

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

DIAGNÓSTICO DO ESTADO NUTRICIONAL E PREVISÃO
DA PRODUTIVIDADE DA BATATEIRA

Danilo dos Reis Cardoso Passos

Engenheiro Agrônomo

2023

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

DIAGNÓSTICO DO ESTADO NUTRICIONAL E PREVISÃO
DA PRODUTIVIDADE DA BATATEIRA

Danilo dos Reis Cardoso Passos

Orientador: Prof. Dr. Arthur Bernardes Cecílio Filho

Coorientador: Prof. Dr. Glauco de Souza Rolim

Coorientador: Prof. Dr. Rogério Peres Soratto

**Tese apresentada à Faculdade de
Ciências Agrárias e Veterinárias – Unesp,
Câmpus de Jaboticabal, como parte das
exigências para a obtenção do título de
Doutor em Agronomia (Produção
Vegetal).**

2023

P289d Passos, Danilo dos Reis Cardoso

 Diagnóstico do estado nutricional e previsão da produtividade da
batateira / Danilo dos Reis Cardoso Passos. --Jaboticabal, 2023

 151 p. : il., tabs., mapas

 Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

 Orientador: Arthur Bernardes Cecílio Filho

 1. Olericultura. 2. Minerais na nutrição de plantas. 3.

 Aprendizado do computador. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo
autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: DIAGNÓSTICO DO ESTADO NUTRICIONAL E PREVISÃO DA PRODUTIVIDADE DA BATATEIRA

AUTOR: DANILO DOS REIS CARDOSO PASSOS

ORIENTADOR: ARTHUR BERNARDES CECILIO FILHO

COORIENTADOR: ROGÉRIO PERES SORATTO

COORIENTADOR: GLAUCO DE SOUZA ROLIM

Aprovado como parte das exigências para obtenção do Título de Doutor em Agronomia (Produção Vegetal), pela Comissão Examinadora:

Documento assinado digitalmente
 ARTHUR BERNARDES CECILIO FILHO
Data: 17/08/2023 08:02:53-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. ARTHUR BERNARDES CECILIO FILHO (Participação Virtual)
Departamento de Ciências da Produção Agrícola / FCAV UNESP Jaboticabal

Documento assinado digitalmente
Prof. Dr. DANILO RICARDO YAMANE (Participação Virtual)  DANILO RICARDO YAMANE
Data: 21/08/2023 08:32:08-0300
Verifique em <https://validar.iti.gov.br>
Universidade de Araraquara (UNIARA) / Araraquara/SP

Prof. Dr. ALEXANDRE DAL PAI (Participação Virtual)  Assinado de forma digital por
Departamento de Engenharia de Bioprocessos e Biotecnologia / FCA UNESP Botucatu Alexandre Dal Pai:20067154883
Dados: 2023.08.17 16:37:40 -03'00'

Documento assinado digitalmente
Prof. Dr. ROGÉRIO TEIXEIRA DE FARIA (Participação Virtual)  ROGÉRIO TEIXEIRA DE FARIA
Data: 18/08/2023 11:52:57-0300
Verifique em <https://validar.iti.gov.br>
Departamento de Engenharia e Ciências Exatas DECEX / FCAV UNESP

Pesquisador Dr. ROBERTO BOTELHO FERRAZ BRANCO (Participação Virtual)
Instituto Agrônomo de Campinas (APTA) / Ribeirão Preto/SP

Documento assinado digitalmente
 ROBERTO BOTELHO FERRAZ BRANCO
Data: 17/08/2023 16:51:40-0300
Verifique em <https://validar.iti.gov.br>

Jaboticabal, 16 de agosto de 2023

DADOS CURRICULARES DO AUTOR

DANILO DOS REIS CARDOSO PASSOS – nascido em 17 de janeiro de 1992 na cidade de Santo Antônio de Jesus, Bahia. Filho de Edgar Rocha Passos e Luciana dos Reis Cardoso Passos; e irmão de Rafael dos Reis Cardoso Passos. Em 2011 ingressou no curso de Agronomia da Universidade Federal do Recôncavo da Bahia (UFRB). Durante a graduação participou do Programa de Educação Tutorial: Mata Atlântica, Conservação e Desenvolvimento, realizou intercâmbio na França, onde estudou na AgroSup Dijon: *Institut National Supérieur des Sciences Agronomiques*, estagiou no *Institut National de la Recherche Agronomique* (INRA) e na Embrapa Mandioca e Fruticultura Tropical. Em julho de 2017 recebeu o título de Engenheiro Agrônomo pela UFRB. Em agosto de 2019 obteve o título de Mestre no Programa de Pós-Graduação em Agronomia (Produção Vegetal) da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de Jaboticabal sob a orientação do Prof. Dr. Arthur Bernardes Cecílio Filho. Em agosto de 2019, iniciou o curso de Doutorado pelo mesmo programa de Pós-graduação na Unesp/FCAV, sob a orientação do Prof. Dr. Arthur Bernardes Cecílio Filho.

“O objetivo é transformar dados em informação e informação em *insights*.”

Carly Fiorina

Aos meus pais, Edgar e Luciana, por todo incentivo e comprometimento na minha formação profissional. Ao meu irmão, Rafael, pelo companheirismo. Ao meu orientador, Prof. Arthur, por acreditar no meu potencial.

Dedico

AGRADECIMENTOS

À Universidade Estadual Paulista “Júlio de Mesquita Filho”, UNESP, Campus de Jaboticabal por proporcionar a execução deste trabalho e fomentar o meu desenvolvimento profissional.

Ao programa de Pós-Graduação em Agronomia (Produção Vegetal), pela qualidade e pelo zelo com os estudantes.

Ao professor Dr. Arthur Bernardes Cecilio Filho, por cada ensinamento e colaboração no meu crescimento profissional e pessoal. Muito obrigado.

Aos professores Dr. Glauco de Souza Rolim e Dr. Rogério Peres Soratto por contribuírem no desenvolvimento deste trabalho.

Aos servidores do Departamento de Produção Vegetal e do Setor de Olericultura pela amizade, atenção e apoio.

Agradeço aos colegas e amigos da Unesp Jaboticabal pelo apoio crucial durante minha jornada acadêmica. Levo comigo memórias valiosas e agradecimento por tornarem essa fase especial. Que possamos manter os laços e continuar apoiando uns aos outros no futuro.

Por fim, expresso minha gratidão a todos que desempenharam um papel em minha jornada acadêmica, depositando uma confiança significativa em mim e contribuindo de diversas maneiras para a conclusão deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO	xi
ABSTRACT	xii
LISTA DE TABELAS	xiii
LISTA DE FIGURAS	xiv

CAPÍTULO 1 CONSIDERAÇÕES GERAIS	1
1.1. INTRODUÇÃO	1
1.2. REVISÃO DE LITERATURA.....	3
1.2.1 Batata: Panoramas agrônomo e socioeconômico	3
1.2.2 Método CND para avaliação do estado nutricional.....	5
1.2.3 Modelagem para previsão da produtividade da batata	8
1.2.4 Parâmetros climáticos para a modelagem	10
1.2.5 Parâmetros nutricionais para a modelagem	12
1.3 REFERÊNCIAS.....	16

CAPÍTULO 2 – PADRÕES NUTRICIONAIS PARA A BATATEIRA PELO MÉTODO CND.....	25
RESUMO.....	25
ABSTRACT	26
2.1 INTRODUÇÃO	27
2.2 MATERIAL E MÉTODOS	29
2.2.1 Composição e caracterização do banco de dados	29
2.2.2 Obtenção das normas pelo método CND- <i>clr</i>	32
2.2.3 Obtenção das normas pelo método CND- <i>ilr</i>	34
2.3 RESULTADOS E DISCUSSÃO.....	38
2.3.1 CND- <i>clr</i> , normas, ordem de nutrientes limitantes e faixa de e suficiência.....	39
2.3.2 CND- <i>ilr</i> , normas, partição binária sequencial e partição de Cate Nelson.....	45
2.4 CONCLUSÕES	52
2.4.1 CND- <i>clr</i>	52
2.4.2 CND- <i>ilr</i>	52
2.5 REFERÊNCIAS.....	53

CAPÍTULO 3 - PREVISÃO DA PRODUTIVIDADE DA CULTURA DA BATATA A PARTIR DE MODELOS DE MACHINE LEARNING.....	60
RESUMO	60
ABSTRACT	61
3.1 INTRODUÇÃO	62
3.2 MATERIAL E MÉTODOS	63
3.2.1 Caracterização do banco de dados	64
3.2.2 Dados meteorológicos	67
3.2.3 Caracterização dos estádios fenológicos da batateira	69

3.2.4	Padronização.....	73
3.2.5	Seleção de variáveis.....	74
3.2.6	Modelos de <i>Machine Learning</i>	75
3.2.6.1	Regressão Linear Múltipla (RLM).....	76
3.2.6.2	K-Nearest Neighbors (KNN).....	77
3.2.6.3	Support Vector Machine (SVM).....	77
3.2.6.4	eXtreme Gradient Boosting.....	77
3.2.6.5	Redes neurais artificiais (RN).....	78
3.2.7	Biblioteca <i>Caret</i>	79
3.2.8	Treinamento e teste, parâmetros e validação cruzada	80
3.2.9	Avaliação de Modelos Preditivos.....	82
3.2.10	Resultados predictors	84
3.2.11	Aplicação R Shiny para análise dos modelos de <i>Machine Learning</i>	84
3.3	RESULTADOS E DISCUSSÃO.....	102
3.3.1	Heatmap de correlação.....	105
3.3.2	Clusterização aglomerativa.....	107
3.3.3	Análise de componentes principais.....	109
3.3.4	Random Forest	111
3.3.5	Resultados dos parâmetros	113
3.3.6	Resultados dos modelos de <i>Machine Learning</i>	114
3.4	CONCLUSÕES	126
3.5	REFERÊNCIAS.....	128

DIAGNÓSTICO DO ESTADO NUTRICIONAL E PREVISÃO DA PRODUTIVIDADE DA BATATEIRA

RESUMO – A avaliação do estado nutricional das plantas e as mudanças nas variáveis climáticas têm um impacto significativo na cultura da batata. Com o objetivo de otimizar a produtividade da cultura, melhorar a qualidade do produto e reduzir o impacto ambiental, ferramentas tecnológicas estão cada vez mais sendo utilizadas na agricultura. Entre elas, destaca-se o método de diagnose da composição nutricional (CND), que considera as relações entre todos os nutrientes e a previsão da produtividade. Os modelos de *Machine Learning*, com base em dados climáticos e nutricionais, são ferramentas úteis para auxiliar os produtores a tomarem decisões durante o cultivo. Os objetivos do Capítulo 2 foram: a) obter as normas dos métodos CND-*clr* e CND-*ilr* para avaliação do estado nutricional da batateira; b) identificar as amostras equilibradas ou não-equilibradas nutricionalmente, em situações de alta e baixa produtividade pelo método CND-*ilr*; c) obter as faixas de teores adequados dos nutrientes para a cultura da batata com alta produtividade; d) conhecer a ordem dos nutrientes limitantes nas amostras classificadas como nutricionalmente desequilibradas e com baixa produtividade. Os objetivos do capítulo 3 foram: a) prever a produtividade da batateira; b) avaliar o desempenho de diferentes modelos de *Machine Learning* na previsão da produtividade da batateira no estado de São Paulo, Brasil; e c) desenvolver uma aplicativo on-line que faça a previsão da produtividade da batateira. O banco de dados foi composto por 752 amostras que relacionaram os teores foliares de nutrientes e produtividades de culturas da batata, de áreas localizadas no estado de São Paulo, Brasil. A produtividade que separou os grupos de alta e baixa produtividade foi 48.993,24 kg ha⁻¹. Foram obtidas faixas de suficiência dos nutrientes, em geral, mais estreitas que as reportadas na literatura. A ordem de limitação dos nutrientes foi: N, B, S, Ca e Mn. Os parâmetros encontrados no CND-*ilr* para acurácia, sensibilidade, especificidade, NPV e PPV foram 96,9; 97,1; 93,6; 64,4 e 99,6 %, respectivamente. Para identificar as variáveis que mais impactam na produtividade da batateira, foram desenvolvidos cinco modelos de *Machine Learning* para análise dos dados. Foi desenvolvida uma aplicação e as opções de modelos utilizados foram: Regressão Linear Múltipla (RLM), *K-nearest neighbors* ou “K-vizinhos mais próximos” (KNN), *Support Vector Machine* (SVR), *eXtreme Gradient Boosting* (XGB) e Redes Neurais Artificiais (RN). Para a construção dos modelos, o banco de dados foi dividido em três grupos: variáveis nutricionais, climáticas e top 10 (as principais variáveis que influenciam no modelo pelo método Random Forest). Dentre os 15 modelos, o SVR top 10 obteve os melhores resultados de R² ajustado (0.76) e RMSE (5646.80 kg ha⁻¹).

Palavras-chave: CND batata, nutrição mineral da batata, balanço nutricional, *Machine Learning* na batata, inteligência artificial na batata.

DIAGNOSIS OF NUTRITIONAL STATUS AND PREDICTION OF POTATO PRODUCTIVITY

ABSTRACT – The assessment of the nutritional status of plants and changes in climatic variables have a significant impact on potato cultivation. In order to optimize crop productivity, improve product quality and reduce environmental impact, technological tools are increasingly being used in agriculture. Among them, the CND stands out, which considers the relationships between all nutrients, and the prediction of productivity from *Machine Learning* models based on climatic and nutritional data are useful tools to help producers make decisions during cultivation. The objectives of chapter 2 were: a) to obtain the norms of the CND-*clr* and CND-*ilr* methods for evaluating the nutritional status of the potato plant; b) identification of nutritionally balanced or non-balanced samples, in situations of high and low productivity by the CND-*ilr* method; c) obtaining the ranges of adequate levels of nutrients for the potato crop with high productivity; d) knowing the order of limiting nutrients in samples classified as nutritionally unbalanced and with low productivity. The objectives of chapter 3 were: a) to predict potato productivity; b) to evaluate the performance of different *Machine Learning* models in predicting potato productivity in the state of São Paulo, Brazil; c) to develop an online application that forecast the productivity of agricultural crops. The database consisted of 752 samples that related leaf nutrient contents and yields of potato crops from areas located in the state of São Paulo, Brazil. The productivity that separated the high and low productivity groups was 48,993.24 kg ha⁻¹. Nutrient sufficiency ranges were generally narrower than those reported in the literature. The order of nutrient limitation was: N, B, S, Ca and Mn. The parameters found in the CND-*ilr* for accuracy, sensitivity, specificity, NPV and PPV were 96.9; 97.1; 93.6; 64.4 and 99.6%, respectively. To identify the variables that most impact potato productivity, five *Machine Learning* models were developed for data analysis. An application was developed to perform calculations based on user interaction. The model options used were: Multiple Linear Regression (RLM), K-nearest neighbors or “K-nearest neighbors” (KNN), Support Vector Machine (SVR), eXtreme Gradient Boosting (XGB) and Artificial Neural Networks (RN). for the construction of the models, the database was divided into three groups: nutritional, climatic and top 10 variables (the main variables that influence the model according to the Random Forest method). Among the 15 models, the top 10 SVR obtained the best adjusted R² (0.76) and RMSE (5646.80 kg ha⁻¹) results.

Keywords: Potato CND, potato mineral nutrition, nutritional balance, potato *Machine Learning*, potato artificial intelligence.

LISTA DE TABELAS

CAPÍTULO 2 – PADRÕES NUTRICIONAIS PARA A BATATEIRA PELO MÉTODO CND

Tabela 1.	Caracterização dos experimentos que compõem o banco de dados	31
Tabela 2.	Normas CND- <i>clr</i> (média e desvio padrão) das variáveis multnutriente (Vn) da batateira na população de alta produtividade pelo método para produtividades de 41.666,66 kg ha ⁻¹	39
Tabela 3.	Faixa de suficiência das concentrações dos nutrientes na folha da batata.....	43
Tabela 4.	Partição binária sequencial (PBS) para os 11 balanços ortonormais.....	46
Tabela 5.	Partição das amostras do banco de dados nos quadrantes verdadeiro negativo (VN), falso positivo (FP), falso negativo (FN) e verdadeiro positivo (VP)	47
Tabela 6.	Resultados do teste t comparando as médias de amostras classificadas como verdadeiro negativo (VN) e das classes VP, FN e FP de balanços ortonormais	48
Tabela 7.	Normas CND- <i>ilr</i> (matriz multnutriente e média) calculadas a partir das coordenadas CND- <i>ilr</i> da população de referência (VN)	49

CAPÍTULO 3 – PREVISÃO DA PRODUTIVIDADE DA CULTURA DA BATATA A PARTIR DE MODELOS DE MACHINE LEARNING

Tabela 1.	Caracterização dos experimentos utilizados no banco de dados	66
Tabela 2.	Parâmetros dos modelos.....	81
Tabela 3.	Valores ajustados para os parâmetros	113
Tabela 4.	R ² ajustado e RMSE (raiz do erro quadrático médio) do treinamento e teste das variáveis nutricionais para os modelos de Machine Learning.....	116
Tabela 5.	R ² ajustado e RMSE (raiz do erro quadrático médio) do treinamento e teste das variáveis climáticas para os modelos de Machine Learning.....	119
Tabela 6.	R ² ajustado e RMSE (raiz do erro quadrático médio) do treinamento e teste das variáveis top 10 para os modelos de Machine Learning.....	121

LISTA DE FIGURAS

CAPÍTULO 2 – PADRÕES NUTRICIONAIS PARA A BATATEIRA PELO MÉTODO CND

Figura 1.	Mapa de localização dos municípios em que foram realizados os experimentos.....	30
Figura 2.	Classificação binária dos dados com índices Distância de Mahalanobis e R^2	37
Figura 3.	Índices CND- <i>clr</i> dos nutrientes da população de alta produtividade. Barras em vermelho, verde e azul indicam que o nutriente está em deficiência, em excesso e adequado, respectivamente.....	40
Figura 4.	Balancos ortonormais dos teores de nutrientes foliares da batateira utilizando a biblioteca Seaborn do Python.....	45
Figura 5.	Distribuição das 587 amostras da batateira (estado nutricional e produtividade) nos quadrantes verdadeiro negativo (VN), falso positivo (FP), falso negativo (FN) e verdadeiro positivo (VP) em função da partição de Cate-Nelson.....	46
Figura 6.	Média móvel e intervalo de confiança dos balancos ortonormais obtidas com amostras das classes de amostras de alta produtividade e equilibradas nutricionalmente (VN) e de baixa produtividade e desequilibradas nutricionalmente (VP).....	50

CAPÍTULO 3 – PREVISÃO DA PRODUTIVIDADE DA CULTURA DA BATATA A PARTIR DE MODELOS DE MACHINE LEARNING

Figura 1.	Mapa de localização dos municípios em que foram realizados os experimentos.....	64
Figura 2.	Fluxograma de cálculo de balanço Hídrico segundo modelo proposto por Thorthwaite e Mather (1955), conforme Rolim et al. (2020)	69
Figura 3.	Apresentação dos estádios fenológicos da batateira	72
Figura 4.	Esquema da divisão de dados em 75% para treinamento e 25% para teste e demonstração da validação cruzada em quatro vezes	80
Figura 5.	Exemplo de construção do Diagrama de Taylor. (Adaptado de Pereira et al., 2014).....	83
Figura 6.	Tela inicial da aplicação – menu Início	86

Figura 7.	Tela para carregar o arquivo de dados – menu Banco de dados > Upload.....	88
Figura 8.	Tela que seleciona a variável resposta, filtra e define as variáveis independentes – menu Banco de dados > variáveis > Variáveis do banco de dados.....	90
Figura 9.	Tela que visualiza as variáveis independentes selecionadas com e sem padronização – menu Resultados > Dados > Tabela de dados	92
Figura 10.	Tela que visualiza as variáveis independentes selecionadas com e sem padronização – menu Resultados > Dados > Tabela de dados tratados.....	93
Figura 11.	Tela que seleciona a variável resposta, filtra e define as variáveis independentes – menu Resultados > Modelagem > Resultados	94
Figura 12.	Tela que visualiza os parâmetros – menu Resultados > Modelagem > Resultados.....	96
Figura 13.	Tela que visualiza os valores dos predictors – menu Resultados > Modelagem > Resultado predictor	97
Figura 14.	Tela que visualiza as médias do R^2 ajustado e do RMSE – menu Resultados > Modelagem > Médias.....	99
Figura 15.	Tela que visualiza as tabelas de dados para treinamento e teste – menu Resultados > Modelagem > Divisão dos dados	100
Figura 16.	Tela que visualiza o conteúdo de ajuda – menu Resultados > Modelagem > Ajuda	101
Figura 17.	Boxplots referente as variáveis nutricionais e climática padronizados.....	102
Figura 18.	Boxplot da variável produtividade em relação a todos os experimentos.....	104
Figura 19.	Heatmap da correlação entre a produtividade e as variáveis nutricionais e climáticas.....	105
Figura 20.	Dendrograma representando a similaridade entre variáveis com base na distância de Pearson	107
Figura 21.	Análise de componentes principais entre a produtividade e as variáveis nutricionais e climáticas.....	110
Figura 22.	Grupos das variáveis que foram selecionadas para as análises e a variável resposta	115

Figura 23.	Aplicação do Diagrama de Taylor representando graficamente as estatísticas e a tabulação cruzada, referentes aos interpoladores utilizados.....	123
-------------------	---	-----

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

1.1 INTRODUÇÃO

A batata é um dos principais alimentos utilizados pela população, considerada fonte de carboidratos e com alto valor nutritivo, devido ao seu ótimo balanço de aminoácidos, nutrientes, compostos antioxidantes (Brown *et al.*, 2005), carotenoides, flavonoides e outros (Navarre *et al.*, 2019). O que a torna um dos vegetais mais consumidos no mundo, devido as suas características, facilidade de preparo e às diversas possibilidades de consumo. Além, de ser um alimento com valor acessível a toda população (Marcomini, 2021).

Em relação à cultura da batata, como em outras espécies agrícolas, a qualidade e a produtividade são influenciadas pelas condições climáticas e o estado nutricional do vegetal (Rozane *et al.*, 2016; Tito *et al.*, 2018). Como pondera Amaral *et al.* (2023), atualmente existe uma maior preocupação com a produção e transformação dos alimentos, originando estudos com intuito de disponibilizar tecnologias que permitam o aumento de produtividade, garantia da segurança alimentar, aporte nutricional e uma maior vida de prateleira.

De tal modo, a tecnologia se torna um elemento essencial para o manejo dos cultivos agrícolas. Nesse sentido, é possível verificar o avanço do uso da chamada tecnologia digital na agricultura, que visa auxiliar os agricultores nos desafios existentes na produção agrícola (Javaid *et al.*, 2023). Corroborando Yang *et al.* (2020), afirmam que a tecnologia digital permite aprimorar a eficiência produtiva, aumentar a produtividade e a qualidade do alimento produzido, além de promover a sustentabilidade ambiental.

Dentre as possibilidades das ferramentas digitais disponibilizadas ao agricultor, existe uma técnica que usa o método *Compositional Nutrient Diagnosis* (CND). O CND tem sido proposto para avaliação nutricional da planta, uma vez que emprega análise multivariada, o que possibilita melhor performance (Parent e Dafir, 1992). Segundo Parent *et al.* (2013b), existem duas abordagens comuns no campo: a relação log centralizada (*clr*), baseada nos estudos de Aitchison (1986); e a relação log isométrica (*ilr*), desenvolvida por Egozcue *et al.* (2003) e Egozcue e Pawlowsky-Glahn (2005).

Esse tipo de metodologia, vem sendo aprimorada, com o desenvolvimento de técnicas para análise de dados composicionais, o que motiva diversos estudos em diferentes locais e períodos. Em relação às pesquisas sobre CND, na cultura da batata, existem diversos estudos mundiais, destacando a realizada por Parent *et al.* (1994), conduzida na região de St. Lawrence Lowlands, no Canadá, por três anos. Já Khiari *et al.* (2001) coletaram dados no Quebec, Canadá, no período compreendido entre os anos de 1988 e 1995. O estudo mais recente foi realizado por Ganeshamurthy *et al.* (2019) correspondendo a sete estados da Índia. Entretanto, no Brasil não existem pesquisas sobre a aplicação do CND na cultura da batata.

Ainda, concomitantemente ao uso do *Compositional Nutrient Diagnosis* é possível a utilização de modelos matemáticos, cada vez mais empregados para a compreensão das variáveis climáticas e seus impactos na cultura da batata (Mellaku e Sebsibe, 2022). Esses modelos são ferramentas que consideram para análise as variáveis ambientais (radiação solar, velocidade do vento, temperatura, umidade do ar e outros), permitindo aos produtores estimarem a produtividade das safras com base nas condições climáticas vigentes.

No caso brasileiro, a batata pode ser cultivada durante todo o ano, conforme a microrregião. Assim, o modelo agroclimático permite realizar estimativas com base nos dados climáticos da microrregião com potencial de cultivo. Esses modelos podem ser mais complexos, envolvendo, por exemplo, parâmetros climáticos e nutricionais. Nesse sentido, a utilização dos modelos agroclimáticos e a avaliação dos teores nutricionais a partir do CND podem ser uma abordagem promissora para a previsão de produtividade e o manejo eficiente da cultura.

Com dados precisos e confiáveis, com base nas condições climáticas e nos teores nutricionais da planta, os modelos de *Machine Learning* podem identificar padrões complexos e fazer previsões precisas de produtividade da batata, permitindo uma melhor tomada de decisão sobre manejo da cultura, redução dos custos, maximização da produção e redução no uso de insumos. Assim, a utilização de modelos de ML na agricultura pode ser uma ferramenta valiosa para os produtores ao prever a produtividade da cultura da batata com base nas condições climáticas e nos teores nutricionais da planta.

Diante do exposto, os objetivos do presente estudo foram: a) obter as normas dos métodos CND-*clr* e CND-*ilr* para avaliação do estado nutricional da batateira e identificação de amostras equilibradas e não-equilibradas nutricionalmente, em situações de alta e baixa produtividade pelo método CND-*ilr*; b) obter as faixas de teores adequados dos nutrientes para a cultura da batata com alta produtividade; c) conhecer a ordem dos nutrientes limitantes nas amostras classificadas como nutricionalmente desequilibradas e com baixa produtividade; d) desenvolver modelos para previsão da produtividade da batateira em função das variáveis meteorológicas e nutricionais usando algoritmos de *machine learning*; e) desenvolver um aplicativo interativo e *on-line* capaz de realizar previsões com base em modelos de *Machine Learning* de regressão.

1.2 REVISÃO DE LITERATURA

1.2.1 Batata: Panoramas agrônômico e socioeconômico

A batata (*Solanum tuberosum* L.) é originária da Cordilheira dos Andes, na América do Sul, pertencente à família Solanaceae, gênero *Solanum*, subespécie *tuberosum*, apresenta uma grande variabilidade genética, aproximadamente 200 espécies e 5.000 variedades conhecidas (Petropoulos *et al.*, 2019). No Brasil, a cultura foi implantada no século XIX, especificamente na região Sul, onde as condições edafoclimáticas são favoráveis para a produção (Silva e Lopes, 2015).

Considerada uma planta anual, herbácea em relação à parte aérea, composta por caules angulosos e ramificados, variando de coloração verde a arroxeada. As folhas são compostas por três ou mais folíolos laterais, um folíolo apical (formato arredondado) e as flores são hermafroditas e apresenta três tipos de caules (aéreo, os estólons e os tubérculos). O destaque cabe ao tubérculo, parte comestível da planta, que apresenta características distintas dependendo da cultivar (Filgueira, 2008).

A espécie é caracterizada pela enorme diversidade fenotípica, o que confere a diferenciação na forma, cor, tamanho, textura da casca, polpa e composição química (matéria seca, proteínas, amido e teor de glicocalóides), o que permite sua ampla utilização (Calliope *et al.*, 2018). O cultivo é melhor adaptado em regiões com

temperaturas amenas (18 a 22 °C), pois temperaturas elevadas reduzem o rendimento e matéria seca dos tubérculos, bem como afetam o crescimento das raízes, parte importante para a absorção de nutrientes. O cultivo exige solos férteis para o crescimento radicular e boa aeração (Silva e Lopes, 2015).

A batateira é uma das culturas mais importantes, depois do arroz, trigo e milho, e a primeira nos denominados não-grãos (Ortiz, 2020), com uma produção global de 368 milhões de toneladas no ano de 2018 (FAOSTAT, 2019). A importância também é devida a sua contribuição para a segurança alimentar e redução da pobreza, já que ela é um substituto para os alimentos básicos, além de fornecer carboidratos na forma de amido, fonte de fibras e nutrientes (niacina, vitamina C e B, proteínas, minerais como potássio, magnésio e ferro) (Sawicka *et al.*, 2018).

No Brasil, a batateira é cultivada em uma área estimada em 118,1 mil hectares, com uma produção de 3,6 milhões de toneladas (ABBA, 2020; IBGE, 2021), o que torna o país um tradicional produtor, conferindo o 21º lugar no ranking mundial em 2020 (Salvador e Pereira, 2021). Nos últimos dez anos houve um aumento de 28,1% na produtividade da batata no país, devido às melhorias nas técnicas de cultivo e uso de cultivares mais produtivas (FAO, 2023). Ainda, se destaca como a segunda hortaliça mais importante no país, cultivada, principalmente, nas regiões Sul e Sudeste, com destaque para os estados de Minas Gerais, São Paulo e Paraná, responsáveis por 32, 22 e 19%, respectivamente (IBGE, 2021).

No Brasil, o plantio ocorre durante todo o ano, durante três safras anuais, sendo: a primeira safra denominada como “Batata das Águas” plantada entre agosto e dezembro, com colheita nos meses de novembro e fevereiro; a segunda “Batata das Secas”, com plantio nos meses de janeiro e março, colheita entre abril e julho; e a terceira safra “Batata de inverno” plantada entre abril e julho, sendo colhida entre julho e outubro (Embrapa Hortaliças, 2023).

O desenvolvimento da batateira compreende, de acordo com Silva e Lopes (2015, p. 6), as seguintes etapas:

I - Brotação a pré-emergência: estende-se por três a seis dias. Os brotos se desenvolvem a partir do tubérculo-semente e começam a emergir do solo, enquanto as raízes começam a se desenvolver; II - Crescimento vegetativo: esta fase se estende por 15 a 30 dias, dependendo da cultivar e das condições ambientais. A parte aérea é formada, enquanto as raízes e estolões se desenvolvem a partir das gemas subterrâneas; III - Início da

tuberização: esta fase se estende por 10 a 15 dias. Inicia-se a formação dos tubérculos nas extremidades dos estolões, como resultado do armazenamento dos fotoassimilados na forma de amido; IV - Crescimento dos tubérculos: o desenvolvimento da folhagem é finalizado enquanto grande quantidade de amido é armazenado, aumentando o tamanho dos tubérculos; e V - Maturação: neste momento, todos os fotoassimilados são direcionados aos tubérculos, e a matéria seca acumulada atinge o máximo, as hastes tendem a prostrar, e as folhas se tornam amareladas até o secamento total da parte aérea, enquanto a película dos tubérculos se torna mais firme.

O florescimento e a frutificação estão relacionados com o fotoperíodo, temperatura, radiação e genótipo. Para o florescimento é necessário um fotoperíodo longo (16 horas), alta luminosidade e temperaturas noturnas moderadas (Pereira *et al.*, 2016). As temperaturas elevadas e dias curtos provocam a abscisão do botão floral e a redução na emissão dos botões florais (Gopal, 1994), o florescimento pode ocorrer concomitantemente à tuberização. O manejo adequado das fases do desenvolvimento é fator determinante para uma alta produtividade e qualidade dos tubérculos (Bisognin e Streck, 2009).

A produtividade e a qualidade dos tubérculos são afetadas por uma ampla gama de fatores, como a cultivar, práticas agronômicas, condições climáticas, solo, doenças, pragas (Trani *et al.*, 1997), condições nutricionais da planta (Fernandes *et al.*, 2011) e as condições meteorológicas (Fortin *et al.*, 2011). Em condições ideais de cultivo, a produtividade pode ser estimada em até 80 toneladas por hectare (Trani *et al.*, 1997). Entretanto, em áreas com baixa tecnologia e condições climáticas adversas, a produtividade pode ser significativamente menor, chegando a menos de 10 toneladas por hectare (Sarkar e Mondal, 2005).

1.2.2 Método CND para avaliação do estado nutricional

As interações entre nutrientes na planta e no solo podem ter efeitos sinérgicos ou antagônicos e essas interações podem levar à diluição ou concentração de nutrientes nos tecidos das plantas. Phillips (2014) afirma que a nutrição mineral e o manejo dos fertilizantes visam garantir a produtividade das culturas.

Conforme Parent (2011), para a otimização da produção agrícola, é fundamental avaliar o estado nutricional das plantas. Assim, a análise foliar é uma ferramenta importante, já que ela identifica o estado nutricional da planta, complementando as análises químicas do solo, o que permite ajustar a recomendação

de fertilizantes. Desse modo, pelas limitações das análises de solo, a diagnose foliar se torna essencial.

O método de diagnose da composição nutricional (CND) permite avaliar as possíveis interações entre os nutrientes, reconhecendo inclusive aquelas de alta ordem, que outros métodos não conseguem explicar, tornando-se assim uma ferramenta importante na diagnose nutricional (Parent e Dafir, 1992). Conforme Holland (1966), o CND visa aprimorar os métodos existentes e, assim, obter diagnósticos nutricionais mais consistentes. Ainda, para o autor, isso porque a razão entre todos os nutrientes em análise, considerando a influência de outros componentes que afetam a estrutura dos tecidos vegetais, tende a representar melhor a indicação do estado nutricional das culturas do que os diagnósticos que avaliam as concentrações dos nutrientes de forma isolada ou por meio de relações duais.

Inicialmente, o CND foi utilizado na avaliação nutricional da cultura do tomate para investigar a relação entre o índice DRIS (*Diagnosis and Recommendation Integrated System*) e o CND, avaliando a aplicabilidade do CND para o diagnóstico multivariado utilizando a análise de componentes principais (ACP). Os autores encontraram alta correlação entre os índices DRIS e CND, demonstrando que a estreita correlação entre os dois métodos permite que a análise bivariada do DRIS possa ser expandida para uma abordagem multivariada usando o CND (Parent *et al.*, 1993). Outros trabalhos são encontrados na literatura que utilizaram o CND para avaliação nutricional das culturas, sendo eles: manga (Parent *et al.*, 2013, Rozane *et al.*, 2007), laranja (Dias *et al.*, 2017; Yamane *et al.*, 2022), milho (Modesto *et al.*, 2014) e tomate (Nowaki *et al.*, 2017).

Ao utilizar métodos de diagnose foliar, é necessário estabelecer uma população de referência para geração das normas nutricionais, que servirão como base para avaliar o equilíbrio nutricional de uma cultura. Para a maioria dos estudos, recomenda-se o uso da população de alta produtividade como referência para a diagnose foliar (Wadt, 1996; Silva *et al.*, 2004; 2005; Dias *et al.*, 2013; 2017; Geiklooi *et al.*, 2021). Isso se deve ao fato de que essas populações possuem condições ótimas de nutrição e outros fatores de produção (Rozane *et al.*, 2009).

As normas CND-clr são obtidas a partir das médias e desvios-padrão das variáveis multinutrientes da população de alta produtividade, e correspondem uma

empregando a transformação logarítmica centrada (*clr*) proposta por Aitchison (1986). Conforme a teoria da análise composicional, essa soma deve ser igual a 1,0 para frações e 100% para dados em porcentagem. Portanto, a transformação logarítmica dos dados é utilizada para controlar essa perturbação e manter as propriedades estatísticas dos dados composicionais (Parent e Natale, 2008).

O método é considerado um dos mais sofisticados de diagnose nutricional, destacando-se pela sua abordagem multivariada, que permite avaliar a interação entre todos os nutrientes e não apenas as interações entre um nutriente em relação a outro (relações duais) ou o seu efeito isolado (Parent e Natale, 2008; Silva *et al.*, 2020). Dessa forma, as normas CND geradas a partir da população de alta produtividade são consideradas balanceadas e equilibradas nutricionalmente, servindo como referência para a diagnose nutricional do restante da população de baixa produtividade (Parent e Dafir, 1992).

Além disso, é observado que, em populações de alta produtividade, os teores de nutrientes tendem a estar mais próximos dos valores ótimos fisiológicos (Wadt, 1996), apresentando menor variabilidade entre eles e maior precisão em suas interações nutricionais (Galdino, 2015). Entretanto, apesar das diversas vantagens oferecidas pelo método CND, é importante destacar suas limitações. Uma delas diz respeito à necessidade de informação sobre os teores de todos os nutrientes em diagnose para a aplicação do método, uma vez que a ausência de qualquer um dos elementos resulta na redução do banco de dados disponível para análise e da quantidade de interações entre os nutrientes, que possivelmente poderia reduzir a acurácia do resultado (Rozane *et al.*, 2007; Rozane *et al.*, 2009).

Outra limitação, seria a falta de consenso quanto ao tamanho amostral ideal para a obtenção de resultados mais precisos. Adicionalmente, há carência de estudos que correlacionem os resultados obtidos pela diagnose nutricional utilizando o CND com as recomendações de adubação. Ademais, ao explorar os distintos métodos propostos para definir a população de alta produtividade, é fundamental reconhecer as especificidades de cada abordagem. O método CND-*clr*, baseado na função cumulativa, e o método CND-*ilr*, fundamentado na otimização de Cate-Nelson, apresentam nuances que demandam uma compreensão aprofundada. (Rozane *et al.*, 2007; Rozane *et al.*, 2009).

Nesse cenário, considerando as vantagens e limitações do método, é possível aferir que ele possui grande potencial de uso na diagnose nutricional das culturas (N, P, K, Ca, Mg, S, B, Cu, Fe, Mn e Zn), principalmente podendo ser aplicado tanto em experimentos controlados quanto em campos comerciais. A avaliação do estado nutricional por meio do método pode resultar em diagnósticos precisos e essenciais para o gerenciamento das lavouras e aumento da produtividade, tornando o CND uma ferramenta importante para a tomada de decisão em manejo nutricional de culturas.

1.2.3 Modelagem para previsão da produtividade da batata

A previsão da produtividade nas culturas é um elemento importante para o agricultor e um desafio complexo, já que envolve informações genotípicas, aspectos ambientais, estrutura física do solo, química do solo, práticas de manejo cultural e outros (Orlandini *et al.*, 2008). O setor agrícola precisa de ideias inovadoras e avanços tecnológicos que auxiliem no aumento da produtividade (Singh *et al.*, 2021).

A inteligência artificial tem conquistado espaço na agricultura pela sua capacidade de análise e interpretação de dados, o que torna os processos mais rápidos e eficazes (Moreti *et al.*, 2021). Dentre as possibilidades tecnológicas, os métodos de aprendizado de máquina, ou *Machine Learning*, têm sido empregados. O *Machine Learning* consiste em análise automatizada de dados através da construção de modelos gerais (Storm *et al.*, 2019). As técnicas de *Machine Learning* apresentam a capacidade de resolver problemas não-lineares de forma autônoma, utilizando conjuntos de dados provenientes de diversas fontes (Chlingaryan *et al.*, 2018).

Nesta técnica, o algoritmo é construído para aprender padrões e reaprender revendo seus próprios padrões observados, utilizando um conjunto de regras matemáticas e suposições estatísticas para desenvolver um modelo baseado em cálculos matemáticos, entre recursos de um determinado conjunto de dados (Janiesch *et al.*, 2021; Gholami e Booij, 2022). As técnicas de análise de dados permitem ao computador aprender a partir dos dados de entrada, sendo capazes de analisar grandes conjuntos de dados, identificando as variáveis mais importantes e criando modelos preditivos precisos (Janiesch *et al.*, 2021).

Para Khaki e Wang (2019), as técnicas de aprendizado de máquina (regressão multivariada, árvore de decisão e redes neurais artificiais) têm sido empregadas na

previsão da produtividade das culturas agrícolas, onde, os modelos - Deep Neural Network, Lasso, Shallow Neural Networks e Regression Tree – apresentaram acurácias variadas, registrando respectivamente 85%, 23%, 60% e 70%. Característica distintiva desses modelos está na abordagem do tratamento em relação à produtividade da cultura, compreendida como uma função implícita das variáveis de entrada, como o clima, o solo e as características da cultura. Essa função pode ser não linear e muito complexa, o que torna a aplicação das técnicas de aprendizado de máquina particularmente adequadas, o que permite a obtenção de previsões precisas e confiáveis (Bengio, 2000). Assim, cada abordagem tem suas vantagens e desvantagens, sendo escolhida de acordo com o objetivo da previsão, o conjunto de dados disponíveis e a capacidade computacional.

Liu e Tian (2013) empregaram a rede neural com uma camada oculta para prever a produção de milho, usando dados do solo, clima e manejo, enquanto Drummond *et al.* (2013) compararam a eficácia de três métodos, a regressão linear múltipla *stepwise*, regressão de busca de projeção e redes neurais, para prever a produtividade das culturas. Na pesquisa, os autores concluíram que o modelo de rede neural foi o mais adequado na previsão da produção. Por fim, Romero *et al.* (2016) utilizaram a árvore de decisão aplicada a mineração de regras de associação para classificar os componentes da produção de trigo.

Em relação às aplicações de *Machine Learning* na agricultura existem diversos outros trabalhos, como exemplo as pesquisas realizadas por Pant *et al.* (2021) e Luciano *et al.* (2021) sobre modelos preditivos para safras; Zareiforoush *et al.* (2015) e Vithu e Moses (2016) com estudos sobre qualidade de grãos; Shah *et al.* (2016) e Chung *et al.* (2016) com modelos para detecção de doenças; e Talaviya *et al.* (2020) para a aplicação de herbicidas e otimização de irrigação, entre outros. Ainda, os modelos podem ser usados para prever a produtividade em certas regiões e épocas do ano, o que pode ajudar no planejamento das safras.

Dessa forma, a modelagem pode ser uma ferramenta valiosa na previsão da produtividade e na segurança alimentar no cultivo da batata (Divya *et al.*, 2021). Ao integrar variáveis, como clima, solo, práticas agrícolas, composição nutricional e doenças, os modelos podem fornecer previsões mais precisas sobre a cultura, isso

em uma determinada região ou safra, maximizando a produtividade da batateira (Haverkort e Struik, 2015; Mishra *et al.*, 2016).

No entanto, a coleta criteriosa dos dados, bem como a correta análise de dados precisos é fundamental para a precisão das previsões dos modelos de *Machine Learning*. Para que os modelos sejam treinados adequadamente é necessário coletar informações detalhadas sobre as condições climáticas e nutricionais da batateira ao longo do tempo. Além disso, é importante que os dados sejam verificados antes de serem utilizados no treinamento dos modelos, para garantir que não haja erros ou inconsistências que possam afetar a qualidade das previsões.

Com dados precisos e confiáveis, com base nas condições climáticas e nos teores nutricionais da planta, os modelos de *Machine Learning* podem identificar padrões complexos e fazer previsões precisas de produtividade da batata, permitindo uma melhor tomada de decisão sobre manejo da cultura e maximização da produção. A utilização de modelos que consideram as variáveis climáticas e nutricionais oferecem vantagens ao modelo climático, sendo elas: a) tomada de decisão mais precisa, prevendo o impacto das variações climáticas e nutricionais nas culturas, auxiliando na tomada das decisões; b) redução de custos e a otimização do uso de recursos; c) aumento da produtividade, identificando as melhores práticas agrícolas para cada tipo de cultura e região; e d) redução do impacto ambiental ao otimizar o uso de recursos (Queiroz *et al.*, 2022).

O que pode permitir a adoção de práticas de manejo mais adequadas e o planejamento mais eficiente do cultivo da cultura. Ainda, essa previsibilidade permite melhor planejamento das operações e a tomada de decisões informadas sobre a gestão da produção, contribuindo para a eficiência no uso de recursos e a maximização da produtividade, resultando em maior segurança alimentar e sustentabilidade da atividade.

1.2.4 Parâmetros climáticos para a modelagem

Os modelos climáticos são ferramentas voltadas para a gestão e planejamento agrícola, permitindo tomadas de decisões com base em informações e dados climáticos (Holzkämper, 2017). A importância desses modelos reside na sua capacidade de prever e mitigar os riscos associados à produção agrícola (variações

climáticas extremas, doenças, pragas e outros), bem como auxiliam na programação do plantio, colheita e melhores práticas de manejo, maximizando a produtividade e qualidade dos produtos agrícolas (Jones *et al.*, 2017).

O modelo comum é baseado no uso dos dados climáticos, tendo em vista que a produtividade de uma cultura é afetada diretamente pela temperatura, precipitação e umidade. O uso de modelos que utilizam dados climáticos históricos, para preverem a produtividade, tem sido amplamente usado, especialmente naquelas regiões com condições climáticas estáveis e previsíveis (Ochieng *et al.*, 2016). Entretanto, a aplicação desses modelos para a cultura da batata ainda é limitada, o que pode ser justificado pelas características únicas da cultura, que apresenta fenologia complexa, influenciada por uma série de fatores, o que torna a modelagem mais desafiadora.

Ademais, a produção de batata apresenta diferenças regionais nas práticas agrônômicas, como época de semeadura e colheita, o que dificulta a criação de um modelo universal para a cultura (Lopes *et al.*, 2011). Os modelos são criados a partir de dados meteorológicos e históricos da cultura da batata, bem como informações sobre o solo e práticas agrônômicas específicas. A Embrapa Hortaliças (2023) pondera que na cultura da batata os fatores climáticos influenciam no ciclo da planta, da germinação até a maturação dos tubérculos. Struik (2007) ressalta a temperatura como fator crucial para a germinação, emergência, floração, tuberização e maturação da batata. Já Ragassi *et al.* (2009) apontam a umidade do solo, pois o déficit hídrico afeta negativamente o crescimento das plantas e reduz a produtividade da batateira.

A duração do ciclo da batata também é influenciada pelas condições climáticas, em regiões mais quentes o ciclo será mais curto, enquanto nas regiões mais frias o ciclo será mais longo. A fase de tuberização corresponde a uma das mais importantes e as condições climáticas, especialmente, a temperatura, influenciam significativamente a formação e o crescimento dos tubérculos (Erpen *et al.*, 2013). As temperaturas elevadas durante a fase de tuberização podem reduzir o tamanho dos tubérculos, enquanto temperaturas mais baixas favorecem o crescimento dos tubérculos (Fagundes *et al.*, 2010).

Em relação à radiação solar, ela é um fator essencial para a fotossíntese, responsável pela produção de energia para a planta. Já precipitação garante a disponibilidade de água para as plantas, porém, o excesso de chuva pode ocasionar

problemas, tais como a inundação das lavouras e o desenvolvimento de doenças fúngicas (Heldwein *et al.*, 2012).

Diante desse contexto, as condições climáticas têm sido preocupação crescente na produção da batata, pois têm efeitos em todas as etapas do desenvolvimento da planta, o que acaba afetando a produtividade e a qualidade dos tubérculos. Assim sendo, os modelos climáticos, podem ajudar a prever a produtividade da cultura em diferentes condições climáticas, auxiliando na tomada de decisões de manejo.

1.2.5 Parâmetros nutricionais para a modelagem

A cultura da batata é altamente responsiva à aplicação de nutrientes no solo, devido ao seu ciclo curto e elevado potencial produtivo, ela requer grandes quantidades de fertilizantes por unidade de área, em um curto período (Fernandes e Soratto, 2013). Dentre os cultivos agrícolas, no Brasil, a batateira se destaca por apresentar o maior consumo relativo de fertilizantes aplicados por hectare. No entanto, é importante ressaltar que a quantidade de nutrientes necessária pode variar dependendo da taxa de crescimento da cultura e da eficiência na conversão dos nutrientes absorvidos em biomassa (Nomura *et al.*, 2008).

A aplicação inadequada de fertilizantes no solo, seja por excesso, seja por insuficiência, pode resultar em baixa produtividade, danos ambientais e prejuízos ao solo. Além disso, a aplicação excessiva pode representar custos desnecessários. Portanto, é fundamental determinar as doses adequadas conforme as necessidades nutricionais da planta, considerando sua fase fenológica e a composição do solo (Fernandes *et al.*, 2011).

Segundo Silva *et al.* (2013), entre todos os nutrientes, o nitrogênio (N) é o que apresenta o maior impacto na produtividade da batateira. A deficiência deste nutriente na planta pode resultar na clorose das folhas mais velhas, além de diminuir o crescimento e a produção de tubérculos, conforme observado por Rosen *et al.* (2014). Zebarth e Rosen (2007) afirmam que o aumento da oferta de N pode contribuir para aumentar a proporção de tubérculos de grande porte.

O fósforo (P) desempenha diversos papéis metabólicos na planta. A sua presença é crítica para a transferência de energia celular, a respiração e a

fotossíntese, atuando também como componente estrutural de fosfolipídios, ácidos nucleicos, coenzimas e fosfoproteínas (Rosen *et al.*, 2014). O P é crucial para os processos de divisão celular, metabolismo de carboidratos, síntese e alocação do amido nos tubérculos (Houghland, 1960). A deficiência de P pode afetar negativamente o crescimento radicular, reduzindo o tamanho dos tubérculos. Os sintomas de excesso de P em plantas de batata são raros, mesmo quando doses elevadas do nutriente são fornecidas. Cabe destacar, que a maioria dos solos brasileiros existe certa deficiência de P, a exceção pode ocorrer quando a adubação for constante e o solo apresente um P residual, exemplo na produção intensiva de hortaliças.

A cultura é altamente dependente do potássio (K) disponível no solo, sendo esse o primeiro macronutriente em ordem de extração. Os níveis adequados de K favorecem a formação e translocação dos carboidratos e o uso eficiente da água pela planta. Esse nutriente desempenha papel crucial na formação do tubérculo, equilibrando a aplicação de N (Filgueira, 2003). Quando ocorre deficiência de K é possível verificar uma redução significativa na produtividade e maior formação de raízes pequenas e irregulares (Barreira, 1986; Mendonça, 1991).

Em relação ao cálcio (Ca), ele é essencial na formação da parede celular e no transporte de auxinas (Tariq e Mott, 2007). Quando ocorre em excesso, pode levar à deficiência de outros nutrientes, como o magnésio, P e micronutrientes catiônicos, isso devido à competição entre essas substâncias (Barker e Pilbeam, 2015). Esse desequilíbrio nutricional pode impactar negativamente a produtividade.

Já o magnésio (Mg) possui função na fotossíntese, pois seu átomo encontra-se localizado no centro de cada molécula de clorofila, pigmento responsável pela absorção da luz solar pelas plantas. Ademais, o Mg desempenha participação relevante em diversas etapas fundamentais da síntese de açúcares e proteínas, bem como na produção de trifosfato de adenosina (Barker e Pilbeam, 2015).

A batata é uma cultura sensível à deficiência de Mg, o que pode levar à redução na produção de tubérculos e afetar a qualidade dos mesmos, resultando em menor teor de amido e maior teor de açúcares. Já o excesso pode ser prejudicial, causando desequilíbrio de nutrientes, levando à deficiência de outros nutrientes importantes (Barker e Pilbeam, 2015).

O enxofre (S) exerce importante papel no funcionamento de vários metabólitos celulares, incluindo a metionina, cisteína e cistina (Koprivova e Kopriva, 2016). Em comparação a outras culturas olerícolas, como as espécies da família Brassicaceae, a batata apresenta demanda comparativamente baixa por S (Barczak e Nowak, 2015). Contudo, o uso continuado de fertilizantes com alta concentração N, P e K negligência a reposição de S, resultando em baixos teores do nutriente no solo. As plantas com deficiência de S apresentam clorose, semelhante à deficiência de N, afetando principalmente a porção mais jovem da planta. Embora a falta de S não afete diretamente a clorofila, ela inibe a síntese das membranas dos tilacoides, podendo resultar em um déficit de clorofila (Hawkesford *et al.*, 2012).

O boro (B) está diretamente associado ao crescimento e desenvolvimento celular, além de compor diversos compostos presentes na parede celular (Hawkesford *et al.*, 2012). A deficiência desse elemento pode resultar em diversos sintomas, incluindo a formação de tubérculos pequenos com fendas nas cascas (Hayyawawi *et al.*, 2020). Esses sintomas são um indicativo da importância do B para a qualidade e produtividade das culturas.

Na planta, o manganês (Mn) é importante para a estrutura de proteínas e enzimas fotossintéticas (Hawkesford *et al.*, 2012). A presença de Mn é crucial para o sistema de defesa da planta. Portanto, manter os teores adequados desse nutriente pode ajudar na prevenção e manejo de doenças. As concentrações excessivas podem ser prejudiciais ao aparelho fotossintético, uma vez que interferem em vários processos como a atividade enzimática, absorção, translocação e utilização de outros elementos minerais (P, Ca, Mg e Fe), resultando em estresse oxidativo (Millaleo *et al.*, 2010).

As quantidades de nutrientes extraídos e exportados pela cultura da batata, geralmente, variam conforme o nível de produtividade alcançado. Embora, não exista relação direta entre esses fatores, é possível observar divergências na eficiência de utilização dos nutrientes, influenciada por diversos elementos, tais como: clima, tipo de solo, teor de água no solo, sanidade da planta, cultivar utilizada e espaçamento (Fernandes e Soratto, 2015). Os teores de nutrientes considerados adequados para a produção da batata se encontram na Tabela 1.

Tabela 1. Faixa de suficiência das concentrações dos nutrientes na folha da batata.

	Padrão 1 ¹	Padrão 2 ²	Padrão 3 ³
Macronutrientes	g kg ⁻¹		
N	40 - 50	45 - 60	30
P	2,5 - 5	2,9 - 5	3,5
K	40 - 65	93 - 115	50
Ca	10 - 20	7,6 - 10	20
Mg	3 - 5	1 - 1,2	7,5
S	2,5 - 5	-	3,5
Micronutrientes	mg kg ⁻¹		
B	25 - 50	25 - 50	40 - 50
Cu	7 - 20	7 - 20	5 - 8
Fe	50 - 100	50 - 100	800 - 1000
Mn	30 - 250	30 - 250	-
Zn	20 - 60	45 - 250	-

¹ amostragem da terceira folha a partir do tufo apical, aos 30 dias (Lorenzi *et al.*, 1997);

² amostragem da folha mais desenvolvida, na amontoa (Martinez *et al.*, 1999); ³ amostragem do pecíolo da quarta folha a partir da ponta, entre 35 e 45 dias (Malvolta *et al.*, 1997).

A elevada amplitude das faixas de suficiência da cultura da batata se deve à gama de cultivares, locais de plantio, fotoperíodo e outros fatores que podem interferir na necessidade nutricional da planta. Nesse sentido, é importante a utilização de técnicas que considerem as especificações genéticas e ambientais da batata, estreitando as faixas de suficiência e garantindo uma melhor produção.

1.3 REFERÊNCIAS

- Abba (2020). Previsão Consumo de Batata. **Associação Brasileira de Batata** 58:1-90.
- Abbas F, Afzaal H, Farooque AA, Tang S (2020) Crop Yield Prediction through Proximal Sensing and *Machine Learning* Algorithms. **Agronomy**, 10(7), 1046. <https://doi.org/10.3390/agronomy10071046>
- Aitchison J (1992) The statistical analysis of compositional data. **Journal of the Statistical Sociaty**. United Kingdom 44(2):139-177.
- Amaral TN, Costa RA, de Resende Andrade PC (2023). Análise de dados em experimentos envolvendo processamento de batatas: mapeamento sistemático. **Sigmae** 12(1):149-157.
- Barker VA, Pilbeam JD (2015). **Handbook of Plant Nutrition**. Boca Raton: CRC Press p. 773.
- Barreira P (1986) **Batata-doce: uma das doze mais importantes culturas do mundo**. São Paulo: Ícone. 91p.
- Bengio Y (2000) Gradient-Based Optimization of Hyperparameters. *Neural Computation*, 12(8), 1889-1900. <https://doi.org/10.1162/089976600300015187>
- Bisognin DA, Streck NA (2009) Desenvolvimento e manejo das plantas para alta produtividade e qualidade da batata. Associação Brasileira da Batata. Itapetininga/SP. **Publicação Técnica ABBA**, 30 p.
- Brown CR (2005) Antioxidants in potato. **American Journal of Potato Research**. 82, 163–172. <https://doi.org/10.1007/BF02853654>
- Calliope SR, Lobo MO, Sammán NC (2018). Biodiversity of Andean potatoes: Morphological, nutritional and functional characterization. **Food chemistry** 238:42-50.
- Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ (2018) Nextgeneration *Machine Learning* for biological networks. **Cell**, v. 173, n. 7, p. 15811592.
- CEPEA - Centro de Estudos Avançados em Economia Aplicada. PIB do Agronegócio. Piracicaba: CEPEA/ESALQ/USP. Disponível em: <<https://www.cepea.esalq.usp.br/br/categoria/batata-cepea.aspx>> Acesso em: 30 mar. 2022.
- Chlingaryan A, Sukkarieh S, Whelan B (2018) *Machine Learning* approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. **Computers and electronics in agriculture**, 2018, 151:61-69.
- Dias JRM, Tucci CAF, Wadt PGS, Partelli FL, Perez DV, Espindula MC, Tomio DB (2013). Antecipação do período de diagnose foliar em laranja 'pêra' no Amazonas. **Pesquisa Agropecuária Brasileira** 48(7):757-764.
- Dias JRM, Wadt PGS, Partelli FL, Espindula MC, Perez DV, Souza FR, Bergamin AC, Delarmelinda EA (2017) Normal nutrient ranges and nutritional monitoring of 'Pêra'

- orange trees based on the CND method in different fruiting stages. **Pesquisa Agropecuária Brasileira** 52(9):645-654. <https://doi.org/10.1590/S0100-204X2017000900010>.
- Divya KL, Mhatre PH, Venkatasalam EP (2021) Crop Simulation Models as Decision-Supporting Tools for Sustainable Potato Production: a Review. **Potato Research**. 64, 387–419. <https://doi.org/10.1007/s11540-020-09483-9>
- Drummond ST, Sudduth KA, Joshi A, Birrell SJ, Kitchen NR (2013) Statistical and neural methods for site specific yield prediction, **American Society of Agricultural Engineers**, v. 46.
- Egozcue JJ, Pawlowsky-Glahn V (2005). Groups of parts and their balances in compositional data analysis. **Mathematical Geology**. 37:795-828.
- Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003). Isometric log-ratio transformations for compositional data analysis. **Mathematical Geology** 35:279-300.
- Embrapa Hortaliças. A cultura da batata. Sistemas de Produção, 8. ISSN 1678-880X Versão Eletrônica. Disponível em: <https://www.embrapa.br/hortalicas/batata/clima>. Acesso em: 16-03-2023.
- Erpen L, Streck NA, Uhlmann LO, Freitas CPO, Andriolo JL (2013) Tuberização e produtividade de batata-doce em função de datas de plantio em clima subtropical. **Bragantia**, 72(4), 369-376. <https://doi.org/10.1590/brag.2013.050>
- Esteves E, Locatelli G, Bou NA, Ferrarezi RS (2021) Sap Analysis: A Powerful Tool for Monitoring Plant Nutrition. **Horticulturae** 7:426. <https://doi.org/10.3390/horticulturae7110426>
- Fagundes JD, de Paula GM, Lago I, Streck, NA, Bisognin, DA (2010) Aquecimento global: efeitos no crescimento, no desenvolvimento e na produtividade de batata. **Ciência Rural**, 40(6), 1376-1382. <https://doi.org/10.1590/S0103-84782010005000091>
- FAO (2023). Food and agriculture data. Acessado em 10 mai. 2023, de <http://faostat3.fao.org/home/e>
- FAOSTAT (2019). Statistical database. Acessado em 9 mai. 2023, de <http://faostat3.fao.org/home/e>
- Feng S, Cao Y, Xu T, Yu F, Zhao D, Zhang G (2021) Rice Leaf Blast Classification Method Based on Fused Features and One-Dimensional Deep Convolutional Neural Network. Remote Sensing, 13(16), 3207. <https://doi.org/10.3390/rs13163207>
- Fernandes AM, Soratto RP, Evangelista, RM (2015) Qualidade de tubérculos de batata da cultivar 'atlantic' afetada pela adubação fosfatada. **Ciência Rural**, v. 45, p. 1401-1407.
- Fernandes AM, Soratto RP, Silva BL (2011) Extração e exportação de nutrientes em cultivares de batata: I - Macronutrientes. **Revista Brasileira de Ciência do Solo**, 35, 2039-2056.

- Fernandes FM (2021b) **Produtividade de tubérculos e acúmulo, partição e exportação de nutrientes em duas cultivares de batata sob influência do manejo do nitrogênio**. 150 f. Tese (Doutorado em Agronomia - Agricultura) – Unesp, Botucatu.
- Fernandes FM, Soratto RP, Fernandes AM, Souza EFC (2021a) Chlorophyll meter-based leaf nitrogen status to manage nitrogen in tropical potato production. **Agronomy Journal** 113:1733-1746. doi.org/10.1002/AGJ2.20589
- Filgueira FAR (2003) Novo Manual de Olericultura: Agrotecnologia moderna na produção e comercialização de hortaliças. 2. ed. Viçosa: UFV, 412p.
- Filgueira FAR (2008) **Novo manual de olericultura: Agrotecnologia moderna na produção e comercialização de hortaliças**. Viçosa: UFV, 402p.
- Fortin JG, Ancil F, Parent LÉ, *et al.* (2011) Site-specific early season potato yield forecast by neural network in Eastern Canada. **Precision Agriculture**, 12, 905-923. DOI: 10.1007/s11119-011-9233-6.
- Galdino MP. (2015) **Valores de referência e faixas de suficiência nutricional para avaliação do estado nutricional da cultura do eucalipto no Brasil**. Tese (Doutorado em Solos e Nutrição de Plantas), Universidade Federal de Viçosa, Viçosa.
- Ganeshamurthy AN, Govindakrishnan P, Raghupathi HB, Mahendra Kumar MB (2019) Compositional Nutrient Diagnosis (CND) Norms and Indices for Potato (*Solanum tuberosum* L.). **Journal of Horticultural Sciences**, 14(2), 142-148.
- Geiklooi A, Reyhanitabar A, Najafi N (2021). Critical indexes of Compositional nutrient diagnosis (CND) and its validation in wheat Fields. **Revista de la Facultad de Agronomía** 38(3):480-504.
- Gholami, V., & Booij, M. J. (2022). Use of *Machine Learning* and geographical information system to predict nitrate concentration in an unconfined aquifer in Iran. **Journal of Cleaner Production**, 360, 131847. <https://doi.org/10.1016/j.jclepro.2022.131847>
- Gómez D, Salvador P, Sanz J, Casanova JL (2021) New spectral indicator Potato Productivity Index based on Sentinel-2 data to improve potato yield prediction: a *Machine Learning* approach. *International Journal of Remote Sensing*, 42(9), 3426-3444. DOI: 10.1080/01431161.2020.1871102
- Gopal J (1993) Flowering behaviour, male sterility and berry setting in tetraploid *Solanum tuberosum* germplasm. **Euphytica**, 72, 133-142.
- Harsányi E, Bashir B, Arshad S, Ocwa A, Vad A, Alsalman A, Bácskai I, Rátonyi T, Hijazi O, Széles A, Mohammed S (2023) Data Mining and *Machine Learning* Algorithms for Optimizing Maize Yield Forecasting in Central Europe. **Agronomy**, 13(5), 1297. DOI: 10.3390/agronomy13051297
- Hassanzadeh A, Zhang F, van Aardt J, Murphy SP, Pethybridge SJ (2021) Broadacre Crop Yield Estimation Using Imaging Spectroscopy from Unmanned Aerial

- Systems (UAS): A Field-Based Case Study with Snap Bean. **Remote Sensing**, 13(16), 3241. DOI: 10.3390/rs13163241
- Haverkort AJ, Struik PC (2015) Yield levels of potato crops: Recent achievements and future prospects. **Field Crops Research**, 182, 76-85. DOI: 10.1016/j.fcr.2015.06.002
- Hayyawi WAA, Estabraq HOA (2020) Effect of Urea and Nano-Nitrogen Fertigation and Foliar Application of Nano-Boron and Molybdenum on some Growth and Yield Parameters of Potato. **Al-Qadisiyah Journal For Agriculture Sciences**, 10, 253-263. DOI: 10.33794/qjas.2020.167074
- Heldwein, A. B., Maldaner, I. C., Bosco, L. C., Trentin, G., Grimm, E. L., Radons, S. Z, Lucas DDP (2012) Saldo de radiação diurno em dosséis de batata como função da radiação solar global. **Revista Ciência Agronômica**, 43(1), 86-93. DOI: 10.1590/S1806-66902012000100012
- Holland DA (1966) The interpretation of leaf analysis. J. **Horticultural Science**, 41, 311-329.
- Holzkämper A (2017) Adapting Agricultural Production Systems to Climate Change—What's the Use of Models? **Agriculture**, 7(10), 86. DOI: 10.3390/agriculture7100086
- Houghland GVC (1960) The influence of phosphorus on the growth and physiology of the potato plant. **American Potato Journal**, 37, 127-138.
- IBGE: Levantamento Sistemático da Produção Agrícola. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/periodicos/2415/epag_2021_dez.pdf Acesso em: 30 mar. 2022.
- Ingh G, Singh A, Kaur G (2021) Role of artificial intelligence and the internet of things in agriculture. In: Kaur G, Tomar P, Tanque M (Eds.) **Artificial intelligence to solve pervasive internet of things issues**. Cambridge: Academic Press, p. 317-330.
- Janiesch C, Zschech P, Heinrich K (2021) *Machine Learning* and deep learning. **Electron Markets** 31, 685–695. DOI: 10.1007/s12525-021-00475-2
- Javaid M, Haleem A, Khan IH, Suman R (2023) Understanding the potential applications of Artificial Intelligence in Agriculture Sector. **Advanced Agrochem**, 2(1), 15-30. DOI: 10.1016/j.aac.2022.10.001
- Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, Keating BA (2017) Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. **Agricultural systems**, 155, 269-288. DOI: 10.1016/j.agsy.2016.09.021
- Khaki S, Wang L (2019) Crop Yield Prediction Using Deep Neural Networks. **Frontiers in Plant Science**, 10, 621. DOI: 10.3389/fpls.2019.00621
- Khiari L, Parent L-É, Tremblay N (2001) The Phosphorus Compositional Nutrient Diagnosis Range for Potato. **Agronomy Journal**, 93(4), 815-819. DOI: 10.2134/agronj2001.934815x

- Koch M, Naumann M, Pawelzik E, Gransee A, Thiel H (2020) The Importance of Nutrient Management for Potato Production Part I: Plant Nutrition and Yield. **Potato Research**. 63, 97-111. DOI: 10.1007/s11540-019-09431-2
- Koprivova A, Kopriva S (2016) Hormonal control of sulfate uptake and assimilation. *Plant Molecular Biology*, 91, 617-627. DOI: 10.1007/S11103-016-0438-YLiu J., G. C. E.; Tian, L. A neural network for setting target corn yields. **ASAE**, v. 44, 2013.
- Lopes CA, Silva GO, Cruz EM, Assad ED, Pereira AS (2011) Uma análise do efeito do aquecimento global na produção de batata no Brasil. **Horticultura Brasileira**, 29(1), 7-12. DOI: 10.1590/S0102-05362011000100002
- Lorenzi JO, Monteiro PA, Miranda Filho HS, Raij, B van (1997) Raízes e tubérculos. In: Raij, B. van.; Cantarella, H.; Quaggio, J.A.; Furlani, A.M.C. (Eds.) **Recomendações de adubação e calagem para o Estado de São Paulo**. Campinas: Instituto Agronômico de Campinas, p.221-229. (Boletim Técnico, 100).
- Malavolta E, Vitti GC, Oliveira SA (1997) **Avaliação do estado nutricional das plantas: princípios e aplicações**. 2. ed. Piracicaba: Potafos, 319p.
- Marcomini, GR (2021). Vantagens competitivas na produção de batata no Brasil e Estados Unidos. **Administração de Empresas em Revista**, 4(22):246-270.
- Martinez HEP, Carvalho JG, Souza RB (1999) Diagnose Foliar. In: Ribeiro AC, Guimarães PTG, Alvarez VVH (Eds.) **Recomendações para o uso de corretivos e fertilizantes em Minas Gerais**. 5ª Aproximação. Viçosa: UFV, p. 143-168.
- Mellaku MT, Sebsibe AS (2022) Potential of mathematical model-based decision making to promote sustainable performance of agriculture in developing countries: A review article. **Heliyon**, 8(2), e08968. doi: <https://doi.org/10.1016/j.heliyon.2022.e08968>
- Mendonça ATC; peixoto N. 1991. Efeitos do espaçamento e de níveis de adubação em cultivares de batata-doce. **Horticultura Brasileira**, 9: 80-82.
- Millaleo R, Reyes-Díaz M, Ivanov AG, Mora ML, Alberdi M (2010) Manganese as essential and toxic element for plants: transport, accumulation and resistance mechanisms. *Journal of Soil Science and Plant Nutrition* 10:470-481. DOI: 10.4067/S0718-95162010000200008
- Mishra S, Mishra D, Santra GH (2016) Applications of *Machine Learning* techniques in agricultural crop production: A review paper. **Indian Journal of Science and Technology**, 9(38):1-9. DOI: 10.17485/ijst/2016/v9i38/95032
- Modesto VC, Parent SEE, Natale W, Parent LE (2014) Foliar Nutrient balance standards for Maize (*Zea mays* L.) at high-yield level. **American Journal of Plant Sciences** 5:497-507.
- Moreti MP, Oliveira T, Sartori R, Caetano W (2021) Inteligência artificial no agronegócio e os desafios para a proteção da propriedade intelectual. **Cadernos de Prospecção**, 14(1):60-60.

- Navarre DA, Brown CR, Sathuvalli VR (2019) Potato Vitamins, Minerals and Phytonutrients from a Plant Biology Perspective. **American Journal of Potato Research**, 96:111-126. DOI: 10.1007/s12230-018-09703-6
- Nomura ES *et al.* (2008) Crescimento de mudas micropropagadas da bananeira cv. Nanicão em diferentes substratos e fontes de fertilizante. **Acta Scientiarum Agronomy**, 30(3):359-363.
- Nowaki RHD, Parent S-É, Cecílio Filho AB, Rozane DE, Meneses NB, dos Santos da Silva JA, Natale W, Parent LE (2017) Phosphorus Over-Fertilization and Nutrient Misbalance of Irrigated Tomato Crops in Brazil. **Frontiers in Plant Science** 8:825. DOI: 10.3389/fpls.2017.00825
- Ochieng J, Kirimi L, Mathenge M (2016) Effects of climate variability and change on agricultural production: The case of small scale farmers in Kenya. NJAS – Wageningen. **Journal of Life Sciences**, 77:71-78. DOI: 10.1016/j.njas.2016.03.005
- Orlandini S, Nejedlik P, Eitzinger J, Alexandrov V, Toullos L, Calanca P, Trnka M, Olesen JE (2008) Impacts of Climate Change and Variability on European Agriculture: Results of Inventory Analysis in COST 734 Countries. **Annals of the New York Academy of Sciences**, 1146(1):338-353. DOI: 10.1196/annals.1446.013
- Ortiz DA (2020) O alimento supernutritivo que se adaptou a todas as culturas e 'mudou o mundo'. BBC News. Acesso em 10 de Maio de 2023 <<https://www.bbc.com/portuguese/vert-tra-52905548>>.
- Pahalvi HN, Rafiya L, Rashid S, Nisar B, Kamili AN (2021) Chemical Fertilizers and Their Impact on Soil Health. In: Dar GH, Bhat RA, Mehmood MA, Hakeem KR (eds) **Microbiota and Biofertilizers**, Vol 2.
- Parent LE (2011) Diagnosis of the nutrient compositional space of fruit crops. **Revista Brasileira de Fruticultura** 33:321–334.
- Parent LE, Cambouris AN, Muhawenimana A (1994) Multivariate Diagnosis of Nutrient Imbalance in Potato Crops. **Soil Science Society of America Journal** 58(5):1432. DOI: 10.2136/sssaj1994.03615995005800050022x
- Parent LE, Dafir M (1992) A Theoretical Concept of Compositional Nutrient Diagnosis. **Journal of the American Society for Horticultural Science** 117(2):239-242. DOI: 10.21273/JASHS.117.2.239
- Parent LE, Karam A, Visser AS (1993) Compositional nutrient diagnosis of the greenhouse tomato. **Hortscience** 28(10):1041-1042.
- Parent LE, Natale W (2008) CND: vantagens e benefícios para culturas de alta produtividade. In: Prado RM, Rozane DE, Vale DW, Correia MAR, Souza HA (Eds.). **Nutrição de plantas: diagnose foliar em grandes culturas**. Jaboticabal: Universidade Estadual Paulista, p. 105-114.

- Parent SE, Parente L, Rozane DE, Natale W (2013) Plant ionome diagnosis using sound balances: case study with mango (*Mangifera Indica*). **Frontiers in Plant Science** 4:01-12.
- Pereira AS, Silva GO, Castro CM (2016) Melhoramento de batata. In: Nick C, Borém A (Org.). **Melhoramento de hortaliças**. Viçosa: Editora UFV p.128-157.
- Petropoulos SA, Sampaio SL, Di Gioia F, Tzortzakis N, Rouphael Y, Kyriacou MC, Ferreira I (2019) Grown to be blue - Antioxidant properties and health effects of colored vegetables. **Part I: Root vegetables. Antioxidants** 8(12):617.
- Phillips S (2014) Indústria de fertilizantes e agricultura de precisão: apoio à segurança alimentar mundial. **Informações Agrônômicas** 146, ISSN 2311-5904.
- Queiroz DM, Valente DSM, Pinto F de A de C, Borém A (Eds.) (2022) **Agricultura digital**. Oficina de Textos. 224p.
- Ragassi CF, Favarin JL, Shiraishi FA, Moita AW, Sako H, de Melo PCT (2009) Efeito da descompactação profunda de solo na produção da cultura da batata [Effect of deep soil compaction alleviation on the production of potato crop]. **Horticultura Brasileira** 27(4). <https://doi.org/10.1590/S0102-05362009000400014>
- René PJJ, Rietra MH, Chistian OD, Prem SB (2017) Effects of Nutrient Antagonism and Synergism on Yield and Fertilizer Use Efficiency. **Communications in Soil Science and Plant Analysis** 48(16):1895-1920.
- Ribeiro WR, Rodrigues RR, Pizetta SC, Gonçalves MS, Gonçalves D da C, Sales RA, Martins CA da S, dos Reis EF (2021) Fator crítico de disponibilidade hídrica para o crescimento inicial do cafeeiro Conilon. **Agrarian**, 14(51), 71–81. <https://doi.org/10.30612/agrarian.v14i51.11391>
- Romero JR, Roncallo PF, Akkiraju PC, Ponzoni I, Echenique VC, Carballido JA (2016) Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. **Comput Electron Agric**, 96, 173–179.
- Rozane DE, Natale W, Prado RM, Barbosa JC (2009) Tamanho da amostra foliar para avaliação do estado nutricional de goiabeiras com e sem irrigação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 13, 233–239.
- Rozane DE, Natale W, Prado RM, Barbosa JCJC (2007) Amostragem para diagnose do estado nutricional de mangueiras. **Revista Brasileira de Fruticultura**, 29(2).
- Rozane DE, Parent LE, Natale W (2016) Evolution of the predictive criteria for the tropical fruit tree nutritional status. **Revista Científica Jaboticabal**, 44(1), 102–112. doi: <http://dx.doi.org/10.15361/1984-5529.2016v44n1p102-112>.
- Salvador CA, Pereira JR (2021) Prognóstico agropecuário de Olericultura 2021/2022. Governo do Estado do Paraná, 113(37).
- Sarkar B, Mondal SS (2005) Effect of integrated nutrient management on the growth and productivity of potato (*Solanum tuberosum*). **Environment and Ecology**, Special Issue 3, 387–391.

- Sawicka B, Mohammed A, Umachandran K (2018) Food safety of potato processed in the aspect of acrylamide risk. **Food Processing & Technology**, 6(1), 96–102.
- Silva GGC, Neves JCL, Alvarez VH, Leite FP (2004) Nutritional diagnosis for eucalypt by DRIS, M-DRIS, and CND. **Scientia Agricola**, Piracicaba, 61(5), 507–515.
- Silva GGC, Neves JCL, Alvarez VH, Leite FP (2005) Avaliação da universalidade das normas DRIS, M-DRIS E CND. **Revista Brasileira de Ciência do Solo**, 29, 755-761.
- Silva GGC, Neves JCL, Alvarez VH, Leite FP (2005) Avaliação da universalidade das normas DRIS, M-DRIS E CND. **Revista Brasileira de Ciência do Solo**, v.29, p. 755-761.
- Silva GO, Lopes CA (2015) Origem e Botânica. In: **Embrapa hortaliças Sistema de Produção da Batata**. Brasília: EMBRAPA, 4–6. Acesso em 13 mai. 2023, de <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/132923/1/Sistema-de-Producao-da-Batata.pdf>.
- Silva GP, Prado RM, Wadt PGS, Moda LR, Caione G (2020) Accuracy of nutritional diagnostics for phosphorus considering five standards by the method of diagnosing nutritional composition in sugarcane. **Journal of Plant Nutrition**, 1-13.
- Storm H, Baylis K, Heckelee T (2019) *Machine Learning* in agricultural and applied economics. **European Review of Agricultural Economics**, 47(3), 849–892. <https://doi.org/10.1093/erae/jbz033>
- Struik PC (2007) Responses of the Potato Plant to Temperature. In *Potato Biology and Biotechnology: Advances and Perspectives*, 367-393. <https://doi.org/10.1016/B978-044451018-1/50060-9>
- Svoikin F, Zhuk K, Svoikin V, Ugryumov S, Bacherikov I, Veas Iniesta D, Ryapukhin A (2023) Classification of Tree Species in the Process of Timber-Harvesting Operations Using Machine-Learning Methods. **Inventions**, 8(2), 57. <https://doi.org/10.3390/inventions8020057>
- Tariq M, Mott CJB (2007) Effect of Boron on the Behavior of Nutrients in Soil-Plant Systems-A Review. **Asian Journal of Plant Sciences**, 6, 195-202. doi.org/10.3923/ajps.2007.195.202
- Tito R, Vasconcelos HL, Feeley KJ (2018) Global climate change increases risk of crop yield losses and food insecurity in the tropical Andes. **Glob Change Biol**, 24, e592-e602. <https://doi.org/10.1111/gcb.13959>
- Trani PE, Raij B (1997) HORTALIÇAS. In: Raij B *et al.* (Eds.). **Recomendação de adubação e calagem para o estado de São Paulo**, 54-57. Campinas, SP: IAC.
- Wadt PGS (1996) **Os métodos da Chance Matemática e do Sistema Integrado de Diagnóstico e Recomendação (DRIS) na avaliação nutricional de plantios de eucalipto**. (Tese de Doutorado) – Universidade Federal de Viçosa, Viçosa.
- Wei HE, Grafton M, Bretherton M, Irwin M, Sandoval E (2021) Evaluation of Point Hyperspectral Reflectance and Multivariate Regression Models for Grapevine

Water Status Estimation. **Remote Sensing**, 13(16), 3198.
<https://doi.org/10.3390/rs13163198>

Yamaç, S. S., & Todorovic, M. (2020). Estimation of daily potato crop evapotranspiration using three different *Machine Learning* algorithms and four scenarios of available meteorological data. **Agricultural Water Management**, 228, 105875. doi:10.1016/j.agwat.2019.105875.

Yamane DR, Parent S-É, Natale W, Cecílio Filho AB, Rozane DE, Nowaki RHD, de Mattos Junior D, Parent LE (2022) Site-Specific Nutrient Diagnosis of Orange Groves. **Horticulturae**, 8(12), 1126. <https://doi.org/10.3390/horticulturae8121126>

Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, Xiong L, Yan J (2020) Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. **Molecular Plant**, 13(2), 187-214. doi: <https://doi.org/10.1016/j.molp.2020.01.008>

CAPÍTULO 2 – PADRÕES NUTRICIONAIS PARA A BATATEIRA PELO MÉTODO CND

RESUMO – A avaliação do estado nutricional da planta objetiva aperfeiçoar a fertilização, maximizar a produtividade da cultura e a qualidade do produto, além de ser importante para um menor impacto ambiental. Entre os métodos, destaca-se o CND, pois ele considera as relações entre todos os nutrientes, o que possibilita melhor análise da condição nutricional da planta. Os objetivos deste estudo foram: a) obter as normas dos métodos CND-*clr* e CND-*ilr* para avaliação do estado nutricional da batateira; b) identificar de amostras equilibradas e não-equilibradas nutricionalmente, em situações de alta e baixa produtividade pelo método CND-*ilr*; c) obter as faixas de teores adequados dos nutrientes para a cultura da batata, com alta produtividade; e d) conhecer a ordem dos nutrientes limitantes nas amostras classificadas como nutricionalmente desequilibradas e com baixa produtividade. O banco de dados foi composto por 752 amostras, que relacionaram os teores foliares de nutrientes e produtividade da cultura da batata, de áreas localizadas no estado de São Paulo, Brasil. A produtividade que separou os grupos de alta e baixa produtividade foi 48.993,24 kg ha⁻¹. Das amostras, 29, 2, 540 e 16 foram classificadas como: alta produtividade e nutricionalmente equilibradas, alta produtividade e nutricionalmente desequilibradas; baixa produtividade e nutricionalmente desequilibradas; e baixa produtividade e nutricionalmente equilibradas, respectivamente. Foram obtidas faixas de suficiência dos nutrientes, em geral, mais estreitas que as reportadas na literatura. A ordem de limitação dos nutrientes foi: N, B, S, Ca e Mn. Os parâmetros encontrados no CND-*ilr* para acurácia, sensibilidade, especificidade, NPV e PPV foram 96,9; 97,1; 93,6; 64,4 e 99,6 %, respectivamente.

Palavras-chave: CND batata, análise multivariada, nutrição mineral da batata, balanço nutricional.

NUTRITIONAL STANDARDS FOR POTATOES BY THE CND METHOD

ABSTRACT – The evaluation of the nutritional status of the plant aims to improve fertilization, maximize crop productivity and product quality, in addition to being important for less environmental impact. Among the methods, the CND stands out, as it considers the relationships between all the nutrients, which allows for a better analysis of the condition of the plant. The objectives of this study were: a) to obtain the norms of the CND-*clr* and CND-*ilr* methods for evaluating the nutritional status of the potato plant; b) identification of nutritionally balanced or non-balanced samples, in situations of high and low productivity by the CND-*ilr* method, c) obtaining the ranges of adequate levels of nutrients for the potato crop with high productivity, d) knowing the order of limiting nutrients in samples classified as nutritionally unbalanced and with low productivity. The database consisted of 752 samples that related leaf nutrient contents and yields of potato crops from areas located in the state of São Paulo, Brazil. The productivity that separated the high and low productivity groups was 48,993.24 kg ha⁻¹. Of the samples, 29, 2, 540 and 16 were classified as high yielding and nutritionally balanced, high yielding and nutritionally unbalanced, low yielding and nutritionally unbalanced, and low yielding and nutritionally balanced, respectively. Nutrient sufficiency ranges were generally narrower than those reported in the literature. The order of nutrient limitation was: N, B, S, Ca and Mn. The parameters found in the CND-*ilr* for accuracy, sensitivity, specificity, NPV and PPV were 96.9; 97.1; 93.6; 64.4 and 99.6%, respectively.

Keywords: CND potato, multivariate analysis, potato mineral nutrition, nutritional balance.

2.1 INTRODUÇÃO

A batata é uma das hortaliças mais importantes do Brasil, sendo a segunda mais cultivada no país, (IBGE, 2021). Estima-se que sua área de produção seja de 118,1 mil hectares, com uma produção de 3,6 milhões de toneladas (ABBA, 2020). Nos últimos dez anos, houve um aumento significativo de 28,1% na produtividade da batata no país, resultado de melhorias nas técnicas de cultivo e no uso de variedades mais produtivas (FAO, 2023).

Diversos fatores podem ser responsáveis pela produtividade e rentabilidade da cultura da batata, com destaque para fertilização, que contribui com 19% do custo de produção (CEPEA, 2021). Assim, quando a fertilização é realizada com base em conhecimentos de fertilidade e na avaliação do estado nutricional da planta, esta afeta positivamente a produtividade e a qualidade dos tubérculos, refletindo na rentabilidade do cultivo (Fernandes *et al.*, 2021).

O conhecimento das necessidades nutricionais das plantas e da fertilidade do solo é essencial para garantir o bom desempenho das culturas. Uma das formas de adquirir esse conhecimento é através do monitoramento do estado nutricional, que pode ser realizado por meio da análise da folha diagnose em estádios específicos da cultura. Essa metodologia permite identificar os nutrientes que podem estar limitando o crescimento e a produção das culturas, seja por falta ou excesso desses nutrientes (Faquin, 2002).

Além disso, o monitoramento nutricional também é importante para avaliar o manejo da fertilização adotado e se ele foi adequado e consegue identificar a necessidade de ajustes no manejo nutricional, visando alcançar maiores produtividades. Dessa forma, é possível garantir uma nutrição equilibrada das plantas e maximizar o potencial produtivo das culturas (Esteves *et al.*, 2021).

Desse modo, são diversos os procedimentos e métodos empregados para interpretar os resultados obtidos por meio da análise foliar. Entre eles, o método conhecido como Diagnose da Composição Nutricional (CND) se destaca. Proposto originalmente por Parent e Dafir (1992), este método representa uma abordagem mais recente, que busca aprimorar e melhorar os diagnósticos realizados por meio de métodos convencionais, como o Nível Crítico (NC) e o Sistema Integrado de Diagnose e Recomendação (DRIS). Comumente duas

abordagens no CND são mais utilizadas, a relação log centralizada (*clr*), com base dos estudos em análise de dados composicionais (Aitchison, 1986) e a relação log isométrica (*ilr*) (Egozcue *et al.*, 2003; Egozcue e Pawlowsky-Glahn, 2005).

Com base na teoria da Análise de Dados Composicionais de Aitchison (1982), o método CND reconhece a existência de uma estrutura de dependência intrínseca entre os nutrientes presentes nas plantas. Nesse contexto, a ocorrência de uma mudança, em determinadas proporções de nutrientes, em um tecido vegetal pode provocar alterações nas demais proporções de nutrientes, visto que a composição das plantas está restrita a 100% do teor da matéria seca (PARENT *et al.*, 1993).

Devido a sua abordagem, que considera a interdependência entre todos os nutrientes, o método CND representa uma expansão multivariada dos métodos univariados (como o NC) e bivariados (como o DRIS). Nesse sentido, sua aplicação na avaliação do estado nutricional de povoamentos florestais pode resultar em diagnósticos mais precisos, como afirmado por Holland (1966), visto que o aumento do número de nutrientes analisados, simultaneamente, tende a aumentar a consistência do diagnóstico, especialmente se abranger todos os nutrientes. É válido ressaltar, a interação entre os nutrientes é um aspecto fundamental no diagnóstico nutricional de culturas, uma vez que a disponibilidade de um determinado nutriente pode afetar a absorção e a utilização de outros.

O monitoramento nutricional em plantios de batata, por meio da análise foliar, tem recebido cada vez mais atenção por parte das empresas agrícolas nos últimos anos. Isso se deve ao fato de que essas análises, quando interpretadas corretamente, fornecem informações complementares sobre a fertilidade do solo e a necessidade da planta, o que ajuda a melhorar os programas de fertilização (Koch *et al.*, 2020).

Ao longo dos anos, vários estudos têm sido realizados utilizando diferentes metodologias para interpretação da análise foliar da batateira, tais como Lorenzi *et al.* (1997), Malvolta *et al.* (1997) e Martinez *et al.* (1999) e poucos são os estudos que realizaram a diagnose nutricional da batateira pelo CND, Parent *et al.* (1994), Khiari *et al.* (2001), Ganeshamurthy *et al.* (2019). Aliás, nenhum dos trabalhos citados, envolvendo o CND, foi desenvolvido no Brasil.

Ademais, a maioria dos estudos de diagnóstico nutricional presente na literatura, independentemente da metodologia adotada, se restringiu apenas a avaliação exclusiva dos macronutrientes (N, P, K, Ca, Mg, e, ocasionalmente, S) em seus diagnósticos, negligenciando a análise de micronutrientes (B, Fe, Cu, Mn e Zn), os quais são raramente incluídos nessas avaliações.

Atualmente, devido ao avanço no desenvolvimento das novas cultivares de batata, aqueles materiais genéticos com elevadas taxas de produtividade de tubérculos têm sido prioritariamente selecionados para os plantios. Entretanto, essas novas cultivares apresentam uma maior exigência nutricional, demandando não apenas os macronutrientes, mas também os micronutrientes.

É importante ressaltar, que as doses elevadas de nutrientes nem sempre resultam em ganhos proporcionais de produtividade. Isso ocorre porque os nutrientes interagem entre si, podendo haver sinergismo ou antagonismo entre eles (René *et al.*, 2017). Dessa forma, o excesso de um nutriente pode afetar a disponibilidade de outro, causando um desequilíbrio nutricional, podendo prejudicar a produtividade da batateira. Nesse sentido, a aplicação de metodologias que avaliam os macronutrientes e micronutrientes, como o CND, podem ser úteis para aprimorar a gestão nutricional da cultura.

Com o exposto, os objetivos foram: a) obter as normas dos métodos CND-*clr* e CND-*ilr* para avaliação do estado nutricional na cultura da batata e identificação de amostras equilibradas ou não nutricionalmente, em situações de alta e baixa produtividade, pelo método CND-*ilr*; b) obter as faixas de teores considerados adequados dos nutrientes para a cultura da batata; e c) conhecer a ordem dos nutrientes limitantes nas amostras classificadas como nutricionalmente desequilibradas e com baixa produtividade.

2.2 MATERIAL E MÉTODOS

2.2.1 Composição e caracterização do banco de dados

O banco de dados foi composto por amostras contendo teores foliares de nutrientes, correspondentes a produtividade da batateira, provenientes de 752 áreas (unidades experimentais), obtidas em 21 experimentos, realizados entre 2011 a 2019, nos municípios: Avaré, Bernardino de Campos, Botucatu,

Cerqueira César, Itaí, Jaboticabal, São Manuel e Taquarituba, localizados no estado de São Paulo, Brasil (Figura 1).

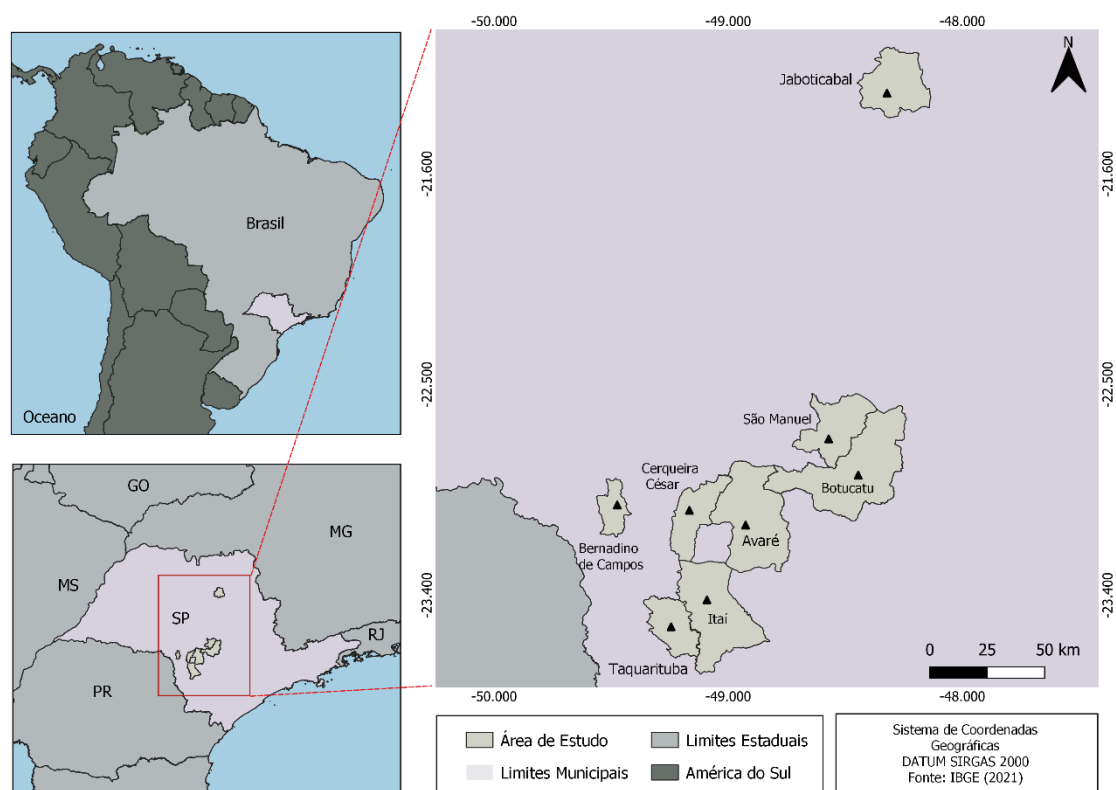


Figura 1. Mapa de localização dos municípios em que foram realizados os experimentos.

Os experimentos (Tabela 1) tiveram como objetivo avaliar a resposta das cultivares de batateira as doses de nitrogênio, fósforo e potássio, aplicados em solos de baixa, média e alta fertilidade, de acordo com classificação de Trani *et al.* (1997).

Tabela 1. Caracterização dos experimentos que compõem o banco de dados.

Id. ¹	Municípios	Cultivares	Nutrientes	Ciclos
1	São Manuel	Agata	N	09-05 a 20-08-2011
2	São Manuel	Agata	N	09-05 a 20-08-2011
3	Avaré	Agata	N	09-05 a 20-08-2011
4	São Manuel	Agata	N	09-05 a 20-08-2011
5	São Manuel	Agata	N	09-05 a 20-08-2011
6	Avaré	Agata	N	09-05 a 20-08-2011
7	Botucatu	Agata	N	13-05 a 20-08-2016
8	Botucatu	Agata	N	06-08 a 06-11-2016
9	Botucatu	Agata	N	27-07 a 04-11-2017
10	Botucatu	Electra	N	13-05 a 20-08-2016
11	Botucatu	Electra	N	06-08 a 06-11-2016
12	Botucatu	Electra	N	27-07 a 04-11-2017
13	Avaré	Agata, Asterix, Atlantic, Markies e Mondial	P	28-04 a 18-08-2011
14	Itaí	Agata, Asterix, Atlantic, Markies e Mondial	P	20-04 a 23-07-2011
15	Cerqueira	Agata, Asterix, Atlantic, Markies e Mondial	P	20-05 a 25-08-2011
16	Taquarituba	Agata e Mondial	P	17-06 a 24-09-2012
17	Taquarituba	Agata	K	15-06 a 22-09-2012
18	Bernardino	Agata	K	18-04 a 20-07-2013
19	Botucatu	Agata	K	12-07 a 13-11-2013
20	Jaboticabal	Asterix	N	16-04 a 11-07-2019
21	Jaboticabal	Agata	N	16-04 a 04-07-2019

¹ Id: Identificação dos experimentos: 1 a 6 – Souza (2014); 7 a 12 – Fernandes (2021b); 13 a 16 – Fernandes (2013); 17 a 19 – Job (2014); 20 e 21 – Nascimento (2021).

Os solos das áreas experimentais corresponderam a Latossolo Vermelho distrófico, Latossolo Vermelho distrófico, Latossolo Vermelho distrófico típico e Latossolo Vermelho eutrófico típico. A classificação climática dos experimentos de 1 a 19 é do tipo Cfa (Köppen-Geiger, 1948; Alvares et al., 2013), clima temperado quente (mesotérmico) úmido, e a temperatura média do mês mais quente é superior a 22 °C (Vianello; Alves, 2000). Já os experimentos 20 e 21 apresentam clima do tipo Aw (Köppen-Geiger, 1948), tropical, com seca no inverno (Lombardi Neto; Drugowich, 1994). A irrigação foi realizada por pivô central ou aspersão do tipo convencional e o manejo fitossanitário foi realizado segundo recomendações técnicas para a cultura (EMBRAPA, 2022). As cultivares de batata (Tabela 1) são as principais utilizadas no estado de São Paulo, sendo elas: ‘Agata’ possui alta produtividade e é a cultivar mais plantada no Brasil, ocupando aproximadamente 50% da área (Ribeiro et al., 2017); ‘Asterix’ é a cultivar mais utilizada na indústria de batata palito pré-frita e congelada (Pereira et al., 2020); ‘Atlantic’ responde pela maior demanda do mercado nacional de batata tipo *chips* (Pereira et al., 2020); ‘Electra’ é indicada para o mercado *in natura*; ‘Markies’ é indicada para cozimento e fritura (Gonçalves et al., 2020); e a ‘Mondial’ apresenta alta produtividade com aptidão para mercado *in natura*.

Diagnose foliar e produtividade

Em cada unidade experimental, foi realizada a amostragem das folhas diagnósticas, sendo coletado aos 30 dias a terceira folha, a partir do tufo apical de trinta plantas, como preconiza Lorenzi *et al.* (1997). O material amostrado foi lavado, seco em estufa de ventilação forçada a 65°C, até massa constante e moído (< 2,0 mm) para realização das determinações dos teores dos nutrientes (N, P, K, Ca, Mg, S, B, Cu, Fe, Mn e Zn), segundo metodologia proposta por Malavolta *et al.* (1997). A produtividade de tubérculos, em cada unidade experimental, foi obtida e estimada a produtividade em kg ha⁻¹.

2.2.2 Obtenção das normas pelo método CND-clr

Através da Diagnose da Composição Nutricional, em relação log centralizada (CND-clr), as amostras do banco de dados foram ordenadas de modo decrescente quanto à produtividade. Na elaboração das normas, os teores de todos os nutrientes foram expressos em mg kg⁻¹ e a produtividade em kg ha⁻¹. Antes de realizar os cálculos CND, foram empregadas técnicas de detecção de valores discrepantes em cada variável por meio da identificação de *outliers*.

A remoção dos *outliers*, a partir da variação interquartil (IQR), foi realizada para cada nutriente, no qual foi feito gráfico de *boxplot*, utilizado para exibir a distribuição para cada nutriente. O *boxplot* foi dividido em quartis e definido o limite inferior e superior. O IQR foi a diferença entre os quartis Q1 e Q3, sendo assim, todos os valores que ultrapassaram os limites foram considerados *outliers*, conforme proposto por Rousseeuw e Croux (1992), utilizando a distância de Mahalanobis (Parent *et al.*, 2009).

Como a normalidade amostral é uma condição exigida para a realização de muitas inferências válidas a respeito de parâmetros populacionais, os dados foram submetidos ao teste de normalidade Shapiro-Wilk, como recomendam Hair *et al.* (2009). Assim, foram calculados o fator de enchimento (*Fv*) para cada amostra, deduzindo de 1.000.000 o somatório das concentrações dos nutrientes. Calculou-se a média geométrica (*g*) das concentrações dos nutrientes de cada amostra do banco de dados (Aitchison, 1986), conforme:

$$g = (N \times P \times \dots \times Zn \times Fv)^{\frac{1}{(d+1)}} \quad (1)$$

em que, d é quantidade de nutrientes avaliados.

A representação *clr* de dados de composição – variáveis multinutrientes (Vn) – foi calculada da seguinte forma (Khiari *et al.*, 2001):

$$Vn = \ln \frac{x_j}{g(x)} \quad (2)$$

em que, x_j é a concentração de cada nutriente e g é a média geométrica da respectiva amostra, ambas em mg kg^{-1} .

A soma dos valores das razões logarítmicas centradas nas variáveis nutrientes (Vn), incluindo o Fv (valor de enchimento), devem ser iguais a zero (Khiari *et al.*, 2001):

$$\ln_N + \ln_P + \ln_K + \dots + \ln_{Fv} = 0 \quad (3)$$

Foram calculadas as médias dos $\ln$ de cada nutriente. Para cada relação entre os nutrientes foi calculada a covariância e a matriz de covariância inversa. Na sequência, para cada amostra, a distância de Mahalanobis (D^2) foi calculada através das variáveis multinutrientes (Vn) da seguinte forma:

$$D^2 = \sqrt{(Vn_j - Vn_j^*)^T COV^{-1}(Vn_j - Vn_j^*)} \quad (4)$$

onde Vn_j^* é o baricentro de toda a população, COV é a matriz da covariância de toda a população, ilr_j é a amostra que está sendo calculada.

Para cada D^2 calculou-se a probabilidade (p) bicaudal da distribuição qui-quadrada (x , graus de liberdade), sendo $x = D^2$ e graus de liberdade igual a d (número de nutrientes), se $p \geq 0,05$ a amostra foi classificada em normal. As amostras consideradas *outliers* ($p < 0,05$) foram removidas (Parent *et al.*, 2009).

Foi delineada a organização da produtividade do conjunto amostral em ordem decrescente, foi estabelecida a função cúbica $(-b/3a)$ pela relação entre a produtividade ($t\ ha^{-1}$) do conjunto amostral e a variância acumulada (função cumulativa) dos valores de D^2 . A determinação da derivada segunda, identificando o ponto de inflexão, permitiu a demarcação das subpopulações de baixa e alta produtividade, conforme descrito por Khiari et al. (2001).

Os índices de balanço (I) de cada nutriente e Fv , de cada amostra, foram calculados pela fórmula:

$$I_x = \frac{Vn - Vn^*}{\sigma_i^*} \quad (5)$$

na qual: I_x é o índice de balanço do nutriente i da amostra x ; Vn é a variável multinutriente i ; Vn^* é a média das variáveis multinutrientes das amostras da população de alta produtividade para o nutriente i e σ_i^* é o desvio-padrão das variáveis multinutrientes das amostras da população de alta produtividade para o nutriente i .

Os índices de balanço dos nutrientes e de Fv foram utilizados para calcular o índice de equilíbrio ($CND - r^2$) de cada amostra:

$$CND - r^2 = I_N^2 + I_P^2 + I_K^2 + [...] + I_{Fv}^2 \quad (6)$$

As normas $CND-clr$ corresponderam a média e ao desvio padrão da população de alta produtividade. A ordem de limitação (OL) foi calculada a partir dos índices $CND-clr$ para cada nutriente das amostras da população de baixa produtividade. A ordem de limitação obedeceu ao critério, mais limitante está em deficiência, sendo a ordem decrescente de limitação do mais negativo para o menos negativo. Depois, entre os limitantes por excesso, tem-se a ordem do mais positivo para o menos positivo.

2.2.3 Obtenção das normas pelo método CND-*ilr*

Para os cálculos do CND-*ilr* (relação log isométrica) foi utilizado o mesmo banco de dados utilizado no CND-*clr*. Com as informações dos teores nutricionais, em mg kg⁻¹, foi realizada a eliminação dos *outliers* e os valores de *Fv* de cada amostra. Com os 12 componentes (D), correspondentes aos nutrientes e *Fv*, foram estabelecidos os 11 balanços (D-1) ou partição binária sequencial (PBS) (Parent *et al.*, 2013a, 2013b).

Para a definição dos balanços ortonormais, utilizou-se a interpretação de dados segundo *Seaborn*, que é uma biblioteca de visualização de dados do programa Python, estabelecendo uma correlação de alto nível entre os nutrientes. Utilizando-se a função *clustermap*, com parâmetros pré-definidos [parâmetros “spearman” (método), distância euclidiana (métrica) e o padrão de escala igual a 1], foi elaborado um *heatmap*, apresentado em forma de dendrograma. A vantagem intrínseca do uso de computadores reside na capacidade de realizar análises sem o viés humano, uma vez que compreendem padrões nos dados e geram recomendações de forma imparcial. Isso contrasta com o método tradicional, conduzido por especialistas na cultura em questão. A singularidade da proposta, portanto, consiste em introduzir uma abordagem inédita que pode contribuir significativamente para o avanço do método CND, oferecendo uma perspectiva inovadora e promissora para futuras pesquisas.

A razão logarítmica isométrica (*ilr*) foi calculada conforme proposto por Egozcue e Pawlowsky-Glahn (2005):

$$ilr_j = \sqrt{\frac{n_j^+ n_j^-}{n_j^+ + n_j^-} \ln \frac{g(c_j^+)}{g(c_j^-)}} \quad (7)$$

onde: n_j^+ é o número de componentes no numerador (r); n_j^- é o número de componentes no denominador (s), determinados a partir da partição binária sequencial; $g(c_j^+)$ é a média geométrica entre os componentes no numerador; e $g(c_j^-)$ é a média geométrica entre os componentes no denominador.

Desse modo, foi calculada a média da população de alta produtividade, para cada ilr calculado na fórmula anterior e, em seguida, foi construída a matriz de covariância da população de alta produtividade. A distância de Mahalanobis (D^2) foi calculada para separar as amostras nutricionalmente equilibradas e desequilibradas (Parent *et al.*, 2009), a partir do espaço Euclidiano (Parent *et al.*, 2013b), como segue:

$$D^2 = \sqrt{(ilr_j - ilr_j^*)^T COV^{-1}(ilr_j - ilr_j^*)} \quad (8)$$

onde ilr_j^* é o baricentro de uma população de referência, COV é a matriz de covariância da população de referência (VN) e ilr_j é a amostra que está sendo calculada.

O necessário procedimento do método Cate-Nelson maximiza a soma dos quadrados entre duas partições da seguinte forma:

$$R^2 = \left[\frac{(\sum_{i=1}^k Y_i)^2}{k} + \frac{(\sum_{j=k}^n Y_j)^2}{(n-k)} \right] - CF \quad (9)$$

em que: Y é a produtividade; k é uma contagem elementar que começa com a primeira observação ordenada acima de n ; e CF é um fator de correção calculado como:

$$\frac{(\sum_{k=1}^n Y_k)^2}{n}$$

Na amostra que obteve o maior valor encontrado para R^2 , observou-se o valor de D^2 e a sua produtividade correspondente (Figura 2).

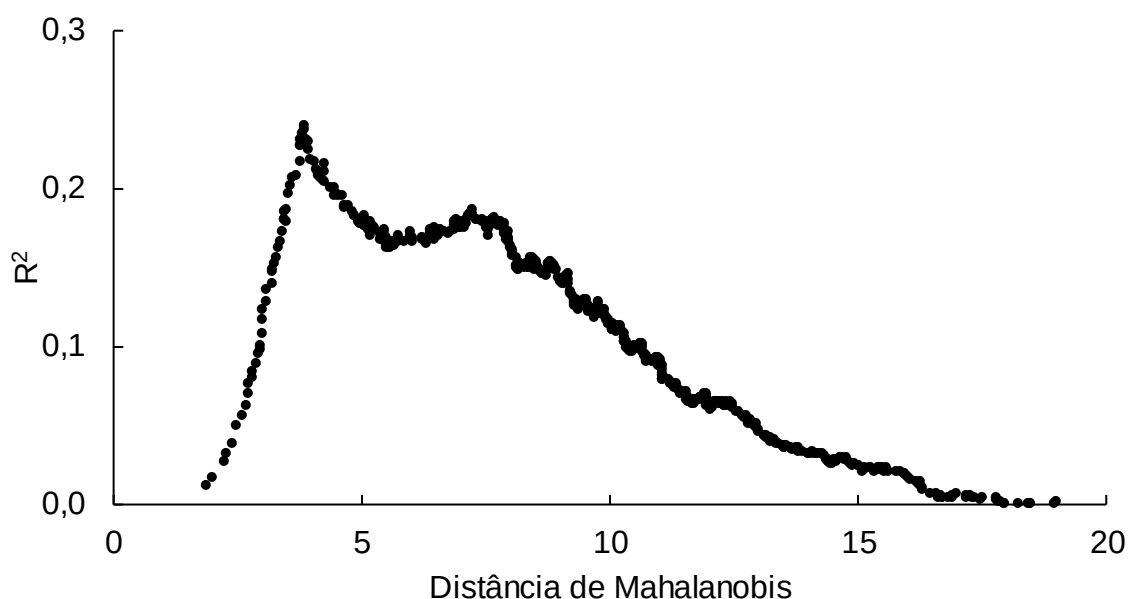


Figura 2. Classificação binária dos dados com índices Distância de Mahalanobis e R^2 .

O valor de D^2 e a produtividade representam os dois pontos de corte para a determinação da partição de Cate-Nelson, das amostras em quatro quadrantes, os quais foram denominados por Parent *et al.* (2013b) em:

a) Verdadeiro Negativo (VN): Amostras com alta produtividade e equilibradas nutricionalmente. Estado nutricional adequado, classificada como “população de referência”.

b) Falso Positivo (FP): Erro tipo I. Amostras com alta produtividade e desequilibradas nutricionalmente. FP representa amostras com “consumo de luxo” de nutrientes ou alta eficiência de uso dos nutrientes.

c) Falso Negativo (FN): Erro tipo II. Amostras com baixa produtividade e equilibradas nutricionalmente. FN corresponde a amostras com influência de outros fatores de produção (por exemplo, clima, pragas e doenças) sobre o desempenho da cultura.

d) Verdadeiro Positivo (VP): Amostras com baixa produtividade e desequilibradas nutricionalmente. Pelo menos um nutriente está causando o desequilíbrio.

Os parâmetros para avaliar as classificações das amostras nos quadrantes foram calculados de acordo com Parent *et al.* (2013b):

a) Acurácia (ACC): é a probabilidade de uma amostra ser corretamente identificada como equilibrada ou desequilibrada, calculado por $(VN + VP) / (VN + FN + VP + FP)$.

b) Sensibilidade (SEN): é a probabilidade de uma amostra de baixa produtividade estar desequilibrada, calculado por $VP / (VP + FN)$.

c) Valor Predito Positivo (PPV): é a probabilidade de uma amostra desequilibrada retornar a um baixo desempenho, calculado por $VP / (VP + FP)$.

d) Especificidade (SPC): é a probabilidade de uma amostra de alta produtividade ser equilibrada, calculado por $VN / (VN + FP)$.

e) Valor Predito Negativo (NPV): é a probabilidade de uma amostra equilibrada retornar a um alto desempenho, calculado por $VN / (VN + FN)$.

As normas CND-*ilr* corresponderam à média e à matriz de multiplicação, que foram calculadas a partir das coordenadas CND-*ilr* da população de referência (VN).

2.3 RESULTADOS E DISCUSSÃO

O banco de dados inicial ($n = 752$, $W = 0,8897$, $p = 0,999^1$) apresentou normalidade. Entretanto, 165 *outliers* foram removidos, sendo 140 *outliers* pelo método IQR e 25 pela probabilidade (p) bicaudal da distribuição qui-quadrada ($p < 0,05$), melhorando a acurácia. A análise foi refeita e os dados, novamente, apresentaram normalidade. A remoção dos *outliers* é uma técnica comumente utilizada nos trabalhos envolvendo o CND.

Essa técnica foi utilizada por Parent *et al.* (2013a), Parent *et al.* (2013b), Rozane *et al.* (2015) e Rozane *et al.* (2020), possibilita melhoria nos parâmetros de avaliação das classificações das amostras e qualidade das normas CND. Com a eliminação de *outliers*, o novo banco de dados apresentou $n = 587$, $W = 0,9206$ e $p = 0,999$.

¹ O valor p é uma medida estatística que auxilia na avaliação se uma diferença observada em um estudo é estatisticamente significativa. Já o W , corresponde ao teste de Shapiro-Wilk, utilizado para avaliar se uma amostra de dados segue uma distribuição normal (Miot, 2017).

2.3.1 CND-*clr*, normas, ordem de nutrientes limitantes e faixa de suficiência

As normas do CND-*clr* (média e desvio padrão) para cada variável multinutriente (V_n) são apresentados na Tabela 2.

Tabela 2. Normas CND-*clr* (média e desvio padrão) das variáveis multinutrientes (V_n) da batateira na população de alta produtividade pelo método para produtividades de 41.666,66 kg ha⁻¹.²

Variável	Normas CND- <i>clr</i> *
N	3,23 ± 0,15
P	0,50 ± 0,39
K	3,24 ± 0,34
Ca	1,56 ± 0,27
Mg	0,92 ± 0,32
S	0,30 ± 0,20
B	-4,34 ± 0,18
Cu	-3,48 ± 1,20
Fe	-1,63 ± 0,41
Mn	-2,81 ± 0,47
Zn	-3,52 ± 0,67

* As normas CND-*clr* são representadas pela média e pelo desvio padrão da população de alta produtividade, respectivamente.

Assim, foi possível encontrar os índices para N, P, K, Ca, Mg, S, B, Cu, Fe, Mn e Zn, com base na concentração média dos teores da população de baixa produtividade (Figura 3), comparada com as normas determinadas para a população de alta produtividade (Tabela 2) e avaliados por meio de padrões CND-*clr*. Desse modo, os resultados mostraram um CND- $r^2 = 3,09$.

As concentrações de N (-0,72), B (-0,65) e S (-0,56) apresentaram deficiência relativa com base nos padrões estabelecidos para o método CND-*clr*, ou seja, seguindo as recomendações, encontraríamos escassez desses elementos em comparação com a utilizando-se a população de alta produtividade. Por outro lado, as concentrações dos nutrientes, Ca (+1,12) e Mn (+0,48), se encontram em excesso relativo (Figura 3). Esse resultado, aponta que as recomendações utilizadas nos experimentos podem induzir o excesso de adubação com esses nutrientes. Os nutrientes, Cu (-0,33), P (+0,29), Zn (+0,31), K (-0,23), Fe (-0,11) e Mg (+0,08) estão dentro da faixa adequada nutricionalmente.

² O valor de 41.666,66 kg ha⁻¹ corresponde ao ponto de corte.

Na determinação da ordem de limitação, foi inicialmente considerada a presença de nutrientes limitantes devido à deficiência e, em seguida, os nutrientes limitantes devido ao excesso.

A ordem de nutrientes limitantes (OL) para a população de baixa produtividade (Figura 3) foi dividida em dois grupos, os limitantes por deficiência e os limitantes por excesso. A OL foi $N > B > S > Ca > Mn$. Os demais nutrientes estão em concentrações adequadas no tecido vegetal. Quanto mais próximo os índices estão de zero, mais equilibrado está a relação dos multinutrientes.

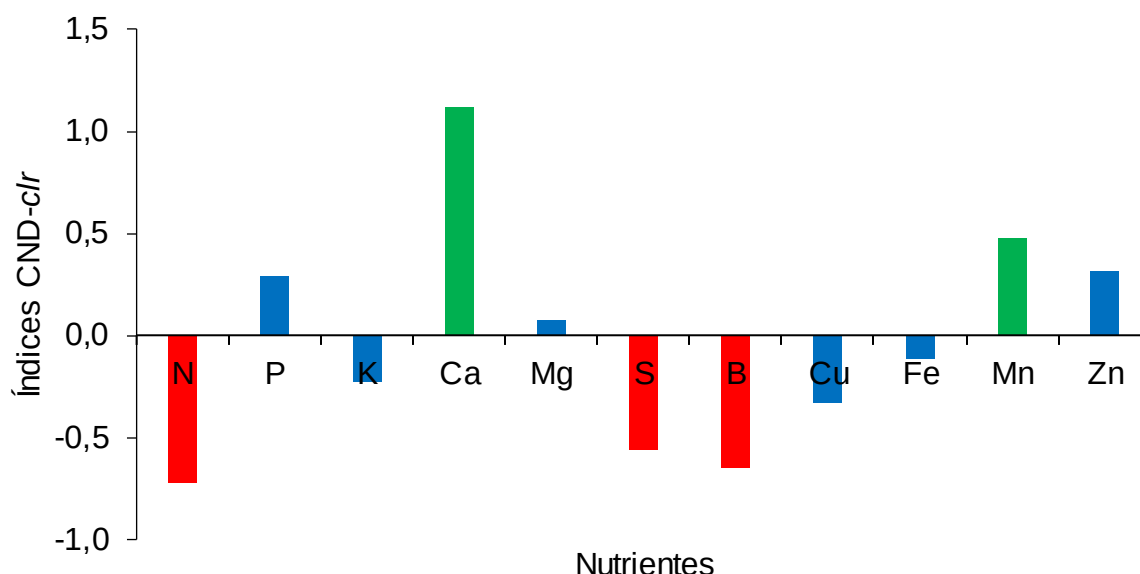


Figura 3. Índices CND-clr dos nutrientes da população de alta produtividade. Barras em vermelho, verde e azul indicam que o nutriente está em deficiência, em excesso e adequado, respectivamente.

O N foi o nutriente mais limitante (Figura 3), pois tem o maior impacto na produtividade da batateira dentre todos os nutrientes (Silva *et al.*, 2013). A sua deficiência na planta causa clorose das folhas mais velhas, diminuição do crescimento e da produção de tubérculos (Rosen *et al.*, 2014). Já o aumento da oferta pode aumentar a proporção de tubérculos de grande porte (Zebarth e Rosen, 2007). A grande quantidade de amostras com o N abaixo do teor adequado, pode ser resultado: da lixiviação de nutrientes, baixa quantidade de reposição de N por fertilização e alta extração do nutriente, resultado da agricultura intensiva (Puga *et al.*, 2019). Segundo Westerman (2005), a batata extrai até 235 kg N ha⁻¹ do solo durante o ciclo.

O B foi o segundo nutriente limitante para a produtividade da batateira por deficiência (Figura 3). Este nutriente está diretamente associado ao crescimento e desenvolvimento celular, além de fazer parte de compostos presentes na parede celular (Hawkesford *et al.*, 2012). Um dos sintomas de deficiência de B, na batata, é a formação de tubérculos pequenos com fendas nas cascas (Hayyawari *et al.*, 2020).

O S foi o terceiro e último nutriente limitante para a batateira por deficiência (Figura 3). Atua no funcionamento de vários metabólitos celulares, como a metionina, cisteína e cistina (Koprivova e Kopriva, 2016). Em comparação as outras culturas olerícolas, como as espécies família Brassicaceae, a batata possui demanda comparativamente baixa por S (Barczak e Nowak, 2015), mas devido ao uso contínuo de fertilizantes com alta concentração de N, P e K, a reposição de S tem sido negligenciada e baixos teores do nutriente têm sido observadas nos solos.

As plantas que possuem deficiência de S apresentam clorose, semelhantes ao de N, porém, na porção mais jovem da planta. Ainda, que a deficiência de S não cause impacto direto na clorofila, ela inibe a síntese das membranas dos tilacóides e, deste modo, pode provocar deficiência de clorofila (Hawkesford *et al.*, 2012).

O Ca foi o quarto nutriente mais limitante e o primeiro por excesso (Figura 3). O nutriente participa da formação da parede celular e no transporte de auxinas (Tariq e Mott, 2007). Esse excesso pode causar a deficiência de outros nutrientes tais como Mg, P e micronutrientes catiônicos por competição (Barker e Pilbeam, 2015), o que, por consequência, causará desequilíbrio nutricional e impactar negativamente na produtividade. O alto teor de Ca encontrado nas amostras pode estar relacionado com a calagem.

O Mn foi o quinto nutriente que limitou a produtividade da batateira e o segundo por excesso (Figura 3). Nas plantas, ele participa da estrutura de proteínas e enzimas fotossintéticas (Hawkesford *et al.*, 2012). A presença do Mn é vital para o sistema de defesa da planta, portanto, manter o Mn ao nível ideal na planta pode auxiliar na prevenção e no manejo de doenças. A concentração de Mn para a batateira é, aproximadamente, 100 vezes o que a sarna (*Streptomyces scabies*) necessita. Assim, quando planta é fertilizada com o nível adequado de Mn, a concentração se torna letal para a bactéria, podendo reduzir

a incidência do patógeno (Lazarovits *et al.*, 2007). Entretanto, concentrações excessivas de Mn, na planta, parecem ser prejudiciais ao aparelho fotossintético, alterando vários processos, tais como a atividade enzimática, absorção, translocação e utilização de outros elementos minerais (P, Ca, Mg e Fe), causando estresse oxidativo (Millaleo *et al.*, 2010).

As faixas de suficiência (FS) dos nutrientes encontram-se na Tabela 3. As faixas, ou teores críticos, encontrados na literatura são, antigo e obtido com cultivares já substituídas e com um manejo de baixa produtividade. Considerando-se a amostragem realizada, segundo critério de Lorenzi *et al.* (1997), as FS de P, K, Ca S, B e Mn obtidas no presente estudo, apresentam-se mais estreitas, exceto N e Mg, que apresentaram FS acima daqueles propostos por Lorenzi *et al.* (1997), citado como padrão 1 na Tabela 3. A FS de N obtida encontra-se um pouco superior ao citado por Lorenzi *et al.* (1997) e dentro da faixa proposta por Martinez *et al.* (1999). As FS com amplitudes menores também foram obtidas por Urano *et al.* (2007), Serra *et al.* (2010b), Camacho *et al.* (2012) e Kurihara *et al.* (2013).

Para micronutrientes, ora a FS foi mais ampla, ora mais estreita (Tabela 3). Para Cu, Fe, Mn e Zn, as concentrações foram muito altas e, exceto para o Mn, as FS dos três micronutrientes encontram-se muito acima do citado na literatura (Tabela 3). Possivelmente, pulverizações com fungicidas (Lorenzetti *et al.*, 2020), que contêm Cu, Mn e Zn, podem ter contribuído para concentrações tão altas. Na literatura, concentrações acima de 150, 250 e 250 mg kg⁻¹ para Cu, Mn e Zn, respectivamente, são consideradas potencialmente tóxicas à batateira (Farias *et al.*, 2013; Martinez *et al.*, 1999). Contudo, as altas concentrações foliares foram observadas em amostras com alta produtividade (superiores a 41.666,66 kg ha⁻¹).

Os limites inferiores do intervalo de confiança, que estabeleceram as FS, corresponderam às concentrações críticas no tecido foliar, nada próximas às citadas por Malavolta *et al.* (1997), que utilizaram pecíolo como parte da planta ao invés de folha adotada por Lorenzi *et al.* (1997), critério também usado neste estudo. O N, Cu, Fe, Mn e Zn apresentaram concentrações críticas muito superiores aos citados por Malavolta *et al.* (1997), enquanto para os demais nutrientes as concentrações críticas foram menores.

As divergências nos teores de nutrientes entre este estudo e as concentrações críticas mencionadas por Malavolta et al. (1997) são atribuíveis a diversos fatores. O progresso no melhoramento genético e práticas agrícolas intensivas ao longo do tempo pode ter aumentado as exigências nutricionais das plantas, resultando em teores mais elevados de certos nutrientes. A escolha de partes específicas da planta para amostragem, variações nas técnicas de coleta de dados e condições ambientais distintas também influenciam as concentrações de nutrientes nos tecidos foliares.

Além disso, avanços contínuos na metodologia analítica podem ter contribuído para medições mais precisas, impactando as discrepâncias nos resultados. Essa complexidade ressalta a importância de considerar múltiplos fatores ao interpretar as diferenças nos teores de nutrientes em estudos agrícolas.

Tabela 3. Faixa de suficiência das concentrações dos nutrientes na folha da batata.

	CND	Padrão 1 ¹	Padrão 2 ²	Padrão 3 ³
	g kg ⁻¹			
N	49,7 - 50,8	40 - 50	45 - 60	30
P	4,1 - 4,3	2,5 - 5	2,9 - 5	3,5
K	52,6 - 55,2	40 - 65	93 - 115	50
Ca	13,6 - 14,2	10 - 20	7,6 - 10	20
Mg	5,7 - 6,0	3 - 5	1 - 1,2	7,5
S	2,7 - 2,8	2,5 - 5	-	3,5
	mg kg ⁻¹			
B	25,9 - 27,1	25 - 50	25 - 50	40 - 50
Cu	121,7 - 155,9	7 - 20	7 - 20	5 - 8
Fe	433,1 - 464,3	50 - 100	50 - 100	800 - 1000
Mn	205,9 - 235,3	30 - 250	30 - 250	-
Zn	85,3 - 93,7	20 - 60	45 - 250	-

¹ Amostragem da terceira folha a partir do tufo apical, aos 30 dias (Lorenzi *et al.*, 1997); ² amostragem da folha mais desenvolvida, na amontoa (Martinez *et al.*, 1999); ³ amostragem do pecíolo da quarta folha a partir da ponta, entre 35 e 45 dias (Malavolta *et al.*, 1997).

O balanço nutricional multnutrientes (multivariado) possibilita uma melhor percepção do estado nutricional, quando comparado com a análise univariada e bivariada, pois a principal vantagem do método CND estaria na utilização do conceito de balanço nutricional, considerando todos os nutrientes da planta

(Parent *et al.*, 2013a; 2013b). Neste sentido, o diagnóstico leva em consideração a relação existente entre todos os nutrientes, fator que possibilita identificar os desequilíbrios problemas, o que nem sempre ocorre quando se considera um nutriente de forma isolada (Mccray *et al.*, 2016).

O nível crítico (NC) refere-se ao teor de nutrientes na folha que, quando abaixo dele, resulta em redução na produção (Ulrich e Hills, 1967). Essa medida também pode ser representada pelo limite inferior da fertilidade do solo. O NC para o N, encontrado no método CND-*clr*, é maior que o NC dos três padrões apresentados (padrão 1: +24.25%; padrão 2: +10.22%; padrão 3: +65.67%) e demais macronutrientes, ora o NC CND-*clr* foi maior, ora foi menor que os demais padrões. Para os micronutrientes, o NC CND-*clr* foi maior em relação aos outros padrões, com exceção do B para o padrão 3 (Tabela 3).

Embora, haja espaço para melhorias no estabelecimento de padrões nutricionais, já existem dados disponíveis sobre os níveis críticos e faixa de suficiência para as principais culturas no Brasil, conforme relatado por Lorenzi *et al.* (1997), Malavolta *et al.* (1997) e Martinez *et al.* (1999) (Tabela 3).

Os valores apresentados de NC e FS apresentados nessas referências têm sido utilizados como guia básico para interpretação do estado nutricional das culturas. Esses métodos apresentam como desvantagem os nutrientes serem interpretados individualmente, não considerando as interações entre eles, isto é, o equilíbrio nutricional. A nutrição adequada da planta não pode ser considerada apenas pelos teores individuais dos nutrientes, deve estar pautado nas relações existentes entre eles. O CND é um método de interpretação da análise foliar, que considera o equilíbrio nutricional, portanto, apresenta uma grande vantagem em relação ao NC e a FS.

Apesar dos desafios significativos e dos custos associados à implementação desses métodos, o que resultou na falta de valores atualizados para NC e FS provenientes da calibração local (Mostashari *et al.*, 2018), eles são marcados por uma abordagem univariada. Em outras palavras, a interpretação é isolada para cada nutriente, sem considerar o equilíbrio nutricional na planta, o que é considerado uma desvantagem (Mccray *et al.*, 2016).

Esses métodos apresentam uma notável complexidade operacional, contribuindo para a escassez de dados recentes e, adicionalmente, não abordam

as interações entre os nutrientes, destacando a necessidade de uma abordagem mais abrangente na avaliação do estado nutricional das culturas.

2.3.2 CND-*ilr*, normas, partição binária sequencial e partição de Cate-Nelson

Na Figura 4 pode ser observado, no *heatmap* que agrupa hierarquicamente, os nutrientes que mais se correlacionam, segundo o estudo utilizando a ferramenta Seaborn-Python. Com base nestas relações, foram elaborados 11 balanços ou partição binária sequencial (Tabela 4).

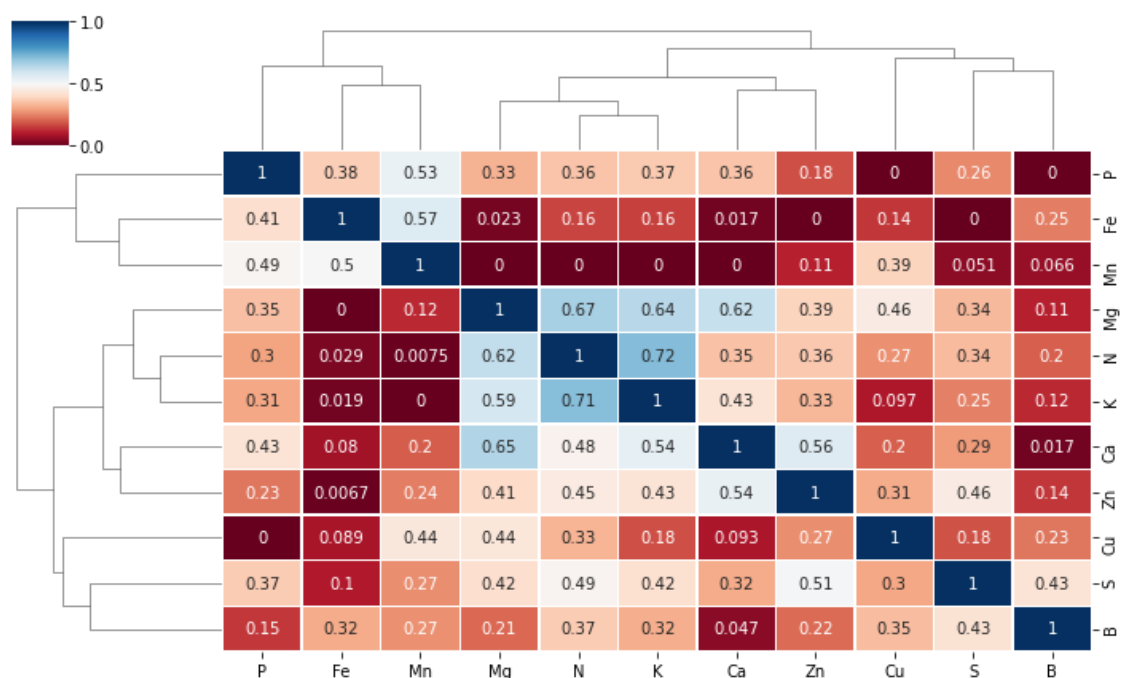


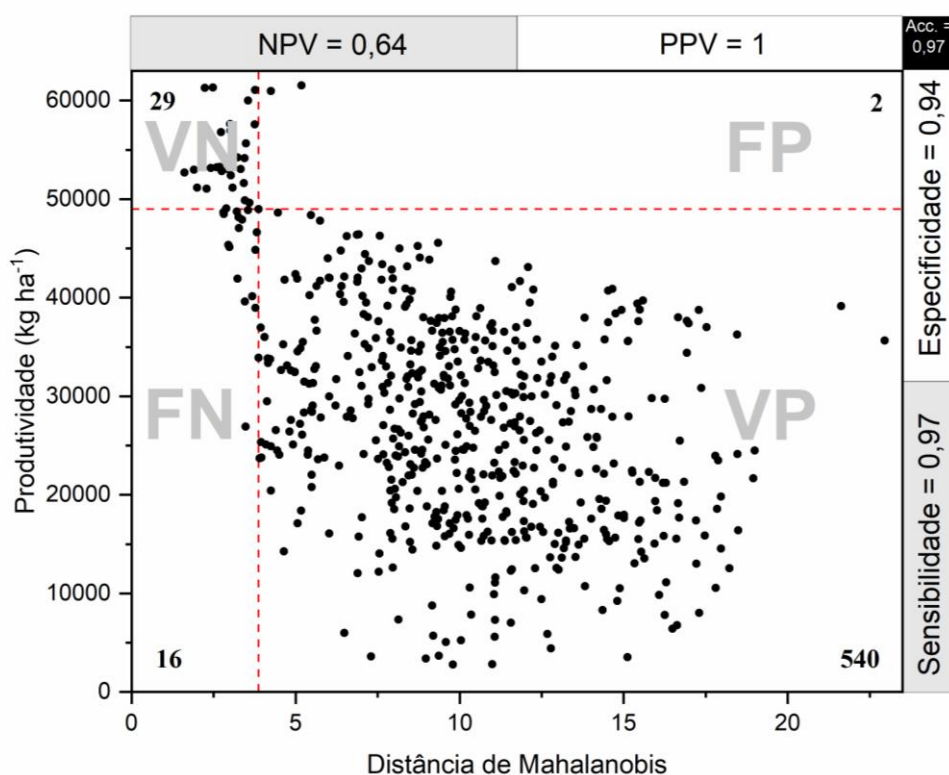
Figura 4. Balanços ortonormais dos teores de nutrientes foliares da batateira utilizando a biblioteca Seaborn do Python.

Tabela 4. Partição binária sequencial (PBS) para os 11 balanços ortonormais.

<i>ilr</i>	Balanços	N	K	P	S	Ca	Mg	B	Cu	Zn	Mn	Fe	<i>Fv</i>	<i>r</i>	<i>s</i>
1	[N K]	-1	1	0	0	0	0	0	0	0	0	0	0	1	1
2	[N, K Mg]	-1	-1	0	0	0	1	0	0	0	0	0	0	1	2
3	[Ca Zn]	0	0	0	0	-1	0	0	0	1	0	0	0	1	1
4	[Fe Mn]	0	0	0	0	0	0	0	0	0	1	-1	0	1	1
5	[Mg, N, K Ca, Zn]	-1	-1	0	0	1	-1	0	0	1	0	0	0	2	3
6	[S B]	0	0	0	-1	0	0	1	0	0	0	0	0	1	1
7	[P Fe, Mn]	0	0	-1	0	0	0	0	0	0	1	1	0	2	1
8	[Cu S, B]	0	0	0	1	0	0	1	-1	0	0	0	0	2	1
9	[Mg, N, K, Ca, Zn Cu, S, B]	-1	-1	0	1	-1	-1	1	1	-1	0	0	0	3	5
10	[P, Fe, Mn Mg, N, K, Ca, Zn, Cu, S, B]	1	1	-1	1	1	1	1	1	1	-1	-1	0	8	3
11	[<i>Fv</i> P, Fe, Mn, Mg, N, K, Ca, Zn, Cu, S, B]	1	1	1	1	1	1	1	1	1	1	1	-1	11	1

Fv: valor de enchimento; *r* e *s*: quantidade de positivos (numerador) e negativos (denominador), respectivamente, da fórmula de cálculo das relações multinutrientes (ilr_j).

As D^2 obtidas para as amostras foram de 1,61 a 22,95. Considerando o maior valor de R^2 das amostras, foi selecionada a amostra cuja D^2 foi 3,87 e produtividade de 48.993,24 kg ha⁻¹ (63.31% acima da produtividade média brasileira, que é de 30 t ha⁻¹). Utilizando-se destas duas informações, foram estabelecidos os quadrantes: verdadeiro negativo (VN), falso positivo (FP), verdadeiro positivo (VP) e falso negativo (FN) (Figura 5).

**Figura 5.** Distribuição das 587 amostras da batateira (estado nutricional e produtividade) nos quadrantes verdadeiro negativo (VN), falso

positivo (FP), falso negativo (FN) e verdadeiro positivo (VP) em função da partição de Cate-Nelson.

A produtividade média obtida pelos bataticultores brasileiros corresponde a 30.000 a 40.000 kg ha⁻¹ (IBGE, 2021). Nessa pesquisa, a produtividade foi de 48.993,24 kg ha⁻¹ correspondendo ao limite inferior da população de alta produtividade, cuja máxima foi de 61.327,78 kg ha⁻¹, ou seja, o grupo VN estão acima da média nacional. A população classificada como desequilibrada nutricionalmente e com baixa produtividade (VP) apresentou amplitude de 45.859,34 kg ha⁻¹, com produtividades mínima e máxima de 2.778,33 e 48.637,67 kg ha⁻¹, respectivamente. Mesmo desequilibradas nutricionalmente, existem 92 produtividades deste grupo que se situam acima da média brasileira, segundo o IBGE (2021).

Do total de amostras, 29 amostras (5%) foram classificadas como VN, ou seja, como população de referência. Destas, 20 e 9 amostras são dos experimentos de N e P, respectivamente. Modesto *et al.* (2014), Parent *et al.* (2013b) e Marchand *et al.* (2013) encontraram 10, 13 e 15% das amostras classificadas como VN, respectivamente. Quanto ao grupo VP, este foi representado por 540 (92%) das amostras, sendo 148, 318 e 74 amostras dos experimentos de N, P e K (Tabela 5).

Tabela 5. Partição das amostras do banco de dados nos quadrantes verdadeiro negativo (VN), falso positivo (FP), falso negativo (FN) e verdadeiro positivo (VP).

Quadrantes	Nº de amostras dos experimentos			Total de amostras	%
	N	P	K		
VN	20	9	0	29	5,0
VP	148	318	74	540	92,0
FP	2	0	0	2	0,3
FN	6	10	0	16	2,7
Total	176	337	74	587	100,0

VN = população de referência = equilibrada nutricionalmente e com alta produtividade, VP = população desequilibrada nutricionalmente e com baixa produtividade, FP = população desequilibrada nutricionalmente e com alta produtividade e FN = população equilibrada nutricionalmente e com baixa produtividade.

Os resultados encontrados para os parâmetros, acurácia, sensibilidade, especificidade, NPV e PPV foram 96,9; 97,1; 93,6; 64,4 e 99,6%, respectivamente. Os altos valores dos parâmetros podem ser atribuídos ao fato do banco de dados ser constituído por amostras oriundas de experimentos, que, pelo rigor experimental, permitem melhor associação entre os teores nutricionais nas folhas e as respectivas produtividades.

Para o banco de dados inicial, sem a eliminação dos *outliers*, foram obtidos 70,6; 70,7; 50,0; 0,45 e 99,8% de acurácia, sensibilidade, especificidade, NPV e PPV, respectivamente. Portanto, com a remoção dos *outliers*, os parâmetros de acurácia, sensibilidade, especificidade e NPV melhoraram em 37, 37, 87 e 14.200%, respectivamente, e somente o parâmetro PPV diminuiu 0,2%. Comparando os balanços de nutrientes entre VN e as classes VP, FN e FP, os testes *t* detectaram diferenças significativas (Tabela 6).

Tabela 6. Resultados do teste *t* comparando as médias de amostras classificadas como verdadeiro negativo (VN) e das classes VP, FN e FP de balanços ortonormais.

<i>l/r</i>	Balanços	VN	VP	FN	FP
1	[N K]	-0,10	0,03**	-0,01 ^{ns}	-0,27 ^{ns}
2	[N, K Mg]	-1,91	-1,80**	-1,88 ^{ns}	-1,86 ^{ns}
3	[Ca Zn]	-3,54	-3,66 ^{ns}	-3,30 ^{ns}	-3,77 ^{ns}
4	[Fe Mn]	-0,71	-0,66 ^{ns}	-0,76 ^{ns}	-0,68 ^{ns}
5	[Mg, N, K Ca, Zn]	-3,78	-3,46**	-3,44*	-3,95 ^{ns}
6	[S B]	-3,29	-3,28 ^{ns}	-3,38*	-3,03 ^{ns}
7	[P Fe, Mn]	-1,99	-2,25*	-2,33 ^{ns}	-1,64 ^{ns}
8	[Cu S, B]	0,56	1,46**	0,96 ^{ns}	0,63 ^{ns}
9	[Mg, N, K, Ca, Zn Cu, S, B]	-4,49	-5,30**	-4,96*	-4,75 ^{ns}
10	[P, Fe, Mn Mg, N, K, Ca, Zn, Cu, S, B]	1,62	1,37**	1,84 ^{ns}	0,89*
11	[Fv P, Fe, Mn, Mg, N, K, Ca, Zn, Cu, S, B]	-6,28	-6,27 ^{ns}	-6,22*	-6,33 ^{ns}

ns = não significativo; * e ** = significativo a 5 e 1 % de probabilidade pelo teste F; VN = população de referência = equilibrada nutricionalmente e com alta produtividade, VP = população desequilibrada nutricionalmente e com baixa produtividade, FP = população desequilibrada nutricionalmente e com alta produtividade e FN = população equilibrada nutricionalmente e com baixa produtividade.

Ao contrastar as classes VN e VP, é possível verificar as diferenças nos balanços 1, 2, 5, 7, 8, 9 e 10 (Tabela 6). Tais resultados indicaram as principais fontes de potenciais desequilíbrios nutricionais que prejudicaram a produtividade da cultura no grupo VP. Assim, ocorreram quatro diferenças nos balanços entre

as classes VN e FN (Tabela 6), ainda que as amostras de ambas as classes sejam equilibradas nutricionalmente.

As variações ocorridas nas concentrações dos nutrientes podem estabelecer diferenças significativas em um determinado balanço, como, por exemplo, nos balanços [Mg, N, K | Ca, Zn] e [S | B]. Entretanto, quando estes fazem parte de outro balanço mais abrangente, considerando mais nutrientes, por exemplo, [P, Fe, Mn | Mg, N, K, Ca, Zn, Cu, S, B], tais teores são compreendidos como em equilíbrio (Tabela 5), conforme reportam Parent *et al.* (2013b).

Nas comparações, entre as classes VN e FP, houve diferença significativa apenas no balanço 10 (Tabela 6), apontando desequilíbrio nutricional, apesar de alta produtividade. Provavelmente, isso se deve a consumo de luxo, ou contaminação da amostra foliar com o nutriente não pertencente ao conteúdo foliar, ou excepcionalmente, a alta eficiência de uso de nutrientes (Parent *et al.*, 2013b). As normas CND-*ilr* corresponderam a matriz de multiplicação (multinutriente) e a média calculada a partir das coordenadas CND-*ilr* da população de referência (VN) (Tabela 7).

Tabela 7. Normas CND-*ilr* (matriz multinutriente e média) calculadas a partir das coordenadas CND-*ilr* da população de referência (VN).

	Ilr1	Ilr2	Ilr3	Ilr4	Ilr5	Ilr6	Ilr7	Ilr8	Ilr9	Ilr10	Ilr11
Ilr1	0,0307										
Ilr2	0,0097	0,0209									
Ilr3	0,0533	0,0371	0,1758								
Ilr4	-0,0121	-0,0026	-0,0280	0,0391							
Ilr5	0,0545	0,0362	0,1537	-0,0126	0,1659						
Ilr6	-0,0120	-0,0085	-0,0425	-0,0010	-0,0373	0,0332					
Ilr7	-0,0782	-0,0524	-0,2000	0,0363	-0,1844	0,0421	0,3075				
Ilr8	0,1078	0,0629	0,2691	-0,0479	0,2572	-0,0707	-0,3835	0,6397			
Ilr9	-0,1058	-0,0714	-0,2590	0,0425	-0,2568	0,0484	0,3757	-0,5579	0,5553		
Ilr10	0,0651	0,0411	0,1493	-0,0122	0,1562	-0,0442	-0,2097	0,2538	-0,2549	0,2095	
Ilr11	0,0084	0,0041	0,0249	-0,0062	0,0252	-0,0032	-0,0258	0,0318	-0,0345	0,0221	0,0059
Média*	-0,0974	-1,9180	-3,5445	-0,7164	-3,7686	-3,2896	-2,0332	0,6681	-4,5701	1,5875	-6,2811

* Média da população de referência (VN) do CND-*ilr*.

Os balanços podem ser representados metaforicamente, a partir de um diagrama móvel, autônomo, com fulcros e pratos de pesagem, onde as

concentrações dos nutrientes impactam diretamente os balanços nutricionais nos fulcros após a mudança (Parent *et al.*, 2013b). A Figura 6 apresenta um dendrograma de equilíbrio derivado de SBP, com valores médios gerais de *ilr* nos fulcros e intervalos de confiança univariados de 0,05 para as classes VN e VP.

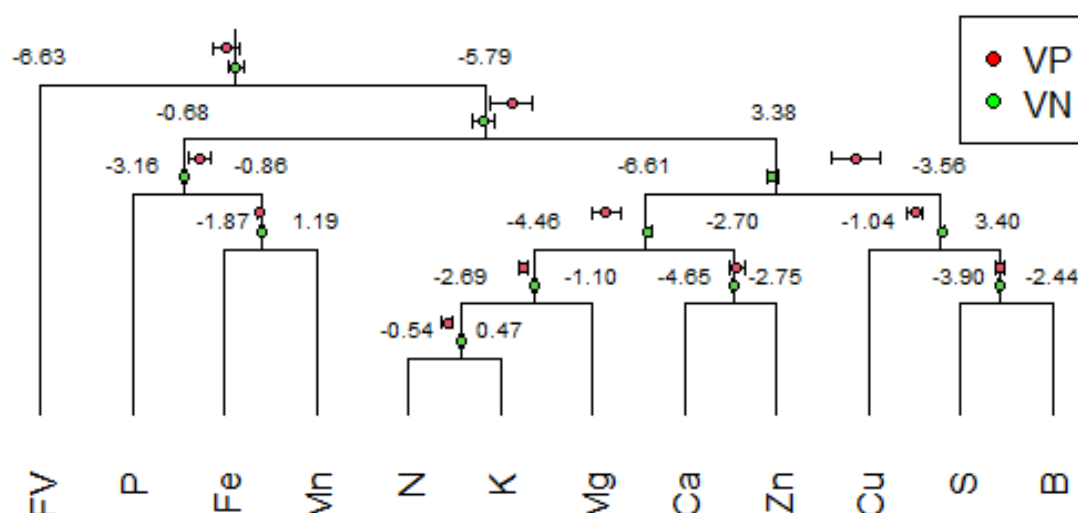


Figura 6. Média móvel e intervalo de confiança dos balaços ortonormais obtidas com amostras das classes de amostras de alta produtividade e equilibradas nutricionalmente (VN) e de baixa produtividade e desequilibradas nutricionalmente (VP).

Os balanços que foram significativos (Figura 6) corroboram com os resultados encontrados na Tabela 6, cujos *ilrs* foram 1, 2, 5, 7, 8, 9 e 10. Entre as classes VN e VP, percebe-se que houve diferença significativa em todos os balanços que N está presente, com exceção do balanço que compara todos os nutrientes com Fv , que não foi significativo (Figura 6). Ainda, verifica-se que os balanços $[N | K]$, $[N, K | Mg]$, $[Mg, N, K, | Ca, Zn]$, $[Mg, N, K, Ca, Zn | Cu, S, B]$ apresentaram médias negativas nas amostras em VP, as quais foram -0,54; -2,69, -4,46 e -6,61, respectivamente.

O N foi o principal responsável por causar desequilíbrio. A deficiência de N de amostras em VP, retratada pelo valor -0,54 no balanço $[N | K]$ e de -2,69 no balanço $[N, K | Mg]$ (Figura 6), encontrada nas amostras da população de VP, possui relação direta com a produtividade da cultura, pois ele é absorvido em quantidades elevadas (Andrews *et al.*, 2013), possui relação com o crescimento

foliar, fotossíntese e produção de tubérculos (Silva *et al.*, 2013; Ahmed *et al.*, 2015). No balanço [Mg, N, K | Ca, Zn] também existem interações competitivas entre os cátions K, Ca e Mg, conforme reportado por Hawkesford (2012).

No balanço [Fe | Mn] não houve diferença entre as amostras de VN e VP. O Fe está diretamente relacionado a gênese do Latossolo Vermelho, das áreas experimentais, disponibilizando o Fe necessário para a planta, o Mn é um dos compostos presentes na maioria dos fungicidas. Entretanto, no balanço [P | Fe, Mn], houve um desequilíbrio causado pelo excesso de P, podendo ser justificado pelas doses muito altas avaliadas nos experimentos, inclusive para solos com alto teor do nutriente.

No balanço [S | B], os índices das amostras de VP não se diferenciaram dos índices das amostras de VN. Provavelmente, signifique que o balanço [S | B] em VP está semelhante ao observado para população de alta produtividade e equilibrada nutricionalmente, ou seja, não parece ser este o problema das amostras VP.

No entanto, o balanço [Cu | S, B] está desequilibrado nutricionalmente, portanto, o fator limitante no segundo balanço foi a deficiência de Cu. A interação negativa do Cu com excesso de P pode reduzir a disponibilidade do Cu para as plantas e como a população apresenta excesso de P (Hopkins *et al.*, 2014) provavelmente esse foi o motivo do baixo teor de Cu no diagnóstico foliar, já que a deficiência de Cu não é comum em solos argilosos.

O desequilíbrio no balanço [Mg, N, K, Ca, Zn | Cu, S, B] foi causado por N e Cu, ambos por deficiência. No balanço [P, Fe, Mn | Mg, N, K, Ca, Zn, Cu, S, B], os principais nutrientes responsáveis pelos desequilíbrios nutricionais foram P do lado esquerdo por excesso, N e Cu do lado direito por deficiência.

Não houve diferenças significativas entre os grupos VN e VP para os seguintes balanços: [Ca | Zn], [Fe | Mn], [S | B] e [*Fv* | P, Fe, Mn, Mg, N, K, Ca, Zn, Cu, S, B]. De fato, tais balanços, provavelmente, não afetaram a produção de batata. Todos os nutrientes avaliados foram confrontados com o valor de enchimento (*Fv*), como acúmulo de nutrientes nos tecidos foliares ou como medida de diluição (Koch *et al.*, 2019).

A partir dos balanços observados (Figura 6), foi constatado que os balanços nutricionais foram mais importantes do que os próprios valores absolutos, pois não houve diferença significativa entre os grupos VN e VP para

o $[Fv \mid P, Fe, Mn, Mg, N, K, Ca, Zn, Cu, S, B]$. Nesse sentido, se deve considerar as interações nutricionais presentes no conceito do equilíbrio de nutrientes limitantes ao crescimento e desenvolvimento vegetal, onde os grupos de nutrientes precisam ser otimamente equilibrados, ao invés de apenas considerar às concentrações individuais dos nutrientes (Parent *et al.*, 2013a).

2.4 CONCLUSÕES

2.4.1 CND-clr

As normas CND-clr foram: N ($3,23 \pm 0,15$), P ($0,50 \pm 0,39$), K ($3,24 \pm 0,34$), Ca ($1,56 \pm 0,27$), Mg ($0,92 \pm 0,32$), S ($0,30 \pm 0,20$), B ($-4,34 \pm 0,18$), Cu ($-3,48 \pm 1,20$), Fe ($-1,63 \pm 0,41$), Mn ($-2,81 \pm 0,47$) e Zn ($-3,52 \pm 0,67$).

As FS foram: N (49,7 - 50,8 kg kg⁻¹), P (4,1 - 4,3 kg kg⁻¹), K (52,6 - 55,2 kg kg⁻¹), Ca (13,6 - 14,2 kg kg⁻¹), Mg (5,7 - 6,0 kg kg⁻¹), S (2,7 - 2,8 kg kg⁻¹), B (25,9 - 27,1 mg kg⁻¹), Cu (121,7 - 155,9 mg kg⁻¹), Fe (433,1 - 464,3 mg kg⁻¹), Mn (205,9 - 235,3 mg kg⁻¹) e Zn (85,3 - 93,7 mg kg⁻¹).

O método CND-clr classificou amostras com produtividade superior a 41.666,66 kg ha⁻¹ como população de alta produtividade. Os nutrientes N, S e B são os principais nutrientes limitantes por deficiência. O Ca e Mn são os nutrientes que estão limitando a produtividade por excesso nas principais áreas de cultivo de batata do estado de São Paulo, Brasil.

2.4.2 CND-llr

As normas CND-llr foram: llr1 (-0,0974), llr2 (-1,9180), llr3 (-3,5445), llr4 (-0,7164), llr5 (-3,7686), llr6 (-3,2896), llr7 (-2,0332), llr8 (0,6681), llr9 (-4,5701), llr10 (1,5875) e llr11 (-6,2811).

Os balaços significativos entre VN e VP formam: (N | K, N, K | Mg), (Mg, N, K | Ca, Zn), (Cu | S, B), (Mg, N, K, Ca, Zn | Cu, S, B).

A Distância de Mahalanobis, que separa as amostras nos grupos equilibrada e desequilibrada nutricionalmente, foi de 3,87 e a produtividade que define os grupos de alta e baixa produtividade foi de 48.993,24 kg ha⁻¹. As amostras, 29, 2, 540 e 16 foram classificadas como alta produtividade e nutricionalmente equilibradas, alta produtividade e nutricionalmente

desequilibradas, baixa produtividade e nutricionalmente desequilibradas, e baixa produtividade e nutricionalmente equilibradas, respectivamente. Os parâmetros encontrados no CND-*ilr* para acurácia, sensibilidade, especificidade, NPV e PPV foram 96,9; 97,1; 93,6; 64,4 e 99,6%, respectivamente.

2.5 REFERÊNCIAS

- Ahmed A, Zaki M, Shafeek M, Helmy Y, El-Baky MA (2015) Integrated use of farmyard manure and inorganic nitrogen fertilizer on growth, yield and quality of potato (*Solanum tuberosum* L.). **International Journal of Current Microbiology and Applied Sciences** 4:325-349. doi.org/10.1080/23311932.2020.1724385
- Aitchison, J. (1986) **The statistical analysis of compositional data**. London: Chapman and Hall, 416p.
- Andre RGB, Garcia A (2015) Alguns aspectos climáticos do município de Jaboticabal, São Paulo. **Nucleus** 12:263-269. doi.org/10.3738/1982.2278.1543
- Andrews M, Raven JA, Lea PJ (2013) Do plants need nitrate? The mechanisms by which nitrogen form affects plants. **Annals of Applied Biology** 163:174-199. doi.org/10.1111/aab.12045
- Anuário brasileiro de Horti & Fruti (2021). Santa Cruz do Sul: Gazeta Santa Cruz do Sul. 104 p. Disponível em: <https://www.editoragazeta.com.br/sitewp/wp-content/uploads/2021/04/HORTIFRUTI_2021.pdf> Acesso em: 30 mar. 2022.
- Barczak B, Nowak K (2015) Effect of sulphur fertilization on the content of macroelements and their ionic ratios in potato tubers. **Journal of Elementology** 20:37-47. doi.org/10.5601/jelem.2014.19.1.471
- Barker VA, Pilbeam JD (Eds.) (2015) Handbook of Plant Nutrition. Boca Raton: CRC Press, 773p. doi.org/10.1201/b18458
- Camacho MA, Silveira MV, Camargo RA, Natale W (2012) Faixas normais de nutrientes pelos métodos ChM, DRIS e CND e nível crítico pelo método de distribuição normal reduzida para laranja-Pera. **Revista Brasileira de Ciência do Solo** 36:193-200. doi.org/10.1590/S0100-06832012000100020
- Cangelosi R, Goriely A (2007) Component retention in principal component analysis with application to cDNA microarray data. **Biology Direct** 2:2-17. doi.org/10.1186/1745-6150-2-2
- CEPEA - Centro de Estudos Avançados em Economia Aplicada. PIB do Agronegócio. Piracicaba: CEPEA/ESALQ/USP. Disponível em: <<https://www.cepea.esalq.usp.br/br/categoria/batata-cepea.aspx>> Acesso em: 30 mar. 2022.
- CNN Brasil. Disponível em: <<https://www.cnnbrasil.com.br/business/produtores-de-sao-paulo-e-minas-sao-os-mais-dependentes-de-fertilizante-russo/>> Acesso em: 30 mar. 2022.
- Cunha AR, Martins D. Classificação climática para os municípios de Botucatu e São Manuel, SP. **Irriga** 14:1-11. doi.org/10.15809/irriga.2009v14n1p1-11
- Cunha MLP, Aquino LA, Novais RF, Clemente JM, Aquino PM, Oliveira TF (2016) **Diagnosis of the nutritional status of garlic crops**. Revista Brasileira de Ciência do Solo 40:1-14. doi.org/10.1590/18069657rbcS20140771

- Egozcue JJ, Pawlowsky-Glahn V (2005) Groups of parts and their balances in compositional data analysis. **Mathematical Geology** 37:795-828. doi.org/10.1007/s11004-005-7381-9
- Esteves, E.; Locatelli, G.; Bou, N.A.; Ferrarezi, R.S. Sap Analysis: A Powerful Tool for Monitoring Plant Nutrition. *Horticulturae* 2021, 7, 426. <https://doi.org/10.3390/horticulturae7110426>.
- Farias JG, Antes FLG, Nunes PAA, Nunes ST, Schaich G, Rossato LV, Miotto A, Girotto E, Tiecher TL, Dressler VL, Nicoloso FT (2013) Effects of excess copper in vineyard soils on the mineral nutrition of potato genotypes. **Food and Energy Security** 2:49-69. doi.org/10.1002/FES3.16
- Fernandes AM (2013) **Adubação fosfatada em cultivares de batata (*Solanum tuberosum* L.)**. 163 f. Tese (Doutorado em Agronomia - Agricultura) – Unesp, Botucatu.
- Fernandes FM (2021b) **Produtividade de tubérculos e acúmulo, partição e exportação de nutrientes em duas cultivares de batata sob influência do manejo do nitrogênio**. 150 f. Tese (Doutorado em Agronomia - Agricultura) – Unesp, Botucatu.
- Fernandes FM, Soratto RP, Fernandes AM, Souza EFC (2021a) Chlorophyll meter-based leaf nitrogen status to manage nitrogen in tropical potato production. **Agronomy Journal** 113:1733-1746. doi.org/10.1002/AGJ2.20589
- Gonçalves DN, Pereira AM, Soares LG, Guimarães MES, Petrucci KPOS, Oliveira LS, Mendonça Neto AB, Finger FL (2020) Post-Harvest Quality of cv. Markies Potatoes Submitted to Mechanical Damages. **Journal of Agricultural Science** 12:124-129. doi.org/10.5539/jas.v12n2p124
- Hair JJF, Black WC, Babin BJ, Anderson RE, Tatham RL (Eds.) (2009) Análise multivariada de dados. Porto Alegre: Bookman, 688p.
- Hawkesford M, Horst W, Kichey T, Lambers H, Schjoerring J, Møller IS, White P. (2012) Functions of macronutrients. In: Marschner H (3ed.) **Mineral Nutrition of Higher Plants**. Amsterdam: Academic Press p. 135-189.
- Hayyawi WAA, Estabraq HOA (2020) Effect of Urea and Nano-Nitrogen Fertigation and Foliar Application of Nano-Boron and Molybdenum on some Growth and Yield Parameters of Potato. **Al-Qadisiyah Journal For Agriculture Sciences** 10:253-263. doi.org/10.33794/qjas.2020.167074
- Hopkins BG, Horneck DA, MacGuidwin AE (2014) Improving phosphorus use efficiency through potato rhizosphere modification and extension. **American Journal of Potato Research** 91:161-174. doi.org/10.1007/s12230-014-9370-3
- IBGE: Levantamento Sistemático da Produção Agrícola. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/2415/epag_2021_dez.pdf> Acesso em: 30 mar. 2022.
- Job ALG (2014) **Doses e parcelamento da adubação potássica na cultura da batata (*Solanum tuberosum* L.)**. 70 f. Dissertação (Mestrado em Agronomia - Agricultura) – Unesp, Botucatu.

- Jones CA (1981) Proposed modifications of the diagnosis and recommendation integrated system (DRIS) for interpreting plant analyses. **Communications in Soil Science and Plant Analysis** 12:785-794. doi.org/10.1080/0010362810936719
- Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. **Psychometrika** 23:187-200. doi.org/10.1007/BF02289233
- Khiari L, Parent LE, Tremblay N (2001) The phosphorus compositional nutrient diagnosis range for potato. **Production Agriculture** 93:815-819. doi.org/10.2134/agronj2001.934815x
- Koch M, Busse M, Naumann M, Jákli B, Smit I, Cakmak I, Hermans C, Pawelzik E (2019) Differential effects of varied potassium and magnesium nutrition on production and partitioning of photoassimilates in potato plants. **Physiologia plantarum** 166:921-935. doi.org/10.1111/ppl.12846
- Koch, M., Naumann, M., Pawelzik, E. et al. The Importance of Nutrient Management for Potato Production Part I: Plant Nutrition and Yield. *Potato Res.* 63, 97–119 (2020). https://doi.org/10.1007/s11540-019-09431-2
- Köppen W, Geiger R (1928) **Klimate der Erde**. Gotha: Verlag Justus Perthes. Wall-map: 150cmx200cm.
- Koprivova A, Kopriva S (2016) Hormonal control of sulfate uptake and assimilation. **Plant Molecular Biology** 91:617-627. doi.org/10.1007/S11103-016-0438-Y
- Kurihara CH, Venegas VHA, Neves JCL, Novais RF, Staut LA (2013) Sufficiency range for nutrient concentration in cotton and soybean leaves, defined through DRIS indexes. **Ceres** 60:412-419. doi.org/10.1590/S0034-737X2013000300015
- Lazarovits G, Hill J, Patterson G, Conn KL, Crump NS (2007) Edaphic soil levels of mineral nutrients, pH, organic matter, and cationic exchange capacity in the geocaulosphere associated with potato common scab. **Phytopathology** 97:1071-1082. doi.org/10.1094/PHYTO-97-9-1071
- Lorenzetti E, Tartaro J, Stangarlin, JR, Kuhn OJ (2020) Agronomic characters and disease management in second season corn using calcium, copper, manganese and zinc products. **Ciência e Agrotecnologia** 44:1-8. doi.org/10.1590/1413-7054202044025019
- Lorenzi JO, Monteiro PA, Miranda Filho HS, Raij, B van (1997) Raízes e tubérculos. In: Raij, B. van.; Cantarella, H.; Quaggio, J.A.; Furlani, A.M.C. (Eds.) **Recomendações de adubação e calagem para o Estado de São Paulo**. Campinas: Instituto Agronômico de Campinas, p.221-229. (Boletim Técnico, 100).
- Malavolta E, Vitti GC, Oliveira SA (1997) Avaliação do estado nutricional das plantas: princípios e aplicações. 2. ed. Piracicaba: Potafos, 319p.
- Marchand S, Parent SE, Deland JP, Parent LE (2013) Nutrient signature of Quebec (Canada) cranberry (*Vaccinium macrocarpon* AIT.). **Revista Brasileira de Fruticultura** 35:199-209. doi.org/10.1590/S0100-29452013000100034

- Martinez HEP, Carvalho JG, Souza RB (1999) Diagnose Foliar. In: Ribeiro AC, Guimarães PTG, Alvarez VVH (Eds.) **Recomendações para o uso de corretivos e fertilizantes em Minas Gerais**. 5ª Aproximação. Viçosa: UFV, p. 143-168.
- Mccray JM, Rice R W, Ezenwa IV, Lang TA, Baucum L (2016) Sugarcane plant nutrient diagnosis. Florida: Sugarcane Handbook. 10 p. Disponível em: <<https://edis.ifas.ufl.edu/pdf%5CSC%5CSC07500.pdf>> Acesso em: 30 mar. 2022.
- Millaleo R, Reyes-Díaz M, Ivanov AG, Mora ML, Alberdi M (2010) Manganese as essential and toxic element for plants: transport, accumulation and resistance mechanisms. **Journal of Soil Science and Plant Nutrition** 10:470-481. doi.org/10.4067/S0718-95162010000200008
- Miot, H. A. (2017) Avaliação da normalidade dos dados em estudos clínicos e experimentais. *Jornal Vascular Brasileiro* [online]. 16(2), pp. 88-91. doi.org/10.1590/1677-5449.041117.
- Miyazawa M, Pavan MA, Muraoka T, Carmo CAFS, Melo WJ (2009) Análise química de tecido vegetal. In: SILVA, F. C. DA (Ed.). **Manual de análises químicas de solos, plantas e fertilizantes**. 2. ed. Brasília: EMBRAPA, p. 191–233.
- Modesto VC, Parent SE, Natale W, Parent LE (2014) Foliar nutrient balance standards for maize (*Zea mays* L.) at high-yield level. **American Journal of Plant Sciences** 5:497-507. doi.org/10.4236/ajps.2014.54064
- Mostashari M, Khosravinejad A, Golmohammadi M (2018) Comparative study of DOP and CND methods for leaf nutritional diagnosis of vitis vinifera in Iran. **Communications in Soil Science and Plant Analysis** 49, 5:576-584. DOI:10.1080/00103624.2018.1432633.
- Nascimento CS (2021) **Yield and quality of potato tubers in response to nitrogen management**. 121 f. Tese (Doutorado em Agronomia – Produção Vegetal) – Unesp, Jaboticabal.
- Parent LE (2011) Diagnosis of the nutrient compositional space of fruit crops. **Revista Brasileira de Fruticultura** 33:321-334. doi.org/10.1590/S0100-29452011000100041
- Parent LE, Dafir M (1992) A theoretical concept of compositional nutrient diagnosis. **Journal of the American Society for Horticultural Science** 117:239-242. doi.org/10.21273/JASHS.117.2.239
- Parent LE, Natale W, Ziadi N (2009) Compositional nutrient diagnosis of corn using the mahalanobis distance as nutrient imbalance index. **Canadian Journal of Soil Science** 89:383-390. doi.org/10.4141/cjss08050
- Parent SE, Parent LE, Egozcue JJ, Rozane DE, Hernandez A, Lapointe L, Hébert-Gentile V, Naess K, Marchand S, Lafond J, Mattos D Jr, Barlow P, Natale W (2013) The plant ionome revisited by the nutrient balance concept **Frontiers in Plant Science** 22:4-39. doi.org/10.3389/fpls.2013.00039

- Parent SE, Parent LE, Rozane DE, Natale W (2013b) Plant ionome diagnosis using sound balances: case study with mango (*Mangifera Indica*). **Frontiers in Plant Science** 4:1-12. doi.org/10.3389/fpls.2013.00449
- Pereira AM, Guimarães MES, Galdino AGS, Gomes MP, Cruz RRP, Ribeiro FCS, Ribeiro WS, Finger FL (2020) Potato cultivars evaluation for processing industry. **Research. Society and Development**. 9:1-10. doi.org/10.33448/RSD-V9I12.10803
- Puga AP, Queiroz MCA, Ligo MAV, Carvalho CS, Pires AMM, Marcatto JOS, Andrade CA (2019) Nitrogen availability and ammonia volatilization in biochar-based fertilizers. **Archives of Agronomy and Soil Science** 66:992-1004. doi.org/10.1080/03650340.2019.1650916
- René P. J. J. Rietra, Marius Heinen, Chistian O. Dimkpa & Prem S. Bindraban (2017) Effects of Nutrient Antagonism and Synergism on Yield and Fertilizer Use Efficiency, *Communications in Soil Science and Plant Analysis*, 48:16, 1895-1920, DOI: 10.1080/00103624.2017.1407429
- Ribeiro GHMR, Samartini CQ, Silva LFL, Vieira SD, Resende LV (2017) Cultivares. In: Nick, C.; Borém, A (Eds.) **Batata do plantio à colheita**. Viçosa: Editora UFV, v.1, p. 77-93.
- Rosen CJ, Kelling KA, Stark JC, Porter GA (2014) Optimizing phosphorus fertilizer management in potato production. **American Journal of Potato Research** 91:145-160. doi.org/10.1007/s12230-014-9371-2
- Rozane DE, Mattos D, Parent SE, Natale W, Parent LE (2015) Meta-analysis in the Selection of Groups in Varieties of Citrus. **Communications in Soil Science and Plant Analysis** 46:1948-1959. doi.org/10.1080/00103624.2015.1069307
- Rozane DE, Paula BV, Melo GWB, Santos EMH, Trentin E, Marchezan C, Silva LOS, Tassinari A, Dotto L, Oliveira FN, Natale W, Baldi E, Toselli M, Brunetto G (2020) Compositional Nutrient Diagnosis (CND) Applied to Grapevines Grown in Subtropical Climate Region. **Horticulturae**. 6:1-13. doi.org/doi.org/10.3390/horticulturae6030056
- Serra AP, Marchetti ME, Vitorino ACT, Novelino JO, Camacho MA (2010) Determinação de faixas normais de nutrientes no algodoeiro pelos métodos ChM, CND e Dris. **Revista Brasileira de Ciência do Solo** 34:105-113. doi.org/10.1590/S0100-06832010000100011
- Silva JG, França MGC, Gomide FTF, Magalhaes JR (2013) Different nitrogen sources affect biomass partitioning and quality of potato production in a hydroponic system. **American Journal of Potato Research** 90:179-185. doi.org/10.1007/s12230-012-9297-5
- Souza EFC (2014) **Fontes e manejo de nitrogênio na cultura da batata em solos arenosos**. 134 f. Tese (Doutorado em Agronomia - Agricultura) – UNESP, Botucatu.
- Tariq M, Mott CJB (2007) Effect of Boron on the Behavior of Nutrients in Soil-Plant Systems-A Review. **Asian Journal of Plant Sciences** 6:195-202. doi.org/10.3923/ajps.2007.195.202

- Trani PE, Raij B van (1997) Hortaliças. In: RAIJ, B van, Cantarella H, Ouaggio JÁ, Furlani AMC. **Recomendação de adubação e calagem para o estado de São Paulo**. Campinas: IAC, 285 p.
- Ulrich A, Hills FJ (1967) Principles and Practices of Plant Analysis. **Soil Testing and Plant Analysis** 2:11-24.
- Urano EOM, Kurihara CH, Maeda S, Vitorino ACT, Gonçalves MC, Marchetti ME (2007) Determinação de teores ótimos de nutrientes em soja pelos métodos chance matemática, sistema integrado de diagnose e recomendação e diagnose da composição nutricional. **Revista Brasileira de Ciência do Solo** 31:63-72. doi.org/10.1590/S0100-06832007000100007
- Wadt PGS, Alvarez VVH, Novais RF, Fonseca S, Barros NF (1998) O método da chance matemática na interpretação de dados de levantamento nutricional de eucalipto. **Revista Brasileira de Ciência do Solo**, 22:773-778. doi.org/10.1590/S0100-06831998000400023
- Westermann DT Nutritional requirements of potatoes (2005) **American Journal of Potato Research** 82:301-307. doi.org/10.1007/BF02871960
- Zebarth BJ, Rosen CJ (2007) Research perspective on nitrogen bmp development for potato. **American Journal of Potato Research** 84:3-18. doi.org/10.1007/BF02986294

CAPÍTULO 3 - PREVISÃO DA PRODUTIVIDADE DA CULTURA DA BATATA A PARTIR DE MODELOS DE *MACHINE LEARNING*

RESUMO – As variabilidades nas condições climáticas afetam várias atividades, incluindo a agricultura e a cultura da batata. A previsão da produtividade da batateira com base em modelos de *Machine Learning*, usando dados climáticos e nutricionais, é uma ferramenta útil para auxiliar os produtores a tomar decisões durante o cultivo, evitando desperdício de recursos e reduzindo o impacto climático. Os objetivos foram: a) prever a produtividade da batateira, b) avaliar o desempenho de diferentes modelos de *Machine Learning* na previsão da produtividade da batateira no estado de São Paulo, Brasil e c) desenvolver uma aplicação on-line que faça a previsão da produtividade de culturas agrícolas. O banco de dados foi composto por 752 amostras que relacionaram os teores foliares de nutrientes e produtividades de culturas da batata, de áreas localizadas no estado de São Paulo, Brasil. Para identificar as variáveis que mais impactam na produtividade da batateira, foram desenvolvidos cinco modelos de *Machine Learning* para análise dos dados. As opções de modelos utilizados foram: Regressão Linear Múltipla (RLM), K-nearest neighbors ou “K-vizinhos mais próximos” (KNN), Support Vector Machine (SVR), eXtreme Gradient Boosting (XGB) e Redes Neurais Artificiais (RN). Para a construção dos modelos, o banco de dados foi dividido em três grupos: variáveis nutricionais, climáticas e top 10 (as principais variáveis que influenciam no modelo pelo método Random Forest). Dentre os 15 modelos, o SVR top 10 obteve os melhores resultados de R^2 ajustado (0.76) e RMSE (5646.80 kg ha⁻¹).

Palavras-chave: agrometeorologia, inteligência artificial, algoritmos na agricultura.

PREDICTING THE PRODUCTIVITY OF POTATO CROPS USING *MACHINE LEARNING* MODELS

ABSTRACT – Changes in climatic conditions affect several activities, including agriculture and potato cultivation. Predicting potato productivity based on *Machine Learning* models, using climate and nutritional data, is a useful tool to help producers make decisions during cultivation, avoiding waste of resources and reducing climate impact. The objectives were: a) to predict potato productivity, b) to evaluate the performance of different *Machine Learning* models in predicting potato productivity in the state of São Paulo, Brazil and c) to develop an online application that predicts productivity of agricultural crops. The database consisted of 752 samples that related leaf nutrient contents and yields of potato crops from areas located in the state of São Paulo, Brazil. To identify the variables that most impact potato productivity, five *Machine Learning* models were developed for data analysis. The model options used were: Multiple Linear Regression (RLM), K-nearest neighbors or “K-nearest neighbors” (KNN), Support Vector Machine (SVR), eXtreme Gradient Boosting (XGB) and Artificial Neural Networks (RN). For the construction of the models, the database was divided into three groups: nutritional, climatic and top 10 variables (the main variables that influence the model by the Ramdon Forest method). Among the 15 models, the top 10 SVR obtained the best adjusted R^2 (0.76) and RMSE (5646.80 kg ha⁻¹) results.

Keywords: agrometeorology, artificial intelligence, algorithms in agriculture.

3.1 INTRODUÇÃO

A cultura da batata é altamente sensível à temperatura e desenvolve-se melhor em clima mais frio. Consequentemente, o aquecimento global tem um efeito substancial na produção de batata em todo o mundo, uma vez que temperaturas moderadamente elevadas podem reduzir drasticamente a quantidade de tubérculos produzidos (Singh *et al.*, 2013).

A compreensão dos efeitos do clima na produção da batateira ressalta a relevância da gestão nutricional adequada, tendo em vista que as variações climáticas na agricultura, muitas vezes, não podem ser completamente controladas, enquanto as questões nutricionais da planta e solo podem ser gerenciadas com sucesso (Raymundo *et al.*, 2018). Tanto o excesso quanto a deficiência de nutrientes podem causar danos severos, perda de produtividade e reduzir a qualidade dos produtos (Pandey *et al.*, 2016).

Ao determinar de forma precisa o estado nutricional da planta evitam-se perdas no processo da produção, como também se promove o uso racional dos fertilizantes agrícolas, um dos princípios da agricultura de precisão (Carletto *et al.*, 2015). Desse modo, a previsão da produtividade da batateira é uma ferramenta para auxiliar os produtores a tomarem decisões rápidas durante o período de cultivo (Piekutowska, 2021), evitando o desperdício de recursos financeiros e reduzindo o impacto ambiental.

Essas previsões têm múltiplas utilidades, incluindo suporte ao comércio agrícola, elaboração de políticas de desenvolvimento e ajuda humanitária relacionada à segurança alimentar. Portanto, a melhora da precisão das previsões de produtividade da batateira pode ser decisiva no enfrentamento do desafio das mudanças climáticas globais, garantia da segurança alimentar e maior sustentabilidade nas áreas agrícolas (Webber *et al.*, 2020; Guo *et al.*, 2021).

Nesse sentido, as ferramentas computacionais para gestão e monitoramento nutricional podem ser disponibilizadas como parte do suporte à decisão e gerenciamento agrícola, inclusive para os agricultores que não têm acesso à consultoria especializada. Entretanto, Webber *et al.* (2020) apontam que os modelos baseados em processos têm limitações intrínsecas relacionadas aos parâmetros de entrada, calibração, procedimentos de avaliação e validação,

bem como aos métodos de simulação de respostas das culturas a diversos fatores ambientais. Essas limitações podem levar a incertezas na previsão das produtividades dos cultivos e dificultar a identificação de medidas adequadas para a adaptação.

Da mesma forma, os modelos estatísticos possuem entraves, principalmente relacionados à quantidade e qualidade dos dados de entrada, gerando incertezas nas previsões realizadas (Roberts *et al.*, 2017). Já os métodos estatísticos tradicionais, em geral, consideram apenas as correlações lineares entre a produtividade do cultivo e as variáveis independentes, resultando em previsões imprecisas, pois não refletem com precisão as relações não lineares existentes (Dormann *et al.*, 2013).

Nos últimos anos há um aumento significativo no uso de modelos de *Machine Learning* (ML) como uma ferramenta para avaliar os efeitos das variáveis nutricionais e climáticas na produção agrícola (Mishra *et al.*, 2016; Barbedo 2019). O ML se revela com potencial para melhoria da produtividade e da eficiência na agricultura, com capacidade de analisar grandes quantidades de dados em tempo real. Os modelos podem auxiliar os agricultores a identificarem padrões e tendências, além de preverem a produtividade das colheitas, o rendimento do solo, o uso de recursos e até mesmo as condições meteorológicas futuras (Zheng *et al.*, 2019). Desse modo, os parâmetros são ajustados conforme o conjunto de dados e o problema específico que está sendo abordado (Yang e Shami, 2020).

Os modelos de ML utilizam as interações entre os preditores e sem nenhum pré-requisito para utilização, além de capturar as relações não lineares (Kan, 2015). Um modelo acurado para prever a produtividade da batateira beneficia muitos aspectos o gerenciamento de áreas produtivas de batata. Portanto, a hipótese é a possibilidade do uso de algoritmos de ML para a previsão da produtividade da batateira a partir de variáveis nutricionais e climáticas no estado de São Paulo, Brasil.

Portanto, os objetivos foram: 1- prever a produtividade da batateira, 2 - avaliar o desempenho dos modelos de *Machine Learning* na previsão da produtividade da batateira no estado de São Paulo, Brasil e 3 – desenvolver uma aplicação on-line que faça a previsão da produtividade de culturas agrícolas.

3.2 MATERIAL E MÉTODOS

3.2.1 Caracterização do banco de dados

Para a construção do banco de dados foram utilizadas as informações sobre teores de nutrientes na folha diagnose do estado nutricional e produtividade da batata, provenientes de 21 experimentos realizados nos municípios de: São Manuel, Avaré, Botucatu, Itai, Cerqueira César, Taquarituba, Bernardino e Jaboticabal, São Paulo (Figura 1), no período compreendido entre 2011 e 2019.

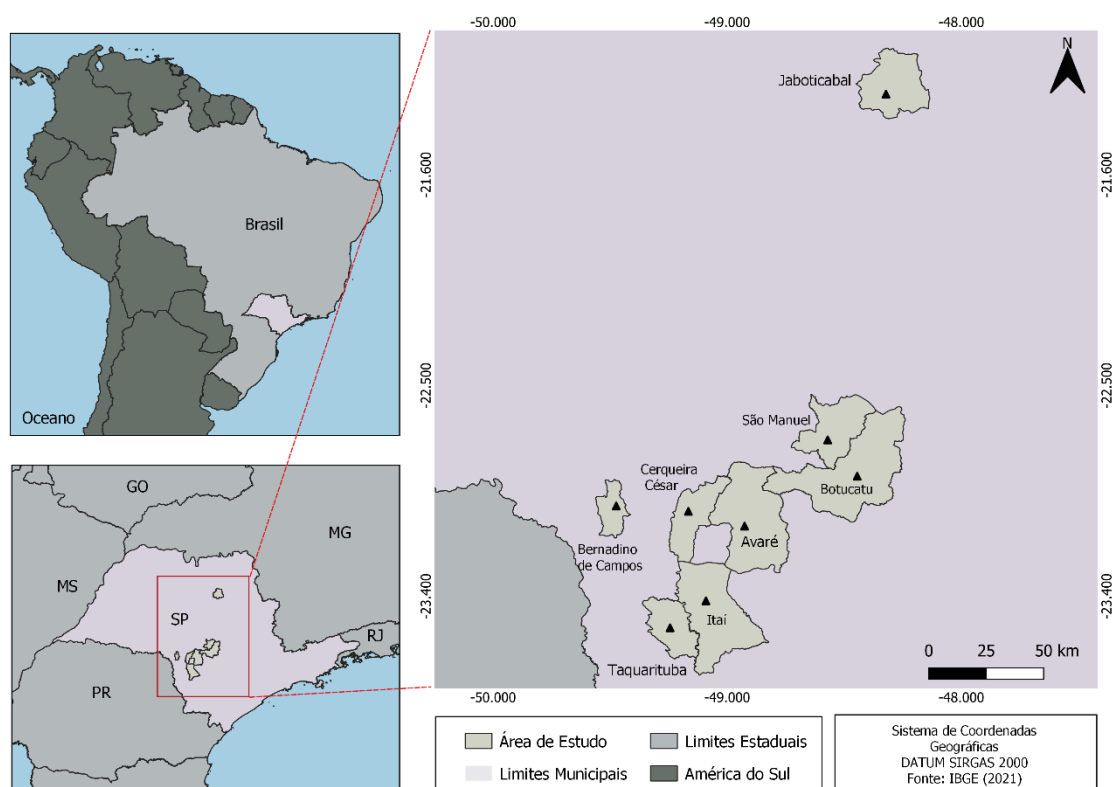


Figura 1. Mapa de localização dos municípios em que foram realizados os experimentos.

Esses municípios, geograficamente, se encontram próximos, o que acaba conferindo características semelhantes, correspondendo a duas classificações, sendo elas: 1. do tipo Cfa, caracterizado pelo clima tropical de altitude, com inverno seco e verão quente e chuvoso; e 2. tipo Aw, tropical, com seca no inverno. De forma mais ampla, pode ser aferido que a região apresenta temperaturas moderadas, chuvas bem distribuídas e verões quentes. Nesse cenário, a similaridade das condições climáticas pode reduzir a relevância do peso das variáveis climáticas em comparação com as variáveis nutricionais.

Em relação à formação dos solos, das áreas experimentais são classificadas como: latossolo vermelho distroférico, latossolo vermelho distrófico e latossolo vermelho eutroférico. Estes solos apresentam características semelhantes quanto aos atributos físicos (densidade do solo, resistência à penetração, retenção e armazenamento de água) e químicos, tais como pH do solo, teor de matéria orgânica, capacidade de troca catiônica (CTC), teores de nutrientes e salinidade. Os atributos químicos têm um impacto direto na produção da cultura, influenciando sua fertilidade, disponibilidade de nutrientes e capacidade de retenção de água, afetando, assim, o crescimento e desenvolvimento das plantas.

Nos experimentos (Tabela 1) foram avaliadas as respostas das cultivares de batata às doses de Nitrogênio (N), Fósforo (P) e Potássio (K), aplicadas nos solos dos 21 municípios. Para as dosagens de NPK foram consideradas os tipos de tubérculos (Ágata, Asterix, Atlantic, Electra, Markies e Mondial), o clima e o solo, pois esses fatores interferem diretamente na qualidade e na produtividade da cultura. A Tabela 1 é possível verificar os dados dos 21 experimentos, constando a altitude de cada cidade, o cultivo das espécies de batata, o tratamento e preparo do solo com o acréscimo de N, P e K, bem como a efetivação da colheita.

Ao variar as doses de N, P e K, é possível determinar qual a quantidade ideal desses nutrientes para otimizar a produtividade e qualidade da batata cultivada em cada região. Considerando que diferentes cultivares de batata foram testadas, assim como variações climáticas e tipos de solo, os resultados dos experimentos permitirão identificar quais combinações de cultivares e doses de nutrientes são mais adequadas para cada condição específica.

Tabela 1. Caracterização dos experimentos utilizados no banco de dados.

ID	Município	Altitude (m)	Cultivar	Tratamento	Colheita
1	São Manuel	750	Ágata	Nitrogênio	20/08/2011
2	São Manuel	750	Ágata	Nitrogênio	17/08/2012
3	Avaré	744	Ágata	Nitrogênio	27/07/2012
4	São Manuel	750	Ágata	Nitrogênio	20/08/2011
5	São Manuel	750	Ágata	Nitrogênio	17/08/2012
6	Avaré	744	Ágata	Nitrogênio	27/07/2012
7	Botucatu	786	Ágata	Nitrogênio	20/08/2016
8	Botucatu	778	Ágata	Nitrogênio	06/11/2016
9	Botucatu	830	Ágata	Nitrogênio	04/11/2017
10	Botucatu	786	Electra	Nitrogênio	20/08/2016
11	Botucatu	778	Electra	Nitrogênio	06/11/2016
12	Botucatu	830	Electra	Nitrogênio	04/11/2017
13	Avaré	744	Ágata, Asterix, Atlantic, Markies e Mondial	Fósforo	18/08/2011
14	Itaí	649	Ágata, Asterix, Atlantic, Markies e Mondial	Fósforo	23/07/2011
15	Cerqueira César	737	Ágata, Asterix, Atlantic, Markies e Mondial	Fósforo	25/08/2011
16	Taquarituba	676	Ágata e Mondial	Fósforo	24/09/2012
17	Taquarituba	676	Ágata	Potássio	22/09/2012
18	Bernardino	695	Ágata	Potássio	20/07/2013
19	Botucatu	740	Ágata	Potássio	13/11/2013
20	Jaboticabal	615	Asterix	Nitrogênio	11/07/2019
21	Jaboticabal	615	Ágata	Nitrogênio	04/07/2019

¹ ID: Identificação dos experimentos: 1 a 6 – Souza (2014); 7 a 12 – Fernandes (2021b); 13 a 16 – Fernandes (2013); 17 a 19 – Job (2014); 20 e 21 – Nascimento (2021).

Cada linha da Tabela representa um experimento específico, identificado pelo ID, fornecendo informações importantes sobre o município em que o experimento foi conduzido, bem como a altitude desse município, a cultivar de batata utilizada, o tipo de tratamento realizado e a data da colheita. Esses dados são essenciais para entender a influência das doses de nutrientes e das características específicas de cada localidade nos resultados relacionados à produtividade e qualidade da batata.

A classificação climática dos experimentos de 1 a 19 é do tipo Cfa (Köppen-Geiger, 1948; Alvares *et al.*, 2013), clima temperado quente (mesotérmico) úmido, e a temperatura média do mês mais quente é superior a 22 °C (Vianello; Alves, 2000). Já os experimentos 20 e 21 apresentam clima do tipo Aw (Köppen-Geiger, 1948), tropical, com seca no inverno (Lombardi Neto; Drugowich, 1994). O clima tem impacto direto na produtividade do cultivo da batata, podendo: reduzir o desenvolvimento vegetativo; estolões curtos; supressão do florescimento; tuberação mais precoce; enchimento rápido dos

tubérculos; maturação precoce; e anomalias ou desordens fisiológicas, provocando a redução na comercialização.

A produtividade foi obtida na colheita e na pesagem dos tubérculos das plantas de cada parcela, sendo calculada a produtividade por área (kg ha^{-1}). Para avaliação do estado nutricional da batata, foi realizada a amostragem de folhas com coleta da terceira folha a partir do tufo apical, de acordo com Lorenzi *et al.* (1997). O material amostrado foi lavado, secado em estufa e moído para realização das determinações dos teores dos nutrientes (N, P, K, Ca, Mg, S, B, Cu, Fe, Mn e Zn), segundo metodologia por Miyazawa *et al.* (2009).

3.2.2 Dados meteorológicos

Os dados climáticos das áreas experimentais de 2011 a 2019 foram coletados diariamente pela estação meteorológica automatizada *Prediction of Worldwide Energy Resources – NASA POWER* (<https://power.larc.nasa.gov>), uma fonte confiável, que apresenta informações precisas e de forma gratuita. Importante destacar que esses dados climáticos são amplamente utilizados em pesquisas científicas devido à sua acurácia e relevância para as condições brasileiras (Giovanella *et al.*, 2021), tendo como fonte dados de satélites.

A escolha, pelos dados de estações meteorológicas automáticas, se justifica devido à possibilidade da redução nos erros de estimativas dos experimentos, isso quando analisados com os índices R^2 ajustado e RMSE. Tendo em vista, que os dados de uma estação meteorológica física, em comparação com dados de satélite, possuem menores chances de erro. No entanto, cabe salientar que essas estações apresentam resoluções baixas. Desse modo, os parâmetros utilizados foram: temperatura média (T_m , °C), temperatura máxima (T_{max} , °C), temperatura mínima (T_{min} , °C), radiação global (QG, $\text{MJ m}^{-2} \text{ dia}^{-1}$), evapotranspiração (PET, mm dia^{-1}), evapotranspiração real (AET, mm dia^{-1}), precipitação + irrigação (P+I, mm dia^{-1}), armazenamento de água no solo (STO, mm), deficiência hídrica (DEF, mm dia^{-1}) e excedente hídrico (EXC, mm dia^{-1}), dados que impactam diretamente na produtividade da cultura de batata.

Os dados de temperatura e radiação global, obtidos por meio da estação meteorológica da plataforma NASA POWER, foram utilizados na estimativa da evapotranspiração utilizando o método proposto por Penman-Monteith. Esse

método é amplamente reconhecido e adotado na literatura científica (equações 1 a 4) para calcular a taxa de evapotranspiração com base em variáveis climáticas relevantes. Assim:

$$PET = \frac{0,408 \times s \times (Rn - G) + \frac{\gamma \times U_2 \times (e_s - e_a)}{T + 273}}{s + \gamma \times (1 + 0,34 \times U_2)} \quad (1)$$

$$s = \frac{4098 \times e_s}{(T + 237,3)^2} \quad (2)$$

$$e_s = 0,6108 \times 10^{\frac{7,5 \times T}{237,3 + T}} \quad (3)$$

$$e_a = \frac{e_s \times UR}{100} \quad (4)$$

em que: Rn é a radiação líquida total diária ($\text{MJ m}^{-2} \text{ d}^{-1}$); G é o fluxo de calor no solo ($\text{MJ m}^{-2} \text{ d}^{-1}$); γ é a constante psicrométrica ($0,063 \text{ kPa } ^\circ\text{C}^{-1}$); T é a temperatura média do ar ($^\circ\text{C}$); U_2 é a velocidade do vento a 2 m (m s^{-1}); e_s é a pressão de saturação de vapor (kPa); e_a é a pressão atual de vapor (kPa); s é a declividade da curva de pressão de vapor (kPa).

Para calcular o saldo de radiação ($\text{MJ m}^{-2} \text{ dia}^{-1}$) – Rn , calcula-se o somatório do balanço de ondas curtas (BOC) mais o balanço de ondas longas (BOL), como segue:

$$\begin{aligned} Rn &= BOC + BOL \\ BOC &= Qg - r \times Qg & BOL &= Qatm - r \times Qsup & Qsup &= \epsilon \times \sigma \times T_k^4 \end{aligned} \quad (5)$$

nas quais $r = 0,25$ (albedo); $\epsilon = 0,95$ (poder emissivo); $\sigma = 4,90 \times 10^{-9}$ ($\text{MJ m}^{-2} \text{ dia}^{-1} \text{ K}^{-4}$) (constante de stefan-Boltzmann); $T_k = T$ ($^\circ\text{C}$) + 273,15; sendo que T , Qo , Qg e $Qatm$ são dados extraídos do site NASA POWER.

O armazenamento mensal de água no solo (STO), a evapotranspiração real (AET) e a deficiência de água (DEF) foram calculados conforme descrito por Thornthwaite e Mather (1955) (Figura 2), com água disponível na capacidade de armazenamento de 100 mm.

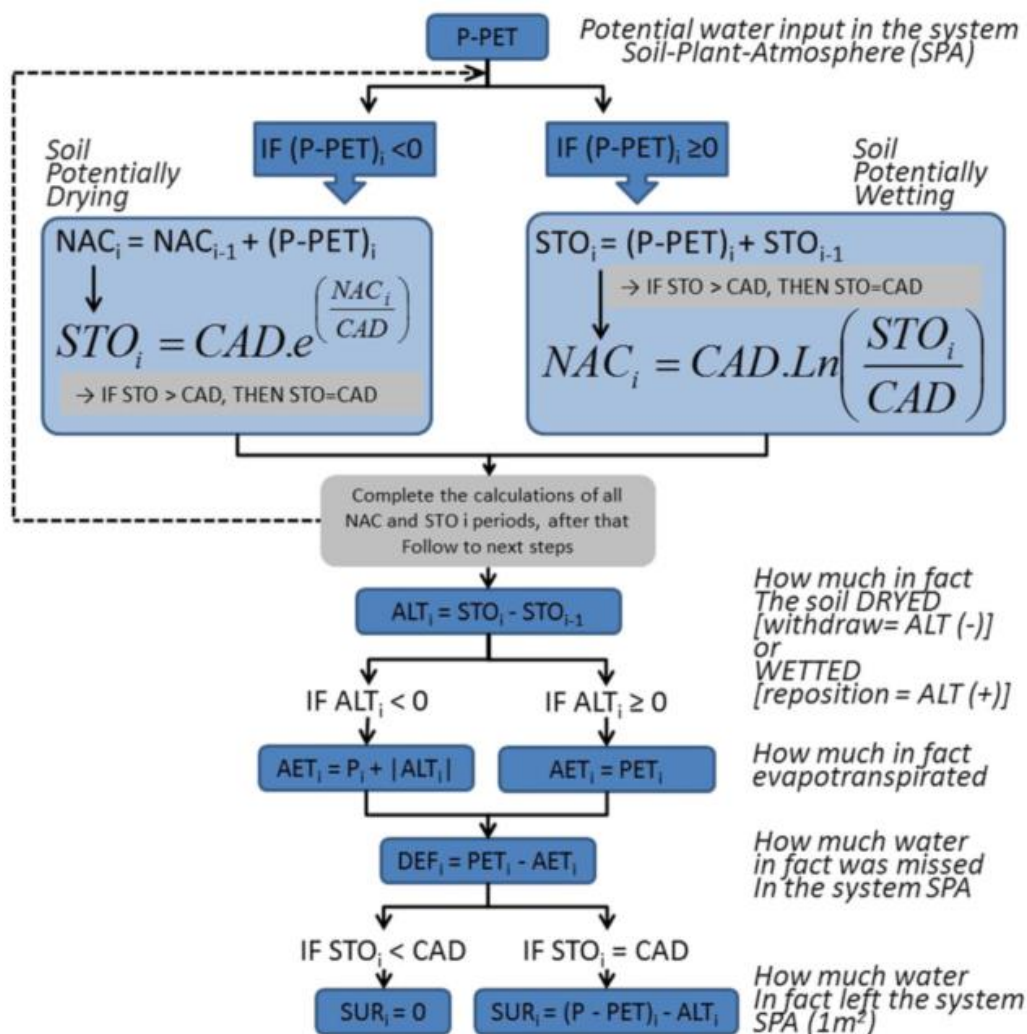


Figura 2. Fluxograma de cálculo de balanço Hídrico segundo modelo proposto por Thorthwaite e Mather (1955), conforme Rolim *et al.* (2020).

Simbologia: *PET* é a evapotranspiração (mm dia⁻¹); *WC* é a capacidade de água disponível (mm dia⁻¹); *STO* é o armazenamento de água no solo (mm); *NAC* = soma da precipitação - evapotranspiração potencial; *P* é a precipitação (mm dia⁻¹); *DEF* é a deficiência de água no sistema solo-planta-atmosfera (mm dia⁻¹); *AET* é a evapotranspiração real (mm dia⁻¹); *ALT* é o armazenamento de água no solo do mês atual - o armazenamento de água no solo do mês anterior (mm) e, *i* é o período do mês.

3.2.3 Caracterização dos estádios fenológicos da batateira

Os estádios fenológicos da batateira são cinco, sendo eles: emergência; desenvolvimento das folhas e raízes; estolinização e tuberização; florescimento e crescimento dos tubérculos; senescência e maturação. Ao finalizar o ciclo dos primeiros 45 dias, existe o fechamento dos espaços entre linhas, o que corresponde a primeira parte do ciclo de plantio da batata. Momento, onde é

finalizado a tuberização, iniciando o florescimento e crescimento dos tubérculos (Thomas; Bredemeier; Vian, 2016).

O segundo ciclo do desenvolvimento da planta de batata é composto pelas fases de florescimento e crescimento, onde ocorre a fecundação das flores e o desenvolvimento de frutos contendo sementes (Thomas; Bredemeier; Vian, 2016). Durante essa fase, há um acúmulo máximo de biomassa na parte aérea da planta, um crescimento exponencial dos tubérculos e uma redução no crescimento das raízes. Após o florescimento, não ocorre mais crescimento das folhas, pois ocorre o amarelecimento das folhas e hastes, um processo conhecido como senescência. Por fim, a maturação é alcançada quando as folhas e hastes ficam secas e estão prontas para cair.

Na cultura da batata o tubérculo é a parte visada agronomicamente. Assim, torna-se necessário pontuar especificamente o estágio fenológico do desenvolvimento do tubérculo semente. Esse estágio é dependente das condições ambientais, fertilidade do solo e a idade cronológica do tubérculo, esses fatores interferem na produção. Assim, os estádios fenológicos do tubérculo semente, são: dormência, dominância apical, plena brotação e senescência, cada um desses estádios contribui para o desenvolvimento da planta (Thomas; Bredemeier; Vian, 2016).

Em determinadas circunstâncias, recomenda-se o plantio da batata na fase de dominância apical, dependendo do objetivo da produção. Esse tipo de plantio é indicado quando se deseja obter campos com plantas de poucas hastes, visando a produção de material para processamento ou consumo in natura. Por outro lado, o plantio nessa fase pode resultar em plantas com apenas uma haste, o que consequentemente resulta em uma lavoura com baixa densidade de hastes (Silva; Hirano, 2015; Thomas; Bredemeier; Vian, 2016). No entanto, quando o objetivo é obter plantas com muitas hastes para a produção de sementes, é recomendado o plantio de batatas com brotação múltipla (Silva; Hirano, 2015).

No contexto desta pesquisa, foram considerados os cinco estágios para caracterizar o ciclo das plantas de batata, com variações nas amplitudes climáticas, de acordo com os tratamentos e experimentos realizados (Figura 3). A amplitude climática tem um impacto direto nos tratamentos e experimentos da batata, uma vez que a cultura requer temperaturas amenas para o processo de tuberização. Mesmo em um ambiente com a temperatura ideal, entre 15 °C e 20 °C, é importante considerar que uma alta amplitude térmica, com temperaturas diurnas elevadas e temperaturas noturnas amenas, pode ser favorável para a produção de batatas (Lopes, 2011).

A amplitude climática tem um impacto direto no metabolismo das plantas e interage com outros fatores ambientais, desempenhando um papel significativo no seu desenvolvimento. No caso específico da batata, temperaturas elevadas têm efeitos negativos, pois reduzem a síntese de fotoassimilados, essenciais

para o desenvolvimento da planta e sua alocação aos tubérculos. Isso resulta em queda de rendimento e diminuição da matéria seca dos tubérculos (Lopes, 2011).

Desse modo, para avaliação de cada variável climática, considerou-se a média de cinco dias e caracterizada por três argumentos: 1. Tm_EIII_35d , a variável climática (Temperatura média - Tm); 2. o estágio (E) em que o parâmetro está sendo avaliado; e 3. refere-se ao período de dias (d) da avaliação. No citado exemplo, o valor obtido para Tm_EIII_35d corresponde a temperatura média dos cinco dias no intervalo de 31 a 35 dias do ciclo no estágio de desenvolvimento de folhas e raízes. Esse procedimento foi realizado para todas as variáveis climáticas, totalizando 28 períodos para cada variável climática (Figura 3).

DAE	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100								
Estádios	Emergência - I			Desenvolvimento das folhas e raízes - II						Estolinização - III				Florescimento e crescimento dos tubérculos - IV										Senescência e maturação - V				
Variáveis climáticas	Tm_EM5	Tm_EM10	Tm_EM15	Tm_DV20	Tm_DV25	Tm_DV30	Tm_ES35	Tm_ES40	Tm_FT45	Tm_FT50	Tm_FT55	Tm_FT60	Tm_FT65	Tm_FT70	Tm_FT75	Tm_FT80	Tm_SM85	Tm_SM90	Tm_SM95	Tm_SM100								
	Tmax_EM5	Tmax_EM10	Tmax_EM15	Tmax_DV20	Tmax_DV25	Tmax_DV30	Tmax_ES35	Tmax_ES40	Tmax_FT45	Tmax_FT50	Tmax_FT55	Tmax_FT60	Tmax_FT65	Tmax_FT70	Tmax_FT75	Tmax_FT80	Tmax_SM85	Tmax_SM90	Tmax_SM95	Tmax_SM100								
	Tmin_EM5	Tmin_EM10	Tmin_EM15	Tmin_DV20	Tmin_DV25	Tmin_DV30	Tmin_ES35	Tmin_ES40	Tmin_FT45	Tmin_FT50	Tmin_FT55	Tmin_FT60	Tmin_FT65	Tmin_FT70	Tmin_FT75	Tmin_FT80	Tmin_SM85	Tmin_SM90	Tmin_SM95	Tmin_SM100								
	QG_EM5	QG_EM10	QG_EM15	QG_DV20	QG_DV25	QG_DV30	QG_ES35	QG_ES40	QG_FT45	QG_FT50	QG_FT55	QG_FT60	QG_FT65	QG_FT70	QG_FT75	QG_FT80	QG_SM85	QG_SM90	QG_SM95	QG_SM100								
	PET_EM5	PET_EM10	PET_EM15	PET_DV20	PET_DV25	PET_DV30	PET_ES35	PET_ES40	PET_FT45	PET_FT50	PET_FT55	PET_FT60	PET_FT65	PET_FT70	PET_FT75	PET_FT80	PET_SM85	PET_SM90	PET_SM95	PET_SM100								
	AET_EM5	AET_EM10	AET_EM15	AET_DV20	AET_DV25	AET_DV30	AET_ES35	AET_ES40	AET_FT45	AET_FT50	AET_FT55	AET_FT60	AET_FT65	AET_FT70	AET_FT75	AET_FT80	AET_SM85	AET_SM90	AET_SM95	AET_SM100								
	P+I_EM5	P+I_EM10	P+I_EM15	P+I_DV20	P+I_DV25	P+I_DV30	P+I_ES35	P+I_ES40	P+I_FT45	P+I_FT50	P+I_FT55	P+I_FT60	P+I_FT65	P+I_FT70	P+I_FT75	P+I_FT80	P+I_SM85	P+I_SM90	P+I_SM95	P+I_SM100								
	STO_EM5	STO_EM10	STO_EM15	STO_DV20	STO_DV25	STO_DV30	STO_ES35	STO_ES40	STO_FT45	STO_FT50	STO_FT55	STO_FT60	STO_FT65	STO_FT70	STO_FT75	STO_FT80	STO_SM85	STO_SM90	STO_SM95	STO_SM100								
	DEF_EM5	DEF_EM10	DEF_EM15	DEF_DV20	DEF_DV25	DEF_DV30	DEF_ES35	DEF_ES40	DEF_FT45	DEF_FT50	DEF_FT55	DEF_FT60	DEF_FT65	DEF_FT70	DEF_FT75	DEF_FT80	DEF_SM85	DEF_SM90	DEF_SM95	DEF_SM100								
	EXC_EM5	EXC_EM10	EXC_EM15	EXC_DV20	EXC_DV25	EXC_DV30	EXC_ES35	EXC_ES40	EXC_FT45	EXC_FT50	EXC_FT55	EXC_FT60	EXC_FT65	EXC_FT70	EXC_FT75	EXC_FT80	EXC_SM85	EXC_SM90	EXC_SM95	EXC_SM100								

Figura 3. Estádios fenológicos da batateira.

Simbologia: Tm = temperatura média, Tmax = temperatura máxima, Tmin = temperatura mínima, QG = radiação global, PET = evapotranspiração, AET = evapotranspiração real, P+I = precipitação + irrigação, STO = armazenamento de água no solo, DEF = deficiência hídrica, EXC = excedente hídrico, EM = emergência, DV = Desenvolvimento das folhas e raízes, ES = Estolinização, FT = Florescimento e crescimento dos tubérculos, SM = senescência e maturação.

Como pode ser visualizado, temos a subdivisão dos cinco estágios fenológicos da batateira em intervalos de cinco dias, a fim de compreender a influência das variáveis climáticas em cada estágio. Essa subdivisão foi adotada devido à imprecisão na duração exata de cada estágio fenológico, uma vez que cada cultivar de batata possui um intervalo específico para cada fase. Da mesma forma, as variáveis climáticas também foram subdivididas em intervalos de cinco dias.

Os dados foram, posteriormente, utilizados para a modelagem de algoritmos de *Machine Learning*, buscando relacionar as variáveis climáticas com os estágios fenológicos da batateira. Essa abordagem visa obter modelos preditivos mais precisos e adequados para compreender e prever o desenvolvimento das culturas de batata com base nas condições climáticas.

3.2.4 Padronização

A padronização de dados é uma etapa fundamental no processo de preparação de dados para modelagem de *Machine Learning*. Essa técnica é especialmente importante em algoritmos baseados em distância, pois a diferença na escala das variáveis pode levar a uma influência desproporcional na medida de distância (El Naqa *et al.*, 2018).

Os algoritmos de aprendizado de máquina utilizam gradientes de peso para minimizar a função de perda, que corresponde a uma medida para verificação do modelo. Quando as variáveis apresentam diferentes escalas, isso pode afetar a magnitude dos gradientes de peso, levando a problemas de convergência ou divergência (Silva e Timo, 2022).

A função `scale` no R é uma ferramenta útil para padronizar dados, pois permite que os dados sejam centralizados na média e escalados pela variação, resultando em dados com média zero e desvio padrão unitário. Essa padronização tem o efeito de igualar a importância das variáveis na modelagem, permitindo que os algoritmos de aprendizado de máquina sejam treinados de forma mais eficiente. Além disso, a padronização também pode ajudar a evitar problemas com gradientes de peso em algoritmos de aprendizado de máquina, o que pode levar a resultados não convergentes ou até mesmo divergentes (Jones, 2019).

3.2.5 Seleção de variáveis

O banco de dados é oriundo dos vinte e um experimentos, conforme já mencionado, composto pelo grupo das variáveis nutricionais, com onze nutrientes. O passo seguinte foi analisar todas as coordenadas geográficas e o tempo em que esses experimentos foram conduzidos, isso tendo como base as produções científicas existentes.

Após essa apuração, as coordenadas e datas específicas foram inseridas no *Prediction of Worldwide Energy Resources*, o que permitiu encontrar os dados climáticos para cada período e experimento. Assim, tendo as variáveis climáticas e as variáveis nutricionais, foi calculado a média diária, precipitação diária e velocidade do vento, com esses resultados foi realizado o cálculo da média de cinco dias, obtendo assim, o resultado da média de cada variável climática, a cada cinco dias. Desse modo, ao final havia um total de 280 variáveis, sendo divididos em grupos de cinco dias. A partir disso, apresentam-se as variáveis nutricionais, as variáveis climáticas e a produtividade estimada.

A produtividade é a variável resposta e é a partir dela que a modelagem é realizada. As variáveis independentes utilizadas neste estudo foram compostas por variáveis climáticas e nutricionais. Para fins de análise, foram criados três grupos distintos: grupo contendo somente variáveis nutricionais, outro grupo contendo somente variáveis climáticas e o terceiro grupo composto pelas top 10 variáveis mais relevantes dentre todas as variáveis nutricionais e climáticas. Essa abordagem permite avaliar o impacto de cada grupo de variáveis no modelo e identificar quais variáveis possuem maior influência na produtividade da batateira.

Para definir as top 10 variáveis mais importantes, utilizou-se a técnica de Random Forest (RF), que permite avaliar a relevância de cada variável no modelo. O RF é um modelo de aprendizado de máquina eficiente para selecionar as variáveis mais importantes em um grande banco de dados. Isso ocorre porque o modelo é capaz de lidar com muitas variáveis e é capaz de identificar interações complexas entre essas variáveis, o que pode ser útil para prever a produtividade com base em dados climáticos e nutricionais (Saleem *et al.*, 2021).

Em seguida, foi realizado um processo de otimização de parâmetros para selecionar os melhores parâmetros para o modelo. Essa abordagem garante que

o modelo construído seja o mais adequado possível para os dados analisados (Demir e Şahin, 2022). A otimização de parâmetros é uma tarefa crucial na construção de modelos de aprendizado de máquina, uma vez que os parâmetros influenciam diretamente no comportamento dos algoritmos (Qolomany *et al.*, 2017). Escolher os valores adequados podem impactar significativamente a performance do modelo (Al-Sahaf *et al.*, 2019). Existem diversas técnicas para otimização de parâmetros, e uma delas é o Tunegrid.

Tunegrid é uma técnica avançada que utiliza uma estratégia de busca eficiente para encontrar a melhor combinação de parâmetros. Em vez de testar todas as combinações possíveis de valores, a técnica utiliza um algoritmo de busca para explorar o espaço de parâmetros. Tunegrid é particularmente útil quando há muitos parâmetros e valores possíveis para serem testados, sendo mais rápida e sofisticada que outras técnicas, como o GridSearch (Boeschoten *et al.*, 2023).

Para otimizar os parâmetros de modelo Random Forest, foi ajustado o *mtry* e o *ntree*. O *mtry* representa o número de variáveis amostradas aleatoriamente como candidatas em cada divisão das árvores de decisões calculadas, enquanto o *ntree* representa o número de árvores de decisão que serão construídas na floresta aleatória. Assim, foi realizado um processo de otimização de parâmetros utilizando a técnica de *tunegrid* para selecionar os melhores valores para as variáveis *ntree* e *mtry*.

O processo de otimização de parâmetros foi realizado testando 160 combinações de valores para *ntree* e *mtry*. A lista de valores para *ntree* variou de 50 a 500, com incrementos de 50 unidades, enquanto a lista de valores para *mtry* variou de 50 a 200, com incrementos de 10 unidades. Essa busca permitiu encontrar a melhor combinação de valores para esses parâmetros, resultando em um modelo com melhor performance.

3.2.6 Modelos de *Machine Learning*

Com o intuito de identificar as variáveis que mais impactam na produtividade da batateira, foram desenvolvidos cinco modelos de *Machine Learning* para análise dos dados. As opções de modelos utilizados foram: Regressão Linear Múltipla (RLM), K-nearest neighbors ou “K-vizinhos mais próximos” (KNN), Support Vector Machine (SVR), eXtreme Gradient Boosting

(XGB) e Redes Neurais Artificiais (RN). Esses modelos foram escolhidos pela capacidade de adaptação e melhoramento do comportamento em relação às respostas do aprendizado, conjunto de dados e a busca por padrões.

Além disso, a escolha de diferentes modelos permite avaliar qual técnica apresenta o melhor desempenho na predição da variável resposta e pode fornecer insights valiosos sobre a relação entre as variáveis independentes e dependentes. Cada modelo de aprendizado de máquina tem suas vantagens e desvantagens que podem influenciar na escolha do modelo a ser utilizado em determinado problema. A seguir, serão discutidos os prós e contras dos modelos utilizados.

3.2.6.1 Regressão Linear Múltipla (RLM)

A RLM é um modelo simples e fácil de ser interpretado. É indicado para problemas onde existe uma relação linear entre as variáveis independentes e a variável resposta. No entanto, esse modelo assume que a relação é linear, o que pode não ser verdade para todos os casos (Rodríguez del Águila *et al.*, 2011).

Na agricultura, a regressão linear múltipla se mostra importante para realizar análises preliminares de dados e prever tendências futuras, além de traçar gráficos entre as diversas variantes (precipitação, temperatura, uso de fertilizantes no rendimento de colheitas, qualidade do cultivo, incidência de doenças e/ou pragas, outros). Entretanto, pode não ser a técnica mais adequada para analisar a relação entre as variáveis climáticas e a produtividade das culturas, tendo em vista que muitas dessas variáveis climáticas estão inter-relacionadas, ou seja, estão altamente correlacionadas entre si (Prabhu e Dakshayini, 2018).

A multicolinearidade pode afetar a interpretação dos resultados, pois torna difícil identificar a contribuição individual de cada variável no modelo. Além disso, a multicolinearidade pode levar a uma instabilidade nos coeficientes das variáveis independentes, o que pode tornar o modelo menos confiável (Garg e Tai, 2013). Por esse motivo, optamos por recorrer à combinação de modelos que garantam um maior nível de confiabilidade para a interpretação e transparência dos resultados, visando garantir sua aplicabilidade na produtividade do cultivo analisado.

3.2.6.2 K-Nearest Neighbors (KNN)

O KNN regressão é um modelo não paramétrico, que não assume nenhuma forma particular para a relação entre as variáveis independentes e a variável resposta. É indicado para problemas em que a relação é complexa e não-linear. Contudo, pode ser computacionalmente pesado para grandes conjuntos de dados e pode ser sensível a valores extremos e pode produzir resultados imprecisos caso não haja uma quantidade suficiente de dados para treinamento do modelo (Song, 2017).

O KNN regressão na área agrícola pode ser utilizado nos problemas que envolvem múltiplas variáveis ambientais, tais como o clima, solo e nutrientes. Na pesquisa de Shakoore *et al.* (2017), os autores utilizaram esse modelo para auxiliar os agricultores nas tomadas de decisões, buscando o aumento da margem de lucro. Para tanto, foram utilizados os dados de produção, clima e região para prever o resultado mais rentável.

3.2.6.3 Support Vector Machine (SVM)

O SVM é um modelo que busca maximizar a margem entre as classes, sendo indicado para problemas onde elas são bem separáveis. Contudo, apresenta certo nível de dificuldade para a escolha do kernel adequado e pode ser sensível à escolha do parâmetro C (Fan *et al.*, 2018).

Na agricultura, o SVM pode ser útil em problemas de classificação e regressão, como identificação de doenças em plantas, classificação de tipos de solo e a influência das variáveis na produtividade. Outra vantagem, está na capacidade de lidar com um conjunto de dados com alta dimensionalidade, comum em análises agrícolas que envolvem muitas variáveis (Anwar *et al.*, 2021).

3.2.6.4 eXtreme Gradient Boosting

O eXtreme Gradient Boosting é um modelo que combina vários modelos fracos para formar um modelo forte. É indicado para problemas com grande quantidade de dados e muitas variáveis independentes. Entretanto, pode ser sensível à escolha dos parâmetros e pode ser computacionalmente pesado (Sahin *et al.*, 2020; Zhang *et al.*, 2021), exigindo recursos de computação

avançados para treinar e ajustar o modelo adequadamente, o que pode impactar significativamente a qualidade dos resultados. Assim, devido à complexidade do modelo, pode ser difícil interpretar os resultados e entender como as variáveis independentes estão contribuindo para a variável resposta (Zhang *et al.*, 2021).

O XGB tem sido amplamente utilizado na agricultura para prever rendimentos de culturas e auxiliar na tomada de decisões. Uma das principais vantagens está na capacidade de lidar com grandes conjuntos de dados, tornando-o adequado para análises de grande escala na agricultura. Além disso, o modelo pode lidar com diferentes tipos de variáveis, incluindo variáveis climáticas, nutricionais e de solo (Anwar *et al.*, 2021; Mariadass *et al.*, 2022).

3.2.6.5 Redes neurais artificiais (RN)

As RN podem ser utilizadas para problemas complexos e não-lineares (Chen *et al.*, 1990). No entanto, podem exigir um grande conjunto de dados e ser sensível à escolha da arquitetura da rede. O Multi-Layer Perceptron (MLP) é uma rede neural artificial que consiste em uma série de camadas de neurônios interconectados (Nestor, 2006). É capaz de lidar com problemas complexos e não-lineares, sendo amplamente utilizado em várias áreas, incluindo a agricultura (Choubin *et al.*, 2016).

Desse modo, o MLP pode aprender a partir de um conjunto de dados de entrada e de fornecer previsões para novos dados (Choubin *et al.*, 2016). Uma das principais vantagens é a sua capacidade de lidar com dados não-lineares e complexos. Além disso, ele pode lidar com várias entradas e saídas, permitindo a modelagem de problemas mais complexos. Apresenta uma alta flexibilidade, aceitando a escolha da arquitetura da rede, bem como a configuração de parâmetros para otimizar a performance (Hemalatha *et al.*, 2017).

Assim, como outras redes neurais, o MLP pode ser suscetível a *overfitting* em casos de sobreajuste dos dados de treinamento, o que pode resultar em desempenho insatisfatório para novos dados. Desse modo, a escolha da arquitetura da rede pode ser complexa e requer conhecimento especializado, bem como a otimização de parâmetros pode ser computacionalmente intensiva (Huang, 2009).

Ao utilizar somente uma camada no modelo de RN, cria-se uma rede do tipo Perceptron Simples, que é um modelo mais simples e menos complexo que

as redes de múltiplas camadas, como o MLP. Essa abordagem pode ser útil em situações em que o problema é relativamente simples e não exige um modelo muito complexo (Choubin *et al.*, 2016). Uma das principais vantagens de utilizar apenas uma camada é que o modelo é mais fácil de ser interpretado e treinado, já que há menos parâmetros a serem ajustados. Ainda, ele tende a ser mais leve, o que pode ser vantajoso em casos em que a velocidade de processamento é um fator crítico (Nestor, 2006).

Todavia, utilizar somente uma camada pode ser limitante em situações em que o problema é complexo e exige uma representação mais sofisticada das relações entre as variáveis. Nesses casos, é possível que o modelo não consiga capturar todas as nuances do problema e, portanto, não apresente uma performance tão boa quanto um modelo de múltiplas camadas (Livingstone *et al.*, 1997).

Na agricultura, as RN podem ser utilizadas para modelar problemas complexos, como a relação entre variáveis climáticas e a produtividade das culturas. Uma das principais vantagens é que as redes neurais podem lidar com dados não-lineares e não paramétricos, o que as torna adequadas para modelar as relações complexas entre variáveis. Ademais, as redes neurais podem ser treinadas com grandes conjuntos de dados, o que é comum em estudos agrícolas (Singh e Panda, 2009).

3.2.7 Biblioteca Caret

Todos os modelos de aprendizado de máquina foram executados usando a biblioteca Classificação e Treinamento de Regressão (caret) no ambiente de programação R. A biblioteca “caret” é um conjunto de funções que buscam simplificar o processo de criação de modelos preditivos no R. O pacote contém ferramentas que permitem dividir os dados, pré-processamento, seleção de recursos, ajuste de modelo usando reamostragem e estimativa de importância de variáveis, além de outras funcionalidades (Kuhn, 2008).

O pacote oferece diversas vantagens para a implementação de modelos de aprendizado de máquina. O destaque desse modelo está na facilidade de uso, já que a biblioteca fornece funções que automatizam tarefas comuns, como pré-processamento dos dados, seleção de variáveis, ajuste de parâmetros e avaliação do modelo (Kuhn, 2008).

3.2.8 Treinamento e teste, parâmetros e validação cruzada

A configuração de um modelo de *Machine Learning* é uma tarefa importante e delicada que envolve diversas etapas, dentre as quais se destacam a definição da porcentagem de treinamento e teste, a escolha dos parâmetros do modelo e a realização da validação cruzada (*cross validation*). A primeira etapa consiste em definir como dividir os dados entre treinamento e teste. Essa divisão permite que o modelo seja treinado em uma parcela dos dados, enquanto outra parcela é utilizada para avaliar o seu desempenho. É importante escolher uma proporção que permita que o modelo seja treinado em dados suficientes, mas que também forneça dados suficientes para teste.

Para realizar as análises propostas para os fins da pesquisa, o banco de dados foi composto por 587 amostras, dividido em 75% para treinamento e 25% para teste. Na Figura 4, é possível visualizar a demonstração métrica de cada iteração/treino/teste, que possibilitam gerar os resultados para avaliação:



Figura 4. Esquema da divisão de dados em 75% para treinamento e 25% para teste e demonstração da validação cruzada em quatro vezes.

Durante a utilização do *cross validation*, os dados foram divididos em quatro grupos, sendo que cada grupo possui duas partes. A primeira parte é utilizada para teste e o restante são utilizadas para treinamento do modelo, gerando uma métrica de avaliação. Esse processo é repetido quatro vezes, até que toda a base passe pelo processo de treino e teste, garantindo assim, uma métrica de avaliação média para o modelo (Figura 4). Posteriormente, foram definidos os parâmetros, considerando a inclusão dos fatores com maior influência na produtividade da batata. A escolha correta dos parâmetros pode influenciar significativamente o desempenho do modelo (Tabela 2).

Tabela 2. Parâmetros dos modelos.

Modelo	Parâmetro	Valores testados	Intervalo
RF	<i>ntree</i>	50 a 500	50
	<i>mtry</i>	50 a 200	10
RLM	<i>intercept</i>	-10 a 3	0.1
KNN	<i>k</i>	1 a 50	1
SVR	<i>sigma</i>	0.01 a 2,	0.1
	<i>C</i>	0.01; 0.1; 0.5; 1; 5 e 10	-
XGB	<i>nrounds</i>	3; 5 e 10	-
	<i>lambda</i>	0.1 a 0.5	0.1
	<i>alpha</i>	0.1 e 0.5	0.1
	<i>eta</i>	0.1 e 0.5	0.1
RN	<i>sigma</i>	1 a 30	1

Como pode ser visualizado, teremos os parâmetros utilizados nos modelos de *Machine Learning*, juntamente com os valores testados e o intervalo de variação de cada parâmetro. Os valores dos parâmetros são testados durante cada iteração do *cross validation*, sendo ajustados a cada iteração. O modelo é treinado e testado com esses valores específicos. Ao repetir o processo para todas as iterações do *cross validation*, é possível avaliar o desempenho do modelo com diferentes combinações de parâmetros e selecionar aqueles que fornecem os melhores resultados médios.

A partir dos dados apresentados é possível visualizar os seis modelos utilizados, os parâmetros de avaliação, os testes das variantes e os intervalos de tempo, para a análise dos dados sobre produtividade. Cada modelo foi ajustado em uma base treino e, em seguida, aplicado em uma base teste para calcular medidas de performance para o objeto de estudo. Cabe destacar, que os parâmetros variam de acordo com as especificidades de cada modelo, porém, ao final para cada modelo é possível verificar uma amostragem da extensão dos dados do algoritmo na tarefa de previsão de ML, permitindo investigar o nível de importância que cada fonte de dados.

Assim sendo, o modelo RF foi utilizado para selecionar as top 10 variáveis, climáticas e nutricionais, utilizando como parâmetros o *ntree* (número de árvores) e o *mtry* (número de variáveis selecionadas em cada nó). Já no RLM, o parâmetro utilizado foi o *intercept* que definiu a relação linear entre as variáveis, com o intuito de estipular uma taxa média de mudança no desempenho da produtividade.

Portanto, a validação cruzada permite que o modelo seja testado em várias combinações diferentes de treinamento e teste, o que ajuda a garantir que o modelo não esteja superajustando ou subajustando aos dados de treinamento, bem como uma estimativa mais precisa do desempenho geral do modelo. No período de treinamento a validação cruzada foi realizada com *k-fold* igual a 4 (Figura 4). É importante destacar que essas etapas são interdependentes e podem ser iterativas, possibilitando refinar o modelo e melhorar o seu desempenho.

3.2.9 Avaliação de Modelos Preditivos

A avaliação do desempenho dos modelos de *Machine Learning* é um aspecto fundamental para garantir a qualidade das previsões e aprimorar as estratégias de manejo nas culturas agrícolas. Para essa avaliação, são utilizadas diversas métricas, como o R^2 ajustado e o RMSE e a análise do diagrama de Taylor.

O R^2 ajustado é uma métrica que avalia a qualidade do ajuste do modelo aos dados, indicando a proporção da variabilidade desses dados, que é explicada pelo modelo. Quanto mais próximo de 1 for o valor do R^2 ajustado, melhor será o desempenho do modelo. Já o RMSE é uma medida de erro que avalia a diferença entre os valores observados e previstos pelo modelo, quanto menor o valor do RMSE, menor será a discrepância entre os dados observados e previstos pelo modelo.

Além dessas métricas, se tem o diagrama de Taylor, uma ferramenta gráfica que permite comparar os modelos de previsão em relação às estatísticas dos dados observados. Nele, é possível visualizar o grau de concordância entre as previsões e os dados observados, bem como a variabilidade e o viés dos modelos em relação aos dados e identificar aquele que apresenta a melhor performance e maior precisão na previsão dos dados.

Coeficiente de determinação ajustado:

$$R_a^2 = 1 - \frac{(n-1)}{(n-1)-k} (1 - R^2)$$

na qual n é a quantidade total de amostras, k é o número de variáveis independentes na regressão e R^2 é o coeficiente de determinação.

Raiz do erro quadrático médio:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Yest_i - Yobs_i)^2}$$

sendo: n é o número de amostras, $Yest_i$ é a variável estimada e, $Yobs_i$ é a variável observada.

Diagrama de Taylor:

Taylor (2001) propôs a construção de um diagrama para comparação de modelos que contemplam algumas estatísticas, como: valor e desvios-padrão dos dados estimados e observados e o coeficiente de correlação. Dessa forma, evita-se a construção de extensas tabelas.

Na Figura 5, o modelo é representado pelo círculo preto, em que a distância radial do ponto de origem é proporcional ao desvio-padrão dos dados interpolados e, em azimuth do ponto, é o valor de R dos dados observados e estimados pelo método. A correlação do ponto de referência com ele próprio é igual a 1 e, por conseguinte, verifica-se no eixo x.

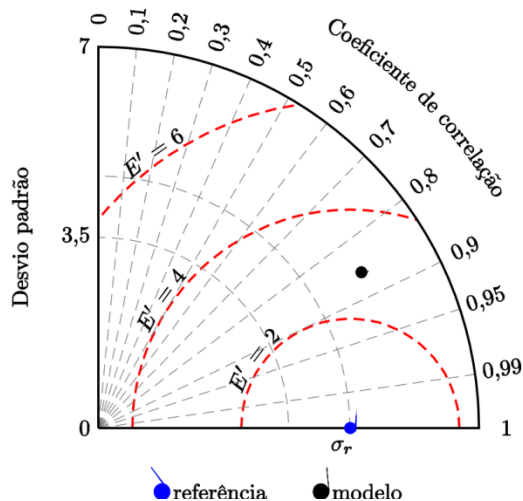


Figura 5. Exemplo de construção do Diagrama de Taylor. (Adaptado de Pereira *et al.*, 2014).

No Diagrama de Taylor, a distância radial do ponto de previsão até o ponto de referência indica o erro de previsão do modelo. Já o ângulo formado entre o eixo horizontal e a linha que conecta os pontos de referência e previsão, indica a correlação entre as duas séries (Pereira *et al.*, 2014). Desse modo, se o ponto de previsão estiver próximo ao ponto de referência e a linha que os conecta

estiver próxima ao eixo horizontal, isso indica que o modelo tem uma boa precisão e correlação com os dados reais. Por outro lado, caso o ponto de previsão estiver distante do ponto de referência ou a linha que os conecta estiver em um ângulo elevado em relação ao eixo horizontal, isso indica que o modelo tem baixa precisão e correlação com os dados reais (Pereira *et al.*, 2014).

3.2.10 Resultados *predictors*

Os *predictors* (preditores) podem ser construídos para prever valores numéricos (regressão) com base em um conjunto de entradas. O modelo aprende a relação entre as entradas e saídas fornecidas e pode ser usado para prever a saída de novos exemplos de entrada. Posteriormente, esses modelos podem ser avaliados utilizando o diagrama de Taylor, que é construído a partir das previsões geradas pelo modelo e comparadas com as observações.

3.2.11 Aplicação R Shiny para análise dos modelos de *Machine Learning*

Para uma análise dos modelos foi desenvolvida uma aplicação interativa capaz de realizar a predição de modelos de aprendizado de máquina, incluindo RLM, KNN, SVR, XGB e RN para regressão. A aplicação é desenvolvida com uso do pacote shiny (Chang *et al.*, 2018), no ambiente R de computação estatística (R Core Team, 2017).

A aplicação ficará disponível no servidor da UNESP para toda a comunidade acadêmica e o algoritmo será disponibilizado no GitHub, que é uma plataforma de hospedagem e colaboração para desenvolvimento de software baseado em Git, um sistema de controle de versão distribuído. A plataforma é amplamente utilizada em projetos de *software* livre e de código aberto, bem como em projetos de universidade, empresas e organizações.

A seguir, serão apresentadas as telas da aplicação e descrito como foi desenvolvida a análise dos algoritmos utilizados neste trabalho. Essa aplicação foi dividida em cinco telas: início, upload, variáveis, dados e modelagem, entretanto, dentro de algumas telas existem sub telas.

Tela 1. Início

Existem três possibilidades de fazer *upload* de diferentes tipos de arquivos de dados, incluindo arquivos CSV, Excel e Colar, tornando a aplicação flexível e versátil. A opção “Passos *et al.* (2023)” carrega os dados referentes ao trabalho (Figura 6). O botão “Instruções de Uso” serve como um manual para orientar a utilização da aplicação e a organização correta dos dados na planilha. Já o botão “Download do trabalho” funciona como um link direto para o repositório da Unesp onde o trabalho estará disponível. Já na parte inferior da tela, é possível encontrar informações sobre a versão da aplicação, bem como detalhes sobre os autores do trabalho e contato do desenvolvedor da aplicação.

Machine Learning

Início

Banco de dados ▾

Resultados ▾

O banco de dados analisado será obtido via:

☐ Importação de arquivo xlsx

☐ Importação de arquivo csv

☐ Colar

☒ Tese - Passos et al. (2023)

INSTRUÇÕES DE USO

DOWNLOAD DA TESE

> AVANÇAR

PPG

Agronomia



Produção Vegetal

FCAV - UNESP

Jaboticabal

unesp

Tese | Danilo Passos

Figura 6. Tela inicial da aplicação – menu Início.

Tela 2. Upload

No caso de o usuário escolher a opção Excel ou CSV, ele poderá carregar o arquivo de dados e especificar o separador de valores e o tipo do decimal. Já para a escolha da opção “Colar”, aparecerá um campo para colar os valores. A opção “Tese - Passos *et al.* (2023)” carregou os dados referentes a esta pesquisa, isso sem precisar modificar a estrutura do banco de dados. Quando os dados forem carregados, eles aparecerão na tabela para que o usuário possa conferir as variáveis escolhidas e a quantidade de amostras do banco de dados (Figura 7).

Machine Learning
Início
Banco de dados ▾
Resultados ▾

Escolha o arquivo:

PROCURAR...
Dados não selecionados

Separador de valores

☐ Virgula (arquivo csv separado por virgula)
☐ Ponto e virgula (normalmente arquivos csv)
☐ Tabulação (normalmente quando você copia e cola)

Separador decimal

☐ Ponto
☐ Virgula

< VOLTAR
AVANÇAR >

Show 10 ▾ entries

Produtividade ▾	Cultivar ▾	N ▾	P ▾	K ▾	Ca ▾	Mg ▾	S ▾	B ▾	Cu ▾	Fe ▾	Mn ▾	Zn ▾	Tm_El_5d ▾
13015.7	Agata	33.6	5.2	53.7	17.7	4.8	3.1	19.4	5.9	810.8	563.8	75.5	17.002
15032.5	Agata	38.1	6.8	55.8	16.2	4.7	2.6	23.1	3.9	430.7	508	65.3	17.002
11119.4	Agata	38.2	6.5	50	16.4	5	3	21.1	5.9	886.8	528.6	69.2	17.002
13055.9	Agata	41.2	7.8	43.5	16.6	5.5	2.7	20.7	7.8	532.1	537.4	92	17.002
24861.8	Agata	50.3	5.6	35	14.3	6.4	2.8	25.7	7.8	988.2	490.4	62.6	17.002
19358.7	Agata	44.2	7.2	45.1	13.7	5.3	3.7	15.7	5.9	836.2	406.7	44.4	17.002
19375.7	Agata	46.3	7.6	36.2	10.8	5.4	3.3	15	3.9	278.7	487.4	62.6	17.002
21198.7	Agata	44.4	9	32.8	17	5.9	3.1	16	5.9	734.8	462.5	80.9	17.002
17885.4	Agata	49.4	5.6	37.8	14	8.6	2.8	24	7.8	591.2	472.8	59.2	17.002
17680.3	Agata	38.8	8.3	45.5	11.7	4.9	2.6	15	3.9	785.5	450.7	43.1	17.002

Showing 1 to 10 of 587 entries

Figura 7. Tela para carregar o arquivo de dados – menu Banco de dados > Upload.

Tela 3. Variáveis

No menu esquerdo é possível selecionar a variável dependente (variável resposta) e filtrar por grupo, caso seja necessário (Figura 8). Nesta tela seleciona-se quais são as variáveis independentes que devem compor o modelo, sendo possível selecionar as cinco ou as dez mais importantes de acordo com o modelo de *Machine Learning* RF, realizado na subtela “Seleção por Random forest”. A descrição de como o algoritmo de RF foi construído e de como ele seleciona o top 10, está descrito no item “Seleção de variáveis”. Para este estudo foram escolhidos três grupos de variáveis independentes: variáveis climáticas, variáveis nutricionais e top 10.

Machine Learning

Início

Banco de dados ▾

Resultados ▾

Selecione a variável resposta

Produtividade ▾

☐ Filtro por grupo

< VOLTAR AVANÇAR >

Variáveis do banco de dados [Seleção por Random forest](#)

Copy CSV Excel Print PDF

Variáveis	Pesos
QG_EV_95d	10113381501.8864
P.I_EIV_75d	5652186511.13355
P	3683910667.06736
Tmin_EIV_65d	3423503956.75753
N	2800369429.29644
QG_EIV_65d	2637685459.89301
Ca	2192215397.38874
Zn	1614539452.25917
K	1405578617.00326
Fe	1401039030.02195

Showing 1 to 10 of 10 entries

Previous 1 Next

Figura 8. Tela que seleciona a variável resposta, filtra e define as variáveis independentes – menu Banco de dados > variáveis > Variáveis do banco de dados.

Tela 4. Dados

Na tela dos dados é possível visualizar duas tabelas, uma somente com as variáveis selecionadas (Figura 9) e outra com as mesmas variáveis, entretanto, nessa os dados foram padronizados com a função nativa *scale* do R (Figura 10).

Tela 5. Modelagem

No menu esquerdo é possível selecionar quais são os modelos de *Machine Learning* que deseja utilizar para analisar os dados, existem cinco opções: Regressão linear múltipla, KNN regressão, Support vector machine, Redes neurais e eXtreme Gradient Boosting. Logo abaixo é possível definir a porcentagem dos dados que serão utilizados para treinamento e definir o número de vezes que a validação cruzada (*cross-validation*) deve ser feita (Figura 11).

Para realizar as análises, o banco de dados – 587 amostras – foi dividido em 75% para treinamento e 25% para teste. No período de treinamento a validação cruzada foi realizada com *k-fold* igual a 4. Na guia “resultados”, se pode visualizar os resultados do R^2 ajustado e do RMSE para o treinamento e o teste, e o Diagrama de Taylor para os cinco modelos (Figura 11).

Machine Learning Início Banco de dados ▾ Resultados ▾

< VOLTAR

AVANÇAR >

Tabela de dados Tabela de dados tratados

Copy

CSV

Excel

Print

PDF

Search:

	N ▾	P ▾	Ca ▾	Mg ▾	Fe ▾	K ▾	QG_EIV_75d ▾	Tmin_EIV_65d ▾	QG_EV_95d ▾	P.I_EIV_75d ▾
1	33.6	5.2	17.7	4.8	810.8	53.7	154.72	9.758	147.88	6.3
2	38.1	6.8	16.2	4.7	430.7	55.8	154.72	9.758	147.88	6.3
3	38.2	6.5	16.4	5	886.8	50	154.72	9.758	147.88	6.3
4	41.2	7.8	16.6	5.5	532.1	43.5	154.72	9.758	147.88	6.3
5	50.3	5.6	14.3	6.4	988.2	35	154.72	9.758	147.88	6.3
6	44.2	7.2	13.7	5.3	836.2	45.1	154.72	9.758	147.88	6.3
7	46.3	7.6	10.8	5.4	278.7	36.2	154.72	9.758	147.88	6.3
8	44.4	9	17	5.9	734.8	32.8	154.72	9.758	147.88	6.3
9	49.4	5.6	14	8.6	591.2	37.8	154.72	9.758	147.88	6.3
10	38.8	8.3	11.7	4.9	785.5	45.5	154.72	9.758	147.88	6.3

Showing 1 to 10 of 587 entries

Previous

1

2

3

4

5

...

59

Next

Figura 9. Tela que visualiza as variáveis independentes selecionadas com e sem padronização – menu Resultados > Dados > Tabela de dados.

Machine Learning Início Banco de dados ▾ Resultados ▾						
< VOLTAR AVANÇAR >						
Tabela de dados Tabela de dados tratados						
Copy CSV Excel Print PDF Search:						
N	P	Ca	Mg	Fe	K	
-2.48358323095642	0.657248161536453	0.988664952308785	-0.561392491637684	1.87907712176152	-0.010842503752214	
-1.81084051896042	1.68554646790833	0.597754049031635	-0.616078418055095	-0.0933226703328905	0.120708228288677	
-1.79589068091607	1.4927405354636	0.649875502801922	-0.452020638802861	2.2734532969764	-0.242622364967118	
-1.3473955395854	2.32823290939075	0.701996956572209	-0.178591006715805	0.43285817396696	-0.649803202236542	
0.0130397224509535	0.914322738129422	0.102600238213912	0.313582331040897	2.79963414127625	-1.18227045097348	
-0.898900398254735	1.9426210445013	-0.0537641230969484	-0.287962859550628	2.01088179084649	-0.54957407306253	
-0.584953799323269	2.19969562109427	-0.809525202766105	-0.233276933133216	-0.882075020762647	-1.10709860409297	
-0.869000722166024	3.09945663916966	0.806239864112782	0.0401526989538407	1.48470094654664	-1.32008550358775	
-0.121508819948246	0.914322738129422	0.0244180575584818	1.51667271222395	0.73953754179853	-1.00686947491896	
-1.70619165264994	2.64957613013196	-0.574978660799815	-0.506706565220273	1.74779136869657	-0.524516790769027	
Showing 1 to 10 of 587 entries Previous 1 2 3 4 5 ... 59 Next						

Figura 10. Tela que visualiza as variáveis independentes selecionadas com e sem padronização – menu Resultados > Dados > Tabela de dados tratados.

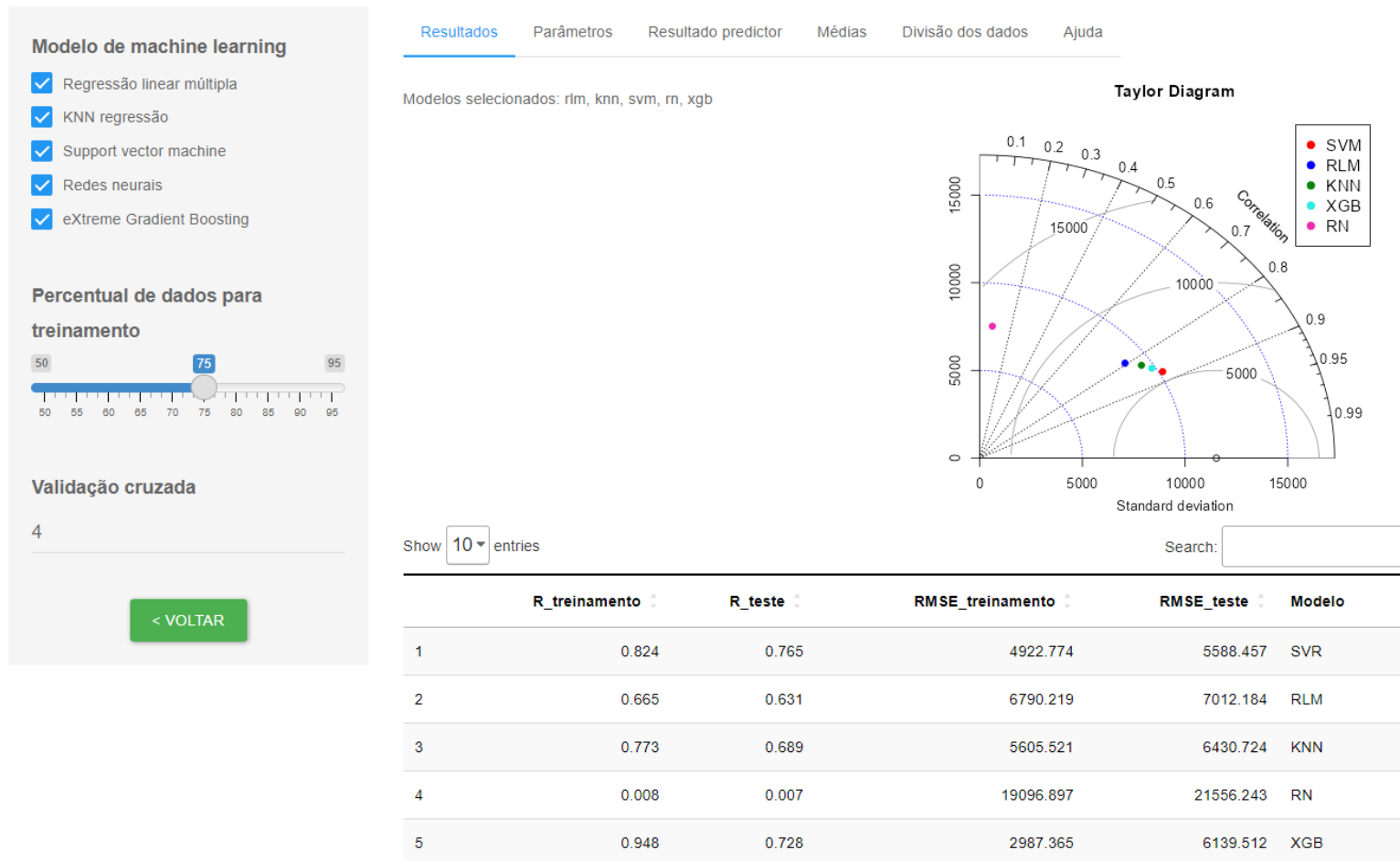


Figura 11. Tela que seleciona a variável resposta, filtra e define as variáveis independentes – menu Resultados > Modelagem > Resultados.

Em *Machine Learning* os parâmetros são as variáveis que o modelo de aprendizado de máquina utiliza para fazer previsões a partir dos dados. Os parâmetros são as configurações do modelo que não são aprendidas a partir dos dados, mas que precisam ser definidas pelo usuário antes do treinamento.

A escolha dos parâmetros pode ser feita através de tentativa e erro, ou utilizando técnicas mais sofisticadas, como a validação cruzada. O objetivo está em encontrar a melhor combinação de parâmetros que minimize a função de perda do modelo, obtendo um modelo com bom desempenho em dados novos. Na guia parâmetros, é possível visualizar os melhores parâmetros para cada modelo analisado (Figura 12), essa escolha foi realizada pelo TuneGrid.

Em seguida, ele usa a validação cruzada para testar todas as combinações possíveis desses valores e selecionar aqueles que fornecem o melhor desempenho do modelo (Figura 4). A metodologia está descrita no item “Treinamento e teste, parâmetros e validação cruzada”. Os resultados dos *predictors* de cada modelo podem ser visualizados e salvos para que posteriormente sejam utilizados no Diagrama de Taylor, comparando os 15 modelos (Figura 13). A metodologia está descrita no item “Avaliação de Modelos Preditivos”.

Machine Learning

Início

Banco de dados ▾

Resultados ▾

Modelo de machine learning

☒ Regressão linear múltipla

☒ KNN regressão

☒ Support vector machine

☒ Redes neurais

☒ eXtreme Gradient Boosting

Percentual de dados para treinamento

507595

50556065707580859095

Validação cruzada

4

< VOLTAR

Resultados

Parâmetros

Resultado predictor

Médias

Divisão dos dados

Ajuda

Melhores parâmetros encontrados pelo tuning:

Modelo: SVR

sigma C

10 0.11 1

Modelo: KNN

k

13 13

Modelo: RLM

intercept

1 -5

Modelo: Redes Neurais

size

10 9

Modelo: XGB

nrounds lambda alpha eta

11 10 0.5 0.01 0.01

Modelo: MLP

Figura 12. Tela que visualiza os parâmetros – menu Resultados > Modelagem > Resultados.

Machine Learning Início Banco de dados - Resultados -

Modelo de machine learning

- ☒ Regressão linear múltipla
- ☒ KNN regressão
- ☒ Support vector machine
- ☒ Redes neurais
- ☒ eXtreme Gradient Boosting

Percentual de dados para treinamento

50 75 95

50 55 60 65 70 75 80 85 90 95

Validação cruzada

4

< VOLTAR

Resultados Parâmetros **Resultado predictor** Médias Divisão dos dados Ajuda

Valores Predictors:

	Valor_real	SVM	RLM	KNN	RN	XGB
1	15032.5	21282.188	18653.049	20718.485	1.652388e+04	18074.836
2	11119.4	16829.361	15610.353	20948.823	1.652388e+04	17027.373
3	19358.7	21966.839	21143.870	22091.338	1.652388e+04	24013.693
4	21198.7	17830.008	20220.314	19877.454	1.652388e+04	16579.457
5	17394.2	20601.172	26554.691	20939.331	1.976895e+04	22917.607
6	17576.8	23493.609	28963.621	20979.638	1.976895e+04	21080.068
7	16404.2	19381.944	17784.611	19947.208	1.976895e+04	21286.076
8	23955.8	21010.712	20409.801	19655.969	1.976895e+04	21568.285
9	22626.0	21520.188	23052.587	20565.800	1.652388e+04	23103.094
10	19799.1	21034.833	21731.232	20385.423	1.976895e+04	21214.143
11	16413.3	13513.961	9112.491	12955.338	1.652388e+04	14523.229
12	15860.0	12319.914	10442.647	12529.985	1.652388e+04	10105.410
13	26499.9	18624.264	16124.352	14960.723	1.652388e+04	18359.412
14	12557.7	16482.091	14310.608	15153.031	1.652388e+04	19327.555
15	14520.0	14829.323	13580.474	14453.546	1.652388e+04	13085.521
16	22749.9	17298.631	11841.995	15413.115	1.652388e+04	11177.729
17	11613.3	16692.811	16778.508	19218.154	1.652388e+04	18413.818
18	14566.6	14740.343	19112.180	15099.385	1.652388e+04	18873.906
19	17241.7	15752.885	14720.295	17269.869	1.652388e+04	15688.937
20	27591.7	19803.196	21420.281	22541.023	1.652388e+04	19231.514
21	35658.3	25016.657	25076.979	23512.492	1.652388e+04	21052.633
22	16225.0	22467.150	23977.970	23920.523	1.652388e+04	20445.273
23	22208.3	24053.890	24671.181	24694.238	1.652388e+04	22692.691
24	22341.7	22025.610	21311.125	21308.331	1.652388e+04	20624.879

Figura 13. Tela que visualiza os valores dos *predictors* – menu Resultados > Modelagem > Resultado *predictor*.

Na guia de médias é possível visualizar dois gráficos de barras referentes ao R^2 ajustado e ao RMSE, com o auxílio dos gráficos verifica-se o desempenho dos modelos escolhidos (Figura 14). A metodologia está descrita no item “Avaliação de Modelos Preditivos”. Na guia de divisão dos dados, existem duas tabelas que apresentam a separação dos dados entre treinamento e teste. Além da exibição das amostras contidas em cada conjunto, assim pode ser visualizado o tamanho de cada um, que corresponde à porcentagem escolhida no momento da definição da divisão entre treinamento e teste (Figura 15).

Na guia de Ajuda, encontram-se as informações acerca do pacote Caret utilizado para construir os algoritmos, bem como a definição dos modelos disponíveis na aplicação. Além disso, é possível obter informações sobre o percentual adequados de treinamento e validação cruzada, além de informações sobre a definição dos parâmetros e análise dos modelos (Figura 16).



Figura 14. Tela que visualiza as médias do R^2 ajustado e do RMSE – menu Resultados > Modelagem > Médias.

Machine Learning
Início
Banco de dados
Resultados

Modelo de machine learning

- ☒ Regressão linear múltipla
- ☒ KNN regressão
- ☒ Support vector machine
- ☒ Redes neurais
- ☒ eXtreme Gradient Boosting

Percentual de dados para treinamento

50 75 95

Validação cruzada

4

< VOLTAR

Resultados
Parâmetros
Resultado predictor
Médias
Divisão dos dados
Ajuda

Treinamento
Teste

Show 10 entries
Search:

	Produtividade	N	P	Ca	Mg	Fe	K	QG_EIV_75d	Tmin_EIV_65d	QG_EV_95d	P.I_EIV_75d
1	13015.7	33.6	5.2	17.7	4.8	810.8	53.7	154.72	9.758	147.88	6.3
4	13055.9	41.2	7.8	16.6	5.5	532.1	43.5	154.72	9.758	147.88	6.3
5	24861.8	50.3	5.6	14.3	6.4	988.2	35	154.72	9.758	147.88	6.3
7	19375.7	46.3	7.6	10.8	5.4	278.7	36.2	154.72	9.758	147.88	6.3
9	17885.4	49.4	5.6	14	8.6	591.2	37.8	154.72	9.758	147.88	6.3
10	17680.3	38.8	8.3	11.7	4.9	785.5	45.5	154.72	9.758	147.88	6.3
11	16616.8	49.3	7.4	11.3	5.1	354.7	32.4	154.72	9.758	147.88	6.3
13	26934.7	53.2	8	10.4	4.9	658.8	36.2	154.72	9.758	147.88	6.3
14	22325.2	41.7	8.4	10.2	4.7	810.8	48.8	154.72	9.758	147.88	6.3
15	15828.1	45.1	7.9	13.7	5.8	278.7	53.1	154.72	9.758	147.88	6.3

Showing 1 to 10 of 440 entries
Previous
1
2
3
4
5
...
44
Next

Figura 15. Tela que visualiza as tabelas de dados para treinamento e teste – menu Resultados > Modelagem > Divisão dos dados.

Machine Learning

Início

Banco de dados ▾

Resultados ▾

Modelo de machine learning

☒ Regressão linear múltipla

☒ KNN regressão

☒ Support vector machine

☒ Redes neurais

☒ eXtreme Gradient Boosting

Percentual de dados para treinamento

50

75

95

50556065707580859095

Validação cruzada

4

< VOLTAR

Resultados

Parâmetros

Resultado predictor

Médias

Divisão dos dados

Ajuda

=====

Ajuda com modelos de machine learning

=====

----- Hiperparâmetros gerais -----

Validação cruzada (n_folds): número de vezes que um conjunto de dados é dividido em conjuntos de treinamento e validação durante a validação cruzada.

tuneGrid: define um conjunto de parâmetros (sigma e C) que serão testados durante a busca de hiperparâmetros.

Sigma: controla a largura do kernel Gaussiano usado para transformar os dados de entrada em um espaço de maior dimensionalidade.

C: define a importância relativa que o modelo dá para o ajuste correto aos dados de treinamento em relação à generalização para novos dados.

=====

Support Vector Regression (SVR)

=====

Figura 16. Tela que visualiza o conteúdo de ajuda – menu Resultados > Modelagem > Ajuda.

3.3 RESULTADOS E DISCUSSÃO

Durante a análise dos dados dos teores foliares dos nutrientes nos experimentos com doses crescentes de N, P e K, foram observados valores superiores em alguns nutrientes como: P, B, Cu, Fe e Zn (Figura 17).

É importante salientar que pontos discrepantes nos resultados não devem ser automaticamente considerados como *outliers*. Em experimentos agrícolas, a presença de variações naturais é uma realidade, e essas variações podem estar intrinsecamente ligadas a fatores específicos relacionados aos valores observados.

A variação dos teores nutricionais no solo é um fator relevante a ser considerado, pois diferentes regiões possuem características edafoclimáticas distintas, como composição química do solo, clima e manejo agrícola, que podem influenciar diretamente a disponibilidade de nutrientes para as plantas. Portanto, é natural que as análises de solo e as práticas de adubação adotadas em cada experimento tenham sido adaptadas às condições específicas de cada local.

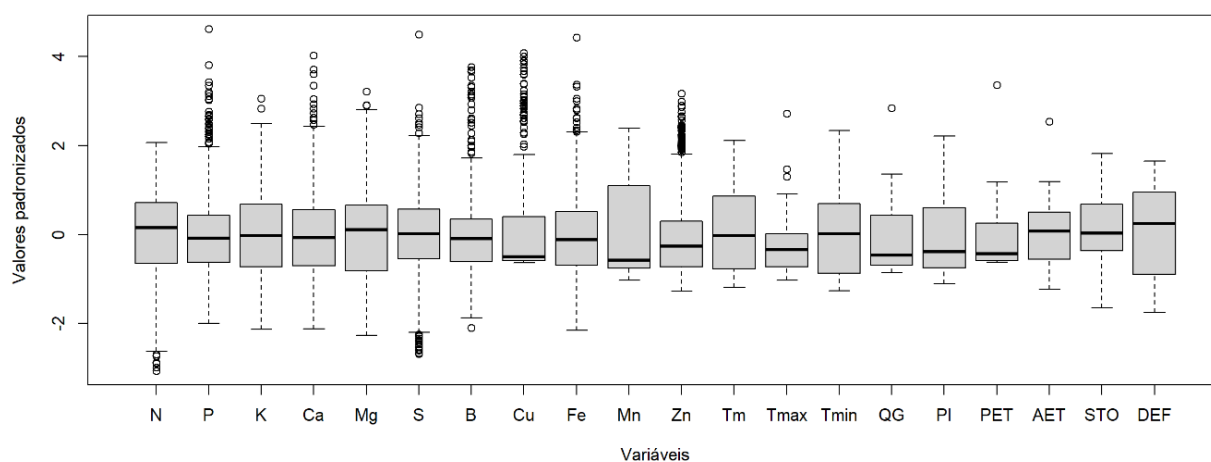


Figura 17. Boxplots referente as variáveis nutricionais e climática padronizados. Simbologia: N = nitrogênio, P = fósforo, K = potássio, Ca = cálcio, Mg = magnésio, S = enxofre, B = boro, Cu = cobre, Fe = ferro, Mn = manganês, Zn = zinco, Tm = temperatura média, Tmax = temperatura máxima, Tmin = temperatura mínima, QG = radiação global, PET = evapotranspiração, AET = evapotranspiração real, P+I = precipitação + irrigação, STO = armazenamento de água no solo, DEF = deficiência hídrica.

Desse modo, foram observados negativos nos nutrientes N e S, o que pode ser justificado já que os valores padronizados podem ser negativos, mas estes não

indicam correlação negativa, somente que os valores estão abaixo da média de zero. É importante considerar a interação entre N e S ao recomendar a adubação, pois a maioria do N está presente em proteínas, enquanto o S é um componente essencial para dois aminoácidos importantes, a cisteína e metionina. A deficiência de S pode reduzir a produção desses aminoácidos e as proteínas que os contêm não serão sintetizadas. O equilíbrio adequado entre as concentrações de N e S, no solo e na planta, é essencial para o crescimento e estado nutricional adequado das plantas (Marschner, 2012).

Os resultados sugerem que a adubação corretiva, realizada antes dos experimentos, pode ter sido eficaz para corrigir as deficiências nutricionais do solo. Entretanto, a presença de *outliers* indica a necessidade de mais estudos para avaliar a eficácia das doses de adubação utilizadas e a relação entre os teores foliares de nutrientes e a produtividade das plantas.

Esse distanciamento ou discrepância em relação aos resultados da adubação corretiva pode configurar um erro amostral, ou fazer parte da variabilidade do experimento. De toda forma, esse dado chama a atenção para a necessidade da busca de uma observação que coopere com mais precisão para redução das interferências relacionadas, nesse caso, com a adubação, permitindo um planejamento mais adequado para a produtividade do cultivo.

Os valores encontrados não são extremos, o que significa que não distorcem a análise estatística ou a interpretação dos resultados encontrados. Essa situação pode ser positiva, pois indica que os dados estão dentro do esperado e não há valores extremos distorcendo a análise estatística ou a interpretação dos resultados (Figura 18).

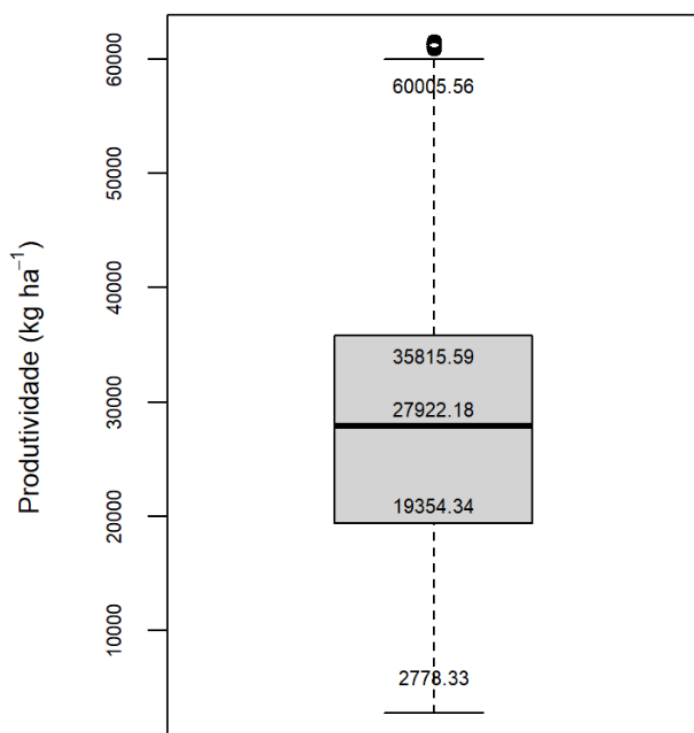


Figura 18. *Boxplot* da variável produtividade em relação a todos os experimentos.

Assim, foi observado que a produtividade apresentou poucos *outliers*, sendo que a mediana dos valores registrados foi de 27922 kg ha⁻¹. A produtividade mínima e máxima registrada foi de 2778 e 61516 kg ha⁻¹, respectivamente. Os intervalos interquartis de 25% a 75% da distribuição foram de 19354 e 35815 kg ha⁻¹, respectivamente. Esses intervalos representam a variação dos valores observados. O primeiro quartil corresponde aos 25% dos valores abaixo do número indicado, enquanto o terceiro quartil representa os 75% abaixo dele.

A produtividade apresentou uma distribuição relativamente homogênea, com poucos pontos discrepantes em relação aos demais valores e sem anomalias, sugerindo que a produtividade foi consistente e estável ao longo do período observado. O que não afeta a interpretação dos resultados. A identificação de *outliers*, em uma análise estatística descritiva através do gráfico *boxplot*, nem sempre modifica a confiabilidade da análise dos dados, uma vez que outros modelos foram utilizados para supervisionar as variáveis (Jonathan *et al.*, 2015).

Ao identificar discrepâncias, podem ser feitos ajustes para corrigir os resultados e torná-los mais confiáveis. Assim, será possível proporcionar maior qualidade e

precisão dos dados, colaborando com a tomada de decisão do agricultor, evitando excesso e/ou déficit nutricionais no solo. O que permite um melhor manejo ambiental, redução nos gastos e aumento da produtividade da cultura.

3.3.1 Heatmap de correlação

O *heatmap* foi criado com base na produtividade e nas variáveis nutricionais e climáticas voltada para o cultivo da batata (Figura 19). As cores azuis e vermelhas indicam, respectivamente, correlações positivas e negativas entre as variáveis, o tamanho dos círculos representa a mesma informação - quanto maior o diâmetro, maior é a correlação.

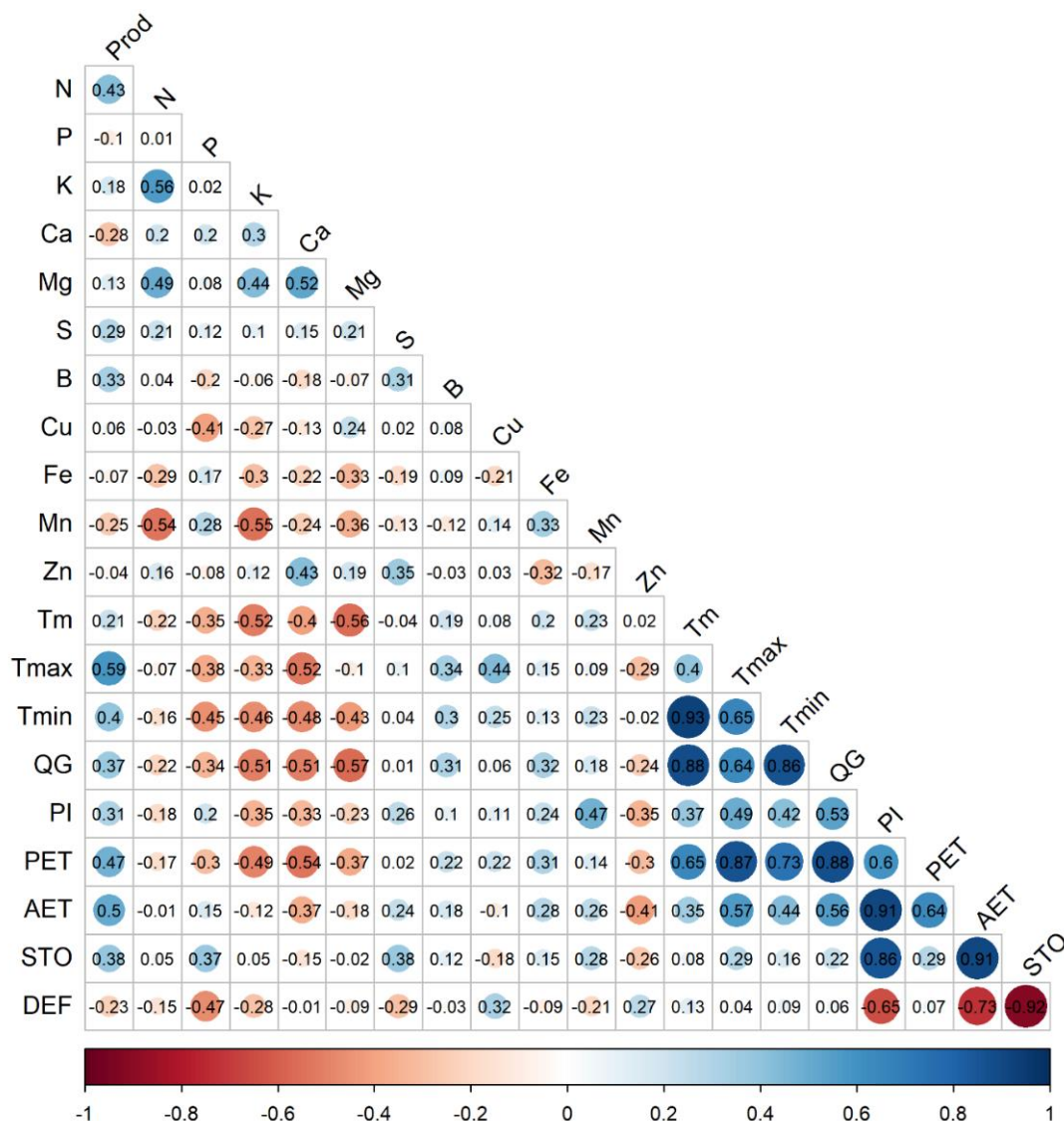


Figura 19. Heatmap da correlação entre a produtividade e as variáveis nutricionais e climáticas.

Simbologia: N = nitrogênio, P = fósforo, K = potássio, Ca = cálcio, Mg = magnésio, S = enxofre, B = boro, Cu = cobre, Fe = ferro, Mn = manganês, Zn = zinco, Tm = temperatura média, Tmax = temperatura máxima, Tmin = temperatura mínima, QG = radiação global, PET = evapotranspiração, AET = evapotranspiração real, P+I = precipitação + irrigação, STO = armazenamento de água no solo, DEF = deficiência hídrica.

Na figura 19, pode-se verificar que a produtividade apresenta alta correlação (superior a 40%) com o N, TMáx, Tmín, AET e PET. Conforme visualizado, a correlação entre produtividade, as variantes nutricionais e climáticas podem subsidiar a adoção de calendários fixos de irrigação, viabilizando o manejo das variantes apropriados para cada etapa do cultivo. A correlação, entre a produtividade e as variáveis, sugere que elas possuem forte influência sobre a produção agrícola, uma vez que as condições ambientais (solo, clima, adubação e outros) afetam a composição química da batata e o seu potencial de mercado (Quadros *et al.*, 2009).

Desse modo, para executar um manejo eficiente o produtor precisa estar atendo ao início da formação dos tubérculos até próximo ao final do ciclo, pois é fase crítica para a cultura onde os nutrientes são exigidos em altas quantidades. Na fase de enchimento dos tubérculos, a disponibilidade de nutrientes e água no solo deve ser elevada, pois é o período de acumulação da matéria seca e a absorção de nutrientes ocorre de forma rápida. Além disso, as temperaturas máxima e mínima podem influenciar o desenvolvimento e a maturação da cultura.

A AET e a PET são importantes para compreender a demanda hídrica da cultura, contribuindo para o planejamento da irrigação, tendo em vista que a batata é uma cultura sensível ao déficit e ao excesso de água. O ápice da demanda hídrica ocorre no início da tuberização, quando a matéria seca produzida pela fotossíntese passa a ser direcionada, quase que totalmente, para promover o crescimento dos tubérculos (Marouelli; Braga; Guimarães, 2013).

Nesse cenário, os dados provenientes podem auxiliar na calibragem das necessidades dos fertilizantes e água, corrigindo as potenciais deficiências, ainda, durante o ciclo de crescimento do plantio. Dessa forma, há possibilidade e viabilidade de melhorar os níveis do potencial produtivo do cultivo, bem como melhorar a qualidade e a uniformidade no tamanho dos tubérculos, favorecendo uma maior rentabilidade final do produto.

3.3.2 Clusterização aglomerativa

O processo de construção do dendrograma envolve a formação de agrupamentos hierárquicos. Inicialmente, cada variável é considerada um agrupamento individual, em seguida, os agrupamentos mais similares são fundidos, formando novos agrupamentos. Esse processo é repetido iterativamente até que reste apenas um único agrupamento, representando a similaridade total entre as variáveis (Kassambara, 2017).

O dendrograma (Figura 20) permite visualizar a estrutura de similaridade entre as variáveis, auxiliando na identificação de padrões e relações entre elas. As ramificações mais curtas (P e K; Mg e S; B e Cu; Mn e Zn; Tmax e Tmin; QG e PI; PET e AET; STO e DEF) correspondem à identificação de dois agrupamentos mais similares entre si. Esses dois agrupamentos são fundidos para formar um novo agrupamento, resultando em uma redução de um agrupamento no total. Esse processo de ramificação é repetido iterativamente até que reste apenas um único agrupamento (Kassambara, 2017). O eixo y representa a distância de Pearson, uma medida de similaridade entre as variáveis. Quanto mais próximo de zero, maior a similaridade entre as variáveis.

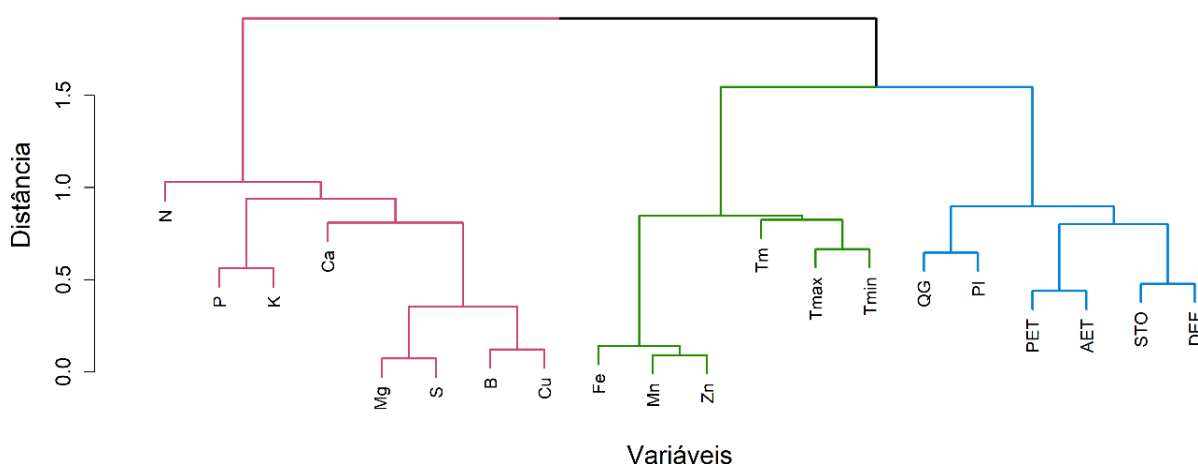


Figura 20. Dendrograma representando a similaridade entre variáveis com base na distância de *Pearson*.

Simbologia: N = nitrogênio, P = fósforo, K = potássio, Ca = cálcio, Mg = magnésio, S = enxofre, B = boro, Cu = cobre, Fe = ferro, Mn = manganês, Zn = zinco, Tm = temperatura média, Tmax = temperatura máxima, Tmin = temperatura mínima, QG = radiação global, PET = evapotranspiração, AET = evapotranspiração real, P+I =

precipitação + irrigação, STO = armazenamento de água no solo, DEF = deficiência hídrica.

À medida que as ramificações se unem, formam-se agrupamentos maiores que são menos similares entre si. No final do dendrograma, resta apenas um agrupamento único, representando todas as variáveis em conjunto. No dendrograma, é possível identificar a formação de três grandes grupos. O grupo vermelho consiste exclusivamente em variáveis nutricionais, incluindo todos os macronutrientes e os nutrientes B e Cu. O grupo verde no dendrograma é composto por variáveis relacionadas à temperatura e aos micronutrientes Fe (ferro), Mn (manganês) e Zn (zinco). Já o grupo azul é constituído exclusivamente por variáveis climáticas, com destaque para aquelas relacionadas à disponibilidade de água no solo e à evapotranspiração.

Observa-se que essas variáveis climáticas estão mais próximas umas das outras no dendrograma. A proximidade dessas variáveis climáticas indica uma relação intrínseca entre a disponibilidade de água no solo e o processo de evapotranspiração. O dendrograma e a distância de Pearson se formam com base nas diferentes variáveis do grupo vermelho, verde e azul. A formação obtida na relação entre o dendrograma e a distância de Pearson resultam em um processo de agrupamento das variáveis que estão representadas, por sua vez, pela medida de distância ou de semelhança.

A evapotranspiração, é o processo combinado de evaporação da água do solo e transpiração das plantas, ela descreve a quantidade de água que é perdida para a atmosfera durante o ciclo hidrológico. Desse modo, ela está intimamente relacionada à disponibilidade de água no solo, uma vez que a quantidade de água disponível afeta diretamente a taxa de evapotranspiração.

De acordo com um estudo conduzido por Tindall *et al.* (1990), que avaliou a absorção de nutrientes pela planta de tomate sob diferentes temperaturas na zona radicular (10; 15,6; 21,1; 26,7; 32,2; e 37,8 °C) em uma câmara de crescimento, foi constatado que a temperatura influenciou a disponibilidade desses nutrientes no solo. Isso resultou em dificuldades na absorção desses nutrientes pelas raízes do tomateiro. É importante ressaltar que a relação entre temperatura e nutrientes pode variar dependendo da espécie de planta, do tipo de solo e das condições ambientais específicas.

3.3.3 Análise de componentes principais

A análise de componentes principais (PCA, do inglês Principal Component Analysis) é uma técnica de redução de dimensionalidade que busca encontrar as principais combinações lineares das variáveis originais que melhor explicam a variabilidade dos dados (Kherif e Latypova *et al.*, 2020). A análise exploratória de dados é uma importante ferramenta utilizada em diversas áreas da agricultura.

Nesse sentido, técnicas como *boxplot*, PCA e *heatmap* de correlação podem fornecer informações valiosas sobre a estrutura e as relações entre as variáveis em um conjunto de dados. O *boxplot* pode ser usado para avaliar a distribuição dos dados e identificar *outliers*. A PCA pode ser usada para reduzir a dimensionalidade dos dados e identificar as principais fontes de variabilidade. O *heatmap* de correlação pode ser usado para identificar relações e padrões específicos entre as variáveis. Juntas, essas técnicas podem ajudar a compreender melhor os dados e fornecer insights para análises posteriores, tornando a modelagem de *Machine Learning* mais precisa e eficiente.

O primeiro componente principal (Figura 21) explicou 32.1% da variância total dos dados originais, enquanto o segundo componente principal explicou 17.6% da variância total. Juntos, eles explicam um total de 49.7% da variância dos dados originais. Essas porcentagens correspondem aos níveis de contribuição dos componentes principais para a batata, pois tais componentes são ativos e quando expostos às condições nutricionais e climáticas adequadas contribuem para a produtividade. No entanto, é válido considerar, ainda, que em âmbito visual, é considerado a uniformidade de tamanho e quantidade de tubérculos, pois os tubérculos medianos possuem um coeficiente baixo de variância, isso quando testados (LEPRE, 2009). Logo, se as variáveis nutricionais e climáticas possuem baixo teor de componentes principais, isso afeta a boa produção e venda da batata no mercado.

A dormência e a dormência apical são fatores essenciais para os tubérculos de batata e sua comercialização. A quebra da dormência resulta na perda da oportunidade de manter a qualidade dos tubérculos, uma vez que eles passam a competir no espaço do solo devido ao aumento no número de brotos (BISOGNIN *et*

al., 2008). Muller *et al.* (2010) corroboram ao enfatizar que a brotação dos tubérculos leva a perdas qualitativas e quantitativas devido ao aumento da atividade metabólica.

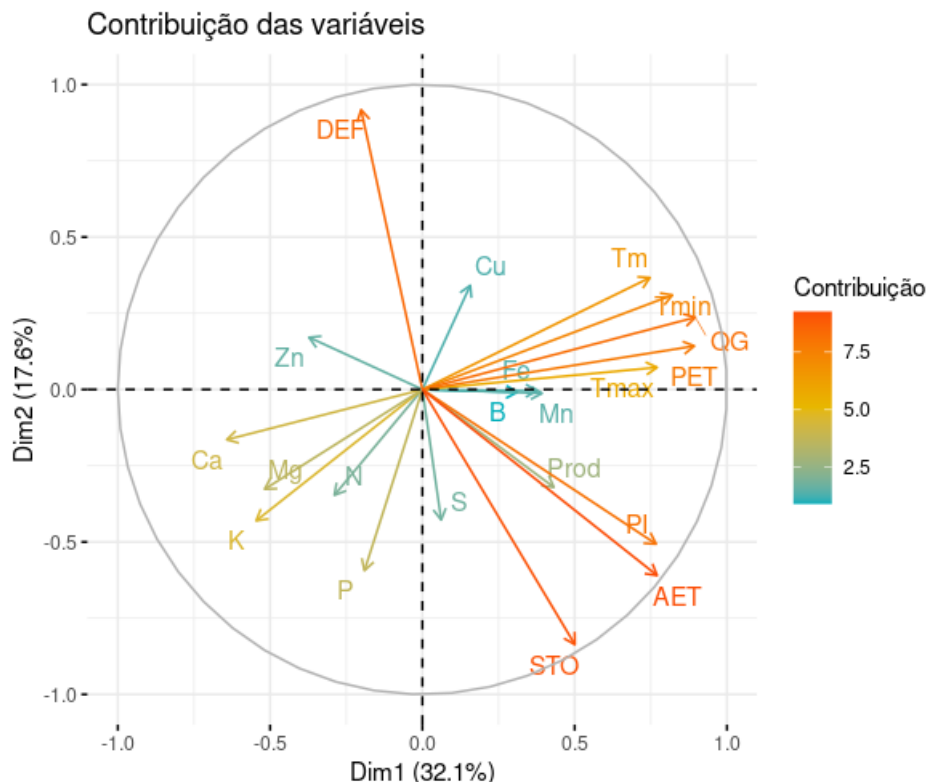


Figura 21. Análise de componentes principais entre a produtividade e as variáveis nutricionais e climáticas.

Simbologia: N = nitrogênio, P = fósforo, K = potássio, Ca = cálcio, Mg = magnésio, S = enxofre, B = boro, Cu = cobre, Fe = ferro, Mn = manganês, Zn = zinco, Tm = temperatura média, Tmax = temperatura máxima, Tmin = temperatura mínima, QG = radiação global, PET = evapotranspiração, AET = evapotranspiração real, P+I = precipitação + irrigação, STO = armazenamento de água no solo, DEF = deficiência hídrica.

O eixo 1 (Dim1) é composto pelas variáveis dos nutrientes e operacionalização da absorção de nutrientes (relacionado à evapotranspiração), já o eixo 2 (Dim2) está relacionado aos componentes de armazenamento de água no solo (STO) e DEF. De acordo com os parâmetros de análise de Kherif e Latypova *et al.* (2020), conforme visualizado, o primeiro componente principal obteve um maior índice de contribuição, isso devido às variáveis climáticas DEF, AET e STO. Já o segundo componente principal estava relacionando as variáveis com um maior nível de correlação, predispostas entre PET, Mn e B.

Nesse sentido, as flechas próximas tendem a ter uma relação positiva entre si, ou seja, quando uma variável aumenta, a outra também tende a aumentar. Por outro lado, flechas opostas tendem a ter uma relação negativa entre si, ou seja, quando uma variável aumenta, a outra tende a diminuir. As flechas relacionadas aos macronutrientes estão todas próximas na análise de componentes principais, isso pode indicar que essas variáveis estão altamente correlacionadas entre si.

Essa correlação pode ser um indicativo que quando alta as cultivares de temporalidade mais tardia terão um maior potencial produtivo. Por outro lado, variáveis com alta correlação negativa, sugere-se uma possível suscetibilidade do genótipo a chamada murcha-bacteriana, fator que incide diretamente na produtividade da batata (Pereira, 2018).

Além desses fatores, no estudo de Haverkort e Struik (2015) foi avaliada a influência das condições meteorológicas sobre o desenvolvimento da cultura da batata. Observou-se que a temperatura do ar pode afetar diferentes aspectos do desenvolvimento da cultura, tais como o crescimento do dossel, o volume dos tubérculos e a duração do ciclo de crescimento. Timlin *et al.* (2006) destacaram o papel da radiação como fator condicionante da produtividade total da batata. A quantidade de matéria seca produzida por unidade de luz interceptada foi de aproximadamente $2,62 \text{ g m}^{-2} \text{ MJ}^{-1}$ para a testemunha, com a adição da radiação fotossinteticamente ativa (PAR) a quantidade de matéria seca foi de $10,2 \text{ g m}^{-2} \text{ MJ}^{-1}$, assumindo 12 plantas por m^2 em câmara de crescimento de plantas.

3.3.4 Random Forest

O uso de modelos de *Machine Learning* tem se mostrado uma ferramenta promissora na previsão da produtividade agrícola. Dentre os modelos disponíveis, o Random Forest tem sido amplamente utilizado devido à sua eficiência em lidar com conjuntos de dados complexos e heterogêneos. Entretanto, a configuração adequada dos parâmetros é crucial para obter um bom desempenho do modelo. Além disso, a seleção das variáveis mais importantes para a previsão da produtividade é outro aspecto relevante a ser considerado.

Nesse contexto, o presente estudo investigou o desempenho de um modelo de Random Forest na previsão da produtividade da cultura de batata, identificando as 10

principais variáveis que influenciam a produtividade. Após a aplicação do *tunegrid*, os melhores parâmetros foram encontrados: *ntree* com valor de 200 e *mtry* com valor de 90, posteriormente, esses valores foram utilizados no modelo de Random Forest.

O modelo fez a seleção das 10 principais variáveis que influenciam na produtividade. O top 10 variáveis é composto por: N, P, K, Ca, Mg, Fe, P+I_EIV_75d, Tmin_EIV_65d, QG_EIV_65d, QG_EV_95d (Figura 4). Para este modelo, o R^2 ajustado e o RMSE foram de 0,82 e 5260,39, respectivamente. Isso significa que o modelo ajustado explicou cerca de 82% da variação dos dados, e o RMSE indica que os erros de predição apresentaram uma média de 5260,39 kg ha⁻¹.

No estudo conduzido por Kuradusenge *et al.* (2023), foram empregados três modelos de *Machine Learning*: Random Forest Regressor, Regressor Polinomial e Regressor de Vetor de Suporte, com o objetivo de prever a produtividade de batatas na Irlanda. A avaliação do desempenho dos modelos foi realizada utilizando medidas de acurácia, tais como R^2 e RMSE, além da técnica de validação cruzada k-fold com $k = 5$. Os resultados obtidos demonstraram que o modelo Random Forest Regressor apresentou o melhor desempenho, com um valor de R^2 igual a 0,875 e um valor de RMSE igual a 510,817 kg ha⁻¹.

Assim, como em outras pesquisas que compararam diferentes modelos de *Machine Learning*, tais como as realizadas por Devaux *et al.* (2014), Raimundo *et al.* (2014) e Fleisher *et al.* (2017). O Random Forest (RF) foi selecionado para realizar a identificação das variáveis mais relevantes em relação à previsão de produtividade agrícola. Tal escolha se deve ao fato de que o RF é amplamente utilizado como modelo de *Machine Learning* em diversas áreas, incluindo a previsão de produtividade agrícola.

Ao utilizar o RF para a seleção das variáveis mais importantes, espera-se obter um melhor entendimento das relações entre as variáveis envolvidas na produção agrícola, permitindo uma análise mais precisa dos fatores que afetam a produtividade das culturas. Dessa forma, os resultados obtidos podem ser utilizados para aprimorar as estratégias de manejo da produção agrícola, contribuindo para a obtenção de cultivos mais produtivos e sustentáveis.

3.3.5 Resultados dos parâmetros

A Tabela 3 apresenta os resultados dos parâmetros ajustados para os modelos nos grupos 1, 2, 3 e 4, correspondentes aos 4 *folds* utilizados na análise.

Tabela 3. Valores ajustados para os parâmetros.

Modelo	Parâmetro	Valores ajustados			
		Grupo 1	Grupo 2	Grupo 3	Grupo 4
RF	<i>Ntree</i>	200	200	200	200
	<i>Mtry</i>	90	90	90	90
RLM	<i>Intercept</i>	-5	-5	-5	-5
KNN	<i>K</i>	13	13	12	13
SVR	<i>Sigma</i>	0.11	0.11	0.11	0.11
	<i>C</i>	1	1	1	1
XGB	<i>Nrounds</i>	10	10	10	10
	<i>Lambda</i>	0.5	0.5	0.5	0.5
	<i>Alpha</i>	0.2	0.3	0.2	0.2
	<i>Eta</i>	0.1	0.1	0.1	0.1
RN	<i>Sigma</i>	9	8	10	9

No modelo RF, todos os grupos utilizaram 200 árvores (*ntree*) e 90 variáveis selecionadas aleatoriamente (*mtry*). Para o modelo RLM, o valor do intercepto foi mantido constante em -5 em todos os grupos. No caso do modelo KNN, o número de vizinhos mais próximos (*k*) variou ligeiramente entre os grupos, sendo 13 no grupo 1, 2 e 4, e 12 no grupo 3.

Os modelos SVR mantiveram os valores de parâmetros em todos os grupos, com um valor de 0.11 para *sigma* e 1 para *C*. Para o modelo XGB, todos os grupos utilizaram 10 rodadas (*nrounds*) e um valor de 0.1 para *eta*. No entanto, o parâmetro *lambda* variou ligeiramente, sendo 0.5 nos grupos 1, 3 e 4, mas 0.3 no grupo 2. Além disso, o valor do parâmetro *alpha* foi de 0.2 nos grupos 1, 3 e 4, enquanto no grupo 2 foi ajustado para 0.3.

Por fim, para o modelo RN, o valor de *sigma* variou ligeiramente entre os grupos, sendo 9 no grupo 1, 8 no grupo 2, 10 no grupo 3 e retornando para 9 no grupo 4. Esses resultados evidenciam as configurações dos parâmetros ajustados para cada modelo nos diferentes grupos, fornecendo informações relevantes para a compreensão das escolhas feitas na análise estatística e modelagem dos dados.

Desse modo, pode ser aferido nessa pesquisa que RF e apresentaram alto índice de similaridade. Ao contrário do estudo realizado Jeong *et al.* (2016), onde os

autores apontam resultados entre RF e RLM discrepantes. A pesquisa foi realizada com os no cultivo de batata, trigo e milho, onde o modelo RF se sobrepôs ao modelo RLM, isso pela existência da possibilidade de os resultados serem tangencialmente mais acurados, uma vez que eles apresentaram baixos valores de RMSE.

3.3.6 Resultados dos modelos de *Machine Learning*

Existem diversas classificações para modelos de *Machine Learning*, e, nesse sentido, optou-se por empregar algoritmos de regressão (RLM e SVR), baseados em instância (KNN), aprendizado conjunto (XGB) e aprendizado profundo (RN), a fim de contemplar algumas dessas categorias. No trabalho de revisão realizado por Sharma *et al.* (2020), foram selecionados 173 artigos científicos sobre o uso de *Machine Learning* na agricultura, publicados entre os anos de 2002 a 2019. A partir dessa revisão, foi elaborado um ranking com os nove principais tipos de algoritmos de *Machine Learning* empregados nesse contexto e os 5 tipos de algoritmos desse estudo estão nesse ranking.

Os modelos de *Machine Learning* selecionados para o presente trabalho foram escolhidos com base na sua adequação para a análise das principais variáveis nutricionais e climáticas que influenciam na produtividade da batata. As variáveis utilizadas nos modelos de *Machine Learning* foram divididas em três grupos distintos, de acordo com sua natureza e relevância para a análise em questão. No primeiro grupo, foram incluídas somente as variáveis nutricionais, sendo elas: N, P, K, Ca, Mg, S, B, Fe, Mn e Zn. Essas variáveis são de extrema importância para a produtividade da cultura da batata e, por isso, foram incluídas nesse grupo.

O segundo grupo de variáveis incluiu as variáveis climáticas, sendo elas: Tm, Tmax, Tmin, QG PET, AET, P+I, STO e DEF. Cada uma dessas variáveis foi dividida em subvariáveis a cada cinco dias, de acordo com o estágio fenológico da cultura (Figura 22). Esse grupo de variáveis climáticas totalizou 28 subvariáveis para cada variável climática considerada. Por fim, o terceiro grupo de variáveis incluiu as 10 principais variáveis que influenciam na produtividade da batata, de acordo com o modelo de random forest. Essas variáveis foram identificadas como as mais relevantes para a análise em questão e são: N, P, K, Ca, Mg, Fe, P+I_EIV_75d, Tmin_EIV_65d, QG_EIV_65d e QG_EV_95d.

A divisão dos dados em grupos distintos, de acordo com a sua natureza e relevância para a análise em questão, permite uma análise mais aprofundada do banco de dados disponível. Assim, comparamos 15 modelos com diferentes opções de pré-processamento para avaliar o desempenho de 5 algoritmos de aprendizado de máquina. A técnica de validação cruzada k-fold foi usada para avaliar o desempenho preditivo dos algoritmos. Em geral, a maioria dos modelos propostos obteve valores de RMSE $< 7.5 \text{ t ha}^{-1}$, o que representa cerca de 25% de erro em relação ao rendimento médio de produtividade.

No contexto agrônomo, a relação percentual de erro com significância em relação à produtividade pode estar diretamente ligada a demanda de nutrientes do solo. De acordo com Borchart *et al.* (2011), uma maneira de aumentar a produtividade é através do manejo adequado de técnicas de plantio e adubação que contribuam para melhorar a produção de tubérculos de batata, utilizando fertilizantes de origem mineral ou orgânica, desde que atendam aos requisitos da cultura.

Neste estudo, ao aplicar os modelos de *Machine Learning* para a cultura da batata, observou-se, dentro do contexto analítico adotado, que os valores de RMSE resultaram em uma taxa de acerto de aproximadamente 75% (Figura 22).

Variáveis nutricionais	Variáveis climáticas					Top 10	Resposta
	Estadio I	Estadio II	Estadio III	Estadio IV	Estadio V		
N						N	Produtividade
P	Tm ₁	Tm ₂	Tm ₃	Tm ₄	Tm ₅	P	
K	Tmax ₁	Tmax ₂	Tmax ₃	Tmax ₄	Tmax ₅	K	
Ca	Tmin ₁	Tmin ₂	Tmin ₃	Tmin ₄	Tmin ₅	Ca	
Mg	QG ₁	QG ₂	QG ₃	QG ₄	QG ₅	Mg	
S	PET ₁	PET ₂	PET ₃	PET ₄	PET ₅	Fe	
B	AET ₁	AET ₂	AET ₃	AET ₄	AET ₅	Tmin_FT65	
Cu	P+I ₁	P+I ₂	P+I ₃	P+I ₄	P+I ₅	QG_FT75	
Fe	STO ₁	STO ₂	STO ₃	STO ₄	STO ₅	QG_SM95	
Mn	DEF ₁	DEF ₂	DEF ₃	DEF ₄	DEF ₅	P+I_FT75	
Zn	EXC ₁	EXC ₂	EXC ₃	EXC ₄	EXC ₅		

Figura 22. Grupos das variáveis que foram selecionadas para as análises e a variável resposta.

Simbologia: N = nitrogênio, P = fósforo, K = potássio, Ca = cálcio, Mg = magnésio, S = enxofre, B = boro, Cu = cobre, Fe = ferro, Mn = manganês, Zn = zinco, Tm = temperatura média, Tmax = temperatura máxima, Tmin = temperatura mínima, QG = radiação global, PET = evapotranspiração, AET = evapotranspiração real, P+I =

precipitação + irrigação, STO = armazenamento de água no solo, DEF = deficiência hídrica. Os números sobrescritos referem-se ao estágio fenológico.

Os resultados de desempenho de diferentes modelos de *Machine Learning* para previsão de variáveis nutricionais (Tabela 4), avaliados por meio dos coeficientes de determinação ajustados (R^2 ajustado) e da raiz do erro quadrático médio (RMSE) para o treinamento e teste dos modelos. É importante mencionar, que a discussão das tabelas seguirá mediante a apresentação de dados, tomando como referência a literatura vigente sobre cada modelo.

Tabela 4. R^2 ajustado e RMSE (raiz do erro quadrático médio) do treinamento e teste das variáveis nutricionais para os modelos de *Machine Learning*.

Modelo	R_a^2 treinamento	R_a^2 teste	RMSE (kg ha ⁻¹) treinamento	RMSE (kg ha ⁻¹) teste
RLM	0.43	0.42	8833.85	9008.22
KNN	0.82	0.75	4971.75	5928.08
SVR	0.90	0.74	3776.71	6088.82
XGB	0.99	0.68	1051.94	6659.37
RN	0.07	0.05	19140.22	19618.32

RLM: Regressão Linear Múltipla; KNN: k-Nearest Neighbors; SVR: Support Vector Regression; XGB: Extreme Gradient Boosting; RN: Rede Neural.

Os resultados mostraram que o modelo RLM obteve um R^2 ajustado de 0,43 no treinamento e 0,42 no teste, com um RMSE de 8833,85 no treinamento e 9008,22 no teste. O modelo RLM possui uma melhor precisão quando utiliza dados relacionados à localização geográfica de incidências climáticas, bem como as condições climáticas dessas mesmas incidências.

Corroborando Wrege *et al.* (2004), afirmam que o fator geográfico interfere na questão do clima, o que está intrinsecamente relacionado ao melhor cultivo da batata. Ainda, observaram a latitude e longitude na intenção de obter dados referentes às temperaturas máximas e mínimas para o uso do modelo. Desse modo, as regiões produtoras de batata estão predominantemente localizadas em áreas de altitude elevada, com poucas exceções.

O modelo KNN apresentou um R^2 ajustado de 0,82 no treinamento e 0,75 no teste, com um RMSE de 4971,75 no treinamento e 5928,08 no teste. Em uma pesquisa publicada recentemente sobre a agricultura no Canadá, Halal (2023) aponta o modelo KNN como sendo o mais eficaz para a análise de conjunto de dados

menores. Entretanto, o autor utilizou variáveis diferentes das utilizadas na presente pesquisa. Assim, estudo serve apenas como parâmetro ampliador da discussão, e não como norteador. No que tange o manejo da batata, tomando como base RMSE e R^2 , Farooque *et al.* (2020) enfatiza que possíveis condutividades elétricas no solo podem contribuir para uma correlação com a produtividade da batata, também no recorte geográfico do Canadá.

O modelo SVR obteve um R^2 ajustado de 0,90 no treinamento e 0,74 no teste, com um RMSE de 3776,71 no treinamento e 6088,82 no teste. Com base nos dados apresentados, é evidente que a variação dos dados entre o conjunto de treinamento e o conjunto de teste não apresenta uma proporção que torne o modelo explanatório viável. Os valores observados são discrepantes e apresentam pouca proximidade.

Vale ressaltar, que conjuntos de dados maiores têm o potencial de gerar resultados mais precisos ao utilizar qualquer um dos modelos. Já o modelo XGB alcançou o melhor desempenho, com um R^2 ajustado de 0,99 no treinamento e 0,68 no teste, e um RMSE de 1051,94 no treinamento e 6659,37 no teste. Por fim, o modelo de Rede Neural apresentou um desempenho inferior, com um R^2 ajustado de apenas 0,07 no treinamento e 0,05 no teste, e um RMSE de 19140,22 no treinamento e 19618,32 no teste. O modelo KNN apresentou o melhor desempenho, com um R^2 ajustado 0,75 no teste, e RMSE de 5928,08 no teste.

Diante de todos os modelos analisados e os resultados obtidos nos treinamentos e teste, com base em R^2 ajustado e RMSE, teremos que o modelo KNN foi capaz de explicar uma maior proporção da variância nos dados, apresentando menor erro médio de previsão tanto no treinamento quanto no teste. Os modelos RLM, SVR e XGB também apresentaram bom desempenho, com R^2 ajustado e RMSE intermediários. Já o modelo de RN teve desempenho insatisfatório em ambos os critérios avaliados. Assim, considera-se que o desempenho mais adequado e explicativo com relação à batata foi o modelo KNN, conforme já explicitado anteriormente.

Durante o treinamento, o modelo KNN buscou as observações mais próximas (ou os "vizinhos") do novo ponto a ser previsto com base nas quantidades de nutrientes nas plantas, estimando a probabilidade de alta produtividade, com base nas quantidades de nutrientes dos vizinhos mais próximos. Com isso, o modelo pode ser

utilizado para identificar quais nutrientes estão mais fortemente associados a altas produtividades.

Corroborando com a perspectiva defendida neste trabalho, onde o modelo KNN, para variáveis climáticas, melhor se destacou com relação à cultura da batata, Garofalo *et al.* (2015) reforçam a relevância de trabalhos que considerem as técnicas de classificação não paramétricas, os chamados “vetores de suporte”. Assim, os modelos SVM e KNN são os que mais se destacam na classificação orientada de objetos.

Ainda, para os autores supracitados, para a cultura da batata, o método KNN (K-Nearest Neighbors) apresentou resultados de classificação de alta qualidade para diferentes classes (área urbana, pastagens, cultivos agrícolas, capoeira, eucalipto, água e solo destinado à agricultura), com taxas de acerto próximas ou superiores a 80%. Esses resultados indicam a eficácia do modelo KNN na classificação de diferentes tipos de áreas em relação à cultura da batata. Assim, pode ser aferido que as melhores classificações podem ser encontradas nos solos de cultivo agrícola, naqueles espaços onde a cultura da batata pode se destacar, pois é uma área que possui para atividades agropastoris.

Diante disso, a eficácia do uso do algoritmo KNN na classificação da cultura da batata tem sido amplamente comprovada, conforme mencionado por Borba (2020). Esse algoritmo tem sido empregado em diversos estudos no setor agrícola e, mais especificamente, na indústria de alimentos. Essa ampla utilização do KNN destaca sua relevância e eficiência como uma ferramenta para classificação na área agrícola.

A Tabela 5 apresenta os resultados do treinamento e teste das variáveis climáticas para os modelos de *Machine Learning*. Os modelos avaliados foram RLM, KNN, SVR, XGB e RN. Observou-se que todos os modelos apresentaram valores de R^2 ajustado muito semelhante, variando de 0,62 a 0,63 no conjunto de teste, com exceção do RN, que apresentou 0,17. O modelo com melhor desempenho em termos de RMSE de teste foi o XGB, com valor de 7235,56, enquanto o pior desempenho foi apresentado pela RN, com valor de 15210.81. Os resultados obtidos para as variáveis climáticas foram inferiores aos resultados obtidos para as variáveis nutricionais, sugerindo que prever as variáveis climáticas podem ser mais difíceis do que a previsão das variáveis nutricionais.

Tabela 5. R^2 ajustado e RMSE (raiz do erro quadrático médio) do treinamento e teste das variáveis climáticas para os modelos de *Machine Learning*.

Modelo	R_a^2 treinamento	R_a^2 teste	RMSE (kg ha ⁻¹) treinamento	RMSE (kg ha ⁻¹) teste
RLM	0.62	0.63	7167.90	7235.88
KNN	0.62	0.63	7167.90	7235.88
SVR	0.62	0.62	7240.41	7330.84
XGB	0.62	0.63	7167.95	7235.56
RN	0.16	0.17	15111.58	15210.81

RLM: Regressão Linear Múltipla; KNN: k-Nearest Neighbors; SVR: Support Vector Regression; XGB: Extreme Gradient Boosting; RN: Rede Neural.

Pode ser observado pelos dados que todos os algoritmos apresentaram resultados muito próximos em relação ao R^2 ajustado e ao RMSE de treinamento e teste das variáveis climáticas. O modelo XGB teve um desempenho um pouco melhor em termos de R^2 ajustado e RMSE de teste, com valores de 0,63 e 7235,56, respectivamente.

Entretanto, é importante notar que os dados climáticos foram subdivididos em várias colunas a cada 5 dias e muitas amostras do mesmo experimento foram submetidas às mesmas condições climáticas, o que pode ter levado a uma repetição de dados. Além disso, durante a divisão dos dados em treinamento e teste e no momento da validação cruzada, muitos dos dados se repetiram, resultando em uma falta de variação amostral. Portanto, essa falta de variação amostral pode explicar por que os modelos apresentaram resultados semelhantes.

Conforme destacado por Lopes *et al.* (2011), os municípios produtores de batata, localizados em regiões mais frias, não serão afetados de forma significativa pelas mudanças climáticas. Contudo, as épocas do plantio podem sofrer alterações, o que destaca a importância de considerar as particularidades climáticas de cada região para a tomada de decisões relacionadas ao cultivo.

Trecha *et al.* (2016), no estudo sobre a interferência das variáveis climáticas em relação ao cultivo da batata, tomando como ponto geográfico de referência o plantio da batata no Rio Grande do Sul, enfatizam que as condições climáticas da região, durante o outono, proporcionam que as horas de incidência de luz durante o dia reduzam simultaneamente com a queda na temperatura. Esses fatores, interferem no crescimento do tubérculo, comprometendo as variáveis nutricionais, já que não foi possível o desenvolvimento adequado da batata.

Segundo os autores supracitados, o menor crescimento dos tubérculos da batata é atribuído ao fotoperíodo mais curto, juntamente com baixas temperaturas e alta umidade relativa. Essas condições desfavoráveis dificultam a expansão da área foliar, por consequência, impactam negativamente na taxa fotossintética da planta. Essa constatação destaca a influência direta dos fatores climáticos no desenvolvimento da cultura da batata.

Ao investigar sobre a influência da radiação solar global no rendimento da cultura da batata, Heldwein *et al.*, (2012) indicaram que a radiação solar global é a variável ambiental que desempenha um papel determinante nas diferenças de rendimento dos tubérculos. Essa constatação ressalta a importância da disponibilidade e intensidade da radiação solar para o desenvolvimento adequado da cultura da batata, destacando a sua dependência em relação a esse fenômeno meteorológico.

Caso ocorra interferências durante a formação do tubérculo se tem um impacto direto na produção da planta. Isso se deve ao fato de que o tubérculo desempenha duas funções principais: o armazenamento de reservas e a reprodução vegetativa da planta. Assim, qualquer perturbação nesse órgão afetará tanto a capacidade de armazenamento de nutrientes quanto a habilidade de se reproduzir vegetativamente (Furlaneto, 2014)

De acordo com as considerações de Radons *et al.* (2019), em relação aos possíveis impactos do aquecimento global na cultura da batata, atrasar as datas de plantio pode resultar em exposição da cultura a temperaturas do ar mais elevadas, durante a segunda metade do ciclo, o que pode levar a uma redução na produtividade. Nesse contexto, uma possível alternativa para mitigar os efeitos negativos das altas temperaturas é antecipar a data de plantio, especialmente em regiões áridas. Essa estratégia visa minimizar a exposição das plantas a condições climáticas desfavoráveis e, conseqüentemente, contribuir para o aumento da produtividade da cultura da batata.

Nesse sentido, o manejo do período em que o plantio será realizado pode ser uma possível alternativa para os produtores de batata na região Sul do Brasil. Uma vez que o aumento no aquecimento global está afetando as safras e ocasionando prejuízos financeiros aos produtores rurais.

Na Tabela 6, o modelo RLM apresentou um R^2 de 0,64 no teste, e um RMSE de 6905,04, que representa a diferença média entre os valores previstos e os observados. O modelo KNN obteve um R^2 de 0,73 e um RMSE de 5968,82. O modelo XGB apresentou o R^2 de 0,72, mas um RMSE de 6171,19, indicando que as previsões estão mais distantes dos valores reais em comparação aos outros modelos. O pior modelo foi a Rede Neural, com um R^2 de 0,27 e um RMSE de 14707,58, indicando que as previsões apresentaram grande variação em relação aos valores observados.

O modelo com o melhor desempenho de acordo com as métricas R^2 ajustado e RMSE teste é o SVR, que obteve um R^2 ajustado de 0.76 e um RMSE teste de 5646.80. Essas métricas indicam que o modelo é capaz de explicar 76% da variação nos dados e tem um erro médio de previsão de aproximadamente 5646.80 para as variáveis TOP 10.

Tabela 6. R^2 ajustado e RMSE (raiz do erro quadrático médio) do treinamento e teste das variáveis top 10 para os modelos de *Machine Learning*.

Modelo	R_a^2 treinamento	R_a^2 teste	RMSE (kg ha ⁻¹) treinamento	RMSE (kg ha ⁻¹) teste
RLM	0.67	0.64	6798.68	6905.04
KNN	0.78	0.73	5474.57	5968.82
SVR	0.84	0.76	4671.51	5646.80
XGB	0.98	0.72	1699.97	6171.19
RN	0.24	0.27	15094.25	14707.58

RLM: Regressão Linear Múltipla; KNN: k-Nearest Neighbors; SVR: Support Vector Regression; XGB: Extreme Gradient Boosting; RN: Rede Neural.

No caso específico do modelo XGB utilizado nas tabelas de nutrição e top 10, observou-se um valor de R^2 ajustado de 0,99 no conjunto de treinamento para a tabela de nutrição e 0,98 para a tabela top 10, o que indica que o modelo ajustou-se muito bem aos dados de treinamento, mas apresentou um desempenho inferior no conjunto de teste. Esse resultado pode ser explicado pelo fato de que o modelo XGB é altamente parametrizado e pode ser facilmente ajustado aos dados de treinamento, resultando em um modelo com baixo viés e alta variância (Saxena e Aggarwal, 2020). Isso significa que o modelo pode ajustar excessivamente aos dados de treinamento, tornando-se muito complexo e, portanto, incapaz de generalizar bem para novos dados.

O modelo RN é o algoritmo Multi-Layer Perceptron (MLP), que é um tipo de rede neural artificial *feedforward*, ou seja, as informações são transmitidas somente em uma direção, sem ciclos ou loops. O MLP é composto por várias camadas de neurônios, sendo que cada camada está conectada com a camada seguinte por meio de pesos sinápticos (Jain *et al.*, 1996).

O MLP é um algoritmo bastante utilizado em problemas de classificação e regressão em diversas áreas do conhecimento (Cinar, 2020). No caso dos resultados da RN nas tabelas apresentadas, o MLP apresentou o pior desempenho em relação aos outros modelos, mesmo com uma camada única. Isso pode indicar que, para esse conjunto de dados e problema específico, outras abordagens de modelagem são mais adequadas e que o MLP pode não ter sido capaz de aprender de forma satisfatória as relações entre as variáveis de entrada e a variável de saída.

O Diagrama de Taylor foi utilizado para comparar e selecionar o melhor modelo entre as tabelas nutricionais, climáticas e top 10 (Figura 23). Cada cor no gráfico representa um modelo, sendo vermelho para RLM, verde para KNN, azul para SVM, laranja para XGB e rosa para RN. As formas também são diferentes, sendo círculos para Top 10, triângulos para nutricionais e quadrados para climáticas.

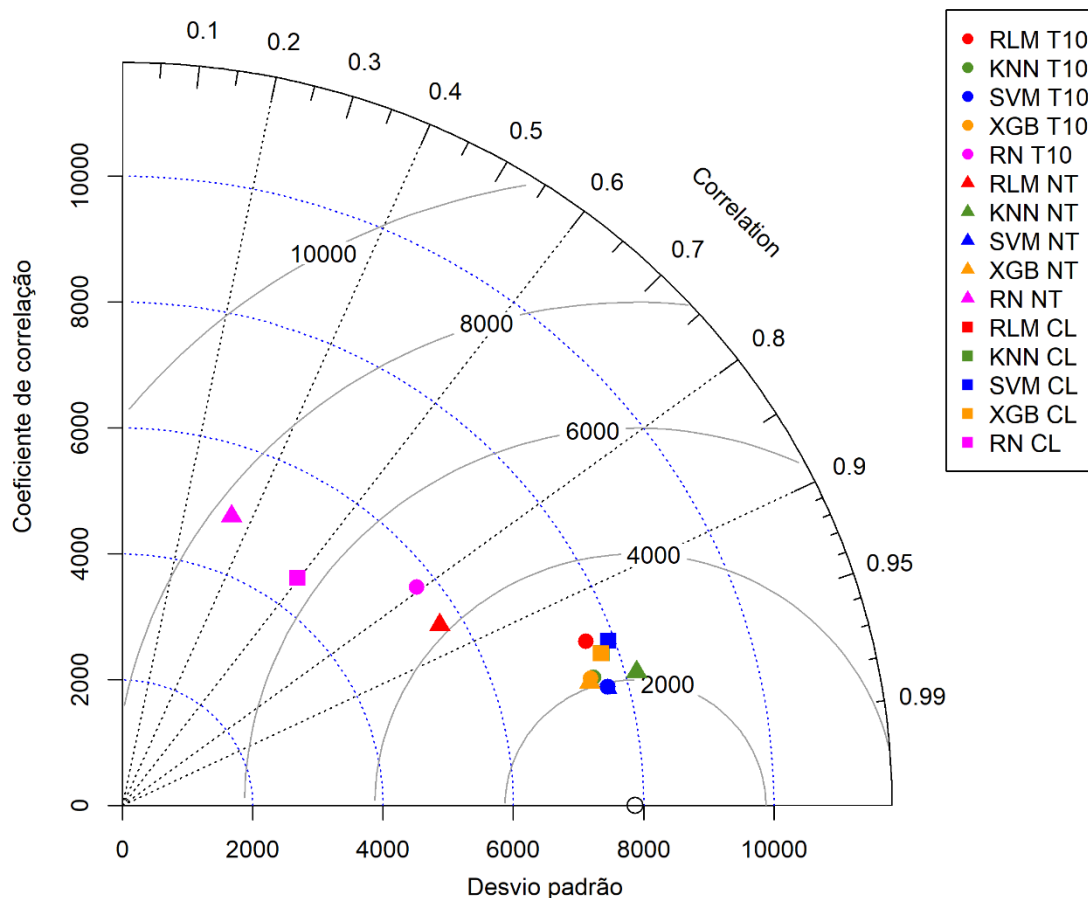


Figura 23. Aplicação do Diagrama de Taylor representando graficamente as estatísticas e a tabulação cruzada, referentes aos interpoladores utilizados.

Simbologia: RLM T10: Regressão Linear Múltipla com variáveis top 10; KNN T10: k-Nearest Neighbors com variáveis top 10; SVR T10: Support Vector Regression com variáveis top 10; XGB T10: Extreme Gradient Boosting com variáveis top 10; RN T10: Rede Neural com variáveis top 10; RLM NT: Regressão Linear Múltipla com variáveis nutricionais; KNN NT: k-Nearest Neighbors com variáveis nutricionais; SVR NT: Support Vector Regression com variáveis nutricionais; XGB NT: Extreme Gradient Boosting com variáveis nutricionais; RN NT: Rede Neural com variáveis nutricionais; RLM CL: Regressão Linear Múltipla com variáveis climáticas; KNN CL: k-Nearest Neighbors com variáveis climáticas; SVR CL: Support Vector Regression com variáveis climáticas; XGB CL: Extreme Gradient Boosting com variáveis climáticas; RN CL: Rede Neural com variáveis climáticas.

A imagem funciona como um resumo das tabelas que aparecem com o R^2 ajustado (Tabelas 1, 2 e 3), com o top 10 e as variáveis nutricionais e climáticas. Nesse sentido, teremos a síntese dos 15 modelos pesquisados e a sua representação – círculo, triângulo e quadrado. Vale chamar atenção para a sobreposição de alguns elementos, sobretudo do quadrado que simboliza a variável climática. Essa sobreposição ocorre porque é apresentado os valores muito próximos entre os modelos.

Primeiramente, é importante considerar que quando o coeficiente de correlação está em 2000, há a presença de um número maior de modelos (RLM, T10, SVM CL, XGB CL, KNN NT, XGB T10, KNN T10, SVM T10), por outro lado, quando o índice de coeficiente de correlação vai para 8000, a incidência de modelos é reduzida, portanto, a presença apenas do modelo RN NT. Diante deste cenário, da aplicação do Diagrama de Taylor, representado por estatísticas e tabulação cruzada, é possível notar que o teor de correlação é maior no arco de 2000 em relação aos demais arcos.

A análise comparativa dos modelos de RN e RLM nos grupos nutricionais e climáticos revelou que esses modelos tiveram um desempenho inferior em relação aos demais modelos avaliados. Além disso, a comparação entre os grupos climáticos e nutricionais indicou uma performance semelhante entre eles, sugerindo que prever variáveis climáticas pode ser tão desafiador quanto prever variáveis nutricionais.

Por outro lado, o modelo SVR, selecionado com base no algoritmo de Random Forest, apresentou o melhor desempenho entre todos os modelos avaliados, destacando-se no Diagrama de Taylor. Esse resultado sugere que a seleção de apenas as 10 principais variáveis, com base no algoritmo de Random Forest, foi bem-sucedida em identificar as variáveis mais importantes para a construção de um modelo com melhor desempenho. Portanto, a análise comparativa entre os modelos permitiram visualizar e comparar o desempenho de cada modelo em um gráfico, facilitando a seleção do melhor modelo. Os resultados indicam que o SVR top 10 é um modelo promissor para previsão da produtividade da batateira.

O SVR se baseia em encontrar o hiperplano que melhor separa as amostras de treinamento em relação à variável de saída. Essa abordagem é adequada para dados que possuem uma relação linear com a variável de saída (Kuter, 2021). Contudo, caso a relação seja não-linear, a utilização de *kernel functions* pode melhorar o desempenho do modelo (Ma *et al.*, 2020). Além disso, o SVR tem a vantagem de ser menos sensível a valores discrepantes (*outliers*) e ruídos nos dados, o que pode ser uma explicação para o seu bom desempenho em relação aos outros modelos avaliados (Üstün, *et al.*, 2005).

Outro fator que pode ter contribuído para o desempenho do modelo de SVR é o ajuste dos parâmetros. O SVR possui vários parâmetros que precisam ser ajustados adequadamente para obter um bom desempenho. Uma técnica comum para ajustar

os parâmetros seria a validação cruzada, que permite testar diferentes combinações de parâmetros e escolher aquela que apresenta o melhor desempenho em dados não vistos (Chuang *et al.*, 2002).

Os modelos de SVM usam proteção de *overfitting* que pode ser ajustada através do parâmetro de regularização C. Este é um parâmetro de penalidade do termo de erro que determina o *trade-off* entre o limite de decisão suave e a classificação correta dos pontos de treinamento. Alguns trabalhos descrevem os benefícios da aplicação de procedimentos de seleção de recursos para melhorar o desempenho preditivo (Blum, 1997; Guyon, *et al.*, 2002), o que corrobora com esse trabalho. Isso pode ser explicado pelo tamanho da amostra do conjunto de dados, conforme sugerido por Jain e Zongker (1997), ou pelos parâmetros internos de regularização que os algoritmos de máquinas de vetores de suporte possuem (Joachims, 1989). Para o SVM, o RMSE foi de 20.2% em relação à média da produtividade. Normalmente, erros de previsão em torno de 15-25% RMSE são encontrados na literatura de previsão de rendimento de culturas (Baez-Gonzalez *et al.*, 2005; Launay *et al.*, 2005).

Uma possível explicação para o fato dos modelos do top 10 ter apresentado o melhor desempenho entre os três grupos pode estar relacionado às características dos dados de entrada. No grupo top 10, os dados selecionados para treinamento e teste foram os que apresentaram as maiores correlações com a variável de saída.

A seleção das variáveis é uma etapa importante no processo de modelagem de dados, uma vez que o uso de variáveis irrelevantes pode levar a uma perda de desempenho do modelo. O processo de seleção das variáveis levou em consideração o valor de R^2 ajustado e o RMSE, onde o modelo de Randomforest apresentou valores de 0,82 e 5260,39, respectivamente. Após a seleção das 10 principais variáveis, todas as que contribuem negativamente para o modelo foram descartadas, permitindo que somente as melhores variáveis fossem incluídas no modelo final.

Com isso, o grupo top 10 variáveis foi criada, permitindo um melhor desempenho dos modelos de *Machine Learning* na predição do desempenho da produtividade com base em variáveis nutricionais e climáticas. A utilização dessa abordagem contribui para um melhor entendimento das variáveis que são mais importantes para a predição do desempenho e para o modelo.

Abbas *et al.* (2020) também realizaram um estudo sobre previsão da produção de tubérculos de batata, utilizando dados de propriedades do solo e da cultura, coletados por meio de sensoriamento proximal. Eles aplicaram quatro algoritmos de aprendizado de máquina (regressão linear, elastic net, k-vizinhos mais próximos e regressão vetorial de suporte) em quatro conjuntos de dados, formados pela combinação de dois locais no Canadá nos anos de 2017 e 2018. As variáveis utilizadas no estudo incluíram condutividade elétrica do solo, teor de umidade do solo, declive do solo, índice de vegetação de diferença normalizada (NDVI) e química do solo. Os resultados indicaram que a Regressão Vetorial de Suporte apresentou melhor desempenho do que os demais modelos, com um RMSE de 5,97, 4,62, 6,60 e 6,17 t ha⁻¹ para NB-2017, NB-2018, PE-2017 e PE-2018, respectivamente.

Nesse sentido, o presente estudo demonstra a importância de considerar diversos fatores para a previsão de produção de tubérculos de batata, bem como diferentes algoritmos de aprendizado de máquina para aprimorar a precisão da previsão. Entretanto, alguns eventos climáticos anormais ou condições adversas as plantas, como doenças e surtos de pragas, podem influenciar os rendimentos modelados, causando subestimação do rendimento (Newton *et al.*, 2011; Gaitán, 2020).

3.4 CONCLUSÕES

Através do algoritmo Random Forest, foi identificado as 10 variáveis principais que exercem maior influência na produtividade. Essas variáveis incluem N, P, K, Ca, Mg, Fe, P+I_EIV_75d, Tmin_EIV_65d, QG_EIV_65d e QG_EV_95d. Os resultados obtidos revelaram R² ajustado de 0,82 e RMSE de 5260,39.

Após a avaliação de diferentes modelos de regressão, incluindo Regressão Linear Múltipla (RLM), K-nearest neighbors (KNN), Support Vector Machine (SVR), eXtreme Gradient Boosting (XGB) e Redes Neurais Artificiais (RN), utilizando três conjuntos de dados distintos (variáveis nutricionais, climáticas e o conjunto das 10 principais variáveis), o modelo SVR top 10 apresentou os melhores resultados. Com um R² ajustado de 0,76 e um RMSE de 5646,80 kg ha⁻¹.

Foi desenvolvida uma aplicação para a previsão da produtividade da batateira, que também pode ser utilizada para outras culturas agrícolas. Essa aplicação

proporcionará aos agricultores e profissionais do setor agrícola uma ferramenta valiosa para tomar decisões fundamentadas e estratégicas, contribuindo para o aumento da eficiência e otimização da produção agrícola.

3.5 REFERÊNCIAS

- Abbas F, Afzaal H, Farooque AA, Tang S (2020) Crop Yield Prediction through Proximal Sensing and *Machine Learning* Algorithms. **Agronomy**, 10(7), 1046. <https://doi.org/10.3390/agronomy10071046>
- Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. D. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische zeitschrift*, 22(6), 711-728.
- Al-Sahaf H, Bi Y, Chen Q, Lensen A, Mei Y, Sun Y, Tran B, Xue B, Zhang M (2019) A survey on evolutionary *Machine Learning*. **Journal of the Royal Society of New Zealand**, 49(2), 205-228. <https://doi.org/10.1080/03036758.2019.1609052>
- Anwar MT, Winarno E, Hadikurniawati W, Novita M (2021) Rainfall prediction using Extreme Gradient Boosting. **Journal of Physics: Conference Series**, 1869(1), 012078. <https://doi.org/10.1088/1742-6596/1869/1/012078>
- Baez-Gonzalez AD, Kiniry JR, Maas SJ, Tiscareno ML, Macias CJL, Mendoza JL, Manjarrez JR (2005) Previsão de produção de milho em grandes áreas usando o modelo de produção baseado no índice de área foliar. **Agronomy Journey**. 97, 418–425.
- Barbedo JGA (2019) Detection of nutrition deficiencies in plants using proximal images and *Machine Learning*: A review. **Computers and Electronics in Agriculture**, 162, 482-492.
- Bisognin DA *et al.* (2008) Envelhecimento fisiológico de tubérculos de batata produzidos durante o outono e a primavera e armazenados em diferentes temperaturas. **Bragantia**, Campinas, v.67, n.1, p.59-65.
- Blum AL, Langley P (1997) Seleção de recursos e exemplos relevantes em aprendizado de máquina. **Artif. Intel.** 97, 245-271.
- Boeschoten S, Catal C, Tekinerdogan B, Lommen A, Blokland M (2023) The automation of the development of classification models and improvement of model quality using feature engineering techniques. **Expert Systems with Applications**, 213(Part A), 118912. <https://doi.org/10.1016/j.eswa.2022.118912>
- Borba KR (2020) **Métodos espectroscópicos para classificação e análise de qualidade de tomates** – Dissertação de Mestrado.
- Borchardt L *et al.* (2011) Adubação orgânica da batata com esterco bovino no município de Esperança-PB. **Revista Ciência Agronômica**, v. 42, p. 482-487.
- Carletto G, Ruel M, Winters P, Zezza A (2015) Farm-Level Pathways to Improved Nutritional Status: Introduction to the Special Issue. **The Journal of Development Studies**, 51(8), 945-957. doi: 10.1080/00220388.2015.1018908

- Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J (2018) **Shiny: Web Application Framework for R**. R Package Version 0.14. Disponível em: <https://CRAN.Rproject.org/package=shiny>
- Chen S, Billings SA, Grant PM (1990) Non-linear system identification using neural networks. **International Journal of Control**, 51(6), 1191-1214. DOI: 10.1080/00207179008934126
- Choubin B, Khalighi-Sigaroodi S, Malekian A, Kişi Ö (2016) Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. **Hydrological Sciences Journal**, 61(6), 1001-1009. DOI: 10.1080/02626667.2014.966721
- Chuang CC, Su S-F, Jeng J-T, Hsiao C-C (2002) Robust support vector regression networks for function approximation with outliers. **IEEE Transactions on Neural Networks**, 13(6), 1322-1330. DOI: 10.1109/TNN.2002.804227
- Cinar AC (2020) Training Feed-Forward Multi-Layer Perceptron Artificial Neural Networks with a Tree-Seed Algorithm. **Arab J Sci Eng** 45, 10915–10938. DOI: 10.1007/s13369-020-04872-1
- Demir S, Şahin EK (2022) Liquefaction prediction with robust *Machine Learning* algorithms (SVM, RF, and XGBoost) supported by genetic algorithm-based feature selection and parameter optimization from the perspective of data processing. **Environ Earth Sci** 81, 459. DOI: 10.1007/s12665-022-10578-4
- Devaux A, Kromann P, Ortiz O (2014) Batatas para a segurança alimentar global sustentável. **Batata Res.** 57, 185–199.
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, García Márquez JR, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. **Ecography**, 36(1), 27-46. DOI: 10.1111/j.1600-0587.2012.07348.x
- El Naqa I, Ruan D, Valdes G, Dekker A, McNutt T, Ge Y, Wu QJ, Oh JH, Thor M, Smith W, Rao A, Fuller C, Xiao Y, Manion F, Schipper M, Mayo C, Moran JM, Ten Haken R (2018) *Machine Learning* and modeling: Data, validation, communication challenges. **Medical Physics**, 45(10), e834-e840. DOI: 10.1002/mp.12811
- Fan J, Wang X, Wu L, Zhou H, Zhang F, Yu X, Lu X, Xiang Y (2018) Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. **Energy Conversion and Management**, 164, 102-111. DOI: 10.1016/j.enconman.2018.02.087

- Farooque AA, *et al.* (2020) Forecasting potato tuber yield using a soil electromagnetic induction method. **European Journal of Soil Science**, 71(5), 880–897. DOI: 10.1111/ejss.12923
- Fleisher DH, Condori B, Quiroz R, Alva A, Asseng S, Barreda C, Bindi M, Boote KJ, Ferrise R, Franke AC (2017) An intercomparison of potato models across contrasting climates and productivity levels. **Global Change Biology**, 23, 1258–1281. DOI: 10.1111/gcb.13442
- Furlaneto KA, Lima PFFS, Daiuto ÉR, Job ALG, Mendonça VZ, Vieites RL, Carvalho LD (2014) Qualidade de batata cv. Ágata influenciada por diferentes concentrações de adubação potássica. **Revista Iberoamericana de Tecnología Postcosecha**, 15(2), 187-192.
- Gaitán CF (2020) *Machine Learning* applications for agricultural impacts under extreme events. In **Climate Extremes and Their Implications for Impact and Risk Assessment** (pp. 119-138). DOI: 10.1016/B978-0-12-814895-2.00007-0
- Garg A, Tai K (2013) Comparison of statistical and *Machine Learning* methods in modeling of data with multicollinearity. **International Journal of Modelling, Identification and Control**, 18(4), 295-312. DOI: 10.1504/IJMIC.2013.053535
- Garofalo DFT, *et al.* (2015) Análise comparativa de classificadores digitais em imagens do Landsat-8 aplicados ao mapeamento temático. **Pesquisa Agropecuária Brasileira**, 50, 593-604.
- Giovanella TH, Oliveira FC, Marchi VAA, Tluszcz J (2021) Desempenho de Métodos de Preenchimento de Falhas em Dados de Evapotranspiração de Referência para Região Oeste do Paraná. **Revista Brasileira de Meteorologia**, 36(3), 415-422. DOI: 10.1590/0102-7786363000
- Guo M, *et al.* (2020) Multi-level system modelling of the resource-food-bioenergy nexus in the global south. **Energy**, 197, 117196.
- Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. **Machine Learning**, 46, 389–422.
- Halal GA (2023) AGROBAYES - **Pacote R para agricultura de precisão com redes Bayesianas**.
- Hatfield JL, Boote KJ, Kimball BA, Ziska LH, Izaurralde RC, Ort D, Thomson AM, Wolfe D (2011) Climate Impacts on Agriculture: Implications for Crop Production. **Agronomy Journal**, 103(2), 351-370. DOI: 10.2134/agronj2010.0303
- Haverkort AJ, Struik PC (2015) Yield levels of potato crops: Recent achievements and future prospects. **Field Crops Research**, 182, 76–85.

- Heldwein AB, Maldaner IC, Bosco LCR, Trentin G, Grimm EL, Radons SZ, Lucas DDP (2012) Daytime net radiation on potato canopies as a function of global solar radiation. **Revista Ciência Agronômica**, 43(1), 96-104.
- Hemalatha K, Usha Rani K (2017) Advancements in Multi-Layer Perceptron Training to Improve Classification Accuracy. **International Journal on Recent and Innovation Trends in Computing and Communication**, 5(6), 353-357.
- Huang Y (2009) Advances in Artificial Neural Networks – Methodological Development and Application. **Algorithms**, 2(3), 973-1007. DOI: 10.3390/algor2030973
- IPCC (2021) Climate change 2021: the physical Science Basis Sixth Assessment Report of intergovernmental panel on climate change.
- Jain A, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 19, 153-158.
- Jain AK, Mao J, Mohiuddin KM (1996) Artificial neural networks: a tutorial. **Computer**, 29(3), 31-44. DOI: 10.1109/2.485891
- Jeong JH, Resop JP, Mueller ND, *et al.* (2016) Random forests for global and regional crop yield predictions. **PLoS ONE**, 11(5), e0156571. DOI: 10.1371/journal.pone.0156571
- Joachims T (1982) Categorização de texto com máquinas de vetores de suporte: aprendendo com muitos recursos relevantes. In **Proceedings of the European conference on Machine Learning**, Viena, Áustria, 12-14 de julho de 1982, Springer: Berlim/Heidelberg, Alemanha, pp. 137-142.
- Jonathan Richetti, Uribe-Opazo MA, Bastiani F, Johann JA (link: <https://www.scielo.br/j/eagri/a/GJR4q9Sq3HJhbkGVzCqz8kG/?lang=pt>) - Técnicas para detecção de pontos influentes em variáveis contínuas regionalizadas.
- Jones DT (2019) Setting the standards for *Machine Learning* in biology. **Nat Rev Mol Cell Biol**, 20, 659-660. DOI: 10.1038/s41580-019-0176-5
- Kan MS, Tan ACC, Mathew J (2015) A review on prognostic techniques for non-stationary and non-linear rotating systems. **Mechanical Systems and Signal Processing**, 62-63, 1-20.
- Kassambara A (2017) **Practical Guide To Cluster Analysis in R**. 1st ed. USA: STHDA.
- Kherif F, Latypova A (2020) Principal component analysis. In **Machine Learning: Methods and Applications to Brain Disorders**, 209-225. Laboratory of Research in Neuroimaging (LREN), Department of Clinical Neurosciences, CHUV,

- University of Lausanne, Lausanne, Switzerland: Elsevier. DOI: 10.1016/B978-0-12-815739-8.00012-2
- Kuhn M (2008) Building Predictive Models in R Using the caret Package. **Journal of Statistical Software**, 28(5), 1-26. URL: <https://doi.org/10.18637/jss.v028.i05>.
- Kuradusenge M, Hitimana E, Hanyurwimfura D, Rukundo P, Mtonga K, Mukasine A, Uwitonze C, Ngabonziza J, Uwamahoro A (2023) Crop Yield Prediction Using *Machine Learning* Models: Case of Irish Potato and Maize. **Agriculture**, 13(1), 225. DOI: 10.3390/agriculture13010225
- Kuter S (2021) Completing the *Machine Learning* saga in fractional snow cover estimation from MODIS Terra reflectance data: Random forests versus support vector regression. **Remote Sensing of Environment**, 255, 112294. DOI: 10.1016/j.rse.2021.112294.
- Launay M, Guerif M (2005) Assimilando dados de sensoriamento remoto em um modelo de cultura para melhorar o desempenho preditivo para aplicações espaciais. *Agrícola Ecosistema*. **Ambiente**, 111, 321-339.
- Lepre AL (2009) **Avaliação de componentes relacionados à aparência externa de tubérculos de batata**. Lavras: UFLA, 71 p.
- Livingstone D, Manallack D, Tetko I (1997) Data modelling with neural networks: Advantages and limitations. **J Comput Aided Mol Des**, 11, 135-142. DOI: 10.1023/A:1008074223811
- Lombardi Neto F, Drugowich MI (coords.) (1994) **Manual técnico de manejo e conservação de solo e água**. V. III CATI. Campinas: CATI, p.121-156. Manual Técnico, 40.
- Lopes CA, *et al.* (2011) Uma análise do efeito do aquecimento global na produção de batata no Brasil. **Horticultura Brasileira**, 29, 7-15.
- Lorenzi JO, Miranda Filho HS, Van Raij B (1997) Raízes e tubérculos. In **Recomendações de adubação e calagem para o Estado de São Paulo**, Tech Bull 100, 2nd edn, eds. Van Raij B, Cantarella H, Quaggio JA, Furlani AMC, 221–229. Campinas: Instituto Agrônômico.
- Ma D, Maki H, Neeno S, Zhang L, Wang L, Jin J (2020) Application of non-linear partial least squares analysis on prediction of biomass of maize plants using hyperspectral images. **Biosystems Engineering**, 200, 40-54. DOI: 10.1016/j.biosystemseng.2020.09.002.
- Manrique LA, Kinry JR, Hodges T, Axness DS (1991) Produção de matéria seca e interceptação de radiação em batata. **Crop Sci**, 31, 1044-1049.

- Mariadass DA-L, Mounq EG, Sufian MM, Farzamnia A (2022) Extreme Gradient Boosting (XGBoost) Regressor and Shapley Additive Explanation for Crop Yield Prediction in Agriculture. 2022 12th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, Islamic Republic of, pp. 219-224. DOI: 10.1109/ICCKE57176.2022.9960069.
- Marin FR, Jones JW, Singels A, *et al.* (2013) Climate change impacts on sugarcane attainable yield in southern Brazil. **Climatic Change**, 117, 227-239. DOI: 10.1007/s10584-012-0561-y
- Marouelli WA, Braga MB, Guimarães TG (2013) **Irrigação na cultura da batata**. Circular Técnica - EMBRAPA. ISSN 1415-3033.
- Marschner H (2012) **Mineral nutrition of higher plants**. 3rd ed. London: Elsevier.
- Mishra S, Mishra D, Santra GH (2016) Applications of *Machine Learning* techniques in agricultural crop production: A review paper. **Indian Journal of Science and Technology**, 9(38), 1-9. DOI: 10.17485/ijst/2016/v9i38/95032
- Miyazawa M, Pavan MA, Muraoka T, Carmo CAFS, Melo WJ (2009) Análise química de tecido vegetal. In: Silva FC (Ed.) **Manual de análises químicas de solos, plantas e fertilizantes**. 2nd ed. Brasília: EMBRAPA, p. 191-233.
- Müller DR, Bisognin DA, Morin GR, Gnolato FS (2010) Dormência e dominância apical de diferentes tamanhos de tubérculos de batata. **Ciência Rural**, 40, 2454-2459.
- NASA/POWER. Portal NASA Prediction Of Worldwide Energy Resources. Disponível em: <https://power.larc.nasa.gov/>. Acesso em: 05 jun. 2021.
- Nestor L. Sy (2006) Modelling the infiltration process with a multi-layer perceptron artificial neural network. **Hydrological Sciences Journal**, 51(1), 3-20. DOI: 10.1623/hysj.51.1.3
- Newton AC, Johnson SN, Gregory PJ (2011) Implications of climate change for diseases, crop yields and food security. **Euphytica**, 179, 3-18. DOI: 10.1007/s10681-011-0359-4
- Pandey VL, Dev SM, Jayachandran U (2016) Impact of agricultural interventions on the nutritional status in South Asia: A review. **Food Policy**, 62, 28-40. DOI: 10.1016/j.foodpol.2016.05.002
- Pereira DP, Lima JSS, Xavier AC, Passos RR (2014) Aplicação do diagrama de Taylor para avaliação de interpoladores espaciais em atributos de solo em cultivo com eucalipto. *Revista Árvore*, 38(5). DOI: 10.1590/S0100-67622014000500014
- Pereira GE (2018) Avaliação agrônômica e curva de crescimento de genótipos de batata para a região produtora do Centro-Oeste – Tese de doutorado.

- Piekutowska M, Niedbała G, Piskier T, Lenartowicz T, Pilarski K, Wojciechowski T, Pilarska AA, Czechowska-Kosacka A (2021) The Application of Multiple Linear Regression and Artificial Neural Network Models for Yield Prediction of Very Early Potato Cultivars before Harvest. **Agronomy**, 11(5), 885. DOI: 10.3390/agronomy11050885
- Prabhu BVB, Dakshayini M (2018) Performance analysis of the regression and time series predictive models using parallel implementation for agricultural data. **Procedia Computer Science**, 132, 198-207. DOI: 10.1016/j.procs.2018.05.187
- Qolomany B, Maabreh M, Al-Fuqaha A, Gupta A, Benhaddou D. Parameters optimization of deep learning models using Particle swarm optimization. In: **2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)**. Valencia, Spain; 2017. p. 1285-1290. doi: 10.1109/IWCMC.2017.7986470.
- Quadros DA, lung MC, Ferreira SMR, Freitas RJS (2009) Composição química de tubérculos de batata para processamento, cultivados sob diferentes doses e fontes de potássio. **Ciência e Tecnologia de Alimentos**, 29(2), 316-323.
- Radons SZ, *et al.* (2014) Ciclo total e subperíodos da batateira na região central do Rio Grande do Sul. **Acta Iguazu**, 8(3), 18-29.
- Raimundo R, Asseng S, Cammarano D, Quiroz R (2014) Modelos de batata, batata-doce e inhame para mudanças climáticas: uma revisão. **Colheita de Campo. Res. Rev.**, 166, 173-185.
- Raymundo R, *et al.* (2018) Climate change impact on global potato production. **European Journal of Agronomy**, 100, 87-98. DOI: 10.1016/j.eja.2017.11.008
- Roberts MJ, *et al.* (2017) Comparing and combining process-based crop models and statistical models with some implications for climate change. **Environmental Research Letters**, 12(9), 095010. DOI: 10.1088/1748-9326/aa836e
- Rodríguez del Águila MM, Benítez-Parejo N (2011) Simple linear and multivariate regression models. **Allergologia et Immunopathologia**, 39(3), 159-173. DOI: 10.1016/j.aller.2011.02.001
- Sahin EK (2020) Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. **SN Applied Sciences**, 2(7), 1308. DOI: 10.1007/s42452-020-3060-1
- Saleem MH, Potgieter J, Arif KM (2021) Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments. **Precision Agriculture**, 22, 2053-2091. DOI: 10.1007/s11119-021-09806-x

- Saxena V, Aggarwal A (2020) Comparative study of select non-parametric and ensemble *Machine Learning* classification techniques. In 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 110-115). **IEEE**. DOI: 10.1109/ICACCCN51052.2020.9362798
- Shakoor MT, Rahman K, Rayta SN, Chakrabarty A (2017) Agricultural production output prediction using Supervised *Machine Learning* techniques. In 2017 1st International Conference on Next Generation Computing Applications (NextComp) (pp. 182-187). **IEEE**. DOI: 10.1109/NEXTCOMP.2017.8016196
- Sharma R, Kamble SS, Gunasekaran A, Kumar V, Kumar A (2020) A systematic literature review on *Machine Learning* applications for sustainable agriculture supply chain performance. **Computers & Operations Research**, 119, 104926. DOI: 10.1016/j.cor.2020.104926
- Silva DHC, Timo EMN (2022) *Machine Learning* applied to home care for predicting passing away conditions. *Research, Society and Development*, 11(14), e20311143606. DOI: 10.33448/rsd-v11i14.36078
- Silva GO, Hirano E (2015) Implantação da cultura. In: **Sistema de produção da Batata**. 2ª ed. - Embrapa: São Paulo, 2015, p. 28-31.
- Singh BP, Dua VK, Govindakrishnan PM, Sharma S (2013) Impact of Climate Change on Potato. In: **Singh H, Rao N, Shivashankar K** (eds) *Climate-Resilient Horticulture: Adaptation and Mitigation Strategies*. Springer, India. DOI: 10.1007/978-81-322-0974-4_12
- Singh G, Panda RK (2011) Daily Sediment Yield Modeling with Artificial Neural Network using 10-fold Cross Validation Method: A small agricultural watershed, Kapgari, India. **International Journal of Earth Sciences and Engineering**, 04(06 SPL), 443-450.
- Song Y, Liang J, Lu J, Zhao X (2017) An efficient instance selection algorithm for k nearest neighbor regression. **Neurocomputing**, 251, 26-34. DOI: 10.1016/j.neucom.2017.04.018
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. **Journal of Geophysical Research**, 106(D7), 7183-7192.
- Thomas AL, Bredemeier C, Vian AL (2016) Desenvolvimento da planta de batata. In: **Desenvolvimento das plantas de batata, mandioca, fumo e cana-de-açúcar**. Porto Alegre: UFRGS, 2016, p. 6-17.
- Timlin D, Rahman L, Baker SM, Reddy J, Fleisher D, Quebedeaux B (2006) Whole plant photosynthesis, development, and carbon partitioning in potato as a function of temperature. **Agronomy Journal**, 98, 1195-1203.

- Tindall JA, Mills HA, Radcliffe DE (1990) The effect of root zone temperature on nutrient uptake of tomato. **Journal of Plant Nutrition**, 13(8), 939-956. DOI: 10.1080/01904169009364127
- Trecha CO, *et al.* (2016) Influência da safra de verão e outono no dano de larvas da vaquinha em tubérculos de batata cultivadas em casa de vegetação. **Revista de la Facultad de Agronomía**, 115(1), 1-8.
- Trenberth KE, Fasullo JT, Branstator G, Phillips AS (2014) Seasonal aspects of the recent pause in surface warming. **Nature Climate Change**, 4(11), 911-916. DOI: 10.1038/nclimate2341
- Üstün B, Melssen WJ, Oudenhuijzen M, Buydens LMC (2005) Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. **Analytica Chimica Acta**, 544(1-2), 292-305. DOI: 10.1016/j.aca.2004.12.024
- Vianello RL, Alves AR (2000) **Meteorologia básica e aplicações**. Viçosa, UFV: Imprensa Universitária. 449p.
- Webber H, *et al.* (2020) No perfect storm for crop yield failure in Germany. **Environmental Research Letters**, 15(10), 104012.
- Wrege MS, *et al.* (2004) Caracterização climática das regiões produtoras de batata no Brasil. **Caderno Embrapa**.
- Yang L, Shami A (2020) On hyperparameter optimization of *Machine Learning* algorithms: Theory and practice. **Neurocomputing**, 415, 295-316. DOI: 10.1016/j.neucom.2020.07.061
- Zhang W, Wu C, Zhong H, Li Y, Wang L (2021) Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. **Geoscience Frontiers**, 12(1), 469-477. DOI: 10.1016/j.gsf.2020.03.007
- Zhao Z, Qiao K, Liu Y, Chen J, Li C (2022) Data Analysis: The Heilongjiang Duobaoshan Area (China) Case Study. **Minerals**, 12(8), 1035. DOI: 10.3390/min12081035
- Zheng YY, Kong JL, Jin XB, Wang XY, Su TL, Zuo M. cropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture. **Sensors**. 2019;19(5):1058. doi:10.3390/s19051058