	ARVI
3	VARI	NDI	RED_EDGE_NDVI	SLAVI	ARVI	NDI
4	GLI	MTCI	ARVI	GCI	LAI	RED_EDGE_NDVI
5	ARI		NDI	GRNDVI	EVI	
6	MAVI				VARI	
7					GRVI	
8						
R^2 Trein.	0,63	0,63	0,63	0,50	0,63	0,66
RMSE Trein.	6,27	6,24	6,22	6,17	6,79	6,23
R^2 Valid.	0,62	0,60	0,63	0,50	0,61	0,64
RMSE Valid.	6,48	6,17	5,73	6,54	4,31	5,80

Fonte: Elaborado pela autora, 2022.

Tabela 12 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição do teor de boro foliar em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Boro	Cubist	Earth	GBM	Lm	Ranger	Random Forest
1	NCI	VARI	NCI	NCI2	NCI	NCI
2	ARI	ARI	ARVO	PSRI	ARI	MTCI
3	RED_EDGE_NDVI	NCI	VARI	GRNDVI		RED_EDGE_NDVI
4		MTCI	GRNDVI	MCARI		
5			MTCI	RBNDVI		
6				RED_EDGE_NDVI		
7						
8						
R^2 Trein.	0,88	0,89	0,88	0,62	0,90	0,90
RMSE Trein.	2,32	2,15	2,23	3,97	2,36	2,22
R^2 Valid.	0,88	0,87	0,88	0,61	0,89	0,89
RMSE Valid.	2,46	2,36	2,12	4,11	2,27	2,24

Fonte: Elaborado pela autora, 2022.

Após identificar os índices mais promissores, os resultados da correlação contribuíram para compreender como elas interagem entre si. Esse diagrama foi elaborado empregando-se a correlação de Spearman (r), sendo que quanto maior o valor de r , maior é a intensidade das cores azul (correlação positiva entre as variáveis), e vermelho indica correlação negativa.

Nas figuras 5 e 6, é possível notar que há elevado grau de interação entre SLAVI com NDI45, VARI, GRVI, PSSR, MGVRI, MCARI, são índices de vegetação sensíveis à detecção de clorofila, que são pigmentos fotossintetizantes com fortes picos de absorbância no vermelho e no azul, regiões do espectro eletromagnético. Portanto, utilizar diversos índices para verificar qual melhor se ajusta a realidade da cultura faz-se necessário, principalmente considerando que os teores relativamente baixos de clorofila são suficientes para saturar a absorção na região de 660-680 nm, reduzindo assim a sensibilidade a altos teores de clorofila de variáveis espectrais com base nesses comprimentos de onda (LICHTENTHALER, 1987; OLIVEIRA *et al.*, 2017). Dentre os algoritmos, os que apresentaram maior grau de correlação entre os índices foram o *Cubist* e *Ranger*, tornando-se importantes modelos para predição de N foliar.

No algoritmo *Ranger* (Figura 6), é possível perceber alta correlação entre os índices ARVI e SLAVI. É importante frisar que o ARVI é mais sensível quando comparado ao NDVI, por considerar na sua fórmula as correções de efeitos atmosféricos na região do vermelho. A interação entre ARVI e SLAVI além de ser importante para avaliar o vigor da vegetação também contribui para estimar a área foliar e acompanhar a condição fotossintética dos plantios, além disso as interações destes com o NDWI_SWIR, indicam o potencial em conseguir extrair informações relacionadas com o teor de água na folha. Esse monitoramento pode contribuir de forma muito expressiva com as estratégias do manejo nutricional dos plantios. O refinamento das informações permite escolher de forma técnica quais as áreas que podem ter suas adubações postergadas por certo período sem perdas ou grandes perdas na produtividade e quais aquelas que necessitam ter aporte de nutrientes para evitar quedas na produção.

Estudos realizados por Wright *et al.* (2004) em cultivos de trigo submetidos a diferentes concentrações de N, observaram que os maiores valores de reflectância no espectro visível se relacionavam com os tratamentos submetidos a menores taxas de adubação nitrogenada e que a adição de N implica na diminuição da reflectância espectral no visível e no aumento da reflectância espectral entre 700 e 1300 nm e que permitiu utilizar índices de vegetação para monitorar o N na cultura de trigo.

Diversas pesquisas mostram que o MCARI e MCARI2 também apresentam uma resposta muito positiva para estimar a clorofila e variações de IAF em dosséis agrícolas e se destacou com melhor coeficiente determinação (0,82) para mapeamento de vegetação, superando os resultados do NDVI (ZHANG *et al.*, 2014).

O P foi o elemento que apresentou a menor correlação entre os índices, apenas o *Random Forest* (Figura 10), obteve algumas interações entre índice de área foliar e detecção do

vigor da lavoura. Estudos realizados por Silva *et al.*, 2021, utilizaram diferentes algoritmos para predição de crescimento de árvores de eucalipto e obtiveram o *Random Forest* como melhor modelo de precisão. Isso foi apresentado por Sousa 2019, onde o *Random Forest* mostrou maior potencial para modelagem do conjunto de dados objetivando determinar o volume de madeira em plantios de eucalipto.

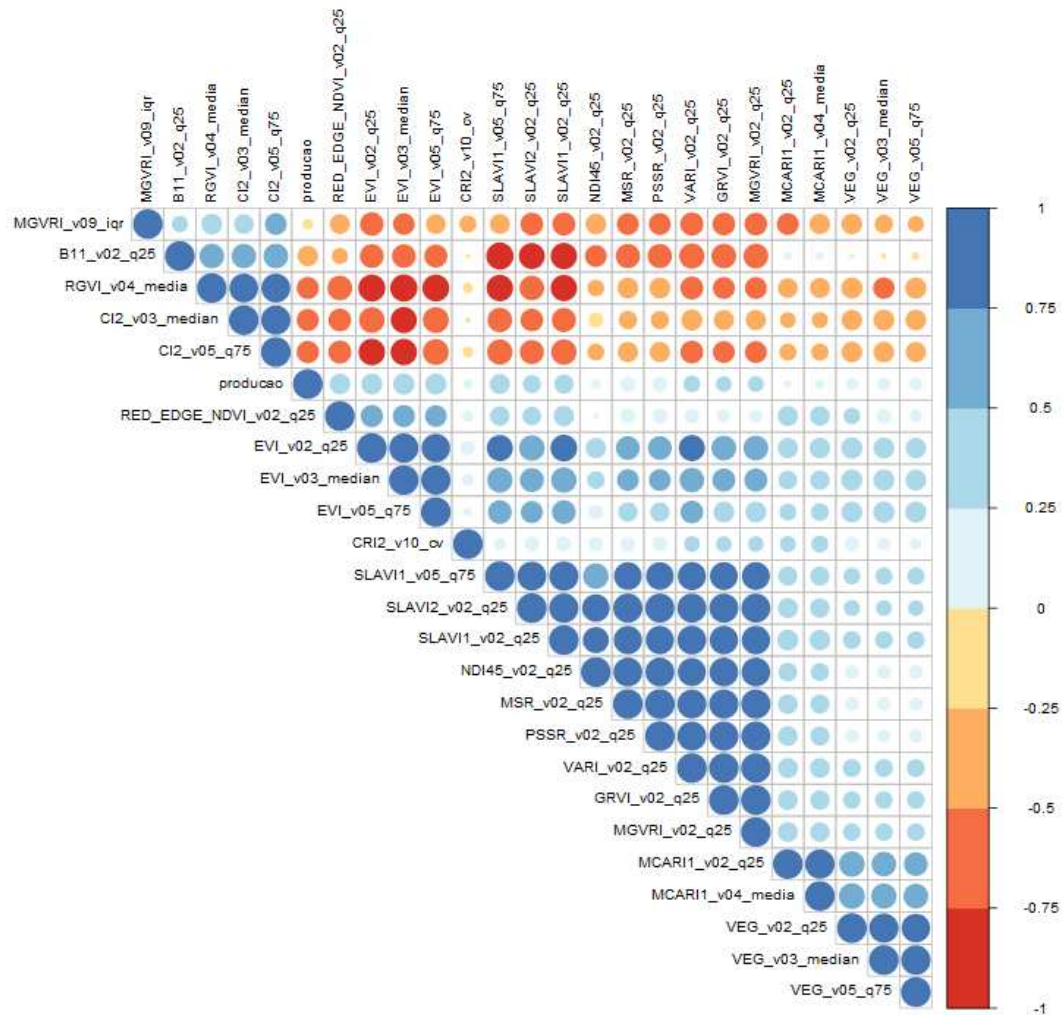
No caso do K (Figura 11), o *Cubist* foi o modelo com maior interação entre os índices, mostrando que os valores relacionados com a presença de antocianinas refletem em teores detectados de clorofila. Indicando que ambos podem ser utilizados como modelos complementares para melhor compreensão dos fatores que podem influenciar os balanços nutricionais. Diferentemente do que ocorreu com o *Cubist*, os modelos *Ranger* e *Random Forest* destacam-se nas interações entre CRI e ARI.

Para o Ca (Figura 14) e Mg (Figura 17), os modelos que tiveram melhor performance foram *Cubist* e *Random Forest*. O NDWI, VARI e SLAVI foram os que mais contribuíram para a predição do teor de Ca. No caso do S (Figuras 20, 21 e 22), há certo padrão de interação entre os índices, apresentando comportamento similar aos de N. O índice NDVI e suas variações GNDVI, DVI, GRNDI, MSAVI foram os que mais se correlacionaram nos algoritmos *Cubist* e *Linear Model*.

Para os micronutrientes, o Cu apresentou índices muito baixos de correlação, como pode ser observado nas Figuras 23 e 24. Na Figura 25 houve correlação mais expressiva somente entre o índice ARI. Diferentemente do que ocorreu com o ferro, onde os modelos *Cubist* (Figura 26), e *Linear Model* (Figura 28), apresentaram altas correlações com diversos índices. O mesmo comportamento pode ser observado nas Figuras 29, 30 e 31, mostrando a alta correlação entre índices que detectam teor de clorofila e área foliar, importante parâmetro para monitoramento do Mn.

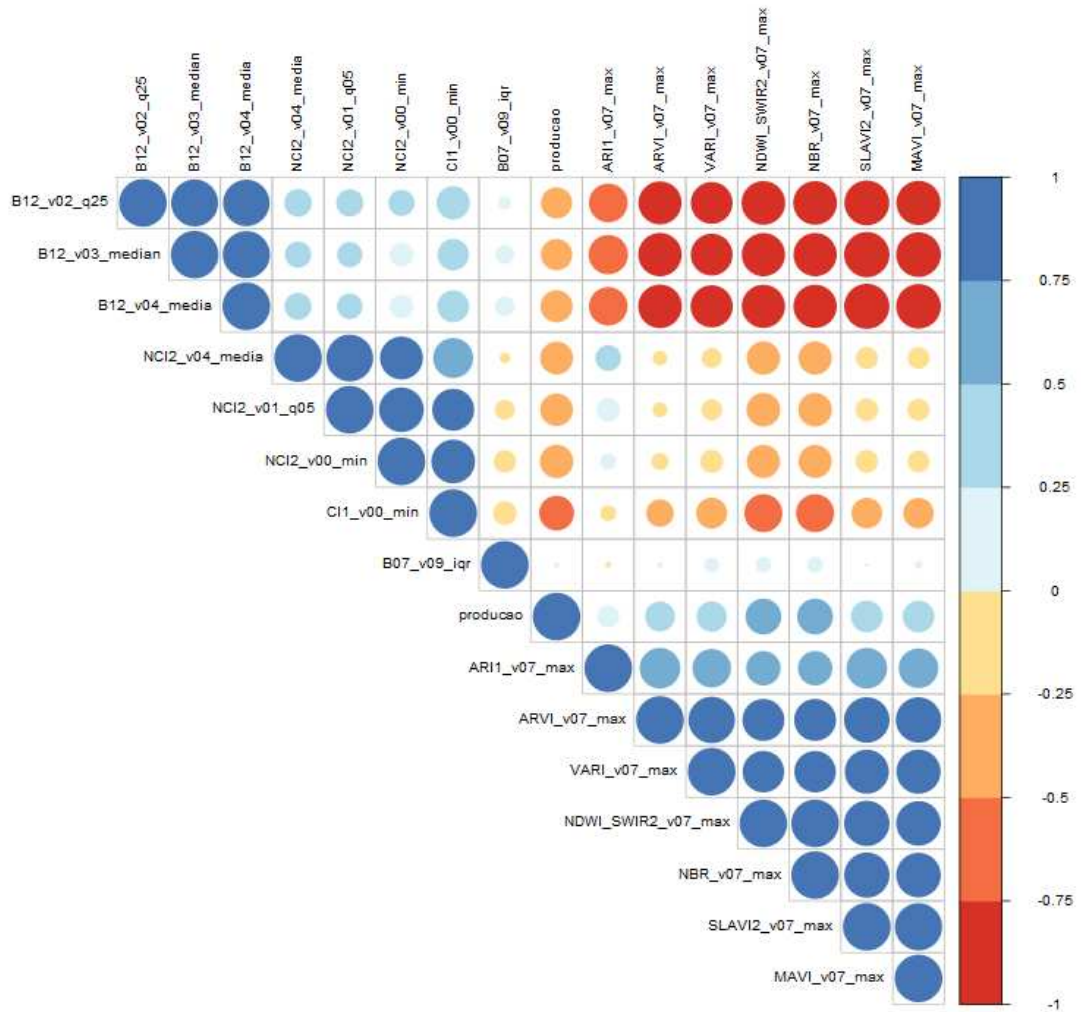
Em relação ao Zn, os modelos *Earth* (Figura 32) e *SVLinear* (Figura 33), se comportaram de forma similar. No caso do B, o ARI foi o índice que mais se destacou nas correlações, principalmente no modelo *Random Forest* (Figura 34), onde apresentou alta correlação entre si, o mesmo ocorreu para o *Earth* (Figura 36), comprovando a influência do ARI em detectar teores de boro. Em contrapartida, no modelo *Linear Model* (Figura 35), as fracas interações não demonstram influência de outros índices.

Figura 21 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento nitrogênio, modelo *Cubist*.



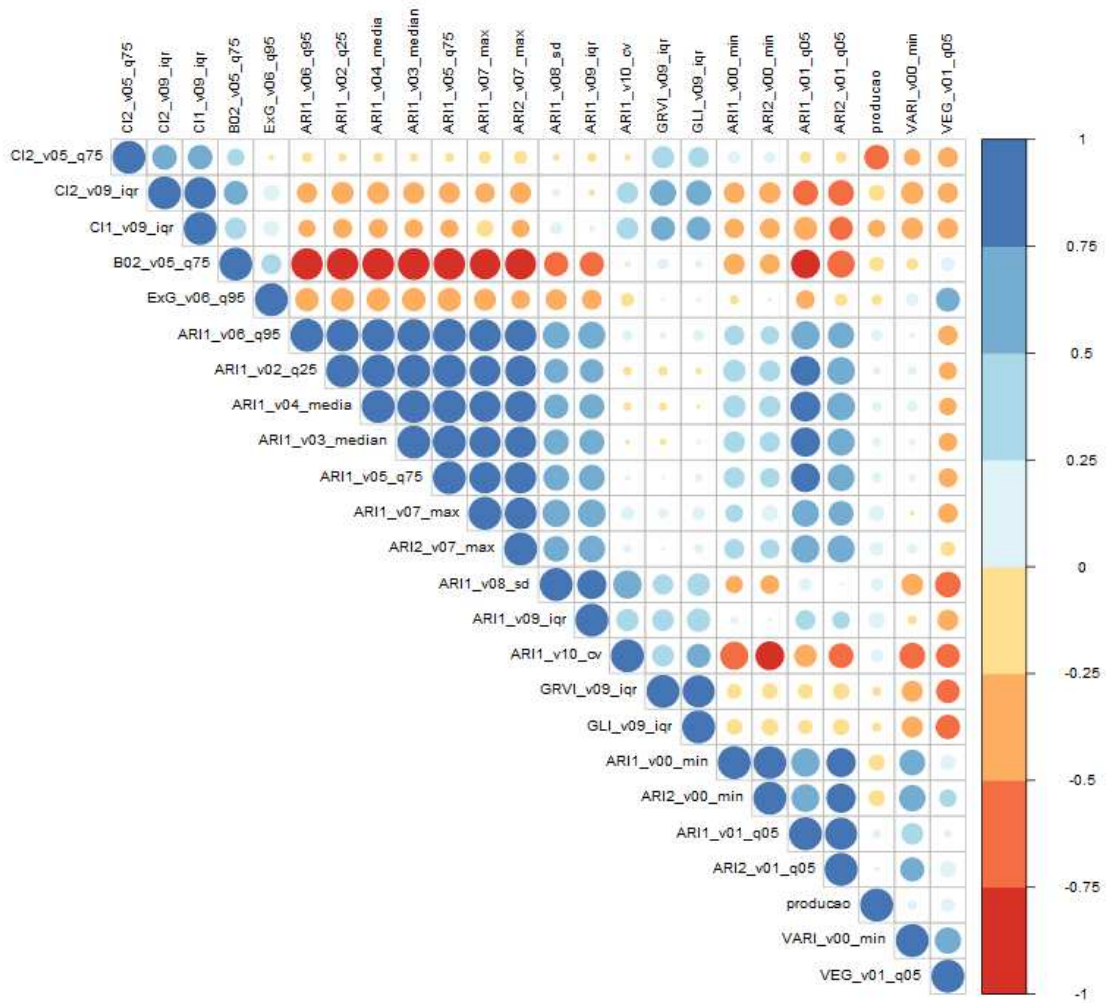
Fonte: Elaborado pela autora, 2022.

Figura 22 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento nitrogênio, modelo *Ranger*.



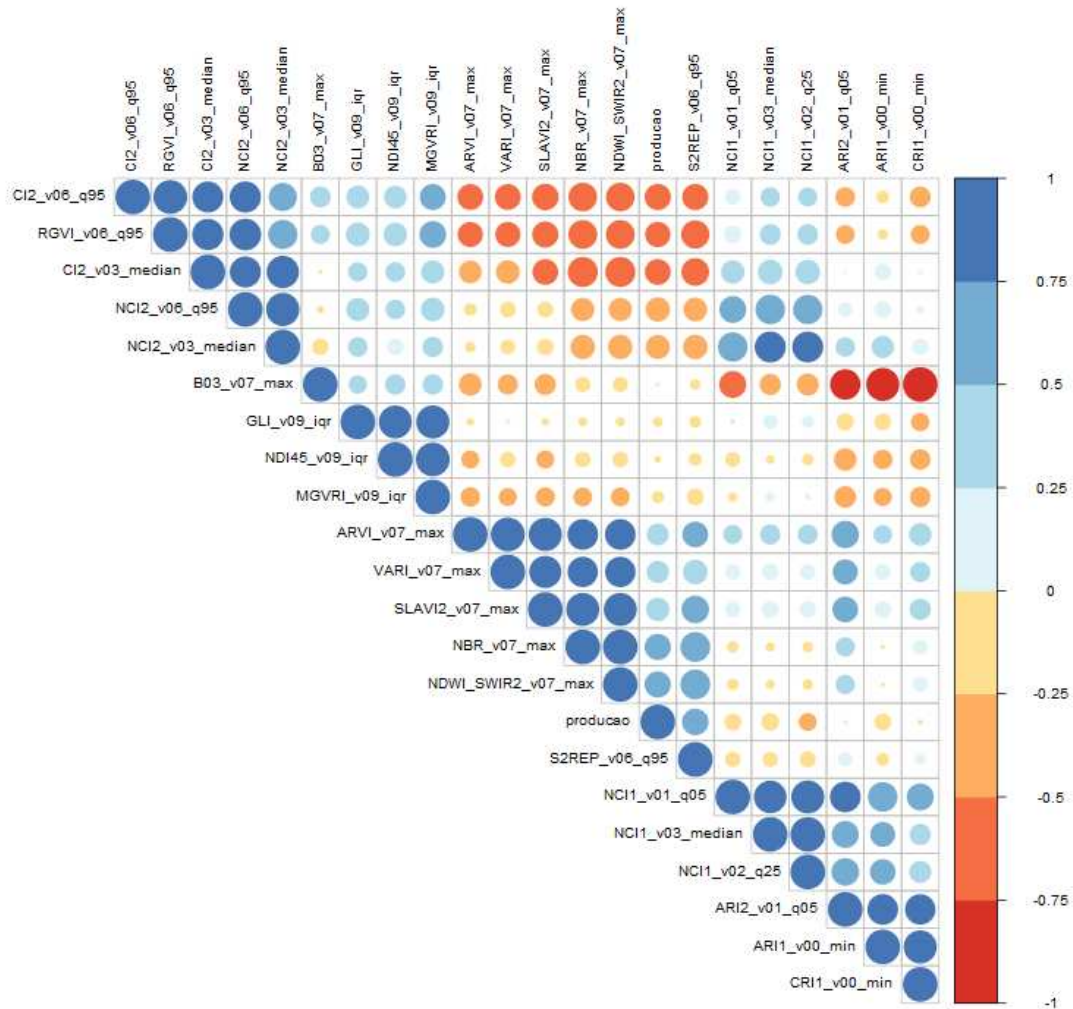
Fonte: Elaborado pela autora, 2022.

Figura 23 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento nitrogênio, modelo *Earth*.



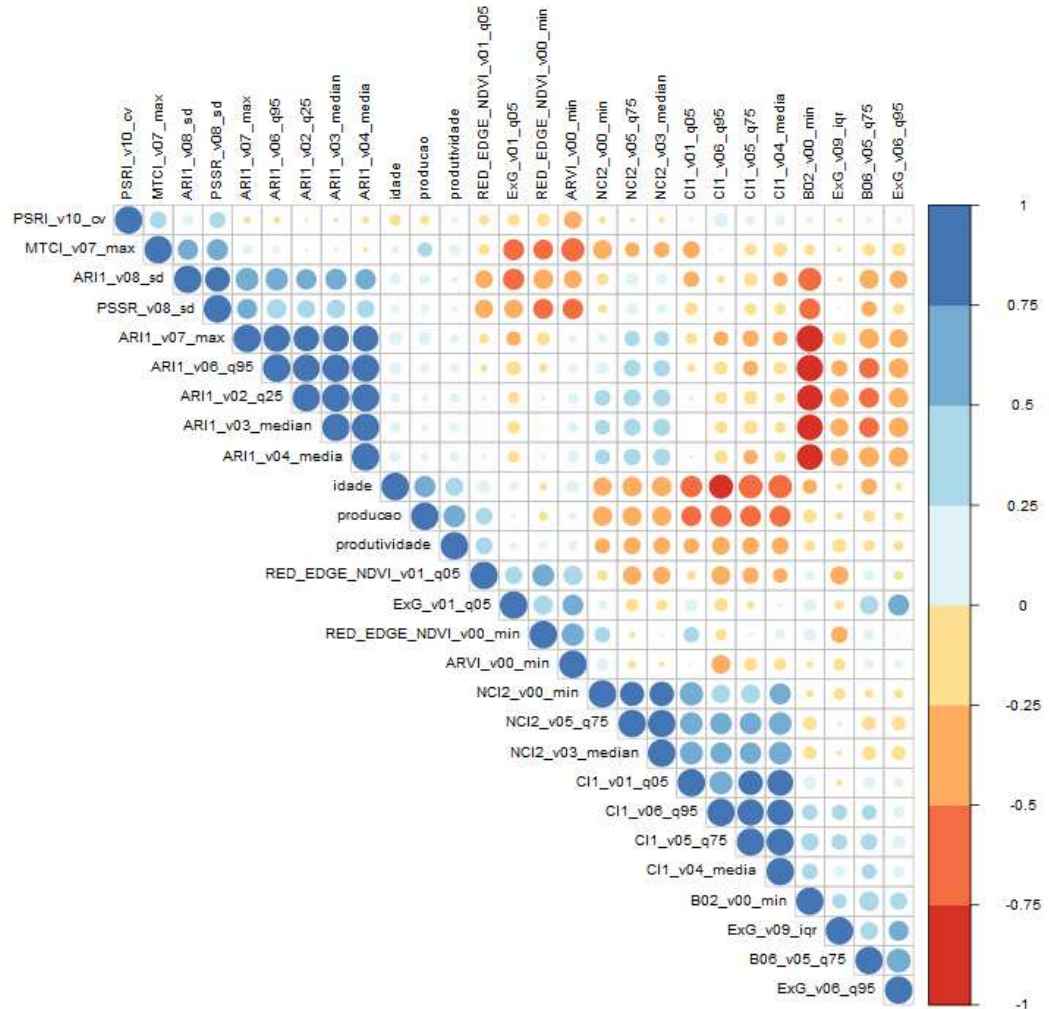
Fonte: Elaborado pela autora, 2022.

Figura 24 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento nitrogênio, modelo *Random Forest*.



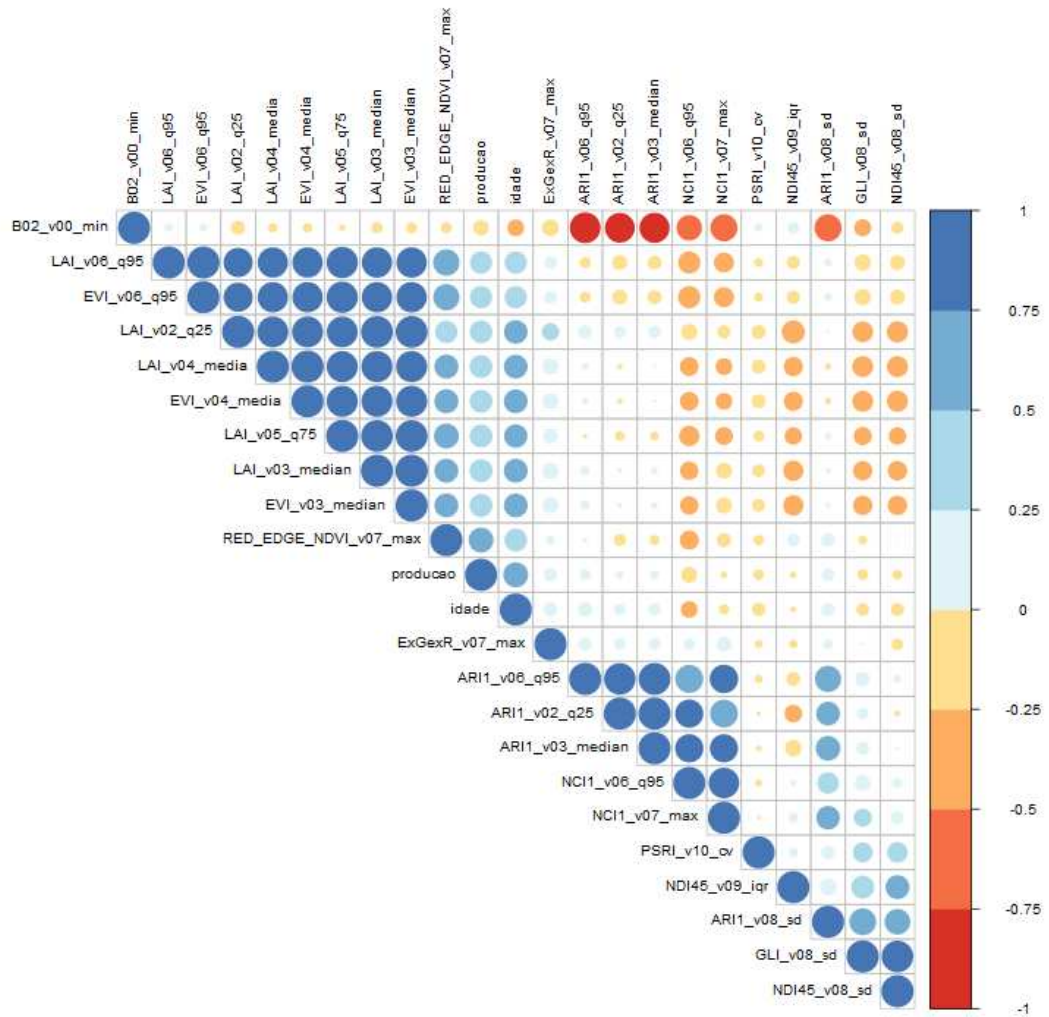
Fonte: Elaborado pela autora, 2022.

Figura 25 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento fósforo, modelo GBM.



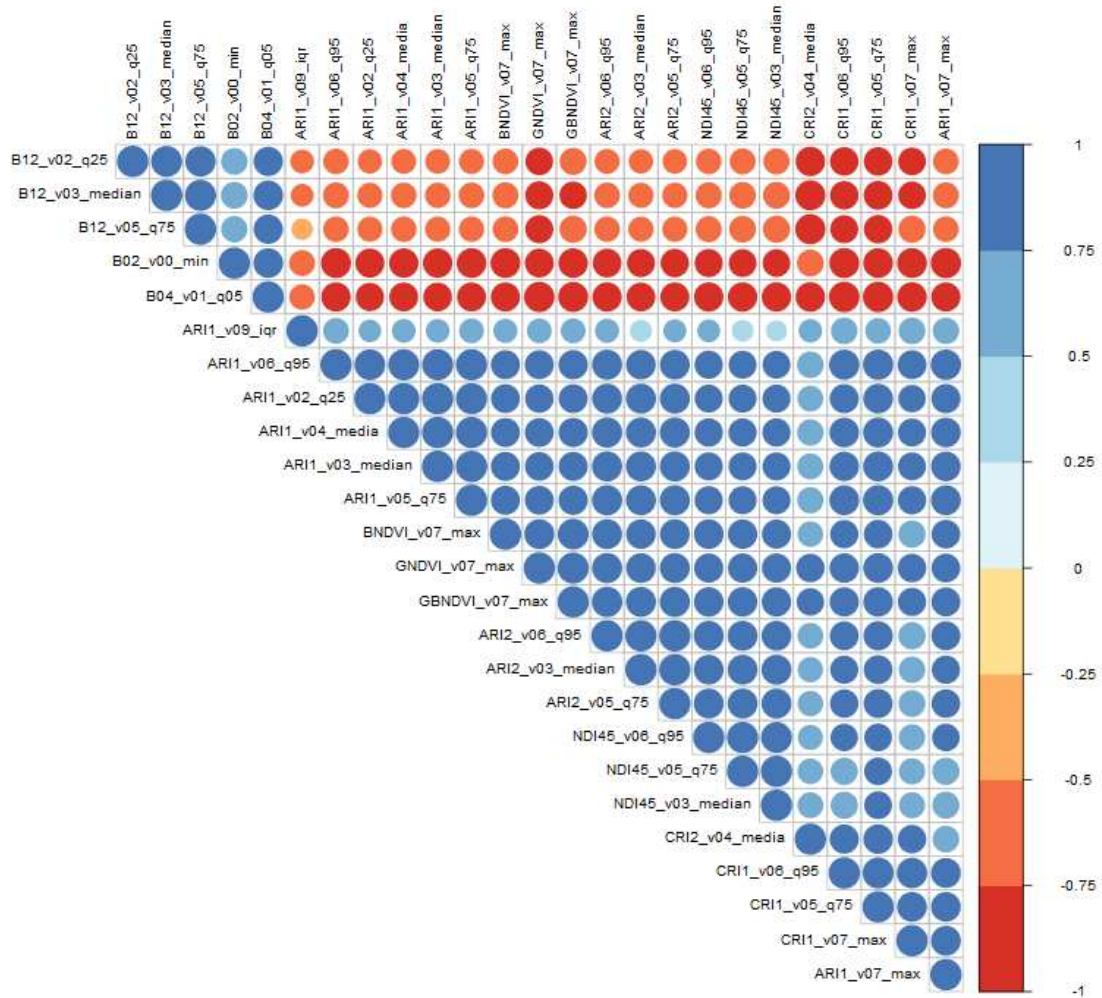
Fonte: Elaborado pela autora, 2022.

Figura 26 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento fósforo, modelo *Random Forest*.



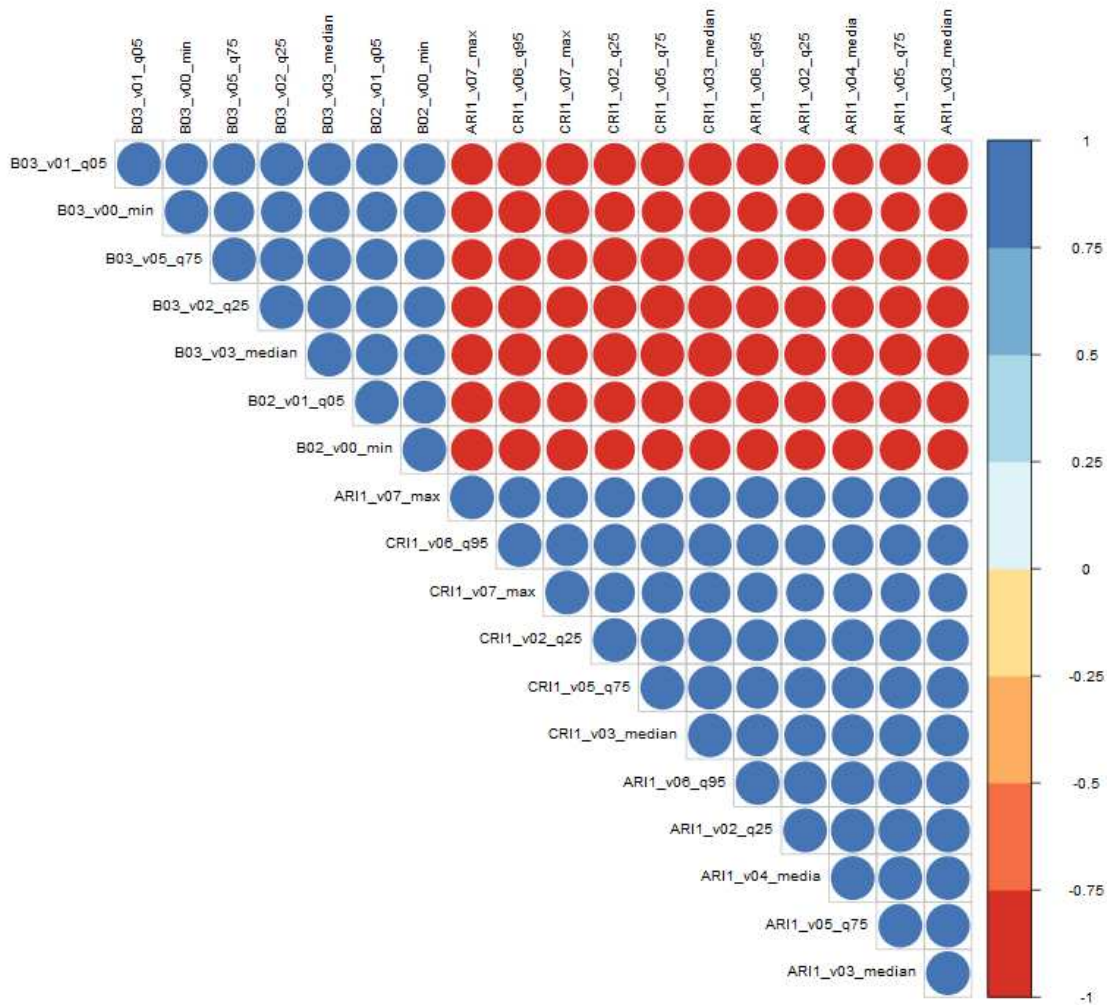
Fonte: Elaborado pela autora, 2022.

Figura 27 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento potássio, modelo *Cubist*.



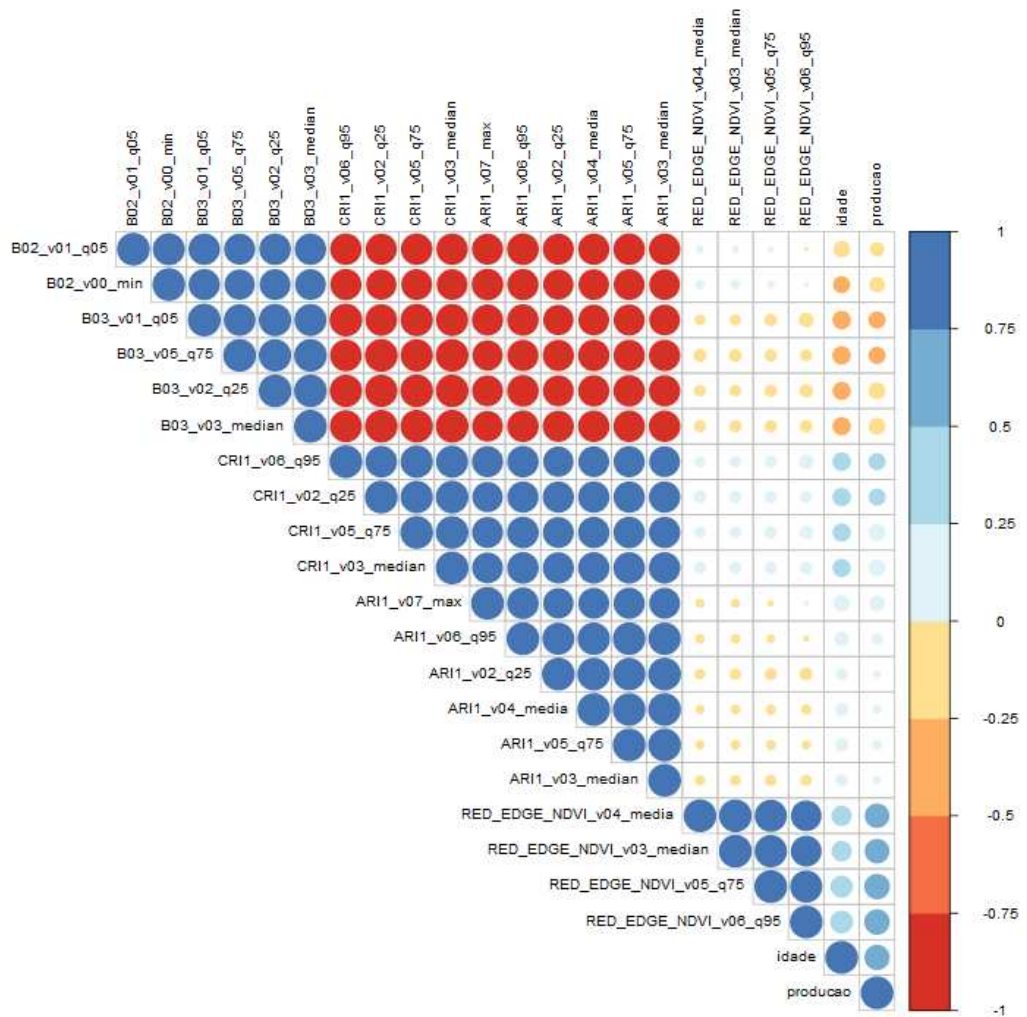
Fonte: Elaborado pela autora, 2022.

Figura 28 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento potássio, modelo *Ranger*.



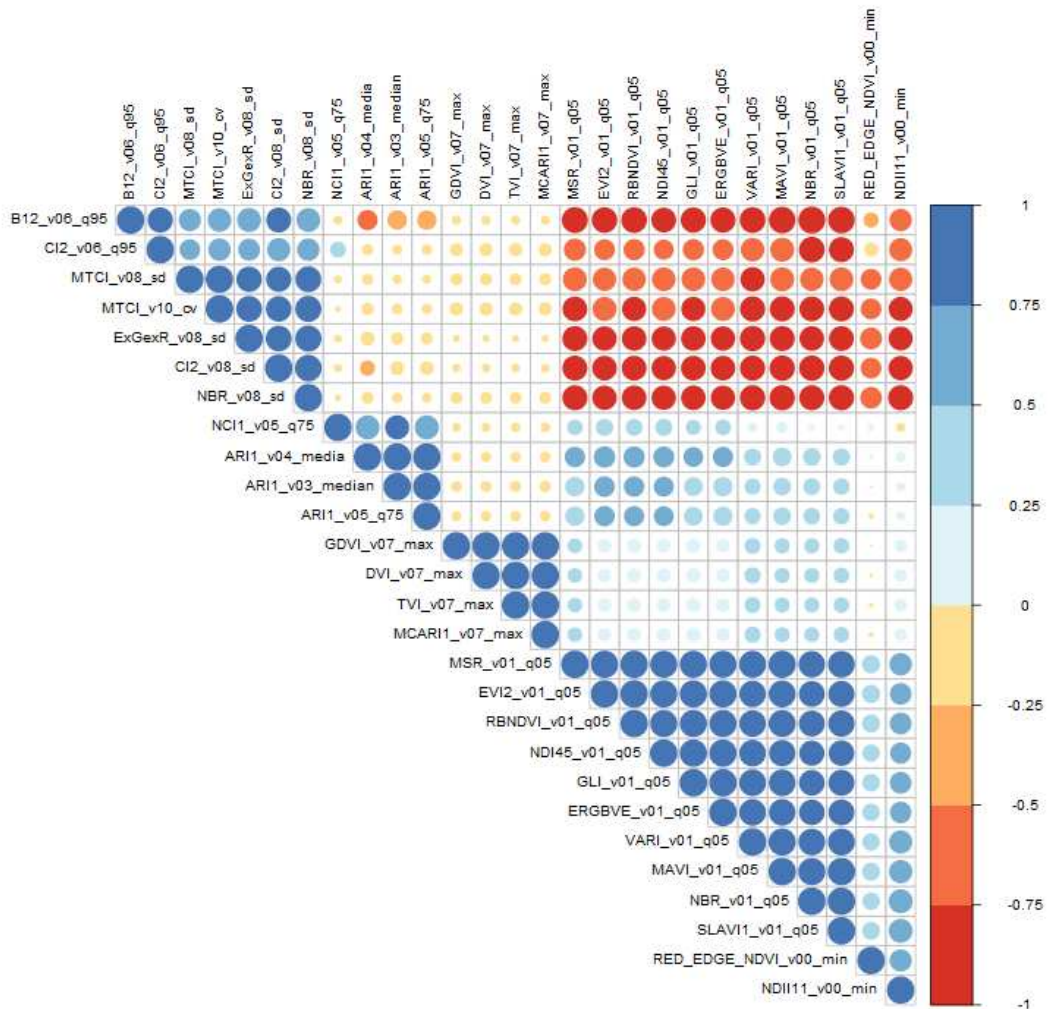
Fonte: Elaborado pela autora, 2022.

Figura 29 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento potássio, modelo *Random Forest*.



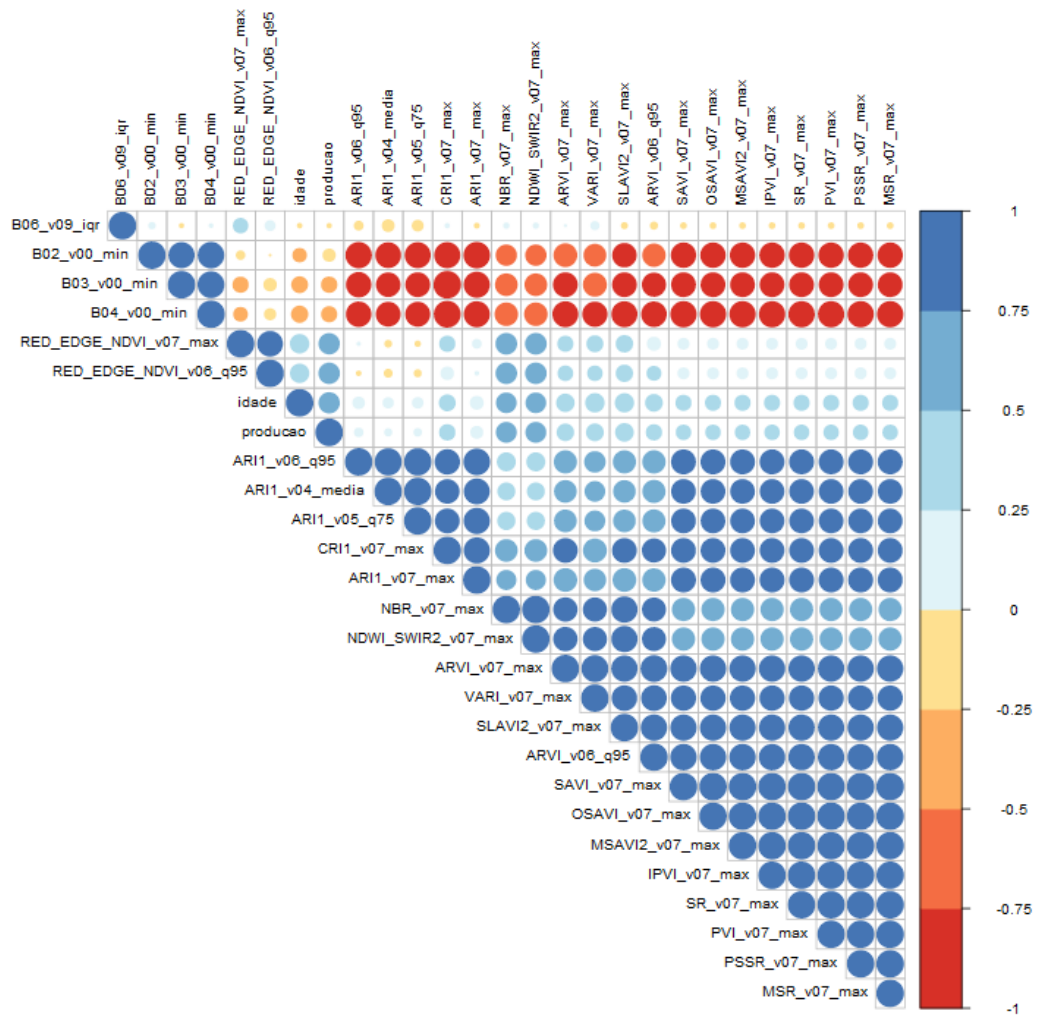
Fonte: Elaborado pela autora, 2022.

Figura 30 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cálcio, modelo *Cubist*.



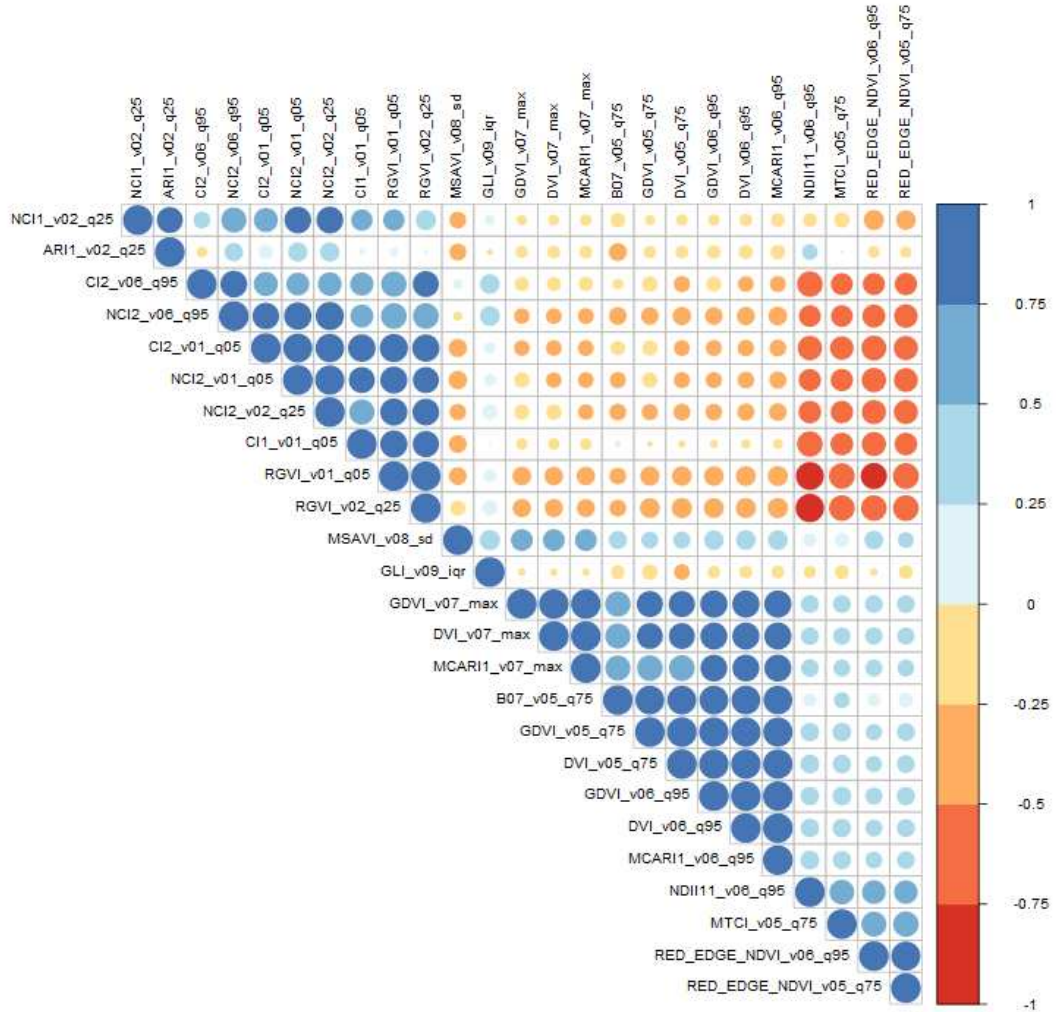
Fonte: Elaborado pela autora, 2022.

Figura 31 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cálcio, modelo *Ranger*.



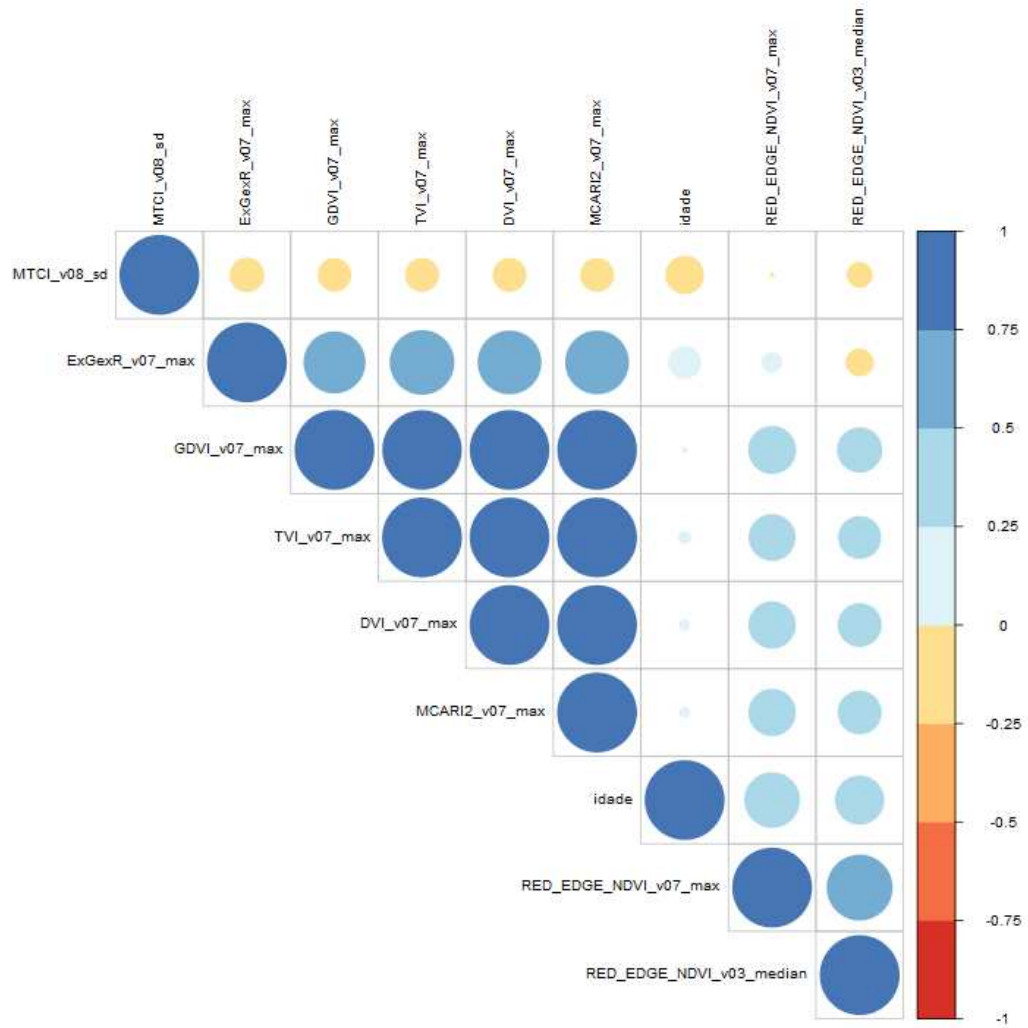
Fonte: Elaborado pela autora, 2022.

Figura 32 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento magnésio, modelo *Cubist*.



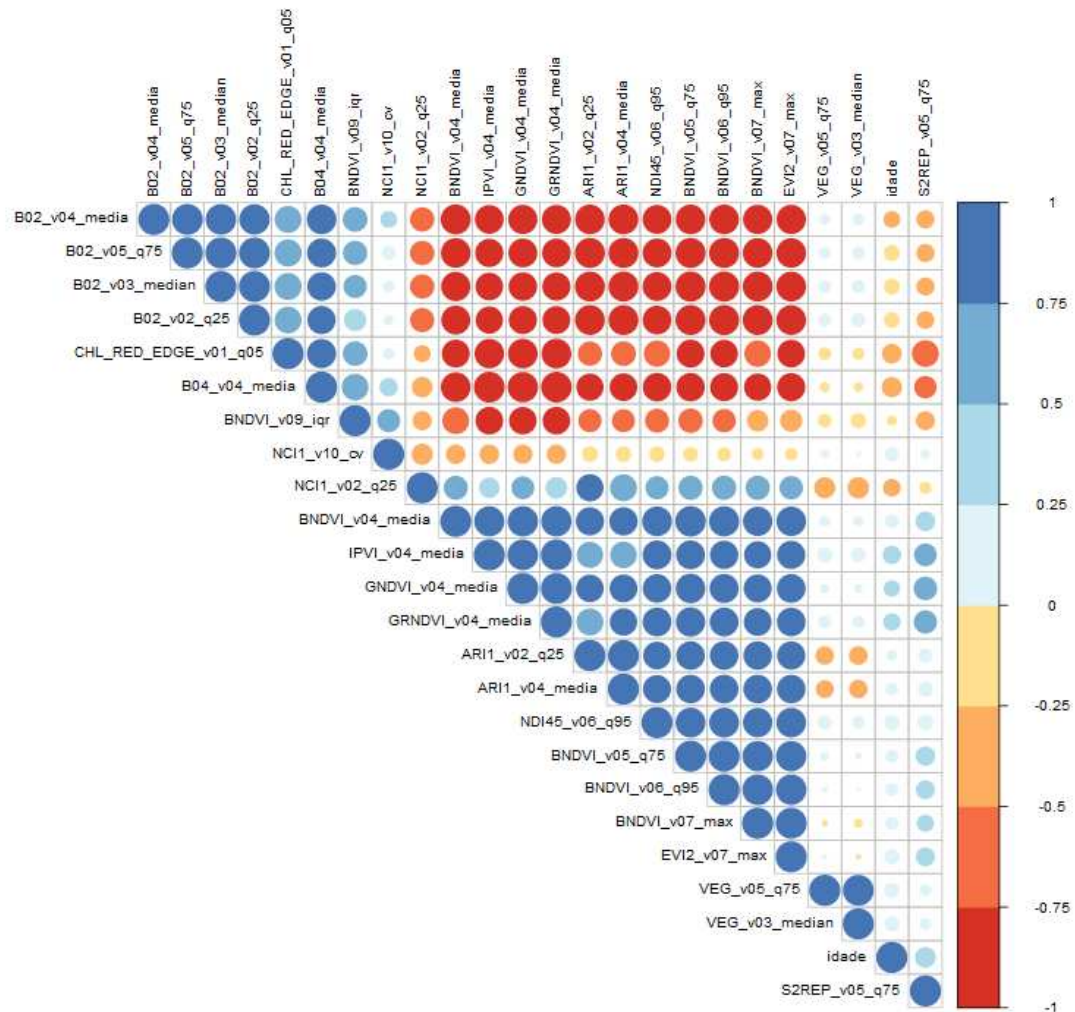
Fonte: Elaborado pela autora, 2022.

Figura 33 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento magnésio, modelo *Random Forest*.



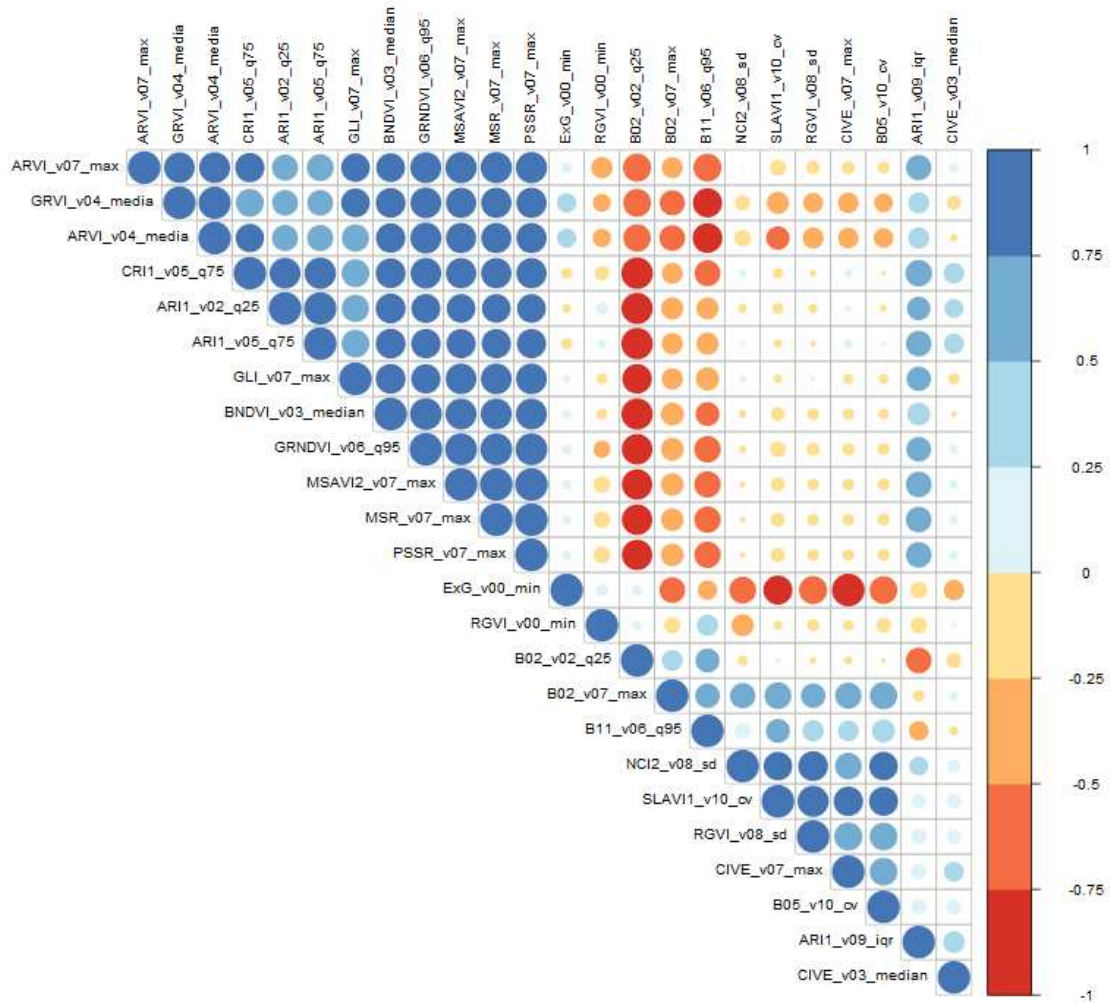
Fonte: Elaborado pela autora, 2022.

Figura 34 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Enxofre, modelo *Cubist*.



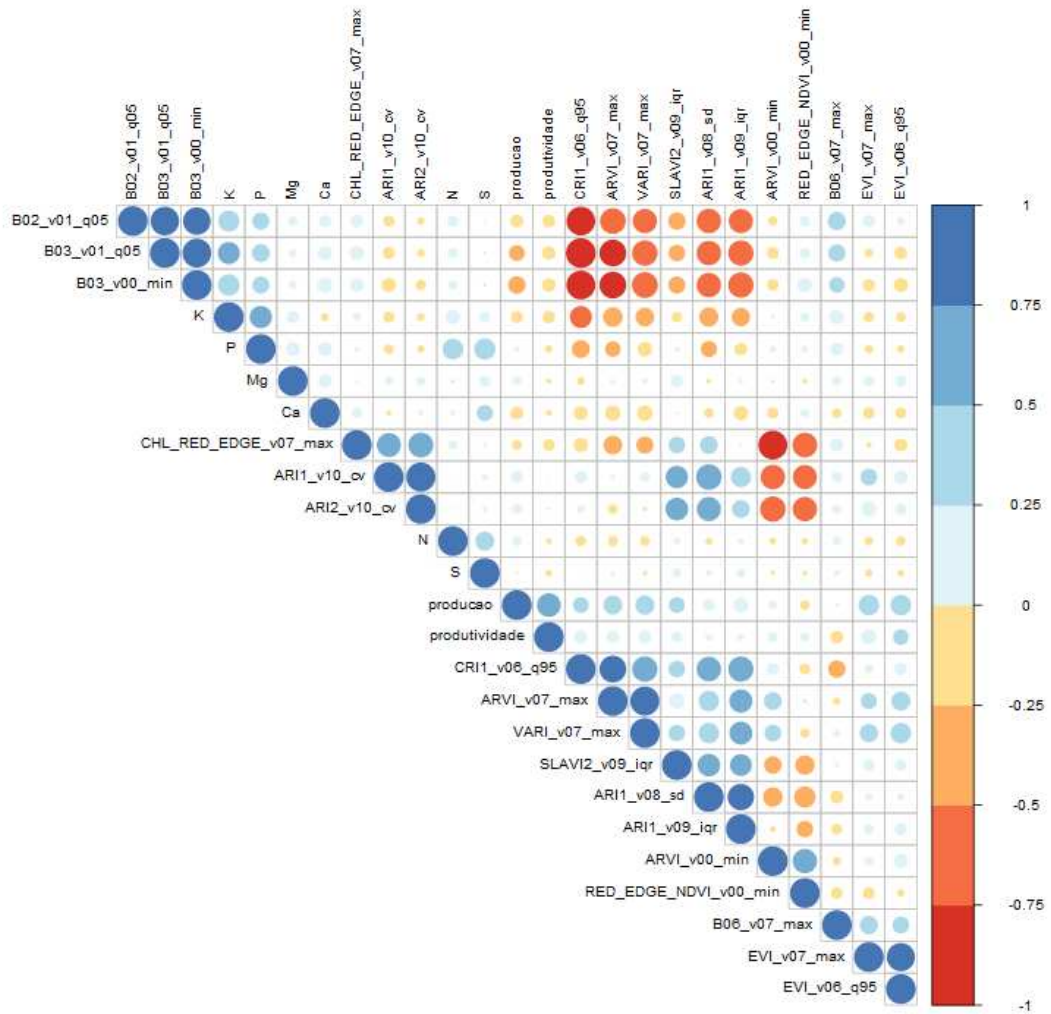
Fonte: Elaborado pela autora, 2022.

Figura 35 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Enxofre, modelo *Linear Model*.



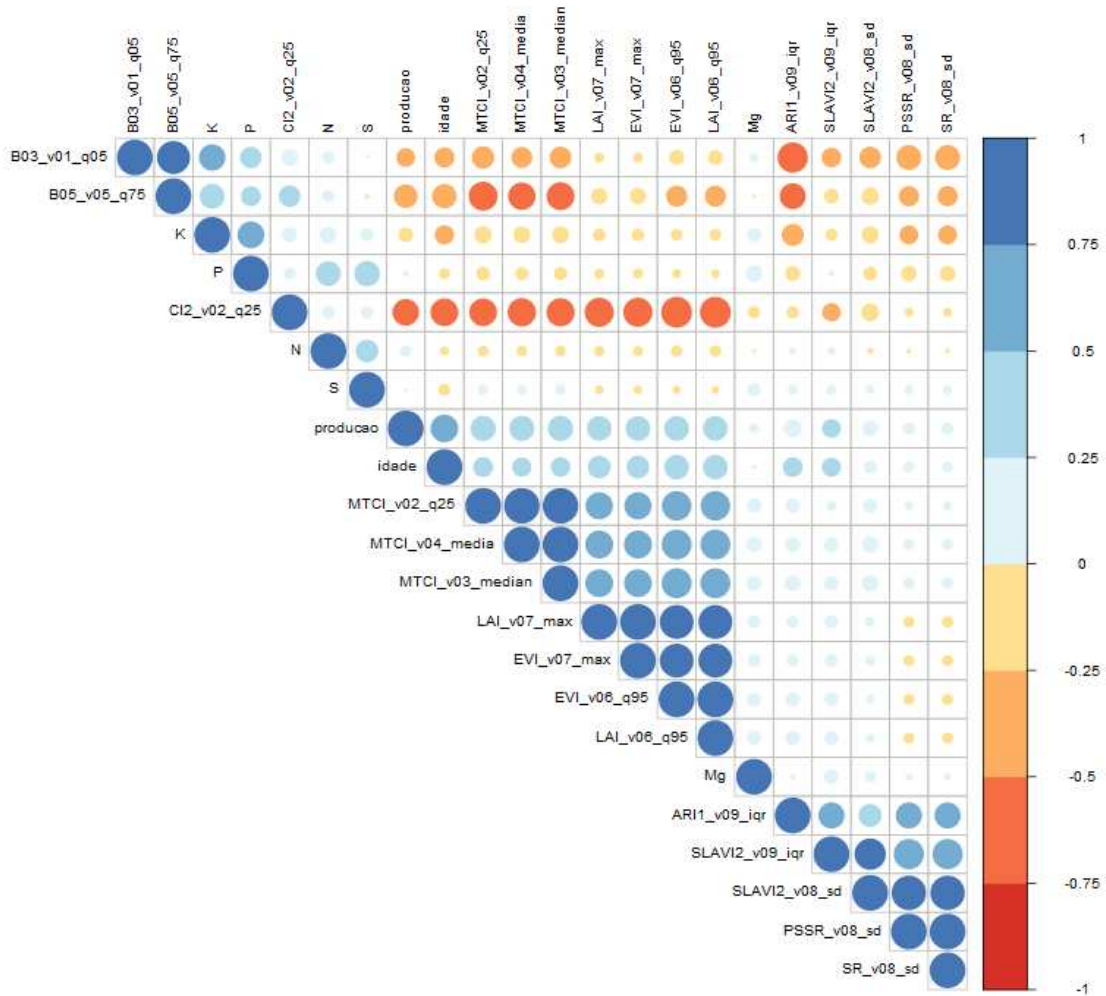
Fonte: Elaborado pela autora, 2022.

Figura 36 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cobre, modelo GBM.



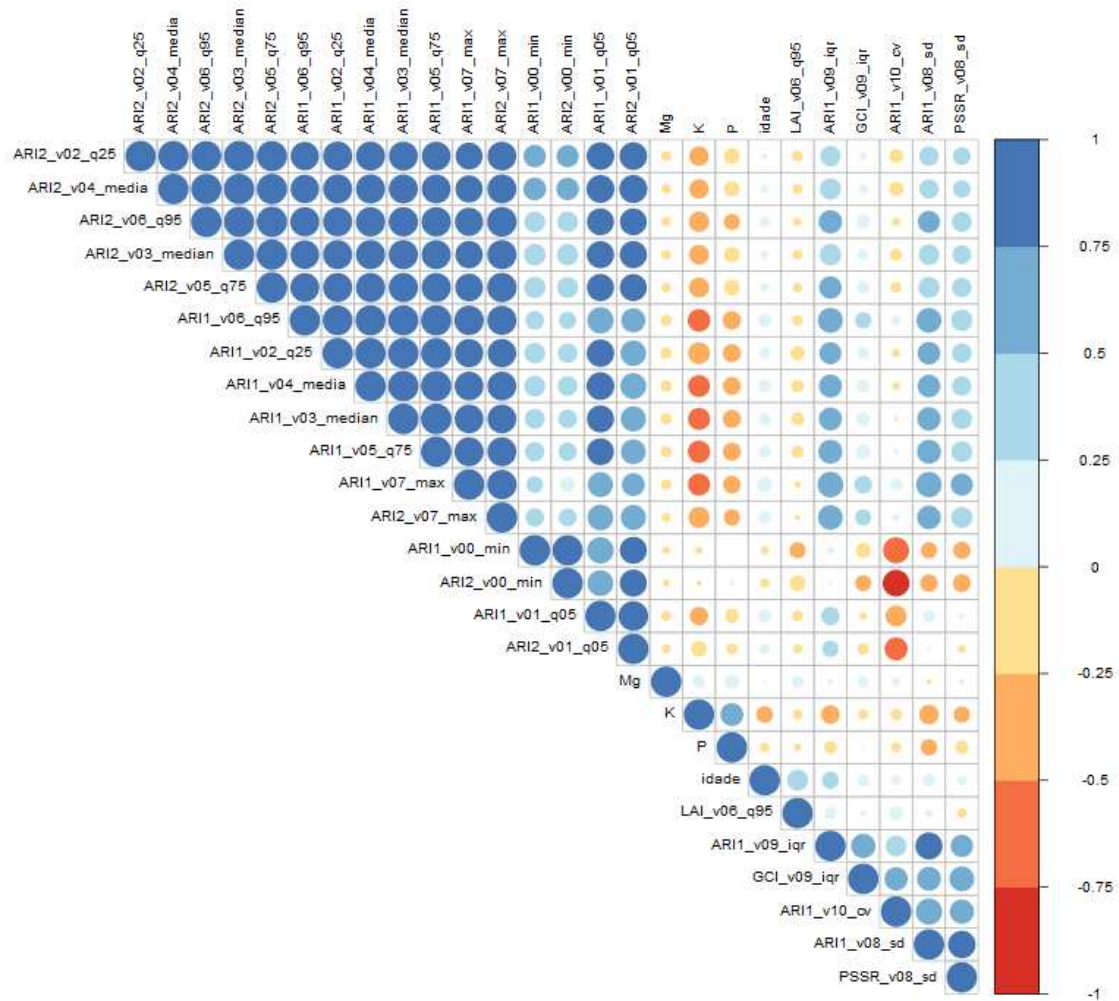
Fonte: Elaborado pela autora, 2022.

Figura 37 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cobre, modelo *Random Forest*.



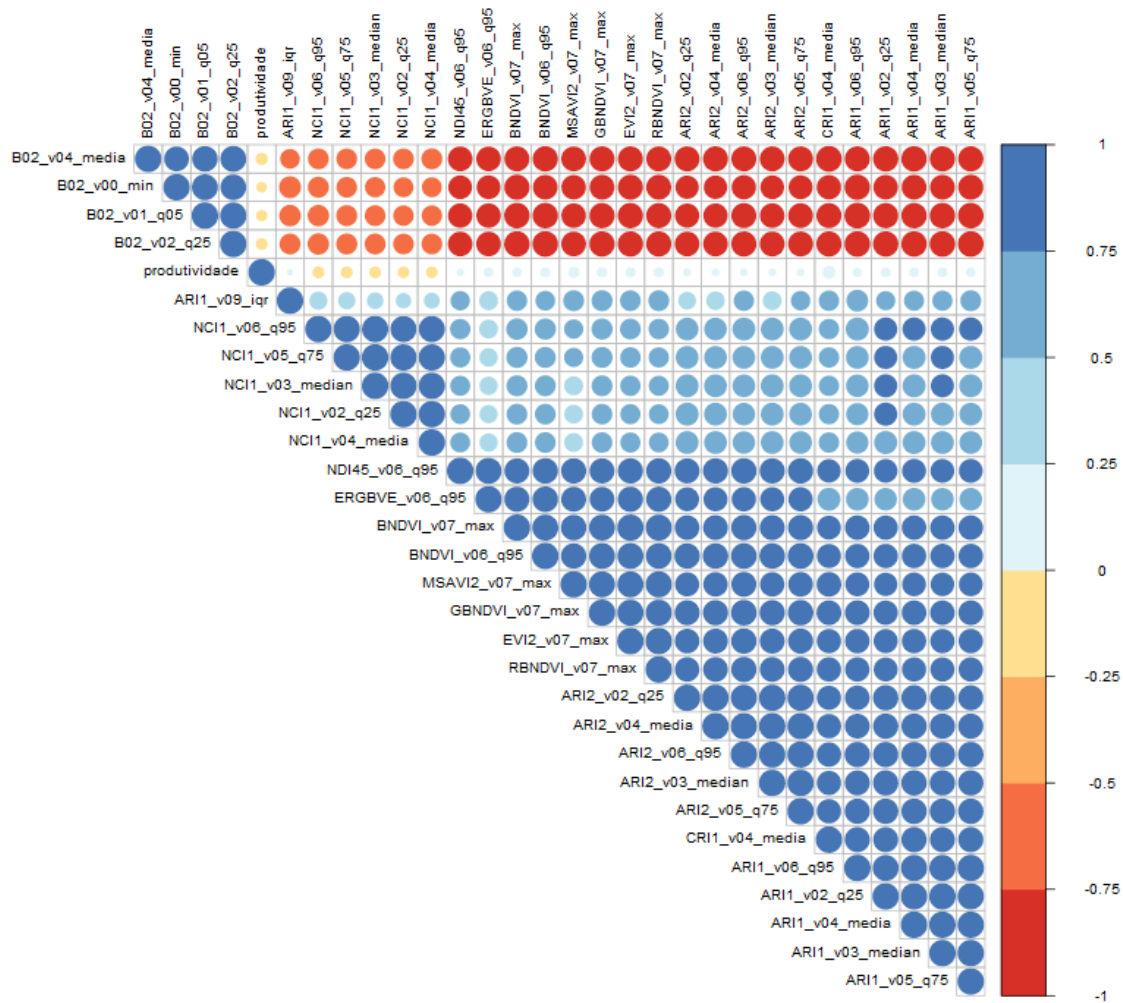
Fonte: Elaborado pela autora, 2022.

Figura 38 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento cobre, modelo *Earth*.



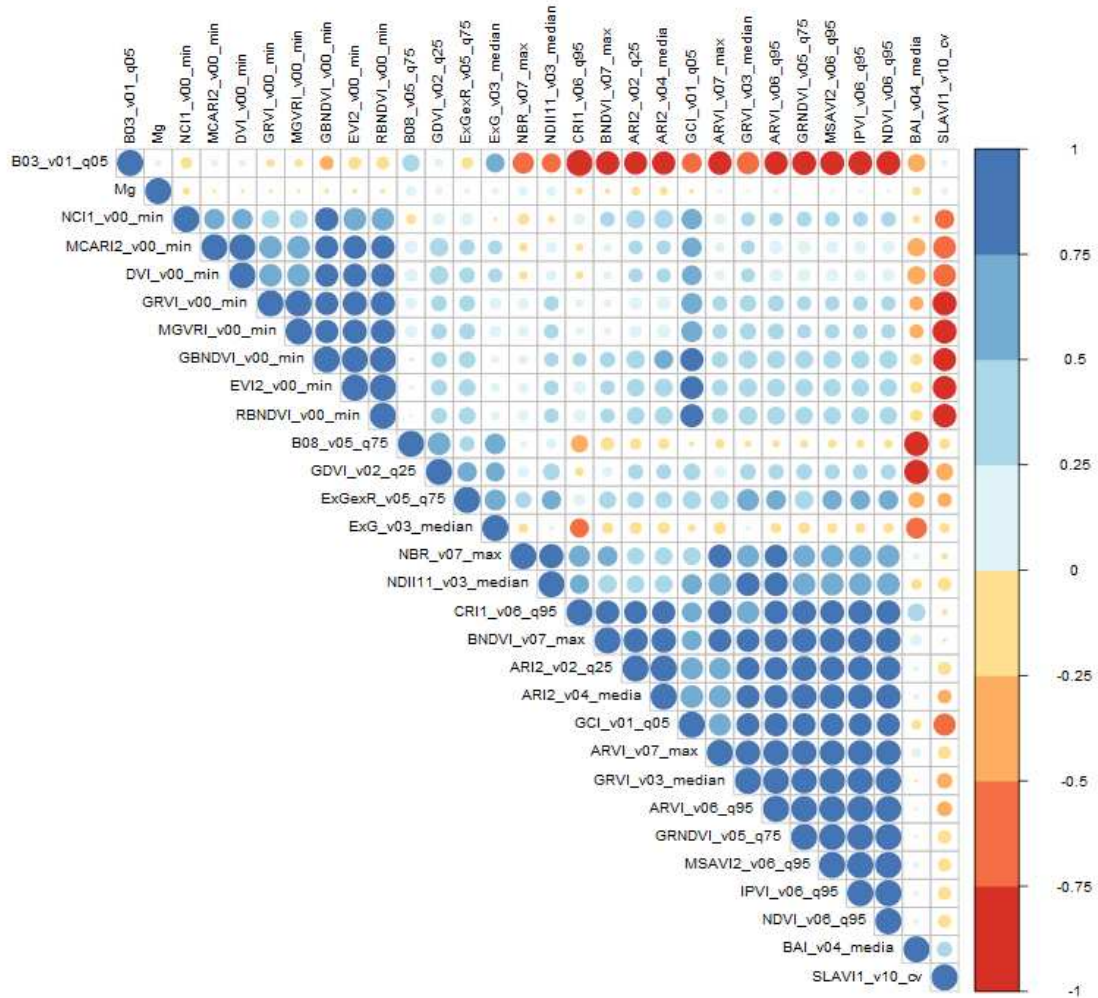
Fonte: Elaborado pela autora, 2022.

Figura 39 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento ferro, modelo *Cubist*.



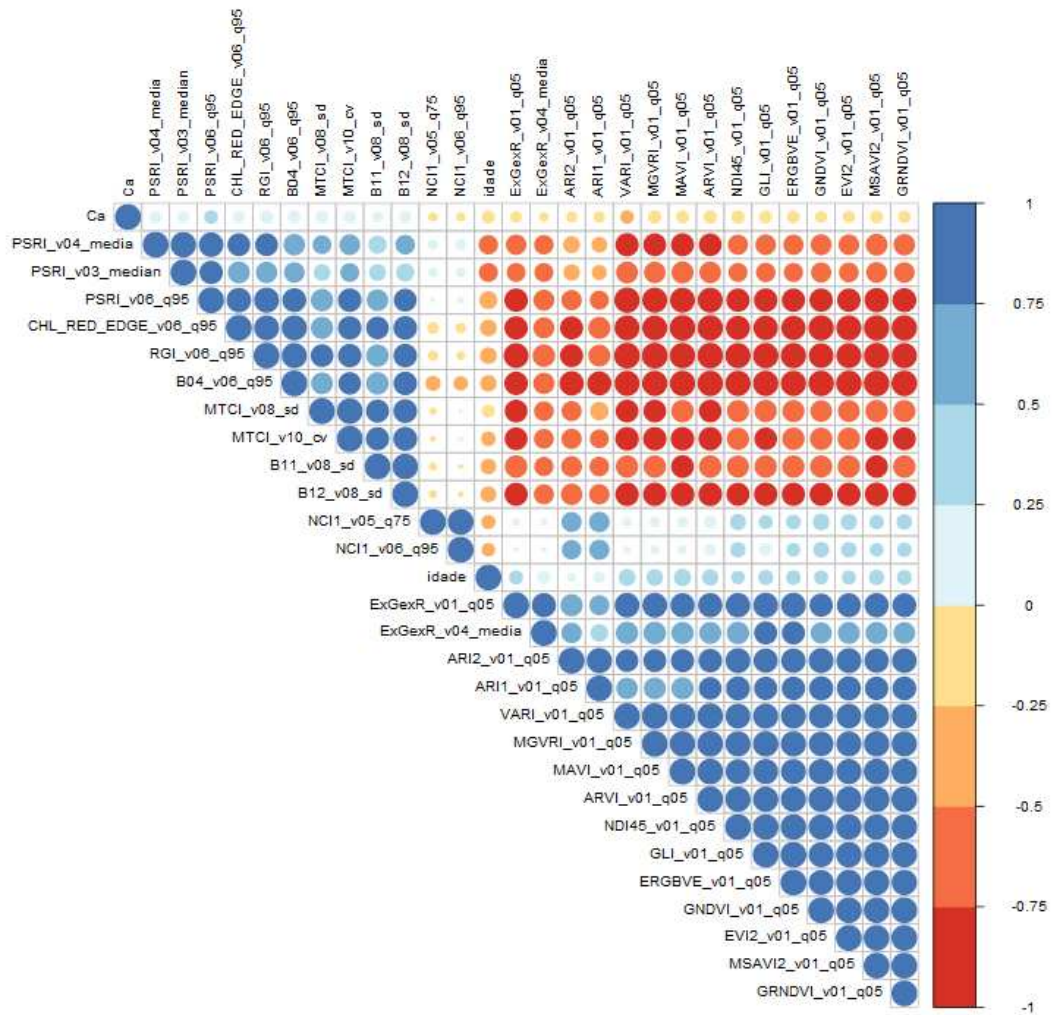
Fonte: Elaborado pela autora, 2022.

Figura 40 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento ferro, modelo *Linear Model*.



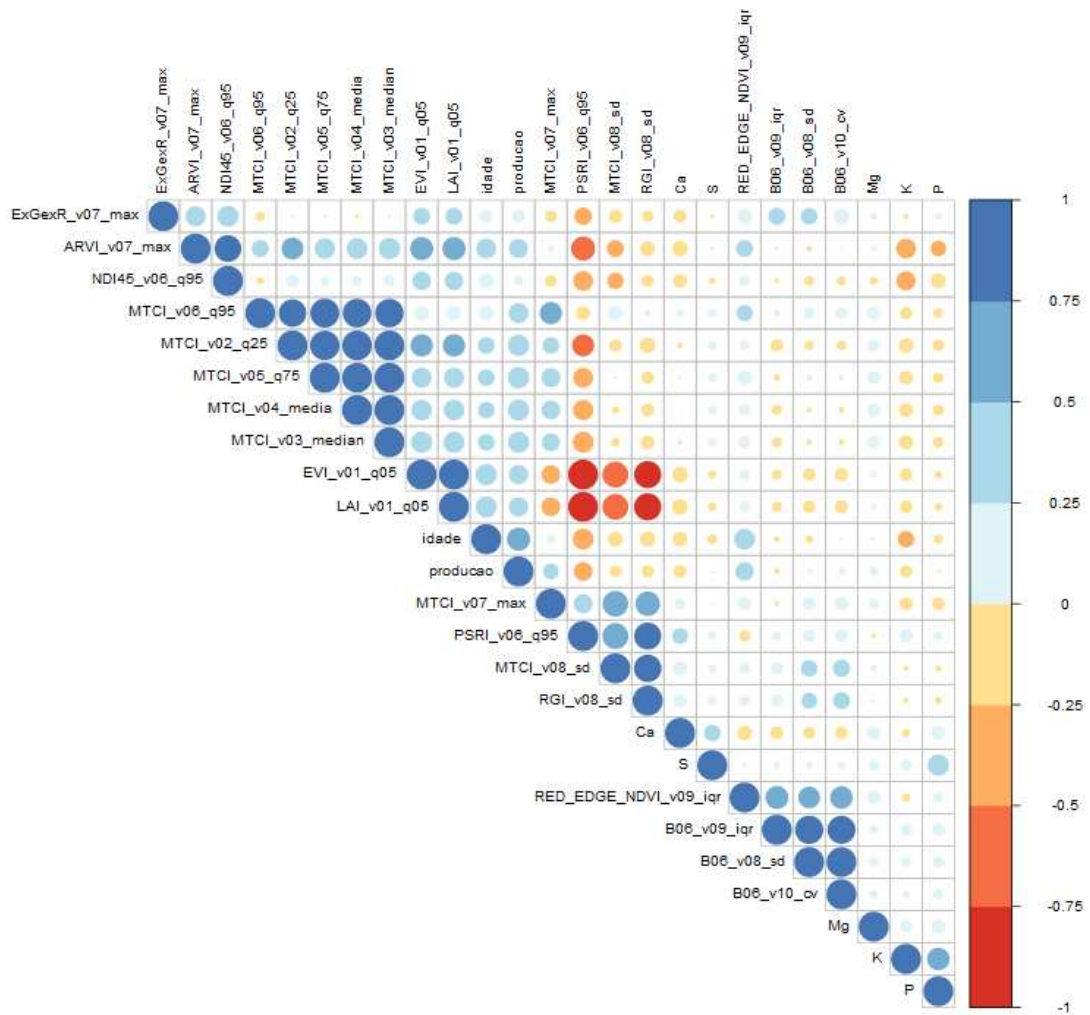
Fonte: Elaborado pela autora, 2022.

Figura 41- Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento manganês, modelo *Cubist*.



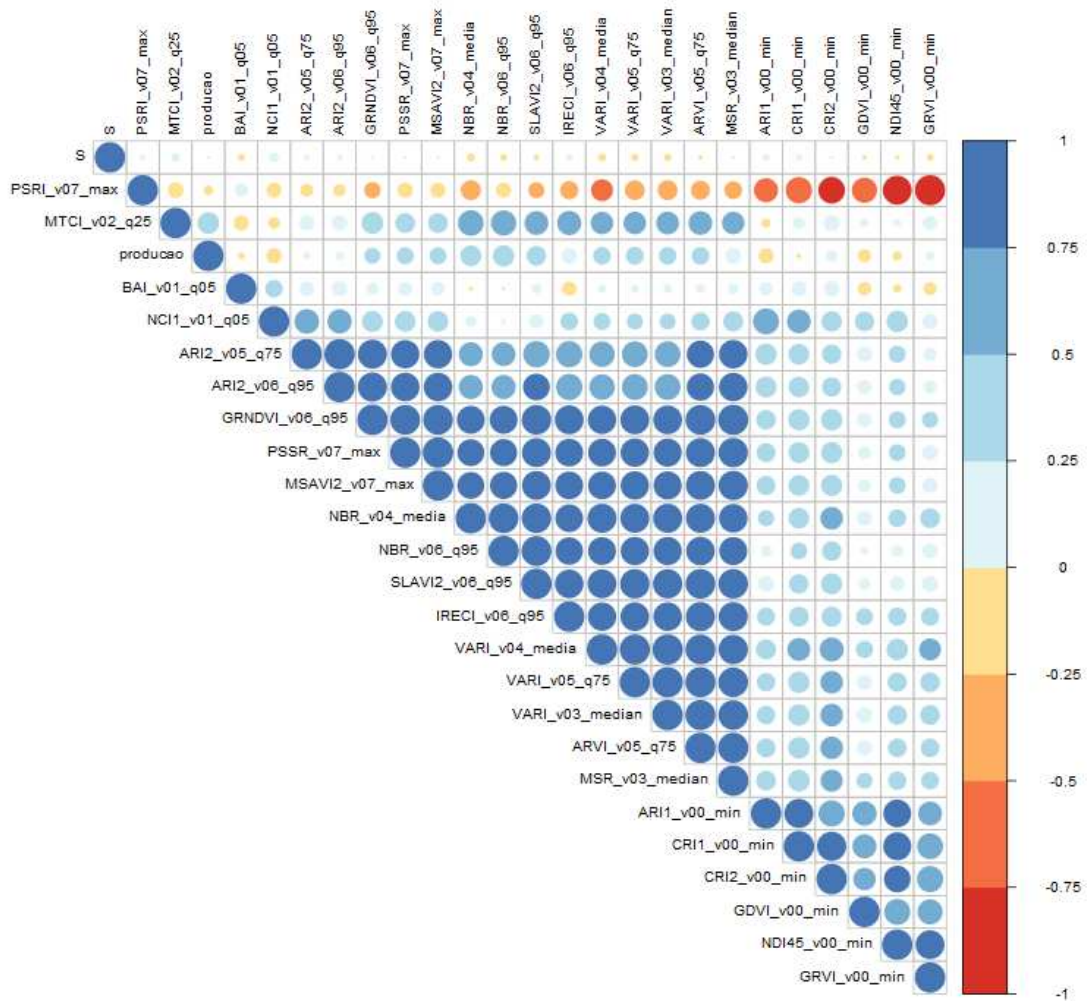
Fonte: Elaborado pela autora, 2022.

Figura 42 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento manganês, modelo *Random Forest*.



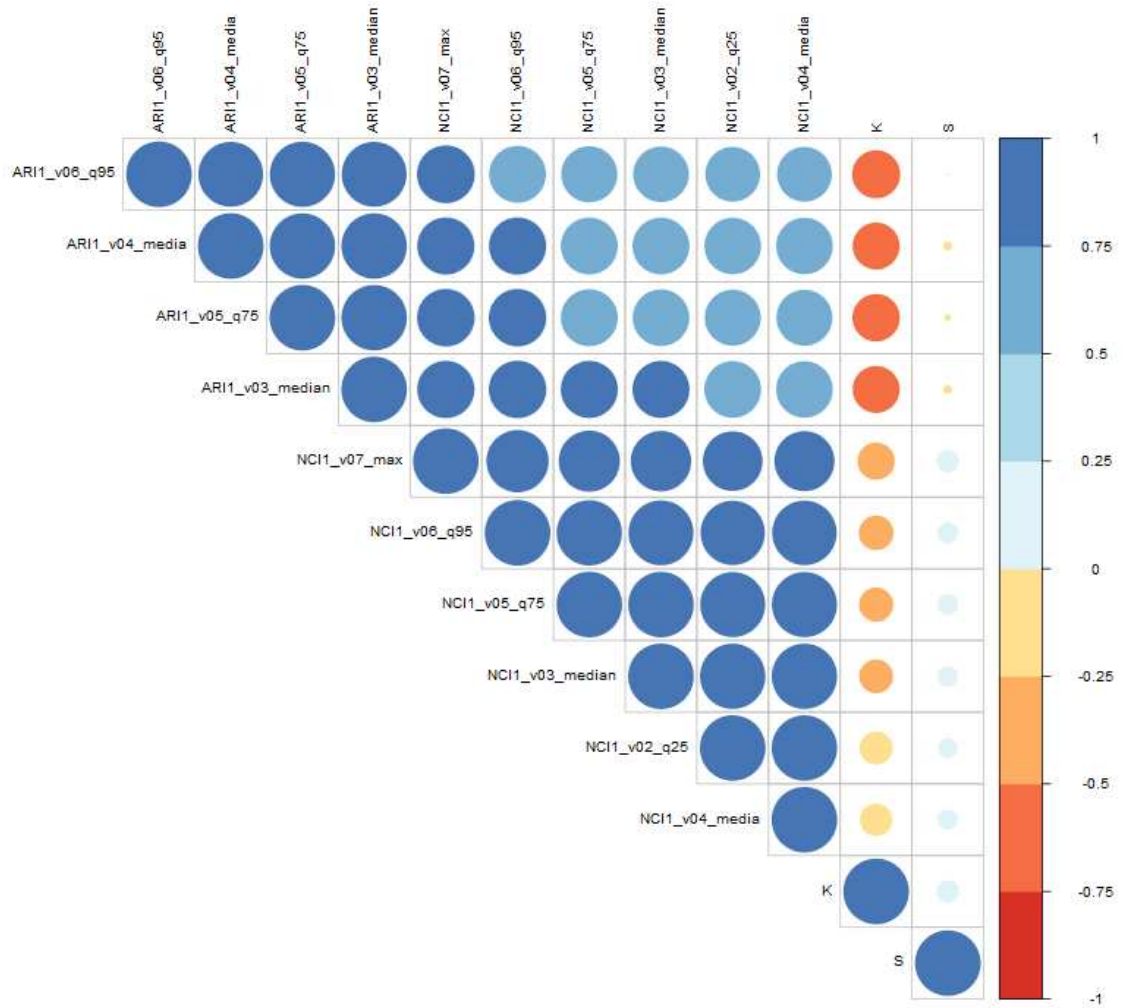
Fonte: Elaborado pela autora, 2022.

Figura 43 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento zinco, modelo *Earth*.



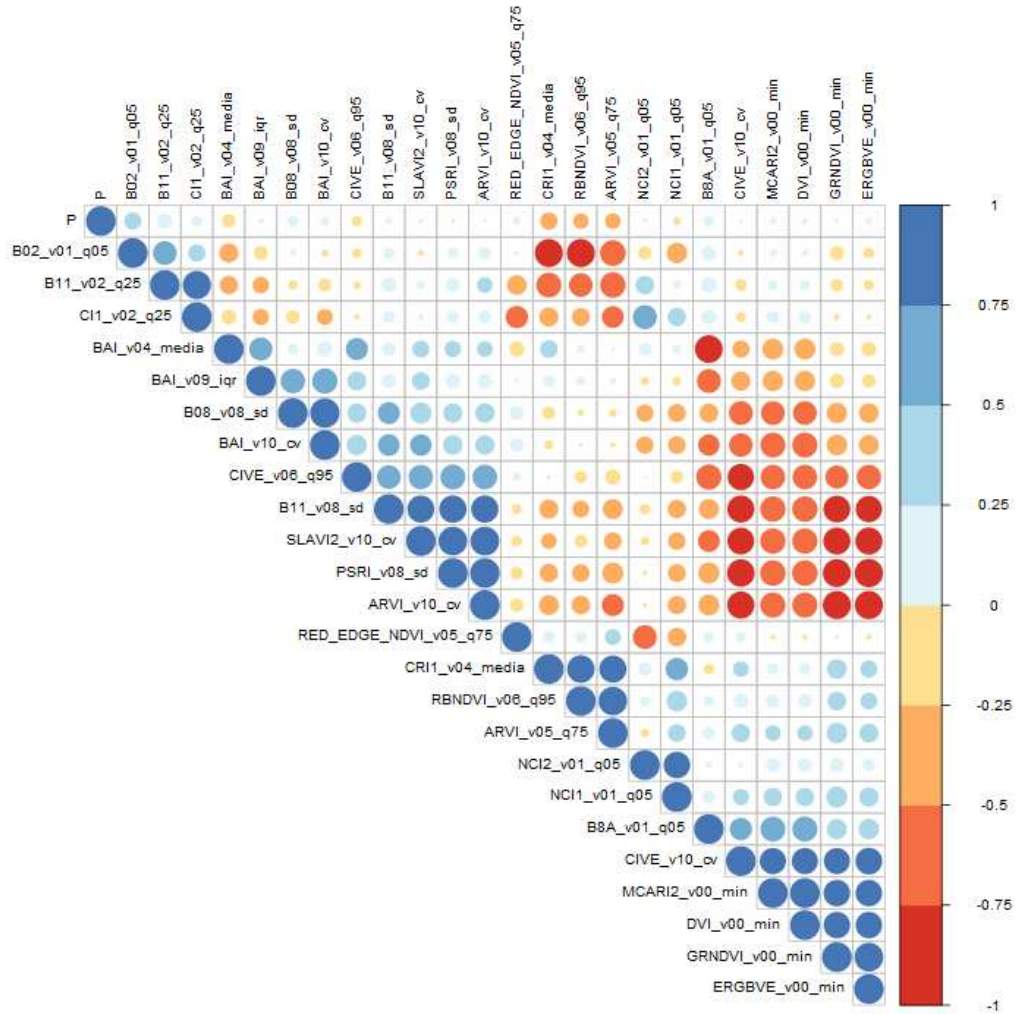
Fonte: Elaborado pela autora, 2022.

Figura 44 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Boro, modelo *Random Forest*.



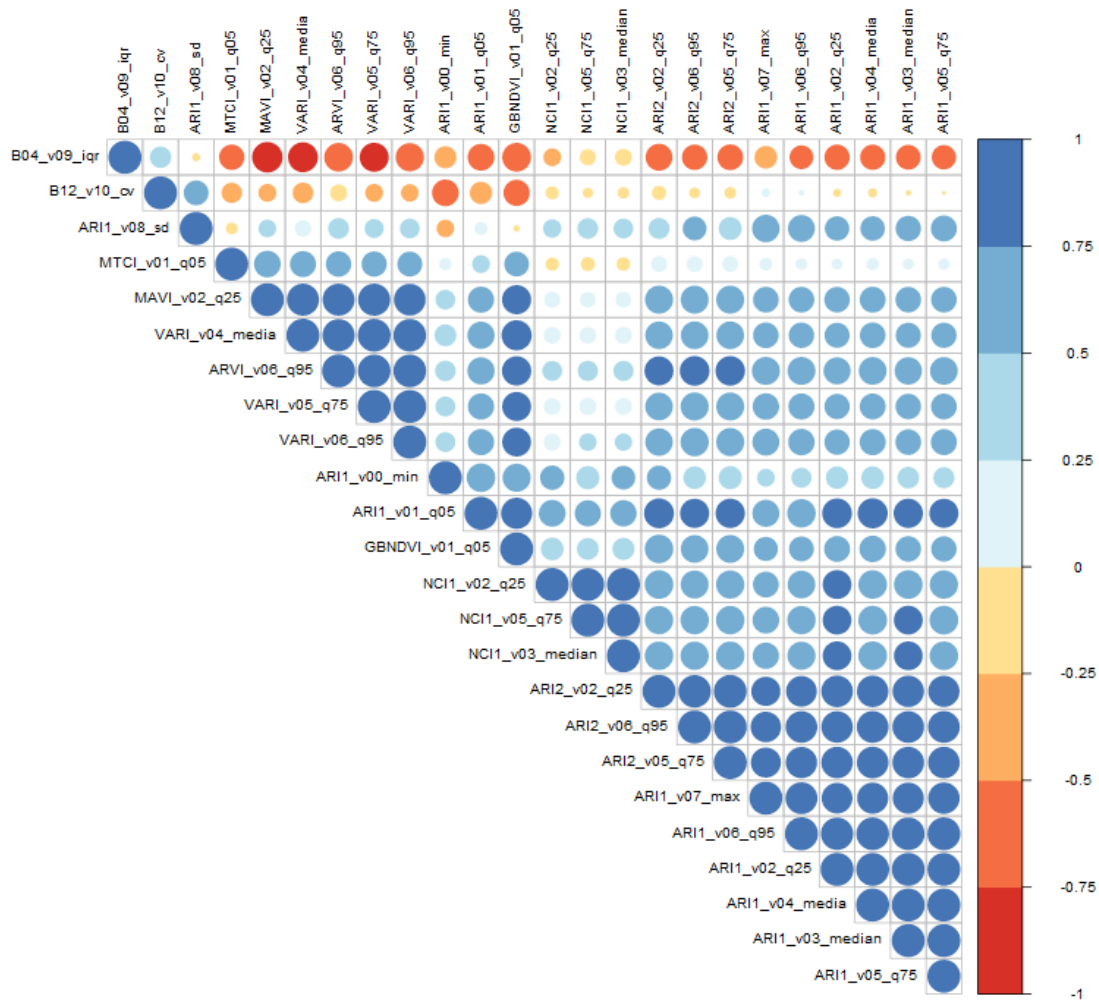
Fonte: Elaborado pela autora, 2022.

Figura 45 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Boro, modelo *Linear Model*.



Fonte: Elaborado pela autora, 2022.

Figura 46 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição do elemento Boro, modelo *Earth*.



Fonte: Elaborado pela autora, 2022

4.2 Produtividade

A Figura 37 indica que há variação na distribuição dos dados entre os modelos avaliados. Essa variabilidade pode estar relacionada às diferentes idades dos plantios, estágios fenológicos e materiais genéticos que influenciam nos picos dos valores de produtividades. Os algoritmos *Earth*, *Ranger* e *Random Forest* foram os que apresentaram melhor distribuição interquartil e entre eles o *Ranger* destacou-se com melhor valor de R^2 , com aproximadamente 45% de acerto.

A Tabela 14 apresenta os IVs que mais influenciaram nos resultados dos algoritmos, nota-se uma variedade de índices, com exceção do modelo *Earth* que possui ARI e ARVI como os mais relevantes, representando um grande de acerto de 33%. Percebe-se que grande parte dos IVs são derivados do NDVI, ou seja, conseguem captar mudanças de refletância seja através das alterações na concentração de clorofila ou expansão da área foliar.

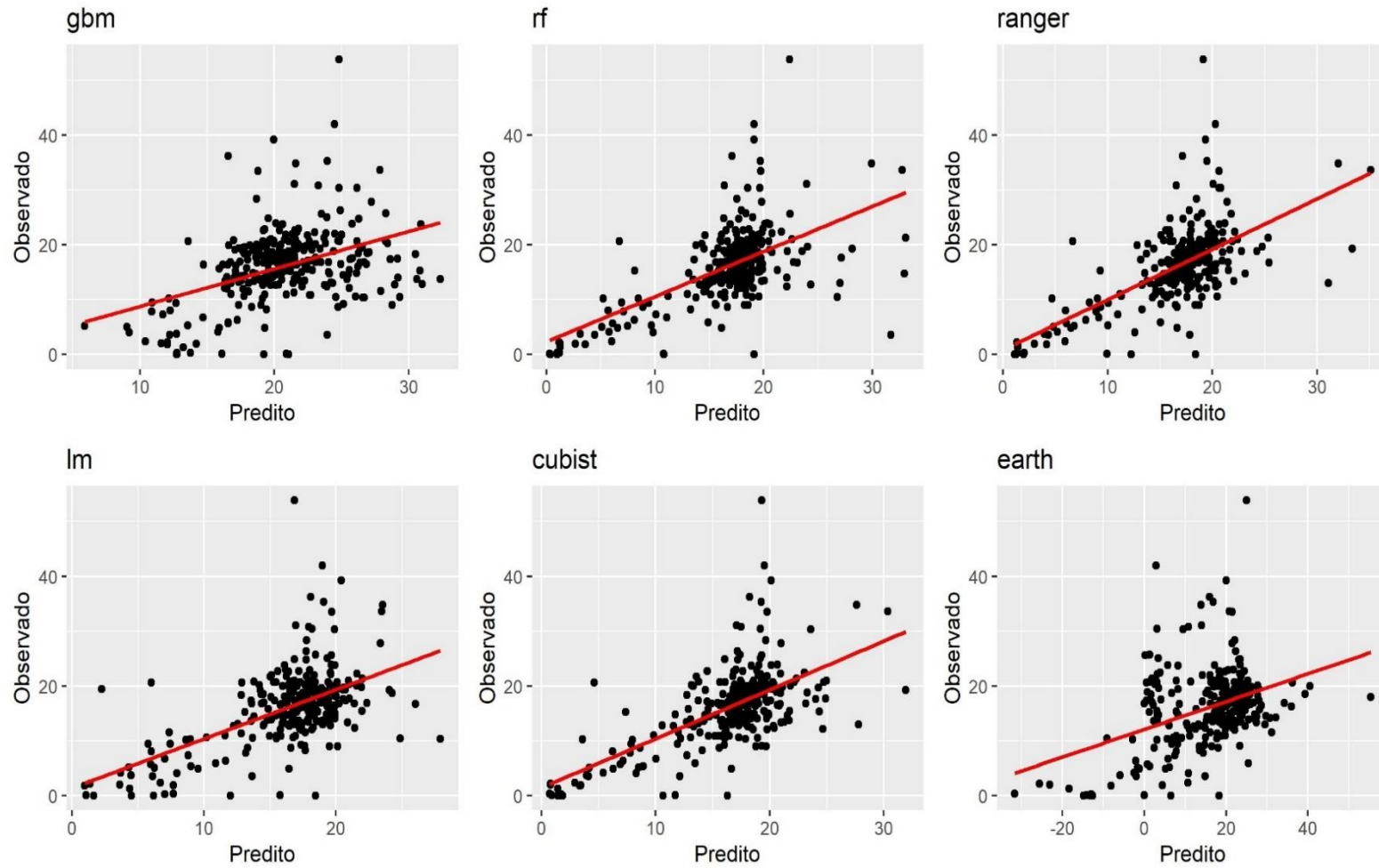
O RMSE calculado para a produtividade de treinamento variaram de 4,70 a 5,88 e na validação de 4,62 a 5,83. O RMSE indica a falta de precisão de uma medida, e quanto menor o seu valor, melhor é a capacidade do algoritmo em apresentar os valores das predições mais próximas do valor real. Sendo assim, o Random Forest se destaca como o melhor modelo para predição de produtividade, pois apresenta valor de R^2 de 41% e o menor RMSE de 4,62. Estudos realizados com palma de óleo por Soares et al., (2020) obtiveram valor de RMSE 20,89% para predição de produtividade.

Tabela 13 - Resultado da análise de importância dos índices de vegetação que mais influenciaram nos algoritmos para a predição de produtividade em palma de óleo e seus respectivos valores de R^2 e RMSE de treinamento e validação.

Produtividade	Cubist	Earth	GBM	LM	Ranger	Random Forest
1	CI2	ARI	CI2	BNDVI	CI2	CI2
2	RED_EDGE _NDVI	ARVI	CI1	BAI	CI1	RED_EDGE _NDVI
3	RGVI		EVI	GLI	RGVI	LAI
4	PSRI		RED_EDGE _NDVI	GRNDVI	MTCI	MTCI
5			MTCI		NCI2	EVI
6						GLI
7						PSRI
R^2 Trein.	0,40	0,32	0,33	0,26	0,41	0,41
RMSE Trein.	5,23	5,95	5,40	5,30	5,48	4,70
R^2 Valid.	0,39	0,33	0,33	0,25	0,43	0,40
RMSE Valid.	5,33	5,56	4,89	5,83	4,85	4,62

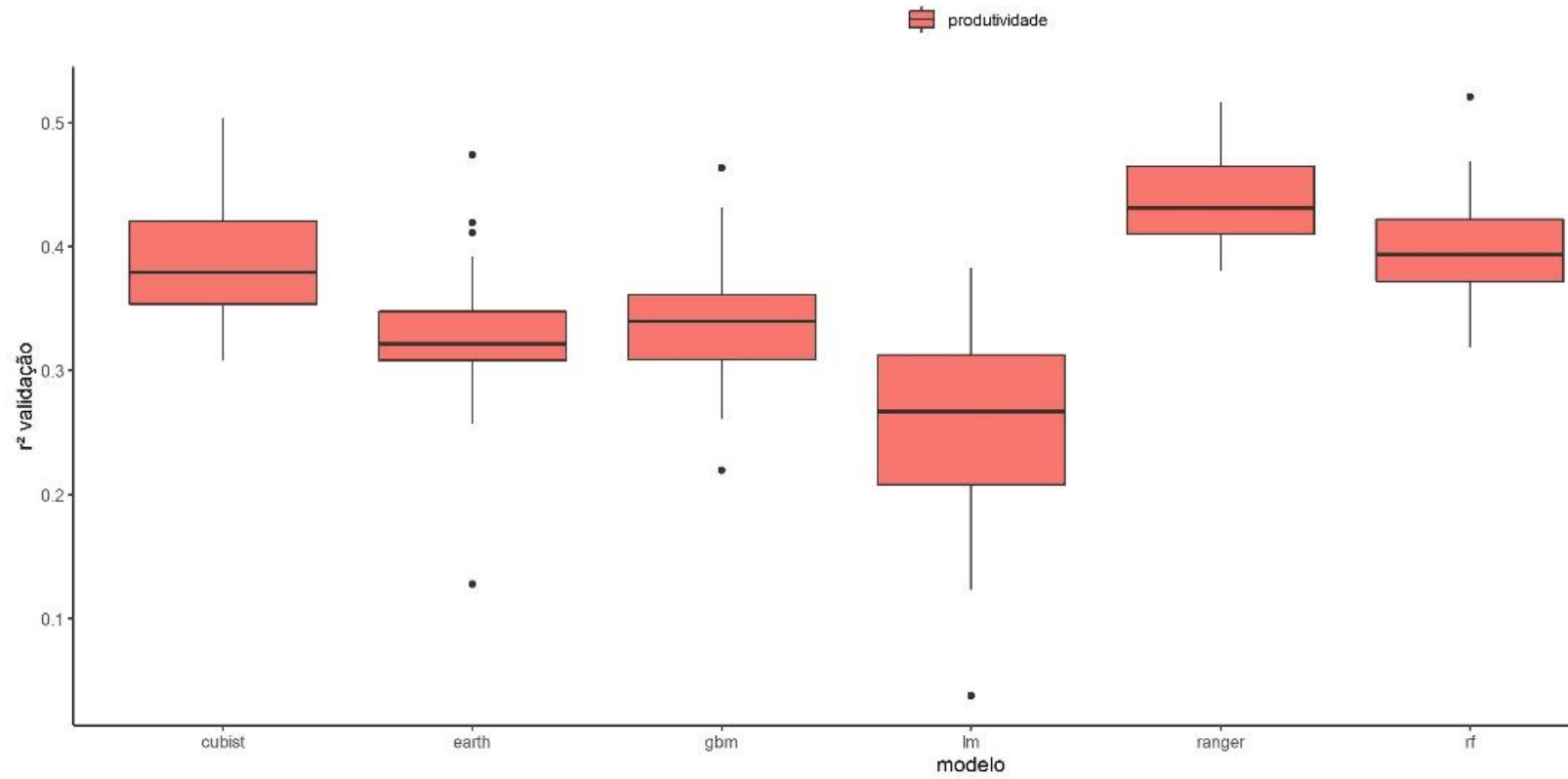
Fonte: Elaborado pela autora, 2022.

Figura 47 - Estimativa de produtividade nas idades de 15 a 25 anos para a cultura da palma de óleo, de acordo com os algoritmos *GradientBoostingMachine* (GBM), *Random Forest* (RF), *Ranger*, Regressão Linear (lm), *Cubist* e *Earth*.



Fonte: Elaborado pela autora, 2022.

Figura 48 - Boxplot dos valores de R^2 validação dos algoritmos utilizados na predição de produtividade de cachos de frutos frescos (CFF) na cultura palma de óleo.



Fonte: Elaborado pela autora, 2022.

É importante destacar que no banco de dados há informações de plantios com diferentes idades, portanto, em fases de desenvolvimento distintas, ou seja, o maior recobrimento da área pode influenciar na sensibilidade de alguns índices em detectar as variações de refletância no tecido foliar, como no caso do NDVI, e este fato está associado com a saturação da cor verde a partir do acúmulo de biomassa.

Estudos realizados por Gitelson (2004); Povh *et al.* (2008), Zanzarini *et al.* (2013) reportam que o NDVI se torna insensível as variações espectrais da cultura quando a refletância no comprimento de onda do infravermelho próximo (faixa do invisível) é maior que a absorção no comprimento de onda no espectro do vermelho (faixa do visível). Estes resultados aliados ao que foram observados neste estudo indicam que ao incluir no banco de dados os diferentes modelos, há maior potencial de selecionar o índice que mais reflete a realidade da lavoura.

Os algoritmos *Cubist*, *Ranger* e *Random Forest* foram os que apresentaram os melhores resultados de validação, com 39%, 43% e 40% (Tabela 1), respectivamente, e há similaridade entre os índices contidos em cada um deles, como é o caso do CI (*Canopy Index*) utilizado para caracterizar a distribuição espacial das folhas de um dossel, bem como a fotossíntese e processos hidrológicos. Os demais estão diretamente associados com fatores de crescimento e são derivações do NDVI, como é o caso do Red_Edge_NDVI, é importante ressaltar as variações no teor de nitrogênio foliar que ocorre nos plantios, portanto, esse resultado pode estar relacionado à forte associação do comprimento de onda com o teor de clorofila das plantas e conseqüentemente ao status deste nutriente na folha.

Outros promissores índices para predição da produtividade são, o RGVI e LAI, capazes de monitorar o crescimento e índice de área foliar. A correlação positiva entre eles indica que os modelos se tornam mais sensíveis a detecção de mudanças da área foliar, e, portanto, indicados para monitoramento das lavouras. Carvalho (2019) e Ferreira (2020) utilizaram diversos IVs para monitorar a produtividade do milho, o primeiro autor e obteve o NDVI, GRNDVI e EVI como os mais satisfatórios, com grau de acerto de 44%. Enquanto o SAVI foi o índice mais eficiente para predição de produtividade de biomassa e grãos de milho, com 68% de probabilidade de acerto (Ferreira, 2020).

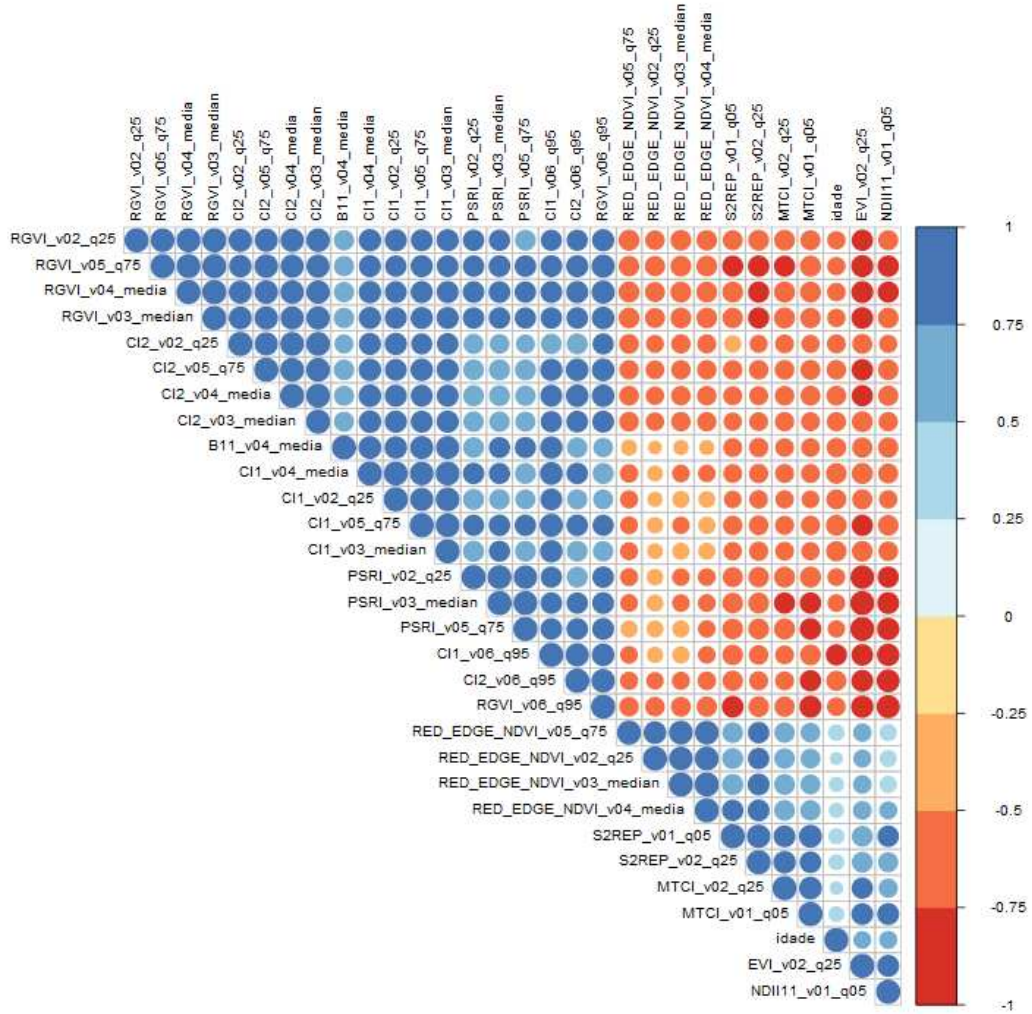
No geral, os gráficos da correlação de Spearman mostram que os algoritmos *Cubist* (Figura 38), *Ranger* (Figura 42) e *Random Forest* (Figura 43) tiveram comportamentos muito similares, apresentando maior interação entre os índices RGVI e CI e pouco ou nenhuma correlação com os outros índices avaliados. No modelo LM (Figura 41), as correlações entre os

índices derivados no NDVI foram as mais expressivas, como é o caso do BNDVI e GRNDVI. É importante ressaltar que as diferentes fases fenológicas da cultura, idade, estresses hídricos, nutricionais, material genético, entre os demais fatores que interferem na produtividade podem influenciar nos ajustes dos modelos preditivos.

Pesquisa realizada por Soares *et al.* (2020) em uma fazenda desta mesma área de estudo, com plantios nas idades de 3 a 6 anos, mostram que os índices NDVI, EVI, ARVI e RENDVI foram os que possibilitaram correlações positivas entre valores médios dos índices e as respectivas produtividades, sendo o EVI e NDVI os que melhor se correlacionaram, como valores de R^2 de 0,66 e 0,59, respectivamente.

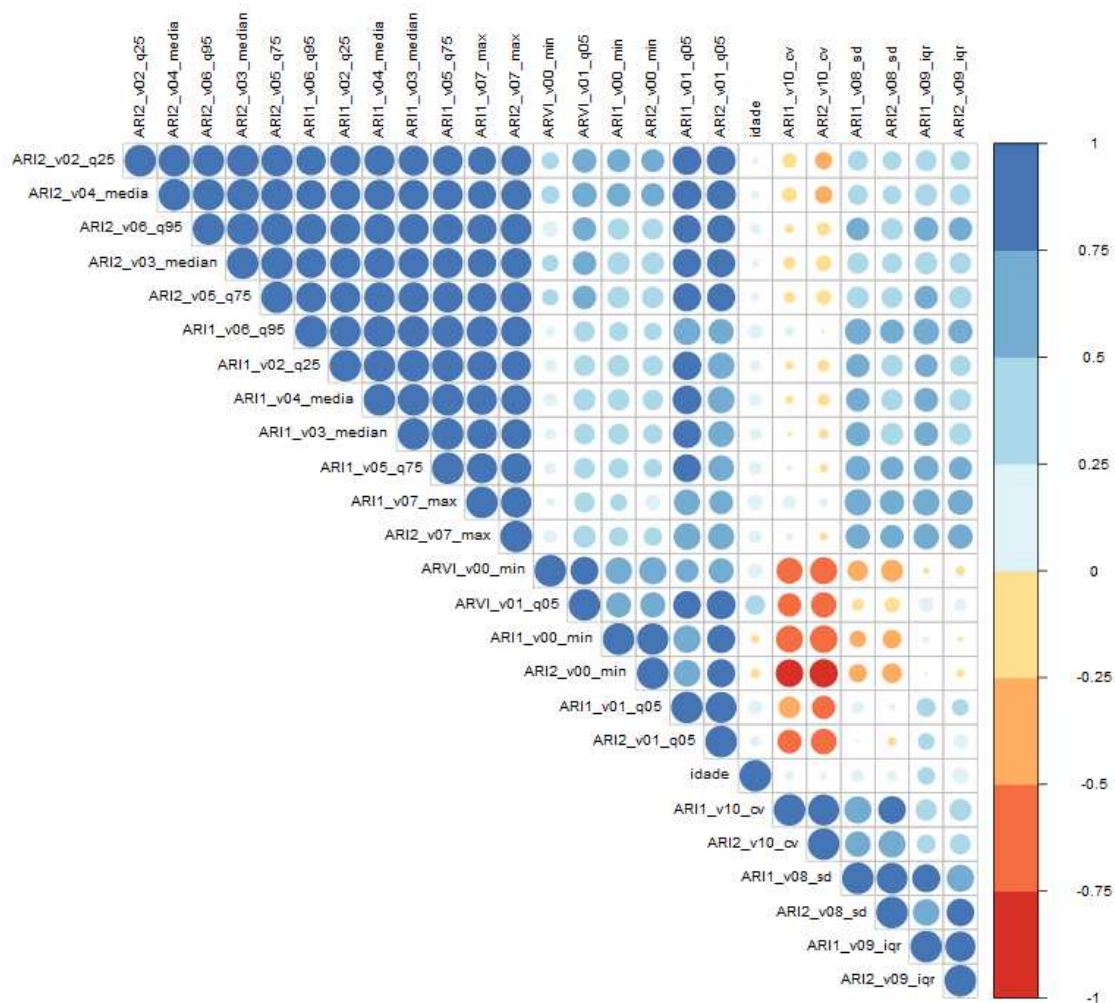
É importante destacar que ainda são poucos os estudos relacionados com modelagem da produtividade utilizando imagens espectrais para a cultura da palma de óleo. Sendo assim, os resultados aqui apresentados se mostram muito promissores e a partir da segregação desses dados por classes de idades, características semelhantes de clones, períodos de produção etc. pode-se ter modelos cada vez mais responsivos para determinada fase de desenvolvimento da cultura.

Figura 49 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo *Cubist*.



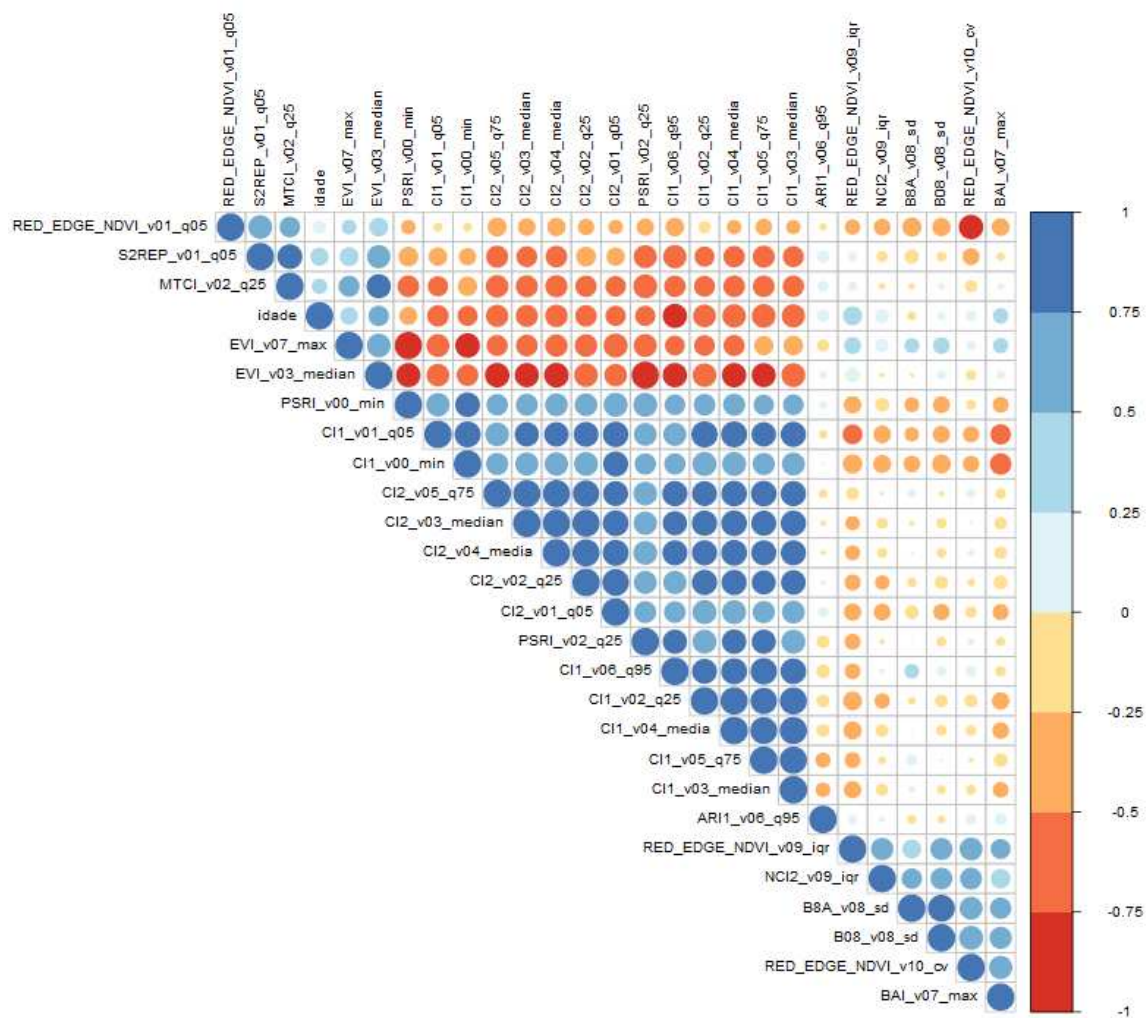
Fonte: Elaborado pela autora, 2022.

Figura 50 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo Earth.



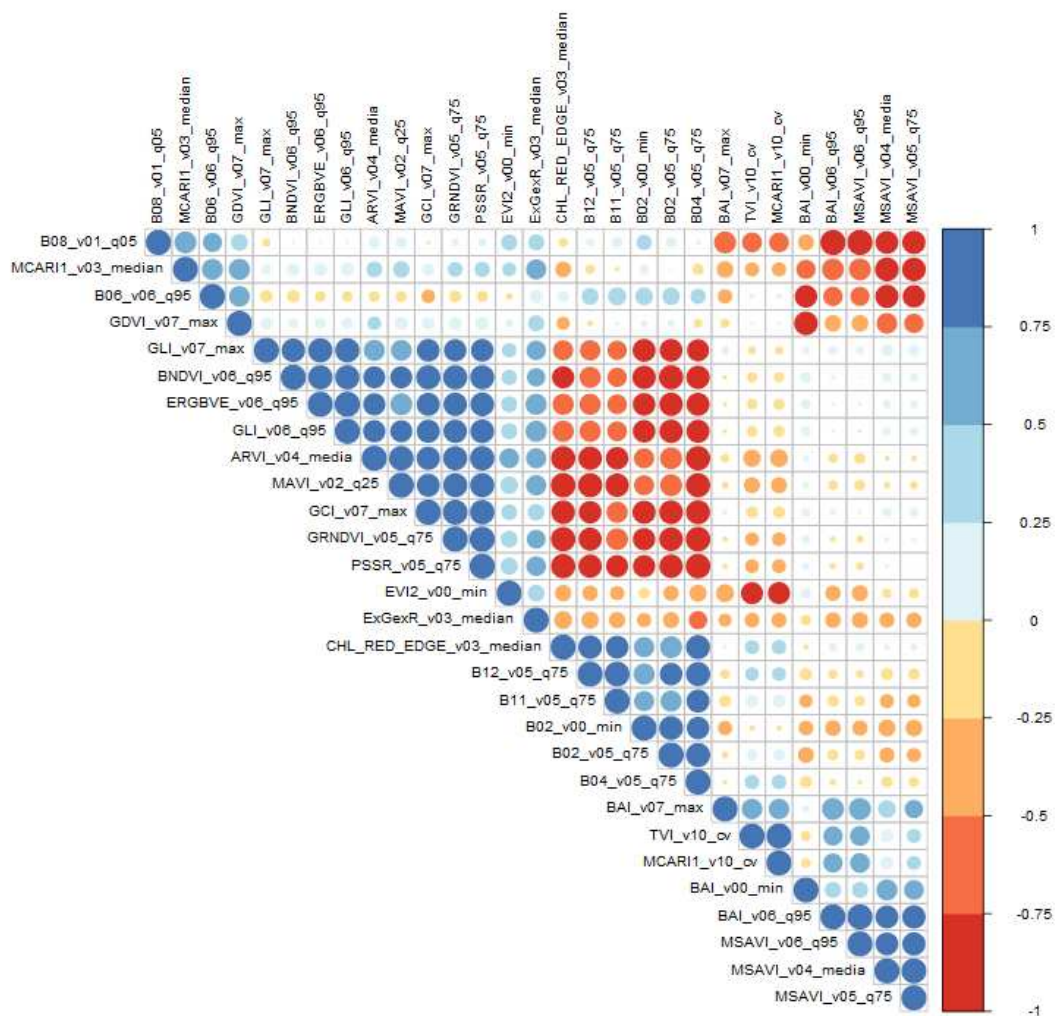
Fonte: Elaborado pela autora, 2022.

Figura 51 - - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo GBM.



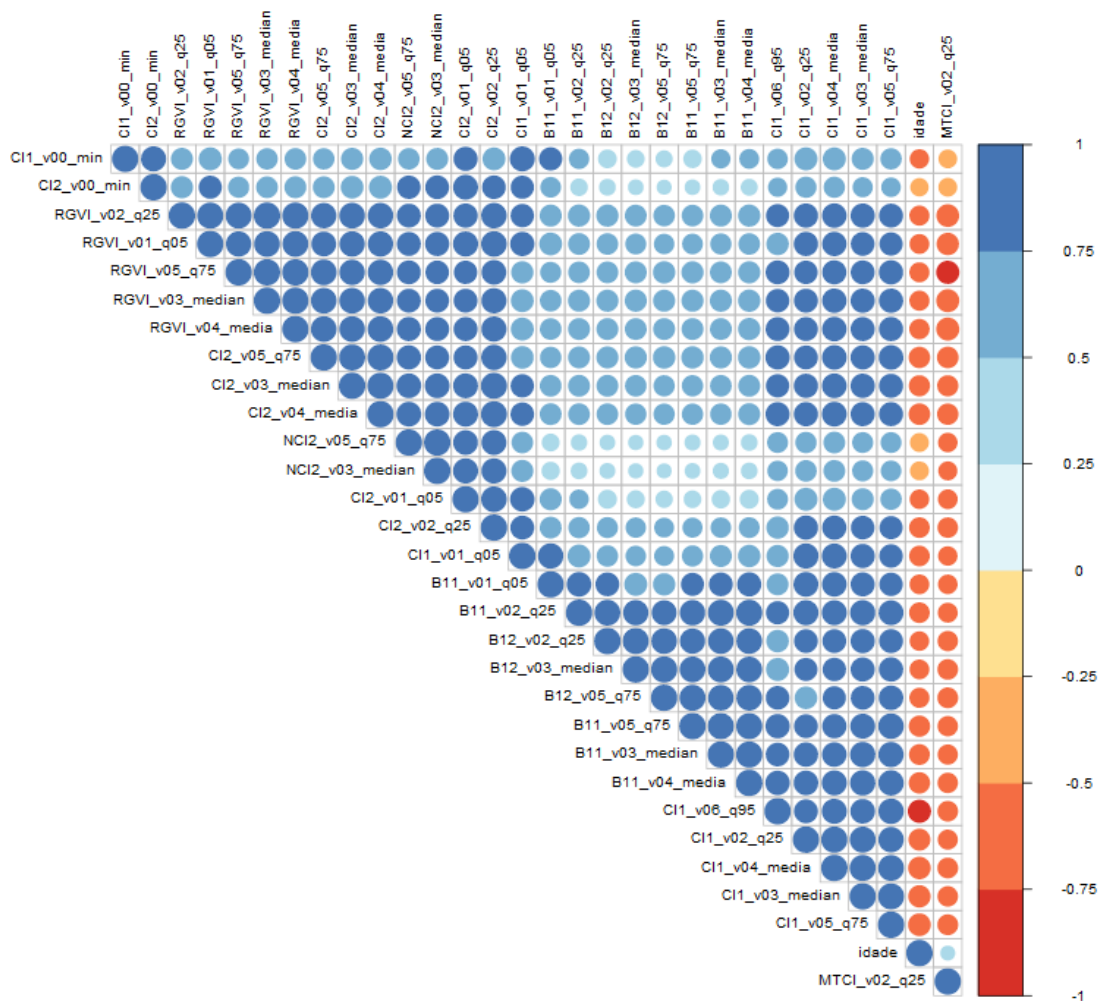
Fonte: Elaborado pela autora, 2022.

Figura 52 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo *Linear Model*.



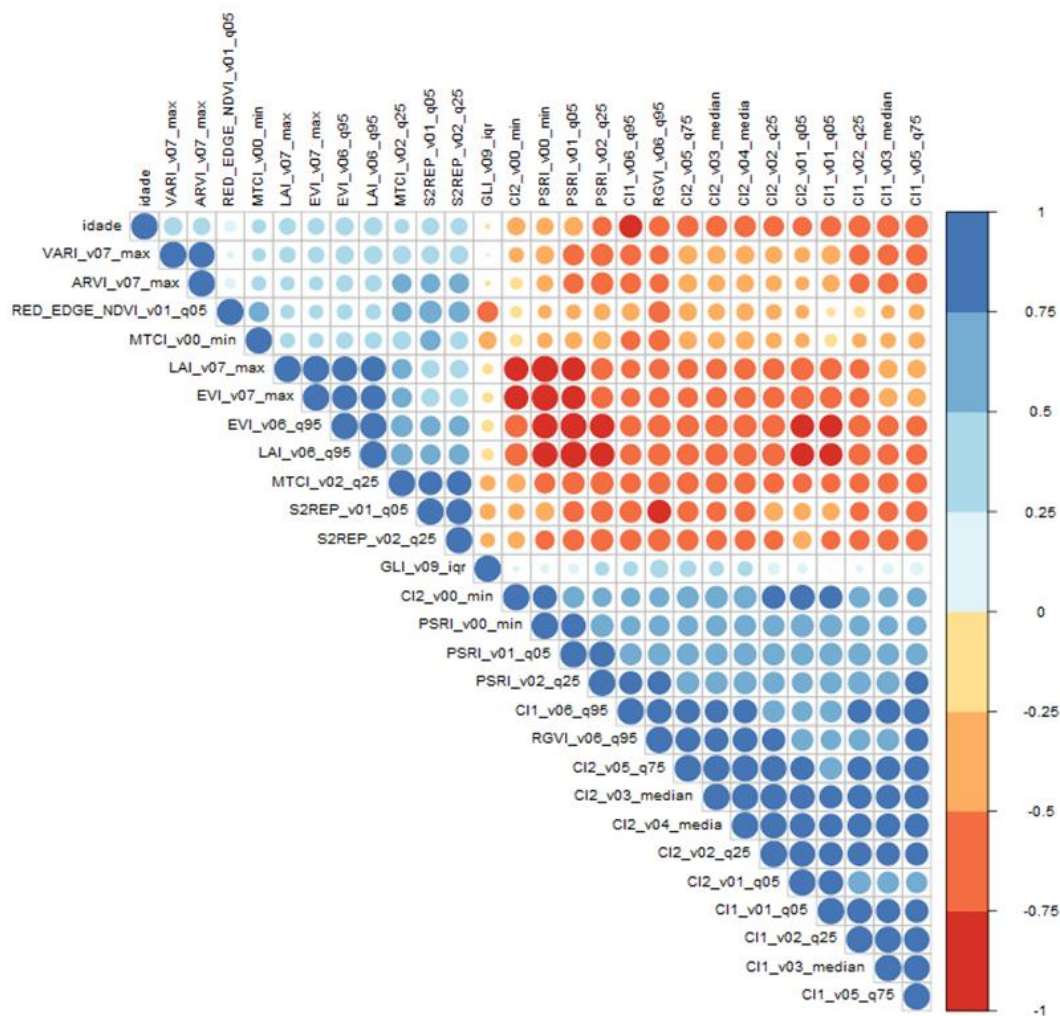
Fonte: Elaborado pela autora, 2022.

Figura 53 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo *Ranger*.



Fonte: Elaborado pela autora, 2022.

Figura 54 - Análise de correlação de Spearman entre os índices de vegetação com maior grau de importância para predição de produtividade, modelo *Random Forest*.



Fonte: Elaborado pela autora, 2022.

5 CONCLUSÃO

Índices de vegetação derivados da refletância espectral, associados à aprendizagem de máquina, são úteis no monitoramento do estado nutricional e de produtividade na cultura da palma de óleo.

A modelagem do estado nutricional com base nos dados do satélite Sentinel 2A se mostrou eficaz, visto que foi possível definir os índices de vegetação que mais se adequaram as informações nutricionais da palma de óleo. Entretanto, o Fe não respondeu à modelagem, sendo necessários outros estudos sobre as técnicas de predição para este elemento;

Além do NDVI, o GRNDVI, MSAVI, ARI, VARI e NDWI foram os índices de vegetação mais representativos para monitoramento nutricional de palma de óleo;

Os algoritmos *Cubist*, *Ranger* e *Random Forest* são promissores para prever a produtividade da cultura palma de óleo, sendo o *Random Forest* o modelo que apresentou os melhores resultados;

Através dos índices de vegetação é possível fazer o monitoramento das lavouras quanto à produtividade, sendo o Red_edge_NDVI, RGVI, GRNDI e LAI os que melhor se ajustaram na base de dados deste estudo.

Estratificar as informações por material genético e idade pode contribuir com melhor ajuste dos modelos e seleção dos índices mais apropriados para monitorar o estado nutricional e a produtividade da cultura da palma de óleo.

REFERÊNCIAS

- ABRAHÃO, S. A.; PINTO, F.; QUEIROZ, D. M.; SANTOS, N. T.; GLERIANI, J. M.; ALVES, E. A. Índices de vegetação de base espectral para discriminar doses de nitrogênio em Capim-Tanzânia. **Revista Brasileira de Zootecnia**, v. 38, n. 9, p. 1637-1644, 2009.
- AGÊNCIA ESPACIAL EUROPEIA (ESA). Software Sentinel Application Platform (SNAP). Versão 8.0. ESA, 2020. Disponível em: <https://step.esa.int/>. Acesso em: 15 jul. 2021.
- AMARAL, L. R., MOLIN, J. P., PORTZ, G., FINAZZI, F. B., & CORTINOVE, L. Comparison of crop canopy reflectance sensors used to identify sugarcane biomass and nitrogen status. **Precision Agriculture**, v. 16, n. 1, p. 15 – 28, 2015.
- BARET, F.; GUYOT, G. Potential and limitations of vegetation indices for LAI and APAR assessment. **Remote Sensing of Environment**, v. 35, p. 161-173, jan. 1991.
- BENDIG, J.; YU, K.; AASEN, H.; BEMMERTS, S.; BROSCHEIT, J.; GNYP, M. L.; BARETH, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. **International Journal of Applied Earth Observation and Geoinformation**, v. 39, p. 79-87, jul. 2015.
- BLACKBURN, G. A. Quantifying chlorophylls and carotenoids at leaf and canopy scales: an evaluation of some hyperspectral approaches. **Remote Sensing of Environment**, v. 66, n. 3, p. 273-285, 1998.
- BOEGH, E. Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. **Remote Sensing of Environment**, v. 81, n. 2-3, p. 179–193, ago. 2002.
- BROGE, N. H.; LEBLANC, E. Comparing predicting power and stability of broadband and hyperspectral vegetation indices for estimation of green area index and canopy chlorophyll density. **Remote Sensing of Environment**, v. 76, n. 2, 156–172, 2000.
- CAMARGO-NETO, J. A. **Combined Statistical—Soft Computing Approach for Classification and Mapping Weed Species in Minimum Tillage Systems**. 2004. Dissertation, University of Nebraska, Lincoln, 2004.
- CARVALHO, L.F.D. Índices de vegetação de dossel obtidos por sensoriamento remoto próximas e embarcado em aeronave remotamente pilotada e sua relação com a produtividade. Universidade Federal de Santa Maria. Tese de Doutorado. p. 91. Santa Maria, RS, 2019.
- CHEN, J. Evaluation of vegetation indices and modified simple ratio for boreal applications. **Can. J. Remote Sens**, v. 22, n. 3, p. 229–242, 1996.
- CHUVIECO, E.; MARTÍN, M. P.; PALACIOS, A. Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination. **International Journal of Remote Sensing**, v. 23, n. 23, p. 5103-5110, 2002.

COOPS, N. C.; JOHNSON, M.; WULDER, M. A.; WHITE, J. C. Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. **Remote Sensing of Environment**, v. 103, n. 1, p. 67–80, jul. 2006.

CRIPPEN, R. E. Calculating the vegetation index faster. **Remote Sensing of Environment**, v. 34, n. 1, p. 71-73, out. 1990.

DASH, J.; CURRAN, P. J. Evaluation of the MERIS Terrestrial Chlorophyll Index (MTCI). **Advances in Space Research**, v. 39, n. 1, p. 100-104, 2007.

DAUGHTRY, C. S. T.; WALTHALL, C. L.; KIM, M. S.; COLSTOUN, E. B. de; MCMURTREY, J. E. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. **Remote Sensing of Environment**, v. 74, n. 2, p. 229-239, nov. 2000.

FALKOWSKI, M. J.; GESSLER, P. E.; MORGAN, P.; HUDAK, A. T.; SMITH, A. M. S. Characterizing and mapping forest fire fuels using ASTER imagery and gradient modeling. **Forest Ecology and Management**, v. 217, n. 2-3, p. 129-146, out. 2005.

FERNANDEZ-MANSO, A.; FERNANDEZ-MANSO, O.; QUINTANO, C. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. **International Journal of Applied Earth Observation and Geoinformation**, v. 50, p. 170-175, ago. 2016.

FERREIRA, E.J.S. Avaliação de índices de vegetação para estimativa de produtividade do milho em sistemas de manejo de solos no cerrado. Dissertação de Mestrado. Universidade Estadual Paulista. p.193. Ilha Solteira, SP, 2020.

FOREIGN AGRICULTURE SERVICE, USDA. Palm Oil 2022: World Production. Disponível em <https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=4243000> Acesso em: 15 fev. 2022.

FRAMPTON, W. J.; DASH, J.; WATMOUGH, G.; MILTON, E. J. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *Journal of Photogrammetry and Remote Sensing*, v. 82, p. 83-92, ago. 2013.

GAMON, J. A.; SURFUS, J. S. Assessing leaf pigment content and activity with a reflectometer. **New Phytologist**, v. 143, p. 105–117, mar. 1999.

GAO, B. C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote Sensing of Environment**, v. 58, n. 3, 257-266, dez. 1996.

GITELSON, A. A. Wide Dynamic Range Vegetation Index for remote quantification of biophysical characteristics of vegetation. **Journal of Plant Physiology**, v. 161, n. 2, p. 165-173, fev. 2004.

GITELSON, A. A.; GRITZ, Y.; MERZLYAK, M. N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. **J. Plant Physiol**, v. 160, n. 3, p. 271-282, mar. 2003.

GITELSON, A. A.; MERZLYAK, M. N.; CHIVKUNOVA, O. B. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. **Photochemistry and Photobiology**, v. 71, n. 1, p. 38-45, jul. 2001.

GITELSON, A. A.; ZUR, Y.; CHIVKUNOVA, O. B.; MERZLYAK, M. N. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. **Photochemistry and Photobiology**, v. 75, n. 3, p. 272-281, mar. 2002.

GITELSON, A.A.; KAUFMAN, Y. J.; MERZLYAK, M .N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sensing of Environment**, v. 58, n. 3, 289–298, dez. 1996.

GOEL, N. S.; QIN, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. **Remote Sensing of Environment**, v. 10, n. 4, p. 309–347, out. 1994.

GÖKKAYA, K.; THOMAS, V.; NOLAND, T. L.; MCCAUGHEY, H.; MORRISON, I; TREITZ, P. Prediction of macronutrients at the canopy level using space-borne imaging spectroscopy and LiDAR data in a mixed wood boreal forest. **Remote Sensing**, v. 7, n. 7, p. 9045-9069, jul. 2015.

GUZMÁN, C. P. H.; 2014. Teores referências de nutrientes em folhas de dendê (*Elaeis guineenses*. Jacq) para as condições da Amazônia. Dissertação de mestrado, Viçosa, Universidade Federal de Viçosa. 74 p.

HABOUDANE, D.; MILLER, J. R.; PATTEY, E.; ZARCO-TEJADA, P. J.; STRACHAN, I. B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. **Remote Sensing of Environment**, v. 90, n. 3, p. 337-352, abr. 2004.

HAGUE, T.; TILLET, N. D.; WHEELER, H. Automated Crop and Weed Monitoring in Widely Spaced Cereals. **Precision Agriculture**, v. 7, n. 1, p. 21-32, mar. 2006.

HARDISKY, M. A.; KLEMAS, V.; SMART, R. M. The Influence of Soil Salinity, Growth Form, and Leaf Moisture on the Spectral Radiance of *Spartina alterniflora* Canopies. **Photogrammetric Engineering and Remote Sensing**, v. 49, n. 1, p. 77-83, jan. 1983.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, n. 3, p. 295-309, ago, 1988.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). Banco de Dados Meteorológicos. São José dos Campos: 2022. Disponível em: <http://bancodedados.cptec.inpe.br/estatisticas/>. Acesso em: 15 jan. 2022.

JIANG, Z.; HUETE, A. R.; DIDAN, K.; MIURA, T. Development of a Two-Band Enhanced Vegetation Index without a Blue Band. **Remote Sensing of Environment**, v. 112, n. 10, p. 3833-3845, 2008.

JORDAN, C. F. Derivation of leaf area index from quality of light on the forest floor. **Ecology**, v.50, n. 4, p. 663–666, jul. 1969.

KATAOKA, T.; KANEKO, T.; OKAMOTO, H.; HATA, S. Crop growth estimation system using machine vision. **Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)**, v. 2, p. 1079-1083, set. 2003.

KAUFMAN, Y. J.; TANRE, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. **IEEE Transactions on Geoscience and Remote Sensing**, v. 30, n. 2, p. 261-270, mar. 1992.

KER, J.C.; NEVES, J.C.L. Relatório Técnico: Agropalma (Consultoria Técnica). Universidade Federal de Viçosa, Minas Gerais, 2010, 21 p.

KEY, C. H.; BENSON, N. C. Landscape Assessment (LA) sampling and analysis methods. **USDA Forest Service - General Technical Report RMRS-GTR**, n. 164 RMRS-GTR, 2006.

LADEIA, C. A.; 2020. Uso de sensores ativos e passivos na estimativa de teores de nitrogênio e do desenvolvimento fitotécnico e produtivo do cafeeiro. Dissertação de Mestrado, Vitória da Conquista, Universidade Estadual do Sudoeste da Bahia. p. 66.

LICHTENTHALER, H. K.; Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. **Methods in Enzymology**. v. 148, p. 350 – 382, 1987.

LOUHAICHI, M.; BORMAN, M. M.; JOHNSON D. E. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. **Geocarto International**, v. 16, n. 1, p. 65-70, mar. 2001.

LYMBURNER, L.; BEGGS, P. J.; JACOBSON, C. R. Estimation of Canopy-Average Surface-Specific Leaf Area Using Landsat TM Data. **Photogrammetric Engineering & Remote Sensing**, v. 66, n. 2, p. 183-191, fev. 2000.

MERZLYAK, M. N.; GITELSON, A. A.; CHIVKUNOVA, O. B.; RAKITIN, V. Y. Non-Destructive Optical Detection of Pigment Changes during Leaf Senescence and Fruit Ripening. **Physiologia Plantarum**, v. 106, n. 1, p. 135-141, 1999.

NUARSA, I. W.; NISHIO, F.; HONGO, C. Spectral Characteristics and Mapping of Rice Plants Using Multi-Temporal Landsat Data. **Journal of Agricultural Science**, v. 3, n. 1, p. 54-67, mar. 2011.

OLIVEIRA, L. F. R.; OLIVEIRA, M. L. R.; GOMES, F. S.; SANTANA, R. C. Estimating foliar nitrogen in Eucalyptus using vegetation indexes. **Scientia Agrícola**, v. 74. n. 2, p. 142-147, mar./ abr. 2017.

OSCO, L. P.; FURUYA, E. D. E. G.; GONÇALVES, W. N.; SANTANA, D. C.; LI, J.; BAILO, F. H. R.; A random forest ranking approach to predict in maize with uav-based vegetation spectral indices. **Computers and Eletronicns in Agriculture**, v. 178, n. 105971, nov. 2020.

PEDRALI, L. D.; MARCHESAN, J.; QUEIROZ, T. B.; ALBA, E.; PEREIRA, R. S.; SCHUMACHER, M. V. Índices de vegetação para análise nutricional de *Eucalyptus saligna* Smith usando espectrorradiometria. **Scientia Forestalis**, v. 50, p. 1 – 12, 2022.

POVH, F. P.; MOLIN, J. P.; GIMENEZ, L.M.; PAULETTI, V.; SALVI, J.V.; Comportamento do NDVI obtido por sensor ótico ativo em cereais. Pesquisa agropecuária brasileira, Brasília, v.43, n.8, p.1075-1083, 2008.

QI, J.; CHEHBOUNI, A.; HUETE, A. R.; KERR, Y. H.; SOROOSHIAN, A. A modified soil adjusted vegetation index. **Remote Sensing of Environment**, v. 48, n. 2, 119-126, maio 1994.

RICHARDSON, A. J.; WIEGAND, C. L. Distinguishing vegetation from soil background information. **Photogrammetric Engineering & Remote Sensing**, v. 43, n. 12, p. 1541–1552, dez. 1977.

RONDEAUX, G.; STEVEN, M.; BARET, F. Optimization of soil-adjusted vegetation indices. **Remote Sensing and Environment**, v. 55, n. 2, p. 95–107, fev. 1996.

ROUJEAN, J. L.; BREON, E. M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. **Remote Sensing of Environment**, v. 51, n. 3, p. 375-384, mar. 1995.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring Vegetation Systems in the Great Plains with ERTS. **Nasa Special Publication**, Washington DC, v. 1, p. 309-317, jan. 1974.

SECRETÁRIA DE ESTADO DE DESENVOLVIMENTO AGROPECUÁRIA E DA PESCA, ESTADO DO PARÁ. Óleo de palma 2023 - Produção Mundial. Disponível em: <https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=4243000>. Acesso em: 04 fev. 2023.

SOARES, J.A.C.; FARIAS, P. R.S.; SANTOS, A. V. F.; MOUTINHO, M.T.; ROCHA; A. J. S. Estimativa da produtividade em plantios jovens de palma de óleo convencional por sensor orbital no município de Tailândia-Pa. **International Journal of Development Research**, v.10, p.40509-40514, set. 2020

SOUSA, J. S.; 2020. Teores e conteúdos foliares de nutrientes modelados com base em componentes de produção da palma de óleo (*Elaeis guineensis* Jacq.). Tese de Doutorado. Universidade Federal de Viçosa. p. 106. Viçosa, MG. 2020.

TEIXEIRA, R.O. NDVI e sua relação com a produtividade e estado nutricional da cultura palma de óleo no estado do Pará. Dissertação de Mestrado. Universidade Federal de Viçosa. p. 34. Viçosa, MG, 2018.

THEMISTOCLEOUS, K. DEM modeling using RGB-based vegetation indices from UAV images. **Proceedings of the SPIE**, v. 11174, jun. 2019.

TUCKER, C. J. A spectral method for determining the percentage of green herbage material in clipped sample. **Remote Sensing of Environment**, v. 9, n. 2, p. 175-181, mar. 1980.

TUCKER, C. J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, v. 8, n. 2, p. 127-150, maio 1979.

UNITED STATES GEOLOGICAL SURVEY (USGS). **USGS EROS Archive - Sentinel-2**. [S. l.], 2021. Disponível em: <https://earthexplorer.usgs.gov/>. Acesso em: 05 abr. 2021.

VESCOVO, L.; GIANELLE, D. Using the MIR bands in vegetation indices for the estimation of grassland biophysical parameters from satellite remote sensing in the Alps region of Trentino (Italy). **Adv. Space Res**, v. 41, n. 11, p. 1764-1772, 2008.

WANG, F. M.; HUANG, J. F.; TANG, Y. L.; WANG, X. Z. New vegetation index and its application in estimating leaf area index of rice. **Rice Science**, v. 14, n. 3, p. 195-203, set. 2007.

WOEBBECKE, D. M.; MEYER, G. E.; VON BARGEN, K.; MORTENSEN, D. A. Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions. **Transactions of the ASAE**, v. 38, n. 1, p. 259-269, 1995.

WRIGHT, D. L.; RASMUSSEN, V. P.; RAMSEY, R.D.; BAKER, D. J.; ELLSWORTH, J.W.; Canopy reflectance estimation of wheat nitrogen content for grain protein management. **GIScience Remote Sens**. v. 41, p. 287–300, 2004.

WU, W. The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization. **Remote Sens**, v. 6, n. 2, p. 1211-1233, jan. 2014.

YANG, K.; PINKER, R. T.; MA, Y.; KOIKE, T.; WONSICK, M. M.; COX, S. J.; ZHANG, Y.; STACKHOUSE, P. Evaluation of satellite estimates of downward shortwave radiation over the Tibetan Plateau. **Journal of Geophysical Research**, v. 113, n. 17, set. 2008.

ZANZARINI, F. V; PISSARA, T. C. T.; BRANDÃO, F. J. C.; TEIXEIRA, D. D. B.; Correlação espacial do índice de vegetação (NDVI) de imagem Landsat/ETM+ com atributos do solo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 17, n. 6, p. 608-614, 2013.

ZARCO-TEJADA, P. J.; MILLER, J. R.; NOLAND, T. L.; MOHAMMED, G. H.; SAMPSON, P. H. Scaling-up and model inversion methods with narrow-band optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. **IEEE Transactions on Geoscience Remote Sensing**, v. 39, n. 7, p. 1491-1507, 2001.

ZHANG, X.; LIU, F.; XIANGYANG, G.; Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. **Biosystems Engineering**, v. 115, n. 1, p. 56 – 65, 2014.

ZHU, G.; LU, W.; CHEN, J. M.; LIU, Y. A Novel Moisture Adjusted Vegetation Index (MAVI) to Reduce Background Reflectance and Topographical Effects on LAI Retrieval. **Plos One**, v. 9, n. 7, jul. 2014.