

UNIVERSIDADE ESTADUAL DE GOIÁS
ANÁPOLIS DE CIÊNCIAS EXATAS E TECNOLÓGICAS: CÂMPUS CENTRAL
MESTRADO *STRICTO SENSU* EM ENGENHARIA AGRÍCOLA

JEFFERSON SOUZA E SILVA

**ESTIMATIVA DE ATRIBUTOS QUÍMICOS DOS SOLOS DO CERRADO
EMPREGANDO REDES NEURAIS ARTIFICIAIS E O ALGORITMO FLORESTA
ALEATÓRIA**

ANÁPOLIS-GO
2023

JEFFERSON SOUZA E SILVA

**ESTIMATIVA DE ATRIBUTOS QUÍMICOS DOS SOLOS DO CERRADO
EMPREGANDO REDES NEURAIS ARTIFICIAIS E O ALGORITMO FLORESTA
ALEATÓRIA**

Dissertação apresentada à
Universidade Estadual de Goiás
Campus Anápolis de Ciências
Exatas e Tecnológicas, para
obtenção do título de Mestre em
Engenharia Agrícola.

Área de concentração CNPQ:
Tecnologia Agrícola.

Orientador: Prof. Dr. Francisco
Ramos de Melo.

ANÁPOLIS-GO
2023

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UEG
com os dados fornecidos pelo(a) autor(a).

Se	Souza e Silva, Jefferson ESTIMATIVA DE ATRIBUTOS QUÍMICOS DOS SOLOS DO CERRADO EMPREGANDO REDES NEURAIS ARTIFICIAIS E O ALGORITMO FLORESTA ALEATÓRIA / Jefferson Souza e Silva; orientador Francisco Ramos de Melo. -- Anápolis, 2023. 71 p. Dissertação (Mestrado - Programa de Pós-Graduação Mestrado Acadêmico em Engenharia Agrícola) -- Câmpus Central - Sede: Anápolis - CET, Universidade Estadual de Goiás, 2023. 1. Inteligência artificial. 2. Atributos químicos do solo. 3. processos agrícolas. I. Ramos de Melo, Francisco , orient. II. Título.
----	---

**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DE TESES E DISSERTAÇÕES
NA BIBLIOTECA DIGITAL (BDTD)**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Estadual de Goiás a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UEG), regulamentada pela Resolução, **CsA n.1087/2019** sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

Dados do autor (a)Nome Completo Jefferson Souza e SilvaE-mail jeffersonarar@gmail.com**Dados do trabalho**

Título Estimativa de Atributos Químicos dos Solos do Cerrado Empregando Redes Neurais Artificiais e o Algoritmo Floresta Aleatória

Tipo() Tese (☒) Dissertação

Curso/Programa Mestrado Acadêmico em Engenharia Agrícola

Concorda com a liberação documento [☒] SIM [] NÃO¹

_____Anápolis-GO____, 17/_03_/ 2024____
Local Data



Assinatura do autor (a)

Assinatura do orientador (a)**¹ Casos de impedimento:**

- Período de embargo é de um ano a partir da data de defesa
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da Dissertação/Tese em livro.

JEFFERSON SOUZA E SILVA

**ESTIMATIVA DE ATRIBUTOS QUÍMICOS DOS SOLOS DO CERRADO
EMPREGANDO REDES NEURAIS ARTIFICIAIS (RNAs) E O ALGORITMO
FLORESTA ALEATÓRIA**

Dissertação apresentada à
Universidade Estadual de Goiás,
Campus Central de Ciências
Exatas e Tecnológicas para
obtenção do título de Mestre em
Engenharia Agrícola.

Aprovado em: 31/08/2023

Banca Examinadora



Documento assinado digitalmente
FRANCISCO RAMOS DE MELO
Data: 04/03/2024 10:33:23-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Francisco Ramos de Melo – UEG

Orientador



Documento assinado digitalmente
MARIA JOSELMA DE MORAES
Data: 04/03/2024 19:40:30-0300
Verifique em <https://validar.iti.gov.br>

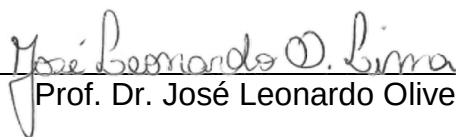


Prof.^a Dra. Maria Joselma de Moraes – UEG

Membro Interno



Documento assinado digitalmente
JOSE LEONARDO OLIVEIRA LIMA
Data: 04/03/2024 11:56:55-0300
Verifique em <https://validar.iti.gov.br>



Prof. Dr. José Leonardo Oliveira Lima – UEG

Membro Externo

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus por me guiar e me fortalecer ao longo dessa trajetória. Sem Sua presença em minha vida, nada disso seria possível.

À minha família, em especial aos meus pais, sou imensamente grato pelo apoio incondicional, amor e incentivo que sempre me proporcionaram. Vocês foram minha base, minha força motriz e fonte de inspiração.

Gostaria de expressar um agradecimento especial aos meus amigos Willian, Ronnie e Samuel. Vocês merecem um reconhecimento sincero por acreditar inabalavelmente em meu potencial desde o início deste percurso.

À minha amiga Alessandra, não tenho palavras para agradecer por sua ajuda inestimável em minha pesquisa. Sua expertise e orientação foram fundamentais para o sucesso deste trabalho.

Aos demais amigos e colegas que estiveram ao meu lado durante esse percurso, agradeço pela amizade, parceria e apoio mútuo. Suas palavras de encorajamento e motivação foram essenciais para enfrentar os desafios.

Agradeço à minha colega de sala, Thathelle, por compartilhar experiências e conhecimentos ao longo do mestrado, tornando essa jornada ainda mais enriquecedora.

Agradeço ao meu amigo, Fernando, pela compreensão e flexibilidade que me concedeu durante esse período de dedicação ao mestrado, permitindo conciliar as atividades profissionais com os estudos.

À equipe docente do programa de mestrado, meus professores em especial ao meu orientador Prof. Dr. Francisco Ramos de Melo que muito me ajudou, Prof^a. Dra. Maria Joselma de Moraes, Prof. Dr. André José de Campos, Prof^a. Dra. Roberta Passine e Prof^a. Dra. Gisele Carneiro da Silva Teixeira, agradeço por compartilharem seus conhecimentos e experiências, contribuindo para o meu crescimento acadêmico e profissional.

Não posso deixar de agradecer também à Universidade Estadual de Goiás (UEG) e à Fundação de Amparo à Pesquisa do Estado de Goiás (Fapeg) pelo suporte financeiro e pela estrutura de ensino, pesquisa e desenvolvimento que me proporcionaram durante o mestrado.

Este momento não seria possível sem cada uma dessas pessoas e as instituições que mencionei. A todos vocês, minha eterna gratidão!

RESUMO

A Inteligência Artificial (IA) tem desempenhado um papel significativo na globalização, impactando profundamente em diversos setores, incluindo a agricultura. No contexto agrícola, a IA por meio de técnicas como Redes Neurais Artificiais (RNAs) e Aprendizado de Máquina, como o algoritmo Floresta Aleatória, tem desempenhado um papel crucial na otimização dos processos agrícolas, tais como a gestão de cultivos, a redução de custos e o aumento da precisão das operações. Neste trabalho, a aplicação de Redes Neurais Artificiais (RNAs) e do algoritmo Floresta Aleatória (RF) teve como objetivo a estimativa dos atributos químicos do solo, visando a redução da quantidade de amostras de material e a comparação das metodologias utilizadas no treinamento para determinar qual delas obteve os melhores resultados nas previsões. Com isso foi selecionada uma base de dados que continha amostras de solo com informações sobre cálcio (CA), pH, fósforo (P), potássio (K), magnésio (MG), soma de cálcio e magnésio (CA+MG), matéria orgânica (MO) e capacidade de troca catiônica (CTC). O estudo empregou três abordagens distintas: Modelo A, Modelo B e Modelo C. Os Modelos A e B utilizaram o algoritmo Floresta Aleatória (RF), enquanto o Modelo C empregou Redes Neurais Artificiais (RNAs). Cada um desses modelos foi treinado e validado com base nas amostras da base de dados selecionada. A análise comparativa dos modelos revelou que o Modelo B apresentou resultados promissores, avaliados por meio de métricas de desempenho de modelos de regressão, tais como o coeficiente de determinação (R^2) e o erro médio quadrático (EMQ). Portanto, foi possível demonstrar que a aplicação da IA, por meio de técnicas como Redes Neurais Artificiais (RNAs) e o algoritmo Floresta Aleatória, teve um impacto positivo na estimativa de atributos químicos do solo. O Modelo B destacou-se como a abordagem mais eficaz, apresentando resultados sólidos com base em métricas de avaliação de desempenho.

Palavras-chaves: Inteligência artificial, atributos químicos do solo, processos agrícolas.

ABSTRACT

Artificial Intelligence (AI) has played a significant role in globalization, profoundly impacting several sectors, including agriculture. In the agricultural context, AI, through techniques such as Artificial Neural Networks (ANNs) and Machine Learning, such as the Random Forest algorithm, has played a crucial role in optimizing agricultural processes, such as crop management, reducing costs and increased precision of operations. In this work, the application of Artificial Neural Networks (ANNs) and the Random Forest (RF) algorithm aimed to estimate soil chemical attributes, aiming to reduce the number of soil samples and compare the methodologies used in training to determine which one obtained the best results in the predictions. With this, a database was selected that contained soil samples with information on calcium (CA), pH, phosphorus (P), potassium (K), magnesium (MG), sum of calcium and magnesium (CA+MG), matter organic matter (MO) and cation exchange capacity (CEC). The study employed three distinct approaches: Model A, Model B and Model C. Models A and B used the Random Forest (RF) algorithm, while Model C employed Artificial Neural Networks (ANNs). Each of these models was trained and validated based on samples from the selected database. The comparative analysis of the models revealed that Model B presented promising results, evaluated using regression model performance metrics, such as the coefficient of determination (R^2) and the mean squared error (EMQ). Therefore, it was possible to demonstrate that the application of AI, through techniques such as Artificial Neural Networks (ANNs) and the Random Forest algorithm, had a positive impact on the estimation of soil chemical attributes. Model B stood out as the most effective approach, delivering solid results based on performance assessment metrics.

Keywords: Artificial intelligence, Chemical attributes of the soil, agricultural processes.

SUMÁRIO

1. INTRODUÇÃO	13
2. REVISÃO DE LITERATURA.....	17
2.1. Atributos dos solos do cerrado	17
2.2. Redes Neurais Artificiais (RNAS)	18
2.3. Perceptron Multicamadas (MLP)	19
2.4. Redes Neurais Artificiais (RNAS) aplicadas na agricultura.....	19
2.5. Algoritmo Floresta Aleatória (RF)	20
2.6. Algoritmo Floresta Aleatória (RF) e inteligência artificial aplicados na agricultura	21
3. MATERIAL E MÉTODOS.....	23
3.1. Descrição e localização da área do estudo	23
3.2. Mapeamento e georrefenciamento dos pontos amostrais	23
3.3. Atributos físicos e químicos do solo	24
3.4. Distribuição das amostras no plano cartesiano	24
3.5. Definição das variáveis de entrada.....	26
3.6. Aplicação da Inteligência artificial	26
3.6.1. <i>Treinamento das Redes Neurais Artificiais (RNAS) e do Algoritmo Floresta Aleatória (RF).....</i>	<i>26</i>
3.7. Aplicação das Redes Neurais Artificiais (RNAS)	28
3.7.1. <i>Redes Neurais Artificiais (RNAS) – Perceptrons de Multicamadas (MLP)</i>	<i>29</i>
3.7.1.1. <i>Conjunto de Treinamento e Teste:</i>	<i>30</i>
3.7.2. <i>Estrutura do Algoritmo Floresta Aleatória (RF)</i>	<i>33</i>
3.7.2.1. <i>Entradas do Modelo de Regressão com Floresta Aleatória.....</i>	<i>33</i>
3.7.2.2. <i>Saídas do Modelo de Regressão com Floresta Aleatória.....</i>	<i>34</i>
3.7.2.3. <i>Índice de GINI.....</i>	<i>34</i>
3.7.2.4. <i>Divisão do Conjunto de Dados em Treinamento e Teste:.....</i>	<i>35</i>
3.7.2.5. <i>Validação das Redes Neurais Artificiais (RNAS) e do Algoritmo Floresta Aleatória (RF).....</i>	<i>36</i>
3.7.3. <i>Distâncias das amostras em relação aos pontos amostrais</i>	<i>37</i>
3.7.4. <i>Análise Estatística descritiva e exploratória</i>	<i>39</i>

4. RESULTADOS E DISCUSSÃO	40
4.1. Análise descritiva dos dados	40
4.1.1. <i>Análise do Modelo A: Estimativas de atributos químicos (5 pontos)</i>	42
4.1.1.1. <i>Potencial hidrogeniônico (Ph) - Modelo A</i>	44
4.1.1.2. <i>Análise do Cálcio (Ca) - Modelo A</i>	45
4.1.1.3. <i>Análise do Magnésio (Mg) - Modelo A</i>	45
4.1.1.4. <i>Análise do Cálcio + magnésio (Ca+Mg) - Modelo A</i>	45
4.1.1.5. <i>Análise da Capacidade de troca catiônica (CTC) - Modelo A</i>	45
4.1.1.6. <i>Análise do Fósforo (P) - Modelo A</i>	46
4.1.1.7. <i>Análise do Potássio (K) - Modelo A</i>	46
4.1.1.8. <i>Análise da Matéria orgânica (MO) - Modelo A</i>	46
4.1.2. <i>Modelo B: Estimativa de atributos químicos do solo:</i>	47
4.1.2.1. <i>Análise do Potencial hidrogeniônico (Ph) - Modelo B</i>	49
4.1.2.2. <i>Análise do Cálcio (Ca) - Modelo B</i>	49
4.1.2.3. <i>Análise do Magnésio (Mg) - Modelo B</i>	50
4.1.2.4. <i>Análise do Cálcio + magnésio (Ca+Mg) - Modelo B</i>	51
4.1.2.5. <i>Análise da Capacidade de troca catiônica (CTC) - Modelo B</i>	51
4.1.2.6. <i>Análise do Fósforo (P) - Modelo B</i>	52
4.1.2.7. <i>Análise do Potássio (K) - Modelo B</i>	52
4.1.2.8. <i>Análise da Matéria orgânica (MO) - Modelo B</i>	53
4.2. <i>Modelo C – Aplicação de Redes Neurais Artificiais (RNAS)</i>	53
4.2.1. <i>Avaliação das previsões dos atributos do Modelo C.</i>	53
4.3. <i>Modelo selecionado</i>	56
 5. CONCLUSÃO	 60
 6. REFERÊNCIAS BIBLIOGRÁFICAS	 61

LISTA DE FIGURAS

Figura 1 - Localização da área de coleta dos dados experimentais, na Fazenda Palmital, Morrinhos-GO.....	23
Figura 2 - Plano cartesiano das posições das amostras em relação aos pontos fixos	24
Figura 3 - Gráfico da posição dos dados e amostras do modelo A no plano cartesiano.....	27
Figura 4 - Gráfico da posição dos dados e amostras do modelo B no plano cartesiano.....	27
Figura 5 - Gráfico da posição dos dados e amostras do modelo C no plano cartesiano.	28
Figura 6 - Modelo Perceptrons de Multimarcas (MLP)	29
Figura 7 - Conjunto de treinamento com 80% dos dados e conjunto de teste com 20% ...	31
Figura 8 - Representação do K-fold	32
Figura 9 - Modelo de Regressão com Floresta Aleatória (Random Forest Regressor).....	33
Figura 10 - Divisão do conjunto de dados em treinamento	35
Figura 11 - Representações das distâncias das amostras previstas em relação aos 5 pontos fixos (Modelo A).....	38
Figura 12 - Representações das distâncias das amostras previstas em relação aos 9 pontos fixos (Modelo B).....	39
Figura 13 - Gráfico de Correlação do Modelo A	43
Figura 14 - Gráfico do Coeficiente de determinação do Modelo A.	43
Figura 15 - Gráfico do Erro Médio do Modelo A	44
Figura 16 - Gráfico do coeficiente de correlação do Modelo B	48

Figura 17 - Gráfico do coeficiente de determinação do Modelo B	48
Figura 18 - Gráfico do Erro Médio do Modelo B	49
Figura 19 - Gráfico do Coeficiente de Correlação do Modelo C	54
Figura 20 - Gráfico do Coeficiente de determinação do Modelo C	55
Figura 21 - Gráfico do Erro médio do Modelo C	55
Figura 22 - Gráfico dos Valores de erro médio quadrático dos modelos A, B e C	57
Figura 23 - Gráfico do Coeficiente de Correlação dos modelos	58
Figura 24 - Gráfico do Coeficiente de determinação dos modelos	59

LISTA DE TABELAS

Tabela 1 - Análise descritiva dos atributos químicos do solo	40
Tabela 2 - Análise estatística dos Valores Observados x Valores Previstos do Modelo A.....	42
Tabela 3 - Análise estatística dos Valores Observados x Valores Previstos do Modelo B.....	47
Tabela 4 - Análise estatística dos Valores Observados x Valores Previstos do Modelo C.....	54

LISTA DE SIGLAS E ABREVIATURAS

Ca - Cálcio
Ca+Mg - Cálcio + magnésio
CTC - Capacidade de troca catiônica
K - Potássio
Mg - Magnésio
MLP – Perceptron de Multicamadas
MO - Matéria orgânica
P - Fósforo
Ph - Potencial Hidrogeniônico
RF - Algoritmo de Floresta Aleatória
RNAS - Redes Neurais Artificiais

1. INTRODUÇÃO

A agricultura de precisão no solo representa uma revolução significativa na forma como os agricultores realizam o cultivo de suas terras, pois buscam maximizar a eficiência e produtividade agrícola, enquanto minimiza o uso excessivo de recursos naturais, como água e fertilizantes. Para que essa abordagem seja eficaz, a compreensão profunda do solo é um papel crucial (NANNI et al., 2011).

A compreensão profunda do solo significa explorar os elementos do solo, como materiais minerais e orgânicos originados de complexas interações ao longo do tempo, na qual os pesquisadores podem adquirir uma visão mais ampla das características do solo. Esse substrato não apenas contém elementos vitais para o crescimento das plantas, mas também incorpora influências humanas para que ocorre esse desenvolvimento dos vegetais (CAPOANE et al., 2017).

Sendo assim, a pesquisa do solo envolve a análise de diversos componentes, incluindo composição mineral por meio da determinação da sua capacidade de retenção de nutrientes e água, bem como sua textura a matéria orgânica no solo que fornece nutrientes essenciais para as plantas e melhora a estrutura do solo. Além disso, os microrganismos presentes no solo atuam na decomposição da matéria orgânica, na fixação de nitrogênio e na ciclagem de nutrientes. Atuam ainda nas propriedades físicas do solo, como a textura (areia, silte, argila), densidade, porosidade e capacidade de retenção de água, que afetam a capacidade do solo de fornecer água e nutrientes às plantas (PELUCO, 2013).

Nesse contexto, o solo com determinados atributos químicos e físicos é resultado da atuação de processos de formação que atuam de forma diferentes com intensidades diferentes para cada ponto geográfico (CAPOANE et al., 2017). Sendo assim, as propriedades do solo e sua variabilidade geográfica permitem um planejamento mais assertivo no manejo de safras (NANNI et al., 2011).

O Cerrado brasileiro é uma das região que vem passando por notáveis transformações econômicas, garantindo a expansão das atividades agrícolas e pecuárias e abrindo novas perspectivas de negócios. Contudo, essa evolução também desencadeou a degradação da vegetação nativa, o que vem proporcionando desafios ambientais de restauração de áreas, como apontado por Klink e Machado (2005).

Para resolver essa problemática, ao considerar os elementos e processos do solo, é possível desenvolver abordagens mais equilibradas que permitam a

prosperidade econômica, ao mesmo tempo em que se preserva o ambiente natural buscando a sustentabilidade a longo prazo. Uma dessas abordagens tecnológicas é a agricultura de precisão que utiliza a caracterização da variabilidade do solo para fazer um manejo adequado de cada área. A agricultura de precisão tem crescido de forma exponencial e fortalecido ainda mais o agronegócio brasileiro (PELUCO, 2013).

De acordo com Reghini et al. (2020) a agricultura de precisão é definida como uma tecnologia cujo propósito central é ampliar a produção agrícola por meio de uma gestão mais eficiente e sustentável dos recursos econômicos e ambientais.

Com base no que foi mencionado anteriormente, uma abordagem promissora para a compreensão e otimização do solo na agricultura é a utilização da Inteligência Artificial (IA). A Inteligência Artificial é um campo da ciência da computação que se concentra na criação de sistemas computacionais inteligentes, ou seja, sistemas capazes de demonstrar características que se assemelham ao comportamento humano inteligente. Essas características englobam habilidades como aprendizado, raciocínio e resolução de problemas, conforme destacado por Fernandes (2003).

A aplicação da IA na agricultura pode ser especialmente valiosa para analisar dados complexos do solo, prever as necessidades das plantas, otimizar a aplicação de fertilizantes e outros insumos, bem como desenvolver estratégias de manejo personalizadas para diferentes tipos de solo, contribuindo assim, para uma agricultura mais eficiente, sustentável e produtiva (MALUF; FLEXOR, 2017).

Neste estudo, foram utilizadas técnicas de Inteligência Artificial, como Redes Neurais Artificiais (RNAs), que são modelos matemáticos inspirados no funcionamento do cérebro humano. Essas redes têm a capacidade de serem treinadas para identificar padrões em conjuntos de dados. Ao aplicar RNAs na modelagem das características do solo, elas podem aprender a estabelecer relações entre os dados de sobrevivência disponíveis e o tamanho das amostras coletadas. Isso permite uma análise eficiente e precisa das propriedades do solo, ajudando a compreender melhor as correlações e os fatores que afetam sua sobrevivência.

Segundo Furtado (2019), as RNAs têm aplicação nas mais diversas áreas do conhecimento, em aplicações grandes e pequenas, como em áreas de engenharia (FÉLIX et al., 2017; LORENZI et al., 2017), economia (TAVARES et al., 2019; TAVARES e PENEDO, 2018), agronomia (ABRAHAM et al., 2019; LAFETRÁ et al., 2018), medicina (BARRETO et al., 2018; LIMA et al., 2019), dentre outras.

Em estudo de Freitas (2018) foi analisado a eficácia da RNA desenvolvida por

Bittar (2016), concentrando-se na avaliação da taxa de acerto. Os resultados revelaram que a rede neural apresentou-se como uma alternativa ágil para prover *insights* acerca das características do solo, mesmo quando a quantidade de amostras era limitada. Como sugestão para avanços futuros, o autor indicou a necessidade de pesquisas mais aprofundadas com o propósito de aperfeiçoar ainda mais as RNAs.

O algoritmo de aprendizagem tem a função de modificar os pesos sinápticos da rede objetivando o ajuste na classificação da situação problema (HAYKIN, 2001). As RNAs de múltiplas camadas (RNAs MLP) apresentam em sua topologia a camada de entrada, uma ou mais camadas intermediárias e a camada de saída. As RNAs MLP têm capacidade para resolver problemas de extração de características, classificação, categorização/clusterização, estimativa, previsão, otimização, aproximação de funções, dentre outras.

Outro ramo da IA utilizado neste estudo foram as técnicas de aprendizado de máquina por meio do Algoritmo Floresta Aleatória (RF) em que o algoritmo utiliza a divisão do espaço de amostras, a estruturação hierárquica dos nós e a combinação das decisões de múltiplas árvores para criar um modelo robusto e geralmente mais preciso. Essa abordagem permite lidar com complexidade nos dados, mitigar o sobreajuste (*overfitting*) e fornecer resultados mais consistentes e confiáveis (ZHOU et al., 2014).

Com base no exposto, surgiu o interesse em utilizar a IA, empregando RNAs e o Aprendizado de Máquina por meio do RF, para prever atributos químicos do solo e determinar qual dos modelos utilizados apresenta o melhor desempenho. Essa abordagem visa aproveitar as capacidades de aprendizado e generalização dessas técnicas para obter estimativas mais eficientes.

Para isso, foi de suma importância considerar a relação entre as distâncias geográficas das coordenadas x e y entre os pontos amostrais conhecidos e aqueles que estão sendo alvo de estimativa. Após o treinamento e a seleção das RNAs, a expectativa foi que elas pudessem modelar a variabilidade espacial com o uso de uma quantidade reduzida de amostras de solo.

O propósito foi desenvolver a capacidade de prever atributos químicos do solo, como cálcio (Ca), potencial hidrogeniônico (pH), fósforo (P), potássio (K), magnésio (Mg), cálcio + magnésio (Ca+Mg), matéria orgânica (MO) e um atributo fundamental para a fertilidade do solo, conhecido como capacidade de troca catiônica (CTC). Para alcançar esse propósito, foi estabelecido um modelo de entrada de dados para o

treinamento das RNAs e para o Aprendizado de Máquina RF, considerando tanto a autocorrelação quanto as distâncias entre as amostras de solo.

Portanto, as RNAs e o RF aplicados neste trabalho objetivou estimar os atributos químicos do solo com intuito de reduzir a quantidade de amostras do solo, comparando quais das metodologias utilizadas no treinamento obtiveram melhores resultados nas previsões.

2. REVISÃO DE LITERATURA

2.1. Atributos dos solos do cerrado

O Cerrado brasileiro ocupa quase toda área central do país, com cerca de 2.045.000 km², o que corresponde a 2,04 milhões de hectares, tomando maior parte na região Centro-Oeste (EMBRAPA, 2021). A maior parte dos solos sob vegetação de cerrado possui baixa fertilidade natural, indicando interferência direta na disponibilidade de nutrientes para as plantas se desenvolverem (EMBRAPA, 2017). Essa baixa fertilidade dos solos pode ser corrigida, mesmo não sendo um obstáculo para o cultivo em áreas agrícolas (KLINK; MACHADO, 2005).

A partir da década de 1970, o Cerrado se transformou em uma nova e importante fronteira agrícola brasileira. Essa transformação modificou os aspectos socioeconômicos regionais e impulsionou a produtividade agropecuária, tornando o Brasil um dos principais produtores mundiais de commodities agrícola (PRODES, 2018).

Nesse sentido, a heterogeneidade é uma característica intrínseca dos solos e sua variabilidade (espacial, horizontal e vertical) é dependente dos fatores de formação, podendo ser influenciada pelo tipo de manejo adotado (EMBRAPA, 2019). O cultivo resulta em alterações, aumentando ainda mais a variabilidade dos atributos do solo (SANTOS et al., 2006; SOUZA et al., 2007). Para Rosa Junior et al. (2012), as práticas de manejo podem provocar alterações em algumas características físicas e/ou químicas do solo.

O uso indevido de recursos naturais geralmente resulta em um ecossistema insustentável. Como resultado, solos inadequadamente usados acabam sendo degradados (BONINI et al., 2015). O Cerrado brasileiro apresenta grande parte de sua vegetação original convertida em produção agrícola, principalmente nas áreas de planícies (BROSSARD; BARCELLOS, 2005; FERREIRA *et al.*, 2009), resultando em rápida perda principalmente do carbono orgânico (RANGEL; SILVA, 2007).

Com isso, a estrutura do solo se tornou alvo de pesquisas para impulsionamento das plantações, especialmente avaliando o arranjo das partículas do solo e do espaço poroso, dentre elas o tamanho, forma e o arranjo dos agregados que resultam do agrupamento de partículas primárias em unidades separáveis (FERREIRA, 2010).

A qualidade do solo está associada com os componentes físicos-químicos e

biológicos, assim entender a distribuição espacial dos atributos dos solos é de suma importância (ROSA et al., 2015). Carneiro et al. (2013) aponta que ter conhecimento dos atributos de qualidade dos solos agrícolas e monitorá-los são imprescindíveis para prever as degradações e evitá-las, garantindo manutenção da biodiversidade no bioma.

2.2. Redes Neurais Artificiais (RNAs)

As RNAs são sistemas computacionais compostos por unidades de processamento, também chamado neurônios artificiais ou nodos, que são conectadas mutuamente de maneira exclusiva para executar uma tarefa estabelecida (BINOTI, 2010; BINOTI et al., 2012). Já os neurônios artificiais podem ser definidos como modelos matemáticos simplificados de neurônios biológicos e acionam informações recebidas e gerando respostas únicas (HAYKIN, 2001; BRAGA et al., 2007).

Na busca por compreender como as propriedades do solo impactam a produtividade e na aplicação desses dados para prever o rendimento de grãos, visando identificar locais específicos para planejar investimentos, aumentar a produtividade e prever lucros, uma alternativa promissora é o uso das RNAs (LEAL et al., 2013).

A aplicação de RNAs nesse contexto permite a análise de dados complexos do solo e sua relação com a produção de grãos de uma forma que pode ser difícil de alcançar por meio de métodos tradicionais. As RNAs têm a capacidade de aprender padrões e relações não-lineares nos dados, o que é crucial para a compreensão das nuances das propriedades do solo e seu impacto na produtividade agrícola. Ao treinar uma RNA com dados históricos de solo e rendimento de grãos, é possível criar um modelo que pode ser usado para fazer previsões futuras de rendimento com base nas características do solo em um local específico (BINOTI et al., 2013).

Isso torna as RNAs uma ferramenta valiosa para o planejamento de investimentos agrícolas, pois permitem tomar decisões informadas sobre como otimizar o uso do solo e maximizar a produtividade, o que, por sua vez, pode levar a previsões mais precisas de lucros. Assim, o uso de RNAs representa uma abordagem promissora para a gestão inteligente de recursos agrícolas e o aumento da eficiência na agricultura (LEAL et al., 2013).

Contudo, compreender os fatores que influenciam a produtividade é essencial para alcançar o sucesso produtivo em sistemas modernos de produção em larga escala, especialmente no Cerrado (SOARES et al., 2014). Esse sistema carece de introdução

de arranjos diferenciais, também conhecido como métodos de agricultura de precisão (MALUF; FLEXOR, 2017). Nesse sentido, a agricultura de precisão é definida como um conjunto de técnicas que permitem o manejo localizado e seu sucesso depende de três elementos: informação, tecnologia e manejo (SRINIVASAN, 2006).

2.3. Perceptron Multicamadas (MLP)

O MLP é um tipo de rede neural artificial amplamente utilizado em tarefas de processamento de máquinas e reconhecimento de padrões. É uma das arquiteturas mais básicas e fundamentais de redes neurais e serve como base para outras estruturas mais complexas (SILVA NETTO, 2006). A estrutura da MLP é inspirada na organização do cérebro humano e é composta por várias camadas de neurônios interconectados, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (AFFONSO, 2013).

Sendo assim, as MLP são compostas pela camada de entrada, camadas ocultas, função de ativação, camada de saída, propagação direta (*forward pass*), e o treinamento. O MLP pode ser usado para resolver uma ampla variedade de problemas, incluindo classificação de imagens, reconhecimento de fala, previsão de séries temporais e muito mais (SANTOS, 2022).

No entanto, ela pode apresentar algumas restrições em problemas com conjuntos de dados grandes e complexos, devido ao risco de *overfitting* ou a uma necessidade de poder computacional significativo para o treinamento (SILVA NETTO, 2006).

2.4. Redes Neurais Artificiais (RNAS) aplicadas na agricultura

As RNAS têm o potencial de mudar a agricultura, sendo que parte dessa mudança já está sendo vista com o uso crescente da inteligência artificial para analisar os dados climáticos obtidos por observações diretas, imagens de satélites e outras fontes de informação (TAVARES et al., 2019; TAVARES; PENEDO, 2018). Esta análise ajuda as autoridades agrícolas a tomar decisões inteligentes sobre como usar os recursos disponíveis para obter os melhores resultados (LOPES; FICHEMAN; SAGGIO, 2023).

A inteligência artificial tem o potencial de aumentar significativamente a eficiência na agricultura, auxiliando os produtores a monitorar o clima, as condições do solo, a presença de insetos e outros fatores (BASSO et al., 2019). Isso possibilita que os

agricultores tomem medidas proativas para atingir a máxima produtividade com o mínimo esforço e recursos, promovendo uma agricultura mais sustentável e eficaz (ROCHA, 2021).

As RNAs estão emergindo como uma nova ferramenta significativa para os agricultores modernos (ANDRADE et al., 2010). O conceito de armazenar e processar dados informáticos para auxiliar na produção de alimentos/agricultura vem sendo desenvolvido há algum tempo, mas as redes neurais, que são mais recentes, permitem que os agricultores identifiquem padrões e tendências complexos a partir de seus dados (FERLIN; REZENDE, 2019).

Outra vantagem importante das RNAs na agricultura é o uso de imagens para planejar a colheita de alimentos (LINHARES; MAYKLYNS, 2014). Essas imagens são capturadas por satélites e processadas por meio de inteligência artificial para gerar informações detalhadas sobre as condições de crescimento das plantas e o quanto rentável cada área poderá ser (DOMINGOS, 2017). Com isso, os agricultores podem identificar problemas quase que imediatamente e decidir se já é hora de plantar outra cultura, para que a produção seja maximizada (FREITAS, 2016).

Em estudos de Freire Capuchinho (2021) foi realizada uma investigação voltada para a utilização de RNAs baseadas em *Deep Learning*, com o intuito de estimar as temperaturas máximas (T_{max}), mínimas (T_{min}) e médias (T_{med}) do ar em escala mensal na região do Centro-Oeste do Brasil. Os resultados desta pesquisa apontam para indícios promissores quanto à viabilidade do emprego de RNAs, em particular àquelas fundamentadas em técnicas de *Deep Learning*, demonstrando um grau aceitável de precisão nas estimativas das temperaturas mensais máximas, médias e mínimas na mencionada região geográfica.

2.5. Algoritmo Floresta Aleatória (RF)

O RF é uma técnica de aprendizado de máquina que consiste em agrupar várias árvores de decisão criadas a partir de uma base de treinamento (SANTOS, 2023). O modelo resultante consolida os resultados de todas as árvores para fazer previsões (DA COSTA FILHO et al., 2019). O método foi proposto por Leo Breiman em 2001 e amplamente utilizado em tarefas de ciência de dados e aprendizado de máquina. A criação das árvores é realizada por meio da seleção aleatória de subconjuntos de atributos, formando uma coleção de árvores de decisão (SILVA et al., 2019).

Em geral, as Random Forests alcançam boa acurácia preditiva em comparação com outros métodos supervisionados de aprendizado de máquina (RIQUETI et al., 2018). Esse tipo de modelo é comumente utilizado não apenas para classificar, mas também para regressão, estudo de importância e seleção de variável, além de detecção de outliers (DE ALVARENGA JÚNIOR, 2018).

A estimativa de atributos químicos do solo no Cerrado utilizando o RF é uma aplicação prática e promissora no campo das ciências do solo. A conquista dos atributos químicos do solo é fundamental para entender a fertilidade do solo e para tomar decisões adequadas em práticas agrícolas, como adubação e manejo das culturas (CANTERL, 2020).

O RF é uma técnica de aprendizado de máquina que se enquadra na categoria de "*ensemble learning*" (aprendizado de conjunto). São construídas várias decisões em paralelo durante o treinamento, cada uma delas usando diferentes subconjuntos dos dados de treinamento e recursos. Em seguida, o RF combina as previsões das árvores para produzir uma previsão final mais robusta e precisa (LUIZ, 2022).

2.6. Algoritmo Floresta Aleatória (RF) e inteligência artificial aplicados na agricultura

A pesquisa de Ferreira (2023) representa um exemplo significativo da aplicação bem-sucedida do algoritmo de aprendizado de máquina RF na agricultura. Nesse estudo, o autor explorou a capacidade do algoritmo em prever a produtividade da cana-de-açúcar com base em atributos do solo e práticas de manejo em áreas comerciais no noroeste paulista. Os resultados obtidos oferecem *insights* valiosos sobre a eficácia desses modelos preditivos gerados pelo RF, demonstrando que eles se aproximam significativamente das observações reais. Esta pesquisa destaca a relevância e o potencial da aplicação da aprendizagem de máquina na agricultura para aprimorar a tomada de decisões e a produtividade nas atividades agrícolas.

Em estudos de Soares et al. (2014) teve como objetivo apresentar uma proposta de metodologia para estimativa da curva de retenção de água em solos do Rio Grande do Sul, utilizando redes neurais artificiais. Para isso, foi montado um banco de dados com informações disponíveis na literatura, de textura e estrutura dos solos. Utilizou-se o *software* Matlab para as previsões, no qual foram treinadas diferentes arquiteturas, variando os números de neurônios na camada de entrada e camada intermediária. Os resultados foram coerentes para utilização na área agrícola.

Em estudos de Meirelles, Moreira e Coelho (2004) utilizou-se técnicas de inteligência computacional, especificamente Redes Bayesianas e as Redes Neurais, como ferramenta de análise espacial em projetos de agricultura de precisão para o manejo da fertilidade do solo e nutrição de plantas. O estudo permitiu um ambiente de manipulação simultânea das variáveis georreferenciadas de solo e planta, juntamente com a produtividade da cultura, o que é extremamente laborioso por métodos usuais.

Outro estudo mais recente de Vitória et al. (2022) utilizou RNAs e regressão não linear para estimar a emissão de CO₂ em função do manejo do solo, e de suas propriedades físicas e químicas. Os resultados mostraram que as RNAs são mais precisas na determinação das relações entre a emissão de CO₂ e a temperatura, teor de água no solo, densidade do solo e carbono orgânico, quando comparado com os resultados de regressão linear.

3. MATERIAL E MÉTODOS

3.1. Descrição e localização da área do estudo

Os dados utilizados no desenvolvimento deste trabalho foram os atributos físicos químicos oriundos da dissertação de Reis (2011). Esses dados foram obtidos em uma área localizada na Fazenda Palmital, numa extensão irrigada por pivô central, localizada no município de Morrinhos, em Goiás, conforme localização na Figura 1, com altitude de 813 m, latitude 17°45'S e longitude 49°10'W. A área apresenta latossolo vermelho escuro, com textura franco argilo arenosa.

Figura 1 - Localização da área de coleta dos dados experimentais, na Fazenda Palmital, Morrinhos-GO.



Fonte: Google (2023)

3.2. Mapeamento e georrefenciamento dos pontos amostrais

Segundo Reis (2011) os dados foram coletados e georreferenciados em um talhão de 23 hectares em uma grade amostral de 50 x 50 m, totalizando 120 pontos amostrais. Cada ponto amostral foi georreferenciado com o um aparelho de GPS (Sistema de Posicionamento Global) Garmin modelo Etrex 36 Legend RoHs com o sistema de correção diferencial em tempo real via satélite e com o datum ajustado ao sistema SAD

69.

3.3. Atributos físicos e químicos do solo

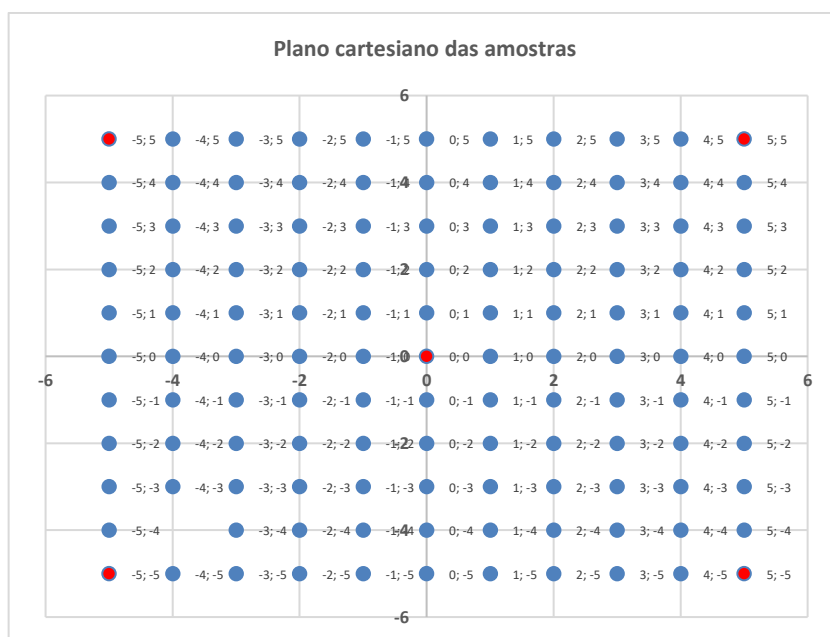
Conforme Reis (2011) foram coletadas, em um raio de um metro em torno de cada ponto, amostras deformadas, na camada de 0,0 a 0,2 m, com trado de rosca. Estas amostras foram embaladas com suas correspondentes identificações e encaminhadas para um laboratório de análise de solos certificado. Determinou-se os seguintes atributos:

- Químicos: Potencial Hidrogeniônico (pH); cálcio (Ca); magnésio (Mg); cálcio + magnésio (Ca+Mg); alumínio trocável (Al); acidez potencial (H+Al); potássio (K); fósforo (P); saturação por alumínio (m%), e; matéria orgânica (MO).;
- Fertilidade: capacidade de troca catiônica (CTC) e saturação por bases (V%);
- Físicos (composição granulométrica do solo): areia, silte e argila.

3.4. Distribuição das amostras no plano cartesiano

Neste estudo foi necessário identificar as amostras em coordenadas cartesianas X e Y que representam a posição das amostras em um plano bidimensional conforme Figura 2.

Figura 2 - Plano cartesiano das posições das amostras em relação aos pontos fixos



Fonte: Autor (2023)

Para calcular a distância entre cada amostra e os pontos fixos (dados com as coordenadas X e Y), foi utilizada a fórmula da distância euclidiana. A distância euclidiana é uma medida de distância entre dois pontos em um espaço euclidiano, como um plano cartesiano.

A equação utilizada para calcular a distância entre duas coordenadas (x1, y1) e (x2, y2) no plano cartesiano é a fórmula da distância euclidiana de acordo com Cunha (2022).

Distância euclidiana Equação 1:

$$Distância = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

onde:

- (x1, y1) são as coordenadas da primeira amostra ou ponto fixo;
- (x2, y2) são as coordenadas da segunda amostra ou ponto fixo;
- (x2 - x1) é a diferença entre as coordenadas x da segunda e primeira amostra ou ponto fixo;
- (y2 - y1) é a diferença entre as coordenadas y da segunda e primeira amostra ou ponto fixo;
- (x2 - x1)² é o quadrado da diferença entre as coordenadas x;
- (y2 - y1)² é o quadrado da diferença entre as coordenadas y;
- (x2 - x1)² + (y2 - y1)² é a soma dos quadrados das diferenças;
- $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ é a raiz quadrada da soma dos quadrados, que representa a distância euclidiana.

Em seguida, a menor distância é utilizada para identificar o ponto fixo mais próximo de cada amostra, e os valores de nutrientes correspondentes a esse ponto fixo são usados para prever os valores de nutrientes da amostra.

No estudo, a fórmula da distância euclidiana foi aplicada dentro de um laço para calcular a distância entre cada amostra e os pontos fixos. A amostra mais próxima de cada ponto fixo foi encontrada calculando-se a distância para todos os pontos fixos e selecionando o índice do ponto fixo com a menor distância.

Assim, a posição de cada amostra foi relacionada aos pontos fixos, e os valores dos nutrientes associados ao ponto fixo mais próximo foram utilizados como rótulos (targets) para treinar as RNAs e o RF.

3.5. Definição das variáveis de entrada

Os seguintes atributos cálcio (Ca), Potencial Hidrogeniônico (pH), fósforo (P), potássio (K), magnésio (Mg), cálcio + magnésio (Ca+Mg), matéria orgânica (MO) e catiônica (CTC) foram utilizados como parâmetros de entrada para as RNAs e para o RF, coordenadas cartesianas X e Y de cada amostra e distância euclidiana em relação as pontos fixos.

3.6. Aplicação da inteligência artificial

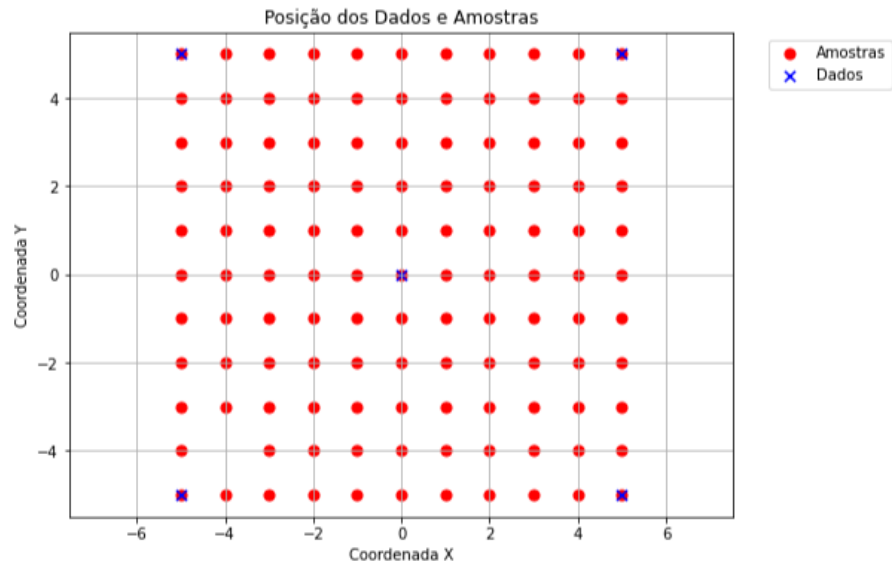
3.6.1. Treinamento das Redes Neurais Artificiais (RNAs) e do Algoritmo Floresta Aleatória (RF)

Para o treinamento das RNAs e do RF, foram desenvolvidos três modelos denominados A, B e C cada um utilizando diferentes conjuntos de amostras do solo. O modelo A foi criado com cinco pontos amostrais utilizando o RF, enquanto o modelo B utilizou nove pontos amostrais utilizando também o RF e o Modelo C utilizou as RNAs MLP, sendo que esses pontos amostrais foram denominados pontos fixos, os quais serviram como entrada para o algoritmo RF e para as RNAs. A distribuição dos pontos foi considerada para abranger uma distribuição espacial, com um ponto central e pontos laterais.

Cada amostra utilizada no estudo é composta por oito atributos químicos do solo: Cálcio (CA), pH do solo (PH), Fósforo (P), Potássio (K), Magnésio (Mg), Matéria Orgânica (MO) e Capacidade de Troca Catiônica (CTC). Além disso, cada amostra possui informações de coordenadas espaciais representando seus respectivos eixos X e Y. Esses atributos foram utilizados como entradas para os modelos A (Figura 3), B (Figura 4) e C (Figura 5) permitindo o processamento e análise dos dados das amostras por meio de redes neurais. Essa abordagem possibilitou a realização de previsões obtenção

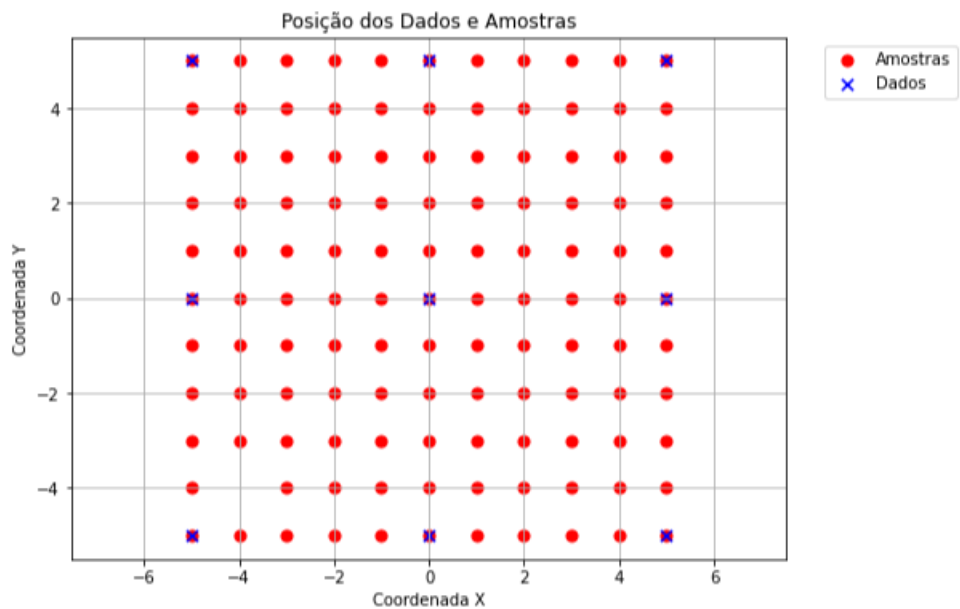
de resultados relevantes e contextualizados.

Figura 3 - Gráfico da posição dos dados e amostras do modelo A no plano cartesiano



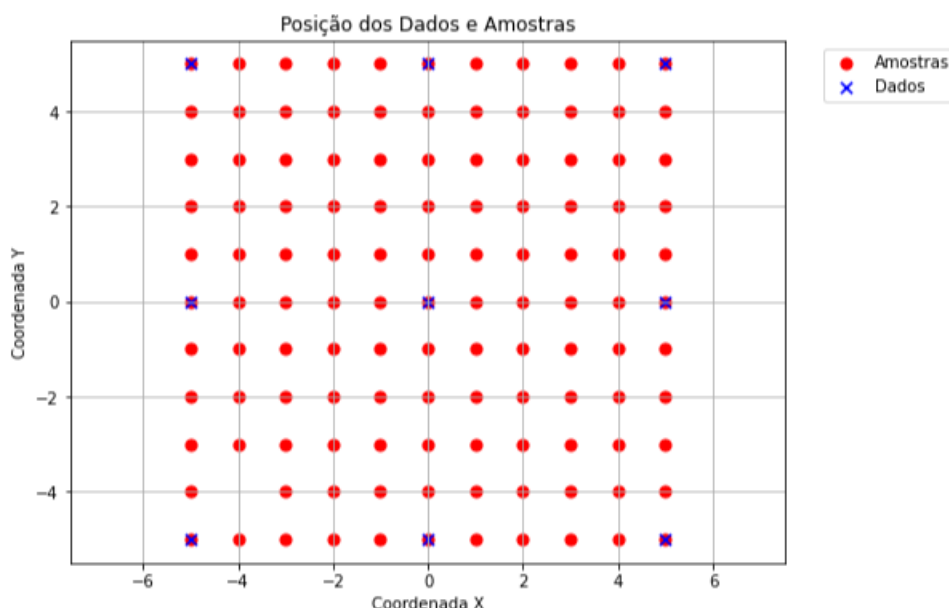
Fonte: Autor (2023).

Figura 4 - Gráfico da posição dos dados e amostras do modelo B no plano cartesiano



Fonte: Autor (2023)

Figura 5 - Gráfico da posição dos dados e amostras do modelo C no plano cartesiano.



Fonte: Autor (2023)

As RNAs e o algoritmo RF foram arquitetados para processar os dados dessas amostras, permitindo a identificação e extração de padrões e relações entre os atributos presentes em cada uma delas. Por meio desse processamento, foram realizadas previsões com base nas características dos atributos, proporcionando resultados relevantes e contextualizados para cada conjunto de amostras.

A inclusão das coordenadas dos eixos X e Y de cada amostra foi uma abordagem crucial, visto que permitiu que a RNAs e o algoritmo RF levasse em consideração a localização espacial das amostras durante o processo de previsão. Essa consideração é particularmente relevante em contextos em que a posição geográfica das amostras desempenha um papel fundamental na caracterização dos dados e na interpretação dos resultados.

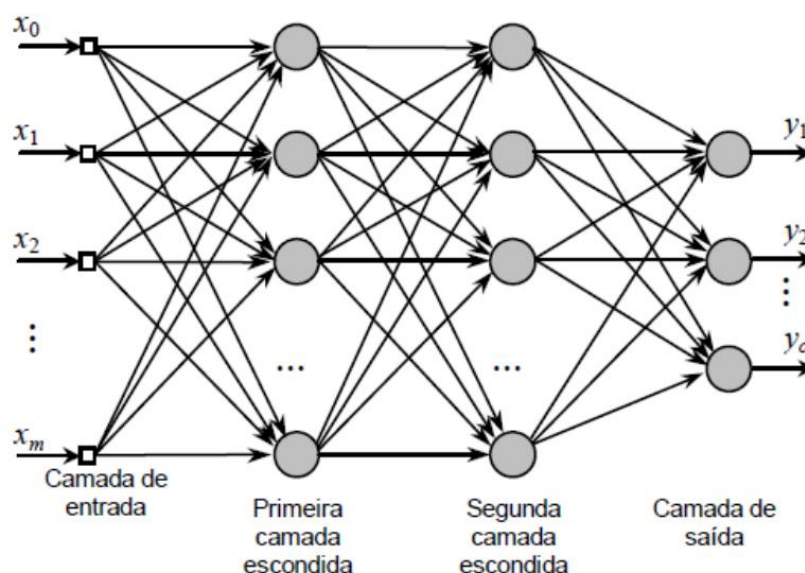
Além disso, durante o processo de treinamento das redes neurais, foram calculadas as distâncias das amostras em relação aos pontos fixos. Essa abordagem visou melhorar e auxiliar na previsão dos atributos químicos do solo, permitindo que as redes neurais levassem em consideração a proximidade espacial das amostras com os pontos fixos, o que pode contribuir para resultados mais precisos e confiáveis.

3.7. Aplicação das Redes Neurais Artificiais (RNAs)

3.7.1. Redes Neurais Artificiais (RNAs) – Perceptrons de Multicamadas (MLP)

O modelo utilizado para treinamento das RNAs foi o Perceptrons de Multicamadas (MLP) como demonstrado na Figura 6. Esse modelo é amplamente utilizado para tarefas de regressão em que a relação entre as variáveis de entrada (características) e a variável de saída (alvo) é complexa e não linear. Bittar (2018) utilizou esta metodologia para preparação, treinamento e aplicação da rede neural artificial proposta pelos autores Braga et al. (2012) que corroboram com Russel e Norvig (2013).

Figura 6 - Modelo Perceptrons de Multimarcas (MLP)



Fonte: Google (2023)

Além disso, Boccato e Attux (2023) citam a equação geral para o funcionamento de um neurônio em uma camada oculta ou de saída, e pode ser descrita como Cálculo do valor de ativação do neurônio (antes da função de ativação):

Cálculo do valor de ativação do neurônio Equação 2:

$$z_j = \sum_i = 1 n w_{ij} \cdot x_i + b_j$$

Sendo que:

- z_j é o valor de ativação do neurônio j na camada atual.

- w_{ij} é o peso sináptico entre o neurônio i na camada anterior e o neurônio j na camada atual.
- x_i é o valor de entrada do neurônio i na camada anterior.
- b_j é o viés (bias) do neurônio j na camada atual.

Aplicação da função de ativação:

Função de ativação Equação 3:

$$a_j = \text{ativacao}(z_j)$$

Sendo que:

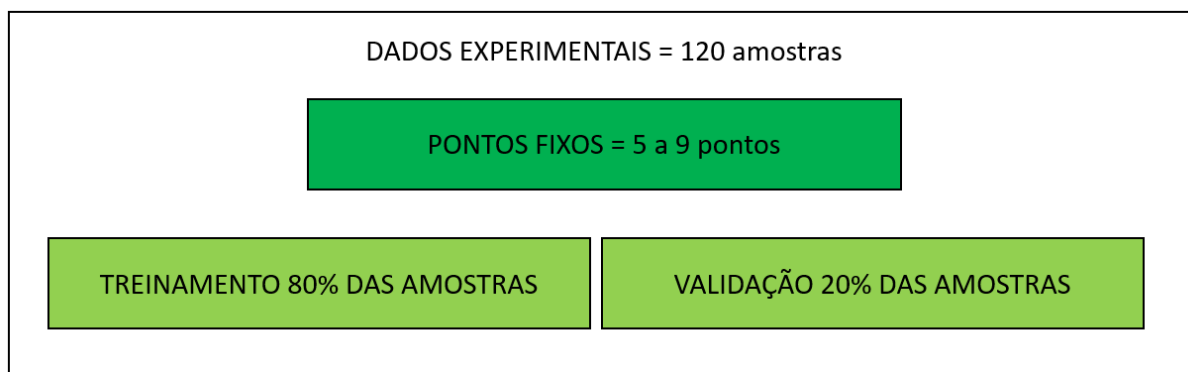
- a_j é a saída ativada (output) do neurônio j na camada atual.
- $\text{ativacao}(\cdot)$ é a função de ativação utilizada foi a função ReLU (Rectified Linear Unit).

Para a implementação da MLP no código, foi utilizado o modelo MLPRegressor da biblioteca Scikit-learn. Neste modelo, a função de ativação padrão é a função ReLU. A arquitetura da MLP é definida pela quantidade de neurônios nas camadas ocultas, que foi configurada como (100, 50) no estudo, indicando duas camadas ocultas com 100 e 50 neurônios, respectivamente.

3.7.1.1. Conjunto de Treinamento e Teste

As amostras e seus respectivos valores de nutrientes são divididos em dois conjuntos: conjunto de treinamento com 80% dos dados e conjunto de teste com 20%. O conjunto de treinamento é usado para treinar o modelo MLP, enquanto o conjunto de teste é usado para avaliar o desempenho do modelo após o treinamento, conforme divisão apontada na Figura 7.

Figura 7 - Conjunto de treinamento com 80% dos dados e conjunto de teste com 20%



Fonte: Autor (2023)

Abaixo está o resumo dos dados amostrais:

- Total de amostras: 120 amostras
- Treinamento: 96 amostras
- Validação: 24 amostras
- Pontos fixos: de 5 a 9 pontos.

3.7.1.2. Validação Cruzada (K-Fold)

Para evitar o *overfitting*, sendo um fenômeno comum em aprendizado de máquina e estatística que ocorre quando um modelo se adapta excessivamente aos dados de treinamento, em vez de generalizar a partir deles, deve-se garantir a generalização do modelo. Para isso, a validação cruzada K-Fold foi empregada com intuito de treinar e avaliar o modelo de MLP com diferentes combinações de dados, garantindo uma melhor estimativa do desempenho geral do modelo a representação da validação cruzada, que pode ser observada na Figura 8.

Figura 8 - Representação do K-fold

DADOS					
TREINAMENTO					TESTE
FOLD 1	FOLD 2	FOLD 3	FOLD 4	FOLD 5	
FOLD 1	FOLD 2	FOLD 3	FOLD 4	FOLD 5	
FOLD 1	FOLD 2	FOLD 3	FOLD 4	FOLD 5	
FOLD 1	FOLD 2	FOLD 3	FOLD 4	FOLD 5	
FOLD 1	FOLD 2	FOLD 3	FOLD 4	FOLD 5	
Métrica	Métrica	Métrica	Métrica	Métrica	

Fonte: Autor (2023)

O processo de validação cruzada K-Fold é realizado da seguinte maneira:

1. O conjunto de treinamento é dividido em K partes, ou "folds".
2. O modelo é treinado K vezes, utilizando K-1 partes como conjunto de treinamento e a parte restante como conjunto de validação. Ou seja, em cada iteração, um fold diferente é usado como conjunto de validação, enquanto os demais K-1 folds são usados como conjunto de treinamento.
3. Para cada iteração, o modelo é treinado com os dados de treinamento e avaliado com os dados de validação. As métricas de avaliação, como o Erro Quadrático Médio (MSE), são calculadas para cada interação.
4. Ao final das K interações, as métricas de avaliação são coletadas e a média é calculada para obter uma estimativa mais confiável do desempenho do modelo.

Neste estudo, a validação cruzada K-Fold foi realizada com 5 folds, o que significa que o conjunto de treinamento é dividido em 5 partes e o modelo é treinado e avaliado 5 vezes. O parâmetro **n_splits=5** é usado para definir o número de folds na função **KFold**.

A fórmula para calcular o número de folds é dada por:

Equação 4:

$$K = N/M$$

onde:

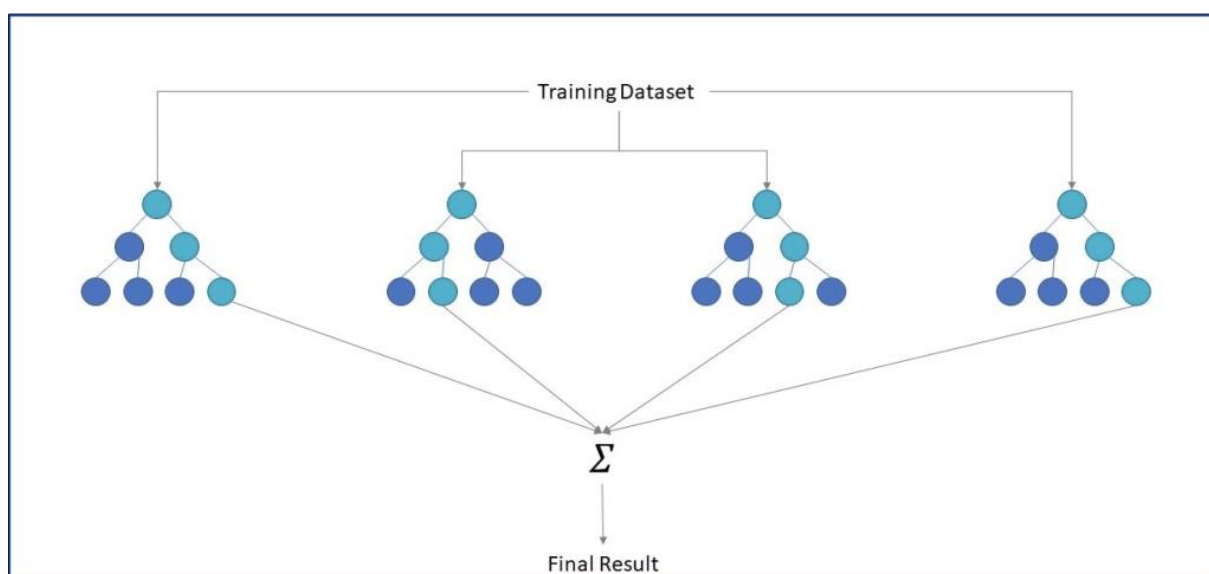
- K é o número de folds (partes) em que o conjunto de treinamento será dividido.
- N é o número total de amostras no conjunto de treinamento.
- M é o tamanho de cada fold, ou seja, o número de amostras em cada parte.

A validação cruzada K-Fold é uma técnica útil para avaliar a performance do modelo de forma mais robusta, pois permite utilizar todos os dados disponíveis tanto para treinamento como para validação, evitando possíveis vieses de seleção de dados.

3.7.2. Estrutura do Algoritmo Floresta Aleatória (RF)

No estudo foi utilizado o modelo de Regressão com Floresta Aleatória (*Random Forest Regressor*), mostrado na Figura 9. Nesse caso, o modelo árvores de decisão (definido pelo parâmetro "n_estimators" ao inicializar o `RandomForestRegressor`). Cada árvore é treinada com um subconjunto aleatório dos dados de treinamento.

Figura 9 - Modelo de Regressão com Floresta Aleatória (Random Forest Regressor)



Fonte: Arora (2021)

3.7.2.1. Entradas do Modelo de Regressão com Floresta Aleatória

Em relação às camadas, o modelo de Floresta Aleatória não tem camadas no sentido tradicional, como redes neurais. Em vez disso, é uma técnica de aprendizado de máquina baseada em árvores de decisão, em que cada árvore é uma unidade independente.

Nas entradas do modelo foram utilizadas as coordenadas X e Y, representadas pelas colunas 'X' e 'Y' dos dados de amostras, respectivamente. Essas coordenadas são

usadas para determinar quais são os pontos mais próximos no conjunto de dados de treinamento a partir de um cálculo de distância euclidiana. Com base nas coordenadas mais próximas encontradas, o modelo fez a previsão dos atributos de solo representados pelas colunas 'CA', 'PH', 'P', 'K', 'Mg', 'Ca+Mg', 'MO' e 'CTC'.

3.7.2.2. Sidas do Modelo de Regressão com Floresta Aleatória

As saídas do modelo são os valores previstos para os atributos de solo (CA, PH, P, K, Mg, Ca+Mg, MO e CTC) para as amostras de teste. O objetivo é estimar os valores desses atributos com base nas coordenadas X e Y das amostras de teste e nas informações do conjunto de dados de treinamento.

A fórmula geral do Algoritmo Floresta Aleatória (RF) para regressão é uma média das previsões de todas as árvores construídas durante o treinamento. Dada uma entrada X, a previsão y_{pred} é calculada como:

Equação 5:

$$y_{pred} = (1/n) * \sum(y_{pred_i}),$$

onde:

- A. y_{pred} é a previsão final para a entrada X;
- B. n é o número total de árvores construídas durante o treinamento;
- C. y_{pred_i} é a previsão da i-ésima árvore para a entrada X.

Cada árvore individual, por sua vez, realiza uma série de decisões com base nos atributos do conjunto de dados de entrada X para chegar a uma previsão. A previsão é geralmente a média (ou a moda no caso de classificação) das previsões feitas em cada uma das árvores.

É importante destacar que o Floresta Aleatória utiliza a randomização para criar diversidade entre as árvores, evitando o overfitting e melhorando a precisão geral do modelo. Essa randomização ocorre ao construir cada árvore, onde apenas um subconjunto de atributos é considerado para cada divisão de nó, e também na amostragem dos dados de treinamento para construção de cada árvore.

3.7.2.3. Índice de GINI

Durante a construção da árvore de decisão em uma Floresta Aleatória, o algoritmo busca dividir os dados em nós de forma que a impureza seja reduzida, utilizando critérios como o índice de Gini. A divisão que resulta em uma maior redução da impureza é escolhida para cada nó da árvore (CAPELO, 2022).

A fórmula para o cálculo do índice de Gini é a seguinte:

Equação 6:

$$Gini(t) = 1 - \sum (p_i^2)$$

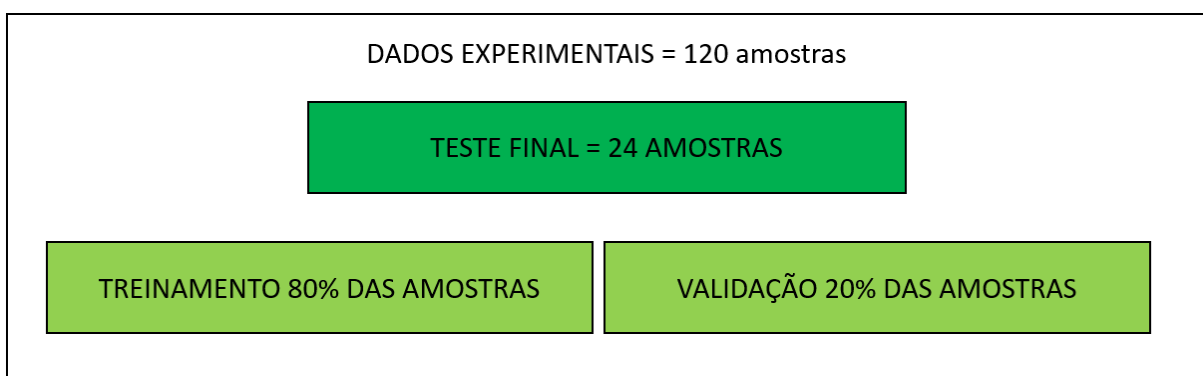
onde:

- Gini(t) é o índice de Gini do nó da árvore t,
- p_i é a proporção de amostras da classe i no nó t.

3.7.2.4. Divisão do Conjunto de Dados em Treinamento e Teste

Assim como no modelo de MLP com uso das redes neurais foram utilizados 80% dos dados para treinamento e 20% para teste (Figura 10).

Figura 10 - Divisão do conjunto de dados em treinamento



Fonte: Autor (2023)

Abaixo encontra-se o resumo das amostras:

- Total de amostras: 120 amostras
- Treinamento: 96 amostras
- Validação: 24 amostras

- Pontos fixos: de 5 a 9 pontos.

3.1.1. Ferramentas para construção dos algoritmos

Para construir o RF e RNAs, foram necessárias algumas bibliotecas e ferramentas em Python. Abaixo estão as principais ferramentas que foram usadas neste estudo:

- NumPy: É uma biblioteca fundamental para computação científica em Python.
- Pandas: É uma biblioteca de análise de dados que fornece estruturas de dados flexíveis e eficientes, para manipular e analisar dados de forma poderosa.
- scikit-learn: Também conhecido como sklearn, é uma das bibliotecas mais populares para aprendizado de máquina em Python.
- Matplotlib e Seaborn: Essas são bibliotecas para visualização de dados em Python.
- Jupyter Notebook: Utilizado para construir o algoritmos.

3.7.2.5. Validação das Redes Neurais Artificiais (RNAs) e do Algoritmo Floresta Aleatória (RF)

Os resultados das RNAs e o RF foram desnormalizados para se obter os índices estatísticos dos modelos. A finalidade da validação das redes propostas foi de analisar o ajuste dos valores estimados pelas RNAs. Para tal, foram calculados os índices estatísticos da Raiz Quadrática de Erro Médio (RQEM), do Erro Médio Absoluto (EMA), Erro Quadrático Médio (EQM), o coeficiente de correlação de Pearson®. Estes índices foram obtidos para as parcelas (20%) que não serão utilizadas na etapa de treino. Todos os índices estatísticos foram calculados conforme as Equações (7 a 10) no programa MSOffice Excel.

Raiz Quadrática de Erro Médio na Equação 7:

$$RQEM = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}$$

Erro Médio Absoluto na Equação 8:

$$EMA = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i|$$

Erro Quadrático Médio na Equação 9:

$$EQM = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

Coeficiente de correlação de Pearson na Equação 10:

$$r = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})(Y_i - \bar{Y})}{\sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2}{n}} * \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}}$$

Em que:

Y_i = atributos de entrada, em °C;

$\hat{Y}_i$ = valor estimado pelas RNAs, em °C;

$\bar{Y}$ = média dos atributos de entrada, em °C;

n = número de dados utilizados;

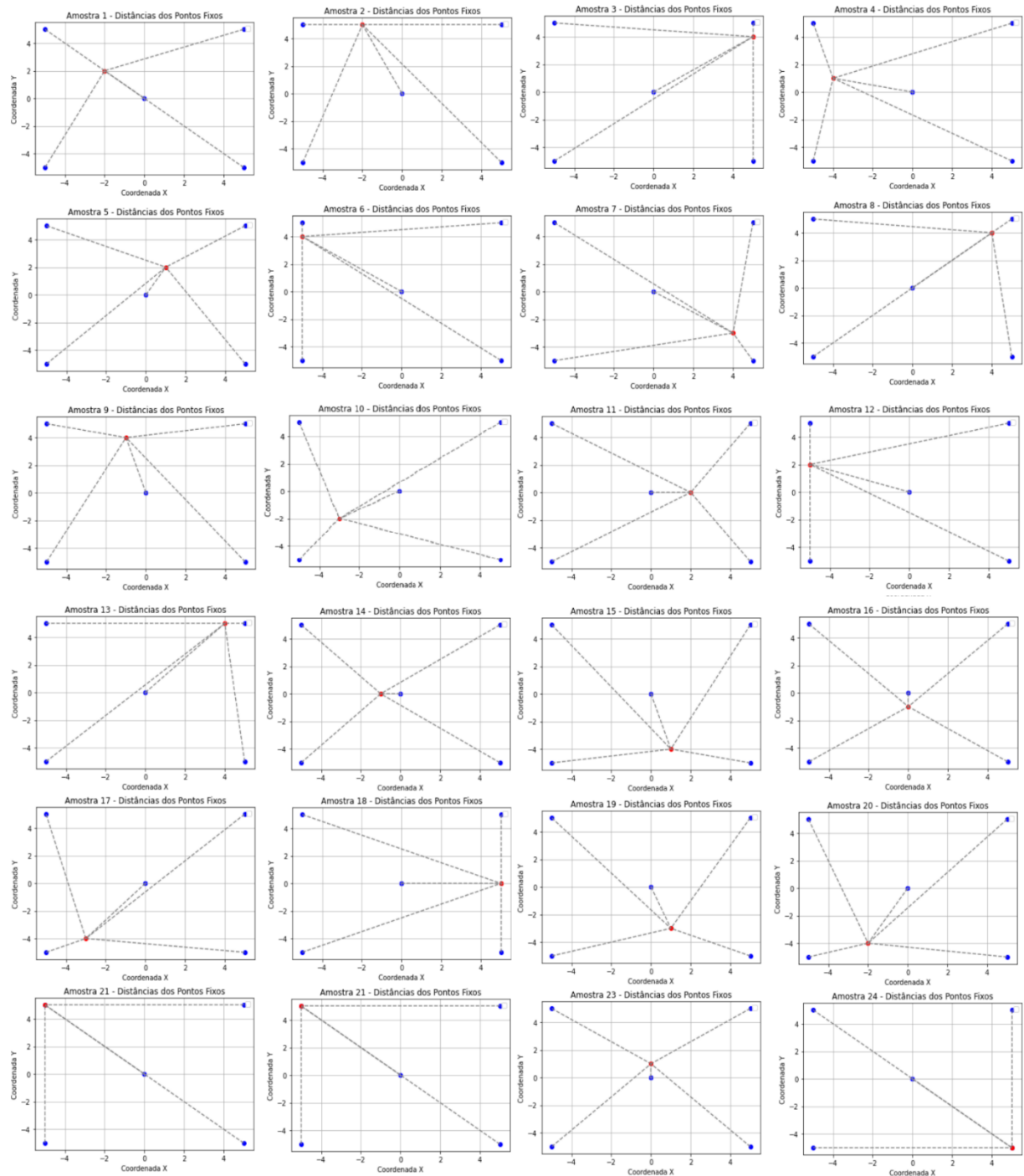
3.7.3. Distâncias das amostras em relação aos pontos amostrais

Nas Figuras 11 e 12 apontam uma representação visual de como cada amostra prevista pelo modelo se relaciona com os pontos fixos (dados) utilizados para o treinamento e teste do modelo. As linhas tracejadas indicam as distâncias entre as amostras previstas e os pontos fixos, permitindo uma melhor compreensão da disposição espacial das amostras previstas em relação aos dados reais.

Essa visualização é especialmente útil em análises de regressão, pois facilita uma avaliação visual de como o modelo está se comportando em relação aos dados de entrada e como as previsões estão distribuídas no plano cartesiano. Segue as Figuras

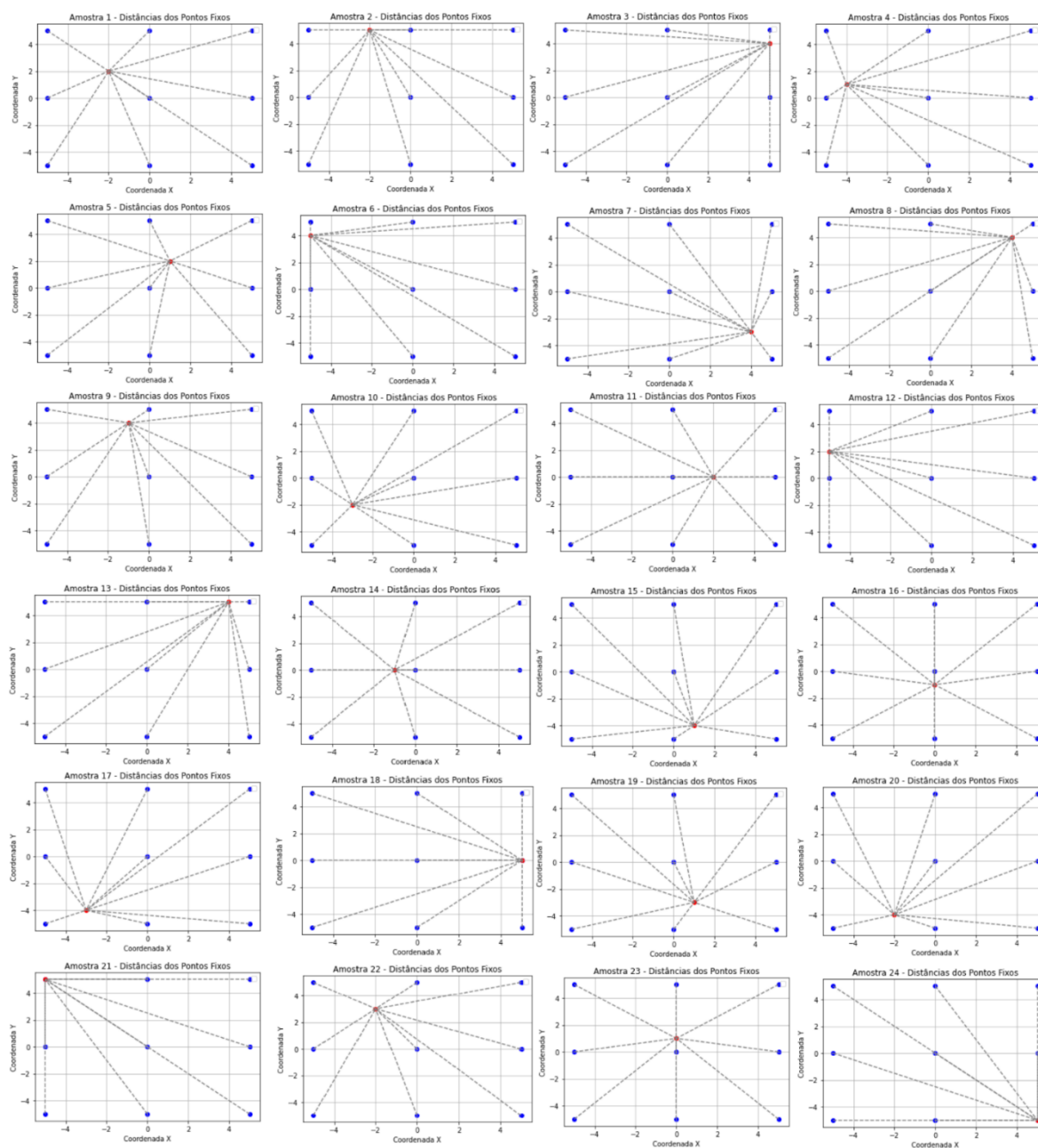
11 e 12.

Figura 11 - Representações das distâncias das amostras previstas em relação aos 5 pontos fixos (Modelo A)



Fonte: Autor (2023)

Figura 12 - Representações das distâncias das amostras previstas em relação aos 9 pontos fixos (Modelo B).



Fonte: Autor (2023)

3.7.4. Análise Estatística descritiva e exploratória

Os dados das características químicas do solo obtidos da base de dados em estudo, quanto os dados previstos pelas RNAs, foram submetidos à estatística descritiva por meio da análise exploratória dos dados, com o intuito de entender e visualizar a tendências de dispersão e distribuição.

Foram realizados os cálculos das seguintes medidas descritivas: média, mediana, curtose, variância, desvio padrão, assimetria e coeficiente de variação. Para verificar a normalidade dos dados ($P < 0,05$), foi realizado o teste de Shapiro e Wilk (1965). Para as análises estatísticas foi utilizado o software SISVAR.

4. RESULTADOS E DISCUSSÃO

4.1. Análise descritiva dos dados

Os resultados das análises descritivas para os atributos químicos do solo avaliados estão apresentados na Tabela 1, em que são mostradas as estatísticas descritivas, incluindo a média, mediana (ou valor central), moda (valor mais frequente), variância, desvio padrão, curtose, coeficiente de variação (CV%) e p-value para cada atributo.

Tabela 1 - Análise descritiva dos atributos químicos do solo

Atributo	MÉDIA	MD	MODA	VAR	DP	CURTOSE	CV%	P-VALUE
PH	5,0133	5,0000	4,7000	0,1086	0,3295	0,8811	6,5722	0,0000
CA	2,4758	2,2500	2,1000	0,8472	0,9204	1,1239	37,1773	0,0000
MG	0,2858	0,2000	0,2000	0,0259	0,1610	0,3397	56,3386	0,0000
Cama	2,7617	2,5000	2,3000	1,0844	1,0413	0,3845	37,7071	0,0000
Al	0,0158	0,0000	0,0000	0,0020	0,0444	8,4553	280,6151	0,0000
CTC	5,6503	5,6850	5,0900	1,1127	1,0549	0,0229	18,6694	0,0000
P	12,9000	9,0000	8,0000	93,8723	9,6888	3,6823	75,1067	0,0000
K	0,3118	0,3070	0,3070	0,0068	0,0826	-0,7906	26,4781	0,0000
MO	1,9358	2,0000	2,0000	0,2447	0,4946	0,0243	25,5519	0,0000

pH - Potencial Hidrogeniônico; Ca – Cálcio (cmolc dm^{-3}); Mg - Magnésio (cmolc dm^{-3}); Cálcio + Mg (cmolc dm^{-3}) + Magnésio (cmolc dm^{-3}); Al – Alumínio (cmolc dm^{-3}); CTC – Capacidade de Troca Catiônica; P – Fósforo (mg dm^{-3}); K – Potássio (cmolc dm^{-3}); MO - Matéria Orgânica (g dm^{-3}); MD – Mediana; Var - Variância; DP – desvio padrão; CV(%) - Coeficiente de Variação; P-Value: teste de normalidade Shapiro Wilk a 5%.

Fonte: Autor (2023)

Na Tabela 1 observa-se que a média do pH é de aproximadamente 5,01. A mediana é 5,00 e a moda é 4,70, o que sugere que a distribuição dos dados pode estar inclinada para a esquerda (assimetria negativa). O baixo valor do coeficiente de variação (CV%) indica baixa dispersão dos dados em relação à média. O p-value de 0,0000

sugere que o valor médio do pH é significativamente diferente de zero.

No cálcio (Ca) tem a média de 2,48 a mediana é 2,25 e a moda é 2,10, o que sugere uma distribuição assimétrica à direita (assimetria positiva). O valor relativamente alto do coeficiente de variação (CV%) indica maior dispersão dos dados em relação à média. O p-value de 0,0000 sugere que o valor médio do cálcio é significativamente diferente de zero.

No magnésio (Mg) a média é aproximadamente 0,29. A mediana e a moda são ambos 0,20, o que sugere uma distribuição relativamente simétrica. O alto valor do coeficiente de variação (CV%) indica maior dispersão dos dados em relação à média. O p-value de 0,0000 sugere que o valor médio do magnésio é significativamente diferente de zero.

O cálcio e magnésio a média é de aproximadamente 2,7617. A mediana e a moda têm valores próximos, indicando uma distribuição simétrica. A variância e o desvio padrão são relativamente altos, sugerindo uma maior dispersão dos dados em relação à média. A curtose é positiva, indicando uma distribuição um pouco mais concentrada do que a distribuição normal. O coeficiente de variação é de 37%, indicando uma variabilidade relativa considerável em relação à média. O valor de p-value é 0,0000, sugerindo que a média da soma de cálcio e magnésio é significativamente diferente de algum valor de referência.

A média do alumínio (Al) é de aproximadamente 0,02. A mediana e a moda são ambos 0, o que sugere uma distribuição fortemente assimétrica à direita. O alto valor do coeficiente de variação (CV%) indica grande dispersão dos dados em relação à média. O p-value de 0,0000 sugere que o valor médio do alumínio é significativamente diferente de zero.

A média da capacidade de troca catiônica (CTC) é de aproximadamente 5,65. A mediana é 5,69 e a moda é 5,09, o que sugere uma distribuição relativamente simétrica. O valor moderado do coeficiente de variação (CV%) indica moderada dispersão dos dados em relação à média. O p-value de 0,0000 sugere que o valor médio da C.

Os dados mostram que a média de fósforo no solo é de aproximadamente 12,9000. A mediana e a moda têm valores diferentes, indicando uma distribuição assimétrica. A variância e o desvio padrão são altos, sugerindo uma grande dispersão dos dados em relação à média. A curtose é positiva, indicando que a distribuição é mais concentrada (pico) e com caudas mais pesadas do que a distribuição normal. O coeficiente de variação é alto, indicando uma grande variabilidade relativa em relação à

média. O valor de p-value é 0,0000, sugerindo que a média do fósforo no solo é significativamente diferente de algum valor de referência.

Em relação a média de potássio no solo, o valor é de aproximadamente 0,3118. A mediana e a moda têm os mesmos valores, indicando uma distribuição simétrica, e a variância e o desvio padrão são baixos, sugerindo uma baixa dispersão dos dados em relação à média. A curtose é negativa, indicando que a distribuição é menos concentrada (achatada) e com caudas mais leves do que a distribuição normal. O coeficiente de variação é moderado, indicando uma variabilidade relativa considerável em relação à média. O valor de p-value é 0,0000, sugerindo que a média do potássio no solo é significativamente diferente de algum valor de referência.

Os dados mostram que a média da matéria orgânica no solo é de aproximadamente 1,9358. A mediana e a moda têm os mesmos valores, indicando uma distribuição simétrica. A variância e o desvio padrão são moderados, sugerindo uma dispersão moderada dos dados em relação à média. A curtose é próxima de zero, indicando que a distribuição é aproximadamente similar à distribuição normal. O coeficiente de variação é moderado, indicando uma variabilidade relativa moderada em relação à média. O valor de p-value é 0,0000, sugerindo que a média da matéria orgânica no solo é significativamente diferente de algum valor de referência.

4.1.1. Análise do Modelo A: Estimativas de atributos químicos (5 pontos).

Neste modelo foram utilizados cinco pontos de amostragem fixos que a partir desses pontos, o algoritmo de floresta aleatória fez a predição. Foram obtidos os seguintes resultados para cada atributo conforme Tabela 2, evidenciado nas Figuras 13, 14 e 15.

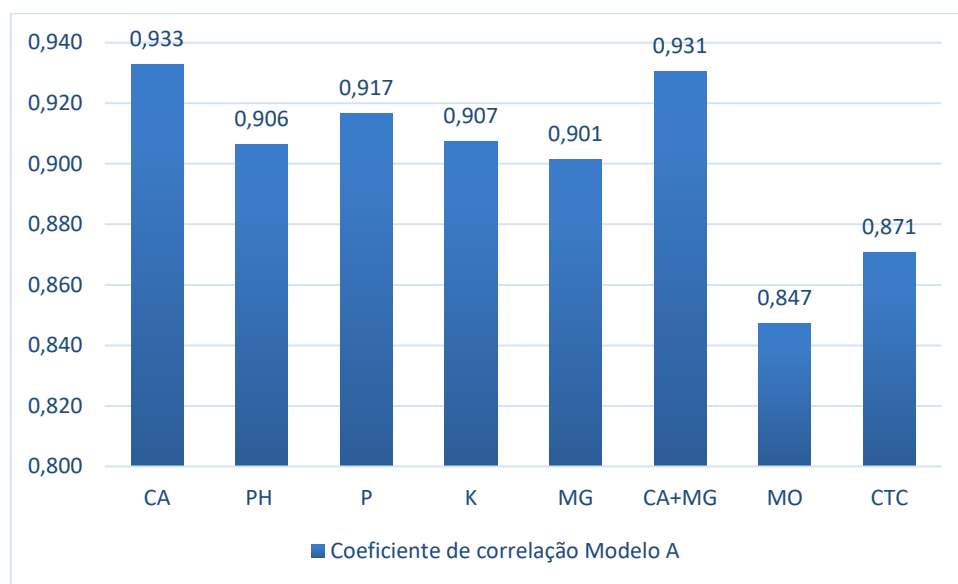
Tabela 2 - Análise estatística dos Valores Observados x Valores Previstos do Modelo A

	Média	EMQ	Coeficiente de correlação	R ²	Desvio Padrão	Erro médio
CA	2,844	0,068	0,933	0,870	0,667	0,096
PH	5,078	0,014	0,906	0,821	0,261	0,038
P	15,073	19,099	0,917	0,840	10,154	1,466
K	0,273	0,001	0,907	0,823	0,061	0,009
MG	0,390	0,005	0,901	0,813	0,149	0,021

CA+MG	3,234	0,102	0,931	0,866	0,809	0,117
MO	2,084	0,015	0,847	0,718	0,211	0,030
CTC	6,521	0,036	0,871	0,758	0,359	0,052

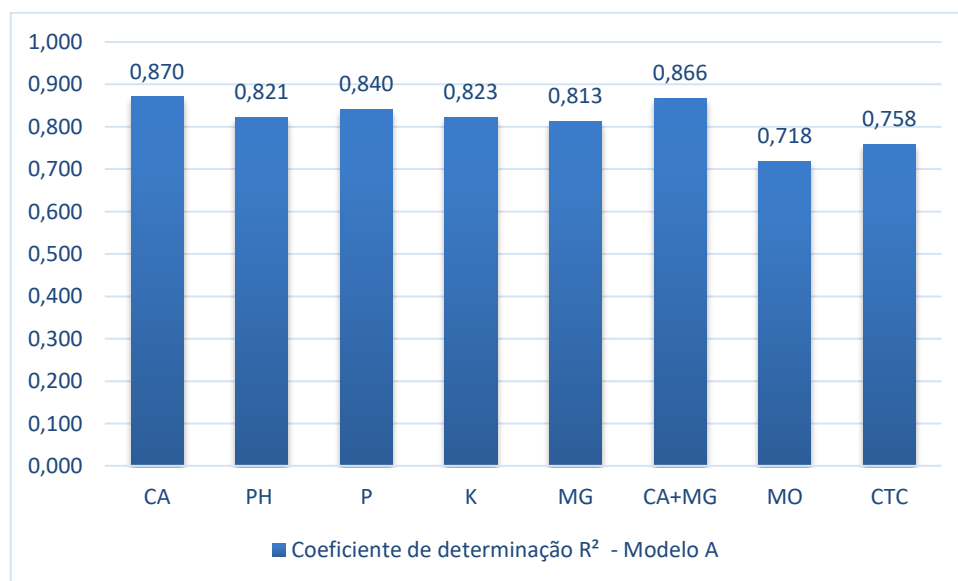
Fonte: Autor (2023)

Figura 13 - Gráfico de Correlação do Modelo A



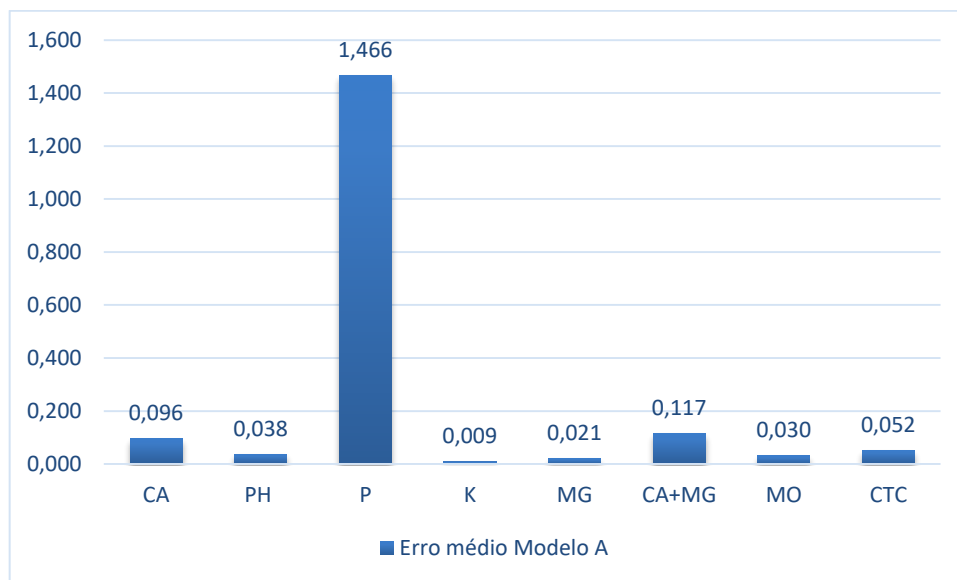
Fonte: Autor (2023)

Figura 14 - Gráfico do Coeficiente de determinação do Modelo A.



Fonte: Autor (2023)

Figura 15 - Gráfico do Erro Médio do Modelo A



Fonte: Autor (2023)

4.1.1.1. Potencial hidrogeniônico (pH) - Modelo A

No experimento, verificou-se que o Modelo A exibiu um EMQ de valor consideravelmente baixo. Esse resultado sugere que o Modelo A apresenta um ajuste satisfatório em relação aos dados observados do pH do solo. Além disso, o coeficiente de correlação do Modelo A, respectivamente (Figura 13), avaliado em 0,906, denota uma forte associação linear entre as estimativas geradas pelo Modelo A e os valores reais (observados) do pH do solo.

O valor do coeficiente de correlação varia de -1 a 1. Neste caso, o valor de 0,9063 indica uma forte correlação positiva entre os valores previstos e os valores observados do pH do solo. E o coeficiente de determinação (R^2), respectivamente (Figura 14), do Modelo A foi de 0,8213. O coeficiente de determinação (R^2) representa a proporção da variância do pH do solo que é explicada pelos valores previstos pelo Modelo A. Quanto mais próximo de 1 o valor de R^2 , melhor é a capacidade do modelo de explicar a variabilidade dos dados observados. Neste caso, o valor de 0,8213 indica que aproximadamente 82% da variância do pH do solo é explicada pelo Modelo A, o que é considerado um ajuste razoavelmente bom.

4.1.1.2. Análise do Cálcio (Ca) - Modelo A

Os valores de cálcio indicam que o modelo possui um bom ajuste aos dados observados de cálcio no solo, com uma forte correlação positiva e uma capacidade de explicar aproximadamente 87% da variabilidade dos dados. Isso sugere que o modelo é capaz de fazer boas previsões do teor de cálcio no solo com base nas variáveis independentes utilizadas no modelo.

4.1.1.3. Análise do Magnésio (Mg) - Modelo A

Os resultados mostram que o atributo "MG" (Magnésio) no solo possui uma média de 0,39. O modelo apresenta um ajuste preciso aos dados observados, como indicado pelo Erro Médio Quadrático (EMQ) de 0,0048.

Há uma forte correlação positiva entre os valores previstos pelo modelo e os valores reais do atributo "MA", com um coeficiente de correlação de 0,901, respectivamente (Figura 13). Isso sugere uma relação linear consistente entre as variáveis. O coeficiente de determinação (R^2) é de aproximadamente 0,8126, indicando que cerca de 81% da variância do atributo "MA" no solo é explicada pelas variáveis independentes incluídas no modelo de regressão.

4.1.1.4. Análise do Cálcio + magnésio (Ca+Mg) - Modelo A

Os resultados mostram que o modelo possui uma correlação positiva forte com os dados observados do atributo combinado "Ca+Mg" no solo e é capaz de explicar aproximadamente 87% da variância desses dados. Isso sugere que o modelo é capaz de fazer boas previsões do teor combinado de cálcio e magnésio no solo com base nas variáveis independentes utilizadas no modelo.

4.1.1.5. Análise da Capacidade de troca catiônica (CTC) - Modelo A

Os resultados indicam que o atributo "CTC" (Capacidade de Troca Catiônica) no solo tem uma média de 6,52. O Erro Médio Quadrático (EMQ) é calculado como 0,0357, o que sugere um ajuste relativamente bom do modelo aos dados observados de "CTC".

O coeficiente de correlação é estimado em 0,8708, respectivamente (Figura 13), o que significa que há uma correlação positiva significativa entre os valores previstos pelo modelo e os valores reais de "CTC". Além disso, o coeficiente de determinação (R^2) para o modelo é de aproximadamente 0,7583, indicando que cerca de 76% da variabilidade do atributo "CTC" no solo é explicada pelas variáveis independentes incluídas no modelo de regressão.

4.1.1.6. Análise do Fósforo (P) - Modelo A

Em síntese, os resultados revelam que o modelo A apresenta uma correlação positiva significativa com os dados observados do atributo "P" (fósforo) no solo, sendo capaz de explicar aproximadamente 84% da variância desses dados. Entretanto, o valor relativamente elevado do Erro Médio Quadrático (EMQ) sugere que o modelo ainda pode ser refinado para uma melhor adaptação aos dados observados.

4.1.1.7. Análise do Potássio (K) - Modelo A

Os resultados mostram que o atributo "K" (Potássio) no solo possui uma média de 0,27. O modelo de regressão apresenta um ajuste preciso aos dados observados, conforme indicado pelo baixo valor do Erro Médio Quadrático (EMQ) de 0,0007.

O Coeficiente de Correlação é estimado em 0,9072, evidenciando uma forte correlação positiva entre os valores previstos pelo modelo e os valores reais do atributo "K" no solo. Esse valor próximo a 1 sugere uma relação linear consistente entre as variáveis.

Além disso, o Coeficiente de Determinação (R^2) do modelo é de aproximadamente 0,8231. Esse valor indica que aproximadamente 82% da variância do atributo "K" no solo é explicada pelas variáveis independentes incluídas no modelo de regressão.

4.1.1.8. Análise da Matéria orgânica (MO) - Modelo A

Com base nos resultados obtidos, constatamos que o modelo revela uma correlação positiva significativa com os dados observados do atributo "MO" (Matéria Orgânica) no solo, demonstrando uma relação linear entre as variáveis previstas e os valores reais. Além disso, o coeficiente de determinação (R^2) do modelo é de

aproximadamente 0,72, indicando que cerca de 72% da variabilidade do atributo "MO" no solo é explicada pelas variáveis independentes incorporadas no modelo de regressão. Isso sugere que o modelo possui uma capacidade razoável de fazer previsões do teor de Matéria Orgânica no solo com base nas variáveis selecionadas.

4.1.2. Modelo B: Estimativa de atributos químicos do solo (9 pontos)

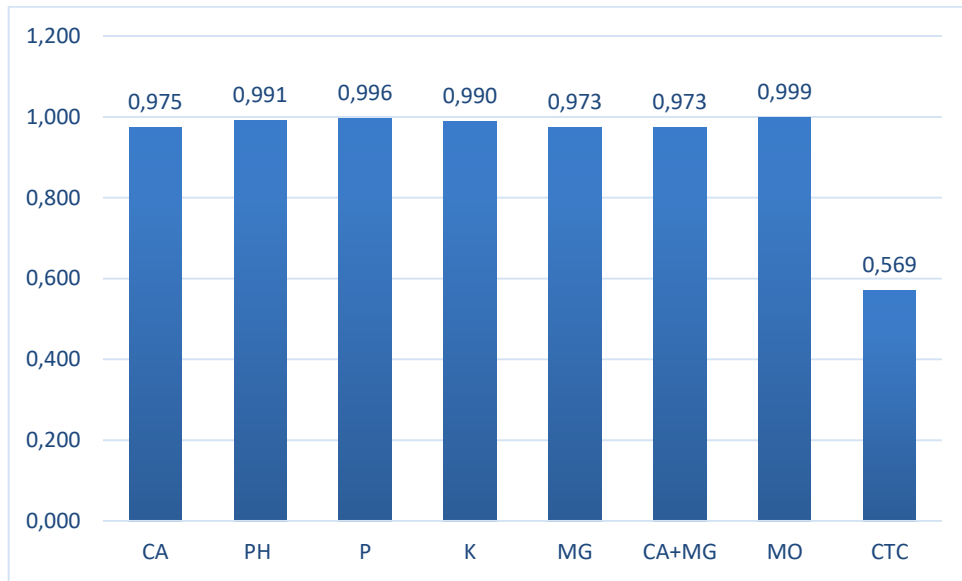
Neste escopo, foram empregados nove pontos fixos de amostragem como base para o Algoritmo Floresta Aleatória (RF) realizar suas previsões. Os resultados obtidos para cada atributo a partir desta abordagem são apresentados na Tabela 3. Os resultados podem ser evidenciados nas Figuras 16, 17 e 18.

Tabela 3 - Análise estatística dos Valores Observados x Valores Previstos do Modelo B

	Média	EMQ	Coefficiente de correlação	R²	Desvio Padrão	Erro médio
CA	2,713	0,068	0,975	0,950	1,002	0,145
PH	5,020	0,003	0,991	0,981	0,336	0,049
P	13,079	0,956	0,996	0,991	9,756	1,408
K	0,277	0,005	0,990	0,980	0,084	0,012
MA	0,314	0,002	0,973	0,946	0,155	0,022
CA+MG	3,027	0,090	0,973	0,947	1,121	0,162
MO	2,130	0,000	0,999	0,999	0,286	0,041
CTC	6,006	1,833	0,569	0,324	1,453	0,210

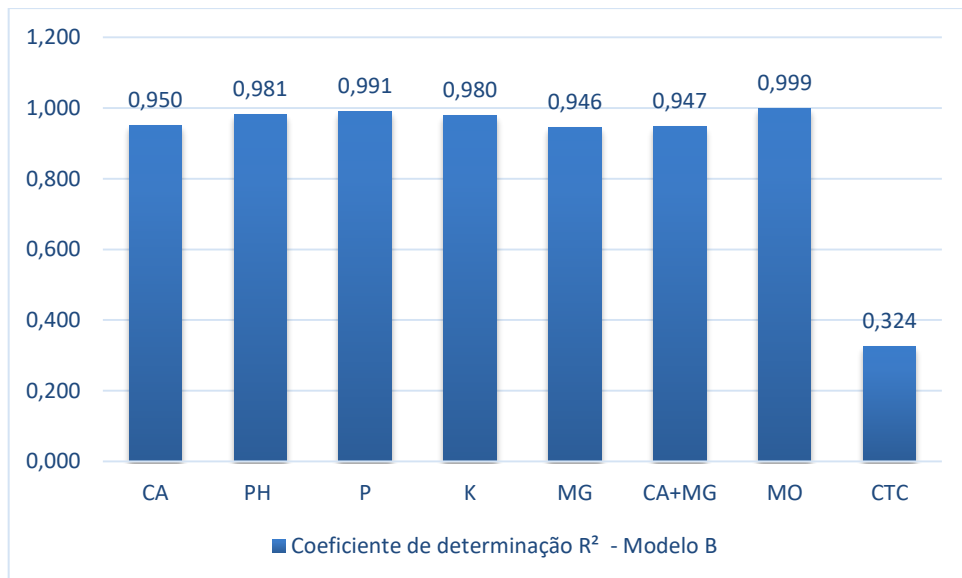
Fonte: Autor (2023)

Figura 16 - Gráfico do coeficiente de correlação do Modelo B



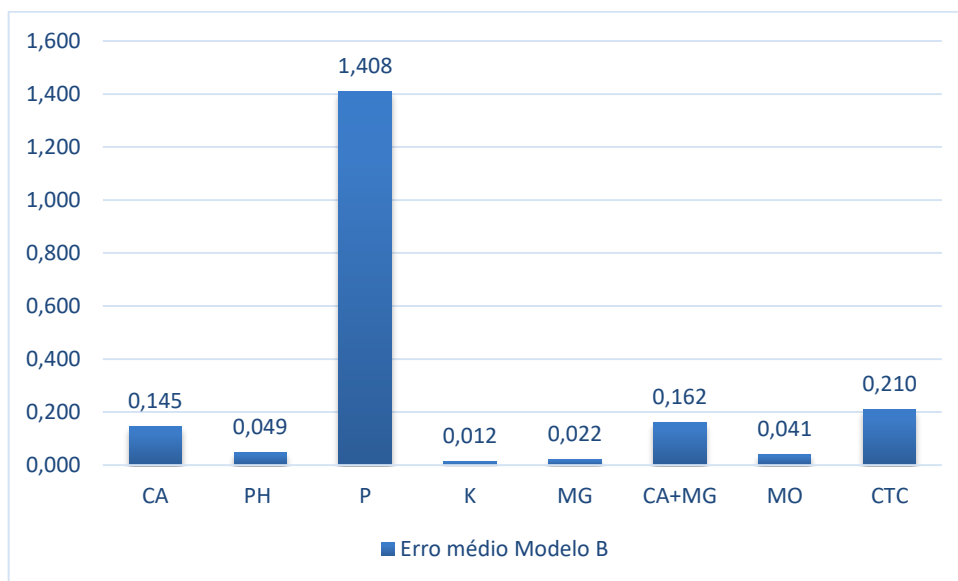
Fonte: Autor (2023)

Figura 17 - Gráfico do coeficiente de determinação do Modelo B



Fonte: Autor (2023)

Figura 18 - Gráfico do Erro Médio do Modelo B



Fonte: Autor (2023)

4.1.2.1. Análise do Potencial Hidrogeniônico (pH) - Modelo B

Os resultados indicam que o modelo B possui um alto desempenho em relação aos dados observados do atributo "pH" do solo. O coeficiente de correlação de 0,991 indica uma correlação positiva forte entre os valores previstos e observados, respectivamente (Figura 16). Além disso, o valor de R^2 de 0,981 sugere que aproximadamente 98,1% da variância dos dados de pH no solo é explicada pelo modelo, o que demonstra uma capacidade de ajuste.

O erro médio quadrático (EMQ) de 0,003 é bastante baixo, o que indica que as previsões do modelo são próximas dos valores reais. O desvio padrão de 0,336 também é relativamente pequeno, reforçando a precisão das previsões.

4.1.2.2. Análise do Cálcio (Ca) - Modelo B

Os resultados do modelo B possui um desempenho bastante satisfatório na previsão do atributo "CA" (Cálcio) do solo em relação aos valores observados. O coeficiente de correlação de 0,975 indica uma correlação muito forte entre os valores previstos pelo modelo e os valores reais observados.

Além disso, o valor elevado do R^2 de 0,950 indica que o modelo é capaz de explicar aproximadamente 95% da variabilidade dos dados, ou seja, a maioria das

variações no atributo "CA" pode ser atribuída às previsões feitas pelo modelo.

O EMQ (Erro Médio Quadrático) de 0,068 indica a média dos erros entre os valores previstos e os valores reais, respectivamente (Figura 18). Quanto menor esse valor, melhor é o desempenho do modelo, indicando que ele está fazendo previsões próximas aos valores reais.

O desvio padrão de 1,002 mostra a dispersão dos dados em torno da média, indicando a variabilidade dos resultados previstos em relação aos valores reais. Portanto, com base nessas métricas, podemos concluir que o modelo B é bastante preciso na previsão do atributo "CA" do solo, fornecendo resultados confiáveis e próximos aos valores observados.

4.1.2.3. Análise do Magnésio (Mg) - Modelo B

Os resultados obtidos por meio do Modelo B revelam uma correlação positiva consideravelmente alta entre os valores previstos e os valores observados da característica "MA" do solo. Isso é enfatizado pelo coeficiente de correlação notavelmente alto de 0,973, respectivamente (Figura 17), juntamente com um coeficiente de determinação (R^2) bastante significativo, atingindo o valor de 0,946.

Esses coeficientes servem como indicadores robustos da excepcional capacidade preditiva do Modelo B em relação ao teor de Magnésio (MG) presente no solo. Essa capacidade preditiva é fundamentada na utilização cuidadosa das variáveis independentes incorporadas durante o processo de treinamento do modelo. Como resultado, essas descobertas corroboram e fortalecem a confiança científica na habilidade do Modelo B em fornecer prognósticos notavelmente precisos dentro desse domínio específico.

Adicionalmente, o baixo valor do Erro Médio Quadrático (EMQ) sugere que as previsões do modelo estão geralmente em concordância com os valores observados, e o desvio padrão de 0,155 denota uma consistência nas previsões realizadas pelo modelo.

Em síntese, os resultados indicam que o modelo B é eficaz na previsão do teor de Magnésio no solo, tendo a capacidade de explicar aproximadamente 94,6% da variação nos dados. Isso ressalta a utilidade do modelo para realizar previsões confiáveis do atributo "MA", considerando tanto as características químicas como as coordenadas espaciais das amostras utilizadas no treinamento.

4.1.2.4. Análise do Cálcio + magnésio (Ca+Mg) - Modelo B

Na previsão dos atributos químicos CA+MG o Modelo B demonstra um elevado desempenho. O coeficiente de correlação de 0,973 aponta para uma sólida relação positiva entre os valores previstos e observados. Ademais, o valor de R^2 de 0,947 indica que aproximadamente 94,7% da variância dos dados é explicada pelo modelo, representando um resultado notável.

O Erro Médio Quadrático (EMQ) de 0,090 é uma medida que reflete a precisão das previsões em relação aos valores reais. Neste caso, um EMQ baixo sugere que as previsões do modelo se aproximam dos valores observados, o que é desejável em análises preditivas.

O desvio padrão de 1,121 representa a dispersão dos dados em relação à média e quanto menor o desvio padrão, mais próximos os valores estão da média. O valor relativamente baixo do desvio padrão sugere que as previsões do modelo estão bem agrupadas em torno da média.

Os resultados indicam que o Modelo B possui uma sólida capacidade de previsão para o atributo químico CA+MG, com alta correlação entre valores previstos e observados, além de uma boa explicação da variabilidade dos dados.

4.1.2.5. Análise da Capacidade de troca catiônica (CTC) - Modelo B

Os resultados obtidos para o atributo "CTC" no Modelo B revelam um bom desempenho da rede neural em relação aos valores previstos e observados. A média dos valores previstos (6,138) está próxima da média dos valores observados, indicando uma boa capacidade da rede em fazer previsões próximas aos dados reais. O Erro Médio Quadrático (EMQ) apresenta um valor baixo de 0,158, o que sugere que as previsões da rede estão bastante próximas dos valores reais, resultando em pouca variação entre eles.

O Coeficiente de Correlação 0,974 é próximo de 1, indicando uma correlação positiva forte entre as previsões da rede e os valores reais do atributo "CTC". Esse valor significa que as previsões da rede estão alinhadas com os dados observados, fortalecendo a confiabilidade do modelo.

O R^2 (0,948) também é próximo de 1, representando a proporção da variabilidade

dos dados que é explicada pelas previsões da rede. Nesse caso, cerca de 94,8% da variância dos valores do atributo "CTC" pode ser explicada pelas previsões da rede neural. Além disso, o Desvio Padrão 1,193 indica a dispersão dos valores em relação à média. Um valor baixo sugere que os valores previstos estão bem agrupados e próximos à média, aumentando a precisão do modelo.

Os resultados indicam que o Modelo B apresenta uma capacidade de previsão para o atributo "CTC", com correlação positiva forte, alta explicação da variância e baixo erro entre os valores previstos e observados. Isso mostra que a rede neural é capaz de realizar previsões acuradas para o atributo "CTC" com base nas informações das amostras do solo e suas coordenadas espaciais.

4.1.2.6. Análise do Fósforo (P) - Modelo B

O Modelo B apresenta resultados precisos e confiáveis para o atributo "P" (Fósforo) no solo. Os valores previstos estão correlacionados com os valores reais, o que é evidenciado pelo alto coeficiente de correlação (0,996) e pelo coeficiente de determinação (R^2) de 0,991, indicando que cerca de 99,1% da variância dos dados é explicada pelo modelo. Além disso, o erro médio é baixo (1,408), mostrando que as previsões estão próximas dos valores reais, e o EMQ (Erro Médio Quadrático) é pequeno (0,956), evidenciando que os erros das previsões são baixos.

Esses resultados indicam que o Modelo B é capaz de realizar previsões acuradas para o teor de Fósforo no solo, o que é de extrema importância para estudos agrícolas e ambientais. A precisão das previsões fornecidas pelo modelo pode ser fundamental para a tomada de decisões e o manejo adequado do solo, auxiliando na melhoria da produtividade agrícola e na sustentabilidade do ecossistema.

4.1.2.7. Análise do Potássio (K) - Modelo B

Os resultados do modelo B para o atributo "K" (Potássio) do solo apresentam uma correlação muito alta com os dados observados, com um coeficiente de correlação de 0,990 e um coeficiente de determinação (R^2) de 0,980. O coeficiente de correlação próximo de 1 indica uma relação forte e positiva entre as previsões do modelo e os valores reais observados para o atributo "K".

Além disso, o erro médio quadrático (EMQ) para o atributo "K" foi igual a zero, o

que significa que o modelo conseguiu prever exatamente os valores reais para esse atributo. O desvio padrão também é baixo (0,083), indicando que as previsões estão próximas dos valores reais. Esses resultados sugerem que o modelo B é preciso na previsão do teor de Potássio no solo com base nas informações disponíveis e características das amostras.

4.1.2.8. Análise da Matéria orgânica (MO) - Modelo B

Os resultados apresentados para o atributo "MO" (Matéria Orgânica) do solo no Modelo B mostram que o modelo possui um desempenho bom na previsão desse atributo. A média dos valores previstos é de aproximadamente 2.130, e o Erro Médio Quadrático (EMQ) é zero, o que indica que as previsões estão praticamente iguais aos valores observados.

O Coeficiente de Correlação (R) é de aproximadamente 0.999, e o coeficiente de determinação (R^2) também é de 0.999, ambos muito próximos de 1. Esses valores de correlação são extremamente altos, indicando uma correlação positiva forte entre os valores previstos e os valores observados. Além disso, o desvio padrão é de aproximadamente 0.286, o que sugere que os valores previstos estão próximos dos valores observados, com baixa dispersão.

Em resumo, os resultados indicam que o modelo B tem um bom desempenho na previsão da Matéria Orgânica do solo, apresentando alta precisão e acurácia nas previsões. Isso significa que o modelo é capaz de fazer previsões próximas aos valores reais, tornando-o uma ferramenta confiável para a estimativa desse atributo em estudos futuros.

4.2. Modelo C – Aplicação de Redes Neurais Artificiais (RNAS)

Neste modelo foram utilizados nove pontos de amostragem fixos que a partir desses pontos a rede neural artificial fez a predição, obtendo os seguintes resultados para cada atributo abaixo especificado.

4.2.1. Avaliação das previsões dos atributos do Modelo C

No modelo C, foram empregados nove pontos fixos de amostragem como base

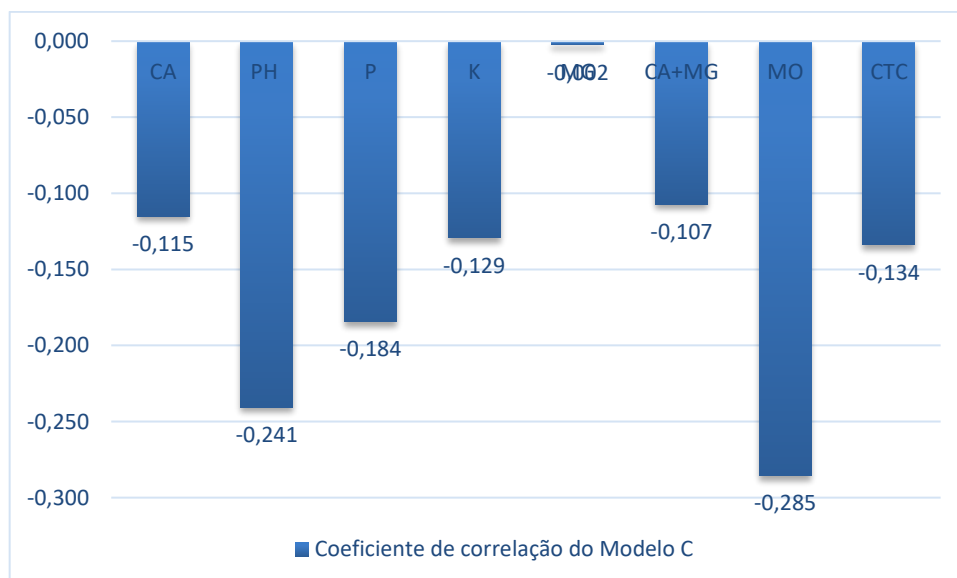
para o Redes Neurais Artificiais realizar suas predições. Os resultados obtidos para cada atributo a partir desta abordagem são os seguintes conforme Tabela 4. Pode ser observado nas Figuras 19, 20 e 21, os resultados em gráficos

Tabela 4 - Análise estatística dos Valores Observados x Valores Previstos do Modelo C

	Média	EMQ	Coefficiente de correlação	R ²	Desvio Padrão	Erro médio
CA	2,636	1,723	-0,115	0,013	0,892	0,129
PH	4,819	1,310	-0,241	0,058	0,743	0,107
P	12,183	172,911	-0,184	0,034	8,646	1,248
K	0,281	0,031	-0,129	0,017	0,118	0,017
MA	0,302	0,042	-0,002	0,000	0,146	0,021
CA+MG	2,890	2,126	-0,107	0,011	0,992	0,143
MO	2,060	0,262	-0,285	0,081	0,319	0,046
CTC	5,886	4,198	-0,134	0,018	1,364	0,197

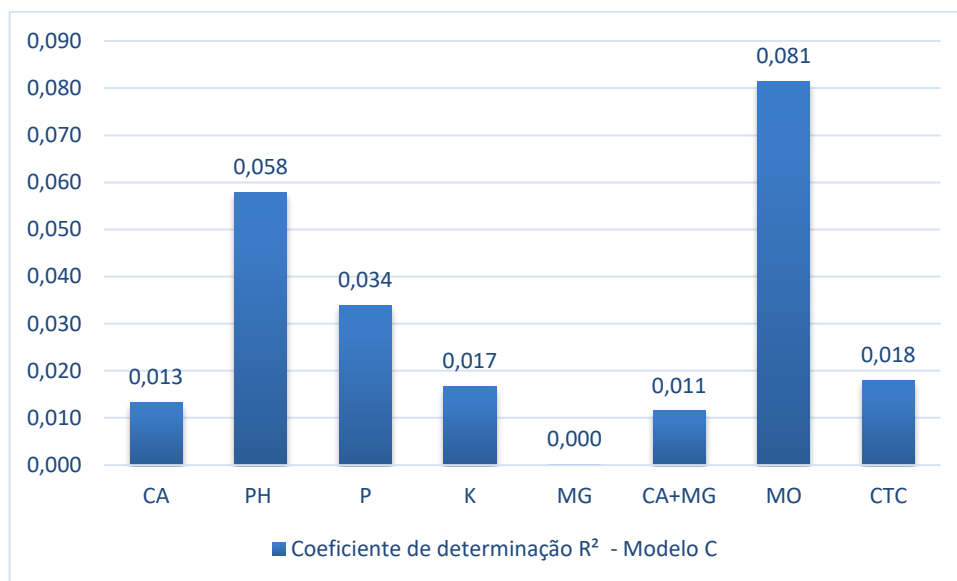
Fonte: Autor (2023)

Figura 19 - Gráfico do Coeficiente de Correlação do Modelo C



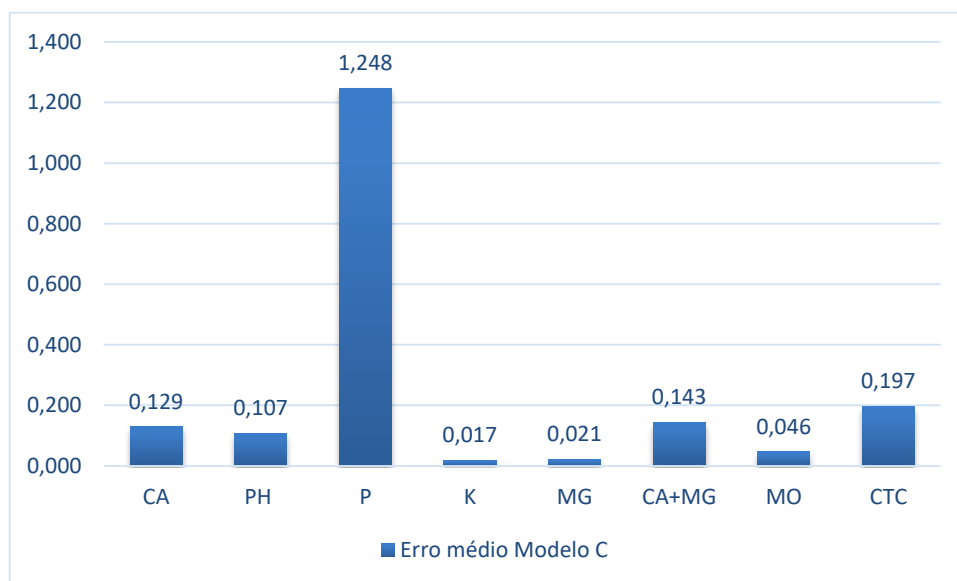
Fonte: Autor (2023)

Figura 20 - Gráfico do Coeficiente de determinação do Modelo C



Fonte: Autor (2023)

Figura 21 - Gráfico do Erro médio do Modelo C



Fonte: Autor (2023)

O modelo de regressão de redes neurais apresentou desempenhos distintos na previsão das variáveis do solo. Na previsão do Cálcio (CA), o modelo apresentou desempenho limitado, indicado por um coeficiente de correlação próximo de zero conforme (Figura 19) e um baixo valor de R^2 conforme (Figura 20), sugerindo que não conseguiu capturar adequadamente os padrões nos dados, resultando em previsões

imprecisas e uma dispersão significativa entre as previsões e os valores reais.

Já na previsão do pH do solo, os valores de coeficiente de correlação e R^2 negativos indicam que o modelo não se ajustou perfeitamente aos dados, mas ainda foi capaz de capturar parte da variação. O valor médio de erro quadrático médio (EMQ) foi de 1.31, sugerindo uma dispersão média em relação aos valores reais, e o desvio padrão e erro médio indicam que as previsões estão relativamente próximas dos valores reais em média, mas com alguma variabilidade nas previsões.

No entanto, o modelo apresentou desempenho insatisfatório na previsão do fósforo (P) no solo, indicado pelos valores de coeficiente de correlação e R^2 negativos, bem como pelo alto valor de erro quadrático médio (EMQ) de 172.91, o que sugere previsões com dispersão significativa e erros consideráveis.

Para o potássio (K) no solo, o modelo apresentou uma correlação fraca, com valores próximos de zero para o coeficiente de correlação e R^2 . No entanto, o baixo valor de erro quadrático médio (EMQ) de 0.0305 e o desvio padrão de 0.1182 indicam previsões relativamente próximas dos valores reais.

Na previsão do magnésio (MG) no solo, o modelo conseguiu capturar parcialmente a relação entre as variáveis de entrada e o magnésio no solo, embora a correlação seja fraca. Os baixos valores de erro quadrático médio (EMQ) de 0.042 e desvio padrão de 0.1461 mostram que as previsões estão próximas dos valores reais, indicando um bom desempenho do modelo nessa variável.

Para a soma de cálcio e magnésio (CA+MG) no solo, o modelo não conseguiu capturar bem a relação entre as variáveis de entrada e a soma, conforme indicado pelos valores próximos de zero para o coeficiente de correlação e R^2 . O alto valor de erro quadrático médio (EMQ) de 2.1257 e desvio padrão de 0.9915 mostram previsões com alto grau de dispersão e erros significativos.

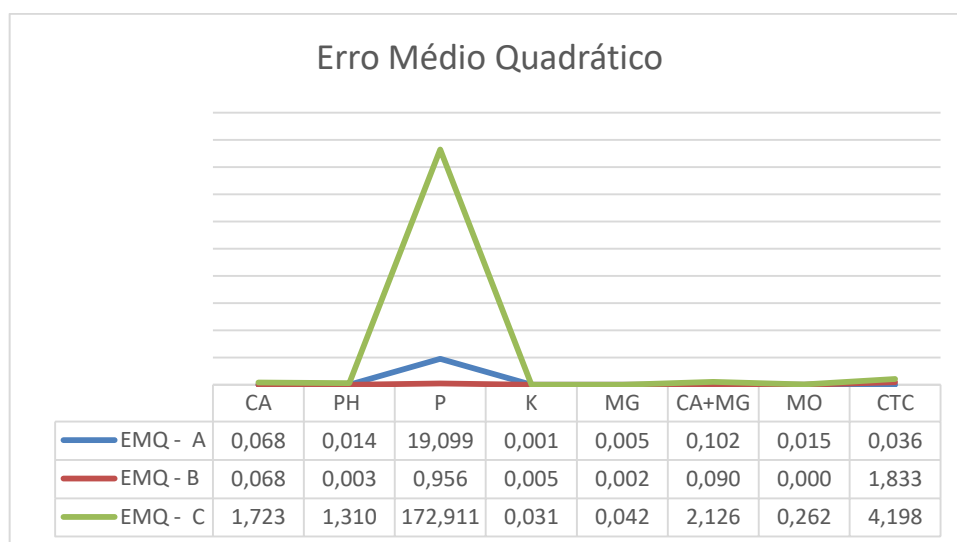
Finalmente, na previsão da Capacidade de Troca Catiônica (CTC) do solo, o modelo apresentou um desempenho razoável, porém não satisfatório. O coeficiente de correlação e R^2 próximos de zero indicam que o modelo não conseguiu capturar bem a relação entre as variáveis de entrada e a CTC. O alto valor de erro quadrático médio (EMQ) de 4.1985 e desvio padrão de 1.3638 mostram previsões com alto grau de dispersão e erros significativos. O erro médio de 0.1968 também é relativamente alto, indicando previsões imprecisas em média.

4.3. Modelo selecionado

Comparando os três modelos de regressão aplicados para prever as variáveis do solo, realizados por meio do Algoritmo Floresta Aleatória (RF) (Modelo A com 5 pontos e Modelo B com 9 pontos) e Redes Neurais MLP (Modelo C com 9 pontos), foram analisadas diversas métricas de desempenho para cada variável.

O Modelo B (Algoritmo Floresta Aleatória (RF) com 9 pontos) apresentou o melhor desempenho geral em relação às métricas de avaliação. Este modelo exibiu médias superiores para todas as variáveis do solo em comparação aos demais modelos. Notavelmente, o Erro Médio Quadrático (EMQ) do Modelo B foi o mais baixo conforme (Figura 22), indicando que as previsões possuem menor dispersão em relação aos valores reais.

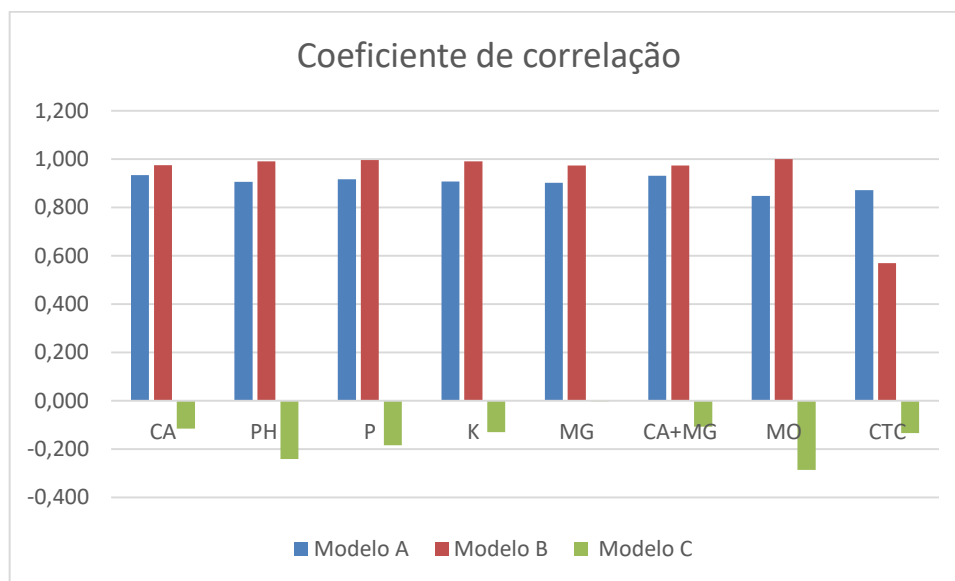
Figura 22 - Gráfico dos Valores de erro médio quadrático dos modelos A, B e C



Fonte: Autor (2023)

Na Figura 23 é possível identificar que o coeficiente de correlação (R) no Modelo B variou de 0,973 a 0,999, o que reflete uma correlação positiva forte entre as previsões e os valores reais das variáveis do solo. Juntamente com isso, o coeficiente de determinação (R^2) do Modelo B foi notavelmente alto, variando de 0,946 a 0,999, indicando que o modelo foi capaz de explicar uma parcela significativa da variabilidade nos dados.

Figura 23 - Gráfico do Coeficiente de Correlação dos modelos

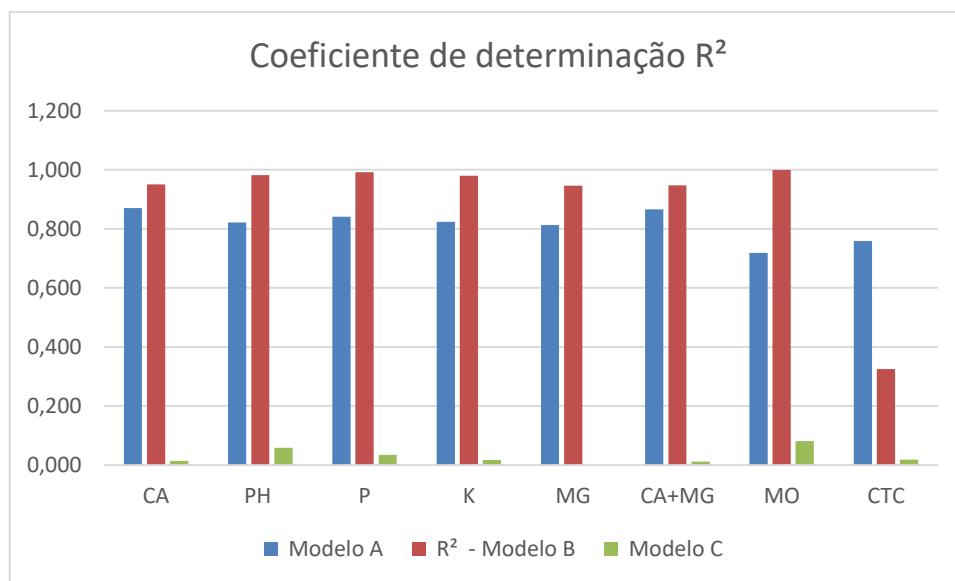


Fonte: Autor (2023)

Em contraste, o Modelo C (Redes Neurais MLP com 9 pontos) exibiu o pior desempenho em termos das métricas de avaliação. Suas médias foram menores para todas as variáveis do solo em comparação aos outros modelos. Especificamente, o EMQ do Modelo C foi consideravelmente maior, sugerindo que as previsões apresentam maior dispersão em relação aos valores reais, culminando em erros mais expressivos nas previsões.

Além disso, na Figura 24 demonstra o coeficiente de correlação (R) que do Modelo C variou de -0,285 a -0,002, evidenciando uma correlação fraca ou até mesmo negativa entre as previsões e os valores reais das variáveis do solo.

Figura 24 - Gráfico do Coeficiente de determinação dos modelos



Fonte: Autor (2023)

O coeficiente de determinação (R^2) do Modelo C também foi substancialmente baixo, variando de 0,011 a 0,081, o que indica que o modelo teve dificuldades em explicar a variabilidade dos dados.

Considerando todas as métricas avaliadas, o Modelo B (Algoritmo Floresta Aleatória (RF) com 9 pontos) demonstrou ser o mais preciso e adequado para prever as variáveis do solo em comparação ao Modelo A e ao Modelo C. Suas estimativas mais precisas e maior capacidade de explicar a variação nos dados o indicam a escolha preferencial entre os modelos examinados neste estudo.

5. CONCLUSÃO

Com a base de dados selecionada, contendo 120 pontos, a aplicação dos modelos A e B usando o algoritmo Floresta Aleatória (RF) obteve resultados promissores. Isso indica que, mesmo com um conjunto de dados relativamente pequeno, a abordagem com o algoritmo RF foi eficaz na modelagem e previsão das características do solo, destacando a utilidade dessa técnica para estudos futuros relacionados ao solo e agricultura de precisão.

Os Modelos A e B, empregando o Algoritmo Floresta Aleatória (RF), mostraram-se eficazes na previsão dos atributos químicos do solo. O Modelo A, com cinco pontos amostrais fixos, e o Modelo B, com nove pontos, destacaram-se por sua capacidade de ajuste aos dados observados. Por outro lado, o Modelo C, baseado em Redes Neurais Artificiais (RNAs) apresentou desempenhos variados na previsão dos atributos do solo, o que pode ser explicado com a baixa quantidade de dados para testes e treinamento.

Comparando os modelos, o Modelo B se destacou como a escolha preferencial para prever atributos químicos do solo. Com um desempenho geral superior, evidenciado por coeficientes de correlação elevados, coeficientes de determinação expressivos e baixos erros médios quadráticos, esse modelo revelou sua robustez e capacidade de oferecer previsões precisas.

A sugestão para trabalhos futuros envolve a aplicação de redes neurais artificiais com um conjunto de dados mais extenso. Essa sugestão poderia melhorar a capacidade de generalização do modelo, reduzir o *overfitting*, aumentar a robustez e a precisão.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAHAM, E.R.; REIS, J.G.M; TOLOI, R.C.; SOUZA, A.E.; COLOSSETTI, A.P. Estimativa da produção da soja brasileira utilizando Redes Neurais Artificiais (RNAS) . **Revista Agrarian**, Dourados, v.12, n.44, p.261-271, 2019.
- AFFONSO, C. D. O. **Redes neurais aplicadas em processos de usinagem da madeira**. UNESP, São Paulo, 2013.
- AGRIANUAL: **Anuário da agricultura brasileira**. São Paulo: FNP Consultoria & Agroinformativos, 2020.
- ALVES, M. C. Geostatistical analysis of the spatial variation of the berry borer and leaf miner in a coffee agroecosystem. **Precision Agriculture**, Dordrecht, v. 10, n. 12, p. 1-14, Dec. 2009.
- ALVES, G.R. **Estimativa da produtividade da soja com Redes Neurais Artificiais (RNAS)**. 2016. 76p. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual de Goiás, Anápolis, 2016.
- ANDRADE, L. N.; VIEIRA, T. G. C.; ALVES, H. M. R.; VOLPATO, M. M. L.; SOUZA, V. C. O. In.: CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA, 39. A engenharia Agrícola e o Desenvolvimento das pequenas propriedades rurais. **Congresso Latinoamericano Y Del Caribe De Ingeniería Agrícola**, v. 9, 2010.
- ARCHANA, S.; YADAV, A.; RANA, A. **K-means with Three different Distance Metrics**. **International Journal of Computer Applications** (0975 – 8887) Volume 67–No.10, April 2013.
- ARORA, A. **Analytics Arora: Florestas aleatorias**. São Paulo, 2021. Disponível em: <https://analyticsarora.com/8-unique-machine-learning-interview-questions-about-random-forests/>. Acesso em: 1 ago. 2023.
- BABUJI, Y.; WOODARD, A.; LI, Z.; KATZ, D. S.; CLIFFORD, B.; KUMAR, R.; CHARD, K. Parsl: Pervasive parallel programming in python. In **Proceedings of the 28th International Symposium on High-Performance Parallel and Distributed Computing**, (pp. 25-36), June, 2019.
- BAGHSHAH, M. S.; SHOURAKI, S. B. Learning low-rank kernel matrices for constrained clustering. **Neurocomputing**, v. 74, p. 2201-2211, 2011
- BARANAUSKAS, J. A. **Clustering (Agrupamento)**. São Paulo: USP, 2018. Disponível em: <https://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-Clustering.pdf>. Acesso em: 2 maio 2023.
- BASSOI, L. H.; INAMASU, R. Y.; B. A.C. C.; VAZ, C. M. P.; SPERANZA, E. A.; CRUVINEL, P. E. Agricultura de precisão e agricultura digital. In: TECCOGS – **Revista Digital de Tecnologias Cognitivas**, n. 20, jul./dez. 2019, p. 17-36.
- BHOLOWALIA, P.; KUMAR, A. Ekb-means: A clustering technique based on elbow method and k-means in wsn. **International Journal of Computer Applications**,

Citeseer, v. 105, n. 9, 2014.

BONINI, C. S. B.; ALVES, M. C.; MONTANARI, R. **Recuperação da estrutura de um latossolo vermelho degradado utilizando iodo de esgoto**. *Agrária* (Recife. Online), v. 10, p. 34-42, 2015.

BARRETO, R. G.; MARINHO, G. M. G. A.; BARRETO, G. F. M.; BARRETO, R. G.; AVERSARI, L. O. C.; DANTAS, B. L. Utilizando Redes Neurais Artificiais (RNAS) para o diagnóstico de câncer cervical. **Revista Saúde & Ciência**, Campina Grande, v. 7, n. 2, p. 59-67, 2018.

BINOTI, D. H. B. *et al.* Redução dos custos em inventário de povoamentos equiâneos utilizando Redes Neurais Artificiais (RNAS). **Agrária**, v. 8, p. 125- 129, 2012.

BINOTI, M. L. M. S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de Redes Neurais Artificiais (RNAS) para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, Viçosa, v. 37, n. 4, p. 639-645, 2013.

BITTAR, R. D. **Redes Neurais Artificiais (RNAS) aplicadas a modelagem da variabilidade espacial de atributos físico-químicos de solos do cerrado**. 2016. 112p. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual de Goiás, Anápolis, 2016.

BOCCATO, L.; ATTUX, R. Redes Neurais Artificiais (RNAS). **Redes Neurais**, São Paulo, v. 2, n. 2, p. 1-47, 2021.

BROSSARD, M.; BARCELLOS, A. D. O. Conversão do cerrado em pastagens cultivadas e funcionamento de latossolos. **Cadernos de Ciência & Tecnologia**, Brasília, v. 22, n. 1, p. 153-168, 2005.

CAPOANE, V.; TIECHER, T.; SANTOS, D. R. Variação de atributos do solo ao longo de três topossequências no planalto do Rio Grande do Sul. **Revista Brasileira de Geografia Física**, Recife, v.10, n.05, p. 1435-1454, 2017.

CANTERL, K. F. F. **Aprendizado de máquina na modelagem temporal da emissão de CO₂ do solo em áreas agrícolas no Cerrado brasileiro**. Unesp, São Paulo, 2020.

CAPUCHINHO, F. F. **Deep Learning para estimativa na escala mensal das temperatura do ar para a região centro-oeste do Brasil**. 2021. 90p. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual de Goiás, Anápolis, 2021.

CARDOSO, C. F. de S.dos.; CARVALHO, A. W. De. **ANAIS DA XIII SEMANA DE INFORMÁTICA**. Instituto Federal Goiano, 2021.

CAPELO, P. R. A. **Random forests and classification and regression trees in the analysis of censored data (Doctoral Dissertation)**. UM, Portugal, 2022.

CARNEIRO, M. A. C.; SOUZA, E. D. de; PAULINO, H. B.; SALES, ; L. E. de O.; VILELA, L. A. F. Atributos Indicadores de Qualidade em Solos de Cerrado no Entorno do Parque Nacional Das Emas, Goiás. **Bioscience Journal**, Uberlândia-MG, v. 29, ed. 6, p. 1857-1868, 2013.

CARVALHO, A. X. Y. D.; ALBUQUERQUE, P. H. M.; ALMEIDA JUNIOR, G. R. D.; GUIMARÃES, R. D. **Clusterização hierárquica espacial**. 2009.

CHERUBIN, M. R.; SANTI, A. L.; EITELWEIN, M. T.; AMADO, T. J. C.; SIMON, D. H.; DAMIAN, J. M. Dimensão da malha amostral para caracterização da variabilidade espacial de fósforo e potássio em Latossolo Vermelho. **Pesquisa Agropecuária Brasileira**, Brasília, v. 50, n. 02, p.168-177, 2015.

CUNHA, E. G. S. **Quantificação de biomassa em floresta estacional semidecidual por meio de Redes Neurais Artificiais (RNAs)**. UFVJM- Minas Gerais, 2015.

CHERUBIN, M. R.; SANTI, A. L.; EITELWEIN, M. T.; MENEGOL, D. R.; DA ROS, C. O.; PIAS, O. H. C.; BERGHETTI, J. Eficiência de malhas amostrais utilizadas na caracterização da variabilidade espacial de fósforo e potássio. **Ciência Rural**, Santa Maria, v. 44, n. 03, p. 425-432, 2014.

CHLINGARYAN, A.; SUKKARIEH, S.; WHELAN, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. **Computers and Electronics in Agriculture**, v. 151, p. 61-69, 2018.

COELHO FILHO, O. P.; MARTINHON, C. A.; CABRAL, L. Uma Abordagem Melhorada do Algoritmo de Otimização por Enxame de Partículas para o Problema de Clusterização de Dados. **Simpósio Brasileiro de Pesquisa Operacional**. Natal:[sn], p. 2135-2146, 2013.

CONAB. Companhia Nacional de Abastecimento. Tomate: Análise dos Indicadores da Produção e Comercialização no Mercado Mundial, Brasileiro e Catarinense. **Compêndio de Estudos Conab**, v. 21, 2019 – Brasília.

COUTINHO, E. R.; SILVA, R. M.; DELGADO, A. R. S. Utilização de inteligência computacional na predição de dados meteorológicos. **Revista Brasileira de Meteorologia**, São José dos Campos, v. 31, n. 1, p. 24-36, 2016.

COUTINHO, E. R.; SILVA, R. M.; MADEIRA, J. G. F.; COUTINHO, P. R. O. S.; BOLOY, R. A. M.; DELGADO, A. R. S. **Application of Artificial Neural Networks (ANNs) in the gap**, 2018.

DA COSTA FILHO, S. V. S.; ARCE, J. E.; MONTANO, R. A. N. R.; PELISSARI, A. L. Configuração de algoritmos de aprendizado de máquina na modelagem florestal: um estudo de caso na modelagem da relação hipsométrica. **Ciência Florestal**, v. 29, n. 4, p. 1501–1515, 2019.

DE ALVARENGA JUNIOR, W. J. **Métodos de otimização hiper paramétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária**. Tese de Doutorado – Universidade Federal de Minas Gerais, 2018.

DE QUEIROZ, D. M.; VALENTE, D. S. M.; DE CARVALHO PINTO, F. D. A.; BORÉM, A. **Agricultura digital**. Oficina de Textos, 2ª ed. São Paulo, 2022.

DI MININ, E.; BROOKS, T. M.; TOIVONEN, T.; BUTCHART, S. H.; HEIKINHEIMO, V.; WATSON, J. E.; MOILANEN, A. Identificar centros globais de colheita comercial insustentável de espécies. **Science Advances**, 2019.

DOMINGOS, P. **O algoritmo mestre: como a busca pelo algoritmo de machine learning definitivo recriará nosso mundo**. Novatec Editora, 2017.

DOSSA, D.; FUCHS, F. Tomate: **análise técnico-econômica e os principais indicadores da produção nos mercados mundiais, brasileiro e paranaense**. Boletim Técnico 03 Tomate, Curitiba, ago. 2017. Disponível em: http://www.ceasa.pr.gov.br/arquivos/File/BOLETIM/Boletim_Tecnico_Tomate1.pdf.

EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. **Cerrado**. Brasília-DF, 25 jan. 2023. Disponível em: <https://www.embrapa.br/contando-ciencia/bioma-cerrado#:~:text=O%20segundo%20maior%20bioma%20do,a%20204%20milh%C3%B5es%20de%20hectares!>. Acesso em: 12 jan. 2023.

EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. **Agricultura de precisão no gerenciamento da fertilidade do solo sob plantio direto no cerrado**. Sete Lagoas-MG, 2004. ISSN 0101-560'5. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/488639/1/Com112.pdf>. Acesso em: 14 mar. 2023.

EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. **Sistema Brasileiro de Classificação de Solos**. Rio de Janeiro-RJ, 2019.

FARINA, E. M. M. Q.; DE AZEVEDO, P. F.; SAES, M. S. M. **Competitividade: mercado, estado e organizações**. Editora Singular, 2022.

FÉLIX, E. F.; CARRAZEDO, R.; POSSAN, E. Análise paramétrica da carbonatação em estruturas de concreto armado via Redes Neurais Artificiais (RNAs). **Revista ALCONPAT**, Mérida, v. 07, n. 03, p. 302-316, 2017.

FERLIN, E. P.; REZENDE, D. A. *Big Data* aplicado à cidade digital estratégica: estudo sobre o volume de dados das aplicações *smart city*. **Revista Gestão & Tecnologia**, Pedro Leopoldo, v. 19, n. 2, p. 175-194, 2019.

FERREIRA, J. A. B. L. **Desenvolvimento de rede neuronal artificial para o diagnóstico de doenças neurodegenerativas**. Tese de Doutorado - Universidade do Minho, Portugal, 2013.

FERREIRA, L. G.; FERREIRA, M. E.; ROCHA, G. F.; NEMAYER, M.; FERREIRA, N. C. Dinâmica agrícola e desmatamentos em áreas de cerrado: uma análise a partir de dados censitários e imagens de resolução moderada. **Revista Brasileira de Cartografia**, Rio de Janeiro, v. 61, n. 2, p. 117-127, 2009.

FERREIRA, M. M. Caracterização física do solo. In: LIER, Q. J. v. **Física do solo**. Viçosa: SBCS, 2010. p.1-27.

FREITAS, E. F. M. **Redes neural artificial na predição de atributos físico e químicos do solo**. 2018. 112p. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual de Goiás, Anápolis, 2016.

GALLO, A.de S. **Variabilidade espacial das formas de fósforo em solo sob manejo conservacionista no sistema integrado de produção agroecológica**. Universidade Rural do Rio de Janeiro, Rio de Janeiro, 2020.

GUIDINI, M. N.; NASCIMENTO, A. M.; BONE, R. B.; ALVES, T. W. **Aplicação do k-means cluster para a classificação de estilos gerenciais**. Rio de Janeiro, 2008.

GUO, P. T.; LI, M. F.; LUO, W.; TANG, Q. F.; LIU, Z. W.; LIN, Z. M. Mapeamento digital da matéria orgânica do solo para plantação de borracha em escala regional: uma aplicação da abordagem de floresta aleatória mais krigagem de resíduos. **GEODERMA**, v. 237, p. 49-59, 2015.

GIUFFRIDA, M. V.; DOEMES, P.; TSAFTARIS, S. A. Pheno-DEEEP Counter: a unified and versatile deep learning architecture for leaf counting. **The Plant Journal**, Hoboken, v. 96, p. 880-890, 2018.

HAYKIN, S. **Redes neurais: princípios e prática**. Tradução Paulo Martins Engel. 2º Edição. Porto Alegre: Bookman, 2001. 937p.

HAYKIN, S. **Redes neurais: princípios e prática**. Porto Alegre: 2001. 900p.

HILLIER, F. S.; LIEBERMAN, G. J. **Introdução à pesquisa operacional**. McGraw Hill Brasil, 2013.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. COORDENAÇÃO GERAL DE OBSERVAÇÃO DA TERRA. **PRODES – Incremento anual de área desmatada no Cerrado Brasileiro**. 2018. Disponível em: <http://www.obt.inpe.br/cerrado>. Acesso em: 03 dez. 2023.

IRMAK, A.; JONES, J. W.; BATCHELOR, W. D.; IRMAK, S.; BOOTE, K. J.; PAZ, J. O. Artificial neural network model as a data analysis tool in precision farming. **Transactions of the ASABE**, v. 49, n. 6, p. 2027-2037, 2006.

JALLAL, M. A.; GONZÁLEZ-VIDAL, A.; SKARMETA, A. F.; CHABBA, S.; ZEROUAL, A. A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting Applied to energy consumption prediction. **Applied Energy**, v. 268, p. 1-19, 2020.

JOHARI, A.; JAVADI, A. A.; HABIBAGAH, G. Modelling the mechanical behaviour of unsaturated soils using a genetic algorithm-based neural network. **Computers and Geotechnics**, v. 38, n. 1, p. 2-13, 2011.

KLINK, C. A.; MACHADO, R. B. A conservação do cerrado brasileiro. **Revista Megadiversidade**, v. 01, n. 01, p. 147-155, 2005.

LEAL, A. J. F.; LAZARINI, E.; RODRIGUES, L. R.; MARCANDALLI, L. H. Adubação nitrogenada para milho com o uso de plantas de cobertura e modos de aplicação de calcário. **Revista Brasileira de Ciencia do Solo**, 37, 491-50, 2013.

LEAL, A. J. F.; MIGUEL, E. F.; BAIO, F. H. R.; NEVES, D. C.; LEAL, U. A. S. Redes Neurais Artificiais (RNAS) na predição da produtividade de milho e definição de sítios de manejo diferenciado por meio de atributos do solo. **Solos e Nutrição de Plantas – Bragantia**, v. 74, n. 4, 2015.

LEVI, T.; SHEPARD, Jr. G. H.; OHL-SCHACHERER, J.; WILMERS, C. C.; PERES, C. A.; YU, D. W. Ferramentas espaciais para modelar a sustentabilidade da caça de subsistência em florestas tropicais. **Aplicações Ecológicas**, v. 21, n. 5, pág. 1802-1818, 2011.

LIMA, J. S. S.; SOUZA, G. S.; SILVA, S. A. Amostragem e variabilidade espacial de atributos químicos do solo em área de vegetação natural em regeneração. **Revista Árvore**, v. 34, n. 1, p. 127-136, 2010.

LIMA, H. N.; KERN, D. C.; MADARI, B. E.; WOODS, W. T. W. G.; WENCESLAU G. T. C. P. A. A.; HEDINALDO N. L. **As terras pretas de índio da Amazônia: sua caracterização e uso deste conhecimento na criação de novas áreas**. Manaus: Embrapa Amazônia Ocidental, 2009.

LIMA, L. M. M.; VIANNA, R. P. T.; MORAES, R. M. Identificação das anomalias congênitas baseado em um modelo de decisão a partir de Redes Neurais Artificiais (RNAS). **Journal of Helth Informatics**, São Paulo, v. 11, n. 01, p. 8-12, 2019.

LINDEN, R. Técnicas de agrupamento. **Revista de Sistemas de Informação da FSMA**, v. 4, n. 4, p. 18-36, 2009.

LINHARES, M. M. A. *et al.* Vant's aplicados ao monitoramento da evolução vegetativa de culturas agrícolas. In: Congresso Brasileiro De Agricultura De Precisão- **CONBAP** 2014, 10., 2014, São Pedro – Sp. Anais. São Pedro – Sp: Sbea, 2014. p. 1 – 8.

LOPES, K. A. L.; PINTO JÚNIOR, F. F.; AGUIAR, F. I. S.; SOUSA, A. E. S.; OLIVEIRA, I. R.; DANTAS, J. S. Diferentes densidades amostrais na caracterização da variabilidade espacial de atributos granulométricos de um Argissolo Amarelo Distrófico Típico. **Revista de Ciências Agrônômicas**, Ilha Solteira, v. 29, n. 01, p. 50-60, 2020.

LOPES, R. D.; FICHEMAN, K. I.; SAGGIO, E. **Feira Brasileira de Ciências e Engenharia**. LIMA, B. F. G.; SANTOS, M. D. da S.; SILVA, J. O.B. da. In: Construção de Redes Neurais Artificiais (RNAS) para mapeamento de inteligências múltiplas dentro de jogos digitais. Anais.2023, p. 19-20.

LORENZI, A.; SILVA, B. V.; BARBOSA, M. P.; SILVA FILHO, L. C. P. Aplicação de Redes Neurais Artificiais (RNAS) na predição da aderência aço-concreto em ensaios do tipo “pull-out”. **Revista IBRACON de Estruturas e Materiais**, São Paulo, v. 10, n. 05, p. 1051-1074, 2017.

LUIZ, T. B. P. **Previsão de dados de níveis de água subterrânea utilizando modelos baseados em aprendizado de máquina**. Doctoral - Universidade Federal de

Santa Maria), Santa Maria, 2022.

MACEDO, R. D. C.; ALMEIDA, C. M. D.; SANTOS, J. R. D.; RUDORFF, B. F. T. Modelagem dinâmica espacial das alterações de cobertura e uso da terra relacionadas à expansão canavieira. **Boletim de Ciências Geodésicas**, 2013 p313-337.

MALUF, R. S.; FLEXOR, G. **Questões agrárias, agrícolas e rurais: Conjunturas e políticas públicas**. Rio de Janeiro: E-papers, 2017. 330 p.

MARCAL, L. A. B. **Estudo de propriedades estruturais de nanomembranas semicondutoras**. 2015.

MARINHO, J. L. M.; CID, Y. P. M. **Aplicação da geotecnologia para identificar fragmentos de vegetação do município de Belém-Pará no ano de 2020**. Belém, 2021.

MARTINS, M. R. **Sistema de coleta, armazenamento e interpretação de dados de colheita para geração de mapas de produtividade em culturas perenes**. Master's thesis, Universidade Tecnológica Federal do Paraná, Paraná, 2020.

MEIRELLES, M. S. P.; MOREIRA, M.; COELHO, F. B. N. Uso de Inteligência Artificial em Agricultura de Precisão: Redes Bayesianas e Neurais. **Agricultura de precisão para o manejo da fertilidade do solo em sistema plantio direto**, p. 175, 2004.

MELO, M. S. de. **Avaliação de potencial hidroelétrico com aplicação de sistema de informações geográficas**. UnB, Brasília, 2013.

MENDES, A.; SANTOS, G. R.; EMILLIANO, P. C.; KALEITA, A. L.; FERREIRA, M. P. Estimação do tamanho amostral na geoestatística usando um modelo de variograma gaussiano na presença de outliers. **Energia na Agricultura**, Botucatu, v.34, n.03, p.429-440, 2019.

MENDES, W. S.; DEMATTÊ, J. A. M.; MOLIN, J. P.; CORRÊDO, L. P.; TAVARES, P. A. Geoestatística na variabilidade espacial dos atributos físicos e químicos do solo em solo tropical. In: CONGRESSO INTERNACIONAL DAS CIÊNCIAS AGRÁRIAS, 3, 2018, João Pessoa-PB. **Anais**. Recife-PE: Cointer PDVAgro. p.1-12.

MENDES, R. B. **Predição da retenção de água em solos de Santa Catarina através de Redes Neurais Artificiais (RNAS)**. 2014. p.147. Dissertação (Mestrado em Ciência do solo), Universidade do Estado de Santa Catarina, Centro de Ciências Agrovetenárias, Programa de Pós-Graduação em Ciência do Solo, Lages, SC, 2014.

NANNI, M. R.; POCH, F. P.; DEMATTÊ, J. A. M.; OLIVEIRA, R. B.; CHICATI, M. L.; CEZAR, E. Optimum size in grid soil sampling for variable rate application in site-specific management. **Scientia Agricola**, Piracicaba, v. 68, n. 03, p. 386-392, 2011.

NETO, R. F. de O. **Inteligência Artificial**. 3. ed. Fundação Universidade Estadual do Vale do São Francisco, 2013.

OLIVEIRA, M. C. N.; BARBIN, D.; SILVA, E. F.; MACEDO, J. R.; CAPECHE, C. L.; HISSA, H. R. Estimação dos parâmetros de uma função de covariância no estudo da variabilidade espacial do teor de cálcio no solo. **Revista de Matemática e Estatística**,

São Paulo, v. 24, n. 04, p. 117-132, 2006.

OLUDARE, I. A.; AMAN, J.; ABIODUN, E. O.; KEMI, V. D.; NACHAAT, A. M.; HUMAIRA, A. State of the art in: artificial neural network applications: A survey. **Heliyon**, 2018.

PATIL, N. G.; SINGH, S. K. Pedotransfer functions for estimating soil hydraulic properties: A review. **Pedosphere**, v. 26, n. 4, p. 417-430, 2016.

PEIXOTO, J. V. M.; DE MORAES, E. R.; PEIXOTO, J. L. M.; DOS REIS NASCIMENTO, A.; NEVES, J. G. Tomaticultura: Aspectos morfológicos e propriedades físico-químicas do fruto. **Revista Científica Rural**, 2017.

PELUCO, R. G. **Técnicas indiretas na identificação de áreas de manejo específico para cana-de-açúcar**. Mestrado em Produção Vegetal. UNESP-SP, 2013.

PINHEIRO, M. P.; MEDEIROS. **Avaliação do potencial hidrológico da estação ecológica de Avaré (SP) para segurança hídrica local**. UNESP, São Paulo, 2018.

PUC Rio - MÉTODOS de otimização. **Certificação digital, nº0310953/CA**. Rio de Janeiro, 2011. Disponível em: https://www.maxwell.vrac.puc-rio.br/7603/7603_4.PDF. Acesso em: 2 mar. 2023.

REIS, J. D. S. **Determinação de zonas de manejo para adubação nitrogenada em lavoura de tomate industrial**. 2011. 67p. (Mestrado em Engenharia Agrícola) - Universidade Estadual de Goiás, Anápolis, 2011.

RIBEIRO, A. C.; GUIMARÃES, P. T. G.; ALVAREZ, V. H. **Recomendações para o uso de corretivos e fertilizantes em Minas Gerais**, 5ª aproximação. Viçosa, MG: Comissão de Fertilidade do Solo do Estado de Minas Gerais, 1999. 359p.

RIBEIRO JUNIOR, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. Tese de Doutorado. Universidade de São Paulo.

RIQUETI, G.; RIBEIRO, C.; ZARATE, L. **Classificando perfis de longevidade debases de dados longitudinais usando floresta aleatoria**, 2018.

ROCHA, E. T. B. da. **Agricultura 4.0 nas lavouras: Estudo multicaso para caracterização em propriedades rurais**. 91p. Dissertação – Universidade Estadual Paulista, 2021.

RODRIGUES, M. S.; CORÁ, J. E.; CASTRIGNANÒ, A.; MUELLER, T. G.; RIENZI, E. A spatial and temporal prediction model of corn grain yield as a function of soil attributes. **Agronomy Journal**, v. 105, p. 1878-1887, 2013.

ROSA JUNIOR, E. J.; ROSA, Y. B. C. J.; ROSA, C. D. C. J.; RAMOS, A. Z. Manejo do solo e de culturas e seu efeito sobre soja e atributos de um latossolo. In: REUNIÃO BRASILEIRA DE MANEJO E CONSERVAÇÃO DO SOLO E DA ÁGUA, 15., Santa Maria, 2004. **Anais...** Santa Maria, RS. 2012.

ROSA, M. G.; KLAUBERG FILHO, O.; BARTZ, M. L. C.; MAFRA, A. L.; SOUSA, J. P.

F. A.; BARETTA, D. Macrofauna edáfica e atributos físicos e químicos em sistemas de uso do solo no planalto catarinense. **Revista Brasileira de Ciência do Solo**, Viçosa, v.39, n.6, p.1544-1553, 2015.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Campus, 2013. 1016p.

SANTOS, M. L.; CARVALHO, M. P.; RAPASSI, R. M. A.; MURAISHI, C. T.; MALLER, A.; MATOS, F. A. Correlação linear e espacial entre produtividade de milho e atributos físicos de um Latossolo Vermelho distroférrico sob plantio direto do cerrado brasileiro. **Acta Scientiarum Agronomy**, Maringá, v. 28, n. 3 , p. 313-321, 2006.

SANTOS, A. M. R. T.; SANTOS, G. R.; EMILIANO, P. C.; MEDEIROS, N. G.; KALEITA, A. L.; PRUSKI, L. O. S. Detection of inconsistencies in geospatial data with geostatistics. **Boletim de Ciências Geodésicas**, v. 23, n. 2, 2017.

SANTOS, A. da S. **Geoestatística aplicada na avaliação da qualidade de dados geoespaciais**. 148p. Dissertação – Universidade Federal de Viçosa, 2013.

SANTOS NETO, L. **Otimização da extração por solventes de B-caroteno em amostras de tomates (*Solanum lycopersicum*), abóboras (*Cucurbita* sp) e beterrabas (*Beta vulgaris* L)**. 50p. Trabalho de Conclusão de Curso, Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Piauí, 2019.

SANTOS, M. C. D. **Monitoramento em linha baseado em AI e espectroscopia da concentração enantiomérica de praziquantel**. 77p. Trabalho de Conclusão de Curso, UFRJ, Rio de Janeiro, 2022.

SANTOS, E. D. S. D. **Aplicação de algoritmos de aprendizagem supervisionada de máquina em sistemas embarcados no auxílio à aplicação de defensivos agrícolas**. 157p. Master's thesis, Universidade Tecnológica Federal do Paraná, Paraná, 2023.

SAIZ-RUBIO, V.; ROVIRA-MÁS, F. Da agricultura inteligente para a agricultura 5.0: uma revisão sobre o gerenciamento de dados de culturas. **Agronomia** , v. 10, n. 2, pág. 207, 2020.

SCHUCH, K. E. F. **Análise preditiva com Redes Neurais Artificiais (RNAS) para o planejamento de sistemas de irrigação**. 13p. Trabalho de Conclusão de Curso, PUC, Goiânia, 2021.

SHAFEEQ, A.; HAREESHA, K. S. Dynamic clustering of data with modified k-means algorithm. In: **Proceedings of the 2012 conference on information and computer networks**. 2012. p. 221-225.

SEIDEL, E. J.; OLIVEIRA, M. S. Medidas de dependência espacial baseadas em duas perspectivas do semivariograma paramétrico. **Ciência e Natura**, Santa Maria, v.37, n. 04, p. 20-27, 2015.

SEVEGNAN, K. B.; SILVA, S. H. M. G.; FERNANDES, D. P. B.; SILVA, I. J. O. Estudo da influência de fatores ambientais em variáveis fisiológicas de frangos de corte com Redes Neurais Artificiais (RNAS). **Energia na Agricultura**, Botucatu, v. 32, n. 02, p.

148-152, 2017.

SHAPIRO, S. S.; WILK, M. B. Na analysis of variance test for normality (complete sample). **Biometrika**, London, v. 52, n. 3/4, p. 591-611, 1965.

SILVA, A. R.; DIAS JUNIOR M. S.; LEITE F. P. Avaliação da intensidade de tráfego e carga de um forwarder sobre a compactação de um Latossolo Vermelho-Amarelo. **Revista Árvore**, v. 35, n. 03, p. 547-554, 2011.

SILVA, L. O. **Hortaliças frescas todos os dias na sua casa**. Editora IFG, Goiânia, 2019.

SILVA, J.; VIEIRA, R. Introdução as Redes Neurais Profundas com Python. In: **Introdução as Redes Neurais Profundas com Python**. São Paulo: [s. n.], 2019. cap. 4. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/110/498/771-1?inline=1>. Acesso em: 6 fev. 2023.

SILVA NETTO, R. **Detecção de intrusão utilizando Redes Neurais Artificiais (RNAs) no reconhecimento de padrões de ataque**. 94p. Dissertação - Universidade Federal Itajubá, Itajubá-MG, 2006.

SILVA NETO, V. L.; CARDOSO, J. A. F.; SILVEIRA JÚNIOR, O.; BATISTA, E.; ARAÚJO, A. M. M.; LIMA, I. C. S. Avaliação de interpoladores espaciais para melhoria do mapeamento de atributos químicos do solo na cultura da banana no Estado do Tocantins, Brasil. **Research, Society and Development**, Vargem Grande Paulista, v. 10, n. 01, p. 1-12, 2021.

SILVA, R. J. H.; PAZOTI, M. A.; DA SILVA, F. A.; PEREIRA, D. R.; ALMEIDA, L. L. Autenticação biométrica para sistemas por meio da dinâmica da digitação. In **Colloquium Exactarum**. ISSN: 2178-8332, volume 11, pages 26–33, 2019.

SINGH, S. S; CHAUHAN, N. C. K-means v/s k-medoids: A comparative study. In: **National Conference on Recent Trends in Engineering & Technology**. [S.l.: s.n.], v. 13, 2011.

SOARES, E. A. **Sistemas evolutivos baseados em regras para previsões de séries temporais meteorológicas**. 2017. 78f. Dissertação (Mestrado em Engenharia de Sistemas e Automação), Universidade Federal de Lavras, Lavras, 2017.

SOARES, F. C.; ROBAINA, A. D.; PEITER, M. X.; RUSSI, J. L.; VIVAN, G. A. Redes Neurais Artificiais (RNAs) na estimativa da retenção de água no solo. **Ciência Rural**, Santa Maria, v. 44, n. 2, p. 293-300, 2014.

SOARES, R. V.; BATISTA, A. C. **Meteorologia e Climatologia Florestal**. Curitiba: UFPR. 2004. 195p.

SOUSA NETO, R. de. O mercado de tomate em Goiás: **estudo sobre o comportamento da cadeia e a evolução da atividade produtiva no setor in natura**.

Dissertação (Mestrado em Agronegócio) - Universidade Federal de Goiás, Goiânia, 2019.

SOUSA, N. B. D. **Uma metodologia para identificação de perdas não técnicas em unidades consumidoras irrigantes de café do estado de São Paulo**. 148p. Dissertação, Universidade Federal de Santa Maria, 2022.

SOUZA, Z. M.; BARBIERI, D. M.; MARQUES JÚNIOR, J.; PEREIRA, G. T.; CAMPOS, M. C. C. Influência da variabilidade espacial de atributos químicos de um latossolo na aplicação de insumos para cultura da cana-de-açúcar. **Ciência e Agrotecnologia**, Lavras, v.31, n.2, p. 371-377, 2007.

SRINIVASAN, A. **Handbook of precision agriculture: principles and applications**. New York: CRC. 683 p. (2006).

TAVARES, V. B.; MAESTRI, C. O. N. M.; PENEDO, A. S. T.; PEREIRA, V. S. O grau de internacionalização influencia a previsão da rentabilidade de ações? Evidências do mercado de ações brasileiro por meio de redes neurais artificiais. **Revista de Gestão do Unisalle**, Canoas, v. 08, n. 01, p. 137-148, 2019.

TAVARES, V. B.; PENEDO, A. S. T. Níveis de governança corporativa da B3: interesse e desempenho das empresas – uma análise por meio de Redes Neurais Artificiais (RNAS). **Revista Contabilidade, Gestão e Governança**, Brasília, v. 21, n. 01, p. 40-62, 2018.

TUKEY, J. W. A survey of sampling from contaminated distribution. In: OLKIN, I. **Contributions to probability and statistics**. California: Stanford University Press, 1960. Chapter 39, p.448-485.

VITÓRIA, E. L. *et al.* Estimativas de emissão de CO₂ em solos cultivados por meio de redes neurais artificiais e modelo linear de regressão. **Revista Caatinga**, v. 35, p. 964-973, 2022.

WANG, J.; XINDONG, W.; CHENGQI, Z. Support Vector Machines Based on K-Means Clustering for Real-Time Business Intelligence Systems. **International Journal of Business Intelligence and Data Mining**. Inderscience Publishers: 54–64, 2005.

YADAV, A. K.; MALIK, H.; CHANDEL, S. S. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. **Renewable and Sustainable Energy Reviews**, Amsterdam, v. 31, p. 509-519, 2014.